



**The significance of clinico-
pathological and molecular sub-
groups in Malignant Rhabdoid
Tumours**

Yura Grabovska

Thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

Newcastle University

Faculty of Medical Sciences

Wolfson Childhood Cancer Research Centre

Northern Institute for Cancer Research

September 2019

Abstract

Introduction Malignant Rhabdoid Tumours (MRT) are aggressive early childhood tumours characterised by biallelic inactivation of *SMARCB1*. Having the potential to arise in an array of distinct tissues (CNS-located atypical teratoid rhabdoid tumours, ATRT; extra-cranial rhabdoid tumours, ECRT) they are often treated as distinct entities therapeutically and in clinical/biological studies although emerging sub-groups of MRT have provided new understanding of the disease heterogeneity. Lack of consensus on sub-group number and biology can be seen as a hurdle to future studies.

Methods Gene expression and methylation array profiling of primary MRT was performed on clinico-pathologically annotated tumour profiles from UK cancer centres and combined with published MRT data in a meta-analysis. To characterise the common biological features of MRT, regardless of location, differential expression, methylation, gene and pathway analyses were compared to other paediatric embryonal tumour expression and methylation profiles (i.e. Medulloblastoma, Ewings Sarcoma, Rhabdomyosarcoma, Wilms tumour and Neuroblastoma). Survival analysis was carried out on UK MRT samples to identify novel DNA methylation markers associated with disease outcome. Based on evidence suggesting immune system involvement in at least one MRT subgroup, “methylCibersort” a novel in-silico method was used to estimate immune cell infiltration in a large cohort of pan-CNS tumours including MRT.

Results Clustering all MRT together recapitulates the subgroups observed in ATRT alone; broadly overlapping with recently published ATRT and ECRT subgroup models. A putative expanded subgrouping model encompassing all MRT highlights additional heterogeneity and defines novel subgroup characteristics. Subgroup differences were shown to better explain differences in MRT biology than tumour location alone. Survival analysis identified a number of novel survival associations with DNA methylation state. Immune infiltration estimation using methylCibersort identified differences in immune interactions across a large dataset of different CNS tumours, and presented novel prognostic feature.

Conclusion MRT is a complex disease owing both to the rarity of the tumour, resulting in lack of comprehensive genomic profiling, and heterogeneity observed in the tumour biology. This thesis presents evidence to support the definition of MRT as

a related tumour type with differences arising due to disease subgroups. In addition, a meta-analysis comparing published subgrouping schemes seeks to direct future research by providing a subgroup consensus encompassing all MRT. Novel survival associations and immune infiltration estimates provide new avenues for further research.

Word count: 372

Declaration

I declare that the content of this thesis represents my own work, carried out by myself during the course of my degree unless otherwise acknowledged. No part of this thesis has previously been submitted for any qualification at any university or institution. Every effort has been made to cite and reference the work of others.

Yura Grabovska

September 2019

Dedication

To my supervisors Dr Daniel Williamson and Prof Steven Clifford for giving me the opportunity to be part of the Newcastle University Paediatric Brain Tumour Group and for their ongoing support along the way over these last four years. To my panel Dr Vikki Rand, Prof Josef Voormoor and Dr Simon Cockell for steering me right...

To Dr Daniel Williamson, Dr Edward Schwalbe and Dr Matt Bashton for helping me get off the ground in bioinformatics...

To everyone in the group who helped me chase, collect, annotate, extract and profile the UK MRT cohort that is the backbone of this thesis – Prof Simon Bailey, Dr Steve Crosier, Dr Claire Keeling, Dr Debbie Hicks, Amanda Smith and Jemma Castle, Dr Jessica Pickles, Dr Amy Fairchild and Dr Patricia O'Hare...

To the MRT group, past and present – Dr Martina Finnetti, Dr Matthew Selby, Dr Azira Ramli and (future Dr) Emma Lishman. To everyone else in the Paediatric Brain Tumour Group, with whom I have always felt I am among friends...

To friends, old and new, who propped me up or helped me fall with style, and to The Dads (you know who you are)...

...Thank you

Dan, your patience with me, at various points in this project, bordered on the supernatural.

Table of Contents

Abstract	i
Declaration	iii
Dedication	iv
Table of Contents	v
List of Tables	x
List of Figures	xi
1 Introduction	14
1.1 Significance of Study	15
1.2 Malignant Rhabdoid Tumours	15
1.2.1 Atypical Teratoid / Rhabdoid Tumours	16
1.2.2 Extra-cranial Rhabdoid Tumours	17
1.2.3 MRT Histology	17
1.2.4 Incidence	18
1.2.5 Survival	21
1.2.6 Therapy	21
1.2.7 MRT Subgrouping	24
1.3 Genetics of MRT	26
1.3.1 SMARCB1 loss	26
1.3.2 SMARCA4 mutation	28
1.3.3 Genetic predisposition	28
1.4 SWI/SNF chromatin remodelling complex	28
1.4.1 Introduction	28
1.4.2 Normal function of SWI/SNF	29
1.5 Molecular profiling platforms	30
1.5.1 Methylation and expression microarrays	30
1.5.2 RNA-Sequencing	34

1.6	Dimensionality reduction and clustering approaches	35
1.6.1	Dimensionality reduction	35
1.6.2	Clustering approaches	37
1.6.3	Measuring cluster robustness	38
1.7	Cancer immune interactions	40
1.7.1	Immune interactions with cancer	40
1.7.2	Known immune interactions in the CNS	42
1.8	Summary	42
1.9	Project aims	43
2	Materials and Methods	44
2.1	Study Cohorts	45
2.2	Processing of primary patient tumours	45
2.2.1	DNA extraction from FFPE material	45
2.2.2	DNA extraction from fresh-frozen material	45
2.2.3	RNA extraction from fresh-frozen material	45
2.2.4	Quantification of genomic DNA and RNA	45
2.3	Bioinformatic analysis	46
2.3.1	Methylation array	46
2.3.2	Expression array	47
2.3.3	RNA-Sequencing	47
2.3.4	Clustering analysis	48
2.3.5	Gene pathway analysis	49
2.3.6	Survival analysis	49
2.3.7	Estimation of immune infiltration using DNA methylation	50
3	Investigating the biological relationship between ATRT and ECRT	52
3.1	Introduction	53
3.2	Aims	55
3.3	Results	55

3.3.1	ATRT and ECRT share methylation and expression features across anatomically distinct sites.....	55
3.3.2	Analysis of differences between ATRT and ECRT highlights location-specific features	60
3.3.3	Combined MRT sub-grouping strategies recapitulate previous ATRT-only models	63
3.4	Discussion.....	66
4	Generating a molecular signature of MRT sub-groups encompassing methylation and expression features	68
4.1	Introduction	69
4.2	Aims	72
4.3	Differential methylation and expression analysis of MRT sub-groups reveals group-specific changes	73
4.3.1	Meta-analysis of gene expression array	73
4.3.2	Meta-analysis of methylation array data	78
4.3.3	MRT methylation analysis reveals additional heterogeneity	80
4.4	Discussion.....	84
5	Survival analysis of MRT using sub-group specific and hypothesis-free testing	86
5.1	Aims	87
5.2	Summary of MRT cohort	87
5.3	Survival analysis MRT using clinicopathological association	90
5.4	Discussion.....	97
6	Investigating the immunological landscape of CNS tumours.....	99
6.1	Introduction	100
6.2	Aims	102
6.3	Results	102
6.3.1	Validation and benchmarking of signature matrix	102

6.3.2	Tumour Immune Microenvironment in Malignant Rhabdoid Tumours is associated with subtype and prognosis in a Tumour location dependent manner.	105
6.3.3	methyLCIBERSORT analysis of a pan-CNS Tumour methylation cohort shows significant differences in Tumour immune microenvironment related to Tumour type and grade	109
6.3.4	Tumour Immune Microenvironment in Medulloblastoma is related to molecular subtype but provides independent prognostic information	113
6.3.5	Differences in proportion of immune cell infiltration in HGG are associated with subtype, Histone/MAPK mutation, clinicopathological characteristics and prognosis	118
6.4	Discussion	123
7	Summary and Discussion	126
7.1	Malignant Rhabdoid Tumours.....	127
7.2	Investigating the biological relationship between ATRT and ECRT	128
7.3	Generating a molecular signature of MRT subgroups	129
7.4	Survival analysis in MRT	130
7.5	Estimating immune infiltration in CNS tumours.....	130
7.6	Limitations	131
7.7	Future work.....	132
7.7.1	Expanding survival analysis cohort	132
7.7.2	Further identification of subgroup-specific features.....	133
7.7.3	Novel 'omics profiling	133
7.8	Final summary	133
	Bibliography	1
8	Appendix.....	1
8.1	A) Primary tumour cohort used as part of this study	1
8.2	B) HGU133Plus2 cohort used as part of this thesis.....	4
8.3	C) Cell populations used for the generation of the CIBERORT signature matrix	

8.4	D) Cohorts of primary CNS tumours analysed by CIBERSORT.....	33
8.5	E) Validation cell populations used in benchmarking CIBERSORT	34

List of Tables

Table 1. Summary of CNS localisation of ATRT at diagnosis based on available publications.....	17
Table 2. IHC staining protocols recommended for different localisations of MRT.....	18
Table 3. Comparison of agreement and disagreement between 3 clustering approaches:.....	66
Table 4. Summary of subgrouping approaches from recent MRT subgrouping publications.....	69
Table 5. A summary table of the MRT cohort assembled as part of this thesis. Summary statistics and numbers of samples that are annotated for various clinical features are provided.....	88
Table 6. Summary of Cox proportional hazard regression univariate analyses	93

List of Figures

Figure 1. Distribution of extra-renal ECRT cases from each localisation as reported by USA Surveillance Epidemiology and End Results (SEER) (Sultan et al., 2010b).....	20
Figure 2. Data obtained from the Catalogue of Somatic Mutations in Cancer	27
Figure 3. A)Diagram of steps in bisulphite conversion of methylated	32
Figure 4. Usual process for tumour immune interaction and immune-mediated cancer cell killing.....	41
Figure 5. 2D t-SNE embedding plot of 824 embryonal tumour U133Plus2.0 expression array profiles.....	56
Figure 6. Venn diagram of the resulting significantly differentially expressed probes in the U133Plus2.0 gene expression array between multiple tumour types and MRT. .	57
Figure 7. Heatmap of relative gene expression between MRT and other embryonal tumour types.....	59
Figure 8. Heatmap of relative gene expression between ATRT and ECRT.	61
Figure 9. Heatmap of relative gene expression between ATRT and ECRT RNA-Seq.	62
Figure 10. NMF cluster quality metrics for	64
Figure 11. Sankey plot comparison between different NMF clustering cohorts.....	65
Figure 12.Consensus NMF cluster metrics from clustering HGU133Plus2.0 gene expression array data.....	74
Figure 13. Sub-group clinicopathological features as defined by consensus NMF for HGU133Plus2.0 MRT data.....	76
Figure 14. Radar plot showing the genesets identified to be significantly enriched across MRT sub-groups in HGU133Plus2.0.....	78
Figure 15. Consensus NMF cluster metrics from clustering DNA methylation array data	79
Figure 16. Sankey diagram of consensus NMF sub-group assignments	81
Figure 17. Sub-group characteristics as defined by consensus NMF sub-group assignment from clustering methylation array data;	82
Figure 18. Distribution of average beta-value across sub-groups defined from the k = 5 consensus NMF solution.	83
Figure 19. Kaplan-Meier curves showing survival differences for extent of resection in MRT	91
Figure 20. Kaplan-Meier curves showing survival differences for age categories in ATRT between patients under 3 years of age and over 3 years of age.....	92

Figure 21. Kaplan-Meier curves showing survival differences for age categories in ATRT between patients under 1 years of age and over 3 years of age	93
Figure 22. Forest plot from multivariate cox proportional hazard analysis for MRT overall survival	95
Figure 23. ROC curves for Cox proportional hazard ratio 2-year survival predictions for multiple combinations of age and subgroup categories	96
Figure 24.A) Heatmap of the resulting 2215 CpG probe beta-value signature matrix as generated by the methylCIBERSORT package.	104
Figure 25.A) Comparison of estimated proportions and type of non-cancer cells in MRT by sub-group and by location.....	106
Figure 26. A) Heatmap showing row-scaled relative levels of immune cell infiltration in 229 Malignant Rhabdoid Tumors (MRT).....	108
Figure 27. Barplots of the estimated median infiltration of specific cell types as a proportion of all non-cancer cell types (range scaled from 0-1) in 3,763 CNS tumour samples from the panCNS tumor cohort.....	110
Figure 28. t-SNE plot showing clustering of the panCNS cohort by immune cell estimates.	112
Figure 29. A) Barplots of the estimated median infiltration of specific cell types as a proportion of all non-cancer cell types (range scaled from 0-1) in 2,325 Medulloblastoma by subgroup (classic 4 medulloblastoma consensus subgroups) by SHH subtype and by 10 group consensus.....	115
Figure 30.A) t-SNE plot showing clustering of the MB cohort by immune cell estimates.	117
Figure 31. A) Barplots of the estimated median infiltration of specific cell types as a proportion of all non-cancer cell types (range scaled from 0-1) in 401 pHGG.	120
Figure 32.A) t-SNE plot showing clustering of the pHGG cohort by immune cell estimate large panel shows four immune clusters (pHGGIC1-3),	122

1 Introduction

1.1 Significance of Study

Malignant Rhabdoid Tumours (MRT) are a paediatric malignancy with a dismal prognosis and pose a number of clinical and biological challenges. The tumours originate in very young children, and rapidly progress following diagnosis, while diagnosis can be challenging due to a heterogeneous and ambiguous pathology, expressing markers from multiple lineages and a variable amount of classical rhabdoid tumour cell features. The tumours are rare and the majority of current and past studies heavily rely on historical and archived material preventing modern genomic tools being utilised, and are typically hindered by poor clinical annotation. The main molecular feature of these tumours, bi-allelic inactivation or loss of *SMARCB1* is not directly targetable by therapeutic intervention; while young age of the patients limits other therapy intensification strategies.

Efforts to characterise MRT have generated a number of sub-grouping and stratification schemes, although it is currently not clear whether these schemes are compatible, both methodically and on the basis of their biological characterisation of MRT, or if there is a clear link between subgroup and patient prognosis. There is currently no consensus on the number or content of molecular subgroups in MRT.

This study was developed to further characterise the heterogeneity seen in the disease, both in terms of extending current understanding in MRT biology and to further interrogate the disease to identify novel features. By expanding the current profiling cohorts through collecting previously unpublished cases and applying contemporary analysis tools to newly profiled and published data, this study seeks to continue to characterise the features of molecular subgroups and to develop a subgroup consensus, providing meaningful recommendations for future studies and the wider research community.

1.2 Malignant Rhabdoid Tumours

MRT are a group of related aggressive embryonal malignancies that can occur across a wide range of tissues. These malignancies occur rarely in the general population (age standardised >2 per 1000000) (Woehrer *et al.*, 2010; Brennan *et al.*, 2013), but represent a significant unmet clinical need due to a highly aggressive and rapidly progressing nature, with a dismal outcome for most patients and lack of effective therapeutic options or standardised therapy (Reinhard *et al.*, 2008; Woehrer *et al.*, 2010; Lafay-Cousin *et al.*, 2012; Brennan *et al.*, 2013; Ostrom *et al.*, 2014).

MRT comprise several sup-types, defined by the tumour localisation. Rhabdoid tumours of the kidney (RTK) are of renal origin (initially identified as an aggressive sub-type of Wilm's tumour and later classified as a distinct entity) (Beckwith and Palmer, 1978; Haas *et al.*, 1981) but more generally MRT can occur in a wide array of soft tissues and organs in the body, including the central nervous system (CNS) where they are named atypical teratoid / rhabdoid tumours (ATRT) (Rorke *et al.*, 1996). Importantly, synchronous tumours have been described localised both within the CNS and extra-cranially, as well as metastases that encompass multiple distal sites. (Szymanski *et al.*, 2013; Abu Arja *et al.*, 2018; Pinto *et al.*, 2018)

In the literature, MRT of different sites have been designated by various names. For the purpose of consistency this thesis will employ the following nomenclature: *MRT* - general term referring to all malignant rhabdoid tumours; *ATRT* - CNS atypical teratoid / rhabdoid tumours; *ECRT* - extra-cranial MRT, indicated as extra-renal where necessary; *RTK* - extra-cranial MRT occurring in the kidney.

1.2.1 Atypical Teratoid / Rhabdoid Tumours

MRT arising in the CNS were first identified in 1987, later defined as a distinct entity in 1996 and recognised by the World Health Organisation (WHO) in 2000 (Kleihues and Sobin, 2000). The name 'atypical teratoid / rhabdoid tumour' refers to the "unusual combination of mixed cellular elements similar to but not typical of teratomas and the rhabdoid cells" (Rorke *et al.*, 1996). ATRT present with a complex immune-staining phenotype, noting the presence of cells with immunophenotypes typical of glial, mesenchymal or neuronal alongside typical rhabdoid cells. Less than 20% of histology sample fields are predominated by rhabdoid cells, leading to ATRT being historically prone to misdiagnosis.

ATRT can present within any part of the CNS. Infratentorial tumours (occurring below the boundary of the *tentorium cerebelli*), including the structures of the posterior fossa (cerebellum, tectum, 4th ventricle) and the brain stem (pons, medulla) have been reported to occur in 33-61% of cases. Supratentorial tumours including the cerebrum, pineal gland, choroid plexus, hypothalamus and ventricles have been reported in 26-50% of cases. Isolated spinal tumours occur in <10% of cases, and up to around 15% of cases have a complex localisation spanning across the tentorial boundary or involving the spine (Table 1).

Publication	Lafay-Cousin <i>et al.</i> (2012)	Woehrer <i>et al.</i> (2010)	Athale <i>et al.</i> (2009)	Warmuth-Metz <i>et al.</i> (2008)	Oka and Scheithauer (1999)
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Total cases	50	19 (100)	147 (100)	33 (100)	133 (100)
Infratentorial	26 (52)	11 (57.9)	54 (36.7)	11 (33.3)	81 (60.9)
Supratentorial	22 (44)	5 (26.3)	74 (50.3)	16 (48.5)	41 (30.8)
Spine	2 (4)	2 (10.5)	9 (6.1)	1 (3)	1 (0.8)
Multiple sites†	NA	1 (5.3)	7 (4.8)	5 (15.2)	7 (5.3)
Other information			3 cases NOS; 3 cases synchronous renal		3 cases NOS

Table 1. Summary of CNS localisation of ATRT at diagnosis based on available publications. NOS - not otherwise stated, NA - not applicable; † tumours that are not localised to one CNS site

1.2.2 Extra-cranial Rhabdoid Tumours

RTK were the first type of MRT to be described and characterised (Beckwith and Palmer, 1978; Haas *et al.*, 1981). They represent roughly 2% of all paediatric kidney tumours and possess the archetypal histological characteristics of rhabdoid tumours. RTK is more commonly associated with germline mutations in *SMARCB1*, tend to originate in younger patients and commonly progress within the CNS (Vujanic *et al.*, 1996; Tomlinson *et al.*, 2005).

Extra-renal ECRT are less common than RTK but can be found in almost any part of the body (Wick *et al.*, 1995; Sultan *et al.*, 2010b). Unlike the renal tumours, other ECRT often have a complex histological appearance due to involvement of numerous tissues and organ structures, with little to no presence of typical rhabdoid cell and has historically led to difficulties with diagnosing MRT. Extra-renal ECRT tend to originate in older patients (Sultan *et al.*, 2010b).

