

Investigation of the Molecular Genetics of Renal Ciliopathies

A thesis submitted for the degree of Doctor of Philosophy

Miguel Barroso Gil

Translational and Clinical Research Institute

Newcastle University

September 2023

Abstract

Primary ciliopathies are a group of inherited diseases caused by dysfunction or absence of the primary cilia. They exhibit genetic heterogeneity, genetic pleiotropy and a wide spectrum of clinical phenotypes, often including renal disease. The main objective of this project is to investigate the underlying molecular disease mechanisms and phenotypic and genotypic complexity of renal ciliopathies, and to increase the genetic diagnosis yield in patients with primary ciliopathies. To achieve this, I used a combination of *in vitro* and *in silico* techniques, focussing on a selection of pleiotropic genes associated with overlapping ciliopathy phenotypes. This approach covers a broad perspective to study renal ciliopathies, including: genetic pleiotropy, heterogeneity of clinical phenotypes, genetic diagnosis yield, the use of functional analyses to validate disease-causing variants and gene therapies.

The spectrum of clinical phenotypes in primary ciliopathies may reflect the tissue specificity of expression of disease genes. Consequently, gene expression of two pleiotropic genes associated with overlapping rare primary ciliopathies were characterised: *CEP120* and *ARL3*, using the MRC Wellcome Trust Human Developmental Biology Resource (HDBR) and RNAscope, an RNA *in situ* detection technique. Expression patterns of both genes revealed brain, eye and kidney expression, which correlates with known clinical phenotypes. *In silico* approaches were evaluated to explore genotype-phenotype correlations in two primary ciliopathy genes: *CEP120* and *CC2D2A*. These *in silico* approaches allowed to investigate potential exon skipping genetic therapies in these two genes. Approximately 30% of patients with a primary ciliopathy phenotype lack a genetic diagnosis. Using the data available in the Genomics England 100,000 Genomes Project, a novel primary multisystem ciliopathy syndrome in a patient with biallelic pathogenic variants in *TOGARAM1* was identified. Also, within this dataset, in patients with cystic kidney disease, underlying disease-causing genetic variants in *DNAJB11*, *ALG5*, *ALG8* and *ALG9* were identified, supporting these genes as ciliopathy genes associated with the phenotypic spectrum of autosomal dominant polycystic kidney and liver disease.

I used human urine-derived renal epithelial cells (hURECs) to characterise the ciliary phenotype of a patient with cystic kidney disease and proposed hURECs as a valuable non-invasive source of primary cells to validate candidate genes for novel renal ciliopathies. hURECs may also be used as an *ex vivo* model to study patient-specific genetic variants, disease mechanisms and complement therapeutic studies.

Acknowledgment

I would like to thank my supervisors Prof. John Sayer and Dr. Colin Miles for its guidance, support and motivation throughout my PhD project. I would also like to thank Kidney Research UK and Northern Counties Kidney Research fund for funding this project. To all the members of Sayer and Miles Laboratory for all the advice, support and positive energy and attitude, specially to Elisa, Simon, Eric, Laura Devlin, Laura Powell, Becky and Ruxandra.

To the MRC Wellcome Trust Human Developmental Biology Resource (HDBR), the Bioimaging Unit and Administration Services and staff at Newcastle University and Genomics England for their support. To José Luis, Miguel, María and Elena as their supervision they provided during my Bachelor's degree and Master's degree was essential for me to start this PhD project.

To my friends in Newcastle and Durham, especially to Thais, Mónica, Erandi, Magomet and Marina, which made these years living in the North East England enjoyable and unforgettable.

To my friends in Madrid and Turégano, especially to Daniel Almendros, Víctor, Marcos, Antonio, Alberto, Esther, Elisa and Daniel for their support and empathy every time I went back to Spain.

To my family, my grandparents, uncles, aunts, cousins and specially to my parents Mercedes and Emilio, and my sister Marta, which have given and give me support, strength and enthusiasm to accomplish everything, and I would not have been able to do this project without them.

Table of Contents

Chapter 1	Introduction.....	1
1.1.	Primary Ciliopathies.....	1
1.1.1.	Primary cilia.....	2
1.1.2.	Centrosome.....	7
1.2.	Phenotypic spectrum of primary ciliopathies.....	8
1.2.1.	Joubert syndrome	12
1.2.2.	Jeune syndrome	14
1.2.3.	Renal ciliopathies	15
1.3.	Molecular genetics of renal ciliopathies.....	18
1.3.1.	Molecular genetics of Joubert syndrome	19
1.3.2.	Molecular genetics of Jeune syndrome.....	19
1.3.3.	Molecular genetics of renal ciliopathies.....	22
1.4.	Molecular disease pathogenesis of ciliopathy genes studied in this project.	32
1.4.1.	<i>CEP120</i>	35
1.4.2.	<i>CC2D2A</i>	42
1.4.3.	<i>ARL3</i>	43
1.4.4.	<i>TOGARAM1</i>	44
1.4.5.	Other genes associated with Joubert syndrome.....	44
1.4.6.	Endoplasmic reticulum pathway and Asparagine-Linked Glycosylation (ALG) proteins.....	46
1.5.	Current and future perspectives for the study of primary ciliopathies.....	48
1.5.1.	Human urine-derived renal epithelial cells (hURECs).....	51
1.5.2.	Next generation sequencing and the Genomics England 100,000 Genomes Project.....	53
1.5.3.	Antisense oligonucleotide mediated exon skipping as an emerging therapy in ciliopathies.....	58
1.6.	Aims of the project.....	60
Chapter 2	Materials and Methods.....	62
2.1.	Software tools and online resources.....	62
2.2.	Cell culture.....	64
2.2.1.	Cell culture of mouse Shh-LIGHT2 cells.....	64
2.2.2.	Cell culture of human primary skin fibroblasts.....	64

2.2.3. Cell culture of Human Urine-Derived Renal Epithelial Cells (hURECs).....	64
2.2.4. Subculturing, serum starvation and freezing of cells.....	67
2.3. Immunofluorescence (IF) staining	68
2.3.1. Antibodies.....	68
2.3.2. Fixation of cells.....	68
2.3.3. Immunofluorescence staining in Shh-LIGHT2 cells and primary human fibroblasts.....	68
2.3.4. Immunofluorescence staining in hURECs.....	70
2.4. Use of the RNA in situ detection platform, RNAscope.....	71
2.5. Analysis of the reported clinical phenotypes and protein sequences of <i>CEP120</i> and <i>ARL3</i>	73
2.6. Immunohistochemistry (IHC) staining.....	74
2.7. Review and annotation of <i>CEP120</i> and <i>CC2D2A</i> genetic variants and clinical presentations reported in the literature.....	76
2.8. Microscopy.....	78
2.8.1. Microscopes used.....	78
2.8.2. Processing of microscopy images.....	78
2.9. Genetic and gene expression studies.....	80
2.9.1. Polymerase chain reaction (PCR).....	80
2.9.2. Agarose Gel electrophoresis.....	80
2.9.3. RNA extraction and clean-up.....	80
2.9.4. Reverse transcription-PCR (RT-PCR).....	81
2.9.5. Quantitative PCR (qPCR).....	83
2.10. Protein extraction and Western blotting.....	84
2.11. <i>Cep120/CEP120</i> gene knockdown.....	85
2.11.1 Small interfering RNA (siRNA) <i>Cep120</i> knockdown experiment in mouse Shh LIGHT2 cells.....	85
2.11.2 SAG stimulation of Shh-LIGHT2 cells and dual luciferase assay.....	87
2.11.3 Antisense oligonucleotides (ASO) morpholino <i>CEP120</i> knockdown experiment in primary human fibroblasts.....	87
2.12. Alamar Blue cell viability assay.....	88
2.13. Cell-based assays using Human Urine-Derived Renal Epithelial Cells (hURECs) to complement genetic studies.....	90
2.14. Plasmid transfection efficiency into mammalian cells.....	90

2.15. Use of the Genomics England 100,000 Genomes Project database.....	100
2.15.1 Search for patients presenting with biallelic pathogenic variants in <i>CEP120</i>	102
2.15.2 Search for patients presenting with biallelic pathogenic variants in <i>TOGARAM1</i>	104
2.15.3 Search for patients presenting with monoallelic pathogenic variants in genes encoding for proteins functioning in the endoplasmic reticulum: <i>DNAJB11</i> , <i>GANAB</i> , <i>ALG5</i> , <i>ALG8</i> and <i>ALG9</i>	105
2.15.4 Estimation of the proportion of loss of function (LoF) variants in <i>ALG</i> genes.....	108
Chapter 3 Characterisation of <i>CEP120</i> expression patterns in human embryonic and fetal tissue using the MRC-Wellcome Human Developmental Biology Resource (HDBR), and comparison with <i>ARL3</i> expression patterns.....	111
3.1. Introduction and aims.....	111
3.2. Results.....	117
3.2.1. Comparison of the clinical phenotypes of <i>CEP120</i> - and <i>ARL3</i> -affected patients and RNA expression data from public databases.....	117
3.2.2. Expression of <i>CEP120</i> and <i>ARL3</i> in early human brain development..	126
3.2.3. Expression <i>CEP120</i> and <i>ARL3</i> in the developing cerebellum.....	130
3.2.4. Expression of <i>CEP120</i> and <i>ARL3</i> in the developing eye.....	134
3.2.5. Expression of <i>CEP120</i> and <i>ARL3</i> in the developing dorsal root ganglia.....	137
3.2.6. Expression of <i>CEP120</i> and <i>ARL3</i> in the developing kidney.....	140
3.2.7. Expression of <i>CEP120</i> and <i>ARL3</i> in other major organs.....	144
3.2.8. Immunohistochemistry of <i>OLIG2</i> and <i>PAX6</i> transcription factors in human developing brain.....	149
3.3. Discussion.....	154
Chapter 4 Investigation of Joubert syndrome (JBTS) and JBTS candidate genes in the Genomics England 100,000 Genomes Project, the potential use of exon therapies and the use of cell-based assays to complement genetic studies	160
4.1. Introduction and aims.....	160
4.2. Results.....	169

4.2.1. <i>In silico</i> investigation of genotype-phenotype correlations in <i>CEP120</i> and <i>CC2D2A</i>	169
4.2.2. <i>In silico</i> analysis of gene expression and tissue-specific basal exon skipping in <i>CEP120</i> and <i>CC2D2A</i>	174
4.2.3. <i>In silico</i> investigation of <i>CEP120</i> and <i>CC2D2A</i> as potential candidates for targeted exon skipping therapies.....	182
4.2.4. Use of WGS data from the Genomics England 100,000 Genomes Project to find novel patients presenting with biallelic <i>CEP120</i> variants.....	193
4.2.5. Use of WGS data from the Genomics England 100,000 Genomes Project to find patients with known <i>CEP120</i> variants.....	207
4.2.6. Use of WGS data from the Genomics England 100,000 Genomes Project to find novel patients presenting with biallelic variants in <i>TOGARAM1</i> , a gene recently associated with Joubert syndrome.....	210
4.2.7. Analysis of <i>CCDC28B</i> as a candidate gene for Joubert syndrome.....	213
4.2.8. Expression of <i>CEP120</i> Shh-LIGHT2 cells and primary human fibroblasts.....	215
4.2.9. Effect on hedgehog signalling upon <i>Cep120</i> silencing on Shh-LIGHT2 cells.....	224
4.2.10. Site directed mutagenesis of WT <i>CEP120</i> plasmid and optimisation of transfection efficiency.....	234
4.3. Discussion.....	243
4.3.1. <i>CEP120</i> and <i>CC2D2A</i> in silico analyses and its combination with <i>ex vivo</i> detection of basal exon skipping.....	243
4.3.2. Use of 100,000 Genomes Project to find patients presenting pathogenic variants in <i>CEP120</i> and <i>TOGARAM1</i>	246
4.3.3. Design of a pipeline to study patient-specific <i>CEP120</i> mutations using cell-based assays.....	249
Chapter 5 Investigation of genes involved in the endoplasmic reticulum (ER) and use of Human Urine-Derived Renal Epithelial Cells (hURECs) to complement genetic studies.....	253
5.1. Introduction and aims.....	253
5.2. Results.....	256
5.2.1. Novel patients with pathogenic mutations in <i>DNAJB11</i>	257
5.2.2. Novel patients with pathogenic mutations in <i>GANAB</i>	261

5.2.3. Novel patients with pathogenic mutations in <i>ALG5</i>	265
5.2.4. Novel patients with pathogenic mutations in <i>ALG8</i>	267
5.2.5. Novel patients with pathogenic mutations in <i>ALG9</i>	269
5.2.6. Enrichment of loss-of-function (LoF) variants in <i>ALG1</i> , <i>ALG8</i> , <i>ALG9</i> and <i>ALG12</i> in a population with kidney and/or liver cysts.....	270
5.2.7. Selection of control hURECs for genetic studies.....	285
5.2.8. Characterisation of the ciliary and molecular phenotype of hURECs from a patient with a nonsense mutation in <i>ALG8</i>	291
5.3. Discussion.....	309
Chapter 6 Final discussion and concluding remarks.....	322
6.1. Complexity of renal ciliopathies.....	322
6.2. Results from this study, its limitations and future work.....	324
6.3. Future perspectives: the use of the Genomics England 100,000 Genomes Project and hURECs to study renal ciliopathies.....	332
6.4. Final conclusions.....	336
References.....	338
Appendices.....	384
Appendix 1. Extract variants by coordinate script (modified from Genomics England).....	384
Appendix 2. List of genes associated with ciliopathies (Gold Standard SYSCILIA).386	386
Appendix 3. List of terms used to select case population.....	389
Appendix 4. Distribution of the patients selected for the case population (n=1922), in terms of their diagnosed disease.....	391
Appendix 5. Distribution of the patients selected for the control population (n=16773), in terms of their diagnosed disease.....	392
Appendix 6. Genomic position of the loss-of-function (LoF) variants found in cases in the genes <i>ALG1</i> , <i>ALG8</i> , <i>ALG9</i> and <i>ALG12</i>	394
Appendix 7. Genomic position of the loss-of-function (LoF) variants found in controls in the genes <i>ALG1</i> , <i>ALG8</i> , <i>ALG9</i> and <i>ALG12</i>	395
Appendix 8. List of publications.....	396

List of Figures

Figure 1.1. Structure of the primary cilium.....	3
Figure 1.2. Centriole duplication cycle.....	8
Figure 1.3. Phenotypes and syndromes associated with primary ciliopathies	9
Figure 1.4. Representative example of the “molar tooth sign” (MTS) distinctive brain malformation, which is characteristic of Joubert syndrome (JBTS).....	12
Figure 1.5. Genetic heterogeneity and genetic pleiotropy of the genes associated with Joubert syndrome (JBTS).....	14
Figure 1.6. Variability and overlap in the age of disease onset in patients with renal ciliopathies.....	16
Figure 1.7. Representation of cystic kidneys in ADPKD.....	17
Figure 1.8. Genetic heterogeneity and genetic pleiotropy of the genes associated with short-rib polydactyly (SRP)/Jeune asphyxiating thoracic dystrophy (JATD)....	22
Figure 1.9. Genetic heterogeneity and genetic pleiotropy in genes associated with ADPKD, ADPLD and ADTKD.....	24
Figure 1.10. Representation of the domain structure of CEP120 protein (NP_694955.2).....	36
Figure 1.11. Cellular compartments associated with renal ciliopathies with a focus on the endoplasmic reticulum (ER).....	47
Figure 1.12. Representative example of immunofluorescence microscopy imaging of human urine-derived renal epithelial cells (hURECs) from a healthy control individual.....	52
Figure 2.1. Representative pipeline of the collection, processing and culture of human urine-derived renal epithelial cells (hURECs).....	65
Figure 2.2. Summary diagram of the types of formalin fixed paraffin embedded (FFPE) sections of human tissue utilised in this study and the corresponding human developmental stage of each of them	72
Figure 2.3. Representative image of a sagittal section of a human embryo at Carnegie Stage 23 (CS23) and location of mayor organs.....	72

Figure 2.4. Pipeline followed for the <i>in silico</i> identification of skippable exons in <i>CEP120</i> and <i>CC2D2A</i> genes which are suitable for an ASOs-based exon skipping therapy.....	77
Figure 2.5. RT-PCR using cDNA, reverse transcribed from total RNA isolated from Shh-LIGHT2 cells.....	82
Figure 2.6. RT-PCR using cDNA, reverse transcribed from total RNA isolated from human urine-derived renal epithelial cells (hURECs) from an ADPKD patient.....	83
Figure 2.7. Sequence alignment of the siRNA targeting <i>CEP120/Cep120</i> (D-016493-01; Dharmacon).....	86
Figure 2.8. Sequence alignment of the ASO morpholino targeting <i>CEP120</i> (GeneTools).....	88
Figure 2.9. Schematic workflow of a rescue experiment to study patient-specific <i>CEP120</i> mutations.....	89
Figure 2.10. Plasmid map of WT <i>CEP120</i> plasmid (hCEP120-EGFP, Addgene plasmid # 50382).....	93
Figure 2.11. WT <i>CEP120</i> plasmid (hCEP120-EGFP) restriction mapping.....	94
Figure 2.12. Plasmid map of pIRES2-EGFP plasmid (pIRES2-EGFP, Clontech # 6029-1, Addgene # 3178).....	95
Figure 2.13. pIRES2-EGFP plasmid (pIRES2-EGFP) restriction mapping.....	96
Figure 2.14. Plasmid map of EGFP plasmid.....	97
Figure 2.15. EGFP plasmid (EGFP) restriction mapping.....	98
Figure 2.16. Pipeline to detect candidate disease-causing variants in <i>DNAJB11</i> , <i>GANAB</i> , <i>ALG5</i> , <i>ALG8</i> and <i>ALG9</i>	106
Figure 2.17. Pipeline to investigate if there is an enrichment of heterozygous alleles in any of the <i>ALG</i> genes in a population of patients with cysts in kidney or liver, compared to a non-cystic control population.....	109
Figure 3.1. Representative figure of the RNAscope procedure.....	114
Figure 3.2. Amino acid sequences of the domains of the human <i>CEP120</i> protein aligned against different <i>CEP120</i> orthologs using BLASTP.....	121

Figure 3.3. Expression of <i>CEP120</i> (red) in a human fetal brain of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green.....	127
Figure 3.4. Expression of <i>ARL3</i> (red) in a human fetal brain of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green.....	128
Figure 3.5. Expression of <i>Ki67</i> (red) in a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green.....	129
Figure 3.6. Expression of <i>CEP120</i> (red) in sagittal sections of human hindbrain at 14 PCW (A) and 19 PCW (B) stained using RNAscope, counterstained with Methyl Green.....	131
Figure 3.7. Expression of <i>ARL3</i> (red) in sagittal sections of human hindbrain at 14 PCW (A) and 19 PCW (B) stained using RNAscope, counterstained with Methyl Green.....	132
Figure 3.8. Expression of <i>Ki67</i> (red) in a sagittal section of human hindbrain at 18 PCW stained using RNAscope, counterstained with Methyl Green.....	133
Figure 3.9. Schematic diagrams of the <i>CEP120</i> and <i>ARL3</i> expression in developing cerebellum at (f) 14 PCW and (g) 19 PCW.....	133
Figure 3.10. Diagram showing the development of the retina layers.....	135
Figure 3.11. Expression of <i>CEP120</i> (red) in sagittal sections of human eye at 8 PCW (A) and 14 PCW (B) stained using RNAscope, counterstained with Methyl Green.....	136
Figure 3.12. Expression of <i>ARL3</i> (red) in sagittal sections of human eye at 8 PCW (A) and 14 PCW (B) stained using RNAscope, counterstained with Methyl Green.....	137
Figure 3.13. Expression of <i>CEP120</i> (red) in the developing human dorsal root ganglia of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green	138
Figure 3.14. Expression of <i>ARL3</i> (red) in the developing human dorsal root ganglia of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green	139

Figure 3.15. Expression of <i>KI67</i> (red) in the developing human dorsal root ganglia of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green.....	140
Figure 3.16. Expression of <i>CEP120</i> (red) in sagittal sections of human kidney at 8 PCW (A), 14 PCW (B) and 18 PCW (C) stained using RNAscope, counterstained with Methyl Green.....	142
Figure 3.17. Expression of <i>ARL3</i> (red) in sagittal sections of human kidney at 8 PCW (A), 14 PCW (B) and 18 PCW (C) stained using RNAscope, counterstained with Methyl Green.....	143
Figure 3.18. Expression of <i>KI67</i> (red) in the developing human kidney of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green.....	144
Figure 3.19. Expression of <i>CEP120</i> (red) in sagittal sections of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green.....	145
Figure 3.20. Expression of <i>ARL3</i> (red) in sagittal sections of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green.....	146
Figure 3.21. Expression of <i>KI67</i> (red) in sagittal sections of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green.....	147
Figure 3.22. Expression of <i>dapB</i> (red) in sagittal sections of human hindbrain at 14 PCW (A) and 19 PCW (B) and an eye section at 14 PCW (C) stained using RNAscope, counterstained with Methyl Green.....	148
Figure 3.23. Immunohistochemistry staining of OLIG2 in developing human brain. Cortical section of 8 PCW-stage human embryo stained with OLIG2 (brown) using immunohistochemistry counterstained with Toluidine Blue.....	150
Figure 3.24. Immunohistochemistry staining of PAX6 in developing human brain. Sagittal section of 8 PCW-stage human embryo stained with PAX6 (brown) using immunohistochemistry counterstained with Toluidine Blue.....	151

Figure 3.25. Immunohistochemistry staining of PAX6 in developing human hindbrain. Sagittal section of 8 PCW-stage human embryo stained with PAX6 (brown) using immunohistochemistry counterstained with Toluidine Blue.....	152
Figure 3.26. Immunohistochemistry staining of PAX6 in developing human hindbrain. Sagittal section of human hindbrain at 19 PCW stained with PAX6 (brown) using immunohistochemistry counterstained with Toluidine Blue.....	153
Figure 3.27. Immunohistochemistry staining in developing human hindbrain (negative control in which the sample was not treated with primary antibody).....	154
Figure 4.1. Phenotypes and genotypes in patients carrying biallelic disease-causing <i>CEP120</i> variants reported in the literature.....	170
Figure 4.2. Phenotypes and genotypes in patients carrying biallelic disease-causing <i>CC2D2A</i> variants reported in the literature.....	172
Figure 4.3. Tissue specific transcript expression of <i>CEP120</i> and its exon structure.....	176
Figure 4.4. Tissue specific transcript expression of <i>CC2D2A</i> and its exon structure.....	178
Figure 4.5. RT-PCR showing basal exon skipping of <i>CEP120</i> exon 2 in kidney and whole blood.....	179
Figure 4.6. RT-PCR showing basal exon skipping of <i>CC2D2A</i> exon 30 in human kidney and human urine-derived renal epithelial cells (hURECs).....	181
Figure 4.7. Distribution of exons and the prevalence of kidney disease in patients with truncating <i>CC2D2A</i> per exon.....	182
Figure 4.8. Schematic representation of the 21 exons corresponding to the <i>CEP120</i> human transcript NM_153223.3 (also called ENST00000328236.9).....	184
Figure 4.9. Schematic representation of the 21 exons corresponding to the <i>CEP120</i> human transcript NM_153223.3 (also called ENST00000328236.9) and the location of the reported <i>CEP120</i> mutations.....	185
Figure 4.10. Schematic representation of the 38 exons corresponding to the <i>CC2D2A</i> human transcript NM_001080522 (also called ENST00000503292.5).....	187

Figure 4.11. Schematic representation of the 38 exons corresponding to the CC2D2A human transcript NM_001080522.2 (also called ENST00000503292.5) and the location of the reported CC2D2A mutations.....	188
Figure 4.12. Identified CC2D2A exon targets for potential exon skipping therapies and predicted consequences of exon skipping.....	192
Figure 4.13. Pipeline to search novel patients with pathogenic CEP120 variants in the Genomics England 100,000 Genomes Project (main release v9, dated 2nd April 2020).....	194
Figure 4.14. Pedigree diagram showing the segregation of a missense CEP120 variant in a consanguineous family.....	207
Figure 4.15. Pedigree diagram showing the families presenting CEP120 mutations already reported in the literature.....	208
Figure 4.16. Variant segregations and sequencing visualisation of the in a family presenting a novel disease-causing TOGARAM1 variants leading to JBTS.....	212
Figure 4.17. Immunofluorescence microscopy in human fibroblasts.....	217
Figure 4.18. Immunofluorescence microscopy in human fibroblasts.....	218
Figure 4.19. Immunofluorescence microscopy in mouse Shh-LIGHT2 cells.....	218
Figure 4.20. Immunofluorescence microscopy in mouse Shh-LIGHT2 cells.....	219
Figure 4.21. Immunofluorescence microscopy in mouse Shh-LIGHT2 cells.....	219
Figure 4.22. Immunofluorescence microscopy in mouse Shh-LIGHT2 cells.....	220
Figure 4.23. Immunofluorescence microscopy in human fibroblasts.....	222
Figure 4.24. Immunofluorescence microscopy in mouse Shh-LIGHT2 cells.....	223
Figure 4.25. Western blotting showing Cep120 in protein lysates from mouse kidney tissue.....	224
Figure 4.26. Cep120 expression upon siRNA knockdown in Shh-LIGHT2.....	225
Figure 4.27. Smo expression upon siRNA knockdown in Shh-LIGHT2.....	226
Figure 4.28. Gli expression upon siRNA knockdown in Shh-LIGHT2.....	227
Figure 4.29. Patch1 expression upon siRNA knockdown in Shh-LIGHT2.....	227

Figure 4.30. <i>Cep120</i> expression upon siRNA knockdown in Shh-LIGHT2.....	228
Figure 4.31. <i>Smo</i> expression upon siRNA knockdown in Shh-LIGHT2.....	229
Figure 4.32. Luciferase assay reveals SAG stimulation impairment upon <i>Smo</i> siRNA knockdown in Shh-LIGHT2 cells.....	230
Figure 4.33. <i>CEP120</i> expression upon morpholino knockdown in human fibroblasts.....	232
Figure 4.34. List of the mutated plasmids from the <i>CEP120</i> plasmid (Mut hCEP120-EGFP) restriction mapping.....	235
Figure 4.35. Sequencing of the each of the Mut hCEP120-EGFP plasmids.....	336
Figure 4.36. Transfection of EGFP plasmid or pIRES2-EGFP into Shh-LIGHT2 cells using different transfection reagents.....	241
Figure 4.37. Transfection of EGFP plasmid into Shh-LIGHT2 cells.....	242
Figure 4.38. Transfection of EGFP plasmid into human fibroblasts.....	242
Figure 5.1. Family 5-DNAJB11 presenting a disease-causing variant in <i>DNAJB11</i>	261
Figure 5.2. Families 1-ALG5 and 2-ALG5 presenting disease-causing <i>ALG5</i> variants.....	265
Figure 5.3. Distribution of the participants in the Rare Disease Program (main release v14, dated 27 th of January 2022) to visualise where the case and control populations are selected from.....	272
Figure 5.4. Distribution of the participants in the Rare Disease Program (main release v14, dated 27 th of January 2022).....	273
Figure 5.5. Distribution of the case population created from the Genomics England 100,000 Genomes Project according to the diagnosed disease of the patients and their status regarding to if they are solved or not by Genomics England.....	276
Figure 5.6. Distribution of the control population created from the Genomics England 100,000 Genomes Project according to the diagnosed disease of the patients and their status regarding to if they are solved or not by Genomics England.....	277

Figure 5.7. Enrichment of predicted loss-of-function alleles (LoF) alleles in <i>ALG1</i> , <i>ALG8</i> and <i>ALG12</i> in the population of cases compared to the population of controls.....	278
Figure 5.8. Bar plot showing the percentage of loss-of-function alleles (LoF) alleles in the case and control populations	280
Figure 5.9. Enrichment of predicted loss-of-function alleles (LoF) alleles in <i>ALG9</i> and <i>ALG12</i> in the population of cases compared to the population of controls that are genetically unsolved by Genomics England.....	281
Figure 5.10. Increased odds ratio (OR) of having a predicted loss-of-function alleles (LoF) allele in cases against having a LoF allele in controls, in <i>ALG1</i> , <i>ALG8</i> , <i>ALG9</i> and <i>ALG12</i>	282
Figure 5.11. Lack of enrichment missense alleles in cases against missense alleles in controls in <i>ALG1</i> , <i>ALG8</i> , <i>ALG9</i> and <i>ALG12</i>	284
Figure 5.12. Representative light microscopy images of human urine-derived renal epithelial cells (hURECs) that grew well (were passaged for at least one more time) from healthy individuals (Controls).....	287
Figure 5.13. Representative images of human urine-derived renal epithelial cells (hURECs) that resembled fibroblasts rather than epithelial cells from healthy individuals (Controls).....	288
Figure 5.14. Representative images of human urine-derived renal epithelial cells (hURECs) that look fibroblastic (they did not survive the corresponding passage in which the corresponding picture was taken) from healthy individuals (Controls)....	289
Figure 5.15. Diagram showing the Passage number (P) in which each the hURECs of each control individual stopped growing.....	290
Figure 5.16 Pedigree diagram showing Family 7- <i>ALG8</i> presenting a disease-causing <i>ALG8</i> variant	291
Figure 5.17. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from a Family 7- <i>ALG8</i> patient with cystic kidney disease.....	292
Figure 5.18. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from a Family 7- <i>ALG8</i> patient with cystic kidney disease.....	293

Figure 5.19. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from a healthy control (Control A).....	294
Figure 5.20. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from a healthy control (Control A).....	295
Figure 5.21. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from a healthy control (Control A).....	296
Figure 5.22. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from Family 7- <i>ALG8</i> patient with cystic kidney disease.....	297
Figure 5.23. Increased cilia length in human urine-derived renal epithelial cells (hURECs) in passage 2 (P2), from a Family 7- <i>ALG8</i> patient with an <i>ALG8</i> mutation.....	298
Figure 5.24. Increased cilia length in human urine-derived renal epithelial cells (hURECs) in passage 3 (P3), from a Family 7- <i>ALG8</i> patient with an <i>ALG8</i> mutation.....	299
Figure 5.25. Increased cilia length in human urine-derived renal epithelial cells (hURECs) from Family 7- <i>ALG8</i> patient with an <i>ALG8</i> mutation.....	300
Figure 5.26. Increased mean cilia length in human urine-derived renal epithelial cells (hURECs) from Family 7- <i>ALG8</i> patient with an <i>ALG8</i> mutation, from 3 different experiments.....	301
Figure 5.27. <i>ALG8</i> expression in human urine-derived renal epithelial cells (hURECs) from Family 7- <i>ALG8</i> patient.....	302
Figure 5.28. <i>PKD1</i> expression in human urine-derived renal epithelial cells (hURECs) from a patient with an <i>ALG8</i> mutation.....	303
Figure 5.29. <i>PKD2</i> expression in human urine-derived renal epithelial cells (hURECs) from a patient with an <i>ALG8</i> mutation.....	304
Figure 5.30. Differences in proliferation rate of the human urine-derived renal epithelial cells (hURECs) from a patient with an <i>ALG8</i> mutation (<i>ALG8</i> patient) compared with a healthy control (Control A).....	305

Figure 5.31. Cell viability study of the human urine-derived renal epithelial cells (hURECs) from Family 7-*ALG8* patient with an *ALG8* mutation (*ALG8* patient) and the hURECs from a healthy control (Control A) in different passages (P).....307

Figure 5.32. Representative images of hURECs from Family 7-*ALG8* patient with an *ALG8* mutation (*ALG8* patient) in different passages (P).....308

List of Tables

Table 1.1. Summarizing table of different types of primary and motile cilia found in different tissues and cell types.....	5
Table 1.2. List of Joubert syndrome (JBTS) genes and JBTS candidate genes.....	20
Table 1.3. Genes associated and gene candidates to be associated with cystic kidney disease.....	25
Table 1.4. Genes associated with nephronophthisis (NPHP).....	29
Table 1.5. Comparison of the different the human CEP120 protein and CEP120 protein domains against CEP120 orthologues.....	37
Table 1.6. List of patients with biallelic <i>CEP120</i> variants and their associated phenotypes reported in the literature (ranked chronologically by publication).....	40
Table 1.7. List of reported genetic variants in <i>CEP120</i> (ranked by genomic position).....	41
Table 2.1. List of software tools used.....	62
Table 2.2. List of online resources used.....	63
Table 2.3. List of antibodies with its corresponding dilution/s used, sorted by type of antibody (primary or secondary) and technique in which they were used.....	69
Table 2.4. List of autosomal dominant polycystic kidney disease (ADPKD) genes and ADPKD candidates, including the rest of <i>ALG</i> genes.....	108
Table 3.1. Comparison of the known phenotypes associated with <i>CEP120</i> and <i>ARL3</i> mutations.....	119
Table 3.2. Comparison of the homology of <i>CEP120</i> and <i>ARL3</i> amino acid sequences.....	120
Table 3.3. Comparison of the ciliary localisation, protein domains, function and protein interactors of the <i>CEP120</i> and <i>ARL3</i> proteins.....	123
Table 3.4. RNA expression values of various ciliopathy genes, obtained from two public databases: Genotype-Tissue Expression (GTEx) Project and Human Protein Atlas (HPA).....	124

Table 4.1. Patients with truncating <i>CC2D2A</i> variants in potentially skippable exons and associated phenotypes (ranked by publication).....	190
Table 4.2. Phenotypic information of the families investigated to find novel patients with pathogenic <i>CEP120</i> variants in the Genomics England 100,000 Genomes Project.....	196
Table 4.3. Variants from the families investigated in the Genomics England 100,000 Genomes Project to find novel disease-causing <i>CEP120</i> variants.....	200
Table 4.4. Allele frequency of the <i>CEP120</i> mutations published in the literature up to date found in the rare disease cohort of the Genomics England 100,000 Genomes Project.....	210
Table 4.5. Clinical features of JAS-50, a patient with <i>TOGARAM1</i> -related JBTS...	211
Table 4.6. Predicted pathogenicity and allele frequency of <i>TOGARAM1</i> variants found in JAS-50 patient with <i>TOGARAM1</i> -related JBTS.....	211
Table 4.7. Predicted pathogenicity of <i>CCDC28B</i> variant: c.73C>T; p.(Arg25Trp)...	215
Table 4.8. Allele frequency of <i>CCDC28B</i> variant: c.73C>T; p.(Arg25Trp).....	215
Table 5.1. Clinical features of the unrelated <i>DNAJB11</i> -affected individuals identified in the Genomics England 100,000 Genomes Project database.....	258
Table 5.2. Detailed clinical features of the <i>DNAJB11</i> -affected proband presenting the variant: c.730A>T; p.Lys244Ter (Patient 5-DNAJB11) and her affected relatives...	259
Table 5.3. Variants found in the <i>DNAJB11</i> -affected individuals found in the Genomics England 100,000 Genomes Project database.....	260
Table 5.4. Clinical features of the unrelated individuals with rare heterozygous variants in <i>GANAB</i> identified in the Genomics England 100,000 Genomes Project database.....	262
Table 5.5. Variants found in <i>GANAB/PKD1/PKD2</i> -affected individuals found in the Genomics England 100,000 Genomes Project database.....	263
Table 5.6. Clinical features of the <i>ALG5</i> -affected patients identified in the Genomics England 100,000 Genomes Project database.....	266

Table 5.7. Variants found in <i>ALG5</i> -affected individuals found in the Genomics England 100,000 Genomes Project database.....	267
Table 5.8. Clinical features of the unrelated <i>ALG8</i> -affected patients identified in the Genomics England 100,000 Genomes Project database.....	268
Table 5.9. Variants found in <i>ALG8</i> -affected individuals found in the Genomics England 100,000 Genomes Project database.....	268
Table 5.10. Clinical features of the unrelated <i>ALG9</i> -affected patients identified in the Genomics England 100,000 Genomes Project database.....	269
Table 5.11. Variants found in <i>ALG9</i> -affected individuals found in the Genomics England 100,000 Genomes Project database.....	269
Table 5.12. Distribution of the participants in the rare disease program diagnosed with at least one disease in terms of their disease group.....	274
Table 5.13. Distribution of the participants included in rare disease group: Ciliopathies, in terms of their diagnosed disease.....	274
Table 5.14. Distribution of the participants included in rare disease group: Renal and urinary tract disorders, in terms of their diagnosed disease.....	275
Table 5.15. Enrichment of loss-of-function (LoF) alleles in <i>ALG1</i> , <i>ALG8</i> , <i>ALG9</i> and <i>ALG12</i> in the case population.....	279
Table 5.16. Non-enrichment of missense alleles in <i>ALG1</i> , <i>ALG8</i> , <i>ALG9</i> and <i>ALG12</i> in the population of cases compared to the population of controls.....	284
Table 5.17. Loss of function (LoF) variants in <i>ALG1</i> and <i>ALG12</i> found the case population in the Genomics England 100,000 Genomes Project database.....	285
Table 5.18. Description of the survival the human urine derived renal epithelial cells (hURECs) samples used from Control individuals and the <i>ALG8</i> patient.....	286

List of abbreviations

AC-TUB	Acetylated Tubulin
ASO	Antisense oligonucleotide
ADPKD	Autosomal dominant polycystic kidney disease
ADPLD	Autosomal recessive polycystic liver disease
ALG	Asparagine-Linked Glycosylation
ARL3	ADP-ribosylation factor-like 3
ARPKD	Autosomal recessive polycystic kidney disease
BBS	Bardet-Biedl syndrome
CAKUT	Congenital anomalies of kidney and urinary tract
CC2D2A	Coiled-coil and C2 domain containing 2A
CEP120	Centrosomal protein of 120 kDa
CGNP	Cerebellar granule neuron progenitors
CNS	Central nervous system
CNV	Copy number variation
eGFR	Estimated glomerular filtration rate
EGL	External granule layer
ER	Endoplasmic reticulum
ESKD	End-stage kidney disease
FBS	Foetal bovine serum
GMC	Genomic Medicine Centres
gnomAD	Genome Aggregation database
GTEX	Genotype-Tissue Expression (Project)
HDBR	MRC-Wellcome Human Developmental Biology Resource
hUREC	Human Urine-derived Renal Epithelial Cell

JATD	Jeune asphyxiating thoracic dystrophy
JBTS	Joubert syndrome
JSRD	Joubert syndrome and related disorders
LoF	Loss-of-function
IF	Immunofluorescence
IFT	Intraflagellar transport
IGL	Internal granule layer
LCA	Leber congenital amaurosis
MKS	Meckel-Gruber syndrome
ML	Molecular layer
MRI	Magnetic resonance imaging
MTS	Molar tooth sign
NCBI	National Center for Biotechnology Information
NPHP	Nephronophthisis
OFD	Orofaciodigital syndrome
OMIM	Online Mendelian Inheritance In Man
OR	Odds ratio
PBS	Phosphate-buffered saline
PC1	Polycystin-1
PC2	Polycystin-2
PCNT	Pericentrin
PCR	Polymerase chain reaction
PCW	Post conception weeks
PKD	Polycystic kidney disease
qPCR	Quantitative PCR

RD	Retinal dystrophy
RTPCR	Reverse transcription PCR
SHH	Sonic Hedgehog
siRNA	Small interfering RNA
SLS	Senior-Løken syndrome
SRP	Short-rip polydactyly
SVs	Structural variants
TCDOE	Tectocerebellar dysraphia with occipital encephalocele
TOGARAM1	TOG array regulator of axonemal microtubules 1
TZ	Transition zone
VCF	Variant Call Format
VEP	Variant Effect Predictor
VUS	Variant of uncertain significance
WB	Western blot
WES	Whole exome sequencing
WGS	Whole genome sequencing

Chapter 1. Introduction

1.1. Primary Ciliopathies

Ciliopathies are a group of complex inherited diseases, characterised by genetic heterogeneity and a wide spectrum of clinical phenotypes (Shaheen et al., 2016). Ciliopathies are divided into two types: motile ciliopathies, if they are caused by dysfunction of motile cilia, or primary ciliopathies if they are caused by dysfunction of primary cilia. There is genetic and phenotypic overlap between both types of ciliopathies and within primary ciliopathies. There are more than a dozen distinguishable primary ciliopathies (Novarino et al., 2011, Braun and Hildebrandt, 2017, Focsa et al., 2021). Within the spectrum of ciliopathies are rare conditions such as: Joubert syndrome (JBTS), nephronophthisis (NPHP), Senior-Løken syndrome (SLS), Orofaciodigital syndrome (OFD), Jeune asphyxiating thoracic dystrophy (JATD), autosomal dominant and recessive polycystic kidney disease (ADPKD and ARPKD), Leber congenital amaurosis (LCA), Meckel-Gruber syndrome (MKS), Bardet-Biedl syndrome (BBS), Usher syndrome (US) and some forms of retinal dystrophy (RD) (Kempeneers and Chilvers, 2018, Novarino et al., 2011, Wheway et al., 2018).

Patients diagnosed with a primary ciliopathy often develop multi-systemic phenotypes. These clinical phenotypes, such as cystic kidney disease, retinal degeneration, neurologic phenotypes, skeletal phenotypes or obesity often overlap between different ciliopathies (Wheway et al., 2018, Focsa et al., 2021, Braun and Hildebrandt, 2017).

Renal ciliopathies is a clinically and genetically heterozygous group of primary ciliopathies. The clinical phenotypic spectrum of renal ciliopathies varies from cystic kidney disease to interstitial fibrosis. Renal ciliopathies include ADPKD, ARPKD and NPHP (Devlin and Sayer, 2019, McConnachie et al., 2021). The phenotypic spectrum of renal ciliopathies is not limited to kidney phenotypes. In NPHP, around 20% of cases are associated with other extra-renal phenotypes, affecting organs such as brain and eye (Novarino et al., 2011, Reiter and Leroux, 2017). NPHP is a renal phenotype seen in multisystem primary ciliopathies such as JBTS (Bachmann-Gagescu et al., 2015a, Brancati et al., 2010, Sakakibara et al., 2022). Patients diagnosed with PKD may also show extra-renal phenotypes such as hepatic cysts, intracranial aneurysms and hypertension (Cornec-Le Gall et al., 2019, Besse et al.,

2019, Pisani et al., 2022). NPHP and PKD are important components of the phenotypic spectrum of primary renal ciliopathies and they are often found in other primary ciliopathies (McConnachie et al., 2021).

1.1.1. Primary cilia

Cilia (singular: cilium) and flagella (singular: flagellum) are the oldest known organelles, they were discovered by Antony Van Leeuwenhoek in 1675 (Satir, 1995).

Even though it was suspected that cilia could have a sensory function (Moran et al., 1977) and cause disease (Afzelius, 1976), it was not until the 1990s and 2000s when the connection between dysfunction of cilia and primary ciliopathies, such as ARPKD (Moyer et al., 1994), NPHP (Otto et al., 2002), JBTS (Sayer et al., 2006), BBS (Ansley et al., 2003) and JATD (Beales et al., 2007) was established (Watnick and Germino, 2003). Subsequently, the number of studies investigating the role of cilia in human disease and development have been rapidly increasing in the last two decades (Satir, 2017).

Cilia are microtubule-based organelles, cellular hair-like projections, which are conserved in eukaryotes and are found in almost all vertebrate cell types (Marra et al., 2016, Mitchison and Valente, 2017, Choksi et al., 2014).

The cilium protrudes from the cell surface and originates from the basal body. The cell body is a modified centriole, a cylindrical structure formed by nine-triplets of microtubules reinforced by a cartwheel-like structure in the centre (Nigg and Stearns, 2011, Vincensini et al., 2011). The core of the cilium is a microtubule scaffold called the axoneme which extends from the basal body and is surrounded by the ciliary membrane (Gonçalves and Pelletier, 2017) (Figure 1.1).

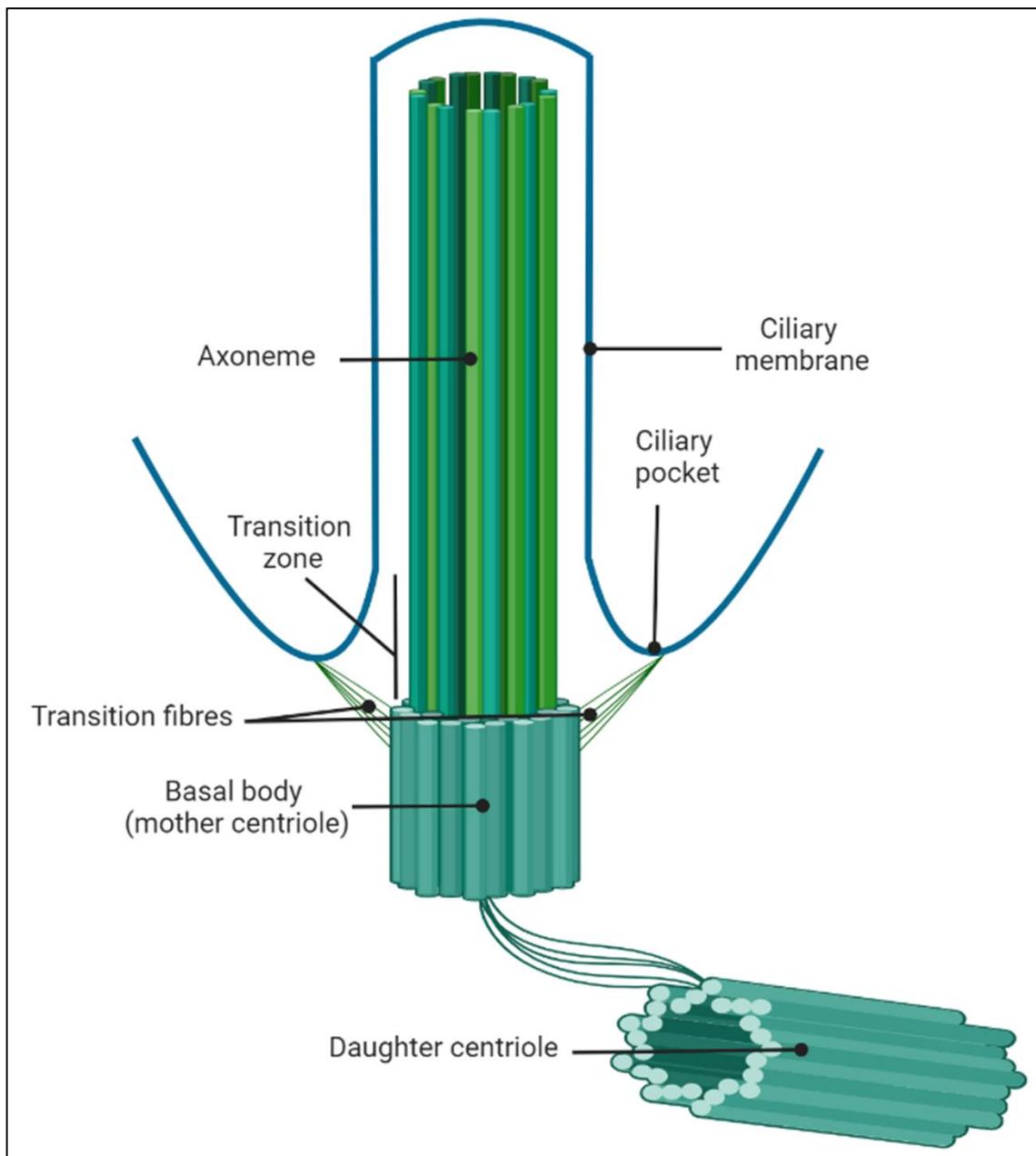


Figure 1.1. Structure of the primary cilium. The core of the primary cilium is the axoneme. The axoneme consists in nine doublets of microtubules. The axoneme is enclosed by the ciliary membrane, which is continuous to the plasma membrane. The axoneme emanates from the basal body, a modified mother centriole. Transition fibres attach the basal body to the plasma membrane at the transition zone (TZ), which regulates protein and lipid entry and exit from the primary cilium. The ciliary pocket has a role as a hub for actin dynamics and trafficking of vesicles (Mahjoub, 2013, Mirvis et al., 2018). Image generated using BioRender online tool.

Unlike the basal body, the axoneme is formed by nine-doublets of microtubules, the change from triplets to doublets of microtubules occurs at a ciliary domain called the transition zone (TZ), which controls the entry and exit of proteins and lipids from the ciliary compartment (Gonçalves and Pelletier, 2017).

The cilium axoneme may have a central pair of microtubules, distinguishing two different configurations of cilia: 9+2 axoneme, when a central pair of microtubules is present, and 9+0 axoneme, when a central pair of microtubules is not present (Choksi et al., 2014, Mirvis et al., 2018).

Traditionally in vertebrates, cilia have been classified into two classes: motile cilia and primary (or non-motile) cilia. The differences between primary and motile cilia generally refer to the structure of their axoneme and the presence of dynein arms which allows motility of cilia (Mirvis et al., 2018, Mitchison and Valente, 2017). Both types of cilia have different functions, which often complement each other, and are found in multiple tissues (Table 1.1).

Motile cilia generally have an axoneme formed by nine-doublets of microtubules and a central doublet of microtubules (9+2 axoneme). Motile cilia can exist as single monocilia on the cell surface (such as the flagella on sperm cells or protozoans) or multicilia (such as the epithelial cells of the respiratory tract or ependymal cells of the central nervous system of mammals) (Choksi et al., 2014).

In contrast, primary cilia typically present an axoneme without a central doublet of microtubules (9+0 axoneme) and exist as single monocilia on the cell surface. Primary cilia are microtubule-based organelles, which emanate from a basal body (a modified centriole). Primary cilia have an essential role in embryonic development and cell homeostasis and allow cells to detect and manage external signals (Malicki and Johnson, 2017, Berbari et al., 2009, Wheway et al., 2018).

Cell type	Location	Axoneme	Motile	Number	Function	Reference
Sperm cells	Reproductive system	9+2	Yes	1	Generation of sperm cells movement	(Choksi et al., 2014)
Fallopian tube epithelial cells	Reproductive system	9+2	Yes	multi	Support of fertilization and early embryogenesis	(Lyons et al., 2006)
Epithelial cells of respiratory tracts and middle ear	Airway epithelia (trachea, lungs, ear canals)	9+2	Yes	multi	Mucociliary clearance	(Choksi et al., 2014)
Ependymal cells	Ventricles of the central nervous system	9+2	Yes	multi	Generation and regulation of cerebrospinal fluid flow	(Omran et al., 2017)
Node cells	Ventral node - central region	9+0	Yes	1	Generation of leftward-directed fluid flow	(Babu and Roy, 2013)
Olfactory neurons	Olfactory epithelium	9+2	No	multi	Detection of odorants	(Bergboer et al., 2018)
Node cells	Ventral node - peripheral region	9+0	No	1	Detection of fluid flow	(Babu and Roy, 2013)
Epithelial cells of nephron tubules and collecting ducts	Kidney	9+0	No	1	Detection of urine flow, composition and osmolality	(Marra et al., 2016)
Cortical progenitor cells and neuronal cells	Brain	9+0	No	1	Detection of chemical and mechanical signals from cell environment	(Mitchison and Valente, 2017)
Neural retina, cone and rod photoreceptors	Retina	9+0	No	1	Detection of light	(Wheway et al., 2014)
Cholangiocytes (epithelial cells of biliary ducts)	Liver	9+0	No	1	Detection of chemical and mechanical signals from cell environment	(Mitchison and Valente, 2017)

Table 1.1. Summarizing table of different types of primary and motile cilia found in different tissues and cell types. Table shows some of the most relevant cell types with motile cilia (shown in green) and primary (non-motile) cilia (shown in blue). Motile cilia are generally characterised by 9+2 axoneme and dynein arms that allow cilia movement (Mitchison and Valente, 2017, Kobayashi and Dynlacht, 2011). Primary cilia are generally characterised by an axoneme without a central doublet of microtubules (9+0 axoneme), they are non-motile and do not have dynein arms (Choksi et al., 2014, Mirvis et al., 2018). An exception to this classification is the cilia in the olfactory epithelium, these have a 9+2 axoneme structure (Bergboer et al., 2018). Another exception to this classification is nodal cilia, found in an embryonic structure called the ventral node, with characteristics of both primary and motile cilia (Babu and Roy, 2013).

Motile cilia are typically involved in cellular locomotion (e.g. allowing movement of sperm cells) and the movement of fluids (e.g. promoting mucus flow in the nasal ciliated epithelial cells) in a variety of tissues such as the respiratory system, the reproductive system and brain ventricles (Lyons et al., 2006, Omran et al., 2017, Choksi et al., 2014). Moreover, motile cilia present dynein arms which allow ciliary beating through ATP hydrolysis, while primary cilia lack these dynein motor proteins. Motile ciliopathies are characterised by failure in generation fluid flow or cell movement, which lead to defects in laterality, brain development, mucus clearance and fertility (Mitchison and Valente, 2017). Primary ciliary dyskinesia is a well-characterised motile ciliopathy causing a wide spectrum of phenotypes such as situs inversus (laterality defect which can also be found in the primary ciliopathies phenotypic spectrum), respiratory and heart diseases and infertility (Mitchison and Shoemark, 2017).

Primary cilia also function as signal transducers in numerous signalling pathways. The primary cilium has been defined as the cell's antenna. It has sensory and transduction roles (Reiter and Singla, 2006, Wheway et al., 2018, Pala et al., 2017). Primary cilia are present in most of mammalian cell types and function as the signalling hub of the cell for various signalling pathways. Even though motile cilia can also sense extracellular environment (e.g. in mammalian respiratory epithelium cells) (Wheway et al., 2018), primary cilia are specialised to sense fluid flow, light, odorants and signalling molecules, for example, kidney epithelial cells sense urine flow and retinal photoreceptor cells sense light (Wheway et al., 2014, Bergboer et al., 2018, Hildebrandt et al., 2011).

In vertebrates, primary cilia have an important role in coordinating the Hedgehog (Hh) signalling pathway (Hynes et al., 2014, Reiter and Leroux, 2017). Moreover, there are other signalling pathways that are coordinated by the primary cilium such as the Wnt, Notch, Hippo, GPCR, PDGF, mTOR, and TGF-beta pathways. Besides, there is a complex cross-talk between these signalling pathways (Wheway et al., 2018, Nishimura et al., 2019, Reynolds et al., 2019).

Transmembrane receptors (such as Ptch1 and Fzd) allow primary cilium to receive diverse external stimuli. Signal transduction and signalling cascades are controlled by regulatory proteins such as Smoothened (Smo), Gli effectors, Disheveled (Dvl), Inversin (Invs) and Ahi1 at diverse locations such as the ciliary pocket (depression of the plasma membrane at the base of the cilium), transition zone or the ciliary

compartment (Hynes et al., 2014, Nishimura et al., 2019, Reynolds et al., 2019, Wheway et al., 2018).

Moreover, the primary cilium has a role in regulating the timing of the cell cycle, as the assembly and disassembly of cilia are cell-cycle dependent processes (Ford et al., 2018, Nigg and Holland, 2018).

The primary cilium is an organelle with important roles in development, especially for organogenesis and embryonic development (Reiter and Singla, 2006, Valente et al., 2014, Wheway et al., 2014). Primary ciliopathies are generally autosomal recessively inherited (Hildebrandt et al., 2011). Even though most of ciliopathies are considered rare diseases (prevalence from 1:1000 to 1:150000 people), collectively, if the entire group of ciliopathies is considered as a whole, their cumulative prevalence increase up to 1:2000 people (Kempeneers and Chilvers, 2018, Huynh et al., 2020, Grochowsky and Gunay-Aygun, 2019). Of note, from now I may refer to ciliopathies and primary ciliopathies indistinctly, this is to avoid confusion and because the underlying molecular causes and pathogenesis of motile ciliopathies are out of the scope of this chapter.

1.1.2. Centrosome

The centrosome is a microtubule-based organelle, which is essential for the correct cell division of animal cells, among other functions. The core of the centrosome is formed by two centrioles, a mother centriole and a daughter centriole, which are different in terms of time in which they were formed and function (Nigg and Stearns, 2011). Centrioles, barrel-shaped structures formed by microtubules, are essential for the formation of centrosomes, cilia and flagella (Mirvis et al., 2018). In cycling cells, the centrosome coordinates the formation of the spindles poles for mitosis to proceed and acts as the major microtubule organizing centre (MTOC) (Mahjoub, 2013).

In interphase, or quiescent (non-dividing) cells, the centrosome migrates to the cell surface, so that the older (mother) centriole matures into the basal body, which attaches to the apical surface of the cell (Kobayashi and Dynlacht, 2011). The attachment of the basal body to the plasma membrane of the cell allows the nucleation of the axoneme, and thus the assembly of cilia. Of note, cilia can be observed not only during quiescence but also during G1, S and G2 phases of the cell cycle (Ford et al., 2018). Centrioles, and the centrosome, are essential cell structures

for ciliogenesis (formation of cilia) and cell division (Nigg and Stearns, 2011, Nigg and Holland, 2018, Kobayashi and Dynlacht, 2011) (Figure 1.2).

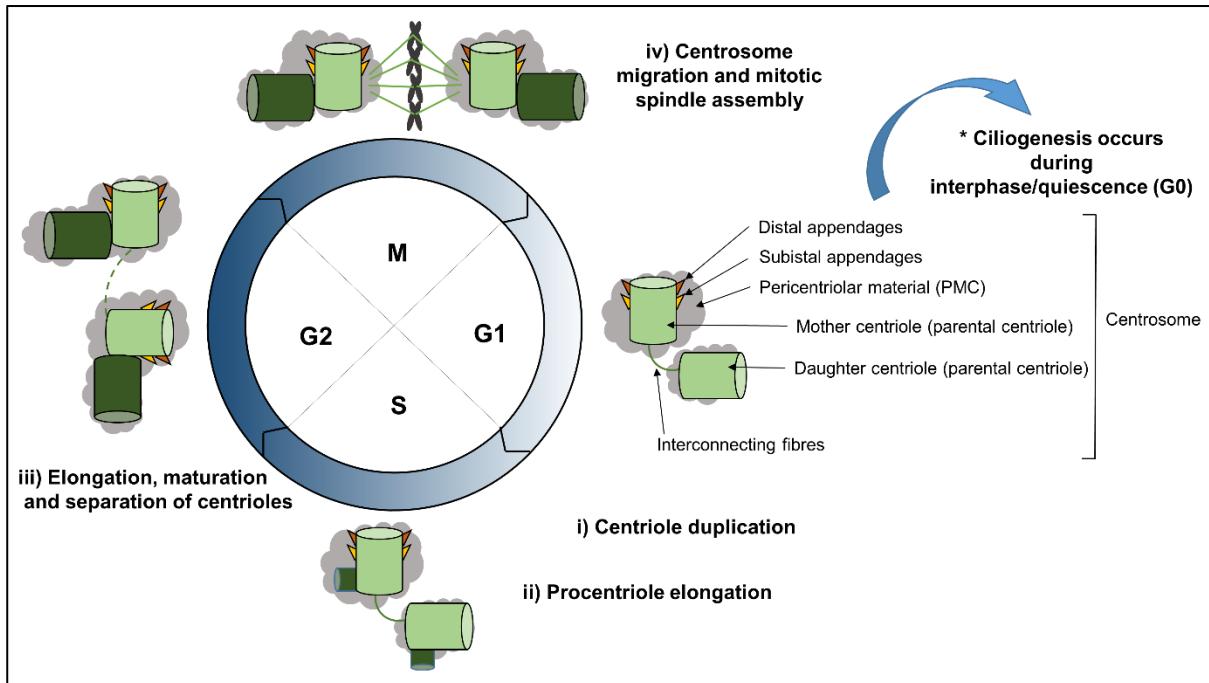


Figure 1.2. Centriole duplication cycle. Centrosome/centriole biogenesis is dependent on cell cycle (Ford et al., 2018, Nigg and Holland, 2018). Centriole duplication starts at the G1-S transition (i). At G1 the mother centriole is distinguishable from the daughter centriole (assembled in the previous cycle), given the presence of distal and subdistal appendages. In S phase, procentriole elongation occurs from the proximal end of two parental centrioles (ii). In G2 phase, procentriole elongation continues and the parental centriole that was the daughter centriole in G1, acquire distal and subdistal appendages (this can also occur during phase M) and separates from the other parental centriole (iii). At the G2-M transition, centrosomes migrate so that they can form the mitotic spindle for mitosis to proceed (iv). Ciliogenesis occurs, often in interphase, when the centrosome migrates to the plasma membrane and the mother centriole matures into the basal body and attaches to the apical surface of the cell. For more detailed information see: (Ford et al., 2018, Nigg and Holland, 2018).

1.2. Phenotypic spectrum of primary ciliopathies

Primary ciliopathies are genetically and phenotypically heterozygous; furthermore, the clinical phenotypes often overlap between ciliopathies. For instance, kidney phenotypes such as NPHP can be found on its own or in patients as additional phenotype in other ciliopathies such as SLS and JBTS syndromes (Novarino et al., 2011, Braun and Hildebrandt, 2017). Similarly, retinal degeneration can be found in multiple ciliopathies such as LCA, SLS and JBTS syndromes (Wang et al., 2018, Grochowsky and Gunay-Aygun, 2019).

Some primary ciliopathies are more severe than others. LCA, for instance, causes retinal degeneration, while other ciliopathies, like BBS, can affect multiple organs causing obesity, polydactyly and retinopathy. MKS, the most severe ciliopathy, causing embryonic lethality, affects multiple organs causing phenotypes such as occipital encephalocele, polycystic kidneys and polydactyly (Novarino et al., 2011, Consugar et al., 2007) (Figure 1.3).

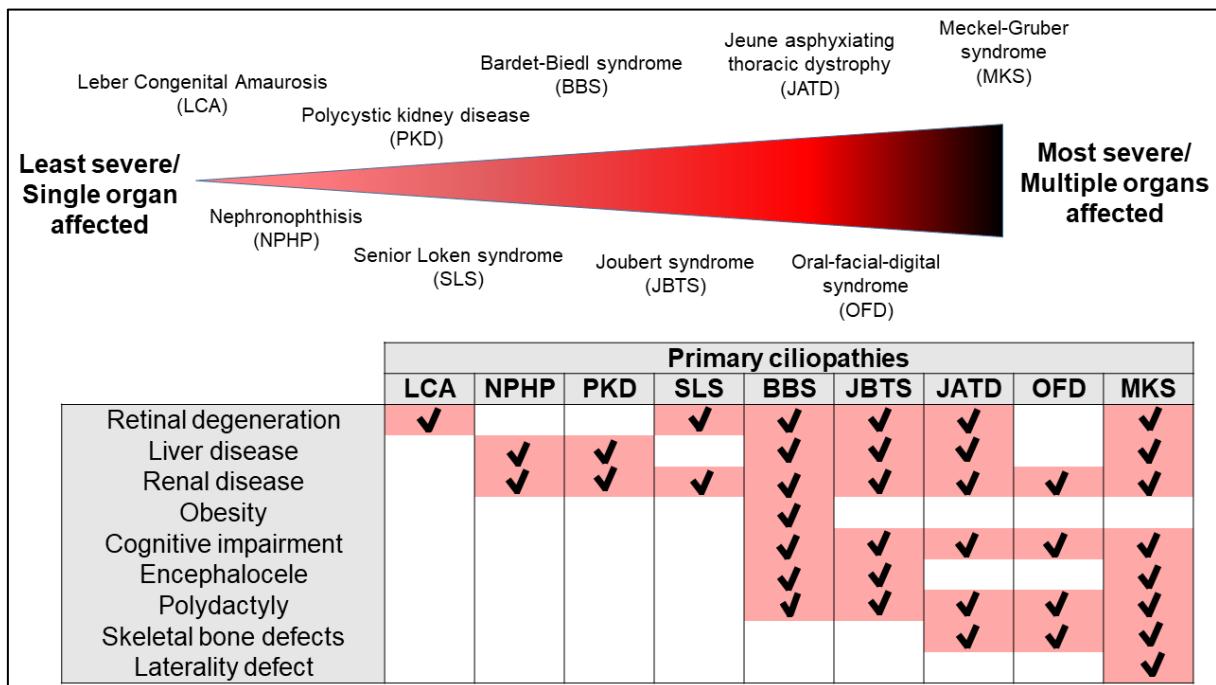


Figure 1.3. Phenotypes and syndromes associated with primary ciliopathies. For simplicity, not all described primary ciliopathies and associated phenotypes are considered. At the upper part of the figure, a diagram is shown with the relative severity of each of the nine ciliopathies considered, being MKS the most severe and LCA the least severe. At the bottom part of the image, it is shown the most relevant clinical phenotypes for each of the primary ciliopathies considered, for more detailed information see: (Mitchison and Valente, 2017, Braun and Hildebrandt, 2017, Wheway et al., 2018, Novarino et al., 2011). BBS: Bardet-Biedl syndrome; JATD: Jeune asphyxiating thoracic dystrophy; JBTS: Joubert syndrome; LCA: Leber congenital amaurosis; MKS: Meckel syndrome; NPHP: nephronophthisis; OFD: Orofaciodigital syndrome; PKD: Polycystic kidney disease; SLS: Senior Løken syndrome.

Ciliopathies are characterised by phenotypic and genetic heterogeneity and overlap. All of this makes ciliopathies difficult to diagnose and challenging for clinicians to provide an early, effective and personalised treatments (Braun and Hildebrandt, 2017, Wheway et al., 2019). Understanding the molecular basis and disease mechanisms of ciliopathies requires the study of complex organs such as brain and kidney as well as the study of cell organelles such as the cilium and the centrosome (Kobayashi and Dynlacht, 2011, Adams et al., 2012).

Apart from the phenotypic complexity and overlap of clinical phenotypes, mutations in different genes can cause the same ciliopathy (Coppieters et al., 2010, Shaheen et al., 2016). There are about 190 genes described to cause ciliopathies, many of them showing genetic pleiotropy (phenomenon in which mutations at the same gene result in different disease phenotypes) (Focsa et al., 2021, Oud et al., 2016). Furthermore, SYSCILIA consortium suggested that there are more than 300 ciliary genes and that most of them are located at the cilium and the centrosome, but not exclusively (Van Dam et al., 2013).

The deleterious effect of a mutation in a gene may or may not correlate with the impairment it causes on the function or production levels of the encoded protein and the severity of the diagnosed disease (Bachmann-Gagescu et al., 2012, Bergmann, 2017). Other factors such as gene disease modifiers (Davis et al., 2011, Ramsbottom et al., 2020) and basal exon skipping (Drivas et al., 2015, Barny et al., 2018) may affect the resulting clinical phenotypes. Furthermore, some ciliary proteins do not exclusively function at the cilium, and may be involved in other cellular processes, which differ from cilia biogenesis. Besides, some proteins associated with ciliopathies may not be functioning at the cilium (Vertii et al., 2015).

This project is focussed on JBTS, JATD and renal ciliopathies. I focussed on these primary ciliopathies because they cover most of the phenotypic and genotypic spectrum and complexity of primary ciliopathies. This selection of primary ciliopathies is suitable to study the molecular genetics of primary ciliopathies from a broad perspective. Considering these primary ciliopathies, I can extensively study two of the main characteristics of primary ciliopathies: genetic and phenotypic heterogeneity and overlap and genetic pleiotropy. The molecular basis of these characteristics of primary ciliopathies remains fully not understood (McConnachie et al., 2021, Novarino et al., 2011, Parisi, 2019).

Besides, JBTS and JATD syndromes, show multiorgan involvement and genetic pleiotropy (Parisi, 2019, Reiter and Leroux, 2017). JBTS syndrome is considered a model ciliopathy and it has been extensively studied in Sayer and Miles Laboratory at Newcastle University (Ramsbottom et al., 2020). The clinical phenotypes in Joubert syndrome are not as severe as in a Meckel syndrome, which is an embryonic lethal ciliopathy, this widens the possibilities of investigating its molecular disease mechanisms, for instance using *in vivo* mouse models. Besides, JBTS shows a specific hallmark phenotype called the “molar tooth sign” (MTS) (a cerebellar

malformation characteristic of JBTS), and this facilitates an *in silico* selection of a defined cohort of JBTS patients from databases with clinical data available (Bachmann-Gagescu et al., 2015a, Latour et al., 2020).

The broad phenotypic spectrum of other groups of primary ciliopathies, such as retinal ciliopathies, including LCA, is out of the scope of this project, as their complexity and overlapping ocular phenotypes are not restricted to primary ciliopathies (e.g. retinal defects can be found in other diseases not related with primary cilia such as Wolfram and Cohen syndromes) (Astuti et al., 2017, Nasser et al., 2020, Perea-Romero et al., 2021).

I have studied JATD because there is phenotypic and genetic overlap between JBTS and JATD syndromes. Some of the genes associated with JBTS are also associated with JATD, and some JBTS patients can show skeletal phenotypes that can also be found in patients with JATD (Lehman et al., 2010, Tuz et al., 2014). Besides, studying JATD, which is a skeletal ciliopathy (Bredrup et al., 2011, Ishida et al., 2021, De Vries et al., 2010), and JBTS, allows to investigate the genetic and phenotypic overlap between two primary ciliopathies and covers a wider phenotypic and genotypic spectrum than studying only one ciliopathy.

Furthermore, in this project renal ciliopathies are extensively investigated because kidney phenotypes can be found in JBTS, JATD and other primary ciliopathies. Moreover, kidney phenotypes offer a wider therapeutic window than cerebellar or skeletal phenotypes, as the kidneys are targetable and treatable via some treatments such as Tolvaptan (Garcia et al., 2022, Devlin et al., 2023). In the last few years Sayer and Miles Laboratory at Newcastle University has investigated the use of ASO morpholinos to treat kidney cysts in a JBTS mouse model and in *in vitro* cell models using patient primary cells (Molinari et al., 2019, Ramsbottom et al., 2018, Srivastava et al., 2017b).

In summary, I will focus on JBTS, JATD and renal primary ciliopathies, to investigate the genetic and phenotypic heterogeneity and overlap of primary ciliopathies and the underlying molecular disease mechanisms leading to these representative primary ciliopathies.

1.2.1. Joubert syndrome

JBTS is a ciliopathy characterised by a cerebellar and brainstem malformation, described as the MTS which is the hallmark for the diagnosis of this disease (Romani et al., 2013, Maria et al., 1997, Parisi, 2019, Radha Rama Devi et al., 2020) (Figure 1.4).

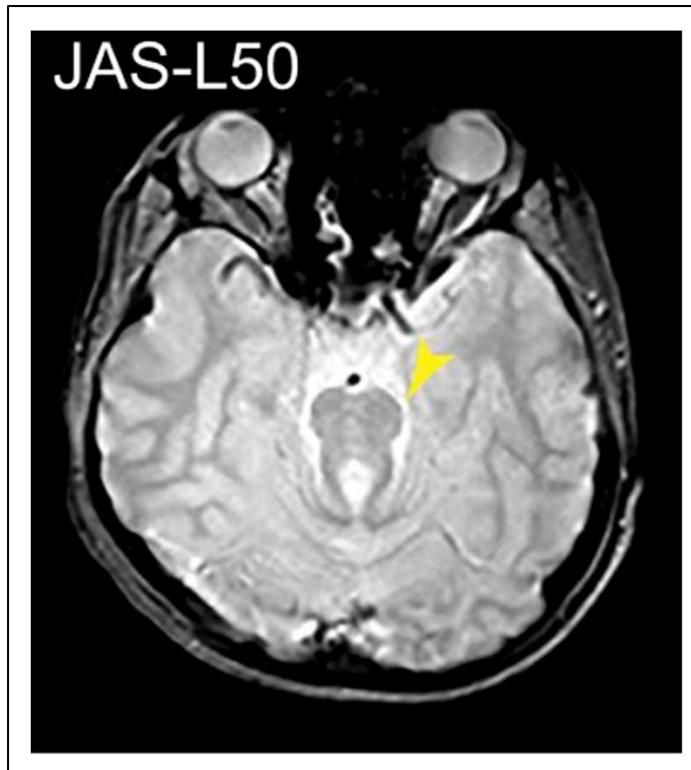


Figure 1.4. Representative example of the “molar tooth sign” (MTS) distinctive brain malformation, which is characteristic of Joubert syndrome (JBTS). Image obtained from Latour et al., 2020 (Latour et al., 2020). The MTS (indicated by an arrow) was visualised using axial T2-weighted magnetic resonance imaging (MRI). The MRI imaging corresponds to a JBTS patient (JAS-L50) presenting compound mutations (c.5023C>T; p.(Arg1675Ter) and c.1112C>A; p.(Ala371Asp)) in a gene associated with JBTS: *TOGARAM1* (Latour et al., 2020).

The incidence of JBTS ranges from 1:80000 to 1:100000 cases (Oud et al., 2016, Romani et al., 2013). JBTS patients are described to have the MTS cerebellar phenotype, develop hypotonia in infancy with later development of ataxia and developmental delay. Moreover, JBTS may involve extra-neurological phenotypes and be grouped as Joubert syndrome and related disorders (JSRD), although currently the terms JSRD and JBTS are often used indistinctly (Parisi, 2019, Parisi et al., 2007). JSRD consists in a group of multisystem diseases showing cerebellar hypoplasia, retinal dystrophy and NPHP, all this together is considered as a cerebello-oculo-renal phenotype (Adams et al., 2012, Romani et al., 2013). Besides,

some patients can show JSRD with hepatic disease, oral-facial-digital features (JBTS-OFD) or Jeune asphyxiating thoracic dystrophy (JBTS-JATD) (Lehman et al., 2010, Parisi, 2009).

JBTS is a model ciliopathy and it has been intensively studied in the last decade (Hynes et al., 2014, Parisi, 2019). There about 40 genes described to be associated with JBTS (Parisi, 2019, Gana et al., 2022, Bachmann-Gagescu et al., 2020). These genes do not encode exclusively for proteins found exclusively at the cilium, some of them may encode centrosomal proteins. For instance, *CEP290* and *CC2D2A* encode for proteins functioning at the transition zone (Cheng et al., 2012, Ojeda Naharros et al., 2017) and *CEP120* encode for a protein functioning at the centrosome (Meka et al., 2022, Wu et al., 2014).

There is an increasing number of functional studies investigating the consequences of specific mutations in different ciliary genes. Besides as explained previously, some of these genes are not exclusively associated to JBTS (Coppieters et al., 2010). Different biallelic mutations in the same gene associated with JBTS may cause different ciliopathies (Figure 1.5) (Coppieters et al., 2010, Hua and Ferland, 2018, Bachmann-Gagescu et al., 2015a).

For example, biallelic mutations in *CEP290* are associated with various ciliopathies such as LCA, JBTS and MKS (Travaglini et al., 2009). Similarly, *CEP120* and *CC2D2A*, which encode proteins functioning at the centrosome and transition zone respectively, are both associated with JBTS as well as skeletal ciliopathies (Roosing et al., 2016, Shaheen et al., 2015, Bachmann-Gagescu et al., 2012, Mougou-Zerelli et al., 2009).

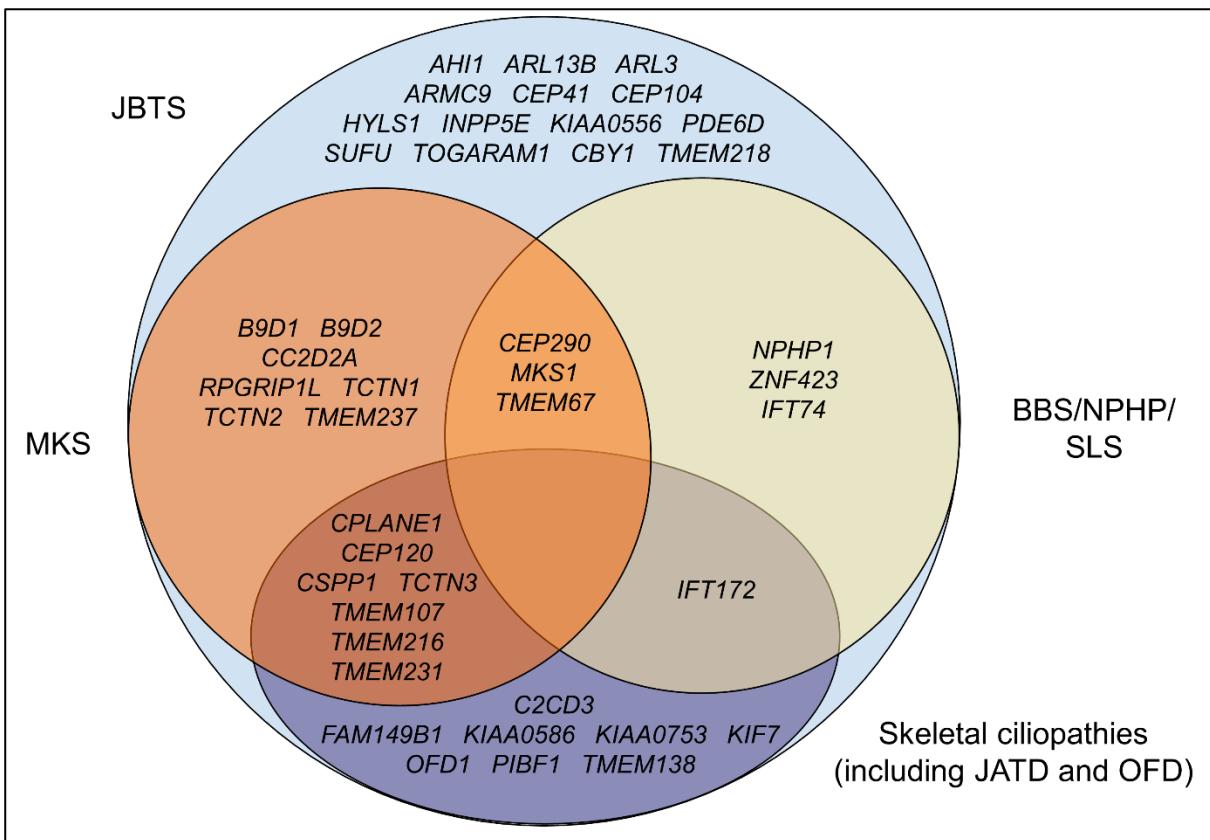


Figure 1.5. Genetic heterogeneity and genetic pleiotropy of the genes associated with Joubert syndrome (JBTS). This diagram shows how JBTS genes can be found to be associated with other ciliopathies as well (Mitchison and Valente, 2017). For simplicity, I did not show other distinct phenotypes such as hydrocephalus, LCA or RCD. Of note, for the genes: *C2CD3* (Thauvin-Robinet et al., 2014), *CBY1* (Epting et al., 2020) and *HYLS1* (Oka et al., 2016), included in this diagram, biallelic variants have been described in more than one family with JBTS but are not included in the OMIM Joubert syndrome phenotypic series (PS213300, status 11/2022). BBS, Bardet-Biedl syndrome; JATD, Jeune asphyxiating thoracic dystrophy; JBTS, Joubert syndrome; LCA, Leber congenital amaurosis; MKS, Meckel syndrome; NPHP, nephronophthisis; OFD, oro-facial-digital syndrome; RCD, Rod-cone dystrophy; SLS: Senior Løken syndrome. Image updated from (Barroso-Gil et al., 2021a).

1.2.2. Jeune syndrome

Jeune asphyxiating thoracic dystrophy (JATD) is a rare autosomal recessive skeletal ciliopathy with multi-organ involvement (Zhang et al., 2018).

JATD is characterised by a small, narrow thorax, and this often causes diverse respiratory complications in the neonatal period and infancy. After infancy, patients may develop renal cyst disease. Like in JBTS, JATD phenotypes may affect other organs such as the pancreas, kidney, liver and eye (De Vries et al., 2010, Zhang et al., 2018).

The severity of JATD is variable, some patients may reach adulthood, however often the clinical phenotypes of JATD cause death in the first week of life, making JATD a

severe ciliopathy (Stembalska et al., 2022, Shaheen et al., 2015, Duran et al., 2016). Moreover, some patients have been described to have a JBTS-JATD ciliopathy with small thorax and respiratory complications (Lehman et al., 2010, Malicdan et al., 2015).

The spectrum of skeletal ciliopathies, including short-rib polydactyly (SRP)/JATD skeletal syndromes is broad with there is extensive phenotypic and genetic overlap (Zhang et al., 2018, Cavalcanti et al., 2011, Duran et al., 2016).

1.2.3. Renal ciliopathies

The genotypic and phenotypic spectrum of renal ciliopathies is heterogeneous (Barroso-Gil et al., 2021b, Devlin and Sayer, 2019). The severity of the clinical phenotypes also varies from very mild symptoms to end-stage kidney disease (ESKD), and may also include extra-renal phenotypes affecting liver, brain, eye, skeletal system and cardiovascular system (Novarino et al., 2011, Reiter and Leroux, 2017). Clinical liver and kidney phenotypes are often found together and the phenotypic and genotypic spectrum each of the renal ciliopathies that will be explained below often overlaps (Cornec-Le Gall et al., 2019, Besse et al., 2019).

There are several renal ciliopathies: autosomal dominant polycystic kidney disease (ADPKD), autosomal recessive polycystic kidney disease (ARPKD) and nephronophthisis (NPHP). There is genetic overlap between these renal ciliopathies as well as clinical overlap, in terms of age of disease onset, and development of end-stage kidney disease (ESKD) (Barroso-Gil et al., 2021b) (Figure 1.6).

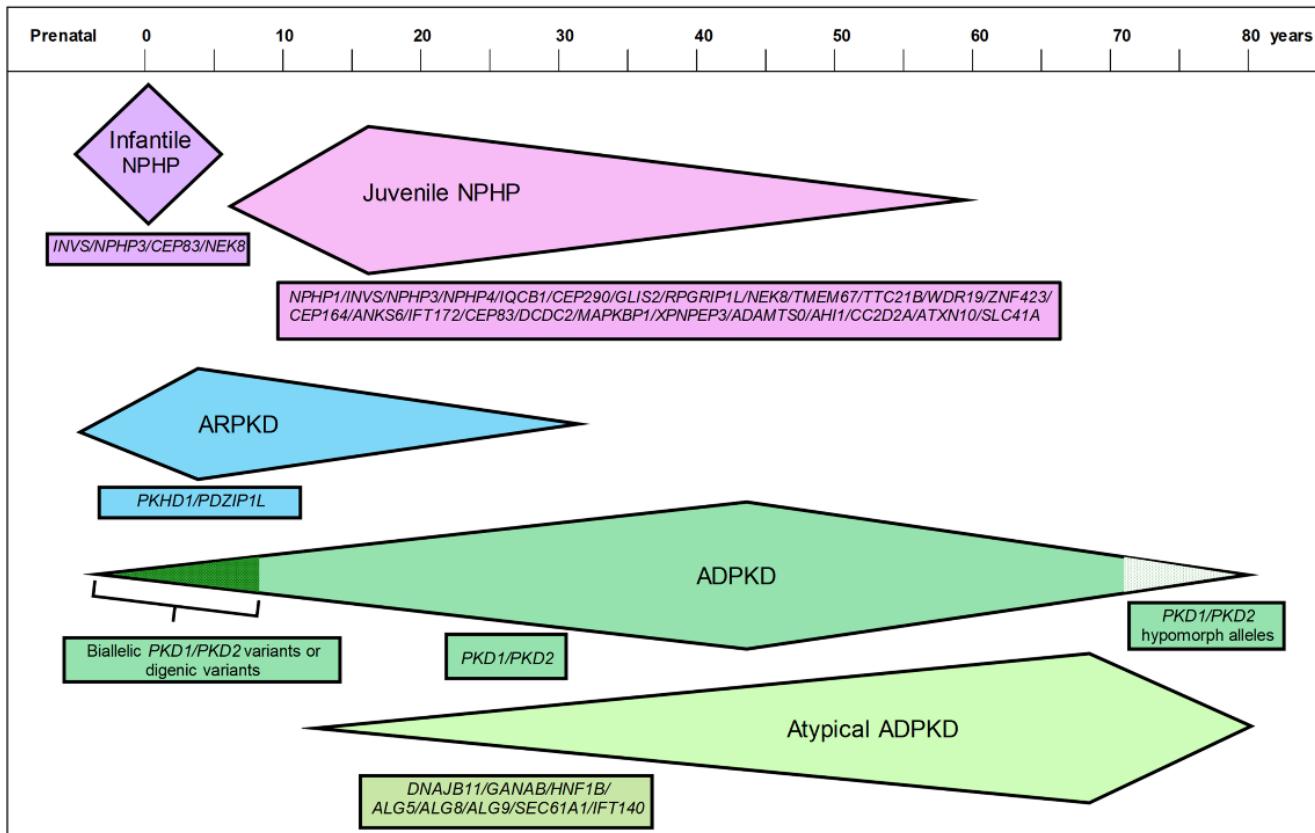


Figure 1.6. Variability and overlap in the age of disease onset in patients with renal ciliopathies. Of note, biallelic mutations in *NPHP1* and *PKD1* are the most common cause of Juvenile NPHP and ARPKD respectively (Bergmann, 2017, McConnachie et al., 2021). Heterozygous mutations in *PKD1* is the most common cause of ADPKD. ADPKD patients with truncating mutations in *PKD1* develop ESKD earlier in life than non-truncating *PKD1* mutations or *PKD2* mutations (Cornec-Le Gall et al., 2019). The renal phenotype in patients with atypical ADPKD is often milder than the one observed in ADPKD caused by *PKD1/PKD2* variants (Besse et al., 2019, Cornec-Le Gall et al., 2018a). ADPKD, autosomal dominant polycystic kidney disease; ADPLD, autosomal dominant polycystic liver disease; ADTKD, autosomal dominant tubulo-interstitial kidney disease; ARPKD, autosomal recessive polycystic kidney disease; NPHP, nephronophthisis. Image modified from (Barroso-Gil et al., 2021b).

Polycystic kidney disease can have an autosomal recessive or an autosomal dominant inheritance pattern. The most common renal ciliopathy is ADPKD, which is also the most common inherited renal ciliopathy (Cornec-Le Gall et al., 2019, McConnachie et al., 2021). It is estimated that its prevalence is 1 in 1000 people. ADPKD is a monoallelic dominant inherited disease and its clinical phenotype typically consists in the development of large cysts and kidney enlargement (Figure 1.7). The development of renal cysts leads to a decrease of kidney function. ADPKD is also a frequent cause of ESKD (Cornec-Le Gall et al., 2018b, Huynh et al., 2020, Porath et al., 2016).

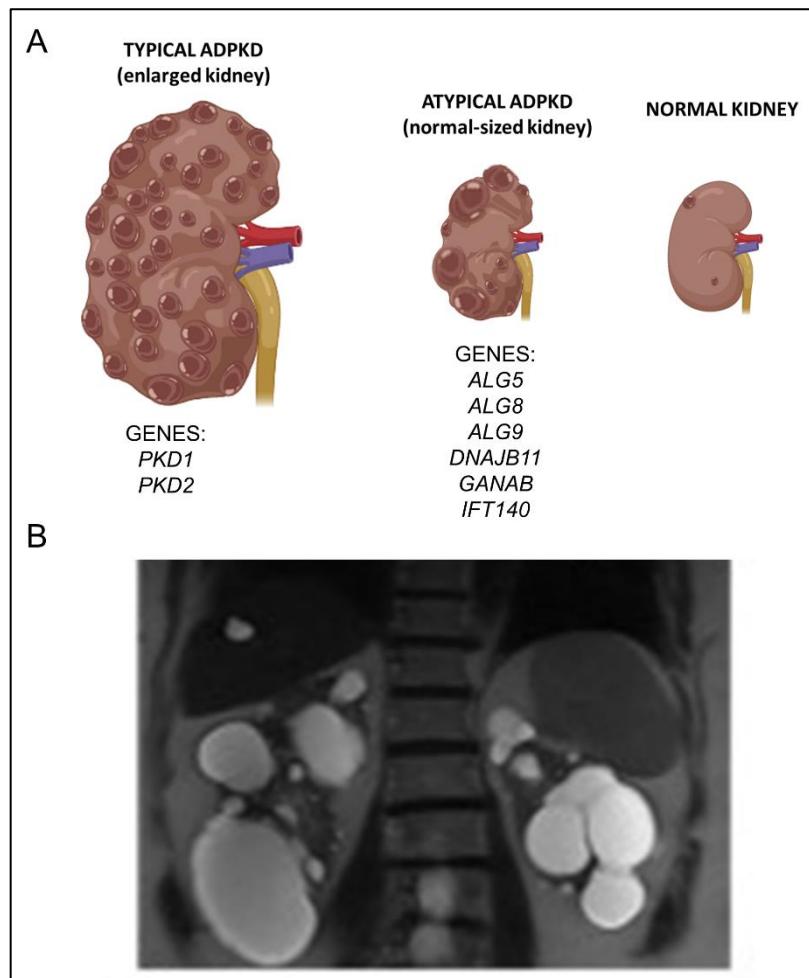


Figure 1.7. Representation of cystic kidneys in ADPKD. A: ADPKD is characterized by the presence of numerous and often large renal cysts, mutations in *PKD1* and *PKD2* genes are responsible for about 90% of the cases, this is shown as typical ADPKD. Some ADPKD patients, who do not have a pathogenic variant in *PKD1* or *PKD2*, show an atypical phenotype of ADPKD. B: Example imaging of large cysts observed in both kidneys and one cyst in liver. This abdominal magnetic resonance imaging (MRI) was obtained from Senum *et al.*, 2022 (Senum *et al.*, 2022). The MRI corresponds to a patient (P1320 II-1) presenting a pathogenic heterozygous loss of function (LoF) *IFT140* mutation and a heterozygous missense *PKD1* variant described as a variant of uncertain significance (VUS) (Senum *et al.*, 2022).

The clinical phenotypes of ARPKD consist in renal cysts and enlargement of the kidney. Unlike ADPKD, in ARPKD renal phenotypes usually occur in early childhood. By the age of 10, 50% of the ARPKD patients develop ESKD (Molinari *et al.*, 2020, Adamiok-Ostrowska and Piekiełko-Witkowska, 2020). ARPKD is not as a common as ADPKD. It is estimated that ARPKD has an incidence of 1:20000 live births (Devuyst *et al.*, 2014), however, this estimate varies between geographic regions (Bergmann *et al.*, 2018, Al Alawi *et al.*, 2020, Bergmann, 2017). Liver phenotypes can also be observed in ARPKD patients, such as liver cysts and ductal plate malformation (Bergmann *et al.*, 2018).

In contrast to polycystic kidney disease, the clinical phenotypes of NPHP are characterised by a disintegration and irregular thickening of the tubular basement membrane, tubulointerstitial fibrosis and tubular atrophy. Kidney cysts are not a hallmark of this renal ciliopathy, in contrast to ADPKD and ARPKD. Kidneys in NPHP patients usually are fibrotic and normal-sized (Srivastava et al., 2017a, Braun and Hildebrandt, 2017).

About 10-20% of patients with NPHP may also have extra-renal phenotypes affecting brain, retina, liver and skeletal system (Srivastava et al., 2017a). Other ciliopathies may include NPHP in their disease phenotypic spectrum, this is, NPHP phenotypes can be found patients diagnosed with other ciliopathies such as SLS and JBTS (Parisi, 2019, Valente et al., 2006, Chaki et al., 2012).

NPHP is not as common as ADPKD, it is estimated that the incidence of NPHP ranges between 1:50,000 to 1:1,000,000 people (Srivastava et al., 2017a). NPHP has been often divided in three clinical subtypes of NPHP: infantile NPHP, juvenile NPHP and adolescent/adult NPHP, being juvenile NPHP the most common subtype of NPHP. Patients with juvenile and adolescent/adult NPHP usually develop ESKD in and after the second decade of life respectively. Some patients with adult NPHP may develop ESKD after the fifth decade of life (Snoek et al., 2018). In contrast, patients with infantile NPHP usually develop ESKD before 5 years of age (Stokman et al., 2021).

1.3. Molecular genetics of primary ciliopathies

There are more than 190 genes encoding for proteins that are predominantly ciliary-localized which are described to be mutated in ciliopathies (Van Dam et al., 2013). Genes associated with ciliopathies, do not necessarily have to encode for proteins located at the cilium. There are multiple genes associated to primary ciliopathies that code for proteins functioning at the centrioles/centrosome (centrosomal proteins) (Braun and Hildebrandt, 2017, Shaheen et al., 2016).

For example, genes such as: *CEP120* (Roosing et al., 2016), *AHI1* (Adams et al., 2012), *CEP104* (Frikstad et al., 2019), *CSPP1* (Frikstad et al., 2019) and *KIAA0586* (Malicdan et al., 2015) are associated with JBTS and encode for proteins functioning (often not exclusively) at the centrosome. Besides, there is a growing interest in the non-ciliary function of cilia proteins, and the interaction of these functions, apart from

cilia assembly/signalling, such as spindle orientation, cell polarity or DNA damage (Vertii et al., 2015, Lovera and Lüders, 2021).

Of note, as an example of genetic heterogeneity and pleiotropy of genes associated with JBTS: mutations in *CEP120* (Shaheen et al., 2015), *CSPP1* (Shaheen et al., 2014) and *KIAA0586* (Malicdan et al., 2015) have also been associated with skeletal ciliopathies. Particularly, biallelic mutations in *CEP120* (Shaheen et al., 2015) and *CSPP1* (Shaheen et al., 2014) have also been associated with MKS.

1.3.1. Molecular genetics of Joubert syndrome

JBTS is a genetically heterogeneous ciliopathy. There are 40 genes described to be associated with JBTS (OMIM Joubert syndrome phenotypic series, PS213300) and JBTS gene candidates according to the Online Mendelian Inheritance In Man (OMIM) database (Hamosh et al., 2000) and the recent literature (Parisi, 2019, Gana et al., 2022, Bachmann-Gagescu et al., 2020) (Table 1.2). The proteins encoded by these genes cover a wide diverse spectrum of functions and locations, most of them related with cilia and centrosome biogenesis (Vertii et al., 2015, Braun and Hildebrandt, 2017, Shaheen et al., 2016).

1.3.2. Molecular genetics of Jeune syndrome

There are over 30 genes described to be associated SRP/JATD syndromes and (OMIM Short-rib thoracic dysplasia phenotypic series, PS208500) and SRP/JATD gene candidates according to OMIM (Hamosh et al., 2000) and the recent literature (Shaheen et al., 2015, Braun and Hildebrandt, 2017, Focsa et al., 2021, Zhang et al., 2018). Some of these genes are: *IFT80* (Beales et al., 2007), *CEP120* (Shaheen et al., 2015), *KIAA0586* (Malicdan et al., 2015), *DYNC2H1* (Emiralioglu et al., 2018) or *TTC21B* (Davis et al., 2011). Some of these genes can cause clinical phenotype consisting SRP/JATD with kidney and/or liver disease, such as *IFT80* (Beales et al., 2007), *DYNC2H1* (Emiralioglu et al., 2018), *KIAA0586* (Malicdan et al., 2015) or *TTC21B* (Davis et al., 2011). Some of them, such as *CEP120* (Roosing et al., 2016) and *KIAA0586* (Malicdan et al., 2015) are also described to be associated with Joubert syndrome (Figure 1.8).

Gene	Ensembl Gene ID	Phenotype in OMIM phenotypic series: Joubert syndrome - PS213300	Locus MIM number
<i>AHI1</i>	ENSG00000135541	JBTS 3	608894
<i>ARL13B</i>	ENSG00000169379	JBTS 8	608922
<i>ARL3</i>	ENSG00000138175	JBTS 35	604695
<i>ARMC9</i>	ENSG00000135931	JBTS 30	617612
<i>B9D1</i>	ENSG00000108641	JBTS 27	614144
<i>B9D2</i>	ENSG00000123810	JBTS 34	611951
<i>C2CD3</i>	ENSG00000168014	Not in JBTS phenotypic series	615944
<i>C5orf42</i> (also known as <i>CPLANE1</i>)	ENSG00000197603	JBTS 17	614571
<i>CC2D2A</i>	ENSG00000048342	JBTS 9	612013
<i>CELSR2</i>	ENSG00000143126	Not in JBTS phenotypic series	604265
<i>CEP41</i>	ENSG00000106477	JBTS 15	610523
<i>CEP104</i>	ENSG00000116198	JBTS 25	616690
<i>CEP120</i>	ENSG00000168944	JBTS 31	613446
<i>CEP164</i>	ENSG00000110274	Not in JBTS phenotypic series	614848
<i>CEP170</i>	ENSG00000143702	Not in JBTS phenotypic series	613023
<i>CEP290</i>	ENSG00000198707	JBTS 5	610142
<i>CSPP1</i>	ENSG00000104218	JBTS 21	611654
<i>FAM149B1</i>	ENSG00000138286	JBTS 36	618413
<i>HYLS1</i>	ENSG00000198331	Not in JBTS phenotypic series	610693
<i>IFT74</i>	ENSG00000096872	JBTS 40	608040
<i>IFT80</i>	ENSG00000068885	Not in JBTS phenotypic series	611177
<i>IFT172</i>	ENSG00000138002	Not in JBTS phenotypic series	607386
<i>INPP5E</i>	ENSG00000148384	JBTS 1	613037
<i>KIAA0556</i> (also known as <i>KATNIP</i>)	ENSG00000047578	JBTS 26	616650
<i>KIAA0586</i>	ENSG00000100578	JBTS 23	610178
<i>KIAA0753</i>	ENSG00000198920	?JBTS 38	617112
<i>KIF7</i>	ENSG00000166813	JBTS 12	611254
<i>LAMA1</i>	ENSG00000101680	Not in JBTS phenotypic series	150320
<i>MKS1</i>	ENSG00000011143	JBTS 28	609883
<i>NME3</i>	ENSG00000103024	Not in JBTS phenotypic series	601817
<i>NPHP1</i>	ENSG00000144061	JBTS 4	607100
<i>OFD1</i>	ENSG00000046651	JBTS 10	300170
<i>PDE6D</i>	ENSG00000156973	JBTS 22	602676
<i>PIBF1</i>	ENSG00000083535	JBTS 33	607532
<i>POC1B</i>	ENSG00000139323	Not in JBTS phenotypic series	614784

<i>RPGRIP1L</i>	ENSG00000103494	JBTS 7	610937
<i>SUFU</i>	ENSG00000107882	JBTS 32	607035
<i>TCTN1</i> (also known as <i>TECT1</i>)	ENSG00000204852	JBTS 13	609863
<i>TCTN2</i>	ENSG00000168778	JBTS 24	613846
<i>TCTN3</i>	ENSG00000119977	JBTS 18	613847
<i>TMEM67</i>	ENSG00000164953	JBTS 6	609884
<i>TMEM107</i>	ENSG00000179029	?JBTS 29	616183
<i>TMEM138</i>	ENSG00000149483	JBTS 16	614459
<i>TMEM216</i>	ENSG00000187049	JBTS 2	613277
<i>TMEM218</i>	ENSG00000150433	JBTS 39	619285
<i>TMEM230</i>	ENSG00000089063	Not in JBTS phenotypic series	617019
<i>TMEM231</i>	ENSG00000205084	JBTS 20	614949
<i>TMEM232</i>	ENSG00000186952	Not in JBTS phenotypic series	620143
<i>TMEM237</i>	ENSG00000155755	JBTS 14	614423
<i>TOGARAM1</i>	ENSG00000198718	JBTS 37	617618
<i>TTC21B</i>	ENSG00000123607	Not in JBTS phenotypic series	612014
<i>WDR72</i>	ENSG00000166415	Not in JBTS phenotypic series	613214
<i>ZNF423</i>	ENSG00000102935	JBTS 19	604557

Table 1.2. List of Joubert syndrome (JBTS) genes and JBTS candidate genes. List of JBTS genes and JBTS candidate genes has been created using OMIM database (accessed on 12/2022) and the available literature (Bachmann-Gagescu et al., 2020, Parisi, 2019). Of note, *TMEM218* and *IFT74* were not in this list at the time of some of this analysis (2019-2020) as they have been recently associated with JBTS. Of note, *TMEM218* and JBTS candidate genes: *CELSR2*, *CEP170*, *NME3*, *LAMA1*, *TMEM230*, *TMEM232*, *POC1B*, *WDR72* are not in the gold standard ciliary gene list (SYSCILIA) (Van Dam et al., 2013).

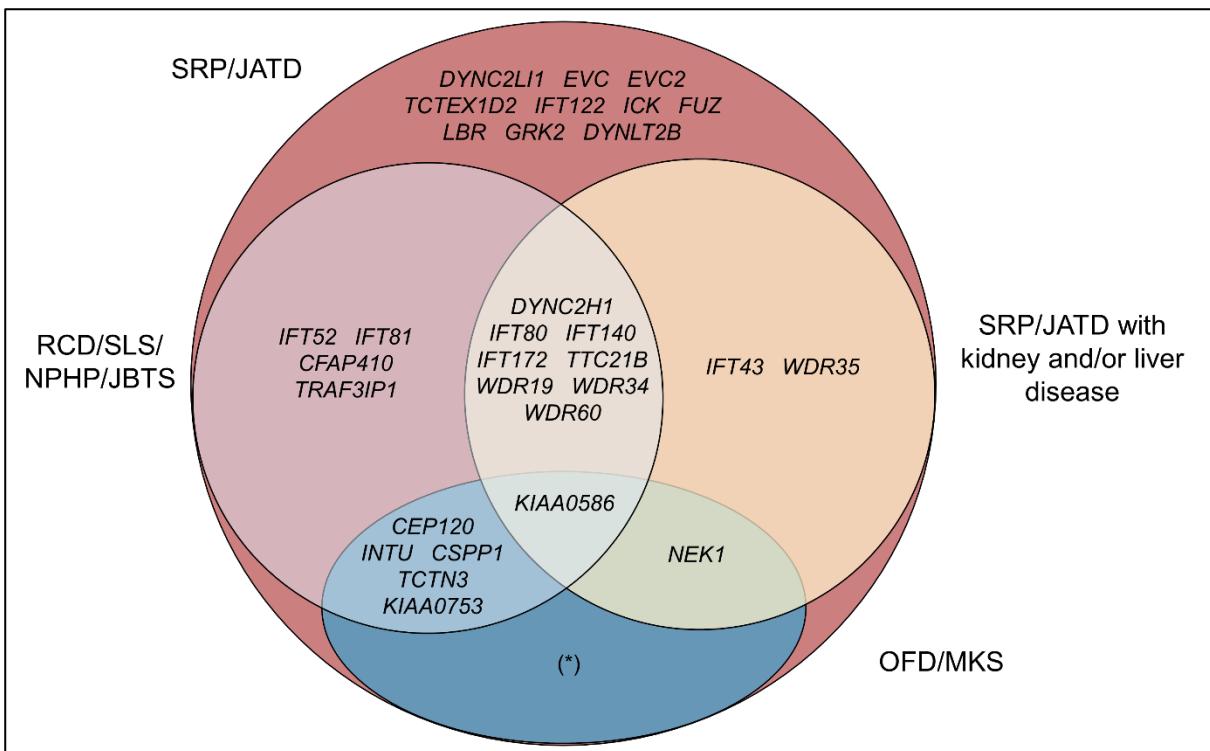


Figure 1.8. Genetic heterogeneity and genetic pleiotropy of the genes associated with short-rib polydactyly (SRP)/Jeune asphyxiating thoracic dystrophy (JATD). The SRP/JATD phenotypic spectrum includes: SRP types I-IV, JATD, EVC syndrome and Sensebrenner syndrome. Other distinct phenotypes such as BBS, male in fertility (spermatogenic dysfunction) or Pelger-Huët anomaly have not been considered for the simplicity of the diagram. (*) For simplicity, I did not include all the genes involved in OFD and MKS. Of note, the following genes have been described to be associated with the SRP/JATD phenotypic spectrum but are not included in the OMIM in Short-rib thoracic dysplasia phenotypic series (PS208500, status 11/2022): *CSPP1* (Tuz et al., 2014), *IFT122* (Walczak-Sztulpa et al., 2010, Silveira et al., 2017), *CILK1* (Paige Taylor et al., 2016), *FUZ* (Zhang et al., 2018), *LBR* (Zhang et al., 2018), *TRAF3IP1* (Zhang et al., 2018), *TCTN3* (Schmidts, 2014), *GRK2* (Bosakova et al., 2020). BBS, Bardet-Biedl syndrome; EVC, Ellis-van Creveld; JATD, Jeune asphyxiating thoracic dystrophy; JBTS, Joubert syndrome; LCA, Leber congenital amaurosis; MKS, Meckel syndrome; NPHP, nephronophthisis; OFD, oro-facial-digital syndrome; RCD, Rod-cone dystrophy; SLS: Senior Løken syndrome; SRP, short-rib polydactyly.

1.3.3. Molecular genetics of renal ciliopathies

ADPKD is age dependent, it is often only found in adult patients (Lanktree et al., 2018). It is also frequently diagnosed in children due to improved screening (Lanktree et al., 2018, Bergmann et al., 2018).

Nevertheless, several factors are involved in the severity of ADPKD, such as the pathogenicity of the mutation of the disease phenotype and the gene in which that mutation is located. Mutations in *PKD1* and *PKD2* genes are responsible for about 90% of the ADPKD cases. Mutations in *PKD1* are the most common cause of

ADPKD, with about 75% of the ADPKD patients presenting a disease-causing mutation in *PKD1* (Cornec-Le Gall et al., 2019, Lanktree et al., 2021).

In the last few years other genes such as: *DNAJB11* (Cornec-Le Gall et al., 2018a, Huynh et al., 2020), *GANAB* (Porath et al., 2016, Besse et al., 2017), *ALG5* (Lemoine et al., 2022) *ALG8* (Besse et al., 2017), *ALG9* (Besse et al., 2019) and *PRKCSH* (Li et al., 2003) have been associated with the autosomal dominant polycystic kidney and liver diseases, including the ADPKD phenotypic spectrum (Figures 1.7 and 1.9). A list of genes associated with PKD can be found in Table 1.3.

Apart from *PKD1* and *PKD2* there are other 7 genes that have been associated with the ADPKD phenotypic spectrum, two of these genes (*IFT140* and *ALG5*) have just been described to be associated with the ADPKD phenotypic spectrum in the last year (Senum et al., 2022, Lemoine et al., 2022) leading to an atypical ADPKD phenotype. In these patients with atypical ADPKD the kidney phenotype is usually milder and without kidney enlargement, but the severity of the liver phenotype is variable (Senum et al., 2022, Lemoine et al., 2022) (Figures 1.7 and 9). Some of the genes associated with atypical ADPKD, such as *ALG5*, *ALG8*, *ALG9*, *GANAB* and *DNAJB11*, code for proteins functioning in the endoplasmic reticulum (ER) biogenesis pathway and are also associated with ADPLD and/or ADTKD (Lemoine et al., 2022, Besse et al., 2017).

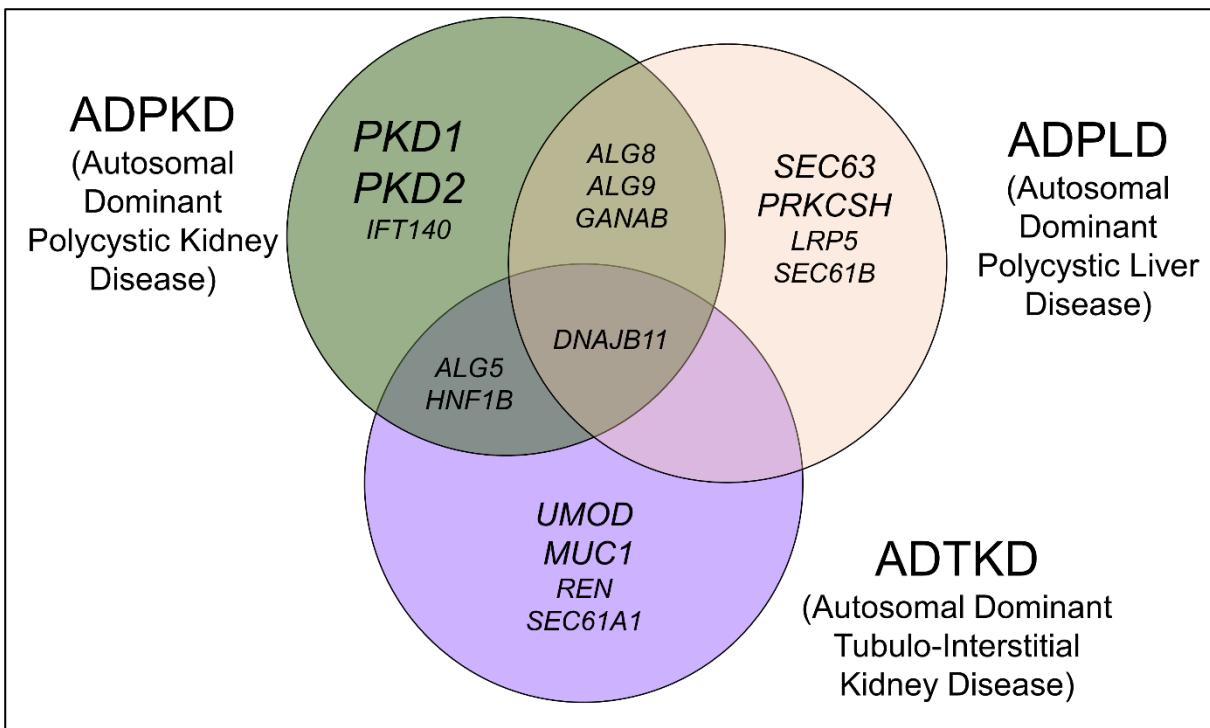


Figure 1.9. Genetic heterogeneity and genetic pleiotropy in genes associated with ADPKD, ADPLD and ADTKD. There is phenotypic and genetic overlap between these diseases leading to a complex and heterogeneous phenotypic spectrum (Cornec-Le Gall et al., 2018b, Besse et al., 2017). For simplicity, the association of *PKD1* with ARPKD and ADPKD is not shown in this diagram (Besse et al., 2017, Bergmann, 2017). Besides, *LRP5* and *SEC61B* can be suggested as candidate genes to be associated with ADPKD (Korvala et al., 2012, Cnossen et al., 2014, Besse et al., 2017). Renal cystic phenotypes have been reported in two ADTKD families presenting heterozygous mutations in *SEC61A1* (Bolar et al., 2016). ADPKD, autosomal dominant polycystic kidney disease; ADPLD, autosomal dominant polycystic liver disease; ADTKD, autosomal dominant tubulo-interstitial kidney disease; ARPKD, autosomal recessive polycystic kidney disease.

Gene	Protein product	Known function	Subcellular localization	Effect on PC1–PC2 complex	Disease phenotype	References
<i>PKD1</i>	Polycystin-1	Transmembrane glycoprotein, forms a complex with PC2	Ciliary membrane	—	ADPKD (1)	(Cai et al., 2014, McConnachie et al., 2021)
<i>PKD2</i>	Polycystin-2	Transmembrane glycoprotein, forms a complex with PC1	Ciliary membrane	—	ADPKD	(Cai et al., 2014, McConnachie et al., 2021)
<i>DNAJB11</i>	DnaJ heat shock protein family member B11	Cofactor of the ER heat shock chaperon BiP, which regulates protein folding and trafficking	ER	PC1 maturation and localization	ADTKD- <i>DNAJB11</i>	(Cornec-Le Gall et al., 2018a, Huynh et al., 2020)
<i>GANAB</i>	Glucosidase II alpha subunit	Catalytic subunit of glucosidase II that has a role in the hydrolysis of glucose residues (a process involved in correct folding, maturation and trafficking of glycoproteins)	ER	PC1 maturation. PC1 and PC2 localization	ADPKD/ADPLD	(Besse et al., 2017, Porath et al., 2016, Xu and Ng, 2015)
<i>IFT140</i>	Intraflagellar Transport Protein 140	Core protein of the ciliary IFT-A complex, involved in retrograde ciliary transport	Primary cilium	NA	ADPKD (2)	(Senum et al., 2022)
<i>ALG5</i>	ALG5 dolichyl-phosphate beta-glucosyltransferase	Participates in the synthesis of the Dol-P-Glc donor substrate	ER	PC1 maturation. PC1 and PC2 localization	ADPKD	(Lemoine et al., 2022)
<i>ALG9</i>	ALG9 Alpha-1,2-mannosyltransferase	Addition of the seventh and ninth mannose molecules to the assembling <i>N</i> -glycan precursors	ER	PC1 maturation	ADPKD/ADPLD (3)	(Besse et al., 2019)
<i>ALG8</i>	ALG8 Alpha-1,3-glucosyltransferase	Addition of the second glucose residue for <i>N</i> -linked-glycosylation (a process involved in correct folding, maturation and trafficking of glycoproteins)	ER	PC1 maturation and localization	ADPKD/ADPLD (3)	(Besse et al., 2017, Apple et al., 2022, Mantovani et al., 2020)
<i>PRKCSH</i>	Glucosidase II beta subunit	Non-catalytic subunit of glucosidase II that has a role in the hydrolysis of glucose residues (involved in correct folding, maturation and trafficking of glycoproteins)	ER	Dosage of functional PC1–PC2 complex	ADPLD	(Fedele et al., 2011, Li et al., 2003, Xu and Ng, 2015, Mallawaarachchi et al., 2021)
<i>SEC63</i>	SEC63 homolog, protein translocation regulator	Folding and translocation of glycoproteins, associates with SEC61 and SEC62	ER	Dosage of functional PC1–PC2 complex	ADPLD	(Perugorria and Banales, 2017, Fedele et al., 2011, Besse et al., 2017)

<i>SEC61B</i>	SEC61 beta subunit	Translocation of polypeptides into the ER, forms part of the translocation pore SEC61	ER	PC1 maturation	ADPLD	(Besse et al., 2017, Perugorria and Banales, 2017)
<i>LRP5</i>	Low density lipoprotein receptor-related protein 5	Transmembrane receptor associated with WNT signalling	Plasma membrane	NA	ADPLD (4)	(Korvala et al., 2012, Cnossen et al., 2014, Wheway et al., 2018)
<i>PKHD1</i>	Fibrocystin/polyductin	Transmembrane protein involved in cell–cell and cell–matrix interactions	Plasma membrane, primary cilium and basal body	PC2 expression (5)	ARPKD (5)	(Besse et al., 2017, Perugorria and Banales, 2017, Kim et al., 2008)
<i>DZIP1L</i>	DAZ interacting protein 1-like	Maintenance of the transition zone diffusion barrier, interaction with Septin2	Centrioles/TZ	PC1 and PC2 and localization	ARPKD	(Lu et al., 2017)
<i>SEC61A1</i>	SEC61 translocon subunit alpha 1	Translocation of polypeptides into the ER, forms part of the translocation pore SEC61	ER	NA	ADTKD-SEC61A1 (6)	(Bolar et al., 2016, Van Nieuwenhove et al., 2020)

Table 1.3. Genes associated and gene candidates to be associated with cystic kidney disease. (1) Biallelic mutations in *PKD1* are associated with severe polycystic disease and embryonic lethality (Cornec-Le Gall et al., 2018b, Bergmann et al., 2018). (2) Biallelic mutations in *IFT140* are associated with skeletal ciliopathy phenotypes (Bayat et al., 2017, Schmidts et al., 2013). (3) Biallelic mutations in *ALG8* or *ALG9* are associated with congenital disorder of glycosylation (Höck et al., 2015, Tham et al., 2016). (4) Heterozygous *LRP5* mutations have been associated with other diseases including primary osteoporosis and osteosclerosis (Cnossen et al., 2014, Korvala et al., 2012). (5) Heterozygous *PKHD1* mutations may also cause the development of multiple small liver cysts (Besse et al., 2017). (6) The ADTKD-SEC61A1 phenotypic spectrum includes glomerulocystic kidney disease with anaemia and neutropenia. ADPKD, autosomal dominant polycystic kidney disease; ADPLD, autosomal dominant polycystic liver disease; ADTKD, autosomal dominant tubulointerstitial kidney disease; ARPKD, autosomal recessive polycystic kidney disease; BiP, Binding immunoglobulin protein; ER, endoplasmic reticulum; NA, not available; PC1, polycystin-1; PC2, polycystin-2; WNT, Wingless/Integrated. Table updated and obtained from (Barroso-Gil et al., 2021b).

Biallelic mutations in *PKHD1* are the most common cause of ARPKD (Ward et al., 2002, Molinari et al., 2020). Regarding genotype-phenotype correlations in ARPKD, it has been shown that loss of function mutations in *PKHD1* have been associated with more severe ARPKD phenotypes than missense *PKHD1* variants (Furu et al., 2003). Additionally, biallelic mutations in *DZIP1L* have been associated with ARPKD (Lu et al., 2017).

PKHD1 encodes the protein Fibrocystin, which is a protein involved in the regulation of planar cell polarity, cell-cell and cell-matrix interactions (Fedele et al., 2014, Besse et al., 2017). Moreover, it has been found an enrichment of rare heterozygous *PKHD1* loss of function variants in a population of patients with polycystic liver disease (Besse et al., 2017). This reinforces the importance of remembering the phenotypic and genetic overlap between renal ciliopathies (Cornec-Le Gall et al., 2019), in this case, the two forms of PKD.

In recent years an additional cause of ARPKD has been found, biallelic mutations in *DZIP1L*. *DZIP1L* encodes for a DAZ interacting protein1-like protein (DZIP1L) protein, which functions at the basal body. Mutations in these genes has been associated with ciliary trafficking defects and abnormal location of polycystin-1 (PC1) and polycystin-2 (PC2) (Lu et al., 2017).

ARPKD generally develops earlier in age than ADPKD, however for some ARPKD patients with mild ARPKD, cystic phenotypes may develop later in life, it is challenging to provide a correct diagnosis, because of the ADPKD and ARPKD phenotypic overlap.

Biallelic mutations in *PKHD1* are the most common cause of ARPKD (Ward et al., 2002, Molinari et al., 2020). Regarding genotype-phenotype correlations in ARPKD, it has been shown that loss of function mutations in *PKHD1* have been associated with more severe ARPKD phenotypes than missense *PKHD1* variants (Furu et al., 2003). Additionally, biallelic mutations in *DZIP1L* have been associated with ARPKD (Lu et al., 2017).

PKHD1 encodes the protein Fibrocystin, which is a protein involved in the regulation of planar cell polarity, cell-cell and cell-matrix interactions (Fedele et al., 2014, Besse et al., 2017). Moreover, it has been found an enrichment of rare heterozygous *PKHD1* loss of function variants in a population of patients with polycystic liver disease (Besse et al., 2017). This reinforces the importance of remembering the

phenotypic and genetic overlap between renal ciliopathies (Cornec-Le Gall et al., 2019), in this case, the two forms of PKD.

In recent years an additional cause of ARPKD has been found, biallelic mutations in *DZIP1L*. *DZIP1L* encodes for a DAZ interacting protein1-like protein (DZIP1L) protein, which functions at the basal body. Mutations in these genes has been associated with ciliary trafficking defects and abnormal location of polycystin-1 (PC1) and polycystin-2 (PC2) (Lu et al., 2017).

ARPKD generally develops earlier in age than ADPKD, however for some ARPKD patients with mild ARPKD, cystic phenotypes may develop later in life, it is challenging to provide a correct diagnosis, because of the ADPKD and ARPKD phenotypic overlap (Bergmann, 2017, Cornec-Le Gall et al., 2019).

Biallelic mutations in *NPHP1*, which may include full-gene deletion, are the most common cause of NPHP, being responsible of about 20% of NPHP cases. Additionally, there are over 20 genes associated with NPHP (Table 1.4) (McConnachie et al., 2021, Srivastava et al., 2017a).

Many of these genes associated with NPHP are also associated with other ciliopathies such as JBTS, SLS and MKS (Braun and Hildebrandt, 2017, Halbritter et al., 2013b). There are several examples that reflect the genetic heterogeneity, overlap and pleiotropy of some of the genes associated with NPHP. For instance, *NPHP1*, encoding a protein functioning at the transition zone, is associated with SLS, NPHP and JBTS phenotypes (Fliegauf et al., 2006, Omran, 2010, Koyama et al., 2017, McConnachie et al., 2021). Moreover, *AHI1*, encoding a protein functioning at the centrosome, is associated with NPHP and JBTS phenotypes (Utsch et al., 2006, Muñoz-Estrada and Ferland, 2019).

Gene	NPHP type	Protein product	Function	Subcellular localization	Disease phenotypes	References
<i>NPHP1</i>	NPHP1	Nephrocystin-1	Interacts with inversin, nephrocystin-3, 4 and 8; docking protein	Microtubular organising center, TZ	NPHP, SLS, JBTS	(Fliegauf et al., 2006, Omran, 2010, Koyama et al., 2017)
<i>INVS</i>	NPHP2	Inversin	Involved in Wnt signaling, cell polarity and migration	Primary cilium	NPHP (with or without situs inversus), SLS	(Veland et al., 2013, Tory et al., 2009)
<i>NPHP3</i>	NPHP3	Nephrocystin-3	Interaction with nephrocystin-1 and Inversin	Primary cilium	NPHP, SLS, MKS	(Tory et al., 2009, Bergmann et al., 2008)
<i>NPHP4</i>	NPHP4	Nephrocystin-4/nephroretinin	Associated with Wnt signaling and the transition zone diffusion barrier	Primary cilium, centrosome, cell-cell junctions	NPHP, SLS, LCA	(Wiik et al., 2008, Otto et al., 2002, Mollet et al., 2005)
<i>IQCB1</i>	NPHP5	IQ motif containing B1 protein/nephrocystin-5	Interaction with calmodulin and RPGR, role in ciliogenesis	Primary cilium	SLS, LCA	(Estrada-Cuzcano et al., 2011, Otto et al., 2005)
<i>CEP290</i>	NPHP6	Centrosomal protein 290	Regulation of trafficking of ciliary proteins at the TZ	Centrosome, TZ, nucleus	JBTS, BBS, MKS, LCA, SLS	(Betleja and Cole, 2010, Sayer et al., 2006, Slaats et al., 2015)
<i>GLIS2</i>	NPHP7	Krüppel-like zinc finger protein Gli-similar 2	Transcription factor	Primary cilium, nucleus	NPHP	(Attanasio et al., 2007)
<i>RPGRIP1L</i>	NPHP8	Nephrocystin-8/RPGRIP1 like	Associated with the transition zone diffusion barrier	TZ/basal body	JBTS, COACH, MKS	(Arts et al., 2007, Williams et al., 2011)
<i>NEK8</i>	NPHP9	NIMA related kinase 8	Associated with the regulation of the Hippo pathway and the regulation of cell cycle progression from G2 to M phase	Primary cilium	NPHP	(Grampa et al., 2016, Hassan et al., 2020)
<i>SDCCAG8</i>	NPHP10	Nephrocystin-10/serologically defined colon cancer antigen 8	Associated with the activation of DDR signaling pathway. Interaction with RABEP2 and OFD1	Centrosome/basal body, cell-cell junctions, nucleus	SLS, BBS	(Airik et al., 2016, Otto et al., 2010, Airik et al., 2014)
<i>TMEM67</i>	NPHP11	Transmembrane protein 67/meckelin	Involved in ciliogenesis and centriole migration. Frizzled-like Wnt receptor. Interaction with filamin A	Plasma membrane, basal body, TZ	NPHP, JBTS, COACH, MKS	(Smith et al., 2006, Otto et al., 2009, Adams et al., 2012, Abdelhamed et al., 2019)

<i>TTC21B</i>	NPHP12	Tetratricopeptide repeat domain 21B	Involved in retrograde IFT transport	Primary cilium	NPHP, JBTS, JATD	(Davis et al., 2011)
<i>WDR19</i>	NPHP13	WD repeat-containing protein 19	Involved in retrograde IFT transport	Primary cilium	NPHP, RP, SLS, sensenbrenner syndrome, Mainzer-Saldino syndrome, JATD, SRTD, asthenoteratospermia	(Bredrup et al., 2011, Ni et al., 2020, Halbritter et al., 2013b, Coussa et al., 2013, Fehrenbach et al., 2014)
<i>ZNF423</i>	NPHP14	Zinc finger protein 423	Transcription factor. DDR signaling pathway	Nucleus, centrosome	NPHP, JBTS	(Chaki et al., 2012, Cheng et al., 2007)
<i>CEP164</i>	NPHP15	Centrosomal protein 164	Involved in vesicular docking to the mother centriole, G2/M checkpoint and DDR signaling pathway	Nucleus, centrosome	NPHP, SLS, JBTS	(Schmidt et al., 2012, Chaki et al., 2012, Maria et al., 2016)
<i>ANKS6</i>	NPHP16	Ankyrin repeat and sterile alpha motif domain-containing protein 6	Connects NEK8 to inversin and nephrocytsin-3	Primary cilia	NPHP	(Hoff et al., 2013, Taskiran et al., 2014, Nakajima et al., 2018)
<i>IFT172</i>	NPHP17	Intraflagellar transport 172	IFT protein (IFT-B module), involved in IFT transport required for cilia assembly and function	Primary cilia	NPHP, JATD, Mainzer-Saldino syndrome, RP, BBS	(Halbritter et al., 2013a, Pruski et al., 2019)
<i>CEP83</i>	NPHP18	Centrosomal protein 83/coiled-coil domain-containing protein 41	Involved in vesicular docking to the mother centriole and anchorage of the centrosome to the apical membrane	Centrosome	NPHP	(Joo et al., 2013, Shao et al., 2020b)
<i>DCDC2</i>	NPHP19	Doublecortin domain-containing protein 2	Interaction with the mediator of Wnt signaling dishevelled	Primary cilium, mitotic spindle fibres	NPHP, liver fibrosis, neonatal sclerosing cholangitis	(Schueler et al., 2015, Girard et al., 2016)
<i>MAPKBP1</i>	NPHP20	Mitogen-activated protein kinase-binding protein 1/JNK-binding protein 1	Involved DDR signaling pathway and regulation of the JNK and NOD2 signaling pathways	Mitotic spindle poles	NPHP	(Lecat et al., 2012, Macia et al., 2017)

<i>XPNP EP3</i>	NPHPL 1	X-prolyl aminopeptidase 3	Recognises N-terminal peptides that have proline in their second position and removes the N-terminal amino acid	Mitochondria	NPHPL, NPHP	(Alizadeh et al., 2020, O'Toole et al., 2010)
<i>ADAM TS9</i>	NPHPL	A disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 9	Metalloprotease involved in the proteolytic modification of ECM components and ciliogenesis	Near basal body	NPHPL	(Choi et al., 2019, Nandadasa et al., 2019)
<i>AHI1</i>	—	Abelson helper integration site 1/Jouberin	Disruption of this gene affects Hh and Wnt signalling pathways	Centrosome, cell-cell junctions	NPHP, JBTS	(Utsch et al., 2006, Muñoz-Estrada and Ferland, 2019)
<i>CC2D 2A</i>	—	Coiled-coil and C2 domain-containing 2A	Involved in ciliogenesis and vesicle docking and fusion at the periciliary membrane (in photoreceptors)	TZ, mother centriole subdistal appendages	JBTS, COACH, MKS	(Tallila et al., 2008, Bachmann-Gagescu et al., 2015b, Mejecase et al., 2019)
<i>ATXN 10</i>	—	Ataxin 10	Dysfunction of this gene affects survival of cerebellar neurons.	Nucleus, cytoplasm	NPHP/spinocerebellar ataxia	(Sang et al., 2011, Wakamiya et al., 2006)
<i>SLC41 A1</i>	—	Solute carrier family 41 member 1	Magnesium transporter. Associated with magnesium homeostasis	Basolateral cell membrane	NPHP-like/primary ciliary dyskinesia	(Arjona et al., 2019, Hurd et al., 2013)

Table 1.4. Genes associated with nephronophthisis (NPHP). BBS, Bardet Biedl syndrome; COACH, Cerebellar vermis defect, oligophrenia, ataxia, coloboma, hepatic fibrosis; cJNK, c-Jun N-terminal kinase; DDR, DNA damage response; ECM, extracellular matrix; JATD, Jeune asphyxiating thoracic dystrophy; JBTS, Joubert syndrome; LCA, Leber congenital amaurosis; MKS, Meckel syndrome; NOD2, nucleotide-binding oligomerization domain-containing protein 2; NPHP, nephronophthisis; NPHPL, nephronophthisis-like nephropathy; RABEP2, RAB GTPase binding effector protein 2; RP, retinitis pigmentosa; RPGR, retinitis pigmentosa GTPase regulator; SLS, Senior-Loken syndrome; TZ, transition zone. Table obtained from (Barroso-Gil et al., 2021b).

1.4. Molecular disease pathogenesis of ciliopathy genes studied in this project

There are more than 40 genes associated with JBTS, more than 30 genes associated with JATD, more than 15 genes associated with polycystic kidney disease and more than 20 genes associated with NPHP (many of them also associated with JBTS and other primary ciliopathies) (Gana et al., 2022, Devlin and Sayer, 2019, Zhang et al., 2018).

I have focussed on studying a subset of ciliopathy genes to investigate the genetic heterogeneity and pleiotropy of primary ciliopathies. Firstly, *CEP120*, *ARL3* and *TOGARAM1* were selected for this project because they are associated with JBTS and because there is a growing interest to further study the molecular genetics of primary ciliopathies with a focus on each of these genes (Joseph et al., 2018, Tsai et al., 2019, Alkanderi et al., 2018, Latour et al., 2019).

CEP120 is associated with JBTS and JATD syndromes (Roosing et al., 2016, Shaheen et al., 2015). Studying *CEP120* allows to investigate how mutations in the same gene can cause different clinical phenotypes, which is a characteristic found in many other ciliopathy genes. Studying *CEP120* also allows to study extra-ciliary functions that may play a role in the wide spectrum of clinical phenotypes (from neurological phenotypes to skeletal phenotypes) arising from mutations in *CEP120* (Roosing et al., 2016, Shaheen et al., 2015).

Recently, other ciliopathy genes such as *KIAA0586* (Pauli et al., 2019), *ARL3* (Alkanderi et al., 2018, Sheikh et al., 2019) and *CEP290* (Ramsbottom et al., 2018), which are similar to *CEP120* in terms of associated phenotypes, have been investigated by Sayer and Miles lab at Newcastle University and other research groups. I have focussed on *CEP120* because the results about its molecular disease pathogenesis can be compared with the results already reported by others, studying the molecular disease pathogenesis of *CEP120* leading the phenotypic spectrum of JBTS and JATD syndromes (Betleja et al., 2018, Comartin et al., 2013a, Joseph et al., 2018, Roosing et al., 2016, Tsai et al., 2019, Wu et al., 2014).

Furthermore, there is an increasing number of studies and growing interest focussing on the molecular pathogenesis of different patient-specific disease-causing *CEP120* mutations reported in the literature (Chang et al., 2021, Joseph et al., 2018, Tsai et al., 2019), this allows new studies about the molecular pathogenesis of *CEP120* to be compared and complemented with previous studies.

ARL3 was included in this project because it can be compared with *CEP120* in several aspects (from molecular function to disease phenotypes). Moreover, *ARL3* has been extensively investigated in Sayer and Miles laboratory (Alkanderi et al., 2018). Furthermore, both genes are suitable to study the molecular genetics of primary ciliopathies using different *in vitro* and *in vivo* functional assays (Joseph et al., 2018, Wu et al., 2014, Alkanderi et al., 2018, Fu et al., 2021).

Furthermore, the comparison of how genes encoding proteins with different functions, as it is the case of *CEP120* and *ARL3*, can cause the same ciliopathy (in this case: JBTS), it is of great interest and relevance to this project, to the ongoing projects in Sayer and Miles laboratory and to the study of primary ciliopathies and other ciliopathy genes. *ARL3* is a pleiotropic gene and there is a growing interest in studying this gene and new disease-causing mutations in *ARL3* have been reported recently (Alkanderi et al., 2018, Fu et al., 2021, Holtan et al., 2019, Ratnapriya et al., 2021, Sheikh et al., 2019).

In summary, *CEP120* and *ARL3* were selected for this project because they allow to study genotypic and phenotypic heterogeneity and overlap in primary ciliopathies. There are patients in the literature reported presenting with biallelic mutations in these genes and their study contributes to further understand genotype-phenotype correlations, the molecular genetics and tissue-specific expression of these two genes and to elucidate why mutations in the same gene can cause different disease phenotypes (Ratnapriya et al., 2021, Sheikh et al., 2019, Roosing et al., 2016).

I also focused on *TOGARAM1*, a gene recently described as a JBTS gene, because investigating novel ciliopathy genes is an important part of studying the molecular genetics of primary ciliopathies (Latour et al., 2019, Latour et al., 2020). Investigating genes recently described as JBTS genes, such as *TOGARAM1*, is important to get insights into the molecular genetics of primary ciliopathies and genetically solve new patients. *TOGARAM1* was selected because I had the opportunity to access the whole genome sequencing data available at the Genomics England 100,000 Genomes Project (Smedley et al., 2021) and search for putative disease-causing mutations in *TOGARAM1* leading to JBTS.

I have studied genes encoding proteins functioning at the endoplasmic reticulum, such as *DNAJB11* gene and *ALG* genes because some of these genes are associated with renal primary ciliopathies, most of them are recently associated (or

are candidate genes to be associated) with ADPKD (Cornec-Le Gall et al., 2018a, Lanktree et al., 2021, Besse et al., 2017). ADPKD is characterised by the development of kidney cysts and it is estimated that has a prevalence of 1:1000, which increases the probability of finding patients with the associated disease phenotypes compared with other primary ciliopathies (Devlin and Sayer, 2019, McConnachie et al., 2021).

In the Genomics England 100,000 Genomes Project, there are more patients diagnosed with cystic kidney disease than any other primary ciliopathy (eg: JBTS) (Wheway et al., 2019). Focusing on patients with cystic kidney disease, allows to genetically solve a higher proportion of patients and to increase the genetic diagnosis yield in patients with cystic kidney disease. This also allows to perform *in silico* analyses using the Genomics England 100,000 Genomes Project, comparing large cohorts of patients and perform robust statistical analyses.

Studying these ciliopathy genes encoding proteins functioning within the endoplasmic reticulum (ER), allows to study how mutations in genes encoding proteins functioning in extra-ciliary locations, can cause the same or similar clinical phenotypes caused by mutations in ciliary genes (eg: *PKD1* and *PKD2*) (Besse et al., 2017, Lemoine et al., 2022).

These is a growing interest in studying genes encoding ER proteins (e.g. *DNAJB11*, *GANAB*, and *ALG5*, *ALG8* and *ALG9* genes) and novel disease-causing genes leading to the phenotypic spectrum of ADPKD (Lemoine et al., 2022, Besse et al., 2019, Besse et al., 2017, Lanktree et al., 2021).

Mutations located in some of these genes encoding proteins functioning within the ER can cause different ciliopathy phenotypes: including renal and liver polycystic disease. This allows to study gene pleiotropy in these genes associated with renal ciliopathies and to study genetic heterogeneity and phenotypic overlap in renal ciliopathies (Besse et al., 2017, Lanktree et al., 2021, Cornec-Le Gall et al., 2019, McConnachie et al., 2021).

In summary, this approach allows to study primary ciliopathies from a broad perspective: from recessive primary ciliopathies such as JBTS and JATD syndromes associated with overlapping clinical phenotypes affecting multiple organs, to the study of recessive and dominant primary ciliopathies leading to a more restricted spectrum of clinical phenotypes affecting kidney and/or liver.

There is a growing interest in studying the genes mentioned above as well as studying extra ciliary functions, novel genotype-phenotype correlations and the analysis of whole genome sequencing (WGS) data. Studying the primary ciliopathies and genes stayed above is a suitable approach to achieve the objective of studying the molecular genetics of primary ciliopathies and increase genetic diagnostic yield in patients with primary ciliopathies.

1.4.1. *CEP120*

Centrosomal protein of 120 kDa (*CEP120*) is a centrosomal protein involved in centriole duplication, assembly (Mahjoub et al., 2010, Wu et al., 2014), elongation (Lin et al., 2013) and maturation (Wu et al., 2014) as well as microtubule stability (Betleja et al., 2018, Meka et al., 2022).

CEP120 encodes for a protein of 986 amino acids (aa). It is predicted that *CEP120* contains three consecutive C2 domains (C2A and C2B and C2C) and one coil-coiled domain (CC) (Mahjoub et al., 2010, Sharma et al., 2018). These three has three consecutive evolutionary conserved C2 domains: C2A (aa 1 - 151), C2B (aa 160 - 340) and C2C (aa 450 - 610), which possess different functions (Sharma et al., 2018, Joseph et al., 2018). The N-terminal C2A domain binds tubulin and microtubules and it promotes microtubule formation (Sharma et al., 2018) (Figure 1.10).

It is important to note that *CEP120* protein sequence is conserved in mammals (>80% homology) (Sharma et al., 2018). Moreover, there is a tendency of the protein domains being slightly more conserved than the *CEP120* full protein (Table 1.5).

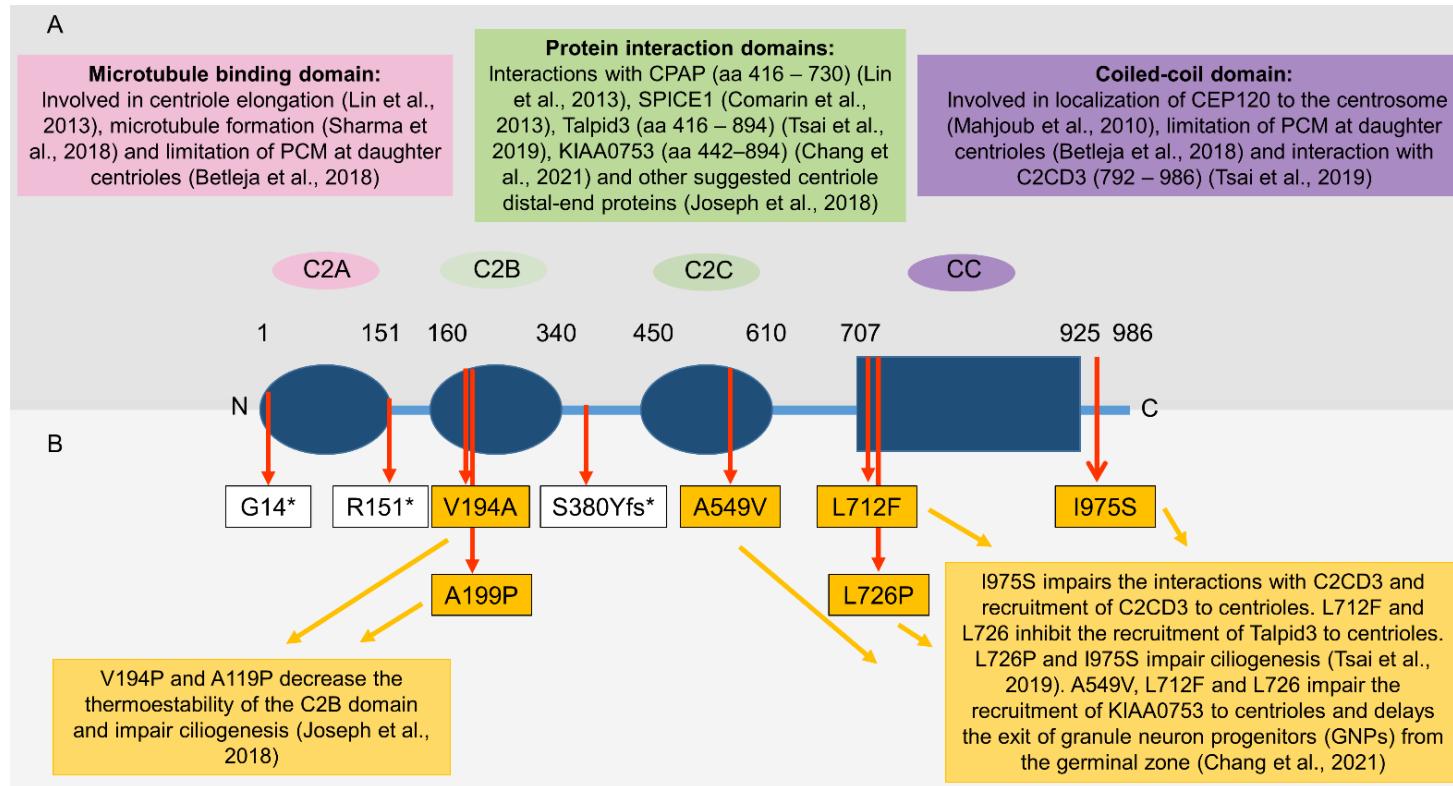


Figure 1.10. Representation of the domain structure of CEP120 protein (NP_694955.2). CEP120 is formed by three C2 evolutionary conserved consecutive C2 domains (C2A, C2B and C2C) and a C-terminal coil-coiled domain (CC) (Sharma et al., 2018). The different domains have different roles. (B) Representation of the nine CEP120 mutations described to be associated with JBTS (G14*, V194A, S380Yfs*, A549V, L712F and L726P), JATD (A199P), MKS/OFD (I949S) and TCDOE (R151* and A199P) syndromes. Some of these mutations (coloured in orange) have been investigated to understand the molecular disease mechanism of them (Chang et al., 2021, Joseph et al., 2018, Tsai et al., 2019). The annotation of the CEP120 mutations (CEP120 transcript: NM_153223.3) is: c.49+5_49+10del; p.(Gly+1AspfsTer14) (G14*), c.451C>T; p.(Arg151Ter) (R151*), c.581T>C; p.(Val194Ala) (V194A), c.595G>C; p.(Ala199Pro) (A199P), c.1138_1139insA; p.(Ser380TyrfsTer19) (S380Yfs*), c.1646C>T; p.(Ala549Val) (A549V), c.2134C>T; p.(Leu712Phe) (L712F), c.2177T>C; p.(Leu726Pro) (L726P) and c.2924T>G; p.(Ile975Ser) (I975S). JATD, Jeune asphyxiating thoracic dystrophy; JBTS, Joubert syndrome; MKS, Meckel syndrome; OFD, oro-facial-digital syndrome; TCDOE, tectocerebellar dysraphia with occipital encephalocele.

	CEP120 full protein			CEP120 C2A domain (1-151aa)		CEP120 C2B domain (160-340aa)		CEP120 C2C domain (450-610aa)		CEP120 CC domain (707-925aa)			
Organism	ID	Query Cover %	Homology %	Query Cover %	Homology %	Query Cover %	Homology %	Query Cover %	Homology %	Query Cover %	Homology %	Database	
<i>Homo sapiens</i>	NP_694955.2	100	100	100	100	100	100	100	100	100	100	NCBI	
<i>Pan troglodytes</i>	XP_009447838.1	100	99.70	100	100	100	100	100	100	100	100	NCBI	
<i>Macaca mulatta</i>	XP_014996146.1	100	98.58	100	98.68	100	99.45	100	98.14	100	99.09	NCBI	
<i>Callithrix jacchus</i>	XP_01782503.1	100	96.86	100	96.69	100	97.79	100	98.76	100	97.72	NCBI	
<i>Sus scrofa</i>	XP_003123925.1	100	94.52	100	97.35	100	95.58	100	96.89	100	94.52	NCBI	
<i>Canis lupus familiaris</i>	XP_531888.3	100	92.83	100	98.01	100	95.03	100	95.65	100	93.61	NCBI	
<i>Capra hircus</i>	XP_017906171.1	100	93.31	100	95.36	100	93.37	100	95.65	100	93.61	NCBI	
<i>Bos Taurus</i>	XP_024850151.1	100	93.15	100	97.35	100	94.48	100	95.03	100	91.52	NCBI	
<i>Marmota marmota marmota</i>	XP_015342220.1	100	92.91	100	96.69	100	95.03	100	95.65	100	94.06	NCBI	
<i>Felis catus</i>	XP_011280127.1	100	92.40	100	96.69	100	93.92	100	95.03	100	94.06	NCBI	
<i>Heteroccephalus glaber</i>	XP_004854535.1	100	91.69	100	94.70	99	95	100	95.65	100	92.69	NCBI	
<i>Erinaceus europaeus</i>	XP_007527673.1	100	90.59	100	96.03	100	95.58	100	94.41	100	92.69	NCBI	
<i>Rattus norvegicus</i>	NP_001178626.1	100	90.39	100	94.70	99	94.44	100	95.65	100	90.41	NCBI	
<i>Mus musculus</i>	NP_848801.2	99	89.17	100	91.39	100	93.92	99	93.75	100	89.04	NCBI	
<i>Gallus gallus</i>	XP_015136305.1	99	72.03	100	80.26	100	79.01	100	80.75	97	77.67	NCBI	
<i>Xenopus tropicalis</i>	NP_001120380.1	92	65.74	100	80.13	100	75.15	99	78.12	92	61.58	NCBI	
<i>Erpetoichthys calabaricus</i>	XP_028661076.1	99	61.36	100	70.20	99	64.44	99	68.12	97	68.84	NCBI	
<i>Salmo salar</i>	XP_014030376.1	99	60.96	100	70.13	100	62.98	100	62.57	98	69.30	NCBI	
<i>Salmo trutta</i>	XP_029578107.1	99	60.59	100	68.83	100	62.98	100	62.57	98	69.30	NCBI	
<i>Danio rerio</i>	XP_017212877.1	99	57.67	100	68.21	98	61.80	99	62.50	100	66.52	NCBI	
<i>Oryzias latipes</i>	XP_011477980.1	99	54.12	100	64.29	99	54.44	99	58.12	98	62.04	NCBI	
<i>Takifugu rubripes</i>	XP_011613676.2	99	52.04	100	64.29	97	56.50	99	56.88	95	59.09	NCBI	
<i>Poecilia reticulata</i>	XP_017162041.1	55	51.39	100	64.29	99	55.25	99	38.60	94	65.28	NCBI	
<i>Callorhinchus milii</i>	XP_007897573.1	98	64.01	96	71.72	98	65.38	99	76.25	96	75.36	NCBI	
<i>Drosophila melanogaster</i>						Not found						FlyBase	
<i>Caenorhabditis elegans</i>						Not found						WormBase (WS72)	
<i>Chlamydomonas reinhardtii</i>						Not found						Phytozome	
<i>Chlamydomonas reinhardtii*</i>	XP_001694685.1	29	36.43		Not hit found		Not hit found		65	23.14	83	35.37	NCBI

	Mammalia
	Tetrapoda (non-mammals)

	Euteleostomi (non-tetrapods)
	Vertebrata (non-euteleostomi)

	Bilateria (non-vertebrate)
	Eukaryota (non-bilateria)

Table 1.5. Comparison of the different the human CEP120 protein and CEP120 protein domains against CEP120 orthologues. The protein sequences corresponding to human CEP120 were aligned against CEP120 orthologues using the NCBI tool BLASTP. It is observed how CEP120 human sequence shares >60 % of homology with the protein sequences of some non-tetrapod organisms (e.g. *Erpetoichthys calabaricus* and *Callorhinchus mili*), and that the percentage of homology increase when only the protein sequence corresponding to the functional domains described in the human CEP120 protein is aligned (Powell et al., 2020, Sharma et al., 2018). Percent homology refers to the similarity between two sequences, in this case, protein sequences from different organisms. Percent homology was obtained from the percent identity, which quantifies the similarity between two aligned sequences. Some of the CEP120 sequences obtained from NCBI database were predicted: XP_017825203.1[*Callithrix_jacchus*], XP_017906171.1[*Capra_hircus*], XP_015342220.1[*Marmota_marmota_marmota*], XP_007527673.1[*Erinaceus_europaeus*], XP_007897573.1[*Callorhinchus_milii*], XP_014030376.1[*Salmo_salar*], XP_017162041.1[*Poecilia_reticulata*], XP_001694685.1[*Chlamydomonas_reinhardtii*] are described as predicted sequences. (*) I included the protein sequence of the Uni2 protein of the ciliated alga *Chlamydomonas reinhardtii* although if this protein is truly a CEP120 homolog is not yet elucidated (Mahjoub et al., 2010, Piasecki and Silflow, 2009).

CEP120 interacts with transforming acidic coiled-coil proteins (TACCs) (Xie et al., 2007) and other centrosomal proteins: CPAP (Comartin et al., 2013b, Lin et al., 2013), SPICE1 (Comartin et al., 2013b), Talpid3 (Wu et al., 2014, Tsai et al., 2019) and C2CD3 (Tsai et al., 2019), which are also required for centriole biogenesis. It was also described that CEP120 and its interactions with other proteins have a role in centriole biogenesis, particularly regulating centriole assembly (Mahjoub et al., 2010, Sharma et al., 2018), duplication (Comartin et al., 2013b) and centriole elongation (Comartin et al., 2013b). Moreover, CEP120 is involved in limiting the amount of pericentriolar material (PCM) at the centrioles, especially at the daughter centriole (Betleja et al., 2018). Depletion of Cep120 causes accumulation of PCM at the daughter centriole (Betleja et al., 2018).

In vivo experiments have been performed to investigate the effect of Cep120 loss in mice (Wu et al., 2014, Xie et al., 2007). Cep120 depletion impaired a microtubule dependent process called interkinetic nuclear migration (INM). INM is an oscillating cell-cycle dependent movement of the nuclei characteristic of vertebrate neural progenitors (Xie et al., 2007). It was suggested that Cep120 has an important role in neocortical development by regulating INM and maintaining the neural progenitor self-renewal (Xie et al., 2007), it was also suggested that the root cause of the impaired INM was due the loss of centrioles (Mahjoub et al., 2010).

The *Cep120* knockout mouse was embryonic lethal. A conditional *Cep120* knockout in the mouse central nervous system resulted in defects in the cerebellar development, in particular, cerebellar hypoplasia and hydrocephalus hydrocephalous and cerebellar hypoplasia (Wu et al., 2014). Lack of cilia was observed in cerebellar granule neuron progenitors (CGNPs) and ependymal cells. It was suggested that failed expansion of CGNPs was due to lack to a response of Hh signalling (Wu et al., 2014).

Biallelic *CEP120* mutations have been found in patients with complex ciliopathy phenotypes, in particular JBTS and JATD syndromes (Roosing et al., 2016, Shaheen et al., 2015) (Table 1.6). The pathogenic homozygous missense mutation (A199P) was found in three JATD patients (Shaheen et al., 2015). Later this mutation and other eight distinct *CEP120* mutations were found in six patients with different ciliopathy phenotypes, including JBTS, tectocerebellar dysraphia with occipital encephalocele (TCDOE) and MKS. These mutations were found in compound heterozygous (three cases) or homozygous (six cases) (Roosing et al., 2016, Barroso-Gil et al., 2021a) (Tables 1.6 and 1.7).

Family ID	Patient ID	Phenotype	Kidney phenotype (1)	Allele 1 (Ex,Int)	Allele 2 (Ex,Int)	Reference
1	Family 1_II:2	JATD (3)	NA	c.595G>C ; p.Ala199Pro (Ex6)	c.595G>C ; p.Ala199Pro (Ex6)	(Shaheen et al., 2015)
2	Family 2_II:4	JATD (3)	NA	c.595G>C ; p.Ala199Pro (Ex6)	c.595G>C ; p.Ala199Pro (Ex6)	(Shaheen et al., 2015)
3	Family 3_II:1	JATD (3)	yes	c.595G>C ; p.Ala199Pro (Ex6)	c.595G>C ; p.Ala199Pro (Ex6)	(Shaheen et al., 2015)
4	COR39 1	JBTS	no	c.581T>C ; p.Val194Ala (Ex6)	c.581T>C ; p.Val194Ala (Ex6)	(Roosing et al., 2016)
5	MTI-143	JBTS	no (2)	c.2177T>C ; p.Leu726Pro (Ex16)	c.2134C>T ; p.Leu712Phe (Ex16)	(Roosing et al., 2016)
6	MTI-991	JBTS	no	c.49+5_49+10del ; p.Gly+1AspfsTer14 (?) (Int2)	c.49+5_49+10del ; p.Gly+1AspfsTer14 (?) (Int2)	(Roosing et al., 2016)
7	MTI-1516	JBTS	no	c.1138_1139insA ; p.Ser380TyrfsTer19 (Ex9)	c.1646C>T ; p.Ala549Val (Ex12)	(Roosing et al., 2016)
8	MKS-2930	MKS/OFD	yes	c.2924T>G ; p.Ile975Ser (Ex21)	c.2924T>G ; p.Ile975Ser (Ex21)	(Roosing et al., 2016)
9	SW-476410	TCDOE	no	c.451C>T ; p.Arg151Ter (Ex5)	c.595G>C ; p.Ala199Pro (Ex6)	(Roosing et al., 2016)

Table 1.6. List of patients with biallelic *CEP120* variants and their associated phenotypes reported in the literature (ranked chronologically by publication). *CEP120* transcript considered: NM_153223.3. (1) Designated as NA, unless renal phenotype clearly stated. (2) For this patient Roosing et al., reported: “grade II-III hydronephrosis was detected at birth but it spontaneously resolved after few months. No renal problems have been reported since then” (Roosing et al., 2016). (3) A fourth JATD case was described in Shaheen et al., 2015, however DNA from the proband was not available. Of note, both parents of this fourth JATD case were heterozygous for the *CEP120*: c.595G>C; p.(Ala199Pro) variant (Shaheen et al., 2015). JATD, Jeune asphyxiating thoracic dystrophy; JBTS, Joubert syndrome; MKS, Meckel syndrome; NA, not available; OFD, oral-facial-digital syndrome; TCDOE, tectocerebellar dysraphia with occipital encephalocele. Table obtained from (Barroso-Gil et al., 2021a).