1.2.3 MRT Histology

MRTs are defined by the presence of rhabdoid cells in the tumour. These cells resemble rhabdomyoblasts and have large, misshapen nuclei with prominent nucleoli, eosinophilic inclusions in the cytoplasm and well-defined cell membranes. Tumours frequently contain areas of mitotic activity or necrosis; calcification and haemorrhages can often also be found. Although rhabdoid cells are a defining characteristic of the tumours, they can constitute < 10% of the tumour, are not exclusive to MRTs and can occur in other malignancies (Tsuneyoshi *et al.*, 1987; Ueyama *et al.*, 1993; Perry *et al.*, 1998). This makes differential diagnoses routinely difficult. Immunohistochemical (IHC) identification of MRTs has often proven inconclusive with the tumours displaying only

some or none of the classic morphology and staining. Typical IHC staining protocols are listed in Table 2, however the type of staining employed is often centre-specific.

Antigen	ATRT	MRT	RTK
Epithelial membrane antigen (EMA)	++	++	+
Vimentin	++	++	+
Smooth muscle antigen (SMA)	+		
Glial fibrillary acidic protein (GFAP)	+		
Neurofilament protein (NFP)	+		
Neuron specific enolase (NSE)		+	+
Synaphophysin	+	+	
Myoglobin		-	
CD34		-	
CD99		+	+
Keratin	++	++	++
Desmin		-	+
S100		+	+
SMARCB1	--	--	--

Table 2. IHC staining protocols recommended for different localisations of MRT. Adapted from The European Rhabdoid Registry (EU-RHAB) Protocol (2016)

A key discovery in MRT was that the main recurring mutation (>85%) of MRT is the biallelic loss of SMARCB1, a component of the SWI/SNF chromatin remodelling complex. Initially identified due to commonly seen monosomy of chromosome 22 (Biegel *et al.*, 1989; Versteeg *et al.*, 1998; Biegel *et al.*, 1999) without other recurrent genetic mutations (McKenna *et al.*, 2008), SMARCB1 was shown to be the main target of inactivation or loss and a defining feature of MRT (Versteeg *et al.*, 1998; Biegel *et al.*, 1999; Sevenet *et al.*, 1999a; Hoot *et al.*, 2004; Jackson *et al.*, 2009; Eaton *et al.*, 2011). It should be noted that tumours with histological appearance incompatible with typical MRT features also display loss of SMARCB1 (Haberler *et al.*, 2006; Bourdeaut *et al.*, 2007); a small percentage of MRT patients also still retain SMARCB1 expression. Here, analysis of cases with wild-type *SMARCB1* identified another member of the SWI/SNF remodelling complex, *SMARCA4*, to be inactivated (Fruhwald *et al.*, 2006; Schneppenheim *et al.*, 2010; Hasselblatt *et al.*, 2011). Loss of *SMARCB1* and *SMARCA4* are now routinely screened for typically by histochemical approaches when MRT may be a potential diagnosis.

1.2.4 Incidence

Results published from the Austrian Brain Tumour Registry between 1996 and 2006 identified 19 (6.1%) ATRT cases out of 311 study eligible tumours, age-standardised rate of 1.38 per 1,000,000 person-years with a median age of 1.44 years. In the 0-2 year age group, ATRT were the most common tumour type analysed in the study and

11 (57.8%) cases were <2 years. 6 patients (31.6%) were older than 3 years and oldest patient in the cohort was 14.4 years. 10 (52.6%) cases were retrospectively diagnosed following central pathology review reaffirming the higher rate of misdiagnosis in cases before the inclusion of ATRT in the WHO brain tumour guidelines (Kleihues and Sobin, 2000) and routine *SMARCB1* screening.

Central Brain Tumour Registry of the United States (CBTRUS) data between 2001 and 2010 identified 586 ATRT cases of 0-19 years of age, representing 1.6% of all CNS tumours diagnosed in that age group (Ostrom *et al.*, 2014). For patients under 1 year of age, ATRT constituted 10.1% of cases of primary CNS tumours, and 65.7% of ATRT cases occurred in patients <2 years old (median age 1 year). Gender distribution was reported to be roughly 1:1 males to females. 35.8% of tumours were supratentorial 28.3% were infratentorial, 27.8% were recorded as 'other brain' or were shown to overlap across the boundary of the tentorium cerebelli, 4.6% were spinal tumours and 3.4% were other CNS. With regards to age, supratentorial localisation was much more likely for older patients (69%, 6-18 years).

Between 1993 and 2010, 106 children under the age of 15 who were diagnosed with ECRT in the UK were identified by the study (Brennan *et al.*, 2013). 56 (61%) diagnosed were younger than 1 year, 15 (14%) between 1-2 years, 17 (16%) 2-4 years and only 9 (8%) were older than 5 years. Of the 106 cases, 51 (48%) were renal and the remainder distributed across extra-renal sites. The proportion of each type of ECRT appears to not be consistent across the age categories, but small numbers prevent any meaningful comparison.

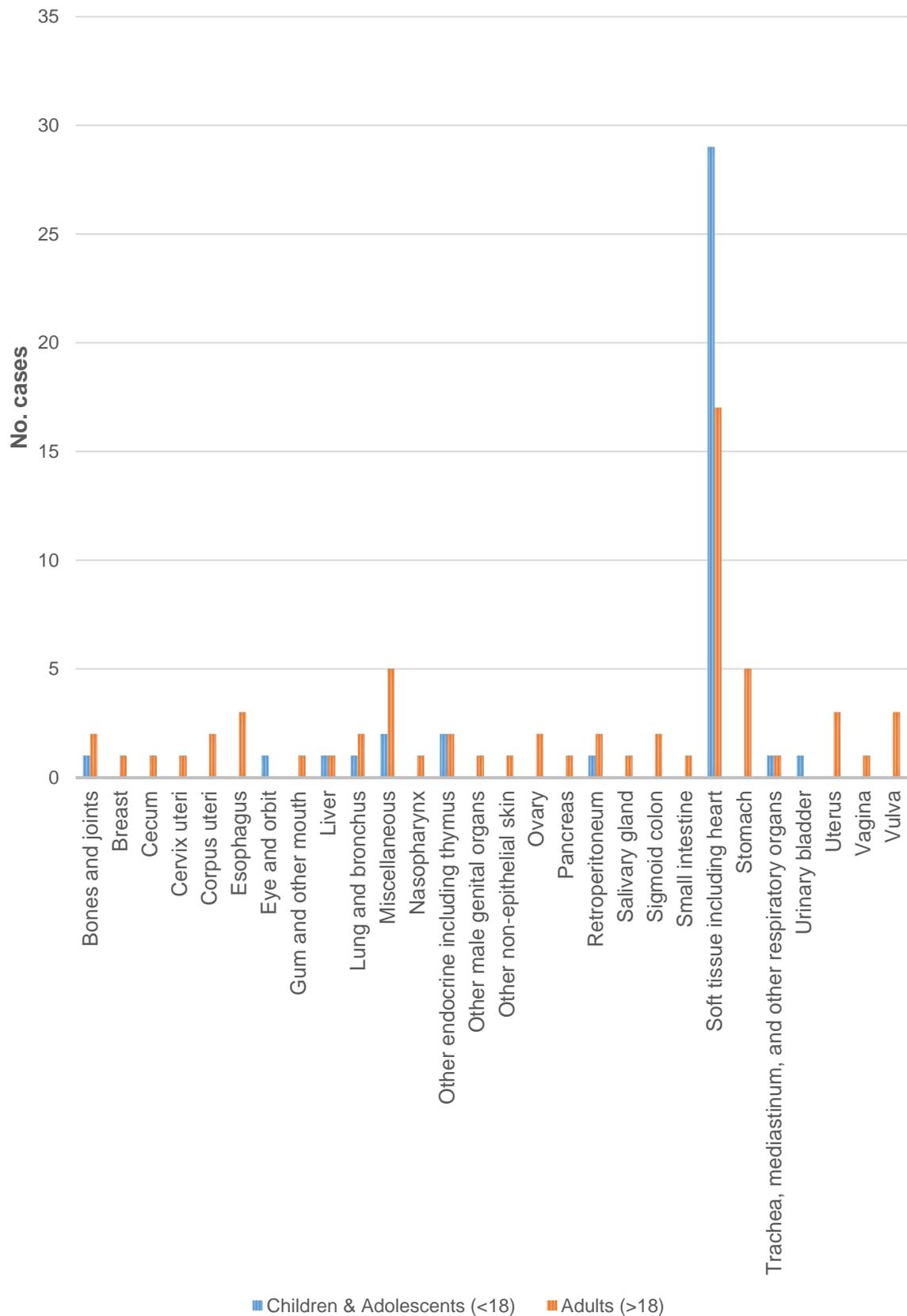


Figure 1. Distribution of extra-renal ECRT cases from each localisation as reported by USA Surveillance Epidemiology and End Results (SEER) (Sultan et al., 2010b)

The USA Surveillance Epidemiology and End Results (SEER) registry reported 3618 of soft-tissue sarcomas diagnosed between 1973 and 2006. 84 were diagnosed as

extra-renal ECRT. For patients younger than 1 year 43 (14.2%) were diagnosed as ECRT, and 41 (1.2%) were between 1-18 years (Sultan *et al.*, 2010a). In a subsequent study focusing on MRT incidence in 1986-2005, Sultan *et al.* (2010b) identified 229 MRT cases of which 45 were RTK and 103 were extra-renal ECRT. 69 (46.6%) were identified to be older than 18 years which suggests that the study may have included soft tissue sarcomas and renal tumours that were not definitive MRT cases. 79 (53%) were younger than 18 years. For extra-renal ECRT the most common localisation was soft-tissue for patients both <18 and >18 years of age (Figure 1).

1.2.5 Survival

Current estimates of MRT survival rely on single-centre reports usually utilising a single therapeutic strategy or retrospective cohort meta-analyses that comprise many different small studies, therapy modalities and cohort structure that may all confound the ability to measure survival.

Early reports of survival suggested that median overall survival (OS) for ATRT was around 6-9 months (Rorke *et al.*, 1996; Bambakidis *et al.*, 2002). More recently, OS can be estimated to be around 12-14 months (Buscariollo *et al.*, 2012; Lafay-Cousin *et al.*, 2012). Patients typically relapse or progress within 6 months of diagnosis with a rate of about 60-80% (Chen *et al.*, 2006; Athale *et al.*, 2009; Biswas *et al.*, 2015).

Long term survival has been reported, typically associated with multimodal and intensified therapy (Woehrer *et al.*, 2010; Buscariollo *et al.*, 2012; Lafay-Cousin *et al.*, 2012; Slavic *et al.*, 2014) although it is difficult to delineate which particular aspect of therapy has a significant contributing effect.

In the UK between 1993 and 2010, ECRT survival after 1 year was reported to be around 31% (Brennan *et al.*, 2013). In the NWTS, survival after 4 years was around 20% with stage I-II tumours almost twice as likely to survive as stage III-IV. Other studies have also identified stage to be a significant factor in survival prediction (Sultan *et al.*, 2010b). The same studies noted that patients under the age of 1 had <20% survival after 1 year. Survival has not improved over time, and in the UK has remained around 30% over the last two decades. When examined by primary site, the worse outcomes are seen for liver and for kidney tumours.

1.2.6 Therapy

Therapy for MRT approaches typically differ between ATRT and ECRT, and there is generally no consensus for standard therapy. A multimodal approach is employed with

surgery, radiotherapy and chemotherapy routinely used in various combinations and protocols and with differing success rates. In certain cases no therapy is noted, either due to extent of progression or other counter-indications (Lau *et al.*, 2015).

1.2.6.1 Surgical intervention

Depending on the tumour site and disease progression, surgical resection is routine for MRT. Surgical intervention may sometimes be employed without curative intent, either to aid diagnosis or to manage symptoms such as raised intracranial pressure. Surgical outcomes are typically defined as gross total resection (GTR) and near-total resection (NTR) if the tumour is almost completely excised although studies describe varying criteria, and subtotal resection (STR) when only a part of the tumour is excised. In certain cases, partial resection or biopsy are the only interventions noted.

In ATRT, multiple studies report that GTR/NTR has a significantly more favourable OS and, in some cases, EFS (Hilden *et al.*, 2004; Ann Zimmerman *et al.*, 2005; Tekautz *et al.*, 2005; Gardner *et al.*, 2008; Chi *et al.*, 2009; Isikay *et al.*, 2019). In certain studies, it is noted that the effects of surgery are difficult to disentangle from age, localisation and other factors, partially due to the small and retrospective nature of most of the available clinical annotation.

In ECRT, the wide range of possible tumour localisations means resection is not always possible and complicates outcome comparison. The 2016 European Paediatric Soft Tissue Sarcoma Study Group Non-Rhabdomyosarcoma Soft Tissue Sarcoma 2005 Study (EpSSG NRSTS 2005) noted no significant benefit to surgical resection (Brennan *et al.*, 2016), while The National Wilms' Tumor Study (NWTs) and SEER programme did not report resection as a factor in outcome, the latter citing incompatible surgical coding across sites (Tomlinson *et al.*, 2005; Sultan *et al.*, 2010b). Although, surgical resection is generally indicated where possible the lack of definitive information in ECRT means that the extent of impact on survival is still unclear.

1.2.6.2 Radiotherapy

Alongside surgery, radiotherapy is a recommended strategy and many differing protocols have been previously described. In ATRT, craniospinal radiation is especially deferred in infant patients under 3 years of age due to the effects on patient development; with patients suffering significant neuro-cognitive deficits due to effects on the developing brain. In ECRT, radiotherapy is more likely to be administered to infants though it is usually low dose (<25Gy) (Tomlinson *et al.*, 2005). Data published

from the SEER programme series reports that radiotherapy was used in 35% of patients overall (lower in infant patients, 23%) and showed no particular site preponderance (Sultan *et al.*, 2010b).

In recent years, advancements in radiotherapeutic approaches have allowed for limited radiotherapy in infants. Proton therapy (PT) has become increasingly popular in the CNS as it is seen to be less damaging to developing brain structures due to the increased precision it offers (Clasie *et al.*, 2012; Vogel *et al.*, 2018). PT protocols have been employed in 3 studies on limited ATRT cohorts. The first included a total of 31 ATRT patients, with a median age at diagnosis and therapy of 19 and 24 months, respectively (McGovern *et al.*, 2014). A second study of 10 patients with a similar median age (28 months) receiving PT between 2004 and 2011 reported that 8 patients had complete response (CR) to therapy, and that following therapy 7 of initially positively-responding patients reporting NED (De Amorim Bernstein *et al.*, 2013). The final reported single-centre study enrolled 16 patients median age 18.5 between 2007 and 2013. The centre reported 11 patients with NED or stable disease following therapy and common toxicity criteria such as nausea, vomiting and skin erythema (Haskins *et al.*, 2015). Although survival was not greatly enhanced, the studies all reported encouraging outcomes with regards to therapy-related effects. A study examining the incidence of therapy-related radiation necrosis in patients treated with PT identified the main risk factors to be the use combination chemotherapy and ATRT pathology (Kralik *et al.*, 2015).

1.2.6.3 Chemotherapy

Adjuvant chemotherapy is often used to supplement radiotherapy and surgery, or in the case of infants with ATRT, is used to defer radiotherapy. No standard therapies exist and to date no large-scale trials examining combination drug protocols have been carried out.

In ATRT, conventional dose chemotherapy has not been successful. Multi-agent therapies containing vincristine, cisplatin or carboplatin, cyclophosphamide or ifosamide and etoposide showed very poor EFS (< 10%) in the CCG-9921 trial which contained a small ATRT cohort (Geyer *et al.*, 2005). A better response was noted, with a 2-year EFS of 53%, using a modified sarcoma regimen that incorporated doxorubicin, dactinomycin and either dacarbazine or temozolomide as well as intrathecal methotrexate, hydrocortisone and cytarabine (Chi *et al.*, 2009).

High-dose chemotherapy (HDCT) is typically highly damaging for normal tissue and so typical high-dose regimens utilise autologous stem cell rescue (ASCR); a process that preserves the patient's own stem cells following treatment. Initially used as a means to defer radiotherapy for infant patients in ATRT, HDCT has become widely utilised in therapy (Fangusaro *et al.*, 2008; Gardner *et al.*, 2008; Shih *et al.*, 2008; Finkelstein-Shechter *et al.*, 2010; Nicolaidis *et al.*, 2010) as well as a means to de-escalate radiotherapy in older patients without affecting survival (Park *et al.*, 2012; Lee *et al.*, 2017). Protocols vary in course length, number and drug combinations but many include high dose methotrexate, and thiotepa as well as typical induction and maintenance agents. A recently closed trial examined the effects of HDCT on EFS. Preliminary results from ACNS0333 showed significant improvements over historical studies especially in infant patients with EFS reaching 39%. The report noted that additional intensification was not feasible and recommended stratification and targeted therapy development as a means to further improve outcomes (Reddy *et al.*, 2016).

ECRT chemotherapy approaches are equally disparate and limited by the lack of comprehensive trial data. The basis for currently employed therapies may be attributed to two studies describing successful treatment of metastatic RTK. The use of ifosfamide, carboplatin, and etoposide and vincristine, doxorubicin, and cyclophosphamide in alternating courses (Waldron *et al.*, 1999; Wagner *et al.*, 2002). The use of doxorubicin was noted as being potentially important for success, although its inclusion in the National Wilm's Tumour Study (NWTS) protocol showed no significant effect on outcome (Tomlinson *et al.*, 2005)

Brennan *et al.* (2013) raises the question of whether experience of HDCT with autologous stem cell rescue (ASCR) in ATRT could inform strategies in ECRT given the lack of any significant consensus of therapeutic approaches. In any eventuality, it is clear that escalation of conventional and currently available therapies is unlikely to significantly improve survival beyond current rates and that there is a clear necessity for novel therapy approaches, likely borne out of additional understanding of the molecular heterogeneity of MRT.

1.2.7 MRT Subgrouping

Initial evidence for the presence of putative sub-groups in MRT came from a 2013 study describing a comparison of gene expression microarray profiles in ATRT and RTK with other tumours and normal controls. The analysis identified 2 ATRT clusters and an RTK cluster which separated MRT from other tumours (Birks *et al.*, 2013) and

highlighted deregulated genes specific to each cluster when compared to other tumours and also normal tissue (906 and 424 genes in ATRT clusters and 453 genes in RTK).

In her PhD thesis, Martina Finetti described the existence of putative sub-groups based on DNA methylation and RNA-sequencing profiling, analysing a combined cohort of 23 RNA-sequencing and 39 DNA methylation array ATRT and ECRT profiles. DNA methylation clustering showed at least 2 sub-groups (Finetti, 2014).

Torchia *et al.* (2015) showed survival differences between 70 ATRT tumours using a combination of gene expression microarray and immunohistochemistry for ASCL1, a regulator of the NOTCH signalling pathway. Expression of ASCL1 correlated with a 15% improvement in survival over ASCL1-negative cases. The study showed that using molecular profiling could aid in risk stratification in ATRT, defining 3 risk categories based on a combination of localisation, evidence of metastases, surgical resection status and ASCL1 expression.

Johann *et al.* (2016) and Chun *et al.* (2016) both carried out a combination of transcriptomic and epigenomic profiling in ATRT and ECRT respectively and suggested a sub-grouping scheme of either 3 groups for ATRT based on DNA methylation and gene expression microarray or 2 groups for ECRT based on RNA-sequencing. The main defining features for the ATRT subgrouping appeared to be differences in age of diagnosis, localisation, gene expression, and an association with different 'super-enhancer' transcriptional regulators namely the tyrosinase (TYR), sonic hedgehog (SHH) and MYC proto-oncogene (MYC) pathways. In addition, Johann suggested that further heterogeneity could be observed in at least one of the putative sub-groups when analysed by DNA methylation clustering, but it was not possible to explore further due to limited cohort size.

Chun *et al.* (2016) presented a sub-grouping based around differences in the expression Homeobox C (*HOXC*) cluster genes. Interestingly, clustering ECRT samples with miRNA from normal tissue and other tumours showed that ECRT readily clustered with normal cerebellum and brain malignancies, despite having originated in either the kidney, liver or soft tissue.

In the same year Han *et al.* (2016) carried out parallel profiling of human primary tumour ATRT and ECRT cases and tumours obtained from mouse in order to develop and measure the effectiveness of an MRT mouse model. The resulting clustering

suggested 3 putative ATRT sub-groups with an additional sub-group that housed the ECRT cases. They identified expression differences between the proposed sub-groups that largely agreed with the previous publications.

Torchia *et al.* (2016) published another sub-grouping publication describing 2 broad ATRT types (Neurogenic/Mesenchymal) further sub-divided into a total of 3 sub-groups. Among the groups, they noted differences in the age and localisation, type of SMARCB1 mutation, differences in the conformation of chromatin and different transcriptional programmes. As before, this new sub-grouping model largely appeared to align with previously suggested models but was not completely homologous with previous sub-grouping strategies.

It is clear from the existing evidence that there are disease sub-groups present in MRT. Notably, only one sub-grouping publication was able to carry out any survival analysis in association with the sub-grouping. While there is significant evidence for sub-groups in MRT, it is yet unclear whether these sub-groups represent clinically relevant disease sub-types that either allow more effective survival prediction and prognostication or whether they can shed light on the underlying biological drivers in these tumours with the potential to identify novel therapeutic targets. While a number of deregulated pathways have been identified in previous sub-grouping efforts, it has not been shown whether any are therapeutically targetable and effective in MRT.

The need for an expanded and robust analysis of the number of sub-groups in these tumours as well as the incorporation of as much profiling data and clinical annotation is evident. A significant number of identified MRT primary material located in biobanks or held locally by cancer centres has yet not been profiled by any platform.

1.3 Genetics of MRT

1.3.1 SMARCB1 loss

The main genetic abnormality of MRT was initially identified due to the commonly observed monosomy of chromosome 22, band 22q11.2 (Biegel *et al.*, 1989; Douglass *et al.*, 1990). The gene *SMARCB1*, a component of the SWI/SNF chromatin remodelling complex, was shown to act as a tumour suppressor and identified as being the main molecular feature of MRT and mutated in a majority of cases (Versteeg *et al.*, 1998; Biegel *et al.*, 1999). While bi-allelic loss or inactivation of *SMARCB1* is directly responsible for the development of MRT, it has also been identified in a number of other cancers including sarcomas, carcinomas and leukaemia (Sevenet *et al.*,

1999b; Mueller *et al.*, 2004; Bourdeaut *et al.*, 2007; Hulsebos *et al.*, 2007; Hadfield *et al.*, 2008; Sullivan *et al.*, 2013; Chatterjee *et al.*, 2018). Despite this, SMARCB1 IHC staining is routinely used to diagnose MRT.

Differences in mutation type have previously been noted. ATRT cases typically show deletions of the whole *SMARCB1* gene, a loss of the 22q11.2 band or loss of heterozygosity (LOH). ECRT, on the other hand, typically presented with homozygous mutations in *SMARCB1* (Biegel *et al.*, 2002; Jackson *et al.*, 2009). Point substitutions resulting nonsense mutations are commonly reported by the Catalogue of Somatic Mutations in Cancer (COSMIC, Figure 2). In addition, different exons of *SMARCB1* are more likely to be affected in ATRT and ECRT. Notably almost no mutations are reported in Exon 1 or 8.

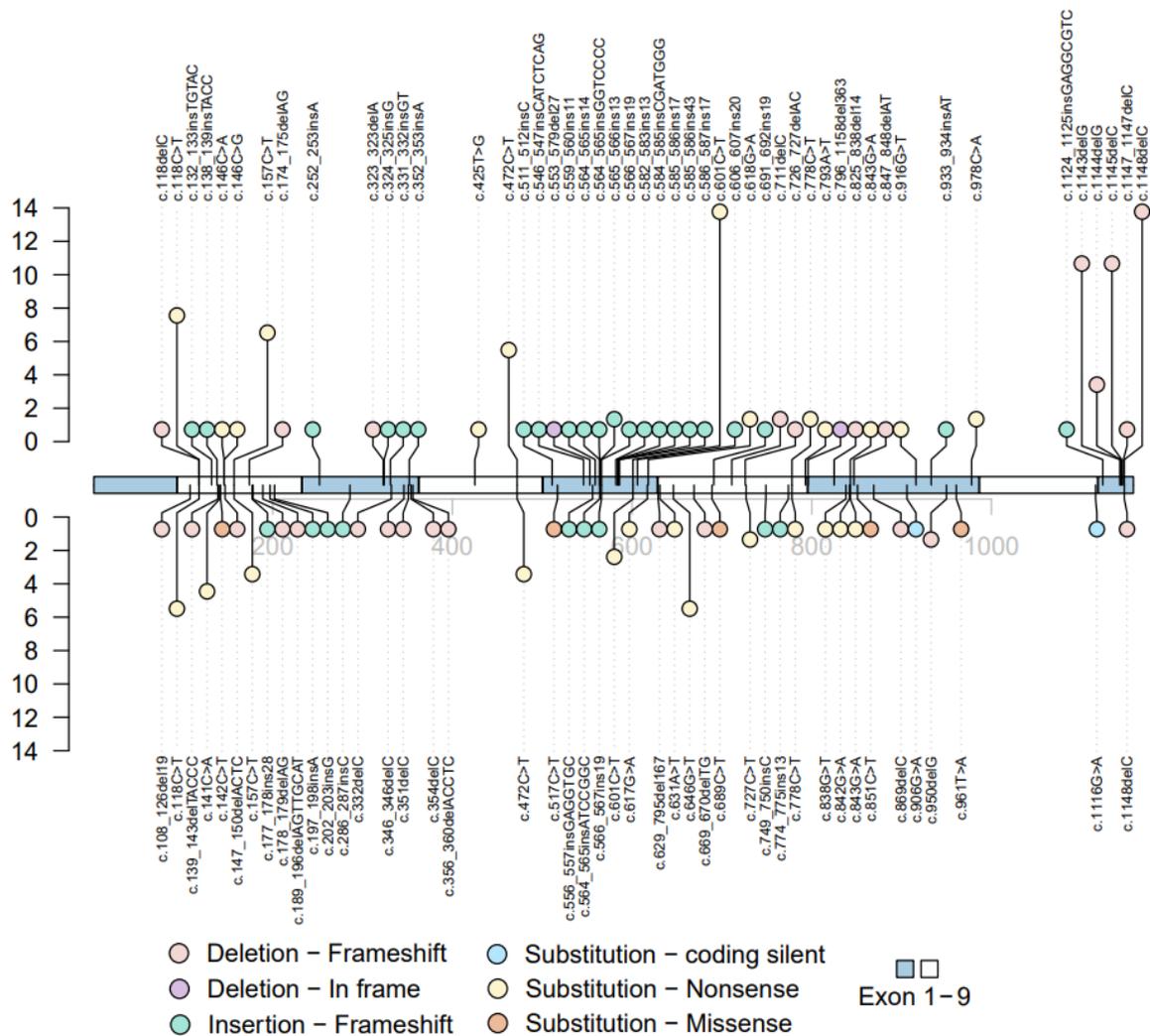


Figure 2. Data obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC; <http://cancer.sanger.ac.uk/>;) database showing absolute counts of each reported *SMARCB1* mutation type in ATRT and ECRT (Tate *et al.*, 2018).

1.3.2 SMARCA4 mutation

In a small sub-set of cases where bi-allelic SMARCB1 expression is retained, it has been shown that normal function of SMARCA4, another member of the SWI/SNF complex, is lost. Mutation in *SMARCA4* is sufficient to cause MRT in the absence of any detectable *SMARCB1* alterations (Schneppenheim *et al.*, 2010). In a limited retrospective study, *SMARCA4* alterations showed association with familial transmission, presenting as germline mutations and typically had worse prognosis (Hasselblatt *et al.*, 2014).

1.3.3 Genetic predisposition

Germline mutations in *SMARCB1* have been shown to predispose to a range of cancers including MRT (Sevenet *et al.*, 1999b). A small proportion of germline mutations are identified as part of a familial transmission pattern as noted in a number of reports and these cases showed more aggressive disease, and multiple primary tumours with dismal outcomes (Proust *et al.*, 1999; Eaton *et al.*, 2011). However, MRT patients simply harbouring germline *SMARCB1* mutations do not necessarily exhibit predisposition syndromes with poor outcome, and have been shown to achieve long-term survival (mean EFS 7 years) with multimodal treatment.

1.4 SWI/SNF chromatin remodelling complex

1.4.1 Introduction

SWI/SNF are a group of ATP-dependent chromatin remodelling complexes. These related complexes play an important role in controlling chromatin structure and regulating gene expression by modulating the accessibility of DNA to transcription machinery. They are one of four major families of complexes involved in chromatin remodelling. Three other chromatin remodelling complex families that rely on ATP hydrolysis are the 'chromo-domain, helicase, DNA binding' (CHD), inositol requiring 80 (INO80) and imitation SWI (ISWI) complexes. These families all function to remodel chromatin for different cellular functions; CHD complexes are primarily associated with transcriptional repression, INO80 regulates expression of DNA-damage repair (DDR) pathways and ISWI associated with transcriptional regulation.

Initially identified in yeast knock-out screening experiments to control mating type switching and sucrose fermentation pathways, SWI/SNF complexes are highly evolutionarily conserved and can be found in all eukaryotes including mammals. The complexes comprise many protein subunits with the combined function of translocating

across DNA to alter the condensed nucleosome structure, change histone dimers and octamer configuration and recruit other remodelling machinery in an ATP-dependent manner.

In mammals, two main complex categories exist: BRG1- or BRM-associated factors (BAF) and polybromo associated factors (PBAF). SMARCB1, SMARCC1, and SMARCC2 are found in every type of SWI/SNF complex, while other subunits vary across complexes. Most of the subunits possess specific domains for interacting with a variety DNA and protein structures including BROMO domains, zinc finger and plant homeodomain (PHD) finger. The large number of potential constituent subunits can generate a vast array of unique complex combinations, largely cell-specific and carrying out distinct regulatory function. It is no surprise that alterations in members of the complex have been implicated in different cancers and deregulation of SWI/SNF function associated with tumorigenesis.

1.4.2 Normal function of SWI/SNF

SWI/SNF complex are involved in many important biological processes including differentiation and development, proliferation and DNA damage repair. The composition of SWI/SNF complexes have been shown to be highly cell-type specific and to also change with developmental stage of the tissue. The complex has been shown to regulate lymphocyte development, maintenance of pluripotency, myogenesis and neural development, and generally be critical for normal growth as shown by mutational studies where loss of SWI/SNF components SMARCB1 and SMARCA4 during embryogenesis was lethal to the embryo. SWI/SNF has been shown to regulate senescence by interacting with p53, p21, p16 and RB1. Interestingly, the two main types of complex, BAF and PBAF have been suggested to act antagonistically to either repress transcriptional activation of genes, or to promote transcription.

SWI/SNF has been shown to be important for normal function of multiple DDR pathways. Loss of the complex has been shown to significantly sensitise cells to DNA-damaging agents. SMARCA2 and SMARCA4 rapidly localise to sites of DNA double strand breaks (DSB) by interacting with RB1 and E2F1. Loss of a number of SWI/SNF subunits can significantly impair the efficiency of DSB repair pathways including both non-homologous end-joining (NHEJ) and homologous repair (HR). In addition to DSBs, loss of SWI/SNF function sensitises cells to ultra-violet light damage and platinum agents which typically trigger nucleotide excision repair (NER). The full involvement of SWI/SNF in DDR is still unclear. The variability of the complex and their multiple

overlapping functionalities are still actively being studied in order to understand the extent of involvement of the complex in normal cell function and in cancer.

Mutations in SWI/SNF components have been found in around 20% of human cancers. Of all the possible components of BAF and PBAF, over 60% have been implicated in different malignancies. *SMARCA4* mutations have been implicated in ovarian cancers, medulloblastoma and melanoma. The involvement of *SMARCB1* in various cancers has already been described in this thesis. *SMARCA2* mutations are found in lung, gastric, breast and bladder cancers as well as sometimes being seen in MRT. *SMARCC1* is mutated in prostate cancer, *ARID1A* in ovarian, breast, liver, lung and bladder cancers. *PBRM1* is mutated in the majority of epithelioid sarcomas as well as various renal cancers. The gene *SS18* is mutated in 100% of synovial cancers.

1.5 Molecular profiling platforms

1.5.1 Methylation and expression microarrays

1.5.1.1 DNA methylation analysis by targeted microarray

Enzymatic methylation of the 5' position of cytosine in DNA is a highly complex form of epigenetic regulation in mammals. Methylation is non-randomly segregated across the genome at sites with cytosine and guanine separated by the DNA phosphate group (CpG). These CpG loci are typically found within high density clusters termed CpG islands (CGI) defined as regions larger than 200 bp, a GC proportion > 50% and a 60% observed-to-expected ratio of CpG sites (Gardiner-Garden and Frommer, 1987). CGI are associated with gene promoter regions where they typically repressed gene expression (Schubeler, 2015). Gene body methylation promotes expression and ensures that transcription is primed correctly (Yang *et al.*, 2014; Neri *et al.*, 2017). DNA methylation at distal regulatory sites has also been shown to control expression (Elliott *et al.*, 2015).

There is a specific interest in DNA methylation and its relation to cancer biology. The involvement of methylation in the development of cancer has been long established (Herman *et al.*, 1994; Jones and Baylin, 2002; Xiao *et al.*, 2016). The need to understand methylation changes in normal and cancer states has led to the development of high-throughput and whole genome approaches.

Bisulphite conversion of DNA for the purpose of screening DNA methylation in the genome was developed by Frommer *et al.* (1992). The basis of the technique is the selective specific denaturation of cytosine and not 5-methylcytosine to uracil in single-

stranded DNA without affecting the rest of the coding sequence. Subsequent polymerase chain-reaction (PCR) amplification converts the uracil to thymine, while any 5-methylcytosines are unaffected. This process is shown in Figure 3A.

For a whole-genome analysis, historically, direct sequencing would be cost-prohibitive and require large amount of high-quality input material. Instead it was favourable to utilise a microarray platform which uses sequence-specific probes in order to detect the presence or absence of a specific target DNA sequence. In the specific case of using retrospective, archival material typically stored as Formalin-Fixed Paraffin-Embedded (FFPE) blocks, DNA becomes fragmented due to the inherent cross-linking of protein molecules by formalin and is usually unsuitable for normal sequencing approaches.

Most recently, the Illumina Infinium HumanMethylation450K (450K) and HumanMethylationEPIC (EPIC) BeadChip microarray provided a high-density microarray platform to measure DNA-methylation downstream from a bisulphite conversion reaction. 450K and EPIC arrays utilise BeadArray technology which relies on randomly self-assembling bead libraries on a purpose-designed silicone substrate to generate a high-density microarray. The beads are algorithmically decoded (Gunderson *et al.*, 2004) to obtain a mapping of the hybridised array probes. Bisulphite-converted DNA is fragmented and hybridised to the array. Primer extension and staining are then used to detect the specific DNA signal treating the specific C/T transversion as a “pseudo” single nucleotide polymorphism (SNP). Two types of probes are used by the 450K and EPIC arrays. Infinium I utilises two single-colour beads specific to either the methylated or unmethylated state of the DNA probe. Infinium II uses a single two-colour bead approach and is tailored to regions of the genome with relatively lower methylation density.

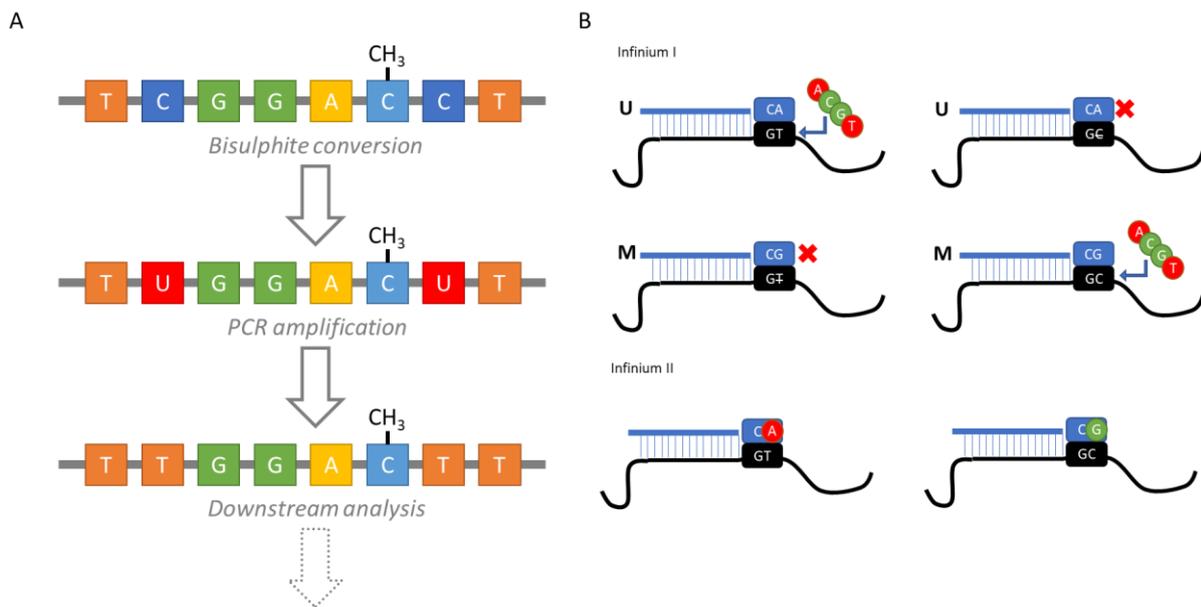


Figure 3. Assay steps part of the Illumina Methylation Array assay A) Diagram of steps in bisulphite conversion of methylated DNA, shown diagrammatically as a single strand; 5' methylation is represented by CH₃; B) The assay difference between Infinium I and II probe types. A – adenosine, C – cytosine, G – guanine, T – thymine, U – uracil.

The assay process involves the hybridisation of DNA fragments to Infinium beads ligated to 50bp oligonucleotide probes and a single-base extension step occurs. Infinium I probes are complementary to either the methylated or the unmethylated sequence at each probe locus and so extension does not take place. Infinium II probes are end base-complementary to the guanine immediately upstream of the assayed cytosine base. During single-base extension the fluorochrome labelled base mixtures are added to the reaction. A/T are red-channel detectable and C/G green channel detectable. For Infinium I, depending on whether extension occurs with the methylated or unmethylated probe, a signal is detected in the correct channel. For Infinium II, a signal can be detected in either channel based on whether a red-labelled adenosine or green-labelled guanine is incorporated (Figure 3B).

As the name would suggest, the 450K array targets roughly 450,000 genomic features (485,577 total) including genes, CpG islands (GCI) and distal features such as shores (< 2kb from GCI) and shelves (2-4kb from GCI) as well as known and predicted regulatory features, sites known to be differentially methylated in health and disease, and features used in assay quality control such as chromosome and non CpG loci. The EPIC array (sometimes referred to 850K array) introduced an additional 413,743 features as well as retaining most of the content of the 450K.

The analysis of data obtained from the 450K/EPIC arrays can be carried in several ways. Illumina provide the GenomeStudio software suite to carry out preprocessing

and differential methylation analysis. However, it has become more common to utilise open-source analysis tools designed for analysis of DNA microarrays of many different types. Software provided as part of the Bioconductor software project using the R programming language has become one of the most common approaches to analysing 450K/EPIC data due to availability, ongoing development and scalability with many tools allowing for high-throughput parallel analysis.

1.5.1.2 Gene expression analysis by targeted microarray

As with DNA-methylation microarrays, gene expression microarrays allow for high-throughput genome-wide coverage by utilising probes bound to high density arrays. Gene expression microarrays equally arose due to the need to understand gene-level differences in health and disease with direct sequencing being technically impossible on a genome-wide level.

One of the most common gene expression microarray platforms has been the Affymetrics GeneChip Human Genome U133 arrays. Of these, the most widely utilised array format, the U133 Plus 2.0 utilises around 54,000 probes to target approximately 47,000 transcripts and designed to cover the whole human genome.