Genomic position (GRCh38)	Nucleotide change	Protein change	rs	gnomAD v2.1.1	SIFT	CADD	Polyphen	ACMG	Location	Allele count (1)
chr5:123422940	c.49+5_49+10del	p.(Gly+1AspfsTer14) (2)	NA	NA	NA	NA	NA	Uncertain significance	Intron 2	2
chr5:123412411	c.451C>T	p.(Arg151Ter)	rs757499322	4.24E-06	NA	8.47	NA	Pathogenic	Exon 5	1
chr5:123399167	c.581T>C	p.(Val194Ala)	rs1554104276	NA	Deleterious (0)	28.1	Probably damaging (0.999)	Uncertain significance	Exon 6	2
chr5:123399153	c.595G>C	p.(Ala199Pro)	rs367600930	2.81E-05	Deleterious (0)	NA	Probably damaging (1.0)	Pathogenic	Exon 6	7
chr5:123390041	c.1138_1139insA	p.(Ser380TyrfsTer19)	rs1554103267	NA	NA	NA	NA	Pathogenic	Exon 9	1
chr5:123385068	c.1646C>T	p.(Ala549Val)	rs775080726	1.99E-05	Deleterious (0.04)	27.4	Benign (0.041)	Uncertain significance	Exon 12	1
chr5:123378398	c.2134C>T	p.(Leu712Phe)	rs114280473	0.003812 (3)	Deleterious (0.03)	26.0	Probably damaging (0.997)	Likely benign	Exon 16	1
chr5:123378355	c.2177T>C	p.(Leu726Pro)	rs1554102026	NA	Deleterious (0)	29.6	Probably damaging (0.999)	Uncertain significance	Exon 16	1
chr5:123346556	c.2924T>G	p.(Ile975Ser)	rs1554098663	NA	Deleterious (0)	27.6	Probably damaging (0.998)	Uncertain significance	Exon 21	2

Table 1.7. List of reported genetic variants in *CEP120* (ranked by genomic position). *CEP120* transcript considered: NM_153223.3. (1) Only family index cases considered. (2) This variant leads to occasional intron retention with use of an alternative splice site predicted to result in a truncated protein after 14 amino acids (Roosing et al., 2016). (3) There are two homozygous individuals for this variant found in gnomAD v2.1.1 database. NA, not available. Table updated from (Barroso-Gil et al., 2021a).

1.4.2. CC2D2A

CC2D2A is one of the genes associated with JBTS. *CC2D2A* encodes for Coiled-coil and C2 domain containing 2A (*CC2D2A*) protein (Gorden et al., 2008). *CC2D2A* functions at the transition zone and have a role in ciliogenesis (in multiple cell types) (Tallila et al., 2008, Veleri et al., 2014, Lewis et al., 2019) and vesicle docking and fusion at the periciliary membrane (in photoreceptors) (Bachmann-Gagescu et al., 2015b, Williams et al., 2011, Ojeda Naharro et al., 2017). *CC2D2A* interacts with other proteins associated with ciliary phenotypes, such as *CEP290* (Gorden et al., 2008) and *TCTN1* (Garcia-Gonzalo et al., 2011).

CC2D2A is expressed in multiple tissues, including adult human tissues such as brain, prostate, pancreas, kidney, lung, liver and retina (Noor et al., 2008).

Biallelic *CC2D2A* mutations are associated with JBTS as well as a spectrum of clinical phenotypes. Biallelic *CC2D2A* mutations can cause rod-cone dystrophy (RCD) (Mejecase et al., 2019) and more severe phenotypes including JBTS (Bachmann-Gagescu et al., 2012, Gorden et al., 2008, Noor et al., 2008) and MKS (Mougou-Zerelli et al., 2009, Szymanska et al., 2012, Tallila et al., 2008, Tallila et al., 2009).

For more than a decade there have been multiple studies reporting patients with biallelic pathogenic mutations in *CC2D2A*. It is estimated that *CC2D2A* mutations account for about 10% of both JBTS and MKS patients (Bachmann-Gagescu et al., 2012, Mougou-Zerelli et al., 2009, Vilboux et al., 2017a).

As there is a large number of variants reported with disease-causing *CC2D2A* mutations it has been possible to study possible genotype-phenotype correlations in JBTS patients with disease-causing *CC2D2A* mutations. It has been shown that the severity of the *CC2D2A* genotype correlates with the severity of the *CC2D2A* disease phenotype (Bachmann-Gagescu et al., 2012, Mougou-Zerelli et al., 2009).

More precisely, patients with biallelic loss of function variants in *CC2D2A* are more likely to develop severe phenotypes that include Meckel (MKS) or Meckel-like (ML) syndromes than patients that present one or two *CC2D2A* missense mutation/s, which may lead less severe phenotype, including JBTS (Bachmann-Gagescu et al., 2012, Mougou-Zerelli et al., 2009). Moreover, it has been shown that *CC2D2A*-

affected JBTS patients are more likely to suffer from seizures and ventriculomegaly (Bachmann-Gagescu et al., 2012).

Similar genotype-phenotype correlations have been suggested for JBTS patients with variants in *TMEM67* and *RPGRIP1L*, in which biallelic loss of function variants are more likely to lead more severe phenotypes (Iannicelli et al., 2010, Delous et al., 2007).

1.4.3. *ARL3*

ARL3 is one of the genes associated with JBTS. *ARL3* encodes ADP-ribosylation factor-like 3 (ARL3), a GTP-binding protein. ARL3 functions at the primary cilium and is involved in releasing lipid-modified proteins into the cilium. Its function is dependent on the cycling of its two states, depending on the addition of GDP (GDP-bound, inactive state) or GTP (GDP-bound, active state) (Gotthardt et al., 2015). To do this, ARL3 interacts with several proteins, such as ARL13B at the primary cilium, which is a guanine nucleotide exchange factor (GEF) (Blacque et al., 2005) and retinitis pigmentosa 2 (RP2) at the basal body of the cilium, which is a GTPase activating protein (GAP) (Veltel et al., 2008, Evans et al., 2010).

A knockout of *Ar3* in mouse displayed disease phenotypes in multiple organs such as kidney, liver and pancreas (Schrick et al., 2006) and rod and retina conditional *Ar3* mouse knockouts revealed defects in photoreceptors (Hanke-Gogokhia et al., 2016).

Biallelic mutations in *ARL3* have been associated with JBTS (Alkanderi et al., 2018), autosomal recessive rod-cone dystrophy (Fu et al., 2021) and autosomal recessive cone rod dystrophy (Sheikh et al., 2019). Monoallelic *ARL3* mutations have been associated with autosomal dominant retinitis pigmentosa (Holtan et al., 2019, Strom et al., 2016) and autosomal dominant retinal degeneration (Ratnapriya et al., 2021).

There are two consanguineous JBTS *ARL3*-affected families described up to date. The JBTS patients in these two families present a homozygous missense mutation in *ARL3*: c.445C>T; p.(Arg149Cys) (in the affected member of Family 1) and c.446G>A; p.(Arg149His) (in the affected members of Family 2) (Alkanderi et al., 2018). These two mutations have an effect on the same ARL3 evolutionary conserved amino acid (Arg149) and are described to affect the ARL13B-ARL3 interaction and ARL3 function (Alkanderi et al., 2018).

1.4.4. TOGARAM1

TOGARAM1 encodes for TOG array regulator of axonemal microtubules 1 (TOGARAM1). TOGARAM1 has 4 protein TOG domains. These domains are important for regulation of microtubules in the cytoplasm and cilia (Das et al., 2015). TOGARAM 1 interacts with ARMC9 which is a protein functioning in centrosome/basal body and is associated with JBTS (Latour et al., 2020, Van De Weghe et al., 2017).

It has been shown that the TOGARAM-ARMC9 complex is required for regulate the tubulin posttranslational modifications (PTMs) of microtubules in cilia. The disruption of any of these two proteins leads to reduced tubulin PTMs, and its disruption cause abnormal ciliary length and abnormal axonemal stability (Latour et al., 2020).

Biallelic mutations in *TOGARAM1* have been associated with JBTS (Latour et al., 2020). In one study, 5 patients from 5 different unrelated families were described. Some of these patients showed phenotypes other than the MTS. Extra cerebellar clinical phenotypes include polydactyly (2 patients) and hepatic (1 patient) and renal phenotypes (2 patients) (Latour et al., 2020).

Regarding the genotype of these patients, 3 presented compound heterozygous *TOGARAM1* mutations and 2 of them presented homozygous *TOGARAM1* mutations. In total, in this study 8 *TOGARAM1* variants were described, 6 of them affect any of the TOG domains. More precisely, 4 of these mutations are located within the sequence encoding the TOG2 domain of TOGARAM1. Moreover, 3 of these 4 mutations are missense mutations. Using *in silico* and *in vitro* (yeast two hybrid direct interaction) analyses it has been shown that these missense *TOGARAM1* mutations lead to the disruption of the interaction of TOGARAM1 with ARMC9 (Latour et al., 2020).

Moreover, recently two additional patients from one family presenting with biallelic *TOGARAM1* variants have been described. These two patients were foetuses presenting clinical phenotypes that included severe MKS phenotypes such as hydrocephalus (Morbidoni et al., 2021).

1.4.5. Other genes associated with Joubert syndrome

As it was mentioned before, there are about 40 genes described to cause JBTS (Bachmann-Gagescu et al., 2020, Gana et al., 2022). These genes may function at

the primary cilia (such as ARL3 and SUFU) (Alkanderi et al., 2018, Pala et al., 2017) or at other locations such as the TZ (such as CC2D2A and *CEP290*) (Veleri et al., 2014, Gorden et al., 2008) and centrosome (such as CEP120, OFD1 and CEP104) (Mahjoub et al., 2010, Frikstad et al., 2019, Rusterholz et al., 2022).

As a final example of the genetic heterogeneity and pleiotropy of genes associated with JBTS, there are several other genes associated with JBTS, that encode proteins involved in the transport of cargo proteins and lipids along the axoneme. This is an evolutionary conserved process called intraflagellar transport (IFT), which is required for the transport of proteins and lipids from basal body to the tip of cilium 2 (anterograde IFT) and from the tip of the cilium to the basal body (retrograde IFT) (Beales et al., 2007, Duran et al., 2016, Satir, 2017, Vertii et al., 2015). It is important to note that protein translation does not take place within the cilium and transport of all proteins within the cilium is essential for ciliary structure and function (Hua and Ferland, 2018, Wheway et al., 2018).

The TZ functions as a ciliary gate between the basal body and the ciliary compartment and thus is essential for regulating the IFT (Anand and Khanna, 2012, Jensen et al., 2018). Dysfunction of TZ proteins and/or IFT may affect signalling pathways in which the primary cilium is involved (Cheng et al., 2012, Wheway et al., 2018). Genes such as *ARL3* (Alkanderi et al., 2018) and *IFT172* (Halbritter et al., 2013a, Pruski et al., 2019) are associated with the correct delivery and transport of cargo into the cilium respectively.

Moreover, there are candidate genes that could be associated with JBTS and genes that are not clearly associated with JBTS yet. For instance, biallelic variants in each of the following genes have been found in only one family with JBTS: *CELSR2* (Vilboux et al., 2017b), *CEP164* (Chaki et al., 2012), *CLUAP1* (Johnston et al., 2017), *EXOC8* (Dixon-Salazar et al., 2012) and *POC1B* (Beck et al., 2014).

Biallelic mutations in other genes such as *DYNC2H1* (El Hokayem et al., 2012) and *IFT140* (Schmidts et al., 2013) that encode for proteins also involved in the IFT transport, are not associated with JBTS, however they are associated with skeletal ciliopathies. Of note, biallelic *DYNC2H1* (Vig et al., 2020) and *IFT140* (Hull et al., 2016) variants are also associated with non-syndromic retinal dystrophy while monoallelic *IFT140* variants have been recently associated with ADPKD (Senum et al., 2022).

1.4.6. Endoplasmic reticulum pathway and Asparagine-Linked Glycosylation (ALG) proteins

The genes associated to renal ciliopathies encode for proteins functioning at diverse cellular compartments (Figure 1.11A). Some of these genes associated with the ADPKD phenotypic spectrum such as *DNAJB11* (Cornec-Le Gall et al., 2018a, Huynh et al., 2020), *GANAB* (Porath et al., 2016, Besse et al., 2017), *ALG5* (Lemoine et al., 2022) *ALG8* (Besse et al., 2017) and *ALG9* (Besse et al., 2019) function at the endoplasmic reticulum, more precisely, the ER protein glycosylation pathway (Figure 1.11B).

The ER protein glycosylation pathway is important for the correct folding, trafficking and maturation of glycoproteins (Besse et al., 2017, Xu and Ng, 2015). The genes mentioned above have been associated with phenotypes associated with polycystic renal and liver phenotypes. There are other genes involved in this pathway that have been associated with ADPLD such as *SEC63*, and *SEC61B* and *PRKCSH* (Li et al., 2003). These genes encode proteins with similar functions, such as it is the case of *GANAB* and *PRKCSH* which encode the α and β subunits of the Glucosidase II respectively (Barroso-Gil et al., 2021b) (Figure 1.11B).

Mutations in these genes can lead to a clinical spectrum that includes autosomal dominant tubulointerstitial kidney disease (ADTKD) and autosomal dominant polycystic liver disease (ADPLD) (Olinger et al., 2020, Besse et al., 2017). It is important to mention that 7-10% of the ADPKD patients remain genetically unsolved, there is not a genetic cause found to be responsible for the ADPKD phenotype of these patients (Lemoine et al., 2022). As the result of its age-dependent penetrance and the phenotypic and genetic overlap of ADPKD, for some clinical cases it is often challenging to find robust phenotype-genotype associations and a unified nomenclature and disease associations between ADPKD, ADPLD and ADTKD.

Other renal diseases are ARPKD and NPHP. These are renal ciliopathies, like ADPKD, however, ARPKD and NPHP are recessive renal ciliopathies.

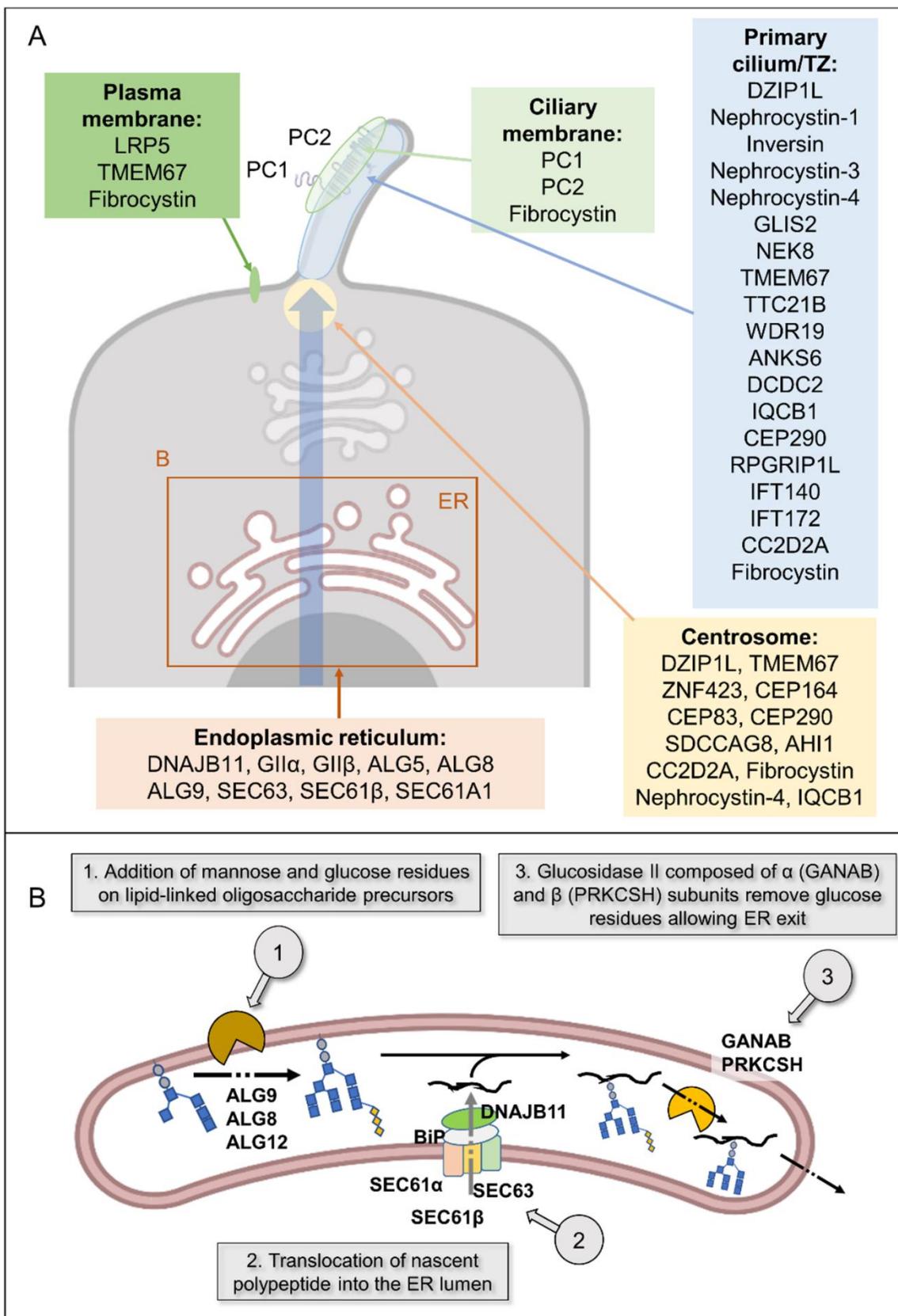


Figure 1.11. Cellular compartments associated with renal ciliopathies with a focus on the endoplasmic reticulum (ER). A: Genes associated with ADPKD, ARPKD and NPHP and the representation of the cellular location of the proteins coded by these genes. Some of these proteins can be found in more than one cell compartment (e.g. CEP290). B: Zoom into the ER protein glycosylation pathway. Figure generated using Biorender and modified from (Barroso-Gil et al., 2021b).

1.5. Current and future perspectives for the study of primary ciliopathies

Current research priorities in primary ciliopathies are: to increase genetic diagnosis yield in patients with primary ciliopathies (Best et al., 2022a, Perea-Romero et al., 2021, De La Vega et al., 2021, Umlai et al., 2022), to find novel ciliopathy genes and genotype-phenotype correlations (Morbidoni et al., 2021, Johnston et al., 2017, Besse et al., 2019), to use cell and animal models to perform *in vitro* and *in vivo* functional analyses (Latour et al., 2020, Bondué et al., 2021) and to develop and validate novel therapies to treat primary ciliopathies (Ramsbottom et al., 2018, Garcia et al., 2022, Yanda et al., 2018, Dulla et al., 2021, Corral-Serrano et al., 2023). The approaches to cover each of these priorities often overlap and complement each other.

The number of genes associated with primary ciliopathies, the number of novel genotype-phenotype correlations and the number of patients with whole genome sequencing (WGS) and whole exome sequencing (WES) available to be analysed has increased in the last few years (Nawaz et al., 2023, Asada et al., 2021). The development of projects such as the Genomics England 100,000 Genomes project in UK reflect the interest of using WGS and WES towards the understanding of the genetics of primary ciliopathies and other rare diseases (Best et al., 2022b, Turro et al., 2020, Wheway et al., 2019). The increase in the amount of sequencing data available, brings the possibility of performing *in silico* analyses focussed on establishing novel genotype-phenotypic correlations, finding novel ciliopathy genes and establishing pipelines for prioritizing putative disease-causing variants (de la Fuente et al., 2023, Best et al., 2022a, Best et al., 2022b).

Variants described as loss of function (such as nonsense or frameshift variants) in genes already associated with ciliopathy phenotypes often require less amount of time and analytical, filtering and validation workload in order to be detected and be classified as disease-causing than genetic variants in candidate genes or missense variants (Best et al., 2021, Umlai et al., 2022, Wheway et al., 2019). Often the second allele (or both alleles) in recessive primary ciliopathies cannot be found because it is located in a deep intronic genomic region, it is hypomorphic, or both variants are located in a gene which has not been associated to any primary ciliopathy or other rare disease. Consequently, it is estimated that 30-50% of patients with primary ciliopathies remain genetically unsolved (Graessner et al., 2021, Smedley et al., 2021). To increase genetic diagnostic yield, some studies have

focussed on designing pipelines with the aim to detect disease-causing variants that are deep intronic variants or synonymous variants with a splicing effect and structural variants, including copy number variants (CNVs) (Best et al., 2022a, Best et al., 2022b, de la Fuente et al., 2023, Vig et al., 2020, Molinari et al., 2020, Olinger et al., 2021).

Additionally, there are multiple studies using functional analyses and *in vivo* and *ex vivo* ciliopathy models, to complement *in silico* studies and investigate molecular disease mechanisms of primary ciliopathies. Mouse (Higgins et al., 2022, Liu et al., 2017, Bashford and Subramanian, 2019) and zebrafish (Trejo-Reveles et al., 2023, Latour et al., 2020, Garcia et al., 2022, Molinari et al., 2018b) are widely used in the study of primary ciliopathies, but also other model organisms such as *Caenorhabditis elegans* (Zhang et al., 2022) and *Xenopus laevis* (Johnston et al., 2017) have been utilised in studies related with the molecular pathogenesis of primary ciliopathies (Delvallée and Dollfus, 2023).

Some examples of these model organisms are conditional gene knockdown mouse models, used to study specific disease phenotypes within the phenotypic spectrum of JBTS and JATD, caused by mutations in ciliopathy genes such as *Cep120* (Wu et al., 2014), *Talpid3* (Bashford and Subramanian, 2019), and a mouse model to study the role of *Barttin* as a modifier gene for Cep290-JBTS disease (Ramsbottom et al., 2020). Zebrafish models have been used to model disease phenotypes associated with some JBTS genes, such as *CEP120* (Shaheen et al., 2015) and *TOGARAM1*, a gene recently described as a JBTS gene (Latour et al., 2020).

There are several studies that use cell models and organoids to find the molecular pathogenesis of genes and disease-causing mutations. Organoids have been recently used as *ex vivo* models for retinal (Georgiou et al., 2022, Corral-Serrano et al., 2023), renal (Zeng et al., 2021) and cerebral disease phenotypes (Zhang et al., 2019). Other cell models use induced pluripotent stem cells (iPSCs) (Delvallée and Dollfus, 2023, Georgiou et al., 2022, Vig et al., 2020), hTERT-RPE1 cells (Latour et al., 2020) and primary cells (Devane et al., 2022, Ziegler et al., 2022, Shaheen et al., 2015). Primary cells from patients are often difficult to obtain and some cell types such as primary skin fibroblasts require to apply a skin biopsy procedure to the patient. In contrast, human urine-derived renal epithelial cells (hURECs) can be obtained by processing an urine sample according to standardised protocols and

offer the possibility of obtaining primary cells from patients, without performing an invasive procedure (Ajzenberg et al., 2015, Ziegler et al., 2022).

There is an increasing interest in non-ciliary functions of ciliopathy genes (Vertii et al., 2015) and describing in detail disease molecular mechanisms caused by mutations in genes encoding proteins that are not located within the primary cilium (Hua and Ferland, 2018), such as proteins functioning in the protein glycosylation pathway at the endoplasmic reticulum (Lemoine et al., 2022, Besse et al., 2017).

The molecular disease pathogenesis in primary ciliopathies often involves the dysfunction in one or more than one cell signalling pathways, including: Hedgehog (Hh), Wnt, Notch signalling, Hippo, GPCR, PDGF, mTOR, TGF-beta and cAMP signalling pathways (Devlin and Sayer, 2019, Wheway et al., 2018).

There is an interest in investigating the molecular disease pathogenesis of primary ciliopathies with a focus on specific cell signalling pathways, using *in vitro* and *in vivo* disease models. There are multiple studies investigating the role of cell signalling pathways in primary ciliopathies, some of those studies have used cell lines such as mouse NIH-3T3 fibroblasts to find gene candidates that affect Hh signalling (Breslow et al., 2018), others have used mouse models to investigate the disease pathogenesis caused by the knockdown of specific ciliopathy genes. For example, it has been studied the effect of *Cep120* or *Tmem67* knockdown on Hh signaling (Betleja et al., 2018, Wu et al., 2014) and Wnt signaling pathways (Abdelhamed et al., 2019), respectively, that could explain phenotypes within the JBTS and JATD phenotypic spectrum. Mouse models have been used to study the role of cAMP signalling pathway in the development of renal cysts leading to PKD (Yanda et al., 2018). Other studies have investigated the activation of ROCK signalling associated with ADPKD using renal epithelial cells (Streets et al., 2020).

There is an increasing number of studies who complement and validate *in silico* approaches using cell models of disease. These cell models are based on the use of cell lines (Betleja et al., 2018, Tsai et al., 2019), primary fibroblasts (Shaheen et al., 2015) or hURECs (Bondué et al., 2021, Devane et al., 2022).

In particular, Sayer and Miles Laboratory at Newcastle University, have focussed on integrating *in silico* analyses with *in vitro* models using primary cells to validate the deleterious effect of putative disease-causing variants, including synonymous variants and splicing variants in ciliopathy genes (Molinari et al., 2020, Molinari et al.,

2018a), the use of renal spheroids (Hynes et al., 2014) and developing *in vivo* mouse models to investigate potential exon skipping therapies (Ramsbottom et al., 2018) and a gene modifier (Ramsbottom et al., 2020).

This interaction between *in silico*, *in vitro* and *in vivo* disease models is used towards the understanding the molecular genetics of primary ciliopathies and is also important for the development of potential therapies, focussed on targeting cell signalling pathways (Hynes et al., 2014, Molinari et al., 2019, Stokman et al., 2021, Devlin et al., 2023), and the use of adeno-associated virus (AAV) (Jacobson et al., 2022) and antisense oligonucleotides (ASOs) (Dulla et al., 2021, Evers et al., 2015, Garanto et al., 2016, Bondu et al., 2022).

There is an increasing interest on the development of ASOs-based therapies and other RNA-based therapies with the objective of bypassing at cellular level, the deleterious effect of mutations located in certain ciliopathy genes (Drivas et al., 2015), and treating some of the clinical phenotypes within the phenotypic spectrum of primary ciliopathies (Bondu et al., 2022, Evers et al., 2015, Jacobson et al., 2022).

Other recent examples contributing to precision and translational medicine, include the study of use of agonists of prostaglandin E2 receptors to treat NPHP and its validation using patient specific immortalised renal urine derived cells and complemented with *in vivo* zebrafish and mouse models (Garcia et al., 2022) and the study of Eupatilin (a flavonoid, which is a compound metabolised in plants) as a therapeutic component, suggested to partially restore the ciliary disease phenotype in (*CEP290*-LCA) patient fibroblasts and model KO *CEP290* RPE1 cell lines and retinal ciliopathy organoids (Corral-Serrano et al., 2023).

Furthermore there is an interest in the use of the gene editing tool CRISPR/Cas9 system to facilitate the study of the undelaying molecular disease mechanisms in ciliopathies and other rare diseases, with a focus cell signalling pathways (Breslow et al., 2018, Pusapati et al., 2018) and gene editing, to create cell and animal models to investigate diseases-causing mutations and ciliopathy genes (Trejo-Reveles et al., 2023, Liu et al., 2017, Cardenas-Rodriguez et al., 2021, Tsai et al., 2019).

1.5.1. Human urine-derived renal epithelial cells (hURECs)

Human urine-derived renal epithelial cells (hURECs) have been widely used in the recent years as a valuable non-invasive source of primary, non-transformed cells to

study molecular and cellular disease mechanism of renal diseases, including renal ciliopathies (Ziegler et al., 2022, Bondu et al., 2021, Molinari et al., 2020, Cardenas-Rodriguez et al., 2021). These renal epithelial cells are suitable to study epithelial cell function of patients with inherited kidney diseases, as the culture of hURECs, allows obtaining primary cells from the kidney (Ziegler et al., 2022).

Moreover, as hURECs are obtained from a urine sample, it is a non-invasive alternative to other sources of primary cells such as a skin biopsy to obtain primary fibroblasts. As the hURECs have primary cilia, they are suitable to study ciliopathies (Oud et al., 2018). They have reproducible epithelial properties characteristics (Ziegler et al., 2022) and can be used cultured to perform a wide variety of *in vitro* experiments such as gene expression studies or immunofluorescence to visualise the primary cilia (Srivastava et al., 2017b) (Figure 1.12).

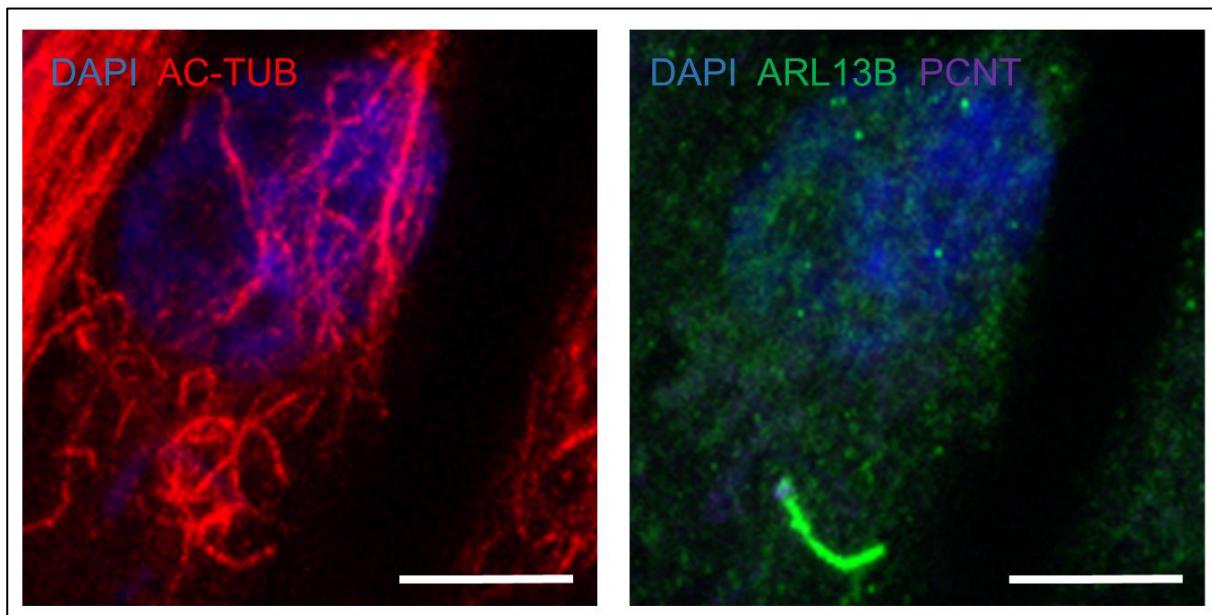


Figure 1.12. Representative example of immunofluorescence microscopy imaging of human urine-derived renal epithelial cells (hURECs) from a healthy control individual. The antibodies used are: a ciliary membrane marker: ARL13B (green), a centrosome marker: PCNT (purple) and a marker for acetylated microtubules AC-TUB (red). DAPI (4, 6-diamidino-2-phenylindole), a nuclear marker, is shown in blue. hURECs were serum starved for 48 hours. Nikon (A1) confocal inverted microscopy was used for visualisation. For the ease of visualisation of the antibodies: DAPI and AC-TUB are shown on the left image, while DAPI, ARL13B and PCNT are shown on the right image. AC-TUB, Acetylated Tubulin; PCNT, Pericentrin. Scale bar: 10 μ m.

The culture of hURECs can be used to study the effect of patient-specific mutations (Srivastava et al., 2017b, Ramsbottom et al., 2018, Molinari et al., 2018a, Oud et al., 2018, Devane et al., 2022).

Some examples of these *in vitro* experiments are the characterization of the ciliary phenotype of a patient with Mainzer–Saldino syndrome, a skeletal ciliopathy that can cause renal failure, presenting with biallelic pathogenic variants in *IFT140* (Oud et al., 2018).

Patient-specific hURECs have been used to confirm defective splicing, in particular to validate the pathogenicity of a synonymous *PKHD1* allele leading to a ARPKD (in a patient which is compound heterozygous of that synonymous allele and a missense *PKHD1* allele) (Molinari et al., 2020). These *in vitro* experiments included RNA studies to detect different splicing isoforms as a result of these mutations, and the characterisation of the ciliary phenotype (in terms of percentage of ciliated cells and cilia length) in the hURECs of this patient with ARPKD (Molinari et al., 2020).

1.5.2. Next generation sequencing and the Genomics England 100,000 Genomes Project

In the last 20 years, advances in next-generation genomic sequencing (NGS) technology has led to an enormous increase in the number of genomes sequenced, the availability of incorporating these genome sequencing technologies into research and the increasing development of the field of -omics data analysis (Bentley et al., 2008, Phillips et al., 2021, Ganini et al., 2021).

In particular generation of WGS and WES data has been increasing since the sequencing of the first human genome. The sequencing of the first human genome was completed in 2003 after the work of multiple laboratories around the world for more than a decade included in the International Human Genome Sequencing Consortium and Celera Genomics (Lander et al., 2001, Waterston et al., 2002, Venter et al., 2001). The remaining gaps in the reference human genome sequence were recently sequenced by the Telomere-to-Telomere (T2T) Consortium (Nurk et al., 2022). Nowadays genome sequencing is significantly more affordable in time and price required per genome than 20 years ago. This has leaded an increase in the number of genomes sequenced and projects aimed to incorporate genome sequencing as a public service funded by national health systems or privately (Phillips et al., 2021, Manolio et al., 2019).

In the last few years various projects have been completed and others are currently on-going or emerging, aimed to recruit more patients and perform WGS/WES. Some examples are the deCODE project in Iceland (Gudbjartsson et al., 2015) or the 1000 Genomes (Auton et al., 2015) and TwinsUK (Moayyeri et al., 2013) in the United Kingdom (UK). Of note, the 1000 Genomes in UK, aimed to study human genetic variation, was completed in 2015 (Auton et al., 2015).

More recently, data from other projects are available such as the UK Biobank (Bycroft et al., 2018) in UK or the All of Us in US (Ramirez et al., 2022), which do not only contain genomic data, but also detailed biological and clinical data are included.

Furthermore, in 2019, the NHS Genomic Medicine Service (NHS GMS) was launched by Genomics England, which is a company owned by the Department of Health of UK. This program aims to promote genomic medicine and incorporate WGS into routinely health care (Barwell et al., 2019).

The NHS Genomic Medicine Service emerged from and as a continuation of the Genomics England 100,000 Genomes project (Turnbull et al., 2018, Turro et al., 2020). Genomics England 100,000 Genomes Project is managed by Genomics England and funded by the National Institute for Health Research and NHS England. Since the project was announced in 2012 and the first participants were recruited in 2014, more than 100,000 genomes have been sequenced (Smedley et al., 2021, Turnbull et al., 2018).

The improvements and development in genome sequencing and genomic medicine have been reflected as the Genomics England 100,000 Genomes Project has achieved its goal of sequencing 100,000 genomes (Smedley et al., 2021, Turnbull et al., 2018, Turro et al., 2020). Currently the number of genomes sequenced in the Genomics England 100,000 Genomes Project (by October 2022, data release v16) is 106,266 genomes (accessed in: <https://research-help.genomicsengland.co.uk/pages/viewpage.action?pagId=72352341>). Similarly, in the last two years, the European commission launched the Beyond One Million Genomes (B1MG) project, which is an initiative to sequence 1 million genomes in all Europe to create a network and large database of genomic and health data in Europe (Saunders et al., 2019).

The Genomics England 100,000 Genomes Project, similarly as the projects mentioned above, aims to investigate the genetic causes of human disease, in

particular: cancer (Prendergast et al., 2020, Degasperi et al., 2022) and rare diseases (Smedley et al., 2021, Turro et al., 2020).

Rare diseases are commonly defined as diseases affecting to fewer than 1 in 2000 people (Haendel et al., 2020). The estimates on the number of different rare disease varies from 6000 to 10000 different rare diseases (Haendel et al., 2020, Turro et al., 2020). Rare diseases are heterogeneous and their prevalence can be variable depending on the geographic region, type of disease and type data considered among other factors (Nguengang Wakap et al., 2020). The global prevalence of rare diseases has been estimated to be about 3.5-5.9% of the global population (Nguengang Wakap et al., 2020). It is estimated that there may be 300 million people globally presenting a rare disease, 3.5 million in the UK, 30 million in EU and 25 million in US (Spencer-Tansley et al., 2022, Nguengang Wakap et al., 2020, Tisdale et al., 2021).

It is estimated that about 80% of these diseases have a genetic component (Haendel et al., 2020). The yields of genetic diagnosis in rare disease is highly variable depending on the molecular genetic basis of the disease and other factors such as mode of inheritance of disease, availability if the genome sequencing of other family members etc (Smedley et al., 2021, Cacheiro et al., 2022). Besides, there are about 200 different diseases added in the Online Mendelian Inheritance in Man (OMIM) database each year (Haendel et al., 2020) and gene panels and clinical data are updated constantly, as more research and data analysis are occurring (Blakes et al., 2022, Smedley et al., 2021).

The Genomics England 100,000 Genomes Project has the potential to provide a genetic diagnosis to patients of a large cohort of patients with renal ciliopathies and a wide variety of rare diseases (Smedley et al., 2021, Wheway et al., 2019).

The pilot study of the Genomics England 100,000 Genomes Project focused on rare diseases, in which 4660 participants from 2183 families were involved, provided genetic diagnoses in 25% of the probands selected for this preliminary study (Smedley et al., 2021). Considering the yields of genetic diagnosis in other recent genomic projects with similarities with the Genomics England 100,000 Genomes Project: one recent study analysed genetic data from a cohort of patients with cystic kidney disease in which genetic diagnosis was provided for 307 patients. In that study, 24% and 7% of those 304 patients were genetically solved with *PKD1* and

PKD2 variants respectively (Groopman et al., 2019). Another study analysed genetic data from a cohort of 3951 unrelated families with inherited retinal dystrophies, also including rare diseases and ciliopathies such as Usher syndrome, BBS and Alström syndrome, providing a genetic diagnosis for 53.2% (2100/3951) of these families (Perea-Romero et al., 2021).

Regarding the participants presenting with a rare disease recruited in the Genomics England 100,000 Genomes Project, about 1% of them are diagnosed with a ciliopathy. Additionally 39% of the participant diagnosed with a rare disease are diagnosed with neurological and neurodevelopmental disorders and about 11% are diagnosed with renal and urinary tract disorders (Wheway et al., 2019). This makes the Genomics England 100,000 Genomes Project a valuable resource to study rare diseases, including the genetic and phenotypically heterogeneous group of ciliopathies (Best et al., 2021, Orr et al., 2023, Devlin et al., 2022, Geraghty et al., 2023).

The availability of these data has allowed diverse studies using a wide diverse of approaches such as genome-wide association studies (GWAS) to investigate susceptibility for posterior urethral valves (PUV), a common cause of end-stage renal disease (ESKD) (Chan et al., 2022) to the systematic analysis of splicing variants to provide a new genetic diagnosis and study the role of non-canonical splicing variants (Blakes et al., 2022).

Some patients recruited within the rare disease program of the Genomics England 100,000 Genomes Project have been genetically solved. These patients presented a diverse range of rare diseases such as Joubert syndrome, Poretti–Boltshauser syndrome, ADPKD, motile ciliopathy-like syndrome including non-cystic fibrosis (CF) bronchiectasis or complex combinations phenotypes such as liver, kidney, and heart degeneration. Some of these patients, are patients who carried disease-causing variants in *IFT140* (Senum et al., 2022), *LAMA1* (Powell et al., 2021), *ALG5* (Lemoine et al., 2022), *TULP3* (Devane et al., 2022), *DNAJB11* (Huynh et al., 2020), *CEP164* (Devlin et al., 2022) or *TOGARAM1* (Latour et al., 2020). Novel disease-causing variants in *CEP164* (Devlin et al., 2022), *IFT140* (Senum et al., 2022) and *TOGARAM1* (Latour et al., 2020) provided novel genotype-phenotype associations and expanded the genetic and phenotypic spectrum associated with these ciliopathy genes.

One of the tools that has helped to provide these genetic diagnoses, is the Genomics England variant prioritization (tiering) pipeline (I will refer to it from now as: tiering data). The tiering data provides, for each of the genomes submitted to the tiering pipeline, a list of variants that are rare in allele frequency in control populations and classifies these variants as Tier 1, Tier 2 or Tier 3 depending on the gene in which each of the variants are located and if the corresponding gene has been associated with the patient's phenotype (Best et al., 2022a, Brown et al., 2022, Smedley et al., 2021).

The tiering prioritization pipeline, per participant's genome, consists in a filtering process of genetic variants and variant classification based on one (or more than one) gene panel applied. The corresponding lists of genes and gene panels are found in PanelApp (Martin et al., 2019). If the variant is located in a gene described as "green gene", the variant can be classified as Tier 1 or Tier 2, depending if it is loss of function or not, respectively. In theory, Tier 1 and Tier 2 are located in genes that are associated with the corresponding disease of the panel, while Tier 3 are not. The tiering pipeline also includes filters of allele frequency and coding impact and it also considers the segregation of the variant with the disease and its mode of inheritance (Smedley et al., 2021, Martin et al., 2019).

Some limitations of the tiering data is that it is mostly focused on exonic variants and splicing donor and splicing acceptor variants. Subsequently, some disease-causing variants are classified as Tier 3 (because of the corresponding gene panel not being updated or because the correct gene panel not being applied). Additionally, some disease-causing variants may be untiered (as some intronic and structural variants can be filter out by the tiering prioritization pipeline), and they may be missed by the genetic variant prioritization pipeline and gene discovery approaches (Powell et al., 2021, Best et al., 2022a).

Recent approaches have genetically solved patients with ciliopathies by a reverse phenotypic approach using a combination of analysis to extract different types of variants, including exonic variants, variants affecting splicing and copy number variation (CNV) and structural variants (SVs), filtered by allele frequency and canonical transcript to genetically solve patients with disease-causing variants in a subset of ciliary genes (Best et al., 2022b).

Additionally, the combination of using, not only the tiering data, but also detecting other untiered pathogenic variants, have been proven to be effective in providing genetic diagnosis of patients diagnosed with ciliopathy phenotypes, detecting disease-causing SVs and other untiered variants (Best et al., 2021). For instance, detecting a disease-causing SV variant in *BBS1* in a BBS patient, which is compound heterozygous for this variant and an untiered missense *BBS1* variant (Best et al., 2021). Furthermore, an untiered pathogenic homozygous synonymous variant in *NPHP3* leading to splicing defects was found in two probands with NPHP with congenital hepatic fibrosis phenotypes (Olinger et al., 2021).

1.5.3. Antisense oligonucleotide mediated exon skipping as an emerging therapy in ciliopathies

Exon skipping therapies are based on removing exons that are dispensable and do not affect protein function. These therapies can consist on the use of antisense oligonucleotides (ASOs) which they have been proven effective in *in vitro* and *in vivo* assays (Aartsma-Rus and van Ommen, 2007, Bennett and Swayze, 2010).

One of the most well-known examples for the success of exon skipping therapies is the treatment of patients with Duchenne muscular dystrophy (DMD, MIM PS310200). This treatment consisted in targeting specific exons of the gene DMD (MIM 300377). DMD is the gene responsible for Duchenne muscular dystrophy and encodes the large protein called dystrophin (Kole and Krieg, 2015, Komaki et al., 2018).

The canonical transcript of DMD (NM_004006.3) has 79 coding exons, that code for the 3685 aa of the protein dystrophin. The large number of exons and protein length increase the possibility of the presence exons that can be skipped without disrupting the reading frame of protein. These therapies on DMD consisted in targeted exon skipping, targeting a dispensable exon in which a specific disease-causing variant is located, restoring the reading frame and obtaining a partially functional dystrophin protein (Aartsma-Rus and van Ommen, 2007, Kole and Krieg, 2015, Komaki et al., 2018, Lee et al., 2018, Servais et al., 2015).

CEP290, a ciliary gene associated with a wide spectrum of ciliary phenotypes, including: JBTS, phenotypes affecting only one organ such as LCA (primary ciliopathy characterized by retinal dystrophy) and severe phenotypes affecting multiple organs such as MKS. *CEP290* is a gene encoding for a large protein

CEP290. The canonical transcript of *CEP290* (NM_025114.4) has 53 coding exons, encoding for the 2479 aa of the protein CEP290.

In vitro assays have revealed that exon skipping is tolerated in *CEP290* and can restore transcript levels of functional CEP290. In particular, skipping *CEP290* exon 36 (using ASOs) in fibroblasts from patients with LCA, can bypass the deleterious effect of the *CEP290* variant located in that exon: c.4723A>T; p.(Lys1575Ter) (Barny et al., 2019).

The deleterious effect of another *CEP290* variant: c.2991+1655A>G leading to LCA phenotypes has effectively been targeted *in vitro*. This intronic *CEP290* variant creates an aberrant pseudoexon as a result of creating a cryptic splice site. This new cryptic splice site leads to the insertion of a premature stop codon in the *CEP290* transcript containing the aberrant pseudoexon (Garanto et al., 2016). The deleterious effect of this *CEP290* variant was rescued using replication-defective recombinant adeno-associated viruses (AAVs) mediated delivery of ASOs in LCA fibroblast cell lines from two unrelated patients with LCA carrying the c.2991+1655A>G mutation. These treatments targeted and skipped the aberrant pseudoexon, restoring the transcript and protein levels as well as restoring the ciliary phenotype, in terms of percentage of ciliated cells and cilium length (Garanto et al., 2016).

Similarly, the ciliary phenotype was characterized in a JBTS patient presenting a homozygous nonsense *CEP290* mutation: (c.5668G>T; p.(Gly1890Ter)) in exon 41 (patient also described as “JBTS-AA”). It has been shown that this exon can be skipped without leading to the disruption of the protein reading frame via an exon skipping therapy using an ASO oligo morpholino. The skipping of the corresponding exon in which the nonsense mutation is located was assayed *in vitro*, leading to the rescue of the ciliary phenotype of the JBTS patient’s hURECs (in terms of cilia length) (Molinari et al., 2019, Ramsbottom et al., 2018).

ASO-mediated exon skipping therapies for *CEP290* variants have been assayed *in vivo*. For example, in *CEP290* defective mouse LCA splicing can be corrected via ASO-mediated exon therapy (Garanto et al., 2016). Moreover in a humanized (of the patient “JBTS-AA”) mouse JBTS model, the ciliary and renal phenotype (cystic index) was partially rescued by restoring functional CEP290 levels via an ASO-mediated exon skipping therapy (Ramsbottom et al., 2018).

Clinical trials of exon skipping therapies for patients with Duchenne muscular dystrophy are more advanced compared to therapies for patients with ciliopathies (Kole and Krieg, 2015). In human clinical trials, LCA patients have been treated with intravitreal injections of ASOs and its administration improved their visual acuity (Cideciyan et al., 2019). RNA-based treatments for kidney diseases have a promising future ahead (Bondué et al., 2022), however kidney molecular therapies are still lagged behind the delivery of therapies targeted for other organs such as eye and liver (Rubin and Barry, 2020).

In the last decade multiple studies using hURECs have been proven effective as a cell renal model (Srivastava et al., 2017b, Ziegler et al., 2022) and several studies have found molecular targets that are suitable for a targeted exon skipping therapies (Molinari et al., 2019, Ramsbottom et al., 2018). The availability and suitability of hURECs can aid studies that investigate the genetic and phenotypic heterogeneity of ciliopathies (Molinari et al., 2018a, Oud et al., 2016), in combination with mouse studies (Ramsbottom et al., 2018) and the widely available and vast genomic data available in large databases such as the Genomics England 100,000 Genomes Project (Best et al., 2022b, Devane et al., 2022, Wheway et al., 2019).

1.6. Aims of the project

Contribute to the understanding of the complexity of the molecular disease pathogenesis of renal ciliopathies. To use a range of cell-based assays and the data available in the Genomics England 100,000 Genomes Project to investigate the molecular genetics of renal ciliopathy genes and to identify underlying disease mechanisms across the disease spectrum.

Specific aims:

1. Expand the current knowledge regarding the developmental spatial expression of ciliopathy genes and the role of tissue expressivity in the phenotypic heterogeneity and overlap in primary ciliopathies. Characterise the expression of two rare recessive ciliopathy genes, *CEP120* and *ARL3*, during human development using RNAscope and the MRC Wellcome Trust Human Developmental Biology Resource (HDBR) at Newcastle University.
2. Contribute to the understanding of the molecular basis behind the genetic pleiotropy of ciliopathy genes and contribute to the design of gene therapies targeting ciliopathy genes to treat patients with renal ciliopathies. This aim will

- be focused on investigating and comparing *in silico* the potential of two primary ciliopathy genes with overlapping phenotypes: *CEP120* and *CC2D2A*, as targets for an antisense-oligonucleotide mediated exon skipping therapy.
3. Increase genetic diagnostic yield and percentage of genetically solved patients with renal ciliopathies. To do this I will use whole genome sequencing (WGS) data within the Genomics England 100,000 Genomes Rare Diseases Project to investigate the molecular genetics of renal ciliopathies and explore genotype-phenotype correlations.
 4. Explore the potential and applicability of functional assays to investigate the molecular pathogenesis of renal ciliopathies and combine *in silico* and *in vitro* approaches to complement genetic studies, with a focus on: patient-specific mutations and primary cells derived from patients. To do this I will use human urine-derived renal epithelial cells (hURECs) as an *ex vivo* kidney cell model of human kidney disease to complement and support findings from the Genomics England 100,000 Genomes Project.

Chapter 2. Materials and Methods

2.1. Software tools and online resources

The software tools and online resources utilised are shown in Tables 2.1 and 2.2. In these tables, their corresponding chapter/s in which each of these tools and resources were used, is also indicated.

Description	Relevant chapter/s in which it was used
Adobe Illustrator 2021 v.25.0	Chapter 3
Adobe Photoshop 2020	Chapter 3
ApE – A plasmid Editor v2.0.61	Chapters 4 and 5
AxioVs40x64 v.4.9.1.0 (for Zeiss AxioVert)	Chapters 4 and 5
EndNote X9 19.2.0.13018	All chapters
Genomics England Research Environment	Chapters 4 and 5
GloMax Multi Detection+ System (Promega)	Chapter 4
GraphPad Prism v 9.1.1 (225)	Chapters 4 and 5
HDBR image server (Leica Biosystems)	Chapter 3
ImageJ 1.52p	Chapters 4 and 5
Image Studio Ver 5.2.lnk (for Odyssey CLx Western Blot reader)	Chapter 4
Integrated Genome Viewer v.2.10.3	Chapter 4
QuantStudio Real-Time PCR Software 1.3	Chapters 4 and 5
Microsoft office - Microsoft 365 MSO	All chapters
NIS Elements 3.20 and 4.20	Chapters 3, 4 and 5
RStudio 2022.07.01 Build 554	Chapter 5
Skanlt Software for microplate readers v.6.1.1.7	Chapter 5
SnapGene Viewer 5.3.2	Chapter 4
VisionWorksLS Acquisition and Analysis v. 7.0.2.RC.1.0 gel reader	Chapters 4 and 5
ZEN 2.3 (for Zeiss Axioimager)	Chapter 4

Table 2.1. List of software tools used.

Description	Link	Relevant chapter/s in which it was used
BioRender	https://biorender.com/	Chapter 1
Ensembl	https://www.ensembl.org/index.html	All Chapters
Ensembl VEP	https://www.ensembl.org/info/docs/tools/vep/index.html	All Chapters
Flybase	https://flybase.org	Chapter 3
GE 100K Genomes Project scripts and workflows	https://research-help.genomicsengland.co.uk/display/GERE/7.+Analysis+Scripts+and+Workflows	Chapters 4 and 5
gnomAD	https://gnomad.broadinstitute.org/	Chapters 1, 4 and 5
GTEx	https://www.gtexportal.org/home/	Chapter 4
HDBR	https://www.hnbr.org/	Chapter 3
HGMD	http://www.hgmd.cf.ac.uk/ac/index.php	Chapters 4 and 5
NCBI ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/	All Chapters
NCBI Gene	https://www.ncbi.nlm.nih.gov/gene	All Chapters
NCBI Primer-BLAST	https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi	All Chapters
NCBI PubMed	https://pubmed.ncbi.nlm.nih.gov/	All Chapters
OMIM	https://omim.org/	All Chapters
Phytozome	https://phytozome.jgi.doe.gov/	Chapters 1 and 3
Primer3Plus	https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi	Chapters 4 and 5
ProteinPaint	https://pecan.stjude.cloud/proteinpaint	Chapter 4
SMART	http://smart.embl-heidelberg.de/	Chapter 4
RNAscope (ACDBio)	https://acdbio.com/manual-assays-rnascop	Chapter 3
The Human Protein Atlas	https://www.proteinatlas.org/	All Chapters
UCSC Genome Browser	https://genome.ucsc.edu/	Chapters 1, 4 and 5
Varsome	https://varsome.com/	Chapters 1, 4 and 5
Wormbase	https://www.wormbase.org/	Chapters 1 and 3

Table 2.2. List of online resources used.

2.2. Cell culture

The corresponding affected individuals (ADPKD and JBTS patients) and control individuals consented to this study. Ethical approval was obtained according to the UK National Research Ethics Service (NRES) Committee North East – Newcastle and North Tyneside 1 (08/H0906/28+5) and the NRES Committee North East (14/NE/1076). All methods were performed in accordance with the relevant ethical guidelines and regulations.

2.2.1. Cell culture of mouse Shh-LIGHT2 cells

Shh-LIGHT2 cells line derived from mouse NIH-3T3 fibroblasts, a Hedgehog (Hh)-responsive cell line originally created in Beachy laboratory (Johns Hopkins University School of Medicine) (Taipale et al., 2000) were utilised for in *vitro* experiments. Shh-LIGHT2 cells were cultured in DMEM medium (High Glucose, L-Glutamine Pyruvate) (Apollo scientific, BINC300) with 10% foetal bovine serum (FBS) and 1% Penicillin-Streptomycin (ThermoFisher; 15070063).

Geneticin (G418), a selection antibiotic, was added to the medium (250 µl of G418 per 50 ml of DMEM medium). G418 is only added during cell expansion (this antibiotic is used because these Shh-LIGHT2 cells were originally established by Taipale et al. (Prof Philip Beachy Laboratory, The Johns Hopkins University School of Medicine, USA) transfecting a pSV-Neo vector encoding G418 resistance with the Gli-luc reporter and TK-Renilla luciferase vectors) (Taipale et al., 2000). Before doing any further experiment (e.g. RNA extraction, serum starvation, SAG stimulation, dual luciferase assay) with the Shh-LIGHT2 cells, the G418 is not added to the DMEM medium, from two passages before these experiments are made.

2.2.2. Cell culture of human primary skin fibroblasts

Human primary fibroblasts were isolated from a skin sample from a control healthy individual according to standardised protocol (Molinari et al., 2019). Human primary fibroblasts were cultured in DMEM medium (High Glucose, L-Glutamine Pyruvate) (Apollo scientific, BINC300) with 10% FBS and 1% Penicillin-Streptomycin (ThermoFisher; 15070063).

2.2.3. Cell culture of Human Urine-Derived Renal Epithelial Cells (hURECs)

Human Urine-Derived Renal Epithelial Cells (hURECs) were isolated from urine from an ADPKD patient and various control healthy individuals according to standardised

protocol (Molinari et al., 2020, Ramsbottom et al., 2018, Ajzenberg et al., 2015, Devane et al., 2022, Srivastava et al., 2017b). This protocol consisted in processing the urine using a series of centrifugations and washes. Urine samples (50-110ml each) can be stored at 4°C for up to 3 hours. The main steps of the protocol are shown in Figure 2.1.

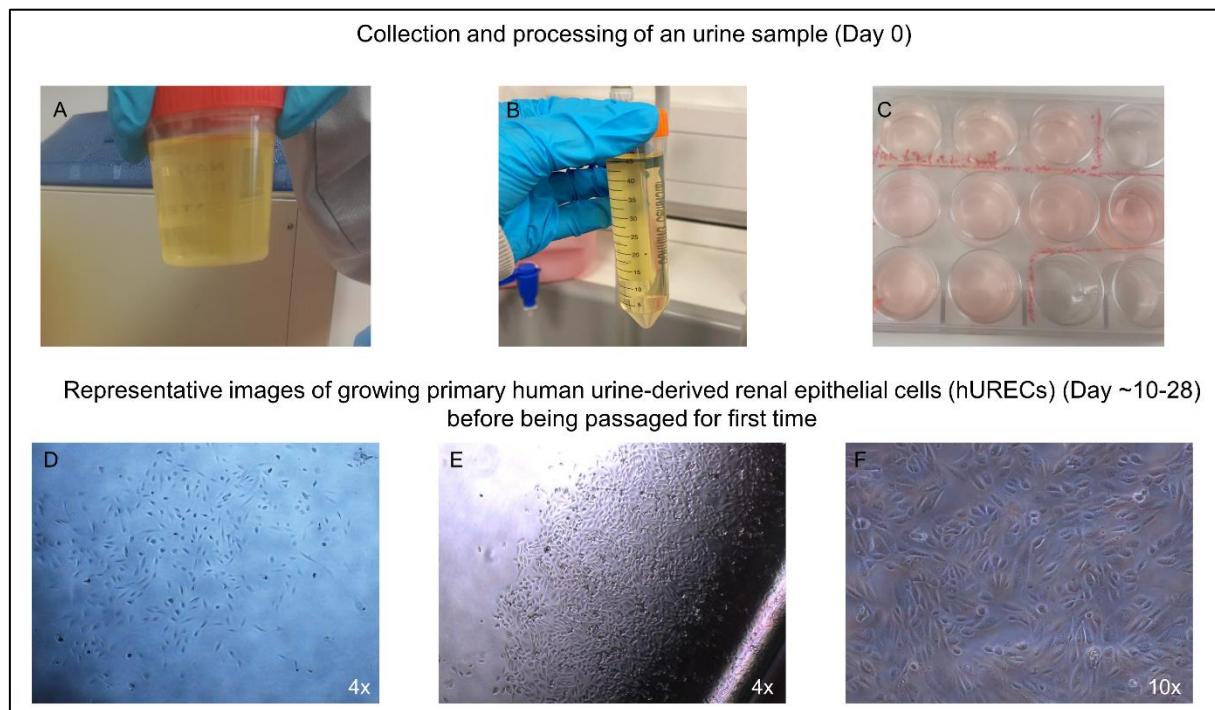


Figure 2.1. Representative pipeline of the collection, processing and culture of human urine-derived renal epithelial cells (hURECs). A: The container with 50-110ml of urine can be kept at cold temperature on ice or on the fridge (about 4°C). The sample should be processed in within 3 hours. B: Following the standard protocol (Ajzenberg et al., 2015, Molinari et al., 2020) the urine sample is distributed into 2 falcon tubes, followed by two centrifugations and one washing steps. C: The resulting cell pellet/s (from the same sample) are resuspended and seeded in primary media. Usually one well of a 12-well plate corresponds to the final processing of one urine sample. Usually the first hURECs are observed after the second week of being seeded (day 14-20). The time at which the hURECs are first observed is variable between different individuals and also within different samples from the same individual. D: Representative image of hURECs from a healthy individual (Control A) after 18 days from urine processing. E and F: Representative images of hURECs from a healthy individual after 11 days from urine processing. The growth of the hURECs' first colony is usually followed by the growth of other colonies that will eventually combine and cover most of the surface of the well. The hURECs observed in D, E and F are from the same individual. In D: Brightness and contrast were increased from original image. Nikon Eclipse TS100 inverted microscope was used for visualisation. The magnification objectives used were Nikon 4x/0.13 and 10x/0.25.

Solutions and Media:

- Washing buffer: Dulbecco's phosphate-buffered saline (DPBS) (Fisher Scientific; 15326239), supplemented with Penicillin-Streptomycin (140 U/ml) (ThermoFisher; 15070063) and Amphotericin B (1 µg/ml) (Sigma-Aldrich; A2942).
- Primary medium: DMEM/F-12 (1:1) (1X) (Gibco; 11320-074) supplemented with Penicillin-Streptomycin (140 U/ml) (ThermoFisher; 15070063) and Amphotericin B (3.5 µg/ml) (Sigma-Aldrich; A2942) and REGM Renal Epithelial Cell Growth Medium BulletKit (Lonza; CC-3190), which includes FBS.
- Proliferation medium: REGM Renal Epithelial Cell Basal Medium (Gibco; CC-3191) supplemented with Penicillin-Streptomycin (140 U/ml) (ThermoFisher; 15070063) and Amphotericin B (3.5 µg/ml) (Sigma-Aldrich; A2942) and REGM Renal Epithelial Cell Growth Medium BulletKit (Lonza; CC-3190), which includes FBS.

Day 1:

1. The urine is distributed into two 50 ml falcon tubes (unless the sample contains less than 50ml of urine, in that case one 50 ml falcon tube can be used) and centrifuged at 400 g for 10 min.
2. Remove the urine without disturbing the pellet and leaving ~2-4ml of urine.
3. Resuspend the pellet by washing it and the internal surfaces of the tubes and container with washing buffer (~7ml per falcon tube and ~7ml to wash the container in which the urine was originally stored).
4. Combine the resuspended pellets (cell suspension plus DPBS) into one of the falcon tubes and centrifuge at 200g for 5 min.
5. Remove as much as liquid as possible trying not to disturb the pellet.
Resuspend the pellet by adding 1ml of primary medium.
6. Add the ~1.5ml cell suspension to a well of a 12-well cell culture plate
7. Add 1ml of primary medium to the well with the cell suspension and store the plate in the cell culture incubator at 37°C and 5% CO₂.

The steps to do during the following the days consist in: In Days 2, 3 and 4, carefully add 1ml of primary medium to the well. Alternatively, 3ml of primary medium can be added on Day 1 or Day 2.

On Day 5 remove 2-3 ml of primary medium and add 1-2ml of proliferation medium. From this step, every other day until hURECs become confluent: remove 1-2 ml of medium and add 1-2ml of proliferation medium (e.g. Day 7, 9, 11 etc). hURECs may be first observed after 1-3 weeks of urine processing. If after 3 weeks hURECs are not observed, the sample is discarded.

2.2.4. Subculturing, serum starvation and freezing of cells

When cells (either mouse Shh-LIGHT2 cells, primary human fibroblasts or hURECs) reach confluence (70-100%) cells are subcultured (also known as passaged). For subculturing cells, cells are washed with DPBS, then DPBS is removed. After that:

- For Shh-LIGHT2 cells and primary human fibroblasts, cells are detached by incubating them with 0.05% Trypsin-EDTA (1X) (Gibco; 25300-054) for 5-10 minutes in the cell incubator at 37°C. Then DMEM medium (at least X2 times the volume of the Trypsin added) is added to inactivate Trypsin. Cells (in suspension) are moved to a falcon tube to be centrifuged 1000 rpm for 10 min. Finally, the DMEM medium-Trypsin liquid is removed leaving the cell pellet undisturbed. Then the ~1ml of DMEM medium is added to resuspend the pellet, and the cell suspension is distributed into well/s with DMEM medium to be cultured as it was explained in the previous sections (2.2.1 and 2.2.2).
- For hURECs, cells are detached by incubating them with TrypLE Select Enzyme (1X; no phenol red) (ThermoFisher; 12563011) for 5-7 minutes in the cell incubator at 37°C. Then proliferation medium (at least X2 times the volume of the TrypLE added) is added to inactivate TrypLE. Cells (in suspension) are transferred to a falcon tube to be centrifuged 200 g for 5 min. Finally, the proliferation medium-Trypsin liquid is removed leaving the cell pellet undisturbed. Then the ~1ml of proliferation medium is added to resuspend the pellet, and the cell suspension is distributed into well/s with proliferation medium and cells are cultured as it was explained in the previous section (2.2.3).

Serum starvation was performed for analysis of cilia and ciliogenesis. Serum starvation is known to induce ciliogenesis (Malicki and Johnson, 2017). When cells were serum starved before being fixed for immunofluorescence staining, serum was not added to the corresponding cell culture medium. In the case of Shh-LIGHT2 cells and primary human fibroblasts, FBS was not added to DMEM medium to prepare a DMEM (-FBS) medium. In the case of hURECs, FBS was not added to proliferation

medium to prepare a proliferation (-FBS) medium. For serum starvation the cells were cultured in their corresponding media without serum, often for about 48h before being fixed.

To store cells at -80°C or in liquid nitrogen, 1 ml of a commercial Cell Freezing Media (Sigma-Aldrich; C6164) was used, or 1 ml of a cell freezing media consisting in DMEM medium (20% FBS) with Dimethyl sulfoxide (10%) (DMSO).

2.3. Immunofluorescence (IF) staining

2.3.1. Antibodies

The antibodies used for the experiments involving immunofluorescence (IF) staining are shown in Table 2.3. In this table I show the primary and secondary antibodies used for IF and also in which chapter they have been used.

2.3.2. Fixation of cells

Cells were fixed using 100% methanol. Firstly, media was removed, then cells were washed twice with cold DPBS. DPBS was removed and 100% methanol is added to fix the cells. The cell plate is left for 10 min on ice. Then the ethanol was removed and cells were washed twice with DPBS. Cells were stored in DPBS in the fridge (4°C) before proceeding with the IF protocols.

2.3.3. Immunofluorescence staining in *Shh-LIGHT2* cells and primary human fibroblasts

Day 1:

1. *Shh-LIGHT2* cells or primary human fibroblasts (which are in 12-well or 6-well cell culture plates) were incubated with blocking buffer (5% BSA, diluted in DPBS) for 30 min. DPBS was removed before the addition of blocking buffer.
2. Incubate cells with primary antibodies (add ~500µl in total per well) overnight at 4°C. Primary antibodies were diluted in blocking buffer, accordingly to the dilutions shown in Table 2.3.

Day 2:

3. Wash with DPBS for 5 min (3X).
4. Incubate cells with secondary antibodies (add ~ 500µl in total per well) for 2 h. Secondary antibodies have been diluted in blocking buffer, accordingly to the

dilutions in Table 2.3. This incubation and the following washes are done with the sample (cell culture plate) protected from light.

5. Wash with DPBS for 5 min (3X). And then wash with DPBS for 10 min.
6. Mount coverslips with mounting medium for fluorescence with DAPI: Vectashield (Vector Laboratories; H-1200).
7. Cells can be visualised in the using the corresponding fluorescent microscope or be stored at 4°C.

Description	Catalogue number ; supplier	Technique in which it was used	Dilutions	Relevant chapter/s in which it was used
Primary antibodies				
Rabbit p. CEP120	orb182544 ; Biorbyt	IF	1:100	Chapter 4
Mouse m. PCNT	ab28144 ; Abcam	IF	1:1000	Chapter 4
Mouse m. ARL13B	66739-1-Ig ; Ptglab	IF	1:400 and 1:1000 (1)	Chapters 4 and 5
Rabbit p. ARL13B	17711-1-AP ; Ptglab	IF	1:400	Chapter 5
Rabbit p. PCNT	Ab4448 ; Abcam	IF	1:1000 (2)	Chapter 5
Mouse m. AC-TUB	T6793 ; Sigma	IF	1:1000	Chapter 5
Mouse m. GFP	sc-9996 ; Santa Cruz	IF	1:50	Chapter 4
Rabbit p. GFP	ab290; Abcam	IF	1:200	Chapter 4
Rabbit p. OLIG2	10141047 ; Merck	IHC	1:300	Chapter 3
Rabbit p. PAX6	2565003 ; Cambridge Bioscience	IHC	1:300	Chapter 3
Rat p. CEP120	Courtesy of Dr Mahjoub (Bettleja et al., 2018)	WB	1:500 and 1:2000	Chapter 4
Rabbit p GAPDH	14C10 ; CST	WB	1:1000	Chapter 4
Secondary antibodies				
Anti-rabbit Alexa Fluor 488	A21206 ; Thermo Fisher	IF	1:1000 (3)	Chapters 4 and 5
Anti-mouse Alexa Fluor 594	A21203 ; Thermo Fisher	IF	1:1000 (3)	Chapters 4 and 5
Zenon Rabbit IgG Alexa Fluor 647	Z25308 ; Thermo Fisher	IF (direct labelling)	NA (4)	Chapter 5
Anti-rabbit ABC-HRP	PK-6100 ; Vector Labs	IHC	1:500	Chapter 3
Anti-rat 680	926-68076 ; LI-COR	WB	1:1000	Chapter 4
Anti-rabbit 800	926-32213 ; LI-COR	WB	1:1000	Chapter 4

Table 2.3. List of antibodies with its corresponding dilution/s used, sorted by type of antibody (primary or secondary) and technique in which they were used. (1) Immunofluorescence corresponding to chapter 4, the dilution for staining of mouse and human fibroblasts 1:1000 dilution was used (using 6-well plates). In chapter 5, for staining hURECs, 1:400 was used (using 8-well chamber slides). (2) In Chapter 5, Rabbit p. PCNT antibody was used 1:400 when using direct labelling with Zenon Rabbit IgG Alexa Fluor 647. (3) Anti-rabbit Alexa Fluor 488 and anti-mouse Alexa Fluor 594 secondary antibodies were used at a dilution of 1:400 when combined with any of the primary GFP antibodies and in the Immunofluorescence corresponding to Chapter 5. AC-TUB, Acetylated Tubulin; IF, immunofluorescence, IHC, immunohistochemistry, m., monoclonal; NA, not applicable; p., polyclonal; PCNT, pericentrin; WB, Western blot. (4) Dilutions were made according to manufacturer's instructions.

2.3.4. Immunofluorescence staining in hURECs

Day 1:

1. hURECs (which are cultured in cell culture 8-well chamber slides: μ -Slide 8 Well ibiTreat, Thistle scientific; 80826) were incubated with blocking buffer (5% BSA, diluted in DPBS) for 1 h. DPBS was removed before the addition of blocking buffer.
2. Incubate cells with primary antibodies (add ~ 300 μ l in total per well), overnight at 4°C. Primary antibodies were diluted in blocking buffer, accordingly to the dilutions in Table 2.3.

Day 2:

3. Wash with DPBS for 15 min.
4. Incubate cells with secondary antibodies (add ~ 300 μ l in total per well) for 1 h. Secondary antibodies have been diluted in blocking buffer, accordingly to the dilutions in Table 2.3. This incubation and the following washes are done with the sample (cell culture chamber slides) protected from light.
5. Wash with DPBS for 15 min.

Option 1:

6. Add NucBlue Fixed Cell Stain ReadyProbes reagent (DAPI special formulation, Invitrogen; R37606) (add ~300 μ l per well) according to manufacturer instructions.
7. Cells can be visualised in the using the corresponding fluorescent microscope or be stored (protected from light) at 4°C.

Option 2: when an additional primary antibody (in this case: PCNT) will be used in the following step via direct immunostaining.

6. Prepare direct labelling solution in an Eppendorf 1.5 ml tube, by adding 1 μ l of Rabbit polyclonal PCNT primary antibody.
7. Add 5 μ l of Zenon Rabbit A 647 to the direct labelling solution and incubate for 5 min.
8. Add 5 μ l of Zenon Blocking reagent B to the direct labelling solution and incubate for 5 min.
9. Add 375 μ l of blocking buffer to (5% BSA, diluted in DPBS).

10. Add the direct labelling solution (~150 µl per well) to the cells and incubate overnight at 4°C.

Day 3 (only for option 2):

11. Wash with DPBS for 15 min.
12. Incubate cells in 4% PFA for 10 min.
13. Wash with DPBS for 15 min.
14. Add NucBlue Fixed Cell Stain ReadyProbes reagent (DAPI special formulation, Invitrogen; R37606) (add ~300 µl per well) according to manufacturer instructions.
15. Cells can be visualised in the using the corresponding fluorescent microscope or be stored (protected from light) at 4°C.

2.4. Use of the RNA *in situ* detection platform, RNAscope

To investigate human *CEP120* and *ARL3* expression patterns we (Dr Laura Powell and I) collaborated with the MRC-Wellcome Human Developmental Biology Resource (HDBR).

The human embryonic tissue samples used in this study were obtained from the HDBR resource. These tissue samples are formalin fixed paraffin embedded (FFPE) sections of human embryonic and foetal tissue (Figures 2.2 and 2.3). The sections were treated using 10% neutral buffered formalin and were fixed for 32 hours at room temperature accordingly with HDBR standardised protocols (<https://www.hdbr.org/>) (Wang et al., 2012, Powell et al., 2020).

Regarding the developmental stages analysed, whole human embryo sections in Carnegie Stage 23 (CS23) were used, which is equivalent to 8 post-conception weeks (PCW) in which multiple organs could be visualised (Figure 2.3) and various sections of the human embryonic hindbrain (14 PCW and 19 PCW), kidney (14 PCW and 18 PCW) and eye (14 PCW) (Powell et al., 2020).

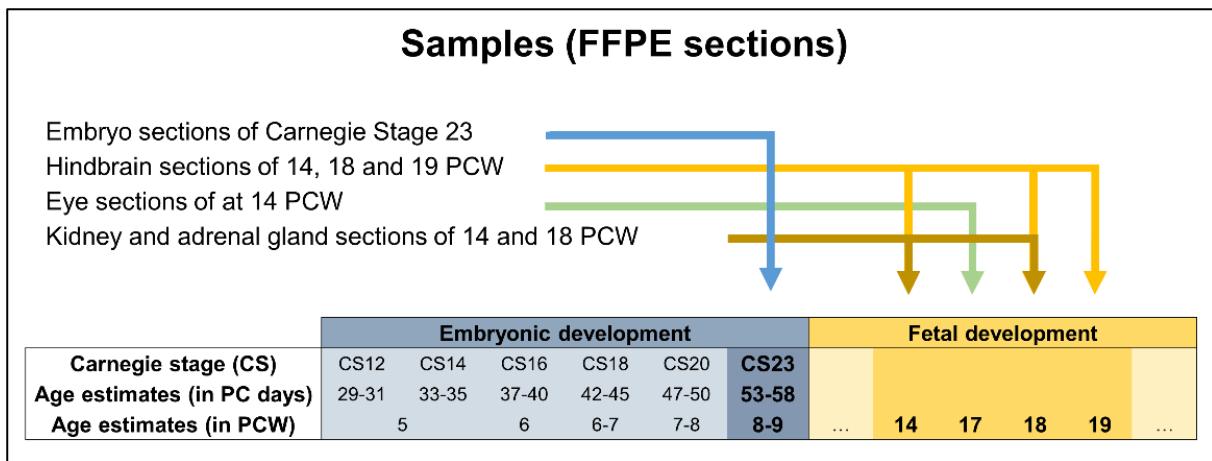


Figure 2.2. Summary diagram of the types of formalin fixed paraffin embedded (FFPE) sections of human tissue utilised in this study and the corresponding human developmental stage of each of them. These sections were obtained and treated by MRC-Wellcome Trust Human Developmental Biology Resource (HDBR) (Lindsay et al., 2016). CS, Carnegie Stage; PCW, post-conception weeks.

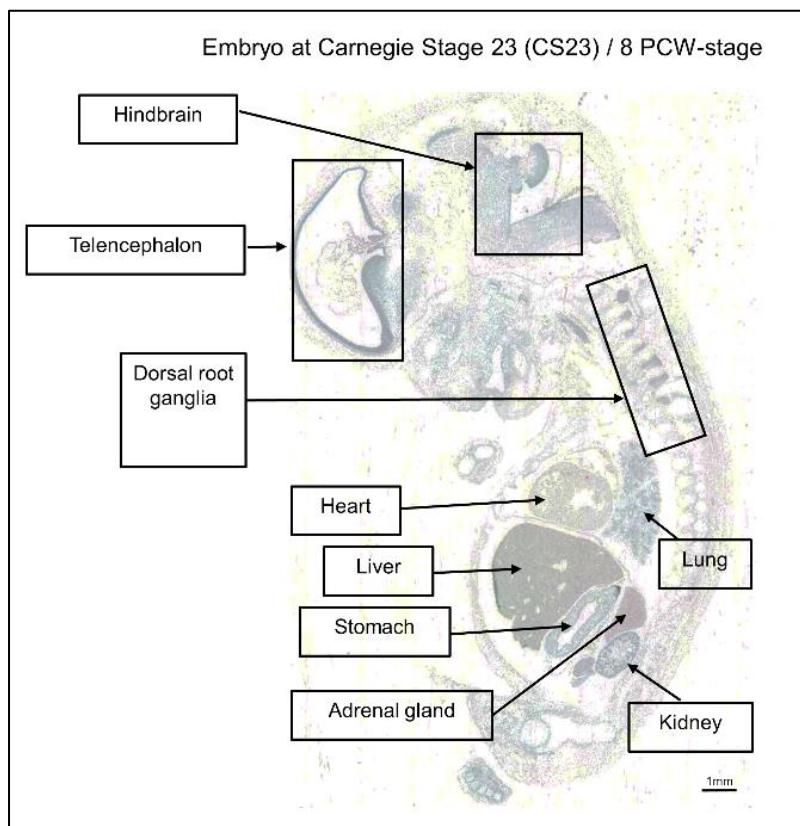


Figure 2.3. Representative image of a sagittal section of a human embryo at Carnegie Stage 23 (CS23) and location of major organs. The age of a human embryo at CS23 stage is approximately 8 post conception weeks (8 PCW). The image shows a formalin fixed paraffin embedded (FFPE) section of a human embryo utilised in this study. This section was obtained and treated by MRC-Wellcome Trust Human Developmental Biology Resource (HDBR) (Lindsay et al., 2016). Sections were visualised using the HDBR image server (Leica Biosystems). Brightness increased and contrast reduced from original image.

RNAscope assays were used according to RNAscope specific guidelines, in order to map *CEP120* and *ARL3* expression and characterise the *CEP120* and *ARL3* expression patterns. This assay is an RNA *in situ* detection technique that detects RNA within intact cells.

The type of RNAscope assay used was: RNAscope 2.5 Assay RED and the probes used targeted the expression of *CEP120*, *ARL3*, *KI67* (positive control) or *dapB* (negative control) (Wang et al., 2012, Robinson et al., 2014). The RNAscope 2.5 Assay RED assays utilised consisted in 20 paired probes. For the *CEP120* and *ARL3* assays the regions targeted are 115-1133 bp (transcript considered: NM_001166226.1) and 169-1570 bp (transcript considered: NM_004311.3), respectively. All the sections were counterstained with Methyl Green.

As a negative control, an RNAscope assay (RNAscope 2.5 HD Assay Red) targeting the expression of *dapB*, which is a bacterial gene, not expressed in human tissues (Voith von Voithenberg et al., 2020) was used. As a positive control, an RNAscope assay (RNAscope 2.5 HD Assay Red) targeting the expression of *KI67*, which is a cell proliferation marker, expressed in cycling cells (Sobecki et al., 2016) was used.

For visualisation of the results of the RNAscope experiment and analysis of the *CEP120* and *ARL3* human embryonic and foetal expression patterns HDBR image server (Leica Biosystems) was used (Powell et al., 2020).

2.5. Analysis of the reported clinical phenotypes and protein sequences of *CEP120* and *ARL3*

To analyse the clinical phenotypes associated with mutations in *CEP120* and *ARL3*, the literature was reviewed using PubMed and OMIM online databases. The orthologues of *CEP120* and *ARL3* were found using NCBI and Ensembl online databases (using the keywords: “cep120” and “arl3”) and the alignment of the ortholog protein sequences was performed using Protein Basic Local Alignment Search Tool, BLASTP.

The query sequences considered for *CEP120* and *ARL3* were NP_694955.2 and NP_004302.1 respectively. To obtain certain ortholog protein sequences from evolutionary distant organisms, additional databases were used, such as: Flybase, Wormbase and Phytozome (Powell et al., 2020).

2.6. Immunohistochemistry (IHC) staining

Immunohistochemistry (IHC) was used to stain Formalin-Fixed Paraffin-Embedded (FFPE) sections of human embryonic tissues obtained from the MRC-Wellcome Human Developmental Biology Resource (HDBR).

These samples consisted in a cortical section and a sagittal section of two 8 PCW-stage human embryos and a sagittal section of human hindbrain at 19 PCW.

I used PAX6 and OLIG2 primary antibodies which are markers of proliferative regions of the developing brain (Alzu'bi et al., 2017, Mo and Zecevic, 2008).

Standardised protocol of immunohistochemistry (Alzu'bi et al., 2017) was used with the primary antibodies OLIG2 and PAX6.

OLIG2 was applied to a section of an 8 PCW-stage human embryo and PAX6 was applied to two sections: one section of a 8 PCW-stage human embryo and one section of a 19 PCW-stage human hindbrain. As negative control I repeated the protocol without primary antibodies in a section of a 19 PCW-stage human hindbrain. Each of the primary and secondary antibodies used and their corresponding working dilutions are shown in Table 2.3.

The sections were incubated with avidin-peroxidase (ABC-HRP, PK-6100, Vector Labs) with biotinylated secondary antibody and developed with diaminobenzidine (DAB) solution (Vector Labs), washed, dehydrated, and mounted using DPX (Sigma-Aldrich). Zeiss AxioPlan microscope was used for visualisation.

Solutions:

- X10 TBS (1.5M NaCl, 0.5M Tris): 87.5 g NaCl (S9888; Sigma-Aldrich), 60 g Trizma base (T1503; Sigma Life Science) and complete up to 1L with ultrapure water.
- 0.01M Citrate Buffer pH 6.0: 2.94g of $(C_6H_5Na_3O_7)2H_2O$. tri-sodium citrate in 1L distilled H_2O , pH to 6.0.
- DAB: ImmPACT DAB peroxidase substrate (2BScientific; SK-4105).

Day 1:

1. Incubate sections in Histoclear (national diagnostics; HS-202) for dewaxing and clearing (10 min).

2. Dehydrate sections with an ethanol series, with decreasing ethanol concentrations: 100%, 100%, 90%, 70% and 50% (3 min each).
3. Block endogenous peroxidase in methanol peroxide solution (10 min).
Methanol peroxide solution is prepared the same day: 3 ml of Hydrogen Peroxide (30%) in 180 ml of methanol (0.5% H₂O₂).
4. Rinse in running tap water.
5. Incubate in citrate buffer (0.01M) antigen retrieval. Microwave for 10 min. Cool it for 15 min at room temperature (RT).
6. Wash in 1XTBS (5 min).
7. Incubate in 10% normal blocking serum (diluted in 1XTBS). ~ 0.5 ml per slide (10 min). Blocking buffer (10% normal blocking serum) is often prepared the same day: 1 ml of normal goat serum in 9 ml of 1XTBS).
8. Tip off excess of normal serum and add primary antibody (diluted in blocking buffer). Antibodies were diluted 1:300 in blocking buffer. ~300 µl of diluted antibody was added per slide section targeting the corresponding tissue.
Incubate overnight at 4°C.

Day 2:

9. 5 min washes in 1XTBS (X2).
10. Incubate the secondary antibody diluted in blocking buffer. ~300µl per slide 30 min. Secondary is diluted 1:500 in blocking buffer. Start preparation of the tertiary complex following manufacturer instructions and keep on ice.
11. 5 min washes in 1XTBS (2X).
12. Incubate section slides with tertiary complex in 1XTBS for 30 min. Add ~300µl per slide. The tertiary complex should have been made at least 30 min before use (1µl of reagent A and 1µl of reagent B in 50 µl of 1XTBS, to allow complex to form at RT).
13. 5 min washes in 1XTBS (2X).
14. Develop with DAB (3:100 using diluent from the DAB kit) (~300 µl per slide) for 10 min at rt.
15. Wash in running tap water for 10 min.
16. Counterstain with 1% toluidine blue (Merck; 89640-5G) (30-40 sec).
17. Wash with ultrapure water.
18. Dehydrate sections with an ethanol series, with increasing ethanol concentrations: 50%, 70%, 90%, 100% and 100% (3 min each).

19. Clear and mount (with DPX). As DPX is toxic, this step should be done in a fume hood and let DPX dry (at least) overnight before imaging.

2.7. Review and annotation of *CEP120* and *CC2D2A* genetic variants and clinical presentations reported in the literature

In order to complete this *in silico* study to find skippable exons in *CEP120* and *CC2D2A*, I first reviewed the literature to annotate all the reported *CEP120* and *CC2D2A* variants in a standardised and organised manner.

To find all the reported patients with biallelic genetic variants in *CEP120* (NG_042125.1) and *CC2D2A* (NG_013035.1) I used PubMed (using the keywords: “cep120” and “cc2d2a”) and HGMD (Stenson et al., 2017) databases (last query 05/2020). After curation to select the relevant publications 33 publications were considered: 31 publications describing patients presenting *CC2D2A* mutations and 2 publications describing patients presenting *CEP120* mutations.

There were situations in which the same patient is reported in more than one publication. When this was detected, the patient was only included once in this annotated database of reported patients and variants. Patients with single heterozygous mutations and patients reported with not enough or incomplete phenotypic information were not included in this study.

For each patient included, I reviewed the phenotypic information available. For some patients the clinical diagnosis was adapted and standardised according to the disease categories (JBTS, ML, MKS) described in Drivas et al., 2015 (Drivas et al., 2015). Based on Drivas et al., 2015, JBTS patients present a phenotype that includes hypoplasia of the cerebellar vermis and/or brain stem abnormalities and intellectual disability with or without extra-CNS manifestations. Patients with Meckel-like syndrome (ML) present a phenotype which is lethal during the first months or years, characterized by cystic kidney disease, CNS malformation (typically Dandy Walker malformation), polydactyly, and hepatic fibrosis. Patients with MKS present a phenotype which is similar to ML, but it is uniformly perinatal lethal with occipital encephalocele and in which the CNS malformation is predominant (Drivas et al., 2015).

I identified the truncating (loss-of-function) variants with a deleterious effect that could be bypassed if the exon in which they are located is skipped. To do that

SMART online tool was used and the information available in the literature to annotate the functional domains of CEP120 and CC2D2A proteins. Then the genotype and phenotype corresponding to each of the patients with reported *CEP120* and *CC2D2A* variants were described (Figures 2.4A and 2.4B).

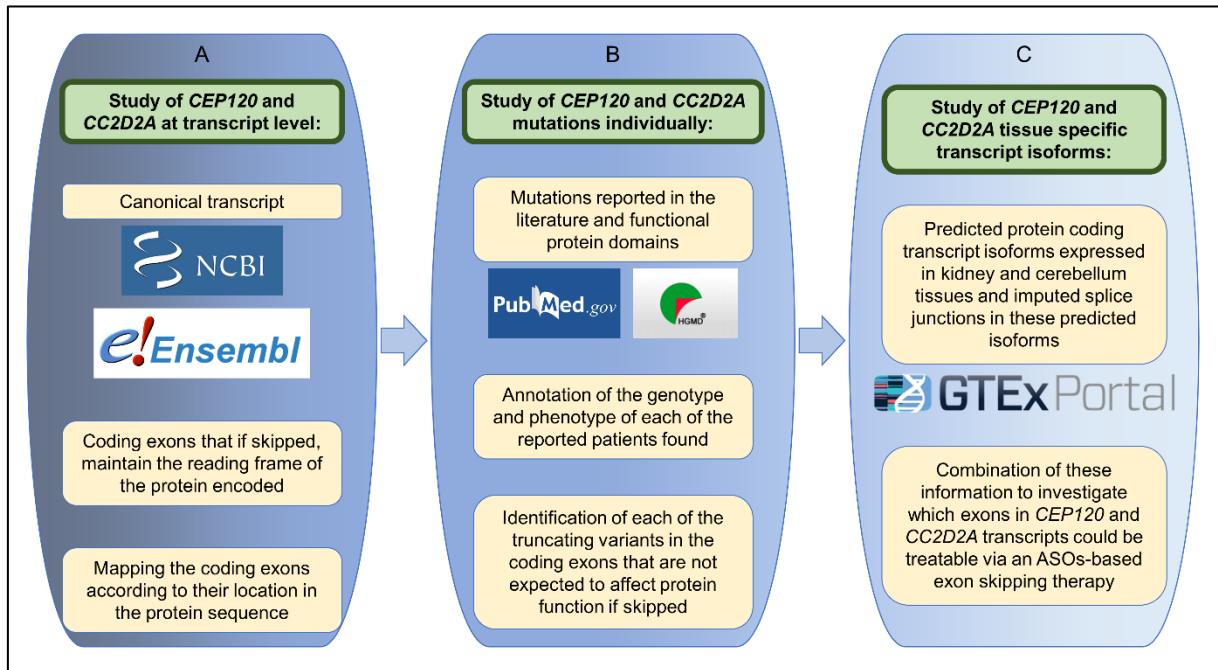


Figure 2.4. Pipeline followed for the *in silico* identification of skippable exons in *CEP120* and *CC2D2A* genes which are suitable for an ASOs-based exon skipping therapy. A: NCBI Gene and Ensembl databases were used to identify the canonical transcript of each of the two genes of interest: ENST00000328236.9 (also called NM_153223.3) for *CEP120* and ENST00000503292.5 (also called NM_001080522.2) for *CC2D2A*. Ensembl was used to identify coding exons corresponding to these transcripts, as well as to determine which of these exons if skipped, maintain the reading frame of the protein. B: At the same time as doing what it was described in panel A: using OMIM (Hamosh et al., 2000), PubMed and the Human Gene Mutation Database online resources, the mutations described in the literature in these two genes were found. C: Genotype-Tissue Expression (GTEx) Project online resource was used to find the predicted protein coding transcript isoforms expressed in kidney and cerebellum tissues and combined the information obtained.

To curate and annotate the genetic variants described in these publications Ensembl Variant Effect Predictor (Ensembl release 100) (Yates et al., 2020), NCBI ClinVar and VarSome (Kopanos et al., 2019) were used. To obtain allelic frequency data in the general population gnomAD v2.1.1 database (Karczewski et al., 2020) was used. The variants were described according to the Human Genome Variation Society (HGVS) nomenclature guidelines (den Dunnen et al., 2016).

Ultimately, the annotation of variants in *CEP120* and *CC2D2A* and the use of Genotype-Tissue Expression (GTEx) Project (The GTEx Consortium, 2020) and Ensembl online resources (accessed on 05/2020) allowed me to find the corresponding skippable exons in *CEP120* and *CC2D2A* genes that are suitable for an ASOs-based exon skipping therapy (Figure 2.4C) (Barroso-Gil et al., 2021a).

If an exon consists in any number of nucleotides that is divisible by 3, the exon is described as skippable unless it is the first or the last coding exon of the transcript. The first and last coding exons of the canonical transcript of *CEP120* and *CC2D2A* are not skippable, as they encode for the corresponding start and stop codons, respectively. Coding exons were mapped according to their location in the protein sequence (Barroso-Gil et al., 2021a). ProteinPaint (Zhou et al., 2016) was used to visualise the transcripts together with their encoded proteins. All the Web resources used were listed in Table 2.1 (in section: 2.1. Software tools and online resources). The GTEx project allowed to find different tissue-specific imputed splice junctions for each of these predicted transcripts. By combining all these pieces of information, it was discussed if which coding exons of the genes *CEP120* and *CC2D2A* could be treatable via an ASOs-based exon skipping therapy (Barroso-Gil et al., 2021a).

2.8. Microscopy

2.8.1. Microscopes used

All the microscopes used are located at the International Centre for Life and are managed by the Bioimaging Unit of Newcastle University (<https://www.ncl.ac.uk/bioimaging/locations/icfl/>). The corresponding training has been received before using these microscopes. The microscopes used were a Nikon (A1) confocal inverted and a Zeiss AxioImager fluorescent microscopes to visualise immunofluorescence staining. For live imaging, Nikon Eclipse TS100 and Zeiss AxioVert 200M inverted microscopes were used. I used Zeiss AxioPlan microscope to visualise tissues treated with RNAScope technique and immunohistochemistry.

2.8.2. Processing of microscopy images

All the images shown in this thesis have been processed using one or several tools shown in Table 2.1.

The images used in Chapter 3 for the characterisation of *CEP120* and *ARL3* expression patterns in human embryonic and fetal tissue were obtained using the

HDBR image server (Leica Biosystems). The size of the images and distribution of the image panels were done using Adobe Illustrator and Adobe Photoshop.

The images for immunofluorescence (IF) staining experiments were processed using the corresponding software of the corresponding microscope. For images obtained using the Nikon (A1) confocal inverted microscope, images were captured using of X40, X60 or X100 magnification objectives and Z-stacks were generated. The resulting ND2 files were exported and processed with NIS Elements 4.20 and Image J. In ImageJ, the Z-stack images were combined to obtain maximum intensity projection and each of the corresponding the channels were split.

To analyse the cilia phenotype of hURECs fixed and treated for immunofluorescence, I used images obtained using the Nikon (A1) confocal inverted microscope. All the images for this quantification analysis were taken using the same magnification objective: Plan Apo VC 60x Oil DIC N2.

Of note, when cilia were analysed the hURECs were serum starved for 48h before being fixed. To measure the cilia length, I used ImageJ (Fiji) to create a maximum intensity projection of a Z-stack, and then measured (using the segmented line tool of ImageJ) the length of the ARL13B (marker of cilia) staining, from the base of the cilia, which limits to the PCNT (pericentrin, marker of centrosome/basal body) to the tip of the cilia staining. Unpaired *t* test statistical test was used to determine if there is any difference in cilia length between a patient and control's hURECs samples, as it has also been used by others in other studies (Molinari et al., 2020, Ramsbottom et al., 2018, Oud et al., 2018). GraphPad software was used for statistical analysis and display of cilia measurements.

For images obtained using the Zeiss AxioImager fluorescent microscope, images were captured using of X20, X40 or X63 magnification objectives. The resulting CZI file was exported and processed with ZEN 2.3 and Image J. In ImageJ, images were processed to split each of the corresponding the channels used. For the images obtained of living cells using the Nikon Eclipse TS100 and Zeiss AxioVert 200M inverted microscopes, the images were analysed using NIS Elements 3.20 and AxioVs40x64 software tools respectively. The images of living cells transfected with a GFP plasmid were visualised using the Zeiss AxioVert 200M inverted microscope and the resulting ZVI files were exported and processed with AxioVs40x64 and

Image J. In ImageJ, images were processed to split each of the corresponding the channels used.

2.9. Genetic and gene expression studies

2.9.1. Polymerase chain reaction (PCR)

Primers for PCR reactions were designed using Primer3Plus and NCBI Primer-BLAST. PCR was performed according to the standardised protocol using GoTaq G2 DNA Polymerase (Promega; M7845). For a fragment with a length between 90-500 bp the following program was used: Initial denaturalisation of 94°C (10 min), followed by 25 cycles a series of denaturalisation, annealing and extension (94°C for 20s, 55°C for 30s and 72°C for 40 s) and a final extension of 72°C (10 min).

A typical 20 µl PCR reaction consisted in 1-2 µl of DNA/cDNA, 1 µl of forward and reverse primers each (10 µM stock, IDT), 0.2 µl of dNTP (10 mM stock, Thermo Fisher Scientific), 0.1 µl GoTaq DNA Polymerase (5 U/µl stock), 4 µl buffer (5X stock, M791A; Promega) and 10.7-11.7 µl ultrapure water (double filtered, with a 0.2µm Syringe Filter).

2.9.2. Agarose Gel electrophoresis

PCR products and a 100bp DNA ladder (Promega, G2101) or a 1kb DNA ladder (Promega; G571A and NEB; N0552G) were typically loaded in 2% Agarose gels. Gels were prepared according to standardised protocols using Agarose Low EEO (NBS Biologicals; NBS-AG500), 1XTAE buffer (Bio-Rad Laboratories Ltd; 1610773) and GelRed Nucleic Acid Gel Stain (Cambridge bioscience; BT41003). To visualise For PCR products larger than 1kb, 1% agarose gels in TAE were used. To clearly separate the corresponding DNA fragments gels were typically run for 30-40 min at 100 V. Gels were visualised using a UV transilluminator (BioSpectrum Imaging System).

2.9.3. RNA extraction and clean-up

RNA was extracted using either one the two following methods: one of them uses TRIzol Reagent (ThermoFisher; 15596018) and the other use RNAeasy mini kit (QIAGEN; 74104).

The method using TRIzol Reagent has been applied for the isolation of all the RNA samples unless stated otherwise. Firstly when cells are confluent, cells are washed

with cold DPBS twice TRIzol Reagent is added (~400 µl per well of a 12-well plate). Using a cell scrapper, cells in TRIzol are transferred into a cryotube (in this step, cells in TRIzol can be stored at -80°C to continue with the protocol another day).

TRIzol Reagent is a phenol-based solution for the isolation of total RNA. I followed manufacturer's instructions in which Tryzol Reagent is used with Chloroform and Isopropanol to separate RNA from DNA and proteins and to precipitate RNA respectively. RNA is also washed with 70% Ethanol and resuspended in RNase-free water. RNA is then clean-up using Lithium Chloride Precipitation Solution (AM9480; Invitrogen) according to manufacturer's instructions using 70% Ethanol and resuspended in RNase-free water.

The method of the RNAeasy mini kit has been applied to isolate RNA from few samples of mouse Shh-LIGHT2 cells. I followed manufacturer instructions of the RNAeasy mini kit, which uses a spin columns technology to isolate total RNA.

RNA concentration and purity were measured using a Ds-11 FX+ Spectrophotometer (DeNovix).

2.9.4. Reverse transcription-PCR (RT-PCR)

To isolate the RNA I used RNeasy mini kit (Qiagen) according to the manufacturer's instructions. For first strand cDNA synthesis (reverse transcription) I used 0.25-1 µg of RNA, SuperScript III Reverse Transcriptase (Invitrogen; 18080044) and Oligo(dT)15 Primer (Promega; C1101) following manufacturers' instructions. Once the RNA was reverse transcribed into cDNA, I reverse transcription-PCR (RT-PCR) was done using the corresponding primers, designed using Primer3Plus and NCBI Primer-BLAST. These gene specific primers are designed so that they do not amplify genomic DNA.

RT-PCR was done using GoTaq DNA Polymerase (Promega) standardised protocol. The 20 µl PCR reactions were performed as it was explained in section: 2.9.1. *Polymerase chain reaction (PCR)*. The RT-PCR products were loaded into a 2% Gel electrophoresis and visualised using UV transilluminator (BioSpectrum Imaging System). RT-PCR was used to ensure cDNA synthesis from total RNA was successful, before using the corresponding cDNA for quantitative and comparative applications such as quantitative PCR (qPCR).

RT-PCR and gel electrophoresis were used to ensure cDNA synthesis occurred using RNA isolated (RNAeasy mini kit method) from mouse Shh-LIGHT2 cells and confirm expression of *Cep120* in mouse Shh-LIGHT2 cells, by amplifying a fragment of cDNA located in *Cep120* (Figure 2.5). I used *Cep120* gene-specific primer pair (5'-CGCCTCCAACCAAGGATGAT-3 and 5'-AGAACCGATGCCCTCACAG-3).

RT-PCR was also used to ensure cDNA synthesis occurred using RNA isolated (TRIzol Reagent method) from hURECs cells, by amplifying fragments of cDNA located in *HPRT1* (Figure 2.6). I used a *HPRT* gene-specific primer pair (5'-TGACACTGGCAAAACAATGCA-3 and 5'-GGTCCTTTCACCAAGCAAGCT-3).

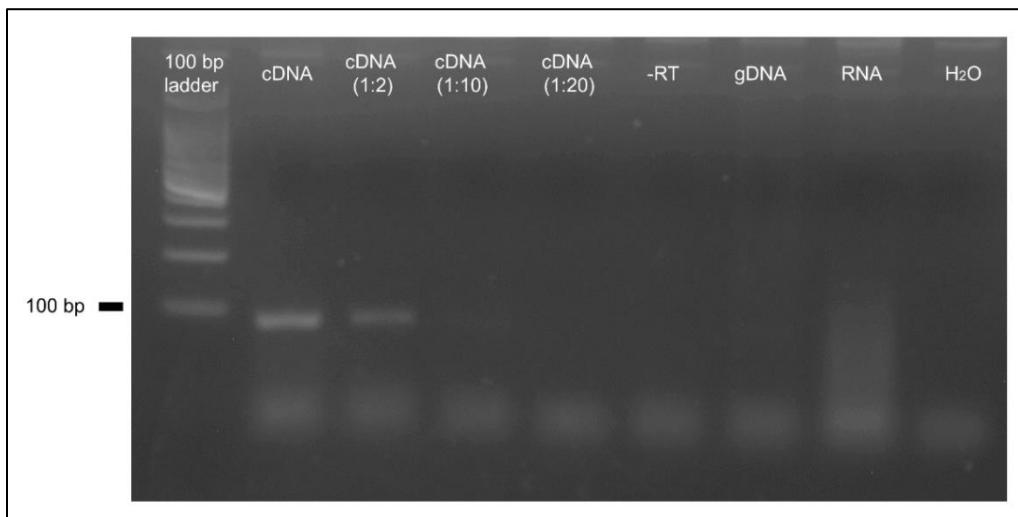


Figure 2.5. RT-PCR using cDNA, reverse transcribed from total RNA isolated from Shh-LIGHT2 cells. It can be observed that the primers utilized used amplified the expected fragment (92 bp located in exons 9 and 10 in *Cep120* mouse transcript (NM_178686.4). RNA concentration from RNA extraction was 187.1 ng/μl. For first strand cDNA synthesis 1 μg of RNA was used. cDNA was diluted in water: 1:2, 1:10 and 1:20. A clear band of ~100 bp can be observed when cDNA is not diluted and when it is diluted 1:2. Negative controls include samples in which reverse transcriptase was not added (-RT), genomic DNA (gDNA), RNA or H₂O were used instead of cDNA. Gel electrophoresis (2% agarose). 100bp ladder used.

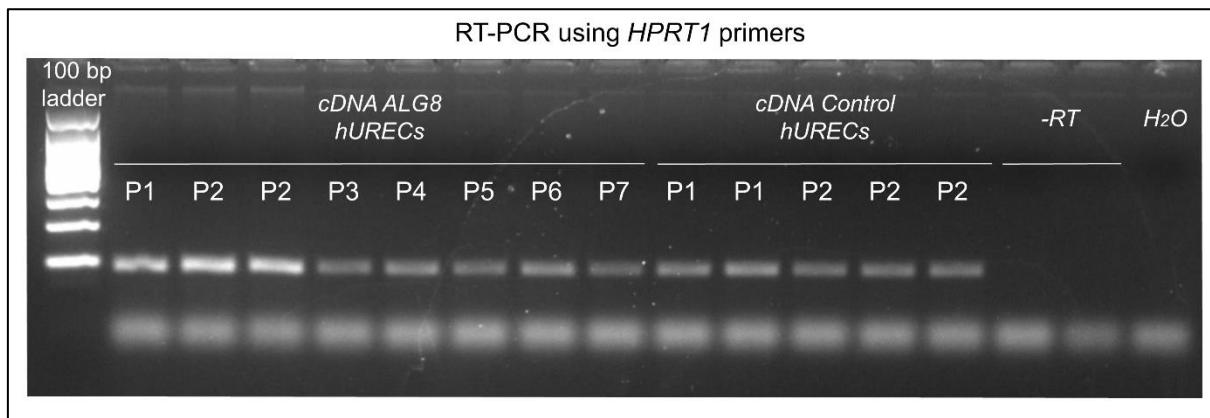


Figure 2.6. RT-PCR using cDNA, reverse transcribed from total RNA isolated from human urine-derived renal epithelial cells (hURECs) from an ADPKD patient. This ADPKD patient presented a heterozygous nonsense mutation in NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter) (*ALG8* hURECs) and from a healthy individual (Control hURECs). RNA was extracted from both individuals was during several passages (P). It can be observed that the amplification of the expected fragment (94 bp located in exons 6 and 7 in *HPRT1* human transcript (NM_000194.3) in all the cDNA samples. RNA concentration from RNA extraction of the samples ranged from 25 ng/ μ l to 290 ng/ μ l. For first strand cDNA synthesis 250 ng of RNA was used. A clear band of ~100 bp can be observed in all samples. Negative controls include (from left to right): samples in which reverse transcriptase was not added (-RT) to the RNA from the *ALG8* and Control hURECs, and a negative control in which H₂O was used instead of cDNA. Gel electrophoresis (2% agarose). 100bp ladder used.

To study tissue-specific transcripts and possible molecular targets for exon skipping in *CEP120* and *CC2D2A* genes, I used total RNA from human kidney (ThermoFisher; AM7976) and total RNA from whole blood samples and total RNA from human urine-derived renal epithelial cells (hURECs). RT-PCR was performed according to the standardised protocol described above.

A *CC2D2A* gene-specific primer pair (5- TGAGAGACACTGGCTGGGAT -3 and 5- AGGCACTGACGATTGGAAAC -3) was used to identify basal skipping of exon 30 and the *CEP120* gene-specific primer pair (amplifying only when the exon 2 is skipped) (5- TACAGCAGTAGTGCGCTTGC -3 and 5- GGGAAATGCCGACCTCACAG -3) was used to identify basal skipping of exon 2. A gene-specific primer pair for *HPRT1* (housekeeping gene) was used (5- TGACACTGGCAAAACAATGCA-3 and 5- GGTCCCTTTCACCAGCAAGCT-3).

2.9.5. Quantitative PCR (qPCR)

Gene expression was measured using PrimeTime quantitative PCR (qPCR) assays. Gene-specific PrimeTime qPCR assays (IDT) were designed by IDT. PrimeTime

qPCR assay is based on a 5' nuclease assay which uses a gene specific primer pair and a probe with a 5' fluorophore and a quencher.

QuantStudio Real-Time PCR Software was used to design the set up the 384-well qPCR plate, and a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems), to perform the qPCR reactions for the standard curves and comparative experiments. Each sample for each gene was assayed in triplicates or quadruplicates following the manufacturer instructions.

Each 10 μ l qPCR reaction consisted in 4 μ l of cDNA (previously diluted 1:10 in nuclease-free water), 5 μ l of PrimeTime Gene Expression Master Mix (1055771; IDT), 0.5 μ l PrimeTime Assay and 0.5 μ l of nuclease-free water.

Results of the qPCR experiments (Ct values) were exported from the QuantStudio Real-Time PCR Software and analysed using the Livak method for relative gene expression analysis ($\Delta\Delta Cq$ method) to find differences in gene expression of genes of interests between a test (treated/patient) sample and a control sample. If cDNA from human cells was used, gene expression levels of target genes were normalised against gene expression levels of *HPRT1*, *GAPDH* and *GUSB* housekeeping genes. If cDNA from mouse cells was used, gene expression levels of target genes were normalised against gene expression levels of *Hprt* and *Gapdh* housekeeping genes.

2.10. Protein extraction and Western blotting

Solutions:

- Lysis buffer: 2.4 g urea (U6504; Sigma Life Science), 2.5 ml 0.5M Tris pH 6.8, 2 ml 20% SDS (L3771; Sigma), 500 μ l β -mercaptoethanol (Sigma-Aldrich; M7522), 2 ml of 50% Glycerol and 10ml of ultrapure water. Loading buffer consist in Lysis Buffer containing bromophenol blue (B5525-5G; Sigma) (~0.02%).
- 10X running buffer: 36.2 g Trizma Base (T1503; Sigma) and 188 g Glycine (G8898; Sigma) up to 1 L with ultrapure water. To make 1X running buffer: 100 ml of 10X running buffer is diluted in ultrapure water up to 1 L and 5 ml of 20% SDS are added.
- 10X Transfer buffer: 58 g Trizma Base and 29 g Glycine up to 1 L with ultrapure water. 100 ml of 100% Methanol and 2 ml of 20% SDS are added to 1L of 1X transfer buffer. 1X transfer buffer needs to be cold before being used.

Protein was extracted from Shh-LIGHT2 cells when cells were confluent. Cells were washed in DPBS, Lysis Buffer was added to the cells (~200 µl per well of a per 12-well plate). Using a cell scrapper, cells in Lysis Buffer are transferred into a cryotube (in this step cells in Lysis Buffer can be stored at -20°C to continue with the protocol another day if needed).

When protein was extracted from a mouse kidney, a mouse kidney was previously snap frozen in liquid nitrogen and stored at -80°C. A small sample (~0.1 g) was cut from this kidney and homogenised with Lysis buffer.

I used standardised protocol for Western blotting (Ramsbottom et al., 2018), using Bio-Rad Power Supplies and buffer tanks (Mini-Protean tetra System, Bio-Rad) and Mini-Protean TGX precast 4-20% gradient gels (456-1094; Bio-Rad).

Protein samples are heated to 95 °C for 5 min and spun at 16000 rcf/maximum speed for 5 min before being loaded for sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Samples with loading buffer (a total of ~20 µl are loaded per well) and 3 µl of ladder (Chameleon Duo Pre-stained Protein Ladder, 928-60000; LI-COR) were loaded into the gel.

The gel was run for a first stage of 80 V for 20 min and a second stage of 200 V for 1 hour. Proteins were transferred into 0.45 µm nitrocellulose membrane (1620115; Bio-Rad). The membrane was washed with TBST and blocked with blocking buffer (5% milk in TBST) for 1 h. TBST is TBT with 0.1% of Tween 20 (P1379; Sigma)

The primary antibodies were added and the membrane was incubated overnight at 4°C. The next day the membrane was washed in TBST, the secondary antibodies were added to the membrane and it was incubated for 90 min at RT (the membrane was protected from light from this step), the membrane was washed in TBST again and it was visualised using Odyssey CLx imaging system (LI-COR) and ImageJ software.

The primary antibodies used for Western blotting were Rat polyclonal CEP120 (Courtesy of Dr Mahjoub (Betleja et al., 2018) and rabbit anti-GAPDH (CST; 14C10) and the secondary antibodies were goat anti-rabbit 800 (926-32213 ; LI-COR) and goat anti-rat 680 (926-68076 ; LI-COR).

2.11. *Cep120/CEP120* gene knockdown

2.11.1. Small interfering RNA (siRNA) Cep120 knockdown experiment in mouse Shh LIGHT2 cells

Shh LIGHT2 cells were seeded one day before the Cep120 knockdown experiment took place. The day the Cep120 knockdown experiment starts, cells are 70-90% confluent. The medium of the cells is changed to a medium without serum (DMEM - FBS) the same day, before the Cep120 knockdown experiment start.

Following standardised protocol, I transfected the corresponding 200 nM of different gene specific siRNA to perform each of the corresponding gene knockdowns. I used siGENOME *CEP120* siRNA (D-016493-01; Dharmacon), siGENOME Smo siRNA (L-041026-00; Dharmacon) and a negative Control siRNA (AM4611, ThermoFisher) using the Lipofectamine RNAiMAX transfection reagent (ThermoFisher, 13778075) following manufacturer's instructions.

To achieve *Cep120* knockdown via siRNA, the corresponding siRNA targets the exon 21 (last coding exon) of human of *CEP120*. There is only one mismatch in this target sequence compared with the mouse sequence of *Cep120* (Figure 2.7). The siRNA sequence targeting *CEP120/Cep120* is: 5-GAUGAGAACGGGUGUGUAU -3.

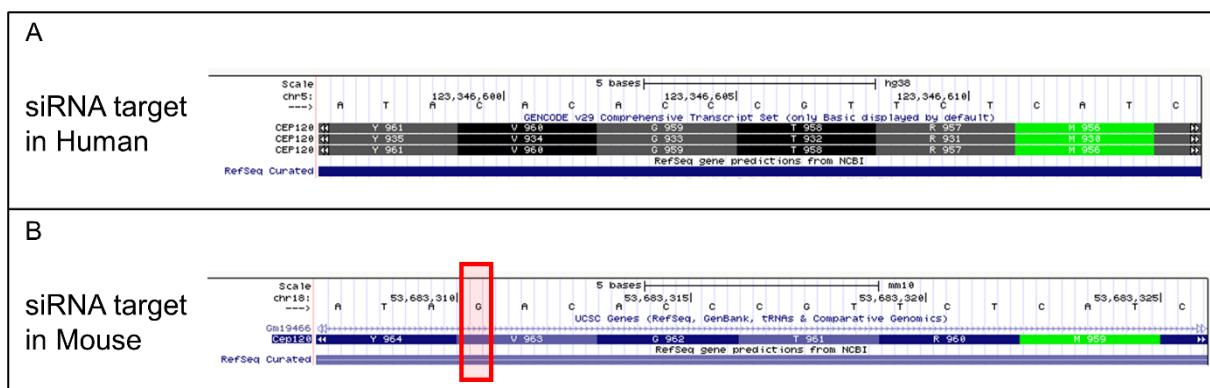


Figure 2.7. Sequence alignment of the siRNA targeting *CEP120/Cep120* (D-016493-01; Dharmacon). siRNA sequence is: 5- GAUGAGAACGGGUGUGUAU -3. A: This siRNA targets a coding region located in exon 21 (last coding exon) of human *CEP120*. B: There is one mismatch between the siRNA target sequence compared with the mouse sequence of *Cep120*. The screenshot in panel A was taken from the UCSC Genome Browser on Human (GRCh38/hg38) to show the genomic position encoding the *CEP120* sequence of the siRNA target. The screenshot in panel B was taken from the UCSC Genome Browser on Mouse (GRCm38/mm10) to show the genomic position encoding the *Cep120* sequence of the siRNA target. Red rectangle in panel B shows the mismatch between the siRNA target sequence and the mouse sequence.

The efficiency of the knockdown experiment via siRNA (Dharmacon) in Shh-LIGHT2 results was evaluated via qPCR, performed according to standard protocols (as explained in 2.9.5).

The qPCR assays (IDT) used to perform the corresponding qPCRs reactions using mouse cDNA are: Mm.PT.58.46061505 (*Cep120*), Mm.PT.58.13243838 (*Smo*), Mm.PT.58.9411934 (*Hprt*), Mm.PT.39a.1 (*Gapdh*) to evaluate siRNA knockdown efficiency.

2.11.2. SAG stimulation of Shh-LIGHT2 cells and dual luciferase assay

To investigate the effect on Hh signalling upon *Cep120* knockdown, when Shh-LIGHT2 were confluent, one day after *Cep120* knockdown via siRNA started, they were stimulated using 250nM SAG Hh signalling agonist. As a negative control 250nM DMSO is added instead of SAG. As *Smo* knockdown is known to affect Hh signalling (Guo et al., 2013, Ogden et al., 2006), I did a *Smo* knockdown as a positive control in the Shh-LIGHT2 cells. Knockdown of *Smo* was performed expecting to impair SAG stimulation in the Shh-LIGHT2 cells.

Of note, the DMEM medium used during at least two passages before the experiments on Shh-LIGHT2 cells are performed, did not have G418. This is to prevent G418 interfering with the dual luciferase assay.

Shh-LIGHT2 cells were lysed, 2 days after SAG or DMSO treatment (3 days after siRNA treatment). Lysed Shh-LIGHT2 cells were stored before I used Promega Dual-Luciferase Reporter Assay System standard protocol (Promega; E1910) and GLOMAX Multi Detection System Luminometer to quantify dual luciferase ratios according to manufacturer's instructions.

SAG stimulation in Shh-LIGHT2 cells is revealed calculating the dual luciferase ratio between the luminescence values reported by the Gli-dependent firefly luciferase and constitutive (control) Renilla luciferase vectors.

2.11.3. Antisense oligonucleotide (ASO) morpholino CEP120 knockdown experiment in primary human fibroblasts

The knockdown of *CEP120* expression in primary human fibroblasts was done using an antisense oligonucleotide (ASO) morpholino (GeneTools). The ASO was designed by Genetools and is predicted to cause the skipping of exon 3 of NM_153223.3 *CEP120* transcript, and subsequently leading to a premature termination codon that might trigger nonsense-mediated decay of the transcript. Its target (5'-TGAAAATGATAATCACCTGTGCTGA-3') includes the splice junction between exon 3 and intron 3 as it is shown in Figure 2.8.

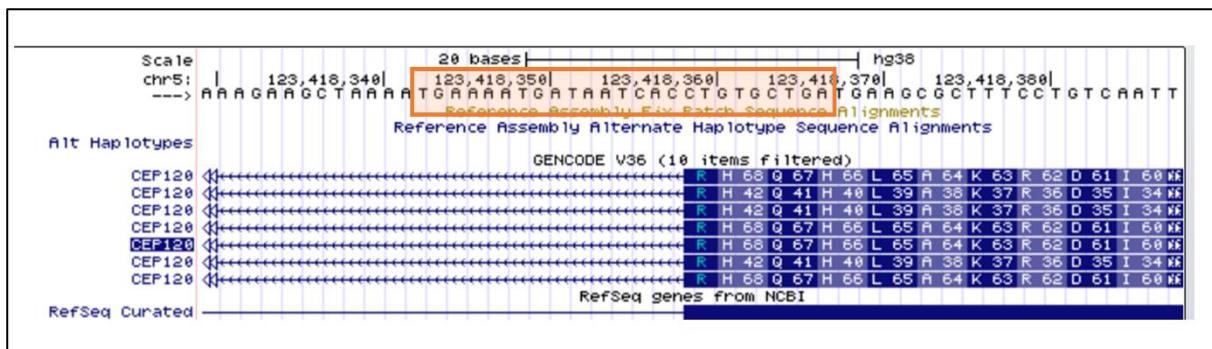


Figure 2.8. Sequence alignment of the ASO morpholino targeting *CEP120* (GeneTools). The ASO morpholino sequence is: 5'-TGAAAATGATAATCACCTGTGCTGA -3. It is observed that this ASO morpholino targets a coding region located in exon 3 and intron 3 of human *CEP120*. The screenshot was taken from the UCSC Genome Browser on Human (GRCh38/hg38) to show the genomic position encoding for the ASO morpholino target in the *CEP120* sequence. Orange rectangle shows sequencing encoding for the ASO morpholino target sequence.

Following manufacturer instructions, I resuspended the freeze-dried ASO morpholino oligo in RNase free water, to make a 1mM oligo stock solution. Once the cells were confluent (70-90% confluence), 2.5 μ M of *CEP120* ASO morpholino oligo were delivered to the cells using the Endo-Porter reagent (in DMSO) (GeneTools). 6 μ l of Endo-Porter reagent were added for every 1 ml media. After 2 days, RNA was extracted from primary human fibroblasts using the TRIzol method for gene expression studies.

As negative controls, primary human fibroblasts were treated only with Endo-Porter reagent (without the ASO morpholino oligo) or not treated (without Endo-Porter reagent and ASO morpholino oligo).

2.12. Alamar Blue cell viability assay

To investigate cell viability of hURECs from a patient (*ALG8* patient) and a healthy individual (Control A) I used alamarBlue HS Cell Viability Reagent (A50100; ThermoFisher) following manufacturer's instructions.

Of note, when added to cells, alamarBlue HS Cell Viability Reagent is modified by the reducing environment of metabolically active (viable) cells, this correlates with a colour change that can be detected using a spectrophotometer to calculate absorbance (detected at 570/610 nm).

Series of an increasing amount of hURECs were seeded to obtain the standard curves for both patient's and Control I's hURECs. I seeded (using 12-well cell culture

plates) an increasing number of cells (e.g: 5K, 10K, 20K and 25K cells) to know the standard curve of each of the individuals.

To seed these corresponding amounts of cells a haemocytometer (Cambridge Bioscience; DHC-N01-50) was used. Knowing the slope of each of the standards, the number of living cells can be estimated throughout seven consecutive days. I initially seeded 10K cells (except in the case of the hURECs of the *ALG8* patient in P1 in which I seeded 15K cells) and using alamarBlue HS Cell Viability Reagent I estimated the number of cells throughout seven consecutive days. I seeded 10K hURECS (using 12-well cell culture plates) from the *ALG8* patient and Control A and treated them with alamarBlue HS Cell Viability Reagent the next day. 40 µl of alamarBlue HS Cell Viability Reagent were added to hURECS in 360 µl of Proliferation Media.

The following day (and each day up to 7 consecutive days), 50 µl of this media containing alamarBlue HS Cell Viability Reagent added the day before is transferred to a well of a 96-well cell culture plates (protected from light) (150 µl were used in total, as I did this experiment in triplicates) to be read by a spectrophotometer (Varioskan Lux; ThermoFisher). The rest of media with alamarBlue HS Cell Viability Reagent was replaced with fresh 40 µl of alamarBlue HS Cell Viability Reagent and 360 µl of Proliferation Media. The following day the same process was done.

Varioskan Lux (ThermoFisher) spectrophotometer was used to obtain the corresponding absorbance values for each of the days in which hURECs were treated with alamarBlue HS Cell Viability Reagent following manufacturer's instructions. This was done for *ALG8* patient and Control A hURECs and was repeated it during several passages.

The resulting SKAK files were obtained with Skanlt Software for microplate readers. and the results were extracted in a form of an excel files. Each of the triplicate values of absorbance was normalised (using the values of a blank result, from a well in which hURECs were not seeded) and they were combined to calculate the slope of the standard curve.

Using the standard curves, I estimated the number of viable hURECs each day and in each passage for the *ALG8* patient and Control A. Becky Dewhurst (Newcastle University) assisted with the design of these cell viability experiments and analysed, interpreted and displayed the data and results using GraphPad software.

2.13. Cell-based assays using Human Urine-Derived Renal Epithelial Cells (hURECs) to complement genetic studies

Cell culture of hURECs, fixation, immunofluorescence (IF) experiments to observe (and compare against hURECs from control individuals) the cilia length in the hURECs of a patient presenting an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*: c.1090C>T; p.(Arg364Ter) were performed as explained in the corresponding previous sections (2.2.3, 2.3.2 and 2.3.4) above, according to standardized protocols (Molinari et al., 2018a, Molinari et al., 2020, Ramsbottom et al., 2018, Srivastava et al., 2017b, Ajzenberg et al., 2015).

Zeiss AxioImager fluorescence microscopy and Nikon (A1) confocal inverted microscopy were used for imaging of the IF experiments. Nikon Eclipse TS100 and Zeiss AxioVert 200M inverted microscopes were used for visualisation of living cells (further details were described in section: 2.8. Microscopy)

Alamar blue cell viability assay (A50100, ThermoFisher) and qPCR was performed according to standard protocols, which were explained in section: 2.12. Alamar Blue cell viability assay).

Before doing the corresponding qPCR experiment, presence of cDNA as a result of first strand cDNA synthesis was confirmed by the expression of *HPRT* via RT-PCR and gel electrophoresis (as it was showed in Figure 2.6 of section: 2.9.4. Reverse transcription-PCR (RT-PCR)). The qPCR assays (IDT) used to perform the corresponding qPCRs reactions using human cDNA are: Hs.PT.58.2872893 (*ALG8*), Hs.PT.58.3602882 (*PKD1*), Hs.PT.58.27017787 (*PKD2*) Hs.PT.39a.22214836 (*GAPDH*), Hs.PT.58v.45621572 (*HPRT1*) and Hs.PT.58.38796939 (*GUSB*).

2.14. Plasmid transfection efficiency into mammalian cells

The aim of using plasmid transfection was to study specific disease-causing *CEP120* mutations. To achieve this, firstly plasmid containing the WT *CEP120* cDNA sequence (hCEP120-EGFP, Addgene plasmid # 50382; called WT *CEP120* plasmid from now) was transfected into mammalian cells, specifically, into mouse Shh-LIGHT2 cells and human primary fibroblasts. A summary of the cell-based pipeline designed to study disease-causing *CEP120* mutations, in which plasmid transfection is performed, is shown in Figure 2.9.

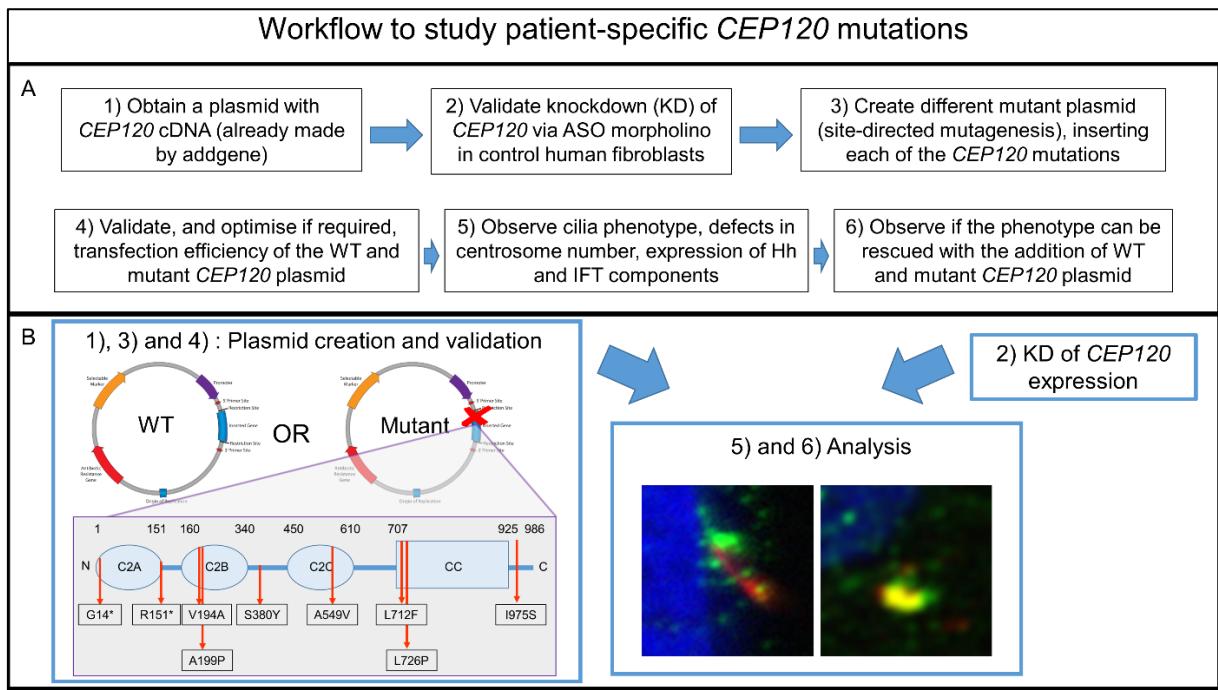


Figure 2.9. Schematic workflow of a rescue experiment to study patient-specific *CEP120* mutations. The top panel (A) shows the suggested steps (1-6). The bottom panel shows the suggested steps (1-6) in a graphic manner. WT *CEP120* plasmid (hCEP120-EGFP, ID: 50382, Addgene) was used to make different mutant plasmids via site-directed mutagenesis (New England BioLabs), each of them presenting a *CEP120* patient-specific mutation. Knockdown (KD) of *CEP120* via a *CEP120* antisense oligo morpholino (Gene Tools) was validated via qRT-PCR. It can be suggested that analysis of the phenotype caused by *CEP120* knockdown plus transfection (rescue) of *CEP120* plasmid (WT or mutant) could be analysed by Immunofluorescence and qRT-PCR. Plasmid images shown are obtained from <http://www.addgene.org/>. The zoomed diagram below the mutant plasmid image is a representation of the domain structure of *CEP120* protein (NP_694955.2), as *CEP120* is formed by three C2 conserved consecutive C2 domains (C2A, C2B and C2C) and a C terminal coil-coiled domain (CC). The nine *CEP120* mutations reported in different patients with Joubert (G14*, V194A, S380Y, A549V, L712F and L726P), Jeune (A199P), Meckel (I975S) and TCDOE (R151* and A199P) syndromes are also shown.

The WT *CEP120* plasmid was purified (Figures 2.10 and 2.11). This is an EGFP tagged plasmid that has been previously used to transfect U2OS and HEK293T cells in Lin et al (Lin et al., 2013). To create different constructs from it, the Mut hCEP120-EGFP plasmids, Site-Directed Mutagenesis (NEB) was used to incorporate each of the *CEP120* mutations reported in the literature. There is one *CEP120* variant (c.49+5_49+10del ; p.Gly+1AspfsTer14 (?)) reported in the literature (Roosing et al., 2016) that cannot be inserted in the WT *CEP120* plasmid, as it is intronic.

Interestingly, the WT *CEP120* plasmid has 4 SNPs (rs6876883, rs6595440, rs1047437, rs1047438) compared to the reference human sequence: NM_153223.

Two of these SNPs are synonymous variants (rs6876883 and rs1047438) and two of them are missense variants (rs6595440 and rs1047437) compared to the reference sequence. The rs6595440 and rs1047437 missense variants have an allele frequency of 0.4187 (24419 homozygotes) and 0.1792 (4854 homozygotes) in gnomAD respectively, which suggest that they are benign. Site directed mutagenesis was used to reverse the amino acid change of these two missense SNPs: c.1804G>C ; p.Val602Leu (reversion of rs6595440) and c.2637C>G ; p.His879Gln (Ex20) (reversion of rs1047437). Lastly, a *CEP120* missense variant was inserted: E401A of uncertain significance, which was found in homozygous state in a patient recruited in the Genomics England 100,000 Genomes Project and diagnosed with developmental macular and foveal dystrophy. Restriction mapping of each mutant plasmids was performed and confirmed by Sanger sequencing in the Mut hCEP120-EGFP plasmids that the mutations were correctly inserted correctly.

It was assessed if plasmids can be transfected in mouse Shh-LIGHT2 cells and primary human fibroblasts. This is, it was assessed if primary human cells can be transfected. I used two different GFP plasmids as GFP is expected to be widely expressed in the cell cytoplasm, which makes transfection efficiency assessment easier than by using WT *CEP120* plasmid as the *CEP120-EGFP* expression is expected to be restricted to the centrosome. One of the GFP plasmids is a pIRES2-EGFP plasmid (pIRES2-EGFP, Clontech # 6029-1, Addgene # 3178) (Figures 2.12 and 2.13 and the other is an EGFP plasmid (courtesy of Dr Magomet Aushev, Newcastle University, UK) (Figures 2.14 and 2.15).

Plasmid map of WT CEP120 plasmid (hCEP120-EGFP)

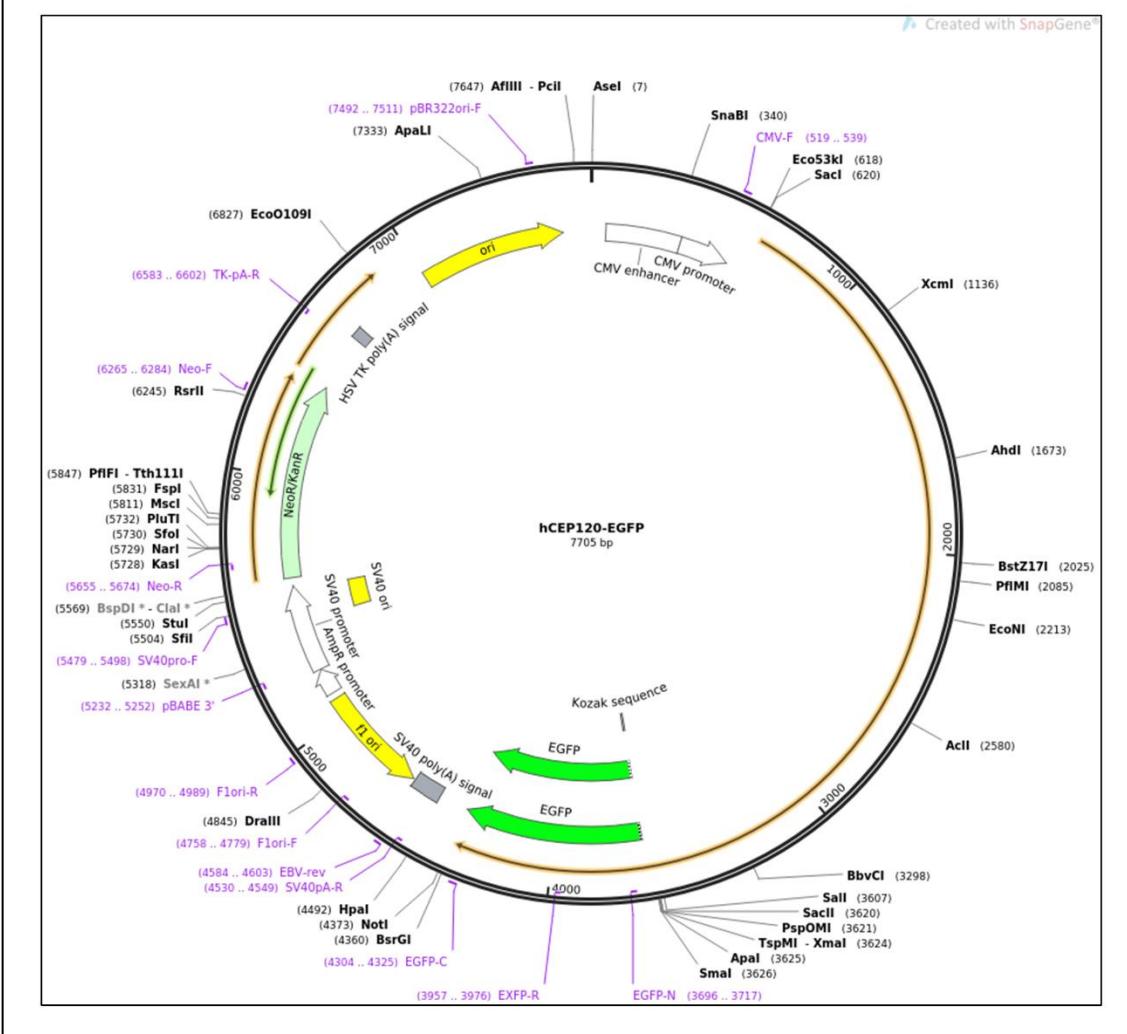


Figure 2.10. Plasmid map of WT CEP120 plasmid (hCEP120-EGFP, Addgene plasmid # 50382). WT CEP120 plasmid consists in a sequence of 7705 bp long, including the human CEP120 cDNA sequence (NM_153223) of 2962 bp. Downstream the CEP120 cDNA sequence, the plasmid contains the sequence corresponding to an EGFP tag. This plasmid, hCEP120-EGFP, was generated by Tang K. Tang et al. Plasmid map was obtained from Addgene.

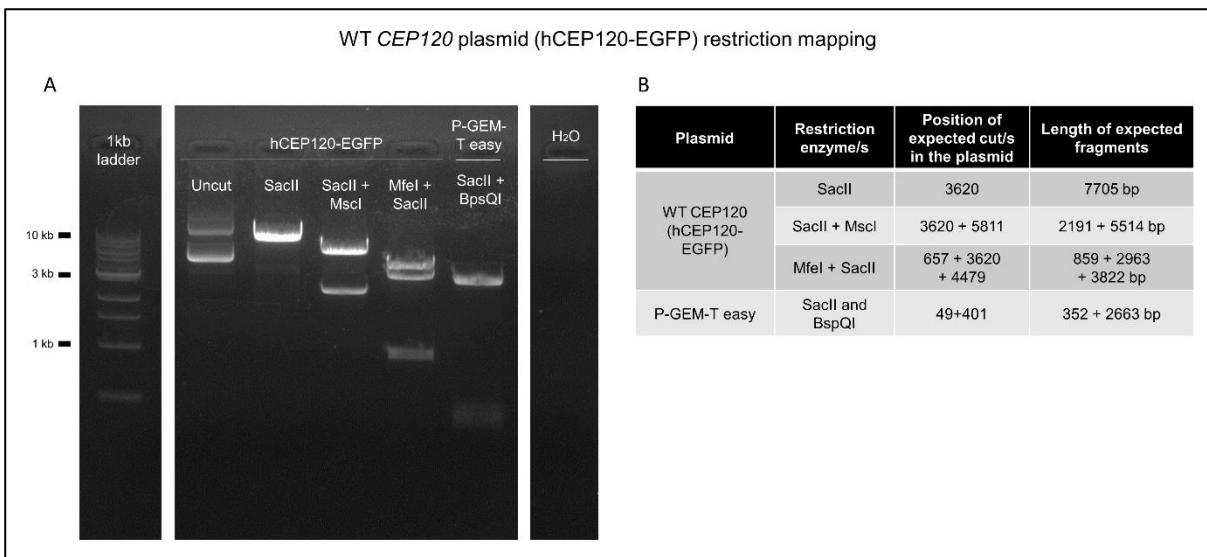


Figure 2.11. WT *CEP120* plasmid (hCEP120-EGFP) restriction mapping. The left image (A) shows gel electrophoresis showing the WT *CEP120* plasmid (hCEP120-EGFP, Addgene plasmid # 50382) treated with different combinations of restriction enzymes. The expected fragments upon the digestion of the WT *CEP120* plasmid with specific combinations of restriction enzymes are observed. When the WT *CEP120* plasmid was treated with SaclI, the fragment of 7705 bp (linearized plasmid) is observed. When the plasmid is treated with SaclI and Mscl, the corresponding two fragments of length 2191 and 5514 bp are observed. And when the plasmid is treated with MfeI + SaclI, the corresponding three fragments of length: 859, 2963 and 3822 bp are observed. As a positive control, the plasmid P-GEM-T easy (which has a length of 3015 bp, Promega) is used and cut using SaclI and BspQI, the corresponding two fragments of 352 and 2663 bp are observed. Uncut corresponds to the WT *CEP120* plasmid not treated with any restriction enzyme. A negative control using the same water used to dilute the plasmids was used to detect any plasmid contamination in the water. Gel electrophoresis (1% agarose gel). 1kb ladder used. For clarity, some lanes between the ladder and the lanes for the restriction mapping experiment and between the restriction mapping experiment and the water negative control were removed. The table on the right (B) shows the expected fragments corresponding to each of the combinations of the restriction analyses performed in the hCEP120-EGFP and P-GEM-T easy plasmids. The combination of SaclI and BspQI was applied to the P-GEM-T easy plasmid. The other three restriction enzyme reactions: SaclI, SaclI+Mscl and MfeI+SaclI were applied to the WT *CEP120* plasmid (hCEP120-EGFP) plasmid.

Plasmid map of pIRES2-EGFP plasmid

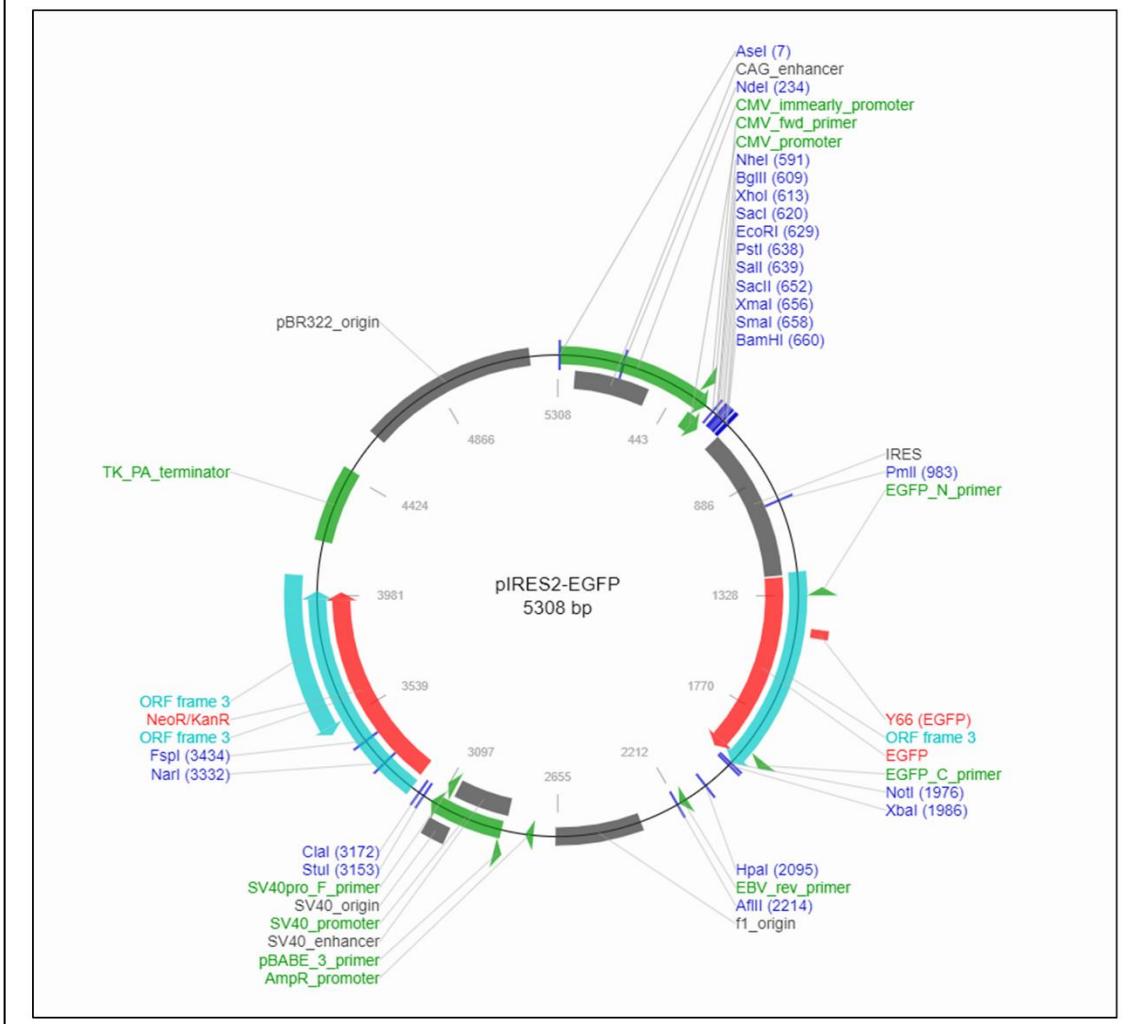


Figure 2.12. Plasmid map of pIRES2-EGFP plasmid (pIRES2-EGFP, Clontech # 6029-1, Addgene # 3178). pIRES2-EGFP plasmid consists in a sequence of 5308 bp long, including an IRES-EGFP tag sequence. Plasmid map was obtained from Addgene.

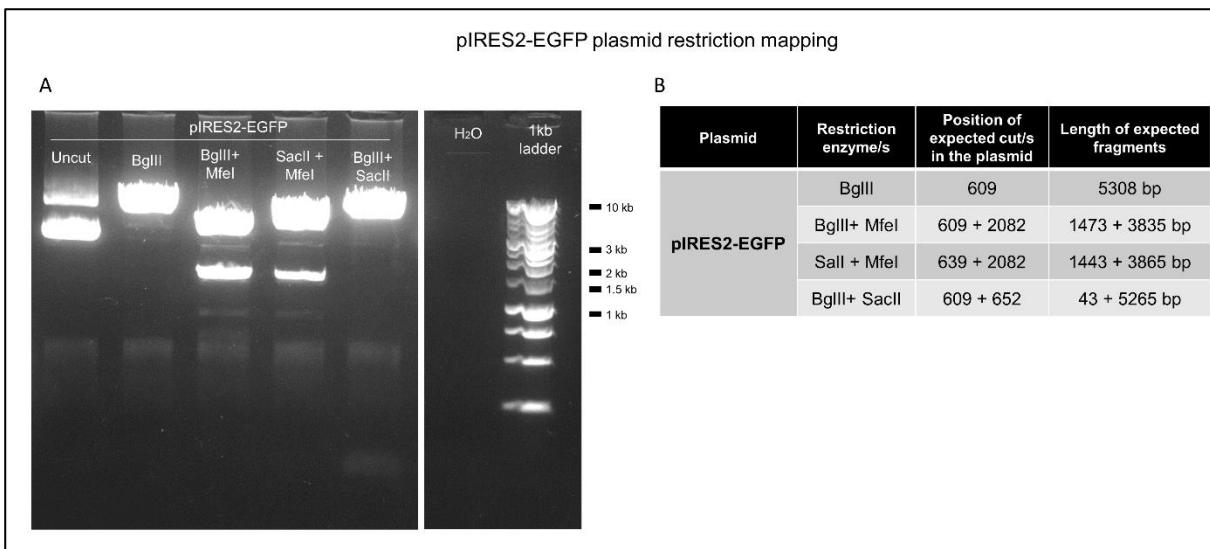


Figure 2.13. pIRES2-EGFP plasmid (pIRES2-EGFP) restriction mapping. The left image (A) shows gel electrophoresis showing the pIRES2-EGFP plasmid (pIRES2-EGFP, Clontech # 6029-1, Addgene # 3178) treated with different combinations of restriction enzymes. The expected fragments upon the digestion of the pIRES2-EGFP plasmid with specific combinations of restriction enzymes are observed. When the pIRES2-EGFP plasmid was treated with BgIII, the fragment of 5308 bp (linearized plasmid) is observed. When the plasmid is treated with BgIII and Mfcl, the corresponding two fragments of length 1473 and 3835 bp are observed. When the plasmid is treated with SacI and Mfcl, the corresponding two fragments of length: 1443 and 3865 bp are observed. And when the plasmid is treated with BgIII and Sacl, the corresponding two fragments of length: 43 + 5265 bp are observed. Uncut corresponds to the pIRES2-EGFP plasmid not treated with any restriction enzyme. A negative control using the same water used to dilute the plasmids was used to detect any plasmid contamination in the water. Gel electrophoresis (1% agarose). 1kb ladder used. For clarity, some lanes of the gel were removed. The table on the right (B) shows the expected fragments corresponding to each of the combinations of the restriction analyses performed in the pIRES2-EGFP plasmid. Brightness has been increased and contrast reduced from the original gel electrophoresis image in order to clearly visualise all the corresponding bands using PowerPoint.

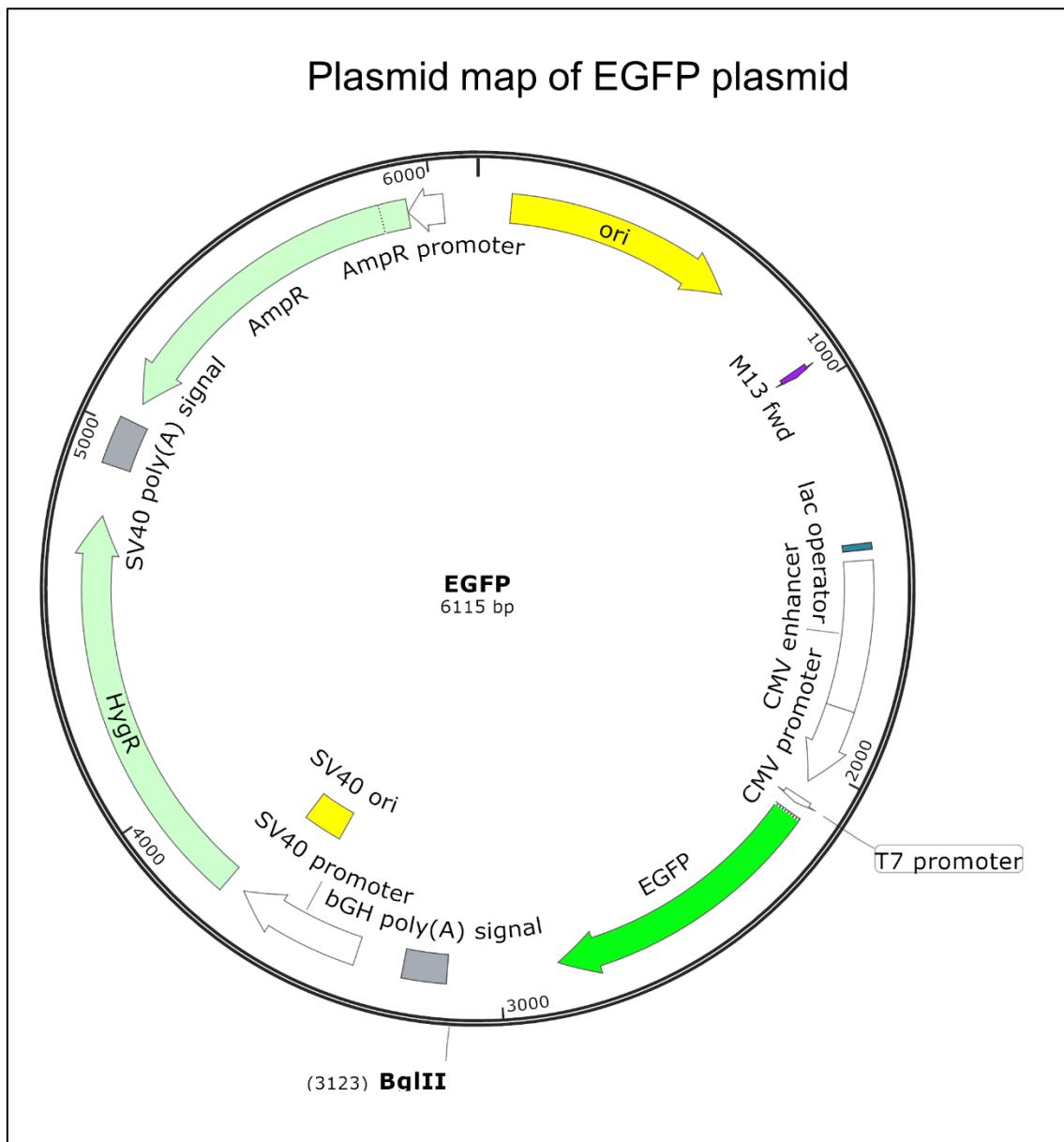


Figure 2.14. Plasmid map of EGFP plasmid. EGFP plasmid (courtesy of Dr Magomet Aushev, Newcastle University, UK) consists in a sequence of 6115 bp long, including an EGFP tag sequence. Plasmid map was generated using SnapGene Viewer software.

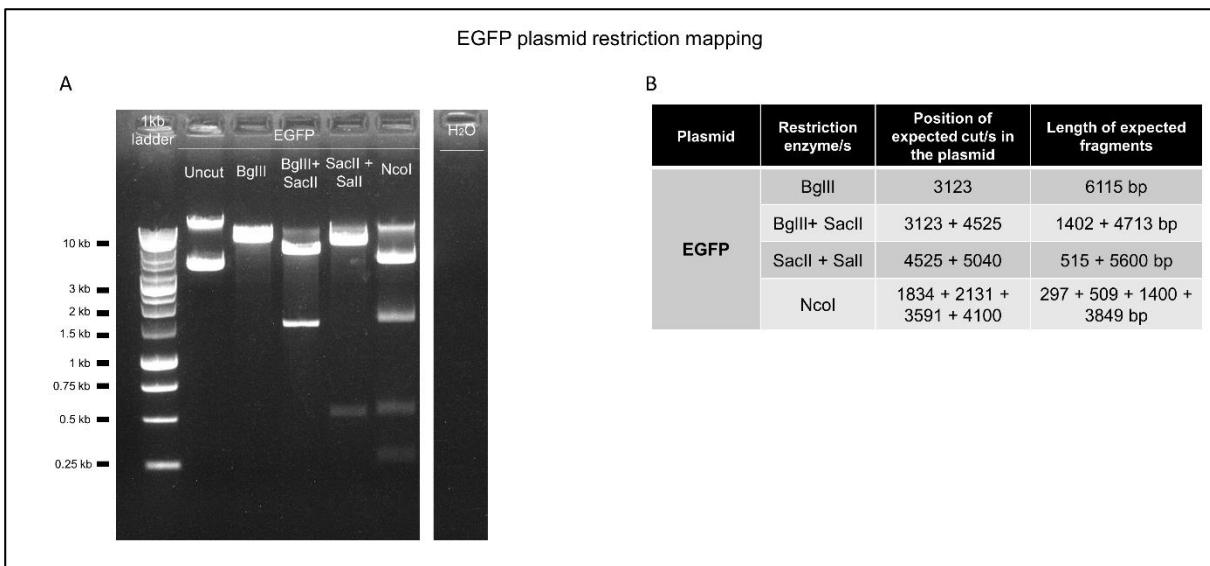


Figure 2.15. EGFP plasmid (EGFP) restriction mapping. The left image (A) shows gel electrophoresis showing the EGFP plasmid (courtesy of Dr Magomet Aushev, Newcastle University, UK) treated with different combinations of restriction enzymes. The expected fragments upon the digestion of the EGFP plasmid with specific combinations of restriction enzymes are observed. When the EGFP plasmid was treated with BgIII, the fragment of 6115 bp (linearized plasmid) is observed. When the plasmid is treated with BgIII and SacII, the corresponding two fragments of length 1402 and 4713 bp are observed. When the plasmid is treated with SacII and Sall, the corresponding two fragments of length: 515 and 5600 bp are observed. And when the plasmid is treated Ncol, the corresponding four fragments of length: 297, 509, 1400 and 3849 bp are observed. Uncut corresponds to the EGFP plasmid not treated with any restriction enzyme. A negative control using the same water used to dilute the plasmids was used to detect any plasmid contamination in the water. Gel electrophoresis (1% agarose). 1kb ladder used. For clarity some lanes of the gel were removed. The table on the right (B) shows the expected fragments corresponding to each of the combinations of the restriction analyses performed in the EGFP plasmid.

Site directed mutagenesis was used to create mutant plasmids. To do this, NEB 5-alpha Competent *E. coli* (NEB; C2987) were used, following the Q5 Site-Directed Mutagenesis Kit Protocol (NEB; E0554).

For bacterial culture for expansion of EGFP plasmid, agar plates and LB broth with the corresponding antibiotic, in this case: Ampicillin (A-9393; Sigma) (Amp) (100µg/ml working concentration) were prepared as standardized protocols (<https://www.addgene.org/protocols>), using Typtone (LP0042; OXOID), NaCl (S9888; Sigma-Aldrich), Yeast extract (Y1625; Sigma Life Extract) and ultrapure water. Agar (BP1423; Fisher Biologicals) was added to the LB broth to make the Agar plates. I used NEB 5-alpha Competent *E. coli* (High Efficiency) bacteria (C2987H; NEB) to expand the EGFP plasmid described above (Figure 2.14). This EGFP plasmid consists in a sequence of 6115 bp long, including an EGFP tag sequence.

These competent bacteria were transformed following manufacturer's instructions. I thaw a vial of NEB 5-alpha Competent *E. coli* cells on ice, then I added ~1ng of EGFP plasmid to the bacteria and incubate for 30 min on ice. A heat shock was done at 42°C for 30 seconds and then bacteria was incubated for 5 min on ice. 250 µl of SOC medium (B9020S; NEB) were added to the bacteria and incubated them at 37°C with shaking (~250 rpm) for 1 h. After this, bacteria were diluted 1:5 with SOC medium and 30 µl of these diluted bacteria were spread onto Agar plates with Amp. The plates were incubated overnight at 37°C.

The next day, 2-3 single colonies were isolated with a tip and each of them was inoculated in 5ml of LB broth and incubated overnight at 37°C. The next day 50 µl of bacterial culture in LB broth was added to 20 ml of LB broth. The next day the EGFP plasmid was isolated using a QIAGEN Plasmid Midi Kit (12143; QIAGEN) following manufacturer's instructions for high copy plasmids using *QIAGEN-tips 100*. I resuspended the plasmid in 200-250 µl of nuclease-free water. DNA concentration and purity were measured using a Ds-11 FX+ Spectrophotometer (DeNovix).

Plasmid restriction enzyme mapping was applied to the isolated EGFP plasmid. Different combinations of restriction enzymes were used. The restriction enzymes used were: BglII (R0144S; NEB), SacII (R0157S; NEB), Sall (R3138S; NEB) and Ncol (R3193S; NEB), following manufacturer's instructions. Typically, the reaction has 10 µl of plasmid DNA, 2 µl of buffer (10X NEBuffer, B7203S or CutSmart,

B7204S; NEB), 6-7 μ l of ultrapure water (depending on if one or two enzymes are used in the reaction) and 1 μ l of each the enzymes used in the reaction.

To analyze the products of the restriction mapping, restriction products were loaded into a 1% Agarose gel. ~100 μ g of the uncut plasmid (untreated) EGFP plasmid was also loaded, ensuring that the sizes of the fragments and the plasmid are the ones expected. 10 μ l of the cut/uncut plasmid DNA with 2 μ l of loading dye (6X loading dye, G190A; Promega) were loaded into the agarose gel (as it was shown in Figure 2.15).

Sanger sequencing was performed by EuroFins GATC Biotech (Germany), the sequences obtained were compared against the sequence of the WT *CEP120* plasmid and the UCSC Genome Browser on Human (GRCh38/hg38) and were visualised using BioEdit software.

Plasmid transfection into mammalian cells using Xfect transfection reagent: The transfected cells (LIGHT2 cells and primary human fibroblasts), were previously cultured until they were confluent before being transfected with the EGFP plasmid. When cells were confluent 2 μ g of EGFP plasmid (250 ng/ μ l stock) using 0.3 μ l of Xfect Polymer (Xfect transfection Reagent, 631318; Takara Bio) was transfected following manufacturer's instructions. Other transfection reagents were used to transfect 2 μ g of EGFP plasmid: jetOPTIMUS DNA Transfection Reagent (101000051; Polyplus) and Fugene HD Transfection Reagent (E2311; Promega) following manufacturer's instructions.

After two days from these transfections, living cells were visualized using Zeiss AxioVert 200M inverted fluorescent microscope. EGFP positive cells (ie: cells transfected with EGFP plasmid) were observed using the Fluorescein (FITC) fluorescent filter and brightfield imaging. Images were obtained and processed using AxioVs40x64 v.4.9.1.0 and ImageJ software.

2.15. Use of the Genomics England 100,000 Genomes Project database

Genomics England 100,000 Genomes Project database was used to perform different *in silico* studies. I accessed the data available in the Genomics England 100,000 Genomes Project. This project is managed by Genomics England (a Department of Health-owned company set up in 2013). Since the first participants were recruited into the project, more than 100,000 genomes have been sequenced.

These genomes have been sequenced using whole-genome sequencing (WGS) technology delivered by Illumina.

The clinical and genetic data in the Genomics England 100,000 Genomes Project can be assessed by an approved community of researchers and clinicians. All the participants recruited in the project provided written consent to access their anonymized clinical and genomic data for research purposes.

WGS was performed by Illumina and Genomics England. Illumina TruSeq DNA PCR-Free sample preparation kit (Illumina, Inc.) was used to prepare samples for WGS and an Illumina HiSeq 2500 sequencer was used for WGS, generating a mean depth of 45x (range from 34x to 72x) and greater than 15x for at least 95% of the reference human genome. WGS reads were aligned to the reference genome: Genome Reference Consortium human genome build 37 (GRCh37) or 38 (GRCh38) using Isaac Genome Alignment Software. The genomic data is supplied in the form of BAM and VCF files provided by Genomics England.

VCF stands for Variant Call Format, a VCF file stores genomic variations of one individual. A BAM binary file is the compressed binary version of a Sequence Alignment/Map (SAM) file, which contains aligned sequences up to 128 Mb. These files have passed initial in-house quality controls of sequencing quality and coverage.

The number of sequenced genomes and the clinical information available has been increasing since the first data release was made available for research users in 2017. Currently new data is being released every 3-5 months by Genomics England.

As a reference for the amount participants recruited and genomes sequenced in the Genomics England 100,000 Genomes Project, the main Program Data Release version 9 (v9, dated 2nd April 2020), provided access to the sequencing of 108431 genomes, 74233 of those genomes are from 71672 participants that are included in the rare disease program. The other 34198 genomes correspond to participants included in the cancer program

(<https://cnfl.extge.co.uk/pages/viewpage.action?pageId=147659370>). I only considered the patients and genomes included in the Rare Disease Program of the Genomics England 100,000 Genomes Project, which includes patients with a wide diversity of different pathologies and their relatives. Regarding the 71672 rare disease participants, 33403 are probands (affected individuals: patients), the rest of

the participants are the relatives of those probands (which can be affected or not). Of note, the genome sequencing of both or one of the parents is not always available.

The tools used to access the genomic and clinical data, were located in the Genomics England Research Environment. One of these tools is Labkey, where the clinical information can be found, including the HPO terms and clinical diagnosis of each of the patients considered. Participant Explorer tool was used to find more detailed clinical information of some patients.

Labkey contains the tiering data and the pathway to access the genomic VCF files to analyze. The tiering data is created by the Genomics England Bioinformatics Rare Diseases Interpretation pipeline.

The genomic variants of the patients are found in each corresponding patient-specific genomic VCF file. I used a script to extract genomic variants from the VCF files. This script (called from now: modified extract variants by coordinate script, and shown in Appendix 1) is a script I modified from the one provided by Genomics England (<https://cnfl.extge.co.uk/display/GERE/Extract+variants+by+coordinate>). The output of this script is a list of variants (SNVs and indels) that are located in a list of specific regions and genomic VCF files selected (ie: the input files are a list of regions and a list of genomic VCF files). Furthermore these variants have a PASS filter, have a coverage >10 and a genotype quality >20. Genomics England airlock tool was used to extract the results from the Research Environment to an external location according to Genomics England guidance.

2.15.1. Search for patients presenting with biallelic pathogenic variants in CEP120

I searched for patients in the Genomics England 100,000 Genomes Project presenting with biallelic *CEP120* mutations responsible for the patient's phenotype. This includes, searching for all biallelic *CEP120* variants in the tiering data and participants presenting any of the *CEP120* variants already reported in the literature using the WGS data available for the participants included in the Rare Disease Program. The main release v9, dated 2nd April 2020 of the Genomics England 100,000 Genomes Project was used.

I searched in the tiering data (which prioritises variants using a gene panel-based method) for biallelic variants in *CEP120*. Ensembl Variant Effect Predictor (VEP)

(Yates et al., 2020) and Genome Aggregation database gnomAD were used to ensure that these *CEP120* variants are rare and affect the coding region of *CEP120*, discarding patients with *CEP120* intronic variants (excluding the ones that are described as splice_donor, splice_acceptor or as splice_acceptor).

Additionally, I used the script: extract variants by coordinate (modified from the Extract variants by coordinate script from Genomics England available <https://research-help.genomicsengland.co.uk/display/GERE/Extract+variants+by+coordinate>), to find the patients presenting any of the 9 different *CEP120* variants described in the literature (Roosing et al., 2016, Shaheen et al., 2015). This script searched these specific variants (SNV and indels < 50bp) in each of the patients (probands) of the Rare Disease Project cohort. The patients that were biallelic for these variants were selected and proceeded as explained before.

Labkey tool was used to find the clinical phenotype (diagnosed disease, HPO terms, age, sex, affected parents and Genomic Medicine Centres (GMC) exit questionnaire) of the patients with these biallelic variants in *CEP120*. The GMC exit questionnaire provides information about if a patient has been genetically solved by Genomics England (ie: if there is a genetic cause found for the patient's phenotype).

For these patients with biallelic variants in *CEP120*, I also looked for variants in other genes that could explain the phenotype of these patients. To do this I selected Tier 1 and Tier 2 variants and any tier variant located in any of the known 302 ciliary genes (according to Syscilia Gold Standard list of ciliary genes available at <http://bioinformatics.bio.uu.nl/john/syscilia/ciliacarta/> and shown in Appendix 2) (Van Dam et al., 2013). Tier 1 and Tier 2 variants are loss-of-function and missense variants in genes associated with the patient's phenotype respectively, these genes are classified as a "green gene" in panel app (<https://panelapp.genomicsengland.co.uk/>), which indicates that they have high level of evidence for disease association.

Exomiser automated variant prioritisation framework (prioritises variants based on predicted pathogenicity, allele frequency and association of phenotypes of diseases and cross-species phenotype comparison) (Robinson et al., 2014) available in the Genomics England Research Environment was used to find other putative disease-causing variants. Exomiser provides a higher score for the variants that are more

likely to be potential disease-causing variants and ranks the variants in way that the variant with the higher score is ranked “1”. Biallelic variants ranked from 1 to 5 were selected.

Variants were analysed in terms of pathogenicity, frequency and segregation (if sequencing information of any relative/s was available). All these variants were annotated and their pathogenicity was analysed using SIFT, PolyPhen, Mutation taster, Combined Annotation Dependent Depletion (CADD) scores, using Ensembl and Varsome (Kopanos et al., 2019).

Of note, SIFT and PolyPhen scores predict pathogenicity of missense variants, SIFT scores range from 0 to 1, the lower the score the more likely the missense variant is affecting protein function and the missense variant is described as damaging if the score is ≤ 0.05 , and is described as tolerated if the score is > 0.05 . PolyPhen scores range from 0 to 1, the higher the score the more likely the missense variant is affecting protein function. CADD scores predicts deleteriousness using a combination of multiple scores (Rentzsch et al., 2019), usually a CADD score higher than 25 indicates the corresponding SNP or indel is deleterious. I discarded from the analysis: single heterozygous variants and biallelic variants in which both alleles were described as benign, tolerated or were common (> 0.05 allele frequency) in the general population.

2.15.2. Search for patients presenting with biallelic pathogenic variants in TOGARAM1

To search for variants in *TOGARAM1*, I used the Genomics England 100,000 Genomes Project (main release v8, dated 28th November 2019). I searched in the tiering for biallelic *TOGARAM1* variants present in patients diagnosed with Joubert syndrome or patients with phenotypes that suggested a primary ciliopathy. To ensure that these variants target the exonic regions and are rare in allele frequency, VEP and gnomAD were used. Labkey was used to find the clinical phenotype and information of the patients with these biallelic variants in *TOGARAM1*.

For these patients with biallelic variants in *TOGARAM1*, I also looked for variants in other genes that could explain the phenotype of these patients. To do this I selected Tier 1 and Tier 2 variants and any tier variant located in any of the known 302 genes ciliary genes (according to Syscilia Gold Standard list of ciliary genes ; <http://bioinformatics.bio.uu.nl/john/syscilia/ciliacarta/> and shown in Appendix 2) (Van

Dam et al., 2013) and a self-curated list of genes associated with Joubert syndrome which is the same list as the one shown Table 1.2, in section: 1.3.1: *Molecular genetics of Joubert syndrome*).

All the variants found were annotated and their pathogenicity and frequency was assayed using SIFT, PolyPhen, CADD values, found using gnomAD, Ensembl, NCBI ClinVar and Varsome. Variants that were described as benign or were common (> 0.05 allele frequency) in the general population were discarded.

BAM files of patients of interest were viewed using Integrated Genome Viewer (IGV) (Latour et al., 2020). IGV viewer (Robinson et al., 2011) allowed to visualize the genomic data of patients within the Genomics England Research Environment.

Once patients presenting with biallelic *TOGARAM1* variants that are or could be disease-causing are found, I searched for other putative disease-causing variants that can be responsible for the phenotype of these patients.

When necessary and possible, referring physician was contacted and Sanger sequencing was requested to confirm the presence of putative disease-causing *TOGARAM1* variants of the corresponding patient.

2.15.3. Search for patients presenting with monoallelic pathogenic variants in genes encoding for proteins functioning in the endoplasmic reticulum: DNAJB11, GANAB, ALG5, ALG8 and ALG9

The approach followed (explained in Figure 2.16) is similar to the approach followed in the previous section (2.15.2) to detect biallelic disease-causing variants in *TOGARAM1* and *CEP120*.

I used the Genomics England 100,000 Genomes Project, in particular, the data release version 7 (dated 25th July 2019) for *DNAJB11* and *GANAB* genes and the data release version 14 (dated 27th January 2022) for the *ALG5*, *ALG8* and *ALG9* genes.

In this section I looked for variants in the genes that are responsible for the disease phenotype in an autosomal dominant pattern (ie: I was looking for single heterozygous pathogenic variants), rather than variants leading to disease phenotypes in an autosomal recessive pattern.

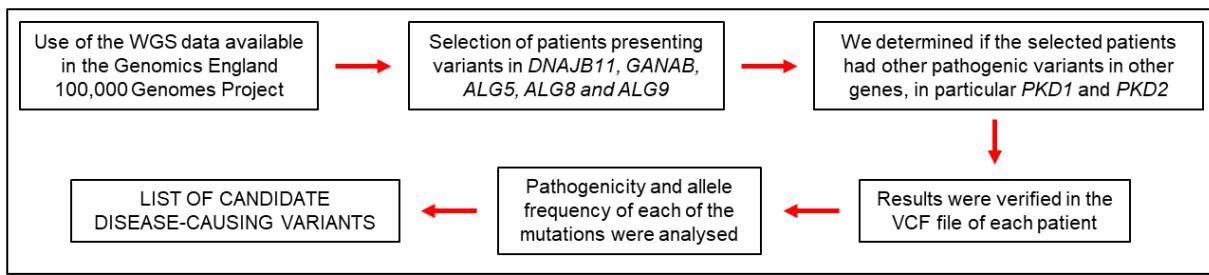


Figure 2.16. Pipeline to detect candidate disease-causing variants in *DNAJB11*, *GANAB*, *ALG5*, *ALG8* and *ALG9*. Whole genome sequencing (WGS) data from Labkey tool of the Genomics England 100,000 Genomes Project was used. I searched in the tiering data tool for variants in these corresponding genes: *DNAJB11*, *GANAB*, *ALG5*, *ALG8* and *ALG9*.

Data available in the Genomics England 100,000 Genomes Project was accessed to find patients with monoallelic pathogenic variants in *DNAJB11*, *GANAB*, *ALG5*, *ALG8* and *ALG9*, which are genes encoding for proteins functioning in the endoplasmic reticulum (ER). The data release version 7 (dated 25th July 2019) was accessed to find pathogenic variants in *DNAJB11* and *GANAB* genes present in any of the 3892 probands with various kidney and urinary tract disorders and/or Ciliopathies. The data release version 14 (dated 27th January 2022) was accessed to find pathogenic variants in *ALG5*, *ALG8* and *ALG9* genes present in any of the 3891 probands with various kidney and urinary tract disorders and/or ciliopathies.

I first used the “tiering data” available in Labkey tool to search for variants in the following genes: *DNAJB11*, *GANAB*, *ALG5*, *ALG8* and *ALG9*. For simplicity, regarding the patients found to present variants in *DNAJB11* or *GANAB* genes, patients whose both parents did not show any clinical phenotype were discarded.

To find disease-causing variants in *ALG5*, *ALG8* and *ALG9*, I exclusively looked for variants described as nonsense, frameshift, splice donor or splice acceptor, while for *DNAJB11* or *GANAB* I also looked for missense variants. Once patients presenting variants of interest in any of these genes were found, I searched for other putative disease-causing variants that can be responsible for the phenotype of these patients. For each of the patients investigated, I searched for other pathogenic variants in other genes, in particular *PKD1* and *PKD2* and other genes that are associated, or could be potentially associated, with ADPKD phenotypes (see Table 2.4).

The phenotype of each of the patients was described using the information available in Labkey. Patients without a kidney or liver phenotype were not included in the analysis. Labkey was used to find information, if available, about if these patients

were already genetically solved by Genomics England and if the proband's parents carry any of the variants investigated.

Variants were annotated according their canonical transcript using VEP. The nomenclature, pathogenicity and allele frequency of these variants was assayed using the Varsome, NCBI ClinVar, gnomAD and Ensembl clinical databases creating a list of candidate disease-causing variants for each of the patients studied.

Gene	Ensembl Gene ID	Phenotype in OMIM phenotypic series: Polycystic kidney disease - PS173900	Locus MIM number	Ensembl transcript ID	Number of coding exons	Transcript length	Protein length
ALG1	ENSG00000033011	Not in PKD phenotypic series	605907	ENST00000262374.10	13	3900 bp	464 aa
ALG2	ENSG00000119523	Not in PKD phenotypic series	607905	ENST00000476832.2	2	2808 bp	416 aa
ALG3	ENSG00000214160	Not in PKD phenotypic series	608750	ENST00000397676.8	9	1552 bp	438 aa
ALG5	ENSG00000120697	Not in PKD phenotypic series	604565	ENST00000239891.4	10	1219 bp	324 aa
ALG6	ENSG00000088035	Not in PKD phenotypic series	604566	ENST00000263440.6	14	3325 bp	507 aa
ALG7	ENSG00000172269	Not in PKD phenotypic series	191350	ENST00000354202.9	9	1913 bp	408 aa
ALG8	ENSG00000159063	Not in PKD phenotypic series	608103	ENST00000299626.10	13	1637 bp	526 aa
ALG9	ENSG00000086848	Not in PKD phenotypic series	606941	ENST00000616540.5	15	6158 bp	618 aa
ALG10	ENSG00000139133	Not in PKD phenotypic series	618355	ENST00000266483.7	3	2913 bp	473 aa
ALG11	ENSG00000253710	Not in PKD phenotypic series	613666	ENST00000521508.2	4	6510 bp	492 aa
ALG12	ENSG00000182858	Not in PKD phenotypic series	607144	ENST00000330817.11	9	5330 bp	488 aa
ALG13	ENSG00000101901	Not in PKD phenotypic series	300776	ENST00000394780.8	27	4113 bp	1137 aa
ALG14	ENSG00000172339	Not in PKD phenotypic series	612866	ENST00000370205.6	4	9375 bp	216 aa

<i>DNAJ B11</i>	ENSG00000090520	PKD 6 with or without polycystic liver disease	6180 61	ENST00000265028.8	10	1640 bp	358 aa
<i>GANAB</i>	ENSG00000089597	PKD 3	6006 66	ENST00000346178.8	25	3906 bp	966 aa
<i>HNF1B</i>	ENSG00000275410	Not in PKD phenotypic series	1899 07	ENST00000617811.5	9	2790 bp	557 aa
<i>IFT140</i>	ENSG00000187535	Not in PKD phenotypic series	6146 20	ENST00000426508.7	29	5232 bp	1462 aa
<i>LRP5</i>	ENSG00000162337	Not in PKD phenotypic series	6035 06	ENST00000294304.12	23	5177 bp	1615 aa
<i>OFD1</i>	ENSG00000046651	Not in PKD phenotypic series	3001 70	ENST00000340096.11	23	3612 bp	1012 aa
<i>PKD1</i>	ENSG00000008710	PKD 1	1739 00	ENST00000262304.9	46	14140 bp	4303 aa
<i>PKD2</i>	ENSG00000118762	PKD 2	6130 95	ENST00000237596.7	15	5089 bp	968 aa
<i>PRKC SH</i>	ENSG00000130175	Not in PKD phenotypic series	1770 60	ENST00000592741.5	16	2100 bp	535 aa
<i>SEC61A1</i>	ENSG00000058262	Not in PKD phenotypic series	6092 13	ENST00000243253.8	12	3568 bp	476 aa
<i>SEC61B</i>	ENSG00000106803	Not in PKD phenotypic series	6092 14	ENST00000223641.5	4	564 bp	96 aa
<i>SEC63</i>	ENSG00000025796	Not in PKD phenotypic series	6086 48	ENST00000369002.9	21	6430 bp	760 aa
<i>UMOD</i>	ENSG00000169344	Not in PKD phenotypic series	1918 45	ENST00000396134.6	11	2408 bp	673 aa

Table 2.4. List of autosomal dominant polycystic kidney disease (ADPKD) genes and ADPKD candidates, including the rest of *ALG* genes. *ALG7* is also known as *DPAGT1*. *DZIP1L* and *FCYT* are genes that are in the PKD phenotypic series but are not included in this list because they are associated with autosomal recessive PKD rather than ADPKD.

2.15.4. Estimation of the proportion of loss-of-function (LoF) variants in *ALG* genes

I used the Genomics England 100,000 Genomes Project to determine if there is an enrichment of heterozygous alleles in any of the *ALG* genes in a population of patients with cysts in kidney or liver, compared to a non-cystic control population. Using Labkey database (data release v14, dated 27th of January 2022) I obtained the corresponding IDs and VCF files corresponding to the participants to create the case

(probands that present cysts in kidney and/or liver) and control (probands that do not present cysts in kidney or liver) populations used in this study. This allowed to calculate the proportion of loss-of-function alleles in each of the populations (Figure 2.17).

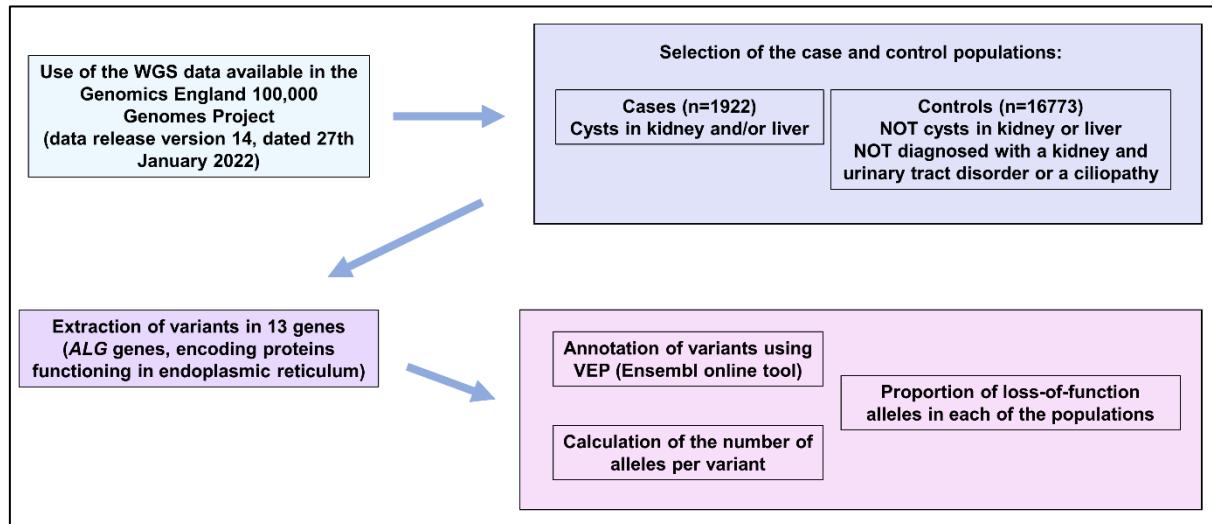


Figure 2.17. Pipeline to investigate if there is an enrichment of heterozygous alleles in any of the *ALG* genes in a population of patients with cysts in kidney or liver, compared to a non-cystic control population. From the literature it is known that heterozygous variants in *ALG8* and *ALG9* have been associated with kidney and liver cystic phenotypes (Besse et al., 2019, Besse et al., 2017). I extracted whole genome sequencing (WGS) data from Labkey and Participant explorer tools of the Genomics England 100,000 Genomes Project (data release version 14, dated 27th January 2022).

Firstly, the participants who are described as “Probands” and are sequenced using the Genome Reference Consortium human genome build 38 (GRCh38).

To create a population of cases and a population of controls, I firstly selected a population of cases consisting in 1922 probands that present cysts in kidney and/or liver. The population of controls consists in 16773 probands without any cysts in kidney or liver, and that are not diagnosed with a renal disease, undiagnosed metabolic disorders a ciliopathy or a disease consisting in similar phenotypes to the ones observed in ciliopathies (e.g. rod-cone dystrophy, hereditary ataxia, unexplained skeletal dysplasia). Using Labkey and R studio I obtained a list of the genomics VCF files corresponding to these participants in the case and control probands, each of the VCF files corresponding to a different proband.

The modified extract variants by coordinate script (shown in Appendix 1, it was modified from the “extract variants by coordinate” script from Genomics England available resources: <https://research->

help.genomicsengland.co.uk/display/GERE/Extract+variants+by+coordinate) was used to find the patients presenting variants in the *ALG* genes. The variants from the 13 *ALG* genes were extracted using this script designed to extract variants from each the genomic VCF file corresponding of each of the probands selected. This script searched exclusively for exonic variants and intronic regions affecting splicing (in the 2 base region at the 3' and 5' ends of a intron) in each of the patients (probands) of the case and control population that were created.

Variants were annotated using VEP according to their canonical transcript. For each of the variants the number of alleles present in the population of cases and in the population of controls were calculated. As the variants are annotated, it can be determined which variants are loss-of-function (LoF) variants and the proportion of LoF alleles in each of the populations can be calculated. The same way, The proportion of missense alleles in each of the populations were calculated.

Statistical tests were performed to evaluate if the difference in the proportion of alleles (LoF vs non-LoF or missense vs non-missense) is significant in the case population compared to the control population. Using R studio and GraphPad I plotted the data and applied a Fisher's exact test to determine if the difference between the observed and expected frequencies of LoF alleles (or missense alleles) in each of the *ALG* genes is significant.

Chapter 3. Characterisation of *CEP120* expression patterns in human embryonic and fetal tissue using the MRC-Wellcome Human Developmental Biology Resource (HDBR), and comparison with *ARL3* expression patterns

3.1. Introduction and aims

Joubert syndrome (JBTS) is a primary ciliopathy characterised by a cerebellar and brainstem malformation, described as the “molar tooth sign” (MTS) (Doherty, 2009). JBTS patients are also described to develop hypotonia in infancy with later development of ataxia and developmental delays. JBTS patients may also develop extra-neurological phenotypes, such as renal and retinal defects (Valente et al., 2008, Parisi, 2019, Radha Rama Devi et al., 2020). The cerebello-retinal-renal phenotype of JBTS can be grouped as Joubert syndrome and related disorders (JSRD), although the terms JSRD and JBTS are often used indistinctly (Alkanderi et al., 2018, Radha Rama Devi et al., 2020).

JSRD and Jeune syndrome (JATD) are multisystem ciliopathy disorders with overlapping phenotypes (Roosing et al., 2016, Focsa et al., 2021, Malicdan et al., 2015). Renal phenotypes in JSRD can vary from renal cysts, interstitial fibrosis to tubulointerstitial kidney disease and they develop in about 25% of JSRD patients (Paprocka and Jamroz, 2012). The renal phenotype in some of these patients may progress and lead to end-stage renal disease (Srivastava et al., 2017b).

There are over 40 genes associated with JSRD (MIM: PS213300) (Gana et al., 2022). The clinical phenotypes in JSRD that can involve multiple organs and there is phenotypic and genetic overlap with other primary ciliopathies such as JATD (MIM: PS208500) and Bardet-Biedl syndrome (MIM: PS209900) (Fansa and Wittinghofer, 2016). Mutations in *CEP120* (Roosing et al., 2016) or *ARL3* (Alkanderi et al., 2018) are some of the genetic causes of JSRD.

Mutations in ciliopathy genes, including *CEP120* and *ARL3*, can lead to a heterogeneous spectrum of clinical phenotypes that can affect multiple organs (Alkanderi et al., 2018, Roosing et al., 2016, Gana et al., 2022, Shaheen et al., 2016). The role of tissue expressivity during human development in the phenotypic heterogeneity of primary ciliopathies is not fully understood (Focsa et al., 2021, Reiter and Leroux, 2017).

CEP120 and *ARL3* have been selected to study the molecular genetics of JBTS, as they are both associated with JBTS and are suitable to study why mutations in different genes can cause the same ciliopathy. Mutations in *CEP120* and *ARL3* can also lead to other primary ciliopathies, such as JATD (Roosing et al., 2016, Shaheen et al., 2015) and retinal dystrophy (Fu et al., 2021, Holtan et al., 2019, Ratnapriya et al., 2021, Sheikh et al., 2019, Strom et al., 2016), respectively.

One area in the study of the molecular genetics of JBTS and other primary ciliopathies is tissue expressivity (Reiter and Leroux, 2017, Malicdan et al., 2015, Gana et al., 2022). I studied the role of tissue expressivity during human development, focussing on *CEP120* and *ARL3*.

Some studies have investigated protein expression of *Cep120* in mouse tissues, but not in human tissues (Wu et al., 2014, Xie et al., 2007). Expression of *Arl3* has been studied in mouse tissues (Schrick et al., 2006, Hanke-Gogokhia et al., 2016). To this date, there are not studies investigating *ARL3* tissue expression during human development at RNA or protein level.

Other ciliopathy genes and pleiotropic genes, which are similar to *CEP120* and *ARL3* in function and associated disease phenotypes, could have been included in this study, such as *KIAA0586* (Bashford and Subramanian, 2019) and *AHI1* (Cheng et al., 2012, Muñoz-Estrada and Ferland, 2019, Adams et al., 2012). However, because of a limitation of available resources and time only two ciliopathy genes: *CEP120* and *ARL3*, were selected for this study of expression patterns in human development.

These two genes exemplify the genetic and phenotypic complexity of JSRD, in which *CEP120* and *ARL3* have different roles and locations: where the former functions at the centrosome/basal body and the latter functions at the primary cilium (Mahjoub et al., 2010, Alkanderi et al., 2018).

Cep120 was first described by Xie et al. (Xie et al., 2007). *CEP120/Cep120* has been described to be expressed in multiple embryonic mouse tissues such as brain, kidney and lungs (Xie et al., 2007). Additionally, in mouse embryonic brain, *Cep120* has been found to be highly expressed compared to mouse adult brain (Xie et al., 2007). Furthermore, it was observed that *CEP120* localises, in cycling cells, along the centrioles, preferentially at the daughter centriole, and it was suggested that *CEP120* has a role in centrosome assembly (Mahjoub et al., 2010). Further studies suggested that this asymmetry was due to the interaction of *Cep120* with the mouse protein

Talpid3 (Talpid3 is known in humans as KIAA0586) (Wu et al., 2014). It was suggested that CEP120 may have a role in inhibiting maturation of daughter centrioles, which it is required to maintain centrosome homeostasis (Betleja et al., 2018).

Only few patients (<10) have been described with disease-causing biallelic mutations in *CEP120* leading to JSRD, JATD and Meckel syndrome (MKS) syndromes (Roosing et al., 2016, Shaheen et al., 2015). These *CEP120* disease-causing mutations lead to JSRD and JATD in 4 and 3 patients respectively, and severe overlapping ciliopathy phenotypes including tectocerebellar dysraphia with occipital encephalocele (TCDOE), MKS and oro-facial-digital (OFD) syndromes (Roosing et al., 2016) in 2 patients. Of note, one of the clinical phenotypes observed in the *CEP120* patient diagnosed with TCDOE is the MTS feature characteristic of JBTS (Roosing et al., 2016).

ADP-ribosylation factor-like 3 (ARL3) is a ciliary protein also associated with JSRD. ARL3, is a low molecular weight GTP-binding protein, which belongs to the RAS superfamily (Kahn et al., 1992). Mutations in *ARL3* have been associated with JSRD (Alkanderi et al., 2018) and retinitis pigmentosa (Holtan et al., 2019, Strom et al., 2016).

The MRC-Wellcome Human Developmental Biology Resource (HDBR) consists in a collection of human embryonic and fetal material (from 3 to 20 weeks of development). HDBR is managed from the Institute of Genetic Medicine in Newcastle and the Institute of Child Health in London. HDBR has recently been involved in the study of gene expression of several ciliopathy genes: *CEP290* (Cheng et al., 2012), *AHI1* (Cheng et al., 2012) and *CEP164* (Devlin et al., 2020) and other projects studying early human cerebellum development and gene expression patterns during embryonic brain development (Lindsay et al., 2016, Alzu'bi et al., 2017, Haldipur et al., 2019).

RNAscope is an RNA *in situ* detection platform from Advanced Cell Diagnostics (ACD) for detection of target RNA within intact cells (Figure 3.1). The workflow of this technique is based on four processes: pre-treatment (fixation and permeabilization), probe hybridization, signal amplification and visualisation (Wang et al., 2012). RNAscope has high specificity given that two RNA probes (described as Z probes) are required to hybridize to the specific target sequence in tandem, for signal

amplification to occur. RNAscope technique has high sensitivity given that only 3 double Z probes (out of the typically 20 double Z probes designed for each RNA target) are required to hybridize for a single RNA molecule to be detected (Wang et al., 2012, Grabinski et al., 2015).

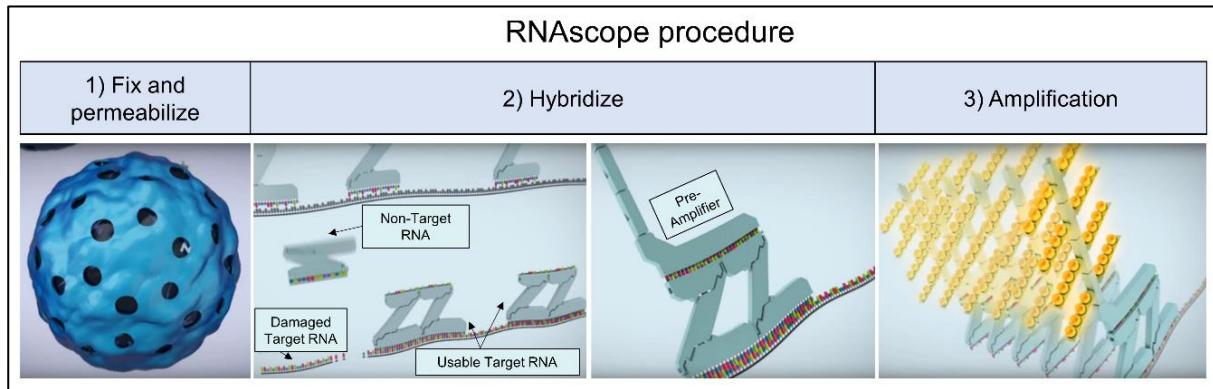


Figure 3.1. Representative figure of the RNAscope procedure. 1) Fixing and permeabilizing of the samples, generally Formalin-Fixed Paraffin-Embedded (FFPE) tissue or cultured cells, so that the designed RNA probes can enter cells. 2) Double-Z RNA target probes hybridize to the specific target sequence and pre-amplifiers bind to double Z probes. 3) Signal amplification occurs as fluorescent molecules or chromogenic enzymes bind to the amplifiers. Results are visualised using fluorescence microscopy (fluorescent label) or bright-field microscopy (enzyme label) (Wang et al., 2012). Images obtained and modified from <https://acdbio.com>.

Protein and RNA expression in different tissues can be studied *in vitro* using various techniques such as Western blot to relatively quantify protein amount from different tissues (Xie et al., 2007, Bashford and Subramanian, 2019) and immunofluorescence (Bashford and Subramanian, 2019) or immunohistochemistry (IHC) (Alzu'bi and Clowry, 2019, Devlin et al., 2020) to reveal the location of a protein at tissue level.

The study of the tissue expressivity via immunohistochemistry of some ciliopathy genes such as: *CEP164* (Devlin et al., 2020), *AHI1* and *CEP290* (Cheng et al., 2012) has provided to the scientific community highly valuable insights into the molecular disease pathogenesis associated with these genes and primary ciliopathies.

In the last few years, the HDBR has optimised the use of IHC to investigate human development. Some of the antibodies HDBR used for IHC include: PAX6 and OLIG2, which are transcription factors associated with proliferation during brain development (Alzu'bi and Clowry, 2019, Alzu'bi et al., 2017, Alzu'bi et al., 2022). JBTS is a developmental disease in which the cerebellar development is affected (Devlin et al., 2020, Bachmann-Gagescu et al., 2015a). Comparing the expression patterns of ciliopathy genes such as *CEP120* or *ARL3* with the antibody stainings of PAX6 and

OLIG2 antibodies can provide insights at tissue level, regarding cerebellar development and its correlation with the expression patterns of these ciliopathy genes (*CEP120* and *ARL3*).

The HDBR resource has extensively studied human brain development via IHC *in vitro* experiments (Devlin et al., 2020, Alzu'bi and Clowry, 2020, Alzu'bi et al., 2017, Lindsay et al., 2016). HDBR members provided me with advice and help to perform the IHC protocol on human developing tissues, which is a unique opportunity for me to investigate expression of ciliopathy genes in human embryonic and foetal tissues, with a focus on the developing organs affected in primary ciliopathies.

RNAscope, is RNA *in situ* technique (Grabinski et al., 2015, Wang et al., 2012) optimised and performed by the HDBR resource at Newcastle University (Alzu'bi et al., 2022). This technique can be applied to characterise the expression of ciliopathy genes, such as *CEP120* and *ARL3*. The methodology and principles of RNAscope are explained in Chapter 2: Materials and Methods (section: 2.4. Use of the RNA *in situ* detection platform, RNAscope).

My initial aim was to study the expression of *CEP120* and *ARL3* at RNA and protein level in human development. Ideally, I would have performed IHC in human developing tissues, with a focus on cerebellar development using *CEP120* and *ARL3* antibodies and comparing these protein expression patterns with the RNA expression patterns obtained from RNA scope technique. However, I was not able to successfully find and use any commercial *CEP120* or *ARL3* antibody suitable to visualize the protein expression of *CEP120* and *ARL3* in mouse or human tissues via immunofluorescence or immunohistochemistry (data not shown). Alternatively, I have focused on the expression of *CEP120* and *ARL3* at RNA level, characterising the expression patterns of *CEP120* and *ARL3* in embryonic and foetal human tissues.

Furthermore, PAX6 and OLIG2 target specific proliferative zones in human developing brain. These proliferative zones can be distinguished and visualised at tissue level via IHC or IF staining of these two transcription factors (PAX6 and OLIG2) (Alzu'bi et al., 2022, Mo and Zecevic, 2008, Alzu'bi and Clowry, 2019). The antibody staining of PAX6 and OLIG2 via IHC can be compared with the RNA expression patterns obtained of *CEP120* and *ARL3* obtained via RNA scope.

The characterisation of the expression patterns of *CEP120* and *ARL3* in human embryonic and foetal tissues can be compared with the staining of PAX6 and OLIG2

in embryonic and foetal human tissues. The staining of PAX6 and OLIG2 provides insights in human brain development and allows to visualise proliferation regions in human developing brain. This complements the characterisation of the RNA expression of *CEP120* and *ARL3* during cerebellar development (which is affected in JBTS).

CEP120 and *ARL3* have been associated with JBTS, but also with other ciliopathy phenotypes. *CEP120* is also associated with skeletal phenotypes, while *ARL3* is associated with retinal dystrophies (Alkanderi et al., 2018, Holtan et al., 2019, Sheikh et al., 2019, Strom et al., 2016, Roosing et al., 2016, Shaheen et al., 2015). It is of interest to investigate, if both genes are expressed during human development in the developing organs often involved in JBTS (brain, eye and kidney) and JATD (skeletal tissues). I am interested in investigating if there is any difference in the expression of *CEP120* and *ARL3* during human development, and if those differences, if any, correlate with the heterogeneity of clinical phenotypes (which represent a large proportion of the wide spectrum of ciliopathy phenotypes) associated with mutations in these two genes.

Investigating gene expression in human tissues from the HDBR resource at Newcastle University (Lindsay et al., 2016, Gerrelli et al., 2015) and with the support and collaboration with Dr Laura Powell from Sayer and Miles Lab at Newcastle University, creates an appropriate scenario to investigate tissue expressivity of two pleiotropic ciliopathy genes: *CEP120* and *ARL3*, which are associated with a broad and overlapping phenotypic spectrum of clinical phenotypes (Alkanderi et al., 2018, Ratnapriya et al., 2021, Roosing et al., 2016, Shaheen et al., 2015). Besides, it allows to study if there is any correlation between tissue expressivity and the disease phenotypes associated with these two ciliopathy genes.

The aim of this chapter is to expand the knowledge regarding the developmental spatial expression of ciliopathy genes, in particular, two genes: *CEP120* and *ARL3* associated with a broad overlapping and heterogeneous spectrum of ciliopathy disease phenotypes. The aim of this chapter is to contribute to define and understand the role of tissue expressivity of ciliopathy genes in the phenotypic heterogeneity and overlap of disease phenotypes associated with primary ciliopathies.

The developmental expression patterns of *CEP120* and *ARL3* in humans are unknown. The aim of this chapter is to characterise *CEP120* expression patterns in

human embryos and foetuses and compare them with the *ARL3* expression patterns. The aim of this chapter is to compare and contrast the developmental spatial expression of these distinct causes of JSRD, focussing on the specific organs involved in JSRD: kidney, eye and brain, and skeletal regions involved in JATD. The results from this chapter will contribute to understand the molecular disease mechanisms and the role of tissue expressivity in primary ciliopathies, with a focus on a model primary ciliopathy: JBTS and two pleiotropic ciliopathy genes: *CEP120* and *ARL3*.

3.2. Results

Using RNAscope technique in collaboration with HDBR, *CEP120* and *ARL3* developmental expression patterns in human embryonic and foetal stages were characterised.

3.2.1. Comparison of the clinical phenotypes of CEP120- and ARL3-affected patients and RNA expression data from public databases

I first compared the disease phenotypes associated with disease-causing mutations in *CEP120* or *ARL3* (Table 3.1).

Biallelic mutations in *CEP120* or *ARL3* can lead to JBTS. Moreover, biallelic, *CEP120* mutations can cause severe phenotypes such as MKS (Barroso-Gil et al., 2021a, Roosing et al., 2016), while monoallelic heterozygous mutations in *ARL3*, can cause retinitis pigmentosa (Strom et al., 2016).

Comparing the human protein sequences of *CEP120* and *ARL3* with their homologs in other organisms, it is observed that the protein sequence of *ARL3* is more evolutionary conserved than the *CEP120* protein sequence. Some percentage of homology (~63-66%) is found between the human *ARL3* protein sequence and its homologs in *D. melanogaster* *C. elegans* and *C. reinhardtii*. In contrast, these homologs for *CEP120* were not found (Table 3.2).

Comparing the protein sequences of the protein functional domains of *CEP120*, it is observed that the amino acid affected by the missense *CEP120* mutations leading to JBTS and JATD phenotypes, are highly evolutionary conserved amino acids (Figure 3.2).

RNA expression data of different ciliopathy genes were obtained and compared. These data were obtained from two public databases: Genotype-Tissue Expression

(GTEx) Project (The GTEx Consortium, 2020) and Human Protein Atlas (HPA) (Thul and Lindskog, 2018) (Table 3.4). 6 tissues, which are involved in the primary ciliopathies studied in this chapter (JBTS and JATD) were considered, and 20 of the ciliopathy genes were included, with a focus on the ciliopathy genes investigated in this project. It can be observed that *CEP120* and *ARL3* are expressed in the tissues affected in patients with biallelic mutations in these genes. *CEP120* is expressed in cerebellum and retina and *ARL3* is expressed in cerebellum, cerebral cortex, retina and kidney. These data correlate with what was expected, as both genes are associated with JBTS, and with retinal dystrophy in the case of *ARL3* (Roosing et al., 2016, Wu et al., 2014, Ratnapriya et al., 2021). Other ciliopathy genes leading to JBTS such as *MKS1* and *CC2D2A* are expressed in cerebellum and kidney and kidney and retina, respectively.

Genes associated with renal ciliopathies, such as *ALG8*, *ALG9*, *DNAJB11* and *PKD2* are expressed in the kidney, as expected. It is also interesting to observe that *PKD1* is highly expressed in cerebral cortex and skeletal muscle and *NPHP1* being expressed in skeletal muscle. It is important to consider that *PKD1* is associated with ADPKD, and with embryonic lethally if two *PKD1* alleles are mutated. Besides, biallelic mutations in *NPHP1* are commonly associated with NPHP, however biallelic mutations in *NPHP1* are also associated with other multisystemic primary ciliopathies such as JBTS (Devlin and Sayer, 2019, Barroso-Gil et al., 2021b).

	Patients with <i>CEP120</i> related ciliopathy	Patients with <i>ARL3</i> related ciliopathy
Number of affected patients reported and presenting phenotypes	4 patients with JSRD, 4 patients with JBTS, 1 foetus with MKS, 1 foetus with TCDOE (Roosing et al., 2016, Shaheen et al., 2015)	4 patients reported with JBTS secondary to biallelic changes (Alkanderi et al., 2018), 4 patients with retinitis pigmentosa secondary to monoallelic changes (Holtan et al., 2019, Strom et al., 2016) 1 patient with autosomal recessive rod-cone dystrophy (Fu et al., 2021), 8 patients (from 2 families) with autosomal recessive cone rod dystrophy (Sheikh et al., 2019), 1 patient with autosomal dominant retinal degeneration (Ratnapriya et al., 2021)
Brain imaging findings	Molar tooth sign	Molar tooth sign
Intellect	Developmental delay, Cognitive impairment	Developmental delay, Psychomotor delay
Skeletal	Severely narrow chest, Skeletal dysplasia, Small and horizontal ribs, Short limbs, Polydactyly, Synpolydactyly	No known associated phenotypes
Mobility	Truncal ataxia Hypotonia	Ataxic gait, Hypotonia
Eye	Microphthalmia, Duane syndrome, Strabismus,	Night blindness, Bilateral vision loss, Retinal dystrophy, Ocular motor apraxia
Kidney	Cystic dysplastic kidney	Cystic dysplastic kidney, Bilateral renal scarring, Recurrent urinary tract infections

Table 3.1. Comparison of the known phenotypes associated with *CEP120* and *ARL3* mutations. JATD, Jeune asphyxiating thoracic dystrophy; JBTS, Joubert syndrome; MKS, Meckel syndrome; TCDOE, tectocerebellar dysraphia with occipital encephalocele.

	ARL3		CEP120		
Organism	ID	Homology %	ID	Homology %	Database
<i>Homo sapiens</i>	NP_004302.1	100	NP_694955.2	100	NCBI
<i>Bos taurus</i>	NP_001033656.1	98.35	NP_001071468.1	88.65	NCBI
<i>Mus musculus</i>	NP_062692.1	98.35	NP_848801.2	89.17	NCBI
<i>Rattus norvegicus</i>	NP_073191.1	97.25	NP_001178626.1	90.39	NCBI
<i>Gallus</i>	XP_421730.1	96.7	XP_015136305.1	72.03	NCBI
<i>Xenopus tropicalis</i>	XP_002938771.2	97.8	NP_001120380.1	65.74	NCBI
<i>Erpetoichthys calabaricus</i>	XP_028651567.1	95.05	XP_028661076.1	61.36	NCBI
<i>Danio rerio</i>	NP_001038373.1 (1)	94.51	XP_017212877.1	57.67	NCBI
<i>Gadus morhua</i>	XP_030195485.1	94.51	XP_030226769.1	55.52	NCBI
<i>Oryzias latipes</i>	XP_011486618.1	93.96	XP_011477980.1	54.12	NCBI
<i>Takifugu rubripes</i>	XP_003978186.2	92.31	XP_011613677.2	52.08	NCBI
<i>Callorhinchus milii</i>	XP_007910451.1	97.8	XP_007897573.1	64.01	NCBI
<i>Drosophila melanogaster</i>	FBpp0083549	65.54	Not found	FlyBase	
<i>Caenorhabditis elegans</i>	F19H8.3	63.39	Not found	WormBase (WS72)	
<i>Chlamydomonas reinhardtii</i>	Cre04.g218250.t1.2	63.84	Not found	Phytozome	

Mammals	Vertebrate (non-euteleostomi)
Tetrapods (non-mammals)	Bilateria (non-vertebrate)
Euteleostomi (non-tetrapods)	Eukaryota (non-bilateria)

Table 3.2. Comparison of the homology of CEP120 and ARL3 amino acid sequences. Percent homology refers to the similarity between two sequences, in this case, protein sequences from different organisms. Percent homology was obtained from the percent identity, which quantifies the similarity between two aligned sequences. (1) There are two arl3 homolog sequences in *Dario rerio*. There are two arl3 transcripts: arl3a NP_001315350.1 (with 95.05% homology with human ARL3) and arl3b NP_001038373.1 (with 94.51% homology with human ARL3), as a result of whole genome duplication events in teleost fish.

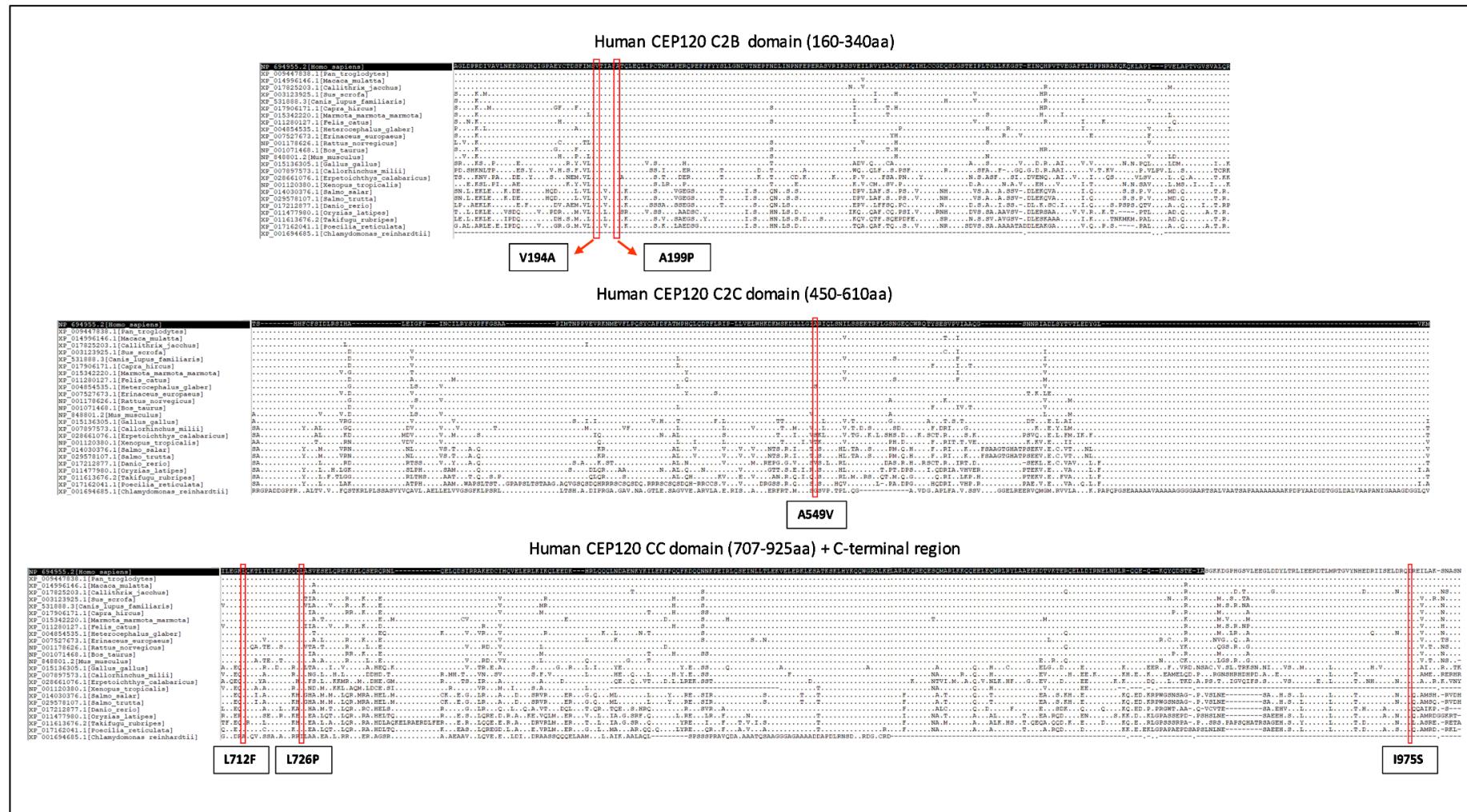


Figure 3.2. Amino acid sequences of the domains of the human *CEP120* protein aligned against different *CEP120* orthologs using BLASTP. The multiple alignments obtained was visualised using bioedit. If an amino acid is the same than the one in the human *CEP120* protein sequence, the position is indicated with a point. The positions coloured black correspond to the C2B domain, C2C domain and coil-coiled (CC) domain. The *CEP120* missense mutations described in the literature (Roosing et al., 2016, Shaheen et al., 2015) are indicated in red. Some of the *CEP120* sequences obtained from NCBI were described as predicted:

XP_017825203.1[*Callithrix_jacchus*], XP_017906171.1[*Capra_hircus*], XP_015342220.1[*Marmota_marmota_marmota*],
XP_007527673.1[*Erinaceus_europaeus*], XP_007897573.1[*Callorhinchus_milii*], XP_014030376.1[*Salmo_salar*],
XP_017162041.1[*Poecilia_reticulata*], XP_001694685.1[*Chlamydomonas_reinhardtii*] are described as predicted sequences. The
protein sequence of the Uni2 protein of the ciliated alga *Chlamydomonas reinhardtii* was included although if this protein is truly a
CEP120 homolog is not yet elucidated (Mahjoub et al., 2010, Piasecki and Silflow, 2009).

CEP120 and ARL3 proteins are encoded by the JBTS genes: *CEP120* and *ARL3*, respectively (Roosing et al., 2016, Alkanderi et al., 2018). CEP120 and ARL3 have different functions and locations. CEP120 functions at the centrosome (Mahjoub et al., 2010) while ARL3 function at the ciliary axoneme (Schwarz et al., 2017, Alkanderi et al., 2018) (Table 3.3).

	CEP120	ARL3
Ciliary / Basal body Localisation	Centrosome (asymmetrically localized to the daughter centriole) (Mahjoub et al., 2010)	Ciliary Axoneme (ARL3-GTP) (Schwarz et al., 2017)
Protein Domains	3 consecutive C2 domains (C2A, C2B and C2C) and a coil-coiled domain (Sharma et al., 2018)	GTPase (Kahn et al., 1992)
Function	Centriole duplication, assembly, elongation and maturation (Mahjoub et al., 2010, Wu et al., 2014, Comartin et al., 2013b)	Releases prenylated, farnesylated, geranylgeranylated and myristolyated cargo from their carriers (e.g. PDE6 δ and UNC119a/b) in the cilium (Gotthardt et al., 2015, Ismail et al., 2011)
Interactors	TACCs (Xie et al., 2007), CPAP (Lin et al., 2013), SPICE1 (Comartin et al., 2013b), Talpid3 (Wu et al., 2014), C2CD3 (Tsai et al., 2019), KIAA0753 (Chang et al., 2021)	ARL13B (GEF) (Gotthardt et al., 2015), RP2 (GAP) (Veltel et al., 2008)

Table 3.3. Comparison of the ciliary localisation, protein domains, function and protein interactors of the CEP120 and ARL3 proteins.

Tissue	Gene	GTEx Median TPM	HPA nTPM	Gene	GTEx Median TPM	HPA nTPM	Gene	GTEx Median TPM	HPA nTPM	Gene	GTEx Median TPM	HPA nTPM
Cerebellum	<i>ALG8</i>	11.03	17.1	<i>CEP120</i>	15.13	11.2	<i>KIAA0586</i>	6.74	7.2	<i>PKD1</i>	577.2	252.1
Cerebral cortex		6.47	19		3.53	4.2		2.36	4.2		84.66	53.8
Kidney		11.14	24.3		5.84	6.7		2.02	3.9		12.06	7.3
Liver		10.02	45.7		2.99	7.7		1.71	6.3		8.09	8.7
Retina		NA	10.2		NA	11.2		NA	7.7		NA	6.6
Skeletal muscle		5.53	11.7		2.98	4.2		2.14	5.6		40.01	46
Cerebellum	<i>ALG9</i>	9.41	9.6	<i>CEP164</i>	33.93	14.2	<i>LAMA1</i>	0.39	0	<i>PKD2</i>	26.4	12.4
Cerebral cortex		4.77	8.6		8.74	5.6		1.46	1		10.75	9.1
Kidney		5.43	9.2		6.54	5		6.8	2.3		27.17	46.8
Liver		3.9	13		5.99	7.1		0.09	0		5.06	14.5
Retina		NA	11.8		NA	25.2		NA	0.3		NA	13.5
Skeletal muscle		3.11	8.7		2.15	3		0.04	0		4.18	5.6
Cerebellum	<i>ARL13B</i>	3.79	1.2	<i>CEP290</i>	9.89	7.6	<i>MKS1</i>	12.11	7.5	<i>SUFU</i>	11.89	8
Cerebral cortex		2.35	1.6		5.58	7.1		5.16	5.5		5.84	6.2
Kidney		4.03	3.4		4.55	6.9		9.97	10.2		5.46	6
Liver		1.75	2.5		1.56	3.5		4.92	8.5		3.02	6.1
Retina		NA	17.6		NA	6.9		NA	6.4		NA	4.5
Skeletal muscle		0.79	0.7		1.35	2.8		4.16	6.5		4.64	8.3
Cerebellum	<i>ARL3</i>	51.74	25.6	<i>DNAJB11</i>	28.13	17.4	<i>NPHP1</i>	0.95	1	<i>TMEV67</i>	3.94	1.8
Cerebral cortex		37.93	30.7		17.37	20.6		2.41	4.3		2.75	2.1
Kidney		18.92	14.9		41.64	32		3.61	7.8		4.11	2.8
Liver		3.07	5.1		45.22	79.7		0.26	0.8		0.48	0.6
Retina		NA	93.4		NA	15.7		NA	10.4		NA	3.1
Skeletal muscle		4.36	5.4		15.17	22.1		4.77	35		0.39	0.6

Cerebellum		3.93	4.4		18.7	12		0.95	4.6		14.46	10
Cerebral cortex		3.52	9.9		14.97	16.5		2.41	4.6		8.75	10.4
Kidney	CC2D2A	5.55	22.7	IFT80	18.29	17.4	OFD1	4.64	8.7	TOGARAM1	4.88	7.6
Liver		0.6	2.7		4.05	7.4		0.26	6		2.33	4.5
Retina		NA	82.1		NA	24.2		NA	2.4		NA	16.3
Skeletal muscle		0.64	2.6		2.18	4.7		4.77	4.7		5.1	12

	Mean GTEx	Min GTEx	Max GTEx	Mean HPO	Min HPO	Max HPO
Cerebellum	42.04	0.39	577.2	21.25	0	252.1
Cerebral cortex	11.59	1.46	84.66	11.25	1	53.8
Kidney	10.13	2.02	41.64	12.30	2.3	46.8
Liver	5.27	0.09	45.22	11.52	0	79.7
Retina	NA	NA	NA	18.48	0.3	93.4
Skeletal muscle	5.42	0.04	40.01	9.51	0	46
Mean of selected tissues	14.89			14.05		

GTEx Median TPM / HPA nTPM		
<10	10-20	>20

Table 3.4. RNA expression values of various ciliopathy genes, obtained from two public databases: Genotype-Tissue Expression (GTEx) Project and Human Protein Atlas (HPA). RNA expression values from GTEx are shown as Median Transcript per Million (TPM) (isoforms are collapsed into a single gene with no further normalisation steps). RNA expression values from HPA are shown as consensus normalized expression TPM (nTPM). GTEx expression data were obtained from GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). The nTPM transcript expression data were obtained from HPA (The HPA version 23.0 and Ensembl version 109), which considers 50 tissues based on transcriptomics data from HPA and GTEx public databases. Only 6 tissues were considered, as they are involved in primary ciliopathies. The bottom table shows the different mean values and the maximum and minimum values of the TPM values. The mean of the TPM values from the 6 tissues selected for each of the two databases (mean of selected tissues). Colour coding is shown below to facilitate visualisation, interpretation and comparison of the expression values shown in the main (top) table.

3.2.2. Expression of *CEP120* and *ARL3* in early human brain development

CEP120 and *ARL3* expression patterns were analysed in a whole human embryo in developmental stage CS23 (8 PCW). It was observed that *CEP120* (Figure 3.3) and *ARL3* (Figure 3.4) have similar expression patterns. At 8 PCW, both genes were expressed in the choroid plexus (Figures 3.3i and 3.4i). In some cells of the choroid plexus epithelium, it seems that the *CEP120* and *ARL3* expression is facing the lumen, especially for *ARL3* (arrows in Figure 3.4i). In contrast, *Ki67*, a cell proliferation marker, does not show this expression pattern at the choroid plexus (Figure 3.5i).

CEP120 and *ARL3* are also widespread expressed in the ganglionic eminences (Figures 3.3ii and 3.4ii) cortical neuroepithelium (Figures 3.3iii and 3.4iii) and developing hindbrain (Figures 3.3iv and 3.4iv) including the rhombic lip (Figures 3.3iv and 3.4iv). In these developing tissues, the expression patterns of both *CEP120* and *ARL3* seem to be facing into the ventricular space, towards the apical surface of these tissues (Figures 3.3ii-iv and 3.4ii-iv). The expression of *Ki67* also at the apical layer of these tissues (Figures 3.5ii-iv) is consistent with these tissues being sites of abundant cell proliferation during brain development (Taverna and Huttner, 2010).

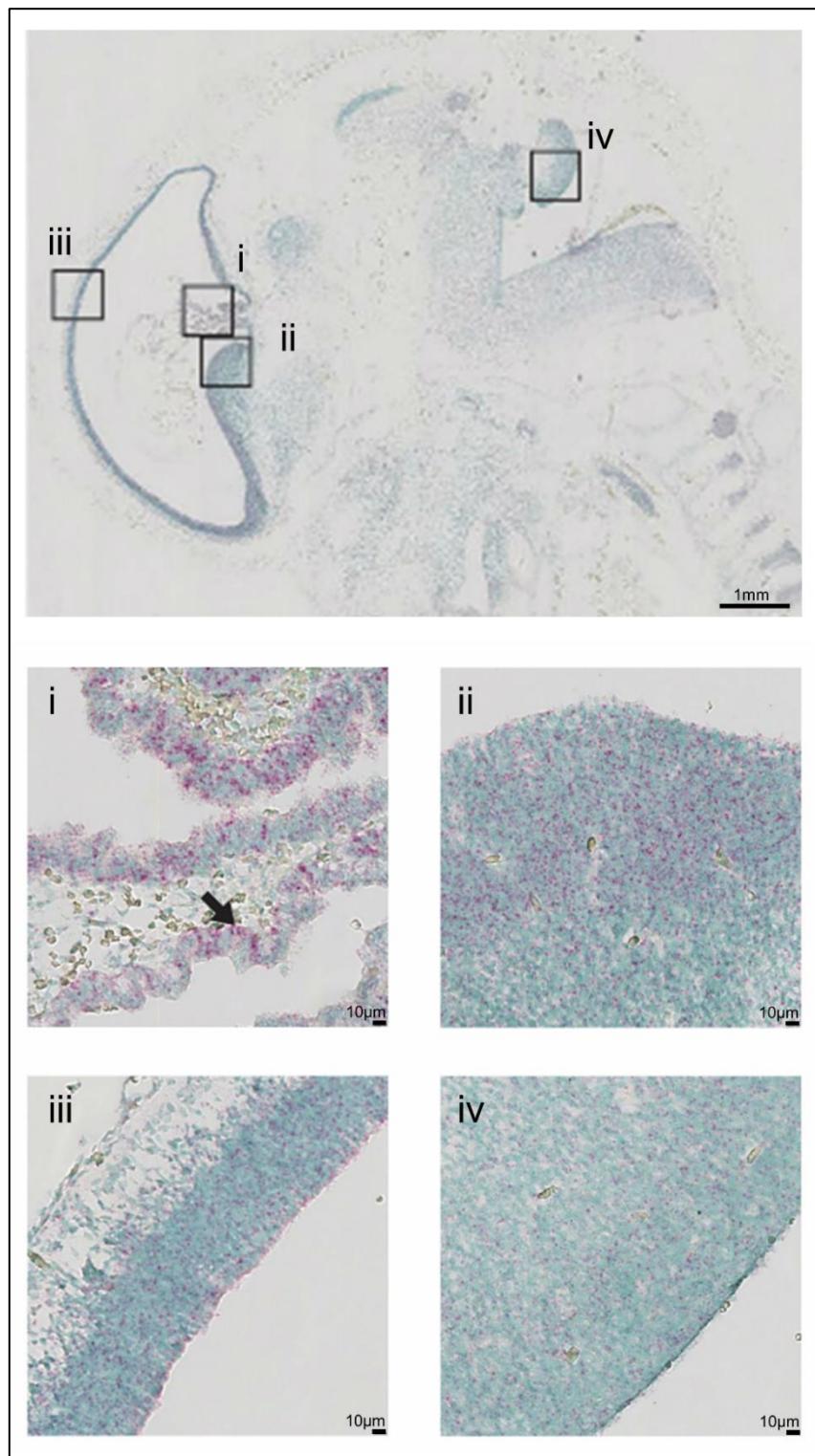


Figure 3.3. Expression of *CEP120* (red) in a human fetal brain of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green. i) *CEP120* is expressed within cells of the choroid plexus (arrow). ii-iv) Expression of *CEP120* is observed in the ventricular radial glia progenitor cells including the ventricular zone of the ganglionic eminences (ii), cerebral cortex (iii), and rhombic lip (iv).

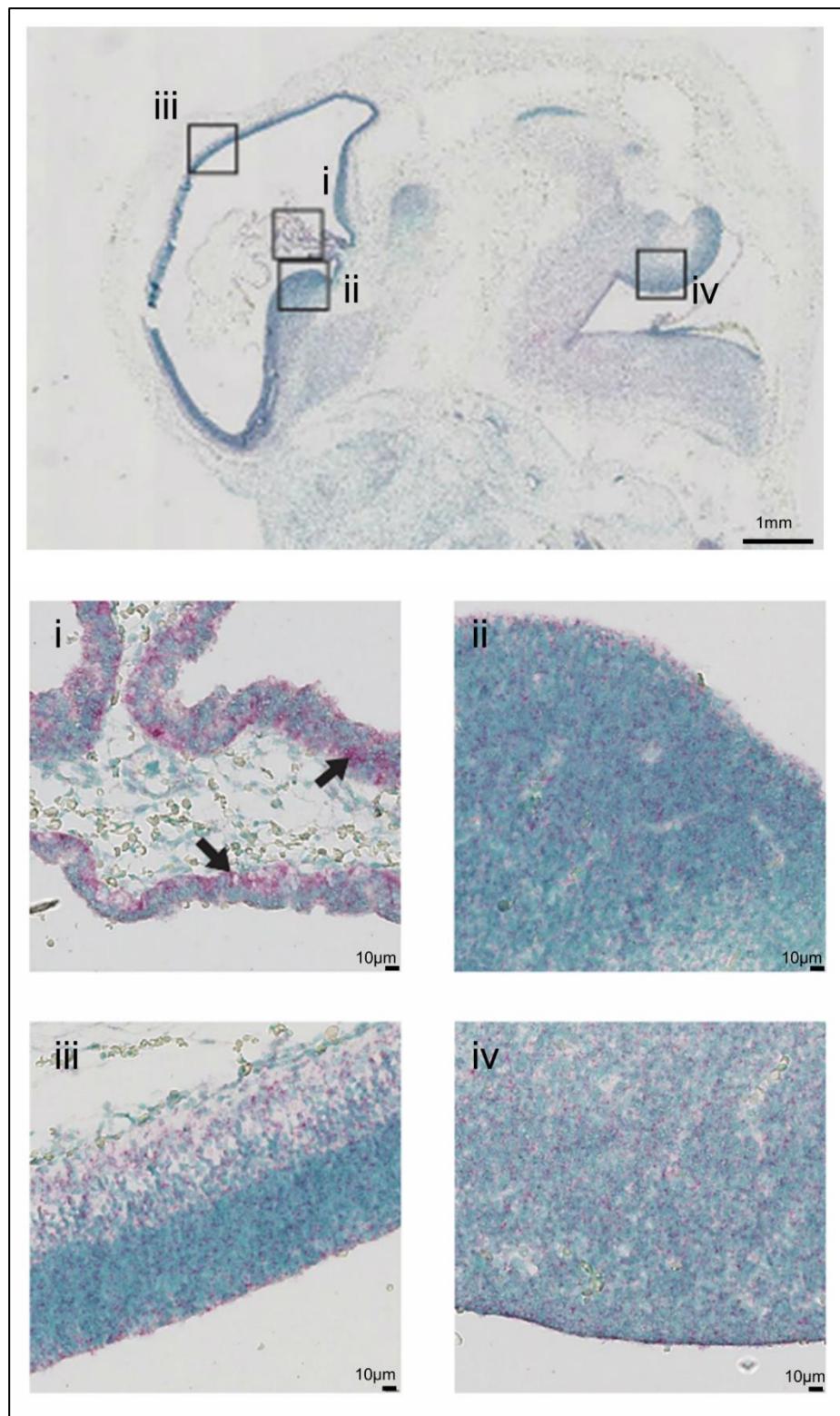


Figure 3.4. Expression of *ARL3* (red) in a human fetal brain of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green. i) *ARL3* is expressed within cells of the choroid plexus (arrows). ii-iv) Expression of *ARL3* is observed in the ventricular radial glia progenitor cells including the ventricular zone of the ganglionic eminences (ii), cerebral cortex (iii), and rhombic lip (iv).

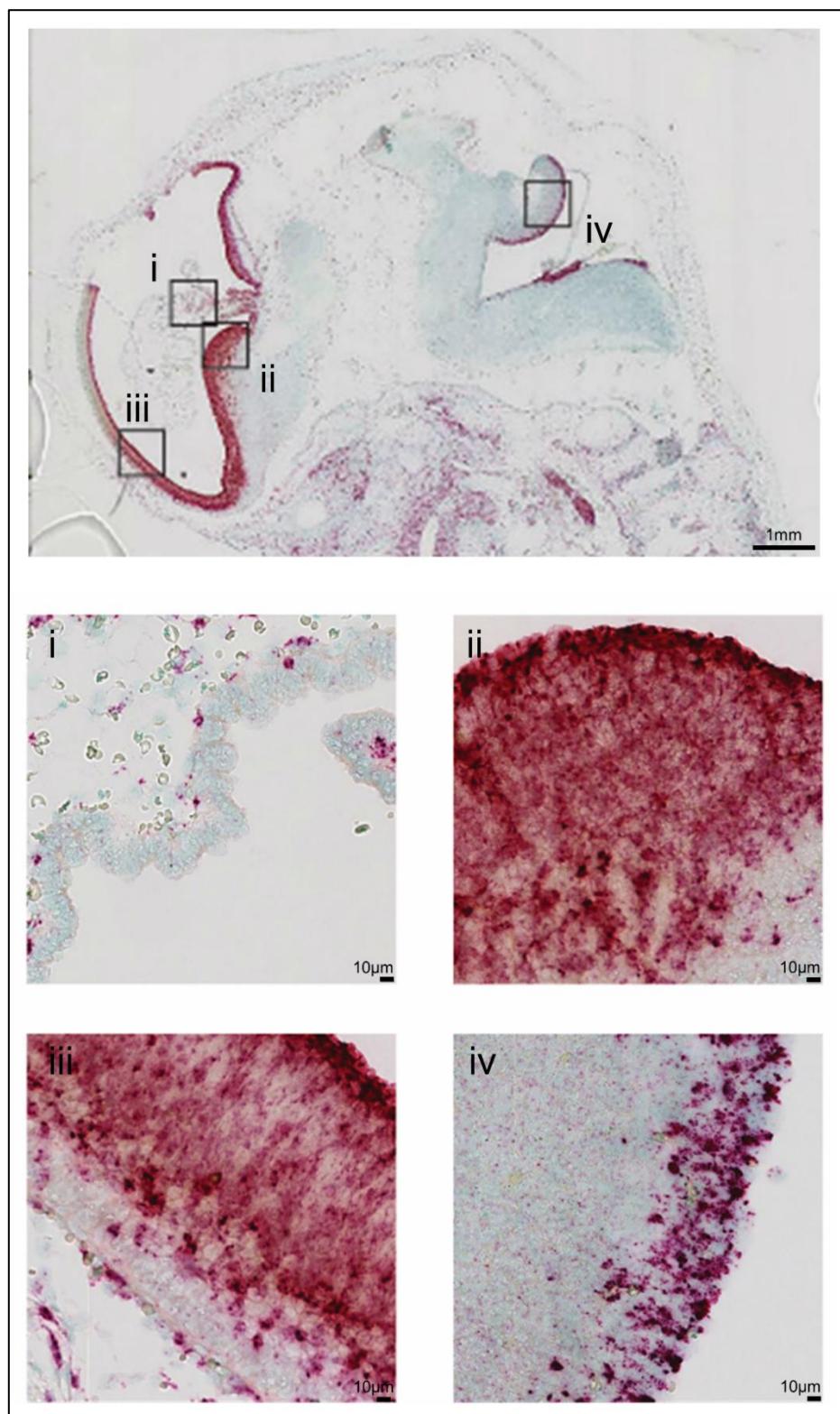


Figure 3.5. Expression of *Ki67* (red) in a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green. Expression of *Ki67* was used, a cell proliferation marker, as a positive control. I) *Ki67* expression is minimal in the choroid plexus. II-IV) Expression of *Ki67* is seen in the ventricular zone of the ganglionic eminences (II), cerebral cortex (III), and hindbrain (IV).

3.2.3. Expression *CEP120* and *ARL3* in the developing cerebellum

The cerebellum is a region of the hindbrain involved in multiple functions such as the regulation of posture, motor movement coordination, balance, equilibrium, motor learning, muscle tone, emotion and cognition (Rahimi-Balaei et al., 2018, Wu et al., 2014). I focussed on the hindbrain because it is a region affected in JBTS and other ciliopathies.

CEP120 and *ARL3* expression was observed during cerebellar development.

Expression of *CEP120* and *ARL3* was analysed in samples of human fetal hindbrain at 14 PCW (Figures 3.6A and 3.7A) and at 19 PCW (Figures 3.6B and 3.7B). Expression of *CEP120* (Figure 3.6A) and *ARL3* (Figure 3.7A) was observed in the hindbrain at 14 PCW, and this expression is maintained at 19 PCW (Figures 3.6B and 3.7B). Similar expression patterns were observed between *CEP120* (Figure 3.6A) and *ARL3* (Figure 3.7A) in the human developing cerebellum.

At 14 PCW both genes are highly expressed in the external and internal granule cell layers (EGL and IGL) of the developing cerebellum (Figures 3.6Ai and 3.7Ai).

The expression of both genes is maintained at 19 PCW in human developing cerebellum. *CEP120* expression was observed at the EGL mainly, and the molecular layer (ML) of the cerebellum (Figure 3.6Bi). In contrast, the *ARL3* expression seems to be localised at the EGL and IGL cerebellar layers of the developing cerebellum (Figure 3.7Bi).

At 18 PCW, *KI67*, a cell proliferation marker, is highly expressed at the EGL and the IGL (Figure 3.8i).

During cerebellar development, the ML is a cerebellar cell layer occupied by the dendrites of the Purkinje cells and interacting parallel fibres of granule cells. (Marzban et al., 2014). Interestingly, I observed specific expression of *CEP120* in the ML (Figures 3.6 and 3.9). This suggest that *CEP120* may have a role in these specific cell types.

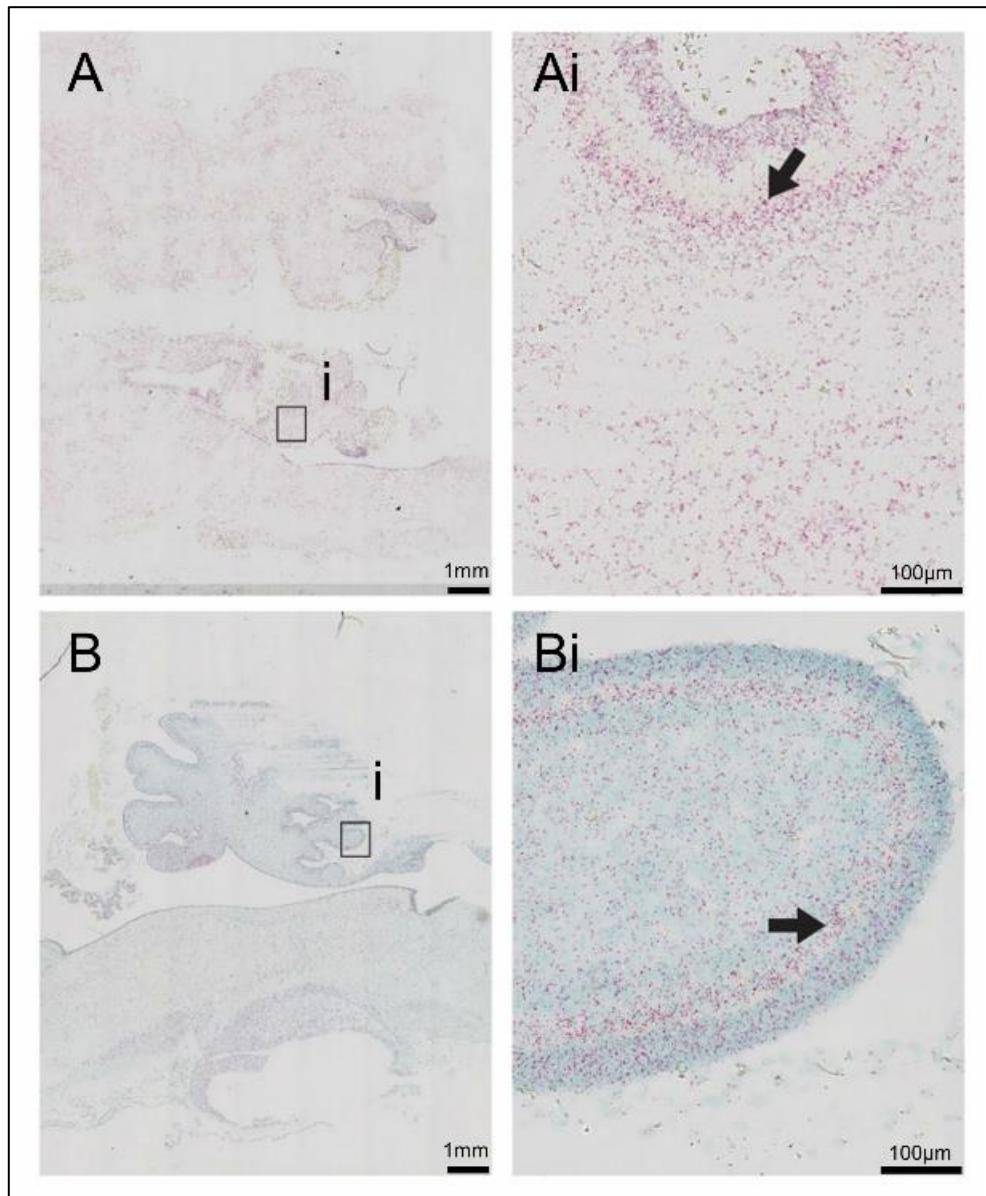


Figure 3.6. Expression of *CEP120* (red) in sagittal sections of human hindbrain at 14 PCW (A) and 19 PCW (B) stained using RNAscope, counterstained with Methyl Green. *CEP120* is expressed in the cerebellum at 14 PCW (Ai) and at 19 PCW (Bi). At 14 PCW *CEP120* expression is observed at the external granular layer (EGL) and internal granular layer (IGL) (arrow) (Ai). At 19 PCW *CEP120* expression is observed at the EGL, IGL and the molecular cell layer (ML) (arrow) (Bi).

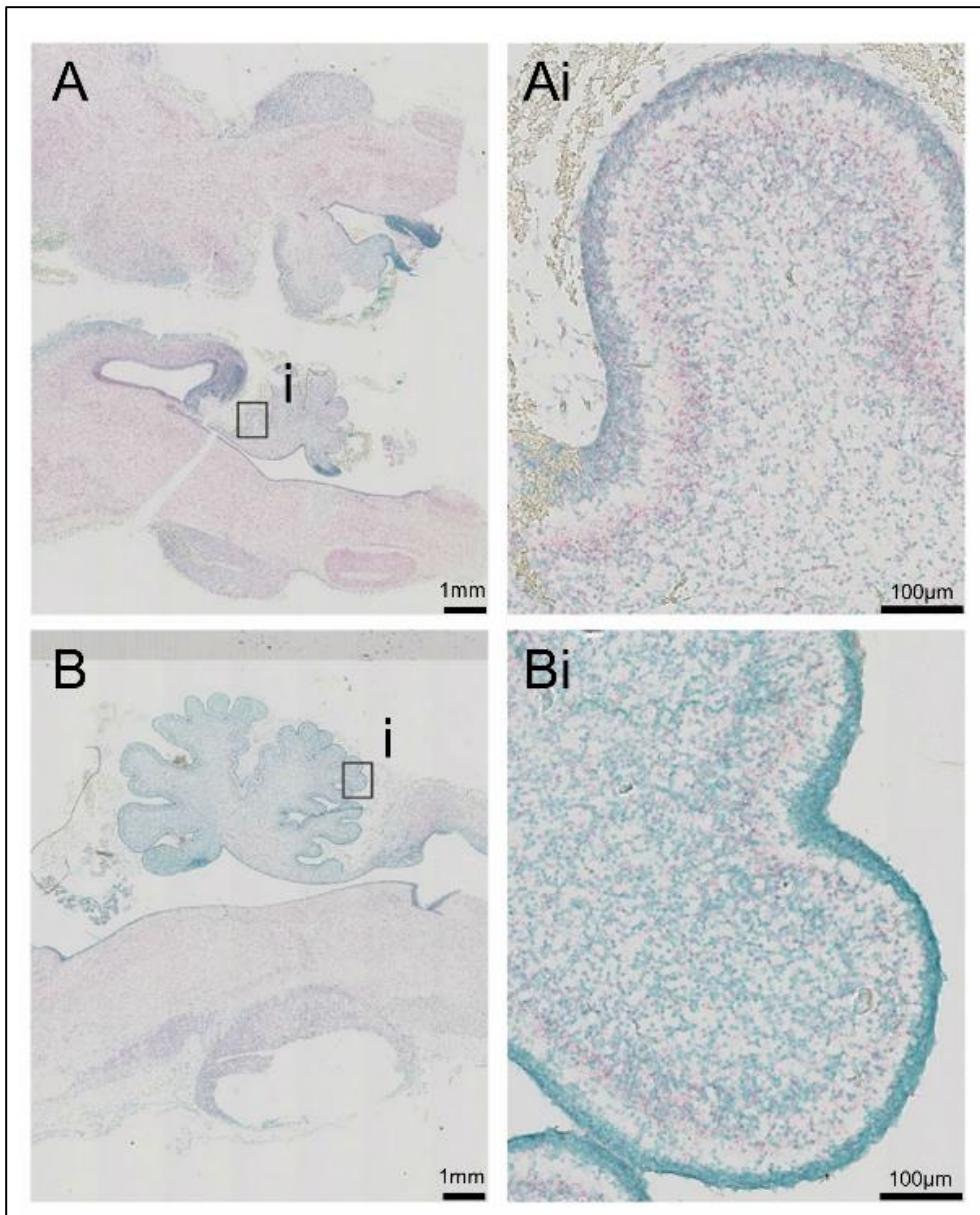


Figure 3.7. Expression of *ARL3* (red) in sagittal sections of human hindbrain at 14 PCW (A) and 19 PCW (B) stained using RNAscope, counterstained with Methyl Green. *ARL3* is expressed in the cerebellum at 14 PCW (Ai) and at 19 PCW (Bi). At 14 PCW (Ai) and 19 PCW (Bi) *ARL3* expression is observed at the external granular layer (EGL) and internal granular layer (IGL).

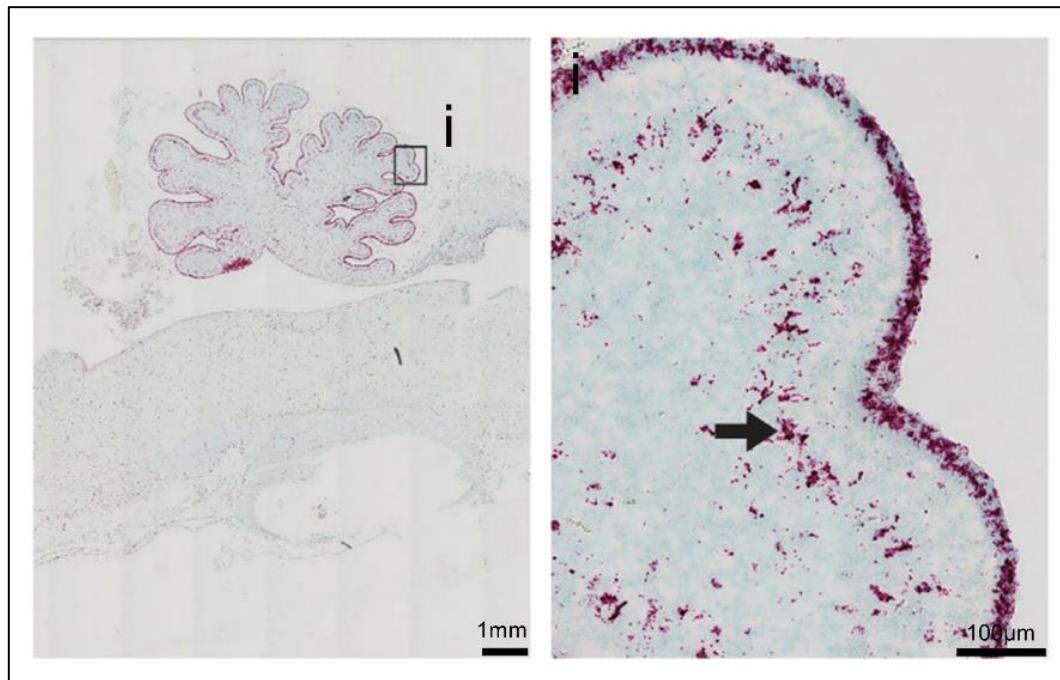


Figure 3.8. Expression of *Ki67* (red) in a sagittal section of human hindbrain at 18 PCW stained using RNAscope, counterstained with Methyl Green. expression of *Ki67*, a cell proliferation marker, as a positive control. *Ki67* Expression is observed in hindbrain at 18 PCW, including the external granular layer (EGL) and internal granular layer (IGL) (arrow), indicating that these tissues in the developing hindbrain are proliferative.

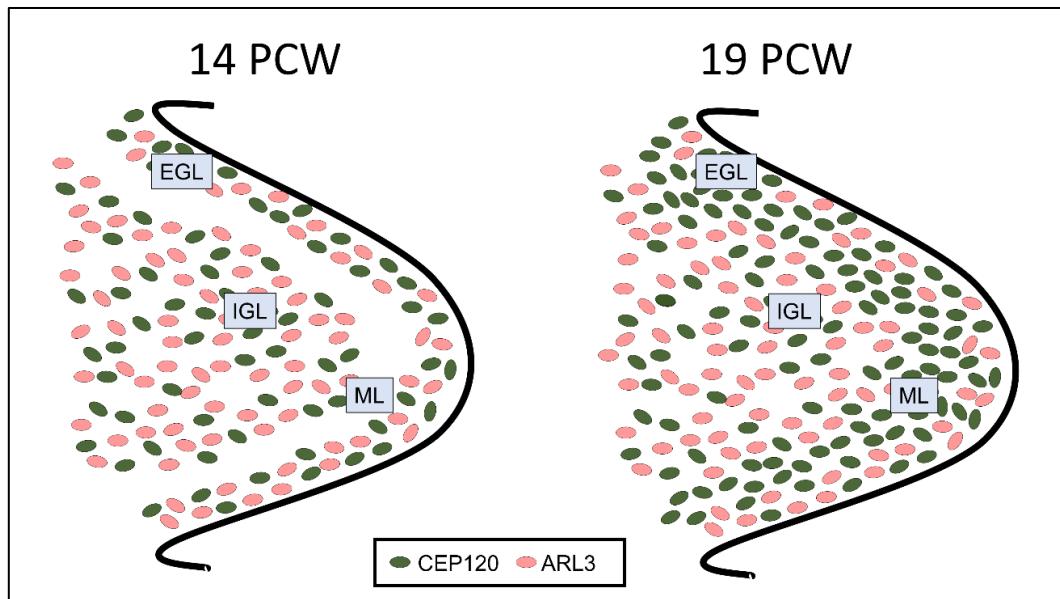


Figure 3.9. Schematic diagrams of the *CEP120* and *ARL3* expression in developing cerebellum at (f) 14 PCW and (g) 19 PCW. Expression of *CEP120* is shown in green and expression of *ARL3* is shown in pink. EGL, external granule layer; IGL, internal granule layer; ML, molecular layer.

3.2.4. Expression of *CEP120* and *ARL3* in the developing eye

Eye defects such as retinal dystrophy are often found in the phenotypic spectrum of primary ciliopathies. Patients presenting JBTS or JATD may also present retinal degeneration (Bachmann-Gagescu et al., 2015a, Brancati et al., 2010).

None of the *CEP120* patients reported in the literature presented eye defects, although it is important to consider that some of the patients described to have biallelic *CEP120* mutations, deceased when they were only few days old, therefore, eye or renal phenotypes may not have been developed and/or detected (Roosing et al., 2016, Shaheen et al., 2015).

Monoallelic mutations in *ARL3* have been associated with retinitis pigmentosa (Holtan et al., 2019). Patients with JBTS caused by biallelic mutations in *ARL3* are described to have rod-cone dystrophy among other extra-cerebellar phenotypes (Alkanderi et al., 2018).

The development of the human retina has unique characteristics and complex structure (Volland et al., 2015). The human retina is formed by multiple cell layers. There are 9 layers in a mature retina based on the cell types that form each layer of the retina (Hendrickson, 2016).

More precisely, these cells layers include the retinal pigment epithelium (RPE) and the photoreceptor cell layer. There are multiple cell types in the cell layers of the mature retina such as photoreceptors, intermediary cell bodies, dendrites and ganglion cells (Hoon et al., 2014). Of note, rod and cone photoreceptors are responsible to respond to light and occupy the photoreceptor cell layer. The outer segment of photoreceptors is a specialized sensory cilium (Chalupa and Gunhan, 2004) (Figure 3.10).

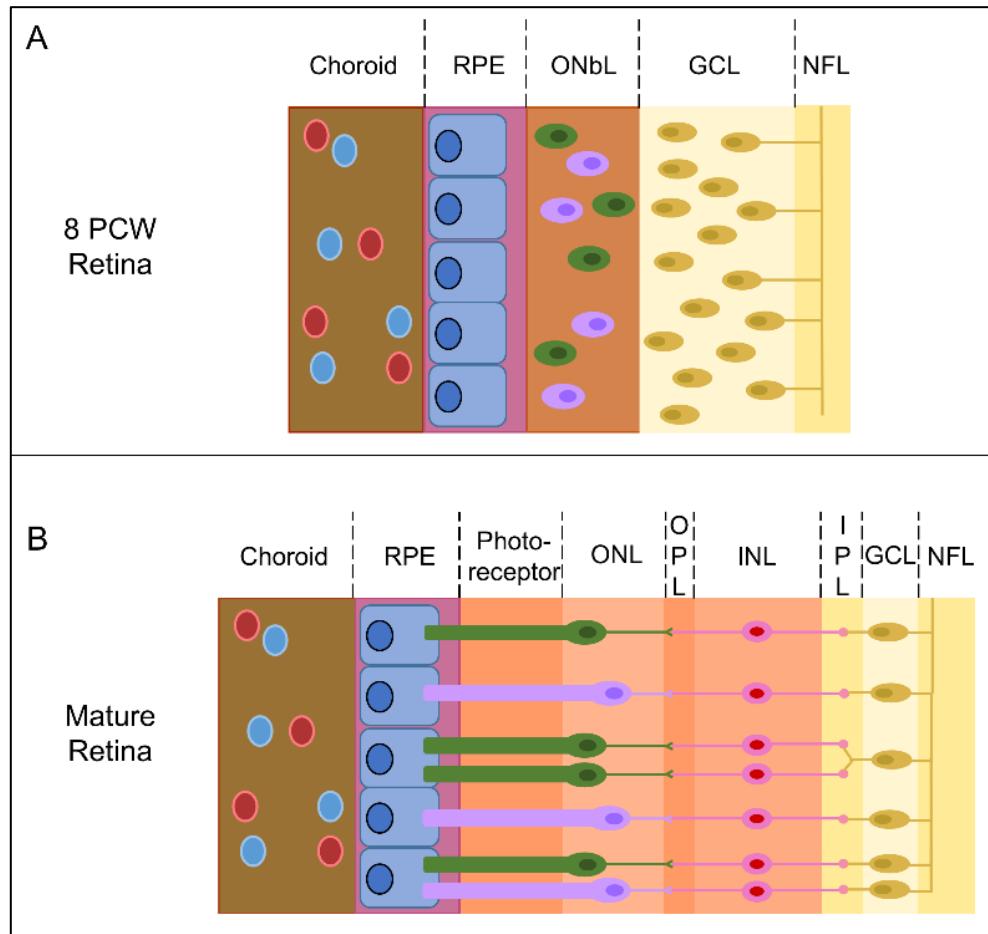


Figure 3.10. Diagram showing the development of the retina layers. In these two diagrams it is shown the retinal layers during eye development, at 8 PCW, compared with a mature retina (adapted from (Hendrickson, 2016). At 8PCW, only few cell layers are present in the retina. Modified from diagram done by Dr Laura Powell (Powell et al., 2020). GCL, ganglion cell layer; INL, inner nuclear layer; IPL, inner plexiform layer; NFL, nerve fibre layer; ONL, outer nuclear layer; OPL, outer plexiform layer; ONbL, outer neuroblastic layer; RPE, retinal pigment epithelium.

During eye development not all the retinal layers that will be present in a mature retina are formed (Hendrickson, 2016). At 8 PCW, during eye development, the retinal layers are not clearly defined. The retinal layers at 8 PCW are the RPE, the nerve fibre layer (NFL), the ganglion cell layer (GCL), and an outer neuroblastic layer (ONbL). The RPE layer locates at the back of the eye and removes waste products from the photoreceptor cells. The ONbL layer is a transitional layer that during eye development will develop into different cell layers such the outer plexiform layer (OPL), inner plexiform layer (IPL) (of note, the IPL can sometimes be observed at 8 PCW), outer nuclear layer (ONL) and inner nuclear layer (INL) (Chalupa and Gunhan, 2004).

It is observed that *CEP120* (Figure 3.11A) and *ARL3* (Figure 3.12A) are expressed in the developing retina of an 8 PCW-stage human embryo. At this stage in particular, in the developing retina at 8 PCW, both genes are expressed in the immature retinal ganglion cells and the photoreceptor layers (Figures 3.11A and 3.12A).

Later in human retinal development, at 14 PCW, the expression of *CEP120* (Figure 3.11B) and *ARL3* (Figure 3.12B) seem to be widespread expressed in the retina cell layers, although expression in some cell layers (plexiform and nerve fibre layers) is not easily observed because of their low cell density (Figures 3.11Bi and 3.12Bi).

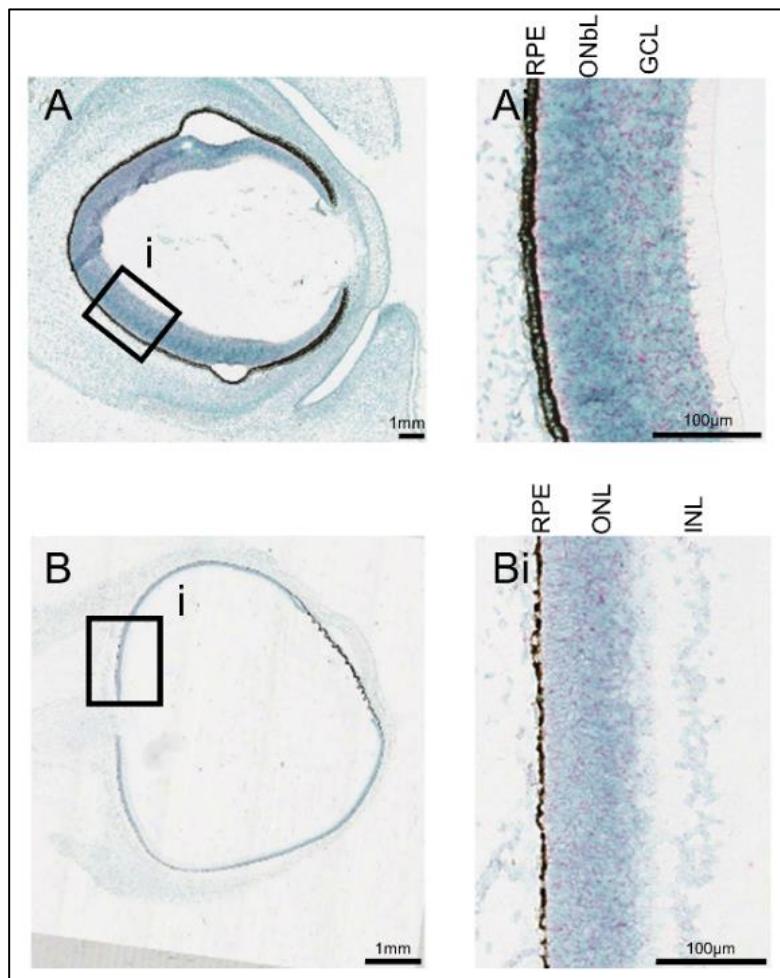


Figure 3.11. Expression of *CEP120* (red) in sagittal sections of human eye at 8 PCW (A) and 14 PCW (B) stained using RNAscope, counterstained with Methyl Green. Ai: At 8 PCW, *CEP120* is expressed across the GCL and the ONbL retinal layers. Bi: at 14 PCW, *CEP120* (Bi) expression is maintained as *CEP120* is expressed in the photoreceptor cell layer including the ONL. GCL, ganglion cell layer; INL, inner nuclear layer; ONL, outer nuclear layer; ONbL, outer neuroblastic layer; RPE, retinal pigment epithelium.

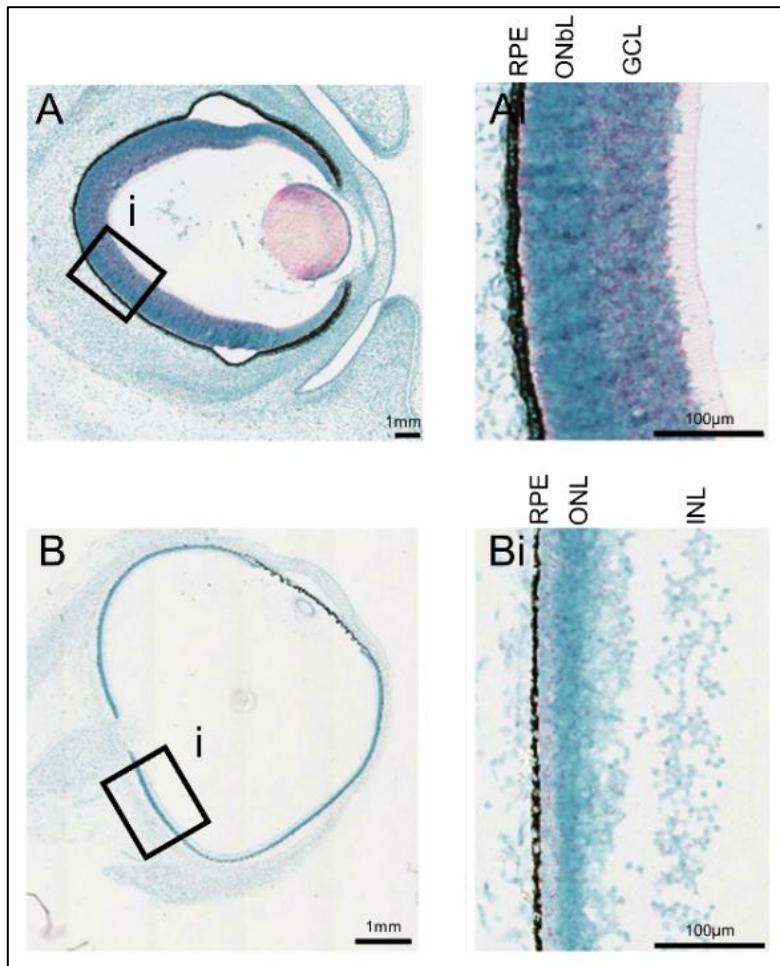


Figure 3.12. Expression of *ARL3* (red) in sagittal sections of human eye at 8 PCW (A) and 14 PCW (B) stained using RNAscope, counterstained with Methyl Green. Ai: At 8 PCW, *ARL3* is expressed across the GCL and the ONbL retinal layers. Bi: at 14 PCW, *ARL3* (Bi) expression is maintained as *ARL3* is expressed in the photoreceptor cell layer including the ONLGCL, ganglion cell layer; INL, inner nuclear layer; ONL, outer nuclear layer; ONbL, outer neuroblastic layer; RPE, retinal pigment epithelium.

3.2.5. Expression of *CEP120* and *ARL3* in the developing dorsal root ganglia

CEP120 has been described to be the disease-causing gene in 3 patients diagnosed with JATD, a skeletal ciliopathy (Shaheen et al., 2015). Biallelic disease-causing *CEP120* mutations have also been reported in 2 foetuses (one diagnosed with MKS and one with TCDOE) presenting narrow thorax with short ribs among other phenotypes (Roosing et al., 2016). This is why I aimed to determine if there is *CEP120* or *ARL3* expression in the skeletal regions of an 8 PCW-stage human embryo.

The dorsal root ganglia (DRG) consist in migrating neural crest cells and it is where most of the body's sensory neurones are located (Krames, 2014, Kalcheim et al., 1987).

CEP120 (Figure 3.13i) and *ARL3* (Figure 3.14i) were observed to be highly expressed in the DRG at 8 PCW. *CEP120* (Figure 3.13ii) and *ARL3* (Figure 3.14ii) expression was not as evident in the surrounding developing musculoskeletal tissue.

KI67 expression in the DRG at 8 PCW is limited to a few number of cells (Figure 3.15i-ii).

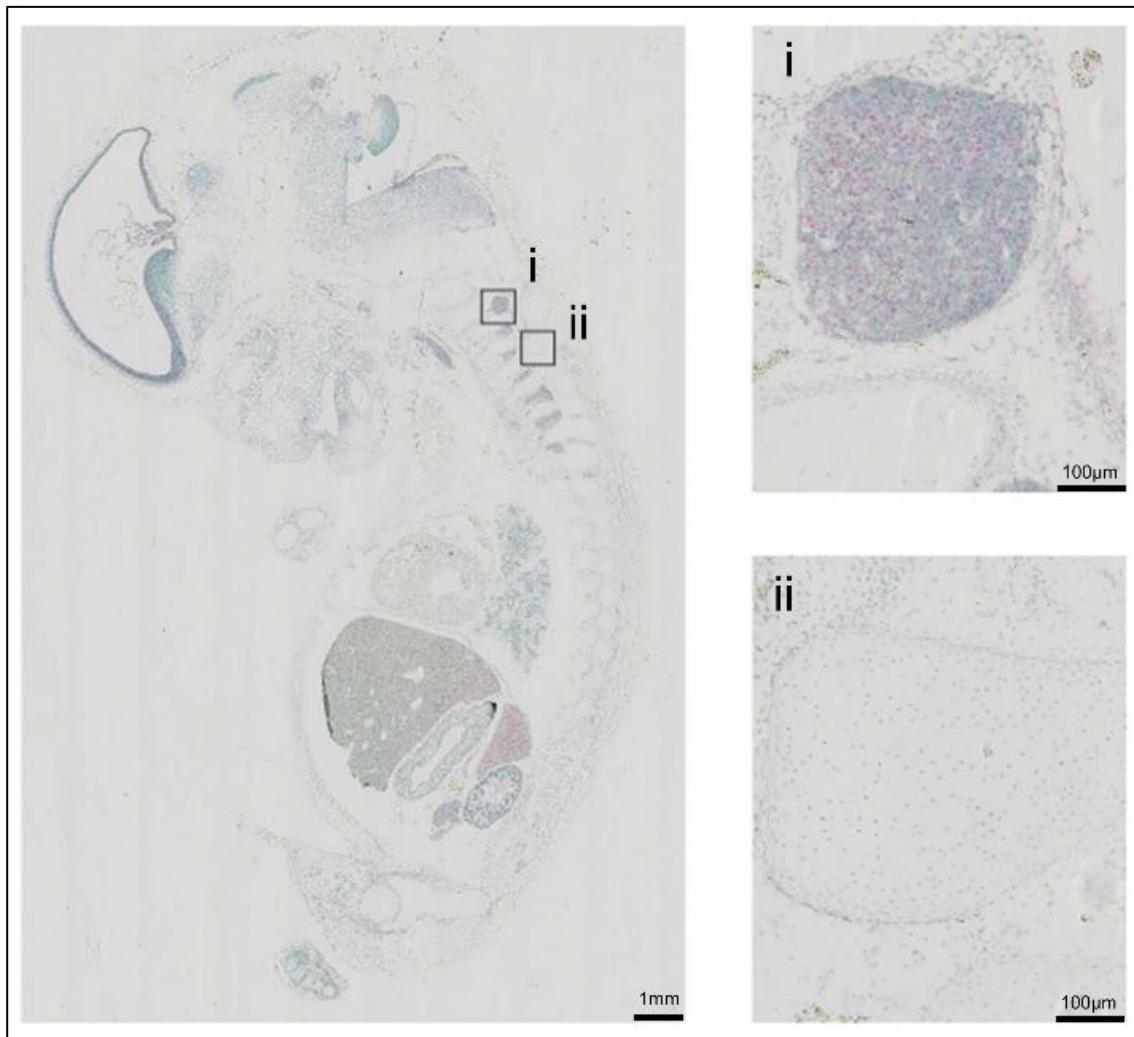


Figure 3.13. Expression of *CEP120* (red) in the developing human dorsal root ganglia of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green. i: *CEP120* expression within the dorsal root ganglia. ii: Low levels of *CEP120* expression within surrounding tissue.

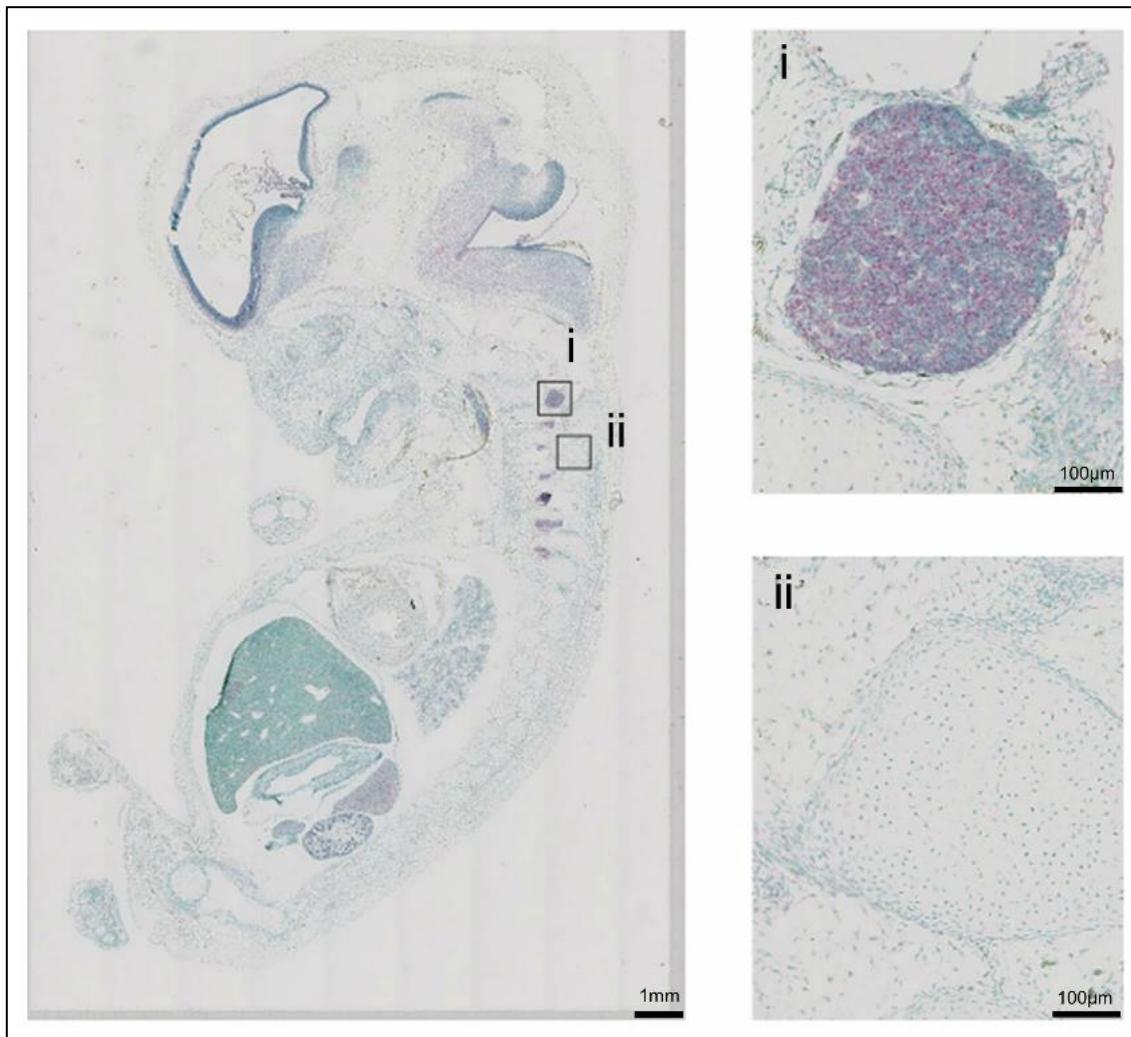


Figure 3.14. Expression of *ARL3* (red) in the developing human dorsal root ganglia of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green. i: *ARL3* expression within the dorsal root ganglia. ii: Low levels of *ARL3* expression within surrounding tissue.

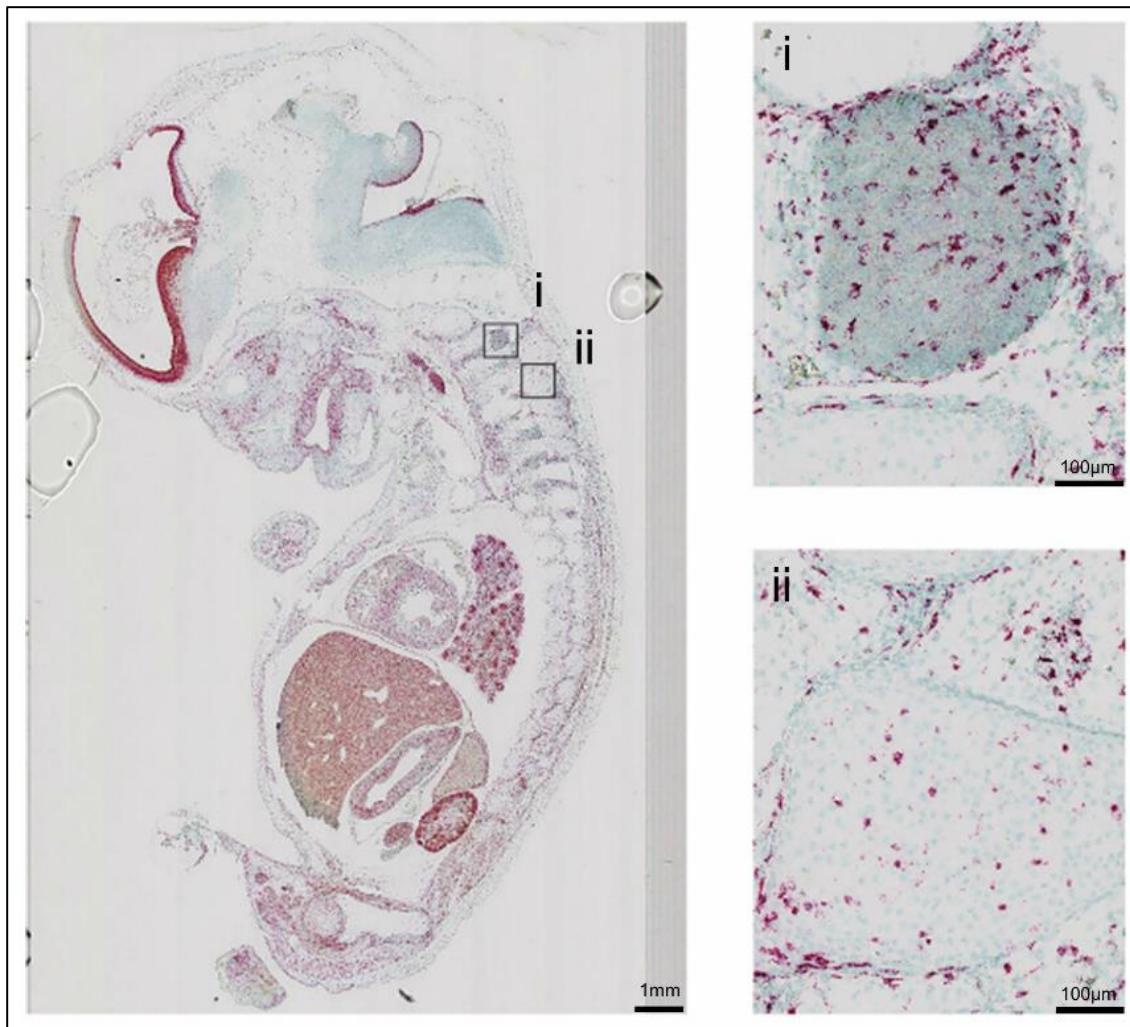


Figure 3.15. Expression of *Ki67* (red) in the developing human dorsal root ganglia of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green. *Ki67*, a cell proliferation marker, was used as a positive control. i: *Ki67* expression within the dorsal root ganglia. ii: *Ki67* expression within surrounding tissue.

3.2.6. Expression of *CEP120* and *ARL3* in the developing kidney

JBTS patients may develop other extra-neurological phenotypes, one of these phenotypes is renal disease, which is often found in the spectrum of ciliopathies (Bachmann-Gagescu et al., 2015a, Brancati et al., 2010).

None of the JBTS patients described in the literature presenting with biallelic *CEP120* mutations were described to have any kidney defects, except for a MKS foetus with cystic dysplastic kidneys (Roosing et al., 2016). In contrast, some of the reported JBTS patients with biallelic *ARL3* mutations presented renal phenotypes (Alkanderi et al., 2018).

It is observed that *CEP120* (Figure 3.16A) and *ARL3* (Figure 3.17A) are expressed in the developing kidney of an embryo in developmental stage CS23 (8 PCW). *Ki67* is highly expressed in the renal cortical regions, indicating cell proliferation in these regions (Figure 3.18).

At 8 PCW, expression of *CEP120* (Figure 3.16Ai) and *ARL3* (Figure 3.17Ai) is observed in developing nephrons are renal cortex, in particular *ARL3* seems to be facing towards the lumen of the developing cortical nephrons (Figure 3.17Ai).

At 14 PCW and 18 PCW, *CEP120* (Figures 3.16B and 3.16C) and *ARL3* (Figures 3.17B and 3.17C) expression in the developing nephrons and renal cortex is maintained, however this expression seems to have decreased at 18 PCW, especially for *CEP120* (Figure 3.16Ci).

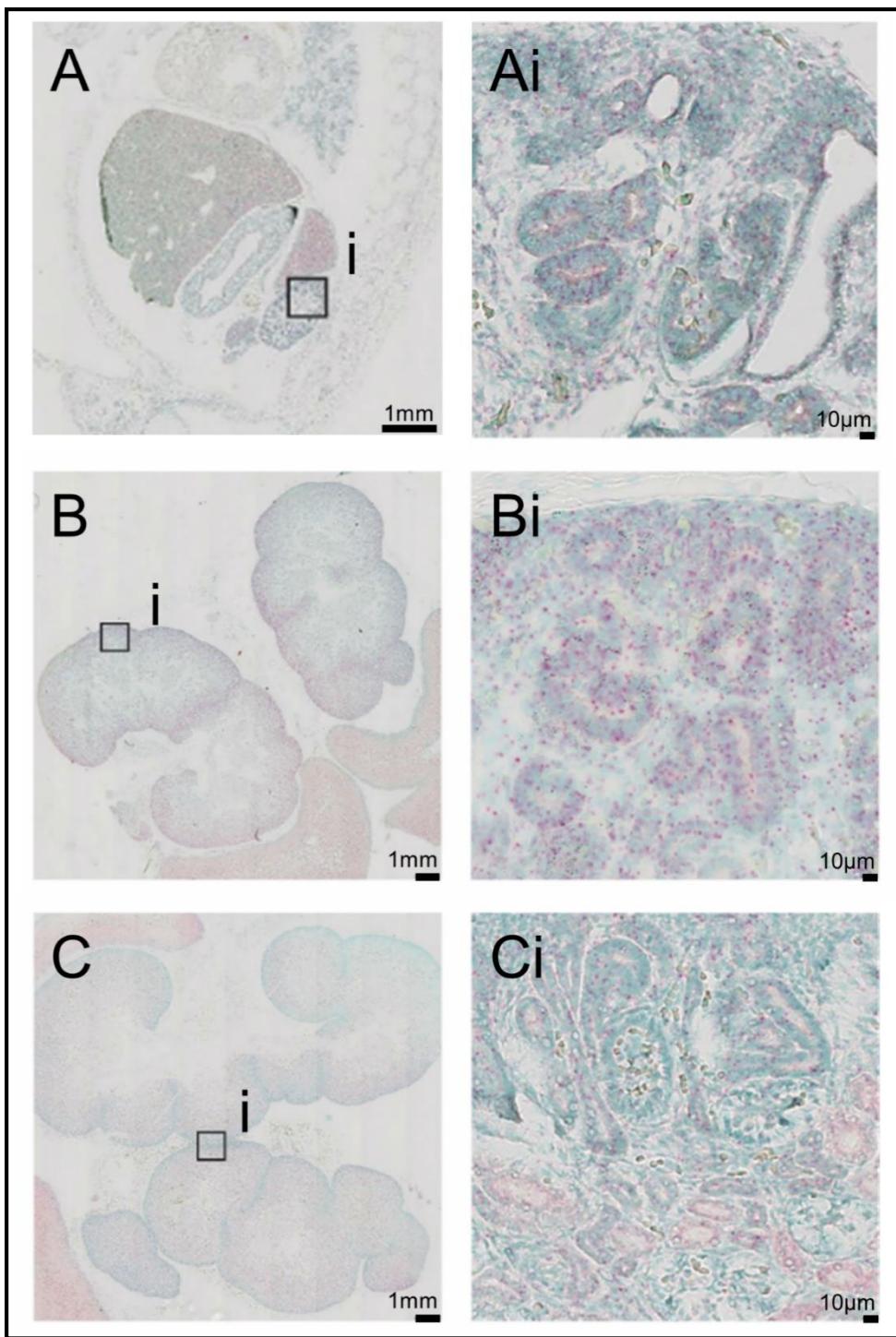


Figure 3.16. Expression of *CEP120* (red) in sagittal sections of human kidney at 8 PCW (A), 14 PCW (B) and 18 PCW (C) stained using RNAscope, counterstained with Methyl Green. Ai: At 8 PCW, *CEP120* expression is observed in the developing kidney cortex. Bi: At 14 PCW, *CEP120* expression is maintained in kidney cortex. Ci: At 18 PCW, *CEP120* expression in kidney cortex seem to have decreased.

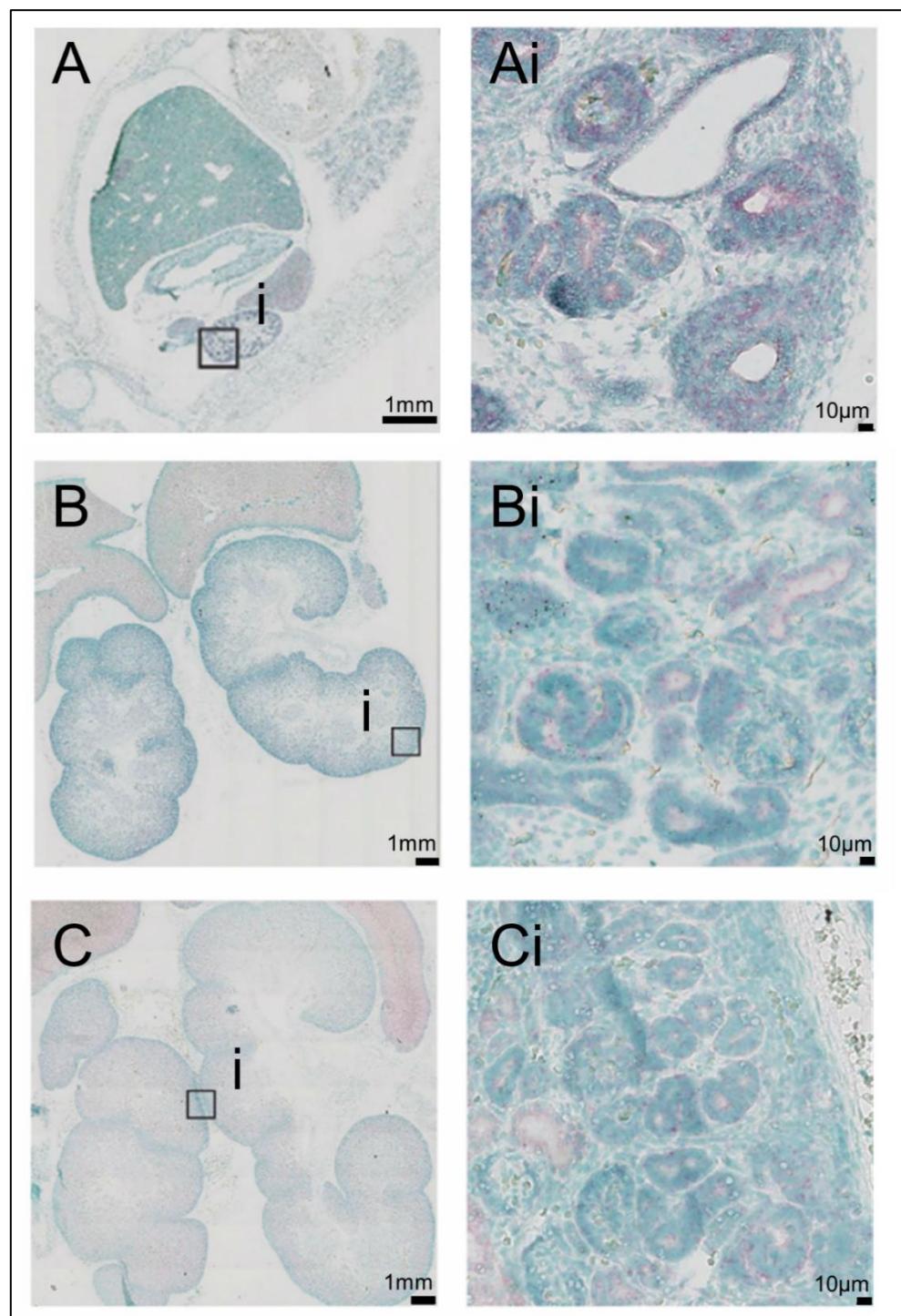


Figure 3.17. Expression of *ARL3* (red) in sagittal sections of human kidney at 8 PCW (A), 14 PCW (B) and 18 PCW (C) stained using RNAscope, counterstained with Methyl Green. At 14 PCW (Bi) and 18 PCW (Ci), *ARL3* expression is maintained in developing cortical nephrons.

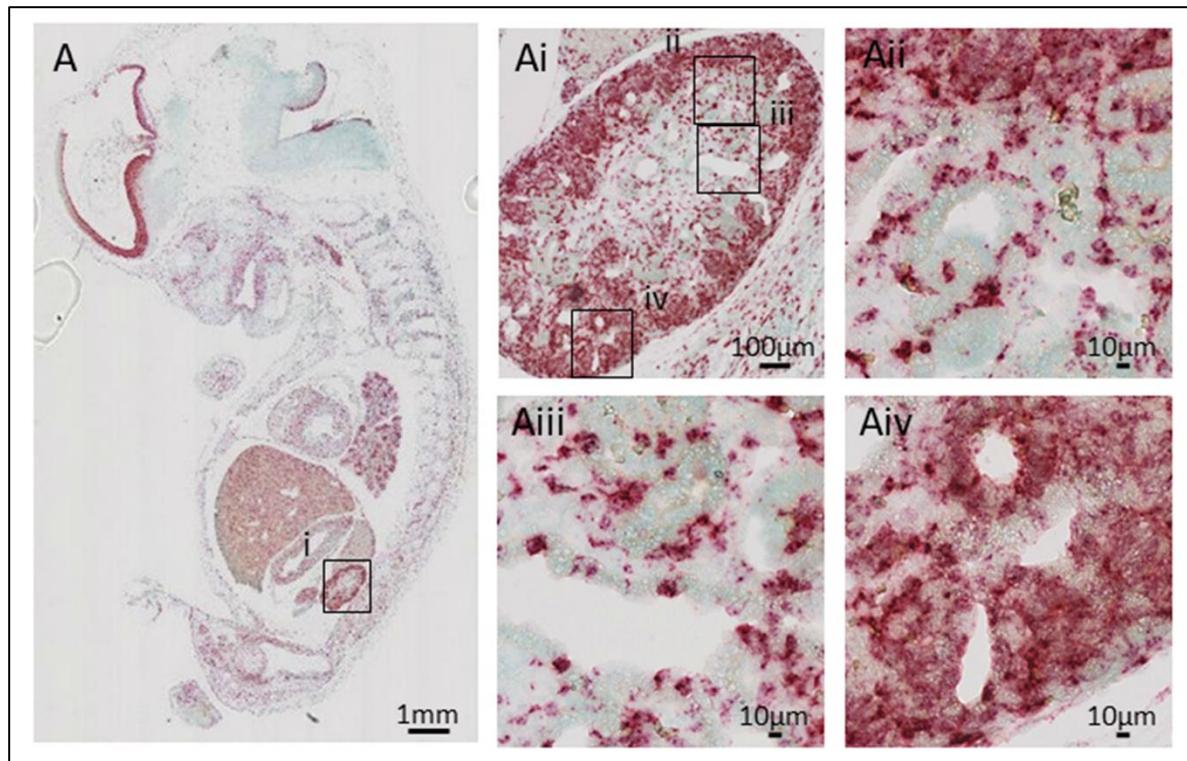


Figure 3.18. Expression of *Ki67* (red) in the developing human kidney of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green. *Ki67*, a cell proliferation marker, was used as a positive control. At 8 PCW (A, Ai-Aiv), *Ki67* expression is observed in the developing kidney cortex (Aiv).

3.2.7. Expression of *CEP120* and *ARL3* in other major organs

Primary ciliopathies can affect multiple organs (Braun and Hildebrandt, 2017), therefore I looked at other major organs with *CEP120* and *ARL3* expression. It is observed that in the developing heart, lung and gut of an 8 PCW-stage human embryo, *CEP120* (Figure 3.19) and *ARL3* (Figure 3.20) are expressed at low levels. Prominent *CEP120* or *ARL3* expression in other organs of an 8 PCW-stage human embryo is not observed.

Ki67 is highly expressed in developing heart, lung and gut (Figure 3.21), indicating cell proliferation in these tissues.

As a negative control, *dapB* (a bacterial gene, not expressed in human tissues) was used. This negative control was used in developing human hindbrain (14 PCW and 19 PCW) (Figures 3.22A and 3.22.B) and eye (14 PCW) (Figure 3.22C), and any expression in the negative control was not observed as expected.

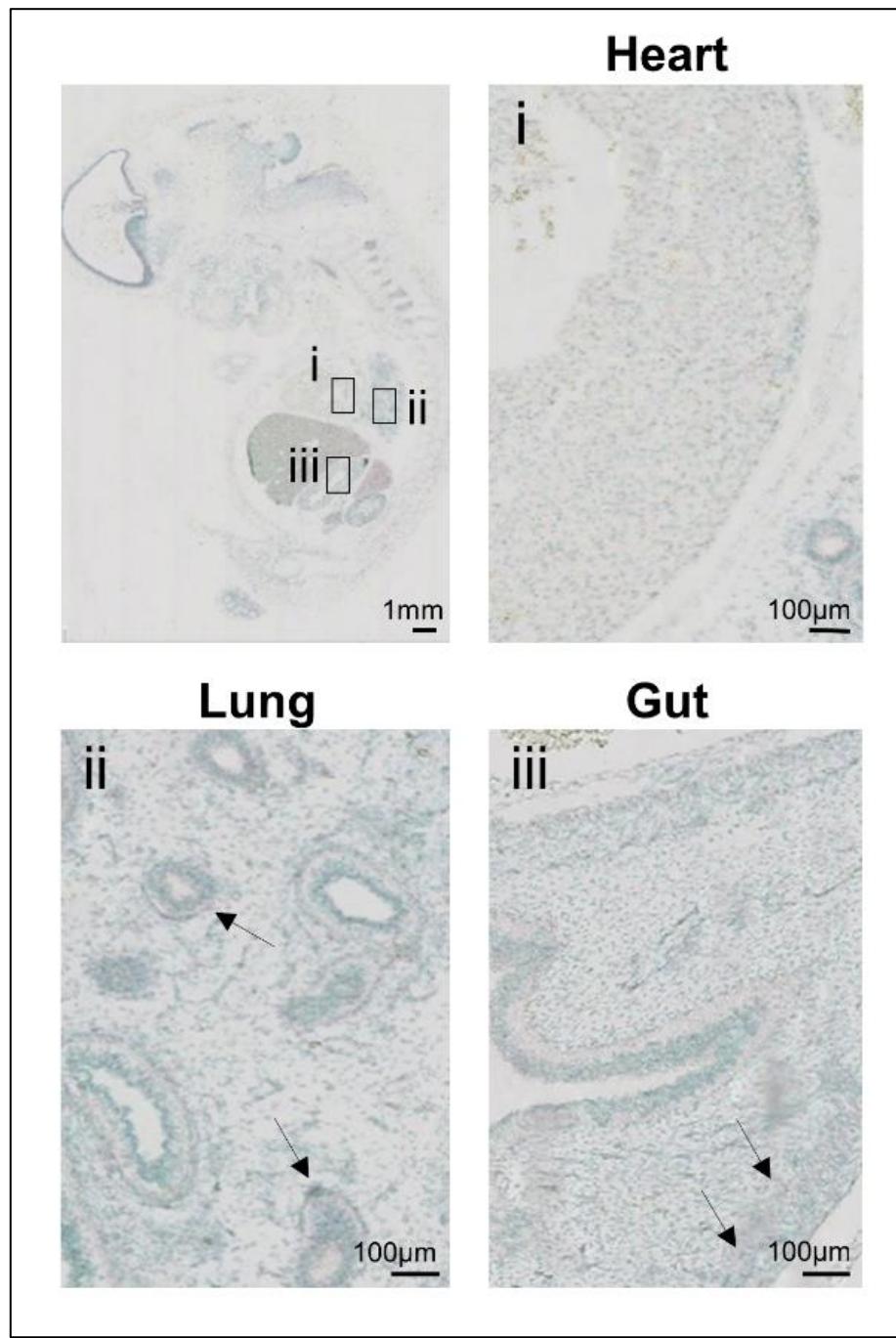


Figure 3.19. Expression of *CEP120* (red) in sagittal sections of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green. With a focus on heart (i) lung (ii) and gut (iii) tissues. It is observed how at 8PCW, *CEP120* is not highly expressed in these developing tissues. i: *CEP120* is weakly expressed in the heart. ii: *CEP120* is slightly expressed in the alveoli (arrows) in the lung. iii: There is some weak *CEP120* expression in the gut (arrows).

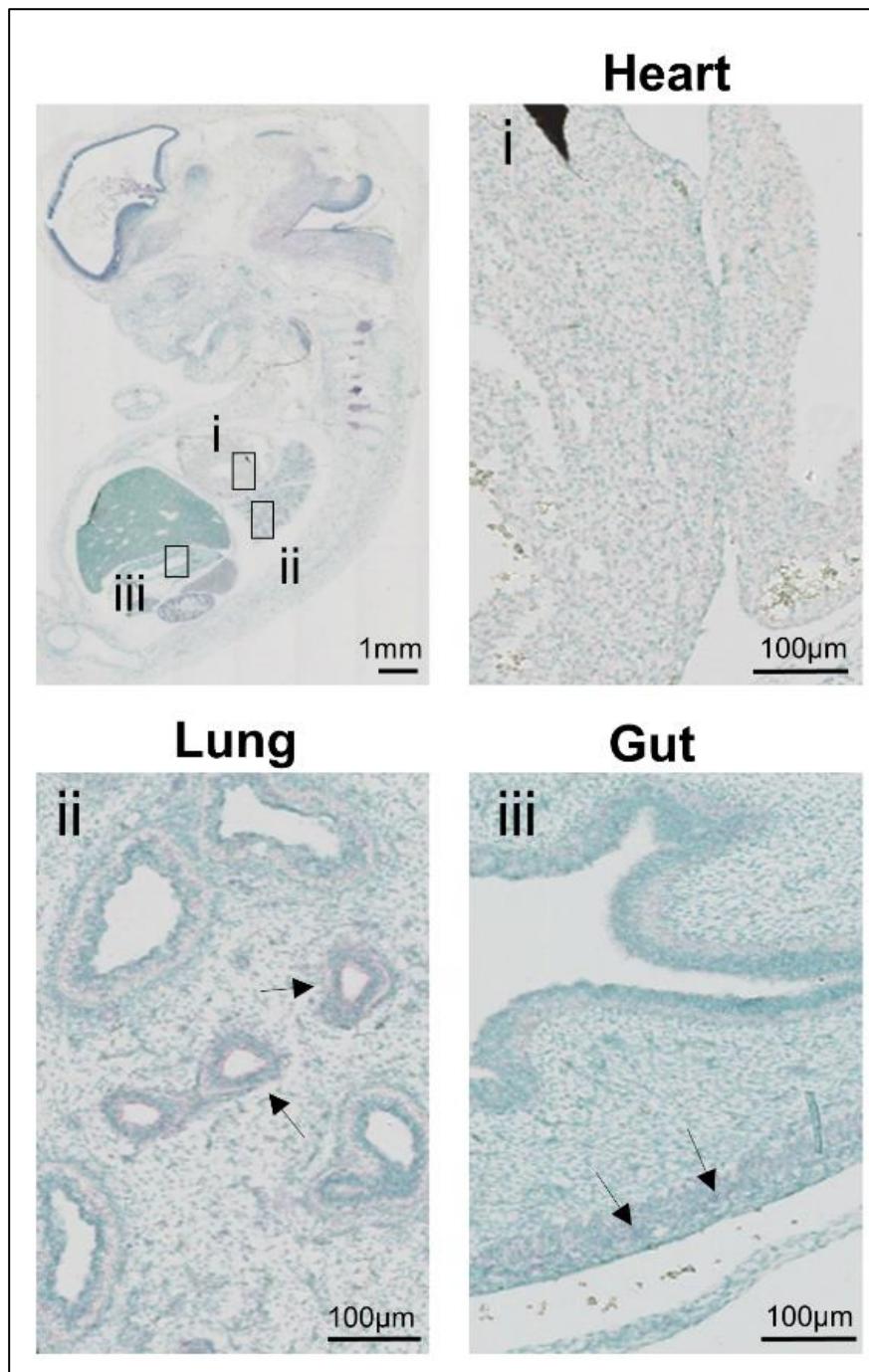


Figure 3.20. Expression of *ARL3* (red) in sagittal sections of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green. With a focus on heart (i) lung (ii) and gut (iii) tissues. It is observed how at 8PCW, *ARL3* is not highly expressed in these developing tissues. i: *ARL3* is weakly expressed in the heart. ii: *ARL3* is slightly expressed in the alveoli (arrows) in the lung. iii: There is some weak *ARL3* expression in the gut (arrows).

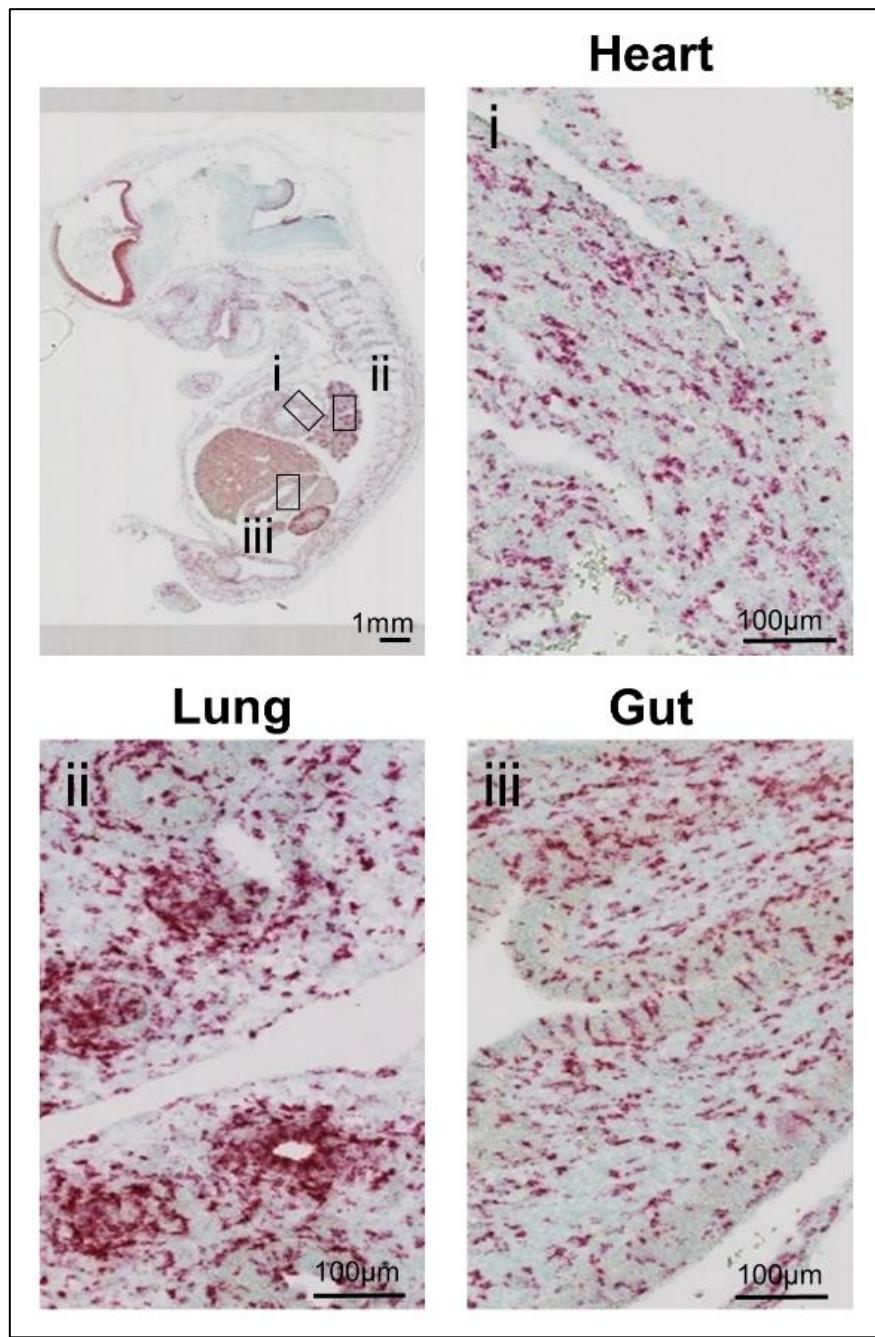


Figure 3.21. Expression of *Ki67* (red) in sagittal sections of a sagittal section of an 8 PCW-stage human embryo stained using RNAscope, counterstained with Methyl Green. With a focus on heart (i) lung (ii) and gut (iii) tissues. *Ki67*, a cell proliferation marker, was used as a positive control. It is observed how at 8PCW, *Ki67* is expressed ubiquitously in these developing tissues. There is *Ki67* expression in the heart (i), lung (ii) and gut (iii).

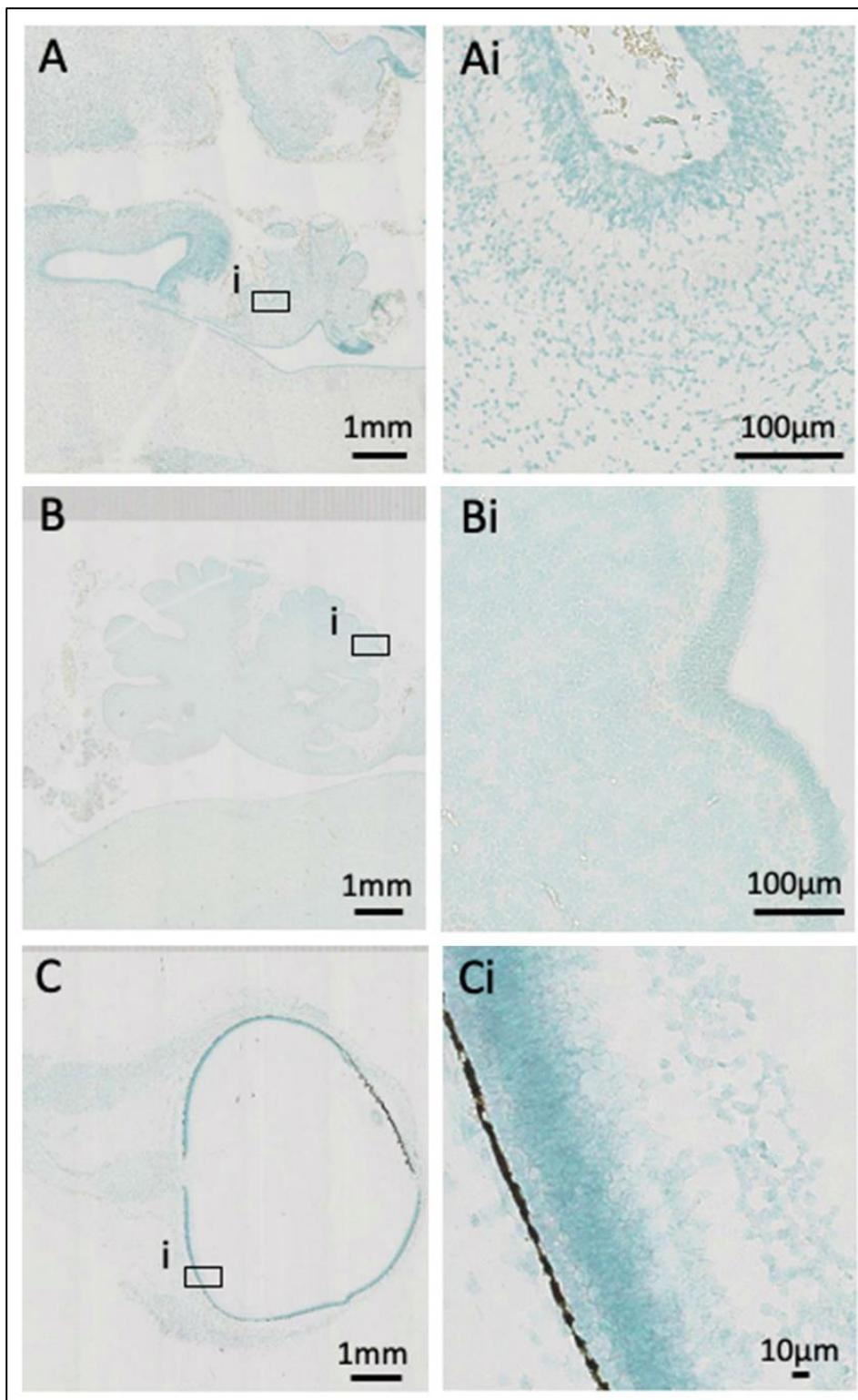


Figure 3.22. Expression of *dapB* (red) in sagittal sections of human hindbrain at 14 PCW (A) and 19 PCW (B) and an eye section at 14 PCW (C) stained using RNAscope, counterstained with Methyl Green. *dapB* was used as a negative control. *dapB* is a bacterial gene, not expressed in human tissue. There is not *dapB* expression in the developing cerebellum at 14 PCW (Ai) and 19 PCW (Bi) and the developing eye at 14 PCW (Ci).

3.2.8. Immunohistochemistry of OLIG2 and PAX6 transcription factors in human developing brain

OLIG2 and PAX6 are two transcription factors that are involved in the development of neural tissues (Mo and Zecevic, 2008, Alzu'bi et al., 2017). Both proteins have been found in proliferative regions of the developing brain (Alzu'bi et al., 2017, Alzu'bi and Clowry, 2019).

Using immunohistochemistry, at 8 PCW it is observed OLIG2 at the ventricular zone (VZ) and subventricular zone (SVZ) of the medial and lateral ganglionic eminences (MGE and LGE respectively) and cerebral cortex (Figure 3.23).

PAX6 at 8 PCW are observed to be expressed in VZ and SVZ of the ganglionic eminence and cerebral cortex and choroid plexus (Figure 3.24).

PAX6 was observed in the hindbrain at 8 PCW (Figure 3.25) and 19 PCW (Figure 3.26). At 8 PCW, PAX6 was observed in the rhombic lip (RL) (Figure 3.25i) and VZ (Figure 3.25ii), which are two zones of neurogenesis in the developing cerebellum (Halidipur et al., 2019). PAX6 was also observed in other regions of the developing hindbrain (Figure 3.25iii-v). At 19 PCW, PAX6 is observed at the external granule layer (EGL), a proliferative layer originally generated from the RL which is essential for the proliferation of cerebellar granule cells (Marzban et al., 2014), and IGL (Figure 3.26).

In developing human hindbrain (at 19 PCW) it was not observed any staining in the negative control (immunohistochemistry as performed with the staining of PAX6 and OLIG2 but without adding primary antibodies) (Figure 3.27).

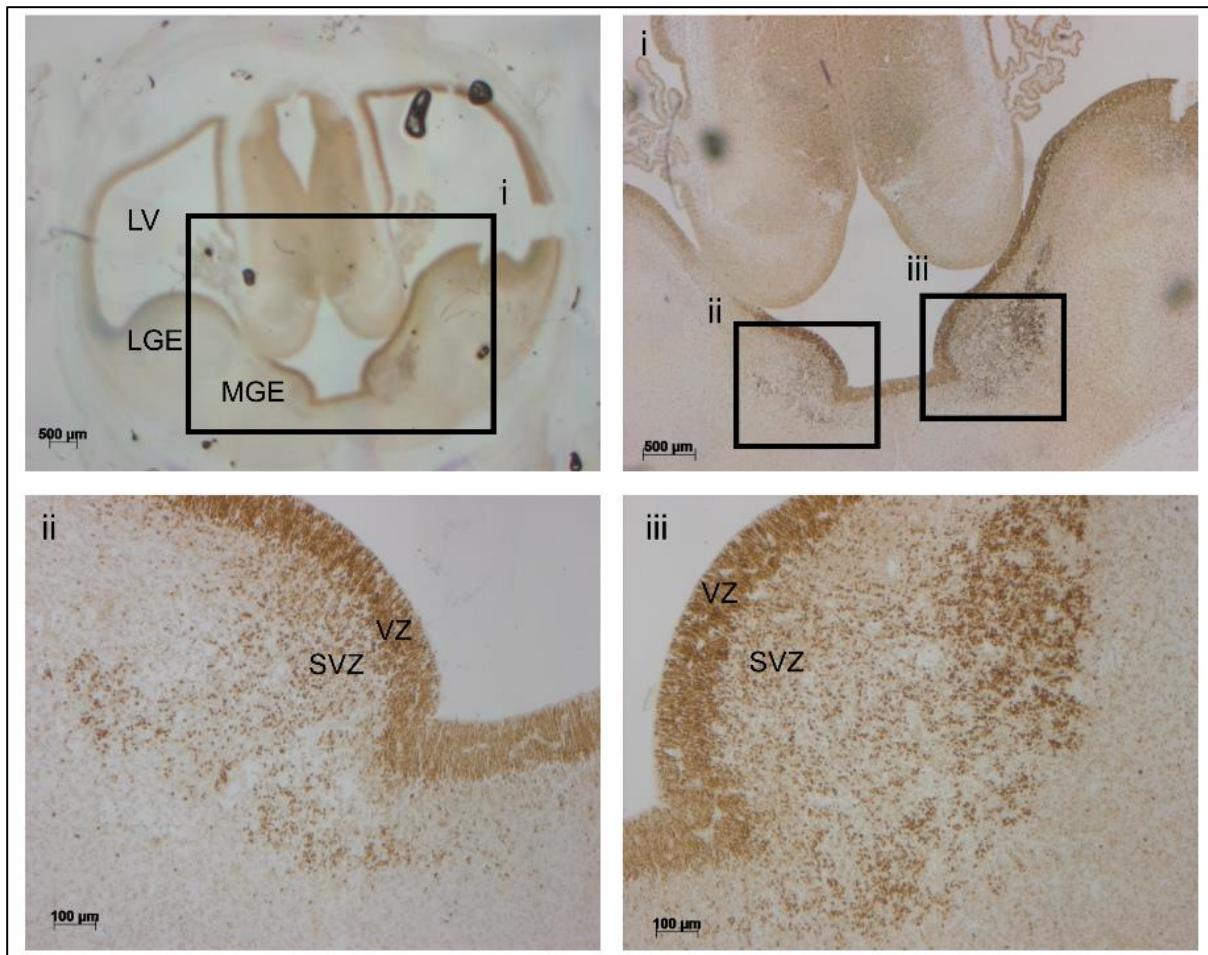


Figure 3.23. Immunohistochemistry staining of OLIG2 in developing human brain. Cortical section of 8 PCW-stage human embryo stained with OLIG2 (brown) using immunohistochemistry counterstained with Toluidine Blue. OLIG2 staining is observed in the VZ and SVZ of the ganglionic eminences and cerebral cortex (i-iii). LGE, lateral ganglionic eminence; LV, lateral ventricle; MGE, medial ganglionic eminence; SVZ, subventricular zone; VZ, ventricular zone. Zeiss AxioPlan microscope was used for visualisation.