The GeneChip utilises paired 25-mer oligonucleotide probes. One sequence in the pair is a perfect match to the target sequence while the other contains a mis-matching nucleotide complement. The signal difference between the two probes is then used to assay the efficiency of binding of the target sequence (Pease *et al.*, 1994). Hybridization is measured in a single colour channel following the addition of a fluorescent compound to the sample labelled with biotin.

1.5.1.3 Future of technologies

While both DNA methylation and gene-expression have historically offered high-quality and reliable analysis methods for assaying different aspects of genomics and epigenomics the technologies are approaching the end of product life. Improvements in both sequencing technology and the rapid decrease in price for whole-genome sequencing as well as changes to how patient material is preserved following sampling resulting in more stability and less fragmentation of genomic material and in recent decades have meant that more powerful approaches are now more widely available.

For DNA methylation, whole-genome bisulphite sequencing has already been shown as an effective replacement and applied to biological questions in health and disease ranging from studying normal immune B cell and T cell development to profiling the

epigenomic landscape of lung and liver malignancies (Kulis *et al.*, 2015; Li *et al.*, 2016; Delacher *et al.*, 2017). For gene expression, RNA-sequencing has become a much more powerful and increasingly accessible technology allowing the identification of gene-fusions, novel transcripts and other features typically not well covered by microarray approaches (Cieřlik and Chinnaiyan, 2017). In addition, RNA-sequencing has been used to analyse the expression of individual, single cells to a high degree of robustness and resolution (Hwang *et al.*, 2018) and even allowed sequencing of FFPE material, previously believed to be far too fragmented and degraded to ever allow whole-genome gene expression sequencing (Li *et al.*, 2018).

1.5.2 RNA-Sequencing

The sequencing of transcribed RNA products has multiple advantages over expression array measurements of gene expression. Rather than relying on probe-based interrogation of the input DNA, RNA-Sequencing (RNA-Seq) directly assays the sequence of complementary DNA (cDNA) transcribed from RNA. This greatly increases the sensitivity of the sequencing to detect low abundance transcripts as well as identify novel features such as mutations, gene fusions and novel transcripts.

Extracted RNA is subject to several steps to select specific RNA populations. Positive selection for poly-adenylated (poly-(A)) RNA is carried out to enrich for protein-coding mRNA. At the same time ribosomal RNA is depleted from the pool since it is unlikely to be informative in typical RNA-Seq experiments. The next step is to obtain short fragments of cDNA complementary to the RNA sequences in the pool. Depending on the protocol, RNA can be fragmented prior to reverse transcription or cDNA can be fragmented following reverse transcription. A typical desired fragment length is around 200-300 bp depending on the sequencing approach but longer fragments can also be used. Adapter sequences are ligated to the cDNA fragments in order to allow identification, incorporation into the sequencing reaction and for multiplexing of multiple samples. For single-end sequencing, the adapter typically contains a priming sequence in the 5'-3' direction. For paired-end sequencing, adapter elements must contain a second 3'-5' priming sequence on the complementary adapter. In addition to this, an index sequencing priming site can be included at the 5'-3' strand of the second priming site if sequence indexing is needed. Following this step, amplification of the library is carried out by PCR to prepare the pool for sequencing.

Fragment libraries are then sequenced. The sequencer returns short reads and an experiment will typically seek to obtain at least 30 million reads, although typically

higher read count (~100m) may be sought for more complex RNA-Seq analyses. Two common approaches for sequencing are employed- single end and paired end. Single-end sequencing will only sequence from one end of a fragment (typically running in the 5'-3' direction of DNA) and so a single end dataset will contain unidirectional reads. Paired-end sequencing will generate reads from both sides of a fragment, firstly sequencing from 5'-3' then priming at the opposite strand of the double-stranded fragment and running in the 3'-5' direction. This the ability to resolve complex structure and, overlapping genes and allows greater sensitivity for mutations and gene-fusions. Paired-end sequencing also allows for the generation of *de-novo* transcriptome assemblies if a reference transcriptome is not available.

Following sequencing, the analysis of the data follows a standard format. Raw data is subject to quality control (QC) measures which seek to identify problems in the sequencing results, variation due to technical error, enrichment of specific sequences or motifs which may confound downstream analysis. Necessary removal of samples can be carried out prior to alignment. The reads are aligned to a reference genome using an alignment program which is typically designed to specifically map short reads to a large genome. Many alignment tools exist and are typically chosen based on performance requirements or different downstream applications. Following alignment, quantification of gene expression is carried out in order to get gene-level counts that can be taken further into various downstream applications.

1.6 Dimensionality reduction and clustering approaches

1.6.1 Dimensionality reduction

Genomic data is typically described as 'high-dimensional'. This type of data is characterised by many variables, or 'features' with unknown correlation state. In a typical experiment, the features would represent the measured genes, or methylated loci and usually grossly outnumber the number of observations. If the goal is to make some inference on the relationship between the features and a biological state, high-dimensional data presents an additional challenge. Commonly, analysing relationships between combinations of variables is carried out by expressing them as points in dimensional space and measuring the distance between points. Many clustering approaches rely on such distance metrics, however in a high-dimension space, combinations of distances between different sets of variables can have non-unique values. This is referred to as the 'curse of dimensionality' and significantly impacts the downstream analysis of such datasets. A way to effectively work with high-dimensional

data is therefore to employ dimensionality reduction methods. There are many methods, but they all seek to map the data to a lower dimension space while preserving as much of the variation observed in the original data as possible. Although many different types of dimensionality reduction approaches exist, here only the approaches utilised in this thesis will be covered in detail.

One of the earliest approaches, principal component analysis (PCA), maps high-dimensional data to a series of lower dimension uncorrelated variables ('components') which maximally represent the variance of the original data (Pearson, 1901; Hotelling, 1933). These components are typically visualised in 2-D or 3-D space with the first principal component summarising the largest proportion of the variance, the second principal component the second-largest, and the others subsequently following the same decreasing pattern, preserving the hierarchy. While these components typically comprise important feature sets from the original dataset and can associate with different phenotypic effects in a biological context, the content of each principal component is mainly cryptic in nature, especially in the case of complex biological data, and may be hard to visualise.

Non-negative matrix factorisation (NMF) is similar to PCA in that it aims to map high-dimensional data to a lower dimensional space but the approach it utilises is to express the original dataset as a combination of two matrices (Paatero and Tapper, 1994; Lee and Seung, 1999; Devarajan, 2008). The NMF 'basis' matrix defines the number and content of variables in the reduced dataset and the 'coefficients' matrix estimates how each observation in the dataset relates to the reduced variables. Unlike PCA, NMF basis matrix outlines the necessary set of features of the original data it describes. Also referred to as a 'metagene' each variable in the basis matrix can be defined as a set of features, and in the case of a biological dataset this would usually be genes or methylation loci, although unlike PCA the metagenes are not ordered by any hierarchy by variance. Because of the way the NMF algorithm operates, the metagenes it derives can also be considered as clusters in the data and the method requires a set cluster number or range ('rank', k) of at least 2. Metagenes in NMF can also be 'projected' onto other data where a coefficients matrix is estimated in the new dataset using the metagenes derived from the original set (Brunet *et al.*, 2004; Tamayo *et al.*, 2007). This is particularly useful in biological data where it allows you to cross-validate results across different cohorts and different data types even if the underlying distributions are not equal.

More recently, techniques for dimensionality reduction have employed non-linear approaches which utilise more abstract concepts such as embedded manifolds in dimensional space. The practical upshot of such methods is that they are much more compatible with non-linear data- which may be represented by a non-linear function; which is a more typical structure for biological data. Two common approaches used in biological analyses are t-distributed stochastic neighbour embedding (t-SNE) and uniform manifold approximation and projection (UMAP). Both methods function well with very large datasets and provide effective visualisation tools. T-SNE uses a probability model to identify neighbouring points and express them as a low dimension map (Maaten and Hinton, 2008). The downside of t-SNE is that the embeddings are probabilistic in nature and the final map is generated to produce the best visualisation, so it often discards most of the original dataset structure. UMAP utilises complex topology in order to map the relationship between points and preserves the structure of the original data to a greater degree than t-SNE (McInnes *et al.*, 2018). Both methods are currently used to a limited extent in biological data, however the advantages they offer for visualising large datasets has been noted in several biological applications (Abdelmoula *et al.*, 2016; Araújo *et al.*, 2017).

1.6.2 Clustering approaches

One of the main applications for genomic sequencing or array data, aside from differential expression or methylation analysis, are class discovery approaches. Here, the aim is to use some subset of the features present in the data to categorise sub-groups in the dataset. There are many downstream applications of sub-grouping such as prognostic significance, understanding biological mechanisms, therapy stratification as well as others. Due to the complexity of data, it is favourable to algorithmically search for and categorise samples into sub-groups by their features. Although a wide array of different clustering approaches exist, only those used in this thesis will be discussed in detail.

Hierarchical clustering is one of the most common approaches to clustering. This approach seeks to generate a 'hierarchy' of clustering by comparing a (dis-)similarity metric between clusters in an unsupervised fashion. Two types of hierarchical clustering are used. Agglomerative clustering begins with each sample being a separate cluster and then these being combined based on similarity. A divisive clustering approach divide a single cluster into smaller groups based on a measure of dissimilarity. A typical metric to use for this purpose is distance between pairs of

observations with some common examples being Euclidean distance or Manhattan distance, or using 1-correlation measures for either Pearson or Spearman correlation coefficients. Hierarchical clustering can be susceptible to noise and can be complex to calculate in a large dataset (Jiang *et al.*, 2004). Despite this hierarchical clustering can still be applied effectively to biological data, though usually with additional steps to measure robustness of clusters and allow resampling to test replicability.

Another unsupervised clustering algorithm is the K-means group of algorithms. K-means assigns objects to a random cluster and iterates the cluster location to minimise the distance metric between the object and the centre of the cluster. K-means requires an initial input of cluster number to generate the cluster result, so typically this algorithm is carried out across a set of cluster numbers and various metrics used to assess robustness and stability. As with hierarchical clustering, K-means can be subject to noise (Jiang *et al.*, 2004) and also to the underlying cluster structure (Wu, 2008).

Previously mentioned, NMF carries out clustering as the typical functionality of dimensionality reduction. It is another example of an unsupervised approach which requires an initial cluster number input and then seeks to minimise a measure of entropy of a probability distribution in the clustering, known as Kullback–Leibler divergence. The main positive feature of NMF is that while it generates clusters, it also generates metagenes for each cluster which seek to describe the features that define that cluster in a meaningful way.

1.6.3 Measuring cluster robustness

The unsupervised nature of many clustering approaches, the need to distinguish between multiple similar resulting cluster models and the fact that a clustering may utilise thousands of features in the result all raise an important issue for the need to effectively quantify and discriminate between clustering results and to also ensure the result is robust and not simply an artefact or error.

Many different measures of robustness are employed, and the strategies can involve one or several specific statistical metrics or an application of the resulting clustering in order to test its performance. Each measures a unique aspect of the clustering, can be method specific and not clearly superior to other approaches.

One of the most common visual measures of cluster consistency is the silhouette value. This value measures the similarity of a data point to its cluster compared to other clusters obtained using any meaningful distance metric (such as Euclidean distance or

1-correlation). For each object in the clustering, silhouette is calculated and compared across the whole dataset (Rousseeuw, 1987). A high silhouette score across the data suggests that the clustering is appropriate and an averaged silhouette for each clustering is readily expressed as a graphic and can be used to identify an optimal number of clusters.

The cophenetic correlation coefficient is another single value method to assess cluster quality by comparing the dendrogram generated from a clustering to the original data pairwise distances. The higher the coefficient value, the more faithfully the original pairwise distances are preserved suggesting that the clustering model is an effective representation of the structure of the data (Sokal and Rohlf, 1962).

The index or coefficient of dispersion can be used to assess whether objects in the cluster are clustered or dispersed compared to a model of a probability distribution or cluster. This measure of dispersion can therefore be used to assess how distributed or dispersed data are within a cluster (Hoel, 1943).

As well as single measures of cluster stability and conformity, it is useful to measure how a particular clustering result may be used to classify novel data. This type of cross-validation usually relies on the use of a 'training' and 'validation' dataset. By generating a clustering based on some known training data typically where some information about the underlying sub-grouping is already available, it is therefore possible to apply the clustering to a novel dataset using several different machine learning approaches and ascertain how effectively the new data are assigned group calls.

However, due to the limitations of certain biological analyses, it is not always possible to have a validation dataset available. In this instance the training data may be used to test the effectiveness of the clustering as a classifier by sampling only parts of the data at random for a set number of iterations and measuring the frequency with which a sample classifies to the same cluster (Dwass, 1957; Efron, 1992) as well as being able to calculate confidence intervals and a p-value estimate for hypothesis testing. This type of internal resampling is compatible with many clustering approaches where cluster order is hierarchical or is otherwise preserved. For methods like NMF, the use of metagene projection onto the new resampled clustering can be used in order to directly compare each iteration, since normal NMF typically does not preserve cluster order. Recently, it was shown that this type of approach could also be used to test non-linear methods such as t-SNE (Sharma *et al.*, 2019b).

1.7 Cancer immune interactions

11 years after they initially defined the archetypal hallmarks of cancer (Hanahan and Weinberg, 2000), Hanahan and Weinberg (2011) refined their original definition to include a number of additional features which had become increasingly recognised as being critical for cancer development and the challenges regarding therapeutic approaches. Among them were the ideas of tumours avoiding destruction by the immune system and the environment surrounding the tumour potentially contributing to cancer development by supporting continued inflammation. Understanding how the immune system interacts with cancer has been shown to be a highly complex and nuanced process.

As well as a growing understanding the underlying biology, studying the immune system in cancer can offer prognostic value potentially allowing clinicians to predict therapy response as well as offer new therapeutic targets by directly allowing the targeting of cancer-promoting immune interactions.

1.7.1 Immune interactions with cancer

It was shown 25 years ago that human tumours can trigger the generation of CD4+ and CD8+ cytolytic T cells specifically able to recognise cancer cells (Boon *et al.*, 1994). This function has been linked to the demonstrable propensity of cancer cells to produce neo-antigens and express combinations of markers not found on normal tissue (Tian *et al.*, 2011). These T-cell interactions typically occur with low affinity due to the normal process of T-cell selection, maturation and removal of high-affinity autoimmune populations (Giraud *et al.*, 2007) through promiscuous expression of tissue-specific antigens in the immune organs. However, high specificity lymphocytes can develop due to several specific causes, namely the presence of viral antigens, mutation in genes expressing antigen or expression germline genes not found in adult somatic tissues (Lennerz *et al.*, 2005; Fujita *et al.*, 2007; Coulie *et al.*, 2014).

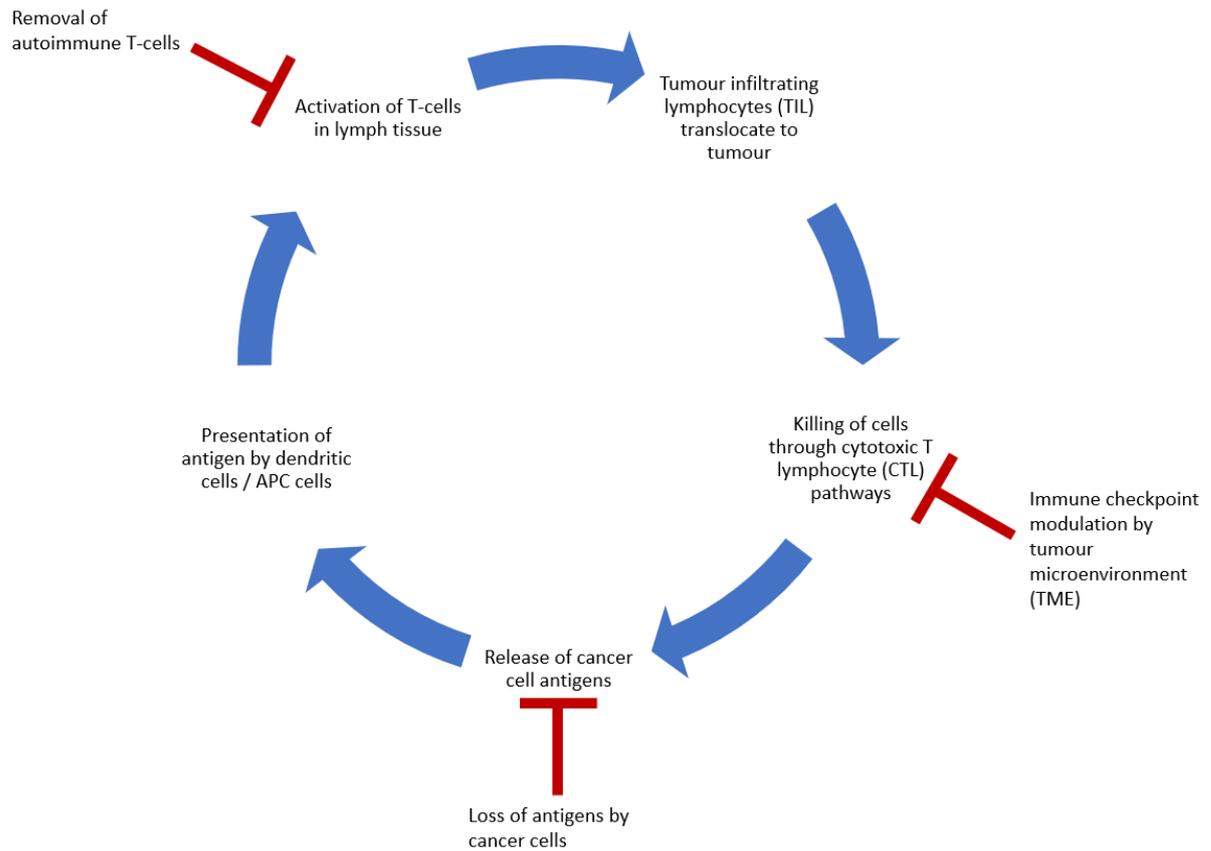


Figure 4. Usual process for tumour immune interaction and immune-mediated cancer cell killing

The ability of cancer cells to actively avoid immune targeting by losing expression of reactive antigens (Dunn *et al.*, 2002) and modulate T cell response by the activation of immune-checkpoint receptor PD-1 (Drake *et al.*, 2006) as well as a number of other regulators add an additional layer of complexity to the tumour immune response (Figure 4). However, ongoing research to elucidate the nature of these types of immune interactions have already led to multiple successful therapeutic targeting strategies for tumour immunology.

Programmed death 1 (PD-1) is an immune checkpoint receptor found on activated T-cells and involved in immune modulation. Activation of PD-1 in peripheral tissues leads to immune suppression and inhibition of this interaction leads to enhanced immune response (Iwai *et al.*, 2002). Ligands of the PD-1 receptor, namely PD-L1/2 are found on somatic tissue but are increasingly being encountered in tumours where interaction with PD-1 causes T-cell apoptosis (Dong *et al.*, 2002). Monoclonal antibodies targeting either CTLA-4 (Hodi *et al.*, 2003), PD-1 or its ligand (Larkin *et al.*, 2015) have been developed and show a significant therapeutic potential although there is a great degree

of variability in patient response with only a small group of patients responding and a high prevalence of therapy-induced side effects (Shen and Zhao, 2018).

Several factors have been shown to be predictive in response to immune therapy. The presence of CD8+ T-cells within the tumour or at the margin (Tumeh *et al.*, 2014), PD-L1 , CTLA-4 associated gene expression on infiltrating lymphocytes (Herbst *et al.*, 2016) and on tumour cells (Larkin *et al.*, 2015). It is therefore clear that understanding the tumour microenvironment and the specific behaviour of tumour infiltrating lymphocytes is necessary for the success of targeted immune therapies.