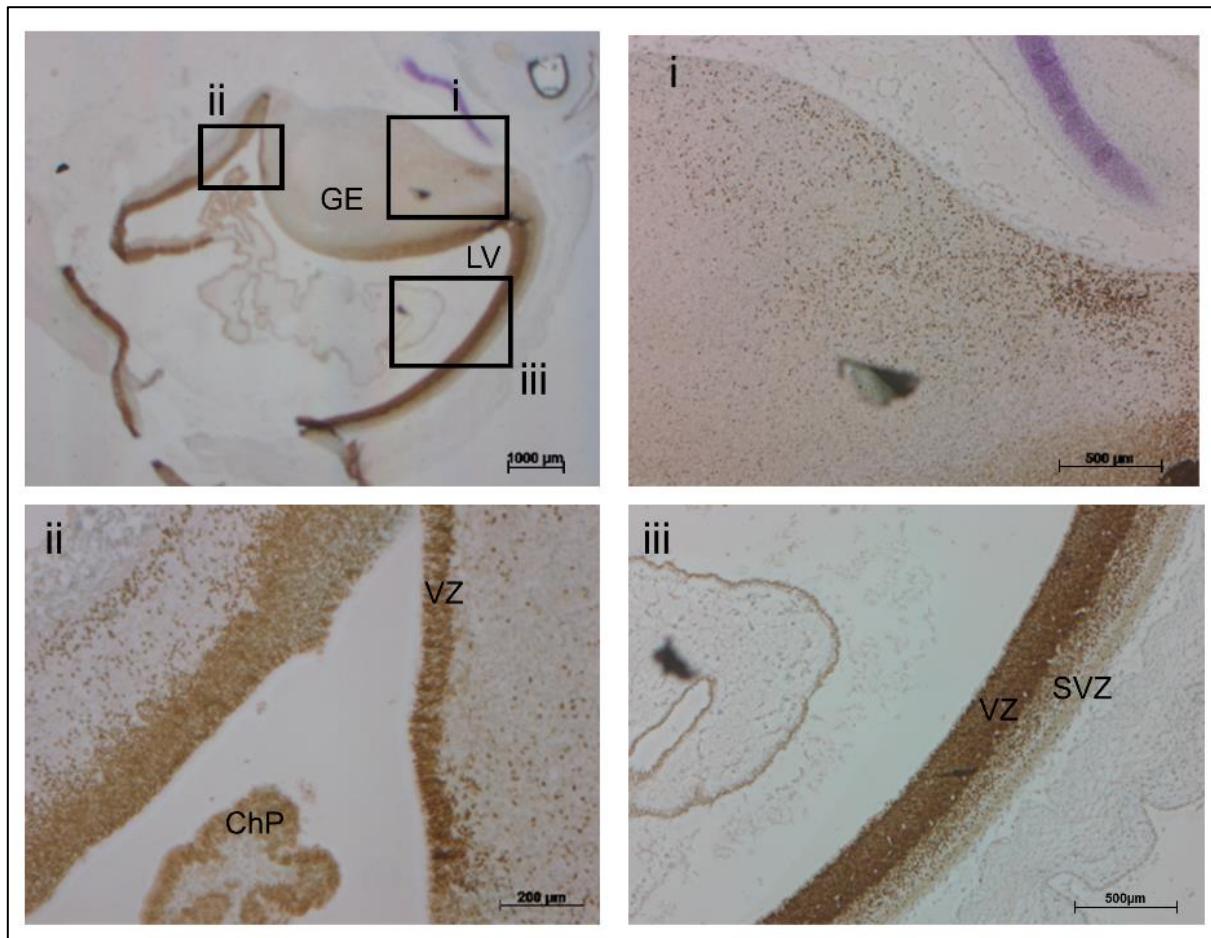


Figure 3.24. Immunohistochemistry staining of PAX6 in developing human brain. Sagittal section of 8 PCW-stage human embryo stained with PAX6 (brown) using immunohistochemistry counterstained with Toluidine Blue. PAX6 staining is observed in the VZ and SVZ of the ganglionic eminences and cerebral cortex and choroid plexus (i-iii). ChP, choroid plexus; LV, lateral ventricle; GE, ganglionic eminence; SVZ, subventricular zone; VZ, ventricular zone. Zeiss AxioPlan microscope was used for visualisation.

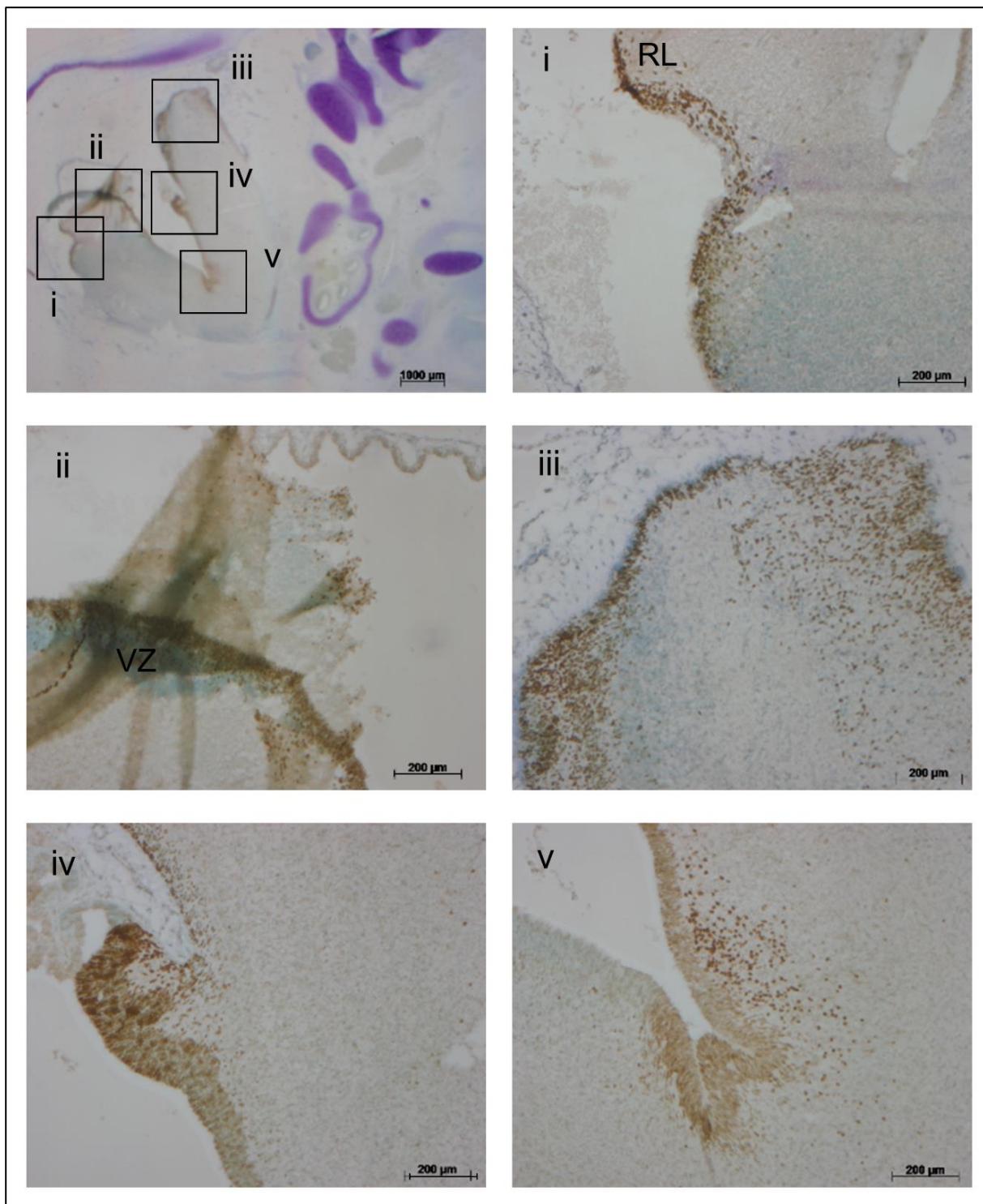


Figure 3.25. Immunohistochemistry staining of PAX6 in developing human hindbrain. Sagittal section of 8 PCW-stage human embryo stained with PAX6 (brown) using immunohistochemistry counterstained with Toluidine Blue. PAX6 staining is observed in the rhombic lip (i) and spreading out into VZ (ii). PAX6 is also observed in other parts of the developing hindbrain (iii-v). RL, rhombic lip, VZ, ventricular zone. Zeiss AxioPlan microscope was used for visualisation.

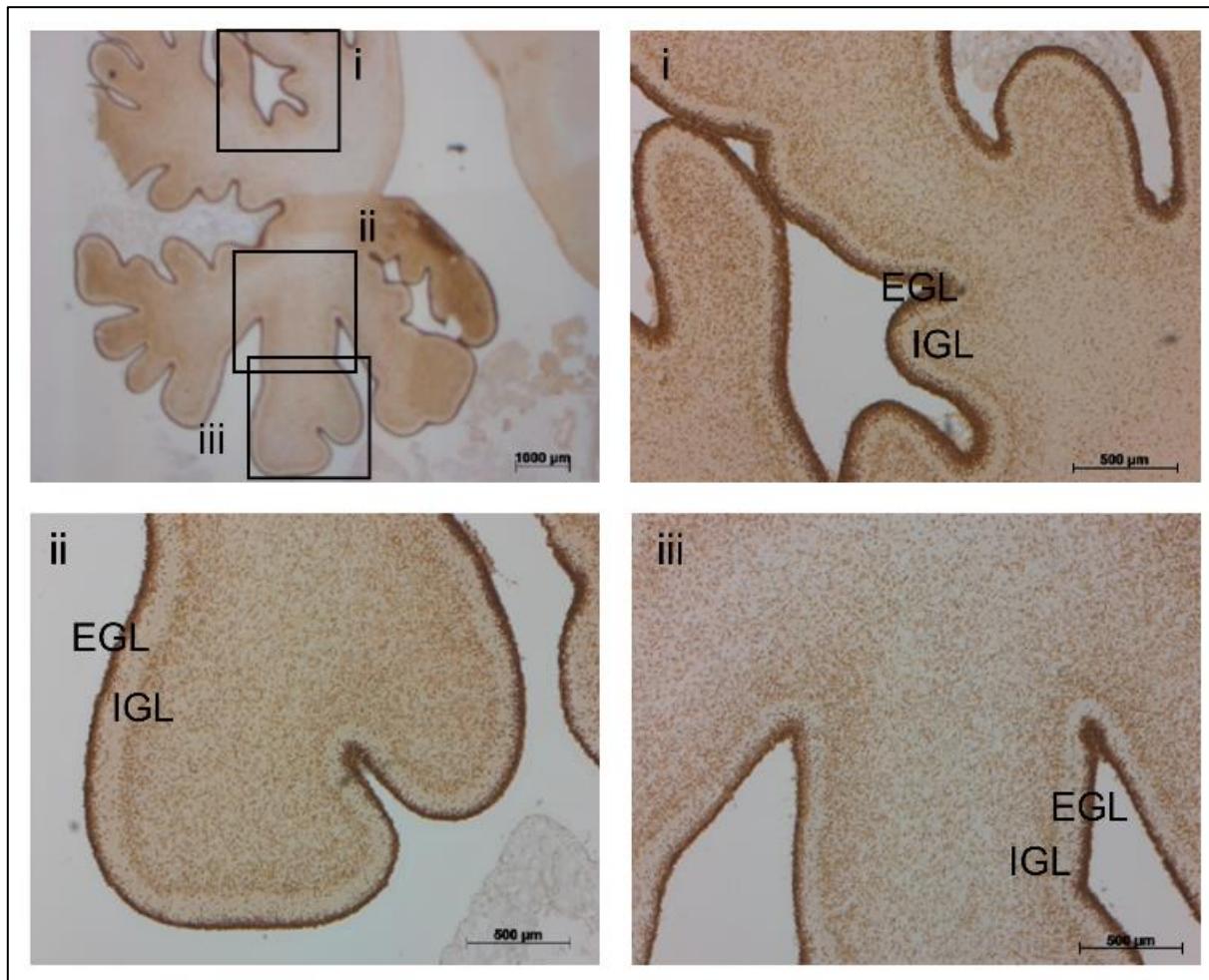


Figure 3.26. Immunohistochemistry staining of PAX6 in developing human hindbrain. Sagittal section of human hindbrain at 19 PCW stained with PAX6 (brown) using immunohistochemistry counterstained with Toluidine Blue. PAX6 is observed widely expressed in the developing cerebellum, including the developing EGL and IGL (i-iii). PAX6 is highly expressed in the developing EGL. EGL, external granule layer; IGL, inner granule layer. Zeiss AxioPlan microscope was used for visualisation.

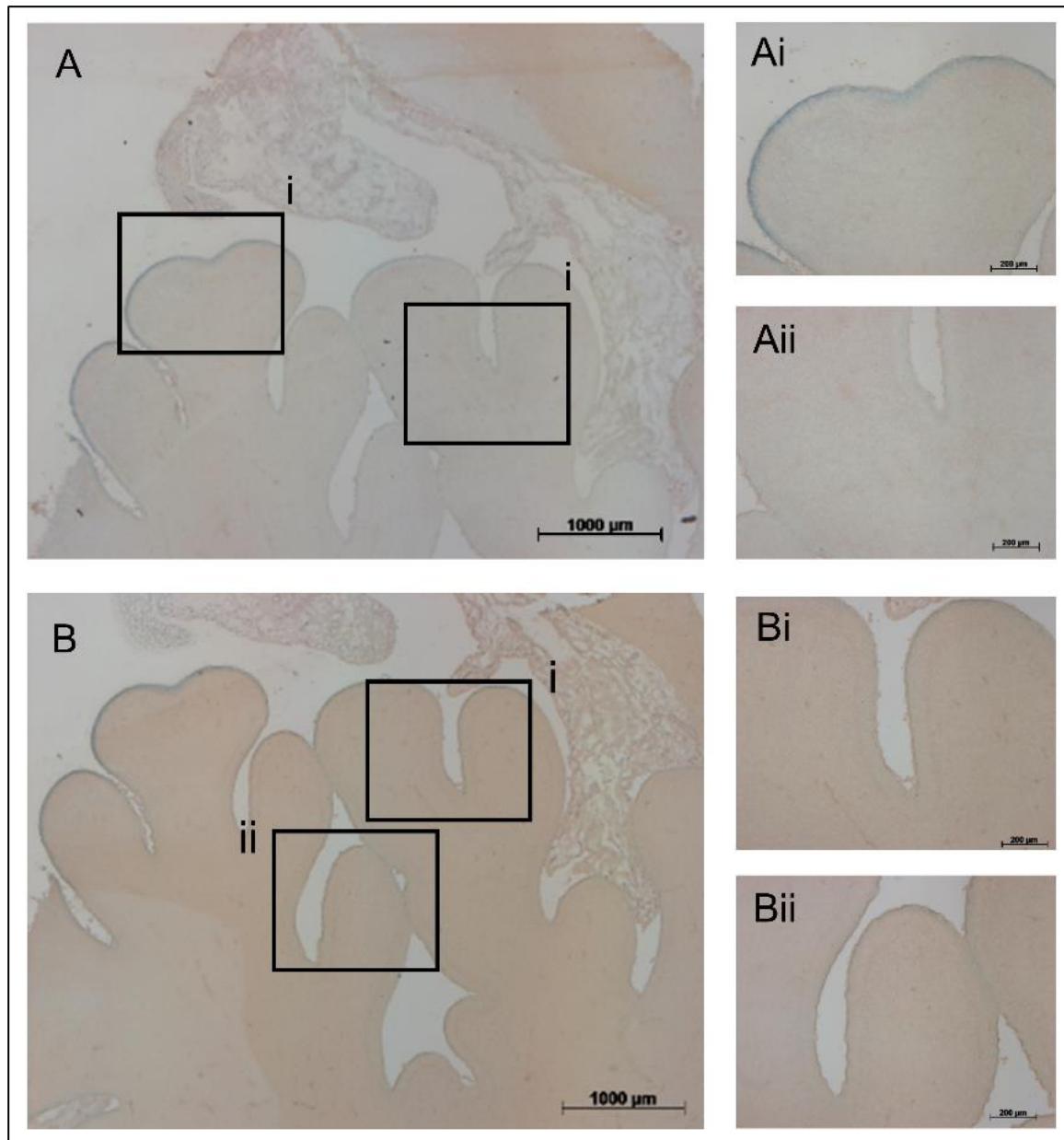


Figure 3.27. Immunohistochemistry staining in developing human hindbrain (negative control in which the sample was not treated with primary antibody). Sagittal section of human hindbrain at 19 PCW, using immunohistochemistry counterstained with Toluidine Blue (no primary antibody was used). There is not staining as expected (A, Ai-Aii, B, Bi-Bii). Zeiss AxioPlan microscope was used for visualisation.

3.3. Discussion

There are over 40 genes associated with JBTS (Gana et al., 2022, Bachmann-Gagescu et al., 2020). Biallelic mutations in *CEP120* and *ARL3* are some of the causes of JBTS (Roosing et al., 2016, Alkanderi et al., 2018). It is unknown how these genes, encoding proteins with different functions and locations are able to cause JBTS.

CEP120 protein is known to be expressed in multiple mouse tissues (Xie et al., 2007), including adult brain, kidney, liver and lung tissues and mouse embryonic brain (Xie et al., 2007). According to the Human Protein Atlas online resource (Uhlén et al., 2015), CEP120 RNA is expressed in brain (<https://www.proteinatlas.org/ENSG00000168944-CEP120/tissue>), however protein CEP120 expression is not available. In protein atlas, ARL3 protein seems to be widely expressed within human tissues, including cerebellum, kidney and retina at RNA and protein levels (<https://www.proteinatlas.org/ENSG00000138175-ARL3/tissue>), which is consistent with the known ARL3 associated phenotypes.

Using RNAscope RNA *in situ* hybridisation assay, in collaboration with HDBR resource (Robinson et al., 2014, Lindsay et al., 2016) CEP120 and ARL3 expression patterns were characterised in human embryos and foetuses. The results explained in this chapter have been published as an article in the journal BMC Developmental Biology (Powell et al., 2020).

CEP120 was observed to be expressed in multiple organs and tissues and compared the expression patterns of CEP120 at different embryonic and fetal stages with the expression patterns of ARL3.

It was observed that both CEP120 and ARL3 genes are expressed in multiple tissues, including developing cerebellum, kidney and retina , which is consistent with the key organs affected and phenotypes observed in patients reported in the literature with disease-causing mutations in CEP120 and ARL3 (Roosing et al., 2016, Alkanderi et al., 2018, Sheikh et al., 2019). CEP120 expression in these key tissues is also consistent with the mouse Cep120 expression previously reported in the literature (Xie et al., 2007).

Both genes are evolutionary conserved in mammals. The protein sequence of ARL3 is more evolutionary conserved than the protein sequence of CEP120. For instance, considering the orthologues both genes in *Danio rerio* (zebrafish), a vertebrate model organism, it can be observed that the human ARL3 protein sequence shares more than 90% of homology with the two arl3 zebrafish protein sequences. In contrast, the human CEP120 protein sequence shares more than 57% of homology with the zebrafish cep120 protein sequence. Moreover, orthologues of ARL3 can be found in the non-

vertebrate model organisms *Drosophila melanogaster*, *Caenorhabditis elegans* and *Chlamydomonas reinhardtii* with a homology percentage higher than 60% with the human ARL3, however these orthologues of CEP120 were not found in these organisms. It was observed that the protein sequence of ARL3 seems to be more evolutionary conserved than the protein sequence of CEP120.

Considering the data available in public databases GTEx and HPA it is observed that there is RNA expression of *CEP120* and *ARL3* in tissues affected in JBTS (Table 3.4). Other genes associated with JBTS such as *CC2D2A* and *CEP164* are also expressed in tissues affected in JBTS.

These databases can be used as a primary assessment of tissue specific RNA expression in ciliopathy genes. It is important to consider that each of these ciliopathy genes may have different expression patterns. Therefore, comparisons between expression levels of different ciliopathy genes must be made with caution.

It is also important to consider that the data obtained from these two public databases is limited to RNA expression data from a representative selection of ciliopathy genes. Besides, these two databases are not focussed on embryonic or foetal tissues human gene expression (The GTEx Consortium, 2020, Thul and Lindskog, 2018). This is particularly relevant as multiple phenotypes found in primary ciliopathies arise during development, such as the MTS cerebellar malformation characteristic of JBTS (Xie et al., 2007, Romani et al., 2013). Ciliopathy genes, in human developing tissues may have different RNA expression patterns compared with tissues from adults. In mice, it was observed that *Cep120* expression in mouse brain decreases during development (Xie et al., 2007). Expression patterns of other ciliopathy genes such as *CEP290* (Cheng et al., 2012, Ramsbottom et al., 2020) and *CEP164* (Devlin et al., 2020), have been characterised during mouse and human development and reflect the importance of investigating gene expression in different developing tissues to study the molecular genetics of JBTS and role of tissue expressivity of these genes in ciliopathy phenotypes.

The characterisation of *CEP120* and *ARL3* expression patterns in embryonic and foetal tissues from HDBR at Newcastle University, allows to study the role of tissue expressivity in JBTS and other ciliopathies from a unique perspective focusing on the expression in the developing tissues associated with primary ciliopathies.

It was observed that the expression patterns of *CEP120* and *ARL3* are very similar during human development, including developing brain, eye, kidney and dorsal root ganglia. I focused on the developing human cerebellum, as it is the main distinct tissue affected in JBTS (Wu et al., 2014, Rahimi-Balaei et al., 2018). Cerebellar neurogenesis during early brain development is important for the proliferation of the neurons in the cerebellum. The cerebellum contains diverse types of neurons: glutamatergic excitatory neurons (such as cerebellar granule cells, the most abundant neurons in the vertebrate brain) and GABAergic inhibitory neurons (such as Purkinje cells of the ML) (Marzban et al., 2014).

In a previous study a conditional *Cep120* knockout (for the central nervous system), was phenotypically characterised, and cerebellar hypoplasia was observed. It was suggested that the lack of proliferation of cerebellar granule neuron progenitors (CGNP) could be responsible for the cerebellar hypoplasia (which is a phenotype within the JBTS phenotypic spectrum) observed in this mouse model (Wu et al., 2014). Moreover, it is known that Sonic hedgehog (Shh) induces CGNP proliferation, and in this study they observed lack of cilia in the CGNP and ependymal cells of the mutant mice.

Subsequently, in this study it was suggested that the failure of expansion of CGNP can be caused by these CGNP being unable to respond to the Shh signalling (Wu et al., 2014). Furthermore, in recent years other studies have investigated the role of *Cep120/CEP120* in the development of the cerebellum and in axon formation (Chang et al., 2021, Meka et al., 2022).

I investigated the *CEP120* and *ARL3* expression in the human developing cerebellum and it was found that both genes were expressed similarly in the EGL and IGL layers at 14 PCW. However, at 19 PCW, *CEP120* expression in the EGL and ML layers was observed, while *ARL3* expression was observed at the IGL predominantly, and the EGL layer of the cerebellum (Figures 3.7-3.9).

These results suggests that the expression of both *CEP120* and *ARL3* have a role in the development of the human cerebellum, which is consistent with the clinical phenotypes caused by mutations in these two JBTS genes (Roosing et al., 2016, Alkanderi et al., 2018) and with the *Cep120* mouse model (Wu et al., 2014).

The differences between the expression of *CEP120* and *ARL3* at 19 PCW, may occur because they are expressed in different cell populations during cerebellum development. These results may suggest that there is specific expression of *CEP120* in immature (in the EGL layer), migratory granule cells and interneurons of the ML (Figures 3.6 and 3.9) (Sotelo, 2015). In contrast, *ARL3* seems to be expressed in immature and mature granule cells (in EGL and IGL layers) (Figures 3.7 and 3.9).

Via immunohistochemistry of PAX6 and OLIG2, the staining of these two transcription factors were observed in proliferative zones in human developing brain, such as the VZ and SVZ of the ganglionic eminences and cerebral cortex (Figures 3.23 and 3.24) and in the developing cerebellum (Figures 3.25 and 3.26), as expected.

Staining of OLIG2 and PAX6 in proliferative zones of the human developing brain (at 8 PCW) (Figures 3.23-3.25). PAX6 staining was observed in the proliferative cell layers: EGL and IGL of the developing cerebellum (at 19 PCW) (Figure 3.26). Staining of PAX6 and OLIG2 in human developing brain tissues (Figures 3.23-3.26) contributes to define these proliferative regions and can be compared with the RNA scope staining of *CEP120* and *ARL3*. This suggests that both *CEP120* and *ARL3* are also expressed in proliferative regions within the human development brain, including the EGL layer of the developing cerebellum (Figures 3.3, 3.4, 3.6 and 3.7).

This also suggests that using RNAscope and immunohistochemistry staining with specific cell markers (for instance with Calbindin 1, a marker of Purkinje cells) (Aldinger et al., 2021), could be compared to determine if at 19 PCW *CEP120* is expressed in the ML, or if on the contrary, it is expressed in the EGL and IGL exclusively.

Additionally, there is some expression *CEP120* and *ARL3* in the dorsal root ganglia, which may indicate that *CEP120* and *ARL3* have a role in the differentiation of primary sensory neurons (Figures 3.13 and 3.14).

Expression of *CEP120* and *ARL3* was also observed in the developing eye, more particularly, in the multiple layers of the retina (Figures 3.11 and 3.12). This may indicate that *CEP120* and *ARL3* have a role in the development of the retinal layers. This is consistent with retinal phenotypes found in patients reported in the literature with

monoallelic and biallelic variants in *ARL3* (Alkanderi et al., 2018, Holtan et al., 2019, Strom et al., 2016).

CEP120 and *ARL3* were also expressed in the developing kidney at 8 PCW, 14 PCW and 18 PCW. *CEP120* expression (Figure 3.16) was observed in the developing nephrons and the developing renal cortex. In contrast, *ARL3* expression seems to be expressed in the nephrons exclusively (Figure 3.17).

Additionally, it is important to note that *CEP120* (Figure 3.19) and *ARL3* (Figure 3.20) expression was minimal in the developing heart, lung and gut tissues. This is consistent with the absence of clinical phenotypes in these tissues as a result of the reported disease-causing *CEP120* and *ARL3* mutations (Roosing et al., 2016, Alkanderi et al., 2018).

It is important to consider that *CEP120* and *ARL3* proteins have different functions. *ARL3* is involved in vesicle trafficking and ciliary signalling (Gotthardt et al., 2015, Ismail et al., 2011). In contrast, *CEP120* does not function at the primary cilia, it is a centrosomal protein involved in centrosome assembly, elongation and duplication (Mahjoub et al., 2010, Wu et al., 2014, Meka et al., 2022). These differences between *CEP120* and *ARL3* protein functions may be reflected in the differences in expression patterns within developing tissues. For instance, as *CEP120* is involved in centriole and cilia biogenesis, it may more widely expressed in ciliated developing tissues (Reiter and Leroux, 2017, Tucker et al., 1979). In contrast, *ARL3* expression may be restricted to developing tissues and time points in which cell signalling is actively occurring, including in granule neuron progenitors of the internal granule cell layer and nephron progenitors in the developing kidney.

In conclusion, human expression patterns of *CEP120* and *ARL3* were characterised in human embryonic and foetal tissues. This study provided insights into the wide spectrum of clinical phenotypes caused by mutations in *CEP120* and *ARL3* genes. This study is a step further towards the understanding of the molecular disease mechanisms of *CEP120* and *ARL3*, two genes associated with the same primary ciliopathy, JBTS.

Chapter 4. Investigation of Joubert syndrome (JBTS) and JBTS candidate genes in the Genomics England 100,000 Genomes Project, the potential use of exon therapies and the use of cell-based assays to complement genetic studies

4.1. Introduction and aims

One of the main characteristics of ciliopathies is their phenotypic heterogeneity and overlap of clinical phenotypes (Novarino et al., 2011, Shaheen et al., 2016). Ciliopathies are also characterised by genetic pleiotropy and genetic heterogeneity. Mutations within the same gene may lead to distinct ciliopathy syndromes (gene pleiotropy) (Coppieters et al., 2010, Roosing et al., 2016, Focsa et al., 2021, Shaheen et al., 2016). In addition, mutations in different genes can cause the same ciliopathy syndrome. These characteristics are clearly observed in JBTS, with more than 40 genes associated, many of those associated with other ciliopathies as well other than JBTS (Focsa et al., 2021, Bachmann-Gagescu et al., 2020, Mitchison and Valente, 2017, Gana et al., 2022, Parisi, 2019). For example, two genes associated with JBTS, and a spectrum of clinical phenotypes, are *CEP120* and *CC2D2A*.

Mutations in *CEP120* are associated with JBTS as well as JATD, TCDOE and MKD/OFD phenotypes (Roosing et al., 2016, Shaheen et al., 2015). Mutations in *CC2D2A* are associated with a wide spectrum of phenotypes ranging from phenotypes restricted to one single organ such as it is the case of isolated rod-cone dystrophy (RCD) (Mejecase et al., 2019) to ciliopathies involving phenotypes affecting multiple organs such as JBTS (Bachmann-Gagescu et al., 2012) and MKS (Mougou-Zerelli et al., 2009, Szymanska et al., 2012). *CEP120* and *CC2D2A* encode proteins with different functions and cellular localizations, and mutations in these genes can lead to distinct clinical phenotypes as well as partially overlapping phenotypes (Roosing et al., 2016, Mejecase et al., 2019).

Regarding genotype-phenotype correlations in *CC2D2A*-related disease, it has been suggested that the genetic pleiotropy of ciliopathies caused by *CEP290* or *CC2D2A* mutations can be due to basal levels of alternative splicing with exon skipping (Drivas et al., 2015). This occurs with pathogenic *CEP290* or *CC2D2A* mutations located in

“skippable” exons (exons whose length is an exact multiple of 3 and it is not the first or last coding exon) that maintain the reading frame if they are skipped. The complete deleterious effect of these mutations is prevented as cells are able to have some residual near-full length protein, as they escape nonsense-mediated mRNA decay (NMD), which is associated with milder clinical phenotypes, than the phenotypes caused by mutations in non-skippable exons (Drivas et al., 2015).

It has also been noted that *CC2D2A* loss-of-function (LoF) variants (nonsense or frameshift) lead to more severe phenotypes such as MKS or ML, compared with patients with at least one *CC2D2A* missense mutation (Bachmann-Gagescu et al., 2012, Mougou-Zerelli et al., 2009). If there is a similar correlation for *CEP120* patients, it is unknown.

To date, there are nine *CEP120* mutations described in patients with JBTS, JATD MKS/ODF and TCDOE syndromes. Describing any genotype-phenotype correlation in patients with pathogenic *CEP120* mutations and understanding why these mutations can cause a broad spectrum of clinical phenotypes is challenging, as there are only nine patients reported with biallelic (homozygous or compound heterozygous) disease-causing variants in *CEP120*, in contrast, more than 100 patients have been described with *CC2D2A* disease-causing variants (Lam et al., 2020, Bachmann-Gagescu et al., 2012). A list of patients with biallelic *CEP120* variants and a list of *CC2D2A* biallelic variants and their associated phenotypes reported in the literature, are published in Barroso-Gil et al., 2021 (Barroso-Gil et al., 2021a).

Exon skipping strategy is widely known as an approach to treat Duchenne muscular dystrophy (DMD) in patients with mutations located in skippable exons of the *DMD* gene (Lee et al., 2018, Komaki et al., 2018, Kole and Krieg, 2015).

ASOs-mediated exon skipping therapies to treat primary ciliopathies, are not as extended and developed as the therapies to treat DMD, however some advances and achievements have occurred in the last few years towards the development of therapies to treat patients with primary ciliopathies (Barny et al., 2019, Garanto et al., 2016, Dulla et al., 2021, Cideciyan et al., 2019, Devlin et al., 2023).

In 2015, Drivas et al. suggested basal exon skipping to be associated with the genetic pleiotropy of *CEP290* or *CC2D2A* ciliopathy genes. It was suggested that milder clinical phenotypes are associated with patients presenting mutations located in skippable exons of *CEP290* or *CC2D2A* (Drivas et al., 2015). They suggested that there is some residual near-full length protein leading to milder clinical phenotypes, compared with patients with mutations in exons that are not skippable (Drivas et al., 2015).

ASOs-mediated exon skipping in a well-known ciliopathy gene: *CEP290*, has already been investigated by Sayer and Miles laboratory at Newcastle University, partially restoring the cystic kidney phenotype in a JBTS mouse model and partially restoring the ciliary phenotype in primary cells from a JBTS patient (Molinari et al., 2019, Ramsbottom et al., 2018).

Other laboratories have also investigated ASOs-mediated exon skipping in ciliopathy genes. ASOs-mediated exon skipping approaches have been applied to treat retinal dystrophies, such as LCA, to bypass the deleterious effect of a *CEP290* deep intronic variant: c.2991 + 1655A > G by modulating the splicing effect caused by this mutation (Garanto et al., 2016). ASOs-mediated exon skipping has also been investigated to treat retinitis pigmentosa by modulating the splicing of exon 13 (Dulla et al., 2021) and exons 30-31 or 39-40 (Schellens et al., 2023) of the gene *USH2A*, to bypass the deleterious effect caused by mutations occurring in these exons.

ASOs-mediated exon skipping approaches have been used in clinical trials to modulate the deleterious splicing effect of the deep *CEP290* intronic variant: c.2991 + 1655A > G to improve vision in patients with LCA, caused by the presence of this deep intronic mutation in homozygosity, via intravitreal injections of ASOs (Cideciyan et al., 2023, Cideciyan et al., 2019).

The studies mentioned above reflect the increasing number of studies investigating the use of ASOs to modulate splicing and ASOs-mediated exon skipping as a therapeutic approach for primary ciliopathies. Besides, it has to be considered that there are multiple ciliopathy genes and disease-causing variants associated with primary ciliopathies, and some of them could potentially be suggested as potential targets for exon skipping therapies.

Exon skipping is not the only therapeutic approach being investigated for primary ciliopathies. CRISPR/Cas9-mediated genome editing has been used to rescue the *CEP290* expression in an *in vitro* model of CEP290-LCA primary ciliopathy (Ruan et al., 2017). Gene therapy using AAV vectors has been investigated to treat Usher syndrome (Whatley et al., 2020, French et al., 2020, Ivanchenko et al., 2023) .

There are therapies associated with the modulation of cell signalling pathways aimed to treat renal ciliopathies: such as mTOR inhibitors, vasopressin receptor 2 (V2R) antagonists (Duong Phu et al., 2021), agonists of prostaglandin E2 receptors (Garcia et al., 2022) and therapeutic targets that modulate Shh (Hynes et al., 2014) and cAMP signalling pathways (Malicki and Johnson, 2017). These therapeutic approaches aim to improve and develop new efficient approaches of translational and personalised medicine with a focus on treating primary ciliopathies (Duong Phu et al., 2021, Devlin et al., 2023, Bondué et al., 2022).

Gene replacement therapies based using AAV viral vectors have an increased efficiency over non-viral vectors, however one disadvantage of the use of AAV viral vectors for gene replacement is their limiting packing capacity to transport genetic material (~5 kb of genetic material) (French et al., 2020, Ivanchenko et al., 2023).

One of the advantages of CRISPR/Cas9 technology is the possibility of performing gene knockouts and incorporating a wide range of different mutations into ciliopathy genes (Pusapati et al., 2018, Rusterholz et al., 2022). However, the efficiency of the incorporation of the specific mutations or gene knockouts is limited. Once the CRISPR/Cas9 gene editing is performed, it is necessary to ensure that the treated cells or tissue (e.g. from mouse or zebrafish) has incorporated the corresponding mutant genotype (Tsai et al., 2019, Jacinto et al., 2020, Liu et al., 2017). Primary cells, such as hURECs or primary fibroblasts, may not maintain their growth for multiple passages, limiting the possibilities of subculturing the primary cells in which the mutant genotype has been confirmed for further functional analyses. Furthermore, CRISPR/Cas9 technology requires certain amount of time for training and optimisation that I could not have during this project.

Considering the whole range of available technologies mentioned above as potential therapeutic technologies for ciliopathies, I have focused on the ASOs-mediated exon

skipping approach because of its high efficiency, rapid application and the possibility of implementation into a wide range of cell lines and primary cells (Cideciyan et al., 2023, Schellens et al., 2023, Barny et al., 2019, Dulla et al., 2021, Rusterholz et al., 2022).

The Genomics England 100,000 Genomes Project allows registered users to search disease-causing variants in a large database, find novel patients presenting variants in a particular gene, such as *CEP120* and then assess the corresponding clinical phenotype of the patients found (genotype-to-phenotype model). It also allows to search for reported variants and disease-causing variants in candidate genes associated with a specific disease, such as JBTS (phenotype-to-genotype model) (Smedley et al., 2021, Best et al., 2022b, Best et al., 2021).

In the last decade, there have been several approaches to find candidate ciliopathy genes based on: analysis of protein-protein interactors (Higgins et al., 2022, Latour et al., 2020), analysis of exome sequencing data combined with homozygosity mapping (Alkanderi et al., 2018, Al Alawi et al., 2021), siRNA-based screening of genes affecting ciliogenesis (Wheway et al., 2015), CRISPR-based screening of genes affecting Hedgehog signaling (Breslow et al., 2018, Pusapati et al., 2018).

The analysis of whole genome sequencing available at the Genomics England 100,000 Genomes Project allows to find and validate candidate genes for primary ciliopathies (Best et al., 2021, Blakes et al., 2022, Latour et al., 2020, Powell et al., 2021, Senum et al., 2022). A robust candidate gene leading to ciliopathy phenotypes is usually characterised by encoding a protein functioning at the cilia or basal body and/or being included in the SYSCILIA Gold Standard gene list (Van Dam et al., 2013).

When putative disease-causing variants are described in patients with ciliopathy phenotypes, these variants need to be pathogenic. Moreover, the allele frequency of the variants needs to be rare (generally lower than 0.01 in allele frequency) and the variant/s described should segregate with the disease according with the mode of inheritance of the corresponding disease (if there is sequencing information of the parents or other relatives) (Richards et al., 2015, Morbidoni et al., 2021, Shaheen et al., 2016, Shamseldin et al., 2020). Previously described genes with similar functions and functioning in the same pathway of the candidate genes can support the analysis of candidate genes (Breslow et al., 2018, Pusapati et al., 2018). Other *in silico* analyses

may involve assessing the tolerance to loss-of-function variants in that particular candidate ciliopathy gene, using gnomAD population database (Karczewski et al., 2020).

As an alternative and complement to defining genotype-phenotype correlations by finding additional patients with mutations in a particular gene, the use *ex vivo* cell-based assays has proven to be appropriate to assess the question: “why mutations in the same gene can cause different phenotypes?”. *Ex vivo* cell-based assays have been widely used to investigate the pathogenic effect of patient-specific mutations and the molecular and cellular disease mechanisms associated with a particular gene (Ramsbottom et al., 2018, Srivastava et al., 2017b, Malicdan et al., 2015).

Ex vivo analyses of candidate genes can involve the use of control and patient cells to assess the cellular location of the protein encoded by the candidate gene and to assess of the proportion of ciliated cells and any defect in primary cilia (Shamseldin et al., 2020, Estrada Mallarino et al., 2020).

One example of a recently described gene associated with JBTS is *TOGARAM1* (Latour et al., 2020). *ARMC9*, which is a gene associated with JBTS (Van De Weghe et al., 2017), encodes a protein that interacts with *TOGARAM1* (Latour et al., 2020). The interaction of *ARMC9* and *TOGARAM1* forming an *ARMC9*-*TOGARAM1* complex was found using protein-protein interaction screens (tandem affinity purification-tagged and Yeast two hybrid analyses), suggesting *TOGARAM1* as novel a candidate JBTS gene (Latour et al., 2020). To further investigate and confirm *TOGARAM1* as a JBTS gene, in this study 5 families were described with disease-causing mutations in *TOGARAM1* leading to JBTS. Using patient fibroblasts, hTERT-RPE1 cell lines and a zebrafish model (genetically modified using CRISPR/Cas9 system) it was found that disruption of *TOGARAM1* affects tubulin posttranslational modifications of ciliary microtubules, ciliary length and ciliary stability and leads to ciliopathy phenotypes (Latour et al., 2020).

CEP120 is a centrosomal protein involved in centriole biogenesis (Mahjoub et al., 2010, Tsai et al., 2019), like other centrosomal proteins associated with ciliopathies, such as *KIAA0586* (Bashford and Subramanian, 2019) *CEP104*, *CSPP1* (Frikstad et al., 2019) or *DZIP1* (Wang et al., 2013), it affects Hedgehog signalling (Wu et al., 2014, Betleja et al., 2018). Previous work has shown that in *Cep120*-depleted mouse embryonic fibroblasts

(MEFs), treated with SAG and *Cep120* siRNA, the amount of Smoothened (Smo) per unit length cilium was significantly reduced compared with the control MEFs (treated with SAG and a control siRNA) (Betleja et al., 2018). Furthermore, in a mouse conditional knockout of *Cep120* for the central nervous system (CNS), it was shown a lack of response to Hedgehog signaling in the cerebellum of a postnatal day 7 (P7) (Wu et al., 2014).

In recent years, the effect of some of *CEP120* patient-specific variants has been investigated using different functional cell-based assays. RPE-1 human retinal pigmented epithelial cells were genome-engineered via CRISPR associated protein (Cas9) (CRISPR/Cas9) targeting system so that they were either homozygous for the variant c.595G>C; p.(Ala199Pro) (from now A199P) or for c.581T>C; p.(Val194Ala) (from now V194A). It was found that both variants caused a reduction on: the levels of *CEP120* protein, the recruitment of distal centriole markers (CEP164 and KIAA0586) and ciliogenesis (Joseph et al., 2018). In another study, using RPE-1 cells in which *CEP120* was knocked out via CRISPR/Cas9, it was suggested that the variants c.2177T>C; p.(Leu726Pro) (from now L726P) and c.2924T>G; p.(Ile975Ser) (from now: I975S) impaired ciliogenesis (as the transfection of constructs with the *CEP120* sequence with any of these mutations could not restore the cilia-null phenotype observed in the *CEP120* knocked out cells) (Tsai et al., 2019).

Similarly, using HEK-293T cells, it was also suggested that I975S impairs the *CEP120* interaction with C2CD3 and recruitment of C2CD3 to centrioles and that L726P and c.2134C>T; p.(Leu712Phe) (from now L712F) inhibit the recruitment of KIAA0586 to centrioles (Tsai et al., 2019). In cultured skin fibroblasts obtained from a JATD patient homozygous for the A199P *CEP120* variant, it was observed a reduced frequency of ciliated cells and an increased number of cells with more than one centrosome (Shaheen et al., 2015). Additionally, mouse oocytes have been used to study another variant: c.2840G>A; p.(Arg947His). This *CEP120* variant, which has not been found leading to ciliopathies, has been associated with an increase in the incidence of aneuploidy and a reduction in spindle microtubule nucleation efficiency (Tyc et al., 2020).

Sonic hedgehog Shh-LIGHT2 cells is a cell line derived from mouse NIH-3T3 cells stably containing a Gli-dependent firefly luciferase and constitutive Renilla luciferase vectors. This is a Hedgehog (Hh)-responsive cell line used to assess effects on Hh protein, which can be quantified via a quantitative assay using a luminometer (Taipale et al., 2000, Chen et al., 2002). In recent years, with the development of the CRISPR/Cas9 system, this and similar cell lines have been used to further study the role of Hh signalling on ciliopathies (Pusapati et al., 2018, Breslow et al., 2018).

There are multiple studies using *in vitro* cell models to complement genetic studies, investigating the pathogenesis of ciliopathy genes and disease-causing mutations (Oud et al., 2018, Molinari et al., 2020, Latour et al., 2020, Garanto et al., 2016). Validating *in vitro* the deleterious effect of patient-specific disease-causing variants contributes to get insights into the disease pathogenesis and genotype-phenotype correlations and gene pleiotropy of multiple ciliopathy genes (Molinari et al., 2018a, Arjona et al., 2019, Barny et al., 2018).

There is a wide array of *in vitro* functional assays to investigate putative disease-causing variants: from rescue experiments using CRISPR/Cas9 technology in RPE-1 and HEK-293T cell lines, with a focus on defects in cilia biogenesis (Tsai et al., 2019), to minigene splicing assays using a HEK-293T cell line (Tamayo et al., 2023). Furthermore, there are approaches investigating the molecular disease mechanisms of patient-specific disease-causing variants using human primary cells, such as hURECs (Molinari et al., 2020) and primary skin-derived fibroblasts (Shaheen et al., 2015).

The limitation of the use of human primary cells is that they can only be cultured up to a very limited number of cell passages, unlike immortalised cell lines. Besides, the use of primary cells from specific patients of interest is dependent on the availability of the patient to provide the corresponding sample to culture primary cells.

It can be suggested that this limitation could be addressed by using primary cells of healthy controls. A gene knockdown experiment for a particular gene of interest can be performed in human primary cells and then be rescued by transfecting a plasmid containing the WT gene sequence (or a mutated gene sequence) of that particular gene of interest. This can potentially allow to observe differences between the rescue effect caused by the transfected plasmid containing specific patient-specific disease-causing

mutations, compared with the rescue effect of a plasmid with the WT gene sequence. Considering this, a cell-based pipeline to study disease-causing *CEP120* mutations was designed (shown in Figure 2.9 of section: 2.14. Plasmid transfection efficiency into mammalian cells).

This pipeline aims to effectively perform a rescue experiment in human primary cells. The rescue experiment consists in a gene knockdown followed by a plasmid transfection, in human primary cells of healthy controls. This approach could potentially have the benefits of a reproducible standardised assay that can be easily modified and modelled for each patient-specific disease-causing mutation of interest, as well as having the advantage of using human non-transformed primary cells, instead of using an immortalised cell line. This rescue experiment approach using human non-transformed primary cells (such as skin-derived primary fibroblasts) and plasmid transfection, can be a more time-efficient approach than using CRISPR/Cas9 engineered primary cells, which require large amounts of optimisation time, workload and involve the risk of cells not surviving the required cell passages to isolate the CRISPR/Cas9 engineered primary cells with the expected genotype.

There are multiple aspects that can be affected when *CEP120/Cep120* is mutated: from Hh signalling pathway (Betleja et al., 2018, Wu et al., 2014) to centrosome and cilia biogenesis (Joseph et al., 2018, Mahjoub et al., 2010). The Hh signalling pathway has not been investigated for *CEP120* patient-specific variant. Moreover, other *CEP120* exonic variants such as: R151X, S380YfsX and A549V have not been studied.

This is why I want to study patient-specific *CEP120* mutations and to do this pipeline was designed consisting in cell-based assays including Shh-LIGH2 cells and primary human fibroblasts. This rescue experiment in human primary cells, can be used to model each of the reported *CEP120* patient-specific disease-causing mutation, and to study Hh signalling in primary cells via gene expression analysis, and cilia biogenesis via immunofluorescence staining.

Shh-LIGHT2 cell line can be used to confirm the effect on Hh signalling upon siRNA *Cep120* knockdown in Shh-LIGHT2 cells and as a preliminary step to optimise and validate the protocols that will be required to complete the designed pipeline in human primary cells.

Here the aim is to investigate genotype-phenotype correlations in two JBTS genes: *CEP120* and *CC2D2A* as exemplars of genetic and phenotypic heterogeneity. The aim is to find novel patients presenting disease-causing variants in JBTS genes, in particular *CEP120*, and the effect of *CEP120/Cep120* knockdown using *in silico* analyses and cell-based assays using mouse Shh-LIGHT2 cells and primary human fibroblasts.

4.2. Results

4.2.1 *In silico* investigation of genotype-phenotype correlations in *CEP120* and *CC2D2A*

In order to study genotype-phenotype correlations in *CEP120* and *CC2D2A*, I first assessed all the *CEP120* and *CC2D2A* variants reported in the literature and their associated phenotypes. The results explained in this section (4.2.1) and the next two sections (4.2.2 and 4.2.3) have been published as an article in the journal Molecular Genetics and Genomic Medicine (Barroso-Gil et al., 2021a).

There are only 9 patients reported in the literature presenting with biallelic (homozygous or compound heterozygous) *CEP120* pathogenic variants (Roosing et al., 2016, Shaheen et al., 2015) (Figure 4.1A).

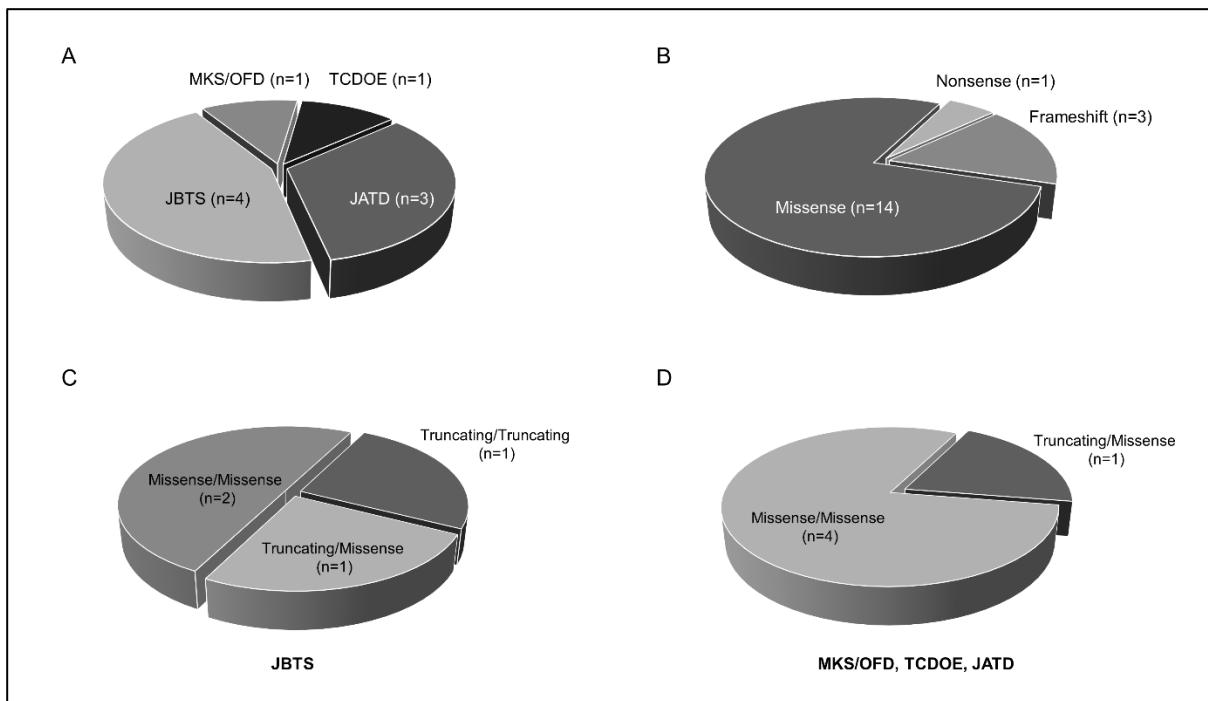


Figure 4.1. Phenotypes and genotypes in patients carrying biallelic disease-causing *CEP120* variants reported in the literature. A: Pie chart indicating how the phenotypes associated with *CEP120* biallelic variants are distributed. n indicates total number of patients. B: Pie chart indicating how the *CEP120* alleles are distributed according to their annotated consequence. n indicates number of alleles. C: Pie chart indicating how the *CEP120* alleles found in JBTS patients are distributed according to their consequence. D: Pie chart indicating how the *CEP120* alleles found in patients with MKS/OFD, TCDOE or JATD are distributed according to their consequence. JBTS, Joubert syndrome; JATD, Jeune asphyxiating thoracic dystrophy; MKS, Meckel syndrome; OFD, oral-facial-digital syndrome; TCDOE, tectocerebellar dysraphia with occipital encephalocele.

Considering these nine *CEP120*-affected patients, there are 14 missense alleles, 3 frameshift alleles and 1 nonsense allele (Figure 4.1B). Of note, 3 of these patients present the *CEP120* variant A199P in homozygosity and 1 patient presented it in compound heterozygosity with another *CEP120* variant (representing 7/14 of the missense alleles reported in the literature) (Roosing et al., 2016, Shaheen et al., 2015).

Of note, one of the JBTS patients reported (Patient MTI-143), presented compound heterozygous *CEP120* missense variants: L712F and L726P (Roosing et al., 2016), one of them: L712F, is also found in 2 homozygous individuals in the gnomAD database, and has a total allelic frequency of ~0.004. In recent years, *in vitro* experiments investigating the pathogenicity of some *CEP120* variants, revealed that this variant: L712F impairs the recruitment of two centrosomal proteins: Talpid3 (Tsai et al., 2019)

and KIAA0753 (Chang et al., 2021) to centrioles, and it delays the exit of granule neuron progenitors (GNPs) from the germinal zone (Chang et al., 2021).

I did not find any correlation between the severity of the *CEP120* genotype and the severity of the associated clinical phenotypes (Figures 4.1C and 4.1D). In other words, it is not clear if the patients with biallelic truncating *CEP120* variants present more severe phenotypes than the patients with biallelic missense *CEP120* variants. However, it is important to note that there are only 9 patients described with disease-causing *CEP120* mutations. This low number of *CEP120* patients makes the establishment of any genotype-phenotype correlation difficult.

In contrast, a higher number of patients has been reported for *CC2D2A*. There are 111 patients from 97 families described with disease-causing *CC2D2A* mutations. (Figure 4.2A).

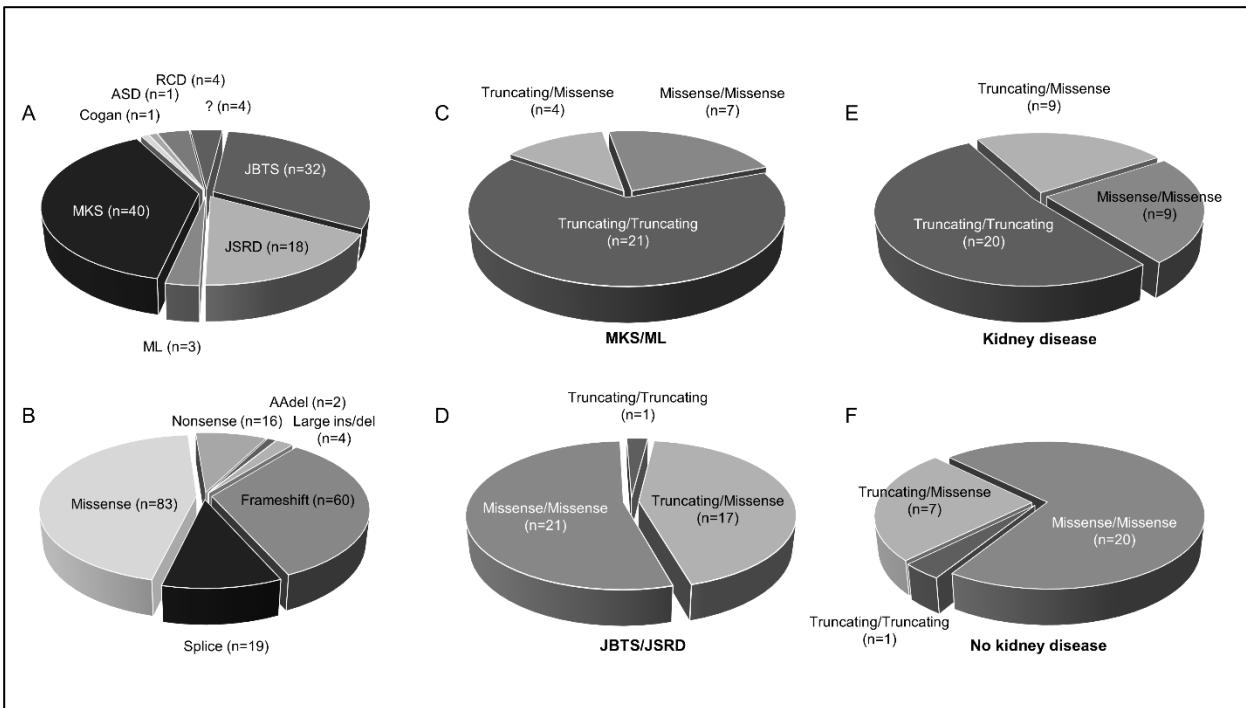


Figure 4.2. Phenotypes and genotypes in patients carrying biallelic disease-causing *CC2D2A* variants reported in the literature. A: Pie chart indicating how the phenotypes associated with *CC2D2A* biallelic variants are distributed. n indicates total number of patients. B: Pie chart indicating how the *CC2D2A* alleles are distributed according to their annotated consequence. n indicates number of alleles. Of note, in one family (385/447 II-4) three different variants were reported with two different compound heterozygote combinations (Srour et al., 2012). C: Pie chart indicating how the *CC2D2A* alleles found in patients with MKS or Meckel-like syndrome are distributed according to their consequence. D: Pie chart indicating how the *CC2D2A* alleles found in JBTS patients are distributed according to their consequence. E: Pie chart indicating how the *CC2D2A* alleles found in patients with kidney disease are distributed according to their consequence. F: Pie chart indicating how the *CC2D2A* alleles found in patients without kidney disease are distributed according to their consequence. The term truncating variants includes nonsense or frameshift variants. AAAdel, single amino acid deletion; ASD, autism spectrum disorder; Cogan, Cogan-type congenital oculomotor apraxia; JBTS, Joubert syndrome; Large ins/del, large insertions/deletion including retrotransposon insertion; MKS, Meckel syndrome; ML, Meckel-like syndrome; RCD, rod cone dystrophy; ?, not unequivocally described.

Considering all the disease-causing *CC2D2A* variants reported in the 111 patients from 97 families, there is a total of 195 pathogenic alleles. Considering the consequence of these 195 alleles: 90 alleles are missense, 60 alleles are frameshift and 20 alleles are alleles affecting splicing, 18 alleles are nonsense and 3 alleles are single amino acid (in frame) deletions and 4 alleles are large insertions/deletions, including one retrotransposon insertion reported in one case (Figure 4.2B).

Of note, there are two *CC2D2A* variants that have also been found in homozygosity in the gnomAD database of genetic variants in healthy individuals. The *CC2D2A* variant: p.(Glu229del), which is found in compound heterozygosity (with another *CC2D2A* variant: p.(Val1298Asp)) in a MKS patient (Otto et al., 2011), was found in 538 homozygous individuals, and it has an allele frequency of 0.062 in gnomAD v.2.1.1. It has an allele frequency of 0.056 (and 311 homozygotes), if gnomAD v.3.1.2 is considered. The *CC2D2A* variant: p.(Pro721Ser) was reported in homozygosity in one JBTS patient (Mougou-Zerelli et al., 2009), moreover it is found in 3 homozygous individuals and it has an allele frequency of 0.002 in the gnomAD v.2.1.1 database. It has an allele frequency of 0.0011 (and 0 homozygotes) if gnomAD v.3.1.2 is considered. It can be suggested that these two *CC2D2A* alleles, especially the p.(Glu229del) allele, can be hypomorphic alleles rather than fully pathogenic alleles (Bachmann-Gagescu et al., 2012).

With this information I wanted to find if truncating variants (loss-of-function variants described as nonsense or frameshift) are associated with more severe clinical phenotypes than non-truncating variants (e.g. missense variants), as it was suggested previously (Mougou-Zerelli et al., 2009). To do this disease-causing *CC2D2A* alleles in the 111 patients reported in the literature were analysed systematically, as well as their associated phenotypes.

It is observed that 66% (21/32) of the patients with MKS or ML carrying biallelic *CC2D2A* mutations, had a genotype (homozygous or compound heterozygous) consisting of truncating *CC2D2A* alleles exclusively. In contrast, only 2% (1/45) of the patients with JBTS carried a genotype consisting of truncating alleles (Fisher's exact test: $p < 0.0001$).

Similarly, it was observed that only 22% (7/32) of the patients with MKS or ML carrying biallelic *CC2D2A* mutations, had a genotype consisting of missense *CC2D2A* alleles. In contrast, 58% (26/45) of the patients with JBTS carried a genotype consisting of missense alleles (Fisher's exact test: $p = 0.0023$) (Figures 4.2C and 4.2D).

This systematic analysis considering the reported patients with disease-causing mutations in *CC2D2A* shows that there is a correlation between the severity of the type (consequence) of each of the *CC2D2A* mutations and the severity of the resulting clinical phenotypes. Moreover, the biallelic *CC2D2A* variants are more likely to be found

in patients with a phenotype including kidney disease. It was found *CC2D2A* affected patients were analysed with a focus on kidney disease. It was found that 50% (20/40) of the patients that have a clinical phenotype that included kidney disease, carried biallelic *CC2D2A* truncating variants. In contrast, only 3% of the patients without a clinical phenotype that included kidney disease carried biallelic *CC2D2A* truncating variants (3%, 1/34) (Fisher's exact test: $p < 0.0001$). This correlates with the concept that missense *CC2D2A* variants are more likely to be found in patients with JBTS without extra-CNS clinical phenotypes (Figures 4.2E and 4.2F).

There are some *CC2D2A* variants that are found in patients with JBTS as well patients with more severe clinical phenotypes such as MKS/ML. This is the case, for example, of the variant: p.(Pro1122Ser), which was found homozygously in patients with JBTS (Gorden et al., 2008) and in a patient with MKS (Al-Hamed et al., 2016) also homozygously. This may indicate that additional modifying factors may have a role in the resulting *CC2D2A* associated phenotype.

4.2.2. In silico analysis of gene expression and tissue-specific basal exon skipping in *CEP120* and *CC2D2A*

I wanted to study *in silico*, the gene expression of two pleiotropic genes associated with JBTS and more severe phenotypes: *CEP120* and *CC2D2A*. For these two genes tissue-specific transcripts, exon skipping events and basal exon skipping were investigated.

With this information I aimed to investigate the applicability of a possible exon skipping therapy to rescue the deleterious effect of truncating variants in these two genes. Having *CEP290* pleiotropic gene as a model example, previous studies have demonstrated that basal exon skipping as well as the targeted skipping of certain exons can be tolerated and resulted in a functional or partially functional protein (Ramsbottom et al., 2018, Drivas et al., 2015).

CEP120 expression has been found in multiple human and mouse tissues (Powell et al., 2020, Xie et al., 2007). According to the Human Protein Atlas (Uhlén et al., 2015) online resource there is RNA expression of *CEP120* and *CC2D2A* found ubiquitously in human tissues, including brain.

Both *CEP120* and *CC2D2A* are associated with JBTS phenotypes (Roosing et al., 2016, Bachmann-Gagescu et al., 2012). Two of the organs that are often affected in JBTS and other primary ciliopathies are the cerebellum and kidney (Badano et al., 2006b, Braun and Hildebrandt, 2017).

In silico tissue-specific expression and splicing of *CEP120* and *CC2D2A* were investigated in these two organs: kidney and cerebellum. To do this *in silico* analysis I used the online resource: Genotype-Tissue Expression (GTEx) project (The GTEx Consortium, 2020), which consists in RNA sequencing data available for different human organs and tissues.

CEP120 has three main coding transcripts with the highest expression levels in cerebellum and/or kidney medulla according to the GTEx Project:

ENST00000306481.10 (“Transcript 1”), ENST00000328236.9 (“Transcript 2”) and ENST00000306467.9 (“Transcript 3”). *CEP120* Transcript 2 and *CEP120* Transcript 3 have the same Consensus Coding Sequence (CCDS) (CCDS4134.2) encoding for the longest *CEP120* protein isoform (986 aa). The protein encoded by *CEP120* Transcript 1 (CCDS54890.1) lacks the first 26 amino acids compared to transcripts 2 and 3. The *CEP120* Transcript 1 is expressed in kidney medulla and cerebellar hemisphere. Transcripts 2 and 3 are expressed in cerebellum. However, Transcripts 2 and 3 are nearly absent in kidney tissues, unlike Transcript 1 (Figure 4.3A).

Alternative splicing events are responsible to generate these transcript isoforms. These events of alternative splicing occur at the pre-mRNA 5-end. The exon 2 (having the canonical transcript ENST00000328236.9, also known as NM_153223.3, as a reference) seems to be skipped in kidney, as *CEP120* transcript 1 (*CEP120* transcript: ENST00000306481.10, Figure 4.3B) is the predominant isoform in kidney medulla (Figure 4.3A), resulting in a shorter protein (Figure 4.3C).

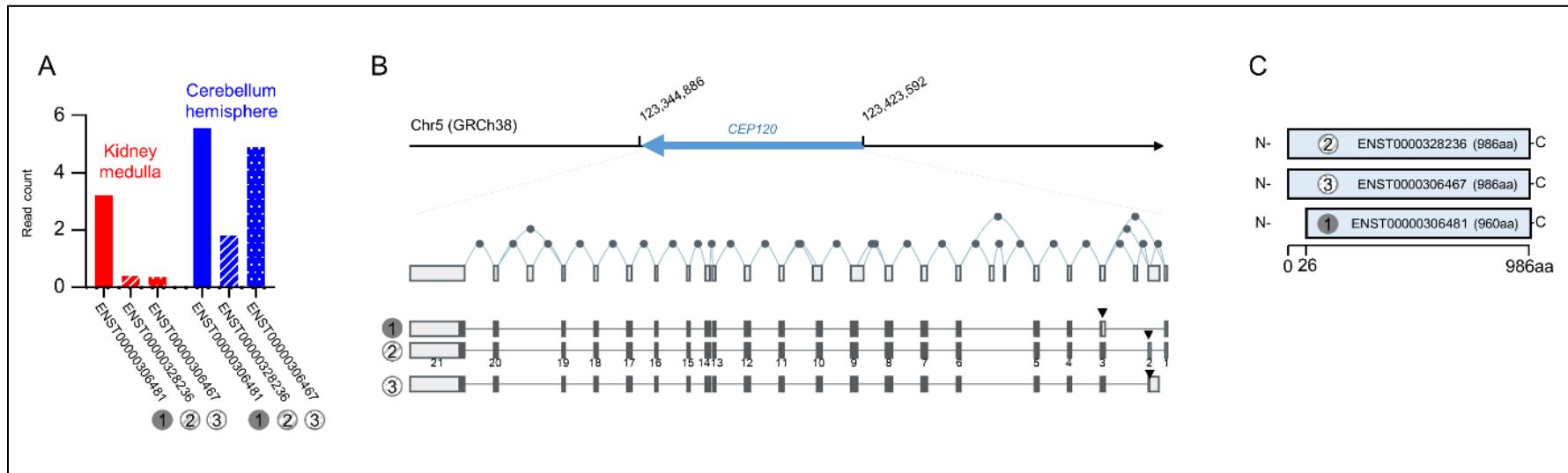


Figure 4.3. Tissue specific transcript expression of *CEP120* and its exon structure. A: RNA expression of the transcript isoforms (protein coding) with the highest expression levels in kidney medulla (red) and cerebellar hemisphere (blue) according to the RNA sequencing data from the Genotype-Tissue Expression (GTEx) Project. B: *CEP120* genomic localization and the different exons of *CEP120* (having transcript ENST00000328236.9 as a reference). Exons and imputed splice junctions corresponding to the three transcripts with the highest expression levels in kidney medulla and cerebellar hemisphere are shown. The open reading frame is shown in dark grey with the start codon marked with an arrowhead. C: Predicted protein length of the protein products encoded by the three *CEP120* transcripts considered. Data shown in this figure were obtained from GTEx Project (The GTEx Consortium, 2020).

CC2D2A has three main coding transcripts with the highest expression levels in cerebellum and/or kidney medulla according to GTEx Project: ENST00000515124.5 (“Transcript 1”), ENST00000503292.5 (“Transcript 2”) and ENST00000389652.9 (“Transcript 3”). Transcript 1 is expressed in kidney medulla and cerebellar hemisphere. Transcript 2 is expressed in kidney medulla and Transcript 3 is expressed in cerebellar hemisphere (Figure 4.4A).

CC2D2A Transcript 1 (consisting in 1,474 bp), does not encode for the functional domains of *CC2D2A* and has an additional exon (that leads to a premature stop codon) after exon 5 (having the longest coding transcript: Transcript 2, ENST00000503292.5, also known as NM_001080522.2, as a reference). Transcript 1 encodes for a 111 amino acid product which shares the first 82 amino acids transcript 2 (Figure 4.4C).

Transcript 1 is the predominant isoform in kidney medulla and its exons’ structure is supported by GTEx splice junction data specific showing splicing events enriched in kidney (Figure 4.4B). Of note, the open reading frame of Transcript 3 is incomplete, as its 5'-end is not fully annotated (ie: the first coding exons of the *CC2D2A* coding sequence are not annotated in GTEx) (Figure 4.4B).

It is interesting to observe a splice junction, which leads to the skipping of exon 30, particularly enriched in kidney medulla (Figure 4.4B).

These results from the RNAseq data provided by GTEx project suggest that there are *CEP120* and *CC2D2A* tissue-specific transcripts, considering kidney and cerebellum tissues.

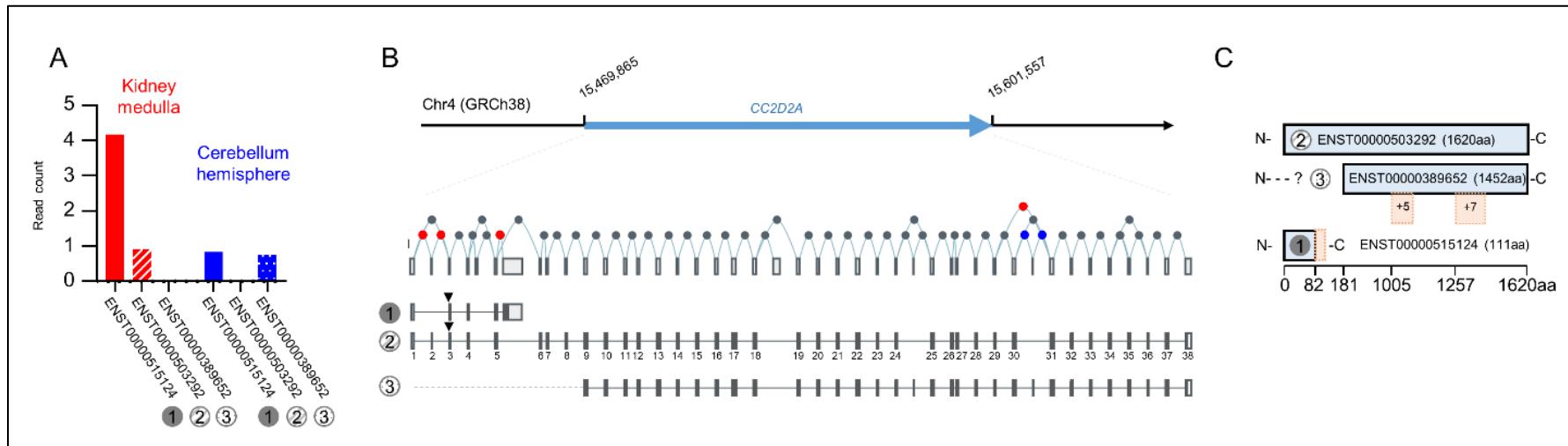


Figure 4.4. Tissue specific transcript expression of *CC2D2A* and its exon structure. A: RNA expression of the transcript isoforms (protein coding) with the highest expression levels in kidney medulla (red) and cerebellar hemisphere (blue) according to the RNA sequencing data from the Genotype-Tissue Expression (GTEx) Project. B: *CC2D2A* genomic localization and the different exons of *CC2D2A* (having transcript ENST00000503292.5 as a reference). Exons and imputed splice junctions corresponding to the three transcripts with the highest expression levels in kidney medulla and cerebellar hemisphere are shown. The open reading frame is shown in dark grey with the start codon marked with an arrowhead. The splice junctions that are enriched in the kidney medulla (compared to cerebellum) are shown in red and the splice junctions that are enriched in the cerebellum (compared to kidney) are shown in blue. C: Predicted protein length of the protein products encoded by the three *CC2D2A* transcripts considered. Data shown in this figure were obtained from GTEx Project (The GTEx Consortium, 2020).

In vitro tissue-specific basal exon skipping occurring in these two genes was confirmed. I also studied if these tissue-specific basal exon skipping have a relationship with the described disease phenotypes of described in the *CEP120*- and *CC2D2A*- affected patients.

Regarding *CEP120* tissue specific basal exon skipping it is important to first mention one *CEP120*-affected patient (patient MTI-991) described in the literature (Roosing et al., 2016). This patient was diagnosed with JBTS and carries homozygously a *CEP120* disease-causing variant: *CEP120* (NM_153223.3): c.49+5_49+10del ; p.Gly+1AspfsTer14 (?), which locates in an exon-intron boundary. This variant affects the 3' boundary of exon 2 (located in intron 2 and exon 2) and leads an intronic retention, and subsequently a premature stop codon. (Roosing et al., 2016). The *in silico* analysis performed using splice junction data from GTEx project, indicates some skipping of exon 2 of *CEP120* in the kidney (Figure 4.3).

This GTEx project prediction was confirmed *in vitro* by during RT-PCR to detect the skipping of this exon. Basal exon skipping of exon 2 of *CEP120* occurring in human kidney tissue was confirmed (Figure 4.5).

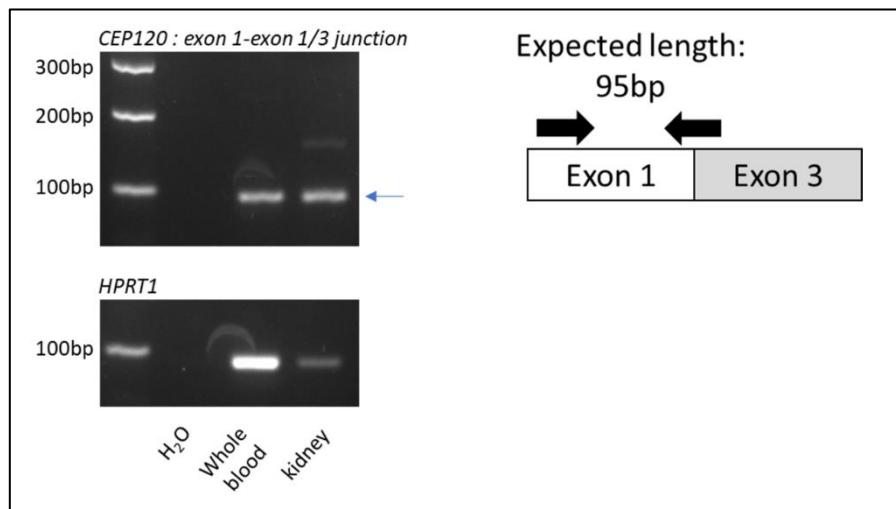


Figure 4.5. RT-PCR showing basal exon skipping of *CEP120* exon 2 in kidney and whole blood. To perform the RT-PCR RNA isolated from human kidney and whole blood was used. Primers (arrows) were designed to amplify a 95 bp RT-PCR product only if the event of the splicing of exon 2 of *CEP120* (creating an exon 1-exon 3 junction) occurs. This RT-PCR product is detected by gel electrophoresis in which a band of the expected size (~100 bp) was observed (arrows) using RNA samples from human kidney and whole blood. RT-PCR and Gel electrophoresis were performed by Dr Eric Olinger (Newcastle University, UK).

This basal exon skipping finding suggest that if exon 2 is skipped in the kidney, a disease-causing variant located in exon 2 or the 3' of exon 2 (exon 2-intron 2) boundary would be bypassed and may be silent and thus not leading to the corresponding disease phenotype. To support this hypothesis, the MTI-991 JBTS patient, presenting a homozygous *CEP120* variant located in the 3' boundary of exon 2, does not have clinical phenotypes affecting the kidney (Roosing et al., 2016).

Regarding *CC2D2A* tissue specific basal exon skipping I focussed on exon 30. The *in silico* analysis performed using splice junction data from GTEx project indicates some skipping of exon 30 of *CC2D2A* in the kidney (Figure 4.4).

This GTEx project prediction was confirmed *in vitro* by during RT-PCR to detect the skipping of this exon. Basal exon skipping of exon 30 of *CC2D2A* (using *CC2D2A* transcript: ENST00000503292.5 as a reference) was confirmed.

Primers were designed to detect a RT-PCR product only if the event of the splicing of exon 30 of *CC2D2A* (creating an exon 29-exon 31 junction) occurs.

RT-PCR was performed using total RNA from whole blood, kidney and human urine-derived renal epithelial cells (hURECs) (Figure 4.6). The predicted product (327 bp) if there is not skipping is detected and kidney and whole blood via gel electrophoresis. Besides a shorter transcript is detected corresponding to the expected size of RT-PCR product in the event the basal skipping of exon 30 (which consists in 177 bp). As this shorter RT-PCR product (~150 bp) is detected in kidney this suggest that tissue specific basal skipping of exon 30 is occurring in the kidney, leading to a transcript lacking exon 30 (Figures 4.4B and 4.6).

Besides, basal skipping of exon 30 was detected using RNA extracted from hURECs. This supports the event of basal skipping of exon 30 in kidney and the utility of the hURECs as a non-invasive source of primary cells and “liquid biopsy” that can be used to study kidney specific splicing events (Molinari et al., 2018a).

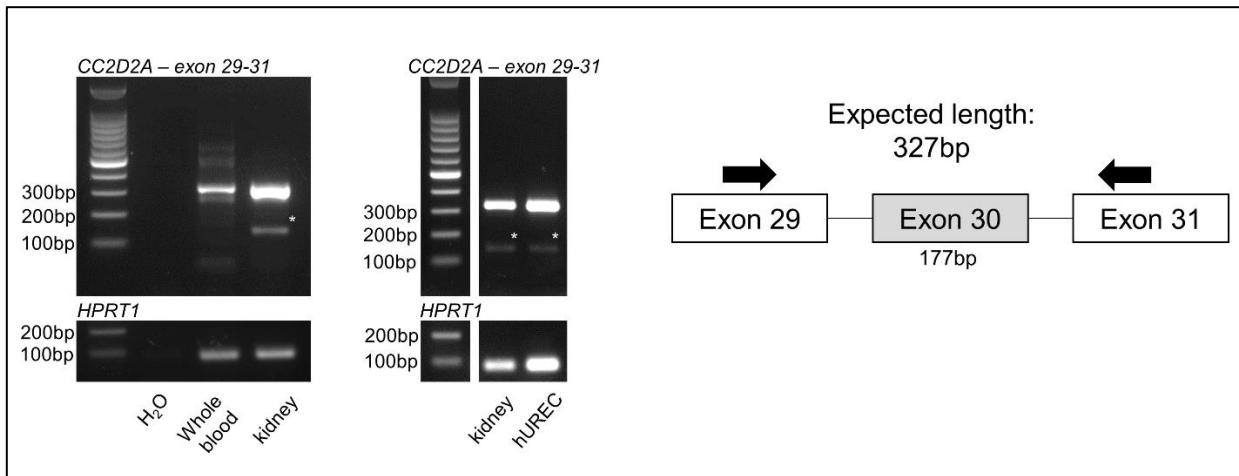


Figure 4.6. RT-PCR showing basal exon skipping of *CC2D2A* exon 2 in human kidney and human urine-derived renal epithelial cells (hURECs). To perform the RT-PCR I used RNA isolated from human kidney, hURECs and whole blood. Primers (arrows) were designed to amplify a 327 bp RT-PCR product if the exon 30 is not skipped. Alternatively, the RT-PCR is 150 bp (327-177) if the event of the splicing of exon 30 of *CC2D2A* (creating an exon 29-exon 31 junction) occurs. This RT-PCR product is detected by gel electrophoresis in which a band of the expected size (~150 bp) was observed (*asterisk) using RNA samples from human kidney and hURECs. RT-PCR and Gel electrophoresis were performed by Dr Eric Olinger (Newcastle University, UK).

In the previous section (section: 4.2.1) it was observed that *CC2D2A* truncating variants are associated with more severe phenotypes and biallelic truncating *CC2D2A* variants are more likely to be found in patients with kidney phenotypes (Figure 4.2). It was also observed that there is some basal exon skipping of the exon 30 of *CC2D2A*, occurring specifically in the kidney (Figure 4.6). This is why I hypothesise that the deleterious effect of truncating variants in exon 30 can be partially rescued if basal exon skipping occurs in exon 30 in kidney. Subsequently, the clinical kidney phenotypes observed in these patients are expected to be milder, as the deleterious effect in the kidney (caused by truncating variants in exon 30) could be bypassed by basal exon skipping of exon 30.

I assessed this hypothesis *in silico* by examining the prevalence of kidney disease in the patients presenting truncating variants in the different exons of *CC2D2A* (Figure 4.7). It is important to note that none of the patients with *CC2D2A* truncating variants in exon 30 of *CC2D2A* have a kidney phenotype. However, the number of patients is very low (there are only 2 patients with *CC2D2A* truncating variants in exon 30) to establish a robust conclusion. However, this is relevant, considering that 33.3%-100% of the patients with *CC2D2A* truncating variants in other exons of *CC2D2A*, have a phenotype

that include kidney disease (Figure 4.7). In total, considering all the exons of *CC2D2A*, there are only 13 (13/48) patients presenting one or two truncating *CC2D2A* mutations that developed a clinical phenotype without kidney disease, including the two patients with *CC2D2A* truncating variants in exon 30.

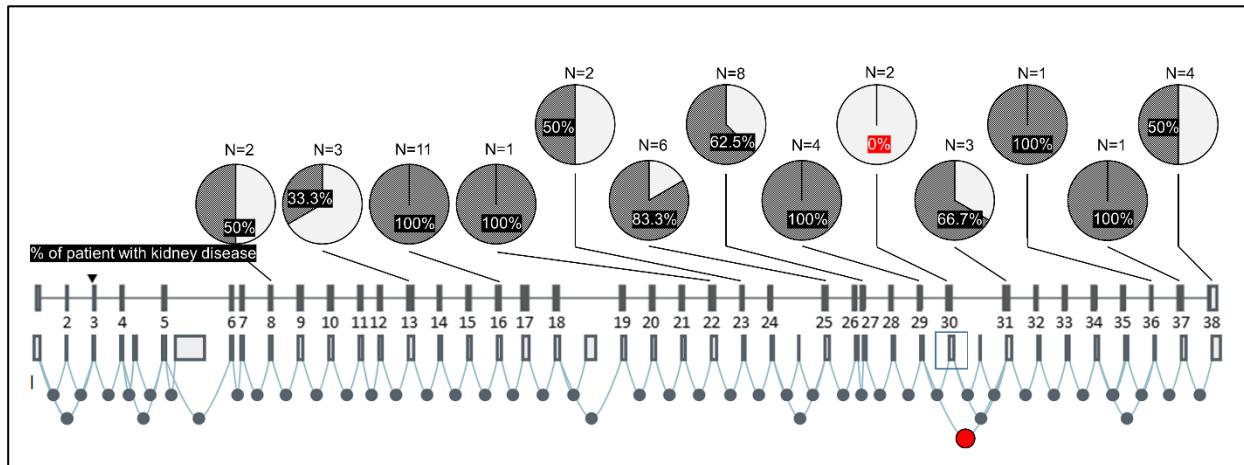


Figure 4.7. Distribution of exons and the prevalence of kidney disease in patients with truncating *CC2D2A* per exon. The transcript ENST00000503292.5 is used for the labelling of the exons. Exons and imputed splice junctions from the provided by GTEx project corresponding to transcript ENST00000503292.5 are shown. The splice junction that is predicted to lead to the basal exon skipping of exon 30, which is enriched in kidney (compared to cerebellum), is shown in red. n indicates the total number of patients (with and without kidney disease) with one or two truncating variants in one in the corresponding exons of *CC2D2A*. Of note, combining all the patients with one or two *CC2D2A* truncating variants: 73% (35/48) of them have a phenotype including kidney disease. Besides, the two patients with truncating *CC2D2A* variants in exon 30 are not described to have a phenotype that includes kidney disease.

4.2.3. *In silico* investigation of *CEP120* and *CC2D2A* as potential candidates for targeted exon skipping therapies

Exon skipping therapies have already been investigated for *CEP290*, a gene associated with JBTS (Garanto et al., 2016, Molinari et al., 2019, Ramsbottom et al., 2018), and for other genes associated with other inherited human diseases, some of these therapies are already in clinical use (Lee et al., 2018, Servais et al., 2015). I propose a multimodal approach to classify exons that can be skipped via exon skipping to bypass the deleterious effects of truncating variants located in these corresponding exons. This study is based on two genes associated with JBTS: *CEP120* and *CC2D2A*.

The first step to find potential candidates for an exon skipping therapy is to find the exons that are skippable (exons that can be skipped without disrupting the reading frame). *CEP120* (having transcript ENST00000328236.9 as a reference), has 20 coding exons. 11 of these coding exons are exons in which the total number of nucleotides is a multiple of 3 (Figures 4.8 and 4.9). Of note, it is considered that the first and last coding exons containing the corresponding start and stop codon respectively and/or exons encoding for protein domains of functional importance are not skippable, this is because if they are skipped it is expected to affect protein function. For *CEP120* analysis, it is also important to consider that only 2 of these 11 exons, do not encode for protein domains of functional importance (coiled-coil and C2 domains). These *CEP120* exons are: exons 14 and 15, which are predicted to maintain the protein reading frame, not causing loss of protein function, if skipped. Furthermore, as these two exons have exon-intron boundaries falling between codons (these exons have flanking introns in phase 0), the skipping of exon 14 and/or 15 would not lead to potential amino acid substitutions.

However, none of the disease-causing *CEP120* mutations reported up to date is located in exons 14 or 15. As a result, it can be suggested that *CEP120* is not a good candidate gene to design and apply targeted exon skipping therapies as there are not patients described that would benefit from them.

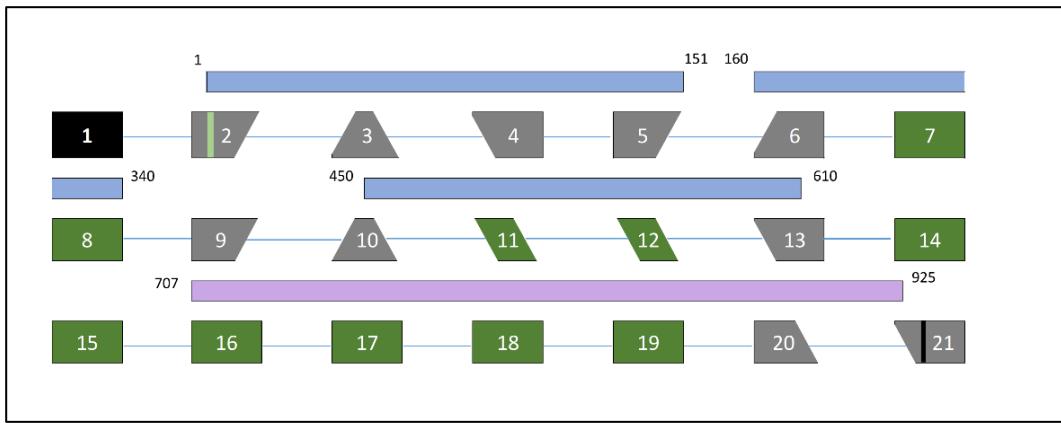


Figure 4.8. Schematic representation of the 21 exons corresponding to the *CEP120* human transcript NM_153223.3 (also called ENST00000328236.9). Exons 1 to 21 are numbered, with exon 1 non-coding shown in black. The start codon (ATG) is represented with a green bar in exon 2. The termination codon (TAA) is represented by a black line and the stop codon (TAA) in exon 21. Green rectangles represent exons with full complement of codons that may be skipped without change in reading frame. Grey trapezoids represent exons in which the exon-intron junction interrupts a codon. The three blue bars indicate the C2 domains (encoded by exons 2-8 and exons 10-13) and the purple bar indicates a coil-coiled domain (encoded by exons 16-21) with the amino acid numbers of each protein domain indicated. Protein domains have been annotated considering SMART online tool.

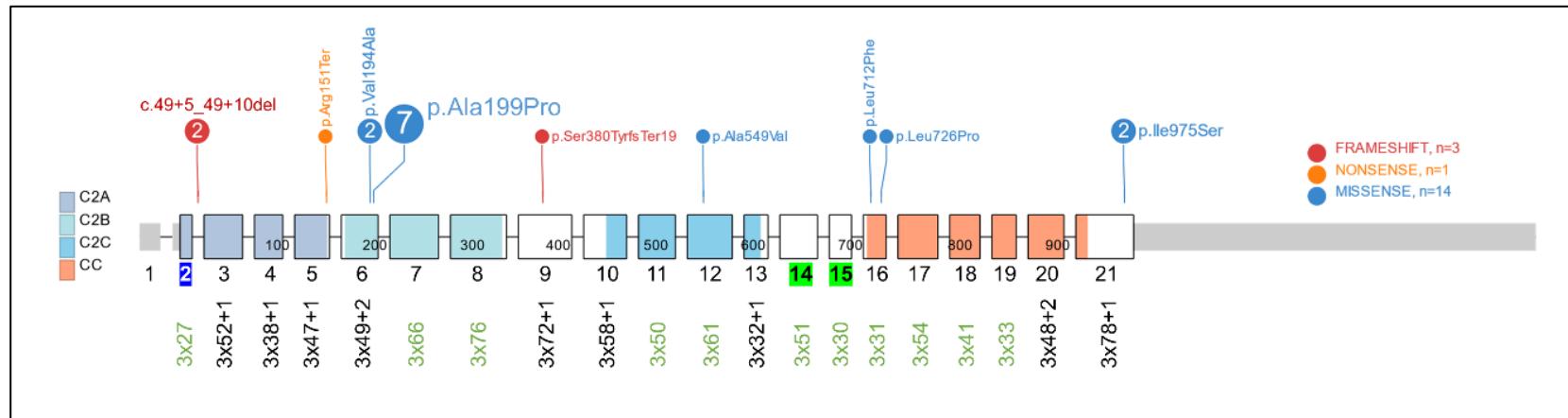


Figure 4.9. Schematic representation of the 21 exons corresponding to the *CEP120* human transcript NM_153223.3 (also called ENST00000328236.9) and the location of the reported *CEP120* mutations. Exons numbers are shown, below it, for each exon, it is shown the number of nucleotides corresponding to each exon (shown in multiples of three). For each exon, the number of nucleotides is coloured in green if the total number of nucleotides is an exact multiple of 3. *CEP120* protein domains: 1 coiled-coil domain (CC) and the 3 C2 domains are coloured with different colours. The *CEP120* variants reported in the literature their corresponding exon and accordingly with their allelic frequency in the patients reported in the literature (disc size reflecting allelic frequency). The variants are colour coded according to their annotated consequence. Exons numbers corresponding to exons that could be suggested as candidates for a targeted exon skipping therapy, based on nucleotide numbers: are shaded in green. Additionally, exons that could be suggested as candidates for a targeted exon skipping therapy, based tissue-specific transcript analysis are shaded in blue. Protein domains have been annotated considering SMART online tool. Diagram has been generated using ProteinPaint online tool.

CC2D2A (having transcript ENST00000503292.5 as a reference), has 36 coding exons. 17 of these coding exons are exons in which the total number of nucleotides is a multiple of 3 (Figures 4.10 and 4.11). Of note, it was considered that the first and last coding exons containing the corresponding start and stop codon respectively and/or exons encoding for protein domains of functional importance are not skippable because if they are skipped it is expected to affect protein function.

For *CC2D2A* analysis, it is also important to consider that 13 of these 36 exons, do not encode for protein domains of functional importance (coiled-coil and C2 domains) (Noor et al., 2008, Bachmann-Gagescu et al., 2012).

These *CC2D2A* exons are: exons 4, 7, 8, 9, 12, 13, 22, 23, 31, 32, 33, 34 and 35, which are predicted to maintain the protein reading frame, not causing loss of protein function, if skipped. Furthermore, as these 13 exons have exon-intron boundaries falling between codons (these exons have flanking introns in phase 0), the skipping of these exons would not lead to potential amino acid substitutions.

Additionally, exon 30 is suggested as skippable, which has a total number of nucleotides multiple of 3. Combining the information according to SMART online tool and the available literature provides, exon 30 would encode for the last 3 amino acids of the C2 domain of the *CC2D2A* protein. Moreover, further functional studies are needed to confirm if skipping 30 would affect the function of *CC2D2A* and if it can be a suitable target for an exon skipping therapy (Bachmann-Gagescu et al., 2012, Noor et al., 2008, Srour et al., 2012).

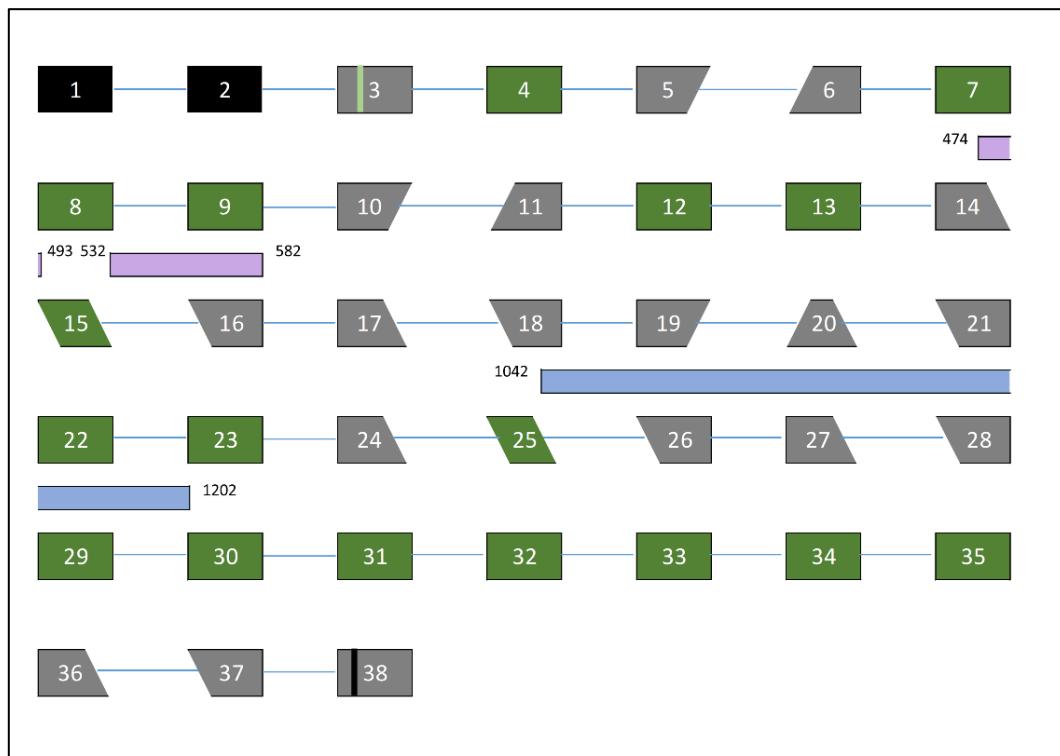


Figure 4.10. Schematic representation of the 38 exons corresponding to the *CC2D2A* human transcript NM_001080522 (also called ENST00000503292.5). Rectangles represent exons (not proportional with their real size) with full complement of codons (may not be the case of the first and last coding exons). Trapezoids represent exons in which the codon in first, last place or both, is split between two adjacent exons, ie: end of exon: “/” represents an exon in which the last codon needs two more nucleotides that will be completed in the next exon, while end of exon: “\” represents an exon in which in the last codon needs one more nucleotide that will be completed in the next exon. The black rectangles corresponds to the non-coding exons 1 and 2. The start codon (ATG) is represented with a green bar in exon 3. The termination codon (TAA) is represented by a black line and the stop codon (TAA) in exon 38. Grey rectangles represent exons which are not skippable, given that the reading frame would be disrupted if skipped. Green rectangles represent skippable exons (ie: exons that when are skipped the reading frame is maintained). Blue rectangle located above exons 25-30, indicate that the exon/s (or part of the exon/s) below encode for the *CC2D2A* C2 domain, while purple rectangles indicate that the exon/s (or part of the exon/s) (exons 14-16) below encode for the two *CC2D2A* coil-coiled domains. Next to each of these rectangles, the amino acid start and end of each of the described *CC2D2A* functional domains is indicated. Protein domains have been annotated considering SMART online tool.

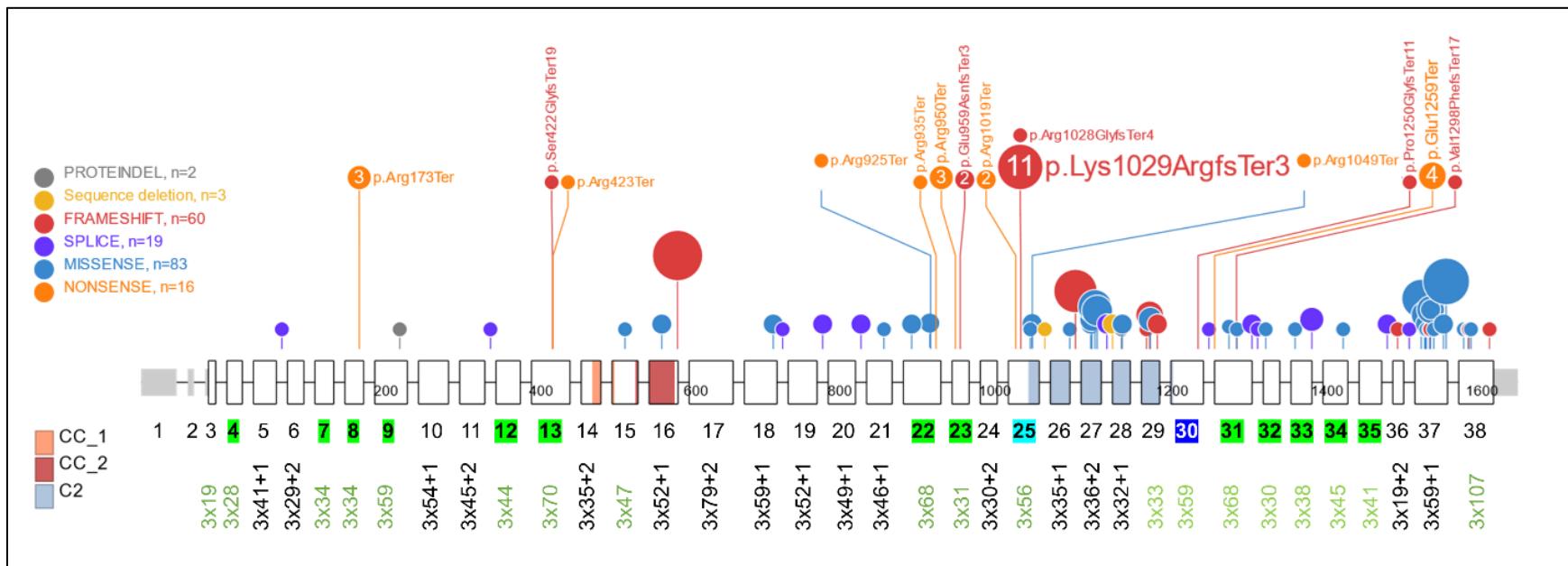


Figure 4.11. Schematic representation of the 38 exons corresponding to the *CC2D2A* human transcript NM_001080522.2 (also called ENST00000503292.5) and the location of the reported *CC2D2A* mutations. Exons numbers are shown, below it, for each exon, it is shown the number of nucleotides corresponding to each exon (shown in multiples of three). For each exon, the number of nucleotides is coloured in green if the total number of nucleotides is an exact multiple of 3. *CC2D2A* protein domains: 2 coiled-coil domains (CC) and 1 C2 domain are coloured with different colours. The *CC2D2A* variants reported in the literature their corresponding exon and accordingly with their allelic frequency in the patients reported in the literature (disc size reflecting allelic frequency). The variants are colour coded according to their annotated consequence. Exons numbers corresponding to exons that could be suggested as candidates for a targeted exon skipping therapy, based on nucleotide numbers: are shaded in green. Additionally, exons that could be suggested as candidates for a targeted exon skipping therapy, based tissue-specific transcript analysis are shaded in blue, and a possible candidate exon based on conflicting domain annotation (between the C2 domain annotation provided by SMART online tool and the C2 annotation found in some studies from the literature) is shaded in turquoise. Protein domains have been annotated considering SMART online tool. Diagram has been generated using ProteinPaint online tool.

There are 7 skippable exons with *CC2D2A* variants reported in the literature. These are exons: 8, 13, 22, 23, 25, 30 and 31. These exons are predicted to maintain the protein reading frame, not causing loss of protein function, if skipped. There are some disease-causing truncating *CC2D2A* mutations reported up to date, that are located in these skippable exons. The patients reported in the literature with one or two truncating *CC2D2A* mutations located in any of these skippable exons are described in Table 4.1.

In total, there are 14 *CC2D2A* truncating variants located in exons 8, 13, 22, 23, 25, 30 and 31, with a pathogenic effect that could potentially be skipped by an exon targeted skipping therapy. The skipping of any of these exons could bypass the deleterious effect of the corresponding truncating mutation/s located in that corresponding exon and is expected to result in a near-full length functional *CC2D2A* protein (Figure 4.12).

Of note, four of these truncating variants: c.517C>T; p.(Arg173Ter) (exon 8), c.2848C>T; p.(Arg950Ter) (exon 23), c.2875del; p.(Glu959AsnfsTer3) (exon 23) and c.3084del; p.(Lys1029ArgfsTer3) (exon 25), have been described in homozygosity. (Table 4.1). As a result, it can be suggested that *CC2D2A* is a good candidate gene to design and apply targeted exon skipping therapies as there are multiple skippable exons that could potentially be used for exon skipping therapy and there are patients with disease-causing truncating *CC2D2A* mutations in these exons.

Family ID (1)	Patient ID	Phe noty pe	Kidney phenot ype (2)	Allele 1 (Ex,Int)	Allele 2 (Ex,Int)	Reference
15	UW 41- IV: 1	GBT S	no	c.2848C>T ; p.(Arg950Ter) (Ex23)	c.2848C>T ; p.(Arg950Ter) (Ex23)	(Gorden et al., 2008)
16	UW 47- II:1	GBT S	no	c.3055C>T ; p.(Arg1019Ter) (Ex25)	c.3288G>C ; p.(Gln1096His) (Ex26)	(Gorden et al., 2008)
20	UM 10	MKS	NA	c.3084del ; p.(Lys1029ArgfsTer3) (Ex25)	c.4179+1del (Int33)	(Tallila et al., 2008)
26	MK S- 54	MKS	yes	c.517C>T ; p.(Arg173Ter) (Ex8)	c.517C>T ; p.(Arg173Ter) (Ex8)	(Mougou-Zerelli et al., 2009)
29	MK S- 977	MKS	yes	c.3084del ; p.(Lys1029ArgfsTer3) (Ex25)	c.3084del ; p.(Lys1029ArgfsTer3) (Ex25)	(Mougou-Zerelli et al., 2009)
33	MK S- 365	MKS	yes	c.2773C>T ; p.(Arg925Ter) (Ex22) (3)	c.2486+1G>C (Int20)	(Mougou-Zerelli et al., 2009)
34	UW 67	GBT S	yes	c.3347C>T ; p.(Thr1116Met) (Ex27)	c.3145C>T ; p.(Arg1049Ter) (Ex25)	(Doherty et al., 2010)
35	F43 4- 21	GBT S	no	c.517C>T ; p.(Arg173Ter) (Ex8)	c.1676T>C ; p.(Leu559Pro) (Ex16)	(Otto et al., 2011)
36	A2 421 -21	MKS	yes	c.3544T>C ; p.(Trp1182Arg) (Ex29)	c.3774dup ; p.(Glu1259Ter) (Ex31)	(Otto et al., 2011)
38	M5 06	MKS	NA	c.3084del ; p.(Lys1029ArgfsTer3) (Ex25)	c.3084del ; p.(Lys1029ArgfsTer3) (Ex25)	(Hopp et al., 2011)
40	UW 75- 3	GBT S	no	c.1676T>C ; p.(Leu559Pro) (Ex16)	c.3892_3893del ; p.(Val1298PhefsTer17) (Ex31)	(Bachmann-Gagescu et al., 2012)
42	UW 78- 3	GBT S	NA	c.3055C>T ; p.(Arg1019Ter) (Ex25)	c.4667 ^a >T ; p.(Asp1556Val) (Ex37)	(Bachmann-Gagescu et al., 2012)
43	UW 79- 3	GBT S	no	c.1263_1264insGGCAT GTTTGCG ; p.(Ser422GlyfsTer19) (Ex13) (4)	c.3452T>C ; p.(Val1151Ala) (Ex28)	(Bachmann-Gagescu et al., 2012)
43	UW 79- 4	GBT S	no	c.1263_1264insGGCAT GTTTGCG ; p.(Ser422GlyfsTer19) (Ex13) (4)	c.3452T>C ; p.(Val1151Ala) (Ex28)	(Bachmann-Gagescu et al., 2012)
51	128	MKS	NA	c.3544T>C ; p.(Trp1182Arg) (Ex29)	c.3774dup ; p.(Glu1259Ter) (Ex31)	(Szymanska et al., 2012)
61		MKS	yes	c.3774dup ; p.(Glu1259Ter) (Ex31)	c.4550C>G ; p.(Thr1517Ser) (Ex37)	(Jones et al., 2014)
62	MTI - 127	GBT S (5)	NA	c.4583G>A ; p.(Arg1528His) (Ex37) (6)	c.3082del ; p.(Arg1028GlyfsTer4) (Ex25) (7)	(Ben-Salem et al., 2014)
72	3	GBT S/M KS (8)	NA	c.2803C>T ; p.(Arg935Ter) (Ex22)	c.3774dup ; p.(Glu1259Ter) (Ex31)	(Watson et al., 2016)

73	4	JBTS/MKS (8)	NA	c.2875del ; p.(Glu959AsnfsTer3) (Ex23)	c.2875del ; p.(Glu959AsnfsTer3) (Ex23)	(Watson et al., 2016)
74	FT-1	MKS	yes	c.3084del ; p.(Lys1029ArgfsTer3) (Ex25)	c.3084del ; p.(Lys1029ArgfsTer3) (Ex25)	(Al-Hamed et al., 2016)
78	FT-15	MKS	yes	c.3084del ; p.(Lys1029ArgfsTer3) (Ex25)	c.3084del ; p.(Lys1029ArgfsTer3) (Ex25)	(Al-Hamed et al., 2016)
79	FT-21	MKS	yes	c.3084del ; p.(Lys1029ArgfsTer3) (Ex25)	c.3084del ; p.(Lys1029ArgfsTer3) (Ex25)	(Al-Hamed et al., 2016)
81	F85 0-21	Cogan	yes	c.1267C>T ; p.(Arg423Ter) (Ex13)	c.4667 ^a >T ; p.(Asp1556Val) (Ex37)	(Schueler et al., 2016)
84		JBTS	yes	c.2581G>A ; p.(Asp861Asn) (Ex21)	c.2848C>T ; p.(Arg950Ter) (Ex23)	(Xiao et al., 2017)
89	44:36:00	JBTS	no	c.3744_3747dup ; p.(Pro1250GlyfsTer11) (Ex30) (9)	c.3989G>A ; p.(Arg1330Gln) (Ex32)	(Vilboux et al., 2017a)
89	45:36:00	JBTS	no	c.3744_3747dup ; p.(Pro1250GlyfsTer11) (Ex30) (9)	c.3989G>A ; p.(Arg1330Gln) (Ex32)	(Vilboux et al., 2017a)

Table 4.1. Patients with truncating *CC2D2A* variants in potentially skippable exons and associated phenotypes (ranked by publication). *CC2D2A* transcript considered: NM_001080522.2. (1) Relates to family ID of complete database published in Barroso-Gil et al., 2021 (Barroso-Gil et al., 2021a). Designated as NA, unless renal phenotype clearly stated. (3) Variant initially reported as c.2673C>T; p.(Arg925Ter). I assumed c.2773C>T is the correct nucleotide change given that it is predicted to give rise to the reported protein change. (4) Variant initially reported as c.1263_4InsGGCATGTTTGGC; c.1268G>A; p.(Ser423Glyfs*19). (5) Study does not state potential extra-CNS manifestations. (6) Variant initially reported as c.4258G>A; p.(Arg1528His). This variant was corrected as c.4583G>A; p.(Arg1528His) (Lam et al., 2020). (7) Variant initially reported as c.1412delG; p.(Lys472Argfs*). This variant was corrected as c.3082del; p.(Arg1028Glyfs*4) (Ben-Salem et al., 2015). (8) In the study of Watson et al., (Watson et al., 2016) patients were referred with a clinical diagnosis of either JBTS (9 patients) or MKS (17 patients). The genetic diagnosis was confirmed in 14 of the 26 cases, a diagnostic yield of 54%. The exact phenotype is not reported. (9) Variant initially reported as c.3743_3746dup; p.(Pro1250Glyfs*11). Cogan, Cogan-type congenital oculomotor apraxia; JBTS, Joubert syndrome; MKS, Meckel syndrome; ML, Meckel-like syndrome.

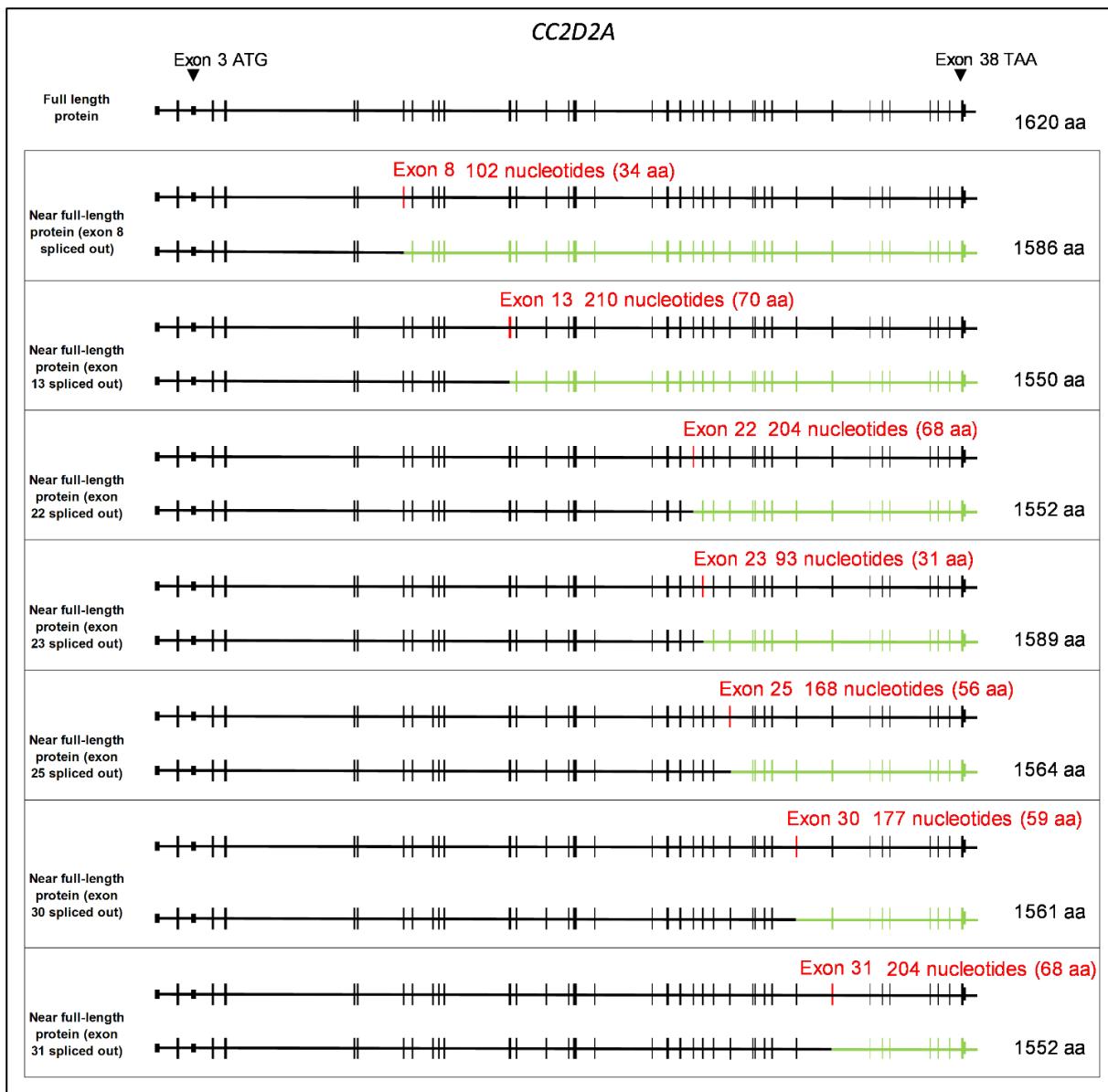


Figure 4.12. Identified CC2D2A exon targets for potential exon skipping therapies and predicted consequences of exon skipping. CC2D2A transcript considered: NM_001080522 (also called ENST00000503292.5). CC2D2A transcript contains 38 exons. The start codon (ATG) is located in exon 3 and the termination codon (TAA) is located in exon 38. The predicted full-length protein contains 1620 amino acids. Below the predicted full-length protein are shown each the different exon skipping scenarios if each of the potentially skippable exons suggested for an exon skipping therapy are skipped. For each of the exon skipping scenarios, in red: it is indicated the corresponding exon that would be skipped, its size, and the size of the protein sequence it encodes. Of note, all exon skipping events maintain the protein reading frame. The sequence downstream each of the exon skipping events is shown in green. The predicted protein size of the protein products (near full length proteins) is indicated next to the corresponding protein product.

4.2.4. Use of WGS data from the Genomics England 100,000 Genomes Project to find novel patients presenting with biallelic *CEP120* variants

Mutations in *CEP120* have been associated with JBTS and complex skeletal phenotypes (JATD, TCDOE, MKS and OFD). However the number of patients with disease-causing pathogenic reported in *CEP120* is low, which is a limitation to establish any robust genotype-phenotype correlation and understand the molecular basis of the pleiotropy observed in *CEP120* (Roosing et al., 2016, Shaheen et al., 2015).

In order to identify novel *CEP120* patients, I used WGS data available from the Genomics England 100,000 Genomes Project database. To search for patients presenting with biallelic *CEP120* variants, the tiering data (Genomics England Rare Disease Tiering Process) was used, consisting in a list of variants provided by this project for most of the participants included in the rare disease program (Figure 4.13A).

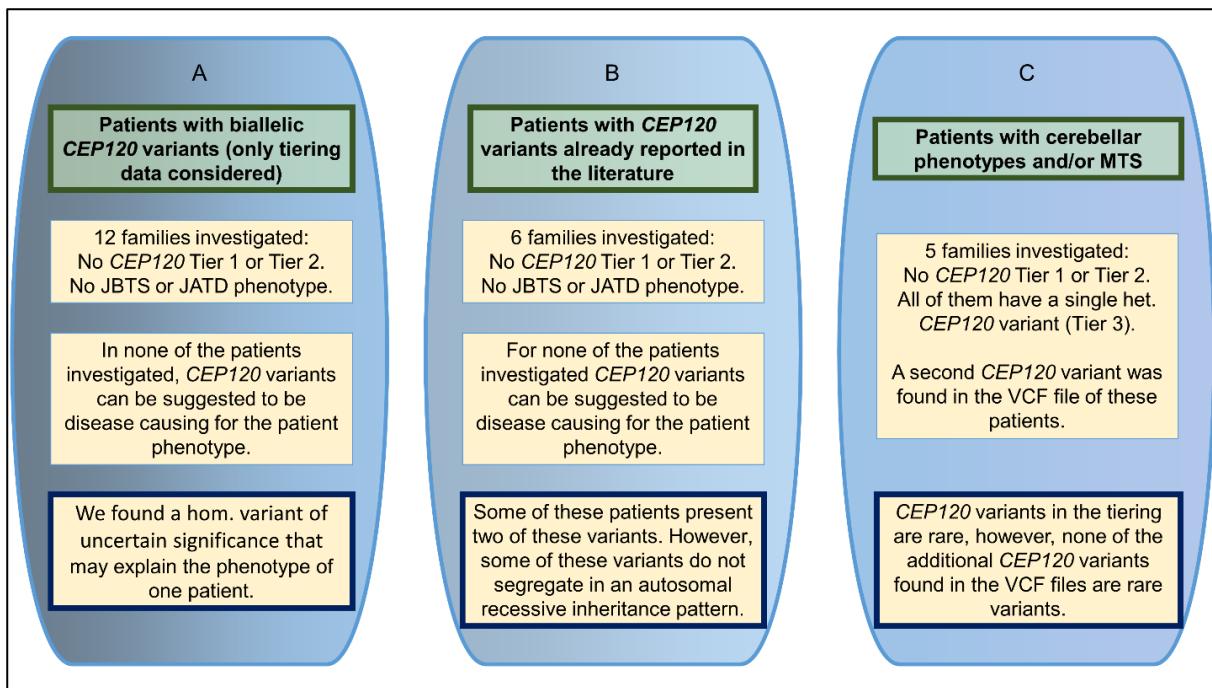


Figure 4.13. Pipeline to search novel patients with pathogenic *CEP120* variants in the Genomics England 100,000 Genomes Project (main release v9, dated 2nd April 2020). In this analysis disease-causing variants in *CEP120* or novel genotype-phenotype associations were not found, neither extend the *CEP120* phenotypic spectrum. A: Search for patients with biallelic *CEP120* mutations in the tiering data available. There are 71161 participants in the tiering data, 33780 of those participants are probands. B: Search for patients with *CEP120* variants already reported in the literature in the participants included in the rare disease program (there are 9 different mutations described in the literature). There are 71672 participants in the Rare Disease Program, 33403 of those participants are probands. C: Search for patients with cerebellar phenotypes and/or MTS and at least one *CEP120* heterozygous variant in the tiering data. For more information, see Tables 4.2 and 4.3. In Tables 4.2 and 4.3: Family 1 and Families 8-18 correspond (12 families) to analysis A, Families 2-7 (6 families) correspond to analysis B and Families 19-23 (5 families) correspond to analysis C.

12 families (Family 1 and families 8-18 in Tables 4.2 and 4.3) were found presenting with biallelic *CEP120* variants (not reported in the literature). These families were included in the rare disease program of Genomics England 100,000 Genomes Project. In these 12 families, I also searched for other putative disease variants in other genes that could explain the patients' phenotype.

None of the patients present a Joubert or Jeune phenotype. Some of these patients present other ciliary phenotypes such as retinal dystrophy or ataxia. None of these families presented *CEP120* variants that could be suggested to be responsible for the patients' phenotypes. One of these family cases (Family 18) is already solved by genomics England, as the patient presents compound heterozygous variants in the gene *PAH*, a gene associated with Phenylketonuria (van Spronsen et al., 2021). The patient has a phenotype consisting in abnormality of metabolism/homeostasis. After discarding deep intronic and common variants, it cannot be suggested any other variant/s as disease-causing for any of the rest of the patients.

Of note, one homozygous missense *CEP120* variant of uncertain significance was found in a patient diagnosed with developmental macular and foveal dystrophy (see Family 1 from Tables 4.2 and 4.3 and Figure 4.14). This homozygous variant is: c.1202A>C ; p.(Glu401Ala) (from now E401A variant) and it is not reported in the literature. Pathogenicity scores: SIFT (0.05, tolerated) and PolyPhen (0.11, benign) indicate that this variant may be benign, the CADD value is 22.4 (which is close to the threshold of being pathogenic: 25), MutationTaster describes it as "disease-causing", and it is rare according to gnomAD (0.000032 of total allele frequency). Its pathogenicity is described by the ClinVar database and American College of Medical Genetics and Genomic (ACMG) guidelines as a variant of uncertain significance.

No other putative pathogenic variants have been found in this family, it cannot be determined if this homozygous *CEP120* variant is the cause of the patient phenotype. In this and the rest of the chapters of this thesis, I use the terms "disease-causing" or "pathogenic" indistinctly for a variant with enough evidence of being responsible for the clinical phenotypic found in the corresponding patient. However, a variant described as disease-causing, pathogenic or deleterious by MutationTaster or other tools for analysis of pathogenicity, may not be described as "disease-causing" or "pathogenic" after combining all the analyses and assessments.

Family ID	Patient ID	Year Of Birth	Participant Ethnic Category	Consanguinity	Cause solved family	Phenotype	HPO_terms	Number of parents affected	Number of siblings affected
1-C P1 20	1-C P1 20	2005 -2010	South Asian	Yes	No	Developmental macular and foveal dystrophy	Visual impairment, Nonprogressive visual loss, Decreased central vision, Central scotoma, Abnormality of color vision, Noninflammatory macular atrophy, Retinal dystrophy	0	1
2-C E P1 20	2-C E P1 20	2005 -2010	Not Stated	No	Yes	Intellectual disability	Preauricular skin tag, Global developmental delay, Generalized hypotonia, Growth abnormality, Proportionate short stature, Localized hirsutism	0	0
3-C E P1 20	3-C E P1 20	1955 -1960	European ancestry	No	NA	Hereditary Motor Sensory Neuropathy	Ataxia, Postural tremor, Falls, Paresthesia, Demyelinating motor neuropathy, Distal lower limb amyotrophy, Proximal muscle weakness in lower limbs, Foot dorsiflexor weakness, Distal lower limb muscle weakness, Peripheral neuropathy	0	0
4-C E P1 20	4-C E P1 20	1960 -1965	European ancestry	No	No	Hereditary haemorrhagic telangiectasia	Abnormality of the bladder, Lip telangiectasia, Epistaxis, Iron deficiency anemia, Migraine with aura, Migraine without aura, Gastrointestinal telangiectasia, Type II diabetes mellitus, Pulmonary arteriovenous malformation, Arteriovenous malformation, Mucosal telangiectasiae	1	0
5-C E P1 20	5-C E P1 20	1965 -1970	Other	No	No	Amyotrophic lateral sclerosis or motor neuron disease	Dysphagia, Fasciculations, Skeletal muscle atrophy, Sensory axonal neuropathy, Sensory impairment, Motor axonal neuropathy, Distal peripheral sensory neuropathy, Distal upper limb amyotrophy, Distal lower limb amyotrophy, Proximal upper limb amyotrophy, Impaired tactile sensation, Impaired proprioception	0	0
6-C E P1 20	6-C E P1 20	1995 -2000	European	No	NA	Epilepsy plus other features	Autism, Cutis marmorata, Seizures, Global developmental delay, Generalized tonic seizures, Cognitive impairment	0	0

7-C E P1 20	7-C E P1 20	19 70 - 19 75	Europ ean ancest ry	No	No	Hereditary spastic paraplegia	Eczema, Spasticity, Myoclonus, Lower limb spasticity, Spastic gait, Upper limb spasticity, Progressive spastic paraplegia, Fatigue, Crohn's disease	0	0
8-C E P1 20	8-C E P1 20	20 10 - 20 15	White	Possi ble	NA	Undiagnosed metabolic disorders	Delayed gross motor development, Failure to thrive, Hepatomegaly, Seizures, Intellectual disability, Delayed fine motor development, Delayed speech and language development, Abnormal levels of creatine kinase in blood, Abnormality of metabolism/homeostasis, Hypocholesterolemia	0	1
9-C E P1 20	9-C E P1 20	19 85 - 19 90	White	No	No	Undiagnosed monogenic disorder seen in a specialist genetics clinic	Abnormality of the immune system, Meconium ileus, Abnormality of metabolism/homeostasis, Abnormality of prenatal development or birth, Pancreatitis, Abnormality of the liver, Abnormality of the genitourinary system, Hydronephrosis, Abnormality of the skeletal system, Constipation, Abnormality of the respiratory system, Abnormality of the gastrointestinal tract	0	NA
10-C E P1 20	10-C E P1 20	19 65 - 19 70	Not Stated	No	No	Early onset dystonia	Head tremor, Craniofacial dystonia	0	0
11-C E P1 20	11-C E P1 20	19 65 - 19 70	Mixed	No	No	Rod-cone dystrophy	Constriction of peripheral visual field, Retinal dystrophy, Visual impairment, Abnormal light-adapted electroretinogram, Abnormality of retinal pigmentation, Rod-cone dystrophy, Progressive visual loss	0	0
12-C E P1 20	12-C E P1 20	19 85 - 19 90	Not Stated	No	No	Extreme early- onset hypertension	Hypertension	0	0
13-C E	13-C E	20 05 - 20 10	Europ ean ancest ry	No	No	Intellectual disability	Delayed gross motor development, Delayed speech and language development, Delayed fine motor development, Intellectual disability, Autistic behavior, Atonic seizures, Global developmental delay, Seizures, Hypoplasia of the corpus callosum, Generalized tonic-clonic seizures, Focal seizures, Ataxia	0	0

P1 20	P1 20									
14 - C E P1 20	14 - C E P1 20	19 65 - 19 70	Asian	No	NA	Hereditary Motor Sensory Neuropathy	Lower limb spasticity, Impaired vibratory sensation, Demyelinating peripheral neuropathy, Bilateral ptosis, Peripheral neuropathy, Spastic paraplegia, Lower limb hyperreflexia	0	0	
15 - C E P1 20	15 - C E P1 20	20 00 - 20 05	Europ ean ancest ry	Un kn ow n	NA	Epilepsy plus other features	Atypical absence seizures, Generalized tonic-clonic seizures, Intellectual disability, Cognitive impairment, EEG abnormality, Autism, Seizures	0	0	
16 - C E P1 20	16 - C E P1 20	19 40 - 19 45	Europ ean ancest ry	No	NA	Familial pulmonary fibrosis	Gastroesophageal reflux, Cough, Dyspnea, Pulmonary fibrosis, Interstitial pulmonary abnormality	0	1	
17 - C E P1 20	17 - C E P1 20	20 00 - 20 05	Europ ean ancest ry	No	No	Early onset and familial Parkinson's Disease / Early onset dystonia	Short stature, Mild microcephaly, Dystonia, Action tremor, Spasticity, Parkinsonism, Limb dystonia, Abnormality of eye movement, Parkinsonism with favorable response to dopaminergic medication, Abnormality of higher mental function	0	0	
18 - C E P1 20	18 - C E P1 20	19 95 - 20 00	Not Stated	No	Yes	Undiagnosed metabolic disorders	Abnormality of metabolism/homeostasis	0	0	
19 - C E P1 20	19 - C E P1 20	NR	NR	NR	NA	Epilepsy plus other features	Cerebellar atrophy and other HPO terms	Unkn ow n	0	

20 - C E P1 20	20 - C E P1 20	NR	NR	NR	NA	Brain channelopathy	Cerebellar atrophy and other HPO terms	Unkno wn	0
21 - C E P1 20	21 - C E P1 20	NR	NR	NR	NA	Hereditary ataxia	Cerebellar atrophy, Progressive cerebellar ataxia and other HPO terms	No	0
22 - C E P1 20	22 - C E P1 20	NR	NR	NR	NA	Hereditary ataxia	Cerebellar hypoplasia and other HPO terms	No	0
23 - C E P1 20	23 - C E P1 20	NR	NR	NR	NA	Complex Parkinsonism (includes pallido- pyramidal syndromes)	Cerebellar atrophy and other HPO terms	No	0

Table 4.2. Phenotypic information of the families investigated to find novel patients with pathogenic *CEP120* variants in the Genomics England 100,000 Genomes Project. The year of birth and some of the clinical phenotypes are not specific or are not included according to the indications of the Airlock tool to extract the results of this analysis from the Research Environment of Genomics England 100,000 Genomes Project. Case solved family column refer to the cases solved by Genomics England (main release v9, dated 2nd April 2020). Two families are solved by Genomics England: Family 2 (variant in *ANKRD11*: GRCh37, chr16:89346157:C>CG) and Family 18 (compound heterozygous for two pathogenic variants in *PAH*: chr12:102840399C>T and chr12:102844359G>C). NR, not relevant; NA, not available.

Family ID	Patient ID	Gene	Transcript	Nucleotide change	Protein change	Genotype	Tier	rs	Exon	Zygosity	Consequence
1-CEP120	1-CEP120	CEP120	ENST00000328236.9	c.1202A>C	p.Glu401Ala	hom	TIER3	rs776114102	9/21	Present in father in het.	missense
		GNA1T1	ENST00000232461.8	c.589G>A	p.Val197Met	het	TIER2	rs751898545	6/9	Not present in father	missense
		PKD1	ENST00000262304.9	c.6622G>A	p.Gly2208Ser	het	TIER3	rs779825346	15/46	Not present in father	missense
		CCDC40	ENST00000374877.7	c.1372G>A	p.Ala458Thr	hom	TIER3	rs372976809	9/18	Present in father in het.	missense
2-CEP120	2-CEP120	CEP120	ENST00000328236.5	c.595G>C	p.Ala199Pro	het	not in the tiering	rs367600930	6/21	Present in father in het., not present in mother	missense
		CEP120	ENST00000328236.5	c.2134C>T	p.Leu712Phe	het	not in the tiering	rs114280473	16/21	Present in father in het., not present in mother	missense
		RABEP2	ENST00000358201.4	c.1383+6G>T	-	het	TIER3	rs376677316	intron 11/12	Present in father in het., not present in mother	splice_region, intronic
		RABEP2	ENST00000358201.4	c.61+181G>T	-	het	TIER3	rs750214730	intron 1/12	Present in mother in het., not present in father	intronic
		NCOA7	ENST00000392477.2	c.1597G>A	p.Gly533Arg	het	TIER3	rs35223550	9/16	Present in father in het., not present in mother	missense
		NCOA7	ENST00000392477.2	c.242G>A	p.Arg81His	het	TIER3	rs746636348	6/16	Present in mother in het., not present in father	missense
		ANKRD11	ENST00000378330.2	c.6792_6793insC	p.Ala2265ArgfsTer8	het	not in the tiering	rs878855327	10/14	Not present in father or mother	frameshift
3-CEP120	3-CEP120	CEP120	ENST00000328236.5	c.2134C>T	p.Leu712Phe	hom	not in the tiering	rs114280473	16/21	NA	missense
		PKD1	ENST00000423118.5	c.8293C>T	p.Arg2765Cys	hom	TIER3	rs144979397	23/46	NA	missense

4-CEP120	4-CEP120	CEP120	ENST00000328236.5	c.595G>C	p.Ala199Pro	het	TIER3	rs367600930	6/21	NA	missense
		CEP120	ENST00000328236.5	c.2134C>T	p.Leu712Phe	het	not in the tiering	rs114280473	16/21	NA	missense
		PKD1	ENST00000423118.5	c.8575G>A	p.Ala2859Thr	het	TIER3	rs201905743	23/46	NA	missense
		ALMS1	ENST00000613296.5	c.6959G>A	p.Arg2319Lys	het	TIER3	rs1281816058	8/23	NA	missense
		ALMS1	ENST00000613296.5	c.10628C>G	p.Thr3543Ser	het	not in the tiering	rs45501594	16/23	NA	missense
		KIF7	ENST00000394412.8	c.2072G>A	p.Arg691His	het	TIER3	rs377712806	10/19	NA	missense
		KIF7	ENST00000394412.8	c.2345G>A	p.Arg782Gln	het	TIER3	rs749948143	11/19	NA	missense
5-CEP120	5-CEP120	CEP120	ENST00000328236.5	c.595G>C	p.Ala199Pro	het	TIER3	rs367600930	6/21	NA	missense
		CEP120	ENST00000328236.5	c.2134C>T	p.Leu712Phe	het	not in the tiering	rs114280473	16/21	NA	missense
		TTLL3	ENST00000426895.9	c.2717C>T	p.Pro906Leu	het	TIER3	rs755665773	13/13	NA	missense
		TTLL3	ENST00000426895.9	c.1155C>T	p.His385=	het	not in the tiering	rs146575088	8/13	NA	splice_region, synonymous
6-CEP120	6-CEP120	CEP120	ENST00000328236.5	c.595G>C	p.Ala199Pro	het	TIER3	rs367600930	6/21	NA	missense
		CEP120	ENST00000328236.5	c.2134C>T	p.Leu712Phe	het	not in the tiering	rs114280473	16/21	NA	missense

7- CEP1 20	7- CEP1 20	CEP 120	ENST0000 0328236.9	c.451C>T	p.Arg151Ter	het	TIER3	rs7574 99322	5/21	NA	stop_gained
		PKD 1L1	ENST0000 0289672.6	c.3530A>G	p.His1177Arg	het	TIER3	rs1288 834296	22/57	NA	missense
		PKD 1L1	ENST0000 0289672.6	c.5125G>C	p.Gly1709Arg	het	TIER3	rs1429 43207	32/57	NA	missense
		CRO CC	ENST0000 0375541.10	c.5237G>A	p.Arg1746Gln	het	TIER3	rs1397 86167	32/37	NA	missense
		CRO CC	ENST0000 0375541.10	c.1892C>T	p.Ala631Val	het	TIER3	rs2001 50662	14/37	NA	missense
		USH 2A	ENST0000 0307340.8	c.4714C>T	p.Leu1572Phe	het	TIER3	rs1110 33333	22/72	NA	missense
		USH 2A	ENST0000 0307340.8	c.2299delG	p.Glu767SerfsTer21	het	TIER3	rs8033 8903	13/72	NA	frameshift
8- CEP1 20	8- CEP1 20	CEP 120	ENST0000 0328236.9	c.2230C>T	p.Arg744Cys	hom	TIER3	rs7637 87298	17/21	Present in full sibling in hom. and in the father in het.	missense
		SAR 1B	ENST0000 0439578.5	c.243_243delA	p.Ala82LeufsTer28	hom	TIER1	-	5/8	Present in full sibling in hom. and in the father in het.	frameshift, splice_region
		MLC 1	ENST0000 0395876.6	c.544G>A	p.Ala182Thr	hom	TIER2	rs5374 57768	7/12	Present in full sibling and father in het.	missense
9- CEP1 20	9- CEP1 20	CEP 120	ENST0000 0328236.9	c.2647C>T	p.Arg883Cys	het	TIER3	rs1403 06974	20/21	NA	missense
		CEP 120	ENST0000 0328236.9	c.2266C>T	p.Arg756Cys	het	TIER3	rs1005 671778	17/21	NA	missense
		PKD 1	ENST0000 0262304.9	c.1159G>A	p.Ala387Thr	het	TIER3	-	5/46	NA	missense
		HTT	ENST0000 0355072.10	c.114-115insCCA	p.Pro49dup	het	TIER3	-	1/67	NA	inframe_insertion
		HTT	ENST0000 0355072.10	c.110_111insGCC	p.Gln38_Pro39i nsGlnProProGln	het	TIER3	-	1/67	NA	inframe_insertion

		IFT80	ENST00000326448.12	c.1516G>A	p.Gly506Arg	het	TIER3	-	14/20	NA	missense,splice_region
10-CEP120	10-CEP120	CEP120	ENST00000328236.9	c.1431-8A>T	-	het	TIER3	rs761915664		NA	splice_region,intronic
		CEP120	ENST00000328236.9	c.779G>A	p.Arg260His	het	TIER3	rs189429890	7/21	NA	missense
		PLP1	ENST00000434483.5	c.-290+4_-290+5insT	-	hom	TIER2	rs57975228	-	NA	splice_region,intronic
11-CEP120	11-CEP120	CEP120	ENST00000328236.9	c.324A>C	p.Ala108=	het	TIER3	rs764578251	5/21	NA	splice_region,synonymous
		CEP120	ENST00000328236.9	c.325C>A	p.Pro109Thr	het	TIER3	rs773832163	5/21	NA	missense
		DNAH6	ENST00000389394.8	c.6025T>A	p.Tyr2009Asn	het	TIER3	rs1297433574	37/77	NA	missense
		DNAH6	ENST00000389394.8	c.7536G>A	p.Glu2512=	het	not in the tiering	rs1455047778	47/77	NA	synonymous
		DNAH5	ENST00000265104.4	c.4258A>G	p.Asn1420Asp	het		rs201841311	27/79	NA	missense
		DNAH5	ENST00000265104.4	c.799-51A>G	-	het	TIER3	rs576246423	-	NA	intron variant
		DNAH5	ENST00000265104.4	c.9694C>T	p.Gln3232Ter	het	TIER3	rs369683202	57/79	NA	stop_gained
		DNAH5	ENST00000265104.4	c.8383C>T	p.Arg2795Ter	het	TIER3	rs560398270	50/79	NA	stop_gained
12-CEP120	12-CEP120	CEP120	ENST00000328236.9	c.2443C>T	p.Arg815Cys	het	TIER3	rs757166785	18/21	NA	missense
		CEP120	ENST00000328236.9	c.1492C>G	p.Arg498Gly	het	TIER3	-	11/21	NA	missense

13-CEP120	13-CEP120	CEP120	ENST00000328236.9	c.779G>A	p.Arg260His	hom	TIER3	rs189429890	7/21	NA	missense
14-CEP120	14-CEP120	CEP120	ENST00000328236.9	c.868T>A	p.Leu290Ile	hom	TIER3	rs774772594	8/21	NA	missense
		PKD1	ENST00000262304.9	c.10094T>C	p.Leu3365Pro	het	TIER3	rs1596503620	31/46	NA	missense
		IFT172	ENST00000260570.8	c.4136G>A	p.Arg1379His	het	TIER3	rs772212247	37/48	NA	missense
		IFT172	ENST00000260570.8	c.4990C>T	p.Arg1664Trp	het	not in the tiering	rs139348179	46/48	NA	missense
		ALM S1	ENST00000484298.5	c.3017G>T	p.Ser1006Ile	het	TIER3	rs543597396	7/22	NA	missense
		ALM S1	ENST00000484298.5	c.2912A>G	p.Tyr971Cys	het	TIER3	rs371893544	7/22	NA	missense
15-CEP120	15-CEP120	CEP120	ENST00000328236.9	c.2711G>T	p.Arg904Ile	het	TIER3	rs763451025	20/21	Present in mother in het. and not present in father	missense
		CEP120	ENST00000328236.9	c.472A>C	p.Ile158Leu	het	TIER3	-	6/21	Present in father in het. and not present in mother	missense
		NSD HL	ENST00000370274.8	c.233T>C	p.Val78Ala	hom	TIER2	rs140327958	3/8	Present in mother in het. and not present in father	missense
16-CEP120	16-CEP120	CEP120	ENST00000328236.9	c.1180C>T	p.Pro394Ser	het	TIER3	rs200450605	9/21	NA	missense
		CEP120	ENST00000328236.9	c.779G>A	p.Arg260His	het	TIER3	rs189429890	7/21	NA	missense
		CCD C40	ENST00000374877.7	c.*20_*82del	-	hom	TIER3	rs71163918	18/18	NA	stop_retained, 3_prime_UTR
17-CEP120	17-CEP120	CEP120	ENST00000328236.9	c.-35G>A	-	het	TIER3	-	1/21	Present in father in het. and not present in mother	splice_region, 5_prime_UTR

		CEP 120	ENST0000 0328236.9	c.2689C>T	p.Arg897Ter	het	TIER3	rs7776 87497	20/21	Present in mother in het. and not present in father	stop_gained
18- CEP1 20	18- CEP1 20	CEP 120	ENST0000 0328236.9	c.1492C>T	p.Arg498Trp	het	TIER3	rs1995 09467	11/21	NA	missense
		CEP 120	ENST0000 0328236.9	c.779G>A	p.Arg260His	het	TIER3	rs1894 29890	7/21	NA	missense
		PAH	ENST0000 0553106.6	c.1042C>G	p.Leu348Val	het	TIER2	rs6251 6092	10/13	NA	missense
		PAH	ENST0000 0553106.6	c.1315+1G>A	-	het	TIER1	rs5030 861	-	NA	splice_donor
19- CEP1 20	19- CEP1 20	CEP 120	ENST0000 0328236.5	c.2690G>A	p.Arg897Gln	het	TIER3	rs5317 28977	20/21	NR	missense
		CEP 120	ENST0000 0328236.5	c.2840G>A	p.Arg947His	het	not tiered	rs2303 720	21/21	NR	missense
20- CEP1 20	20- CEP1 20	CEP 120	ENST0000 0328236.5	c.2303insT	Glu769ArgfsTer 12	het	TIER3	-	17/21	NR	frameshift
		CEP 120	ENST0000 0328236.5	c.1804C>G	p.Leu602Val	homo zygous	not tiered	rs6595 440	13/21	NR	missense
		CEP 120	ENST0000 0328236.5	c.1431-7del	-	homo zygous	not tiered	rs3432 8840	Intron 10/21	NR	splice_region
21- CEP1 20	21- CEP1 20	CEP 120	ENST0000 0328236.5	c.1358A>G	p.His453Arg	het	TIER3	rs1182 201409	10/21	NR	missense
		CEP 120	ENST0000 0328236.5	c.2840G>A	p.Arg947His	het	not tiered	rs2303 720	21/21	NR	missense
22- CEP1 20	22- CEP1 20	CEP 120	ENST0000 0328236.5	c.1040C>G	p.Ser347Cys	het	TIER3	rs7686 96922	9/21	NR	missense
		CEP 120	ENST0000 0328236.5	c.2840G>A	p.Arg947His	het	not tiered	rs2303 720	21/21	NR	missense

23-CEP120	23-CEP120	CEP120	ENST00000328236.5	c.207-4A>G	-	het	TIER3	-	Intron 3/20	NR	splice_region
		CEP120	ENST00000328236.5	c.1804C>G	p.Leu602Val	het	not tiered	rs6595440	13/21	NR	missense
		CEP120	ENST00000328236.5	c.1431-7del	-	het	not tiered	rs34328840	Intron 10/21	NR	splice_region

Table 4.3. Variants from the families investigated in the Genomics England 100,000 Genomes Project to find novel disease-causing *CEP120* variants. *CEP120* variants reported in the literature are highlighted in orange. *CEP120* variant which was further studied as it potentially could be disease-causing is highlighted in blue. The rest of the *CEP120* variants were discarded as potentially disease-causing. Het, heterozygous; hom, homozygous; NR, not relevant; NA, not available.

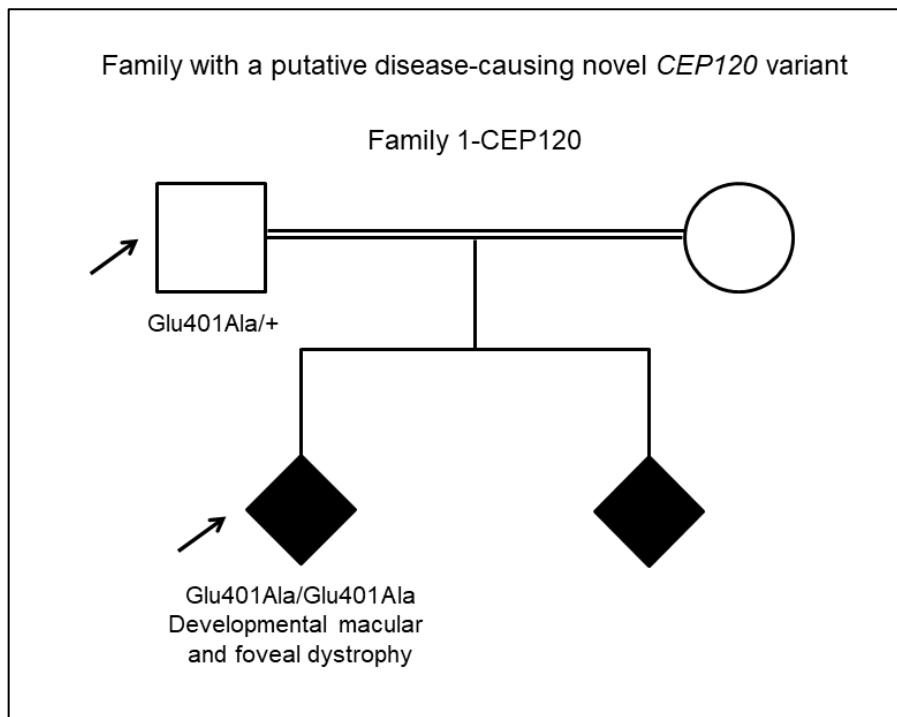


Figure 4.14. Pedigree diagram showing the segregation of a missense *CEP120* variant in a consanguineous family. Patient 1-CEP120 (represented by a black diamond pointed by an arrow) presents a homozygous *CEP120* variant: c.1202A>C; p.(Glu401Ala). This variant is present in the patient's father, in heterozygous state (represented by a white square pointed by an arrow). There is not genome sequencing information from the mother. Affected and unaffected individuals are coloured in black and white respectively.

4.2.5. Use of WGS data from the Genomics England 100,000 Genomes Project to find patients with known *CEP120* variants

To identify novel *CEP120* patients, using WGS data available from the Genomics England 100,000 Genomes Project database, I searched for patients presenting any of the nine variants already reported in the literature. These variants were searched in all the participants included in the rare disease program of the Genomics England 100,000 Genomes Project, regardless if they went through the Genomics England Rare Disease Tiering Process or not (Figure 4.13B).

I found 6 families (families 2-7 in Tables 4.2 and 4.3 and Figure 4.15) presenting *CEP120* variants already reported in the literature. In these 6 families, other putative disease variants were searched in other genes that could explain the patients' phenotype. None of the patients present a Joubert or Jeune phenotype. Some of them present other ciliary phenotypes such retinal dystrophy, ataxia or abnormality of the skeletal system. None of these families presented *CEP120* variant/s in homozygosity or compound heterozygosity that could be suggested to be responsible for any of the patients' phenotype.

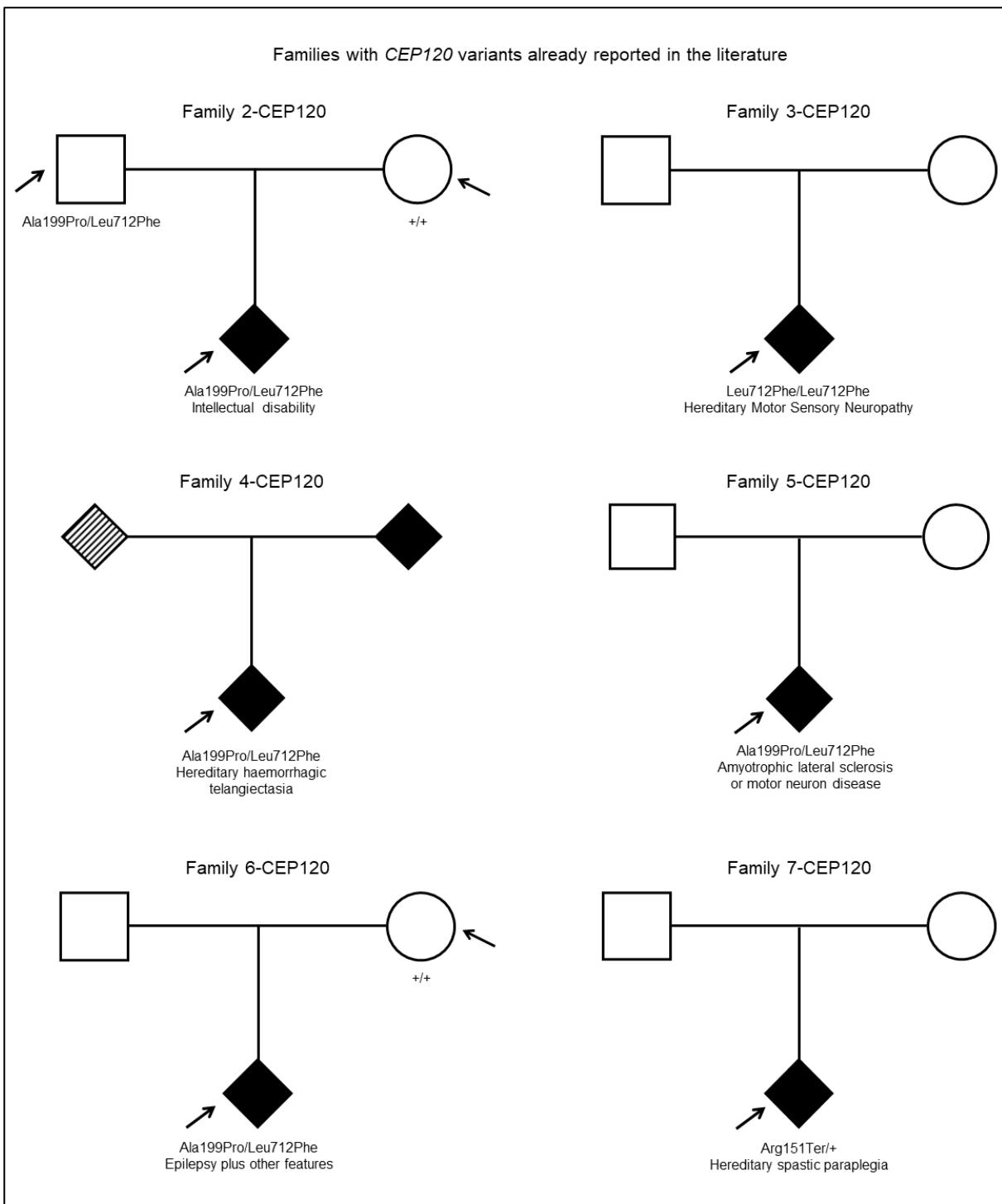


Figure 4.15. Pedigree diagram showing the families presenting *CEP120* mutations already reported in the literature. All these *CEP120* variants shown in this figure were found in heterozygous state, except for Leu712Phe variant present in the proband in Family of 3-CEP120 (which was found homozygously). Arrows indicate the participants from the rare disease cohort in which the genome sequencing is available. Black diamonds indicate affected individuals, while white circles or squares indicate non affected individuals. Lined parent in Family 3 indicates that the parent phenotype is not described as affected or not affected, it is described as unknown.

One of these families (Family 2-CEP120) is already solved by genomics England by heterozygous variant in the gene *ANKRD11*. This proband, diagnosed with intellectual disability, presents a heterozygous frameshift variant in the gene *ANKRD11*: c.6792_6793insC (p.Ala2265ArgfsTer8), which is not present in any of the unaffected parents. *ANKRD11* is a gene associated with KBG syndrome, a rare disease involving intellectual disability and other phenotypes including craniofacial and skeletal phenotypes following an autosomal dominant inheritance pattern (Sirmaci et al., 2011, Breslow et al., 2018). After discarding deep intronic and common variants, it cannot be suggested any other variant/s as disease-causing for any of the rest of the patients.

Of note, in two of these families (Families 2-CEP120 and 6-CEP120, in which genome sequencing of both and one of the parents, respectively, is available) the *CEP120* variants were analysed and excluded as putative disease-causing variants. There is no evidence to support that these variants could be responsible for the patients' phenotype given that in both probands these variants seem to be inherited in cis, from a healthy parent. Furthermore, the variant c.2134C>T; p.(Leu712Phe) was found in 577 participants in the 100,000 Genomes Project, in one of them in homozygosity (Family 3), this is a frequency of ~0.004, which is found (in the 100,000 Genomes Project rare disease participants) to be more than 100 times more frequent than the other two *CEP120* reported variants: c.595G>C; p.(Ala199Pro) and c.451C>T; p.(Arg151Ter) (Table 4.4). Interestingly, the 4 participants presenting the c.595G>C; p.(Ala199Pro) variant in heterozygosity also presented the c.2134C>T; p.(Leu712Phe) variant in heterozygosity.

Mutation	Allele frequency in the participants included in the tiering data		Allele frequency in the participants included in the rare disease program	
	Probands	Total	Probands	Total
c.595G>C ; p.(Ala199Pro)	3/67560 (4.4405E-05)	3/142322 (2.1079E-05)	4/66806 (5.9875E-05)	5/143344 (3.4881E-05)
c.2134C>T ; p.(Leu712Phe)	Not in the tiering data	Not in the tiering data	471/66806 (0.00705)	578/143344 (4.0323E-03)
c.451C>T ; p.(Arg151Ter)	1/67560 (1.4802E-05)	1/142322 (7.0263E-06)	1/66806 (1.4969E-05)	1/143344 (6.9762E-06)

Table 4.4. Allele frequency of the *CEP120* mutations published in the literature up to date found in the rare disease cohort of the 100,000 Genomes Project. Data obtained from 100,000 Genomes Project main release v9 (dated 2nd April 2020). There are 71161 participants in the tiering data, 33780 of those participants are probands. There are 71672 participants in the Rare Disease Program, 33403 of those participants are probands. *CEP120* transcript: NM_153223.3.

Additionally, I searched for *CEP120* variants in patients diagnosed with Joubert syndrome, Cerebellar hypoplasia or rare multisystem ciliopathy ciliopathies (117 participants) and presenting any of the HPO phenotypes that contain the words “cerebellar”, “cerebellum” or “molar tooth sign” (692 participants found) (Figure 4.13C). Firstly, the tiering data was used, finding 5 patients with heterozygous rare *CEP120* variants. I then looked for a second *CEP120* heterozygous variant, not limited to the tiering data, searching directly in the VCF file of each patient to include available annotated variants. However, in all the 5 patients investigated, these second heterozygous variants are common (Families 19-23 in Tables 4.2 and 4.3). None of these families presented *CEP120* variants that could be suggested to be responsible for the patients’ phenotypes.

4.2.6. Use of WGS data from the Genomics England 100,000 Genomes Project to find novel patients presenting with biallelic variants in *TOGARAM1*, a gene recently associated with Joubert syndrome

In order to find JBTS patients presenting novel disease-causing variants in JBTS candidate genes. the tiering data of the 100,000 Genomes Project was used. I searched for any patient presenting pathogenic variants in *TOGARAM1*, a gene recently associated with JBTS. it was found one JBTS patient (JAS-L50) presenting a nonsense variant (c.5023C>T; p.(Arg1675Ter)) and a missense variant (c.1112C>A; p.(Ala371Asp)) in compound heterozygous state (Tables 4.5 and 4.6 and Figure 4.16) (Latour et al., 2020).

Ethnicity	Sex	Age	Molar tooth	Dev disability	Apnea/ta chypnea	Abnl eye mvts	Kidney	Crano-facial	Other
White British	Female	14	Yes	Yes	NA	Yes	Yes	Oculomotor apraxia, bilateral ptosis, left jaw/wink ptosis	Short stature, generalized hypotonia, lumbar hyperlordosis, obesity, joint hypermobility, nonverbal, autistic behaviours, small scared left kidney

Table 4.5. Clinical features of JAS-50, a patient with *TOGARAM1*-related JBTS. The patient does not present retinal, liver, polydactyly or coloboma phenotypes. Abnl, abnormal; Dev, developmental; mvts, movements; NA, not available.

Mutation	C A D D	SIFT	PolyPhen	REVEL	ACMG	Mutati onTas ter	gno mA D	Ge not ype	Zygosity
c.5023C>T ; p.(Arg1675Ter)	38	NA	NA	NA	Pathogenic	Disease causing	1/2 514 14	het	NA
c.1112C>A ; p.(Ala371Asp)	26 .3	Deleterious (0)	Possibly damaging (0.895)	Likely disease causing (0.657)	Likely pathogenic	Disease causing	5/2 825 80	het	Present in the mother in het.

Table 4.6. Predicted pathogenicity and allele frequency of *TOGARAM1* variants found in JAS-50 patient with *TOGARAM1*-related JBTS. gnomAD v2.1.1 was used to find the allelic frequency of both variants. No homozygotes were found in gnomAD for any of the variants. The rs ids corresponding to c.5023C>T ; p.(Arg1675Ter) and c.1112C>A ; p.(Ala371Asp) variants are rs745704336 and rs370676288 respectively. *TOGARAM1* transcript: NM_015091.4. NA, not available.

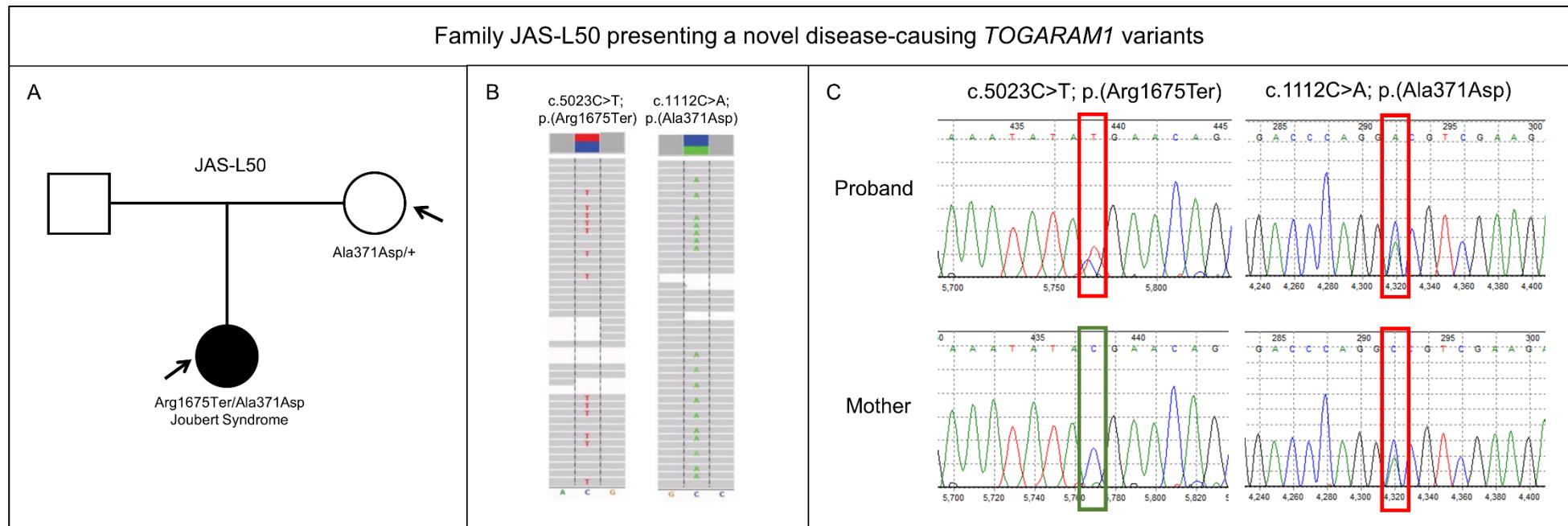


Figure 4.16. Variant segregations and sequencing visualisation of the in a family presenting a novel disease-causing *TOGARAM1* variants leading to JBTS (Latour et al., 2020). A: Pedigree diagram showing a female JBTS patient (represented by a black circle) presenting two variants in *TOGARAM1*. This patient is included in the Genomics England 100,000 Genomes Project (main release v8, dated 28th Nov 2019). The patient presents a nonsense variant: c.5023C>T; p.(Arg1675Ter), possibly inherited from her father (represented by a white rectangle), and a missense variant: c.1112C>A; p.(Ala371Asp), inherited from her mother (represented by a white circle). Genome sequencing from the father was not available. Affected and healthy individuals are coloured in black and white, respectively. B: Visualization of the next-generation sequencing data for the patient using the Integrated Genome Viewer (IGV) confirming that the patient carries both pathogenic *TOGARAM1* variants. Grey bars indicate individual reads, pathogenic variants are shown red (nucleotide base T) and green (nucleotide base A); nucleotides at the bottom of the IGV visualisation corresponds to the reference alleles. C: DNA Sanger sequencing chromatograms provided by referring physician confirming the presence (indicated as a red rectangle) or absence (indicated as a green rectangle) of the corresponding *TOGARAM1* variants in the proband and her mother.

4.2.7. Analysis of *CCDC28B* as a candidate gene for Joubert syndrome

There are more than 40 genes that have been associated with JBTS phenotypes (Focsa et al., 2021, Bachmann-Gagescu et al., 2020, Gana et al., 2022, Dewees et al., 2022). There are many other JBTS candidate genes that have been associated with other ciliopathies, such as *CEP164* and *CELSR2*, but there is not enough evidence to define them as JBTS genes (Chaki et al., 2012, Vilboux et al., 2017b).

I analysed a variant in *CCDC28B*: c.73C>T ; p.(Arg25Trp) reported in the article by Radha Rama Devi et al. (Radha Rama Devi et al., 2020). The pathogenicity of this variant was discussed, as well as if that variant should be considered to suggest *CCDC28B* as a novel candidate JBTS gene (Barroso-Gil et al., 2020).

The authors in Radha Rama Devi et al. provided molecular genetic diagnosis of a large cohort (Radha Rama Devi et al., 2020). This cohort consisted in 59 JBTS patients, from 55 families, and the authors of this study provided a genetic diagnosis in 36 families. Biallelic mutations in *CEP290* were the most common genetic cause of JBTS in this cohort.

The authors also found a putative disease-causing variant in *CCDC28B* ciliary gene, homozygously in a JBTS patient. This gene has not been associated to JBTS before so the author suggests it as a novel candidate JBTS gene.

The family in which the putative disease-causing variant segregates, consist in one affected proband (JBTS patient), an affected foetus (vermis hypoplasia revealed at 20 weeks of gestation and homozygous for the *CCDC28B* variant) and the unaffected parents (carriers of the *CCDC28B* variant in heterozygosity). The JBTS patient presented the molar tooth sign characteristic of JBTS, severe intellectual disability and polydactyly.

CCDC28B is a ciliary gene that was previously associated to the primary ciliopathy Bardet-Biedl syndrome, as a modifier gene (Badano et al., 2006a, Cardenas-Rodriguez et al., 2013). The *CCDC28B* variant was analysed in terms of predicted pathogenicity and allele frequency.

The *CCDC28B* variant was found homozygously in this reported JBTS patient and the affected foetus, and it is a missense variant: *CCDC28B*: c.73C>T; p.(Arg25Trp).

I used various prediction tools available to perform *in silico* analyses. MutationTaster was used, which described the variant as a polymorphism.

According to ACMG guidelines and REVEL pathogenicity score this variant is described as benign (Table 4.7). In contrast, it is described by SIFT and PolyPhen pathogenicity scores as deleterious and possibly damaging, respectively, and its CADD pathogenicity score is 30, which indicates that this variant may be deleterious.

More interestingly, the allele frequency of this variant is very high (too high to suggest is responsible for causing a rare disease as JBTS) in gnomAD, with an allele frequency of 0.03285 and 223 homozygous individuals across a full range of ethnicities (Table 4.8).

The high allele frequency of this variant suggests it is not disease-causing. It is suggested that homozygosity mapping and whole-exome sequencing should be combined, as their utility genes has been demonstrated to genetically solve patients and find putative disease-causing variant in novel JBTS candidate in the last few years (Alkanderi et al., 2018).

It can misleading to consider *CCDC28B* as a new JBTS gene as the pathogenicity of this *CCDC28B* variant: c.73C>T; p.(Arg25Trp) is not clear *in silico* and has not been confirmed.

As a comparison, I suggest reviewing other similar examples in which an allele, of a variant found in other studies, with a relatively high allele frequency is not responsible for JBTS when found homozygously. A similar example is the variant *KIAA0586* (NM_001244189.2): c.428delG. However, this *KIAA0586* variant (gnomAD v.2.1.2. allele frequency: 0.00312, including 2 homozygous individuals) is suggested to be disease-causing in compound heterozygous state (Pauli et al., 2019). Of note, considering the data from the main release v16 (dated 13th October 2022) of the Genomics England 100,000 Genomes Project (Rare Disease), in particular all the participants recruited in the Rare Disease program with a diagnosed disease (n=35936), this variant has an allele frequency of 0.0039 (282/71872) and none of these participants presented this *KIAA0586* variant homozygously.

The analysis of this *CCDC28B* variant reported by Radha Rama Devi et al. (Radha Rama Devi et al., 2020) reinforces the importance of doing a full assessment of all the putative disease-causing variants for each investigated patient, in particular, evaluating the allele frequency of putative disease-causing variants is essential (Barroso-Gil et al., 2020).

CADD	SIFT	PolyPhen	REVEL	ACMG	MutationTaster
30	Deleterious (0)	Possibly Damaging (0.824)	Likely benign (0.192)	Benign	Polymorphism

Table 4.7. Predicted pathogenicity of *CCDC28B* variant: c.73C>T; p.(Arg25Trp). rs id or the variant is rs1407134. *CCDC28B* transcript: NM_024296.5.

gnomAD v2.1.1				gnomAD v3.1.2				100,000 Genomes Project (Rare Disease)	
Min allele freq (1)	Max allele freq (2)	Total allele freq (3)	Total n of hom.	Min allele freq (4)	Max allele freq (5)	Total allele freq (6)	Total n hom.	Total allele freq (7)	Total n hom.
0.0129	0.0647	0.0329	223	0.0228	0.0823	0.0364	140	0.0242	40

Table 4.8. Allele frequency of *CCDC28B* variant: c.73C>T; p.(Arg25Trp). Freq, frequency; hom, homozygotes; n, number of individuals; max, maximum; min, minimum. Rs id or the variant is rs1407134. (1) Allele count/allele number/number of hom: 457/35410/3 (Latino/Admixed American). (2) Allele count/allele number/number of hom: 668/10328/16 (Ashkenazi Jewish). (3) Allele count/allele number/number of hom: 9250/281584/223. (4) Allele count/allele number/number of hom: 1553/68034/22 (European (non-Finnish)). (5) Allele count/allele number/number of hom: 26/316/0 (Middle Eastern). (6) Allele count/allele number/number of hom: 5533/152152/140. (7) Allele count/allele number/number of hom: 1740/71872/40. The data was obtained from the Genomics England 100,000 Genomes Project (Rare Disease) using the main release v16, dated 13th October 2022 and considering the participants recruited in the Rare Disease program with a diagnosed disease (n=35936).

4.2.8. Expression of *CEP120* *Shh-LIGHT2* cells and primary human fibroblasts

I wanted to assess if I was able to observe any difference in Hh signalling upon *Cep120* knockdown. To do this, Shh-LIGHT2 cells were used, as they are Hh-responsive cells used to assess effects on Hh protein (Taipale et al., 2000, Chen et al., 2002). I also wanted to assess if I was able to extract RNA from these cells and analyse the effects of Hh signalling upon *Cep120* knockdown via a dual luciferase assay and at RNA level via qPCR.

Shh-LIGHT2 have already been used to investigate the molecular genetics of primary ciliopathies with a focus on Hh signalling. Previous studies knocked out different genes in Shh-LIGHT2 cells via CRISPR/Cas9 system and used luciferase reporter

assays and other functional analyses to investigate the role of different genes in Hh signalling and ciliary function (Breslow et al., 2018, Pusapati et al., 2018).

One advantage of the use of Shh-LIGHT2 cells is that this mouse cell line grows at a faster pace and for more passages compared with human primary cells such as hURECs or primary fibroblasts, as Shh-LIGHT2 are immortalised mouse cells (Chen et al., 2002, Taipale et al., 2000). The use of Shh-LIGHT2 cells can reduce the time and resources required to perform certain functional analysis, being Shh-LIGHT2 cells suitable for functional assays to investigate ciliary function of specific genes.

The use of Shh-LIGHT2 cells is relevant to assaying ciliary function because Shh-LIGHT2 cells allow to perform multiple functional analyses such as analysis of RNA and protein expression, immunofluorescence staining of ciliary and centrosomal components and also dual luciferase assay to study effects on Hh signalling (Chen et al., 2002, Taipale et al., 2000, Breslow et al., 2018, Pusapati et al., 2018). The use of Shh-LIGHT2 cells allows to study the role of *Cep120* in Hh signalling. Furthermore, it can be suggested that a similar approach could be applied to investigate the molecular genetics of other ciliopathy genes.

Other studies have shown that the amount of Smo per unit length cilium was reduced in *Cep120*-depleted MEFs (Betleja et al., 2018) and the cerebellum of a conditional *Cep120* knockout mouse for the CNS, lacked a response to Hh signalling (Wu et al., 2014). However, the role of *CEP120/Cep120* in Hh signalling is not fully understood.

The use of Shh-LIGHT2 cells allows the analysis of the staining of the CEP120 commercial antibody (from by Biorbyt Ltd). This commercial CEP120 antibody has not been described in the literature, as other studies have used in-house generated CEP120 antibodies (Betleja et al., 2018, Mahjoub et al., 2010, Tsai et al., 2019). The use of an effective CEP120 commercial antibody an alternative to the use of other CEP120 antibodies and reduce the time and resources required for future functional analyses investigating the molecular genetics of *CEP120*.

Shh-LIGHT2 cells allow the study of Hh signalling, evaluate the effectiveness of the CEP120 antibody and validate the *Cep120* knockdown via siRNA.

The use Shh-LIGHT2 cells provides the first step to further understand the molecular disease pathogenesis of *Cep120*. Shh-LIGHT2 cells are suitable to perform functional analyses such as siRNA gene knockdown, analysis of gene expression

and immunofluorescence. Additionally, the use of Shh-LIGHT2 cells will set the basis for further functional studies using human primary cells to study patient-specific *CEP120* mutations (pipeline described in Figure 2.9).

I wanted to observe *CEP120/Cep120* protein at the centrosome, where it is expected. I want to validate its centrosomal location in different cell types: mouse Shh-LIGHT2 cells (NIH-T3T derived cells) and primary human fibroblasts. This will allow to detect any change, if any, in expression and location of *CEP120/Cep120* upon *CEP120/Cep120* knockdown.

It has been observed Cep120/CEP120 protein in different cell types. By doing Immunofluorescence I visualised Cep120/CEP120 at the centrosome and basal body of the primary cilium in control (not treated) cells (Figures 4.17-4.22).

Using primary human fibroblasts of a healthy individual, CEP120 was observed at the centrosome and basal body of the primary cilium, after 24h of serum starvation (Figures 4.17 and 4.18). In immortalised mouse cells, Shh-LIGHT2 cells, Cep120 can also be observed at the centrosome and basal body of the primary cilium, after 24h of serum starvation (Figures 4.19-4.22).

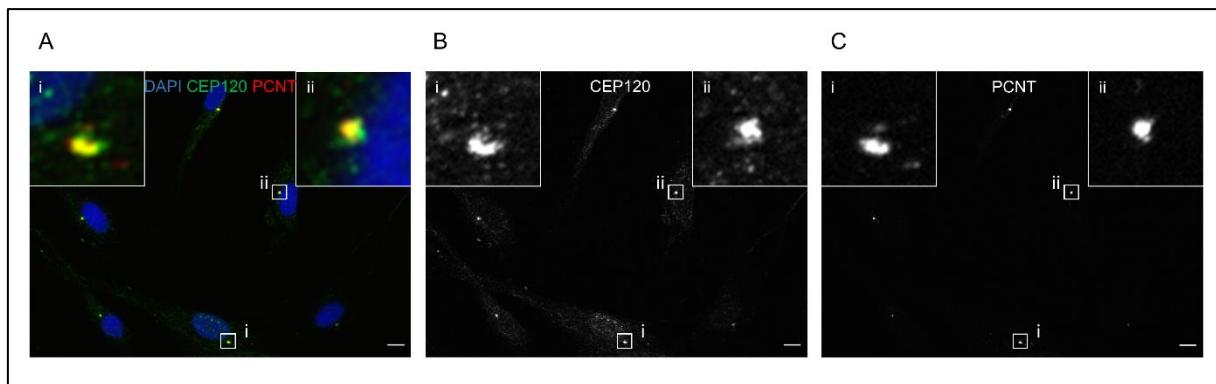


Figure 4.17. Immunofluorescence microscopy in human fibroblasts. CEP120 is observed at the centrosome. Left image (A) shows the antibodies: PCNT (red), which is a centrosome marker, and CEP120 (green). DAPI, a nuclear marker is shown in blue. CEP120 and PCNT antibodies are also shown in the middle (B) and right (C) images respectively, in white. Primary control human fibroblasts were serum starved for 24 hours. Fluorescence microscopy (Zeiss AxioImager) was used for visualisation. PCNT, Pericentrin. Scale bar: 10 μ m.

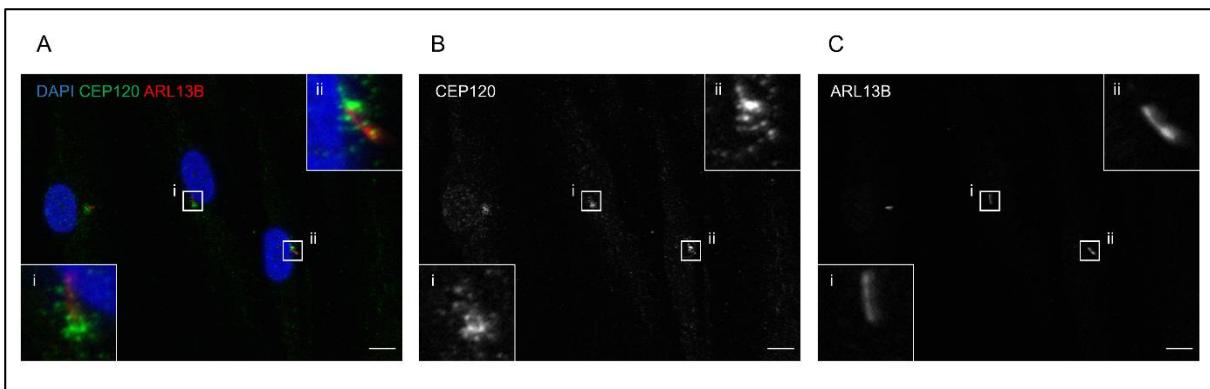


Figure 4.18. Immunofluorescence microscopy in human fibroblasts. CEP120 is observed at the basal body of the cilium and centrosome. Left image (A) shows the antibodies: ARL13B (red), which is a ciliary membrane marker, and CEP120 (green). DAPI, a nuclear marker is shown in blue. CEP120 and ARL13B antibodies are also shown in the middle (B) and right (C) images respectively, in white. Primary control human fibroblasts were serum starved for 24 hours. Fluorescence microscopy (Zeiss AxioImager) was used for visualisation. Scale bar: 10 μ m.

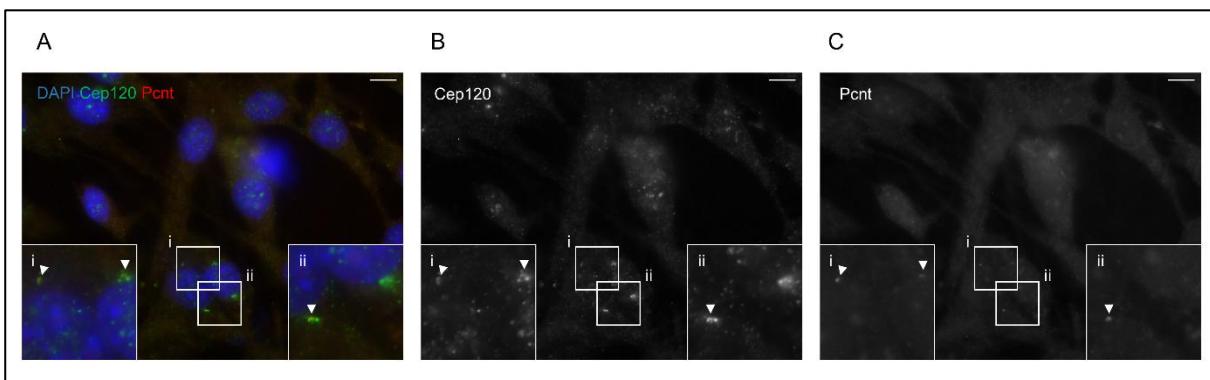


Figure 4.19. Immunofluorescence microscopy in mouse Shh-LIGHT2 cells. Cep120 is observed at the centrosome. Left image (A) shows the antibodies: Pcnt (red), which is a centrosome marker, and Cep120 (green). DAPI, a nuclear marker is shown in blue. Cep120 and Pcnt antibodies are also shown in the middle (B) and right (C) images respectively, in white. Shh-LIGHT2 cells were serum starved for 24 hours. Fluorescence microscopy (Zeiss AxioImager) was used for visualisation. Pcnt, Pericentrin. Scale bar: 10 μ m.

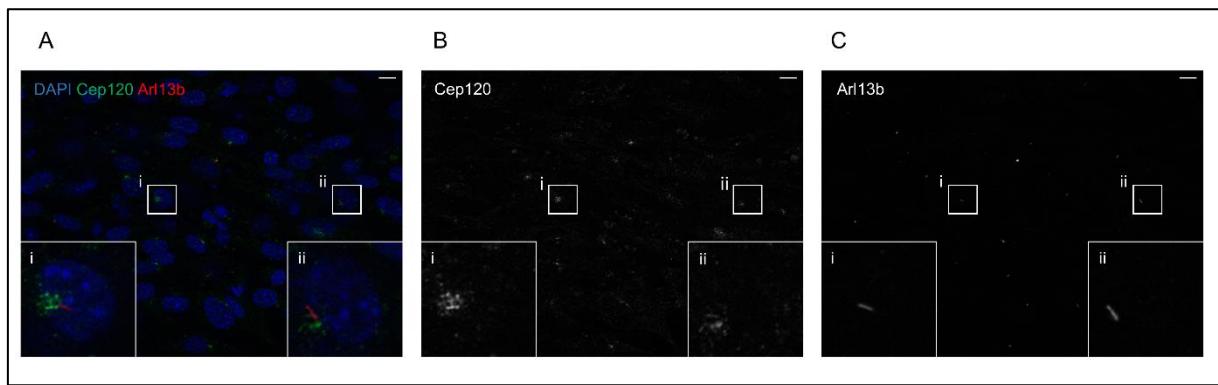


Figure 4.20. Immunofluorescence microscopy in mouse Shh-LIGHT2 cells. Cep120 is observed at the basal body of the cilium and centrosome. Left image (A) shows the antibodies: Arl13b (red), which is a ciliary membrane marker, and Cep120 (green). DAPI, a nuclear marker is shown in blue. Cep120 and Arl13b antibodies are also shown in the middle (B) and right (C) images respectively, in white. Shh-LIGHT2 cells were serum starved for 24 hours. Fluorescence microscopy (Zeiss Axiolmager) was used for visualisation. Scale bar: 10 μ m.

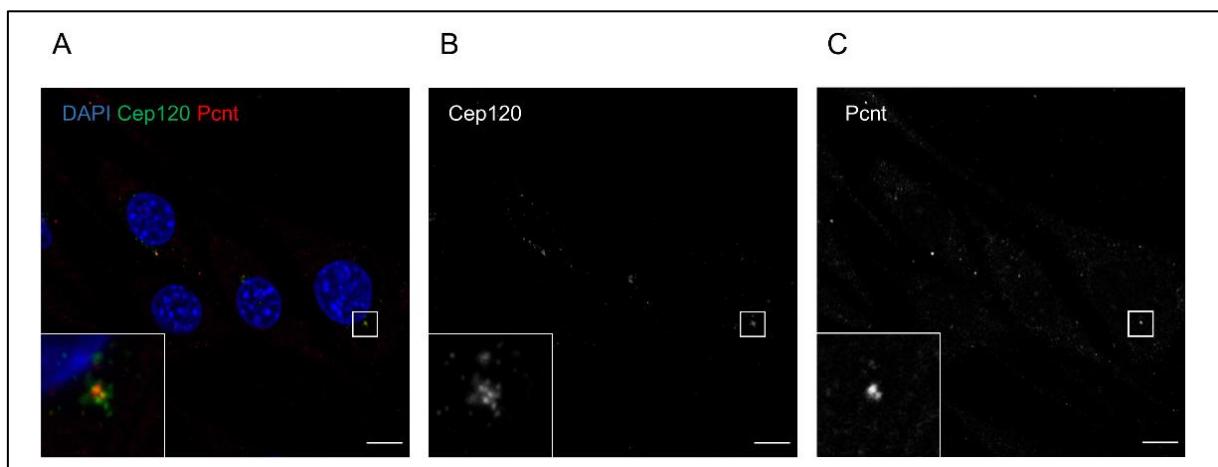


Figure 4.21. Immunofluorescence microscopy in mouse Shh-LIGHT2 cells. Cep120 is observed at the centrosome. Left image (A) shows the antibodies: Pcnt (red), which is a centrosome marker, and Cep120 (green). DAPI, a nuclear marker is shown in blue. Cep120 and Pcnt antibodies are also shown in the middle (B) and right (C) images respectively, in white. Shh-LIGHT2 cells were serum starved for 24 hours. Nikon (A1) confocal inverted microscopy was used for visualisation (image taken by Dr Elisa Molinari, Newcastle University, UK). Pcnt, Pericentrin. Scale bar: 10 μ m.

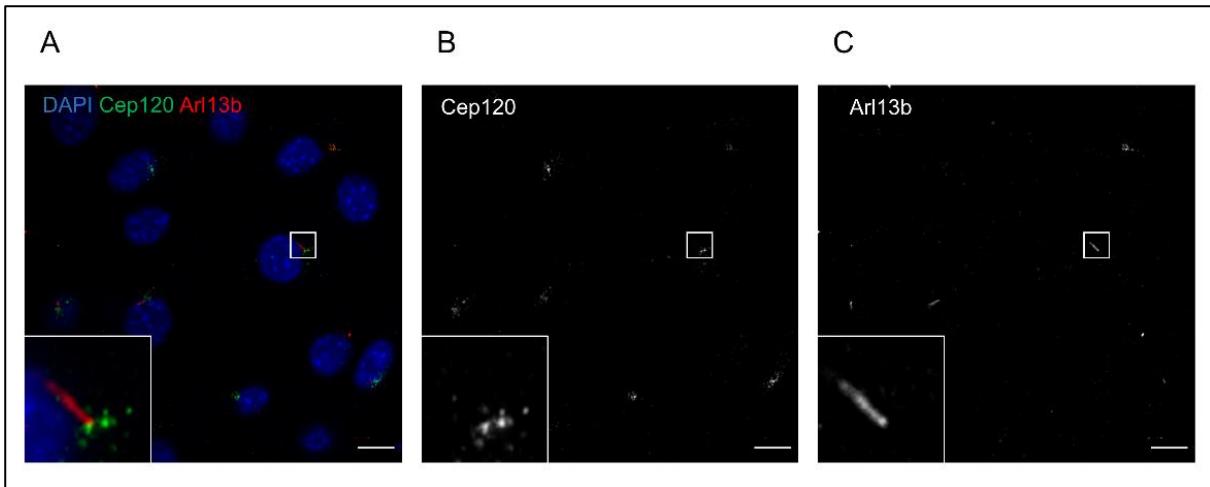


Figure 4.22. Immunofluorescence microscopy in mouse Shh-LIGHT2 cells. Cep120 is observed at the basal body of the cilium and centrosome. Left image (A) shows the antibodies: Arl13b (red), which is a ciliary membrane marker, and Cep120 (green). DAPI, a nuclear marker is shown in blue. Cep120 and Arl13b antibodies are also shown in the middle (B) and right (C) images respectively, in white. Shh-LIGHT2 cells were serum starved for 24 hours. Nikon (A1) confocal inverted microscopy was used for visualisation (image taken by Dr Elisa Molinari, Newcastle University, UK). Scale bar: 10 μ m.

In summary, using a CEP120 commercial antibody from Biorbyt Ltd (catalogue number: orb182544), CEP120 is observed at the expected locations: centrosome/basal body. Of note, the CEP120 staining in primary human fibroblasts is observed with less background (non-specific staining) compared to mouse Shh-LIGHT2 cells (compare Figures 4.17 and 4.18 against Figures 4.19-4.22). It can be suggested that this is because this commercial CEP120 antibody targets the CEP120 human protein sequence which is not identical compared to the mouse one (ie: according to the information from Biorbyt Ltd CEP120 targets the amino acids: 361-440 aa; this human protein sequence has a ~70% homology compared with mouse). It is also observed that this non-specific staining is minimal when a confocal inverted microscope is used (compare Figures 4.19 and 4.20 against Figures 4.21 and 4.22).

Regarding the negative controls, in which only the corresponding secondary antibodies were used in the absence of the primary antibodies, it is not observed any noticeable staining in primary human fibroblasts (Figure 4.23) or Shh-LIGHT2 cells (Figure 4.24).

I also wanted to assess if CEP120 could be observed via Western blot. A Rat polyclonal antibody was assayed in protein extracts from human cells and mouse kidney tissue. A Western blot revealed a defined band located around 125 kDa, which corresponds to CEP120 protein. This band was visualised in protein extracts of

mouse kidney tissue, but not in any of the protein extracts of human cells. The band was easily observed using a dilution of 1:500 of the primary CEP120 antibody (from Dr Mahjoub, Washington University, USA), and using a dilution of 1:5 of the protein extracted from mouse kidney tissue (Figure 4.25). The loading control GAPDH revealed that protein extracts from mice kidney tissue were more abundant in protein amount than protein extracts from human cells, suggesting that I may be able to observe a CEP120 band in protein lysates from cells if I increase the amount of protein loaded in future experiments.

It has to be considered that the amount of protein extracted from mouse kidney tissue is generally higher than the amount of protein extracted from hURECs or other primary cells. This could explain why a band corresponding to CEP120 is not observed in the protein extracts from hURECs in the Western blot experiment shown in Figure 4.25. Western blotting was not optimised to determine if the CEP120 antibody can be visualised in protein extracts from hURECs or other human primary cells via Western blotting, because I only wanted to determine if this CEP120 antibody was suitable for Western blotting, regardless of whether the protein lysates were extracted from tissues or from cells. The amount of CEP120 antibody available was limited and it was not needed for the rest of the experiments planned for this thesis.

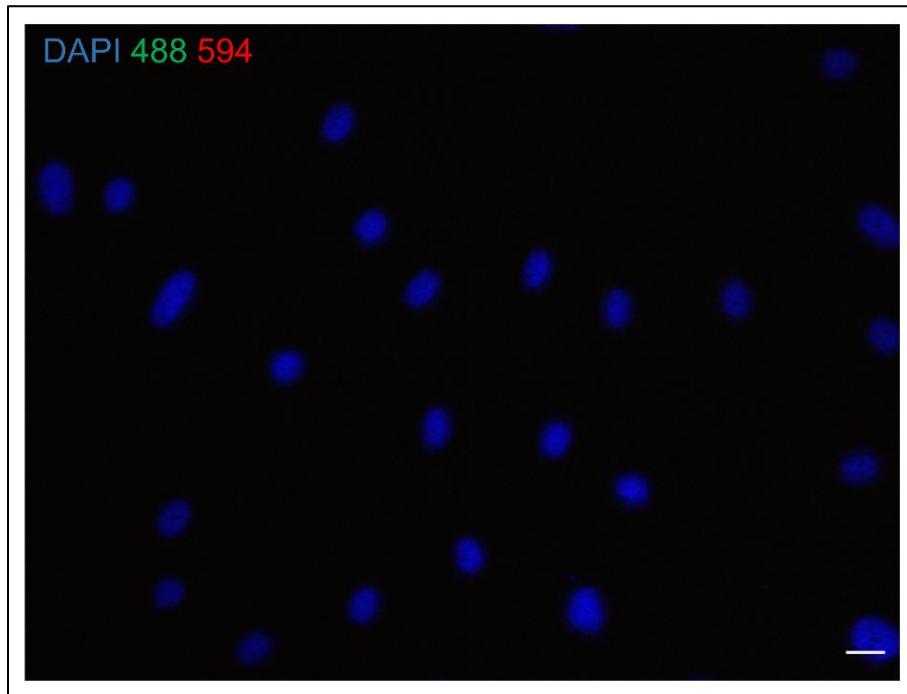


Figure 4.23. Immunofluorescence microscopy in human fibroblasts. It is observed how the cells are not stained in the absence of primary antibodies. The corresponding secondary antibodies were used: anti-rabbit 488 and anti-mouse 594. DAPI, a nuclear marker is shown in blue. Primary control human fibroblasts were serum starved for 24 hours. Fluorescence microscopy (Zeiss AxioImager) was used for visualisation of primary control human fibroblasts. Scale bar: 20 μ m.

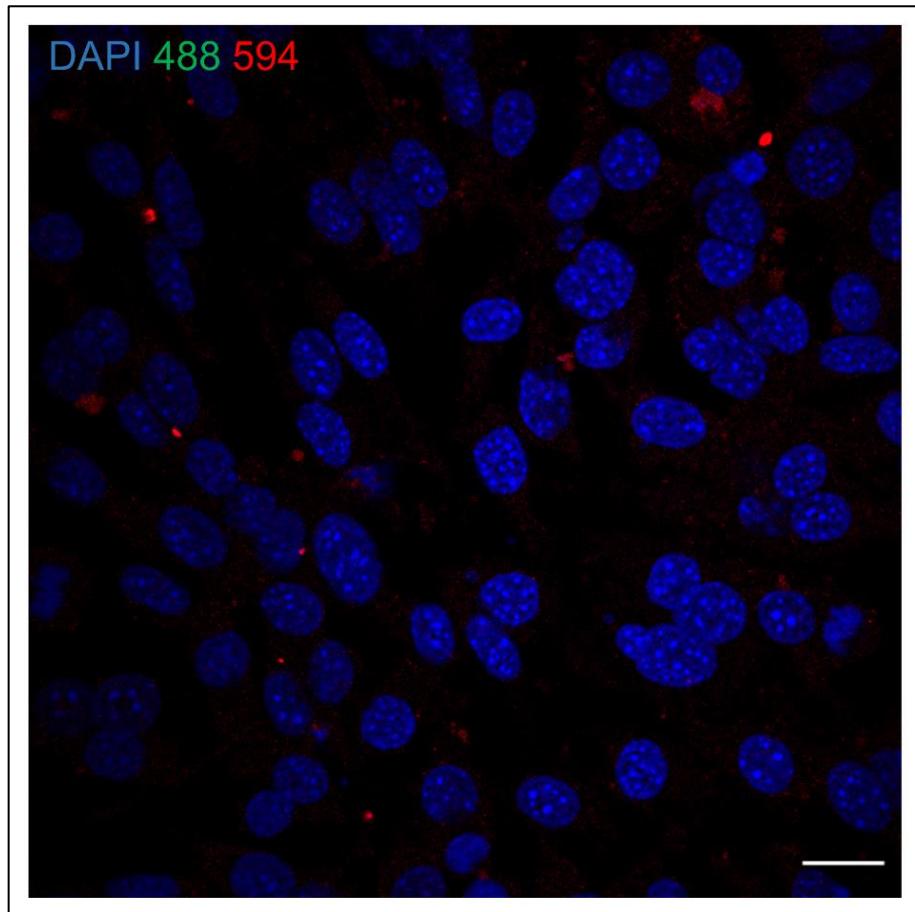


Figure 4.24. Immunofluorescence microscopy in mouse Shh-LIGHT2 cells. It is observed how the cells are not stained in the absence of primary antibodies (left image, A). The corresponding secondary antibodies were used: anti-rabbit 488 and anti-mouse 594. DAPI, a nuclear marker is shown in blue. Shh-LIGHT2 cells were serum starved for 24 hours. Nikon (A1) confocal inverted microscopy was used for visualisation (image taken by Dr Elisa Molinari, Newcastle University, UK). Scale bar: 20 μ m.

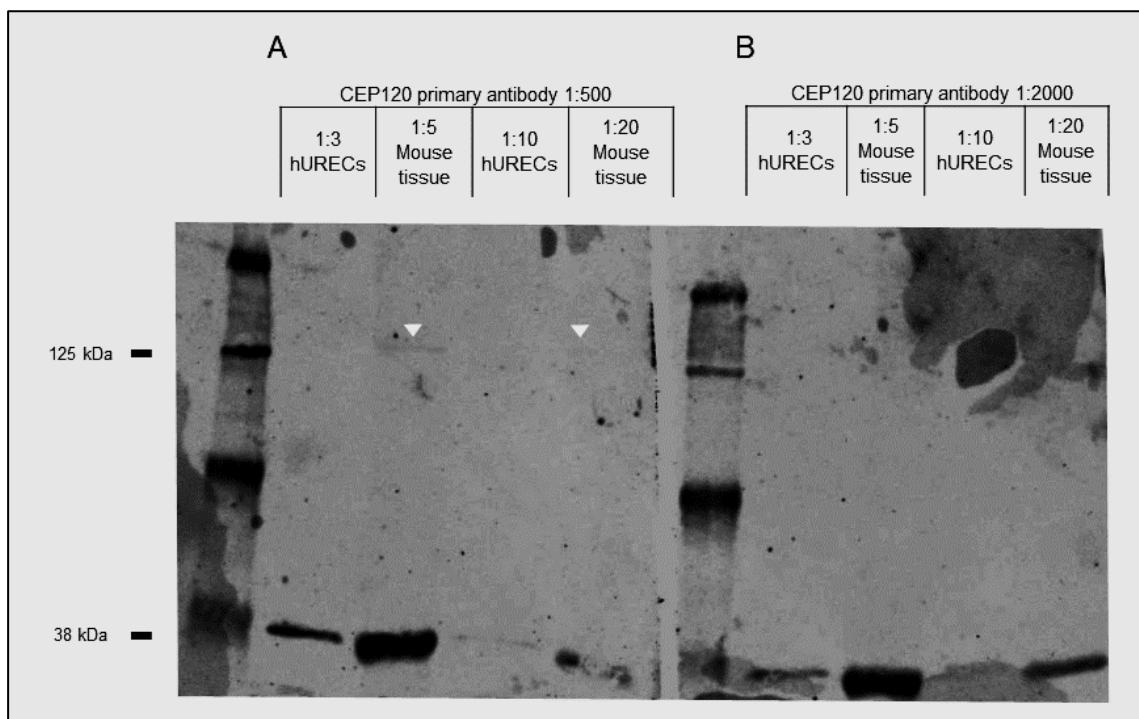


Figure 4.25. Western blotting showing Cep120 in protein lysates from mouse kidney tissue. Western blot shows the presence of Cep120 (~110 kDa) in protein lysates from mice kidney tissue. Left panel (A), shows the four Western blot lanes in which protein extracts were stained with CEP120 primary antibody diluted 1:500, while the right panel (B) shows the four Western blot lanes in which protein extracts were stained with CEP120 primary antibody diluted 1:2000. Cep120 is observed in the second and fourth lanes in panel A corresponding to mouse tissue (see white arrows). In contrast CEP120 is not observed in protein lysates from human urine-derived renal epithelial cells (hURECs). The rat polyclonal primary CEP120 antibody used is courtesy from Dr Moe R. Mahjoub (Washington University, MO, USA). CEP120 primary antibody was used at the dilutions of 1:500 (panel A) and 1:2000 (panel B). Different protein concentrations were loaded as a stock of protein lysate from mice kidney tissue was diluted 1:5 and 1:20 (see second and fourth lanes of panel A and B), and protein lysates from hURECs were diluted 1:3 and 1:10 (see first and third lanes of panel A and B). The nitrocellulose membrane was visualised using an Odyssey CLx imaging system (LI-COR). GAPDH (~37 KDa) was used as a loading control. Chameleon Duo Pre-stained Protein Ladder (928-60000; LI-COR) was loaded into two lanes of the gel that are shown to the left of each panel.

4.2.9 Effect on Hedgehog signalling upon Cep120 silencing on Shh-LIGHT2 cells

I first performed the knockdown of *Cep120* using a *Cep120* siRNA in Shh-LIGHT2 cells. A reduction of more than 50% in *Cep120* expression was observed, comparing the cells treated with the *Cep120* siRNA against the cells treated with a Control siRNA (Ctrl siRNA) (Figure 4.26).

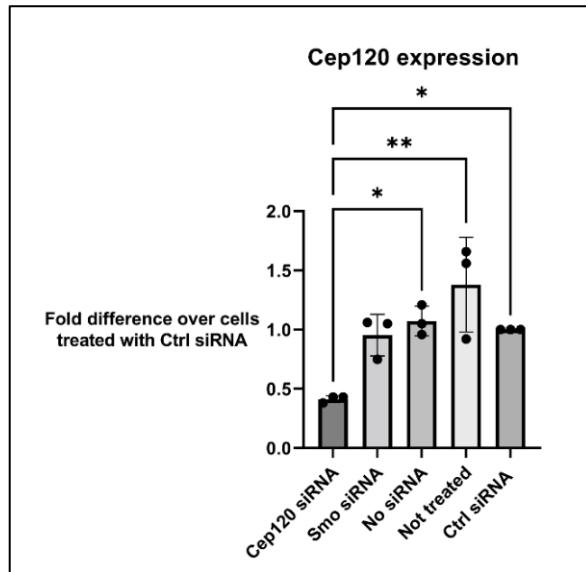


Figure 4.26. *Cep120* expression upon siRNA knockdown in Shh-LIGHT2. *Cep120* expression was normalised over the sample in which a Ctrl (Control) siRNA was used. It is observed a significant decrease in *Cep120* expression upon treatment with *Cep120* siRNA. Mouse Shh-LIGHT2 were serum starved before they were either treated (the same day) with *Cep120* siRNA, *Smo* siRNA, Ctrl siRNA, no siRNA (treated only with RNA delivery reagent, Lipofectamine RNAiMAX Transfection Reagent) or not treated (not treated with siRNA or RNA delivery reagent). Transcript levels of *Cep120* were determined by qRT-PCR. *Gapdh* and *Hprt* were used for normalisation. One-way ANOVA and Tukey's Multiple Comparison Test were used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are not indicated. Number of technical replicates = 3.

I investigated if the cells treated with the *Cep120* siRNA had a different expression of some components of the Hh signalling pathway: *Smo*, *Gli* and *Patch1*. It was observed that gene expression of these genes is not significantly different in the cells treated with the *Cep120* siRNA, compared to the cells treated with the Ctrl siRNA or the rest of the negative controls I have made: No siRNA (cells treated only with RNA delivery reagent) and Not treated (not treated with siRNA or RNA delivery reagent) (Figures 4.27-4.29).

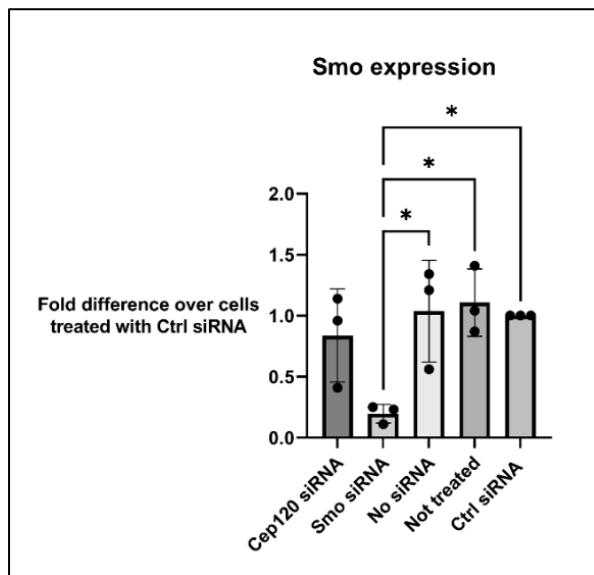


Figure 4.27. *Smo* expression upon siRNA knockdown in Shh-LIGHT2. *Smo* expression was normalised over the sample in which a Ctrl (Control) siRNA was used. It is observed a significant decrease in *Smo* expression upon treatment with *Smo* siRNA. In contrast, there is not a significant decrease in *Smo* expression upon treatment with *Cep120* siRNA. Mouse Shh-LIGHT2 were serum starved before they were either treated (the same day) with *Cep120* siRNA, *Smo* siRNA, Ctrl siRNA, no siRNA (treated only with RNA delivery reagent, Lipofectamine RNAiMAX Transfection Reagent) or not treated (not treated with siRNA or RNA delivery reagent). Transcript levels of *Smo* were determined by qRT-PCR. *Gapdh* and *Hprt* were used for normalisation. One-way ANOVA and Tukey's Multiple Comparison Test were used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are not indicated. Number of technical replicates = 3.

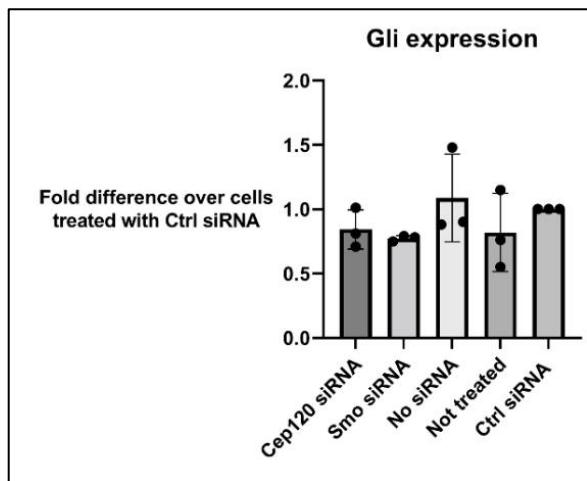


Figure 4.28. *Gli* expression upon siRNA knockdown in Shh-LIGHT2. *Gli* expression was normalised over the sample in which a Ctrl (Control) siRNA was used. It is observed that there is not a significant decrease in *Gli* expression upon treatment with *Cep120* siRNA or *Smo* siRNA. Mouse Shh-LIGHT2 were serum starved before they were either treated (the same day) with *Cep120* siRNA, *Smo* siRNA, Ctrl siRNA, no siRNA (treated only with RNA delivery reagent, Lipofectamine RNAiMAX Transfection Reagent) or not treated (not treated with siRNA or RNA delivery reagent). Transcript levels of *Gli* were determined by qRT-PCR. *Gapdh* and *Hprt* were used for normalisation. One-way ANOVA and Tukey's Multiple Comparison Test were used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are not indicated. Number of technical replicates = 3.

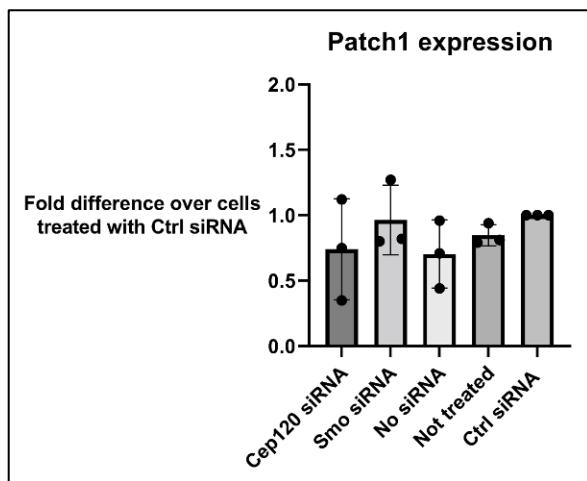


Figure 4.29. *Patch1* expression upon siRNA knockdown in Shh-LIGHT2. *Patch1* expression was normalised over the sample in which a Ctrl (Control) siRNA was used. It is observed that there is not a significant decrease in *Patch1* expression upon treatment with *Cep120* siRNA or *Smo* siRNA. Mouse Shh-LIGHT2 were serum starved before they were either treated (the same day) with *Cep120* siRNA, *Smo* siRNA, Ctrl siRNA, no siRNA (treated only with RNA delivery reagent, Lipofectamine RNAiMAX Transfection Reagent) or not treated (not treated with siRNA or RNA delivery reagent). Transcript levels of *Patch1* were determined by qRT-PCR. *Gapdh* and *Hprt* were used for normalisation. One-way ANOVA and Tukey's Multiple Comparison Test were used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are not indicated. Number of technical replicates = 3.

Smo knockdown is expected to affect Hh signalling. It was observed a reduction of more than 80% in *Smo* expression comparing the cells treated with the *Smo* siRNA against the cells treated with Ctrl siRNA (ie: *Smo* expression was only 20% in cells treated with the *Smo* siRNA) (Figure 4.27). It was observed that gene expression levels of other components of the Hh signalling pathway: *Gli* and *Patch1*, are not significantly different in the cells treated with the *Smo* siRNA, compared to the cells treated with the Ctrl siRNA or the rest of the negative controls No siRNA and Not treated (Figures 4.28 and 4.29).

It can be suggested that treatment with *Cep120* and *Smo* siRNA can efficiently knockdown *Cep120* and *Smo* expression respectively (Figures 4.26 and 4.27). I repeated the experiment and found similar results (Figures 4.30 and 4.31). In this repeated experiment, additionally, some of the cells were stimulated with SAG, an Hh signalling agonist, after being treated with the corresponding siRNA or control treatment. I analysed the SAG simulated cells against unstimulated cells (cells treated with DMSO) using the Promega Dual Luciferase Reporter Assay (Figure 4.32) and it was found that SAG stimulation is partially impaired in the cells treated with *Smo* siRNA.

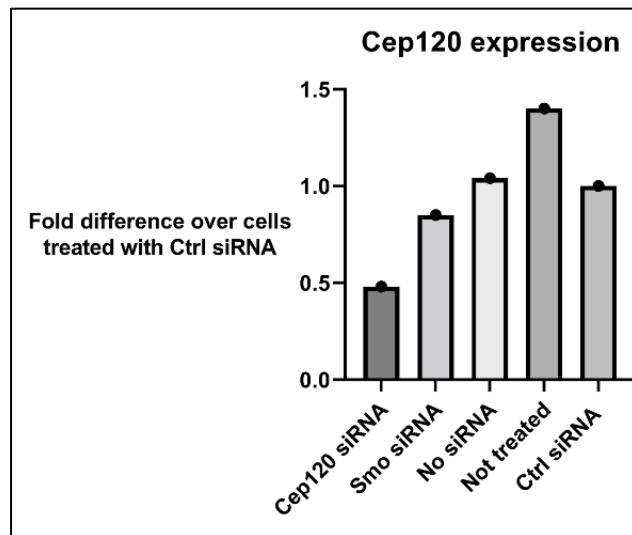


Figure 4.30. *Cep120* expression upon siRNA knockdown in Shh-LIGHT2. *Cep120* expression was normalised over the sample in which a Ctrl (Control) siRNA was used. It is observed a significant decrease in *Cep120* expression upon treatment with *Cep120* siRNA. Mouse Shh-LIGHT2 were serum starved before they were either treated (the same day) with *Cep120* siRNA, *Smo* siRNA, Ctrl siRNA, no siRNA (treated only with RNA delivery reagent, Lipofectamine RNAiMAX Transfection Reagent) or not treated (not treated with siRNA or RNA delivery reagent). Transcript levels of *Cep120* were determined by qRT-PCR. *Gapdh* and *Hprt* were used for normalisation. No statistical test was performed as the number of technical replicates = 1.

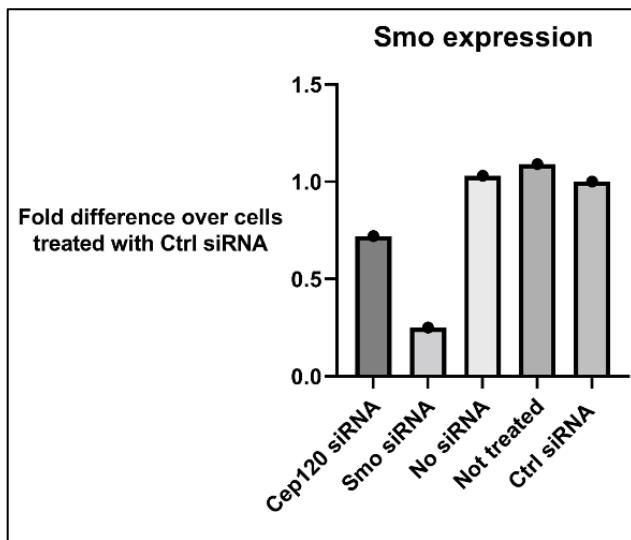


Figure 4.31. *Smo* expression upon siRNA knockdown in Shh-LIGHT2. *Smo* expression was normalised over the sample in which a Ctrl (Control) siRNA was used. It is observed a significant decrease in *Smo* expression upon treatment with *Smo* siRNA. In contrast, there is not a significant decrease in *Smo* expression upon treatment with *Cep120* siRNA. Mouse Shh-LIGHT2 were serum starved before they were either treated (the same day) with *Cep120* siRNA, *Smo* siRNA, Ctrl siRNA, no siRNA (treated only with RNA delivery reagent, Lipofectamine RNAiMAX Transfection Reagent) or not treated (not treated with siRNA or RNA delivery reagent). Transcript levels of *Smo* were determined by qRT-PCR. *Gapdh* and *Hprt* were used for normalisation. No statistical test was performed as the number of technical replicates = 1.

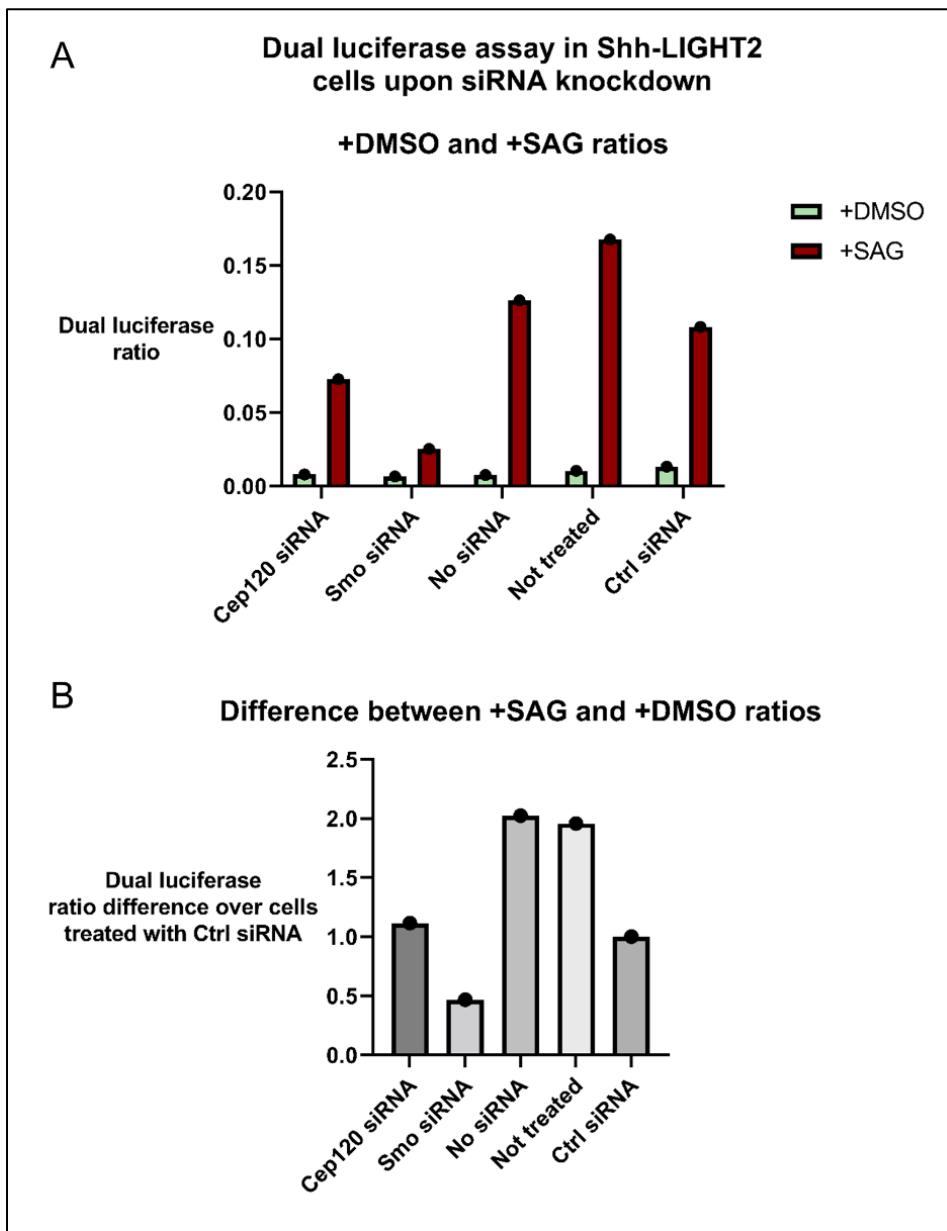


Figure 4.32. Luciferase assay reveals SAG stimulation impairment upon *Smo* siRNA knockdown in Shh-LIGHT2 cells. The top panel (A) shows stimulation with SAG, a Hedgehog signalling agonist, in all the samples. It can be observed that this stimulation is the lowest in the sample treated with *Smo* siRNA. The bottom panel (B), for each of the samples, shows the difference in the dual luciferase ratios (shown in panel A) between samples treated with SAG against the samples treated DMSO (negative control) is normalised against the sample in which no siRNA is added. Mouse Shh-LIGHT2 were serum starved before they were either treated (the same day) with *Cep120* siRNA, *Smo* siRNA, Ctrl siRNA, no siRNA (treated only with RNA delivery reagent, Lipofectamine RNAiMAX Transfection Reagent) or not treated (not treated with siRNA or RNA delivery reagent). Each sample was treated with SAG or DMSO one day after siRNA treatment, lysed after 2 days from SAG or DMSO treatment, stored and quantified with a GLOMAX Multi Detection System Luminometer following Promega Dual-Luciferase Reporter Assay System standard protocol. No statistical test was performed as there are not technical or biological replicates.

When Shh-LIGHT2 cells were stimulated with SAG, it was observed an increase on the dual luciferase ratio in cells treated with siRNA and controls (Figure 4.32A). This increase in dual luciferase ratio was just about 1:2 in the cells treated with *Smo* siRNA compared with the cells treated with Ctrl siRNA (Figure 4.32B). This indicates that knocking down *Smo* in Shh-LIGHT2 cells partially impairs the stimulation of the Hh signalling pathway via SAG. This impairment in the stimulation of the Hh signalling is not observed in the cells treated with *Cep120* siRNA compared with the cells treated with Ctrl siRNA (Figure 4.32B).

However, the limitation of this experiment is that I was able to do only one technical and only biological replicate, this prevents me from establishing any conclusion from this experiment. It can only be suggested that, as expected, knocking down *Smo* in Shh-LIGHT2 has an effect on the activation of Hh signalling pathway.

I did not see any effect on Hh signalling via qPCR at RNA level when *Cep120* was knocked down. This also suggests that: to observe any effect, if any, on the stimulation of Hh signalling (via qPCR of dual luciferase assay), it is necessary to increase the efficiency of the knockdown and to stimulate the Hh signalling pathway (for example with SAG) in the Shh-LIGHT2 cells. Besides, the use of Shh-LIGHT2 cells allowed to observe differences on the stimulation of Hh signalling, via a dual luciferase assay, however this cannot be confirmed unless the appropriate (at least three) technical and biological replicates are performed.

Additionally, using an ASO morpholino, it can be suggested that doing a *CEP120* knockdown in primary human fibroblasts may be also possible and effective (*CEP120* expression was about 30% in cells treated with the *CEP120* ASO morpholino) (Figure 4.33), however this cannot be confirmed unless three technical and biological replicates are performed.

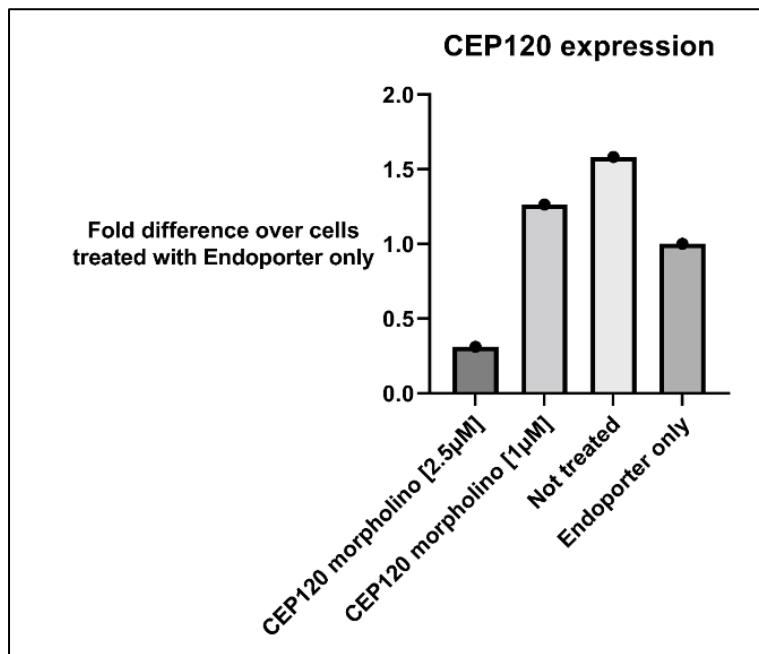


Figure 4.33. *CEP120* expression upon morpholino knockdown in human fibroblasts. *CEP120* expression was normalised over the sample in which only Endoporter was used. It is observed a significant decrease in *CEP120* expression upon treatment with *CEP120* antisense oligo morpholino (Gene Tools). Primary control human fibroblasts were treated with *CEP120* morpholino (of a concentration of [2.5μM] and [1μM] or Endo-Porter only, (treated only with morpholino delivery reagent, Endo-Porter, Gene Tools) or not treated (not treated with morpholino or morpholino delivery reagent). I suggest that only *CEP120* morpholino at a concentration of [2.5μM] was effective to knockdown *CEP120* expression. Transcript levels of *CEP120* were determined by qRT-PCR. *GAPDH*, *HPRT1* and *GUSB* were used for normalisation. No statistical test was performed as there are not technical or biological replicates.

For the experiments consisting in dual luciferase assay using Shh-LIGHT2 cells and ASO morpholino in primary human fibroblasts, there are not technical or biological replicates, this prevents me from establishing any conclusion about this experiment.

Efficiency of the *Cep120* siRNA knockdown has been previously shown by Betleja et al., observing a decrease of 99.7% of Cep120 protein upon *Cep120* siRNA knockdown in MEFs (using same *Cep120* siRNA from Dharmacon), comparing the MEFs treated with the *Cep120* siRNA against the cells treated with a Control siRNA (negative control), via Western blotting. They used two or three independent experiments for the corresponding functional analyses (Betleja et al., 2018).

Regarding the *Smo* siRNA I used, it was previously used and found to be effective at Newcastle University by other members of Sayer and Miles Laboratory (data not available).

I have performed three technical replicates (except for experiment shown in Figure 3.30 and 3.31) of the knockdown of *Cep120* and *Smo* in Shh-LIGHT2. Three technical replicates were considered appropriate for the propose of performing a preliminary study to investigate the efficiency of these siRNAs (Figures 3.26-3.29).

However, I cannot establish any conclusion from the results corresponding to the siRNA knockdown studies (Figures 4.26-4.31), as biological replicates have not been performed. This represents the first step of an approach to study the effect of a *Cep120* siRNA knockdown in Shh-LIGHT2 cells.

If I had the time and available resources to continue with the *Cep120* knockdown study, I would have continued with the experimental approach shown in this section, performing three biological replicates, as well as three technical replicates. I would have completed the dual luciferase assay by doing three technical replicates and three biological replicates.

If I had completed the siRNA knockdown study, it would have provided more details regarding the efficiency of the *Cep120* and *Smo* siRNA knockdown experiments and confirmed the effect of these siRNA knockdown experiments on Hh signalling via a dual luciferase assay and via RNA expression analyses using qPCR.

I suggest this preliminary study using Shh-LIGHT2 cells and siRNA knockdown of *Cep120* can be used to exemplify how I would perform a siRNA knockdown study. In future analyses I would perform functional analyses using human primary cells such as hURECs and primary fibroblasts to study patient specific *CEP120* mutations (pipeline described in Figure 2.9).

The siRNA knockdown experiments in Shh-LIGHT2 cells, the ASO morpholino knockdown experiment in primary human fibroblasts and the dual luciferase assay in Shh-LIGHT2 cells, cannot be used to establish any conclusion. However, these experiments can be used to exemplify the pipeline that I would follow if I had done appropriate number of biological and technical replicates.

In summary, although conclusions cannot be established from these results, I can suggest the pipeline and protocols I would have followed to confirm the *CEP120* morpholino knockdown efficiency and *Cep120* siRNA knockdown effect on Hh signalling. I would have continued this approach by determining the effect on Hh signalling upon *Cep120* siRNA knockdown in Shh-LIGHT2 cells. I would do this via

dual luciferase assay and analysis of the effect on the RNA expression of Hh signalling pathway components (e.g. *Patch*, *Gli*, *Smo*) via qPCR, before and after of SAG expression. Moreover, I would analyse the effect of the *CEP120* knockdown via a *CEP120* ASO morpholino in human primary fibroblasts. For each of these experiments I would perform three technical replicates and three biological replicates in each of the experiments.

The work regarding the siRNA knockdown experiments shown in this chapter are only preliminary experiments. I was not able to continue with these knockdown experiments. In future analyses I would validate these *Cep120* siRNA experiments on Shh-LIGHT2 cells and as a comparison I would validate *CEP120* ASO morpholino knockdown on primary human cells.

I cannot establish any conclusion, including conclusions regarding the role of *Cep120* in Hh signalling, however this preliminary work exemplifies how this study would be approached and analysed in the future, if I had the resources and time available to continue with this approach, to optimise plasmid transfection efficiency and to perform the experiments mentioned above.

4.2.10. Site directed mutagenesis of WT CEP120 plasmid and optimisation of transfection efficiency

Restriction mapping and Sanger sequencing results of the *CEP120* Mutant plasmids are shown in Figures 4.34 and 4.35.

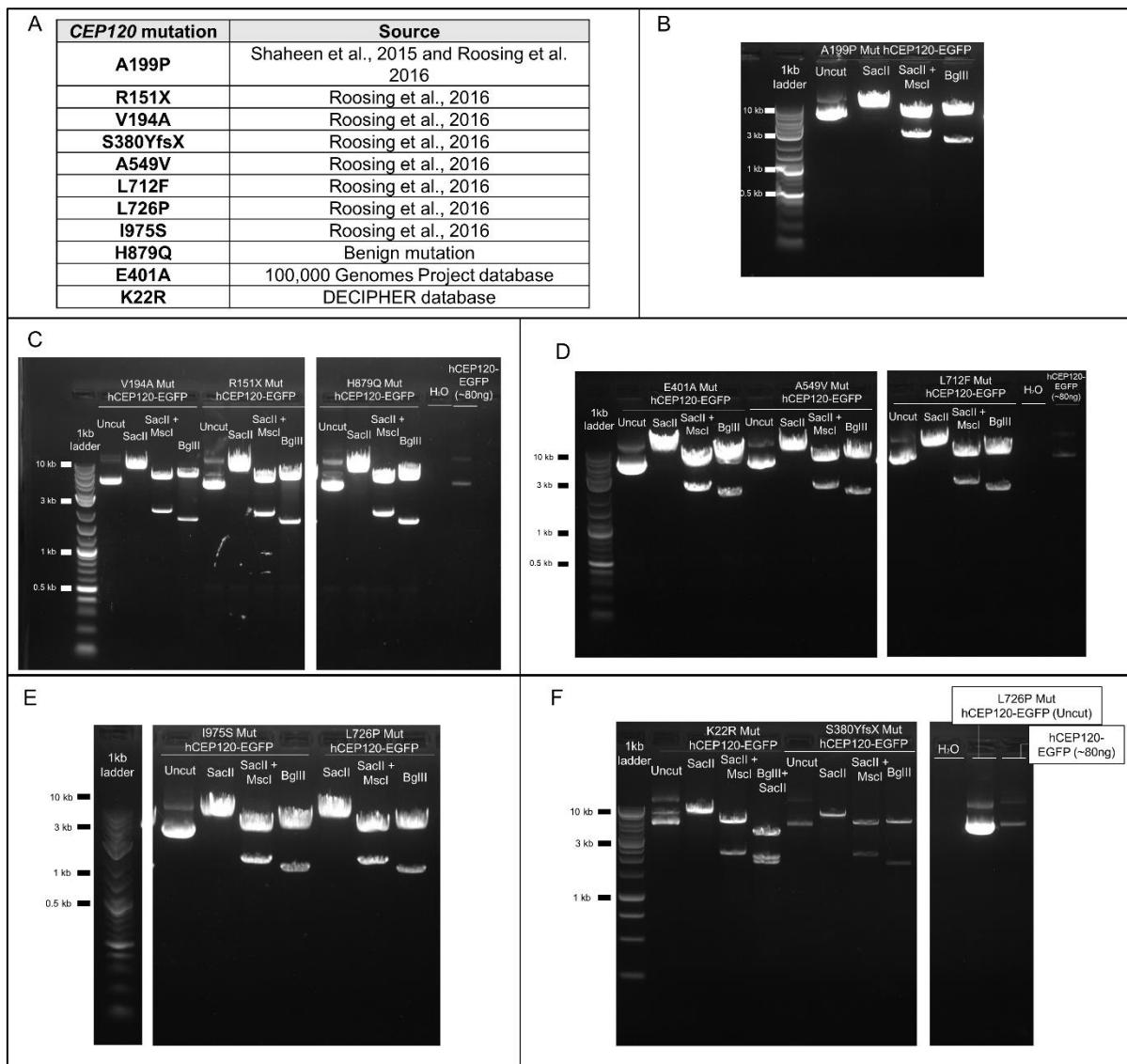


Figure 4.34. List of the mutated plasmids from the *CEP120* plasmid (Mut hCEP120-EGFP) restriction mapping. The top left image (A) shows each of the mutations in each of the Mut hCEP120-EGFP plasmids. Each of these mutations was created via site-directed mutagenesis (New England BioLabs) using the WT *CEP120* plasmid (hCEP120-EGFP). A total of 12 Mut hCEP120-EGFP plasmids were created: 9 of them present a patient-specific mutation already described in the literature. One of them present a H879Q mutation (assumed to be benign) which is present in hCEP120-EGFP compared with the reference *CEP120* cDNA sequence (transcript: NM_153223.4). One of them presents the E401A mutation, which is a putative disease-causing variant found in the 100,000 Genomes Project database and one of them presents the K22R mutation, which is a putative disease-causing variant found in the DECIPHER database. B-F: Gel electrophoresis showing each of the Mut hCEP120-EGFP plasmids treated with different combinations of restriction enzymes. The plasmids presenting the corresponding mutations are shown in B-F. The expected fragments are observed, according to the digestion of the each of the Mut hCEP120-EGFP plasmids with specific combinations of restriction enzymes. When treated with SacII, in each of the Mut hCEP120-EGFP plasmids, the fragment of 7705 bp (linearized plasmid of size 7705 bp) is observed (except in the S380YfsX mutant plasmid, in which the expected size is 7706 bp as the mutation is an insertion: c.1138_1139insA; p.(Ser380TyrfsTer19), however this

difference cannot be noticed by gel electrophoresis). When the plasmid is treated with SacII and Mscl, the corresponding two fragments of length 2191 and 5514 bp are observed. And when the plasmid is treated with BgIII, the corresponding two fragments of length: 1811 and 5194 bp (B-E) are observed. When the plasmid is treated with BgIII + SacII, the corresponding three fragments of length: 1200, 1811 and 4694 bp (F) are observed. Uncut corresponds to each of the Mut hCEP120-EGFP plasmids not treated with any restriction enzyme. As a complementary to the ladder, I used a size reference control ~80ng of WT CEP120 plasmid in C, D and F. A negative control using the same water used to dilute the plasmids was used to detect any plasmid contamination in the water C, D and F. For clarity, some lanes were removed in C-F. Gel electrophoresis (1% agarose). 1kb ladder used.

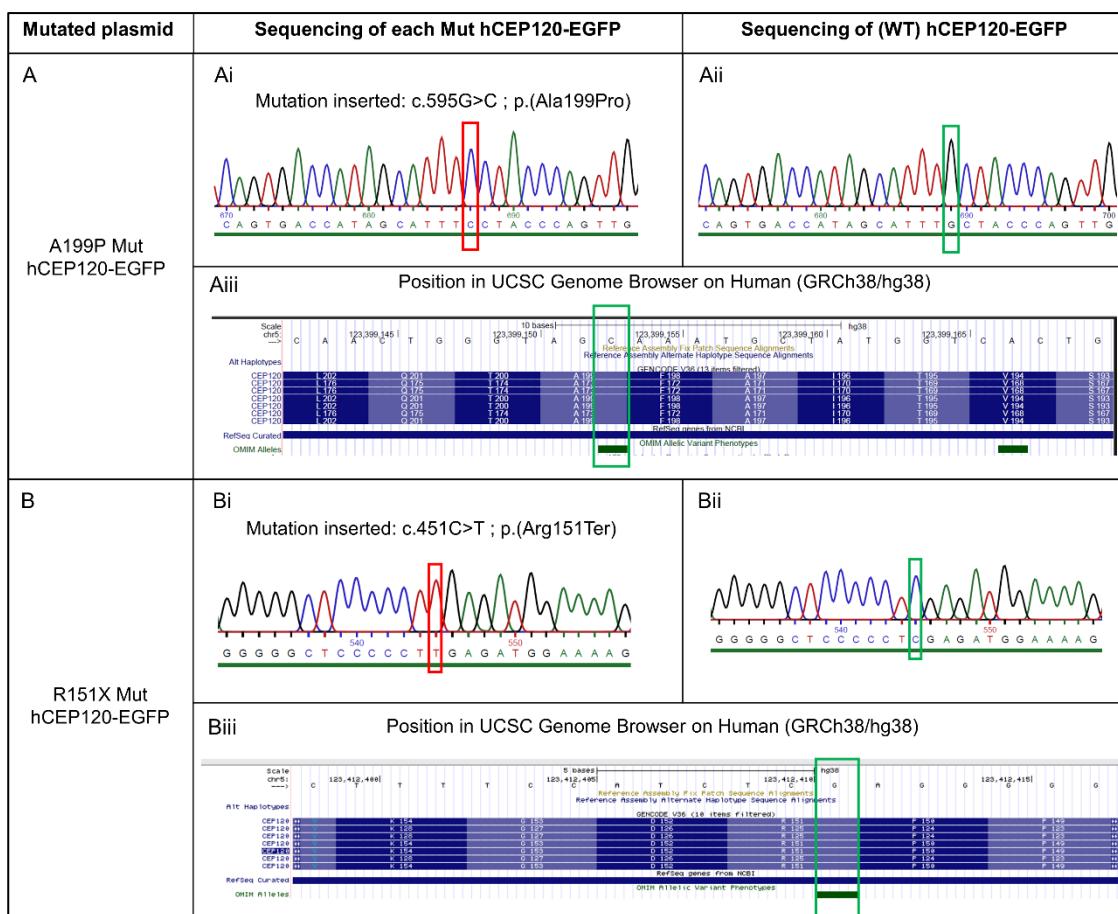


Figure continues in next page

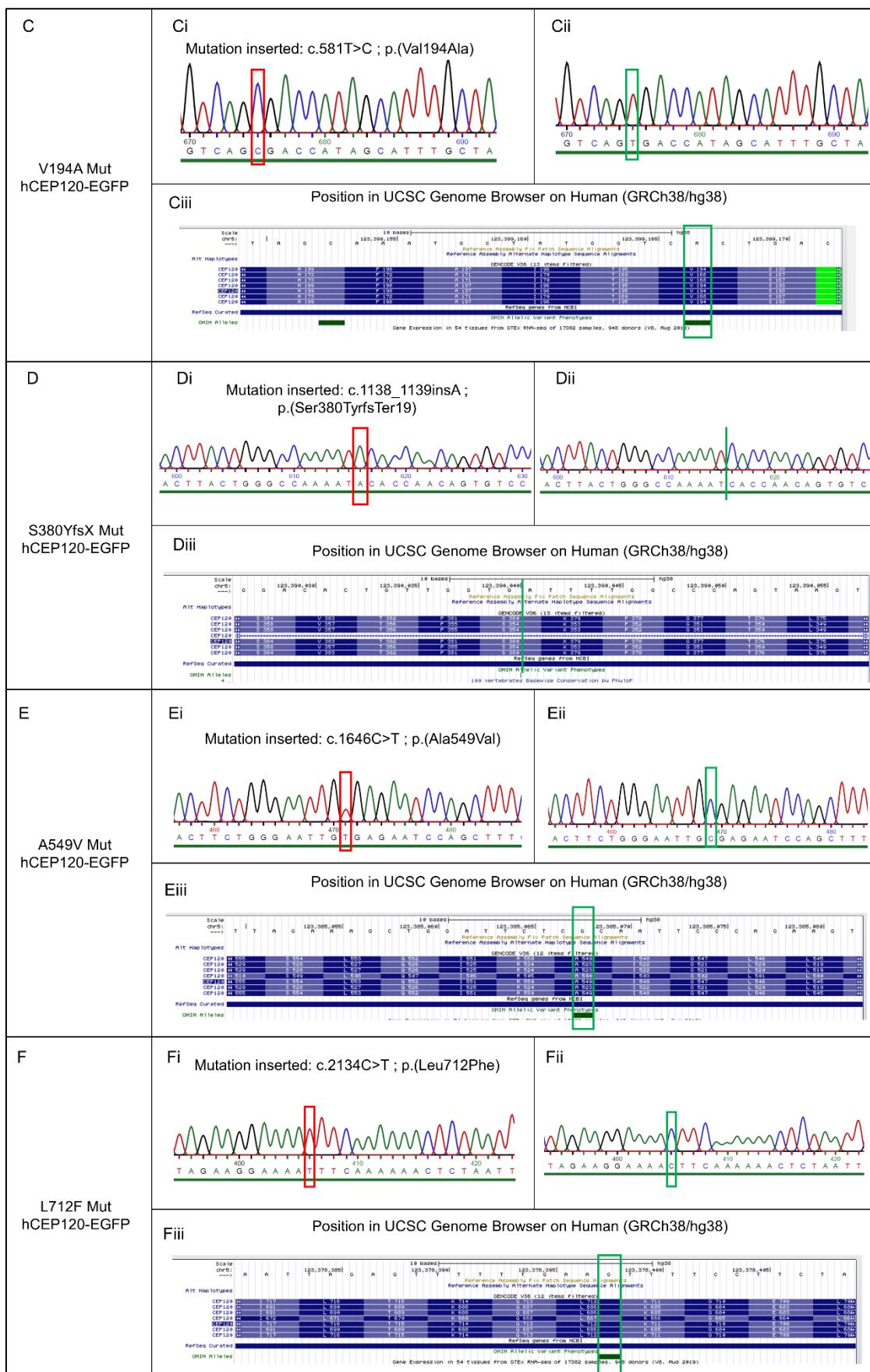


Figure continues in next page

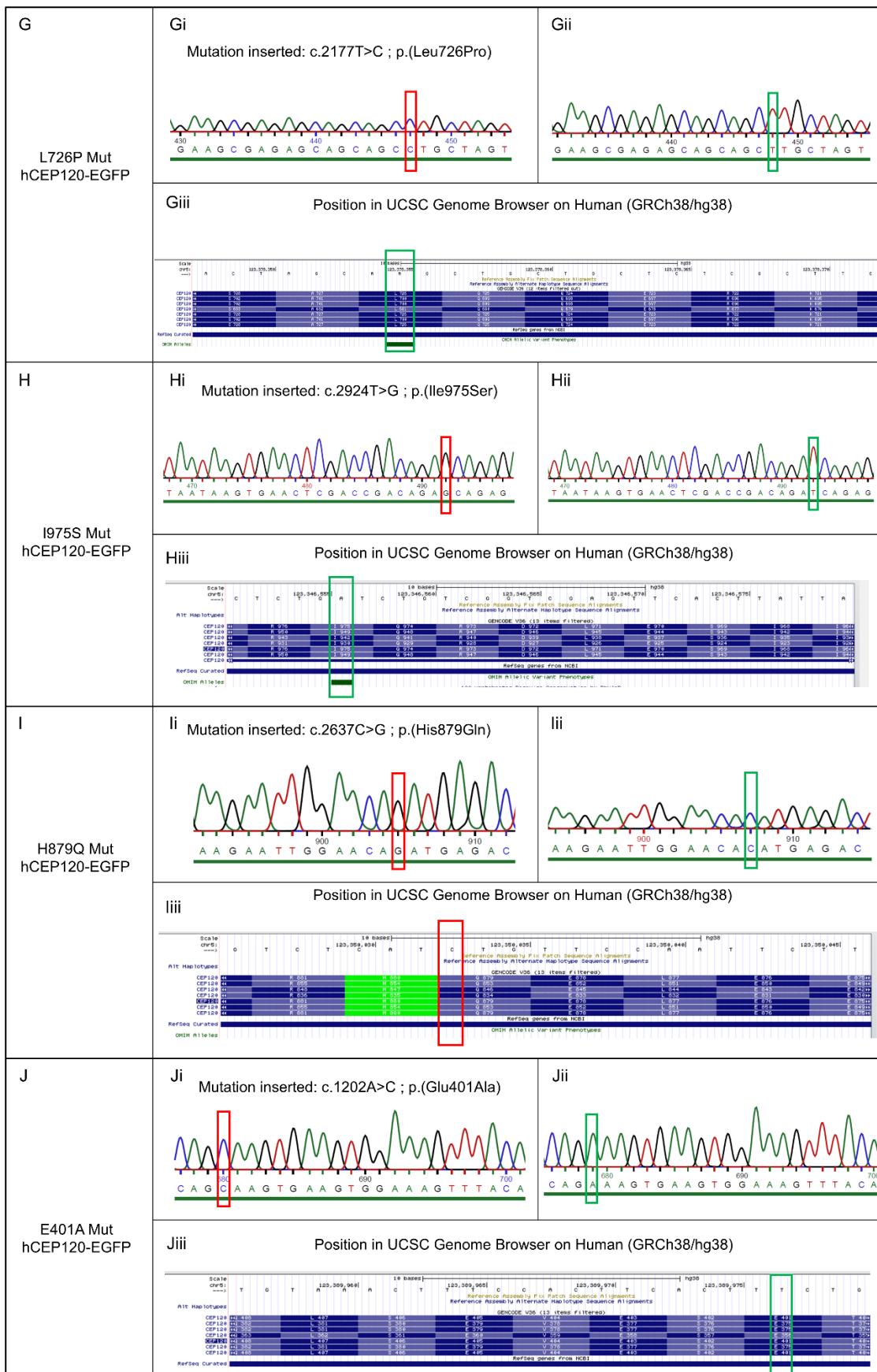


Figure continues in next page

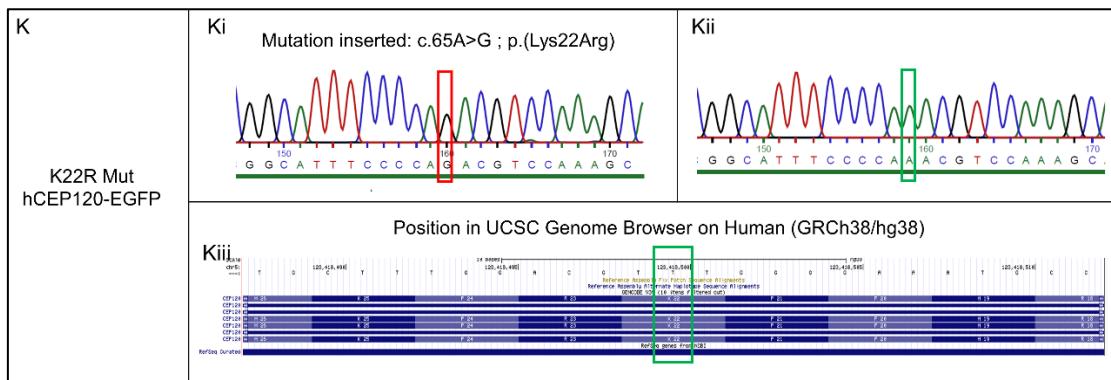


Figure 4.35. Sequencing of the each of the Mut hCEP120-EGFP plasmids. Each of the Mut hCEP120-EGFP plasmids contains a patient-specific mutation (Ai-Ki). Each of these mutations was created via site-directed mutagenesis (New England BioLabs) using the WT *CEP120* plasmid (hCEP120-EGFP). A total of 12 Mut hCEP120-EGFP plasmids were created: 9 of them present a patient-specific mutation already described in the literature (Ai-Hi). One of them presents the E401A mutation (Ji), which is a putative disease-causing variant found in the 100,000 Genomes Project database and one of them presents the K22R mutation (Ki), which is a putative disease-causing variant found in the DECIPHER database. And as a negative control, one plasmid presents a H879Q mutation (li) (assumed to be benign) that reverses a mutation (benign) which is present in hCEP120-EGFP compared with the reference *CEP120* cDNA sequence (transcript: NM_153223.4). The sequencing of the WT *CEP120* plasmid (hCEP120-EGFP) plasmid is shown on the right panels for comparison (Aii-Kii). Screenshots were taken from the UCSC Genome Browser on Human (GRCh38/hg38) to show the position in which each of the mutation is located (Aiii-Kiii). Red rectangles (in left panels: Ai-Ki) show the mutated positions in each of the Mut hCEP120-EGFP plasmids. Green rectangles shows the WT position in the WT *CEP120* plasmid (Aii-Cii and Eii-Kii) and in UCSC Genome Browser on Human (GRCh38/hg38) (Aiii-Ciii and Eiii-Kiii), except in the case of the H879Q mutation in which a red rectangle is shown, given that the position in the UCSC Genome Browser on Human (GRCh38/hg38) matches the sequence of the H879Q hCEP120-EGFP plasmid, rather than the WT *CEP120* plasmid. A green bar in WT *CEP120* plasmid sequencing and the UCSC Genome Browser on Human (GRCh38/hg38), to compare with the sequencing for the S380YfsX Mut hCEP120-EGFP plasmid, represents the position between two bases in which an insertion was introduced: c.1138_1139insA (Di compared with Dii and Diii).

Transfection efficiency into mammalian cells using an EGFP plasmid was assessed. I used Shh-LIGHT2 cells and observed that the transfection of these cells with the EGFP plasmid is possible using the Xfect transfection reagent, as I observed more than 50 cells transfected (GFP positive cells) *in vivo* after two days of transfection (Figures 4.36 and 4.37). Transfected cells were almost absent (no more than 5 GFP positive cells were observed) in Shh-LIGHT2 cells transfected with other transfection reagents or transfecting the pIRE2-EGFP plasmid after one or two days of transfection (instead of the EGFP plasmid). Besides, the WT *CEP120* plasmid was

transfected into Shh-LIGHT2 cells but I did not observe any CEP120-EGFP positive cells.

Additionally, I transfected control human primary fibroblasts with the EGFP plasmid using the Xfect transfection reagent and observed only one GFP positive cell *in vivo* (Figure 4.38).

It is important to note that in the literature there are examples of successful plasmid transfection of similar plasmids in primary cells (Meka et al., 2022), and other cell types to transfect like RPE-1 or U2OS cells can be included in the assay (Tsai et al., 2019, Lin et al., 2013). Patient-specific variants were successfully inserted in the *CEP120* WT cDNA sequence via site directed mutagenesis. However, transfection of the WT *CEP120* plasmid, or GFP plasmids as a test, into primary cells is challenging.

The efficiency of transfection of a GFP plasmid in Shh-LIGHT2 cells is low, and even lower in primary human fibroblasts. This prevented me to continue with the pipeline designed.

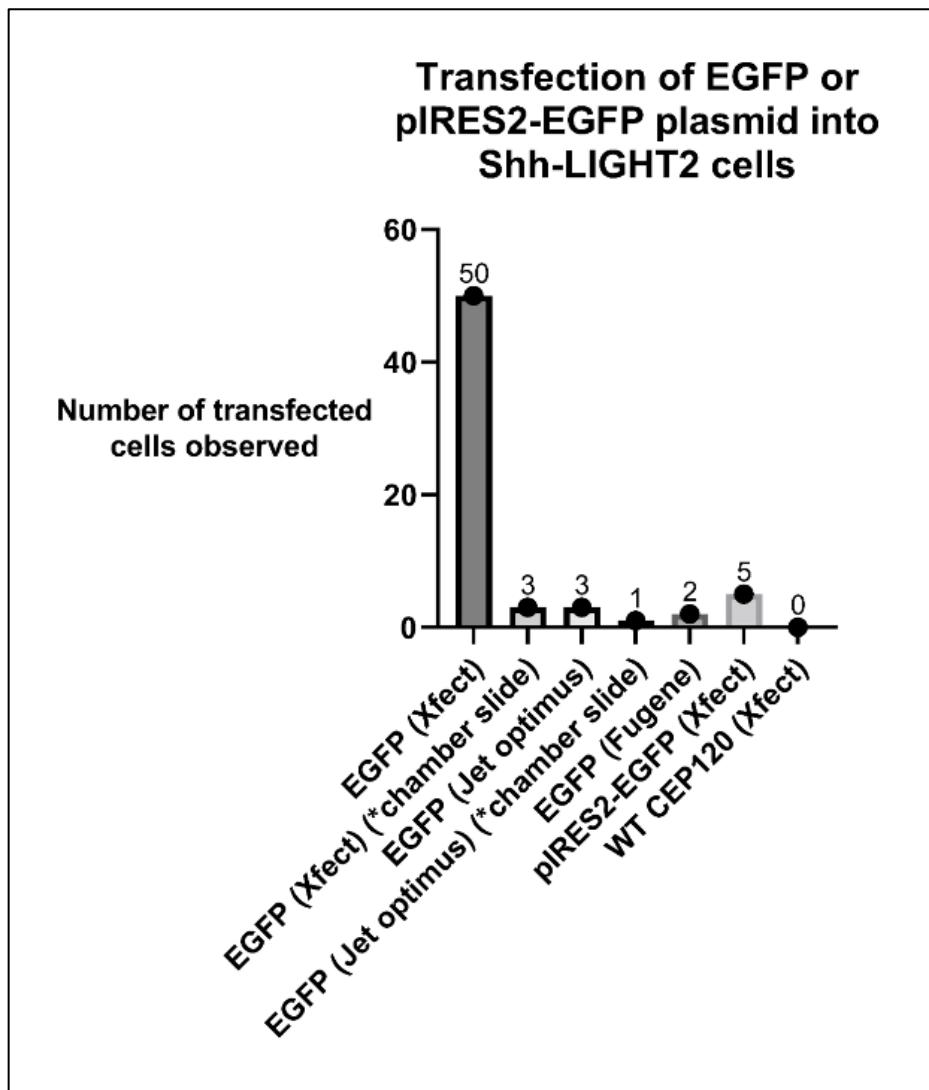


Figure 4.36. Transfection of EGFP plasmid or pIRES2-EGFP into Shh-LIGHT2 cells using different transfection reagents. Diagram showing the number of Shh-LIGHT2 cells transfected (EGFP positive cells). The plasmid transfected was: EGFP plasmid (courtesy of Dr Magomet Aushev, Newcastle University, UK) or pIRES2-EGFP plasmid (pIRES2-EGFP, Clontech # 6029-1, Addgene # 3178). The transfection reagents used were Xfect (Takara bio), JetOptimus (Polyplus) and Fugene HD (Promega). Cells were transfected plated and transfected in 12-well plates when they reached 70-90% confluence. 8-well Nunc chamber slides (Thermo Fisher) were used in two occasions (marked as “(*chamber slide)” in the x-axis label). Shh-LIGHT2 cells were visualised after 2 days from transfection. Zeiss AxioVert 200M inverted fluorescent microscope was used for visualisation. The number of transfected cells with the EGFP plasmid using the Xfect transfection reagent shown (in 12-well plates): 50, it is an approximation of the real number. As negative controls cells were also treated with each of the transfection reagent only (only treated with Xfect, JetOptimus or Fugene) without the addition of any plasmid; no GFP positive cells were observed as expected.

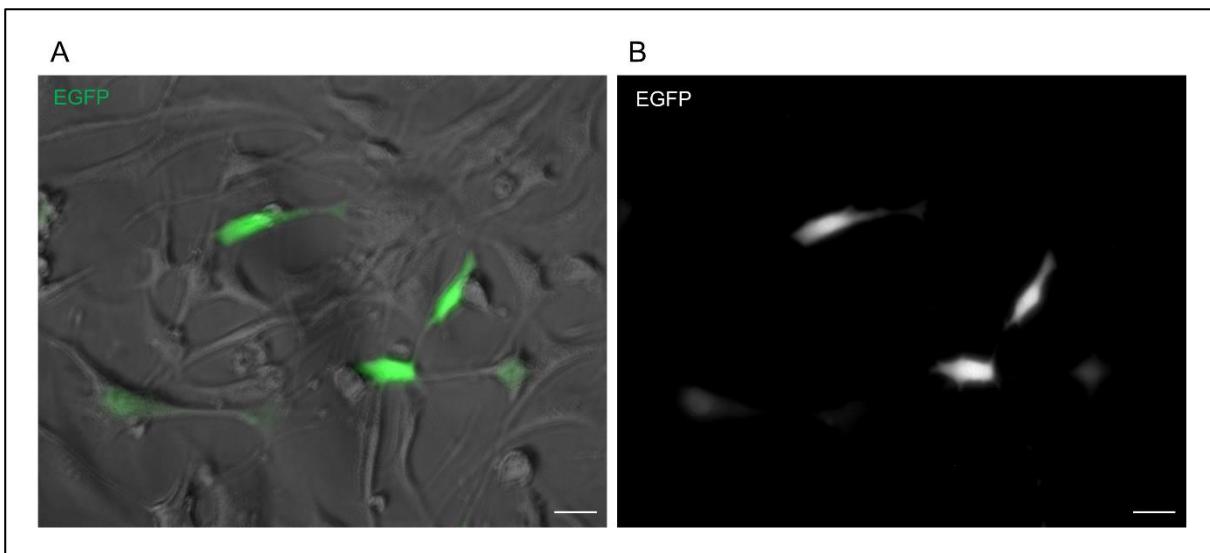


Figure 4.37. Transfection of EGFP plasmid into Shh-LIGHT2 cells. Five cells *in vivo* that have been transfected with the EGFP plasmid (courtesy of Dr Magomet Aushev, Newcastle University, UK) can be observed, as they are fluorescent, observed in green (using the Fluorescein (FITC) fluorescent filter and brightfield imaging) (A). Transfected cells (EGFP positive cells) are also shown in the left image (B) in white. Shh-LIGHT2 cells were transfected using Xfect transfection reagent and visualised after 2 days from transfection. Zeiss AxioVert 200M inverted fluorescent microscope was used for visualisation. Scale bar: 20 μ m.

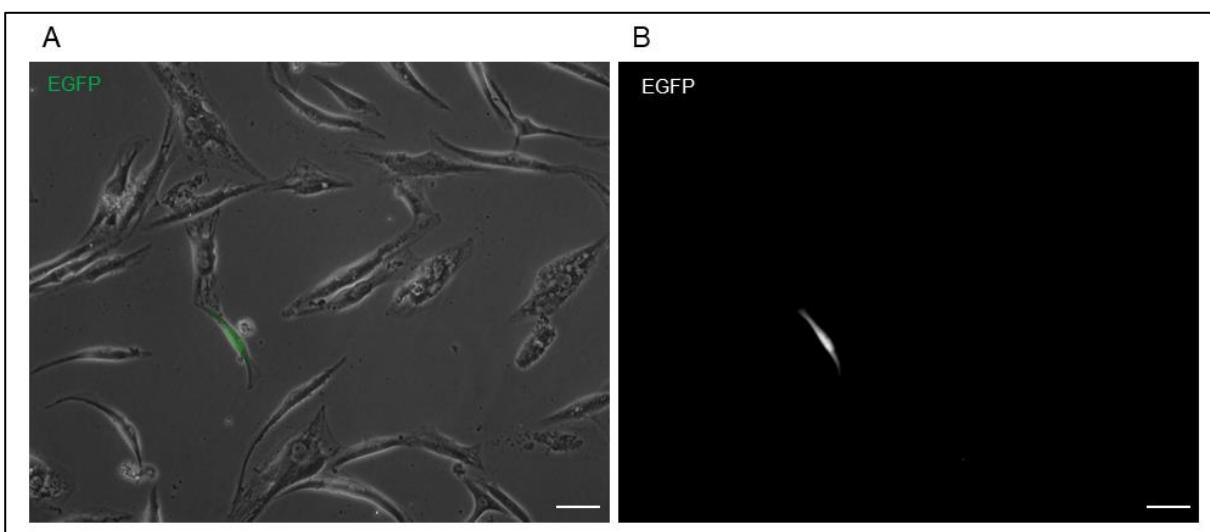


Figure 4.38. Transfection of EGFP plasmid into human fibroblasts. One cell *in vivo* that has been transfected with the EGFP plasmid (courtesy of Dr Magomet Aushev, Newcastle University, UK) can be observed, as it is fluorescent, observed in green (using the Fluorescein (FITC) fluorescent filter and brightfield imaging) (A). The transfected cell (EGFP positive cell) is also shown in the right image (B) in white. Primary control human fibroblasts were transfected using Xfect transfection reagent and visualised after 2 days from transfection. Zeiss AxioVert 200M inverted fluorescent microscope was used for visualisation. Scale bar: 20 μ m.

4.3. Discussion

4.3.1. *CEP120 and CC2D2A in silico analyses and its combination with ex vivo detection of basal exon skipping*

The approach to achieve the objectives of this chapter consisted in a combination of *in silico* analyses to study *CEP120* and *CC2D2A* genes in terms of genotype-phenotype correlations, gene expression and tissue-specific basal exon skipping and suggest potential candidates for targeted exon skipping therapies and also an *ex vivo* approach to validate basal exon skipping.

A database of *CC2D2A* and *CEP120* reported disease-causing variants was generated, which is a curated, updated and annotated database. To create this database, multiple open-access tools and the literature available were used. I also ensured that this database is annotated according to HGVS nomenclature, assessing any discrepancies found.

This database was used to establish genotype-phenotype and identify suitable exons for a potential targeted exon skipping therapy approach.

Regarding genotype-phenotype correlations in *CC2D2A*: I systematically analysed all reported patients with *CC2D2A* mutations and confirmed that biallelic truncating *CC2D2A* variants are associated with more severe phenotypes than biallelic *CC2D2A* non-truncating variants. This was previously suggested (Bachmann-Gagescu et al., 2012). In particular, it was previously suggested that patients with at least one missense *CC2D2A* mutation are more likely to develop JBTS. In contrast, patients with truncating *CC2D2A* variants are more likely to develop more severe phenotypes, including MKS and ML (Bachmann-Gagescu et al., 2012).

The analysis of the reported patients complements this perspective, as only one 1 JBTS patient, out of the 45 patients with JBTS presenting with biallelic variants *CC2D2A*, has biallelic truncating *CC2D2A* variants. In contrast to 21 MKS/ML patients, out of the 32 patients with MKS/ML presenting with biallelic variants *CC2D2A*. It is also observed that the phenotype of patients with biallelic truncating *CC2D2A* is more likely to include kidney disease than the patients with non-truncating *CC2D2A* variants.

The disease molecular mechanism behind this genetic pleiotropy may be associated with the amount of remaining functional protein. In Drivas et al. a combination of *in*

silico and *in vitro* approaches were performed to study the pleiotropy of *CEP290* and *CC2D2A* mutations. In Drivas et al. study it was suggested that the amount of expression of functional protein may be inversely correlated with the severity of the resulting clinical phenotype associated with these two genes. Additionally in this study, it was suggested that basal exon skipping may have a role in modulating the amount of functional protein in these *CEP290*-affected and *CC2D2A*-affected patients, as the pathogenic effect of truncating variants in skippable exons could be bypassed by basal exon skipping (Drivas et al., 2015).

Using the database generated based on the literature and open-access tools it can be suggested that only exons 14 and 15 in *CEP120* are potentially skippable without leading to a disruption of the protein reading frame and without affecting a *CEP120* functional domain, if any of these 2 exons (or both) are skipped. However, even though the skipping of *CEP120* exons 14 and/or 15 may lead to functional near-full length *CEP120* protein, there are not patients described in the literature with disease-causing *CEP120* variants located in these exons. Therefore, to my knowledge there are not patients that would benefit from a potential targeting exon skipping therapy for *CEP120*.

In the case of *CC2D2A*, there are more patients and variants described and it is a gene with more coding exons than *CEP120*. There are 15 *CC2D2A* exons that can be skippable, out of 38 coding exons. In the literature, there are 14 truncating *CC2D2A* variants which were found in patients diagnosed with JBTS and MKS/ML. These variants are located in 7 exons out of the 14 skippable exons. These 7 exons are potentially good candidates for a targeted exon skipping therapy. These 7 exons in which truncating *CC2D2A* variants have been reported, include exon 30, which is an exon in which some degree of basal exon skipping occurs the kidney (shown in Figure 4.5).

The ASOs-based exon skipping approach for *CC2D2A* or *CEP120* was not validated *in vitro*. This is because this study was limited to investigate and compare *in silico* the potential of two primary ciliopathy genes with overlapping phenotypes: *CEP120* and *CC2D2A*, as targets for an antisense-oligonucleotide mediated exon skipping therapy. This study was also limited to find skippable exons that are potentially good candidates for a targeted exon skipping therapy and to investigate *in silico* tissue specificity of exon skipping in these two genes and to validate *in vitro* tissue specific basal exon skipping of one of the *CC2D2A* exons (exon 30) in the kidney.

The principle of ASOs-mediated exon skipping in ciliopathy genes has already been validated in multiple laboratories (Garanto et al., 2016, Schellens et al., 2023, Cideciyan et al., 2023), including Sayer and Miles laboratory at Newcastle University (Molinari et al., 2019, Ramsbottom et al., 2018).

I did not have the allocated time that would be required to validate the ASOs-mediated exon skipping approach *in vitro*. If I had more time, I would have considered to determine if the deleterious effect of the truncating mutations located in the *CC2D2A* skippable exons, can be bypassed via ASOs-mediated exon skipping, I would have considered to use primary cells (e.g. primary fibroblasts or hURECs) from patients with disease-causing *CC2D2A* mutations in these skippable exons, assuming these cells could be available and delivered to Sayer and Miles laboratory.

If I had more time, I would have considered to perform additional *in silico* genetic studies, for instance I could also have searched within the Genomics England 100,000 Genomes Project, for additional patients that could potentially be genetically solved with biallelic *CC2D2A* disease-causing variants, with a focus on the patients with *CC2D2A* disease-causing variants in skippable exons.

Moreover, If I had more time, I would have considered to apply the *in silico* analyses I applied to *CEP120* and *CC2D2A*, to a small selection of ciliopathy genes (such as: *CPLANE1*, *AHI1* or *TMEM67*) (Parisi et al., 2007, Bachmann-Gagescu et al., 2015a) to find potentially skippable exons. I would have considered to perform this approach with a focus on the JBTS genes with the higher number of patients presenting mutations in skippable exons, so that the higher number of patients would benefit from a potential ASOs-mediated exon skipping therapy.

I would have considered to validate the ASOs-mediated exon skipping approach suggested for *CC2D2A* (using primary cells of JBTS patients with biallelic *CC2D2A* mutations, if available), complementing genetic findings. To obtain primary cells from these patients, I would have considered contacting the recruiting clinician, via Genomics England standardised procedure or contacting the corresponding authors of the articles in the literature describing patients with truncating *CC2D2A* variants in potentially skippable exons (described in Table 4.1). Alternatively, if I had more time and primary cells were not available, I could have considered using CRISPR/Cas9-mediated genome editing to introduce *CC2D2A* patient specific mutations to study if

an ASOs-mediated exon skipping approach would be effective to bypass the deleterious effect of disease-causing *CC2D2A* mutations.

4.3.2. Use of 100,000 Genomes Project to find patients presenting pathogenic variants in *CEP120* and *TOGARAM1*

I searched for patients with biallelic *CEP120* variants and found one patient of interest presenting a homozygous missense *CEP120* variant of uncertain significance. This *CEP120* variant is c.1202A>C; p.(Glu401Ala) and it also is found in heterozygosity in the father of the Patient 1-*CEP120* (Figure 4.14). This family (Family 1-*CEP120*) is consanguineous, which may suggest that even though the variant is rare (0.000032 of total allele frequency in gnomAD) both parents contributed with one heterozygous allele of the variant.

The patient, Patient 1-*CEP120*, is diagnosed with developmental macular and foveal dystrophy, which is a phenotype that has not been observed in any of the patients with *CEP120* variants described in the literature (Roosing et al., 2016, Shaheen et al., 2015). It cannot be determined if this *CEP120* variant or any other is responsible for the phenotype observed in the patient. The phenotype of the patient seems to be less severe compared with the other clinical phenotypes (JBTS, JATD and OFD/MKS) found in patients reported in the literature associated with *CEP120* mutations. It can be suggested that doing functional cell-based assays for the patient-specific variant would elucidate insights about the pathogenicity of this variant and aiming to confirm if it is disease-causing and subsequently extend the phenotypic spectrum associated with *CEP120* mutations, or on the contrary discard this variant as disease-causing.

I searched for patients with disease-causing *CEP120* variants reported in the literature and found the variants: R151X, A199P and L712F in 6 families recruited in the Genomics England 100,000 Genomes Project. These variants are not responsible for the phenotype observed in the patients, as no clear segregation following a biallelic autosomal recessive inheritance pattern can be confirmed and the phenotypes of these patients do not match with the phenotypic spectrum of *CEP120* reported disease-causing mutations (Tables and 4.2 and 4.3 and Figure 4.15).

In Family 2-*CEP120* (Figure 4.15) the patient presents the same genotype as the patient's unaffected father. They both present the *CEP120* variants: A199P and

L712F, suggesting that the *CEP120* variants are in the same DNA strand and are inherited from the father and that there is one DNA strand without *CEP120* pathogenic variant, as the mother does not present any of these *CEP120* variants. Besides, Genomics England found a disease-causing in *ANKRD11*, responsible for the phenotype of this patient. It can be suggested a similar situation in which the variants A199P and L712F are not responsible of the phenotype of the patients occurring in Family 2-CEP120, as they do not seem to be inherited from one parent each. In Family 2-CEP120 the patient's mother does not present any of the *CEP120* variants found in the patient's father and patient. WGS data of the parents of these patients, when possible, would allow the possibility of getting insights about the segregation of the *CEP120* variants in these families.

In Family 7-CEP120, the R151X nonsense variant was found in heterozygosis in the patient. Nevertheless, there is no evidence of a second *CEP120* variant or *CEP120* being associated with phenotype of the patient. There is no evidence of *CEP120* mutations leading to clinical phenotypes in an autosomal dominant manner.

In Family 3-CEP120, the *CEP120* L712F variant was found homozygously. In the literature, this variant was reported in compound heterozygosity with the L726P variant, in a JBTS patient (Roosing et al., 2016). In total, the variant L712F was found in 577 participants in the Genomics England 100,000 Genomes Project, only in one of them in homozygosity (patient in Family 3-CEP120), this is a frequency of ~0.004 (out of all the 100,000 Genomes Project rare disease alleles), which is 100 times more frequent than: A199P and R151X *CEP120* variants (Table 4.4). For the L712F *CEP120* variant, population data from public databases indicate an allelic frequency of ~0.004 with 2 homozygous individuals in the normal population (data from gnomAD database). Although *in vitro* experiments indicated that this variant impairs the recruitment of Talpid3 to the centrioles, more functional analysis and reported *CEP120* patients with this mutation would be needed to confirm and understand the molecular disease mechanisms of its pathogenicity (Tsai et al., 2019).

It can be suggested that for the patient in Family 3-CEP120 it cannot be determined if this homozygous *CEP120* variant: L712F, or any or the other variants found in other genes is the cause of the patient phenotype (Tables 4.2 and 4.3 and Figure 4.15). I suggest that it would be appropriate to do additional functional analyses for this variant to complement *in silico* analyses about its pathogenicity and to confirm if this

variant can be suggested as disease-causing in this patient, or rule it out as disease-causing.

These analyses show the importance of recruiting patient's parents for WGS to confirm segregation of any variant that can be suggested to be responsible for the phenotype observed in a particular family case, this would help to provide an accurate and robust genetic diagnosis. On the contrary, the lack of any genome sequencing information from the parents may lead to misleading genetic diagnoses and prevent the validation of the segregation of disease-causing variants.

Assessment of allele frequency in each mutation of interest in any patient subject to a genetic diagnosis is important, as it has been shown in the discussion a variant reported in Radha Rama Devi et al. (Radha Rama Devi et al., 2020, Barroso-Gil et al., 2020). This variant is located in *CCDC28B* and the variant: c.73C>T; p.(Arg25Trp). It can be suggested for this reported variant found in *CCDC28B*, there is not enough evidence to suggest *CCDC28B* as a new JBTS candidate gene (Tables 4.7 and 4.8). The allele frequency of this variant is too common in publicly available databases to be responsible for the JBTS phenotype found in the reported patient, even though this variant has been recently associated with impairment of the immune synapse assembly (Capitani et al., 2022).

TOGARAM1 is a gene encoding TOGARAM1, which a protein with a role in the regulation of axonemal microtubules and ciliary stability (Latour et al., 2020). Biallelic mutations in *TOGARAM1* have been associated with phenotypes within the JBTS (Latour et al., 2020) and MKS phenotypic spectrum (Morbidoni et al., 2021).

The search for novel *TOGARAM1* variants using the tiering data of the Genomics England 100,000 Genomes Project proved to be effective in finding a JBTS patient (JAS-L50) (Figure 4.16) presenting two *TOGARAM1* pathogenic variants in compound heterozygosity. The variants are one heterozygous nonsense variant: c.5023C>T; p.(Arg1675Ter) and one heterozygous missense variant: c.1112C>A; p.(Ala371Asp). WGS data from the mother was available, and that allowed to confirm that one of the heterozygous variants in the patient segregated from the mother, who is unaffected and only presented one heterozygous *TOGARAM1* variant. It was confirmed *in silico*, using publicly online tools, that both *TOGARAM1* variants are rare and pathogenic (Latour et al., 2020).

Recently two foetuses with MKS phenotypes have been described (Morbidoni et al., 2021). These two foetuses presented compound heterozygous *TOGARAM1* variants: NM_015091.4:c.1102C>T; p.(Arg368Trp), and c.3619C>T; p.(Arg1207Ter) (Morbidoni et al., 2021). Of note, the variant c.1102C>T; p.(Arg368Trp) affects the TOG2 protein domain has also been described in homozygosity in a JBTS patient (Latour et al., 2020).

4.3.3. Design of a pipeline to study patient-specific *CEP120* mutations using cell-based assays

The deleterious effect of different patient-specific variants in *CEP120* has been previously investigated by various research groups (Joseph et al., 2018, Tsai et al., 2019), but these studies are limited to few cell types and few functional assays.

The Hh signalling pathway has not been investigated for *CEP120* patient-specific variant. Moreover, other *CEP120* exonic variants such as: R151X, S380YfsX and A549V have not been studied. I wanted to study patient-specific *CEP120* mutations, to do this, a pipeline was designed consisting in cell-based assays including Shh-LIGH2 cells and primary human fibroblast (further information was described in section: 2.14. Plasmid transfection intro mammalian cells, section 4.1. Introduction and aims and section: 4.2.10).

In previous studies, other groups investigated the role of *CEP120* in Hh signalling using cell-based assays with NIH-3T3 or MEF cells (Breslow et al., 2018, Pusapati et al., 2018, Bettleja et al., 2018). Nevertheless, it is unknown if different patient-specific mutations have a different effect on Hh signalling or if a differential effect on Hh signalling may explain why different *CEP120* mutations can lead to different phenotypes.

The cell-based pipeline designed (Figure 2.9) could help to further validate the pathogenicity effect of variants, such as L712F variant. Also this pipeline could be applied to study other *CEP120* variants not reported in the literature that can be candidate variants to be responsible for a patient phenotype, for example the *CEP120* variant of “Uncertain Significance” which was found: E401A (Figure 4.14).

I could not complete this work (aimed to complete the pipeline designed and shown in Figure 2.9) because plasmid transfection efficiency, using a GFP plasmid (which is a plasmid expected to be easier to transfect than the WT-*CEP120* plasmid, as it has

a smaller size) in Shh-LIGHT2 cells, was very low, and nearly absent in primary human fibroblasts.

If the transfection efficiency of the WT and mutant *CEP120* plasmids had been optimised, I would also determine if the transfection of the mutant *CEP120* plasmids can rescue the effect on Hh signalling caused by the *CEP120* knockdown via an *CEP120* ASO morpholino (pipeline shown in Figure 2.9). To study patient-specific *CEP120* mutations I would transfet the WT and Mutant *CEP120* plasmids into primary cells and analyse Hh signalling components via qPCR. I would analyse these components before and after Hh signalling stimulation using SAG.

The low or nearly absent plasmid transfection efficiency (Figure 4.36) prevented me to continue with the pipeline designed. Any attempt to do rescue experiments with such as low transfection efficiency would not detect any difference as the number of transfected cells would be too low, would be irrelevant and without scientific interest. Therefore, the optimisation transfection experiments as well as the ASO morpholino knockdown experiments were stopped.

I did not have the time required to optimise plasmid transfection in Shh-LIGHT2 cells. And even if I had optimised plasmid transfection in Shh-LIGHT2 cells I would also have to optimise transfection in human primary cells, which are described to be more difficult transfect than the Shh-LIGHT2 cells (derived from mouse NIH-3T3 cell line) (Hasan et al., 2021, Kucharski et al., 2021).

If the transfection had been efficient enough, I would aim to continue with the *CEP120* knockdown via the *CEP120* ASO morpholino experiments, using technical and biological triplicate replicates, in human primary fibroblasts. This could confirm if there is an effect in ciliogenesis or Hh signalling. If transfection in human primary fibroblasts or hURECs is not possible, I would transfet the WT and MUT hCEP120-EGFP plasmids into Shh-LIGHT2 cells. However the use of Shh-LIGHT2 cells implies that a new ASO morpholino would have to be designed, and besides, the effect of transfecting human *CEP120* cDNA sequence into mouse cells it is unknown.

To increase transfection efficiency, I would first search for advice and support within the Translational and Clinical Research Institute from collages at Newcastle University who perform routinely plasmid transfections. This is to have a direct reference of what are the most effective procedures to follow for plasmid transfection, including the selection of a suitable cell type to be transfected and the visualisation

method of transfected cells (living, fixed or stained) to assess and optimise transfection efficiency.

Furthermore, I would investigate alternative protocols to increase transfection efficiency (Hasan et al., 2021). I would use the nucleofection method, instead of liposomal transfection reagents such as Lipofectamine or polymer-based transfection reagents such as Xfect or FuGENE. Alternatively, lentiviral vector infection could also be attempted to increase transfection efficiency, however this option would involve more technical complexity and training than the other options mentioned above.

I used a commercial CEP120 antibody that has not been reported in the literature. It was observed that the staining of this CEP120 antibody locates at the expected places: centrosome and basal body as it was previously reported using other antibodies (Betleja et al., 2018, Mahjoub et al., 2010, Meka et al., 2022, Tsai et al., 2019). I used this commercial CEP120 antibody by doing immunofluorescence on two different cell types: Shh-LIGHT2 cells and primary human fibroblasts. Two different types of microscopy were used: Fluorescence microscopy (Zeiss AxioImager) and Nikon (A1) confocal inverted microscopy.

From the immunofluorescence experiments performed, it can be suggested that the quality, in terms of background noise, of the CEP120 staining is different depending on the cell type and microscope used. It can also be suggested that in human cells such as primary human fibroblasts, there is less background/noise (non-specific staining) compared to mouse cells. This is not unexpected, as this commercial antibody targets the CEP120 human protein sequence, which is not identical (~70% of homology) compared to the mouse one. It was found that this background can be avoided using a Nikon (A1) confocal inverted microscopy.

These differences in non-specific staining can be determinant for some experiments, for example: if I had achieved enough transfection efficiency (e.g. 50% of cells are GFP positive cells) transfecting the WT *CEP120* plasmid into mouse cells (such as Shh-LIGHT2 cells) and wanted to analyse the rescue effect of WT or MUT hCEP120-EGFP plasmids by visualising any changes in CEP120 staining (e.g. intensity or localisation), I would suggest to use either a Nikon (A1) confocal inverted microscope or to obtain higher transfection efficiency into human cells.

It was also demonstrated that it is possible to reduce the expression of *CEP120* in Shh-LIGHT2 cells and primary human fibroblasts via siRNA and ASO morpholino,

respectively, and the cells survived until the RNA extraction (after two days) or lysis for dual luciferase assay are done (after 3 days). The siRNA experiment reduced *CEP120* expression up to 50% but did not cause an effect on Smo, Gli and Patch at RNA level (Figures 4.28-4.30). Treating primary human fibroblasts with a *CEP120* ASO morpholino (Figure 4.33) seems to be more effective than *CEP120* siRNA in mouse Shh-LIGHT2 cells (Figure 4.26).

Besides a dual luciferase experiment was performed, in which Shh-LIGHT2 were stimulated with SAG and it was observed that in the cells treated with *Smo* siRNA there was an impairment of Hh stimulation. It can be suggested that stimulating the cells with SAG would allow to observe differences in terms of gene expression of Hh components and stimulation of Hh signalling when *CEP120/Cep120* expression is reduced. If differences in Hh signalling want to be seen, it would be needed to stimulate the cells with SAG and compare them with cells treated with DMSO only, the same way as that was done with the dual luciferase assay (Figure 4.33).

However, the main limitation is that I did not do technical or biological replicates, which prevents to confirm this.

In conclusion, *in silico* tools were used and the literature available to create a database of reported variants of two JBTS genes, including *CC2D2A*, which can be suggested as a good candidate gene for a possible exon skipping therapy. *In vitro* analyses using cells lines and/or primary cells can support genetic findings related with primary ciliopathies allowing, for instance: the visualisation of cilia and centrosome via immunofluorescence and study of the expression via qPCR analysis of some components of signalling pathways upon the knockdown of a particular gene.

I used the WGS data available in the Genomics England 100,000 Genomes Project to find a patient with disease-causing biallelic variants in *TOGARAM1*, a gene recently associated with JBTS. Additionally, several examples were provided of how important the sequencing of parental genomes (when available) is to confirm (or discard) the segregation of putative disease-causing variants and to assess allele frequency of putative disease-causing variants.

Chapter 5. Investigation of genes involved in the endoplasmic reticulum (ER) and use of Human Urine-Derived Renal Epithelial Cells (hURECs) to complement genetic studies

5.1. Introduction and aims

There are multiple renal ciliopathy syndromes, which are characterised by genetic overlap and by genetic and phenotypic heterogeneity (Braun and Hildebrandt, 2017). Some examples of renal ciliopathies include polycystic kidney disease (PKD) and nephronophthisis (NPHP) (Arts and Knoers, 2013). Patients with these renal ciliopathies can show extra-renal phenotypes including brain, eye, liver, skeletal systems and cardiovascular phenotypes (Novarino et al., 2011, Devlin and Sayer, 2019, Reiter and Leroux, 2017). In particular, liver and renal phenotypes can frequently be observed together in patients with renal ciliopathies (Huynh et al., 2020, Besse et al., 2019).

Autosomal dominant polycystic kidney disease (ADPKD) is the most common inherited renal ciliopathy (Cornec-Le Gall et al., 2019, McConnachie et al., 2021), typically ADPKD does not present until adulthood and its prevalence is estimated to be about 1 in 1000 in the general population (Lanktree et al., 2021). However, the prevalence of ADPKD is challenging to estimate because of its age dependent penetrance and variable expression of the disease (Lanktree et al., 2021). ADPKD is not the only the only autosomal dominant ciliopathy in which renal phenotypes can be observed, other autosomal dominant ciliopathies are: autosomal dominant tubulo-interstitial kidney disease (ADTKD) and autosomal dominant polycystic liver disease (ADPLD), furthermore, phenotypic and genetic overlap is often observed between these autosomal dominant ciliopathies (Cornec-Le Gall et al., 2018a, Devuyst et al., 2019).

ADPKD is characterized by the presence of numerous and often large renal cysts. This leads to progressive kidney enlargement and a decline in kidney function over time, mainly in late adulthood. ADPKD patients may also develop other extra-renal phenotypes, some of them are: hepatic cysts, pancreatic cysts and hypertension (Cornec-Le Gall et al., 2019). Regarding the molecular basis of the disease, the 'two-hit model' is the prevalent theory, in which cysts develop from cells that acquire a second somatic mutation, and this deactivates the remaining normal allele (Ta et al., 2020). More precisely, it is thought that cytogenesis and additional PKD phenotypes

are associated with the dysfunction of various interconnected signalling pathways (Cabrita et al., 2020, Vasileva et al., 2021). *PKD1* or *PKD2* mutations are responsible for the ADPKD phenotype in about 90% of patients (Lanktree et al., 2021, Senum et al., 2022). *PKD1* and *PKD2* genes code for the proteins polycystin-1 (PC1) and polycystin-2 (PC2) respectively. PC1 and PC2 form a PC1/PC2 complex at the ciliary membrane. PC1 has features of both an ion channel and a G-protein coupled receptor and PC2 functions as an ion channel in the primary cilium (Wang et al., 2007, DeCaen et al., 2013). The maturity and ciliary localisation of PC1 and PC2 is dependent on the glycosylation pathway at the endoplasmic reticulum (ER) (Besse et al., 2017, Porath et al., 2016). Some patients, who do not have a pathogenic variant in *PKD1* or *PKD2*, show an atypical phenotype of ADPKD. There is an increasing number of new genetic causes of autosomal polycystic kidney and liver phenotypes, such as mutations in *GANAB*, *DNAJB11*, *IFT140*, *ALG5*, *ALG8* and *ALG9* (Senum et al., 2022, Huynh et al., 2020). These genes code for proteins functioning at the ER protein biosynthetic pathway and it has been shown that mutations in some of these genes, or absence of the protein they encode, have an effect on PC1 and/or PC2 (Porath et al., 2016, Besse et al., 2019, Cornec-Le Gall et al., 2018a, Besse et al., 2017).

The ER is a cell organelle, which is involved in protein synthesis, folding and transport, among other functions (Schwarz and Blower, 2016). Multiple co- and post-translational protein modifications occur in the ER, one of them is Asparagine (N)-linked protein glycosylation. The N-linked-glycosylation pathway is an evolutionary conserved process involved in correct folding, maturation and trafficking of membrane and secreted glycoproteins including PC1 and PC2 (Schwarz and Blower, 2016, Aebi, 2013, Hu and Harris, 2020, Lemoine et al., 2022).

The Asparagine-Linked Glycosylation (ALG) genes are a group of 13 genes involved in the ER glycosylation pathway (Schoberer et al., 2018, Aebi, 2013, Lemoine et al., 2022). *ALG5*, *ALG8* and *ALG9* function in the N-linked-glycosylation pathway. In particular, *ALG5* participates in the dolichol cycle, *ALG5* is a glucosyltransferase that participates in the synthesis of the Dol-P-Glc donor substrate (Lemoine et al., 2022). *ALG8* adds the second glucose residue and *ALG9* adds the seventh and ninth mannose residues to the polypeptide (Besse et al., 2017, Xu and Ng, 2015).

Recently *in vitro* experiments based on human or mouse cells revealed that heterozygous loss-of-function (LoF) mutations in *Alg5* or *ALG8* resulted in defective

maturation and trafficking of PC1 (Besse et al., 2017, Lemoine et al., 2022) and absence of Alg9 resulted in defective PC1 maturation (Besse et al., 2019).

Other genes such as: *GANAB* and *DNAJB11* are associated with ADPKD-like or ADPLD phenotypes and are also functioning in the ER glycosylation pathway. In the ADPKD patients with mutations in any of these genes the renal disease was usually mild, however the severity of the liver disease ranged from no cysts to severe polycystic liver disease (PLD). *GANAB* encodes for glucosidase II subunit α (GII α) (Porath et al., 2016). Mutations in *GANAB* are associated with an atypical (usually mild) form of ADPKD with variable, including severe, PLD (Besse et al., 2017, Porath et al., 2016). The renal phenotype in patients with *DNAJB11* mutations is characterised by non-enlarged kidneys, chronic interstitial fibrosis, small renal cysts and variable liver disease severity (Cornec-Le Gall et al., 2018a, Huynh et al., 2020). Of note, mutations in *PRKCSH* are associated with ADPLD and *PRKCSH* encodes GII β , the other subunit of GII (Li et al., 2003).

DNAJB11 is a cofactor of the ER heat shock chaperon BiP, which regulates protein folding and trafficking (Cornec-Le Gall et al., 2018a, Huynh et al., 2020). Loss of *DNAJB11* is associated with defects in the maturation of PC1 and a reduction in PC1 membrane expression (Cornec-Le Gall et al., 2018a). Furthermore, biallelic mutations in some of the ALG genes involved in the glycosylation pathway (including *ALG8* and *ALG9*) lead to a more severe phenotype affecting multiple organs called congenital disorder of glycosylation (AlSubhi et al., 2016).

As it was explained previously in the Introduction (Chapter 1) and in Chapter 4, the Genomics England 100,000 Genomes Project is a valuable resource to find novel patients presenting variants in a list of genes (Best et al., 2021, Best et al., 2022b, Wheway et al., 2019). The Genomics England 100,000 Genomes Project contributed to the description of novel ADPKD patients presenting variants in *DNAJB11* (Huynh et al., 2020), and more recently *ALG5* (Lemoine et al., 2022). Furthermore, this database has been essential to genetically solve patients and investigate other primary ciliopathies and rare diseases (Smedley et al., 2021, Best et al., 2022b, Best et al., 2021, Latour et al., 2020, Olinger et al., 2021, Chan et al., 2022).

Genetic studies of primary ciliopathies and novel ciliopathy genes are often complemented with functional studies using cell-based and models and/or patient's

cells (Latour et al., 2020, Shaheen et al., 2015, Besse et al., 2017, Cornec-Le Gall et al., 2018a, Devlin et al., 2022).

As explained above, heterozygous variants in *ALG5*, *ALG8* and *ALG9* are associated with ADPKD and ADPKD-like phenotypes (Besse et al., 2019, Besse et al., 2017, Lemoine et al., 2022). It can be suggested that there may be other *ALG* genes that could be associated with the overlapping ADPKD and ADPKD-like phenotypes, and furthermore, that the combination of *in silico* experiments and *in vitro* experiments using hURECs can be used to test this hypothesis.

By using the Genomics England 100,000 Genomes Project, I aim to find novel patients (genetically solve patients) presenting disease-causing variants in genes encoding proteins functioning within the endoplasmic reticulum such as *DNAJB11*, *ALG5*, *ALG8* and *ALG9*. I aim to investigate if there is an enrichment of heterozygous loss-of-function alleles in any of the *ALG* genes, in a population of patients with cysts in kidney or liver, compared to a non-cystic control population. I aim to investigate one disease-causing patient-specific mutation and complement genetic studies using hURECs.

5.2. Results

Using the tiering data in the Genomics England 100,000 Genomes project, several putative disease-causing variants were found in the genes: *DNAJB11*, *GANAB*, *ALG5*, *ALG8* and *ALG9* in patients diagnosed with cystic kidney disease or similar phenotypes. These genes code for proteins involved in the ER biogenesis pathway. Of note, none of the pathogenic variants described in this section were homozygous, all of them were heterozygous.

I did not find any other putative disease-causing variants in other genes that could explain the cystic kidney disease or similar phenotypes found in these patients, except for some cases with variants in *GANAB*.

The terms “disease-causing” or “pathogenic” are used for a variant with enough evidence of being responsible for the clinical phenotypic found in the corresponding patient. However, a variant described as disease-causing, pathogenic or deleterious by MutationTaster or other tools for *in silico* analysis of pathogenicity, may not be described in this thesis as “disease-causing” or “pathogenic” after combining all the analyses and assessment shown in this chapter.

5.2.1. Novel patients with pathogenic mutations in *DNAJB11*

Nine *DNAJB11*-affected individuals were found, presenting heterozygous pathogenic variants in *DNAJB11* (Tables 5.1-5.3), a gene recently associated with an atypical form of autosomal dominant polycystic kidney disease (Cornec-Le Gall et al., 2018a). 8 of these patients were diagnosed with cystic kidney disease and one of them diagnosed with Renal tract calcification (Huynh et al., 2020).

For one of these *DNAJB11*-affected individuals: Patient 5-*DNAJB11*, more detailed clinical information was provided by clinician (Table 5.2 and Figure 5.1). Cystic kidney disease is present in Patient 5-*DNAJB11*, as well as in her uncle and sister, however genome sequencing information was unavailable to confirm segregation of the *DNAJB11* disease-causing variant.

8 different disease-causing variants in *DNAJB11* were found. 4 of them are described as nonsense, 2 of them as frameshift, 1 of them as splice donor variant and 1 of them as missense. One of them: c.724C>T; p.(Arg242Ter) is found in two different unrelated patients. I did not find any other putative disease-causing variants in other genes that could explain the cystic kidney disease or similar phenotypes found in these patients (Huynh et al., 2020).

Patient ID	Age	Sex	Phenotype	HPO terms	Affected father	Affected mother
1-DNAJB11	76	M	Cystic kidney disease	Multiple renal cysts (cortical and medullary), HBP, inguinal and umbilical hernias	Yes	No
2-DNAJB11	81	F	Cystic kidney disease	Multiple renal cysts, HBP, CKD, Stroke, Macular degeneration	No	Yes
3-DNAJB11	55	F	Cystic kidney disease	Hypertension, Renal cortical cysts, cone dystrophy	Yes	No
4-DNAJB11	63	F	Cystic kidney disease	Multiple renal cysts, HBP, CKD stage 4	Yes	No
5-DNAJB11 (1)	50	F	Cystic kidney disease	Multiple renal cysts, liver cysts, ascending aortic aneurysm	Unknown	Yes
6-DNAJB11	66	F	Renal tract calcification (or nephrolithiasis or nephrocalcinosis)	Nephrocalcinosis, CKD stage 3	No	Yes
7-DNAJB11	64	F	Cystic kidney disease	Multiple renal cysts, HBP	Unknown	No
8-DNAJB11	64	M	Cystic kidney disease	Multiple renal cysts (cortical), liver cysts, CKD stage 3, gout, HBP, obesity, hernia of the abdominal wall	Unknown	Unknown
9-DNAJB11 (2)	59	M	Cystic kidney disease	Multiple renal cysts, HBP	No	Yes

Table 5.1. Clinical features of the unrelated *DNAJB11*-affected individuals identified in the Genomics England 100,000 Genomes Project database. (1) The phenotype of this patient and affected family members is further described in Table 5.2. (2) Patient 9 has an affected sibling, which is also diagnosed with cystic kidney disease, presenting multiple renal cysts and HBP. CKD Chronic Kidney Disease; HBP High blood pressure. These patients are also described in Huynh et al., 2020 (Huynh et al., 2020).

Patient ID	Relationship with Proband	Age, yr	eGFR or ESRD	Type	Description of the cysts	Kidney length, cm (TKV ml)	HBP	ADPKD classification	Other significant conditions (age, yr)
5-DNAJ B11	-	50	74	MRI	MBSC	R, 10.4; L, 10.8 (360)	No	1A	Thoracic aortic aneurysm (45)
5-DNAJ B11(II. 1)	Uncle	68	ESRD	US	MBSC	Non enlarged	NA	2B	None
5-DNAJ B11 (III.2)	Sister	60	ESRD	CT	MBSC	Non enlarged	NA	2B	Thoracic aortic aneurysm (50)

Table 5.2. Detailed clinical features of the *DNAJB11*-affected proband presenting the variant: c.730A>T; p.Lys244Ter (Patient 5-DNAJB11) and her affected relatives. Patient 5-DNAJB11 (from Family 5-DNAJB11) correspond to “proband III.1” from “Family P Newcastle” described in Huynh et al., 2020 (Huynh et al., 2020). No blood sample were available for Proband’s uncle and sister; the presence of the familial variant: c.730A>T; p.(Lys244Ter) was not confirmed. Further details of this patient are shown in Figure 5.1 and Table 5.1. ADPKD, autosomal dominant polycystic kidney disease; CT, computed tomography; eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; MBSC; multiple bilateral small cysts; MRI, magnetic resonance imaging; NA, not available; TKV, total kidney volume; US, ultrasound.

Patient ID	Mutation	Rs ID	CA DD	SIFT	PolyP hen	ACMG	MutationT aster	gnomA D
1- DNAJB11	c.724C>T; p.(Arg242Ter)	rs765764 090	T:40	NA	NA	Likely Pathogen ic	Disease causing	4.798E -06
2- DNAJB11	c.161C>G; p.(Pro54Arg)	rs155384 9919	G:2 6.7	Deleteri ous (0)	Proba bly dama ging (1)	Pathog enic	Disease causing	NA
3- DNAJB11	c.296_297delA G; p.(Ser100Leuf sTer33)	NA	NA	NA	NA	Pathog enic	Disease causing	NA
4- DNAJB11 (1)	c.100C>T; p.(Arg34Ter)	rs148976 481	T:39	NA	NA	Likely Pathogen ic	Disease causing	4.153E -06
5- DNAJB11	c.730A>T; p.(Lys244Ter)	NA	NA	NA	NA	Pathog enic	Disease causing	NA
6- DNAJB11	c.724C>T; p.(Arg242Ter)	rs765764 090	T:39	NA	NA	Likely Pathogen ic	Disease causing	4.798E -06
7- DNAJB11 (2)	c.68+1del	NA	NA	NA	NA	Likely Pathogen ic	Disease causing	NA
8- DNAJB11	c.400delA, p.(Ile134fsTer3)	rs116919 2999	NA	NA	NA	Pathog enic	Disease causing	4.042E -06
9- DNAJB11	c.616C>T; p.(Arg206Ter)	rs941713 150	T:36	NA	NA	Pathog enic	Disease causing	3.986E -06

Table 5.3. Variants found in the *DNAJB11*-affected individuals found in the Genomics England 100,000 Genomes Project database. *DNAJB11* transcript considered: NM_016306.6. All the variants have been found in heterozygosis. (1) This patient also presents a heterozygous *PKD2* variant: c.2140A>G; p.(Lys714Glu) of unknown significance. (2) This patient also presents a heterozygous *PKD1* variant: c.4085C>T; p.(Ser1362Phe) of unknown significance. NA, not available. These variants are also described in Huynh et al., 2020 (Huynh et al., 2020).

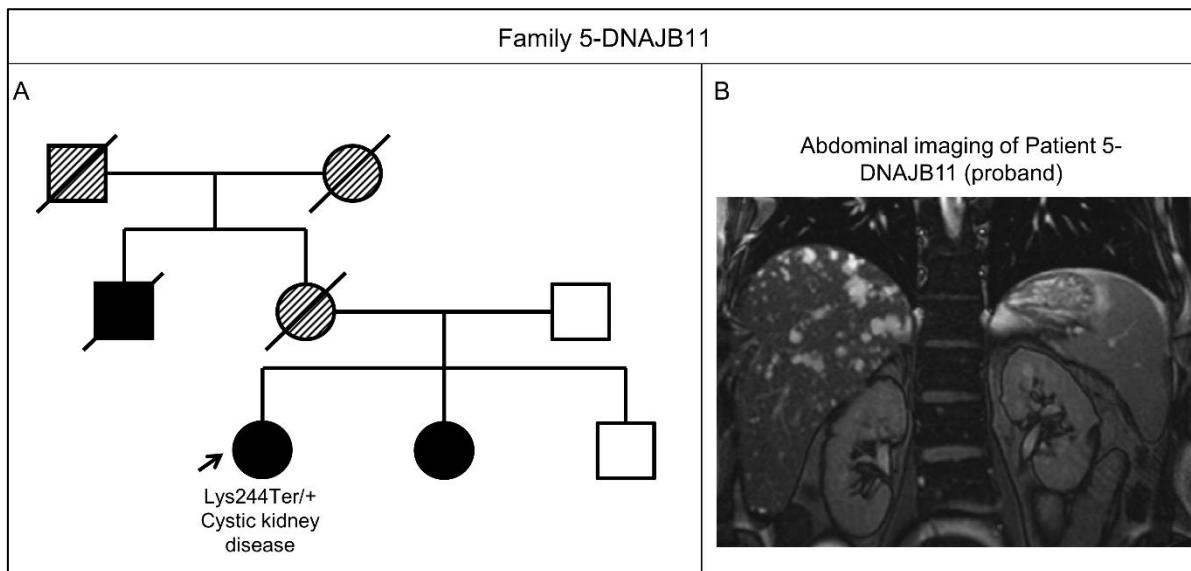


Figure 5.1. Family 5-DNAJB11 presenting a disease-causing variant in *DNAJB11*. A: Pedigree diagram showing the segregation of a nonsense *DNAJB11* variant in Family 5-DNAJB11. The proband (patient with genome sequencing available indicated by an arrow: Patient 5-DNAJB11) presents a heterozygous variant: NM_016306.5 (*DNAJB11*): c.730A>T; p.(Lys244Ter). There is not genome sequencing information available from the other family members. Black circles and squares indicate affected females and males respectively, while white squares indicate unaffected males. Lined individuals indicate that the clinical information of that individual is unavailable. Clinical characteristics proband, and relatives when available, and further annotation of the variant are detailed in Tables 5.1, 5.2 and 5.3. B: Abdominal imaging of Patient 5-DNAJB11 via T2-weighted magnetic resonance imaging. The proband described as “Patient 5-DNAJB11” from Family 5-DNAJB11 is the same individual as “Individual 5” described in Huyn et al., 2020 (Huynh et al., 2020).

5.2.2. Novel patients with pathogenic mutations in *GANAB*

GANAB is a gene associated with ADPKD/ADPLD phenotypes (Porath et al., 2016, Besse et al., 2017). Using the same approach, I looked for *GANAB* variants in-affected individuals with cystic kidney phenotypes.

8 *GANAB*-affected individuals were investigated: 7 of them diagnosed with cystic kidney disease and one of them diagnosed with unexplained kidney failure in young people (Table 5.4).

All 8 of the *GANAB* variants are described as missense (Table 5.5). 5 of these 8 patients also present a different *PKD1* variant that can explain the cystic phenotype in each of them. For the rest of the patients (3 patients without a *PKD1* variant that can explain the renal phenotype), one of them presents a *PKD2* disease-causing variant. For the remaining other two patients (without a *PKD1* or *PKD2* disease-causing variant), further *in silico* and *in vitro* analyses are needed to assess

pathogenicity of these *GANAB* variants and determine if they are responsible for the corresponding patients' phenotype.

Patient ID	Phenotype	HPO terms	Affected father	Affected mother	Suggested disease-causing gene
1-PKD1	Cystic kidney disease	Pancreatic cysts, Hepatic cysts, HBP, Multiple renal cysts	Yes	No	<i>PKD1</i>
2-GANAB	Cystic kidney disease	Multiple renal cysts, Enlarged kidney, Chronic kidney disease, HBP	No	Yes	-
3-PKD1	Cystic kidney disease	Multiple renal cysts, Enlarged kidney	No	Unknown	<i>PKD1</i>
4-PKD1	Cystic kidney disease	Renal cortical cysts, Multiple renal cysts	Yes	No	<i>PKD1</i>
5-GANAB	Cystic kidney disease	Enlarged kidney, HBP, Multiple renal cysts, Chronic kidney disease	Unknown	No	-
6-PKD1	Unexplained kidney failure in young people	HBP, Chronic kidney disease, Gout	Unknown	No	<i>PKD1</i>
7-PKD1	Cystic kidney disease	Hepatic cysts, HBP, Multiple renal cysts, Enlarged kidney	Yes	No	<i>PKD1</i>
8-PKD2	Cystic kidney disease	Hearing impairment, Multiple renal cysts, HBP	Yes	No	<i>PKD2</i>

Table 5.4. Clinical features of the unrelated individuals with rare heterozygous variants in *GANAB* identified in the Genomics England 100,000 Genomes Project database. Most of the patients present a disease-causing variant in *PKD1* or *PKD2*. For patients 2-GANAB and 5-GANAB a putative disease-causing variant was not found in any related cystogene. For more information about the specific variants investigated in these patients see Table 5.5. HBP, high blood pressure.

Patient ID	Mutation	Gene	Rs ID	CA DD	SIFT	PolyP hen	ACMG	Mutatio n Taster	gno mAD	Suggested disease-causing gene
1- PKD 1	c.986C>G; p.(Pro329Arg)	<i>GANAB</i>	rs201 8014 99	C:2 0.4	0.47 (tolerated)	0.015 (benign)	Likely Benign	Polymorphism	1.13 E-04	<i>PKD1</i>
	c.8311G>A; p.(Glu2771 Lys)	<i>PKD1</i>	rs105 7518 897	T:2 6.8	0 (deleterious)	0.998 (probably damaging)	Pathogenic	Disease causing	6.59 E-06 (v.3.1.2)	
2- GAN AB	c.790A>G; p.(Met264Val)	<i>GANAB</i>	rs757 0595 06	C:2 1.8	0.01 (deleterious)	0.023 (benign)	Likely pathogenic	Polymorphism	7.96 E-06	-
3- PKD 1	c.1569G>T; p.(Trp523Cys)	<i>GANAB</i>	rs779 8197 84	A:2 5.5	0.05 (tolerated)	0.887 (possibly damaging)	Uncertain significance	Disease causing	3.98 E-06	<i>PKD1</i>
	c.12691C>T; p.(Gln4231 Ter)	<i>PKD1</i>	rs755 4964 50	A:5 1	NA	NA	Pathogenic	Disease causing	NA	
4- PKD 1	c.2159G>A; p.(Arg720Gln)	<i>GANAB</i>	rs765 0004 69	T:2 6.3	0.01 (deleterious)	0.914 (probably damaging)	Uncertain significance	Disease causing	7.16 E-05	<i>PKD1</i>
	c.984_1014 del; p.(Gly329TrpfsTer126)	<i>PKD1</i>	NA	NA	NA	NA	Likely pathogenic	Disease causing	NA	
5- GAN AB	c.757A>C; p.(Lys253Gln)	<i>GANAB</i>	rs201 6959 01	G:2 2.6	0.09 (tolerated)	0.439 (benign)	Benign	Disease causing	5.66 E-05	-

	c.160C>T; p.(Arg54Trp)	<i>GANAB</i>	rs555 8263 81	A:2 6.2	0.02 (delete rious)	0.663 (possi bly dama ging)	Benig n	Diseas e causing	7.43 E-05	
6- PKD 1	c.8119G>A; p.(Val2707 Met)	<i>PKD1</i>	rs139 5070 58	T:8. 036	0.28 (tolerat ed)	0.031 (beni gn)	Benig n	Polymo rphism	9.33 E-04	<i>PKD1</i>
	c.4151C>T; p.(Thr1384I Ie)	<i>PKD1</i>	rs150 0314 40	A:2 4.3	0 (delete rious)	0.997 (prob ably dama ging)	Likely Benig n	Diseas e causing	1.19 E-04	
	c.2579G>C; p.(Gly860Al a)	<i>GANAB</i>	rs374 8488 41	G:2 0.9	0.13 (tolerat ed)	0.381 (beni gn)	Uncert ain Signifi cance	Diseas e causing	1.59 E-05	
7- PKD 1	c.7467delC; p.(Lys2490 ArgfsTer13 0)	<i>PKD1</i>	NA	NA	NA	NA	Likely Patho genic	Diseas e causing	NA	<i>PKD1</i>
	c.655G>A; p.(Asp219A sn)	<i>GANAB</i>	rs772 8663 33	T:1 5.4 5	0.53 (tolerat ed)	0 (beni gn)	Benig n	Polymo rphism	3.18 E-05	
8- PKD 2	c.637C>T; p.(Arg213T er)	<i>PKD2</i>	rs130 2726 543	T:3 7	NA	NA	Patho genic	Diseas e causing	3.19 E-05	<i>PKD2</i>

Table 5.5. Variants found in *GANAB*/*PKD1*/*PKD2*-affected individuals found in the Genomics England 100,000 Genomes Project database. The transcripts considered for *GANAB*, *PKD1* and *PKD2* are: NM_198335.4, NM_001009944.3 and NM_000297.4 respectively. All the variants have been found in heterozygosity. *GANAB* variant in Patient 2-*GANAB*: c.790A>G; p.(Met264Val) also causes a start loss in an alternative transcript: *GANAB*(NM_001329224.1):c.1A>G ; p.(Met1Val). Patient 1-*PKD1* and Patient 4-*PKD1* also have a relative each diagnosed with cystic kidney disease, presenting the corresponding variants in heterozygosity. For patients 2-*GANAB* and 5-*GANAB* a putative disease-causing variant was not found in any related cystogene.

5.2.3. Novel patients with pathogenic mutations in *ALG5*

ALG5 is a gene that has only been recently associated with ADPKD. Before the work of Lemoine et al., (Lemoine et al., 2022) in which I contributed, only *ALG8* and *ALG9* mutations were associated with ADPKD/ADPLD phenotypes (Besse et al., 2019, Besse et al., 2017, Mantovani et al., 2020). The proteins coded by these three genes have similar roles at the ER glycosylation pathway (Aebi, 2013, Besse et al., 2017, Schoberer et al., 2018). I found 2 *ALG5*-affected individuals diagnosed with cystic kidney disease with *ALG5* heterozygous pathogenic variants (Figure 5.2 and Table 5.6).

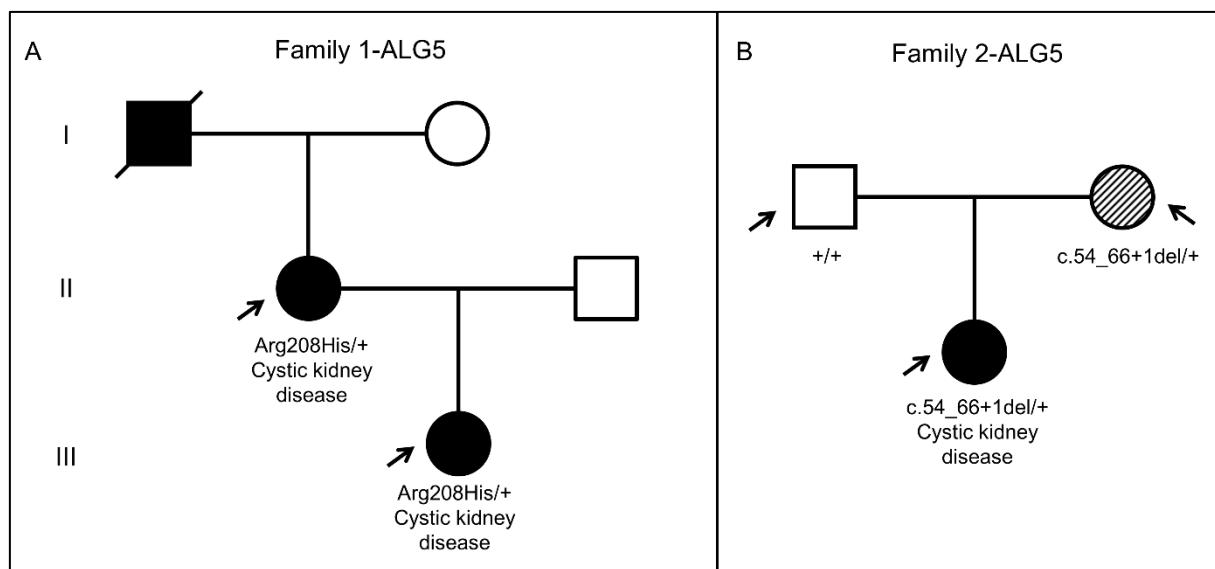


Figure 5.2. Families 1-ALG5 and 2-ALG5 presenting disease-causing *ALG5* variants. A: Pedigree diagram showing the segregation of a missense *ALG5* variant in Family 1-ALG5. The proband (patient with genome sequencing available indicated by an arrow; Patient 1-ALG5(III.1)) presents a heterozygous variant: NM_013338.5 (*ALG5*): c.623G>A; p.(Arg208His). Genome sequencing information is available from the mother (also represented with an arrow), who is also affected, and presents the missense variant in heterozygosity. B: Pedigree diagram showing the segregation of a deletion in *ALG5* affecting a splice donor site (exon 1) in Family 2-ALG5. The proband (patient with genome sequencing available indicated by an arrow; Patient 2-ALG5) and the mother of the proband present the deletion in heterozygosity: NM_013338.5 (*ALG5*): c.54_66+1del. The father of the proband does not present the deletion. Black circles and squares indicate affected females and males respectively, while white circles and squares indicate unaffected females and males respectively. Lined individuals indicate that the clinical information of that individual is unavailable. Clinical characteristics of the probands, and relatives, if available, and further annotation of the variants are detailed in Tables 5.6 and 5.7. The probands in Families 1-ALG5 and 2-ALG5 are the same as: "subject III.1 of Family 4; c.623G>A (p.Arg208His)" and "Singleton, c.54_66+1del" respectively, described in Lemoine et al., 2022 (Lemoine et al., 2022).