1.7.2 Known immune interactions in the CNS

The CNS has historically been suggested to be a largely immune-privileged site – that is largely devoid of immune activity (Streilein, 1993). This was in part due to the presence of the blood-brain barrier (BBB), the lack of typical lymphatic structure and the significant lack of dendritic cell presence (D'Agostino *et al.*, 2012). It has since been shown that immune functions are not absent in the CNS although they occur through alternative pathways. Microglia are the cell responsible for a large proportion of antigen presentation in the brain as well as several other functions including regulation of inflammation (Hayes *et al.*, 1987; Gehrman *et al.*, 1995; Aloisi, 2001). There is also clear evidence for immune cell infiltration in CNS malignancies, with macrophage and lymphocyte infiltrates being described in glioblastoma (Rossi *et al.*, 1987; Yang *et al.*, 2010). Immune therapies in glioblastoma have not been largely successful (Reardon *et al.*, 2017a; Reardon *et al.*, 2017b). Deeper understanding of immune infiltration of CNS tumours and the role of the TME is necessary to further develop therapeutic strategies, improve response and outcome and allow for stratification.

1.8 Summary

There is a clear need to further understand how differences genomic and epigenomic differences between ATRT and ECRT, the presence and clinical relevance of sub-groups and novel aspects such as the immune landscape shape MRT biology and inform our ability to predict patient outcome and the development of new therapies. Ongoing sub-grouping strategies have identified multiple group-specific features but have yet not been able to define a single unifying sub-grouping scheme. In addition, it is not clear whether there is a clear link between molecular subgroups identified in MRT and any significant clinical benefit. Finally, little is known about the immune interactions of many CNS tumours including MRT. Understanding the TME in these cancers may shed a light on new therapeutic avenues.

1.9 Project aims

This project aimed to expand on current understanding in MRT biology, differences between tumour types and the presence of any molecular sub-groups. The focus of the identification of subgroups was identifying the optimum number of subgroups and using available clinical and genomic data to define their clinical and biological relevance. This project aimed to provide guidance on consensus sub-group definitions both in terms of methodology and the nature of subgroups identified, as well as rationale for their adoption into clinical practice. Finally, this project sought to explore novel aspects of MRT biology such as the TME in MRT and any potential for therapeutic targeting.

The major aims for this project were as follows:

- Generate a UK-wide MRT cohort with molecular profiling and high-quality clinical annotation using new and retrospective cases obtained from UK cancer centres. Profiling of additional cases to be combined with published and collaborator datasets to generate a large multi-platform MRT study cohort (Chapter 2)
- Evaluate current understanding of genomic and epigenomic features of ATRT and ECRT and how these differences impact the biology of the tumour to understand whether a combined approach of investigating all MRT regardless of location is a valid strategy for the development of novel therapeutic approaches and understanding of MRT biology (Chapter 3)
- Develop a consensus in current MRT sub-grouping strategies and generate a molecular signature of MRT sub-groups alongside clinicopathological annotation addressing any additional heterogeneity in the disease not currently well characterised.(Chapter 4)
- Define sub-group specific survival effects and investigate survival features independent of sub-group based on clinical annotation of tumour cohort (Chapter 5)
- Examine the characteristics of immune infiltration of CNS tumours including MRT and investigate whether differences in immune infiltration are indicative of differences in underlying biology and disease outcome (Chapter 6)

2 Materials and Methods

2.1 Study Cohorts

Datasets used in this study are detailed in the Thesis Appendix

2.2 Processing of primary patient tumours

2.2.1 DNA extraction from FFPE material

Genomic DNA was extracted from scrolls of material obtained from FFPE blocks. Sections were obtained from multiple sources and were cut at varying thickness. Where possible, scrolls of thickness 20 µm were preferentially used. The number of scrolls used per sample was decided based on visual assessment of the block and the available scrolls. Extraction of DNA was carried out using the Qiagen QIAmp® DNA FFPE tissue extraction kit (Qiagen, Venlo, Netherlands) using the standard manufacturer-supplied protocol. The DNA was eluted in 50-100µl of DNase/RNase free water.

2.2.2 DNA extraction from fresh-frozen material

Genomic DNA was extracted from fresh-frozen material using the Qiagen DNeasy® bloody and tissue extraction kit (Qiagen, Venlo, Netherlands) using the standard manufacturer-supplied protocol. The DNA was eluted in 50-100µl of DNase/RNase free water. Prior to extraction, all handling of material was carried out under refrigeration utilising dry ice to avoid degradation of material due to repeated freeze/thaw cycles. For long-term storage, frozen material was stored at -80°C.

2.2.3 RNA extraction from fresh-frozen material

Patient RNA was extracted from frozen tissue homogenised with TissueLyser II (Qiagen) using Trizol Reagent (Ambion, Life Technologies) using the standard manufacturer-supplied protocol.

2.2.4 Quantification of genomic DNA and RNA

Initial quantity and quality of extracted genomic DNA was carried out using the NanoDrop 1000 Spectrophotometer to analyse the absorbance ratio 260nm/280nm.

For the purposes of microarray analysis, the Qubit® PicoGreen dsDNA broad range assay kit (Invitrogen, Carlsbad, CA, USA) was used using the standard manufacturer-supplied protocol to obtain the quantity of double-stranded DNA.

RNA quality was assessed using a Bioanalyzer (Agilent Technologies) with an Agilent RNA 6000 Nano kit using the standard manufacturer-supplied protocol to assay

general quality of extracted RNA and to obtain the RNA integrity (RIN) number which is a measure of the level of degradation and fragmentation of RNA.

2.3 Bioinformatic analysis

2.3.1 Methylation array

2.3.1.1 Pre-processing of raw 450K/EPIC microarray data

BeadArray IDAT files were obtained and preprocessed using the (R/Bioconductor) package minfi (Aryee *et al.*, 2014). Data from both array platforms was combined at the RedGreenChannelSet level and any probes not shared between the two platforms were discarded prior to normalisation.

2.3.1.2 Quality control of 450K/EPIC microarray data

Quality control of the DNA methylation microarray data was carried out using internal minfi functionality for QC. Detection P-value as well as array-specific control probe intensities were assessed for any evidence of failed or poor-quality arrays. Where detection P-values for probes on an array > 0.05 in 5% of probes or more, the sample was removed from the dataset prior to normalisation.

2.3.1.3 Normalisation of 450K/EPIC microarray data

Normalisation of data was carried out using the normal-exponential out-of-band (noob) method using single sample normalisation as this approach allows a flexible sample processing pipeline and removes the need to preprocess an entire dataset in one batch while still controlling for technical variability (Triche *et al.*, 2013). Datasets intended for use with copy-number estimation were retained at this point and were not subject to additional steps outside of the specific copy-number estimation pipeline.

Following normalisation, the dataset was mapped to the human genome using the Illumina array manifest for genome assembly hg19/GRCh37 and beta-value ratios were obtained for each probe retained in the dataset.

2.3.1.4 Non-specific filtering of CpG probes

Probes were removed based on several filtering criteria. Probes that mapped up to 2 nucleotides away from a known SNP and where a minor allele frequency (MAF) was greater than 0.05 were removed. Sex chromosomes were removed and any probes mapping to non-methylated loci (typically used as control probes). Finally, probes shown to cross-hybridise to multiple loci on the genome were removed based on recommendations from two publications specifically analysing the hybridizing behaviour of the 450K/EPIC arrays (Chen *et al.*, 2013; Pidsley *et al.*, 2016).

2.3.1.5 Estimating Copy-Number Aberrations using methylation array data

450K/EPIC methylation array data was used to estimate DNA copy-number using the package *conumee* (R/Bioconductor). 119 samples in the “Control” group from the Molecular Neuropathology 2.0 (MNP2.0) dataset (Capper *et al.*, 2018) were used as control reference arrays. The analysis was run using default parameters. Gain or loss of individual chromosomal arms was estimated using the method discussed in Schwalbe *et al.* (2017b) with the cut-off for a “Loss” event of -0.22 and a “Gain” event of 0.12.

2.3.2 Expression array

2.3.2.1 Pre-processing of Affymetrix expression microarray data

Affymetrix expression array data was processed using the *affy* package (R/Bioconductor). Raw CEL files were read into *AffyBatch* objects and normalised using Robust Multi-Array Average expression measure (RMA) method (Irizarry *et al.*, 2003)

2.3.2.2 Non-specific filtering of Affymetrix probes

For the purpose of clustering and dimensionality reduction, non-specific filtering of probes was carried out using the *genefilter* package (R/Bioconductor). Probes were temporarily transformed from log₂-scale and retained if they satisfied two criteria: a coefficient of variation > 1 and at least 5% of samples having an intensity of 200 or greater.

2.3.2.3 Differential expression analysis (DE)

Supervised analyses of differential expression were carried out using the *limma* package (R/Bioconductor). The model formula and testing contrasts were constructed using internal functionality and using phenotypic data factors such as sub-group, age, CNS location and other factors relevant for testing. An empirical Bayes method was used to obtain either the moderated t-statistics or F-statistic and the p-value provided was adjusted for multiple testing using the “Benjamini & Hochberg” method.

2.3.3 RNA-Sequencing

Total RNA (RIN > 7) was prepared for RNA-sequencing using Illumina Tru-seq RNA-seq Library Preparation Kit. Library was run on an Illumina HiSeq2500, 4 x multiplex as a 100bp paired-end run at around 90M reads per sample.

2.3.3.1 Pre-processing of raw sequencing data

Fastq files from RNA-seq experiments were subject to quality control check using FastQC. Reads were aligned to HG19 (USCS) genome using RNA-STAR. Read counts were generated by HT-seq-count mapping to the GENCODEv17 library.

2.3.4 Clustering analysis

2.3.4.1 Hierarchical clustering

Hierarchical clustering was performed using the ConsensusClusterPlus package (R/Bioconductor). A distance matrix was generated as 1- Pearson correlation coefficient and clustered using the k-means algorithm for 1000 repetitions and an item resampling rate Of 80%.

2.3.4.2 Non-negative matrix factorisation (NMF)

Non-negative matrix factorisation was carried using the NMF package (R/Bioconductor). The default package parameters were used for all analyses at 256 iterations. Resampling NMF was carried out using the method outline in Schwalbe *et al.* (2017b) for 1000 repetitions and an item resampling rate of 80%. An initial training NMF result was calculated for the range of metagenes (k) being tested. For the range of k , test data was initially subject to resampling without replacement at the specific resampling rate and subsequently clustered again by NMF for 256 iterations and a pseudo-inverse metagene projection was carried out as described in (Tamayo *et al.*, 2007) onto the whole training NMF result. The resulting projection was clustered by K-means for the number of clusters in the range of k . The resulting frequency of assignment to the same cluster group was recorded and an average of the NMF H values for each metagene calculated. Clustering robustness was estimated using the initial metrics provided by the NMF package as well as the Corrected Rand Index (CRAND), Cohen's Kappa coefficient, Average Silhouette calculated from k-means clustering for each combination of k metagenes and k clusters and the proportion of samples in each combination of k metagenes and k clusters that are assigned to the same metagene with a frequency 95% or greater.

Comparison of NMF with hierarchical clustering was carried out on subsets of data. Although not included as part of this thesis, hierarchical clustering and NMF were carried out with different numbers of methylation array probes based on different thresholds of standard deviation. Additionally, different hierarchical clustering and NMF computation algorithms offered by the respective R packages were used. Metrics of subgrouping were inspected alongside the resulting number and content of subgroups.

Although some variability in a subset of the data does exist, with around 10% of samples being assigned different subgrouping calls, the overall subgrouping for the majority of the datasets despite different approaches to thresholding, the clustering method and algorithm. Due to the additional features of NMF such as the basis and coefficient matrices defining metagenes within the data and its overall robustness through the analyses, NMF was chosen as the main method of subgrouping for this thesis.

A more formal comparison of subgrouping approaches was published as part of the international ATRT consensus project parts of this thesis analysis contributed to. The ATRT consensus study text has been accepted for publication and is currently available as a pre-publication document ([doi: 10.1093/neuonc/noz235](https://doi.org/10.1093/neuonc/noz235)).

2.3.4.3 *T-distributed Stochastic Neighbour Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP)*

T-SNE was carried out using the package Rtsne (R/Bioconductor) on a distance matrix generated as 1- Pearson correlation coefficient without PCA. Exact t-SNE was carried out with the *theta* parameter of 0. Perplexity was set to default where the package function allowed, and where the function stated the perplexity was too high, the value was reduced to satisfy the internal threshold and based on subjective judgement of the final visualisation. Each t-SNE analysis was iterated 5000 times.

UMAP was carried out using the package uwot (R/Github) using default parameters and the number of neighbours based on subjective judgement of the final visualisation.

2.3.5 *Gene pathway analysis*

2.3.5.1 *Gene-set enrichment analysis (GSEA)*

GSEA was carried out using version 2.2.4 of the program on pre-ranked datasets with geneset msigdb version 6.2, 1000 permutations and maximum and minimum geneset thresholds set to 500 and 15, respectively.

2.3.6 *Survival analysis*

Overall survival (OS) and progression-free survival (PFS) were calculated from date of diagnosis to date of death, record of progressive disease or censored. Survival curves were generated using the package survminer (R/Bioconductor). Cox proportional hazard regression modelling was carried out using the package survival for both univariate and multivariate analyses. ROC curves used to test survival models were generated using the package survivalROC (R/Bioconductor)

2.3.7 Estimation of immune infiltration using DNA methylation

2.3.7.1 Generation of signature matrix

The custom limma-based function as described by Chakravarthy *et al.* (2018) was used to fit linear models performing a pairwise comparison between each of the cell types. A maximum of 200 top features per pairwise comparison were selected restricting to probes showing a median beta-value difference of 0.2 and FDR of 0.01. Beta-values were scaled to between 0-100 and probe means per cell type calculated to form a signature matrix compatible with CIBERSORT.

2.3.7.2 Running CIBERSORT

Input methylation matrices were created by processing raw .idat files as per above. Data were sourced from published GEO datasets GSE70460, GSE109381, GSE77353, GSE63669, GSE60274; Array Express dataset E-MTAB-5528 and the MRT cohort generated as part of this thesis. CIBERSORT was run in relative mode using the provided R script (<https://cibersort.stanford.edu>) using 1000 permutations without quantile normalization.

2.3.7.3 Validation and benchmarking of signature matrix

The signature matrix was inspected to verify that each cell type was accounted for by specific hypo/hyper-methylated CpGs and not unduly compromised by batch effects. Likewise, t-SNE (package rtsne) was used to visualize the cell-type specificity of the signature matrix. The mean and sd of signature matrix CpGs were inspected in each of the 80 CNS-tumor methylation types represented in dataset GSE109381 to identify possible outlier or confounding effects between immune-cell type specific CpGs and tumor cell types.

Deconvolution performance was benchmarked against 18 gold standards i.e. 6 x methylation profiles of peripheral blood mononuclear cell (PBMC) mixtures with known flow-cytometry and 12 x mixtures of reference pure populations DNA in known proportions (GSE112618). Performance was also benchmarked against simulated mixtures generated to contain known quantities of a given cell type. This was achieved by taking the mean beta-value of each pure cell reference and applying a random uniform distribution such that each simulated mixture contained a fixed amount of a given cell type (100 simulations for each) and a fixed 75% cancer cell signature derived from relevant cancer cell reference lines. Correlation with methylCIBERSORT estimates was tested by the Spearman Rank method.

A Breast cancer dataset (GSE20713, GSE72308) containing 87 samples with parallel Affymetrix HGU133p2 expression profiles and 450K Methylation profiles was used to compare relative cell type estimates from both methylCIBERSORT and the classic expression CIBERSORT run using standard signature matrix LM22 (cibersort.stanford.edu) in relative mode using 1000 permutations with quantile normalization. meTIL score (an independent measure of T Lymphocyte infiltration based upon methylation status of 5 CPGs) was calculated following the method as described by Jeschke *et al.* (2017)

3 Investigating the biological relationship between ATRT and ECRT

3.1 Introduction

As previously discussed, ATRT and ECRT share many common biological and clinical characteristics, appear to have similar aetiology and exhibit the same sole molecular feature of bi-allelic *SMARCB1* loss. Historically, the similarities and differences between the two broad tumour types have been discussed by citing comparable studies or similar outcomes in various small datasets.

The first, and currently only study to directly compare expression in ATRT and ECRT examined 10 RTK and 13 ATRT tumours by gene expression and miRNA profiling array (Gruppenmacher *et al.*, 2013). The authors reported 122 genes significantly differentially expressed between the two tumour types, but failed to identify miRNA differences. Genes downregulated in ATRT included *TBX2*, *HOXA5/9*, *IGFBP5* while genes upregulated in the group included *FABP7*, *SOX2*, *NEUROG2*, and *BMP7*.

In the last 5 years, a number of studies have sought to analyse the molecular features of either ATRT, or ECRT and these can give insights into the overlapping biology of MRT as well as highlight the key features which separate these tumours. Han *et al.* (2016) carried out clustering of gene expression microarray data from human primary tumours alongside a series of tumours derived from mouse-modelling of MRT during embryogenesis, and public datasets of stem cell populations. By perturbing *Smarcb1* using a temporal gene knockout system in mice, they showed high incidence of intra-cranial tumours resembling ATRT and they examined the gene expression and pathway relationships between the different tumour types alongside these murine-derived tumours. For MRT, they identified 3 intra-cranial sub-groups termed (hIC1-3) and a single extra-cranial sub-group (hEC). By comparing each sub-group with populations of stem cells, they showed that hEC as well as hIC1/2 correlated expression with embryonal stem cells (ESC), neuroepithelium, and to a lesser extent neural progenitor cells. Additionally, hEC and hIC1/3 but not hIC2 showed significant correlation with neural crest cells and mesenchymal stem cells.

hIC1,2 and 3 showed high expression of neural gene *ACTL6A* suggesting a neural progenitor lineage, while hEC showed high expression of homeobox genes such as *TBX2* and *HOXC*, although moderate expression was noted in hIC2 and hIC3 for the two genes, respectively. Interestingly, the paper highlighted differential expression between sub-groups of the gene *HMOX1* encoded on chromosome band 22q12, which is relatively close to the position of *SMARCB1* (22q11). An emerging feature of MRT is the different types of chromosome 22 aberrations between sub-groups, with the

MYC sub-group showing the lowest frequency of whole and partial chromosome 22 loss (Johann *et al.*, 2016; Torchia *et al.*, 2016). This could, in part, explain the higher expression levels of *HMOX1* in the hEC and hIC1 sub-groups in the publication as these also share higher expression of *MYC* and lower expression of *ASCL1*, *HES5* and *GLI1*, although this has not been investigated further.

Subsequently, two analyses on ATRT and ECRT respectively were published later in 2016 looking at sub-groups and biological heterogeneity in the tumour types. Chun *et al.* (2016) published an analysis of ECRT transcriptomic data. RNA-Seq data for 40 primary tumours, (comprising 34 RTK and 6 extra-renal ECRT) showed the presence of two expression sub-groups. Sub-group 1 (n = 22) contained all the extra-renal ECRT cases and was showed high expression of immune genes such as immunoglobulins and genes associated with BMP pathway signalling such as *BMP4*. Sub-group 2 showed increased WNT signalling such as *WNT5A*. Importantly, comparing their sub-group definition to the 122 differentially expressed genes highlighted by Illumina HT-12 array in ATRT and RTK (Gruppenmacher *et al.*, 2013) the authors suggested that sub-group 1 could resemble ATRT (11/29 genes up in ATRT) while sub-group 2 RTK (21/92 genes up in RTK). In addition to the transcriptomic analysis, the authors carried out ChIP sequencing (ChIP-seq) on 10 MRT primary tumours, 3 MRT cell lines and 3 human embryonal stem-cell (hESC) lines, identifying significant H3K27 acetylation density at *HOXA,B,C* gene clusters, consistent with the previous findings of Han *et al.* of HOX gene association with hEC.