Renal phenotypes were also present in two relatives of one of ALG5-affected individuals: Patient 1-ALG5(III.1) (Lemoine et al., 2022). I found 2 different disease-causing variants. 1 of them described as splice donor variant and 1 of them as missense (Table 5.7).

Patient ID	Family ID	Relationship with Proband	Variant	Sex	eGFR (age) or ESKD (age)	Aspect of the kidneys	Other significant conditions (age)
1-ALG5(III.1)	1-ALG5	-	c.623G>A; p.(Arg208His)	F	83 (50)	Normal-sized kidneys with multiple cysts	HBP (41), diaphragmatic hernia
1-ALG5 (II.1)	1-ALG5	Mother	c.623G>A; p.(Arg208His)	F	38 (73)	atrophic multicystic kidneys	DMT2
1-ALG5 (I.1) (1)	1-ALG5	Grandfather	-	M	ESKD (85)	NA	Cause of death: uremia
2-ALG5	2-ALG5	-	c.54_66+1del	F	>90 (20)	Multiple kidney cysts	Nephrocalcinosis, hypercalcemia

Table 5.6. Clinical features of the *ALG5*-affected patients identified in the Genomics England 100,000 Genomes Project database. *ALG5* transcript considered: NM_013338.5. (1) Genome sequencing was unavailable in this individual; therefore, the *ALG5* variant was not confirmed. All the variants have been found in heterozygosity. These patients and variants are also described in Lemoine et al., 2022 (Lemoine et al., 2022). DMT2 diabetes mellitus type 2; eGFR, estimated glomerular filtration rate; F, female; HBP, high blood pressure; M, male; NA, not available.

Mutation	Rs ID	CADD	SIFT	PolyPhen	ACMG	Mutation Taster	gnomAD
c.623G>A; p.(Arg208His)	rs749484 470	T:33	0 (deleterious)	0.997 (probably damaging)	Pathogenic	Disease causing	7.28E-06
c.54_66+1del	NA	NA	NA	NA	Likely pathogenic	Disease causing	NA

Table 5.7. Variants found in *ALG5*-affected individuals found in the Genomics England 100,000 Genomes Project database. *ALG5* transcript considered: NM_013338.5. NA, not available.

5.2.4. Novel patients with pathogenic mutations in *ALG8*

5 *ALG8*-affected individuals diagnosed with cystic kidney disease with *ALG8* heterozygous pathogenic variants were found (Tables 5.8 and 5.9). Patient 4-*ALG8* has a sibling, who is also diagnosed with cystic kidney disease and presents the c.1090C>T; p.(Arg364Ter) disease-causing *ALG8* variant.

A total 4 different disease-causing variants were found. 2 of them are described as nonsense and 2 of them as frameshift. One of them: c.1090C>T; p.(Arg364Ter) is found in two different unrelated patients. I did not find any other putative disease-causing variants in other genes that could explain the cystic kidney disease or similar phenotypes found in these patients.

Patient ID	Variant	Phenotype	HPO terms	Affected father	Affected mother
1- ALG8	c.1041del; p.(Ser348LeufsTer19)	Cystic kidney disease	Multiple renal cysts, hepatic cysts	yes	no
2- ALG8	c.1296dupA; (p.Leu433ThrfsTer9)	Cystic kidney disease	Enlarged kidney, multiple renal cysts	no	Unknown
3- ALG8	c.1090C>T; p.(Arg364Ter)	Cystic kidney disease	Abnormality of the eye, abnormality of the corneal thickness, multiple renal cysts	yes	no
4- ALG8 (1)	c.1090C>T; p.(Arg364Ter)	Cystic kidney disease	Renal cortical cysts, haemangioma, multiple renal cysts	Unknown	yes
5- ALG8	c.121C>T; p.(Arg41Ter)	Cystic kidney disease	Renal cortical cysts, HBP, hepatic cysts, ab of nails, ab of spleen, multiple renal cysts, malignant GI tumors, multiple small medullary cysts	Unknown	Unknown

Table 5.8. Clinical features of the unrelated *ALG8*-affected patients identified in the Genomics England 100,000 Genomes Project database. *ALG8* transcript considered: NM_024079.5. All the variants have been found in heterozygosis. (1) Patient 4-ALG8 has a sister also diagnosed with cystic kidney disease. Ab, abnormality, HBP, high blood pressure.

Mutation	Rs ID	CADD	ACMG	MutationTaster	gnomAD
c.1041del; p.(Ser348LeufsTer19)	rs963287770	NA	Pathogenic	Disease causing	3.99E-06
c.1296dupA; (p.Leu433ThrfsTer9)	NA	NA	Pathogenic	Disease causing	NA
c.1090C>T; p.(Arg364Ter)	rs376161880	A:44	Pathogenic	Disease causing	6.37E-05
c.121C>T; p.(Arg41Ter)	rs200888240	A:42	Pathogenic	Disease causing	1.19E-05

Table 5.9. Variants found in *ALG8*-affected individuals found in the Genomics England 100,000 Genomes Project database. *ALG8* transcript considered: NM_024079.5. NA, not available.

5.2.5. Novel patients with pathogenic mutations in *ALG9*

3 *ALG9*-affected individuals diagnosed with cystic kidney disease with *ALG9* heterozygous pathogenic variants were found (Tables 5.10 and 5.11). A total of 3 different disease-causing variants were found. 2 of them are described as nonsense and 1 of them as splice donor variant. I did not find any other putative disease-causing variants in other genes that could explain the cystic kidney disease or similar phenotypes found in these patients.

Patient ID	Variant	Phenotype	HPO terms	Affected father	Affected mother
1-ALG9	c.1472del; p.(Asn491IlefsTer33)	Cystic kidney disease	Multiple renal cysts	No	No
2-ALG9	c.1012_1018+8 delinsATAC	Cystic kidney disease	Multiple renal cysts, ab of the outer ear, hearing impairment, ab of the eye, HBP, ab heart morphology	No	No
3-ALG9	c.1363C>T; p.(Arg455Ter)	Cystic kidney disease	Multiple renal cysts, HBP	No	No

Table 5.10. Clinical features of the unrelated *ALG9*-affected patients identified in the Genomics England 100,000 Genomes Project database. *ALG9* transcript considered: NM_024740.2. All the variants have been found in heterozygosity.

Mutation	Rs ID	CADD	ACMG	Mutation Taster	gnomAD
c.1472del; p.(Asn491IlefsTer33)	NA	NA	Pathogenic	NA	NA
c.1012_1018+8delinsATAC	NA	NA	Likely pathogenic	NA	NA
c.1363C>T; p.(Arg455Ter)	rs782775735	A:38	Pathogenic	Disease causing	4.01E-06

Table 5.11. Variants found in *ALG9*-affected individuals found in the Genomics England 100,000 Genomes Project database. *ALG9* transcript considered: NM_024740.2. NA, not available.

5.2.6. Enrichment of loss-of-function (LoF) variants in *ALG1*, *ALG8*, *ALG9* and *ALG12* in a population with kidney and/or liver cysts

ALG genes encode proteins that function in the endoplasmic reticulum. It is known that heterozygous loss-of-function (LoF) variants in some of these genes, such as *ALG5*, *ALG8* and *ALG9* can be associated with defects on PC1 biogenesis and lead to cystic renal and/or liver phenotypes within the phenotypic spectrum of ADPKD (Apple et al., 2023, Lemoine et al., 2022, Besse et al., 2019, Besse et al., 2017).

The overall aim of this section is to investigate if there is an enrichment of LoF heterozygous alleles in any of the *ALG* genes in a population of patients with cysts in kidney or liver, compared to a non-cystic control population. This enrichment analysis in *ALG* genes may validate known phenotype-genotype associations, such as *ALG8* and *ALG9* with ADPKD and ADPKD-like phenotypes, and it may also suggest new ones in other *ALG* genes with ADPKD and ADPKD-like phenotypes.

To do this I used the WGS and clinical data available in the Genomics England 100,000 Genomes Project to create two populations: one population of patients with cysts in kidney or liver (case population) and a population without cysts in kidney or liver (control population).

Even though there are few missense disease-causing variants in *ALG5*, *ALG8* and *ALG9* described to be associated with ADPKD, this enrichment analysis is focussed on LoF variants (Apple et al., 2023, Lemoine et al., 2022, Besse et al., 2019). This is because the age-dependent and incomplete penetrance of heterozygous variants in some of the *ALG* genes such as *ALG8* and *ALG9* (Besse et al., 2019, Apple et al., 2023) and the complexity of finding and validating disease-causing missense variants leading to ADPKD.

I estimated the burden of different types of variants (LoF variants, and non-LoF variants as a negative control) in the *ALG* genes, in a population of patients with a cystic phenotype in kidney and/or liver and compared it with the burden of the same type of variants in a population without these cystic phenotypes.

Using the data available in the Genomics England 100,000 Genomes Project it is observed an enrichment of LoF alleles in *ALG1*, *ALG8*, *ALG9* and *ALG12* in the population of cases compared to the population of controls. Furthermore, the odds of having a LoF variant in these 4 genes are higher in cases than in controls.

In order to create the case and control populations I first extracted the phenotypic information of the data available of the participants in the Rare Disease Program (main release v14, dated 27th of January 2022).

Firstly, the probands in the Rare Disease Program were selected. There are 72955 participants in the Rare Disease program. It can be observed that 47% (34085 participants) of these participants are described as "Proband" and 53% (38869 participants) are described as "Relative" (Figure 5.3.A). There are 39844 participants with a diagnosis in the Rare Disease Program, it can be observed that 86% (34084) of them are described as "Proband" and 14% (5760) of them are described as "Relative" (Figure 5.3B), in this study only probands were considered (n=34084).

A case and a control population were created based on the phenotype of these probands, 1992 probands were selected for the case population as they have cysts in kidney and/or liver. 16773 probands were selected for the control population, as they do not have cysts in kidney or liver. 15389 probands without cysts in kidney or liver were discarded of this study as they are diagnosed with a renal disease, undiagnosed metabolic disorders, a ciliopathy or a disease similar to a ciliopathy. These 15389 probands were discarded from the final control population to avoid having a control population in which some patients, even though they do not have kidney or liver cysts, present some the clinical phenotypes that could be caused by mutations in the *ALG* genes (Figure 5.3C).

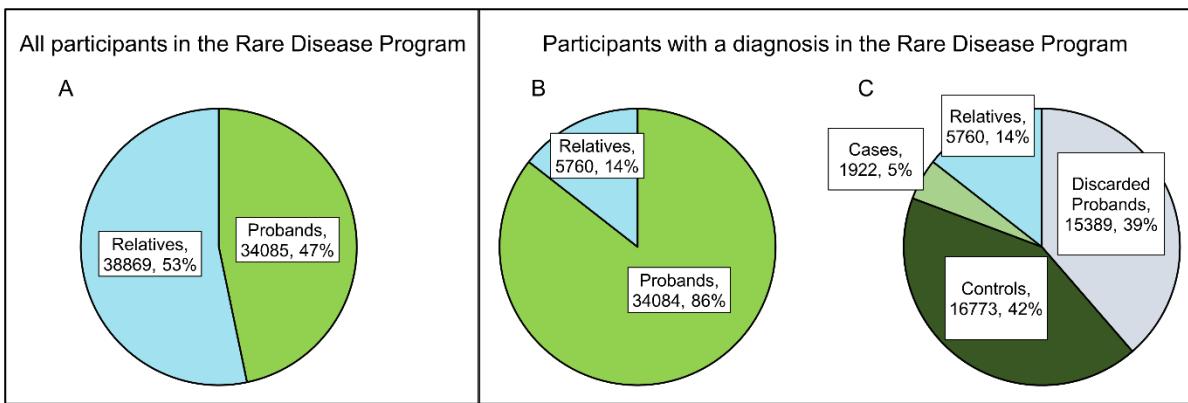


Figure 5.3. Distribution of the participants in the Rare Disease Program (main release v14, dated 27th of January 2022) to visualise where the case and control populations are selected from. A: Pie chart showing the distribution of participants, found in the table from Labkey: “rare_diseases_analysis” corresponding to the participants in the rare disease program (72955 participants) according to their “participant type status”. B: Pie chart showing the distribution of participants, found in the table from Labkey: “rare_diseases_participants_disease” corresponding to the participants in the rare disease program with a diagnosis (39844 participants), according to their “participant type status”. C: Pie chart showing the distribution of participants, found in the table from Labkey: “rare_diseases_participants_disease” corresponding to the participants in the rare disease program with a diagnosis (39844 participants), according to if they were selected for this study (case and control populations) or not (discarded probands and relatives).

Of note, about 10% and 1% of the participants with a diagnosis in the Rare Disease Program, are diagnosed with Renal and Urinary tract disorders and ciliopathies respectively (Figure 5.4 and Table 5.12). More details about the number of patients diagnosed with a particular ciliopathy or renal and urinary tract disorder are shown in Tables 5.13 and 5.14.

The case population (1922 participants), a population of patients with cysts in kidney and/or liver, was selected based on 74 terms that referred to cysts in kidney or liver (Appendix 3). It is observed that most of the cases are diagnosed with a renal disease, as expected.

For example, 65% of the participants (1245) are diagnosed with cystic kidney disease, 9% of the participants (174) are diagnosed with congenital anomaly of the kidney and urinary track (CAKUT) and 3% of participants (53) are diagnosed with unexplained kidney failure in young people (Figure 5.5A and Appendix 4). It is observed that 37% of the cases (717 cases) are genetically solved by Genomics England (Figure 5.5B).

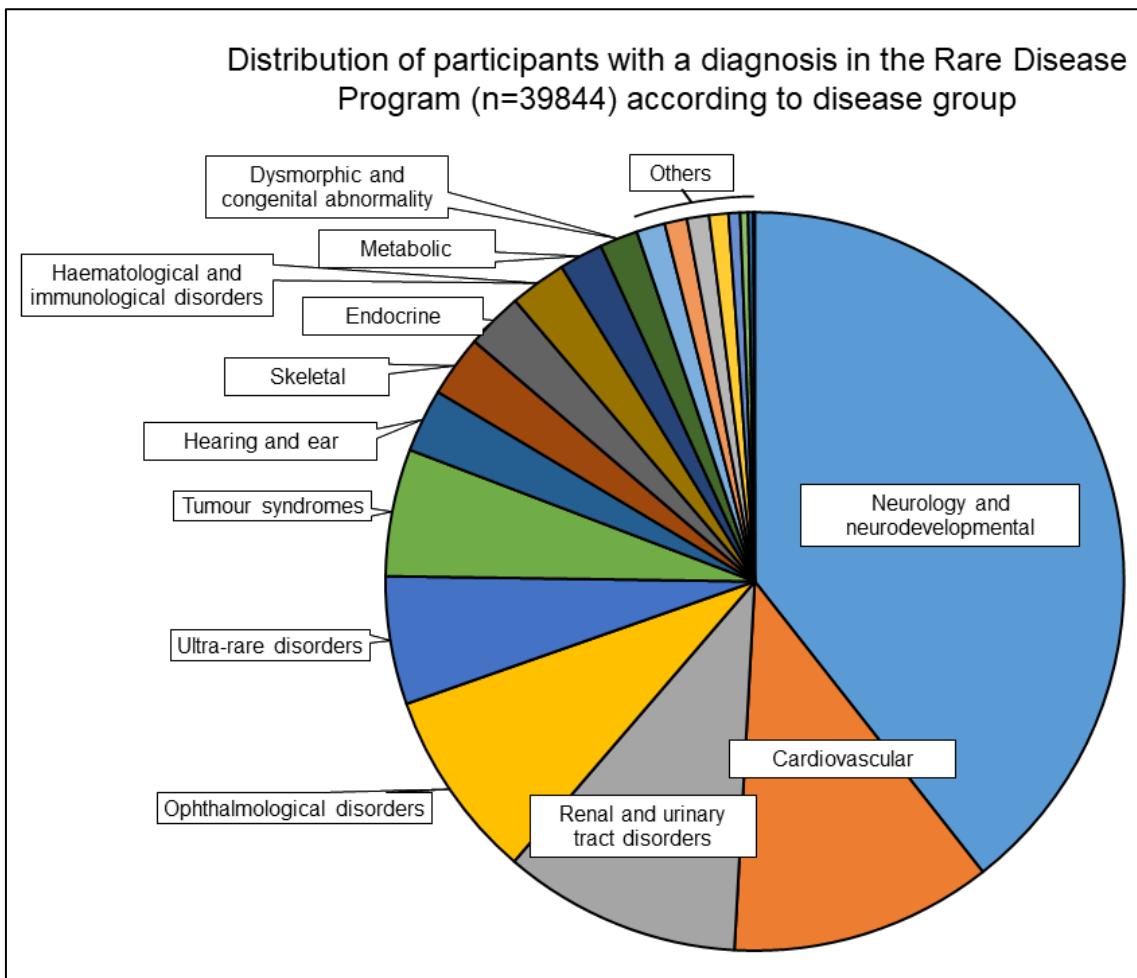


Figure 5.4. Distribution of the participants in the Rare Disease Program (main release v14, dated 27th of January 2022). Pie chart showing the distribution of participants, found in the table from Labkey: “rare_diseases_participants_disease” corresponding to the participants in the rare disease program with a diagnosis (39844 participants), according to their disease group. Of note, 10% of the participants (4234) are diagnosed with a kidney and urinary tract disorder and 1% (391) are diagnosed a ciliopathy. For clarity, of the pie chart shown, a group called: “Others”, includes the following disease groups: Dermatological disorders, Respiratory disorders, Ciliopathies, Rheumatological disorders, Growth disorders, Gastroenterological disorders, Psychiatric disorders, Infectious diseases and NA’s. The number of participants included in each disease group is detailed in Table 5.12.

Disease group	Number of participants	%
Neurology and neurodevelopmental disorders	16014	39.41
Cardiovascular disorders	4664	11.48
Renal and urinary tract disorders	4234	10.42
Ophthalmological disorders	3382	8.32
Ultra-rare disorders	2278	5.61
Tumour syndromes	2266	5.58
Hearing and ear disorders	1127	2.77
Skeletal disorders	1096	2.70
Endocrine disorders	1010	2.49
Haematological and immunological disorders	1006	2.48
Metabolic disorders	778	1.91
Dysmorphic and congenital abnormality syndromes	679	1.67
Dermatological disorders	507	1.25
Respiratory disorders	401	0.99
Ciliopathies	391	0.96
Rheumatological disorders	343	0.84
Growth disorders	206	0.51
Gastroenterological disorders	140	0.34
Psychiatric disorders	95	0.23
Infectious diseases	13	0.03
NA's	9	0.02

Table 5.12. Distribution of the participants in the rare disease program diagnosed with at least one disease in terms of their disease group. This information was obtained from Labkey: main release v14, dated 27th of January 2022. Of note, there are 751 participants in more than one disease group (1 participant in 5 disease groups, 2 participants in 4 disease groups, 37 participants in 3 disease groups and 711 participants in 2 disease groups).

Disease	Number of participants	%
Non-CF bronchiectasis	161	41.18
Primary ciliary dyskinesia	137	35.04
Bardet-Biedl Syndrome	53	13.55
Rare multisystem ciliopathy disorders	26	6.65
Joubert syndrome	14	3.58

Table 5.13. Distribution of the participants included in rare disease group: Ciliopathies, in terms of their diagnosed disease.

Disease	Number of participants	%
Cystic kidney disease	1558	36.80
CAKUT (congenital anomalies of the kidney and urinary tract)	1212	28.63
Renal tract calcification (or Nephrolithiasis or nephrocalcinosis)	301	7.11
Proteinuric renal disease	287	6.78
Unexplained kidney failure in young people	271	6.40
Familial haematuria	240	5.67
Extreme early-onset hypertension	224	5.29
Renal tubular acidosis	60	1.42
Atypical haemolytic uraemic syndrome	45	1.06
Primary membranoproliferative glomerulonephritis and Familial IgA nephropathy and IgA vasculitis	36	0.85

Table 5.14. Distribution of the participants included in rare disease group: Renal and urinary tract disorders, in terms of their diagnosed disease.

Distribution of participants of the case population (n=1922) according to their specific disease and their status regarding to if they are solved or not by Genomics England

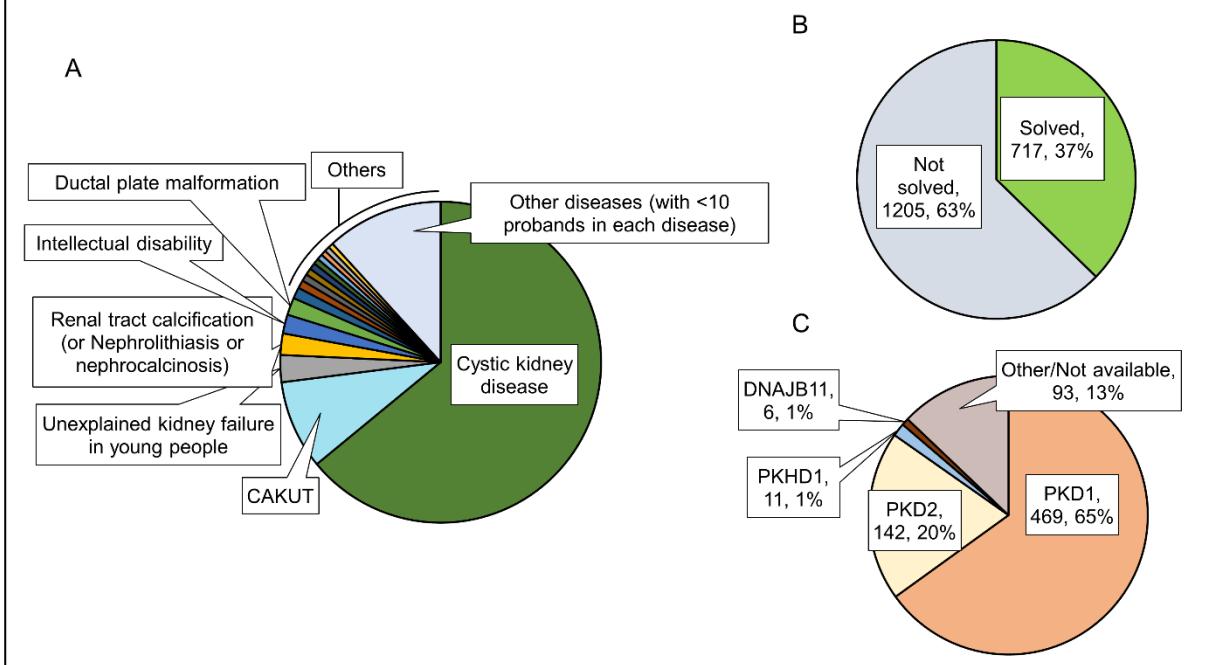


Figure 5.5. Distribution of the case population created from the Genomics England 100,000 Genomes Project according to the diagnosed disease of the patients and their status regarding to if they are solved or not by Genomics England. This information was obtained from Labkey: main release v14, dated 27th of January 2022. The case population consists in 1922 probands with cystic kidney and/or liver disease. A: Pie chart showing the distribution of participants selected corresponding to the case population selected according to their disease. B: Pie chart showing the distribution of participants in the case population (1922 participants), according to if they are solved by Genomics England (using the criteria: *case_solved* family = yes, this information was obtained from the table from Labkey: "gmc_exit_questionnaire"). C: Considering the 717 patients from the case population that are genetically solved by Genomics England (gmc_exit_questionnaire, main release v14), more than 50% of them (n=469) are solved by variants in *PKD1* gene. Of note, for some other patients with cysts in kidney or liver the gene name was not available. Other genes reported in the GMC exit questionnaire to genetically solve these patients are: *HNF1B* (n=4), *CEP290* (n=4), *SEC63* (n=2), *COL4A4* (n=2), *GANAB* (n=1), *LRP5* (n=1), *WDR19* (n=1), *COL1A1* (n=1), *COL4A5* (n=1) among other genes. Some patients present several variants reported in the gmc_exit_questionnaire, for example: there are 4 patients with both *PKD1* and *PKD2* variants, and the patient genetically solved with a pathogenic *LRP5* variant has also a variant of unknown clinical significance in *PKD2*.

In the control population (16773 participants), a population of patients without cysts in kidney or liver and without the diagnosis or a renal disease, a ciliopathy or similar ciliopathy phenotypes. For example, 33% of the participants (5610) are diagnosed with Intellectual disability, 9% of the participants (1361) are diagnosed with Epilepsy plus other features, 3% of participants (1215) are diagnosed with ultra-rare undescribed monogenic disorders (Figure 5.6A and Appendix 5). It is observed that 13% of the controls (2188 controls) are genetically solved by Genomics England (Figure 5.6B).

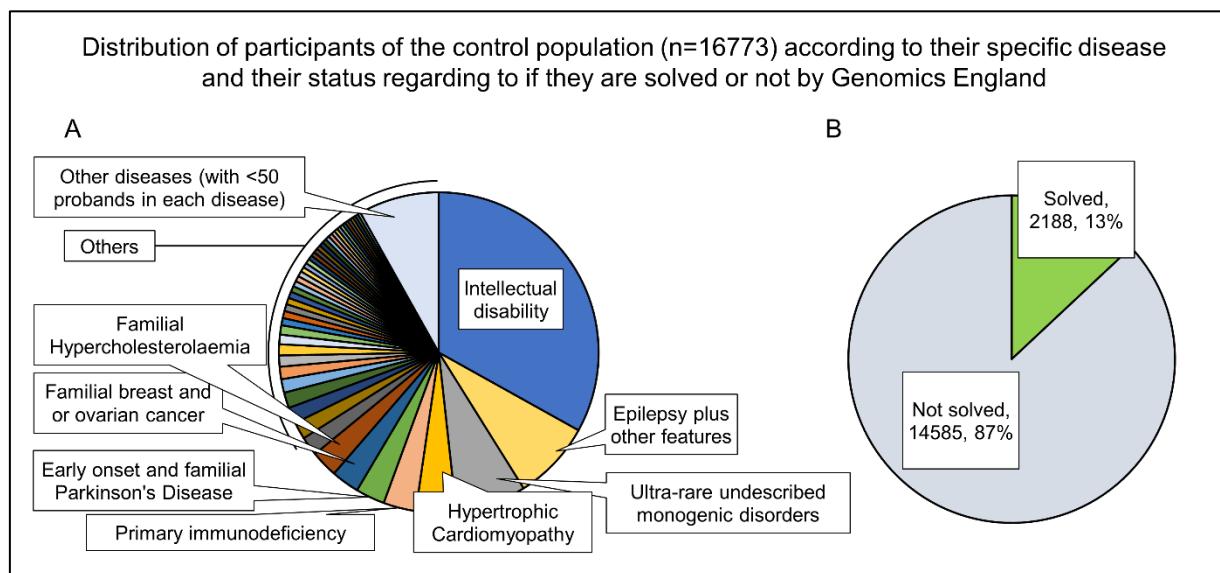


Figure 5.6. Distribution of the control population created from the Genomics England 100,000 Genomes Project according to the diagnosed disease of the patients and their status regarding to if they are solved or not by Genomics England. This information was obtained from Labkey: main release v14, dated 27th of January 2022. The control population consists in 16773 probands without cystic kidney or liver disease. A: Pie chart showing the distribution of participants selected corresponding to the control population, according to their disease. B: Pie chart showing the distribution of participants in the control population selected for this study (16773 participants), according to if they are solved by Genomics England (using the criteria: case_solved family = yes, this information was obtained from the table from Labkey: "gmc_exit_questionnaire").

Using the case and control population created, it was found an enrichment of LoF alleles in *ALG1*, *ALG8*, *ALG9* and *ALG12* in the population of cases compared to the population of controls. I studied the proportion of LoF alleles for each of the population: cases and controls and for each of the *ALG* genes. The proportion of LoF alleles is the number of loss-of-function alleles divided the total alleles.

Three different analyses were performed, based on LoF of variants were described. In the first analysis, I considered as LoF variant any variant that is described as

nonsense or frameshift. It is observed, in the control population, that 0.09% of the alleles in *ALG1* are LoF alleles, while in the case population this percentage is 0.08, which is about 9 times higher. For *ALG8* and *ALG12*, the proportion of LoF alleles in cases is about 3 and 5 times higher than in controls, respectively (Figure 5.7).

Applying a Fisher's Exact Test it was found that the difference in the proportion of LoF alleles in *ALG1*, *ALG8*, and *ALG12* in cases compared to that proportion in controls is statistically significant, with p-values of 0.017, 0.016 and 0.041 respectively (Table 5.15).

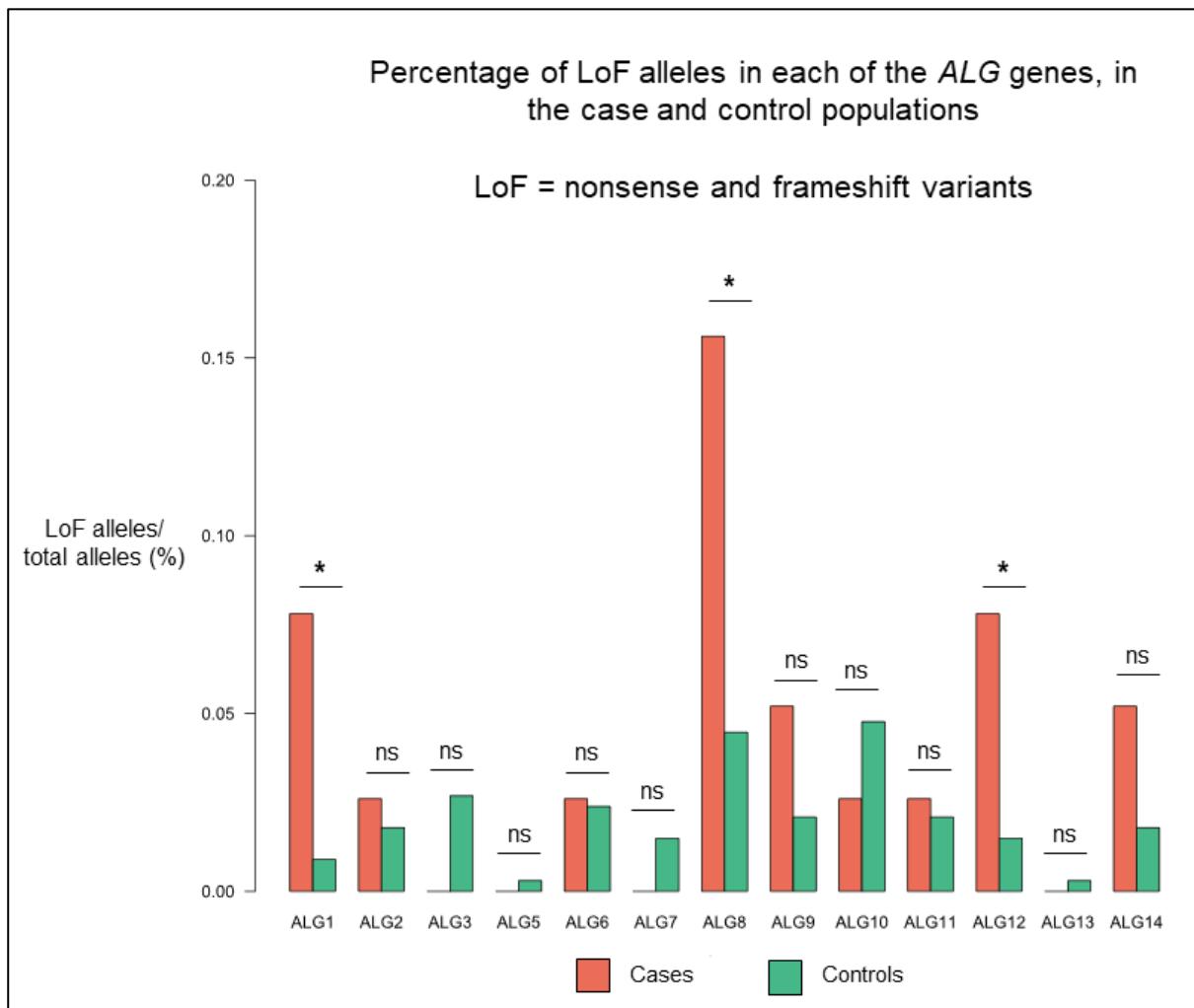


Figure 5.7. Enrichment of predicted loss-of-function alleles (LoF) alleles in *ALG1*, *ALG8* and *ALG12* in the population of cases compared to the population of controls. Bar plot showing the percentage of LoF, considering nonsense and frameshift alleles as LoF alleles in the case population (3844 total number alleles), shown as red bars against the control population (33546 total number alleles), shown as green bars. The corresponding allele counts and p-values for each gene are detailed in Table 5.15. Fisher's Exact Test was used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are indicated as ns.

Gene	LoF = stop and frameshift				LoF = stop, frameshift, splice donor and splice acceptor				LoF = stop, frameshift, splice donor and splice acceptor (unsolved patients only)			
	cases	controls	p-value	OR (95%CI)	cases	controls	p-value	OR (95%CI)	cases	controls	p-value	OR (95%CI)
<i>ALG1</i>	3	3	0.017	8.74 (2.04-37.45)	3	8	0.096	3.28 (0.94-10.95)	1	4	0.328	3.03 (0.25-18.14)
<i>ALG2</i>	1	6	0.532	1.46 (0.13-8.74)	1	6	0.532	1.46 (0.13-8.74)	1	5	0.379	2.42 (0.21-17.36)
<i>ALG3</i>	0	9	0.611	0.00 (0.00-3.32)	1	9	1.000	0.97 (0.09-5.78)	0	8	1.000	0.00 (0.00-5.42)
<i>ALG5</i>	0	1	1.000	0.00 (0.00-78.54)	1	1	0.195	8.73 (0.46-165.90)	1	1	0.147	12.11 (0.64-230.20)
<i>ALG6</i>	1	8	1.000	1.09 (0.10-6.94)	2	18	1.000	0.97 (0.22-3.70)	2	14	0.348	1.73 (0.39-7.19)
<i>ALG7</i>	0	5	1.000	0.00 (0.00-5.84)	0	5	1.000	0.00 (0.00-5.84)	0	5	1.000	0.00 (0.00-8.10)
<i>ALG8</i>	6	15	0.016	3.50 (1.42-8.77)	6	29	0.167	1.81 (0.80-4.26)	5	28	0.103	2.17 (0.90-5.27)
<i>ALG9</i>	2	7	0.235	2.50 (0.52-11.02)	3	8	0.096	3.28 (0.94-10.95)	3	8	0.046	4.55 (1.30-15.19)
<i>ALG10</i>	1	16	1.000	0.55 (0.05-3.10)	1	18	0.714	0.48 (0.05-2.67)	1	12	1.000	1.01 (0.09-6.30)
<i>ALG11</i>	1	7	0.580	1.25 (0.11-8.74)	1	7	0.580	1.25 (0.11-8.74)	1	6	0.426	2.02 (0.18-12.13)
<i>ALG12</i>	3	5	0.041	5.24 (1.39-19.57)	3	12	0.195	2.18 (0.66-7.63)	3	8	0.046	4.55 (1.30-15.19)
<i>ALG13</i>	0	1	1.000	0.00 (0.00-78.54)	0	1	1.000	0.00 (0.00-78.54)	0	0	1.000	NA
<i>ALG14</i>	2	6	0.195	2.91 (0.60-12.11)	2	9	0.315	1.94 (0.42-8.00)	2	8	0.174	3.03 (0.65-12.15)

Table 5.15. Enrichment of loss-of-function (LoF) alleles in *ALG1*, *ALG8*, *ALG9* and *ALG12* in the case population. Loss-of-function (LoF) alleles counts in case and control populations with the p-value of the statistical tests performed and Odds Ratio (OR). A p-value lower than 0.05 is indicated in bold. An OR with the lowest confidence interval (95%CI) higher than 1 is also indicated in bold.

In the second analysis, I considered as LoF variant any variant that is described as nonsense, frameshift, splice donor or splice acceptor. It is observed that, none of the differences in the proportion of LoF alleles in cases compared to the proportion of LoF alleles in controls is statistically significant (Figure 5.8 and Table 5.15).

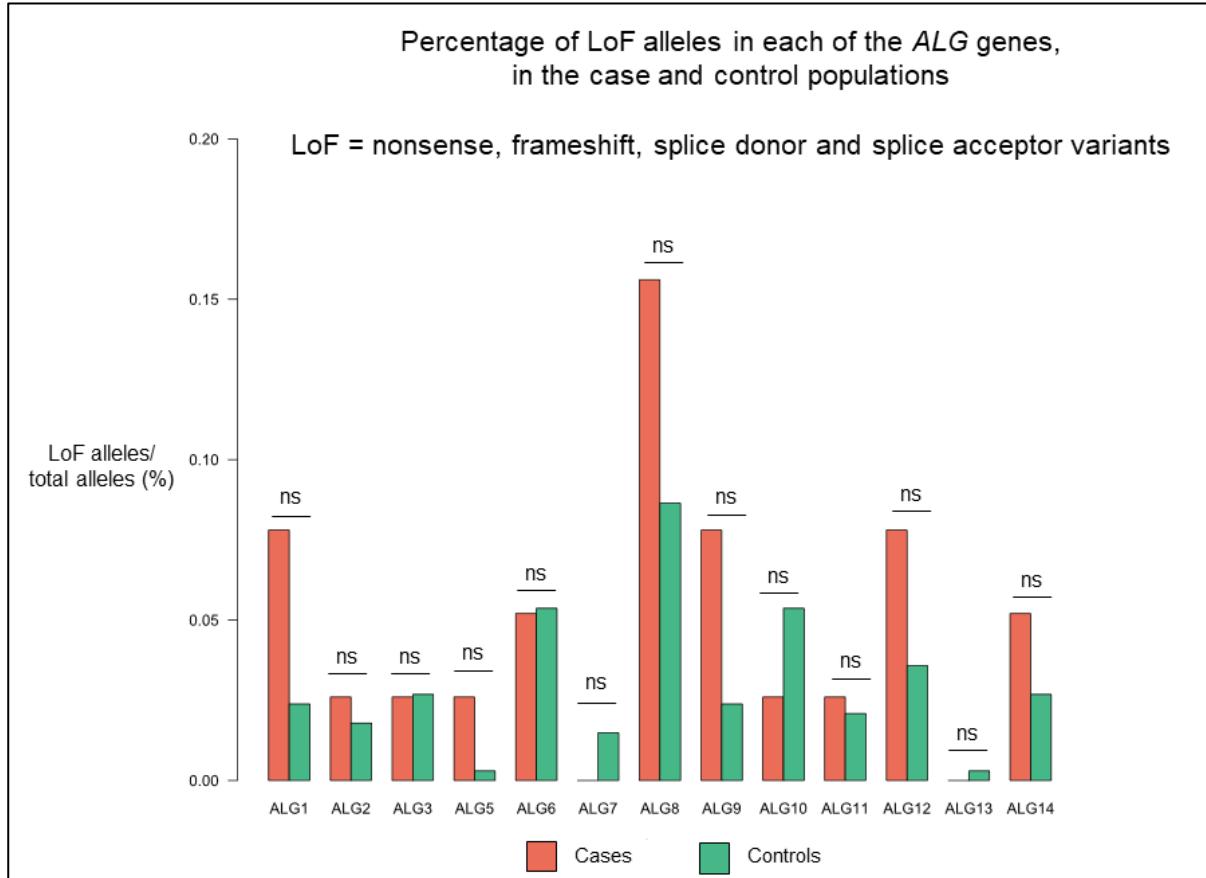


Figure 5.8. Bar plot showing the percentage of loss-of-function alleles (LoF) alleles in the case and control populations. Nonsense, frameshift, splice donor and splice acceptor alleles are considered LoF alleles. The percentage of LoF alleles for each of the *ALG* genes in the case population (3844 total number alleles) is shown as red bars, and in the control population (33546 total number alleles) is shown as green bars. The corresponding allele counts and p-values for each gene are detailed in Table 5.15. Fisher's Exact Test was used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are indicated as ns.

In the third analysis, I considered as LoF variants any variant that is described as nonsense, frameshift, splice donor or splice acceptor. Besides, I only included the genetically unsolved cases (1205 cases) and genetically unsolved controls (14585 controls). It is observed that, for *ALG9* and *ALG12*, the proportion of LoF alleles in cases is about 3 times higher than in controls (Figure 5.9). And Applying a Fisher's Exact Test it was found that the difference in the proportion of LoF alleles in *ALG9* and *ALG12* in cases compared to that proportion in controls is statistically significant, with a p-value of 0.046 in both genes (Table 5.15).

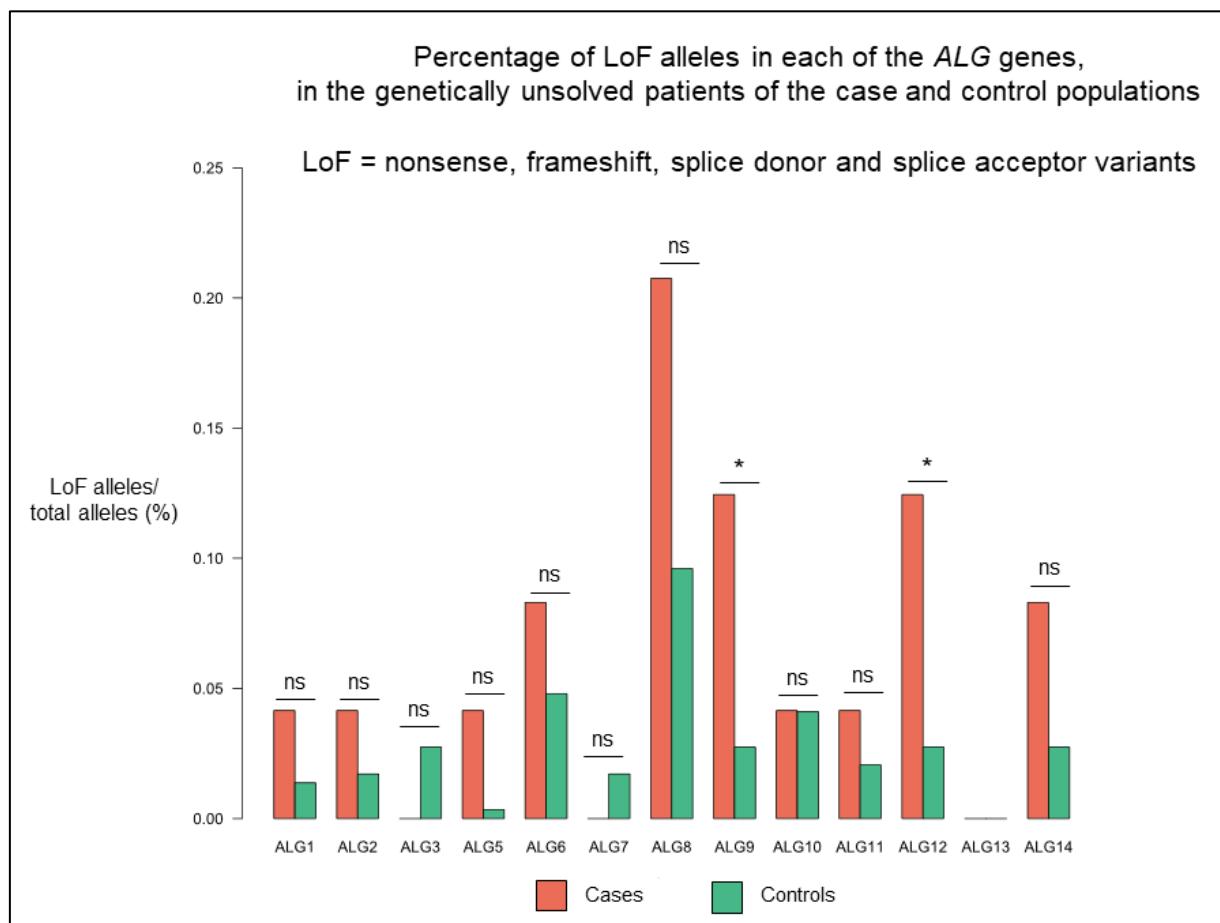


Figure 5.9. Enrichment of predicted loss-of-function alleles (LoF) alleles in *ALG9* and *ALG12* in the population of cases compared to the population of controls that are genetically unsolved by Genomics England. Bar plot showing the percentage of LoF, considering nonsense, frameshift alleles, splice donor and splice acceptor as LoF alleles in the genetically unsolved (based on GMC exit questionnaire, main release v14) case population (2410 total number alleles), shown as red bars against the genetically unsolved control population (29170 total number alleles), shown as green bars. There are 717 cases and 2188 controls genetically solved by Genomics England. These patients have been excluded to create the following bar plot. The corresponding allele counts and p-values for each gene are detailed in Table 5.15. Fisher's Exact Test was used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are indicated as ns.

Each of the three analyses described above was complemented by calculating the odds ratio (OR) of having a LoF variant in cases against having a LoF variant in controls for each of the *ALG* genes. An OR of 1.0 indicates that there is no difference in the odds between the cases and controls while an OR higher than 1.0 indicates an increase in odds among the cases against controls.

In the first analysis it is observed an odds ratio of 8.74, 3.5, and 5.24 in *ALG1*, *ALG8* and *ALG12*, respectively, and in these three genes the lowest confidence interval of these OR is higher than 1 (Figure 5.10A and Table 5.15).

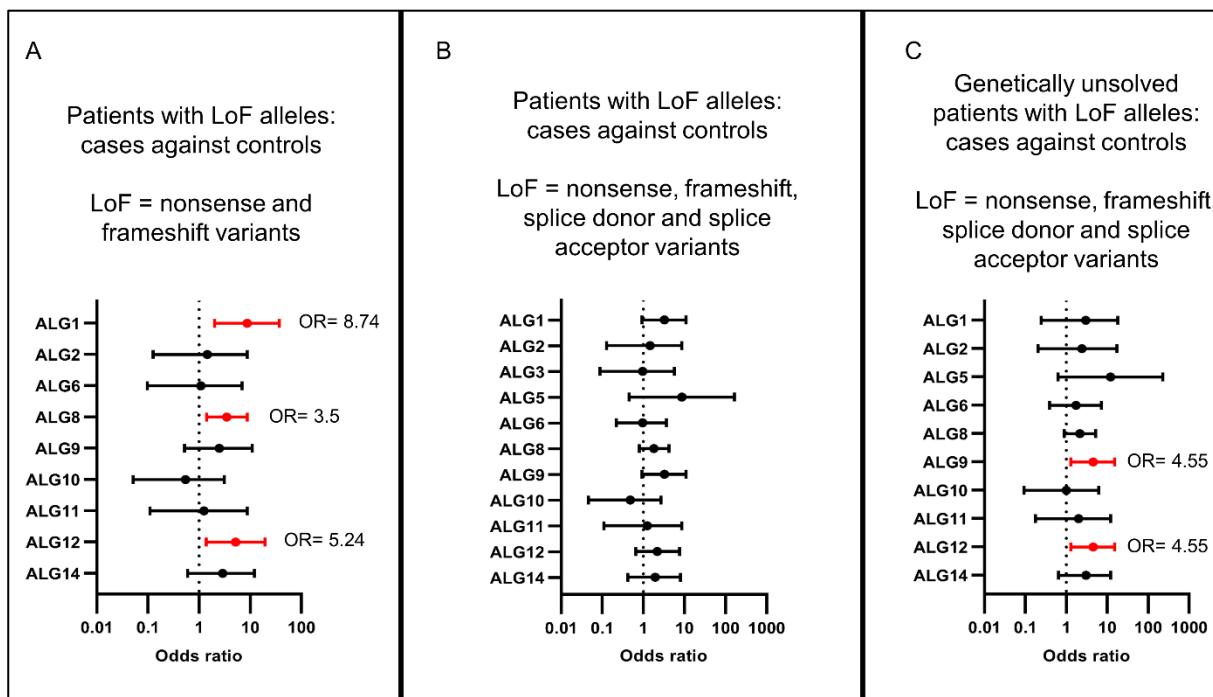


Figure 5.10. Increased odds ratio (OR) of having a predicted loss-of-function alleles (LoF) allele in cases against having a LoF allele in controls, in *ALG1*, *ALG8*, *ALG9* and *ALG12*. Several forest plots have been created showing the odds ratio (OR) with its 95% confidence interval for each of the *ALG* genes. A: Forest plot showing the odds ratio of having LoF allele in cases (1922) against having a LoF allele in controls (16773), for each of the *ALG* genes. I considered nonsense and frameshift alleles as LoF alleles. B: Forest plot showing the odds ratio of having LoF allele in cases (1922) against having a LoF allele in controls (16773), for each of the *ALG* genes. I considered nonsense, frameshift alleles, splice donor and splice acceptor alleles as LoF alleles C: Forest plot showing the odds ratio of having LoF allele in genetically unsolved cases (1205) against having a LoF allele in genetically unsolved controls (14585), for each of the *ALG* genes. I considered nonsense, frameshift alleles, splice donor and splice acceptor alleles as LoF alleles. The corresponding allele counts and OR for each gene are detailed in Table 5.15. When an OR equals to 0 or cannot be calculated, it is not represented here (for example: *ALG3*, *ALG5* *ALG7* and *ALG13* are not represented in panel A). An OR with the lowest confidence interval higher than 1 is indicated in red.

In the second analysis, it is observed that in none of the *ALG* genes the odds ratio had a lowest confidence interval higher than 1 (Figure 5.10B and Table 5.15).

And in the third analysis it is observed an odds ratio of 4.55 in *ALG9* and *ALG12*, and in these two genes the lowest confidence interval of these OR is higher than 1 (Figure 5.10C and Table 5.15).

In summary, combining the three analyses it is found an enrichment of LoF alleles in *ALG1*, *ALG8*, *ALG9* and *ALG12* in the population of cases compared to the population of controls, in other words, the odds of having a LoF variant in these 4 genes are higher in cases than in controls.

As a negative control I calculated the proportion of missense alleles in cases and compared it with the proportion of missense alleles in controls in *ALG1*, *ALG8*, *ALG9* and *ALG12*. The proportion of missense alleles is the number missense alleles divided the total alleles. Missense alleles with an allele frequency higher than 0.02 in the case or control populations were excluded because some missense alleles can be very common in the case and control populations, and subsequently, prevent to observe a real enrichment in missense variants (if any).

It is found that none of the differences in the proportion of missense alleles in cases compared to that proportion in controls is statistically significant (Figure 5.11A and Table 5.16). It is observed that in none of these four *ALG* genes the odds ratio had a lowest confidence interval higher than 1 (Figure 5.11B and Table 5.16).

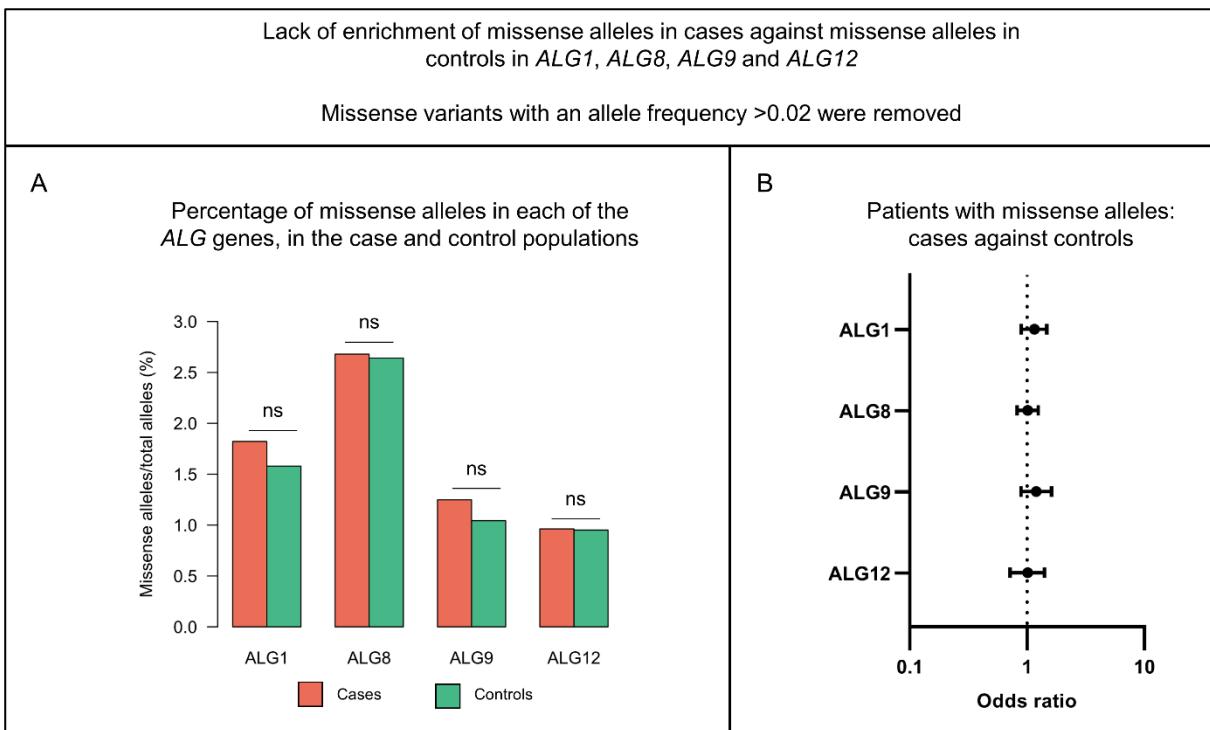


Figure 5.11. Lack of enrichment missense alleles in cases against missense alleles in controls in *ALG1*, *ALG8*, *ALG9* and *ALG12*. A: Bar plot showing the percentage of missense alleles in the case population (3844 total number alleles), shown as red bars against the control population (33546 total number alleles), shown as green bars. Fisher's Exact Test was used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are indicated as ns. B: Forest plot showing the odds ratio (OR) with its 95% confidence interval for each of the *ALG1*, *ALG8*, *ALG9* and *ALG12* genes. This forest plot shows the odds ratio of having missense allele in cases (1922) against having a missense allele in controls (16773), for *ALG1*, *ALG8*, *ALG9* and *ALG12* genes. Missense alleles with an allele frequency higher than 0.02 in the case or control populations were excluded. The corresponding allele counts, p-values and OR for each gene are detailed in Table 5.16.

	Missense alleles in unsolved patients only and excluding variants with an allele frequency >0.02			
Gene	cases	controls	p-value	OR (95%CI)
<i>ALG1</i>	70	530	0.25	1.16 (0.89-1.48)
<i>ALG8</i>	103	886	0.87	1.02 (0.83-1.25)
<i>ALG9</i>	48	350	0.24	1.20 (0.89-1.62)
<i>ALG12</i>	37	319	0.93	1.01 (0.72-1.41)

Table 5.16. Non-enrichment of missense alleles in *ALG1*, *ALG8*, *ALG9* and *ALG12* in the population of cases compared to the population of controls. Missense alleles counts in case and control populations with the p-value of the statistical tests performed and Odds Ratio (OR).

Of note, the *ALG8* and *ALG9* variants found in cases are the same as the ones described in the previous sections 5.2.4 and 5.2.5 (shown in Tables 5.8 and 5.11), and further information about the genomic position of these variants found in cases are shown in Appendix 6.

The genomic position for the *ALG1*, *ALG8*, *ALG9* and *ALG12* variants found in controls is shown in Appendix 7. More detailed information about the *ALG1* and *ALG12* variants found in cases is shown in Table 5.17.

Mutation	Rs ID	CADD	ACMG	Mutation Taster	gnomAD	Gene
c.677del; p.(Pro226LeufsTer10)	NA	NA	Likely pathogenic	Disease causing	NA	<i>ALG1</i>
c.364G>T; p.(Glu122Ter)	NA	NA	Likely pathogenic	Disease causing	NA	<i>ALG1</i>
c.1059C>A; p.(Tyr353Ter)	NA	NA	Likely pathogenic	Disease causing	NA	<i>ALG1</i>
c.1001del; p.(Asn334ThrfsTer15)	rs75924 4819	NA	Pathogenic	Disease causing	2.6E-5 (*v3.1.1)	<i>ALG12</i>
c.732G>A; p.(Trp244Ter)	NA	NA	Pathogenic	Disease causing	NA	<i>ALG12</i>
c.604C>T; p.(Arg202Ter)	rs54764 0376	A:29.6	Likely pathogenic	Disease causing	2.13E-05	<i>ALG12</i>

Table 5.17. Loss-of-function (LoF) variants in *ALG1* and *ALG12* found in the case population in the Genomics England 100,000 Genomes Project database. The transcripts considered for *ALG1* and *ALG12* are: NM_019109.5 and NM_024105.4 respectively. NA, not available.

5.2.7. Selection of control hURECs for genetic studies

Cell cultures of hURECs are a non-invasive source of primary cells that can be used to complement genetic studies (Ajzenberg et al., 2015, Molinari et al., 2020, Ramsbottom et al., 2018). I cultured hURECs from several healthy individuals (controls). It is observed variability between some samples from different control individuals and within different samples of the same control individual. One individual can provide several urine samples and hURECs may only grow in some of them (e.g. Control B), while some individuals may provide different samples in which hURECs grow in all the samples (e.g. Control A), and other individuals may provide one or more samples in which hURECs could not be cultured (e.g. Control H) (Table 5.18).

Individual	Age range (yr)	Sex	Samples	Samples that survived/total samples (%)	Passage (P) in which the sample/s stopped growing
Control A	30-60	M	A (1) and A (2)	2/2 (100)	P4
Control B	18-30	F	B (1), B (2), B (3) and B (4)	2/4 (50)	P1 and P3
Control C	30-60	F	C (1) and C (2)	1/2 (50)	P5
Control D	30-60	F	D	1/1 (100)	P5
Control E	18-30	F	E (1) and E (2)	1/2 (50)	P3
Control F	30-60	F	F	1/1 (100)	P3
Control G	18-30	M	G	1/1 (100)	P1
Control H	18-30	F	H	0/1 (0)	NA
Control I	18-30	M	I	NA	NA
ALG8 Patient	30-60	M	ALG8	1/1 (100)	Frozen at P8 (*)

Table 5.18. Description of the survival the human urine derived renal epithelial cells (hURECs) samples used from Control individuals and the *ALG8* patient. (*) The hURECs of the *ALG8* patient survived to P8 and were frozen to be stored for future analyses. NA, not available.

In terms of cell growth and survival it is observed that at early passages (P1-P3), hURECs from control individuals usually grew well and looked “healthy” (ie: with the expected morphology and shape) (Figure 5.12). I also observed (as early as P2 in some samples) that hURECs may start to look elongated and grow at a slower rate (Figure 5.13) before they stopped growing. At the passage (P) when hURECs stopped growing and subsequently (which can be different depending on the sample and individual) usually the hURECs look elongated and may display extensions of their own cytoplasm (ie: look “unhealthy”) (Figure 5.14).

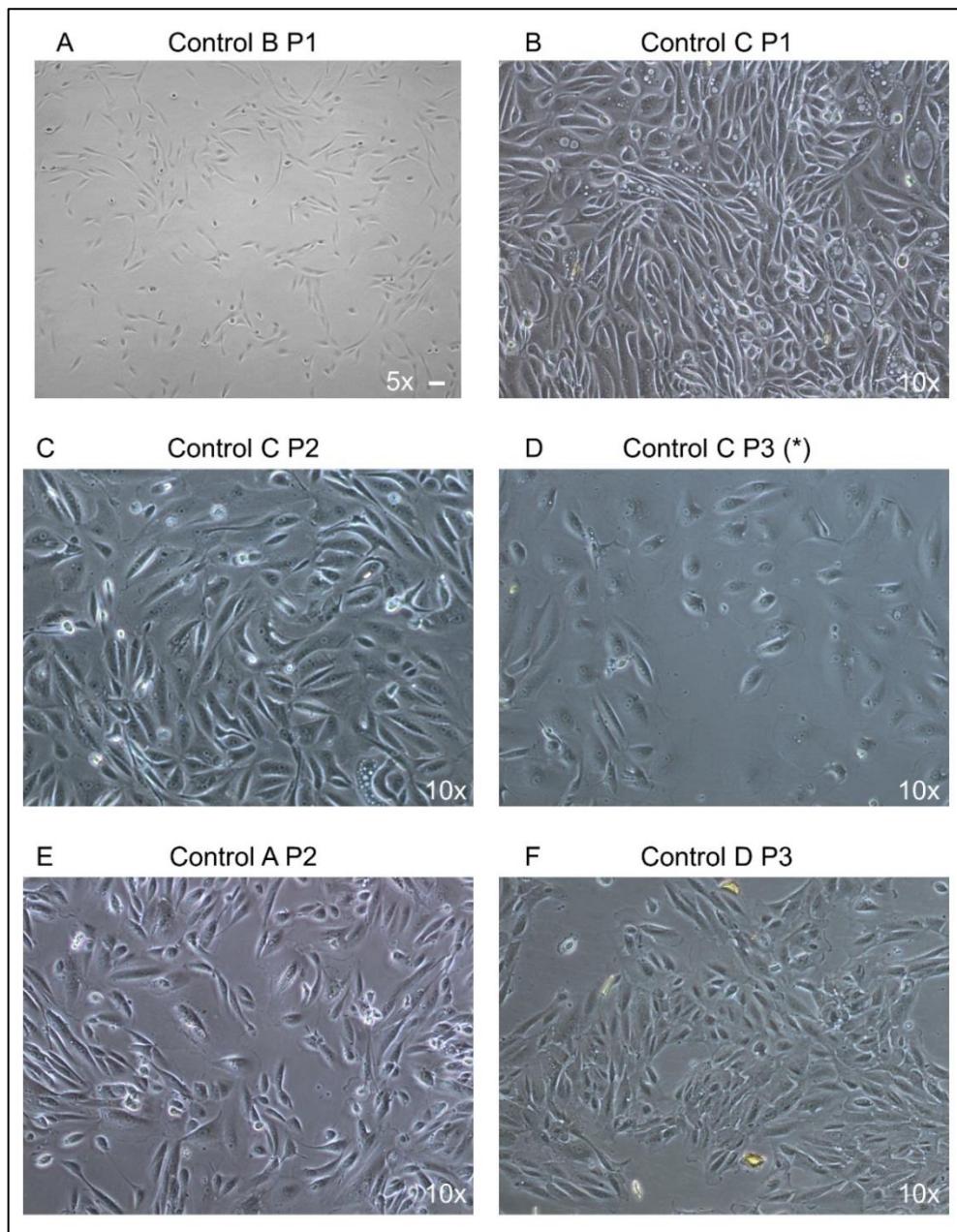


Figure 5.12. Representative light microscopy images of human urine-derived renal epithelial cells (hURECs) that grew well (were passaged for at least one more time) from healthy individuals (Controls). These hURECs continued growing and looked healthy also in the next passage (from the passage in which each of the corresponding pictures was taken). A: Image of hURECs from a healthy individual (Control B) after 24 days of being seeded from urine processing. B-D: Images of hURECs from a healthy individual (Control C) after 26 (B), 30 (C), 34 (D) days of being seeded. (*) hURECs in D do not seem confluent, however this is because there were passaged just three days before the picture was taken. Eventually these hURECs were as confluent as in images B and C. E: Image of hURECs from a healthy individual (Control A) after 30 days of being seeded. F: Image of hURECs from a healthy individual (Control D) after 34 days of being seeded. For image in panel A: Zeiss AxioVert 200M inverted microscope and a magnification objective Plan Neofluar 5x/0.15 Ph1 were used. Scale bar: 20 μ m. The specific control individual and passage number is indicated above each image. For images B-F Nikon Eclipse TS100 inverted microscope was used for visualisation. The magnification objective used was Nikon 10x/0.25.

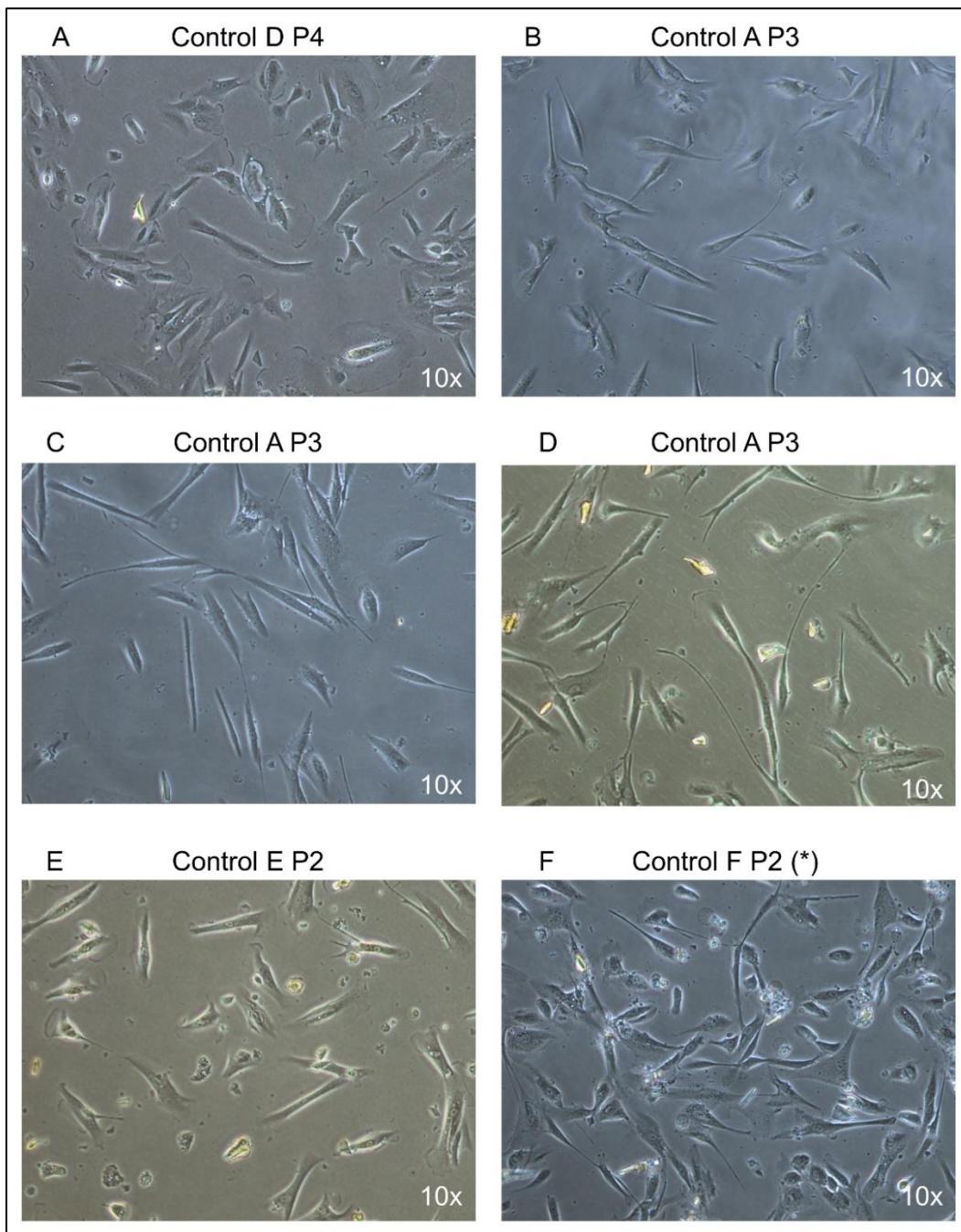


Figure 5.13. Representative images of human urine-derived renal epithelial cells (hURECs) that resembled fibroblasts rather than epithelial cells from healthy individuals (Controls). These hURECs continued growing during the passage in which each of the corresponding pictures was taken however they stop growing in the next passage. A more elongated shape compared to hURECs in Figure 5.12 is observed. A: Image of hURECs from a healthy individual (Control D) after 46 days of being seeded from urine processing. B-D: Images of hURECs from a healthy individual (Control A) after 43 (B and C) and 44 (D) days of being seeded. E: Image of hURECs from a healthy individual (Control E) after 46 days of being seeded. F: Image of hURECs from a healthy individual (Control F) after 45 days of being seeded. (*) Even though these hURECs survived to P3, the slow grow and elongated morphology suggest that they probably would not have survived if they were passaged for P4. The specific control individual and passage number is indicated above each image. Nikon Eclipse TS100 inverted microscope was used for visualisation. The magnification objective used was Nikon 10x/0.25.

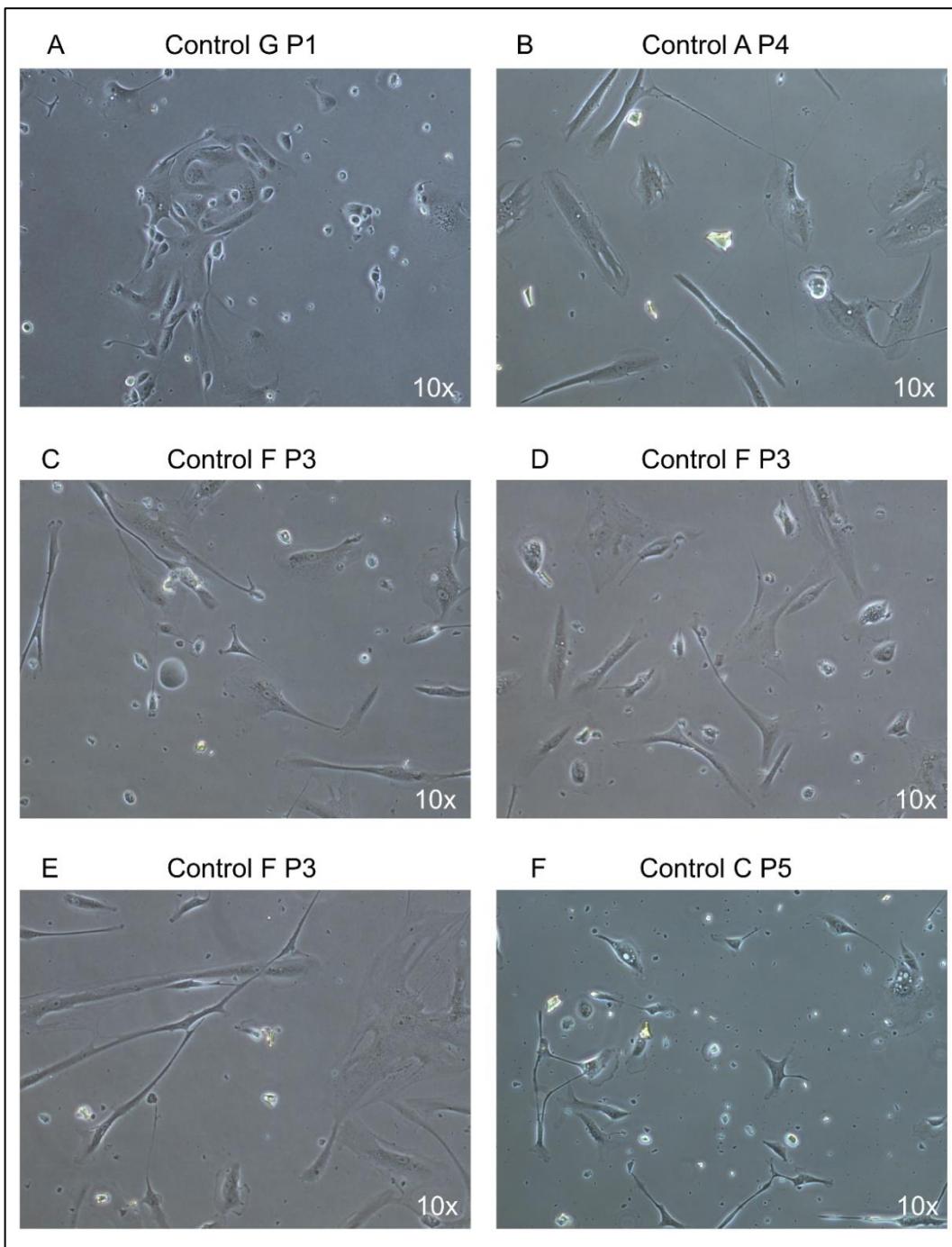


Figure 5.14. Representative images of human urine-derived renal epithelial cells (hURECs) that look fibroblastic (they did not survive the corresponding passage in which the corresponding picture was taken) from healthy individuals (Controls). These hURECs did not continue growing during the passage in which each of the corresponding pictures was taken. A more elongated shape compared to hURECs in Figure 5.12 is observed. Furthermore, these hURECs did not grow anymore, unlike hURECs in Figures 5.10 and 5.11. A: Image of hURECs from a healthy individual (Control G) after 34 days of being seeded from urine processing. B: Image of hURECs from a healthy individual (Control A) after 52 days of being seeded. C-E: Images of hURECs from a healthy individual (Control F) after 52 days of being seeded. F: Image of hURECs from a healthy individual (Control C) after 52 days of being seeded. The specific control individual and passage number is indicated above each image. Nikon Eclipse TS100 inverted microscope was used for visualisation. The magnification objective used was Nikon 10x/0.25.

Considering the hURECs from the nine different controls, it is observed that Control A, Control C and Control D are the best suited to be used for *in vitro* experiments. This is because hURECs from Control A can grow up to passage 4 (P4) and hURECs from Control C and D can grow up to P5 (Figure 5.15 and Table 5.18).

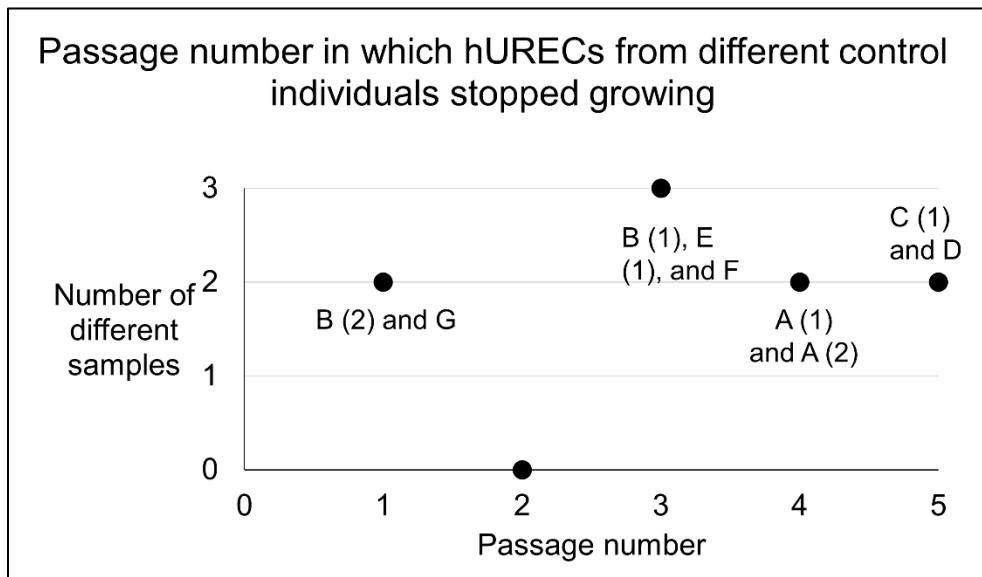


Figure 5.15. Diagram showing the Passage (P) number in which each the hURECs of each control individual stopped growing. For some individuals (A, B, C and E) I was able to culture hURECs from different urine sample obtained at different (ie: different samples from the same individual), this is indicated next to each of the ID of each sample. it can be observed that the hURECs from controls B (2) and G stopped growing at P1. hURECs from controls: B (1), E (1), and F stopped growing at P3. hURECs from control A stopped growing at P4. hURECs from controls C (1) and D stopped growing at P5. Samples in which hURECs did not grow (B (3), B (4), C (2), E (2) and H) are not shown in the diagram for simplicity. Some of the samples shown in the diagram (B (1) and B (2)) may have growth if they had been further passaged however, they were used for other experiments instead.

5.2.8. Characterisation of the ciliary and molecular phenotype of hURECs from a patient with a nonsense mutation in ALG8

I compared the hURECs of a patient with cystic kidney disease (Patient 7-ALG8) who presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter) (Figure 5.16) with the hURECs on Control A (sex-matched control).

The corresponding *in vitro* experiments consisted in immunofluorescence (IF) to stain the cilia in the hURECs from the Family 7-ALG8 patient (Figures 5.17-5.18) to be compared with the cilia in the hURECs of Control A (Figures 5.19 and 5.20).

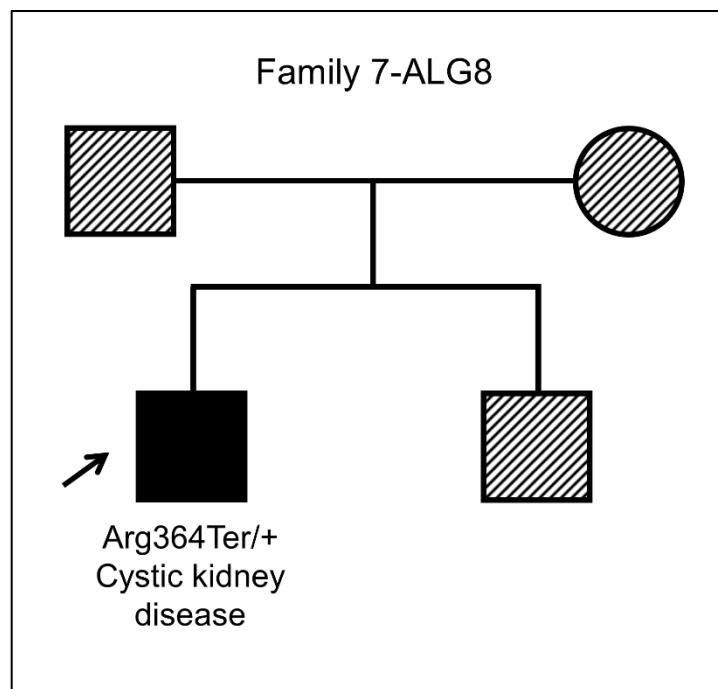


Figure 5.16 Pedigree diagram showing Family 7-ALG8 presenting a disease-causing *ALG8* variant. The proband (patient with genome sequencing available indicated by an arrow; Patient 7-ALG8) presents a heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). There is not genome sequencing information available from the other family members. Clinical information about the parents of the proband is unavailable. Black square indicates affected male. Lined individuals indicate that the clinical information of that individual is unavailable.

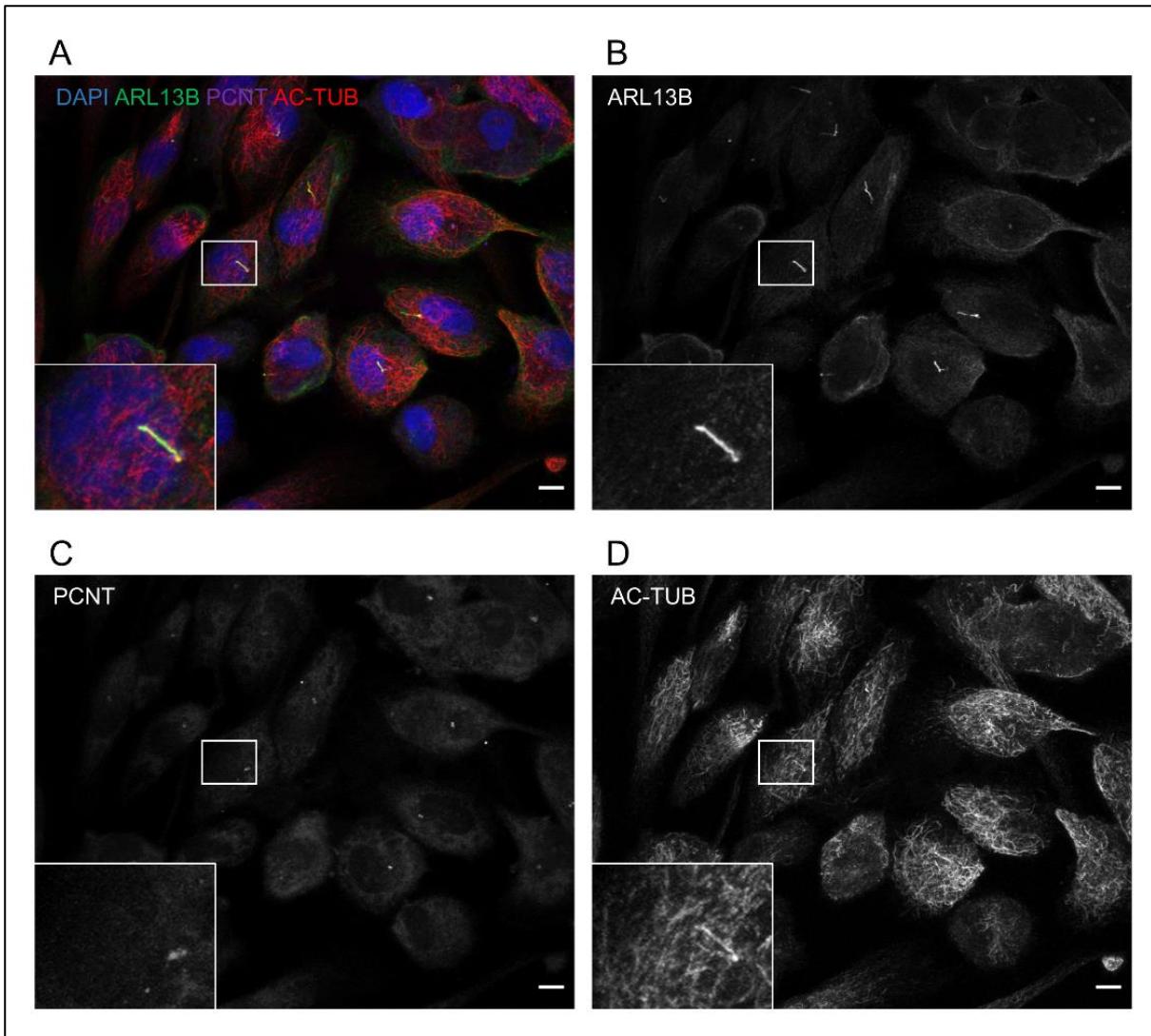


Figure 5.17. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from a Family 7-ALG8 patient with cystic kidney disease. This patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). A: it shows the antibodies: ARL13B (green), which is a ciliary membrane marker, PCNT (purple) which is a centrosome marker and AC-TUB (red) which is a marker for acetylated microtubules. DAPI, a nuclear marker is shown in blue. ARL13B, PCNT and AC-TUB antibodies are also shown in panels B, C and D respectively, in white. hURECs were serum starved for 48 hours. Nikon (A1) confocal inverted microscopy was used for visualisation. AC-TUB, Acetylated Tubulin; PCNT, Pericentrin. Scale bar: 10 μ m.

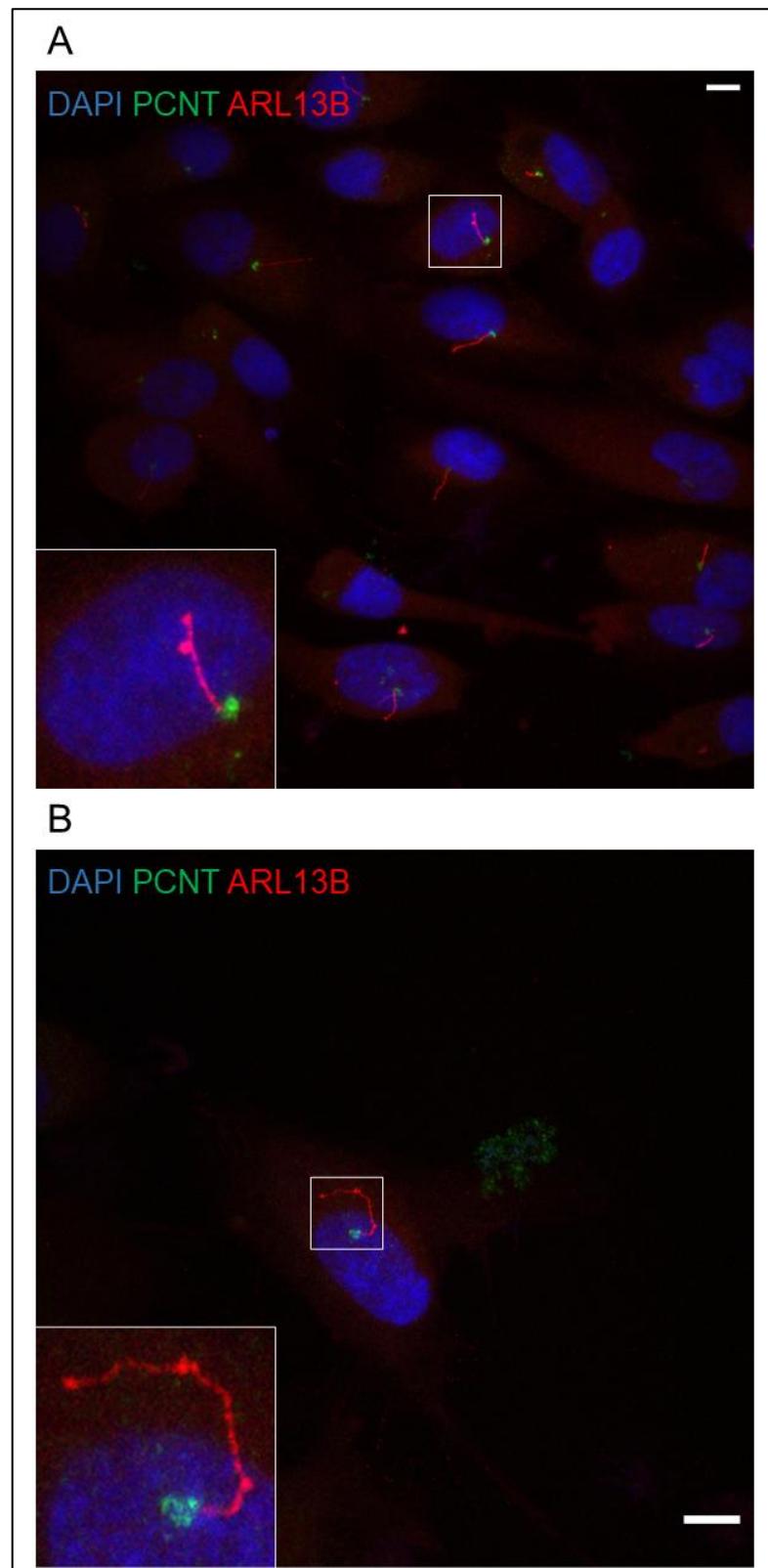


Figure 5.18. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from a Family 7-*ALG8* patient with cystic kidney disease. This patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). A and B: Images show the antibodies: ARL13B (red), which is a ciliary membrane marker and PCNT (green) which is a centrosome marker. DAPI, a nuclear marker is shown in blue. hURECs were serum starved for 48 hours. Nikon (A1) confocal inverted microscopy was used for visualisation. PCNT, Pericentrin. Scale bar: 10 μ m.

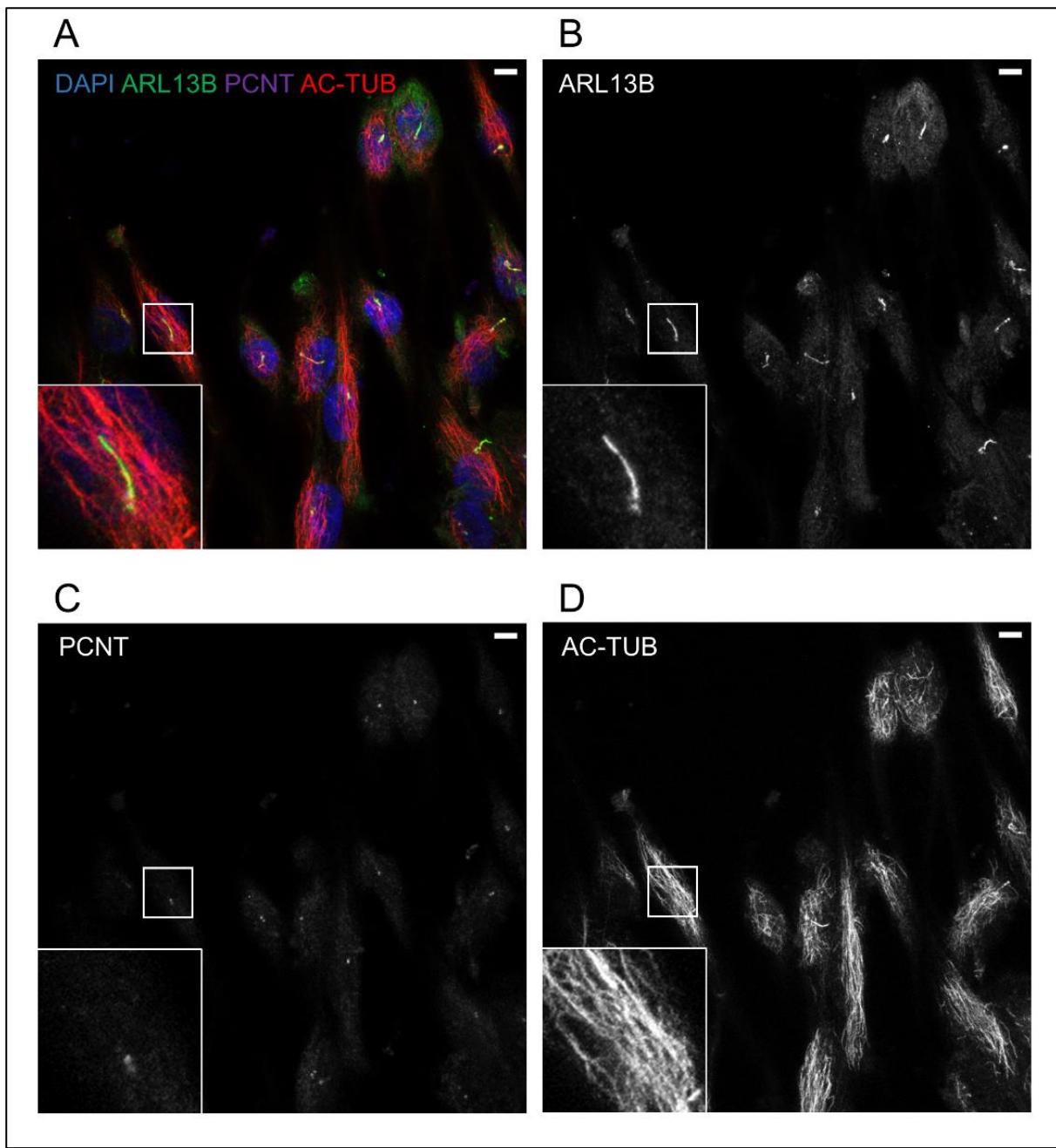


Figure 5.19. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from a healthy control (Control A). A: it shows the antibodies: ARL13B (green), which is a ciliary membrane marker, PCNT (purple) which is a centrosome marker and AC-TUB (red) which is a marker for acetylated microtubules. DAPI, a nuclear marker is shown in blue. ARL13B, PCNT and AC-TUB antibodies are also shown in panels B, C and D respectively, in white. hURECs were serum starved for 48 hours. Nikon (A1) confocal inverted microscopy was used for visualisation. AC-TUB, Acetylated Tubulin; PCNT, Pericentrin. Scale bar: 10 μ m.

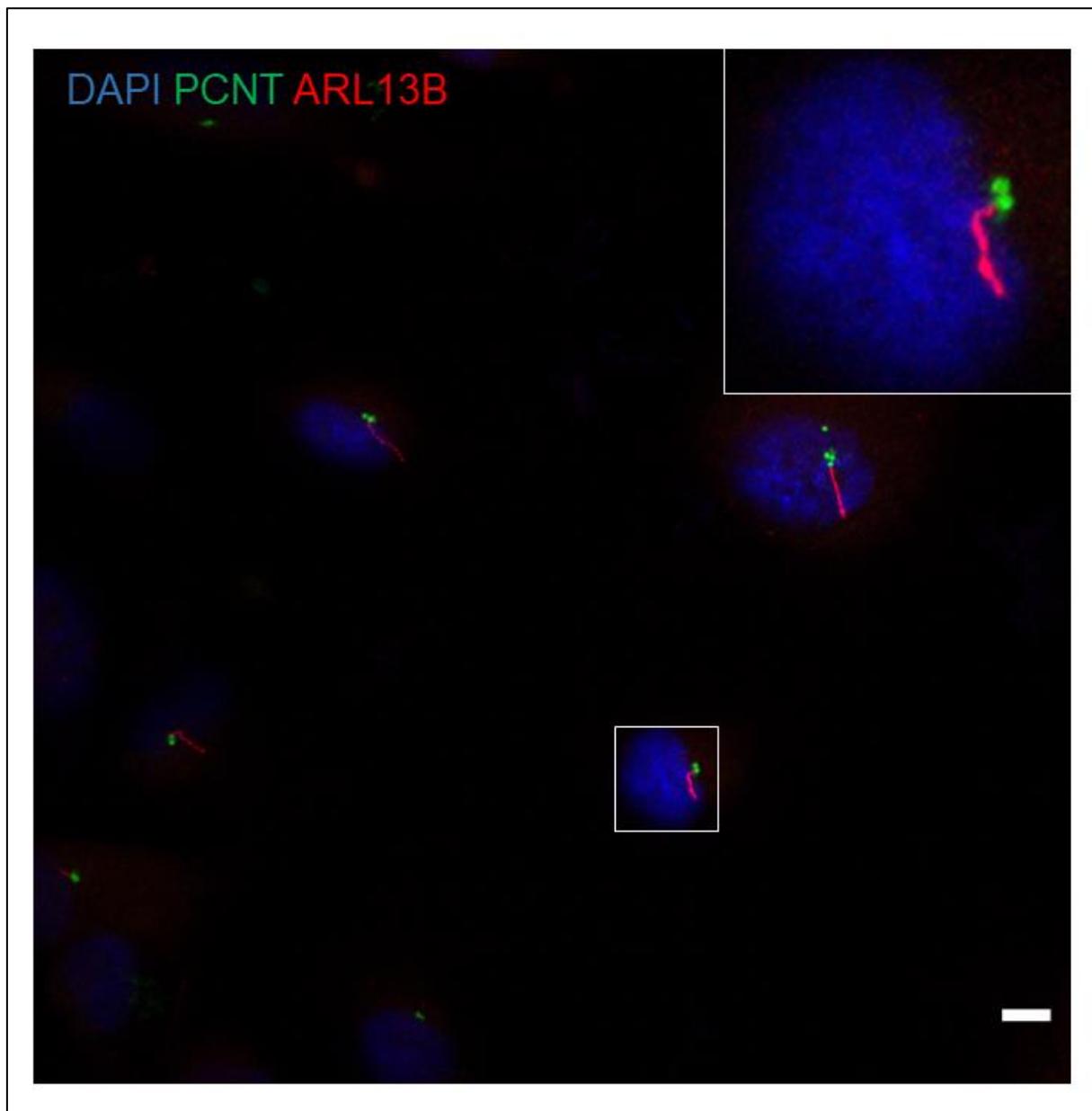


Figure 5.20. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from a healthy control (Control A). It shows the antibodies: ARL13B (red), which is a ciliary membrane marker and PCNT (green) which is a centrosome marker. DAPI, a nuclear marker is shown in blue. hURECs were serum starved for 48 hours. Nikon (A1) confocal inverted microscopy was used for visualisation. PCNT, Pericentrin. Scale bar: 10 μ m.

Regarding the negative controls, in which only the corresponding secondary antibodies were used in the absence of the primary antibodies, I did not observe any noticeable staining in hURECs from Control A (Figure 5.21) or the *ALG8* patient (Figure 5.22).

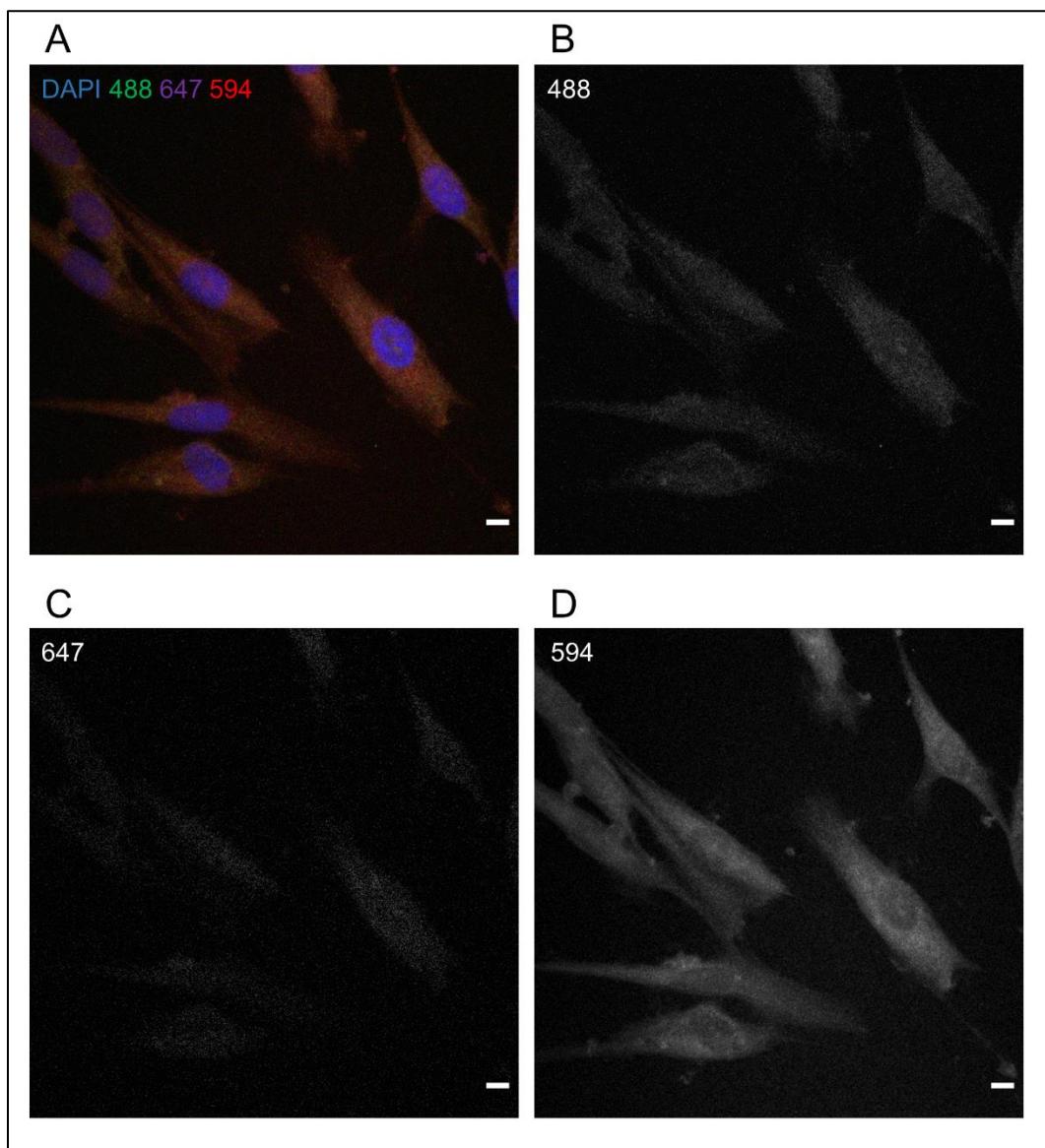


Figure 5.21. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from a healthy control (Control A). It is observed how the cells are not stained in the absence of primary antibodies (A). The corresponding secondary antibodies were used: anti-rabbit 488, anti-rabbit 647, anti-mouse 594. DAPI, a nuclear marker is shown in blue. Secondary antibodies anti-rabbit 488, anti-rabbit 647 and anti-mouse 594 antibodies are also shown in panels B, C and D respectively, in white. hURECs were serum starved for 48 hours. Nikon (A1) confocal inverted microscopy was used for visualisation. Scale bar: 10 μ m.

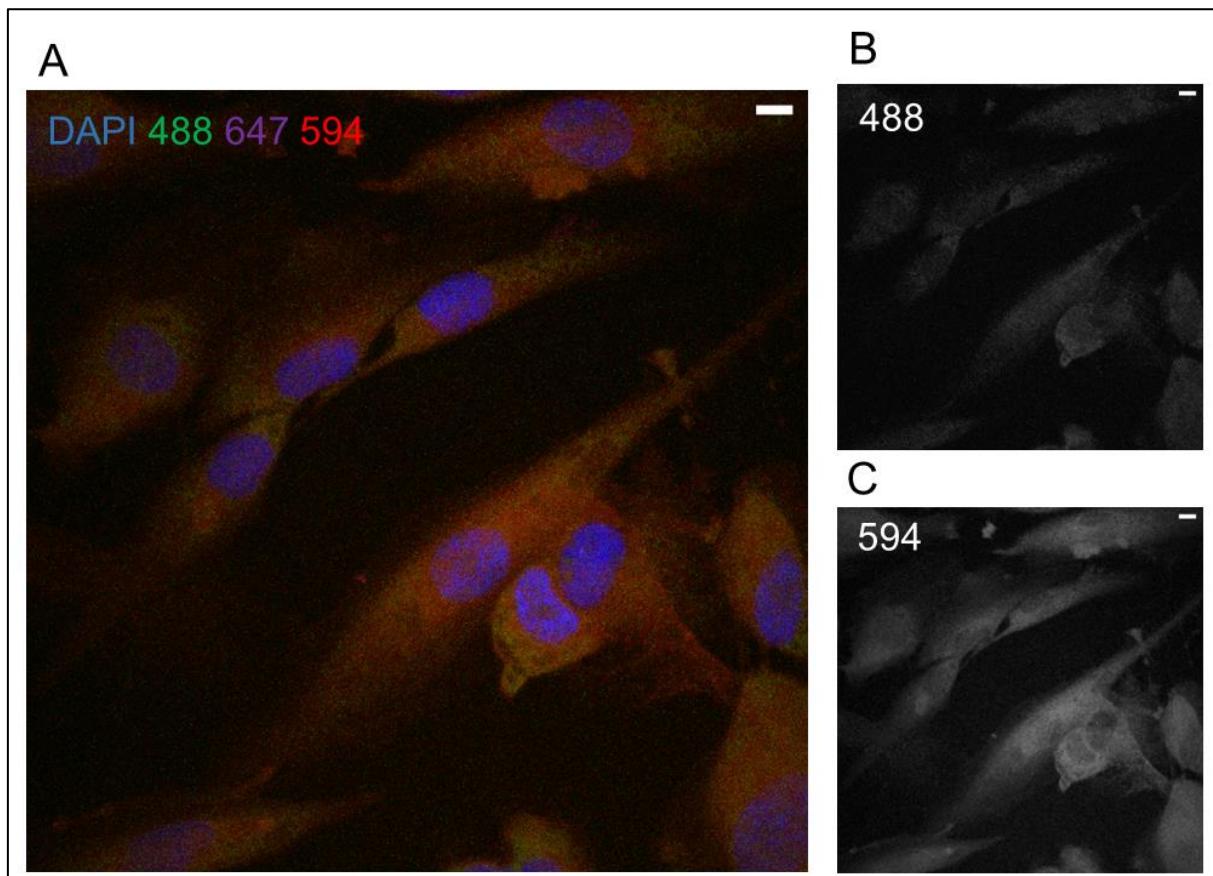


Figure 5.22. Immunofluorescence microscopy in human urine-derived renal epithelial cells (hURECs) from Family 7-*ALG8* patient with cystic kidney disease. This patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). It is observed how the cells are not stained in the absence of primary antibodies (A). The corresponding secondary antibodies were used: anti-rabbit 488 and anti-mouse 594. DAPI, a nuclear marker is shown in blue. Secondary antibodies anti-rabbit 488 and anti-mouse 594 antibodies are also shown in panels B and C, in white. hURECs were serum starved for 48 hours. Nikon (A1) confocal inverted microscopy was used for visualisation. Scale bar: 10 μ m.

In terms of cilia length, using the staining of the cilia marker: ARL13B, Control A was compared with Patient *ALG8* and cilia length was analysed in two different passages: P2 and P3. Using an unpaired *t* test, it is observed that in both P2 and P3 the cilia length of the hURECs of the *ALG8* patient are significantly larger (p-values of 0.0392 and 0.0397 for the comparison in P2 and P3 respectively), compared with the cilia length of the hURECs of Control A (Figures 5.23 and 5.24).

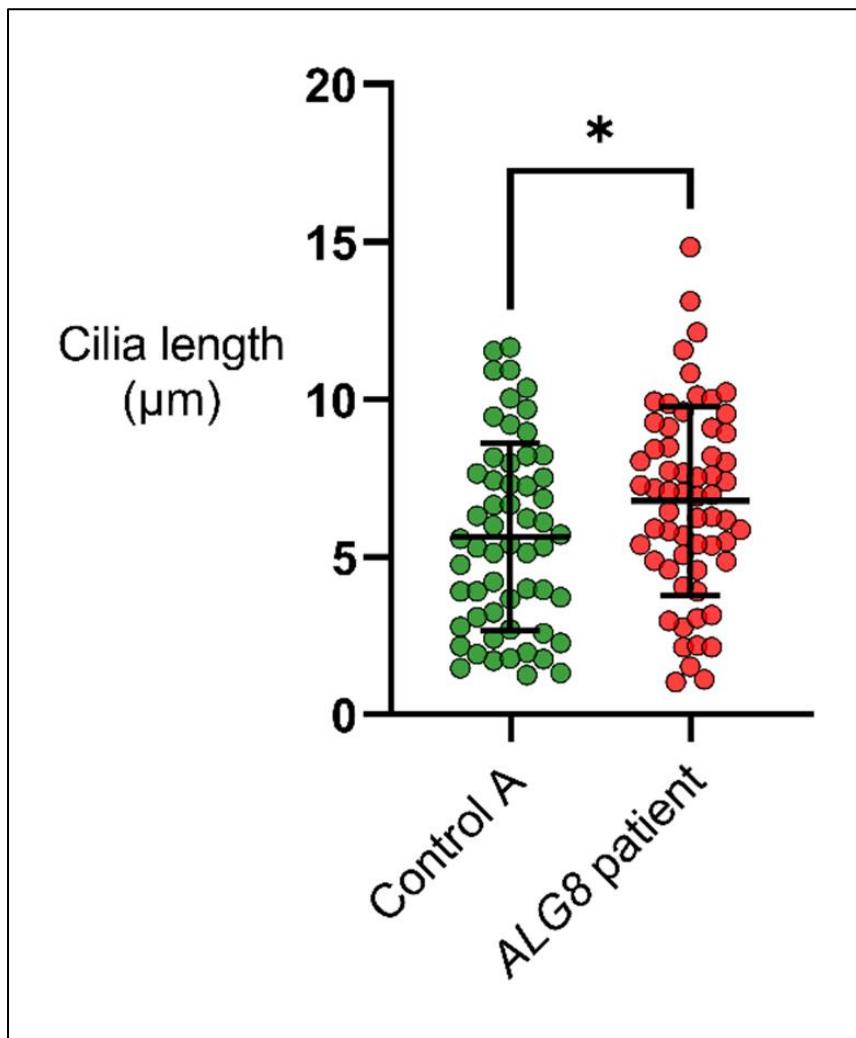


Figure 5.23. Increased cilia length in human urine-derived renal epithelial cells (hURECs) in passage 2 (P2), from a Family 7-ALG8 patient with an *ALG8* mutation. This diagram shows the quantification of ciliary length in the hURECs from a patient with cystic kidney disease (in red) (*ALG8* patient) and a healthy control (Control A). The patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). The mean cilia length in the *ALG8* patient (6.78 μm) is significantly higher compared to the mean cilia length in Control A (5.64 μm) (p-value = 0.04, unpaired *t* test). Cilia length from hURECs of the *ALG8* patient (n=61) and Control A (n=57) was measured using immunofluorescence microscopy and Image J software. hURECs from the *ALG8* patient and Control A were in P2 and serum starved for 48 hours before being fixed. Unpaired *t* test was used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are indicated as ns. Nikon (A1) confocal inverted microscopy was used for visualisation.

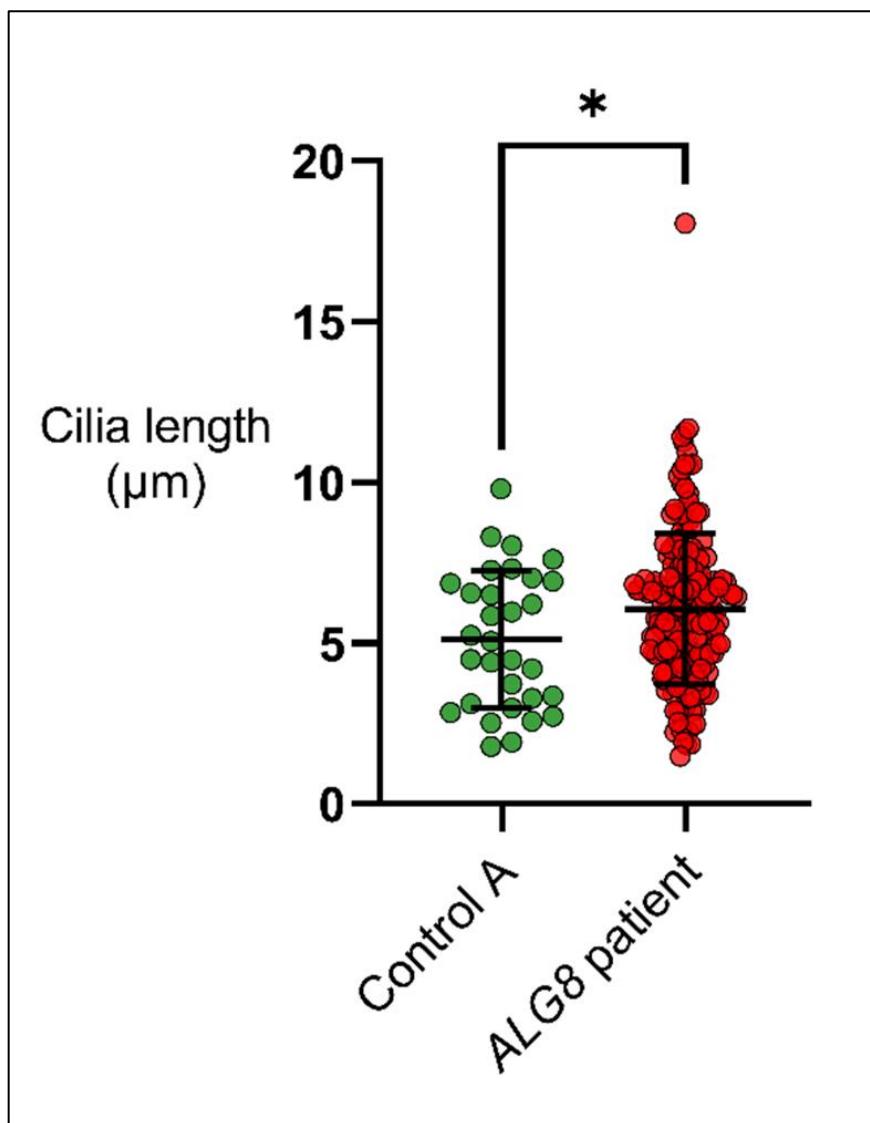


Figure 5.24. Increased cilia length in human urine-derived renal epithelial cells (hURECs) in passage 3 (P3), from a Family 7-*ALG8* patient with an *ALG8* mutation. This diagram shows the quantification of ciliary length in the hURECs from a patient with cystic kidney disease (in red) (*ALG8* patient) and a healthy control (Control A). The patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). The mean cilia length in the *ALG8* patient (6.05 μm) is significantly higher compared to the mean cilia length in Control A (5.12 μm) (p-value = 0.04, unpaired *t* test). Cilia length from hURECs of the *ALG8* patient (n=186) and Control A (n=31) was measured using immunofluorescence microscopy and Image J software. ARL13B staining, which is a ciliary membrane marker was used as a reference. hURECs from the *ALG8* patient and Control A were in P3 and serum starved for 48 hours before being fixed. Unpaired *t* test was used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are indicated as ns. Nikon (A1) confocal inverted microscopy was used for visualisation.

Additionally, Dr Eric Olinger (Newcastle University, UK) provided me with images the hURECs of the *ALG8* patient and Control I. Using the same cilia marker: ARL13B, it is observed that the cilia length of the hURECs of the Family 7-*ALG8* patient are significantly larger (p-value < 0.001, unpaired *t* test), compared with the cilia length of the hURECs of Control I (Figure 5.25).

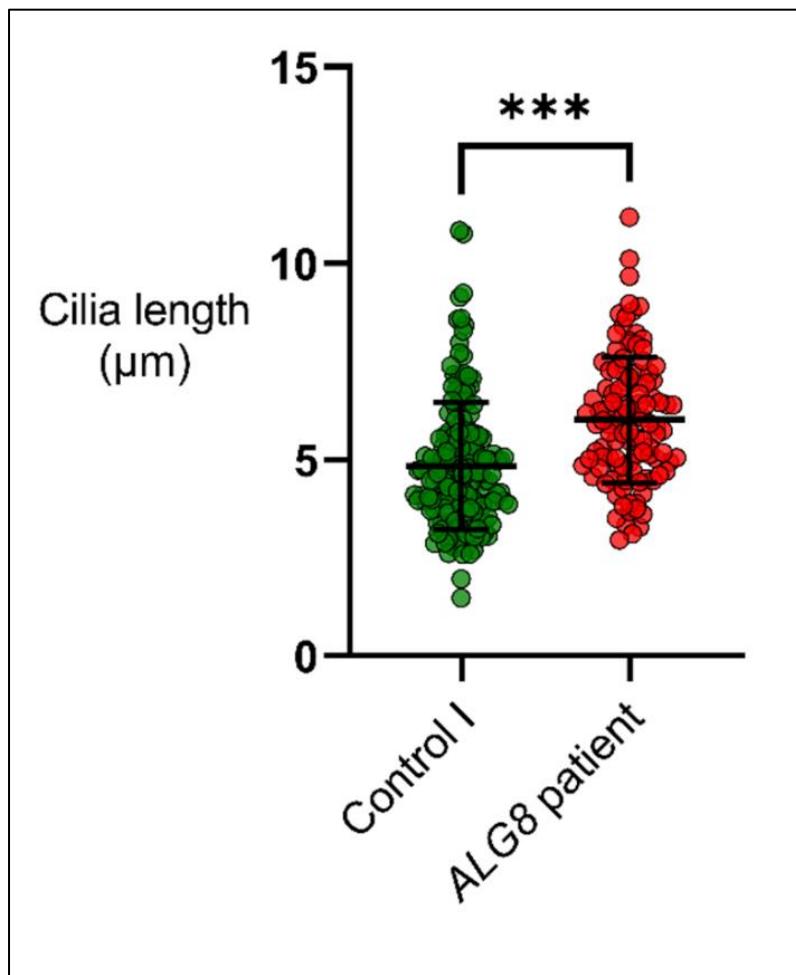


Figure 5.25. Increased cilia length in human urine-derived renal epithelial cells (hURECs) from Family 7-*ALG8* patient with an *ALG8* mutation. This diagram shows the quantification of ciliary length in the hURECs from a patient with cystic kidney disease (in red) (*ALG8* patient) and a healthy control (Control I). The patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). The mean cilia length in the *ALG8* patient (6.01 μ m) is significantly higher compared to the mean cilia length in Control I (4.84 μ m) (p-value < 0.001, unpaired *t* test). Cilia length from hURECs of the *ALG8* patient (n=110) and Control I (n=169) was measured using immunofluorescence microscopy and Image J software. ARL13B staining, which is a ciliary membrane marker was used as a reference. hURECs from the *ALG8* patient and Control I were serum starved for 48 hours before being fixed. Unpaired *t* test was used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are indicated as ns. Nikon (A1) confocal inverted microscopy was used for visualisation. Cell culture and immunofluorescence of the hURECs used for the quantification analysis shown in this Figure 5.25 were performed by Dr Eric Olinger (Newcastle University, UK).

Combining the mean cilia length of these three comparisons performed (Figures 5.23-5.25) between the *ALG8* patient and Controls A and I, it is observed that there is a significant increase (p-value = 0.03, unpaired *t* test) in the mean cilia length in the hURECs of the Family 7-*ALG8* patient compared to Controls A and I (Figure 5.26).

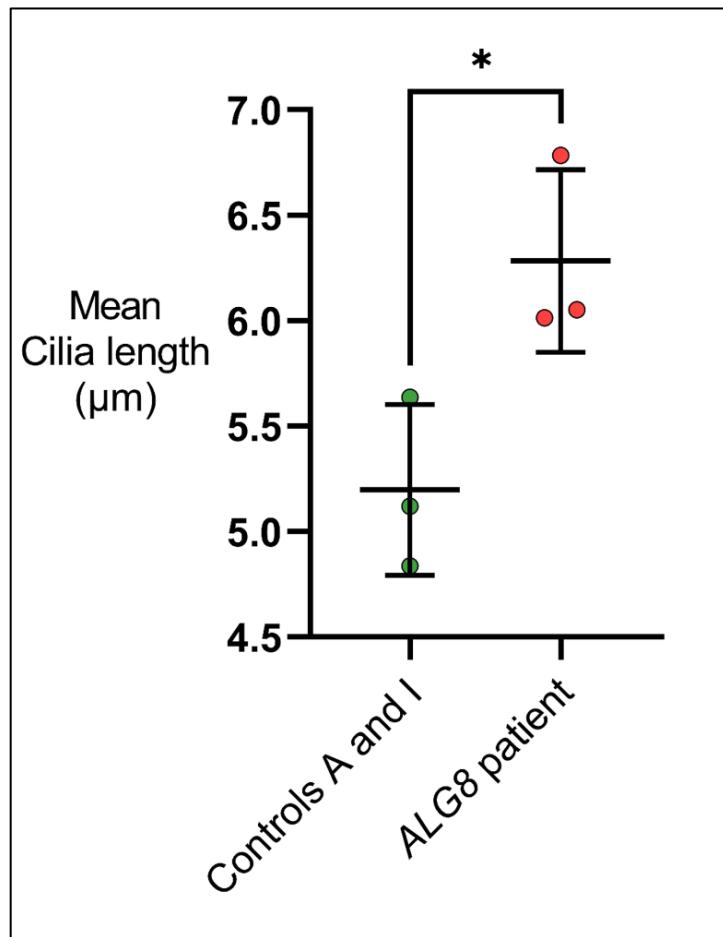


Figure 5.26. Increased mean cilia length in human urine-derived renal epithelial cells (hURECs) from Family 7-*ALG8* patient with an *ALG8* mutation, from 3 different experiments. This diagram shows the quantification of mean cilia length in the hURECs from a patient with cystic kidney disease (in red) (*ALG8* patient) and two healthy controls (Controls A and I). The patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). The diagram shows 3 different mean cilia length in the *ALG8* patient (6.78 μm 6.05 μm 6.01 μm, also shown in Figures 5.23, 5.24 and 5.25 respectively) which are significantly higher compared to the 3 different mean cilia length in Control A and Control I (5.64 μm 5.12 μm 4.84 μm, also shown in Figures 5.23, 5.24 and 5.25 respectively) (p-value = 0.03, unpaired *t* test). Cilia length from hURECs of the *ALG8* patient and Controls A and I was measured using immunofluorescence microscopy and Image J software. ARL13B staining, which is a ciliary membrane marker was used as a reference. hURECs from the *ALG8* patient and Controls A and I were serum starved for 48 hours before being fixed. Unpaired *t* test was used to determine significance. *, ** and *** represent p-value < 0.05, p-value < 0.01 and p-value < 0.001 respectively. Not significant comparisons (p-value > 0.05) are indicated as ns. Nikon (A1) confocal inverted microscopy was used for visualisation.

Cell culture of hURECs also allowed to extract RNA from the hURECs of *ALG8* patient and Control A. RNA from different cell passages was extracted to measure the gene expression of *ALG8* and genes that it is suspected can be affected downstream *ALG8*. I measured the expression of *ALG8*, *PKD1* and *PKD2* comparing gene expression in *ALG8* patient against Control A (Figures 5.27-5.39).

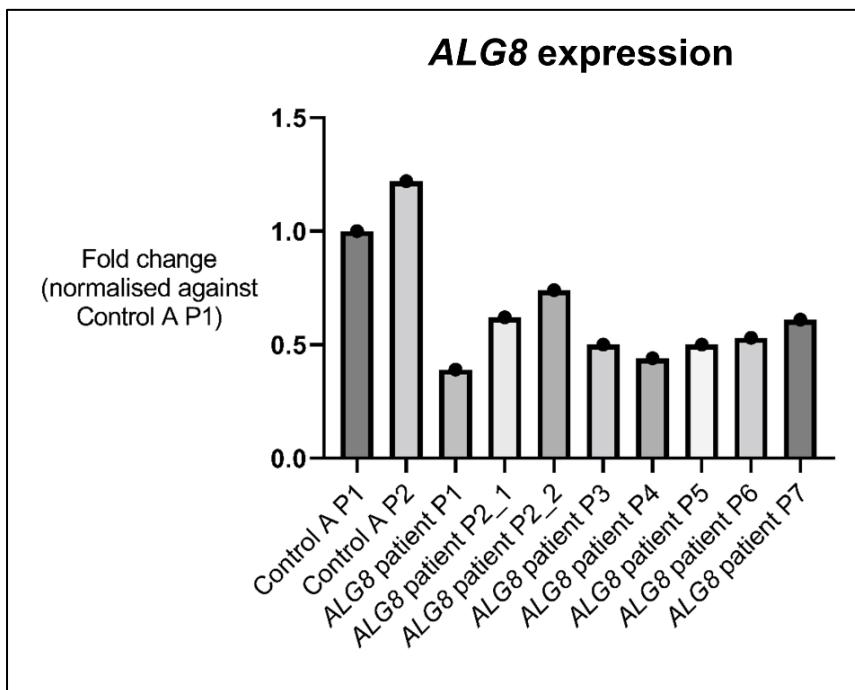


Figure 5.27. *ALG8* expression in human urine-derived renal epithelial cells (hURECs) from Family 7-*ALG8* patient. This diagram shows the *ALG8* gene expression in hURECs from a patient with cystic kidney disease (*ALG8* patient) and a healthy control (Control A) at different passages (P). The patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). *ALG8* expression was normalised over the sample from hURECs of Control A at P1 (Control A P1). It is observed a decrease in *ALG8* expression in the *ALG8* patient's hURECs regardless of the passage in which the RNA was extracted compared to the control samples (Control A P1 and Control A P2). Of note, transcripts levels in *ALG8* patient P2_1 and *ALG8* patient P2_2 are obtained using from two different RNA extractions from hURECs of the same sample and same passage: *ALG8* patient P2. Transcript levels of *ALG8* were determined by qRT-PCR. *GAPDH*, *HPRT1* and *GUSB* were used for normalisation. No statistical test was performed as the number of biological replicates = 1.

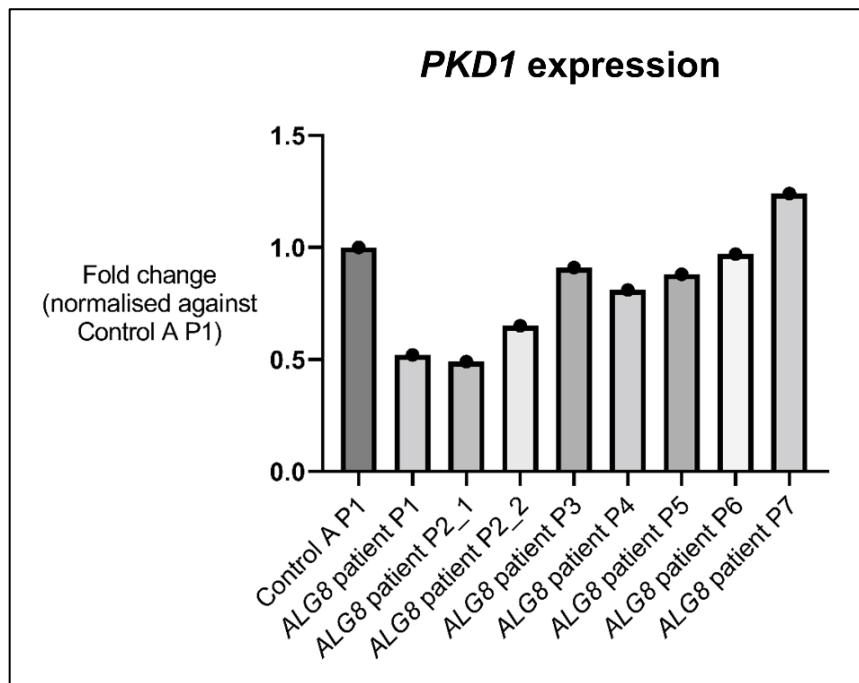


Figure 5.28. *PKD1* expression in human urine-derived renal epithelial cells (hURECs) from a patient with an *ALG8* mutation. This diagram shows the *PKD1* gene expression in hURECs from a patient with cystic kidney disease (*ALG8* patient) and a healthy control (Control A) at different passages (P). The patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). *PKD1* expression was normalised over the sample from hURECs of Control A at P1 (Control A P1). It is observed a decrease in *PKD1* expression in the *ALG8* patient's hURECs at P1 and P2 (*ALG8* patient P1, *ALG8* patient P2_1 and *ALG8* patient P2_2) compared to the control sample (Control A P1). Of note transcripts levels in *ALG8* patient P2_1 and *ALG8* patient P2_2 are obtained using from two different RNA extractions from hURECs of the same sample and same passage: *ALG8* patient P2. Transcript levels of *ALG8* were determined by qRT-PCR. *GAPDH*, *HPRT1* and *GUSB* were used for normalisation. No statistical test was performed as the number of biological replicates = 1.

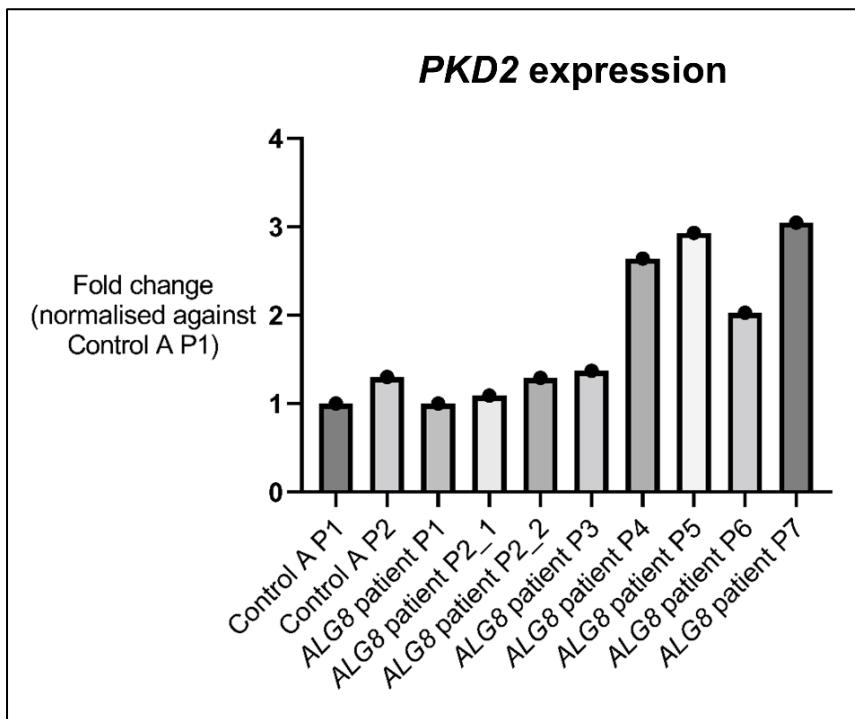


Figure 5.29. *PKD2* expression in human urine-derived renal epithelial cells (hURECs) from a patient with an *ALG8* mutation. This diagram shows the *PKD2* gene expression in hURECs from Family 7-*ALG8* patient with cystic kidney disease (*ALG8* patient) and a healthy control (Control A) at different passages (P). The patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). *PKD2* expression was normalised over the sample from hURECs of Control A at P1 (Control A P1). It is observed a increase in *PKD2* expression in the *ALG8* patient's hURECs at later passages (*ALG8* patient P4, *ALG8* patient P5, *ALG8* patient P6, and *ALG8* patient P7) compared to the control samples (Control A P1 and Control A P2). Of note transcripts levels in *PKD2* patient P2_1 and *ALG8* patient P2_2 are obtained using from two different RNA extractions from hURECs of the same sample and same passage: *ALG8* patient P2. Transcript levels of *PKD2* were determined by qRT-PCR. *GAPDH*, *HPRT1* and *GUSB* were used for normalisation. No statistical test was performed as the number of biological replicates = 1.

ALG8 expression seems to be reduced in the hURECs from *ALG8* patient compared to the hURECs of Control A (Figure 5.27). This may suggest that the *ALG8* mutation: c.1090C>T; p.(Arg364Ter) can affect the *ALG8* expression and suggest haploinsufficiency of the *ALG8* gene. *PKD1* expression seem to be slightly reduced in the hURECs of the *ALG8* patient compared to Control A if only early passages (P1 and P2) are considered (Figure 5.28). However, *PKD1* expression seems to increase in the *ALG8* patient in later passages (P3-P7).

PKD2 expression seems to be similar in the hURECs of the *ALG8* patient compared to Control A if only early passages (P1-P3) are considered. However, *PKD2* expression seems to increase in the *ALG8* patient in later passages (P4-P7) (Figure 5.29). This may indicate that gene expression in the hURECs (in this case of the

ALG8 patient) may be different in early passages compared late passages, however I cannot test this hypothesis as RNA from passages in Control A (other than the ones shown) were unavailable, and more control samples would be needed to support these comparisons. The limitation of this experiment is that I was not able to do technical replicates, this prevents me from establishing any conclusion from this experiment.

It is observed that hURECs of the *ALG8* patient were able to grow at later passages compared to the hURECs of any other control sample (Figure 5.30 and Table 5.18). Comparing Control A and *ALG8* patient, it can be observed how the hURECs of the *ALG8* patient survive for a longer period of time and the number of passages that the hURECs of the patient survive to is higher (P8 in the patient compared to P4 in the control). Furthermore, hURECs of the patient usually require a fewer number of days to reach an appropriate confluence (80-100%) to be passaged (Figure 5.30).

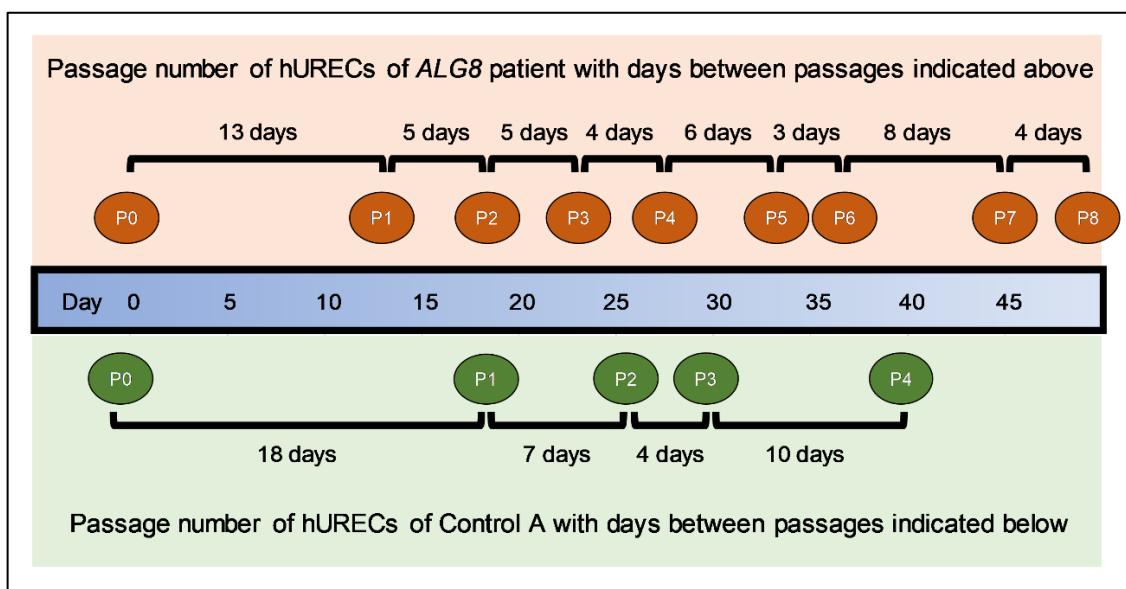


Figure 5.30. Differences in proliferation rate of the human urine-derived renal epithelial cells (hURECs) from a patient with an *ALG8* mutation (*ALG8* patient) compared with a healthy control (Control A). This patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). The day in which urine from *ALG8* patient and Control A was processed, and thus hURECs were first seeded, is the same (day 0). Day 0 is the start of Passage 0 (P0). Bar in the middle shows the specific days in which the hURECs of the *ALG8* patient and Control A were passaged for a next passage. The corresponding passage numbers in which hURECs from the *ALG8* patient and Control A are passaged, are represented above and below the bar respectively. The number of days between each of the passages in the hURECs of the *ALG8* patient and Control A are indicated above and below the passage numbers respectively. hURECs of the *ALG8* patient survived to P8 and were frozen and stored for future analyses. For more information regarding the *ALG8* patient and Control A, see Table 5.18.

I compared the cell viability of hURECs of the *ALG8* patient with the hURECs of Control A. AlamarBlue HS Cell Viability Reagent was used during 7 consecutive days and in different passages. It is observed that at P1 the *ALG8* patient hURECs survive for a longer period of time (Figure 5.31A), while in P2 and P3 the hURECs of Control A survived for a longer period of time (Figure 5.31B and 5.31C). In P4 the hURECs of both *ALG8* patient and Control A survived for consecutive 7 days. Of note, hURECs from the Control A stopped growing at P4, therefore in P5 and P6 cell viability assay could only be done for the *ALG8* patient (Figure 5.31E and 5.31F).

From P5 it can be observed a more elongated morphology of the patient hURECs, which can also display a more disorganised growth pattern, leaving some empty spaces on the surface of the cell culture well (hURECs were cultured using 12-well plates), at the same time than overgrowing at some other regions in the same well (Figures 5.32C-5.32F).

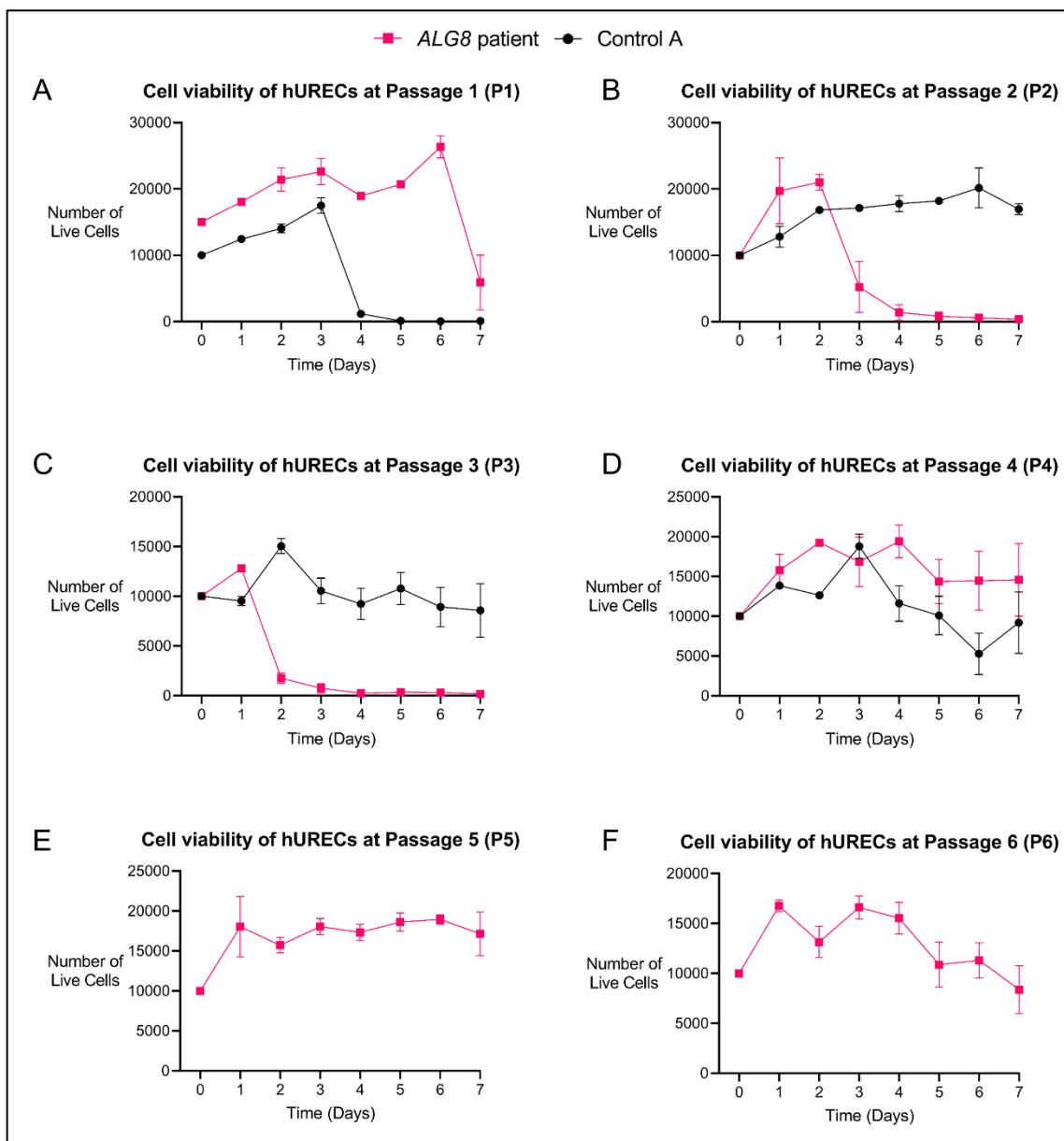


Figure 5.31. Cell viability study of the human urine-derived renal epithelial cells (hURECs) from Family 7-ALG8 patient with an *ALG8* mutation (ALG8 patient) and the hURECs from a healthy control (Control A) in different passages (P). This patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). The cell viability study was done using alamarBlue HS Cell Viability Reagent following the standard protocol. Using a hemocytometer I seeded 10000 cells (except in the case of the hURECs of *ALG8* patient in P1) and using alamarBlue HS Cell Viability Reagent the number of cells was estimated throughout seven consecutive days corresponding to the hURECs of the *ALG8* patient (indicated in red squares) and the Control A (indicated in black circles). The estimated number of live cells in P1, P2, P3, P4, P5 and P6 are shown in panels A, B, C, D, E and respectively. Each value shown as a black circle or red square in the panels (A-F) is the mean of 3 technical replicates. Standard deviation is shown as error bars, some of these cannot be shown if they are shorter than the size of the symbol (circle or square).

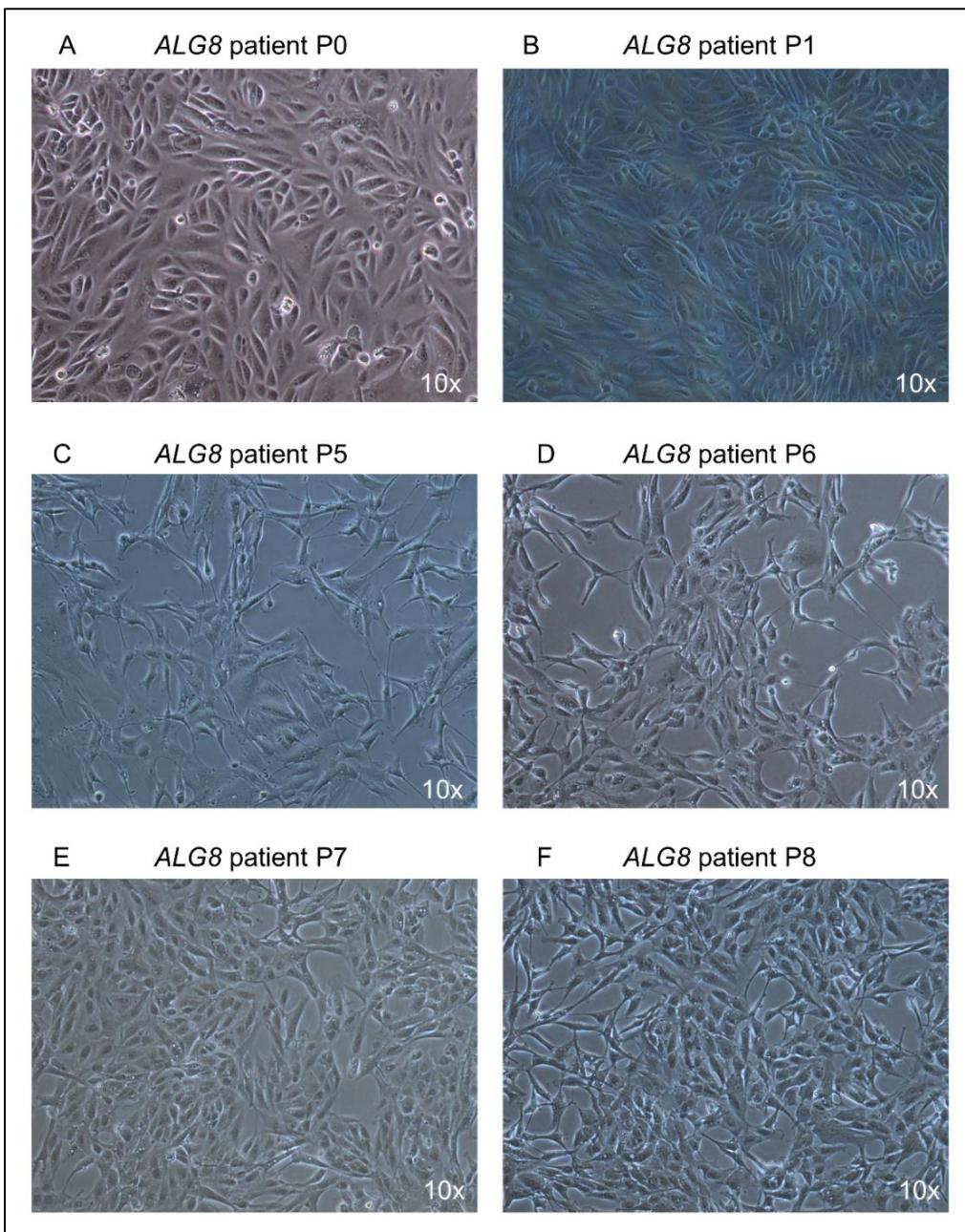


Figure 5.32. Representative images of hURECs from Family 7-*ALG8* patient with an *ALG8* mutation (*ALG8* patient) in different passages (P). This patient is the proband in Family 7 (Figure 5.16), which presents an *ALG8* heterozygous variant: NM_024079.5 (*ALG8*): c.1090C>T; p.(Arg364Ter). A-F: Images of hURECs from *ALG8* patient in P0, P1, P5, P6, P7 and P8 are shown in panels A, B, C, D, E and F respectively. These images were taken after 11 (A), 18 (B), 36 (C), 41 (D), 48 (E) and 53 (F) days of being seeded from urine processing (therefore cells were first seeded at day 0 and the pictures were taken at day 11, 18, 36, 41, 48 and 53 respectively). Passage 1 (P1) started after 13 days from day 0. These hURECs were further passaged on day 18 for P2, on day 23 for P3, on day 27 for P4, on day 33 for P5, on day 36 for P6, on day 44 for P7 and on day 48 for P8 (this also shown in Figure 5.30). hURECs of the *ALG8* patient survived to P8 and were frozen for future analyses (Figure 5.30). The corresponding specific passage number of the hURECs of the *ALG8* patient is indicated above each image. Nikon Eclipse TS100 inverted microscope was used for visualisation. The magnification objective used was Nikon 10x/0.25.

5.3. Discussion

Several genes have been recently described to be associated with the ADPKD phenotypic spectrum such as *GANAB*, *DNAJB11*, *IFT140*, *ALG5*, *ALG8* and *ALG9* (Senum et al., 2022, Huynh et al., 2020). Some of these genes code for proteins functioning in the N-linked-glycosylation pathway. The ADPKD phenotypic spectrum overlaps with the phenotype of other autosomal dominant ciliopathies such as ADPLD and other rare diseases such as ADTKD (Cornec-Le Gall et al., 2018a, Bergmann, 2017).

The Genomics England 100,000 Genomes Project has allowed me to find novel patients associated with ADPKD spectrum, the phenotypic and genotypic description of some of these patients has been published complementing in collaboration with the work of others. 9 patients with heterozygous pathogenic variants in *DNAJB11* (Tables 5.1, 5.2 and 5.3) were described (Huynh et al., 2020), 2 patients with pathogenic variants in *ALG5* (Figure 5.2 and Tables 5.6 and 5.7) (Lemoine et al., 2022), 5 patients with pathogenic variants in *ALG8* (Tables 5.8 and 5.9) and 3 patients with pathogenic variants in *ALG9* (Tables 5.10 and 5.11).

The clinical phenotypes of the patients described in this study with heterozygous pathogenic *DNAJB11* mutations are characterised by non-enlarged kidneys with multiple renal cysts and often also showing extra-renal phenotypes such as with high blood pressure, aortic aneurysms and/or liver cysts (Table 5.1), is consistent with the phenotypes previously described in the literature not included in the Genomics England 100,000 Genomes Project (Cornec-Le Gall et al., 2018a, Huynh et al., 2020). Of note, vascular phenotypes, including aortic aneurysms, were reported in 4 pedigrees (out of the total of 28 pedigrees reported by Huynh et al., 2020) (Huynh et al., 2020). A comparison between the clinical phenotypes found in 27 *DNAJB11*-ADPKD patients (from 6 different families) with the clinical phenotypes found in 42 patients (from 23 different families) with typical ADPKD due to mutations *PKD1* or *PKD2* genes showed that *DNAJB11*-ADPKD patients presented a smaller kidney size and kidney cysts size and had a higher prevalence of cardiovascular disease and a lower prevalence of cardiac valvular defects, among other phenotypic differences (Pisani et al., 2022).

The phenotypes of the cases with heterozygous pathogenic mutations in *ALG8* or *ALG9* found and described in this study show a clinical phenotype characterised by

multiple renal cysts, some of them also showing extra renal phenotypes such as a high blood pressure and/or liver cysts, which also correlates with the patients described in the literature (Besse et al., 2017, Mantovani et al., 2020, Besse et al., 2019).

Several of the variants found have already been described in the literature. Three of the *DNAJB11* pathogenic variants found: c.724C>T; p.(Arg242Ter), c.161C>G; p.(Pro54Arg), c.400delA, p.(Ile134fsTer3) and c.616C>T; p.(Arg206Ter) were also described by others in patients not included in the Genomics England 100,000 Genomes Project, with a clinical phenotype within the ADPKD phenotypic spectrum (Cornec-Le Gall et al., 2018a, Huynh et al., 2020).

The *ALG8* variant: c.1090C>T; p.(Arg364Ter) was previously found in three ADPLD patients with few renal cysts (Besse et al., 2017) and in 85 families the Geisinger-Regeneron DiscovEHR MyCode study (Apple et al., 2023). The *ALG8* variant c.121C>T; p.(Arg41Ter) was found in 7 participants of MyCode study (Apple et al., 2023). The data of MyCode study revealed an increased risk of polycystic kidney disease in participants with loss-of-function variants in *ALG8* (Apple et al., 2023). Two of the *ALG9* mutations that found in patients in the Genomics England 100,000 Genomes Project: c.1472del; p.(Asn491IlefsTer33) and c.1363C>T; p.(Arg455Ter) were also described previously by others in patients with renal disease including renal cysts (Besse et al., 2019). The *ALG12* variant c.1001del; p.(Asn334ThrfTer15) has been described in compound heterozygosis with other pathogenic *ALG12* variants leading to Congenital Disorder of Glycosylation (CDG) (Tahata et al., 2019, Murali et al., 2014).

Regarding the patients with *GANAB* variants, most of them were already genetically solved by variants in *PKD1* or *PKD2*, this reinforces the importance of looking for disease-causing variants in other genes to correctly provide a genetic diagnosis in every patient. To investigate if the *GANAB* variants: c.790A>G; p.(Met264Val) and c.757A>C; p.(Lys253Gln) found in patients without any other putative disease-causing variant in other genes, are responsible for the cystic phenotype found in the corresponding patients, additional *in vitro* and *in silico* analysis would be needed.

It can be suggested that there may be other *ALG* genes that could be associated with the ADPKD phenotypic spectrum, as the proteins they encode function in the same pathway than *ALG8* and *ALG9*, this is why I used the Genomics England 100,000

Genomes Project to detect if there is an enrichment of LoF alleles in any of the *ALG* genes. It can be confirmed that there is an enrichment of LoF alleles in *ALG8* and *ALG9* in a population of with cysts in kidney and or liver (case population).

Considering only nonsense and frameshift alleles as LoF alleles, it is observed a significant enrichment in *ALG1*, *ALG8* and *ALG12* (Figure 5.7 and Table 5.15). When splice donor and splice acceptor alleles are included in the analysis there was not a significant enrichment, this may be because some splice donor and splice acceptor alleles may not lead to a disease phenotype (Figure 5.8 and Table 5.15). When only patients that are genetically unsolved (ie: patients not genetically solved by Genomics England) are included, it is observed a significant enrichment in *ALG9* and *ALG12* (Figure 5.9 and Table 5.15). When selecting only genetically unsolved patients I did not observe a significant enrichment of LoF alleles in *ALG1* or *ALG8*. In this analysis, I may not have observed a significant enrichment in *ALG1* because 2 out of 3 cases were already solved by Genomics England. Regarding *ALG8*, there was not a significant enrichment because the inclusion of splice donor and splice acceptor alleles (13 alleles), all of them were found in the control population. An enrichment on *ALG5* LoF alleles was not detected probably because the low number of cases and controls (<5 patients) with LoF alleles.

The limitations of this enrichment study are: for some patients investigated within the Genomics England 100,000 Genomes Project, their clinical phenotype may be incomplete (Best et al., 2021) where kidney or renal cysts have not been reported. Another limitation is the age dependent penetrance of ADPKD, and the incomplete penetrance of some disease-causing mutations (Barroso-Gil et al., 2021b, Cornec-Le Gall et al., 2018a, Huynh et al., 2020). Subsequently, for some ADPKD patients a disease phenotype is only found in adulthood (Cornec-Le Gall et al., 2018a) or may never display a disease phenotype even when presenting a pathogenic mutation.

Furthermore, for simplicity this study only focussed on the enrichment of LoF alleles, using missense alleles as a negative control. Of note, few heterozygous missense variants in *ALG9* and *ALG5* have been described to be responsible of polycystic renal and liver phenotypes (Lemoine et al., 2022, Besse et al., 2019). Even though biallelic mutations in *ALG1* and *ALG12* have been associated with congenital disorder of glycosylation (Tahata et al., 2019, Murali et al., 2014, Schwarz et al., 2004). To be able to test if these heterozygous *ALG1* and *ALG12* variants are pathogenic, more *in silico* analyses and *in vitro* functional analyses would be needed.

Following the discussion in the previous chapter (Chapter 4), the genome sequencing of the parents to confirm segregation as well as the use of hURECs is proven useful to study putative disease-causing variant and novel candidate genes, and complement genetic studies (Devane et al., 2022, Ramsbottom et al., 2018), as renal epithelial cells are ciliated and there is extensive genetic and phenotypic overlap between primary ciliopathies hURECs can be used to study different primary ciliopathies and different genes associated to one or various primary ciliopathies (Molinari et al., 2018a, Molinari et al., 2020, Srivastava et al., 2017b, Oud et al., 2018).

As a prove of concept of the applicability and benefits of working with hURECs, I characterised the ciliary phenotype and analysed gene expression via immunofluorescence and qPCR respectively, to investigate the molecular disease phenotype in the hURECs of a patient (*ALG8* patient; also described as Patient 7-*ALG8*) with the pathogenic heterozygous *ALG8* mutation: c.1090C>T; p.(Arg364Ter) (Figure 5.16). This patient has been diagnosed with cystic kidney disease which is a phenotype found in other patients with the same pathogenic mutation in heterozygosity and patients with heterozygous mutations in *ALG8* (Besse et al., 2017, Mantovani et al., 2020, Apple et al., 2023).

hURECs of the Family 7-*ALG8* patient and hURECs from different healthy individuals (controls) were cultured. I observed variability in terms of survival between samples, and in term of morphology between passages. Considering Control A as an example, I was not able to culture the hURECs from the samples of Control A for longer than passage 4 (Figure 5.15 and Table 5.18). In addition, before these hURECs stopped growing, the morphology of the cells seemed to be more elongated compared to early passages (e.g. P0, P1 and P2) in which the hURECs had the expected morphology (compare Figure 5.12 against Figures 5.13 and 5.14). This change in morphology can be used as an indicator of cells not being able to become confluent and may stop growing in the passage in which they are observed or the next passage.

Comparing the cilia of the hURECs of the *ALG8* patient against Control A it can be observed longer cilia in the hURECs of the *ALG8* patient (Figures 5.25 and 5.26). Comparing the cilia of the hURECs of the *ALG8* patient with Control I (Figure 5.27).

In the available literature, to my knowledge, cilia length has not been reported in hURECs of ADPKD patients presenting variants in genes recently associated with ADPKD (like *DNAJB11*, *ALG5*, *ALG8*, *ALG9* or *IFT140*) (Besse et al., 2017, Besse et al., 2019, Lemoine et al., 2022, Senum et al., 2022, Huynh et al., 2020).

In mouse cystic cell lines cilia length is longer (Jin et al., 2014) and also in ADPKD mouse kidney (Radadiya et al., 2021). In kidney sections of ADPKD patients, cilia is also longer (Shao et al., 2020a). However, in contrast, in some human cystic kidney cell lines cilia has been described to be shorter (Streets et al., 2020) or there is no change in cilia length (Freedman et al., 2013).

Of note, it has also been described an association between loss of cilia and a reduction of cyst growth in ADPKD (Huang and Lipschutz, 2014). Future work can include assessing percentage of ciliated cells in the hURECs from the *ALG8* patient compared to the percentage of ciliated cells in the hURECs from the healthy control used and from other healthy controls.

In terms of *ALG8* expression, comparing *ALG8* patient against Control A, *ALG8* expression seems to be reduced (Figure 5.29), *PKD1* expression seem to be reduced as well if only P1 and P2 are considered (Figure 5.30). *PKD2* expression seems similar but there an increase in the samples of the *ALG8* patient at later passages (Figure 5.31), which is the time in which the morphology of the *ALG8* hURECs seem to become morphologically different compared to earlier passages (Figure 5.34). These differences between passages of the same individual may reinforce the importance of comparing similar passages of patient and control. It is observed that the cells for of the *ALG8* patient grow for a higher number of passages; however, I only had one control and I did not have technical replicates for these experiments. To my knowledge, persistent proliferation has only been reported in immortalized hURECs (using the thermosensitive SV40 T-antigen) from a NPHP patient with biallelic mutations in *NPHP1* (Garcia et al., 2022).

I cannot establish any conclusion from the gene expression studies shown here, although the reduction of *PKD1* expression observed here would be consistent with other studies performed by others in which there is a reduction of the amount of PC1 protein (the protein coded by *PKD1*) observed in a CRISPR/Cas9 engineered *Alg8*–/– mouse cell line (Fedele et al., 2014, Besse et al., 2017, Corne Le Gall et al., 2018a). It is important to note that PC1 and PC2 biogenesis could be further

investigated via Western Blotting (Besse et al., 2017) to complete the characterisation of these patient-specific hURECs. In particular, future work using hURECs from ADPKD patients can include analysis of protein expression via Western blot to determine if the amount of PC1 and PC2 is affected in the patient-specific hURECs. Additionally, as defects in the maturation of PC1 are associated with ADPKD (Fedele et al., 2014), maturation of PC1 can be assayed visualising via Western blot and Endoglycosidase H treatment, distinguishing the different sizes of the immature and mature fragments of PC1 (Besse et al., 2017, Corne Le Gall et al., 2018a).

The Genomics England 100,000 Genomes Project was used to find patients in *ALG* genes and other genes recently associated with ADPKD such as *DNAJB11*. Nevertheless, more than half of the case cohort of probands with cysts in kidney and/or liver remain genetically unsolved, suggesting that other genes, apart from *PKD1* and *PKD2*, may be responsible for the cystic phenotype of these patients. Alternatively, other *PKD1* and *PKD2* variants may remain undetected by clinicians and researchers with access to the WGS data of these patients.

The enrichment of LoF variants in *ALG8* and *ALG9*, genes previously associated with the ADPKD phenotypic spectrum, in the case cohort of probands with cysts in kidney and/or liver, was confirmed. This approach could be applied in the future, to other genes associated with ADPKD and other renal ciliopathies.

It can be suggested that LoF variants in *ALG1* and *ALG12*, which are genes also encoding proteins functioning in the ER glycosylation pathway (Aebi, 2013, Besse et al., 2017), may be associated with the ADPKD phenotypic spectrum, however further validation using *in silico* and *in vitro* analyses would be needed to confirm this hypothesis.

The use of hURECs can be a valuable non-invasive source of primary cells which can be used to investigate defects in primary cilia in patients with renal ciliopathies (Molinari et al., 2020, Ziegler et al., 2022, Garcia et al., 2022).

The use of hURECs has been suggested as a powerful non-invasive personalised tool for functional studies in inherited renal disorders (Molinari et al., 2020), a useful tool to study regulation of the immune response and regenerative processes in transplanted kidneys (Pizzuti et al., 2023) and a promising experimental model to

study renal epithelial cell functions and renal and metabolic phenotypes caused by certain genetic conditions (Lenzini et al., 2022).

Several studies highlighted the potential of hURECs as a non-invasive source of primary non-transformed cells to study ciliopathies (Ajzenberg et al., 2015), the pathogenicity of disease-causing variants in ciliopathy genes such as *IFT140* (Oud et al., 2018) and *CEP290* (Ramsbottom et al., 2018) and its potential use to investigate strategies based on cell therapy and confirm genetic diagnosis (Pizzuti et al., 2023, Oud et al., 2018, Lang et al., 2013, Pavathuparambil Abdul Manaph et al., 2018).

I used hURECs to investigate the effect of one disease-causing mutation in one patient (*ALG8* patient) with ADPKD, and to complement the genetic findings shown in this chapter regarding a selection of patients with ADPKD within the Genomics England 100,000 Genomes project.

I did not use hURECs from multiple ADPKD patients or patients with other renal ciliopathies because urine samples from these individuals were not available and because using urine samples from multiple patients was not the aim of this project. In contrast, I have achieved to grow and use hURECs from different healthy control individuals and the *ALG8* patient, describing in detail how this was achieved, including the number of passages these hURECs were maintained in cell culture and including information regarding sex and age range of the controls used. Moreover, selecting the control urine samples that were able to provide hURECs able to grow during several passages (ideally up to P4), required a significant amount of time.

From the findings of this results chapter, I can suggest a potential preliminary pipeline to study the molecular disease pathogenesis leading to ADPKD in this patient. Besides, this results chapter provides detailed protocols involved in the use hURECs such as urine processing, cell culture, RNA extraction and immunofluorescence that can complement other protocols already described in the literature (Molinari et al., 2020, Ajzenberg et al., 2015, Srivastava et al., 2017b).

The morphological characteristics in living hURECs have been described in multiple studies (Lang et al., 2013, Molinari et al., 2020, Ziegler et al., 2022, Srivastava et al., 2017b) and correlate with the ones provided in this results chapter. This results chapter provides additional information that also describes in detail the expected morphology of hURECs as well as the expected number of passages hURECs can survive in cell culture.

This results chapter highlights the importance of comparing hURECs in the same cell passage, as it was already done by others comparing hURECs in P2 and P4 (Lenzini et al., 2022, Devane et al., 2022, Oud et al., 2018). The use of hURECs in this results chapter provides images of hURECs in multiple cell passages (images of living hURECs and after being fixed and stained with ciliary and centrosomal antibodies). This is the first study that shows images of living hURECs for 7 different controls, and during multiple passages for 4 of these controls (Controls A, C, D and F).

Regarding the *in vitro* studies based on immunofluorescence to characterise the primary cilia in the hURECs of the *ALG8* patient, I did not use multiple controls and I only used one patient because I did not have the required time and resources to find, process and analyse the hURECs of other ADPKD patients, patients with other renal ciliopathies such as ARPKD or NPHP or healthy age and sex matched controls. In contrast, I was able to compare the mean length of the primary cilia in the hURECs of the *ALG8* patient in three different independent experiments (Figures 5.23-5.26).

There are several studies in which only one control and one patient were used in functional studies using hURECs (Molinari et al., 2020, Srivastava et al., 2017b, Ramsbottom et al., 2018, Molinari et al., 2018a, Devane et al., 2022). Cilia length was compared after a treatment with purmorphamine (a Hh signalling agonist) to rescue the elongated cilia phenotype in the hURECs of a JBTS patient and a healthy age-matched control (Srivastava et al., 2017b). RNA was extracted from hURECs and from whole blood of one patient's parent (who was a heterozygous carrier of a disease-causing synonymous *NPHP3* mutation) and one unrelated control. These studies using RNA, were used to demonstrate the kidney-specific alternate splicing effect of this synonymous *NPHP3* variant leading to end-stage renal disease (Molinari et al., 2018a). Similarly, hURECs have been used to demonstrate the aberrant splicing caused by a synonymous mutation in *PKHD1* in the hURECs of a ARPKD patient, compared with an unrelated healthy control (Molinari et al., 2020).

In another study the effect of a targeted exon skipping therapy was assayed in hURECs from a JBTS patient compared with a healthy gender- and age-matched control. This targeted exon skipping approach was also tested in a mouse JBTS model (Ramsbottom et al., 2018). Another functional study assessed DNA damage response (DDR) and cilia trafficking components in hURECs from a patient with unexplained liver and kidney failure due to *TULP3* mutations, and it was compared with sex and age matched controls at the same passage number, using only one

control in each of these experiments (Devane et al., 2022). hURECs from a patient with NPHP-like nephropathy-1 and an age-matched control were also used to investigate cilia length and the molecular disease pathogenesis of the *XPNPEP3* gene. Those results using hURECs were compared with other results investigating the molecular disease pathogenesis of *XPNPEP3* gene, using patient-derived lymphoblasts and a *Xpnpep3* mutant mice (Tong et al., 2023).

In contrast, other studies have used multiple cases and multiple controls. Garcia et al. used hURECs to investigate the molecular disease mechanisms involved in NPHP-related ciliopathies (NPHP-RC) and to investigate agonists of prostaglandin E2 receptors as potential treatment for NPHP-RC (Garcia et al., 2022). They used hURECs from 11 patients with biallelic mutations in *NPHP1*. They also used hURECs from 5 age-matched control individuals and 4 patients with non-ciliopathy-related chronic kidney disease as control samples. The hURECs from these individuals were immortalized to perform different functional studies, such as transcriptomics, and evaluating percentage of ciliated cells. Those results obtained using hURECs were compared with functional studies performed in *Nphp1* KO mice (Garcia et al., 2022).

Oud et al. compared the ciliary phenotype in hURECs from a Mainzer–Saldino syndrome patient caused by biallelic variants in *IFT140*, with the ciliary phenotype in the hURECs of her unaffected parents and three unrelated and unaffected healthy age-matched controls (Oud et al., 2018). In that study, *in vitro* experiments to investigate cilia biogenesis (analysis of percentage of ciliated cells, mean cilia length, and staining of IFT components) in hURECs were performed in triplicates and were compared against ciliary experiments in CRISPR/Cas9 KO mIMCD3 cell lines (Oud et al., 2018). Moreover, Oud et al., showed that the mean cilia length in hURECs can be variable between different controls and between different passages. The mean cilia length in controls ranged between 3-5 μ m in control individuals (Oud et al., 2018).

hURECs of two *CEP290*-JBTS patients were also used to investigate the molecular disease pathogenesis of *CEP290*. RNA was extracted from hURECS from the two JBTS patients and two unrelated WT controls, and it was found that *UNC119B* expression was upregulated in patients' samples compared with control samples (Cardenas-Rodriguez et al., 2021). Furthermore, hURECs from three patients with Glycogen Storage Disease 1a (GSD1a) and three healthy age-matched controls were used to study the disease pathogenesis of GSD1a (Lenzini et al., 2022).

In 2022, Ziegler et al. used hURECs from 19 patients (10 ARPKD, 4 NPHP and 5 BBS patients) and 39 healthy controls. They suggested that *ex vivo* 3D cell culture of primary hURECs could be used as a non-invasive source to further investigate epithelial cell function in renal diseases, functional consequences of inherited rare diseases and pharmacological interventions (Ziegler et al., 2022).

In this project I used two healthy sex matched controls (Controls A and I) to be compared with the *ALG8* patient, in terms of mean cilia length. The mean cilia length in the hURECs of the *ALG8* patient and Control A, were compared in P2 (6.78 μ m vs 5.64 μ m) (Figure 5.23) and in P3 (6.05 μ m vs 5.12 μ m) (Figure 5.24). These increases in mean cilia length in the hURECs of *ALG8* patient compared with the Control A and Control I were found to be statistically significant. The difference between the *ALG8* patient and Control I were also statistically significant (6.01 μ m vs 4.84 μ m) (Figure 5.25). Considering other studies, the mean cilia length in the two healthy controls, which is about 4-5 μ m, is between the expected range for the ciliary length in hURECs from control individuals, using the same or similar immunofluorescence protocols (Molinari et al., 2020, Srivastava et al., 2017b, Oud et al., 2018).

Differences in mean cilia length between different controls have been reported in the literature (Oud et al., 2018), highlighting the importance of using multiple controls, ideally healthy age, sex and passage matched controls to ensure results obtained are robust and reproducible. Nevertheless, these differences between controls' hURECs do not invalidate all the potentials and advantages of the use hURECs. Moreover, if the molecular disease pathogenesis occurring in ADPKD is associated with differences in mean cilia length it is not fully known (Shao et al., 2020a, Freedman et al., 2013, Streets et al., 2020).

It is known that the disease mechanism occurring in ADPKD is associated with PC1 and PC2 biogenesis (Bergmann et al., 2018, Lanktree et al., 2021, Cornec-Le Gall et al., 2019). The N-linked-glycosylation pathway is involved in the PC1 and PC2 biogenesis (Hu and Harris, 2020, Aeby, 2013, Lemoine et al., 2022). Heterozygous mutations in *ALG8*, cause defects in the maturation and trafficking of PC1, leading to phenotypes within the ADPKD phenotypic spectrum (Apple et al., 2023, Besse et al., 2017).

If I had the required time and resources, I would have performed a robust experimental design that potentially would validate hURECs as a good reproducible *ex vivo* model of renal disease. To do this, I would have investigated the biogenesis of PC1, PC2 and multiple ER N-linked-glycosylation pathway components such as ALG8, DNAJB11, ALG9 and ALG5 via immunofluorescence or Western blot protocols. However, I was not able to optimise and perform these protocols using PC1 or PC2 antibodies, as it was successfully done by others in other studies (Besse et al., 2017, Cornec-Le Gall et al., 2018a). Furthermore, I did not find suitable antibodies for ALG8, DNAJB11, ALG9 and ALG5 staining, to be used in immunofluorescence staining or Western blot protocols.

Regarding the molecular studies of the gene expression of *ALG8*, *PKD1* and *PKD2* (Figures 5.27-5.29) and the Alamar Blue cell viability assay in hURECs of the *ALG8* patient and a healthy control (Figure 5.31), I cannot establish any conclusion from these functional studies because I did not do the corresponding technical or biological replicates because I did not have the required time to do them. However, these results show how RNA was extracted from different passages and suggests the importance of comparing hURECs in the same passage, as it was also suggested by others (Oud et al., 2018, Devane et al., 2022, Lenzini et al., 2022).

This results chapter provides multiple representative images of living hURECs and a detailed description of the protocols for RNA extraction and immunofluorescence staining in hURECs and illustrate the potential uses of hURECS to investigate renal ciliopathies and complement other multiple studies using hURECs (Molinari et al., 2020, Garcia et al., 2022). However, more samples from controls and patients, with at least three technical and biological replicates, are needed to establish any conclusion and to perform a robust experimental design investigating the molecular disease pathogenesis of *ALG8* mutations leading to ADPKD.

Similarly to the approaches followed in other studies (Lenzini et al., 2022, Ziegler et al., 2022), If I had the required time and resources, I would address the variability between controls in various functional analyses, investigate PC1 and PC2 biogenesis, perform gene expression studies of *ALG8*, *PKD1* and *PKD2* and the Alamar blue cell viability assay. I would use hURECs from the *ALG8* patient and at least two additional patients with ADPKD due to heterozygous *ALG8* mutations and at least three healthy age, sex and passage matched controls. I would use three technical replicates in each of the experiments mentioned above. Ideally, I would

apply the same approach to investigate the molecular disease pathogenesis of ADPKD, ARPKD and NPHP, using hURECs from patients with mutations in genes associated with these primary renal ciliopathies (e.g. *PKD1*, *PKD2*, *ALG5*, *NPHP1* or *PKDH1*).

One limitation of the use of hURECs is that these primary cells usually can only grow up to passage 4 or 5 (P4 or P5), unlike immortalised cell lines. This limitation causes an increase in the time and resources required to grow an appropriate number of passages of hURECs and required to have an appropriate number of biological and technical replicates, in comparison with the use of immortalised cell lines.

However, there are remarkable advantages in using hURECs, as they are a non-invasive and easily collectable source of human non-transformed primary cells (Ramsbottom et al., 2018, Bondue et al., 2021). Moreover, their use is cost effective compared with the use of animal models (Ziegler et al., 2022). Splicing assays using patient-specific hURECs can be a less time-consuming and less artefact-prone methodology compared with the use of minigene constructs (Molinari et al., 2020).

One approach used to address the limitations in the use of hURECs is to immortalise patient and control hURECs' samples through the expression of a thermosensitive SV40 T-antigen (Garcia et al., 2022). Other studies have highlighted the potential of complementing and comparing results obtained from hURECs with results obtained from functional studies using reference cell lines, CRISPR/Cas9 knockout cell lines or patient-derived lymphoblasts (Ziegler et al., 2022, Oud et al., 2018, Tong et al., 2023) or mouse or zebrafish models (Tong et al., 2023, Devane et al., 2022, Ramsbottom et al., 2018). These recent studies reflect the increasing use of hURECs and its multiple potential applications to study the molecular genetics and disease mechanisms of renal ciliopathies.

There are multiple questions regarding the use of hURECs that remain unsolved and are out of the scope of this project. Future research studies may address and validate the use of hURECs in established assays as an *ex vivo* cell model of renal diseases, which can be accompanied with additional assays using cell lines or animal models (Ziegler et al., 2022, Tong et al., 2023, Devane et al., 2022). Future studies could also address the molecular mechanisms occurring in hURECs leading to the variability between different passages and between different healthy controls. These molecular mechanisms could be addressed investigating primary cilia characteristics

and gene expression of renal markers (such as megalin and aquaporin 2) and certain genes associated with primary ciliopathies and cell signalling pathways. Future studies could also address the molecular mechanisms leading to the hURECs of some controls growing up to P5 while others are unable to grow to P1. Future studies may lead to advances in precision and preventive medicine such as the development of biomarkers used in hURECs of patients with damaged or transplanted kidneys (Ziegler et al., 2022, Pizzuti et al., 2023).

Apart from to study the biogenesis of primary cilia, hURECs can be used to study expression of certain genes and other non-ciliary phenotypes. It is observed that hURECs from control healthy individuals do not often grow further than P4. To provide an accurate comparison between patient and control hURECs it is important to use hURECs in early passages (not later than P5) before hURECs acquire a fibroblastic morphology and/or stop growing. This correlates with the study of Lenzini et al., in which they suggested that hURECs could be cultured up to P4, maintaining their biological features and could be used as an experimental model to study the molecular mechanisms of GSD1a (Lenzini et al., 2022).

In conclusion, using the Genomics England 100,000 Genomes Project, novel patients and novel pathogenic variants were found in various genes involved in the ER glycosylation pathway: *DNAJB11*, *ALG5*, *ALG8* and *ALG9*.

ALG8 and *ALG9* were validated as genes associated with the ADPKD disease spectrum, with an enrichment of LoF alleles in patients with kidney and/or liver cysts. In addition, it can be suggested that heterozygous LoF alleles in *ALG1* or *ALG12* genes may contribute to genetically solve patients presenting disease phenotypes within the ADPKD phenotypic spectrum.

The use of primary cells allows the possibility of doing a wide array of *in vitro* experiments to study the molecular disease mechanisms in renal ciliopathies at cellular level and can support findings from genomic databases. I suggest that hURECs could be used in future analyses to validate candidate genes for ADPKD such as *ALG1* and *ALG12* and other candidate genes associated with renal ciliopathies.

Chapter 6. Final discussion and concluding remarks

6.1. Complexity of renal ciliopathies

Renal ciliopathies are genetically and phenotypically heterogeneous (Devlin and Sayer, 2019). The phenotypic spectrum of renal ciliopathies is not limited to kidney phenotypes. The phenotypic and genotypic spectrum of some primary ciliopathies may overlap (Lemoine et al., 2022). There are over 30 genes associated with NPHP and autosomal dominant and recessive forms of PKD (McConnachie et al., 2021, Barroso-Gil et al., 2021b, Cornec-Le Gall et al., 2019). In total, there are about 190 genes described to cause ciliopathies. Genetic pleiotropy is occurring in many of the genes associated with NPHP, PKD, JBTS and other primary ciliopathies (McConnachie et al., 2021, Gana et al., 2022, Rozet and Gerard, 2015).

Primary cilia are present in most of the vertebrate cell types and genes associated with primary ciliopathies can be expressed in multiple tissues (Marra et al., 2016, Mitchison and Valente, 2017, Choksi et al., 2014). Expression of these genes can be tissue specific (Devlin et al., 2020, Uhlén et al., 2015). For example, *CEP120* and *ARL3* are two genes associated with JBTS, a model primary ciliopathy, among other multisystemic primary ciliopathies (Roosing et al., 2016, Alkanderi et al., 2018).

These two genes are pleiotropic and show tissue specificity (Powell et al., 2020), In the case of *CEP120/Cep120*, there is expression of *Cep120* in developmental brain in mouse (with lower *Cep120* expression in adult mouse brain compared with mouse embryonic brain) (Xie et al., 2007) and *CEP120* expression in human embryonic and foetal brain.

Some “ciliopathy” genes encode proteins not functioning within the cilia, rather they have non-ciliary roles (Vertii et al., 2015, Hua and Ferland, 2018). Some examples of this are the role of *CEP120* in centrosome biogenesis (Mahjoub et al., 2010) and the role of *DNAJB11* (Cornec-Le Gall et al., 2018a) and *ALG5*, *ALG8* and *ALG9* proteins functioning in the protein glycosylation pathway at the endoplasmic reticulum (Lemoine et al., 2022, Besse et al., 2017, Aebi, 2013).

The non-ciliary roles of ciliopathy genes and the genetic and phenotypic heterogeneity of primary ciliopathies can complicate the establishment of robust genotype-phenotypic correlations for this heterogeneous group of rare inherited diseases (Bachmann-Gagescu et al., 2012, Iannicelli et al., 2010).

Novel primary ciliopathy genes are constantly being described (Shamseldin et al., 2020, Gana et al., 2022, Wheway et al., 2019). Multiple genes have been recently associated with of JBTS and ADPKD, widening the phenotypic and genotypic spectrum of primary ciliopathies (Luo et al., 2021, Lemoine et al., 2022, Lanktree et al., 2021). One of the genes recently associated with ADPKD is *DNAJB11*. The description of the ADPKD-*DNAJB11* phenotype expanded the phenotypic spectrum of renal ciliopathies, in this case, the ADPKD phenotypic spectrum (Cornec-Le Gall et al., 2018a, Huynh et al., 2020, Pisani et al., 2022).

Genotype-phenotype correlations are also challenging to make as the pathogenicity of some variants can be difficult to validate *in silico* and some disease-causing variants are difficult to detect (Best et al., 2021, Wheway et al., 2019). Some disease-causing variants are synonymous variants or deep intronic variants affecting splicing (Olinger et al., 2021, Molinari et al., 2018a, Best et al., 2022b). Hypomorphic alleles have been described in several genes such as *TMEM67* (Otto et al., 2009), *DYNC2H1* (Vig et al., 2020), *PKD1* (Mantovani et al., 2020) and *IFT144* (also called: *WDR19*) (Ishida et al., 2021). For instance, in *IFT144*, which is a gene associated with NPHP, JATD and other ciliary phenotypes (Coussa et al., 2013, Fehrenbach et al., 2014), an hypomorphic variant has been described. This hypomorphic *IFT144* missense variant (L710S) leads to cranoectodermal dysplasia in compound heterozygosis (with a nonsense variant), and it leads to autosomal recessive retinitis pigmentosa in homozygosis (Ishida et al., 2021).

Furthermore, similar genotypes can cause different ciliopathies. For example, in *CEP120*, the missense variant A199P was found in three JATD patients in homozygosis (Shaheen et al., 2015). In contrast, the *CEP120* missense variant V194A, which is located in the same exon, was found in homozygosis in a patient diagnosed with JBTS (a less severe phenotype compared to JATD) (Roosing et al., 2016). Additionally, various studies have proven that the deleterious effect of some nonsense variants, in *CEP290* (Barny et al., 2018, Drivas et al., 2015), *CC2D2A* (Drivas et al., 2015) or *IFT52* (Dupont et al., 2019), can be bypassed via nonsense-associated splicing or basal exon skipping. This mechanism leading to a partially functional protein can help to understand genotype-phenotype correlations and can be exploited to design therapies to bypass the deleterious effect of these mutations.

Often for some genes, such as *CEP120* (Roosing et al., 2016), *CEP164* (Devlin et al., 2022, Chaki et al., 2012), *ARL3* (Alkanderi et al., 2018), *TOGARAM1* (Latour et

al., 2020) or *ALG5* (Lemoine et al., 2022), there is a very low number of patients with disease-causing mutations described. Consequently, the number of studies investigating the disease molecular mechanisms of disease-causing mutations in these genes is limited, as well as the possibility of making robust genotype-phenotype correlations (Bachmann-Gagescu et al., 2015a).

These uncertainties pose many interesting questions. How can the understanding of the genetic and phenotypic complexity of renal ciliopathies be improved? How can the investigation of the molecular disease-mechanisms of ciliopathy genes be improved? How robust genotype-phenotype correlations for ciliopathy genes can be established? How can *in silico* tools and data available be used to find disease-causing variants and novel disease-causing genes and provide a genetic diagnosis to the patients diagnosed with renal ciliopathies that remain genetically unsolved?

6.2. Results from this study, its limitations and future work

Human expression patterns of *CEP120* and *ARL3*, two rare ciliopathy genes associated with JBTS, in human embryonic and foetal tissues were characterised (Powell et al., 2020). My aim to characterise *CEP120*, using the MRC Wellcome Trust Human Developmental Biology Resource (HDBR) at Newcastle University was fulfilled. *CEP120* expression patterns were compared with *ARL3* expression patterns in human embryonic and foetal tissues. It was observed that both genes are expressed in key developmental tissues: brain, kidney and eye, correlating to the clinical phenotypes described in patients with disease-causing mutations in these two genes (Powell et al., 2020).

ARL3 is a protein functioning at the cilium and involved in vesicle trafficking and ciliary signalling (Alkanderi et al., 2018), while *CEP120* is a centrosomal protein involved in centrosome biogenesis (Comartin et al., 2013b, Mahjoub et al., 2010). Some differences in gene expression are found in cerebellum and kidney developmental tissues that may reflect different roles in human development, as a consequence of the different functions of the *CEP120* and *ARL3* proteins. This gene expression study represents a step further towards the understanding of the molecular disease mechanisms of *CEP120* and *ARL3* and the spectrum of clinical phenotypes caused by mutations in these two genes associated with multisystem rare primary ciliopathies (Powell et al., 2020). Analysis of RNA expression of *CEP120* and *ARL3* (Powell et al., 2020) and protein expression of *CEP164* (Devlin et al., 2020) in

human developmental stages have improved the understanding of the molecular disease mechanisms of these ciliopathy genes. There are about 190 ciliopathy genes (Gana et al., 2022, Van Dam et al., 2013), it is important to consider that there are resources publicly available with human and mouse RNA and protein data, including spatial distribution of RNA and proteins, such as the Genotype-Tissue Expression (GTEx) project (The GTEx Consortium, 2020), the Human Protein Atlas Project (Uhlén et al., 2015, Thul and Lindskog, 2018) and the Allen brain map (Daigle et al., 2018).

These databases can be used to further investigate these ciliopathy genes and candidate ciliopathy genes, they can also be used to investigate correlation between RNA and protein expression, which in brain tissues is suggested to be low (Moritz et al., 2019, Tasaki et al., 2022) and integrate the information provided by these resources with additional *in silico* and *in vitro* studies.

It is challenging to find robust genotype-phenotype correlations for ciliopathy genes (Best et al., 2021, Wheway et al., 2019). I investigated, using *in silico* tools and the literature available, *CEP120* and *CC2D2A* ciliopathy genes as potential candidates for targeted exon skipping therapies (Barroso-Gil et al., 2021a). These *in silico* tools to investigate genotype-phenotype correlations in ciliopathy genes include the GTEx Project, Ensembl and the Human Protein Atlas Project. None of the *CEP120* mutations reported in the literature are located in potentially skippable exons, suggesting that *CEP120* is not a good candidate gene to apply exon skipping therapies. In contrast, *CC2D2A* contains seven skippable exons that could be potential targets for exon skipping therapies (with a total of 14 distinct loss-of-function variants reported in the literature). These *CC2D2A* variants are reported in patients diagnosed with JBTS and MKS/ML. Tissue specific exon skipping in the kidney of one of these *CC2D2A* exons (exon 30) was confirmed *in vitro* using RNA isolated from human kidney and human urine-derived renal epithelial cells (hURECs) (Barroso-Gil et al., 2021a).

Novel genes and disease-causing mutations associated with primary ciliopathies are constantly being reported (Shamseldin et al., 2020, Best et al., 2021). The Genomics England 100,000 Genomes Project (Turnbull et al., 2018, Turro et al., 2020) is a valuable resource to find disease-causing variants in patients with rare diseases and find novel genes associated with ciliopathies (Smedley et al., 2021, Devlin et al., 2022, Orr et al., 2023).

Using the Genomics England 100,000 Genomes Project database, I firstly aimed to find novel *CEP120* patients and *CEP120* mutations as this may allow to describe genotype-phenotype associations and extend the *CEP120* phenotypic spectrum. However, I did not find novel patients with biallelic disease-causing variants in *CEP120*. I found heterozygous pathogenic *CEP120* variants, however these variants are not responsible for the phenotype observed in the patients, because no clear segregation following a biallelic autosomal recessive inheritance pattern can be confirmed (Figure 4.15). With the analysis of these heterozygous variants, it has been showed the importance of sequencing parental genomes to confirm (or discard) the segregation of a putative disease-causing genotype.

Secondly, the data available in the Genomics England 100,000 Genomes Project database was used to find patients presenting with biallelic disease-causing variants in genes recently associated with JBTS and monoallelic heterozygous disease-causing variants accounting for clinical phenotypes within the ADPKD phenotypic spectrum. The following patients were genetically solved in this study: 1 patient diagnosed with JBTS with biallelic (compound heterozygous) pathogenic variants in *TOGARAM1* (Figure 4.16) (Latour et al., 2020), 8 patients diagnosed with cystic kidney disease and 1 patient diagnosed with renal tract calcification with heterozygous pathogenic variants in *DNAJB11* (Table 5.1) (Huynh et al., 2020), 2 patients diagnosed with cystic kidney disease with heterozygous pathogenic variants in *ALG5* (Table 5.6) (Lemoine et al., 2022), 5 patients diagnosed with cystic kidney disease with heterozygous pathogenic variants in *ALG8* (Table 5.8) and 3 patients diagnosed with cystic kidney disease with heterozygous pathogenic variants in *ALG9* (Table 5.10).

I found an enrichment of LoF alleles in *ALG1*, *ALG8*, *ALG9* and *ALG12* in a population of patients with cysts in kidney and/or liver compared to a control population of patients without cysts kidney or liver (Table 5.15). *ALG8* and *ALG9* were validated as genes with an enrichment of LoF alleles in patients with kidney and/or liver cysts. Additionally, it can be suggested that LoF alleles in *ALG1* or *ALG12* may contribute to genetically solve patients within the ADPKD phenotypic spectrum.

The description of these patients found in Genomics England 100,000 Genomes Project contributed to investigate the molecular genetics of renal ciliopathies and

explore genotype-phenotype correlations by widening the genetic spectrum of ADPKD.

Additionally, I used human urine-derived renal epithelial cells (hURECs), to complement these findings from the Genomics England 100,000 Genomes Project. In particular, I used the hURECs from a patient presenting a nonsense variant in *ALG8*. It was observed longer cilia in the hURECs of the *ALG8* patient and lower *ALG8* expression compared with the cilia in hURECs from two healthy controls.

The work combining the findings in Genomics England 100,000 Genomes Project with the use of hURECs contributes to a better understanding of the disease mechanisms and pathogenicity of ADPKD. In particular ADPKD caused by mutations in *ALG8*, I suggest that hURECs could be used in future analyses to validate candidate genes for ADPKD, such as *ALG1* and *ALG12* and other candidate genes associated with renal ciliopathies.

There are some limitations in the studies I conducted. Regarding the characterisation of *CEP120* and *ARL3* gene expression patterns in human embryonic and fetal tissue: the study using HDBR and RNAscope technique (Powell et al., 2020), does not include *in vitro* experiments to analyse *CEP120* and *ARL3* expression at protein level and it is limited to certain human developmental stages.

Regarding the approach to investigate potential exon skipping genetic therapies in two JBTS genes, it can be suggested *CC2D2A* as a good candidate for an ASOs-based exon skipping therapy (Barroso-Gil et al., 2021a). It is important to consider that there are some challenges to treat some clinical phenotypes, for instance, the delivery of a treatment aiming to access brain phenotypes would be limited by the brain-blood barrier (Min et al., 2020, Evers et al., 2015). Often phenotypes in ciliopathies occur during development so an early treatment is needed before some tissues, such as the cerebellum, are fully developed (Marzban et al., 2014). Early treatment is more effective in rare diseases such as Duchenne Muscular Dystrophy (Kole and Krieg, 2015, Komaki et al., 2018), in which there are several ongoing exon skipping trials to treat Duchenne Muscular Dystrophy, reflecting the progress, effectiveness and translational potential of exon skipping therapies (Eser and Topaloğlu, 2022). Consequently, potential treatments for ciliopathies with clinical phenotypes occurring during development may need to be delivered within a short therapeutic window in order to be effective (Stokman et al., 2021).

In recent years extensive progress has been made in delivering treatment to tissues like the retina via AAV vectors (Trapani and Auricchio, 2019, Jacobson et al., 2022) and ASOs (Dulla et al., 2021), treating ciliopathies such as LCA (Garanto et al., 2016, Cideciyan et al., 2019, Barny et al., 2019). It has to be considered that even though ASOs can access the kidney (Ramsbottom et al., 2018, Bondu et al., 2022) and retina (Cideciyan et al., 2019, Dulla et al., 2021), the suitability of exon skipping therapies are dependent on the number of patients that are suitable to benefit from the therapy, and the exon (it has to be skippable and dispensable) in which each of the targeted disease-causing mutations is located (Garanto et al., 2016, Ramsbottom et al., 2018). It is challenging to design potential targeted exon skipping therapies for small genes or for genes in which the number of patients with disease-causing variants reported is very low, or for genes in which there are only few skippable or dispensable exons (Barroso-Gil et al., 2021a). Also, for a potential exon skipping genetic therapy to succeed, the functionality of the nearly-full protein needs to be assayed *in vitro* and *in vivo*.

To find and establish robust genotype-correlations in primary ciliopathies is challenging, particularly in the event of very few patients described to be associated with mutations in one particular gene (Best et al., 2021, Smedley et al., 2021), as it has been shown for *CEP120* associated phenotypes (Barroso-Gil et al., 2021a). Some primary ciliopathies such as MKS and JATD involve very severe phenotypes. Consequently, MKS/JATD patients may not be often recruited in WGS databases. Consequently, this may have prevented to find disease-causing biallelic *CEP120* variants accounting for MKS/JATD phenotypes in the Genomics England 100,000 Genomes Project.

At the other side of the severity spectrum of primary ciliopathies, some clinical phenotypes are often so mild that are not detected and/or reported. This could have prevented to find more disease-causing monoallelic heterozygous *ALG8* variants accounting for cystic kidney phenotypes (Besse et al., 2017, Orr et al., 2023). It is important to consider the incomplete penetrance of some disease-causing variants and the age-dependent disease diagnosis of ADPKD (Besse et al., 2017, Besse et al., 2019, Orr et al., 2023). This can also affect to the enrichment of LoF variants study in *ALG* genes. LoF of function variants in *ALG5*, a gene known to be associated with ADPKD, were not significantly enriched in the population with cysts in kidney and/or liver, however this could be because the very low number of patients

with LoF variants in *ALG5*. Also, the approach to detect an enrichment of LoF variants and find disease-causing variants in specific genes does not detect pathogenic hypomorphic or modifier alleles.

The use of hURECs has been suggested as an effective cell model for renal ciliopathies (Srivastava et al., 2017b, Ziegler et al., 2022, Molinari et al., 2018a).

However, it has to be considered that hURECs may not grow in some of the urine samples processed (as it has been shown in Figure 5.15). Besides, even though culturing hURECs from patients and controls often represent an unlimited source of primary cells, the urine of some patients with renal ciliopathies can be limited, furthermore some of these patients may have had a kidney transplant, which can be a limitation for the use of hURECs. Immortalised patient's hURECs have been used by others and could be an alternative and solution to this limitation for certain *in vitro* studies (Garcia et al., 2022).

Some of the laboratory work was temporarily impacted by the disruption caused by the pandemic. During several months in 2020 and 2021, I was not able to perform the planned laboratory experiments, as access to laboratory space and equipment was limited. Besides, the delays in deliveries of some reagents and consumables caused disruption in some of the *in vitro* experiments. The use of mouse models to perform the *in vivo* experiments that were initially planned, had to be substituted by alternative *in vitro* experiments using cell lines and primary cells. The design of alternative *in vitro* and *in silico* experiments during and after the pandemic required a significant amount of time as well as the acquisition of the corresponding skills. This situation was addressed using the guidance provided by the supervisory team and a 3 month extension provided by the Covid-19 Impact Funding Scheme from Newcastle University.

If I had more time and resources, I could have performed a robust experimental design using hURECs, including a wide array of *in vitro* experiments, adding more technical replicates and more samples from controls and patients to the experiments shown in this thesis. Moreover, I could have considered to perform additional *in silico* experiments using the Genomics England 100,000 Genomes Project. Nevertheless, using the guidance provided by my supervisory team, I was able to prioritize experiments and perform a wide array of *in silico* and *in vitro* experiments that led to

the completion of the results chapters of this thesis, combining *in vitro* and *in silico* studies to investigate the molecular genetics of primary ciliopathies.

Future work that can be done to complement the aims achieved. Future work would include, regarding the differences in gene expression of *CEP120* and *ARL3* in human embryonic and fetal tissue (Powell et al., 2020), refining and complementing expression patterns using immunostaining. In particular, to determine if *CEP120* is expressed in the Molecular cell layer (ML), in which Purkinje cells are located (using specific antibodies such as Calbindin 1, a marker of Purkinje cells) (Aldinger et al., 2021), or if it expressed in the external (EGL) and internal (IGL) granule cell layers.

The use primary cells have been extensively used to model diseases, including renal ciliopathies, and complement the findings obtained using cell lines and *in vivo* models (Kaur and Dufour, 2012, Ajzenberg et al., 2015, Ziegler et al., 2022, Ramsbottom et al., 2018, Molinari et al., 2019, Molinari et al., 2020). Others have performed rescue experiments using assays based on cell lines to study *CEP120* disease-causing variants (Joseph et al., 2018, Tsai et al., 2019). Also, the role of *CEP120* in Shh signalling has been previously studied using assays based on cell lines (including NIH-3T3, MEF and hTERT-RPE1) (Breslow et al., 2018, Pusapati et al., 2018, Betleja et al., 2018, Wheway et al., 2015). Consequently, future work can involve investigating, using primary cells, the pathogenic effect of the reported *CEP120* disease-causing variants (a further explanation of this workflow and suggested methodology can be found in Appendices 3 and 4). This work would require optimising plasmid transfection techniques in primary cells, in particular hURECs, using transfection techniques such as electroporation-based methods as a possibly more effective alternative to liposomal and non-liposomal transfection reagents (Hasan et al., 2021, Lin et al., 2013, Tsai et al., 2019).

In addition to ciliogenesis percentage and cilia length, I would study the effect of disease-causing variants on centrosome biogenesis and components of the IFT transport (e.g. IFT88). This *in silico* and *in vitro* approach to investigate the molecular disease mechanisms of disease-causing variants in *CEP120*, can also be applied to other ciliopathy genes associated with similar clinical phenotypes and encoding proteins interacting with *CEP120*, such as *KIAA0586* and *KIAA0753*. Ciliopathy genes may encode proteins with a role in the WNT signalling pathway, such it is the case of *TMEM67* (Abdelhamed et al., 2019). Consequently, this approach could be

used to investigate the effect on the WNT signalling components of *CEP120* disease-causing variants and upon *CEP120* knockdown.

Using the same approach to detect enrichment of LoF variants in *ALG* genes, the Genomics England 100,000 Genomes Project can also be used to validate the enrichment of *ALG1*, *ALG8*, *ALG9* and *ALG12* using a different control population without cysts in kidney or liver. This alternative control population can be created by selecting unaffected relatives of the probands previously selected, to make the “previous” control population of probands without cysts in kidney or liver (only one unaffected relative per family would be selected in families in which there were several unaffected relatives).

Finding novel patients within the Genomics England 100,000 Genomes Project can widen the genotypic and phenotypic spectrum of renal ciliopathies. Multiple aspects of the designed approach can be applied to other genes associated with ADPKD (e.g. *HNF1B*) (following a monoallelic disease inheritance pattern), and to other genes associated with other renal ciliopathies (following a biallelic disease inheritance pattern), such as NPHP genes (e.g. *CEP164*, *CC2D2A* or *CEP290*) or ARPKD genes (e.g. *PKHD1* and *DZIP1L*) (Barroso-Gil et al., 2021b, McConnachie et al., 2021) as well as with other primary ciliopathies and rare diseases.

Using the data available in the Genomics England 100,000 Genomes Project, it can be possible to find additional probands with disease-causing mutations in *CC2D2A*, *ARL3* and *KIAA0586*, which are genes associated with JBTS (Bachmann-Gagescu et al., 2012, Malicdan et al., 2015, Alkanderi et al., 2018), and other ciliopathy genes and genes encoding proteins functioning in the ER glycosylation pathway that can be associated to PKD (Besse et al., 2017). The process of searching disease-causing variants can be automated (systematic), using R and Bash coding languages. This automated process could be possible using the tools, scripts and workflows available in the Genomics England Research Environment (<https://research-help.genomicsengland.co.uk/display/GERE/7.+Analysis+Scripts+and+Workflows>), and having the search of *TOGARAM1* (Latour et al., 2020), *DNAJB11* (Huynh et al., 2020) and *ALG5* (Lemoine et al., 2022) disease-causing variants as a “positive control” of the process.

This potential automated pipeline aims to extract putative disease-causing variants from the VCF files and the variants from the tiering and exomiser prioritization

pipelines of putative disease-causing variants and genes (Best et al., 2022a, Smedley et al., 2021, Turnbull et al., 2018, Robinson et al., 2014). Putative disease-causing variants would be annotated using VEP (Ensembl). Additional information regarding allele frequency (from gnomAD population database and from the Genomics England 100,000 Genomes Project) and ClinVar classification can also be incorporated. This automated pipeline can be applied to a group of patients based on the presence of certain clinical phenotypes (phenotype-to-genotype approach) or to a group of patients based on presence of putative disease-causing variant/s in one particular gene (genotype-to-phenotype approach).

Additional phenotypic information from the GMC exit questionnaire and HPO terms could also be incorporated, and also self-curated lists of disease-causing genes and candidate genes could be considered, to further detect additional putative disease-causing variants. This suggested automated pipeline could be applied to one or multiple patients and would reduce the time required (compared to the lack of use of an automated pipeline) for the detection and validation of the putative disease-causing variants.

Additional *in vitro* functional analyses can be performed if patients' cells were available. For instance, if primary fibroblasts from patients reported in the literature with biallelic disease-causing variants *CEP120* were available. This would allow to investigate how these patient-specific mutations affect cilia and centrosome biogenesis. Additionally, if hURECs from the ADPKD patients with heterozygous putative disease-causing mutations in *ALG12* and *ALG1* were available, these could be used to investigate how these patient-specific mutations affect cilia and PC1 and PC2 biogenesis. As it was mentioned in the discussion of the results chapter 5 (section: 5.3. Discussion), If I had the required time and resources, I would do these experiments using at least three patients and three healthy age, sex and passage matched controls, and three technical and biological replicates in each of the experiments.

6.3. Future perspectives: the use of the Genomics England 100,000 Genomes Project and hURECs to study renal ciliopathies

The Genomics England 100,000 Genomes Project is a large database with WGS and clinical information of patients, this database includes patients with rare diseases, including patients with renal ciliopathies (Best et al., 2021, Wheway et al., 2019). It is

challenging to detect all possible disease-causing variants. For recessive rare primary ciliopathies such as JBTS or ARPKD or NPHP, sometimes the second disease-causing variant is difficult to detect as some of these variants are deep intronic variants or synonymous variants affecting splicing (Olinger et al., 2021, Blakes et al., 2022, Best et al., 2022b). Furthermore, disease-causing variants in some ciliopathy genes, can lead to different ciliopathy phenotypes following both monoallelic and biallelic inheritance disease patterns, such as it is the case of *ARL3* (Ratnapriya et al., 2021), *PKDH1* (Besse et al., 2017) or *IFT140* (Senum et al., 2022).

In the Genomics England 100,000 Genomes Project, there over 1000 of patients diagnosed with cystic kidney disease and about 100 patients diagnosed with rare multisystem ciliopathy disorders, including JBTS and BBS. The analysis of these cohorts of patients with ciliopathy phenotypes allowed to find patients with disease-causing mutations in genes recently associated with ciliopathies such as *TOGARAM1* (Latour et al., 2020) and *ALG5* (Lemoine et al., 2022).

The potential of the Genomics England 100,000 Genomes Project in providing a genetic diagnosis to patients with ciliopathy phenotypes is widely known (Best et al., 2022b, Smedley et al., 2021). Some suggestions to improve the potential of this large database have been made by Best et al. regarding the phenotypic information available and the tiering data, among other areas to facilitate access and analysis of the relevant clinical and genetic data, provide a genetic diagnosis to a higher number of patients, the establishment of genotype-phenotype correlations etc (Best et al., 2022a, Brown et al., 2022).

In line to what it was suggested by Best et al. (Best et al., 2022a), I suggest that the time required to analyse the tiering data would be reduced, if transcript information (Ensembl or NCBI transcript ID) of each of the variants is provided in the tiering data (to facilitate the annotation of each of the variants according to their canonical transcript). Additionally, the prioritisation and segregation analysis of the putative disease-causing variants can be facilitated by the sequencing parental genomes (of the singletons recruited and where it is possible), by providing a more detailed clinical information (more detailed HPO terms) and by facilitating the access (using R scripts) to the more detailed clinical information found in Participant Explorer Research Environment tool.

About 40% of patients with primary ciliopathies do not have a genetic diagnosis. The percentage of genetically solved patients is variable depending on the ciliopathy and sequencing method used (WGS, WES, targeted sequencing), for renal ciliopathies such as ADPKD and ARPKD, this percentage can be higher, up to 80-90% (Lemoine et al., 2022). In summary, the Genomics England 100,000 Genomes Project, including the application has the potential to increase the percentage of ciliopathy patients that are genetically solved. The amount of genetic diagnosis reported, genomes sequenced and WGS and clinical data available is constantly increasing and being updated (Manolio et al., 2019). Also, the pipelines and the advances to provide genetic diagnosis to patients with ciliopathies may be suitable to be applied to other inherited rare diseases.

Recently a systematic pipeline has been developed for the analysis of genetic splicing variants in the Genomics England 100,000 Genomes Project (Blakes et al., 2022). Furthermore, the large database of gnomAD have been used to find genes that are tolerant to LoF variants (Karczewski et al., 2020) and the analysis of the phenotypic and genetic data available in the International Mouse Phenotyping Consortium (IMPC) database identified 32 candidate ciliopathy genes, further studies included the phenotyping of knockout mice of some of these genes (Higgins et al., 2022). An approach based on artificial intelligence have been used to analyse genomic data to detect and prioritize putative disease-causing variants (De La Vega et al., 2021). It is also important to consider that epigenetic factors are associated with multiple rare diseases and biological cell processes, including ciliogenesis (Li and Li, 2021). There are multiple projects that integrate epigenetic and genetic data to further understand the disease mechanisms responsible for cancer and a diverse range of rare diseases, including primary ciliopathies and autoimmune diseases (Asada et al., 2021, Teruel et al., 2021, Cordell et al., 2021, Drivas et al., 2021). These and other studies and projects have the potential to improve the discovery of disease-causing variants and candidate genes.

Focussing on the Genomics England 100,000 Genomes Project, future work would include access to more detailed phenotyping and the design of more automated pipelines, pipelines that could detect multiple types of variants, including structural variants and splicing variants (Best et al., 2022b, Blakes et al., 2022, Orr et al., 2023). Future work may aim to facilitate the communication between the Genomics England Clinical Interpretation Partnership (GeCIP) members (which are researchers

authorised to access and analyse the data available in the Genomics England 100,000 Genomes Project to standardise these pipelines. These pipelines include variant prioritisation, variant annotation (according to the American College of Medical Genetics and Genomics (ACMG) guidelines) and self-curated gene panels of known disease-causing genes and candidate genes. Additionally, updated lists of known pathogenic variants (that may not be classified as pathogenic by ClinVar database) that are intronic, structural or synonymous can be created. These pipelines should be easily available and easy to update and review by authorised members of the Genomics England 100,000 Genomes Project as well as contact with the recruiting clinicians should be enhanced (to confirm and update clinical phenotypes, for instance).

There are multiple examples in which *in silico* studies can be supported and complemented by *in vitro* studies and vice versa (Higgins et al., 2022, Molinari et al., 2019, Devlin et al., 2022, Latour et al., 2020). There are several factors that are assisting and will further assist the development of *in silico*, *in vitro* and *in vivo* studies and their combination. Firstly, the access to large databases with phenotypic, genetic, and/or epigenetic data available (Karczewski et al., 2020, The GTEx Consortium, 2020, Thul and Lindskog, 2018, Higgins et al., 2022, Wheway et al., 2019). Also, the development of organoids such as brain (Zhang et al., 2019), retinal (Georgiou et al., 2022) and kidney (Zeng et al., 2021) organoids as *ex vivo* models that recapitulate some molecular and cellular aspects of key organs affected in ciliopathies. Furthermore, the development of therapies targeting retinal (Cideciyan et al., 2019, Barny et al., 2019, Ramsbottom et al., 2018, Jacobson et al., 2022) and renal phenotypes (Garcia et al., 2022, Ramsbottom et al., 2018).

This together with other increasing developments other areas such as three-dimensional (3D) imaging (Kiewisz et al., 2022), single-cell tracking (Huelsz-Prince et al., 2022) and the development of a wide range of automated pipelines and algorithms: from the prediction of the structure of human proteins (~350,000 proteins, 44% of all known human proteins) and protein interactions (Humphreys et al., 2021, Jumper et al., 2021), to the detection of behavioural changes in animal models (Jia et al., 2022) and robotics (Kanda et al., 2022), can support and increase the efficiency of *in vitro* and *in vivo* experiments. The use and combination of these tools have limitless potential that can answer to multiple biomedical and genomic questions covering multiple fields, from basic science to personalised medicine.

More precisely, the genetic findings related with renal ciliopathies from the Genomics England 100,000 Genomes Project could be complemented using hURECs. The use of hURECs have been proven to be valuable non-invasive source of primary cells and an *ex vivo* model to complement *in vivo* mouse models and *in silico* genetic studies (Ramsbottom et al., 2018, Devane et al., 2022). The use of hURECs can be used to characterise disease-causing mutations (as it has been done for the nonsense *ALG8* mutation accounting for ADPKD). It can be suggested that use of hURECs can be used for *in vitro* experiments to model disease-causing mutations and investigate potential treatments for renal ciliopathies, including, but not exclusively, potential exon skipping therapies (Molinari et al., 2019, Garcia et al., 2022). In recent years, there have been numerous advances in exon skipping therapies (Dulla et al., 2021) and *in vitro* studies using hURECs and other primary cells to study new therapies to treat renal ciliopathies and other primary ciliopathies (Ramsbottom et al., 2018, Devane et al., 2022, Oud et al., 2018).

6.4. Final conclusions

In conclusion, gene expression patterns of two genes associated with rare primary ciliopathies: *CEP120* and *ARL3*, have been characterised in human embryonic and fetal tissue. *CEP120* and *ARL3* have similar expression patterns and are expressed in multiple tissues, including developing cerebellum, kidney and retina (Powell et al., 2020). This is consistent with the clinical phenotypes caused by disease-causing mutations in *CEP120* and *ARL3* reported in the literature (Alkanderi et al., 2018, Roosing et al., 2016).

Genotype-phenotype correlations were explored in two primary ciliopathy genes: *CEP120* and *CC2D2A*, using a self-curated and annotated database. It was confirmed that biallelic truncating *CC2D2A* variants are associated with more severe phenotypes than biallelic *CC2D2A* non truncating variants. An *in silico* approach was used to find the skippable exons in *CEP120* and *CC2D2A* suitable for a potential exon skipping therapy. *CC2D2A* was suggested as a good candidate gene for a potential exon skipping therapy, with 7 exons as potential candidates for targeted exon skipping therapies (Barroso-Gil et al., 2021a).

The Genomics England 100,000 Genomes Project was used to genetically solve patients and find disease-causing mutations accounting for primary ciliopathies. Two *TOGARAM1* pathogenic variants were found in a JBTS patient, supporting the

studies of *TOGARAM1* as a novel JBTS gene (Latour et al., 2020). Moreover, 17 patients diagnosed with cystic kidney disease and 1 patient with diagnosed with renal tract calcification were found to have heterozygous disease-causing variants in genes involved in the ER glycosylation pathway: *DNAJB11* (8 patients) (Huynh et al., 2020), *ALG5* (2 patients) (Lemoine et al., 2022), *ALG8* (5 patients) and *ALG9* (3 patients), supporting these genes as genes associated with clinical phenotypes within the ADPKD phenotypic spectrum.

The WGS and clinical data available in the Genomics England 100,000 Genomes Project has allowed to investigate if there is an enrichment of heterozygous LoF alleles in any of the *ALG* genes, in a population of patients with cysts in kidney or liver, compared to a non-cystic control population. It was found that LoF alleles in *ALG1*, *ALG8*, *ALG9* and *ALG12* are enriched in patients with cystic kidney/liver disease. This validates *ALG8* and *ALG9* as genes associated with the ADPKD phenotypic spectrum and suggests that heterozygous LoF alleles in *ALG1* or *ALG12* genes may contribute to genetically solve patients presenting disease phenotypes within the ADPKD phenotypic spectrum.

The hURECs of a patient with a heterozygous nonsense variant in *ALG8* displayed elongated primary cilia compared with the hURECs of two healthy sex matched controls. It has been shown how hURECs can be used to complement and support findings from the Genomics England 100,000 Genomes Project.

The combination of *in silico* and *in vitro* approaches that assist and improve the understanding of the disease mechanisms caused by disease-causing mutations, can lead to understand the molecular genetics of ciliopathy genes and the complexity of the molecular genetics of primary ciliopathies and design potential targeting gene therapies to treat patients with renal ciliopathies.

References

- AARTSMA-RUS, A. & VAN OMMEN, G. J. 2007. Antisense-mediated exon skipping: a versatile tool with therapeutic and research applications. *Rna*, 13, 1609-24.
- ABDELHAMED, Z. A., ABDELMOTTALEB, D. I., EL-ASRAG, M. E., NATARAJAN, S., WHEWAY, G., INGLEHEARN, C. F., TOOMES, C. & JOHNSON, C. A. 2019. The ciliary Frizzled-like receptor Tmem67 regulates canonical Wnt/β-catenin signalling in the developing cerebellum via Hoxb5. *Sci Rep*, 9, 5446.
- ADAMIOK-OSTROWSKA, A. & PIEKIEŁKO-WITKOWSKA, A. 2020. Ciliary Genes in Renal Cystic Diseases. *Cells*, 9.
- ADAMS, M., SIMMS, R. J., ABDELHAMED, Z., DAWE, H. R., SZYMANSKA, K., LOGAN, C. V., WHEWAY, G., PITT, E., GULL, K., KNOWLES, M. A., BLAIR, E., CROSS, S. H., SAYER, J. A. & JOHNSON, C. A. 2012. A meckelin-filamin A interaction mediates ciliogenesis. *Hum Mol Genet*, 21, 1272-86.
- AEBI, M. 2013. N-linked protein glycosylation in the ER. *Biochim Biophys Acta*, 1833, 2430-7.
- AFZELIUS, B. A. 1976. A human syndrome caused by immotile cilia. *Science*, 193, 317-9.
- AIRIK, R., SCHUELER, M., AIRIK, M., CHO, J., ULANOWICZ, K. A., PORATH, J. D., HURD, T. W., BEKKER-JENSEN, S., SCHRØDER, J. M., ANDERSEN, J. S. & HILDEBRANDT, F. 2016. SDCCAG8 Interacts with RAB Effector Proteins RABEP2 and ERC1 and Is Required for Hedgehog Signaling. *PLoS One*, 11, e0156081.
- AIRIK, R., SLAATS, G. G., GUO, Z., WEISS, A. C., KHAN, N., GHOSH, A., HURD, T. W., BEKKER-JENSEN, S., SCHRØDER, J. M., ELLEDGE, S. J., ANDERSEN, J. S., KISPERT, A., CASTELLI, M., BOLETTA, A., GILES, R. H. & HILDEBRANDT, F. 2014. Renal-retinal ciliopathy gene Sdccag8 regulates DNA damage response signaling. *J Am Soc Nephrol*, 25, 2573-83.
- AJZENBERG, H., SLAATS, G. G., STOKMAN, M. F., ARTS, H. H., LOGISTER, I., KROES, H. Y., RENKEMA, K. Y., VAN Haelst, M. M., TERHAL, P. A., VAN ROOIJ, I. A., KEIJZER-VEEN, M. G., KNOERS, N. V., LILIEN, M. R., JEWETT, M. A. & GILES, R. H. 2015. Non-invasive sources of cells with primary cilia from pediatric and adult patients. *Cilia*, 4, 8.
- AL-HAMED, M. H., KURDI, W., ALSAHAN, N., ALABDULLAH, Z., ABUDRAZ, R., TULBAH, M., ALNEMER, M., KHAN, R., AL-JURAYB, H., ALAHMED, A., TAHIR, A. I., KHALIL, D., EDWARDS, N., AL ABDULAZIZ, B., BINHUMAID, F. S., MAJID, S., FAQUIH, T., EL-KALIOBY, M., ABOUELHODA, M., ALTASSAN, N., MONIES, D., MEYER, B., SAYER, J. A. & ALBAQUMI, M. 2016. Genetic spectrum of Saudi Arabian patients with antenatal cystic kidney disease and ciliopathy phenotypes using a targeted renal gene panel. *J Med Genet*, 53, 338-47.
- AL ALAWI, I., AL RIYAMI, M., BARROSO-GIL, M., POWELL, L., OLINGER, E., AL SALMI, I. & SAYER, J. A. 2021. The diagnostic yield of whole exome sequencing as a first approach in consanguineous Omani renal ciliopathy syndrome patients. *F1000Res*, 10, 207.
- AL ALAWI, I., MOLINARI, E., AL SALMI, I., AL RAHBI, F., AL MAWALI, A. & SAYER, J. A. 2020. Clinical and genetic characteristics of autosomal recessive polycystic kidney disease in Oman. *BMC Nephrol*, 21, 347.
- ALDINGER, K. A., THOMSON, Z., PHELPS, I. G., HALDIPUR, P., DENG, M., TIMMS, A. E., HIRANO, M., SANTPERE, G., ROCO, C., ROSENBERG, A. B., LORENTE-GALDOS, B., GULDEN, F. O., O'DAY, D., OVERMAN, L. M., LISGO, S. N., ALEXANDRE, P., SESTAN, N., DOHERTY, D., DOBYNS, W.

- B., SEELIG, G., GLASS, I. A. & MILLEN, K. J. 2021. Spatial and cell type transcriptional landscape of human cerebellar development. *Nat Neurosci*, 24, 1163-1175.
- ALIZADEH, R., JAMSHIDI, S., KERAMATIPOUR, M., MOEINIAN, P., HOSSEINI, R., OTUKESH, H. & TALEBI, S. 2020. Whole Exome Sequencing Reveals a XPNPEP3 Novel Mutation Causing Nephronophthisis in a Pediatric Patient. *Iran Biomed J*, 24, 405-8.
- ALKANDERI, S., MOLINARI, E., SHAHEEN, R., ELMAGHLOOB, Y., STEPHEN, L. A., SAMMUT, V., RAMSBOTTOM, S. A., SRIVASTAVA, S., CAIRNS, G., EDWARDS, N., RICE, S. J., EWIDA, N., ALHASHEM, A., WHITE, K., MILES, C. G., STEEL, D. H., ALKURAYA, F. S., ISMAIL, S. & SAYER, J. A. 2018. ARL3 Mutations Cause Joubert Syndrome by Disrupting Ciliary Protein Composition. *Am J Hum Genet*, 103, 612-620.
- ALSUBHI, S., ALHASHEM, A., ALAZAMI, A., TLILI, K., ALSHAHWAN, S., LEFEBER, D., ALKURAYA, F. S. & TABARKI, B. 2016. Further Delineation of the ALG9-CDG Phenotype. *JIMD Rep*, 27, 107-12.
- ALZU'BI, A. & CLOWRY, G. J. 2019. Expression of ventral telencephalon transcription factors ASCL1 and DLX2 in the early fetal human cerebral cortex. *J Anat*, 235, 555-568.
- ALZU'BI, A. & CLOWRY, G. J. 2020. Multiple Origins of Secretogogin Expressing Cortical GABAergic Neuron Precursors in the Early Human Fetal Telencephalon. *Front Neuroanat*, 14, 61.
- ALZU'BI, A., LINDSAY, S. J., HARKIN, L. F., MCINTYRE, J., LISGO, S. N. & CLOWRY, G. J. 2017. The Transcription Factors COUP-TFI and COUP-TFII have Distinct Roles in Arealisation and GABAergic Interneuron Specification in the Early Human Fetal Telencephalon. *Cereb Cortex*, 27, 4971-4987.
- ALZU'BI, A., SANKAR, N., CROSIER, M., KERWIN, J. & CLOWRY, G. J. 2022. Tyramide signal amplification coupled with multiple immunolabeling and RNAScope *in situ* hybridization in formaldehyde-fixed paraffin-embedded human fetal brain. *J Anat*, 241, 33-41.
- ANAND, M. & KHANNA, H. 2012. Ciliary transition zone (TZ) proteins RPGR and CEP290: role in photoreceptor cilia and degenerative diseases. *Expert Opinion on Therapeutic Targets*, 16, 541-551.
- ANSLEY, S. J., BADANO, J. L., BLACQUE, O. E., HILL, J., HOSKINS, B. E., LEITCH, C. C., KIM, J. C., ROSS, A. J., EICHERS, E. R., TESLOVICH, T. M., MAH, A. K., JOHNSEN, R. C., CAVENDER, J. C., LEWIS, R. A., LEROUX, M. R., BEALES, P. L. & KATSANIS, N. 2003. Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. *Nature*, 425, 628-33.
- APPLE, B., SARTORI, G., CHINTAM, K., MOORE, B., TRIFFO, W., STRANDE, N., SINGH, G., MIRSHAHI, T. & CHANG, A. 2022. eP077: Atypical polycystic kidney disease in individuals heterozygous for rare ALG8 protein-truncating variants. *Genetics in Medicine*, 24, S50-S51.
- APPLE, B., SARTORI, G., MOORE, B., CHINTAM, K., SINGH, G., ANAND, P. M., STRANDE, N. T., MIRSHAHI, T., TRIFFO, W. & CHANG, A. R. 2023. Individuals heterozygous for ALG8 protein-truncating variants are at increased risk of a mild cystic kidney disease. *Kidney Int*, 103, 607-615.
- ARJONA, F. J., LATTA, F., MOHAMMED, S. G., THOMASSEN, M., VAN WIJK, E., BINDELS, R. J. M., HOENDEROP, J. G. J. & DE BAAIJ, J. H. F. 2019. SLC41A1 is essential for magnesium homeostasis *in vivo*. *Pflugers Arch*, 471, 845-860.
- ARTS, H. H., DOHERTY, D., VAN BEERSUM, S. E., PARISI, M. A., LETTEBOER, S. J., GORDEN, N. T., PETERS, T. A., MÄRKER, T., VOESENEK, K.,

- KARTONO, A., OZYUREK, H., FARIN, F. M., KROES, H. Y., WOLFRUM, U., BRUNNER, H. G., CREMERS, F. P., GLASS, I. A., KNOERS, N. V. & ROEPMAN, R. 2007. Mutations in the gene encoding the basal body protein RPGRIP1L, a nephrocystin-4 interactor, cause Joubert syndrome. *Nat Genet*, 39, 882-8.
- ARTS, H. H. & KNOERS, N. V. 2013. Current insights into renal ciliopathies: what can genetics teach us? *Pediatr Nephrol*, 28, 863-74.
- ASADA, K., KANEKO, S., TAKASAWA, K., MACHINO, H., TAKAHASHI, S., SHINKAI, N., SHIMOYAMA, R., KOMATSU, M. & HAMAMOTO, R. 2021. Integrated Analysis of Whole Genome and Epigenome Data Using Machine Learning Technology: Toward the Establishment of Precision Oncology. *Front Oncol*, 11, 666937.
- ASTUTI, D., SABIR, A., FULTON, P., ZATYKA, M., WILLIAMS, D., HARDY, C., MILAN, G., FAVARETTO, F., YU-WAI-MAN, P., ROHAYEM, J., LÓPEZ DE HEREDIA, M., HERSHY, T., TRANEBJAERG, L., CHEN, J. H., CHAUSSENOT, A., NUNES, V., MARSHALL, B., MCAFFERTY, S., TILLMANN, V., MAFFEI, P., PAQUIS-FLUCKLINGER, V., GEBERHIWOT, T., MLYNARSKI, W., PARKINSON, K., PICARD, V., BUENO, G. E., DIAS, R., ARNOLD, A., RICHENS, C., PAISEY, R., URANO, F., SEMPLE, R., SINNOTT, R. & BARRETT, T. G. 2017. Monogenic diabetes syndromes: Locus-specific databases for Alström, Wolfram, and Thiamine-responsive megaloblastic anemia. *Hum Mutat*, 38, 764-777.
- ATTANASIO, M., UHLENHAUT, N. H., SOUSA, V. H., O'TOOLE, J. F., OTTO, E., ANLAG, K., KLUGMANN, C., TREIER, A. C., HELOU, J., SAYER, J. A., SEELOW, D., NÜRNBERG, G., BECKER, C., CHUDLEY, A. E., NÜRNBERG, P., HILDEBRANDT, F. & TREIER, M. 2007. Loss of GLIS2 causes nephronophthisis in humans and mice by increased apoptosis and fibrosis. *Nat Genet*, 39, 1018-24.
- AUTON, A., BROOKS, L. D., DURBIN, R. M., GARRISON, E. P., KANG, H. M., KORBEL, J. O., MARCHINI, J. L., MCCARTHY, S., MCVEAN, G. A. & ABECASIS, G. R. 2015. A global reference for human genetic variation. *Nature*, 526, 68-74.
- BABU, D. & ROY, S. 2013. Left-right asymmetry: cilia stir up new surprises in the node. *Open Biology*, 3, 130052-130052.
- BACHMANN-GAGESCU, R., DEMPSEY, J. C., BULGHERONI, S., CHEN, M. L., D'ARRIGO, S., GLASS, I. A., HELLER, T., HÉON, E., HILDEBRANDT, F., JOSHI, N., KNUTZEN, D., KROES, H. Y., MACK, S. H., NUOVO, S., PARISI, M. A., SNOW, J., SUMMERS, A. C., SYMONS, J. M., ZEIN, W. M., BOLTSCHAUSER, E., SAYER, J. A., GUNAY-AYGUN, M., VALENTE, E. M. & DOHERTY, D. 2020. Healthcare recommendations for Joubert syndrome. *Am J Med Genet A*, 182, 229-249.
- BACHMANN-GAGESCU, R., DEMPSEY, J. C., PHELPS, I. G., O'ROAK, B. J., KNUTZEN, D. M., RUE, T. C., ISHAK, G. E., ISABELLA, C. R., GORDEN, N., ADKINS, J., BOYLE, E. A., DE LACY, N., O'DAY, D., ALSWAID, A., RAMADEVI, A. R., LINGAPPA, L., LOURENCO, C., MARTORELL, L., GARCIA-CAZORLA, A., OZYUREK, H., HALILOGLU, G., TUYSUZ, B., TOPCU, M., CHANCE, P., PARISI, M. A., GLASS, I. A., SHENDURE, J. & DOHERTY, D. 2015a. Joubert syndrome: a model for untangling recessive disorders with extreme genetic heterogeneity. *J Med Genet*, 52, 514-22.
- BACHMANN-GAGESCU, R., DONA, M., HETTERSCHIJT, L., TONNAER, E., PETERS, T., DE VRIEZE, E., MANS, D. A., VAN BEERSUM, S. E., PHELPS, I. G., ARTS, H. H., KEUNEN, J. E., UEFFING, M., ROEPMAN, R., BOLDT, K.,

- DOHERTY, D., MOENS, C. B., NEUHAUSS, S. C., KREMER, H. & VAN WIJK, E. 2015b. The Ciliopathy Protein CC2D2A Associates with NINL and Functions in RAB8-MICAL3-Regulated Vesicle Trafficking. *PLoS Genet*, 11, e1005575.
- BACHMANN-GAGESCU, R., ISHAK, G. E., DEMPSEY, J. C., ADKINS, J., O'DAY, D., PHELPS, I. G., GUNAY-AYGUN, M., KLINE, A. D., SZCZALUBA, K., MARTORELL, L., ALSWAID, A., ALRASHEED, S., PAI, S., IZATT, L., RONAN, A., PARISI, M. A., MEFFORD, H., GLASS, I. & DOHERTY, D. 2012. Genotype-phenotype correlation in CC2D2A-related Joubert syndrome reveals an association with ventriculomegaly and seizures. *J Med Genet*, 49, 126-37.
- BADANO, J. L., LEITCH, C. C., ANSLEY, S. J., MAY-SIMERA, H., LAWSON, S., LEWIS, R. A., BEALES, P. L., DIETZ, H. C., FISHER, S. & KATSANIS, N. 2006a. Dissection of epistasis in oligogenic Bardet-Biedl syndrome. *Nature*, 439, 326-30.
- BADANO, J. L., MITSUMA, N., BEALES, P. L. & KATSANIS, N. 2006b. The ciliopathies: an emerging class of human genetic disorders. *Annu Rev Genomics Hum Genet*, 7, 125-48.
- BARNY, I., PERRAULT, I., MICHEL, C., GOUDIN, N., DEFOORT-DHELLEMES, S., GHAZI, I., KAPLAN, J., ROZET, J. M. & GERARD, X. 2019. AON-Mediated Exon Skipping to Bypass Protein Truncation in Retinal Dystrophies Due to the Recurrent CEP290 c.4723A > T Mutation. Fact or Fiction? *Genes (Basel)*, 10.
- BARNY, I., PERRAULT, I., MICHEL, C., SOUSSAN, M., GOUDIN, N., RIO, M., THOMAS, S., ATTIE-BITACH, T., HAMEL, C., DOLLFUS, H., KAPLAN, J., ROZET, J.-M. & GERARD, X. 2018. Basal exon skipping and nonsense-associated altered splicing allows bypassing complete CEP290 loss-of-function in individuals with unusually mild retinal disease. *Human Molecular Genetics*, 27, 2689-2702.
- BARROSO-GIL, M., OLINGER, E., RAMSBOTTOM, S. A., MOLINARI, E., MILES, C. G. & SAYER, J. A. 2021a. Update of genetic variants in CEP120 and CC2D2A-With an emphasis on genotype-phenotype correlations, tissue specific transcripts and exploring mutation specific exon skipping therapies. *Mol Genet Genomic Med*, 9, e1603.
- BARROSO-GIL, M., OLINGER, E. & SAYER, J. A. 2021b. Molecular genetics of renal ciliopathies. *Biochem Soc Trans*, 49, 1205-1220.
- BARROSO-GIL, M., POWELL, L. & SAYER, J. A. 2020. RE: Clinical and Molecular Diagnosis of Joubert Syndrome and Related Disorders. *Pediatr Neurol*, 112, 10.
- BARWELL, J., SNAPE, K. & WEDDERBURN, S. 2019. The new genomic medicine service and implications for patients. *Clin Med (Lond)*, 19, 273-277.
- BASHFORD, A. L. & SUBRAMANIAN, V. 2019. Mice with a conditional deletion of Talpid3 (KIAA0586) - a model for Joubert syndrome. *J Pathol*, 248, 396-408.
- BAYAT, A., KERR, B. & DOUZGOU, S. 2017. The evolving craniofacial phenotype of a patient with Sensenbrenner syndrome caused by IFT140 compound heterozygous mutations. *Clin Dysmorphol*, 26, 247-251.
- BEALES, P. L., BLAND, E., TOBIN, J. L., BACCHELLI, C., TUYSUZ, B., HILL, J., RIX, S., PEARSON, C. G., KAI, M., HARTLEY, J., JOHNSON, C., IRVING, M., ELCIOGLU, N., WINEY, M., TADA, M. & SCAMBLER, P. J. 2007. IFT80, which encodes a conserved intraflagellar transport protein, is mutated in Jeune asphyxiating thoracic dystrophy. *Nature Genetics*, 39, 727-729.
- BECK, B. B., PHILLIPS, J. B., BARTRAM, M. P., WEGNER, J., THOENES, M., PANNES, A., SAMPSON, J., HELLER, R., GÖBEL, H., KOERBER, F.,

- NEUGEBAUER, A., HEDERGOTT, A., NÜRNBERG, G., NÜRNBERG, P., THIELE, H., ALTMÜLLER, J., TOLIAT, M. R., STAUBACH, S., BOYCOTT, K. M., VALENTE, E. M., JANECKE, A. R., EISENBERGER, T., BERGMANN, C., TEBBE, L., WANG, Y., WU, Y., FRY, A. M., WESTERFIELD, M., WOLFRUM, U. & BOLZ, H. J. 2014. Mutation of POC1B in a severe syndromic retinal ciliopathy. *Hum Mutat*, 35, 1153-62.
- BEN-SALEM, S., AL-SHAMSI, A. M., GLEESON, J. G., ALI, B. R. & AL-GAZALI, L. 2014. Mutation spectrum of Joubert syndrome and related disorders among Arabs. *Hum Genome Var*, 1, 14020.
- BEN-SALEM, S., AL-SHAMSI, A. M., GLEESON, J. G., ALI, B. R. & AL-GAZALI, L. 2015. Erratum: Mutation spectrum of Joubert syndrome and related disorders among Arabs. *Hum Genome Var*, 2, 15001.
- BENNETT, C. F. & SWAYZE, E. E. 2010. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu Rev Pharmacol Toxicol*, 50, 259-93.
- BENTLEY, D. R., BALASUBRAMANIAN, S., SWERDLOW, H. P., SMITH, G. P., MILTON, J., BROWN, C. G., HALL, K. P., EVERE, D. J., BARNES, C. L., BIGNELL, H. R., BOUTELL, J. M., BRYANT, J., CARTER, R. J., KEIRA CHEETHAM, R., COX, A. J., ELLIS, D. J., FLATBUSH, M. R., GORMLEY, N. A., HUMPHRAY, S. J., IRVING, L. J., KARBELASHVILI, M. S., KIRK, S. M., LI, H., LIU, X., MAISINGER, K. S., MURRAY, L. J., OBRADOVIC, B., OST, T., PARKINSON, M. L., PRATT, M. R., RASOLONJATOVO, I. M., REED, M. T., RIGATTI, R., RODIGHIERO, C., ROSS, M. T., SABOT, A., SANKAR, S. V., SCALLY, A., SCHROTH, G. P., SMITH, M. E., SMITH, V. P., SPIRIDOU, A., TORRANCE, P. E., TZONEV, S. S., VERMAAS, E. H., WALTER, K., WU, X., ZHANG, L., ALAM, M. D., ANASTASI, C., ANIEBO, I. C., BAILEY, D. M., BANCARZ, I. R., BANERJEE, S., BARBOUR, S. G., BAYBAYAN, P. A., BENOIT, V. A., BENSON, K. F., BEVIS, C., BLACK, P. J., BOODHUN, A., BRENNAN, J. S., BRIDGHAM, J. A., BROWN, R. C., BROWN, A. A., BUERMANN, D. H., BUNDU, A. A., BURROWS, J. C., CARTER, N. P., CASTILLO, N., CHIARA, E. C. M., CHANG, S., NEIL COOLEY, R., CRAKE, N. R., DADA, O. O., DIAKOUMAKOS, K. D., DOMINGUEZ-FERNANDEZ, B., EARNSHAW, D. J., EGBUJOR, U. C., ELMORE, D. W., ETCHEIN, S. S., EWAN, M. R., FEDURCO, M., FRASER, L. J., FUENTES FAJARDO, K. V., SCOTT FUREY, W., GEORGE, D., GIETZEN, K. J., GODDARD, C. P., GOLDA, G. S., GRANIERI, P. A., GREEN, D. E., GUSTAFSON, D. L., HANSEN, N. F., HARNISH, K., HAUDENSCHILD, C. D., HEYER, N. I., HIMS, M. M., HO, J. T., HORGAN, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53-9.
- BERBARI, N. F., O'CONNOR, A. K., HAYCRAFT, C. J. & YODER, B. K. 2009. The primary cilium as a complex signaling center. *Curr Biol*, 19, R526-35.
- BERGBOER, J. G. M., WYATT, C., AUSTIN-TSE, C., YAKSI, E. & DRUMMOND, I. A. 2018. Assaying sensory ciliopathies using calcium biosensor expression in zebrafish ciliated olfactory neurons. *Cilia*, 7, 2-2.
- BERGMANN, C. 2017. Genetics of Autosomal Recessive Polycystic Kidney Disease and Its Differential Diagnoses. *Front Pediatr*, 5, 221.
- BERGMANN, C., FLIEGAUF, M., BRÜCHLE, N. O., FRANK, V., OLBRICH, H., KIRSCHNER, J., SCHERMER, B., SCHMEDDING, I., KISPERT, A., KRÄNZLIN, B., NÜRNBERG, G., BECKER, C., GRIMM, T., GIRSCHEK, G., LYNCH, S. A., KELEHAN, P., SENDEREK, J., NEUHAUS, T. J., STALLMACH, T., ZENTGRAF, H., NÜRNBERG, P., GRETZ, N., LO, C., LIENKAMP, S., SCHÄFER, T., WALZ, G., BENZING, T., ZERRES, K. &

- OMRAN, H. 2008. Loss of nephrocystin-3 function can cause embryonic lethality, Meckel-Gruber-like syndrome, situs inversus, and renal-hepatic-pancreatic dysplasia. *Am J Hum Genet*, 82, 959-70.
- BERGMANN, C., GUAY-WOODFORD, L. M., HARRIS, P. C., HORIE, S., PETERS, D. J. M. & TORRES, V. E. 2018. Polycystic kidney disease. *Nat Rev Dis Primers*, 4, 50.
- BESSE, W., CHANG, A. R., LUO, J. Z., TRIFFO, W. J., MOORE, B. S., GULATI, A., HARTZEL, D. N., MANE, S., TORRES, V. E., SOMLO, S. & MIRSHAH, T. 2019. ALG9 Mutation Carriers Develop Kidney and Liver Cysts. *J Am Soc Nephrol*, 30, 2091-2102.
- BESSE, W., DONG, K., CHOI, J., PUNIA, S., FEDELES, S. V., CHOI, M., GALLAGHER, A. R., HUANG, E. B., GULATI, A., KNIGHT, J., MANE, S., TAHVANAINEN, E., TAHVANAINEN, P., SANNA-CHERCHI, S., LIFTON, R. P., WATNICK, T., PEI, Y. P., TORRES, V. E. & SOMLO, S. 2017. Isolated polycystic liver disease genes define effectors of polycystin-1 function. *J Clin Invest*, 127, 1772-1785.
- BEST, S., INGLEHEARN, C. F., WATSON, C. M., TOOMES, C., WHEWAY, G. & JOHNSON, C. A. 2022a. Unlocking the potential of the UK 100,000 Genomes Project-lessons learned from analysis of the "Congenital Malformations caused by Ciliopathies" cohort. *Am J Med Genet C Semin Med Genet*, 190, 5-8.
- BEST, S., LORD, J., ROCHE, M., WATSON, C. M., POULTER, J. A., BEVERS, R. P. J., STUCKEY, A., SZYMANSKA, K., ELLINGFORD, J. M., CARMICHAEL, J., BRITTAIN, H., TOOMES, C., INGLEHEARN, C., JOHNSON, C. A. & WHEWAY, G. 2021. Molecular diagnoses in the congenital malformations caused by ciliopathies cohort of the 100,000 Genomes Project. *J Med Genet*.
- BEST, S., YU, J., LORD, J., ROCHE, M., WATSON, C. M., BEVERS, R. P. J., STUCKEY, A., MADHUSUDHAN, S., JEWELL, R., SISODIYA, S. M., LIN, S., TURNER, S., ROBINSON, H., LESLIE, J. S., BAPLE, E., TOOMES, C., INGLEHEARN, C., WHEWAY, G. & JOHNSON, C. A. 2022b. Uncovering the burden of hidden ciliopathies in the 100 000 Genomes Project: a reverse phenotyping approach. *J Med Genet*.
- BETLEJA, E. & COLE, D. G. 2010. Ciliary trafficking: CEP290 guards a gated community. *Curr Biol*, 20, R928-31.
- BETLEJA, E., NANJUNDAPPA, R., CHENG, T. & MAHJOUB, M. R. 2018. A novel Cep120-dependent mechanism inhibits centriole maturation in quiescent cells. *Elife*, 7.
- BLACQUE, O. E., PERENS, E. A., BOROEVICH, K. A., INGLIS, P. N., LI, C., WARNER, A., KHATTRI, J., HOLT, R. A., OU, G., MAH, A. K., MCKAY, S. J., HUANG, P., SWOBODA, P., JONES, S. J., MARRA, M. A., BAILLIE, D. L., MOERMAN, D. G., SHAHAM, S. & LEROUX, M. R. 2005. Functional genomics of the cilium, a sensory organelle. *Curr Biol*, 15, 935-41.
- BLAKES, A. J. M., WAI, H. A., DAVIES, I., MOLEDINA, H. E., RUIZ, A., THOMAS, T., BUNYAN, D., THOMAS, N. S., BURREN, C. P., GREENHALGH, L., LEES, M., PICHINI, A., SMITHSON, S. F., TAYLOR TAVARES, A. L., O'DONOVAN, P., DOUGLAS, A. G. L., WHIFFIN, N., BARALLE, D. & LORD, J. 2022. A systematic analysis of splicing variants identifies new diagnoses in the 100,000 Genomes Project. *Genome Med*, 14, 79.
- BOLAR, N. A., GOLZIO, C., ŽIVNÁ, M., HAYOT, G., VAN HEMELRIJK, C., SCHEPERS, D., VANDEWEYER, G., HOISCHEN, A., HUYGHE, J. R., RAES, A., MATTHYS, E., SYS, E., AZOU, M., GUBLER, M. C., PRAET, M., VAN CAMP, G., MCFADDEN, K., PEDIADITAKIS, I., PŘISTOUPILOVÁ, A.,

- HODAŇOVÁ, K., VYLEŤAL, P., HARTMANNOVÁ, H., STRÁNECKÝ, V., HŮLKOVÁ, H., BAREŠOVÁ, V., JEDLIČKOVÁ, I., SOVOVÁ, J., HNÍZDA, A., KIDD, K., BLEYER, A. J., SPONG, R. S., VANDE WALLE, J., MORTIER, G., BRUNNER, H., VAN LAER, L., KMOCH, S., KATSANIS, N. & LOEYS, B. L. 2016. Heterozygous Loss-of-Function SEC61A1 Mutations Cause Autosomal-Dominant Tubulo-Interstitial and Glomerulocystic Kidney Disease with Anemia. *Am J Hum Genet*, 99, 174-87.
- BONDUE, T., ARCOLINO, F. O., VEYS, K. R. P., ADEBAYO, O. C., LEVTCHENKO, E., VAN DEN HEUVEL, L. P. & ELMONEM, M. A. 2021. Urine-Derived Epithelial Cells as Models for Genetic Kidney Diseases. *Cells*, 10.
- BONDUE, T., VAN DEN HEUVEL, L., LEVTCHENKO, E. & BROCK, R. 2022. The potential of RNA-based therapy for kidney diseases. *Pediatr Nephrol*, 1-18.
- BOSAKOVA, M., ABRAHAM, S. P., NITA, A., HRUBA, E., BUCHTOVA, M., TAYLOR, S. P., DURAN, I., MARTIN, J., SVOZILOVA, K., BARTA, T., VARECHA, M., BALEK, L., KOHOUTEK, J., RADASZKIEWICZ, T., PUSAPATI, G. V., BRYJA, V., RUSH, E. T., THIFFAULT, I., NICKERSON, D. A., BAMSHAD, M. J., ROHATGI, R., COHN, D. H., KRAKOW, D. & KREJCI, P. 2020. Mutations in GRK2 cause Jeune syndrome by impairing Hedgehog and canonical Wnt signaling. *EMBO Mol Med*, 12, e11739.
- BRANCATI, F., DALLAPICCOLA, B. & VALENTE, E. M. 2010. Joubert Syndrome and related disorders. *Orphanet J Rare Dis*, 5, 20.
- BRAUN, D. A. & HILDEBRANDT, F. 2017. Ciliopathies. *Cold Spring Harb Perspect Biol*, 9.
- BREDRUP, C., SAUNIER, S., OUD, M. M., FISKERSTRAND, T., HOISCHEN, A., BRACKMAN, D., LEH, S. M., MIDTBØ, M., FILHOL, E., BOLE-FEYSOT, C., NITSCHKÉ, P., GILISSEN, C., HAUGEN, O. H., SANDERS, J. S., STOLTE-DIJKSTRA, I., MANS, D. A., STEENBERGEN, E. J., HAMEL, B. C., MATIGNON, M., PFUNDT, R., JEANPIERRE, C., BOMAN, H., RØDAHL, E., VELTMAN, J. A., KNAPPSKOG, P. M., KNOERS, N. V., ROEPMAN, R. & ARTS, H. H. 2011. Ciliopathies with skeletal anomalies and renal insufficiency due to mutations in the IFT-A gene WDR19. *Am J Hum Genet*, 89, 634-43.
- BRESLOW, D. K., HOOGENDOORN, S., KOPP, A. R., MORGENS, D. W., VU, B. K., KENNEDY, M. C., HAN, K., LI, A., HESS, G. T., BASSIK, M. C., CHEN, J. K. & NACHURY, M. V. 2018. A CRISPR-based screen for Hedgehog signaling provides insights into ciliary function and ciliopathies. *Nat Genet*, 50, 460-471.
- BROWN, M. A., WIGLEY, C., WALKER, S., LANCASTER, D., RENDON, A. & SCOTT, R. 2022. Re: Best et al., 'Unlocking the potential of the UK 100,000 Genomes Project - Lessons learned from analysis of the "Congenital malformations caused by ciliopathies" cohort'. *Am J Med Genet A*, 188, 3376-3377.
- BYCROFT, C., FREEMAN, C., PETKOVA, D., BAND, G., ELLIOTT, L. T., SHARP, K., MOTYER, A., VUKCEVIC, D., DELANEAU, O., O'CONNELL, J., CORTES, A., WELSH, S., YOUNG, A., EFFINGHAM, M., MCVEAN, G., LESLIE, S., ALLEN, N., DONNELLY, P. & MARCHINI, J. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562, 203-209.
- CABRITA, I., KRAUS, A., SCHOLZ, J. K., SKOCZYNSKI, K., SCHREIBER, R., KUNZELMANN, K. & BUCHHOLZ, B. 2020. Cyst growth in ADPKD is prevented by pharmacological and genetic inhibition of TMEM16A in vivo. *Nat Commun*, 11, 4320.
- CACHEIRO, P., WESTERBERG, C. H., MAGER, J., DICKINSON, M. E., NUTTER, L. M. J., MUÑOZ-FUENTES, V., HSU, C. W., VAN DEN VEYVER, I. B., FLENNIKEN, A. M., MCKERLIE, C., MURRAY, S. A., TEBOUL, L., HEANEY,

- J. D., LLOYD, K. C. K., LANOUE, L., BRAUN, R. E., WHITE, J. K., CREIGHTON, A. K., LAURIN, V., GUO, R., QU, D., WELLS, S., CLEAK, J., BUNTON-STASYSHYN, R., STEWART, M., HARRISSON, J., MASON, J., HASELI MASHHADI, H., PARKINSON, H., MALLON, A. M. & SMEDLEY, D. 2022. Mendelian gene identification through mouse embryo viability screening. *Genome Med*, 14, 119.
- CAI, Y., FEDELES, S. V., DONG, K., ANYATONWU, G., ONOE, T., MITOBE, M., GAO, J. D., OKUHARA, D., TIAN, X., GALLAGHER, A. R., TANG, Z., XIE, X., LALIOTI, M. D., LEE, A. H., EHRLICH, B. E. & SOMLO, S. 2014. Altered trafficking and stability of polycystins underlie polycystic kidney disease. *J Clin Invest*, 124, 5129-44.
- CAPITANI, N., ONNIS, A., FINETTI, F., CASSIOLI, C., PLEBANI, A., BRUNETTI, J., TROILO, A., D'ELIOS, S., BARONIO, M., GAZZURELLI, L., DELLA BELLA, C., BILLADEAU, D. D., D'ELIOS, M. M., LOUGARIS, V. & BALDARI, C. T. 2022. A CVID-associated variant in the ciliogenesis protein CCDC28B disrupts immune synapse assembly. *Cell Death Differ*, 29, 65-81.
- CARDENAS-RODRIGUEZ, M., AUSTIN-TSE, C., BERGBOER, J. G. M., MOLINARI, E., SUGANO, Y., BACHMANN-GAGESCU, R., SAYER, J. A. & DRUMMOND, I. A. 2021. Genetic compensation for cilia defects in *cep290* mutants by upregulation of cilia-associated small GTPases. *J Cell Sci*, 134.
- CARDENAS-RODRIGUEZ, M., OSBORN, D. P., IRIGOÍN, F., GRAÑA, M., ROMERO, H., BEALES, P. L. & BADANO, J. L. 2013. Characterization of CCDC28B reveals its role in ciliogenesis and provides insight to understand its modifier effect on Bardet-Biedl syndrome. *Hum Genet*, 132, 91-105.
- CAVALCANTI, D. P., HUBER, C., SANG, K. H., BAUJAT, G., COLLINS, F., DELEZOIDE, A. L., DAGONEAU, N., LE MERRER, M., MARTINOVIC, J., MELLO, M. F., VEKEMANS, M., MUNNICH, A. & CORMIER-DAIRE, V. 2011. Mutation in *IFT80* in a fetus with the phenotype of Verma-Naumoff provides molecular evidence for Jeune-Verma-Naumoff dysplasia spectrum. *J Med Genet*, 48, 88-92.
- CHAKI, M., AIRIK, R., GHOSH, A. K., GILES, R. H., CHEN, R., SLAATS, G. G., WANG, H., HURD, T. W., ZHOU, W., CLUCKEY, A., GEE, H. Y., RAMASWAMI, G., HONG, C. J., HAMILTON, B. A., CERVENKA, I., GANJI, R. S., BRYJA, V., ARTS, H. H., VAN REEUWIJK, J., OUD, M. M., LETTEBOER, S. J., ROEPMAN, R., HUSSON, H., IBRAGHIMOV-BESKROVNAYA, O., YASUNAGA, T., WALZ, G., ELEY, L., SAYER, J. A., SCHERMER, B., LIEBAU, M. C., BENZING, T., LE CORRE, S., DRUMMOND, I., JANSSEN, S., ALLEN, S. J., NATARAJAN, S., O'TOOLE, J. F., ATTANASIO, M., SAUNIER, S., ANTIGNAC, C., KOENEKOOP, R. K., REN, H., LOPEZ, I., NAYIR, A., STOETZEL, C., DOLLFUS, H., MASSOUDI, R., GLEESON, J. G., ANDREOLI, S. P., DOHERTY, D. G., LINDSTRAD, A., GOLZIO, C., KATSANIS, N., PAPE, L., ABBOUD, E. B., AL-RAJHI, A. A., LEWIS, R. A., OMRAN, H., LEE, E. Y., WANG, S., SEKIGUCHI, J. M., SAUNDERS, R., JOHNSON, C. A., GARNER, E., VANSELOW, K., ANDERSEN, J. S., SHLOMAI, J., NURNBERG, G., NURNBERG, P., LEVY, S., SMOGORZEWSKA, A., OTTO, E. A. & HILDEBRANDT, F. 2012. Exome capture reveals *ZNF423* and *CEP164* mutations, linking renal ciliopathies to DNA damage response signaling. *Cell*, 150, 533-48.
- CHALUPA, L. M. & GUNHAN, E. 2004. Development of On and Off retinal pathways and retinogeniculate projections. *Prog Retin Eye Res*, 23, 31-51.
- CHAN, M. M. Y., SADEGHİ-ALAVIJEH, O., LOPES, F. M., HILGER, A. C., STANESCU, H. C., VOINESCU, C. D., BEAMAN, G. M., NEWMAN, W. G.,

- ZANIEW, M., WEBER, S., HO, Y. M., CONNOLLY, J. O., WOOD, D., MAJ, C., STUCKEY, A., KOUSATHANAS, A., KLETA, R., WOOLF, A. S., BOCKENHAUER, D., LEVINE, A. P. & GALE, D. P. 2022. Diverse ancestry whole-genome sequencing association study identifies TBX5 and PTK7 as susceptibility genes for posterior urethral valves. *Elife*, 11.
- CHANG, C. H., CHEN, T. Y., LU, I. L., LI, R. B., TSAI, J. J., LIN, P. Y. & TANG, T. K. 2021. CEP120-mediated KIAA0753 recruitment onto centrioles is required for timely neuronal differentiation and germinal zone exit in the developing cerebellum. *Genes Dev*, 35, 1445-1460.
- CHEN, J. K., TAIPALE, J., YOUNG, K. E., MAITI, T. & BEACHY, P. A. 2002. Small molecule modulation of Smoothened activity. *Proc Natl Acad Sci U S A*, 99, 14071-6.
- CHENG, L. E., ZHANG, J. & REED, R. R. 2007. The transcription factor Zfp423/OAZ is required for cerebellar development and CNS midline patterning. *Dev Biol*, 307, 43-52.
- CHENG, Y. Z., ELEY, L., HYNES, A. M., OVERMAN, L. M., SIMMS, R. J., BARKER, A., DAWE, H. R., LINDSAY, S. & SAYER, J. A. 2012. Investigating embryonic expression patterns and evolution of AHI1 and CEP290 genes, implicated in Joubert syndrome. *PLoS One*, 7, e44975.
- CHOI, Y. J., HALBRITTER, J., BRAUN, D. A., SCHUELER, M., SCHAPIRO, D., RIM, J. H., NANDADASA, S., CHOI, W. I., WIDMEIER, E., SHRIL, S., KÖRBER, F., SETHI, S. K., LIFTON, R. P., BECK, B. B., APTE, S. S., GEE, H. Y. & HILDEBRANDT, F. 2019. Mutations of ADAMTS9 Cause Nephronophthisis-Related Ciliopathy. *Am J Hum Genet*, 104, 45-54.
- CHOKSI, S. P., LAUTER, G., SWOBODA, P. & ROY, S. 2014. Switching on cilia: transcriptional networks regulating ciliogenesis. *Development*, 141, 1427-1441.
- CIDECIYAN, A. V., JACOBSON, S. G., DRACK, A. V., HO, A. C., CHARNG, J., GARAFALO, A. V., ROMAN, A. J., SUMAROKA, A., HAN, I. C., HOCHSTEDLER, M. D., PFEIFER, W. L., SOHN, E. H., TAIEL, M., SCHWARTZ, M. R., BIASUTTO, P., WIT, W., CHEETHAM, M. E., ADAMSON, P., RODMAN, D. M., PLATENBURG, G., TOME, M. D., BALIKOVA, I., NERINCKX, F., ZAEYTIJD, J., VAN CAUWENBERGH, C., LEROY, B. P. & RUSSELL, S. R. 2019. Effect of an intravitreal antisense oligonucleotide on vision in Leber congenital amaurosis due to a photoreceptor cilium defect. *Nat Med*, 25, 225-228.
- CIDECIYAN, A. V., JACOBSON, S. G., HO, A. C., SWIDER, M., SUMAROKA, A., ROMAN, A. J., WU, V., RUSSELL, R. C., VIARBITSKAYA, I., GARAFALO, A. V., SCHWARTZ, M. R. & GIRACH, A. 2023. Durable vision improvement after a single intravitreal treatment with antisense oligonucleotide in CEP290-LCA: Replication in two eyes. *Am J Ophthalmol Case Rep*, 32, 101873.
- CNOSEN, W. R., TE MORSCHE, R. H., HOISCHEN, A., GILISSEN, C., CHRISPIJN, M., VENSELAAR, H., MEHDI, S., BERGMANN, C., VELTMAN, J. A. & DRENTH, J. P. 2014. Whole-exome sequencing reveals LRP5 mutations and canonical Wnt signaling associated with hepatic cystogenesis. *Proc Natl Acad Sci U S A*, 111, 5343-8.
- COMARTIN, D., GUPTA, G. D., FUSSNER, E., COYAUD, E., HASEGAN, M., ARCHINTI, M., CHEUNG, S. W., PINCHEV, D., LAWO, S., RAUGHT, B., BAZETT-JONES, D. P., LUDERS, J. & PELLETIER, L. 2013a. CEP120 and SPICE1 cooperate with CPAP in centriole elongation. *Curr Biol*, 23, 1360-6.
- COMARTIN, D., GUPTA, G. D., FUSSNER, E., COYAUD, É., HASEGAN, M., ARCHINTI, M., CHEUNG, S. W. T., PINCHEV, D., LAWO, S., RAUGHT, B.,

- BAZETT-JONES, D. P., LÜDERS, J. & PELLETIER, L. 2013b. CEP120 and SPICE1 Cooperate with CPAP in Centriole Elongation. *Current Biology*, 23, 1360-1366.
- CONSUGAR, M. B., KUBLY, V. J., LAGER, D. J., HOMMERDING, C. J., WONG, W. C., BAKKER, E., GATTONE, V. H., TORRES, V. E., BREUNING, M. H. & HARRIS, P. C. 2007. Molecular diagnostics of Meckel-Gruber syndrome highlights phenotypic differences between MKS1 and MKS3. *Human Genetics*, 121, 591-599.
- COPPIETERS, F., LEFEVER, S., LEROY, B. P. & DE BAERE, E. 2010. CEP290, a gene with many faces: mutation overview and presentation of CEP290base. *Hum Mutat*, 31, 1097-108.
- CORDELL, H. J., FRYETT, J. J., UENO, K., DARLAY, R., AIBA, Y., HITOMI, Y., KAWASHIMA, M., NISHIDA, N., KHOR, S. S., GERVAIS, O., KAWAI, Y., NAGASAKI, M., TOKUNAGA, K., TANG, R., SHI, Y., LI, Z., JURAN, B. D., ATKINSON, E. J., GERUSSI, A., CARBONE, M., ASSELTA, R., CHEUNG, A., DE ANDRADE, M., BARAS, A., HOROWITZ, J., FERREIRA, M. A. R., SUN, D., JONES, D. E., FLACK, S., SPICER, A., MULCAHY, V. L., BYAN, J., HAN, Y., SANDFORD, R. N., LAZARIDIS, K. N., AMOS, C. I., HIRSCHFIELD, G. M., SELDIN, M. F., INVERNIZZI, P., SIMINOVITCH, K. A., MA, X., NAKAMURA, M. & MELLS, G. F. 2021. An international genome-wide meta-analysis of primary biliary cholangitis: Novel risk loci and candidate drugs. *J Hepatol*, 75, 572-581.
- CORNEC-LE GALL, E., ALAM, A. & PERRONE, R. D. 2019. Autosomal dominant polycystic kidney disease. *Lancet*, 393, 919-935.
- CORNEC-LE GALL, E., OLSON, R. J., BESSE, W., HEYER, C. M., GAINULLIN, V. G., SMITH, J. M., AUDRÉZET, M. P., HOPP, K., PORATH, B., SHI, B., BAHETI, S., SENUM, S. R., ARROYO, J., MADSEN, C. D., FÉREC, C., JOLY, D., JOURET, F., FIKRI-BENBRAHIM, O., CHARASSE, C., COULIBALY, J. M., YU, A. S., KHALILI, K., PEI, Y., SOMLO, S., LE MEUR, Y., TORRES, V. E. & HARRIS, P. C. 2018a. Monoallelic Mutations to DNAJB11 Cause Atypical Autosomal-Dominant Polycystic Kidney Disease. *Am J Hum Genet*, 102, 832-844.
- CORNEC-LE GALL, E., TORRES, V. E. & HARRIS, P. C. 2018b. Genetic Complexity of Autosomal Dominant Polycystic Kidney and Liver Diseases. *J Am Soc Nephrol*, 29, 13-23.
- CORRAL-SERRANO, J. C., SLADEN, P. E., OTTAVIANI, D., REZEK, O. F., ATHANASIOU, D., JOVANOVIC, K., VAN DER SPUY, J., MANSFIELD, B. C. & CHEETHAM, M. E. 2023. Eupatilin Improves Cilia Defects in Human CEP290 Ciliopathy Models. *Cells*, 12.
- COUSSA, R. G., OTTO, E. A., GEE, H. Y., ARTHURS, P., REN, H., LOPEZ, I., KESER, V., FU, Q., FAINGOLD, R., KHAN, A., SCHWARTZENTRUBER, J., MAJEWSKI, J., HILDEBRANDT, F. & KOENEKOOP, R. K. 2013. WDR19: an ancient, retrograde, intraflagellar ciliary protein is mutated in autosomal recessive retinitis pigmentosa and in Senior-Loken syndrome. *Clin Genet*, 84, 150-9.
- DAIGLE, T. L., MADISEN, L., HAGE, T. A., VALLEY, M. T., KNOBLICH, U., LARSEN, R. S., TAKENO, M. M., HUANG, L., GU, H., LARSEN, R., MILLS, M., BOSMA-MOODY, A., SIVERTS, L. A., WALKER, M., GRAYBUCK, L. T., YAO, Z., FONG, O., NGUYEN, T. N., GARREN, E., LENZ, G. H., CHAVARHA, M., PENDERGRAFT, J., HARRINGTON, J., HIROKAWA, K. E., HARRIS, J. A., NICOVICH, P. R., MCGRAW, M. J., OLLERENTHAW, D. R., SMITH, K. A., BAKER, C. A., TING, J. T., SUNKIN, S. M., LECOQ, J., LIN, M.

- Z., BOYDEN, E. S., MURPHY, G. J., DA COSTA, N. M., WATERS, J., LI, L., TASIC, B. & ZENG, H. 2018. A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. *Cell*, 174, 465-480.e22.
- DAS, A., DICKINSON, D. J., WOOD, C. C., GOLDSTEIN, B. & SLEP, K. C. 2015. Crescerin uses a TOG domain array to regulate microtubules in the primary cilium. *Mol Biol Cell*, 26, 4248-64.
- DAVIS, E. E., ZHANG, Q., LIU, Q., DIPLAS, B. H., DAVEY, L. M., HARTLEY, J., STOETZEL, C., SZYMANSKA, K., RAMASWAMI, G., LOGAN, C. V., MUZNY, D. M., YOUNG, A. C., WHEELER, D. A., CRUZ, P., MORGAN, M., LEWIS, L. R., CHERUKURI, P., MASKERI, B., HANSEN, N. F., MULLIKIN, J. C., BLAKESLEY, R. W., BOUFFARD, G. G., GYAPAY, G., RIEGER, S., TÖNSHOFF, B., KERN, I., SOLIMAN, N. A., NEUHAUS, T. J., SWOBODA, K. J., KAYSERILI, H., GALLAGHER, T. E., LEWIS, R. A., BERGMANN, C., OTTO, E. A., SAUNIER, S., SCAMBLER, P. J., BEALES, P. L., GLEESON, J. G., MAHER, E. R., ATTÍÉ-BITACH, T., DOLLFUS, H., JOHNSON, C. A., GREEN, E. D., GIBBS, R. A., HILDEBRANDT, F., PIERCE, E. A. & KATSANIS, N. 2011. TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. *Nature Genetics*, 43, 189-196.
- DE LA FUENTE, L., DEL POZO-VALERO, M., PEREA-ROMERO, I., BLANCO-KELLY, F., FERNÁNDEZ-CABALLERO, L., CORTÓN, M., AYUSO, C. & MÍNGUEZ, P. 2023. Prioritization of New Candidate Genes for Rare Genetic Diseases by a Disease-Aware Evaluation of Heterogeneous Molecular Networks. *Int J Mol Sci*, 24.
- DE LA VEGA, F. M., CHOWDHURY, S., MOORE, B., FRISE, E., MCCARTHY, J., HERNANDEZ, E. J., WONG, T., JAMES, K., GUIDUGLI, L., AGRAWAL, P. B., GENETTI, C. A., BROWNSTEIN, C. A., BEGGS, A. H., LÖSCHER, B. S., FRANKE, A., BOONE, B., LEVY, S. E., ÖUNAP, K., PAJUSALU, S., HUENTELMAN, M., RAMSEY, K., NAYMIK, M., NARAYANAN, V., VEERARAGHAVAN, N., BILLINGS, P., REESE, M. G., YANDELL, M. & KINGSMORE, S. F. 2021. Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. *Genome Med*, 13, 153.
- DE VRIES, J., YNTEMA, J. L., VAN DIE, C. E., CRAMA, N., CORNELISSEN, E. A. M. & HAMEL, B. C. J. 2010. Jeune syndrome: Description of 13 cases and a proposal for follow-up protocol. *European Journal of Pediatrics*, 169, 77-88.
- DECAEN, P. G., DELLING, M., VIEN, T. N. & CLAPHAM, D. E. 2013. Direct recording and molecular identification of the calcium channel of primary cilia. *Nature*, 504, 315-8.
- DEGASPERI, A., ZOU, X., AMARANTE, T. D., MARTINEZ-MARTINEZ, A., KOH, G. C. C., DIAS, J. M. L., HESKIN, L., CHMELOVA, L., RINALDI, G., WANG, V. Y. W., NANDA, A. S., BERNSTEIN, A., MOMEN, S. E., YOUNG, J., PEREZ-GIL, D., MEMARI, Y., BADJA, C., SHOOTER, S., CZARNECKI, J., BROWN, M. A., DAVIES, H. R. & NIK-ZAINAL, S. 2022. Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science*, 376.
- DELOUD, M., BAALA, L., SALOMON, R., LACLEF, C., VIERKOTTEN, J., TORY, K., GOLZIO, C., LACOSTE, T., BESSE, L., OZILOU, C., MOUTKINE, I., HELLMAN, N. E., ANSELME, I., SILBERMANN, F., VESQUE, C., GERHARDT, C., RATTENBERRY, E., WOLF, M. T., GUBLER, M. C., MARTINOVIC, J., ENCHA-RAZAVI, F., BODDAERT, N., GONZALES, M., MACHER, M. A., NIVET, H., CHAMPION, G., BERTHELEME, J. P., NIAUDET, P., MCDONALD, F., HILDEBRANDT, F., JOHNSON, C. A.,

- VEKEMANS, M., ANTIGNAC, C., RUTHER, U., SCHNEIDER-MAUNOURY, S., ATTIE-BITACH, T. & SAUNIER, S. 2007. The ciliary gene RPGRIP1L is mutated in cerebello-oculo-renal syndrome (Joubert syndrome type B) and Meckel syndrome. *Nat Genet*, 39, 875-81.
- DELVALLÉE, C. & DOLLFUS, H. 2023. Retinal Degeneration Animal Models in Bardet-Biedl Syndrome and Related Ciliopathies. *Cold Spring Harb Perspect Med*, 13.
- DEN DUNNEN, J. T., DALGLEISH, R., MAGLOTT, D. R., HART, R. K., GREENBLATT, M. S., MCGOWAN-JORDAN, J., ROUX, A. F., SMITH, T., ANTONARAKIS, S. E. & TASCHNER, P. E. 2016. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat*, 37, 564-9.
- DEVANE, J., OTT, E., OLINGER, E. G., EPTING, D., DECKER, E., FRIEDRICH, A., BACHMANN, N., RENSCHLER, G., EISENBERGER, T., BRIEM-RICHTER, A., GRABHORN, E. F., POWELL, L., WILSON, I. J., RICE, S. J., MILES, C. G., WOOD, K., TRIVEDI, P., HIRSCHFIELD, G., PIETROBATTISTA, A., WOHLER, E., MEZINA, A., SOBREIRA, N., AGOLINI, E., MAGGIORE, G., DAHMER-HEATH, M., YILMAZ, A., BOERRIES, M., METZGER, P., SCHELL, C., GRÜNEWALD, I., KONRAD, M., KÖNIG, J., SCHLEVOGT, B., SAYER, J. A. & BERGMANN, C. 2022. Progressive liver, kidney, and heart degeneration in children and adults affected by TULP3 mutations. *Am J Hum Genet*, 109, 928-943.
- DEVLIN, L., DHONDURAO SUDHINDAR, P. & SAYER, J. A. 2023. Renal ciliopathies: promising drug targets and prospects for clinical trials. *Expert Opin Ther Targets*, 1-22.
- DEVLIN, L. A., COLES, J., JACKSON, C. L., BARROSO-GIL, M., GREEN, B., WALKER, W. T., THOMAS, N. S., THOMPSON, J., ROCK, S. A., NEATU, R., POWELL, L., MOLINARI, E., WILSON, I., CORDELL, H. J., OLINGER, E., MILES, C. G., SAYER, J. A., WHEWAY, G. & LUCAS, J. S. 2022. Biallelic variants in CEP164 cause a motile ciliopathy-like syndrome. *Clin Genet*.
- DEVLIN, L. A., RAMSBOTTOM, S. A., OVERMAN, L. M., LISGO, S. N., CLOWRY, G., MOLINARI, E., POWELL, L., MILES, C. G. & SAYER, J. A. 2020. Embryonic and foetal expression patterns of the ciliopathy gene CEP164. *PLoS One*, 15, e0221914.
- DEVLIN, L. A. & SAYER, J. A. 2019. Renal ciliopathies. *Curr Opin Genet Dev*, 56, 49-60.
- DEVUYST, O., KNOERS, N. V., REMUZZI, G. & SCHAEFER, F. 2014. Rare inherited kidney diseases: challenges, opportunities, and perspectives. *Lancet*, 383, 1844-59.
- DEVUYST, O., OLINGER, E., WEBER, S., ECKARDT, K. U., KMOCH, S., RAMPOLDI, L. & BLEYER, A. J. 2019. Autosomal dominant tubulointerstitial kidney disease. *Nat Rev Dis Primers*, 5, 60.
- DEWEES, S. I., VARGOVÁ, R., HARDIN, K. R., TURN, R. E., DEVI, S., LINNERT, J., WOLFRUM, U., CASPARY, T., ELIÁŠ, M. & KAHN, R. A. 2022. Phylogenetic profiling and cellular analyses of ARL16 reveal roles in traffic of IFT140 and INPP5E. *Mol Biol Cell*, 33, ar33.
- DIXON-SALAZAR, T. J., SILHAVY, J. L., UDPA, N., SCHROTH, J., BIELAS, S., SCHAFFER, A. E., OLVERA, J., BAFNA, V., ZAKI, M. S., ABDEL-SALAM, G. H., MANSOUR, L. A., SELIM, L., ABDEL-HADI, S., MARZOUKI, N., BEN-OMRAN, T., AL-SAANA, N. A., SONMEZ, F. M., CELEP, F., AZAM, M., HILL, K. J., COLLAZO, A., FENSTERMAKER, A. G., NOVARINO, G., AKIZU, N., GARIMELLA, K. V., SOUGNEZ, C., RUSS, C., GABRIEL, S. B. & GLEESON,

- J. G. 2012. Exome sequencing can improve diagnosis and alter patient management. *Sci Transl Med*, 4, 138ra78.
- DOHERTY, D. 2009. Joubert syndrome: insights into brain development, cilium biology, and complex disease. *Semin Pediatr Neurol*, 16, 143-54.
- DOHERTY, D., PARISI, M. A., FINN, L. S., GUNAY-AYGUN, M., AL-MATEEN, M., BATES, D., CLERICUZIO, C., DEMIR, H., DORSCHNER, M., VAN ESSEN, A. J., GAHL, W. A., GENTILE, M., GORDEN, N. T., HIKIDA, A., KNUTZEN, D., OZYUREK, H., PHELPS, I., ROSENTHAL, P., VERLOES, A., WEIGAND, H., CHANCE, P. F., DOBYNS, W. B. & GLASS, I. A. 2010. Mutations in 3 genes (MKS3, CC2D2A and RPGRIP1L) cause COACH syndrome (Joubert syndrome with congenital hepatic fibrosis). *J Med Genet*, 47, 8-21.
- DRIVAS, T. G., LUCAS, A., ZHANG, X. & RITCHIE, M. D. 2021. Mendelian pathway analysis of laboratory traits reveals distinct roles for ciliary subcompartments in common disease pathogenesis. *Am J Hum Genet*, 108, 482-501.
- DRIVAS, T. G., WOJNO, A. P., TUCKER, B. A., STONE, E. M. & BENNETT, J. 2015. Basal exon skipping and genetic pleiotropy: A predictive model of disease pathogenesis. *Sci Transl Med*, 7, 291ra97.
- DULLA, K., SLIJKERMAN, R., VAN DIEPEN, H. C., ALBERT, S., DONA, M., BEUMER, W., TURUNEN, J. J., CHAN, H. L., SCHULKENS, I. A., VORTHOREN, L., DEN BESTEN, C., BUIL, L., SCHMIDT, I., MIAO, J., VENSELAAR, H., ZANG, J., NEUHAUSS, S. C. F., PETERS, T., BROEKMAN, S., PENNINGS, R., KREMER, H., PLATENBURG, G., ADAMSON, P., DE VRIEZE, E. & VAN WIJK, E. 2021. Antisense oligonucleotide-based treatment of retinitis pigmentosa caused by USH2A exon 13 mutations. *Mol Ther*, 29, 2441-2455.
- DUONG PHU, M., BROSS, S., BURKHALTER, M. D. & PHILIPP, M. 2021. Limitations and opportunities in the pharmacotherapy of ciliopathies. *Pharmacol Ther*, 225, 107841.
- DUPONT, M. A., HUMBERT, C., HUBER, C., SIOUR, Q., GUERRERA, I. C., JUNG, V., CHRISTENSEN, A., POULIET, A., GARFA-TRAORÉ, M., NITSCHKÉ, P., INJYAN, M., MILLAR, K., CHITAYAT, D., SHANNON, P., GIRISHA, K. M., SHUKLA, A., MECHLER, C., LORENTZEN, E., BENMERAH, A., CORMIER-DAIRE, V., JEANPIERRE, C., SAUNIER, S. & DELOUS, M. 2019. Human IFT52 mutations uncover a novel role for the protein in microtubule dynamics and centrosome cohesion. *Hum Mol Genet*, 28, 2720-2737.
- DURAN, I., TAYLOR, S. P., ZHANG, W., MARTIN, J., FORLENZA, K. N., SPIRO, R. P., NICKERSON, D. A., BAMSHAD, M., COHN, D. H. & KRAKOW, D. 2016. Destabilization of the IFT-B cilia core complex due to mutations in IFT81 causes a Spectrum of Short-Rib Polydactyly Syndrome. *Sci Rep*, 6, 34232.
- EL HOKAYEM, J., HUBER, C., COUVÉ, A., AZIZA, J., BAUJAT, G., BOUVIER, R., CAVALCANTI, D. P., COLLINS, F. A., CORDIER, M. P., DELEZOIDE, A. L., GONZALES, M., JOHNSON, D., LE MERRER, M., LEVY-MOZZICONACCI, A., LOGET, P., MARTIN-COIGNARD, D., MARTINOVIC, J., MORTIER, G. R., PEREZ, M. J., ROUME, J., SCARANO, G., MUNNICH, A. & CORMIER-DAIRE, V. 2012. NEK1 and DYNC2H1 are both involved in short rib polydactyly Majewski type but not in Beemer Langer cases. *J Med Genet*, 49, 227-33.
- EMIRALIOGLU, N., WALLMEIER, J., OLBRICH, H., OMRAN, H. & OZCELIK, U. 2018. DYNC2H1 mutation causes Jeune syndrome and recurrent lung infections associated with ciliopathy. *The Clinical Respiratory Journal*, 12, 1017-1020.

- EPTING, D., SENARATNE, L. D. S., OTT, E., HOLMGREN, A., SUMATHIPALA, D., LARSEN, S. M., WALLMEIER, J., BRACHT, D., FRIKSTAD, K. M., CROWLEY, S., SIKIRIC, A., BARØY, T., KÄSMANN-KELLNER, B., DECKER, E., DECKER, C., BACHMANN, N., PATZKE, S., PHELPS, I. G., KATSANIS, N., GILES, R., SCHMIDTS, M., ZUCKNICK, M., LIENKAMP, S. S., OMRAN, H., DAVIS, E. E., DOHERTY, D., STRØMME, P., FRENGEN, E., BERGMANN, C. & MISCEO, D. 2020. Loss of CBY1 results in a ciliopathy characterized by features of Joubert syndrome. *Hum Mutat*, 41, 2179-2194.
- ESER, G. & TOPALOĞLU, H. 2022. Current Outline of Exon Skipping Trials in Duchenne Muscular Dystrophy. *Genes (Basel)*, 13.
- ESTRADA-CUZCANO, A., KOENEKOOP, R. K., COPPIETERS, F., KOHL, S., LOPEZ, I., COLLIN, R. W., DE BAERE, E. B., ROELEVeld, D., MAREK, J., BERND, A., ROHRSCHNEIDER, K., VAN DEN BORN, L. I., MEIRE, F., MAUMENE, I. H., JACOBSON, S. G., HOYNG, C. B., ZRENNER, E., CREMERS, F. P. & DEN HOLLANDER, A. I. 2011. IQCB1 mutations in patients with leber congenital amaurosis. *Invest Ophthalmol Vis Sci*, 52, 834-9.
- ESTRADA MALLARINO, L., ENGEL, C., İLİK İ, A., MATICZKA, D., HEYL, F., MÜLLER, B., YAKULOV, T. A., DENGJEL, J., BACKOFEN, R., AKHTAR, A. & WALZ, G. 2020. Nephronophthisis gene products display RNA-binding properties and are recruited to stress granules. *Sci Rep*, 10, 15954.
- EVANS, R. J., SCHWARZ, N., NAGEL-WOLFRUM, K., WOLFRUM, U., HARDCASTLE, A. J. & CHEETHAM, M. E. 2010. The retinitis pigmentosa protein RP2 links pericentriolar vesicle transport between the Golgi and the primary cilium. *Hum Mol Genet*, 19, 1358-67.
- EVERS, M. M., TOONEN, L. J. & VAN ROON-MOM, W. M. 2015. Antisense oligonucleotides in therapy for neurodegenerative disorders. *Adv Drug Deliv Rev*, 87, 90-103.
- FANSA, E. K. & WITTINGHOFER, A. 2016. Sorting of lipidated cargo by the Arl2/Arl3 system. *Small GTPases*, 7, 222-230.
- FEDELES, S. V., GALLAGHER, A. R. & SOMLO, S. 2014. Polycystin-1: a master regulator of intersecting cystic pathways. *Trends Mol Med*, 20, 251-60.
- FEDELES, S. V., TIAN, X., GALLAGHER, A. R., MITOBE, M., NISHIO, S., LEE, S. H., CAI, Y., GENG, L., CREWS, C. M. & SOMLO, S. 2011. A genetic interaction network of five genes for human polycystic kidney and liver diseases defines polycystin-1 as the central determinant of cyst formation. *Nat Genet*, 43, 639-47.
- FEHRENBACH, H., DECKER, C., EISENBERGER, T., FRANK, V., HAMPEL, T., WALDEN, U., AMANN, K. U., KRÜGER-STOLLFUß, I., BOLZ, H. J., HÄFFNER, K., POHL, M. & BERGMANN, C. 2014. Mutations in WDR19 encoding the intraflagellar transport component IFT144 cause a broad spectrum of ciliopathies. *Pediatr Nephrol*, 29, 1451-6.
- FLIEGAUF, M., HORVATH, J., VON SCHNAKENBURG, C., OLBRICH, H., MÜLLER, D., THUMFART, J., SCHERMER, B., PAZOUR, G. J., NEUMANN, H. P., ZENTGRAF, H., BENZING, T. & OMRAN, H. 2006. Nephrocystin specifically localizes to the transition zone of renal and respiratory cilia and photoreceptor connecting cilia. *J Am Soc Nephrol*, 17, 2424-33.
- FOCŞA, I. O., BUDIŞTEANU, M. & BĂLGRĂDEAN, M. 2021. Clinical and genetic heterogeneity of primary ciliopathies (Review). *Int J Mol Med*, 48.
- FORD, M. J., YEYATI, P. L., MALI, G. R., KEIGHREN, M. A., WADDELL, S. H., MJOSENG, H. K., DOUGLAS, A. T., HALL, E. A., SAKAUE-SAWANO, A., MIYAWAKI, A., MEEHAN, R. R., BOULTER, L., JACKSON, I. J., MILL, P. &

- MORT, R. L. 2018. A Cell/Cilia Cycle Biosensor for Single-Cell Kinetics Reveals Persistence of Cilia after G1/S Transition Is a General Property in Cells and Mice. *Developmental Cell*, 47, 509-523.e5.
- FREEDMAN, B. S., LAM, A. Q., SUNDSBAK, J. L., IATRINO, R., SU, X., KOON, S. J., WU, M., DAHERON, L., HARRIS, P. C., ZHOU, J. & BONVENTRE, J. V. 2013. Reduced ciliary polycystin-2 in induced pluripotent stem cells from polycystic kidney disease patients with PKD1 mutations. *J Am Soc Nephrol*, 24, 1571-86.
- FRENCH, L. S., MELLOUGH, C. B., CHEN, F. K. & CARVALHO, L. S. 2020. A Review of Gene, Drug and Cell-Based Therapies for Usher Syndrome. *Front Cell Neurosci*, 14, 183.
- FRIKSTAD, K. M., MOLINARI, E., THORESEN, M., RAMSBOTTOM, S. A., HUGHES, F., LETTEBOER, S. J. F., GILANI, S., SCHINK, K. O., STOKKE, T., GEIMER, S., PEDERSEN, L. B., GILES, R. H., AKHMANOVA, A., ROEPMAN, R., SAYER, J. A. & PATZKE, S. 2019. A CEP104-CSPP1 Complex Is Required for Formation of Primary Cilia Competent in Hedgehog Signaling. *Cell Rep*, 28, 1907-1922.e6.
- FU, L., LI, Y., YAO, S., GUO, Q., YOU, Y., ZHU, X. & LEI, B. 2021. Autosomal Recessive Rod-Cone Dystrophy Associated With Compound Heterozygous Variants in ARL3 Gene. *Front Cell Dev Biol*, 9, 635424.
- FURU, L., ONUCHIC, L. F., GHARAVI, A., HOU, X., ESQUIVEL, E. L., NAGASAWA, Y., BERGMANN, C., SENDEREK, J., AVNER, E., ZERRES, K., GERMINO, G. G., GUAY-WOODFORD, L. M. & SOMLO, S. 2003. Milder presentation of recessive polycystic kidney disease requires presence of amino acid substitution mutations. *J Am Soc Nephrol*, 14, 2004-14.
- GANAI, S., SERPIERI, V. & VALENTE, E. M. 2022. Genotype-phenotype correlates in Joubert syndrome: A review. *Am J Med Genet C Semin Med Genet*, 190, 72-88.
- GANINI, C., AMELIO, I., BERTOLO, R., BOVE, P., BUONOMO, O. C., CANDI, E., CIPRIANI, C., DI DANIELE, N., JUHL, H., MAURIELLO, A., MARANI, C., MARSHALL, J., MELINO, S., MARCHETTI, P., MONTANARO, M., NATALE, M. E., NOVELLI, F., PALMIERI, G., PIACENTINI, M., RENDINA, E. A., ROSELLI, M., SICA, G., TESAURO, M., ROVELLA, V., TISONE, G., SHI, Y., WANG, Y. & MELINO, G. 2021. Global mapping of cancers: The Cancer Genome Atlas and beyond. *Mol Oncol*, 15, 2823-2840.
- GARANTO, A., CHUNG, D. C., DUIJKERS, L., CORRAL-SERRANO, J. C., MESSCHAERT, M., XIAO, R., BENNETT, J., VANDENBERGHE, L. H. & COLLIN, R. W. 2016. In vitro and in vivo rescue of aberrant splicing in CEP290-associated LCA by antisense oligonucleotide delivery. *Hum Mol Genet*, 25, 2552-2563.
- GARCIA-GONZALO, F. R., CORBIT, K. C., SIREROL-PIQUER, M. S., RAMASWAMI, G., OTTO, E. A., NORIEGA, T. R., SEOL, A. D., ROBINSON, J. F., BENNETT, C. L., JOSIFOVA, D. J., GARCIA-VERDUGO, J. M., KATSANIS, N., HILDEBRANDT, F. & REITER, J. F. 2011. A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. *Nat Genet*, 43, 776-84.
- GARCIA, H., SERAFIN, A. S., SILBERMANN, F., PORÉE, E., VIAU, A., MAHAUT, C., BILLOT, K., BIRGY, É., GARFA-TRAORE, M., ROY, S., CECCARELLI, S., MEHRAZ, M., RODRIGUEZ, P. C., DELEGLISE, B., FURIO, L., JABOT-HANIN, F., CAGNARD, N., DEL NERY, E., FILA, M., SIN-MONNOT, S., ANTIGNAC, C., LYONNET, S., KRUG, P., SALOMON, R., ANNEREAU, J. P., BENMERAH, A., DELOUS, M., BRISEÑO-ROA, L. & SAUNIER, S. 2022.

- Agonists of prostaglandin E(2) receptors as potential first in class treatment for nephronophthisis and related ciliopathies. *Proc Natl Acad Sci U S A*, 119, e2115960119.
- GEORGIU, M., YANG, C., ATKINSON, R., PAN, K. T., BUSKIN, A., MOLINA, M. M., COLLIN, J., AL-AAMA, J., GOERTLER, F., LUDWIG, S. E. J., DAVEY, T., LÜHRMANN, R., NAGARAJA-GRELLSCHEID, S., JOHNSON, C. A., ALI, R., ARMSTRONG, L., KOROLCHUK, V., URLAUB, H., MOZAFFARI-JOVIN, S. & LAKO, M. 2022. Activation of autophagy reverses progressive and deleterious protein aggregation in PRPF31 patient-induced pluripotent stem cell-derived retinal pigment epithelium cells. *Clin Transl Med*, 12, e759.
- GERAGHTY, R. M., ORR, S., OLINGER, E., NEATU, R., BARROSO-GIL, M., MABILLARD, H., CONSORTIUM, G. E. R., WILSON, I. & SAYER, J. A. 2023. Use of whole genome sequencing to determine the genetic basis of visceral myopathies including Prune Belly syndrome. *J Rare Dis (Berlin)*, 2, 9.
- GERRELLI, D., LISGO, S., COPP, A. J. & LINDSAY, S. 2015. Enabling research with human embryonic and fetal tissue resources. *Development*, 142, 3073-6.
- GIRARD, M., BIZET, A. A., LACHAUX, A., GONZALES, E., FILHOL, E., COLLARDEAU-FRACHON, S., JEANPIERRE, C., HENRY, C., FABRE, M., VIREMOUNEIX, L., GALMICHE, L., DEBRAY, D., BOLE-FEYSOT, C., NITSCHKE, P., PARIENTE, D., GUETTIER, C., LYONNET, S., HEIDET, L., BERTHOLET, A., JACQUEMIN, E., HENRION-CAUDE, A. & SAUNIER, S. 2016. DCDC2 Mutations Cause Neonatal Sclerosing Cholangitis. *Hum Mutat*, 37, 1025-9.
- GONÇALVES, J. & PELLETIER, L. 2017. The Ciliary Transition Zone: Finding the Pieces and Assembling the Gate. *Molecules and Cells*, 40, 243-253.
- GORDEN, N. T., ARTS, H. H., PARISI, M. A., COENE, K. L., LETTEBOER, S. J., VAN BEERSUM, S. E., MANS, D. A., HIKIDA, A., ECKERT, M., KNUTZEN, D., ALSWAID, A. F., OZYUREK, H., DIBOOGLU, S., OTTO, E. A., LIU, Y., DAVIS, E. E., HUTTER, C. M., BAMMLER, T. K., FARIN, F. M., DORSCHNER, M., TOPCU, M., ZACKAI, E. H., ROSENTHAL, P., OWENS, K. N., KATSANIS, N., VINCENT, J. B., HILDEBRANDT, F., RUBEL, E. W., RAIBLE, D. W., KNOERS, N. V., CHANCE, P. F., ROEPMAN, R., MOENS, C. B., GLASS, I. A. & DOHERTY, D. 2008. CC2D2A is mutated in Joubert syndrome and interacts with the ciliopathy-associated basal body protein CEP290. *Am J Hum Genet*, 83, 559-71.
- GOTTHARDT, K., LOKAJ, M., KOERNER, C., FALK, N., GIESSL, A. & WITTINGHOFER, A. 2015. A G-protein activation cascade from Arl13B to Arl3 and implications for ciliary targeting of lipidated proteins. *Elife*, 4.
- GRABINSKI, T. M., KNEYNSBERG, A., MANFREDSSON, F. P. & KANAAN, N. M. 2015. A method for combining RNAscope in situ hybridization with immunohistochemistry in thick free-floating brain sections and primary neuronal cultures. *PLoS One*, 10, e0120120.
- GRAESSNER, H., ZUREK, B., HOISCHEN, A. & BELTRAN, S. 2021. Solving the unsolved rare diseases in Europe. *Eur J Hum Genet*, 29, 1319-1320.
- GRAMPA, V., DELOUD, M., ZAIDAN, M., ODYE, G., THOMAS, S., ELKHARTOUI, N., FILHOL, E., NIEL, O., SILBERMANN, F., LEBRETON, C., COLLARDEAU-FRACHON, S., ROUVET, I., ALESSANDRI, J. L., DEVISME, L., DIEUX-COESLIER, A., CORDIER, M. P., CAPRI, Y., KHUNG-SAVATOVSKY, S., SIGAUDY, S., SALOMON, R., ANTIGNAC, C., GUBLER, M. C., BENMERAH, A., TERZI, F., ATTIE-BITACH, T., JEANPIERRE, C. & SAUNIER, S. 2016. Novel NEK8 Mutations Cause Severe Syndromic Renal Cystic Dysplasia through YAP Dysregulation. *PLoS Genet*, 12, e1005894.

- GROCHOWSKY, A. & GUNAY-AYGUN, M. 2019. Clinical characteristics of individual organ system disease in non-motile ciliopathies. *Transl Sci Rare Dis*, 4, 1-23.
- GROOPMAN, E. E., MARASA, M., CAMERON-CHRISTIE, S., PETROVSKI, S., AGGARWAL, V. S., MILO-RASOULY, H., LI, Y., ZHANG, J., NESTOR, J., KRITHIVASAN, P., LAM, W. Y., MITROTTI, A., PIVA, S., KIL, B. H., CHATTERJEE, D., REINGOLD, R., BRADBURY, D., DIVECCHIA, M., SNYDER, H., MU, X., MEHL, K., BALDERES, O., FASEL, D. A., WENG, C., RADHAKRISHNAN, J., CANETTA, P., APPEL, G. B., BOMBAC, A. S., AHN, W., UY, N. S., ALAM, S., COHEN, D. J., CREW, R. J., DUBE, G. K., RAO, M. K., KAMALAKARAN, S., COPELAND, B., REN, Z., BRIDGERS, J., MALONE, C. D., MEBANE, C. M., DAGAONKAR, N., FELLSTRÖM, B. C., HAEFLIGER, C., MOHAN, S., SANNA-CHERCHI, S., KIRYLUK, K., FLECKNER, J., MARCH, R., PLATT, A., GOLDSTEIN, D. B. & GHARAVI, A. G. 2019. Diagnostic Utility of Exome Sequencing for Kidney Disease. *N Engl J Med*, 380, 142-151.
- GUDBJARTSSON, D. F., HELGASON, H., GUDJONSSON, S. A., ZINK, F., ODDSON, A., GYLASON, A., BESENBACHER, S., MAGNUSSON, G., HALLDORSSON, B. V., HJARTARSON, E., SIGURDSSON, G. T., STACEY, S. N., FRIGGE, M. L., HOLM, H., SAEMUNDSDOTTIR, J., HELGADOTTIR, H. T., JOHANNSDOTTIR, H., SIGFUSSON, G., THORGEIRSSON, G., SVERRISSON, J. T., GRETARSDOTTIR, S., WALTERS, G. B., RAFNAR, T., THJODLEIFSSON, B., BJORNSSON, E. S., OLAFSSON, S., THORARINSDOTTIR, H., STEINGRIMSDOTTIR, T., GUDMUNDSDOTTIR, T. S., THEODORS, A., JONASSON, J. G., SIGURDSSON, A., BJORNSDOTTIR, G., JONSSON, J. J., THORARENSEN, O., LUDVIGSSON, P., GUDBJARTSSON, H., EYJOLFSSON, G. I., SIGURDARDOTTIR, O., OLAFSSON, I., ARNAR, D. O., MAGNUSSON, O. T., KONG, A., MASSON, G., THORSTEINSDOTTIR, U., HELGASON, A., SULEM, P. & STEFANSSON, K. 2015. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet*, 47, 435-44.
- GUO, J., GAO, J., LI, Z., GONG, Y., MAN, X., JIN, J. & WU, H. 2013. Adenovirus vector-mediated Gli1 siRNA induces growth inhibition and apoptosis in human pancreatic cancer with Smo-dependent or Smo-independent Hh pathway activation in vitro and in vivo. *Cancer Lett*, 339, 185-94.
- HAENDEL, M., VASILEVSKY, N., UNNI, D., BOLOGA, C., HARRIS, N., REHM, H., HAMOSH, A., BAYNAM, G., GROZA, T., MCMURRY, J., DAWKINS, H., RATH, A., THAXON, C., BOCCI, G., JOACHIMIAK, M. P., KÖHLER, S., ROBINSON, P. N., MUNGALL, C. & OPREA, T. I. 2020. How many rare diseases are there? *Nat Rev Drug Discov*, 19, 77-78.
- HALBRITTER, J., BIZET, A. A., SCHMIDTS, M., PORATH, J. D., BRAUN, D. A., GEE, H. Y., MCINERNEY-LEO, A. M., KRUG, P., FILHOL, E., DAVIS, E. E., AIRIK, R., CZARNECKI, P. G., LEHMAN, A. M., TRNKA, P., NITSCHKÉ, P., BOLE-FEYSOT, C., SCHUELER, M., KNEBELMANN, B., BURTEY, S., SZABÓ, A. J., TORY, K., LEO, P. J., GARDINER, B., MCKENZIE, F. A., ZANKL, A., BROWN, M. A., HARTLEY, J. L., MAHER, E. R., LI, C., LEROUX, M. R., SCAMBLER, P. J., ZHAN, S. H., JONES, S. J., KAYSERILI, H., TUYSUZ, B., MOORANI, K. N., CONSTANTINESCU, A., KRANTZ, I. D., KAPLAN, B. S., SHAH, J. V., HURD, T. W., DOHERTY, D., KATSANIS, N., DUNCAN, E. L., OTTO, E. A., BEALES, P. L., MITCHISON, H. M., SAUNIER, S. & HILDEBRANDT, F. 2013a. Defects in the IFT-B component IFT172 cause Jeune and Mainzer-Saldino syndromes in humans. *Am J Hum Genet*, 93, 915-25.

- HALBRITTER, J., PORATH, J. D., DIAZ, K. A., BRAUN, D. A., KOHL, S., CHAKI, M., ALLEN, S. J., SOLIMAN, N. A., HILDEBRANDT, F. & OTTO, E. A. 2013b. Identification of 99 novel mutations in a worldwide cohort of 1,056 patients with a nephronophthisis-related ciliopathy. *Hum Genet*, 132, 865-84.
- HALDIPUR, P., ALDINGER, K. A., BERNARDO, S., DENG, M., TIMMS, A. E., OVERMAN, L. M., WINTER, C., LISGO, S. N., RAZAVI, F., SILVESTRI, E., MANGANARO, L., ADLE-BIASSETTE, H., GUIMIOT, F., RUSSO, R., KIDRON, D., HOF, P. R., GERRELLI, D., LINDSAY, S. J., DOBYNS, W. B., GLASS, I. A., ALEXANDRE, P. & MILLEN, K. J. 2019. Spatiotemporal expansion of primary progenitor zones in the developing human cerebellum. *Science*, 366, 454-460.
- HAMOSH, A., SCOTT, A. F., AMBERGER, J., VALLE, D. & MCKUSICK, V. A. 2000. Online Mendelian Inheritance in Man (OMIM). *Hum Mutat*, 15, 57-61.
- HANKE-GOGOKHIA, C., WU, Z., GERSTNER, C. D., FREDERICK, J. M., ZHANG, H. & BAEHR, W. 2016. Arf-like Protein 3 (ARL3) Regulates Protein Trafficking and Ciliogenesis in Mouse Photoreceptors. *J Biol Chem*, 291, 7142-55.
- HASAN, M. M., RAGNARSSON, L., CARDOSO, F. C. & LEWIS, R. J. 2021. Transfection methods for high-throughput cellular assays of voltage-gated calcium and sodium channels involved in pain. *PLoS One*, 16, e0243645.
- HASSAN, S., WOLF, M. T. F., UMAÑA, L. A., MALIK, S., UDDIN, N., ANDERSEN, J. & AQUL, A. 2020. Homozygous NEK8 Mutations in Siblings With Neonatal Cholestasis Progressing to End-stage Liver, Renal, and Cardiac Disease. *J Pediatr Gastroenterol Nutr*, 70, e19-e22.
- HENDRICKSON, A. 2016. Development of Retinal Layers in Prenatal Human Retina. *American Journal of Ophthalmology*, 161, 29-35.e1.
- HIGGINS, K., MOORE, B. A., BERBEROVIC, Z., ADISSU, H. A., ESKANDARIAN, M., FLENNIKEN, A. M., SHAO, A., IMAI, D. M., CLARY, D., LANOUE, L., NEWBIGGING, S., NUTTER, L. M. J., ADAMS, D. J., BOSCH, F., BRAUN, R. E., BROWN, S. D. M., DICKINSON, M. E., DOBBIE, M., FLICEK, P., GAO, X., GALANDE, S., GROBLER, A., HEANEY, J. D., HERAULT, Y., DE ANGELIS, M. H., CHIN, H. G., MAMMANO, F., QIN, C., SHIROISHI, T., SEDLACEK, R., SEONG, J. K., XU, Y., LLOYD, K. C. K., MCKERLIE, C. & MOSHIRI, A. 2022. Analysis of genome-wide knockout mouse database identifies candidate ciliopathy genes. *Sci Rep*, 12, 20791.
- HILDEBRANDT, F., BENZING, T. & KATSANIS, N. 2011. Ciliopathies. *The New England Journal of Medicine*, 364, 1533-1543.
- HÖCK, M., WEGLEITER, K., RALSER, E., KIECHL-KOLENDORFER, U., SCHOLL-BÜRG, S., FAUTH, C., STEICHEN, E., PICHLER, K., LEFEBER, D. J., MATTHJIS, G., KELDERMANS, L., MAURER, K., ZSCHOCKE, J. & KARALL, D. 2015. ALG8-CDG: novel patients and review of the literature. *Orphanet J Rare Dis*, 10, 73.
- HOFF, S., HALBRITTER, J., EPTING, D., FRANK, V., NGUYEN, T. M., VAN REEUWIJK, J., BOEHLKE, C., SCHELL, C., YASUNAGA, T., HELMSTÄDTER, M., MERGEN, M., FILHOL, E., BOLDT, K., HORN, N., UEFFING, M., OTTO, E. A., EISENBERGER, T., ELTING, M. W., VAN WIJK, J. A., BOCKENHAUER, D., SEBIRE, N. J., RITTIG, S., VYBERG, M., RING, T., POHL, M., PAPE, L., NEUHAUS, T. J., ELSHAKHS, N. A., KOON, S. J., HARRIS, P. C., GRAHAMMER, F., HUBER, T. B., KUEHN, E. W., KRAMER-ZUCKER, A., BOLZ, H. J., ROEPMAN, R., SAUNIER, S., WALZ, G., HILDEBRANDT, F., BERGMANN, C. & LIENKAMP, S. S. 2013. ANKS6 is a central component of a nephronophthisis module linking NEK8 to INVS and NPHP3. *Nat Genet*, 45, 951-6.

- HOLTAN, J. P., TEIGEN, K., AUKRUST, I., BRAGADOTTIR, R. & HOUGE, G. 2019. Dominant ARL3-related retinitis pigmentosa. *Ophthalmic Genet*, 40, 124-128.
- HOON, M., OKAWA, H., DELLA SANTINA, L. & WONG, R. O. L. 2014. Functional architecture of the retina: development and disease. *Progress in retinal and eye research*, 42, 44-84.
- HOPP, K., HEYER, C. M., HOMMERDING, C. J., HENKE, S. A., SUNDSBAK, J. L., PATEL, S., PATEL, P., CONSUGAR, M. B., CZARNECKI, P. G., GLIEM, T. J., TORRES, V. E., ROSSETTI, S. & HARRIS, P. C. 2011. B9D1 is revealed as a novel Meckel syndrome (MKS) gene by targeted exon-enriched next-generation sequencing and deletion analysis. *Hum Mol Genet*, 20, 2524-34.
- HU, J. & HARRIS, P. C. 2020. Regulation of polycystin expression, maturation and trafficking. *Cell Signal*, 72, 109630.
- HUA, K. & FERLAND, R. J. 2018. Primary cilia proteins: ciliary and extraciliary sites and functions. *Cellular and Molecular Life Sciences*, 75, 1521-1540.
- HUANG, L. & LIPSCHUTZ, J. H. 2014. Cilia and polycystic kidney disease, kith and kin. *Birth Defects Res C Embryo Today*, 102, 174-85.
- HUELSZ-PRINCE, G., KOK, R. N. U., GOOS, Y., BRUENS, L., ZHENG, X., ELLENBROEK, S., VAN RHEENEN, J., TANS, S. & VAN ZON, J. S. 2022. Mother cells control daughter cell proliferation in intestinal organoids to minimize proliferation fluctuations. *Elife*, 11.
- HULL, S., OWEN, N., ISLAM, F., TRACEY-WHITE, D., PLAGNOL, V., HOLDER, G. E., MICHAELIDES, M., CARSS, K., RAYMOND, F. L., ROZET, J. M., RAMSDEN, S. C., BLACK, G. C., PERRAULT, I., SARKAR, A., MOOSAJEE, M., WEBSTER, A. R., ARNO, G. & MOORE, A. T. 2016. Nonsyndromic Retinal Dystrophy due to Bi-Allelic Mutations in the Ciliary Transport Gene IFT140. *Invest Ophthalmol Vis Sci*, 57, 1053-62.
- HUMPHREYS, I. R., PEI, J., BAEK, M., KRISHNAKUMAR, A., ANISHCHENKO, I., OVCHINNIKOV, S., ZHANG, J., NESS, T. J., BANJADE, S., BAGDE, S. R., STANCHEVA, V. G., LI, X. H., LIU, K., ZHENG, Z., BARRERO, D. J., ROY, U., KUPER, J., FERNÁNDEZ, I. S., SZAKAL, B., BRANZEI, D., RIZO, J., KISKER, C., GREENE, E. C., BIGGINS, S., KEENEY, S., MILLER, E. A., FROMME, J. C., HENDRICKSON, T. L., CONG, Q. & BAKER, D. 2021. Computed structures of core eukaryotic protein complexes. *Science*, 374, eabm4805.
- HURD, T. W., OTTO, E. A., MISHIMA, E., GEE, H. Y., INOUE, H., INAZU, M., YAMADA, H., HALBRITTER, J., SEKI, G., KONISHI, M., ZHOU, W., YAMANE, T., MURAKAMI, S., CARIDI, G., GHIGGERI, G., ABE, T. & HILDEBRANDT, F. 2013. Mutation of the Mg²⁺ transporter SLC41A1 results in a nephronophthisis-like phenotype. *J Am Soc Nephrol*, 24, 967-77.
- HUYNH, V. T., AUDRÉZET, M. P., SAYER, J. A., ONG, A. C., LEFEVRE, S., LE BRUN, V., DESPRÉS, A., SENUM, S. R., CHEBIB, F. T., BARROSO-GIL, M., PATEL, C., MALLETT, A. J., GOEL, H., MALLAWAARACHCHI, A. C., VAN EERDE, A. M., PONLOT, E., KRIBS, M., LE MEUR, Y., HARRIS, P. C. & CORNEC-LE GALL, E. 2020. Clinical spectrum, prognosis and estimated prevalence of DNAJB11-kidney disease. *Kidney Int*, 98, 476-487.
- HYNES, A. M., GILES, R. H., SRIVASTAVA, S., ELEY, L., WHITEHEAD, J., DANILENKO, M., RAMAN, S., SLAATS, G. G., COLVILLE, J. G., AJZENBERG, H., KROES, H. Y., THELWALL, P. E., SIMMONS, N. L., MILES, C. G. & SAYER, J. A. 2014. Murine Joubert syndrome reveals Hedgehog signaling defects as a potential therapeutic target for nephronophthisis. *Proceedings of the National Academy of Sciences*, 111, 9893-9898.

- IANNICELLI, M., BRANCATI, F., MOUGOU-ZERELLI, S., MAZZOTTA, A., THOMAS, S., ELKHARTOUIFI, N., TRAVAGLINI, L., GOMES, C., ARDISSINO, G. L., BERTINI, E., BOLTSCHAUSER, E., CASTORINA, P., D'ARRIGO, S., FISCHETTO, R., LEROY, B., LOGET, P., BONNIERE, M., STARCK, L., TANTAU, J., GENTILIN, B., MAJORE, S., SWISTUN, D., FLORI, E., LALATTA, F., PANTALEONI, C., PENZIEN, J., GRAMMATICO, P., INTERNATIONAL, J. S. G., DALLAPICCOLA, B., GLEESON, J. G., ATTIE-BITACH, T. & VALENTE, E. M. 2010. Novel TMEM67 mutations and genotype-phenotype correlates in meckelin-related ciliopathies. *Hum Mutat*, 31, E1319-31.
- ISHIDA, Y., KOBAYASHI, T., CHIBA, S., KATOH, Y. & NAKAYAMA, K. 2021. Molecular basis of ciliary defects caused by compound heterozygousIFT144/WDR19 mutations found in cranioectodermal dysplasia. *Hum Mol Genet*, 30, 213-225.
- ISMAIL, S. A., CHEN, Y. X., RUSINOVA, A., CHANDRA, A., BIERBAUM, M., GREMER, L., TRIOLA, G., WALDMANN, H., BASTIAENS, P. I. & WITTINGHOFER, A. 2011. Arl2-GTP and Arl3-GTP regulate a GDI-like transport system for farnesylated cargo. *Nat Chem Biol*, 7, 942-9.
- IVANCHENKO, M. V., HATHAWAY, D. M., KLEIN, A. J., PAN, B., STRELKOVA, O., DE-LA-TORRE, P., WU, X., PETERS, C. W., MULHALL, E. M., BOOTH, K. T., GOLDSTEIN, C., BROWER, J., SOTOMAYOR, M., INDZHYKULIAN, A. A. & COREY, D. P. 2023. Mini-PCDH15 gene therapy rescues hearing in a mouse model of Usher syndrome type 1F. *Nat Commun*, 14, 2400.
- JACINTO, F. V., LINK, W. & FERREIRA, B. I. 2020. CRISPR/Cas9-mediated genome editing: From basic research to translational medicine. *J Cell Mol Med*, 24, 3766-3778.
- JACOBSON, S. G., CIDECIYAN, A. V., HO, A. C., ROMAN, A. J., WU, V., GARAFALO, A. V., SUMAROKA, A., KRISHNAN, A. K., SWIDER, M., MASCIO, A. A., KAY, C. N., YOON, D., FUJITA, K. P., BOYE, S. L., PESHENKO, I. V., DIZHOOR, A. M. & BOYE, S. E. 2022. Night vision restored in days after decades of congenital blindness. *iScience*, 25, 105274.
- JENSEN, V. L., LAMBACHER, N. J., LI, C., MOHAN, S., WILLIAMS, C. L., INGLIS, P. N., YODER, B. K., BLACQUE, O. E. & LEROUX, M. R. 2018. Role for intraflagellar transport in building a functional transition zone. *EMBO reports*, 19, e45862-e45862.
- JIA, Y., LI, S., GUO, X., LEI, B., HU, J., XU, X. H. & ZHANG, W. 2022. Selfee, self-supervised features extraction of animal behaviors. *Elife*, 11.
- JIN, X., MUNTEAN, B. S., AAL-AABODA, M. S., DUAN, Q., ZHOU, J. & NAULI, S. M. 2014. L-type calcium channel modulates cystic kidney phenotype. *Biochim Biophys Acta*, 1842, 1518-26.
- JOHNSTON, J. J., LEE, C., WENTZENSEN, I. M., PARISI, M. A., CRENSHAW, M. M., SAPP, J. C., GROSS, J. M., WALLINGFORD, J. B. & BIESECKER, L. G. 2017. Compound heterozygous alterations in intraflagellar transport protein CLUAP1 in a child with a novel Joubert and oral-facial-digital overlap syndrome. *Cold Spring Harb Mol Case Stud*, 3.
- JONES, D., FIOZZO, F., WATERS, B., MCKNIGHT, D. & BROWN, S. 2014. First-trimester diagnosis of Meckel-Gruber syndrome by fetal ultrasound with molecular identification of CC2D2A mutations by next-generation sequencing. *Ultrasound Obstet Gynecol*, 44, 719-21.
- JOO, K., KIM, C. G., LEE, M. S., MOON, H. Y., LEE, S. H., KIM, M. J., KWEON, H. S., PARK, W. Y., KIM, C. H., GLEESON, J. G. & KIM, J. 2013. CCDC41 is

- required for ciliary vesicle docking to the mother centriole. *Proc Natl Acad Sci U S A*, 110, 5987-92.
- JOSEPH, N., AL-JASSAR, C., JOHNSON, C. M., ANDREEVA, A., BARNABAS, D. D., FREUND, S. M. V., GERGELY, F. & VAN BREUGEL, M. 2018. Disease-Associated Mutations in CEP120 Destabilize the Protein and Impair Ciliogenesis. *Cell Rep*, 23, 2805-2818.
- JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ŽÍDEK, A., POTAPENKO, A., BRIDGLAND, A., MEYER, C., KOHL, S. A. A., BALLARD, A. J., COWIE, A., ROMERA-PAREDES, B., NIKOLOV, S., JAIN, R., ADLER, J., BACK, T., PETERSEN, S., REIMAN, D., CLANCY, E., ZIELINSKI, M., STEINEGGER, M., PACHOLSKA, M., BERGHAMMER, T., BODENSTEIN, S., SILVER, D., VINYALS, O., SENIOR, A. W., KAVUKCUOGLU, K., KOHLI, P. & HASSABIS, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583-589.
- KAHN, R. A., DER, C. J. & BOKOCH, G. M. 1992. The ras superfamily of GTP-binding proteins: guidelines on nomenclature. *Faseb j*, 6, 2512-3.
- KALCHEIM, C., BARDE, Y. A., THOENEN, H. & LE DOUARIN, N. M. 1987. In vivo effect of brain-derived neurotrophic factor on the survival of developing dorsal root ganglion cells. *The EMBO journal*, 6, 2871-2873.
- KANDA, G. N., TSUZUKI, T., TERADA, M., SAKAI, N., MOTOZAWA, N., MASUDA, T., NISHIDA, M., WATANABE, C. T., HIGASHI, T., Horiguchi, S. A., KUDO, T., KAMEI, M., SUNAGAWA, G. A., MATSUKUMA, K., SAKURADA, T., OZAWA, Y., TAKAHASHI, M., TAKAHASHI, K. & NATSUME, T. 2022. Robotic search for optimal cell culture in regenerative medicine. *Elife*, 11.
- KARCZEWSKI, K. J., FRANCIOLI, L. C., TIAO, G., CUMMINGS, B. B., ALFOLDI, J., WANG, Q., COLLINS, R. L., LARICCHIA, K. M., GANNA, A., BIRNBAUM, D. P., GAUTHIER, L. D., BRAND, H., SOLOMONSON, M., WATTS, N. A., RHODES, D., SINGER-BERK, M., ENGLAND, E. M., SEABY, E. G., KOSMICKI, J. A., WALTERS, R. K., TASHMAN, K., FARJOUN, Y., BANKS, E., POTERBA, T., WANG, A., SEED, C., WHIFFIN, N., CHONG, J. X., SAMOCHA, K. E., PIERCE-HOFFMAN, E., ZAPPALA, Z., O'DONNELL-LURIA, A. H., MINIKEL, E. V., WEISBURD, B., LEK, M., WARE, J. S., VITTAL, C., ARMEAN, I. M., BERGELSON, L., CIBULSKIS, K., CONNOLY, K. M., COVARRUBIAS, M., DONNELLY, S., FERRIERA, S., GABRIEL, S., GENTRY, J., GUPTA, N., JEANDET, T., KAPLAN, D., LLANWARNE, C., MUNSHI, R., NOVOD, S., PETRILLO, N., ROAZEN, D., RUANO-RUBIO, V., SALTZMAN, A., SCHLEICHER, M., SOTO, J., TIBBETTS, K., TOLONEN, C., WADE, G., TALKOWSKI, M. E., GENOME AGGREGATION DATABASE, C., NEALE, B. M., DALY, M. J. & MACARTHUR, D. G. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581, 434-443.
- KAUR, G. & DUFOUR, J. M. 2012. Cell lines: Valuable tools or useless artifacts. *Spermatogenesis*, 2, 1-5.
- KEMPENEERS, C. & CHILVERS, M. A. 2018. To beat, or not to beat, that is question! The spectrum of ciliopathies. *Pediatr Pulmonol*, 53, 1122-1129.
- KIEWISZ, R., FABIG, G., CONWAY, W., BAUM, D., NEEDLEMAN, D. & MÜLLER-REICHERT, T. 2022. Three-dimensional structure of kinetochore-fibers in human mitotic spindles. *Elife*, 11.
- KIM, I., FU, Y., HUI, K., MOECKEL, G., MAI, W., LI, C., LIANG, D., ZHAO, P., MA, J., CHEN, X. Z., GEORGE, A. L., JR., COFFEY, R. J., FENG, Z. P. & WU, G.