The second published analysis, focusing on ATRT examined 150 methylation array and 49 gene expression array profiles. They identified 3 ATRT sub-groups for which they defined a broad definition of SHH, TYR and MYC using ChIP association with gene enhancer regions. The three groups showed differences in localisation, type of *SMARCB1* mutation and gene expression with SHH being the most neuronal group showing high expression of *GLI2*, *SOX11* and MYC being the most mesenchymal group expressing *MYC* and *HOXC* most highly among the 3 groups.

Taken together these findings show that differences between MRT may not be simply explained by where they originate. They suggest that in both their transcription profile and epigenetic regulation MRT occurring in the CNS may resemble their extra-cranial counterpart, or vice versa and rather than simply focusing on the tumour location, it is important to understand the nature of any disease subgroups that exist. The degree of overlap between ATRT and ECRT therefore still remains difficult to pin down and

requires further elucidation to fully assess the biological overlap between tumours occurring between tissues of vastly different function.

3.2 Aims

The aims of this chapter are to highlight the common features shared by all MRT regardless of localisation by comparing them to other tumour types of embryonal origin as well as examine the specific differences between ATRT and ECRT using expression and methylation profiling.

The degree of overlap will be examined to understand whether a recommendation for future subgrouping and therapeutic targeting strategies should be aimed at MRT as a whole.

3.3 Results

3.3.1 ATRT and ECRT share methylation and expression features across anatomically distinct sites

In order to compare gene expression between different embryonal tumours, relevant Affymetrix HG U133Plus2.0 gene expression datasets were obtained from publications and online public repositories. Although other datasets were available, this platform was selected due to the availability of multiple tumour types and the ability to process and normalise the combined raw dataset. In order to generate the final dataset, samples which were found to have duplicated GEO accessions or not to cluster by t-SNE with their published subgroup assignment were removed. A final table of the samples used is provided in the Appendix 8.2 along with a designation for inclusion and a reason for their removal.

A total of 824 HGU133Plus2.0 profiles were analysed from 7 tumour types: ATRT (n = 111, GSE35493, GSE64019 GSE67851, GSE70678, GSE73038 GEO/NIH), ECRT (n = 20, GSE64019 GEO/NIH), Ewing Sarcoma (EWS; n = 103, GSE34620 GEO/NIH), medulloblastoma (MB; n = 214, GSE10327, GSE12992, GSE37418, GSE73038 GEO/NIH), neuroblastoma (NB; n = 137, GSE1623, GSE16476 GEO/NIH), rhabdomyosarcoma (RMS; n = 101, E-TABM-1202 ArrayExpress/EMBL-EBI) and Wilms tumour (WT; n = 138, TARGET/NIH).

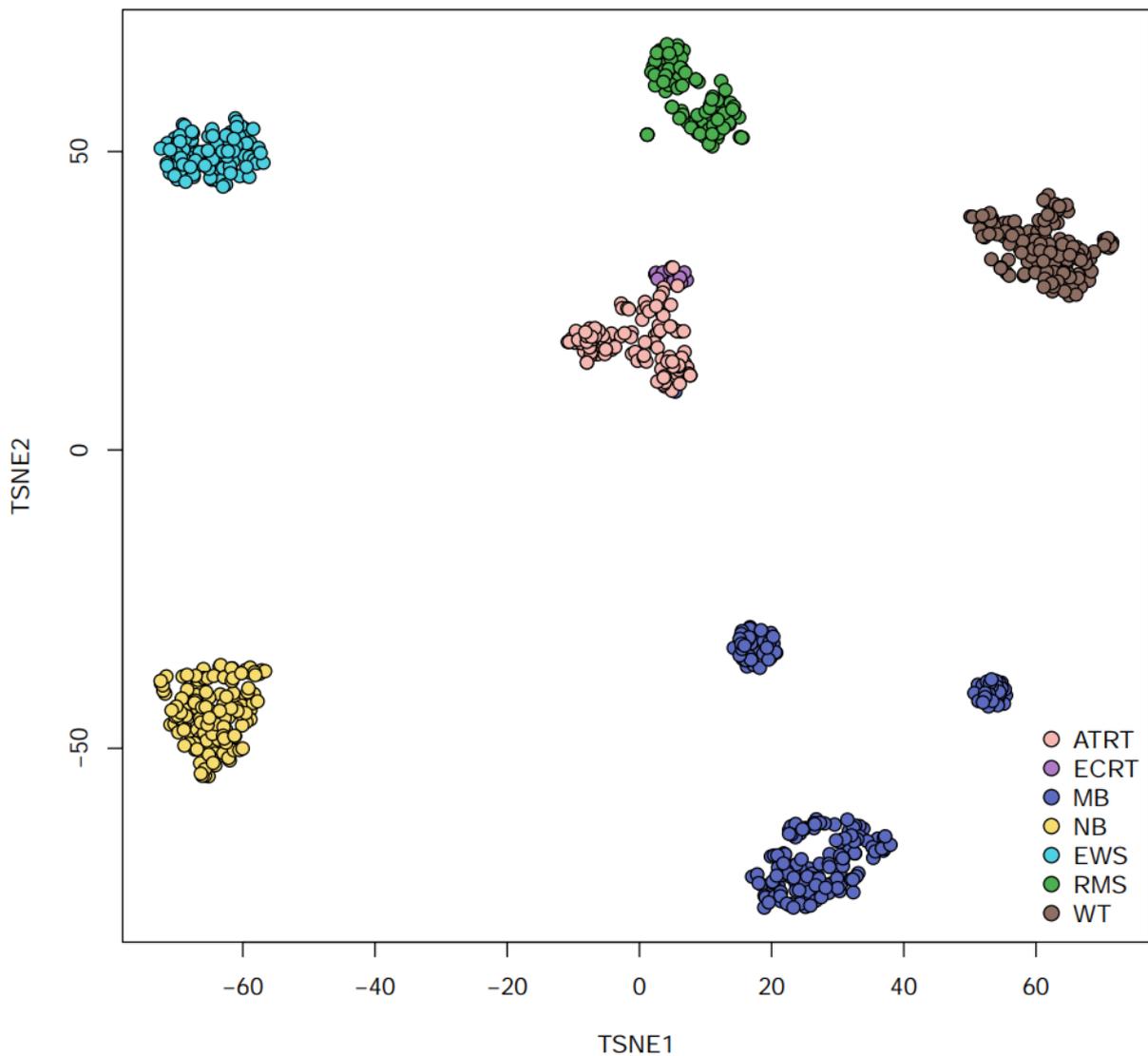


Figure 5. 2D t-SNE embedding plot of 824 embryonal tumour U133Plus2.0 expression array profiles using the log₂-fold expression intensities of 4934 most variable probes. MB: medulloblastoma, NB: neuroblastoma, EWS: Ewing sarcoma, RMS: rhabdomyosarcoma, WT: Wilms tumour.

T-SNE embedding of the 4934 most variable probes (Figure 5) shows each tumour type reliably generates unique clusters. ATRT and ECRT cluster together without visible separation between extra-cranial samples, although the segregated structure suggests ECRT behaving as a sub-type. Importantly, the presence of classical MB sub-groups (WNT, SHH, Grp 3/4) shown by the separation of the MB population into at least 3 distinct clusters suggest that intra- and extra-cranial localisations of MRT have a more related transcriptional programme than sub-groups of MB derived from the infratentorial region of the CNS.

absolute mean $\log_2FC > 1$). The top 100 probes from this comparison are shown in Figure 7.

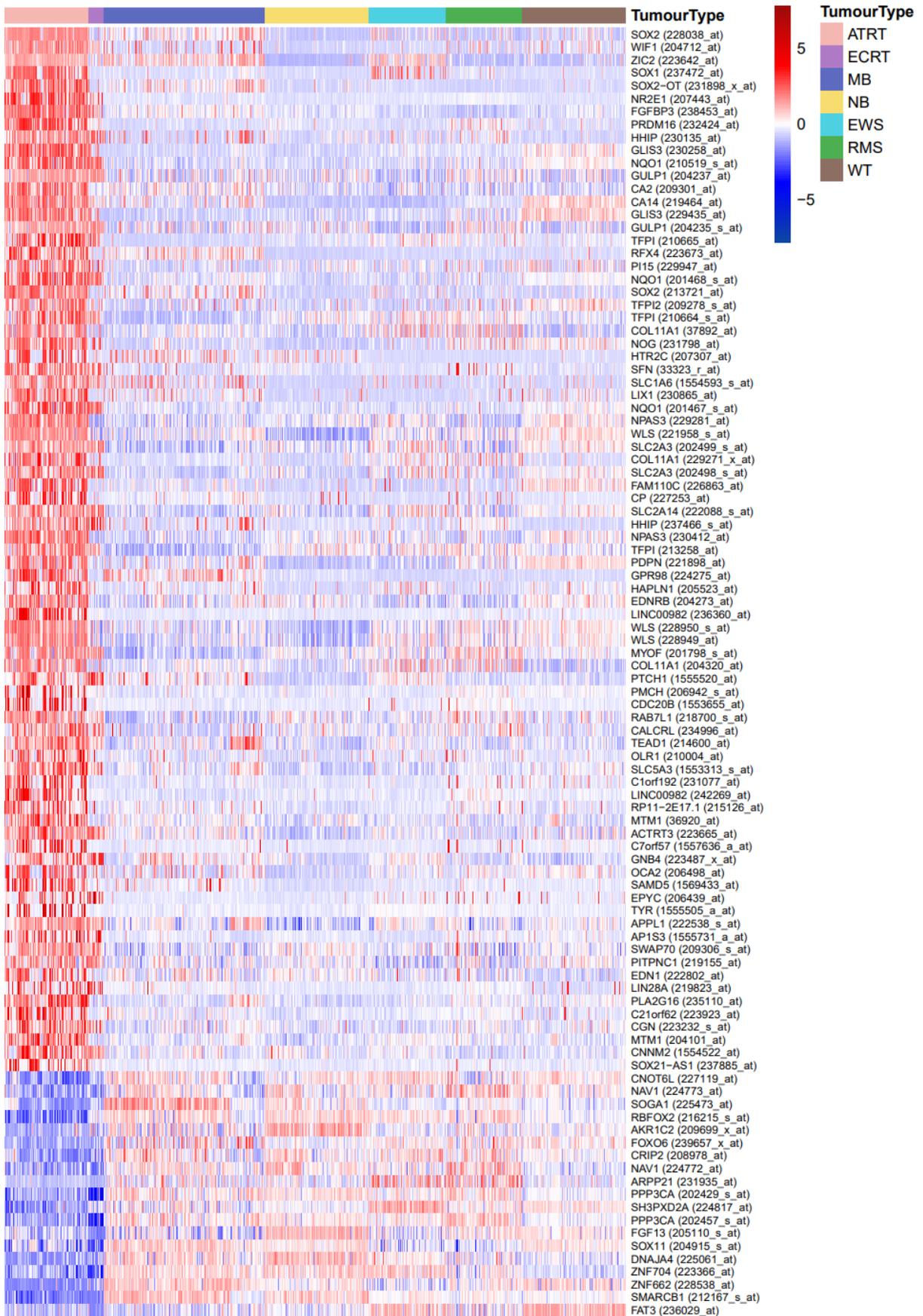


Figure 7. Heatmap of relative gene expression between MRT and other embryonal tumour types. The top 100 differentially expressed probes are shown. Red denotes high expression and blue denotes low expression. The data is scaled by row.

The top differentially expressed genes between MRT and other tumours include *FGFBP3*, *HHIP*, *TFPI2*, *SLC2A3*, *FGF13*, *SOX11* and *FAT3*. Enrichment of genes associated with nervous system development, osteoblast differentiation, and WNT signalling was noted using DAVID highlighting the many lineage markers these tumours express however the highly specific and stringent nature of the analysis likely underestimates differences due to many cancer pathways being shared across multiple tumour types.

3.3.2 Analysis of differences between ATRT and ECRT highlights location-specific features

In order to examine the differences in gene expression between ATRT and ECRT, a supervised differential gene expression analysis was carried out on 131 MRT U133Plus2.0 profiles (ATRT n = 111, ECRT = 20). 553 probes were found to be significantly differentially expressed between the two MRT types ($p < 0.05$, absolute $\log_2FC > 2$). Probes significantly enriched in ATRT included *SOX2*, *FABP7*, *OTX2*, *ASCL1* and *GFAP*, while ECRT expression showed high expression of *HOXA/B/C*, *IGF2* and *MYC*.

Functional annotation using DAVID for the ATRT enriched probes (n = 490) was predominated by pathways/genesets related to neuronal function and neural development. ECRT enriched probes (n = 63) identified a smaller set of pathways/genesets largely associated with embryonal development and skeletal system development. Importantly, among profiles in the ATRT dataset, a small group (n = 12) exhibited expression much more in line with ECRT including high expression of *MYC* and HOC cluster genes.

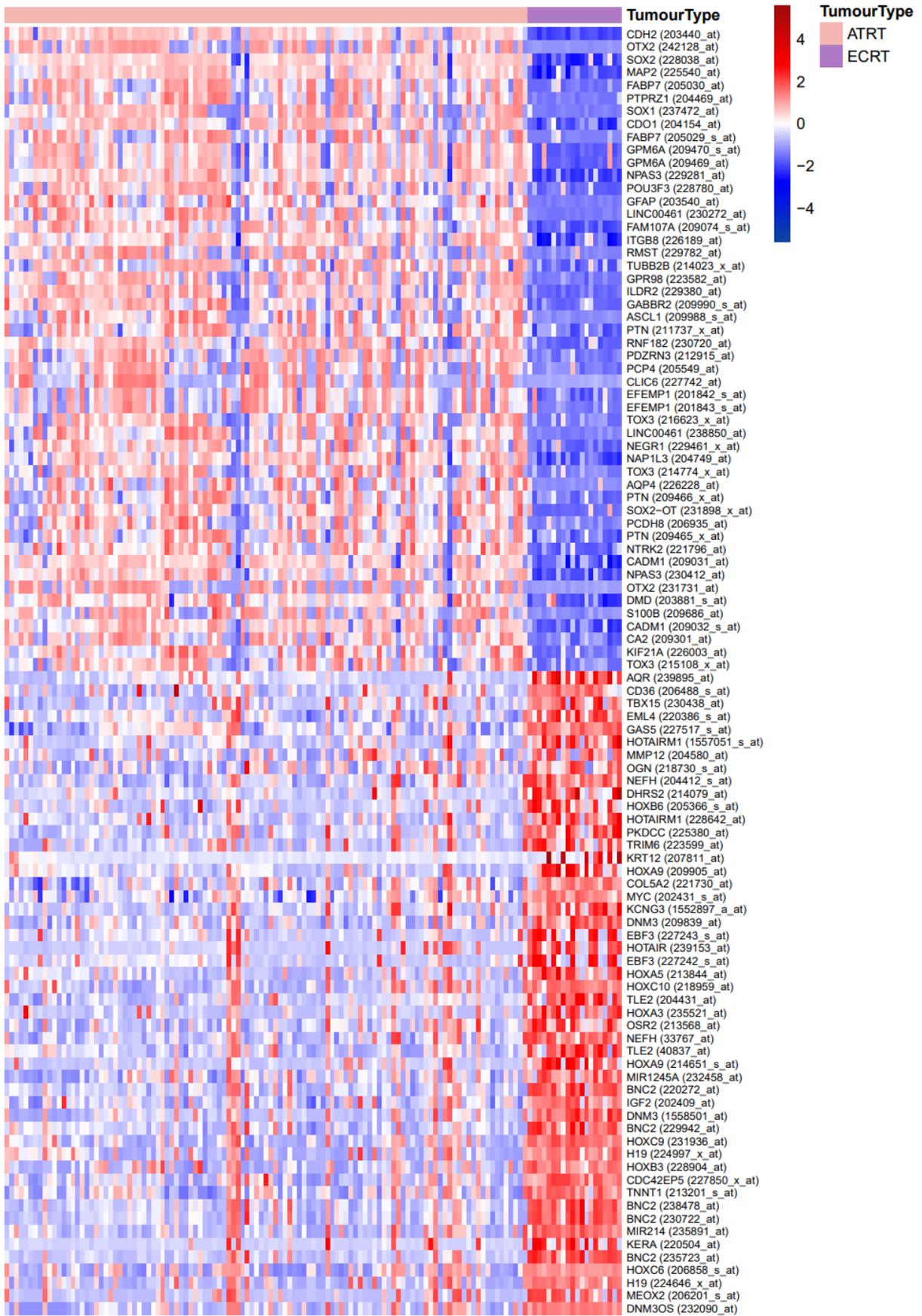


Figure 8. Heatmap of relative gene expression between ATRT and ECRT. The top 100 differentially expressed probes are shown. Red denotes high expression and blue denotes low expression. The data is scaled by row

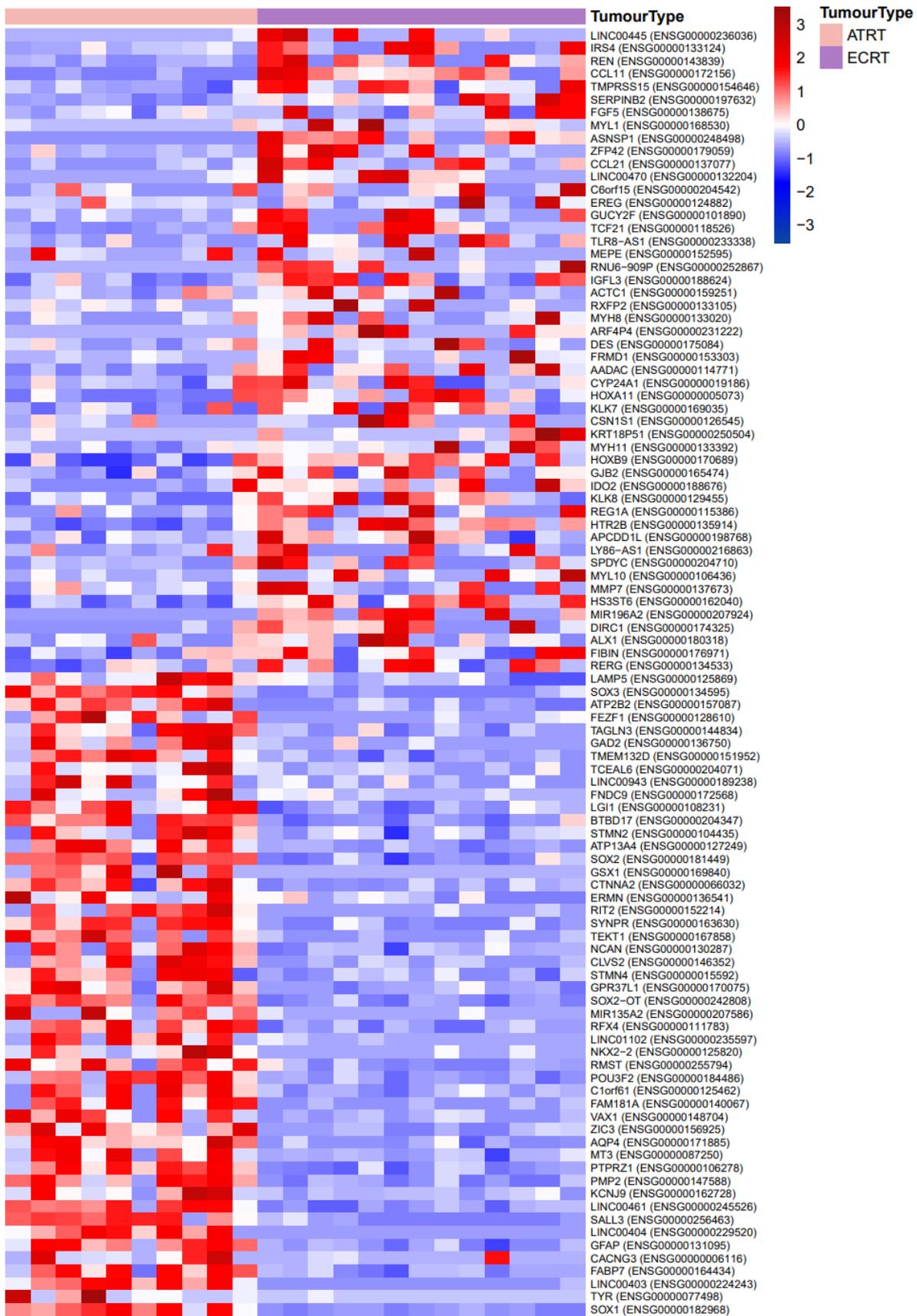


Figure 9. Heatmap of relative gene expression between ATRT and ECRT RNA-Seq. The top 100 differentially expressed genes are shown. Red denotes high expression and blue denotes low expression. The data is scaled by row

Differential expression analysis of an MRT RNA-Seq cohort (ATRT n = 10, ECRT n = 13) was carried out to further interrogate the differences. 2050 genes were found to be differentially expressed between ATRT and ECRT. Genes enriched in ATRT largely supported previous findings including overexpression of *FABP7*, *SOX2*, and *GFAP*, and functional annotation with DAVID of the 1691 genes associated with and overexpressed in ATRT showed a predominance of neural development and function genesets. 359 genes associated with ECRT included HOX cluster genes, and functional annotation showed enrichment of inflammatory response, skeletal and muscle development. A heatmap of the differential expression findings are shown in Figure 9.

3.3.3 Combined MRT sub-grouping strategies recapitulate previous ATRT-only models

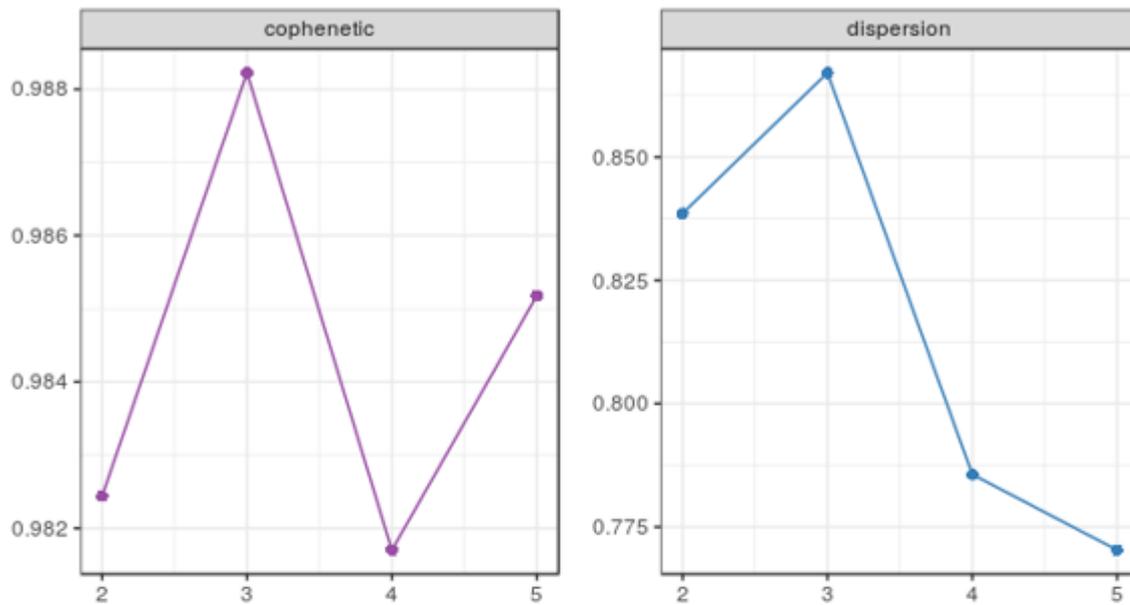
Previous sub-grouping strategies have largely relied on a single MRT type and only one publication has co-clustered ATRT and ECRT tumour profiles (Han *et al.*, 2016). Differential expression between MRT and other embryonal tumours as well as between types of MRT partially supports previous findings that ECRT sub-types may resemble their CNS counterparts and *vice-versa*. In order to investigate the relationship between existing sub-grouping strategies for ATRT and ECRT tumours, unsupervised clustering was carried out using NMF.

Using the published 3 sub-group annotation of ATRT from Johann *et al.* (2016) and ECRT data obtained from public databases, the resulting sub-grouping schemes were compared in order to ascertain whether inclusion of ECRT caused previously defined sample subgroup to drastically change and whether the addition of ECRT would simply create a separate sub-group such as hEC in Han *et al.* (2016).

Figure 10 shows the NMF clustering quality metrics derived from two consensus NMF unsupervised clustering results. Figure 10A shows the results obtained from clustering 49 ATRT profiles from from Johann *et al.* (2016), Figure 10B shows the results from the combined 131 MRT dataset. Both the cophenetic coefficient and dispersion index in A indicate a 3-group result as the most optimal, while in B both measures steadily decline as the number of sub-groups is increased. However, when comparing the resampled group calls between shared samples between the two clustering models (Table 3) only 4% (2/49) of samples were reclassified. In other words, although the measures of cluster robustness by the NMF package do not favour any result above 2 groups in the MRT NMF dataset, the resulting resampled group allocations are almost

identical between the two datasets showing that addition of additional samples including ECRT into the dataset did not significantly alter the existing sub-grouping but allowed to also assign sub-groups to ECRT and ATRT together.

A



B

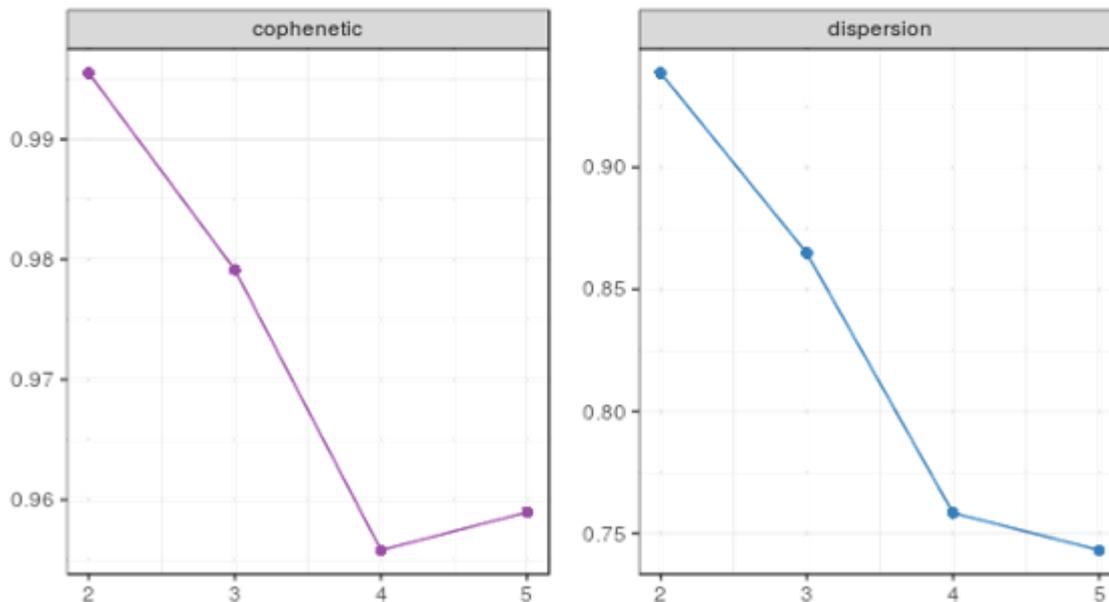


Figure 10. NMF cluster quality metrics for A) clustering of 49 samples from 49 ATRT only B) clustering of 131 MRT profiles. The cophenetic correlation coefficient and the index of dispersion are shown.

GSE70460 NMF

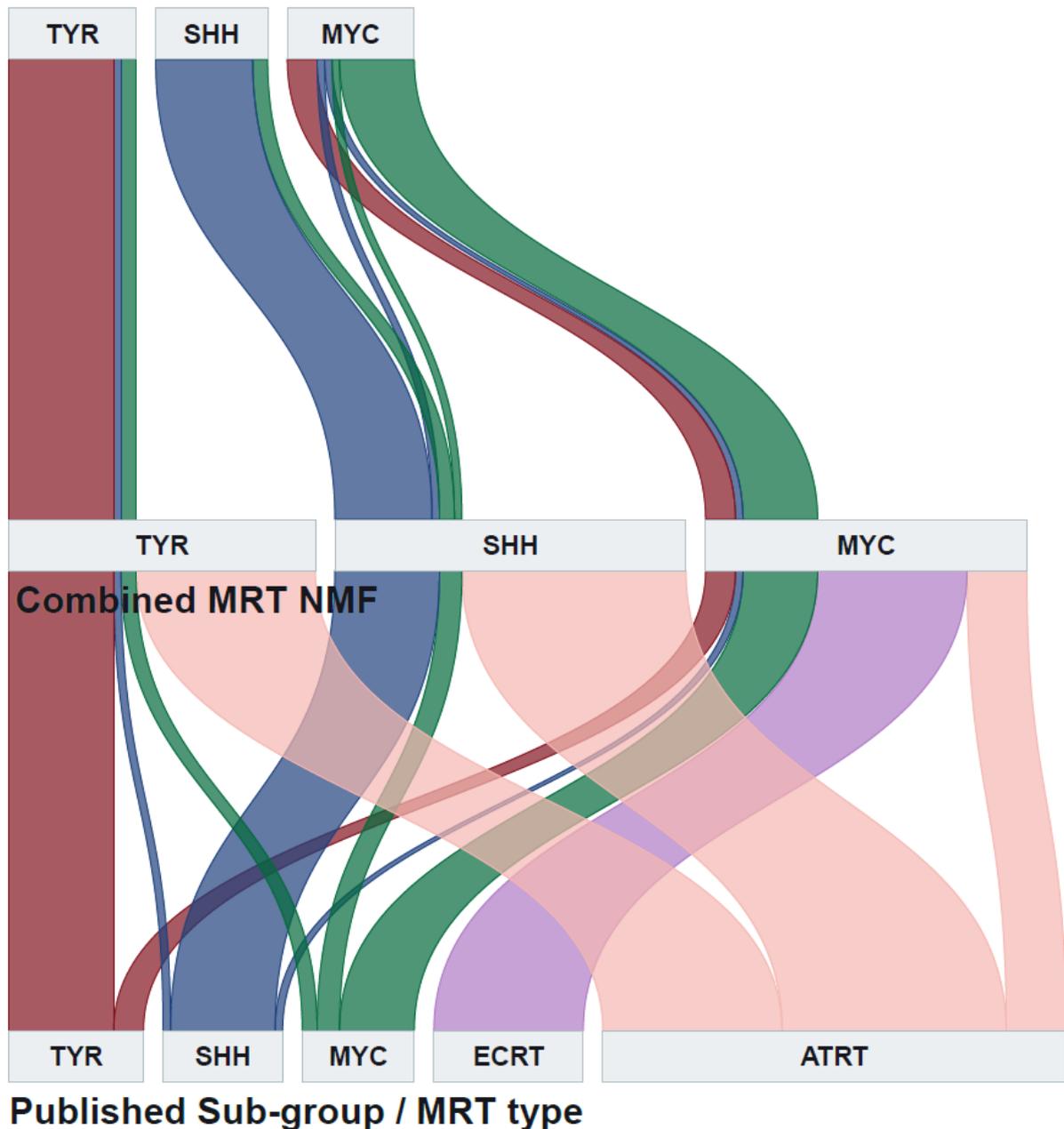


Figure 11. Sankey plot comparison between different NMF clustering cohorts 49 ATRT-only and combined MRT and the published classification from Johann et al. (2016) or the MRT localisation derived from the expression array dataset.

In the combined MRT expression cohort, all additional 20 ECRT samples clustered to the MYC sub-group. Notably, there was more disagreement between the NMF methods used in this analysis and the published group assignments based on hierarchical clustering with 22% (11/49) samples being reassigned to a different sub-group – most commonly moving between MRT and TYR. It can be assumed that the differences are largely caused by the use of different clustering strategies and highlights a need for a consensus approach to sub-grouping in MRT.

However, is possible to combine data from all MRT localisations and still obtain a robust sub-grouping result that is not skewed due to the addition of a different tissue type.

	<i>Published Calls</i>		
<i>49ATRT only NMF</i>	SHH	TYR	MYC
SHH	13	1	2
TYR	0	14	4
MYC	2	2	11
Agreement = 78%, disagreement 22%			

	<i>Published Calls</i>		
<i>131 MRT NMF</i>	SHH	TYR	MYC
SHH	14	1	1
TYR	0	14	4
MYC	3	2	10
Agreement = 78%, disagreement 22%			

	<i>131 MRT NMF</i>		
<i>49ATRT NMF</i>	SHH	TYR	MYC
SHH	15	0	2
TYR	0	17	0
MYC	0	0	15
Agreement = 96%, disagreement 4%			

Table 3. Comparison of agreement and disagreement between 3 clustering approaches: Published calls from Johann et al. (2016), 49ATRT-only NMF and combined MRT NMF.

3.4 Discussion

Currently, there is no definitive sub-grouping strategies that combine both ATRT and ECRT. Previous publications have shown that sub-groups of MRT can resemble their distally localised counterparts in expression profiles and DNA methylation, but despite a number of large genomic studies, MRT sub-types are still analysed and studied separately. The findings in this chapter highlight the relatedness of ATRT and ECRT by showing that differences between the localisations appear to closer resemble disease sub-groups than distinct tumour types. In addition, the expression programme of MRT appears to be more closely related than sub-groups of other CNS tumours

largely derived from a single location in the brain. This chapter presented an MRT-specific expression signature that was derived by comparing MRT against other embryonal tumours in a stringent differential expression analysis. Combined MRT subgrouping schemes were compared with those derived in only a single tumour type and it was shown that despite a reduction in the resulting clustering metrics, resampled subgroup calls from consensus NMF did not differ on the addition of ECRT and the clustering was able to accommodate tumour types from different parts of the body without clustering artifacts

Differences between types of MRT appear to be largely related to the typical tissues they are found in. ATRT overexpress genes associated with neuronal function and neural development while ECRT display a less cohesive expression programme consistent with the differing localisations. Notably, it was not possible to expand the ECRT proportion of the cohort and it can be considered a limitation of the analysis as the differential expression between embryonal tumour types and also between ATRT and ECRT will be skewed towards the larger ATRT component. Nevertheless, it was still shown that ECRT contributed unique expression markers including a strong expression of HOX cluster genes and *MYC*.

The findings highlighted in this chapter suggest that future sub-grouping approaches should include both ATRT and ECRT and that moving forward, there should be a concerted effort to further delineate the relationship between these two tumour sub-types by expanding the size and quality of the available profiling cohorts. In addition, the results point to a common biology shared by all MRT which could play a role in future therapeutic strategies. However, there is considerable heterogeneity between MRT cases and the relatively successful combined clustering approach suggests that rather than focusing on sub-types as defined by tumour location, MRT heterogeneity could perhaps be better defined as the result of disease subgroups, which have already been suggested by a number of publications.

4 Generating a molecular signature of MRT sub-groups encompassing methylation and expression features

4.1 Introduction

Restricting analysis to only examine differences between localisations of MRT does not account for the full extent of the heterogeneity seen in the disease. Chapter 3 already discussed how sub-sets of ECRT can resemble the CNS disease and vice-versa, however differences in markers expressed by the tumours as reported in literature, variability in response to therapies and ultimately patient survival also point towards potential disease sub-groups. As well as explaining the currently observed tumour heterogeneity, sub-grouping can further reveal novel aspects of the tumour biology by highlighting differential gene, pathway, and epigenetic features. It can allow for disease prognostication and even identify new therapeutic targets and better inform our current approach to treating MRT.

Table 4. Summary of subgrouping approaches from recent MRT subgrouping publications

Study	MRT Type	Subgroups			
		Group 1		Group 2	
Torchia <i>et al.</i> , (2015)	ATRT	overexpression of <i>ASCL1</i>			
Han <i>et al.</i> , (2016)	ATRT & ECRT	hIC1	hIC2	hIC3	hEC
		Overexpression of <i>OTX2</i> , <i>ODZ2</i> , <i>BMP4</i> , <i>SMAD7</i>	Overexpression of <i>SOX2</i> , <i>POU3F1/2</i> , <i>ASCL1</i> , <i>HES5</i> , <i>BOC</i> , <i>GLI2</i>	Overexpression of <i>GFAP</i> , <i>FABP7</i>	Overexpression of <i>HOXA/C</i> , <i>TBX2</i> , <i>TGFB</i> , <i>MYC</i>
Johann <i>et al.</i> , (2016)	ATRT	ATRT-SHH	ATRT-TYR	ATRT-MYC	
		Overexpression of SHH pathway genes (<i>GLI2</i> , <i>BOC</i> , <i>PTCHD2</i>), <i>MYCN</i>	Overexpression of melanosomal genes (<i>TYR</i> , <i>TYRP</i> , <i>MITF</i> , <i>OTX2</i>)	Overexpression of <i>MYC</i> and HOX cluster genes	
Chun <i>et al.</i> , (2016)	ECRT	Group 1		Group 2	
		Overexpression of immune genes, BMP pathway signaling (<i>BMP4</i>). Extra-renal ECRT enriched		Overexpression WNT signaling genes (<i>WNT5A</i> , <i>HIC1</i>) RTK enriched	
Torchia <i>et al.</i> , (2016)	ATRT	Group 1	Group 2A	Group 2B	
		Overexpression of NOTCH pathway genes (<i>ASCL1</i> , <i>CBL</i> , <i>HES1</i>)	Overexpression of neuronal genes and mesenchymal genes (<i>OTX2</i> , <i>PDGFRB</i> , <i>BMP4</i>)	Overexpression of <i>HOX</i> cluster genes	

A number of sub-grouping schemes for MRT have been proposed. They differ by the type of MRT used in analysis, analysis carried out, sub-group number, and the sub-group definitions. A summary of the currently available sub-grouping strategies is outlined in (Table 4). Before beginning of this study in 2015, a sub-grouping scheme was proposed for ATRT only by (Torchia *et al.*, 2015). Sub-grouping was carried out on Illumina HT-12 v4.0 expression array analysing 43 primary tumours. The authors reported two groups as identified by unsupervised hierarchical clustering and NMF, noting that one of the groups showed additional heterogeneity but acknowledged small cohort size as a limitation of confidently classifying more than two clusters. Sub-grouping was validated with IHC assay for expression of ASCL1, a member of the NOTCH pathway identified to be differentially expressed between proposed sub-groups. Low expression of this marker had a significantly higher associated risk and worse survival and was the first study to extract prognostic information from ATRT sub-grouping.

Subsequently, in early 2016, a sub-grouping study comprising both ATRT (n = 30) and ECRT (n = 20) was carried out by (Han *et al.*, 2016) as part of a wider comparison between a mouse model of MRT and the human disease. Hierarchical clustering with resampling using the ConsensusClusterPlus R package identified 4 MRT sub-groups separate to medulloblastoma and neuroblastoma comparators. The three described intra-cranial groups (hIC1-3) and one extra-cranial group (hEC) were further explored for differential gene expression as well as correlation with expression profiles for various stem cell and progenitor populations. hIC2 was shown to correlate strongly with neuro-epithelium and other early neuronal lineages. Lower correlation was observed in the other sub-groups, with hIC3 showing weak correlation in all comparisons likely attributed to the fact the group only contained 5 samples. hEC correlated strongly with embryonal stem cell populations. Differential gene expression analysis highlighted similar features in the sub-groups with hIC2 showing high expression of genes associated with early neural development *SOX2*, *POU3F1/2* and *ASCL1*, hIC1 showing lower levels of expression of these genes but high BMP pathway gene expression, and hIC3 showing expression of glial genes such as *GFAP*. The main feature of hEC was the lack of expression of neural progenitor genes, high *MYC* and *HOXA/C* gene expression as well expression of various cytokines such as *TGFBR2*, *TGFBR3*. Although this study did not include any prognostic or survival annotation, it generated a sub-grouping scheme comparing all types of MRT and provided an

overview of expression differences that hint at multiple possible cells of origin of the tumours.