2008. Fibrocystin/polyductin modulates renal tubular formation by regulating polycystin-2 expression and function. *J Am Soc Nephrol*, 19, 455-68.
- KOBAYASHI, T. & DYNLACHT, B. D. 2011. Regulating the transition from centriole to basal body. *Journal of Cell Biology*, 193, 435-444.
- KOLE, R. & KRIEG, A. M. 2015. Exon skipping therapy for Duchenne muscular dystrophy. *Adv Drug Deliv Rev*, 87, 104-7.
- KOMAKI, H., NAGATA, T., SAITO, T., MASUDA, S., TAKESHITA, E., SASAKI, M., TACHIMORI, H., NAKAMURA, H., AOKI, Y. & TAKEDA, S. 2018. Systemic administration of the antisense oligonucleotide NS-065/NCNP-01 for skipping of exon 53 in patients with Duchenne muscular dystrophy. *Sci Transl Med*, 10.
- KOPANOS, C., TSIOLKAS, V., KOURIS, A., CHAPPLE, C. E., ALBARCA AGUILERA, M., MEYER, R. & MASSOURAS, A. 2019. VarSome: the human genomic variant search engine. *Bioinformatics*, 35, 1978-1980.
- KORVALA, J., JÜPPNER, H., MÄKITIE, O., SOCHETT, E., SCHNABEL, D., MORA, S., BARTELS, C. F., WARMAN, M. L., DERASKA, D., COLE, W. G., HARTIKKA, H., ALA-KOKKO, L. & MÄNNIKÖ, M. 2012. Mutations in LRP5 cause primary osteoporosis without features of OI by reducing Wnt signaling activity. *BMC Med Genet*, 13, 26.
- KOYAMA, S., SATO, H., WADA, M., KAWANAMI, T., EMI, M. & KATO, T. 2017. Whole-exome sequencing and digital PCR identified a novel compound heterozygous mutation in the NPHP1 gene in a case of Joubert syndrome and related disorders. *BMC Med Genet*, 18, 37.
- KRAMES, E. S. 2014. The role of the dorsal root ganglion in the development of neuropathic pain. *Pain Med*, 15, 1669-85.
- KUCHARSKI, M., MROWIEC, P. & OCŁOŃ, E. 2021. Current standards and pitfalls associated with the transfection of primary fibroblast cells. *Biotechnol Prog*, 37, e3152.
- LAM, Z., ALBABA, S., STUDY, D. & BALASUBRAMANIAN, M. 2020. Atypical, milder presentation in a child with CC2D2A and KIDINS220 variants. *Clin Dysmorphol*, 29, 10-16.
- LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, Y., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.

- LANG, R., LIU, G., SHI, Y., BHARADWAJ, S., LENG, X., ZHOU, X., LIU, H., ATALA, A. & ZHANG, Y. 2013. Self-renewal and differentiation capacity of urine-derived stem cells after urine preservation for 24 hours. *PLoS One*, 8, e53980.
- LANKTREE, M. B., HAGHIGHI, A., DI BARI, I., SONG, X. & PEI, Y. 2021. Insights into Autosomal Dominant Polycystic Kidney Disease from Genetic Studies. *Clin J Am Soc Nephrol*, 16, 790-799.
- LANKTREE, M. B., HAGHIGHI, A., GUIARD, E., ILIUTA, I. A., SONG, X., HARRIS, P. C., PATERSON, A. D. & PEI, Y. 2018. Prevalence Estimates of Polycystic Kidney and Liver Disease by Population Sequencing. *J Am Soc Nephrol*, 29, 2593-2600.
- LATOUR, B. L., VAN DE WEGHE, J. C., RUSTERHOLZ, T. D., LETTEBOER, S. J., GOMEZ, A., SHAHEEN, R., GESEMANN, M., KARAMZADE, A., ASADOLLAHI, M., BARROSO-GIL, M., CHITRE, M., GROUT, M. E., VAN REEUWIJK, J., VAN BEERSUM, S. E., MILLER, C. V., DEMPSEY, J. C., MORSY, H., BAMSHAD, M. J., NICKERSON, D. A., NEUHAUSS, S. C., BOLDT, K., UEFFING, M., KERAMATIPOUR, M., SAYER, J. A., ALKURAYA, F. S., BACHMANN-GAGESCU, R., ROEPMAN, R. & DOHERTY, D. 2020. Dysfunction of the ciliary ARMC9/TOGARAM1 protein module causes Joubert syndrome. *J Clin Invest*, 130, 4423-4439.
- LATOUR, B. L., VAN DE WEGHE, J. C., RUSTERHOLZ, T. D. S., LETTEBOER, S. J. F., GOMEZ, A., SHAHEEN, R., GESEMANN, M., GROUT, M. E., VAN REEUWIJK, J., VAN BEERSUM, S. E. C., MILLER, C. V., DEMPSEY, J. C., MORSY, H., BAMSHAD, M. J., NICKERSON, D. A., NEUHAUSS, S. C. F., BOLDT, K., UEFFING, M., ALKURAYA, F. S., BACHMANN-GAGESCU, R., ROEPMAN, R. & DOHERTY, D. 2019. ARMC9 and TOGARAM1 define a Joubert syndrome-associated protein module that regulates axonemal post-translational modifications and cilium stability. *bioRxiv*, 817213.
- LECAT, A., DI VALENTIN, E., SOMJA, J., JOURDAN, S., FILLET, M., KUFER, T. A., HABRAKEN, Y., SADZOT, C., LOUIS, E., DELVENNE, P., PIETTE, J. & LEGRAND-POELS, S. 2012. The c-Jun N-terminal kinase (JNK)-binding protein (JNKBP1) acts as a negative regulator of NOD2 protein signaling by inhibiting its oligomerization process. *J Biol Chem*, 287, 29213-26.
- LEE, J. J. A., SAITO, T., DUDDY, W., TAKEDA, S. & YOKOTA, T. 2018. Direct Reprogramming of Human DMD Fibroblasts into Myotubes for In Vitro Evaluation of Antisense-Mediated Exon Skipping and Exons 45-55 Skipping Accompanied by Rescue of Dystrophin Expression. *Methods Mol Biol*, 1828, 141-150.
- LEHMAN, A. M., EYDOUX, P., DOHERTY, D., GLASS, I. A., CHITAYAT, D., CHUNG, B. Y. H., LANGLOIS, S., YONG, S. L., LOWRY, R. B., HILDEBRANDT, F. & TRNKA, P. 2010. Co-occurrence of Joubert syndrome and Jeune asphyxiating thoracic dystrophy. *American Journal of Medical Genetics Part A*, 5, 1411-1419.
- LEMOINE, H., RAUD, L., FOULQUIER, F., SAYER, J. A., LAMBERT, B., OLINGER, E., LEFÈVRE, S., KNEBELMANN, B., HARRIS, P. C., TROUVÉ, P., DESPRÈS, A., DUNEAU, G., MATIGNON, M., POYET, A., JOURDEC-CHICHE, N., GUERROT, D., LEMOINE, S., SERET, G., BARROSO-GIL, M., BINGHAM, C., GILBERT, R., LE MEUR, Y., AUDRÉZET, M. P. & CORNEC-LE GALL, E. 2022. Monoallelic pathogenic ALG5 variants cause atypical polycystic kidney disease and interstitial fibrosis. *Am J Hum Genet*, 109, 1484-1499.
- LENZINI, L., IORI, E., SCANNAPIECO, F., CARRARO, G., AVOGARO, A. & VITTURI, N. 2022. Urine-Derived Epithelial Cells as a New Model to Study

Renal Metabolic Phenotypes of Patients with Glycogen Storage Disease 1a.
Int J Mol Sci, 24.

- LEWIS, W. R., BALES, K. L., REVELL, D. Z., CROYLE, M. J., ENGLE, S. E., SONG, C. J., MALARKEY, E. B., UYTINGCO, C. R., SHAN, D., ANTONELLIS, P. J., NAGY, T. R., KESTERSON, R. A., MRUG, M. M., MARTEENS, J. R., BERBARI, N. F., GROSS, A. K. & YODER, B. K. 2019. Mks6 mutations reveal tissue- and cell type-specific roles for the cilia transition zone. *FASEB J*, 33, 1440-1455.
- LI, A., DAVILA, S., FURU, L., QIAN, Q., TIAN, X., KAMATH, P. S., KING, B. F., TORRES, V. E. & SOMLO, S. 2003. Mutations in PRKCSH cause isolated autosomal dominant polycystic liver disease. *Am J Hum Genet*, 72, 691-703.
- LI, L. X. & LI, X. 2021. Epigenetically Mediated Ciliogenesis and Cell Cycle Regulation, and Their Translational Potential. *Cells*, 10.
- LIN, Y. N., WU, C. T., LIN, Y. C., HSU, W. B., TANG, C. J., CHANG, C. W. & TANG, T. K. 2013. CEP120 interacts with CPAP and positively regulates centriole elongation. *J Cell Biol*, 202, 211-9.
- LINDSAY, S. J., XU, Y., LISGO, S. N., HARKIN, L. F., COPP, A. J., GERRELLI, D., CLOWRY, G. J., TALBOT, A., KEOGH, M. J., COXHEAD, J., SANTIBANEZ-KOREF, M. & CHINNERY, P. F. 2016. HDBR Expression: A Unique Resource for Global and Individual Gene Expression Studies during Early Human Brain Development. *Front Neuroanat*, 10, 86.
- LIU, E. T., BOLCUN-FILAS, E., GRASS, D. S., LUTZ, C., MURRAY, S., SHULTZ, L. & ROSENTHAL, N. 2017. Of mice and CRISPR: The post-CRISPR future of the mouse as a model system for the human condition. *EMBO Rep*, 18, 187-193.
- LOVERA, M. & LÜDERS, J. 2021. The ciliary impact of nonciliary gene mutations. *Trends Cell Biol*, 31, 876-887.
- LU, H., GALEANO, M. C. R., OTT, E., KAESLIN, G., KAUSALYA, P. J., KRAMER, C., ORTIZ-BRÜCHLE, N., HILGER, N., METZIS, V., HIERSCHE, M., TAY, S. Y., TUNNINGLEY, R., VIJ, S., COURTNEY, A. D., WHITTLE, B., WÜHL, E., VESTER, U., HARTLEBEN, B., NEUBER, S., FRANK, V., LITTLE, M. H., EPTING, D., PAPATHANASIOU, P., PERKINS, A. C., WRIGHT, G. D., HUNZIKER, W., GEE, H. Y., OTTO, E. A., ZERRES, K., HILDEBRANDT, F., ROY, S., WICKING, C. & BERGMANN, C. 2017. Mutations in DZIP1L, which encodes a ciliary-transition-zone protein, cause autosomal recessive polycystic kidney disease. *Nat Genet*, 49, 1025-1034.
- LUO, M., LIN, Z., ZHU, T., JIN, M., MENG, D., HE, R., CAO, Z., SHEN, Y., LU, C., CAI, R., ZHAO, Y., WANG, X., LI, H., WU, S., ZOU, X., LUO, G., CAO, L., HUANG, M., JIAO, H., GAO, H., SUI, R., ZHAO, C., MA, X. & CAO, M. 2021. Disrupted intraflagellar transport due to IFT74 variants causes Joubert syndrome. *Genet Med*, 23, 1041-1049.
- LYONS, R. A., SARIDOGAN, E. & DJAHANBAKHCH, O. 2006. The reproductive significance of human Fallopian tube cilia. *Human Reproduction Update*, 12, 363-372.
- MACIA, M. S., HALBRITTER, J., DELOUS, M., BREDRUP, C., GUTTER, A., FILHOL, E., MELLGREN, A. E. C., LEH, S., BIZET, A., BRAUN, D. A., GEE, H. Y., SILBERMANN, F., HENRY, C., KRUG, P., BOLE-FEYSOT, C., NITSCHKÉ, P., JOLY, D., NICOUD, P., PAGET, A., HAUGLAND, H., BRACKMANN, D., AHMET, N., SANDFORD, R., CENGIZ, N., KNAPPSKOG, P. M., BOMAN, H., LINGHU, B., YANG, F., OAKELEY, E. J., SAINT MÉZARD, P., SAILER, A. W., JOHANSSON, S., RØDAHL, E., SAUNIER, S., HILDEBRANDT, F. & BENMERAH, A. 2017. Mutations in MAPKBP1 Cause

- Juvenile or Late-Onset Cilia-Independent Nephronophthisis. *Am J Hum Genet*, 100, 323-333.
- MAHJOUB, M. R. 2013. The importance of a single primary cilium. *Organogenesis*, 9, 61-69.
- MAHJOUB, M. R., XIE, Z. & STEARNS, T. 2010. Cep120 is asymmetrically localized to the daughter centriole and is essential for centriole assembly. *J Cell Biol*, 191, 331-46.
- MALICDAN, M. C., VILBOUX, T., STEPHEN, J., MAGLIC, D., MIAN, L., KONZMAN, D., GUO, J., YILDIRIMLI, D., BRYANT, J., FISCHER, R., ZEIN, W. M., SNOW, J., VEMULAPALLI, M., MULLIKIN, J. C., TORO, C., SOLOMON, B. D., NIEDERHUBER, J. E., GAHL, W. A. & GUNAY-AYGUN, M. 2015. Mutations in human homologue of chicken talpid3 gene (KIAA0586) cause a hybrid ciliopathy with overlapping features of Jeune and Joubert syndromes. *J Med Genet*, 52, 830-9.
- MALICKI, J. J. & JOHNSON, C. A. 2017. The Cilium: Cellular Antenna and Central Processing Unit. *Trends Cell Biol*, 27, 126-140.
- MALLAWAARACHCHI, A. C., LUNDIE, B., HORT, Y., SCHONROCK, N., SENUM, S. R., GAYEVSKIY, V., MINOCHE, A. E., HOLLOWAY, G., OHNESORG, T., HINCHCLIFFE, M., PATEL, C., TCHAN, M., MALLETT, A., DINGER, M. E., RANGAN, G., COWLEY, M. J., HARRIS, P. C., BURNETT, L., SHINE, J. & FURLONG, T. J. 2021. Genomic diagnostics in polycystic kidney disease: an assessment of real-world use of whole-genome sequencing. *Eur J Hum Genet*.
- MANOLIO, T. A., ROWLEY, R., WILLIAMS, M. S., RODEN, D., GINSBURG, G. S., BULT, C., CHISHOLM, R. L., DEVERKA, P. A., MCLEOD, H. L., MENSAH, G. A., RELLING, M. V., RODRIGUEZ, L. L., TAMBURRO, C. & GREEN, E. D. 2019. Opportunities, resources, and techniques for implementing genomics in clinical care. *Lancet*, 394, 511-520.
- MANTOVANI, V., BIN, S., GRAZIANO, C., CAPELLI, I., MINARDI, R., AIELLO, V., AMBROSINI, E., CRISTALLI, C. P., MATTIACCIO, A., PARIALI, M., DE FANTI, S., FALETRA, F., GROSSO, E., CANTONE, R., MANCINI, E., MENCARELLI, F., PASINI, A., WISCHMEIJER, A., SCIASCIA, N., SERI, M. & LA MANNA, G. 2020. Gene Panel Analysis in a Large Cohort of Patients With Autosomal Dominant Polycystic Kidney Disease Allows the Identification of 80 Potentially Causative Novel Variants and the Characterization of a Complex Genetic Architecture in a Subset of Families. *Front Genet*, 11, 464.
- MARIA, B. L., HOANG, K. B. N., TUSA, R. J., MANCUSO, A. A., HAMED, L. M., QUISLING, R. G., HOVE, M. T., FENNELL, E. B., BOOTH-JONES, M., RINGDAHL, D. M., YACHNIS, A. T., CREEL, G. & FRERKING, B. 1997. 'Joubert syndrome' revisited: Key ocular motor signs with magnetic resonance imaging correlation. *Journal of Child Neurology*, 12, 423-430.
- MARIA, M., LAMERS, I. J., SCHMIDTS, M., AJMAL, M., JAFFAR, S., ULLAH, E., MUSTAFA, B., AHMAD, S., NAZMUTDINOVA, K., HOSKINS, B., VAN WIJK, E., KOSTER-KAMPHUIS, L., KHAN, M. I., BEALES, P. L., CREMERS, F. P., ROEPMAN, R., AZAM, M., ARTS, H. H. & QAMAR, R. 2016. Genetic and clinical characterization of Pakistani families with Bardet-Biedl syndrome extends the genetic and phenotypic spectrum. *Sci Rep*, 6, 34764.
- MARRA, A. N., LI, Y. & WINGERT, R. A. 2016. Antennas of organ morphogenesis: the roles of cilia in vertebrate kidney development. *Genesis*, 54, 457-469.
- MARTIN, A. R., WILLIAMS, E., FOULGER, R. E., LEIGH, S., DAUGHERTY, L. C., NIBLOCK, O., LEONG, I. U. S., SMITH, K. R., GERASIMENKO, O., HARALDSDOTTIR, E., THOMAS, E., SCOTT, R. H., BAPLE, E., TUCCI, A.,

- BRITTAIN, H., DE BURCA, A., IBAÑEZ, K., KASPERAVICIUTE, D., SMEDLEY, D., CAULFIELD, M., RENDON, A. & MCDONAGH, E. M. 2019. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet*, 51, 1560-1565.
- MARZBAN, H., DEL BIGIO, M. R., ALIZADEH, J., GHAVAMI, S., ZACHARIAH, R. M. & RASTEGAR, M. 2014. Cellular commitment in the developing cerebellum. *Front Cell Neurosci*, 8, 450.
- MCCONNACHIE, D. J., STOW, J. L. & MALLETT, A. J. 2021. Ciliopathies and the Kidney: A Review. *Am J Kidney Dis*, 77, 410-419.
- MEJECASE, C., HUMMEL, A., MOHAND-SAID, S., ANDRIEU, C., EL SHAMIEH, S., ANTONIO, A., CONDROYER, C., BOYARD, F., FOUSSARD, M., BLANCHARD, S., LETEXIER, M., SARAIVA, J. P., SAHEL, J. A., ZEITZ, C. & AUDO, I. 2019. Whole exome sequencing resolves complex phenotype and identifies CC2D2A mutations underlying non-syndromic rod-cone dystrophy. *Clin Genet*, 95, 329-333.
- MEKA, D. P., KOBLER, O., HONG, S., FRIEDRICH, C. M., WUESTHOFF, S., HENIS, M., SCHWANKE, B., KRISP, C., SCHMUELLING, N., RUETER, R., RUECKER, T., BETLEJA, E., CHENG, T., MAHJOUB, M. R., SOBA, P., SCHLÜTER, H., FORNASIERO, E. F. & CALDERON DE ANDA, F. 2022. Centrosome-dependent microtubule modifications set the conditions for axon formation. *Cell Rep*, 39, 110686.
- MIN, H. S., KIM, H. J., NAITO, M., OGURA, S., TOH, K., HAYASHI, K., KIM, B. S., FUKUSHIMA, S., ANRAKU, Y., MIYATA, K. & KATAOKA, K. 2020. Systemic Brain Delivery of Antisense Oligonucleotides across the Blood-Brain Barrier with a Glucose-Coated Polymeric Nanocarrier. *Angew Chem Int Ed Engl*, 59, 8173-8180.
- MIRVIS, M., STEARNS, T. & JAMES NELSON, W. 2018. Cilium structure, assembly, and disassembly regulated by the cytoskeleton. *Biochemical Journal*, 475, 2329-2353.
- MITCHISON, H. M. & SHOEMARK, A. 2017. Motile cilia defects in diseases other than primary ciliary dyskinesia: The contemporary diagnostic and research role for transmission electron microscopy. *Ultrastructural Pathology*, 41, 415-427.
- MITCHISON, H. M. & VALENTE, E. M. 2017. Motile and non-motile cilia in human pathology: from function to phenotypes. *J Pathol*, 241, 294-309.
- MO, Z. & ZECEVIC, N. 2008. Is Pax6 critical for neurogenesis in the human fetal brain? *Cereb Cortex*, 18, 1455-65.
- MOAYYERI, A., HAMMOND, C. J., HART, D. J. & SPECTOR, T. D. 2013. The UK Adult Twin Registry (TwinsUK Resource). *Twin Res Hum Genet*, 16, 144-9.
- MOLINARI, E., DECKER, E., MABILLARD, H., TELLEZ, J., SRIVASTAVA, S., RAMAN, S., WOOD, K., KEMPF, C., ALKANDERI, S., RAMSBOTTOM, S. A., MILES, C. G., JOHNSON, C. A., HILDEBRANDT, F., BERGMANN, C. & SAYER, J. A. 2018a. Human urine-derived renal epithelial cells provide insights into kidney-specific alternate splicing variants. *Eur J Hum Genet*, 26, 1791-1796.
- MOLINARI, E., RAMSBOTTOM, S. A., SAMMUT, V., HUGHES, F. E. P. & SAYER, J. A. 2018b. Using zebrafish to study the function of nephronophthisis and related ciliopathy genes. *F1000Res*, 7, 1133.
- MOLINARI, E., RAMSBOTTOM, S. A., SRIVASTAVA, S., BOOTH, P., ALKANDERI, S., MCLAFFERTY, S. M., DEVLIN, L. A., WHITE, K., GUNAY-AYGUN, M., MILES, C. G. & SAYER, J. A. 2019. Targeted exon skipping rescues ciliary

- protein composition defects in Joubert syndrome patient fibroblasts. *Sci Rep*, 9, 10828.
- MOLINARI, E., SRIVASTAVA, S., DEWHURST, R. M. & SAYER, J. A. 2020. Use of patient derived urine renal epithelial cells to confirm pathogenicity of PKHD1 alleles. *BMC Nephrol*, 21, 435.
- MOLLET, G., SILBERMANN, F., DELOUS, M., SALOMON, R., ANTIGNAC, C. & SAUNIER, S. 2005. Characterization of the nephrocystin/nephrocystin-4 complex and subcellular localization of nephrocystin-4 to primary cilia and centrosomes. *Hum Mol Genet*, 14, 645-56.
- MORAN, D. T., VARELA, F. J. & ROWLEY, J. C., 3RD 1977. Evidence for active role of cilia in sensory transduction. *Proc Natl Acad Sci U S A*, 74, 793-7.
- MORBIDONI, V., AGOLINI, E., SLEP, K. C., PANNONE, L., ZUCCARELLO, D., CASSINA, M., GROSSO, E., GAI, G., SALVIATI, L., DALLAPICCOLA, B., NOVELLI, A., MARTINELLI, S. & TREVISSON, E. 2021. Biallelic mutations in the TOGARAM1 gene cause a novel primary ciliopathy. *J Med Genet*, 58, 526-533.
- MORITZ, C. P., MÜHLHAUS, T., TENZER, S., SCHULENBORG, T. & FRIAUF, E. 2019. Poor transcript-protein correlation in the brain: negatively correlating gene products reveal neuronal polarity as a potential cause. *J Neurochem*, 149, 582-604.
- MOUGOU-ZERELLI, S., THOMAS, S., SZENKER, E., AUDOLLENT, S., ELKHARTOUI, N., BABARIT, C., ROMANO, S., SALOMON, R., AMIEL, J., ESCULPAVIT, C., GONZALES, M., ESCUDIER, E., LEHEUP, B., LOGET, P., ODENT, S., ROUME, J., GERARD, M., DELEZOIDE, A. L., KHUNG, S., PATRIER, S., CORDIER, M. P., BOUVIER, R., MARTINOVIC, J., GUBLER, M. C., BODDAERT, N., MUNNICH, A., ENCHA-RAZAVI, F., VALENTE, E. M., SAAD, A., SAUNIER, S., VEKEMANS, M. & ATTIE-BITACH, T. 2009. CC2D2A mutations in Meckel and Joubert syndromes indicate a genotype-phenotype correlation. *Hum Mutat*, 30, 1574-82.
- MOYER, J. H., LEE-TISCHLER, M. J., KWON, H. Y., SCHRICK, J. J., AVNER, E. D., SWEENEY, W. E., GODFREY, V. L., CACHEIRO, N. L., WILKINSON, J. E. & WOYCHIK, R. P. 1994. Candidate gene associated with a mutation causing recessive polycystic kidney disease in mice. *Science*, 264, 1329-33.
- MUÑOZ-ESTRADA, J. & FERLAND, R. J. 2019. Ahi1 promotes Arl13b ciliary recruitment, regulates Arl13b stability and is required for normal cell migration. *J Cell Sci*, 132.
- MURALI, C., LU, J. T., JAIN, M., LIU, D. S., LACHMAN, R., GIBBS, R. A., LEE, B. H., COHN, D. & CAMPEAU, P. M. 2014. Diagnosis of ALG12-CDG by exome sequencing in a case of severe skeletal dysplasia. *Mol Genet Metab Rep*, 1, 213-219.
- NAKAJIMA, Y., KIYONARI, H., MUKUMOTO, Y. & YOKOYAMA, T. 2018. The Inv compartment of renal cilia is an intraciliary signal-activating center to phosphorylate ANKS6. *Kidney Int*, 93, 1108-1117.
- NANDADASA, S., KRAFT, C. M., WANG, L. W., O'DONNELL, A., PATEL, R., GEE, H. Y., GROBE, K., COX, T. C., HILDEBRANDT, F. & APTE, S. S. 2019. Secreted metalloproteases ADAMTS9 and ADAMTS20 have a non-canonical role in ciliary vesicle growth during ciliogenesis. *Nat Commun*, 10, 953.
- NASSER, F., KURTENBACH, A., BISKUP, S., WEIDENSEE, S., KOHL, S. & ZRENNER, E. 2020. Ophthalmic features of retinitis pigmentosa in Cohen syndrome caused by pathogenic variants in the VPS13B gene. *Acta Ophthalmol*, 98, e316-e321.

- NAWAZ, H., MUJAHID, KHAN, S. A., BIBI, F., WAQAS, A., BARI, A., FARDOUS, KHAN, N., MUHAMMAD, N., KHAN, A., PARACHA, S. A., ALAM, Q., KAMAL, M. A., RAFEEQ, M. M., MUHAMMAD, N., HAQ, F. U., KHAN, S., MAHMOOD, A., KHAN, S. & UMAIR, M. 2023. Biallelic Variants in Seven Different Genes Associated with Clinically Suspected Bardet-Biedl Syndrome. *Genes (Basel)*, 14.
- NGUENGANG WAKAP, S., LAMBERT, D. M., OLRY, A., RODWELL, C., GUEYDAN, C., LANNEAU, V., MURPHY, D., LE CAM, Y. & RATH, A. 2020. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet*, 28, 165-173.
- NI, X., WANG, J., LV, M., LIU, C., ZHONG, Y., TIAN, S., WU, H., CHENG, H., GAO, Y., TAN, Q., CHEN, B., LI, Q., SONG, B., WEI, Z., ZHOU, P., HE, X., ZHANG, F. & CAO, Y. 2020. A novel homozygous mutation in WDR19 induces disorganization of microtubules in sperm flagella and nonsyndromic asthenoteratospermia. *J Assist Reprod Genet*, 37, 1431-1439.
- NIGG, E. A. & HOLLAND, A. J. 2018. Once and only once: Mechanisms of centriole duplication and their deregulation in diseases. *Nature Reviews Molecular Cell Biology*, 19, 297-312.
- NIGG, E. A. & STEARNS, T. 2011. The centrosome cycle: Centriole biogenesis, duplication and inherent asymmetries. *Nature Cell Biology*, 13, 1154-1160.
- NISHIMURA, Y., KASAHARA, K., SHIROMIZU, T., WATANABE, M. & INAGAKI, M. 2019. Primary Cilia as Signaling Hubs in Health and Disease. *Advanced Science*, 6.
- NOOR, A., WINDPASSINGER, C., PATEL, M., STACHOWIAK, B., MIKHAILOV, A., AZAM, M., IRFAN, M., SIDDIQUI, Z. K., NAEEM, F., PATERSON, A. D., LUTFULLAH, M., VINCENT, J. B. & AYUB, M. 2008. CC2D2A, encoding a coiled-coil and C2 domain protein, causes autosomal-recessive mental retardation with retinitis pigmentosa. *Am J Hum Genet*, 82, 1011-8.
- NOVARINO, G., AKIZU, N. & GLEESON, J. G. 2011. Modeling human disease in humans: the ciliopathies. *Cell*, 147, 70-9.
- NURK, S., KOREN, S., RHIE, A., RAUTIAINEN, M., BZIKADZE, A. V., MIKHEENKO, A., VOLLMER, M. R., ALTEMOSE, N., URALSKY, L., GERSHMAN, A., AGANEZOV, S., HOYT, S. J., DIEKHANS, M., LOGSDON, G. A., ALONGE, M., ANTONARAKIS, S. E., BORCHERS, M., BOUFFARD, G. G., BROOKS, S. Y., CALDAS, G. V., CHEN, N. C., CHENG, H., CHIN, C. S., CHOW, W., DE LIMA, L. G., DISHUCK, P. C., DURBIN, R., DVORKINA, T., FIDDES, I. T., FORMENTI, G., FULTON, R. S., FUNGTAMMASAN, A., GARRISON, E., GRADY, P. G. S., GRAVES-LINDSAY, T. A., HALL, I. M., HANSEN, N. F., HARTLEY, G. A., HAUKNES, M., HOWE, K., HUNKAPILLER, M. W., JAIN, C., JAIN, M., JARVIS, E. D., KERPEDJIEV, P., KIRSCH, M., KOLMOGOROV, M., KORLACH, J., KREMITZKI, M., LI, H., MADURO, V. V., MARSCHALL, T., MCCARTNEY, A. M., MCDANIEL, J., MILLER, D. E., MULLIKIN, J. C., MYERS, E. W., OLSON, N. D., PATEL, B., PELUSO, P., PEVZNER, P. A., PORUBSKY, D., POTAPOVA, T., ROGAEV, E. I., ROSENFELD, J. A., SALZBERG, S. L., SCHNEIDER, V. A., SEDLAZECK, F. J., SHAFIN, K., SHEW, C. J., SHUMATE, A., SIMS, Y., SMIT, A. F. A., SOTO, D. C., SOVIĆ, I., STORER, J. M., STREETS, A., SULLIVAN, B. A., THIBAUD-NISSEN, F., TORRANCE, J., WAGNER, J., WALENZ, B. P., WENGER, A., WOOD, J. M. D., XIAO, C., YAN, S. M., YOUNG, A. C., ZARATE, S., SURTI, U., MCCOY, R. C., DENNIS, M. Y., ALEXANDROV, I. A., GERTON, J. L., O'NEILL, R. J., TIMP, W., ZOOK, J. M., SCHATZ, M. C., EICHLER, E. E.,

- MIGA, K. H. & PHILLIPPY, A. M. 2022. The complete sequence of a human genome. *Science*, 376, 44-53.
- O'TOOLE, J. F., LIU, Y., DAVIS, E. E., WESTLAKE, C. J., ATTANASIO, M., OTTO, E. A., SEELOW, D., NURNBERG, G., BECKER, C., NUUTINEN, M., KÄRPPÄ, M., IGNATIUS, J., UUSIMAA, J., PAKANEN, S., JAAKKOLA, E., VAN DEN HEUVEL, L. P., FEHRENBACH, H., WIGGINS, R., GOYAL, M., ZHOU, W., WOLF, M. T., WISE, E., HELOU, J., ALLEN, S. J., MURGA-ZAMALLOA, C. A., ASHRAF, S., CHAKI, M., HEERINGA, S., CHERNIN, G., HOSKINS, B. E., CHAIB, H., GLEESON, J., KUSAKABE, T., SUZUKI, T., ISAAC, R. E., QUARMBY, L. M., TENNANT, B., FUJIOKA, H., TUOMINEN, H., HASSINEN, I., LOHI, H., VAN HOUTEN, J. L., ROTIG, A., SAYER, J. A., ROLINSKI, B., FREISINGER, P., MADHAVAN, S. M., HERZER, M., MADIGNIER, F., PROKISCH, H., NURNBERG, P., JACKSON, P. K., KHANNA, H., KATSANIS, N. & HILDEBRANDT, F. 2010. Individuals with mutations in XPNPEP3, which encodes a mitochondrial protein, develop a nephronophthisis-like nephropathy. *J Clin Invest*, 120, 791-802.
- OGDEN, S. K., CASSO, D. J., ASCANO, M., JR., YORE, M. M., KORNBERG, T. B. & ROBBINS, D. J. 2006. Smoothened regulates activator and repressor functions of Hedgehog signaling via two distinct mechanisms. *J Biol Chem*, 281, 7237-43.
- OJEDA NAHARROS, I., GESEMANN, M., MATEOS, J. M., BARMETTLER, G., FORBES, A., ZIEGLER, U., NEUHAUSS, S. C. F. & BACHMANN-GAGESCU, R. 2017. Loss-of-function of the ciliopathy protein Cc2d2a disorganizes the vesicle fusion machinery at the periciliary membrane and indirectly affects Rab8-trafficking in zebrafish photoreceptors. *PLoS Genet*, 13, e1007150.
- OKA, M., SHIMOJIMA, K., YAMAMOTO, T., HANAOKA, Y., SATO, S., YASUHARA, T., YOSHINAGA, H. & KOBAYASHI, K. 2016. A novel HYLS1 homozygous mutation in living siblings with Joubert syndrome. *Clin Genet*, 89, 739-43.
- OLINGER, E., ALAWI, I. A., AL RIYAMI, M. S., SALMI, I. A., MOLINARI, E., FAQEIH, E. A., AL-HAMED, M. H., BARROSO-GIL, M., POWELL, L., AL-HUSSAINI, A. A., RAHIM, K. A., ALMONTASHIRI, N. A. M., MILES, C., SHRIL, S., HILDEBRANDT, F., CONSORTIUM, G. E. R., WILSON, I. J. & SAYER, J. A. 2021. A discarded synonymous variant in NPHP3 explains nephronophthisis and congenital hepatic fibrosis in several families. *Hum Mutat*, 42, 1221-1228.
- OLINGER, E., HOFMANN, P., KIDD, K., DUFOUR, I., BELGE, H., SCHAEFFER, C., KIPP, A., BONNY, O., DELTAS, C., DEMOULIN, N., FEHR, T., FUSTER, D. G., GALE, D. P., GOFFIN, E., HODAŇOVÁ, K., HUYNH-DO, U., KISTLER, A., MORELLE, J., PAPAGREGORIOU, G., PIRSON, Y., SANDFORD, R., SAYER, J. A., TORRA, R., VENZIN, C., VENZIN, R., VOGT, B., ŽIVNÁ, M., GREKA, A., DAHAN, K., RAMPOLDI, L., KMOCH, S., BLEYER, A. J., SR. & DEVUYST, O. 2020. Clinical and genetic spectra of autosomal dominant tubulointerstitial kidney disease due to mutations in UMOD and MUC1. *Kidney Int*, 98, 717-731.
- OMRAN, A. J. A., SATERNOS, H. C., ALTHOBAITI, Y. S., WISNER, A., SARI, Y., NAULI, S. M. & ABOUALAIWI, W. A. 2017. Alcohol consumption impairs the ependymal cilia motility in the brain ventricles. *Scientific Reports*, 7, 1-8.
- OMRAN, H. 2010. NPHP proteins: gatekeepers of the ciliary compartment. *J Cell Biol*, 190, 715-7.
- ORR, S., OLINGER, E., IOSIFIDOU, S., BARROSO-GIL, M., NEATU, R., WOOD, K., WILSON, I. & SAYER, J. A. 2023. Molecular genetic diagnosis of kidney ciliopathies: Lessons from interpreting genomic sequencing data and the requirement for accurate phenotypic data. *Ann Hum Genet*.

- OTTO, E., HOEFELE, J., RUF, R., MUELLER, A. M., HILLER, K. S., WOLF, M. T., SCHUERMANN, M. J., BECKER, A., BIRKENHÄGER, R., SUDBRAK, R., HENNIES, H. C., NÜRNBERG, P. & HILDEBRANDT, F. 2002. A gene mutated in nephronophthisis and retinitis pigmentosa encodes a novel protein, nephroretinin, conserved in evolution. *Am J Hum Genet*, 71, 1161-7.
- OTTO, E. A., HURD, T. W., AIRIK, R., CHAKI, M., ZHOU, W., STOETZEL, C., PATIL, S. B., LEVY, S., GHOSH, A. K., MURGA-ZAMALLOA, C. A., VAN REEUWIJK, J., LETTEBOER, S. J., SANG, L., GILES, R. H., LIU, Q., COENE, K. L., ESTRADA-CUZCANO, A., COLLIN, R. W., MCLAUGHLIN, H. M., HELD, S., KASANUKI, J. M., RAMASWAMI, G., CONTE, J., LOPEZ, I., WASHBURN, J., MACDONALD, J., HU, J., YAMASHITA, Y., MAHER, E. R., GUAY-WOODFORD, L. M., NEUMANN, H. P., OBERMÜLLER, N., KOENEKOOP, R. K., BERGMANN, C., BEI, X., LEWIS, R. A., KATSANIS, N., LOPES, V., WILLIAMS, D. S., LYONS, R. H., DANG, C. V., BRITO, D. A., DIAS, M. B., ZHANG, X., CAVALCOLI, J. D., NÜRNBERG, G., NÜRNBERG, P., PIERCE, E. A., JACKSON, P. K., ANTIGNAC, C., SAUNIER, S., ROEPMAN, R., DOLLFUS, H., KHANNA, H. & HILDEBRANDT, F. 2010. Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nat Genet*, 42, 840-50.
- OTTO, E. A., LOEYS, B., KHANNA, H., HELLEMANS, J., SUDBRAK, R., FAN, S., MUERB, U., O'TOOLE, J. F., HELOU, J., ATTANASIO, M., UTSCH, B., SAYER, J. A., LILLO, C., JIMENO, D., COUCKE, P., DE PAEPE, A., REINHARDT, R., KLAGES, S., TSUDA, M., KAWAKAMI, I., KUSAKABE, T., OMRAN, H., IMM, A., TIPPENS, M., RAYMOND, P. A., HILL, J., BEALES, P., HE, S., KISPERT, A., MARGOLIS, B., WILLIAMS, D. S., SWAROOP, A. & HILDEBRANDT, F. 2005. Nephrocystin-5, a ciliary IQ domain protein, is mutated in Senior-Loken syndrome and interacts with RPGR and calmodulin. *Nat Genet*, 37, 282-8.
- OTTO, E. A., RAMASWAMI, G., JANSSEN, S., CHAKI, M., ALLEN, S. J., ZHOU, W., AIRIK, R., HURD, T. W., GHOSH, A. K., WOLF, M. T., HOPPE, B., NEUHAUS, T. J., BOCKENHAUER, D., MILFORD, D. V., SOLIMAN, N. A., ANTIGNAC, C., SAUNIER, S., JOHNSON, C. A. & HILDEBRANDT, F. 2011. Mutation analysis of 18 nephronophthisis associated ciliopathy disease genes using a DNA pooling and next generation sequencing strategy. *J Med Genet*, 48, 105-16.
- OTTO, E. A., TORY, K., ATTANASIO, M., ZHOU, W., CHAKI, M., PARUCHURI, Y., WISE, E. L., WOLF, M. T., UTSCH, B., BECKER, C., NÜRNBERG, G., NÜRNBERG, P., NAYIR, A., SAUNIER, S., ANTIGNAC, C. & HILDEBRANDT, F. 2009. Hypomorphic mutations in meckelin (MKS3/TMEM67) cause nephronophthisis with liver fibrosis (NPHP11). *J Med Genet*, 46, 663-70.
- OUD, M., LAMERS, I. & ARTS, H. 2016. Ciliopathies: Genetics in Pediatric Medicine. *Journal of Pediatric Genetics*, 06, 018-029.
- OUD, M. M., LATOUR, B. L., BAKEY, Z., LETTEBOER, S. J., LUGTENBERG, D., WU, K. M., CORNELISSEN, E. A. M., YNTEMA, H. G., SCHMIDTS, M., ROEPMAN, R. & BONGERS, E. 2018. Cellular ciliary phenotyping indicates pathogenicity of novel variants in IFT140 and confirms a Mainzer-Saldino syndrome diagnosis. *Cilia*, 7, 1.
- PAIGE TAYLOR, S., KUNOVA BOSAKOVA, M., VARECHA, M., BALEK, L., BARTA, T., TRANTIREK, L., JELINKOVA, I., DURAN, I., VESELA, I., FORLENZA, K. N., MARTIN, J. H., HAMPL, A., BAMSHAD, M., NICKERSON, D., JAWORSKI, M. L., SONG, J., KO, H. W., COHN, D. H., KRAKOW, D. & KREJCI, P. 2016. An inactivating mutation in intestinal cell kinase, ICK, impairs hedgehog

- signalling and causes short rib-polydactyly syndrome. *Hum Mol Genet*, 25, 3998-4011.
- PALA, R., ALOMARI, N. & NAULI, S. M. 2017. Primary cilium-dependent signaling mechanisms. *International Journal of Molecular Sciences*, 18.
- PAPROCKA, J. & JAMROZ, E. 2012. [Joubert syndrome and related disorders]. *Neurol Neurochir Pol*, 46, 379-83.
- PARISI, M. A. 2009. Clinical and molecular features of Joubert syndrome and related disorders. *Am J Med Genet C Semin Med Genet*, 151c, 326-40.
- PARISI, M. A. 2019. The molecular genetics of Joubert syndrome and related ciliopathies: The challenges of genetic and phenotypic heterogeneity. *Transl Sci Rare Dis*, 4, 25-49.
- PARISI, M. A., DOHERTY, D., CHANCE, P. F. & GLASS, I. A. 2007. Joubert syndrome (and related disorders) (OMIM 213300). *Eur J Hum Genet*, 15, 511-21.
- PAULI, S., ALTMULLER, J., SCHRODER, S., OHLENBUSCH, A., DREHA-KULACZEWSKI, S., BERGMANN, C., NURNBERG, P., THIELE, H., LI, Y., WOLNIK, B. & BROCKMANN, K. 2019. Homozygosity for the c.428delG variant in KIAA0586 in a healthy individual: implications for molecular testing in patients with Joubert syndrome. *J Med Genet*, 56, 261-264.
- PAVATHUPARAMBIL ABDUL MANAPH, N., AL-HAWWAS, M., BOBROVSKAYA, L., COATES, P. T. & ZHOU, X. F. 2018. Urine-derived cells for human cell therapy. *Stem Cell Res Ther*, 9, 189.
- PEREA-ROMERO, I., GORDO, G., IANCU, I. F., DEL POZO-VALERO, M., ALMOGUERA, B., BLANCO-KELLY, F., CARREÑO, E., JIMENEZ-ROLANDO, B., LOPEZ-RODRIGUEZ, R., LORDA-SANCHEZ, I., MARTIN-MERIDA, I., PÉREZ DE AYALA, L., RIVEIRO-ALVAREZ, R., RODRIGUEZ-PINILLA, E., TAHSIN-SWAFIRI, S., TRUJILLO-TIEBAS, M. J., GARCIA-SANDOVAL, B., MINGUEZ, P., AVILA-FERNANDEZ, A., CORTON, M. & AYUSO, C. 2021. Genetic landscape of 6089 inherited retinal dystrophies affected cases in Spain and their therapeutic and extended epidemiological implications. *Sci Rep*, 11, 1526.
- PERUGORRIA, M. J. & BANALES, J. M. 2017. Genetics: Novel causative genes for polycystic liver disease. *Nat Rev Gastroenterol Hepatol*, 14, 391-392.
- PHILLIPS, K. A., DOUGLAS, M. P., WORDSWORTH, S., BUCHANAN, J. & MARSHALL, D. A. 2021. Availability and funding of clinical genomic sequencing globally. *BMJ Glob Health*, 6.
- PIASECKI, B. P. & SILFLOW, C. D. 2009. The UNI1 and UNI2 genes function in the transition of triplet to doublet microtubules between the centriole and cilium in *Chlamydomonas*. *Mol Biol Cell*, 20, 368-78.
- PISANI, I., ALLINOVYI, M., PALAZZO, V., ZANELLI, P., GENTILE, M., FARINA, M. T., GIULIOTTI, S., CRAVEDI, P., DELSANTE, M., MAGGIORE, U., FIACCADORI, E. & MANENTI, L. 2022. More dissimilarities than affinities between DNAJB11-PKD and ADPKD. *Clin Kidney J*, 15, 1179-1187.
- PIZZUTI, V., DONADEI, C., BALDUCCELLI, E., CONTE, D., GESSAROLI, E., PARIS, F., BINI, C., DEMETRI, M., DI NUNZIO, M., CORRADETTI, V., ALVIANO, F., LA MANNA, G. & COMAI, G. 2023. Urine-Derived Renal Epithelial Cells (URECs) from Transplanted Kidneys as a Promising Immunomodulatory Cell Population. *Cells*, 12.
- PORATH, B., GAINULLIN, V. G., CORNEC-LE GALL, E., DILLINGER, E. K., HEYER, C. M., HOPP, K., EDWARDS, M. E., MADSEN, C. D., MAURITZ, S. R., BANKS, C. J., BAHETI, S., REDDY, B., HERRERO, J. I., BAÑALES, J. M., HOGAN, M. C., TASIC, V., WATNICK, T. J., CHAPMAN, A. B., VIGNEAU, C.,

- LAVAINNE, F., AUDRÉZET, M. P., FEREC, C., LE MEUR, Y., TORRES, V. E. & HARRIS, P. C. 2016. Mutations in GANAB, Encoding the Glucosidase II α Subunit, Cause Autosomal-Dominant Polycystic Kidney and Liver Disease. *Am J Hum Genet*, 98, 1193-1207.
- POWELL, L., BARROSO-GIL, M., CLOWRY, G. J., DEVLIN, L. A., MOLINARI, E., RAMSBOTTOM, S. A., MILES, C. G. & SAYER, J. A. 2020. Expression patterns of ciliopathy genes ARL3 and CEP120 reveal roles in multisystem development. *BMC Dev Biol*, 20, 26.
- POWELL, L., OLINGER, E., WEDDERBURN, S., RAMAKUMARAN, V. S., KINI, U., CLAYTON-SMITH, J., RAMSDEN, S. C., RICE, S. J., BARROSO-GIL, M., WILSON, I., COWLEY, L., JOHNSON, S., HARRIS, E., MONTGOMERY, T., BERTOLI, M., BOLTSHAUSER, E. & SAYER, J. A. 2021. Identification of LAMA1 mutations ends diagnostic odyssey and has prognostic implications for patients with presumed Joubert syndrome. *Brain Commun*, 3, fcab163.
- PRENDERGAST, S. C., STROBL, A. C., CROSS, W., PILLAY, N., STRAUSS, S. J., YE, H., LINDSAY, D., TIRABOSCO, R., CHALKER, J., MAHAMDALLIE, S. S., SOSINSKY, A., FLANAGAN, A. M. & AMARY, F. 2020. Sarcoma and the 100,000 Genomes Project: our experience and changes to practice. *J Pathol Clin Res*, 6, 297-307.
- PRUSKI, M., HU, L., YANG, C., WANG, Y., ZHANG, J. B., ZHANG, L., HUANG, Y., RAJNICEK, A. M., ST CLAIR, D., MCCAIG, C. D., LANG, B. & DING, Y. Q. 2019. Roles for IFT172 and Primary Cilia in Cell Migration, Cell Division, and Neocortex Development. *Front Cell Dev Biol*, 7, 287.
- PUSAPATI, G. V., KONG, J. H., PATEL, B. B., KRISHNAN, A., SAGNER, A., KINNEBREW, M., BRISCOE, J., ARAVIND, L. & ROHATGI, R. 2018. CRISPR Screens Uncover Genes that Regulate Target Cell Sensitivity to the Morphogen Sonic Hedgehog. *Dev Cell*, 44, 113-129.e8.
- RADADIYA, P. S., THORNTON, M. M., DANIEL, E. A., IDOWU, J. Y., WANG, W., MAGENHEIMER, B., SUBRAMANIAM, D., TRAN, P. V., CALVET, J. P., WALLACE, D. P. & SHARMA, M. 2021. Quinomycin A reduces cyst progression in polycystic kidney disease. *Faseb j*, 35, e21533.
- RADHA RAMA DEVI, A., NAUSHAD, S. M. & LINGAPPA, L. 2020. Clinical and Molecular Diagnosis of Joubert Syndrome and Related Disorders. *Pediatr Neurol*, 106, 43-49.
- RAHIMI-BALAEI, M., BERGEN, H., KONG, J. & MARZBAN, H. 2018. Neuronal Migration During Development of the Cerebellum. *Front Cell Neurosci*, 12, 484.
- RAMIREZ, A. H., SULIEMAN, L., SCHLUETER, D. J., HALVORSON, A., QIAN, J., RATSIMBAZAFY, F., LOPERENA, R., MAYO, K., BASFORD, M., DEFLAUX, N., MUTHURAMAN, K. N., NATARAJAN, K., KHO, A., XU, H., WILKINS, C., ANTON-CULVER, H., BOERWINKLE, E., CICEK, M., CLARK, C. R., COHN, E., OHNO-MACHADO, L., SCHULLY, S. D., AHMEDANI, B. K., ARGOS, M., CRONIN, R. M., O'DONNELL, C., FOUAD, M., GOLDSTEIN, D. B., GREENLAND, P., HEBBRING, S. J., KARLSON, E. W., KHATRI, P., KORF, B., SMOLLER, J. W., SODEKE, S., WILBANKS, J., HENTGES, J., MOCKRIN, S., LUNT, C., DEVANEY, S. A., GEBO, K., DENNY, J. C., CARROLL, R. J., GLAZER, D., HARRIS, P. A., HRIPCSAK, G., PHILIPPAKIS, A. & RODEN, D. M. 2022. The All of Us Research Program: Data quality, utility, and diversity. *Patterns (N Y)*, 3, 100570.
- RAMSBOTTOM, S. A., MOLINARI, E., SRIVASTAVA, S., SILBERMAN, F., HENRY, C., ALKANDERI, S., DEVLIN, L. A., WHITE, K., STEEL, D. H., SAUNIER, S., MILES, C. G. & SAYER, J. A. 2018. Targeted exon skipping of a CEP290

- mutation rescues Joubert syndrome phenotypes in vitro and in a murine model. *Proc Natl Acad Sci U S A*, 115, 12489-12494.
- RAMSBOTTOM, S. A., THELWALL, P. E., WOOD, K. M., CLOWRY, G. J., DEVLIN, L. A., SILBERMANN, F., SPIEWAK, H. L., SHRIL, S., MOLINARI, E., HILDEBRANDT, F., GUNAY-AYGUN, M., SAUNIER, S., CORDELL, H. J., SAYER, J. A. & MILES, C. G. 2020. Mouse genetics reveals Barttin as a genetic modifier of Joubert syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 1113-1118.
- RATNAPRIYA, R., JACOBSON, S. G., CIDEKIYAN, A. V., ENGLISH, M. A., ROMAN, A. J., SUMAROKA, A., SHEPLOCK, R. & SWAROOP, A. 2021. A Novel ARL3 Gene Mutation Associated With Autosomal Dominant Retinal Degeneration. *Front Cell Dev Biol*, 9, 720782.
- REITER, J. F. & LEROUX, M. R. 2017. Genes and molecular pathways underpinning ciliopathies. *Nat Rev Mol Cell Biol*, 18, 533-547.
- REITER, J. F. & SINGLA, V. 2006. The Primary Cilium as the Cell's Antenna: Signaling at a Sensory Organelle. *Science*, 313, 629-633.
- RENTZSCH, P., WITTEN, D., COOPER, G. M., SHENDURE, J. & KIRCHER, M. 2019. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*, 47, D886-d894.
- REYNOLDS, K., KUMARI, P., SEPULVEDA RINCON, L., GU, R., JI, Y., KUMAR, S. & ZHOU, C. J. 2019. Wnt signaling in orofacial clefts: crosstalk, pathogenesis and models. *Disease Models & Mechanisms*, 12, dmm037051-dmm037051.
- RICHARDS, S., AZIZ, N., BALE, S., BICK, D., DAS, S., GASTIER-FOSTER, J., GRODY, W. W., HEGDE, M., LYON, E., SPECTOR, E., VOELKERDING, K. & REHM, H. L. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, 17, 405-24.
- ROBINSON, J. T., THORVALDSDÓTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nat Biotechnol*, 29, 24-6.
- ROBINSON, P. N., KOHLER, S., OELLRICH, A., WANG, K., MUNGALL, C. J., LEWIS, S. E., WASHINGTON, N., BAUER, S., SEELOW, D., KRAWITZ, P., GILISSEN, C., HAENDEL, M. & SMEDLEY, D. 2014. Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res*, 24, 340-8.
- ROMANI, M., MICALIZZI, A. & VALENTE, E. M. 2013. Joubert syndrome: congenital cerebellar ataxia with the molar tooth. *Lancet Neurol*, 12, 894-905.
- ROOSING, S., ROMANI, M., ISRIE, M., ROSTI, R. O., MICALIZZI, A., MUSAEV, D., MAZZA, T., AL-GAZALI, L., ALTUNOGLU, U., BOLTSHAUSER, E., D'ARRIGO, S., DE KEERSMAECKER, B., KAYSERILI, H., BRANDENBERGER, S., KRAOUA, I., MARK, P. R., MCKANNA, T., VAN KEIRSBILCK, J., MOERMAN, P., PORETTI, A., PURI, R., VAN ESCH, H., GLEESON, J. G. & VALENTE, E. M. 2016. Mutations in CEP120 cause Joubert syndrome as well as complex ciliopathy phenotypes. *J Med Genet*, 53, 608-15.
- ROZET, J. M. & GERARD, X. 2015. Understanding disease pleiotropy: From puzzle to solution. *Sci Transl Med*, 7, 291fs24.
- RUAN, G. X., BARRY, E., YU, D., LUKASON, M., CHENG, S. H. & SCARIA, A. 2017. CRISPR/Cas9-Mediated Genome Editing as a Therapeutic Approach for Leber Congenital Amaurosis 10. *Mol Ther*, 25, 331-341.

- RUBIN, J. D. & BARRY, M. A. 2020. Improving Molecular Therapy in the Kidney. *Mol Diagn Ther*, 24, 375-396.
- RUSTERHOLZ, T. D. S., HOFMANN, C. & BACHMANN-GAGESCU, R. 2022. Insights Gained From Zebrafish Models for the Ciliopathy Joubert Syndrome. *Front Genet*, 13, 939527.
- SAKAKIBARA, N., NOZU, K., YAMAMURA, T., HORINOUCHI, T., NAGANO, C., YE, M. J., ISHIKO, S., AOTO, Y., ROSSANTI, R., HAMADA, R., OKAMOTO, N., SHIMA, Y., NAKANISHI, K., MATSUO, M., IIJIMA, K. & MORISADA, N. 2022. Comprehensive genetic analysis using next-generation sequencing for the diagnosis of nephronophthisis-related ciliopathies in the Japanese population. *J Hum Genet*, 67, 427-440.
- SANG, L., MILLER, J. J., CORBIT, K. C., GILES, R. H., BRAUER, M. J., OTTO, E. A., BAYE, L. M., WEN, X., SCALES, S. J., KWONG, M., HUNTZICKER, E. G., SFAKIANOS, M. K., SANDOVAL, W., BAZAN, J. F., KULKARNI, P., GARCIA-GONZALO, F. R., SEOL, A. D., O'TOOLE, J. F., HELD, S., REUTTER, H. M., LANE, W. S., RAFIQ, M. A., NOOR, A., ANSAR, M., DEVI, A. R., SHEFFIELD, V. C., SLUSARSKI, D. C., VINCENT, J. B., DOHERTY, D. A., HILDEBRANDT, F., REITER, J. F. & JACKSON, P. K. 2011. Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways. *Cell*, 145, 513-28.
- SATIR, P. 1995. Landmarks in Cilia Research From Leeuwenhoek to Us. *Cell Motility and the Cytoskeleton*.
- SATIR, P. 2017. CILIA: Before and after. *Cilia*. BioMed Central Ltd.
- SAUNDERS, G., BAUDIS, M., BECKER, R., BELTRAN, S., BÉROUD, C., BIRNEY, E., BROOKSBANK, C., BRUNAK, S., VAN DEN BULCKE, M., DRYSDALE, R., CAPELLA-GUTIERREZ, S., FLICEK, P., FLORINDI, F., GOODHAND, P., GUT, I., HERINGA, J., HOLUB, P., HOOYBERGHS, J., JUTY, N., KEANE, T. M., KORBEL, J. O., LAPPALAINEN, I., LESKOSEK, B., MATTHIJS, G., MAYRHOFER, M. T., METSPALU, A., NAVARRO, A., NEWHOUSE, S., NYRÖNEN, T., PAGE, A., PERSSON, B., PALOTIE, A., PARKINSON, H., RAMBLA, J., SALGADO, D., STEINFELDER, E., SWERTZ, M. A., VALENCIA, A., VARMA, S., BLOMBERG, N. & SCOLLEN, S. 2019. Leveraging European infrastructures to access 1 million human genomes by 2022. *Nat Rev Genet*, 20, 693-701.
- SAYER, J. A., OTTO, E. A., O'TOOLE, J. F., NURNBERG, G., KENNEDY, M. A., BECKER, C., HENNIES, H. C., HELOU, J., ATTANASIO, M., FAUSETT, B. V., UTSCH, B., KHANNA, H., LIU, Y., DRUMMOND, I., KAWAKAMI, I., KUSAKABE, T., TSUDA, M., MA, L., LEE, H., LARSON, R. G., ALLEN, S. J., WILKINSON, C. J., NIGG, E. A., SHOU, C., LILLO, C., WILLIAMS, D. S., HOPPE, B., KEMPER, M. J., NEUHAUS, T., PARISI, M. A., GLASS, I. A., PETRY, M., KISPERT, A., GLOY, J., GANNER, A., WALZ, G., ZHU, X., GOLDMAN, D., NURNBERG, P., SWAROOP, A., LEROUX, M. R. & HILDEBRANDT, F. 2006. The centrosomal protein nephrocystin-6 is mutated in Joubert syndrome and activates transcription factor ATF4. *Nat Genet*, 38, 674-81.
- SCHELLENS, R. T. W., BROEKMAN, S., PETERS, T., GRAAVE, P., MALINAR, L., VENSELAAR, H., KREMER, H., DE VRIEZE, E. & VAN WIJK, E. 2023. A protein domain-oriented approach to expand the opportunities of therapeutic exon skipping for USH2A-associated retinitis pigmentosa. *Mol Ther Nucleic Acids*, 32, 980-994.

- SCHMIDT, K. N., KUHNS, S., NEUNER, A., HUB, B., ZENTGRAF, H. & PEREIRA, G. 2012. Cep164 mediates vesicular docking to the mother centriole during early steps of ciliogenesis. *J Cell Biol*, 199, 1083-101.
- SCHMIDTS, M. 2014. Clinical genetics and pathobiology of ciliary chondrodysplasias. *J Pediatr Genet*, 3, 46-94.
- SCHMIDTS, M., FRANK, V., EISENBERGER, T., AL TURKI, S., BIZET, A. A., ANTONY, D., RIX, S., DECKER, C., BACHMANN, N., BALD, M., VINKE, T., TOENSHOFF, B., DI DONATO, N., NEUHANN, T., HARTLEY, J. L., MAHER, E. R., BOGDANOVIĆ, R., PECO-ANTIĆ, A., MACHE, C., HURLES, M. E., JOKSIĆ, I., GUĆ-ŠĆEKIĆ, M., DOBRICIC, J., BRANKOVIC-MAGIC, M., BOLZ, H. J., PAZOUR, G. J., BEALES, P. L., SCAMBLER, P. J., SAUNIER, S., MITCHISON, H. M. & BERGMANN, C. 2013. Combined NGS approaches identify mutations in the intraflagellar transport gene IFT140 in skeletal ciliopathies with early progressive kidney Disease. *Hum Mutat*, 34, 714-24.
- SCHOBERER, J., SHIN, Y. J., VAVRA, U., VEIT, C. & STRASSER, R. 2018. Analysis of Protein Glycosylation in the ER. *Methods Mol Biol*, 1691, 205-222.
- SCHRICK, J. J., VOGEL, P., ABUIN, A., HAMPTON, B. & RICE, D. S. 2006. ADP-ribosylation factor-like 3 is involved in kidney and photoreceptor development. *Am J Pathol*, 168, 1288-98.
- SCHUELER, M., BRAUN, D. A., CHANDRASEKAR, G., GEE, H. Y., KLASSON, T. D., HALBRITTER, J., BIEDER, A., PORATH, J. D., AIRIK, R., ZHOU, W., LOTURCO, J. J., CHE, A., OTTO, E. A., BÖCKENHAUER, D., SEBIRE, N. J., HONZIK, T., HARRIS, P. C., KOON, S. J., GUNAY-AYGUN, M., SAUNIER, S., ZERRES, K., BRUECHLE, N. O., DRENTH, J. P., PELLETIER, L., TAPIA-PÁEZ, I., LIFTON, R. P., GILES, R. H., KERE, J. & HILDEBRANDT, F. 2015. DCDC2 mutations cause a renal-hepatic ciliopathy by disrupting Wnt signaling. *Am J Hum Genet*, 96, 81-92.
- SCHUELER, M., HALBRITTER, J., PHELPS, I. G., BRAUN, D. A., OTTO, E. A., PORATH, J. D., GEE, H. Y., SHENDURE, J., O'ROAK, B. J., LAWSON, J. A., NABHAN, M. M., SOLIMAN, N. A., DOHERTY, D. & HILDEBRANDT, F. 2016. Large-scale targeted sequencing comparison highlights extreme genetic heterogeneity in nephronophthisis-related ciliopathies. *J Med Genet*, 53, 208-14.
- SCHWARZ, D. S. & BLOWER, M. D. 2016. The endoplasmic reticulum: structure, function and response to cellular signaling. *Cell Mol Life Sci*, 73, 79-94.
- SCHWARZ, M., THIEL, C., LÜBBEHUSEN, J., DORLAND, B., DE KONING, T., VON FIGURA, K., LEHLE, L. & KÖRNER, C. 2004. Deficiency of GDP-Man:GlcNAc2-PP-dolichol mannosyltransferase causes congenital disorder of glycosylation type I κ . *Am J Hum Genet*, 74, 472-81.
- SCHWARZ, N., LANE, A., JOVANOVIC, K., PARFITT, D. A., AGUILA, M., THOMPSON, C. L., DA CRUZ, L., COFFEY, P. J., CHAPPLE, J. P., HARDCASTLE, A. J. & CHEETHAM, M. E. 2017. Arl3 and RP2 regulate the trafficking of ciliary tip kinesins. *Hum Mol Genet*, 26, 2480-2492.
- SENUM, S. R., LI, Y. S. M., BENSON, K. A., JOLI, G., OLINGER, E., LAVU, S., MADSEN, C. D., GREGORY, A. V., NEATU, R., KLINE, T. L., AUDRÉZET, M. P., OUTEDA, P., NAU, C. B., MEIJER, E., ALI, H., STEINMAN, T. I., MRUG, M., PHELAN, P. J., WATNICK, T. J., PETERS, D. J. M., ONG, A. C. M., CONLON, P. J., PERRONE, R. D., CORNEC-LE GALL, E., HOGAN, M. C., TORRES, V. E., SAYER, J. A. & HARRIS, P. C. 2022. Monoallelic IFT140 pathogenic variants are an important cause of the autosomal dominant polycystic kidney-spectrum phenotype. *Am J Hum Genet*, 109, 136-156.

- SERVAIS, L., MONTUS, M., GUINER, C. L., BEN YAOU, R., ANNOUSSAMY, M., MORAUX, A., HOGREL, J. Y., SEFERIAN, A. M., ZEHROUNI, K., LE MOING, A. G., GIDARO, T., VANHULLE, C., LAUGEL, V., BUTOIANU, N., CUISSET, J. M., SABOURAUD, P., CANCES, C., KLEIN, A., LETURCQ, F., MOULLIER, P. & VOIT, T. 2015. Non-Ambulant Duchenne Patients Theoretically Treatable by Exon 53 Skipping have Severe Phenotype. *J Neuromuscul Dis*, 2, 269-279.
- SHAHEEN, R., SCHMIDTS, M., FAQEIH, E., HASHEM, A., LAUSCH, E., HOLDER, I., SUPERTI-FURGA, A., CONSORTIUM, U. K., MITCHISON, H. M., ALMOISHEER, A., ALAMRO, R., ALSHIDDI, T., ALZAHHRANI, F., BEALES, P. L. & ALKURAYA, F. S. 2015. A founder CEP120 mutation in Jeune asphyxiating thoracic dystrophy expands the role of centriolar proteins in skeletal ciliopathies. *Hum Mol Genet*, 24, 1410-9.
- SHAHEEN, R., SHAMSELDIN, H. E., LOUCKS, C. M., SEIDAHMED, M. Z., ANSARI, S., IBRAHIM KHALIL, M., AL-YACOUB, N., DAVIS, E. E., MOLA, N. A., SZYMANSKA, K., HERRIDGE, W., CHUDLEY, A. E., CHODIRKER, B. N., SCHWARTZENTRUBER, J., MAJEWSKI, J., KATSANIS, N., POIZAT, C., JOHNSON, C. A., PARBOOSINGH, J., BOYCOTT, K. M., INNES, A. M. & ALKURAYA, F. S. 2014. Mutations in CSPP1, Encoding a Core Centrosomal Protein, Cause a Range of Ciliopathy Phenotypes in Humans. *The American Journal of Human Genetics*, 94, 73-79.
- SHAHEEN, R., SZYMANSKA, K., BASU, B., PATEL, N., EWIDA, N., FAQEIH, E., AL HASHEM, A., DERAR, N., ALSHARIF, H., ALDAHMESH, M. A., ALAZAMI, A. M., HASHEM, M., IBRAHIM, N., ABDULWAHAB, F. M., SONBUL, R., ALKURAYA, H., ALNEMER, M., AL TALA, S., AL-HUSAIN, M., MORSY, H., SEIDAHMED, M. Z., MERIKI, N., AL-OWAIN, M., ALSAHAWAN, S., TABARKI, B., SALIH, M. A., CILIOPATHY, W., FAQUIH, T., EL-KALIOBY, M., UEFFING, M., BOLDT, K., LOGAN, C. V., PARRY, D. A., AL TASSAN, N., MONIES, D., MEGARBANE, A., ABOUELHODA, M., HALEES, A., JOHNSON, C. A. & ALKURAYA, F. S. 2016. Characterizing the morbid genome of ciliopathies. *Genome Biol*, 17, 242.
- SHAMSELDIN, H. E., SHAHEEN, R., EWIDA, N., BUBSHAIT, D. K., ALKURAYA, H., ALMARDAWI, E., HOWAIDI, A., SABR, Y., ABDALLA, E. M., ALFAIFI, A. Y., ALGHAMDI, J. M., ALSAGHEIR, A., ALFARES, A., MORSY, H., HUSSEIN, M. H., AL-MUHAIZEA, M. A., SHAGRANI, M., AL SABBAN, E., SALIH, M. A., MERIKI, N., KHAN, R., ALMUGBEL, M., QARI, A., TULBA, M., MAHNASHI, M., ALHAZMI, K., ALSALAMAH, A. K., NOWILATY, S. R., ALHASHEM, A., HASHEM, M., ABDULWAHAB, F., IBRAHIM, N., ALSHIDI, T., ALOBEID, E., ALENAZI, M. M., ALZAIDAN, H., RAHBEENI, Z., AL-OWAIN, M., SOGATY, S., SEIDAHMED, M. Z. & ALKURAYA, F. S. 2020. The morbid genome of ciliopathies: an update. *Genet Med*.
- SHAO, L., EL-JOUNI, W., KONG, F., RAMESH, J., KUMAR, R. S., SHEN, X., REN, J., DEVENDRA, S., DORSCHEL, A., WU, M., BARRERA, I., TABARI, A., HU, K., HAQUE, N., YAMBAYEV, I., LI, S., KUMAR, A., BEHERA, T. R., MCDONOUGH, G., FURUICHI, M., XIFARAS, M., LU, T., ALHAYAZA, R. M., MIYABAYASHI, K., FAN, Q., AJAY, A. K. & ZHOU, J. 2020a. Genetic reduction of cilium length by targeting intraflagellar transport 88 protein impedes kidney and liver cyst formation in mouse models of autosomal polycystic kidney disease. *Kidney Int*, 98, 1225-1241.
- SHAO, W., YANG, J., HE, M., YU, X. Y., LEE, C. H., YANG, Z., JOYNER, A. L., ANDERSON, K. V., ZHANG, J., TSOU, M. B., SHI, H. & SHI, S. H. 2020b. Centrosome anchoring regulates progenitor properties and cortical formation. *Nature*, 580, 106-112.

- SHARMA, A., GERARD, S. F., OLIERIC, N. & STEINMETZ, M. O. 2018. Cep120 promotes microtubule formation through a unique tubulin binding C2 domain. *J Struct Biol*, 203, 62-70.
- SHEIKH, S. A., SISK, R. A., SCHIAVON, C. R., WARYAH, Y. M., USMANI, M. A., STEEL, D. H., SAYER, J. A., NARSANI, A. K., HUFNAGEL, R. B., RIAZUDDIN, S., KAHN, R. A., WARYAH, A. M. & AHMED, Z. M. 2019. Homozygous Variant in ARL3 Causes Autosomal Recessive Cone Rod Dystrophy. *Invest Ophthalmol Vis Sci*, 60, 4811-4819.
- SILVEIRA, K. C., MORENO, C. A. & CAVALCANTI, D. P. 2017. Beemer-Langer syndrome is a ciliopathy due to biallelic mutations in IFT122. *Am J Med Genet A*, 173, 1186-1189.
- SIRMACI, A., SPILIOPOULOS, M., BRANCATI, F., POWELL, E., DUMAN, D., ABRAMS, A., BADEMCI, G., AGOLINI, E., GUO, S., KONUK, B., KAVAZ, A., BLANTON, S., DIGILIO, M. C., DALLAPICCOLA, B., YOUNG, J., ZUCHNER, S. & TEKIN, M. 2011. Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. *Am J Hum Genet*, 89, 289-94.
- SLAATS, G. G., SALDIVAR, J. C., BACAL, J., ZEMAN, M. K., KILE, A. C., HYNES, A. M., SRIVASTAVA, S., NAZMUTDINOVA, J., DEN OUDEN, K., ZAGERS, M. S., FOLETTI, V., VERHAAR, M. C., MILES, C., SAYER, J. A., CIMPRICH, K. A. & GILES, R. H. 2015. DNA replication stress underlies renal phenotypes in CEP290-associated Joubert syndrome. *J Clin Invest*, 125, 3657-66.
- SMEDLEY, D., SMITH, K. R., MARTIN, A., THOMAS, E. A., MCDONAGH, E. M., CIPRIANI, V., ELLINGFORD, J. M., ARNO, G., TUCCI, A., VANDROVCVA, J., CHAN, G., WILLIAMS, H. J., RATNAIKE, T., WEI, W., STIRRUPS, K., IBANEZ, K., MOUTSIANAS, L., WIELSCHER, M., NEED, A., BARNES, M. R., VESTITO, L., BUCHANAN, J., WORDSWORTH, S., ASHFORD, S., REHMSTRÖM, K., LI, E., FULLER, G., TWISS, P., SPASIC-BOSKOVIC, O., HALSALL, S., FLOTO, R. A., POOLE, K., WAGNER, A., MEHTA, S. G., GURNELL, M., BURROWS, N., JAMES, R., PENKETT, C., DEWHURST, E., GRÄF, S., MAPETA, R., KASANICKI, M., HAWORTH, A., SAVAGE, H., BABCOCK, M., REESE, M. G., BALE, M., BAPLE, E., BOUSTRED, C., BRITTAIN, H., DE BURCA, A., BLEDA, M., DEVEREAUX, A., HALAI, D., HARALDSOTTIR, E., HYDER, Z., KASPERAVICIUTE, D., PATCH, C., POLYCHRONOPOULOS, D., MATCHAN, A., SULTANA, R., RYTEN, M., TAVARES, A. L. T., TREGIDGO, C., TURNBULL, C., WELLAND, M., WOOD, S., SNOW, C., WILLIAMS, E., LEIGH, S., FOULGER, R. E., DAUGHERTY, L. C., NIBLOCK, O., LEONG, I. U. S., WRIGHT, C. F., DAVIES, J., CRICHTON, C., WELCH, J., WOODS, K., ABULHOUL, L., AURORA, P., BOCKENHAUER, D., BROOMFIELD, A., CLEARY, M. A., LAM, T., DATTANI, M., FOOTITT, E., GANESAN, V., GRUNEWALD, S., COMPEYROT-LACASSAGNE, S., MUNTONI, F., PILKINGTON, C., QUINLIVAN, R., THAPAR, N., WALLIS, C., WEDDERBURN, L. R., WORTH, A., BUESER, T., COMPTON, C., DESHPANDE, C., et al. 2021. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med*, 385, 1868-1880.
- SMITH, U. M., CONSUGAR, M., TEE, L. J., MCKEE, B. M., MAINA, E. N., WHELAN, S., MORGAN, N. V., GORANSON, E., GISSEN, P., LILLIQUIST, S., ALIGIANIS, I. A., WARD, C. J., PASHA, S., PUNYASHTHITI, R., MALIK SHARIF, S., BATMAN, P. A., BENNETT, C. P., WOODS, C. G., MCKEOWN, C., BUCOURT, M., MILLER, C. A., COX, P., ALGAZALI, L., TREMBATH, R. C., TORRES, V. E., ATTIE-BITACH, T., KELLY, D. A., MAHER, E. R., GATTONE, V. H., 2ND, HARRIS, P. C. & JOHNSON, C. A. 2006. The

- transmembrane protein meckelin (MKS3) is mutated in Meckel-Gruber syndrome and the wpk rat. *Nat Genet*, 38, 191-6.
- SNOEK, R., VAN SETTEN, J., KEATING, B. J., ISRANI, A. K., JACOBSON, P. A., OETTING, W. S., MATAS, A. J., MANNON, R. B., ZHANG, Z., ZHANG, W., HAO, K., MURPHY, B., REINDL-SCHWAIGHOFER, R., HEINZL, A., OBERBAUER, R., VIKLICKY, O., CONLON, P. J., STAPLETON, C. P., BAKKER, S. J. L., SNIEDER, H., PETERS, E. D. J., VAN DER ZWAAG, B., KNOERS, N., DE BORST, M. H. & VAN EERDE, A. M. 2018. NPHP1 (Nephrocystin-1) Gene Deletions Cause Adult-Onset ESRD. *J Am Soc Nephrol*, 29, 1772-1779.
- SOBECKI, M., MROUJ, K., CAMASSES, A., PARISIS, N., NICOLAS, E., LLÈRES, D., GERBE, F., PRIETO, S., KRASINSKA, L., DAVID, A., EGUREN, M., BIRLING, M. C., URBACH, S., HEM, S., DÉJARDIN, J., MALUMBRES, M., JAY, P., DULIC, V., LAFONTAINE, D., FEIL, R. & FISHER, D. 2016. The cell proliferation antigen Ki-67 organises heterochromatin. *Elife*, 5, e13722.
- SOTELO, C. 2015. Molecular layer interneurons of the cerebellum: developmental and morphological aspects. *Cerebellum*, 14, 534-56.
- SPENCER-TANSLEY, R., MEADE, N., ALI, F., SIMPSON, A. & HUNTER, A. 2022. Mental health care for rare disease in the UK - recommendations from a quantitative survey and multi-stakeholder workshop. *BMC Health Serv Res*, 22, 648.
- SRIVASTAVA, S., MOLINARI, E., RAMAN, S. & SAYER, J. A. 2017a. Many Genes- One Disease? Genetics of Nephronophthisis (NPHP) and NPHP-Associated Disorders. *Front Pediatr*, 5, 287.
- SRIVASTAVA, S., RAMSBOTTOM, S. A., MOLINARI, E., ALKANDERI, S., FILBY, A., WHITE, K., HENRY, C., SAUNIER, S., MILES, C. G. & SAYER, J. A. 2017b. A human patient-derived cellular model of Joubert syndrome reveals ciliary defects which can be rescued with targeted therapies. *Hum Mol Genet*, 26, 4657-4667.
- SROUR, M., HAMDAN, F. F., SCHWARTZENTRUBER, J. A., PATRY, L., OSPINA, L. H., SHEVELL, M. I., DÉSILETS, V., DOBRZENIECKA, S., MATHONNET, G., LEMYRE, E., MASSICOTTE, C., LABUDA, D., AMROM, D., ANDERMANN, E., SÉBIRE, G., MARANDA, B., ROULEAU, G. A., MAJEWSKI, J. & MICHAUD, J. L. 2012. Mutations in TMEM231 cause Joubert syndrome in French Canadians. *J Med Genet*, 49, 636-41.
- STEMBALSKA, A., RYDZANICZ, M., KLANIEWSKA, M., DUDAREWICZ, L., POLLAK, A., BIELA, M., STAWINSKI, P., PLOSKI, R. & SMIGIEL, R. 2022. Prenatal Diagnosis of Jeune Syndrome Caused by Compound Heterozygous Variants in DYNC2H1 Gene-Case Report with Rapid WES Procedure and Differential Diagnosis of Lethal Skeletal Dysplasias. *Genes (Basel)*, 13.
- STENSON, P. D., MORT, M., BALL, E. V., EVANS, K., HAYDEN, M., HEYWOOD, S., HUSSAIN, M., PHILLIPS, A. D. & COOPER, D. N. 2017. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet*, 136, 665-677.
- STOKMAN, M. F., SAUNIER, S. & BENMERAH, A. 2021. Renal Ciliopathies: Sorting Out Therapeutic Approaches for Nephronophthisis. *Front Cell Dev Biol*, 9, 653138.
- STREETS, A. J., PROSSEDA, P. P. & ONG, A. C. 2020. Polycystin-1 regulates ARHGAP35-dependent centrosomal RhoA activation and ROCK signaling. *JCI Insight*, 5.

- STROM, S. P., CLARK, M. J., MARTINEZ, A., GARCIA, S., ABELAZEEM, A. A., MATYNIA, A., PARIKH, S., SULLIVAN, L. S., BOWNE, S. J., DAIGER, S. P. & GORIN, M. B. 2016. De Novo Occurrence of a Variant in ARL3 and Apparent Autosomal Dominant Transmission of Retinitis Pigmentosa. *PLoS One*, 11, e0150944.
- SZYMANSKA, K., BERRY, I., LOGAN, C. V., COUSINS, S. R., LINDSAY, H., JAFRI, H., RAASHID, Y., MALIK-SHARIF, S., CASTLE, B., AHMED, M., BENNETT, C., CARLTON, R. & JOHNSON, C. A. 2012. Founder mutations and genotype-phenotype correlations in Meckel-Gruber syndrome and associated ciliopathies. *Cilia*, 1, 18.
- TA, C. M., VIEN, T. N., NG, L. C. T. & DECAEN, P. G. 2020. Structure and function of polycystin channels in primary cilia. *Cell Signal*, 72, 109626.
- TAHATA, S., GUNDERSON, L., LANPHER, B. & MORAVA, E. 2019. Complex phenotypes in ALG12-congenital disorder of glycosylation (ALG12-CDG): Case series and review of the literature. *Mol Genet Metab*, 128, 409-414.
- TAIPALE, J., CHEN, J. K., COOPER, M. K., WANG, B., MANN, R. K., MILENKOVIC, L., SCOTT, M. P. & BEACHY, P. A. 2000. Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature*, 406, 1005-9.
- TALLILA, J., JAKKULA, E., PELTONEN, L., SALONEN, R. & KESTILA, M. 2008. Identification of CC2D2A as a Meckel syndrome gene adds an important piece to the ciliopathy puzzle. *Am J Hum Genet*, 82, 1361-7.
- TALLILA, J., SALONEN, R., KOHLSCHMIDT, N., PELTONEN, L. & KESTILA, M. 2009. Mutation spectrum of Meckel syndrome genes: one group of syndromes or several distinct groups? *Hum Mutat*, 30, E813-30.
- TAMAYO, A., NÚÑEZ-MORENO, G., RUIZ, C., PLAISANCIE, J., DAMIAN, A., MOYA, J., CHASSAING, N., CALVAS, P., AYUSO, C., MINGUEZ, P. & CORTON, M. 2023. Minigene Splicing Assays and Long-Read Sequencing to Unravel Pathogenic Deep-Intronic Variants in PAX6 in Congenital Aniridia. *Int J Mol Sci*, 24.
- TASAKI, S., XU, J., AVEY, D. R., JOHNSON, L., PETYUK, V. A., DAWE, R. J., BENNETT, D. A., WANG, Y. & GAITERI, C. 2022. Inferring protein expression changes from mRNA in Alzheimer's dementia using deep neural networks. *Nat Commun*, 13, 655.
- TASKIRAN, E. Z., KORKMAZ, E., GUCER, S., KOSUKCU, C., KAYMAZ, F., KOYUNLAR, C., BRYDA, E. C., CHAKI, M., LU, D., VADNAGARA, K., CANDAN, C., TOPALOGLU, R., SCHAEFER, F., ATTANASIO, M., BERGMANN, C. & OZALTIN, F. 2014. Mutations in ANKS6 cause a nephronophthisis-like phenotype with ESRD. *J Am Soc Nephrol*, 25, 1653-61.
- TAVERNA, E. & HUTTNER, W. B. 2010. Neural progenitor nuclei IN motion. *Neuron*, 67, 906-14.
- TERUEL, M., BARTUREN, G., MARTÍNEZ-BUENO, M., CASTELLINI-PÉREZ, O., BARROSO-GIL, M., POVEDANO, E., KERICK, M., CATALÀ-MOLL, F., MAKOWSKA, Z., BUTTGEREIT, A., PERS, J. O., MARAÑÓN, C., BALLESTAR, E., MARTIN, J., CARNERO-MONTORO, E. & ALARCÓN-RIQUELME, M. E. 2021. Integrative epigenomics in Sjögren's syndrome reveals novel pathways and a strong interaction between the HLA, autoantibodies and the interferon signature. *Sci Rep*, 11, 23292.
- THAM, E., EKLUND, E. A., HAMMARSJÖ, A., BENGTSON, P., GEIBERGER, S., LAGERSTEDT-ROBINSON, K., MALMGREN, H., NILSSON, D., GRIGELIONIS, G., CONNER, P., LINDGREN, P., LINDSTRAND, A., WEDELL, A., ALBÅGE, M., ZIELINSKA, K., NORDGREN, A.,

- PAPADOGIANNAKIS, N., NISHIMURA, G. & GRIGELIONIENE, G. 2016. A novel phenotype in N-glycosylation disorders: Gillessen-Kaesbach-Nishimura skeletal dysplasia due to pathogenic variants in ALG9. *Eur J Hum Genet*, 24, 198-207.
- THAUVIN-ROBINET, C., LEE, J. S., LOPEZ, E., HERRANZ-PÉREZ, V., SHIDA, T., FRANCO, B., JEGO, L., YE, F., PASQUIER, L., LOGET, P., GIGOT, N., ARAL, B., LOPES, C. A., ST-ONGE, J., BRUEL, A. L., THEVENON, J., GONZÁLEZ-GRANERO, S., ALBY, C., MUNNICH, A., VEKEMANS, M., HUET, F., FRY, A. M., SAUNIER, S., RIVIÈRE, J. B., ATTÉ-BITACH, T., GARCIA-VERDUGO, J. M., FAIVRE, L., MÉGARBANÉ, A. & NACHURY, M. V. 2014. The oral-facial-digital syndrome gene C2CD3 encodes a positive regulator of centriole elongation. *Nat Genet*, 46, 905-11.
- THE GTEX CONSORTIUM 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, 369, 1318-1330.
- THUL, P. J. & LINDSKOG, C. 2018. The human protein atlas: A spatial map of the human proteome. *Protein Sci*, 27, 233-244.
- TISDALE, A., CUTILLO, C. M., NATHAN, R., RUSSO, P., LARAWAY, B., HAENDEL, M., NOWAK, D., HASCHE, C., CHAN, C. H., GRIESE, E., DAWKINS, H., SHUKLA, O., PEARCE, D. A., RUTTER, J. L. & PARISER, A. R. 2021. The IDeAS initiative: pilot study to assess the impact of rare diseases on patients and healthcare systems. *Orphanet J Rare Dis*, 16, 429.
- TONG, L., RAO, J., YANG, C., XU, J., LU, Y., ZHANG, Y., CANG, X., XIE, S., MAO, J. & JIANG, P. 2023. Mutational burden of XPNPEP3 leads to defects in mitochondrial complex I and cilia in NPHPL1. *iScience*, 26, 107446.
- TORY, K., ROUSSET-ROUVIÈRE, C., GUBLER, M. C., MORINIÈRE, V., PAWTOWSKI, A., BECKER, C., GUYOT, C., GIÉ, S., FRISHBERG, Y., NIVET, H., DESCHÈNES, G., COCHAT, P., GAGNADOUX, M. F., SAUNIER, S., ANTIGNAC, C. & SALOMON, R. 2009. Mutations of NPHP2 and NPHP3 in infantile nephronophthisis. *Kidney Int*, 75, 839-47.
- TRAPANI, I. & AURICCHIO, A. 2019. Has retinal gene therapy come of age? From bench to bedside and back to bench. *Hum Mol Genet*, 28, R108-r118.
- TRAVAGLINI, L., BRANCATI, F., ATTÉ-BITACH, T., AUDOLLENT, S., BERTINI, E., KAPLAN, J., PERRAULT, I., IANNICELLI, M., MANCUSO, B., RIGOLI, L., ROZET, J.-M., SWISTUN, D., TOLENTINO, J., DALLAPICCOLA, B., GLEESON, J. G. & VALENTE, E. M. 2009. Expanding CEP290 mutational spectrum in ciliopathies. *American Journal of Medical Genetics Part A*, 149A, 2173-2180.
- TREJO-REVELES, V., OWEN, N., CHING CHAN, B. H., TOMS, M., SCHOENEBECK, J. J., MOOSAJEE, M. & RANGER, J. 2023. Identification of Novel Coloboma Candidate Genes through Conserved Gene Expression Analyses across Four Vertebrate Species. *Biomolecules*, 13.
- TSAI, J. J., HSU, W. B., LIU, J. H., CHANG, C. W. & TANG, T. K. 2019. CEP120 interacts with C2CD3 and Talpid3 and is required for centriole appendage assembly and ciliogenesis. *Sci Rep*, 9, 6037.
- TUCKER, R. W., PARDEE, A. B. & FUJIWARA, K. 1979. Centriole ciliation is related to quiescence and DNA synthesis in 3T3 cells. *Cell*, 17, 527-35.
- TURNBULL, C., SCOTT, R. H., THOMAS, E., JONES, L., MURUGAESU, N., PRETTY, F. B., HALAI, D., BAPLE, E., CRAIG, C., HAMBLIN, A., HENDERSON, S., PATCH, C., O'NEILL, A., DEVEREAUX, A., SMITH, K., MARTIN, A. R., SOSINSKY, A., MCDONAGH, E. M., SULTANA, R., MUELLER, M., SMEDLEY, D., TOMS, A., DINH, L., FOWLER, T., BALE, M., HUBBARD, T., RENDON, A., HILL, S. & CAULFIELD, M. J. 2018. The

100 000 Genomes Project: bringing whole genome sequencing to the NHS. *Bmj*, 361, k1687.

TURRO, E., ASTLE, W. J., MEGY, K., GRÄF, S., GREENE, D., SHAMARDINA, O., ALLEN, H. L., SANCHIS-JUAN, A., FRONTINI, M., THYS, C., STEPHENS, J., MAPETA, R., BURREN, O. S., DOWNES, K., HAIMEL, M., TUNA, S., DEEVI, S. V. V., AITMAN, T. J., BENNETT, D. L., CALLEJA, P., CARSS, K., CAULFIELD, M. J., CHINNERY, P. F., DIXON, P. H., GALE, D. P., JAMES, R., KOZIELL, A., LAFFAN, M. A., LEVINE, A. P., MAHER, E. R., MARKUS, H. S., MORALES, J., MORRELL, N. W., MUMFORD, A. D., ORMONDROYD, E., RANKIN, S., RENDON, A., RICHARDSON, S., ROBERTS, I., ROY, N. B. A., SALEEM, M. A., SMITH, K. G. C., STARK, H., TAN, R. Y. Y., THEMISTOCLEOUS, A. C., THRASHER, A. J., WATKINS, H., WEBSTER, A. R., WILKINS, M. R., WILLIAMSON, C., WHITWORTH, J., HUMPHRAY, S., BENTLEY, D. R., KINGSTON, N., WALKER, N., BRADLEY, J. R., ASHFORD, S., PENKETT, C. J., FRESON, K., STIRRUPS, K. E., RAYMOND, F. L. & OUWEHAND, W. H. 2020. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*, 583, 96-102.

TUZ, K., BACHMANN-GAGESCU, R., O'DAY, D. R., HUA, K., ISABELLA, C. R., PHELPS, I. G., STOLARSKI, A. E., O'ROAK, B. J., DEMPSEY, J. C., LOURENCO, C., ALSWAID, A., BÖNNEMANN, C. G., MEDNE, L., NAMPOOTHIRI, S., STARK, Z., LEVENTER, R. J., TOPÇU, M., CANSU, A., JAGADEESH, S., DONE, S., ISHAK, G. E., GLASS, I. A., SHENDURE, J., NEUHAUSS, S. C., HALDEMAN-ENGLERT, C. R., DOHERTY, D. & FERLAND, R. J. 2014. Mutations in CSPP1 cause primary cilia abnormalities and Joubert syndrome with or without Jeune asphyxiating thoracic dystrophy. *Am J Hum Genet*, 94, 62-72.

TYC, K. M., EL YAKOUBI, W., BAG, A., LANDIS, J., ZHAN, Y., TREFF, N. R., SCOTT, R. T., TAO, X., SCHINDLER, K. & XING, J. 2020. Exome sequencing links CEP120 mutation to maternally derived aneuploid conception risk. *Hum Reprod*, 35, 2134-2148.

UHLÉN, M., FAGERBERG, L., HALLSTRÖM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, Å., KAMPF, C., SJÖSTEDT, E., ASPLUND, A., OLSSON, I., EDLUND, K., LUNDBERG, E., NAVANI, S., SZIGYARTO, C. A., ODEBERG, J., DJUREINOVIC, D., TAKANEN, J. O., HOBER, S., ALM, T., EDQVIST, P. H., BERLING, H., TEGEL, H., MULDER, J., ROCKBERG, J., NILSSON, P., SCHWENK, J. M., HAMSTEN, M., VON FEILITZEN, K., FORSBERG, M., PERSSON, L., JOHANSSON, F., ZWAHLEN, M., VON HEIJNE, G., NIELSEN, J. & PONTÉN, F. 2015. Proteomics. Tissue-based map of the human proteome. *Science*, 347, 1260419.

UMLAI, U. I., BANGARUSAMY, D. K., ESTIVILL, X. & JITHESH, P. V. 2022. Genome sequencing data analysis for rare disease gene discovery. *Brief Bioinform*, 23.

UTSCH, B., SAYER, J. A., ATTANASIO, M., PEREIRA, R. R., ECCLES, M., HENNIES, H. C., OTTO, E. A. & HILDEBRANDT, F. 2006. Identification of the first AHI1 gene mutations in nephronophthisis-associated Joubert syndrome. *Pediatr Nephrol*, 21, 32-5.

VALENTE, E. M., BRANCATI, F. & DALLAPICCOLA, B. 2008. Genotypes and phenotypes of Joubert syndrome and related disorders. *Eur J Med Genet*, 51, 1-23.

VALENTE, E. M., ROSTI, R. O., GIBBS, E. & GLEESON, J. G. 2014. Primary cilia in neurodevelopmental disorders. *Nature Reviews Neurology*, 10, 27-36.

- VALENTE, E. M., SILHAVY, J. L., BRANCATI, F., BARRANO, G., KRISHNASWAMI, S. R., CASTORI, M., LANCASTER, M. A., BOLTSHAUSER, E., BOCCONE, L., AL-GAZALI, L., FAZZI, E., SIGNORINI, S., LOUIE, C. M., BELLACCHIO, E., RELATED DISORDERS STUDY GROUP, I. J. S., BERTINI, E., DALLAPICCOLA, B. & GLEESON, J. G. 2006. Mutations in CEP290, which encodes a centrosomal protein, cause pleiotropic forms of Joubert syndrome. *Nature Genetics*, 38, 623-625.
- VAN DAM, T. J. P., WHEWAY, G., SLAATS, G. G., HUYNEN, M. A. & GILES, R. H. 2013. The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. *Cilia*, 2, 1-5.
- VAN DE WEGHE, J. C., RUSTERHOLZ, T. D. S., LATOUR, B., GROUT, M. E., ALDINGER, K. A., SHAHEEN, R., DEMPSEY, J. C., MADDIREVULA, S., CHENG, Y. H., PHELPS, I. G., GESEMANN, M., GOEL, H., BIRK, O. S., ALANZI, T., RAWASHDEH, R., KHAN, A. O., BAMSHAD, M. J., NICKERSON, D. A., NEUHAUSS, S. C. F., DOBYNS, W. B., ALKURAYA, F. S., ROEPMAN, R., BACHMANN-GAGESCU, R. & DOHERTY, D. 2017. Mutations in ARMC9, which Encodes a Basal Body Protein, Cause Joubert Syndrome in Humans and Ciliopathy Phenotypes in Zebrafish. *Am J Hum Genet*, 101, 23-36.
- VAN NIEUWENHOVE, E., BARBER, J. S., NEUMANN, J., SMEETS, E., WILLEMSSEN, M., PASCIUTO, E., PREZZEMOLO, T., LAGOU, V., SELDESLACHTS, L., MALENGIER-DEVRIES, B., METZEMAEKERS, M., HARDENTENTEUFEL, S., KERSTENS, A., VAN DER KANT, R., ROUSSEAU, F., SCHYMKOWITZ, J., DI MARINO, D., LANG, S., ZIMMERMANN, R., SCHLENNER, S., MUNCK, S., PROOST, P., MATTHYS, P., DEVALCK, C., BOECKX, N., CLAESSENS, F., WOUTERS, C., HUMBLET-BARON, S., MEYTS, I. & LISTON, A. 2020. Defective Sec61 α 1 underlies a novel cause of autosomal dominant severe congenital neutropenia. *J Allergy Clin Immunol*, 146, 1180-1193.
- VAN SPRONSEN, F. J., BLAU, N., HARDING, C., BURLINA, A., LONGO, N. & BOSCH, A. M. 2021. Phenylketonuria. *Nat Rev Dis Primers*, 7, 36.
- VASILEVA, V. Y., SULTANOVA, R. F., SUDARIKOVA, A. V. & ILATOVSKAYA, D. V. 2021. Insights Into the Molecular Mechanisms of Polycystic Kidney Diseases. *Front Physiol*, 12, 693130.
- VELAND, I. R., MONTJEAN, R., ELEY, L., PEDERSEN, L. B., SCHWAB, A., GOODSHIP, J., KRISTIANSEN, K., PEDERSEN, S. F., SAUNIER, S. & CHRISTENSEN, S. T. 2013. Inversin/Nephrocystin-2 is required for fibroblast polarity and directional cell migration. *PLoS One*, 8, e60193.
- VELERI, S., MANJUNATH, S. H., FARISS, R. N., MAY-SIMERA, H., BROOKS, M., FOSKETT, T. A., GAO, C., LONGO, T. A., LIU, P., NAGASHIMA, K., RACHEL, R. A., LI, T., DONG, L. & SWAROOP, A. 2014. Ciliopathy-associated gene Cc2d2a promotes assembly of subdistal appendages on the mother centriole during cilia biogenesis. *Nat Commun*, 5, 4207.
- VELTEL, S., GASPER, R., EISENACHER, E. & WITTINGHOFER, A. 2008. The retinitis pigmentosa 2 gene product is a GTPase-activating protein for Arf-like 3. *Nat Struct Mol Biol*, 15, 373-80.
- VENTER, J. C., ADAMS, M. D., MYERS, E. W., LI, P. W., MURAL, R. J., SUTTON, G. G., SMITH, H. O., YANDELL, M., EVANS, C. A., HOLT, R. A., GOCAYNE, J. D., AMANATIDES, P., BALLEW, R. M., HUSON, D. H., WORTMAN, J. R., ZHANG, Q., KODIRA, C. D., ZHENG, X. H., CHEN, L., SKUPSKI, M., SUBRAMANIAN, G., THOMAS, P. D., ZHANG, J., GABOR MIKLOS, G. L., NELSON, C., BRODER, S., CLARK, A. G., NADEAU, J., MCKUSICK, V. A., ZINDER, N., LEVINE, A. J., ROBERTS, R. J., SIMON, M., SLAYMAN, C.,

- HUNKAPILLER, M., BOLANOS, R., DELCHER, A., DEW, I., FASULO, D., FLANIGAN, M., FLOREA, L., HALPERN, A., HANNENHALLI, S., KRAVITZ, S., LEVY, S., MOBARRY, C., REINERT, K., REMINGTON, K., ABUTHREIDEH, J., BEASLEY, E., BIDDICK, K., BONAZZI, V., BRANDON, R., CARGILL, M., CHANDRAMOULISWARAN, I., CHARLAB, R., CHATURVEDI, K., DENG, Z., DI FRANCESCO, V., DUNN, P., EILBECK, K., EVANGELISTA, C., GABRIELIAN, A. E., GAN, W., GE, W., GONG, F., GU, Z., GUAN, P., HEIMAN, T. J., HIGGINS, M. E., JI, R. R., KE, Z., KETCHUM, K. A., LAI, Z., LEI, Y., LI, Z., LI, J., LIANG, Y., LIN, X., LU, F., MERKULOV, G. V., MILSHINA, N., MOORE, H. M., NAIK, A. K., NARAYAN, V. A., NEELAM, B., NUSSKERN, D., RUSCH, D. B., SALZBERG, S., SHAO, W., SHUE, B., SUN, J., WANG, Z., WANG, A., WANG, X., WANG, J., WEI, M., WIDES, R., XIAO, C., YAN, C., et al. 2001. The sequence of the human genome. *Science*, 291, 1304-51.
- VERTII, A., BRIGHT, A., DELAVAL, B., HEHNL, H. & DOXSEY, S. 2015. New frontiers: discovering cilia-independent functions of cilia proteins. *EMBO reports*, 16, 1275-1287.
- VIG, A., POULTER, J. A., OTTAVIANI, D., TAVARES, E., TOROPOVA, K., TRACEWSKA, A. M., MOLICA, A., KANG, J., KEHELWATHUGODA, O., PATON, T., MAYNES, J. T., WHEWAY, G., ARNO, G., KHAN, K. N., MCKIBBIN, M., TOOMES, C., ALI, M., DI SCIPIO, M., LI, S., ELLINGFORD, J., BLACK, G., WEBSTER, A., RYDZANICZ, M., STAWIŃSKI, P., PŁOSKI, R., VINCENT, A., CHEETHAM, M. E., INGLEHEARN, C. F., ROBERTS, A. & HEON, E. 2020. DYNC2H1 hypomorphic or retina-predominant variants cause nonsyndromic retinal degeneration. *Genet Med*, 22, 2041-2051.
- VILBOUX, T., DOHERTY, D. A., GLASS, I. A., PARISI, M. A., PHELPS, I. G., CULLINANE, A. R., ZEIN, W., BROOKS, B. P., HELLER, T., SOLDATOS, A., ODEN, N. L., YILDIRIMLI, D., VEMULAPALLI, M., MULLIKIN, J. C., NISC COMPARATIVE SEQUENCING, P., MALICDAN, M. C. V., GAHL, W. A. & GUNAY-AYGUN, M. 2017a. Molecular genetic findings and clinical correlations in 100 patients with Joubert syndrome and related disorders prospectively evaluated at a single center. *Genet Med*, 19, 875-882.
- VILBOUX, T., MALICDAN, M. C., RONEY, J. C., CULLINANE, A. R., STEPHEN, J., YILDIRIMLI, D., BRYANT, J., FISCHER, R., VEMULAPALLI, M., MULLIKIN, J. C., STEINBACH, P. J., GAHL, W. A. & GUNAY-AYGUN, M. 2017b. CELSR2, encoding a planar cell polarity protein, is a putative gene in Joubert syndrome with cortical heterotopia, microophthalmia, and growth hormone deficiency. *Am J Med Genet A*, 173, 661-666.
- VINCENSINI, L., BLISNICK, T. & BASTIN, P. 2011. 1001 Model Organisms To Study Cilia and Flagella. *Biology of the Cell*, 103, 109-130.
- VOITH VON VOITHENBERG, L., FOMITCHeva KHARTCHENKO, A., HUBER, D., SCHRAML, P. & KAIGALA, G. V. 2020. Spatially multiplexed RNA in situ hybridization to reveal tumor heterogeneity. *Nucleic Acids Res*, 48, e17.
- VOLLAND, S., ESTEVE-RUDD, J., HOO, J., YEE, C. & WILLIAMS, D. S. 2015. A comparison of some organizational characteristics of the mouse central retina and the human macula. *PLoS One*, 10, e0125631.
- WAKAMIYA, M., MATSUURA, T., LIU, Y., SCHUSTER, G. C., GAO, R., XU, W., SARKAR, P. S., LIN, X. & ASHIZAWA, T. 2006. The role of ataxin 10 in the pathogenesis of spinocerebellar ataxia type 10. *Neurology*, 67, 607-13.
- WALCZAK-SZTULPA, J., EGGENSCHWILER, J., OSBORN, D., BROWN, D. A., EMMA, F., KLINGENBERG, C., HENNEKAM, R. C., TORRE, G., GARSHASBI, M., TZSCHACH, A., SZCZEPANSKA, M., KRAWCZYNSKI, M.,

- ZACHWIEJA, J., ZWOLINSKA, D., BEALES, P. L., ROPERS, H. H., LATOS-BIELENSKA, A. & KUSS, A. W. 2010. Cranioectodermal Dysplasia, Sensenbrenner syndrome, is a ciliopathy caused by mutations in the IFT122 gene. *Am J Hum Genet*, 86, 949-56.
- WANG, C., LOW, W. C., LIU, A. & WANG, B. 2013. Centrosomal protein DZIP1 regulates Hedgehog signaling by promoting cytoplasmic retention of transcription factor GLI3 and affecting ciliogenesis. *J Biol Chem*, 288, 29518-29.
- WANG, F., FLANAGAN, J., SU, N., WANG, L. C., BUI, S., NIELSON, A., WU, X., VO, H. T., MA, X. J. & LUO, Y. 2012. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn*, 14, 22-9.
- WANG, S., ZHANG, J., NAULI, S. M., LI, X., STARREMAN, P. G., LUO, Y., ROBERTS, K. A. & ZHOU, J. 2007. Fibrocystin/polyductin, found in the same protein complex with polycystin-2, regulates calcium responses in kidney epithelia. *Mol Cell Biol*, 27, 3241-52.
- WANG, S. F., KOWAL, T. J., NING, K., KOO, E. B., WU, A. Y., MAHAJAN, V. B. & SUN, Y. 2018. Review of Ocular Manifestations of Joubert Syndrome. *Genes (Basel)*, 9.
- WARD, C. J., HOGAN, M. C., ROSSETTI, S., WALKER, D., SNEDDON, T., WANG, X., KUBLY, V., CUNNINGHAM, J. M., BACALLAO, R., ISHIBASHI, M., MILLINER, D. S., TORRES, V. E. & HARRIS, P. C. 2002. The gene mutated in autosomal recessive polycystic kidney disease encodes a large, receptor-like protein. *Nat Genet*, 30, 259-69.
- WATERSTON, R. H., LANDER, E. S. & SULSTON, J. E. 2002. On the sequencing of the human genome. *Proc Natl Acad Sci U S A*, 99, 3712-6.
- WATNICK, T. & GERMINO, G. 2003. From cilia to cyst. *Nature Genetics*, 34, 355-356.
- WATSON, C. M., CRINNION, L. A., BERRY, I. R., HARRISON, S. M., LASCELLES, C., ANTANAVICIUTE, A., CHARLTON, R. S., DOBBIE, A., CARR, I. M. & BONTHON, D. T. 2016. Enhanced diagnostic yield in Meckel-Gruber and Joubert syndrome through exome sequencing supplemented with split-read mapping. *BMC Med Genet*, 17, 1.
- WHATLEY, M., FRANCIS, A., NG, Z. Y., KHOH, X. E., ATLAS, M. D., DILLEY, R. J. & WONG, E. Y. M. 2020. Usher Syndrome: Genetics and Molecular Links of Hearing Loss and Directions for Therapy. *Front Genet*, 11, 565216.
- WHEWAY, G., MITCHISON, H. M. & GENOMICS ENGLAND RESEARCH, C. 2019. Opportunities and Challenges for Molecular Understanding of Ciliopathies-The 100,000 Genomes Project. *Front Genet*, 10, 127.
- WHEWAY, G., NAZLAMOVA, L. & HANCOCK, J. T. 2018. Signaling through the Primary Cilium. *Front Cell Dev Biol*, 6, 8.
- WHEWAY, G., PARRY, D. A. & JOHNSON, C. A. 2014. The role of primary cilia in the development and disease of the retina. *Organogenesis*, 10, 69-85.
- WHEWAY, G., SCHMIDTS, M., MANS, D. A., SZYMANSKA, K., NGUYEN, T. T., RACHER, H., PHELPS, I. G., TOEDT, G., KENNEDY, J., WUNDERLICH, K. A., SORUSCH, N., ABDELHAMED, Z. A., NATARAJAN, S., HERRIDGE, W., VAN REEUWIJK, J., HORN, N., BOLDT, K., PARRY, D. A., LETTEBOER, S. J. F., ROOSING, S., ADAMS, M., BELL, S. M., BOND, J., HIGGINS, J., MORRISON, E. E., TOMLINSON, D. C., SLAATS, G. G., VAN DAM, T. J. P., HUANG, L., KESSLER, K., GIESSL, A., LOGAN, C. V., BOYLE, E. A., SHENDURE, J., ANAZI, S., ALDAHMESH, M., AL HAZZAA, S., HEGELE, R. A., OBER, C., FROSK, P., MHANNI, A. A., CHODIRKER, B. N., CHUDLEY, A. E., LAMONT, R., BERNIER, F. P., BEAULIEU, C. L., GORDON, P., PON, R.

- T., DONAHUE, C., BARKOVICH, A. J., WOLF, L., TOOMES, C., THIEL, C. T., BOYCOTT, K. M., MCKIBBIN, M., INGLEHEARN, C. F., STEWART, F., OMRAN, H., HUYNEN, M. A., SERGOUNIOTIS, P. I., ALKURAYA, F. S., PARBOOSINGH, J. S., INNES, A. M., WILLOUGHBY, C. E., GILES, R. H., WEBSTER, A. R., UEFFING, M., BLACQUE, O., GLEESON, J. G., WOLFRUM, U., BEALES, P. L., GIBSON, T., DOHERTY, D., MITCHISON, H. M., ROEPMAN, R. & JOHNSON, C. A. 2015. An siRNA-based functional genomics screen for the identification of regulators of ciliogenesis and ciliopathy genes. *Nat Cell Biol*, 17, 1074-1087.
- WIIK, A. C., WADE, C., BIAGI, T., ROPSTAD, E. O., BJERKÅS, E., LINDBLAD-TOH, K. & LINGAAS, F. 2008. A deletion in nephronophthisis 4 (NPHP4) is associated with recessive cone-rod dystrophy in standard wire-haired dachshund. *Genome Res*, 18, 1415-21.
- WILLIAMS, C. L., LI, C., KIDA, K., INGLIS, P. N., MOHAN, S., SEMENEC, L., BIALAS, N. J., STUPAY, R. M., CHEN, N., BLACQUE, O. E., YODER, B. K. & LEROUX, M. R. 2011. MKS and NPHP modules cooperate to establish basal body/transition zone membrane associations and ciliary gate function during ciliogenesis. *J Cell Biol*, 192, 1023-41.
- WU, C., YANG, M., LI, J., WANG, C., CAO, T., TAO, K. & WANG, B. 2014. Talpid3-binding centrosomal protein Cep120 is required for centriole duplication and proliferation of cerebellar granule neuron progenitors. *PLoS One*, 9, e107943.
- XIAO, D., LV, C., ZHANG, Z., WU, M., ZHENG, X., YANG, L., LI, X., WU, G. & CHEN, J. 2017. Novel CC2D2A compound heterozygous mutations cause Joubert syndrome. *Mol Med Rep*, 15, 305-308.
- XIE, Z., MOY, L. Y., SANADA, K., ZHOU, Y., BUCHMAN, J. J. & TSAI, L. H. 2007. Cep120 and TACCs control interkinetic nuclear migration and the neural progenitor pool. *Neuron*, 56, 79-93.
- XU, C. & NG, D. T. 2015. Glycosylation-directed quality control of protein folding. *Nat Rev Mol Cell Biol*, 16, 742-52.
- YANDA, M. K., LIU, Q. & CEBOTARU, L. 2018. A potential strategy for reducing cysts in autosomal dominant polycystic kidney disease with a CFTR corrector. *J Biol Chem*, 293, 11513-11526.
- YATES, A. D., ACHUTHAN, P., AKANNI, W., ALLEN, J., ALLEN, J., ALVAREZ-JARRETA, J., AMODE, M. R., ARMEAN, I. M., AZOV, A. G., BENNETT, R., BHAI, J., BILLIS, K., BODDU, S., MARUGAN, J. C., CUMMINS, C., DAVIDSON, C., DODIYA, K., FATIMA, R., GALL, A., GIRON, C. G., GIL, L., GREGO, T., HAGGERTY, L., HASKELL, E., HOURLIER, T., IZUOGU, O. G., JANACEK, S. H., JUETTEMANN, T., KAY, M., LAVIDAS, I., LE, T., LEMOS, D., MARTINEZ, J. G., MAUREL, T., MCDOWALL, M., MCMAHON, A., MOHANAN, S., MOORE, B., NUHN, M., OHEH, D. N., PARKER, A., PARTON, A., PATRICIO, M., SAKTHIVEL, M. P., ABDUL SALAM, A. I., SCHMITT, B. M., SCHUILENBURG, H., SHEPPARD, D., SYCHEVA, M., SZUBA, M., TAYLOR, K., THORMANN, A., THREADGOLD, G., VULLO, A., WALTS, B., WINTERBOTTOM, A., ZADISSA, A., CHAKIACHVILI, M., FLINT, B., FRANKISH, A., HUNT, S. E., G, I. I., KOSTADIMA, M., LANGRIDGE, N., LOVELAND, J. E., MARTIN, F. J., MORALES, J., MUDGE, J. M., MUFFATO, M., PERRY, E., RUFFIER, M., TREVANION, S. J., CUNNINGHAM, F., HOWE, K. L., ZERBINO, D. R. & FLICEK, P. 2020. Ensembl 2020. *Nucleic Acids Res*, 48, D682-D688.
- ZENG, Z., HUANG, B., PARVEZ, R. K., LI, Y., CHEN, J., VONK, A. C., THORNTON, M. E., PATEL, T., RUTLEDGE, E. A., KIM, A. D., YU, J., GRUBBS, B. H., MCMAHON, J. A., PASTOR-SOLER, N. M., HALLOWS, K. R., MCMAHON, A.

- P. & LI, Z. 2021. Generation of patterned kidney organoids that recapitulate the adult kidney collecting duct system from expandable ureteric bud progenitors. *Nat Commun*, 12, 3641.
- ZHANG, W., TAYLOR, S. P., ENNIS, H. A., FORLENZA, K. N., DURAN, I., LI, B., SANCHEZ, J. A. O., NEVAREZ, L., NICKERSON, D. A., BAMSHAD, M., LACHMAN, R. S., KRAKOW, D. & COHN, D. H. 2018. Expanding the genetic architecture and phenotypic spectrum in the skeletal ciliopathies. *Hum Mutat*, 39, 152-166.
- ZHANG, W., YANG, S. L., YANG, M., HERRLINGER, S., SHAO, Q., COLLAR, J. L., FIERRO, E., SHI, Y., LIU, A., LU, H., HERRING, B. E., GUO, M. L., BUCH, S., ZHAO, Z., XU, J., LU, Z. & CHEN, J. F. 2019. Modeling microcephaly with cerebral organoids reveals a WDR62-CEP170-KIF2A pathway promoting cilium disassembly in neural progenitors. *Nat Commun*, 10, 2612.
- ZHANG, X., LIU, J., PAN, T., WARD, A., LIU, J. & XU, X. Z. S. 2022. A cilia-independent function of BBSome mediated by DLK-MAPK signaling in *C. elegans* photosensation. *Dev Cell*, 57, 1545-1557.e4.
- ZHOU, X., EDMONSON, M. N., WILKINSON, M. R., PATEL, A., WU, G., LIU, Y., LI, Y., ZHANG, Z., RUSCH, M. C., PARKER, M., BECKSFORT, J., DOWNING, J. R. & ZHANG, J. 2016. Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nat Genet*, 48, 4-6.
- ZIEGLER, W. H., LÜDIGER, S., HASSAN, F., GEORGIADIS, M. E., SWOLANA, K., KHERA, A., MERTENS, A., FRANKE, D., WOHLGEMUTH, K., DAHMER-HEATH, M., KÖNIG, J., DAFINGER, C., LIEBAU, M. C., CETINER, M., BERGMANN, C., SOETJE, B. & HAFFNER, D. 2022. Primary URECs: a source to better understand the pathology of renal tubular epithelia in pediatric hereditary cystic kidney diseases. *Orphanet J Rare Dis*, 17, 122.

Appendices

Appendix 1. Extract variants by coordinate script (modified from Genomics England)

```
#####
#####
#####
##### Extract variants in a cohort by consequence type
#####

# This script subsets VCFs by coordinate and consequence type and prints out sample ID and genotype

# INPUT:
# 1: Tab-delimited list of chromosomal coordinates (regions.txt)
# 2. A list of full-paths to VCF files of interest (vcf_list.txt)

# OUTPUT:
# 1: A tab-delimited file of sample ID, variant information and quality , this includes indels, ie: variants with a DPI score, instead of DP (results.final.txt)
# Intermediate files are generated, if want to be removed use: rm results_no.indels.txt
# results_indels.only.txt results_temporary.file.txt

# STEPS:
# 1: tabix to subset each VCF by the coordinates given
# 2: bcftools to split multi-allelic variants across multiple lines
# 3: bcftools to filter for PASS variants and variants with a coverage >10 and genotype quality >20
# 4: bcftools to print out the sample ID, and variant and quality information as a text file

# IMPORTANT!
# 1: Make sure the chromosome nomenclature is correct for GRCh37 and GRCh38
# 2: Do not mix genome builds in VCF file list
# 3: Change the file paths to suit your working environment
# 4: This script will extract SNVs and indels , this is because we use MIN(FMT/DP)>10 and next sepately MIN(FMT/DPI)>10
# 5: This script will only be function with Illumina germine variant calls - please use the relevent script to process somatic variants
```

```
#!/bin/bash

module load bcftools/1.9

while read -r vcf; do
    tabix -h $vcf -R regions.txt | \
    bcftools norm -m -any | \
    bcftools view -f PASS -i 'MIN(FMT/DP)>10 & MIN(FMT/GQ)>20' | \
    bcftools query -f
'[%SAMPLE]\t%CHROM\t%POS\t%ID\t%REF\t%ALT\t%QUAL\t%FILTER\t[%GT]\t[%GQ]\t[%DP]\t%INFO/CSQ
T\n' | \
    grep /1 >> results_no.indels.txt ;
done < vcf_list.txt

sed -i '1s/^/SAMPLE\tCHROM\tPOS\tID\tREF\tALT\tQUAL\tFILTER\tGT\tGQ\tDP\tCSQT\n/' \
results_no.indels.txt

while read -r vcf; do
    tabix -h $vcf -R regions.txt | \
    bcftools norm -m -any | \
    bcftools view -f PASS -i 'MIN(FMT/DPI)>10 & MIN(FMT/GQ)>20' | \
    bcftools query -f
'[%SAMPLE]\t%CHROM\t%POS\t%ID\t%REF\t%ALT\t%QUAL\t%FILTER\t[%GT]\t[%GQ]\t[%DP]\t%INFO/CSQ
T\n' | \
    grep /1 >> results_indels.only.txt ;
done < vcf_list.txt

sed -i '1s/^/SAMPLE\tCHROM\tPOS\tID\tREF\tALT\tQUAL\tFILTER\tGT\tGQ\tDP\tCSQT\n/' \
results_indels.only.txt

cat results_no.indels.txt results_indels.only.txt > results_temporary.file.txt
sort results_temporary.file.txt | uniq > results.final.txt
#####
#####
```

This is a modified script from the extract variants by coordinate script of Genomics England (<https://cnfl.extge.co.uk/display/GERE/Extract+variants+by+coordinate>).

Appendix 2. List of genes associated with ciliopathies (Gold Standard SYSCILIA)

Ensembl Gene ID	Gene name	Ensembl Gene ID	Gene name	Ensembl Gene ID	Gene name	Ensembl Gene ID	Gene name
ENSG00000 138031	ADCY3	ENSG00000 169676	DRD5	ENSG00000 117650	NEK2	ENSG00000 144451	SPAG16
ENSG00000 135541	AHI1	ENSG00000 107404	DVL1	ENSG00000 114904	NEK4	ENSG00000 155761	SPAG17
ENSG00000 140057	AK7	ENSG00000 187240	DYNC2H1	ENSG00000 160602	NEK8	ENSG00000 077327	SPAG6
ENSG00000 165695	AK8	ENSG00000 146425	DYNL1	ENSG00000 064300	NGFR	ENSG00000 042317	SPATA7
ENSG00000 116127	ALMS1	ENSG00000 256061	DYX1C1	ENSG00000 100503	NIN	ENSG00000 152582	SPEF2
ENSG00000 168374	ARF4	ENSG00000 096093	EFHC1	ENSG00000 101004	NINL	ENSG00000 176101	SSNA1
ENSG00000 169379	ARL13B	ENSG00000 072840	EVC	ENSG00000 112981	NME5	ENSG00000 183473	SSTR3
ENSG00000 138175	ARL3	ENSG00000 173040	EVC2	ENSG00000 143156	NME7	ENSG00000 123473	STIL
ENSG00000 113966	ARL6	ENSG00000 180104	EXOC3	ENSG00000 086288	NME8	ENSG00000 163482	STK36
ENSG00000 153317	ASAP1	ENSG00000 131558	EXOC4	ENSG00000 214513	NOTO	ENSG00000 211455	STK38L
ENSG00000 130638	ATXN10	ENSG00000 070367	EXOC5	ENSG00000 144061	NPHP1	ENSG00000 133115	STOML3
ENSG00000 141577	AZI1	ENSG00000 138190	EXOC6	ENSG00000 113971	NPHP3	ENSG00000 166900	STX3
ENSG00000 108641	B9D1	ENSG00000 144036	EXOC6B	ENSG00000 131697	NPHP4	ENSG00000 107882	SUFU
ENSG00000 123810	B9D2	ENSG00000 170264	FAM161A	ENSG00000 126883	NUP214	ENSG00000 054654	SYNE2
ENSG00000 174483	BBS1	ENSG00000 188878	FBF1	ENSG00000 163002	NUP35	ENSG00000 111490	TBC1D30
ENSG00000 179941	BBS10	ENSG00000 196924	FLNA	ENSG00000 075188	NUP37	ENSG00000 145979	TBC1D7
ENSG00000 181004	BBS12	ENSG00000 133393	FOPN1	ENSG00000 213024	NUP62	ENSG00000 204852	TCTN1
ENSG00000 125124	BBS2	ENSG00000 129654	FOXJ1	ENSG00000 102900	NUP93	ENSG00000 168778	TCTN2
ENSG00000 140463	BBS4	ENSG00000 010361	FUZ	ENSG00000 122126	OCRL	ENSG00000 119977	TCTN3
ENSG00000 163093	BBS5	ENSG00000 141013	GAS8	ENSG00000 136811	ODF2	ENSG00000 092850	TEKT2
ENSG00000 138686	BBS7	ENSG00000 111087	GLI1	ENSG00000 046651	OFD1	ENSG00000 163060	TEKT4
ENSG00000 122507	BBS9	ENSG00000 074047	GLI2	ENSG00000 085840	ORC1	ENSG00000 153060	TEKT5
ENSG00000 160226	C21orf2	ENSG00000 106571	GLI3	ENSG00000 112530	PACRG	ENSG00000 149483	TMEM138
ENSG00000 168014	C2CD3	ENSG00000 126603	GLIS2	ENSG00000 007168	PAFAH1B1	ENSG00000 187049	TMEM216
ENSG00000 179270	C2orf71	ENSG00000 143147	GPR161	ENSG00000 148498	PARD3	ENSG00000 205084	TMEM231
ENSG00000 156172	C8orf37	ENSG00000 164199	GPR98	ENSG00000 102981	PARD6A	ENSG00000 155755	TMEM237
ENSG00000 100211	CBY1	ENSG00000 082701	GSK3B	ENSG00000 150275	PCDH15	ENSG00000 164953	TMEM67
ENSG00000 048342	CC2D2A	ENSG00000 173805	HAP1	ENSG00000 078674	PCM1	ENSG00000 083312	TNPO1

ENSG00000 167131	CCDC 103	ENSG00000 164818	HEAT R2	ENSG00000 156973	PDE6D	ENSG00000 197579	TOPO RS
ENSG00000 105479	CCDC 114	ENSG00000 108753	HNF1 B	ENSG00000 186862	PDZD7	ENSG00000 179636	TPPP2
ENSG00000 157856	CCDC 164	ENSG00000 109971	HSPA 8	ENSG00000 077684	PHF17	ENSG00000 204104	TRAF3 IP1
ENSG00000 160050	CCDC 28B	ENSG00000 081870	HSPB 11	ENSG00000 083535	PIBF1	ENSG00000 160218	TRAPP C10
ENSG00000 163885	CCDC 37	ENSG00000 158748	HTR6	ENSG00000 008710	PKD1	ENSG00000 054116	TRAPP C3
ENSG00000 145075	CCDC 39	ENSG00000 197386	HTT	ENSG00000 158683	PKD1L 1	ENSG00000 167632	TRAPP C9
ENSG00000 141519	CCDC 40	ENSG00000 157423	HYDI N	ENSG00000 118762	PKD2	ENSG00000 119401	TRIM3 2
ENSG00000 173588	CCDC 41	ENSG00000 198331	HYLS 1	ENSG00000 170927	PKHD1	ENSG00000 100815	TRIP11
ENSG00000 103540	CCP1 10	ENSG00000 163913	IFT12 2	ENSG00000 166851	PLK1	ENSG00000 128881	TTBK2
ENSG00000 107736	CDH2 3	ENSG00000 187535	IFT14 0	ENSG00000 164087	POC1A	ENSG00000 149292	TTC12
ENSG00000 151849	CENP J	ENSG00000 138002	IFT17 2	ENSG00000 185920	PTCH1	ENSG00000 123607	TTC21 B
ENSG00000 116198	CEP1 04	ENSG00000 109083	IFT20	ENSG00000 158079	PTPDC 1	ENSG00000 105948	TTC26
ENSG00000 174799	CEP1 35	ENSG00000 100360	IFT27	ENSG00000 103769	RAB11 A	ENSG00000 137473	TTC29
ENSG00000 110274	CEP1 64	ENSG00000 119650	IFT43	ENSG00000 090565	RAB11 FIP3	ENSG00000 197557	TTC30 A
ENSG00000 126001	CEP2 50	ENSG00000 118096	IFT46	ENSG00000 124839	RAB17	ENSG00000 196659	TTC30 B
ENSG00000 198707	CEP2 90	ENSG00000 101052	IFT52	ENSG00000 112210	RAB23	ENSG00000 165533	TTC8
ENSG00000 106477	CEP4 1	ENSG00000 114446	IFT57	ENSG00000 127328	RAB3I P	ENSG00000 112742	TTK
ENSG00000 112877	CEP7 2	ENSG00000 096872	IFT74	ENSG00000 167461	RAB8A	ENSG00000 214021	TTLL3
ENSG00000 121289	CEP8 9	ENSG00000 068885	IFT80	ENSG00000 128581	RABL5	ENSG00000 170703	TTLL6
ENSG00000 182504	CEP9 7	ENSG00000 122970	IFT81	ENSG00000 132341	RAN	ENSG00000 131044	TTLL9
ENSG00000 165376	CLDN 2	ENSG00000 032742	IFT88	ENSG00000 099901	RANBP 1	ENSG00000 167552	TUBA1 A
ENSG00000 103351	CLUA P1	ENSG00000 148384	INPP5 E	ENSG00000 080298	RFX3	ENSG00000 167553	TUBA1 C
ENSG00000 183862	CNGA 2	ENSG00000 164066	INTU	ENSG00000 188026	RILPL1	ENSG00000 127824	TUBA4 A
ENSG00000 132259	CNGA 4	ENSG00000 119509	INVS	ENSG00000 150977	RILPL2	ENSG00000 137267	TUBB2 A
ENSG00000 070729	CNGB 1	ENSG00000 173226	IQCB1	ENSG00000 145491	ROPN1 L	ENSG00000 137285	TUBB2 B
ENSG00000 130545	CRB3	ENSG00000 117245	KIF17	ENSG00000 104237	RP1	ENSG00000 198211	TUBB3
ENSG00000 058453	CROC C	ENSG00000 196169	KIF19	ENSG00000 102218	RP2	ENSG00000 074935	TUBE1
ENSG00000 168036	CTNN B1	ENSG00000 186638	KIF24	ENSG00000 156313	RPGR	ENSG00000 130640	TUBG CP2
ENSG00000 146038	DCDC 2	ENSG00000 165115	KIF27	ENSG00000 092200	RPGR P1	ENSG00000 126216	TUBG CP3
ENSG00000 095397	DFNB 31	ENSG00000 131437	KIF3A	ENSG00000 103494	RPGR P1L	ENSG00000 137822	TUBG CP4

ENSG00000 162946	<i>DISC1</i>	ENSG00000 101350	<i>KIF3B</i>	ENSG00000 160188	<i>RSPH1</i>	ENSG00000 153575	<i>TUBG</i> <i>CP5</i>
ENSG00000 154099	<i>DNAA F1</i>	ENSG00000 084731	<i>KIF3C</i>	ENSG00000 130363	<i>RSPH3</i>	ENSG00000 128159	<i>TUBG</i> <i>CP6</i>
ENSG00000 165506	<i>DNAA F2</i>	ENSG00000 166813	<i>KIF7</i>	ENSG00000 111834	<i>RSPH4</i> A	ENSG00000 112041	<i>TULP1</i>
ENSG00000 167646	<i>DNAA F3</i>	ENSG00000 135338	<i>LCA5</i>	ENSG00000 172426	<i>RSPH9</i>	ENSG00000 078246	<i>TULP3</i>
ENSG00000 114841	<i>DNAH 1</i>	ENSG00000 129295	<i>LRRC</i> 6	ENSG00000 176225	<i>RTTN</i>	ENSG00000 168038	<i>ULK4</i>
ENSG00000 197653	<i>DNAH 10</i>	ENSG00000 163818	<i>LZTFL</i> 1	ENSG00000 156876	<i>SASS6</i>	ENSG00000 006611	<i>USH1C</i>
ENSG00000 105877	<i>DNAH 11</i>	ENSG00000 111837	<i>MAK</i>	ENSG00000 151466	<i>SCLT1</i>	ENSG00000 182040	<i>USH1</i> G
ENSG00000 183914	<i>DNAH 2</i>	ENSG00000 172005	<i>MAL</i>	ENSG00000 054282	<i>SDCC</i> AG8	ENSG00000 042781	<i>USH2A</i>
ENSG00000 039139	<i>DNAH 5</i>	ENSG00000 101367	<i>MAPR</i> E1	ENSG00000 168385	<i>Sep-02</i>	ENSG00000 078668	<i>VDAC3</i>
ENSG00000 115423	<i>DNAH 6</i>	ENSG00000 128285	<i>MCHR</i> 1	ENSG00000 122545	<i>Sep-07</i>	ENSG00000 134086	<i>VHL</i>
ENSG00000 122735	<i>DNAI1</i>	ENSG00000 111554	<i>MDM1</i>	ENSG00000 185900	<i>SGK19</i> 6	ENSG00000 143951	<i>WDPC</i> P
ENSG00000 171595	<i>DNAI2</i>	ENSG00000 125863	<i>MKKS</i>	ENSG00000 164690	<i>SHH</i>	ENSG00000 157796	<i>WDR1</i> 9
ENSG00000 119661	<i>DNAL 1</i>	ENSG00000 011143	<i>MKS1</i>	ENSG00000 180638	<i>SLC47</i> A2	ENSG00000 118965	<i>WDR3</i> 5
ENSG00000 163879	<i>DNALI 1</i>	ENSG00000 178053	<i>MLF1</i>	ENSG00000 128602	<i>SMO</i>	ENSG00000 126870	<i>WDR6</i> 0
ENSG00000 166171	<i>DPCD</i>	ENSG00000 138587	<i>MNS1</i>	ENSG00000 132639	<i>SNAP2</i> 5	ENSG00000 152763	<i>WDR7</i> 8
ENSG00000 092964	<i>DPYS L2</i>	ENSG00000 091536	<i>MYO1</i> 5A	ENSG00000 086300	<i>SNX10</i>	ENSG00000 196236	<i>XPNPE</i> P3
ENSG00000 184845	<i>DRD1</i>	ENSG00000 137474	<i>MYO7</i> A	ENSG00000 064199	<i>SPA17</i>	ENSG00000 102935	<i>ZNF42</i> 3
ENSG00000 149295	<i>DRD2</i>	ENSG00000 137601	<i>NEK1</i>				

This list of 302 genes ciliary genes corresponds to the Syscilia Gold Standard list of ciliary genes (obtained from: <http://bioinformatics.bio.uu.nl/john/syscilia/ciliacarta/>) and their corresponding Ensembl Gene ID.

Appendix 3. List of terms used to select case population

Description of each of terms used to select the case population		
Genomics England Rare Disease Cystic kidney disease	ICD10 Q61.5: Medullary cystic kidney	SNOMED 426978005: Endoscopic de-roofing of multiple cysts of kidney
Genomics England Rare Disease GMS R193 Cystic renal disease	ICD10 Q61.8: Other cystic kidney diseases	SNOMED 431909007: Counselling for polycystic kidney disease
HPO HP:0000003: Multicystic kidney dysplasia	ICD10 Q61.9: Cystic kidney disease, unspecified	SNOMED 53670009: Evacuation of kidney cyst
HPO HP:0000107: Renal cyst	SNOMED 187144000: Hydatid cyst of kidney	SNOMED 5941000119101: Congenital single renal cyst
HPO HP:0000108: Renal corticomedullary cysts	SNOMED 197801000: Single acquired kidney cyst	SNOMED 64119005: Marsupialization of cyst of kidney
HPO HP:0000113: Polycystic kidney dysplasia	SNOMED 204957003: Medullary cystic disease of the kidney	SNOMED 71064009: Acquired polycystic kidney disease
HPO HP:0000800: Cystic renal dysplasia	SNOMED 236376009: Infected renal cyst	SNOMED 712538004: Fluoroscopy guided radiofrequency ablation of renal cyst
HPO HP:0000803: Renal cortical cysts	SNOMED 105999006: Acquired renal cystic disease	SNOMED 712539007: Fluoroscopy guided radiofrequency ablation of cyst of liver
HPO HP:0001407: Hepatic cysts	SNOMED 236439005: Cystic disease of kidney	SNOMED 716196007: Isolated polycystic liver disease
HPO HP:0004734: Renal cortical microcysts	SNOMED 251096000: Unroofing of cyst of kidney	SNOMED 717749002: Bilateral multicystic renal dysplasia
HPO HP:0005209: Intrahepatic bile duct cysts	SNOMED 252053007: Excision of cyst of kidney	SNOMED 722223000: Cyst of kidney
HPO HP:0005562: Multiple renal cysts	SNOMED 253878003: Adult type polycystic kidney disease type 1	SNOMED 72925005: Congenital cystic disease of liver
HPO HP:0006557: Polycystic liver disease	SNOMED 253879006: Adult type polycystic kidney disease type 2	SNOMED 735471001: Complex renal cyst
HPO HP:0006706: Cystic liver disease	SNOMED 253880009: Autosomal dominant polycystic kidney disease in childhood	SNOMED 737562008: Multicystic renal dysplasia

HPO HP:0008659: Multiple small medullary renal cysts	SNOMED 253881008: Cortical cystic disease	SNOMED 765330003: Autosomal dominant polycystic kidney disease
HPO HP:0012581: Solitary renal cyst	SNOMED 253882001: Nephronophthisis - medullary cystic disease	SNOMED 765331004: Autosomal dominant polycystic kidney disease type 1 with tuberous sclerosis
HPO HP:0100611: Multiple glomerular cysts	SNOMED 253883006: Multiple renal cysts	SNOMED 77945009: Simple renal cyst
HPO HP:0100890: Cyst of the ductus choledochal cyst	SNOMED 270910001: Medullary cystic disease, adult type	SNOMED 82525005: Multiple congenital cysts of kidney
ICD10 N28.1: Cyst of kidney	SNOMED 287496005: Diagnostic aspiration of liver cyst or abscess	SNOMED 85057007: Liver cyst
ICD10 Q44.6: Cystic disease of liver	SNOMED 28770003: Polycystic kidney disease, infantile type	SNOMED 86127008: Excision of hydatid cyst of liver
ICD10 Q61: Cystic kidney disease	SNOMED 307620003: Marsupialization of liver cyst	SNOMED 86463003: Solitary multilocular renal cyst
ICD10 Q61.0: Congenital single renal cyst	SNOMED 307621004: Laparoscopic marsupialization of liver cyst	SNOMED 87421006: Evacuation of cyst of liver
ICD10 Q61.1: Polycystic kidney, autosomal recessive	SNOMED 369071000119105: Congenital renal cyst	SNOMED 98371000119102: Acquired complex renal cyst
ICD10 Q61.2: Polycystic kidney, autosomal dominant	SNOMED 37184006: Enucleation of cyst of liver	SNOMED 432582008: Drainage of cyst of kidney using ultrasound guidance
ICD10 Q61.3: Polycystic kidney, unspecified	SNOMED 373599008: Multiple acquired kidney cysts	

These terms were used to create the case population (1922 patients) from Labkey and Participant Explorer tools from the Genomics England 100,000 Genomes Project.

Appendix 4. Distribution of the patients selected for the case population (n=1922), in terms of their diagnosed disease

Disease	Number of participants	%
Cystic kidney disease	1245	64.78
CAKUT	174	9.05
Unexplained kidney failure in young people	53	2.76
Renal tract calcification (or Nephrolithiasis or nephrocalcinosis)	42	2.19
Intellectual disability	37	1.93
Ductal plate malformation	34	1.77
Ultra-rare undescribed monogenic disorders	22	1.14
Classical tuberous sclerosis	16	0.83
Epilepsy plus other features	14	0.73
Hypertrophic Cardiomyopathy	13	0.68
Primary immunodeficiency	13	0.68
Familial Thoracic Aortic Aneurysm Disease	11	0.57
Rod-cone dystrophy	11	0.57
Familial haematuria	10	0.52
Hereditary spastic paraparesis	10	0.52
Proteinuric renal disease	10	0.52
Others	229	11.91

Only the diseases with at least 10 patients are shown in this list. Diseases in which less than 10 patients are diagnosed with are in grouped in the category "Others". There are 20 participants diagnosed with more than one disease (2 participant with 3 diseases and 18 participants with 2 diseases).

Appendix 5. Distribution of the patients selected for the control population (n=16773), in terms of their diagnosed disease

Disease	Number of participants	%
Intellectual disability	5610	33.45
Epilepsy plus other features	1361	8.11
Ultra-rare undescribed monogenic disorders	1215	7.24
Hypertrophic Cardiomyopathy	687	4.10
Primary immunodeficiency	559	3.33
Early onset and familial Parkinson's Disease	500	2.98
Familial breast and or ovarian cancer	481	2.87
Familial Hypercholesterolaemia	460	2.74
Epileptic encephalopathy	327	1.95
Dilated Cardiomyopathy	294	1.75
Mitochondrial disorders	276	1.65
Syndromic congenital heart disease	268	1.60
Multiple Tumours	231	1.38
Exceptionally young adult onset cancer	213	1.27
Familial congenital heart disease	192	1.14
Complex Parkinsonism (includes pallido-pyramidal syndromes)	190	1.13
Brugada syndrome	164	0.98
Early onset dementia	159	0.95
Familial Genetic Generalised Epilepsies	123	0.73
Hereditary haemorrhagic telangiectasia	122	0.73
Multiple bowel polyps	120	0.72
Classical Ehlers-Danlos Syndrome	112	0.67
Inherited bleeding and or platelet disorders	111	0.66
Arrhythmogenic Right Ventricular Cardiomyopathy	101	0.60
Anophthalmia or microphthalmia	95	0.57
Primary lymphoedema	94	0.56
Long QT syndrome	92	0.55
Lipoedema disease	82	0.49
Multiple endocrine tumours	81	0.48
Neurofibromatosis Type 1	81	0.48
Ketotic hypoglycaemia	75	0.45
Glaucoma (developmental)	74	0.44
Cytopenia and pancytopenia	73	0.44
Ectodermal dysplasia without a known gene mutation	73	0.44
Periodic fever syndromes and amyloidosis	73	0.44
Silver Russell syndrome	73	0.44

Disorders of sex development	71	0.42
Severe multi-system atopic disease with high IgE	66	0.39
Congenital anaemias	64	0.38
Familial colon cancer	61	0.36
Familial exudative vitreoretinopathy	59	0.35
Familial Focal Epilepsies	57	0.34
Palmoplantar keratoderma and erythrokeratodermas	53	0.32
Idiopathic hypogonadotropic hypogonadism	52	0.31
Noonan syndrome	52	0.31
Primary Microcephaly - Microcephalic Dwarfism Spectrum	52	0.31
Kyphoscoliotic Ehlers-Danlos syndrome	51	0.30
Meige disease	51	0.30
Congenital hypothyroidism	50	0.30
Others	1383	8.25

Only the diseases with at least 50 patients are shown in this list. Diseases in which less than 10 patients are diagnosed with are in grouped in the category "Others". There are 184 participants diagnosed with more than one disease (7 participants with 3 diseases and 177 participants with 2 diseases).

Appendix 6. Genomic position of the loss-of-function (LoF) variants found in cases in the genes *ALG1*, *ALG8*, *ALG9* and *ALG12*

CHROM	POS	Rs ID	REF	ALT	Gene
chr16	5077952	.	AC	A	<i>ALG1</i>
chr16	5073230	.	G	T	<i>ALG1</i>
chr16	5081043	.	C	A	<i>ALG1</i>
chr11	78104032		G	GT	<i>ALG8</i>
chr11	78106943	rs963287770	AG	A	<i>ALG8</i>
chr11	78106895	rs376161880	G	A	<i>ALG8</i>
chr11	78127411	rs200888240	G	A	<i>ALG8</i>
chr11	111837467	.	CT	C	<i>ALG9</i>
chr11	111844592	.	GCCACTCACCATGAAA	GGTAT	<i>ALG9</i>
chr11	111837577	rs782775735	G	A	<i>ALG9</i>
chr22	49904497	rs759244819	GT	G	<i>ALG12</i>
chr22	49909280	.	C	T	<i>ALG12</i>
chr22	49909954	rs547640376	G	A	<i>ALG12</i>

LoF = stop, frameshift, splice donor and splice acceptor. Some of these variants can be found in several participants. There are not participants with more than one LoF variant in the same *ALG* gene. There are not participants with LoF variants in more than one *ALG* gene. For the genomic position the GRCh38/hg38 human reference genome assembly was considered.

Appendix 7. Genomic position of the loss-of-function (LoF) variants found in controls in the genes *ALG1*, *ALG8*, *ALG9* and *ALG12*

CHR	POS	Rs ID	REF	ALT	Gene
chr16	5080944	.	A	G	<i>ALG1</i>
chr16	5075537	rs752189808	G	A	<i>ALG1</i>
chr16	5079808	rs373355236	G	C	<i>ALG1</i>
chr16	5079087	.	CTGGCAGCTTAGAAAGTAGGTGTG	C	<i>ALG1</i>
chr16	5084828	rs1047747	C	T	<i>ALG1</i>
chr16	5077503	.	C	T	<i>ALG1</i>
chr16	5080999	.	C	T	<i>ALG1</i>
chr11	78101043	.	AC	A	<i>ALG8</i>
chr11	78104032	.	G	GT	<i>ALG8</i>
chr11	78104413	.	GA	G	<i>ALG8</i>
chr11	78127438	.	T	C	<i>ALG8</i>
chr11	78109441	.	C	G	<i>ALG8</i>
chr11	78121064	rs139832787	C	T	<i>ALG8</i>
chr11	78104001	.	G	T	<i>ALG8</i>
chr11	78106895	rs376161880	G	A	<i>ALG8</i>
chr11	78113978	rs533704173	G	A	<i>ALG8</i>
chr11	78119193	rs762811727	G	A	<i>ALG8</i>
chr11	78127411	rs200888240	G	A	<i>ALG8</i>
chr11	111868609	.	GT	G	<i>ALG9</i>
chr11	111837467	.	CT	C	<i>ALG9</i>
chr11	111853733	.	TA	T	<i>ALG9</i>
chr11	111871351	.	C	A	<i>ALG9</i>
chr11	111809748	.	A	C	<i>ALG9</i>
chr11	111836233	.	C	A	<i>ALG9</i>
chr11	111870255	.	C	A	<i>ALG9</i>
chr11	111865230	.	G	A	<i>ALG9</i>
chr22	49904057	rs746988876	CTT	C	<i>ALG12</i>
chr22	49904342	rs770711819	T	TG	<i>ALG12</i>
chr22	49904351	.	AC	A	<i>ALG12</i>
chr22	49904067	.	C	G	<i>ALG12</i>
chr22	49907946	.	T	C	<i>ALG12</i>
chr22	49913519	.	T	G	<i>ALG12</i>
chr22	49913384	rs376314741	C	T	<i>ALG12</i>
chr22	49909954	rs547640376	G	A	<i>ALG12</i>

LoF = stop, frameshift, splice donor and splice acceptor. Some of these variants can be found in several participants. There are not participants with more than one LoF variant in the same *ALG* gene. There are not participants with LoF variants in more than one *ALG* gene. For the genomic position the GRCh38/hg38 human reference genome assembly was considered.

Appendix 8. List of publications

First author articles:

BARROSO-GIL, M., OLINGER, E., RAMSBOTTOM, S. A., MOLINARI, E., MILES, C. G. & SAYER, J. A. 2021a. Update of genetic variants in CEP120 and CC2D2A-With an emphasis on genotype-phenotype correlations, tissue specific transcripts and exploring mutation specific exon skipping therapies. *Mol Genet Genomic Med*, 9, e1603.

POWELL, L., BARROSO-GIL, M., CLOWRY, G. J., DEVLIN, L. A., MOLINARI, E., RAMSBOTTOM, S. A., MILES, C. G. & SAYER, J. A. 2020. Expression patterns of ciliopathy genes ARL3 and CEP120 reveal roles in multisystem development. *BMC Dev Biol*, 20, 26.

Review article:

BARROSO-GIL, M., OLINGER, E. & SAYER, J. A. 2021. Molecular genetics of renal ciliopathies. *Biochem Soc Trans*, 49, 1205-1220.

Comment:

BARROSO-GIL, M., POWELL, L. & SAYER, J. A. 2020. RE: Clinical and Molecular Diagnosis of Joubert Syndrome and Related Disorders. *Pediatr Neurol*, 112, 10.

Contributions in articles as a co-author:

AL ALAWI, I., AL RIYAMI, M., BARROSO-GIL, M., POWELL, L., OLINGER, E., AL SALMI, I. & SAYER, J. A. 2021. The diagnostic yield of whole exome sequencing as a first approach in consanguineous Omani renal ciliopathy syndrome patients. *F1000Res*, 10, 207.

DEVLIN, L. A., COLES, J., JACKSON, C. L., BARROSO-GIL, M., GREEN, B., WALKER, W. T., THOMAS, N. S., THOMPSON, J., ROCK, S. A., NEATU, R., POWELL, L., MOLINARI, E., WILSON, I., CORDELL, H. J., OLINGER, E., MILES, C. G., SAYER, J. A., WHEWAY, G. & LUCAS, J. S. 2022. Biallelic variants in CEP164 cause a motile ciliopathy-like syndrome. *Clin Genet*, 103(3), 330-334.

GERAGHTY, R. M., ORR, S., OLINGER, E., NEATU, R., BARROSO-GIL, M., MABILLARD, H., CONSORTIUM, G. E. R., WILSON, I. & SAYER, J. A. 2023. Use of whole genome sequencing to determine the genetic basis of visceral myopathies including Prune Belly syndrome. *J Rare Dis (Berlin)*, 2, 9.

HUYNH, V. T., AUDRÉZET, M. P., SAYER, J. A., ONG, A. C., LEFEVRE, S., LE BRUN, V., DESPRÉS, A., SENUM, S. R., CHEBIB, F. T., BARROSO-GIL, M., PATEL, C., MALLETT, A. J., GOEL, H., MALLAWAARACHCHI, A. C., VAN EERDE, A. M., PONLOT, E., KRIBS, M., LE MEUR, Y., HARRIS, P. C. & CORNEC-LE GALL, E. 2020b. Clinical spectrum, prognosis and estimated prevalence of DNAJB11-kidney disease. *Kidney Int*, 98, 476-487.

LATOUR, B. L., VAN DE WEGHE, J. C., RUSTERHOLZ, T. D., LETTEBOER, S. J., GOMEZ, A., SHAHEEN, R., GESEMANN, M., KARAMZADE, A., ASADOLLAHI, M., BARROSO-GIL, M., CHITRE, M., GROUT, M. E., VAN REEUWIJK, J., VAN BEERSUM, S. E., MILLER, C. V., DEMPSEY, J. C., MORSY, H., BAMSHAD, M. J., NICKERSON, D. A., NEUHAUSS, S. C., BOLDT, K., UEFFING, M., KERAMATIPOUR, M., SAYER, J. A., ALKURAYA, F. S., BACHMANN-GAGESCU, R., ROEPMAN, R. & DOHERTY, D. 2020. Dysfunction of the ciliary

ARMC9/TOGARAM1 protein module causes Joubert syndrome. *J Clin Invest*, 130, 4423-4439.

LEMOINE, H., RAUD, L., FOULQUIER, F., SAYER, J. A., LAMBERT, B., OLINGER, E., LEFÈVRE, S., KNEBELMANN, B., HARRIS, P. C., TROUVÉ, P., DESPRÈS, A., DUNEAU, G., MATIGNON, M., POYET, A., JOURDE-CHICHE, N., GUERROT, D., LEMOINE, S., SERET, G., BARROSO-GIL, M., BINGHAM, C., GILBERT, R., LE MEUR, Y., AUDRÉZET, M. P. & CORNEC-LE GALL, E. 2022. Monoallelic pathogenic ALG5 variants cause atypical polycystic kidney disease and interstitial fibrosis. *Am J Hum Genet*, 109, 1484-1499.

OLINGER, E., ALAWI, I. A., AL RIYAMI, M. S., SALMI, I. A., MOLINARI, E., FAQEIH, E. A., AL-HAMED, M. H., BARROSO-GIL, M., POWELL, L., AL-HUSSAINI, A. A., RAHIM, K. A., ALMONTASHIRI, N. A. M., MILES, C., SHRIL, S., HILDEBRANDT, F., CONSORTIUM, G. E. R., WILSON, I. J. & SAYER, J. A. 2021. A discarded synonymous variant in NPHP3 explains nephronophthisis and congenital hepatic fibrosis in several families. *Hum Mutat*, 42, 1221-1228.

ORR, S., OLINGER, E., IOSIFIDOU, S., BARROSO-GIL, M., NEATU, R., WOOD, K., WILSON, I. & SAYER, J. A. 2023. Molecular genetic diagnosis of kidney ciliopathies: Lessons from interpreting genomic sequencing data and the requirement for accurate phenotypic data. *Ann Hum Genet*, 00, 1–10.

POWELL, L., OLINGER, E., WEDDERBURN, S., RAMAKUMARAN, V. S., KINI, U., CLAYTON-SMITH, J., RAMSDEN, S. C., RICE, S. J., BARROSO-GIL, M., WILSON, I., COWLEY, L., JOHNSON, S., HARRIS, E., MONTGOMERY, T., BERTOLI, M., BOLTSHAUSER, E. & SAYER, J. A. 2021. Identification of LAMA1 mutations ends diagnostic odyssey and has prognostic implications for patients with presumed Joubert syndrome. *Brain Commun*, 3, fcab163.