Two separate publications published in the same journal issue examined ATRT and ECRT tumours, respectively. (Johann *et al.*, 2016) carried out clustering on 150 ATRT 450K methylation and 49 ATRT HGU133Plus2.0 expression array profiles using ConsensusClusterPlus. The analysis identified 3 sub-groups as the most favourable clustering result with a high degree of concordance between the two platforms (88%, 23/26) for samples with both data types available. They termed the sub-groups “ATRT-SHH”, “ATRT-TYR” and “ATRT-MYC”, with the SHH group showing high SHH signalling with overexpression of *MYCN* and *GLI2*, TYR sub-group containing the majority of patients <1 year old and showing high expression of the *TYR* and *MITF* genes and MYC overexpressing the gene *MYC*. Importantly, unlike the SHH sub-group of MB, ATRT-SHH harboured no aberrations in any SHH-pathway genes. The authors also noted that the ATRT-SHH sub-group could be further clustered into sub-types by methylation, suggesting additional heterogeneity in this cluster. Whole-genome DNA (WGS) and RNA-Seq for 18 samples was carried out but did not identify any additional mutations aside from in *SMARCB1* varied across sub-group. The authors described differences in activating mutations across the proposed sub-groups with ATRT-TYR showing broad chromosome 22q deletions which were not prevalent in the other sub-groups. In addition to this, all sub-groups were shown to harbour high levels of whole-genome and promoter-specific DNA methylation compared to other embryonal tumours, with ATRT-TYR showing the highest of the three groups.

(Chun *et al.*, 2016) carried out RNA-Seq, WGS and microRNA sequencing of ECRT including both RTK and extra-renal tumours. From NMF clustering of 40 RNA-Seq profiles, they identified 2 stable sub-groups. Group 1 comprised tumours from all ECRT locations, contained older patients (50% >1 year old) and had overexpressed genes associated with immune function, and BMP-signalling. Group 2 contained only RTK cases, was enriched for younger patients (72% <1year old) and overexpressed WNT-signalling genes. The study also compared microRNA profiles from ECRT with other tumour and normal cell types identified two groups – one clustering with synovial sarcoma and one that clustered with normal cerebellum and neural crest cell tumours. Taken together, these results suggest that at least a sub-set of ECRT share regulatory and transcriptomic features with ATRT once again showing that sub-groups in MRT

potentially cut across tumour localisations rather than being strictly dictated by anatomical site.

In late 2016, Following on from the 2015 study which identified 2 ATRT sub-groups with differing survival, (Torchia *et al.*, 2016) published an expanded ATRT sub-group scheme identifying 3 ATRT sub-groups through a combined analysis of DNA methylation and gene expression array. The authors noticed similar age distribution, SMARCB1 alterations and gene expression features to (Johann *et al.*, 2016) with Group 1 resembling ATRT-SHH, Group 2A ATRT-TYR and Group 2B ATRT-MYC. The authors went on to demonstrate that Group 2 tumours were sensitive to dasatinib and nilotinib as well as validating other downstream targets.

In a relatively short space of less than 2 years, 5 different sub-grouping approaches have been suggested for MRT. The schemes share some parallels but there is no consensus on the number of sub-groups, their molecular features and if these groups present different prognostic and therapeutic opportunities. This situation is further complicated by examples of subsequent literature utilising varying sub-grouping strategies, as well as a DNA methylation-based molecular sub-group classifier for CNS tumours having been developed (Capper *et al.*, 2018) that uses only one sub-grouping approach.

There appears to be a clear need for a consensus sub-grouping approach in MRT not only to generate a more complete understanding of the different features of sub-groups presented by the array of profiling and analyses carried out, but also to provide a single, unified platform that informs future clinical and biological research in this tumour type. In order to address this issue partially, an international collaboration is currently underway to generate a consensus ATRT-only sub-grouping approach to which the analysis presented here contributed in part and the manuscript currently in peer-review is attached. However, no current such consensus analysis currently exists for MRT as a whole. Therefore in this chapter, an expanded consensus analysis encompassing all types of MRT is presented.

4.2 Aims

The aims of this chapter were to, firstly, carry out a meta-analysis of publicly available and newly-profiled MRT datasets in order to contrast and compare the current sub-grouping schemes in MRT. The optimum number of sub-groups was assessed, as well as analysing the gene expression and DNA methylation array molecular signature and

clinicopathological features of each group in order to provide a recommendation consensus MRT sub-grouping strategy.

Secondly, the inclusion of ECRT and additional newly-profiled ATRT profiles were used to investigate whether additional sub-groups may be identified in MRT, and further defined by clinicopathological features.

4.3 Differential methylation and expression analysis of MRT sub-groups reveals group-specific changes

4.3.1 Meta-analysis of gene expression array

HGU133Plus2.0 gene expression array data comprising 131 primary tumours (ATRT n = 111, extra-renal ECRT n = 16, and RTK n = 4, see Methods 2.1.3) was initially subject to consensus NMF clustering described in Methods 2.3.4.2 in order to generate a single comparable sub-group assignment across datasets. Clustering identified 3 stable clusters based on clustering metrics shown in **Error! Reference source not found.A-E** with 95% (125/131) of samples being assigned a consensus NMF call. 5% (6/131) of samples were assigned as NC due to failing to reliably cluster to a single metagene. Nomenclature for the consensus NMF clusters was chosen as SHH, TYR, MYC and non-classifiable (NC) due to this scheme being the chosen nomenclature proposed in the current ATRT consensus meta-analysis.

Comparing the (Han *et al.*, 2016) sub-grouping to the consensus NMF result for samples obtained from the study shows a considerable degree of overlap between the two schemes. hIC2 completely overlaps with the consensus NMF group SHH, hIC1 has 84.6% (11/13) overlap with group TYR with 2 samples being assigned to hIC2 and the hEC group completely overlaps with MYC. 2 samples from the hIC3 group were assigned to SHH, 1 sample to MYC and the other 2 samples in this group failed to gain a group assignment (**Error! Reference source not found.F**). It is not clear whether group hIC3 is a group of outlier samples, a result spurious clustering or is indeed a genuine sub-group that is simply too small to reliably cluster the data. However the strong concordance between the (Han *et al.*, 2016) sub-groups and the group assignments from the consensus NMF analysis suggests a direct parallel between the two schemes, especially given they were both generated with a method based on NMF clustering.

Comparing the (Johann *et al.*, 2016) sub-groups to the group assignments generated by consensus NMF also highlights shows a high degree of overlap. 87.5% (14/16)

of ATRT-SHH overlap with the consensus NMF SHH group, 77.8% (14/18) of ATRT-TYR overlaps with consensus NMF TYR and 66.7% (10/15) ATRT-MYC overlaps with consensus NMF MYC (Figure 12F). As discussed in Chapter 3, the lower level of concordance likely originates from the difference in clustering method since the authors derived their group assignments using ConsensusClusterPlus. Importantly, only 1 sample from this dataset failed to reliably gain a group assignment with the consensus NMF approach. This highlights a reason for carrying out consensus meta-analysis to generate a robust dataset with reliable calls to facilitate classification of future MRT profiles.

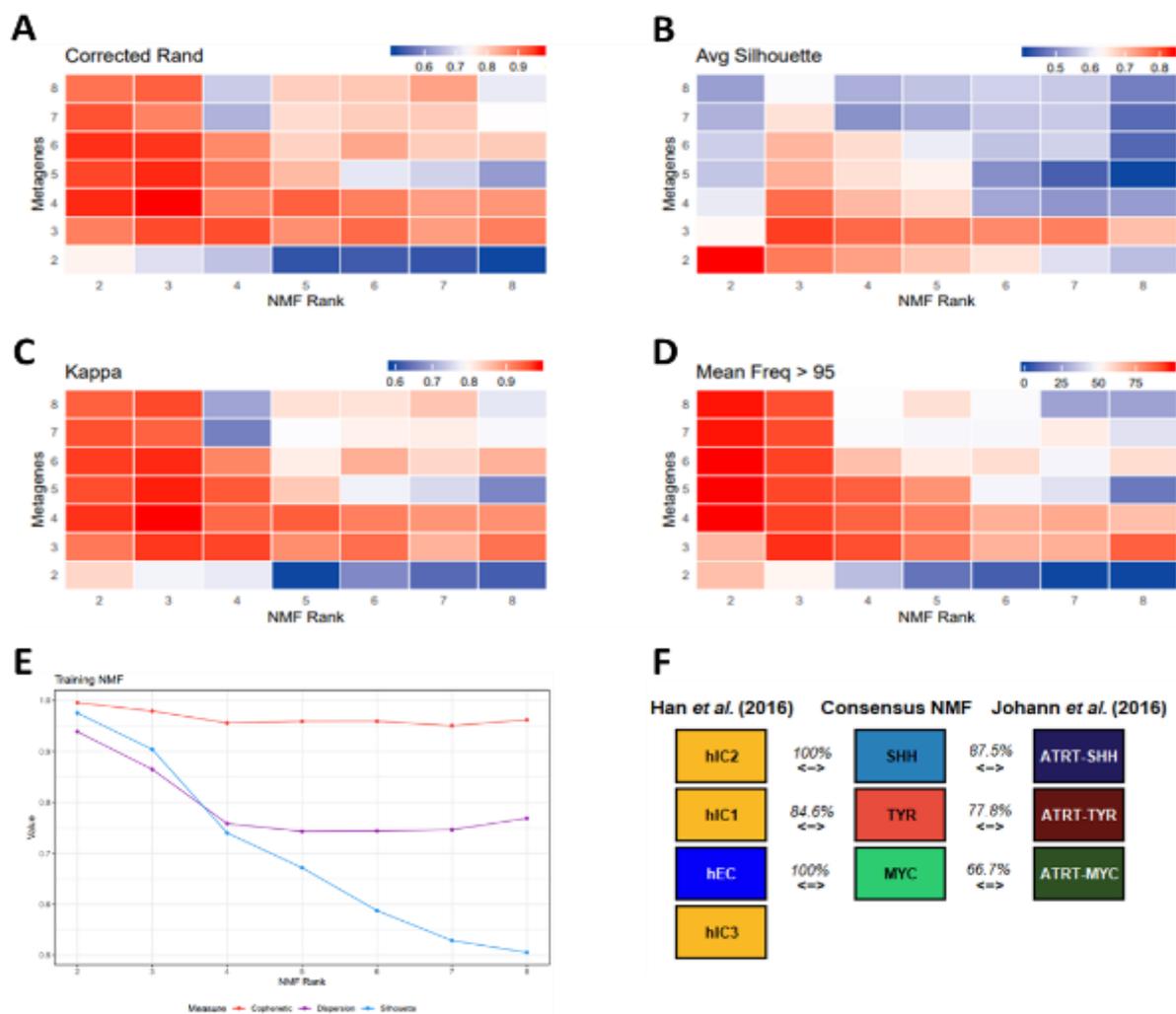


Figure 12. Consensus NMF cluster metrics from clustering HGU133Plus2.0 gene expression array data comparisons are made across every NMF rank and every combination of metagenes A) Corrected Rand index calculated from k-means clustering of projected NMF metagenes, red denotes higher similarity between clustering iterations B) Average silhouette from k-means clustering of projected NMF metagenes red denotes higher silhouette score C) Cohen's kappa calculated from k-means clustering of projected NMF metagenes, red denotes higher level of agreement across iterations D) Percentage of samples in dataset which were assigned a group call with greater than 95% frequency following resampling, red denotes more samples receiving robust call E) Training NMF cluster metrics red denotes cophenetic correlation, purple denotes dispersion index and blue denotes the silhouette score F) Agreement between previously published sub-grouping schemes for comparable samples, percentages indicate the number of samples where the previously published sub-group corresponds to the Consensus NMF result, non-classified samples were excluded

Examining the clinicopathological differences between the consensus NMF sub-groups highlights sub-group specific features (Figure 13). The TYR group shows the lowest age (mean = 17.5 months, maximum = 51.6 months) with SHH showing an intermediate age distribution (mean = 29 months, maximum = 120 months) and MYC showing older patients (mean = 35.2, maximum = 114). Only 60% of the cohort had any patient age data available, 29/131 had continuous data available and 50/131 had discrete data as 3 age categories. In order to carry out statistical analysis between sub-groups, all available data was converted to compatible categorical data with groups <2 years old, 2-5 years old and \geq 5 years old. Chi-square testing did not show a significant difference between sub-groups when analysed in this way.

The MYC sub-group contains all extra-cranial tumours and a sub-set of ATRT (n = 11). Chi-square testing CNS location across sub-groups showed a significant enrichment of infratentorial tumours in the TYR sub-group (n = 17/19, p = 0.025). Supratentorial tumours were distributed approximately equally across SHH and MYC (SHH n = 6/12, MYC n = 5/11 ATRT). Published mutational data was limited but chi-squared testing showed significant enrichment of focal *SMARCB1* deletion (defined as loss of exons 1-9) in the MYC sub-group (n = 20/37, p = 0.023)

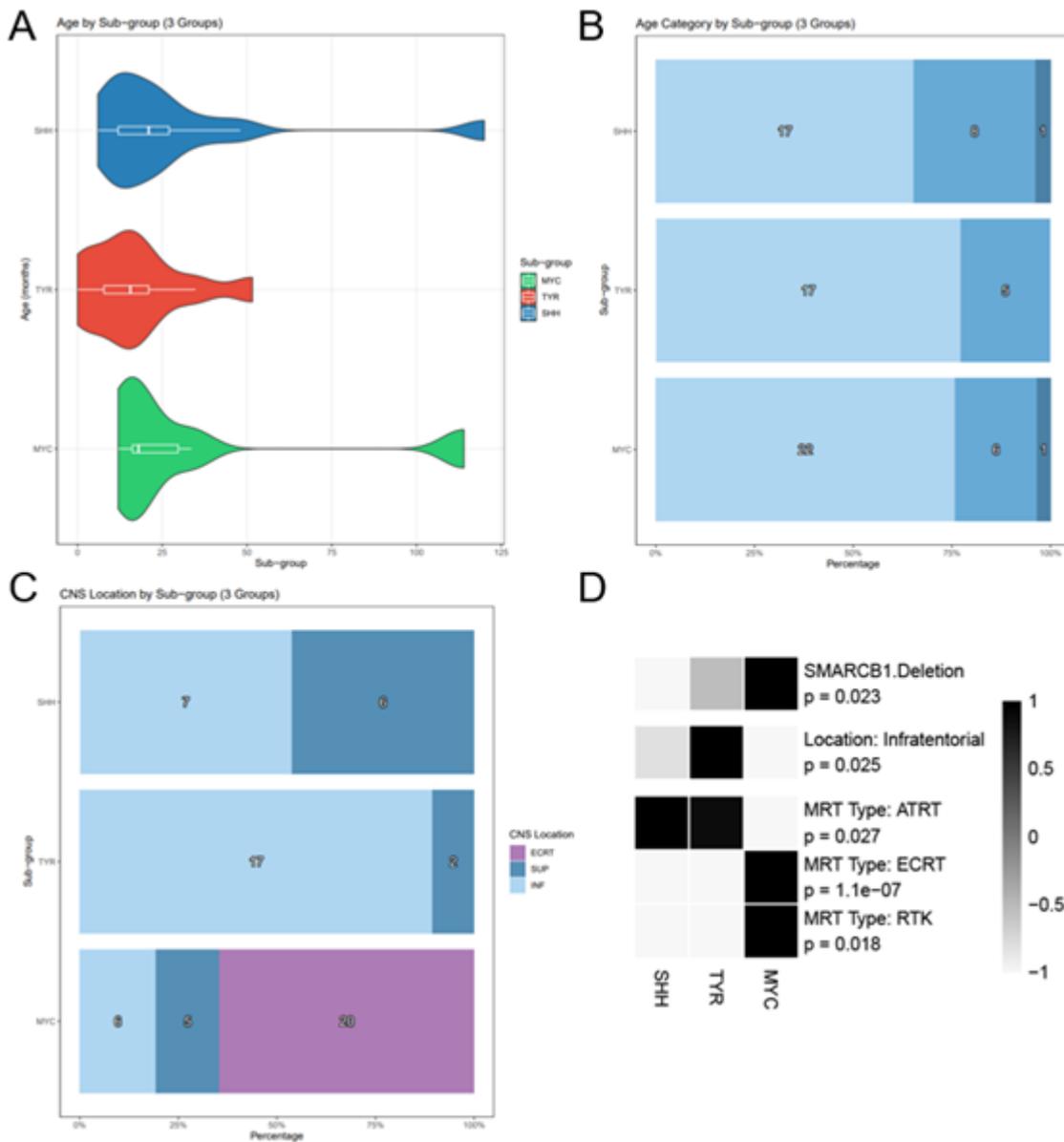


Figure 13. Sub-group clinicopathological features as defined by consensus NMF for HGU133Plus2.0 MRT data A) Violin plot of age distribution across sub-groups B) Age distribution by sub-group; C) CNS location by sub-group INF = infratentorial, SUP = supratentorial. B,C Data is shown as a proportion of the total, missing values are removed. D) Chi-square test residuals for significant comparisons.

Differential expression analysis of gene expression array data was carried out in order to examine differences between MRT sub-groups. Gene expression was compared using the package limma (R/Bioconductor) for each group in relation to the other two. The moderated t-statistic was obtained and used to generate a ranked gene list for the basis of GSEA analysis. A summary of significant GSEA results is shown in Figure 14.

As previously published, the SHH sub-group expresses high levels of neural lineage genes such *ASCL1*, *HRS1*, *DTX1* and *NOTCH1* of the NOTCH pathway; and *GLI2*, *PTCH1*, *BOC* of the SHH pathway, as well as *MYCN*. GSEA highlighted an enrichment of neuronal differentiation genesets (GO SPINAL CORD DEVELOPMENT, NES 2.08,

q = 0.002; GO HIPPOCAMPUS DEVELOPMENT, NES = 2.07, q = 0.002; GO CELL MORPHOGENESIS INVOLVED IN NEURON DIFFERENTIATION, NES = 2.00, q = 0.004) and SHH signalling (HALLMARK HEDGEHOG SIGNALING, NES 1.88, q < 0.001). Interestingly, SHH demonstrated a significant enrichment of genesets associated with active DNA replication and cell division (GO DNA STRAND ELONGATION INVOLVED IN DNA REPLICATION, NES = 2.34, q < 0.001; HALLMARK E2F TARGETS, NES = 2.76, q < 0.001). It is unclear as to the biological significance of this enrichment and further investigation would be required to define the relationship with any clinical associations.

The TYR sub-group shows high expression of the *TYR* gene as highlighted in other publications. In addition, there was noted overexpression in BMP pathway genes such as *BMP4*, developmental transcription factors such as *OTX2* and melanocyte – promoting *MITF*. GSEA highlighted enrichment of epithelium-associated genesets (HALLMARK EPITHELIAL MESENCHYMAL TRANSITION, NES = 2.02, q < 0.001; GO AXONEME ASSEMBLY, NES = 2.35, q < 0.001) suggesting an association of the TYR group with neuroepithelium, a feature previously noted in the hIC1 from (Han *et al.*, 2016).

The MYC sub-group shows high expression of the *MYC* as well as a number of HOX-cluster genes including *HOXC10*. Geneset enrichment highlighted a large number of immune activation associated genesets (GO ACTIVATION OF IMMUNE RESPONSE, NES = 2.70, q < 0.001; GO INFLAMMATORY RESPONSE, NES = 2.63, q < 0.001) as well as highlighting a mesenchymal component in this sub-group (HALLMARK EPITHELIAL MESENCHYMAL TRANSITION, NES = 2.52, q < 0.001). The recurring immune association highlights an interesting feature that is not present in the other MRT sub-groups. It is not clear what biological role is played by the immune system in MRT, and especially in ATRT, therefore it would be an important target for future investigation.

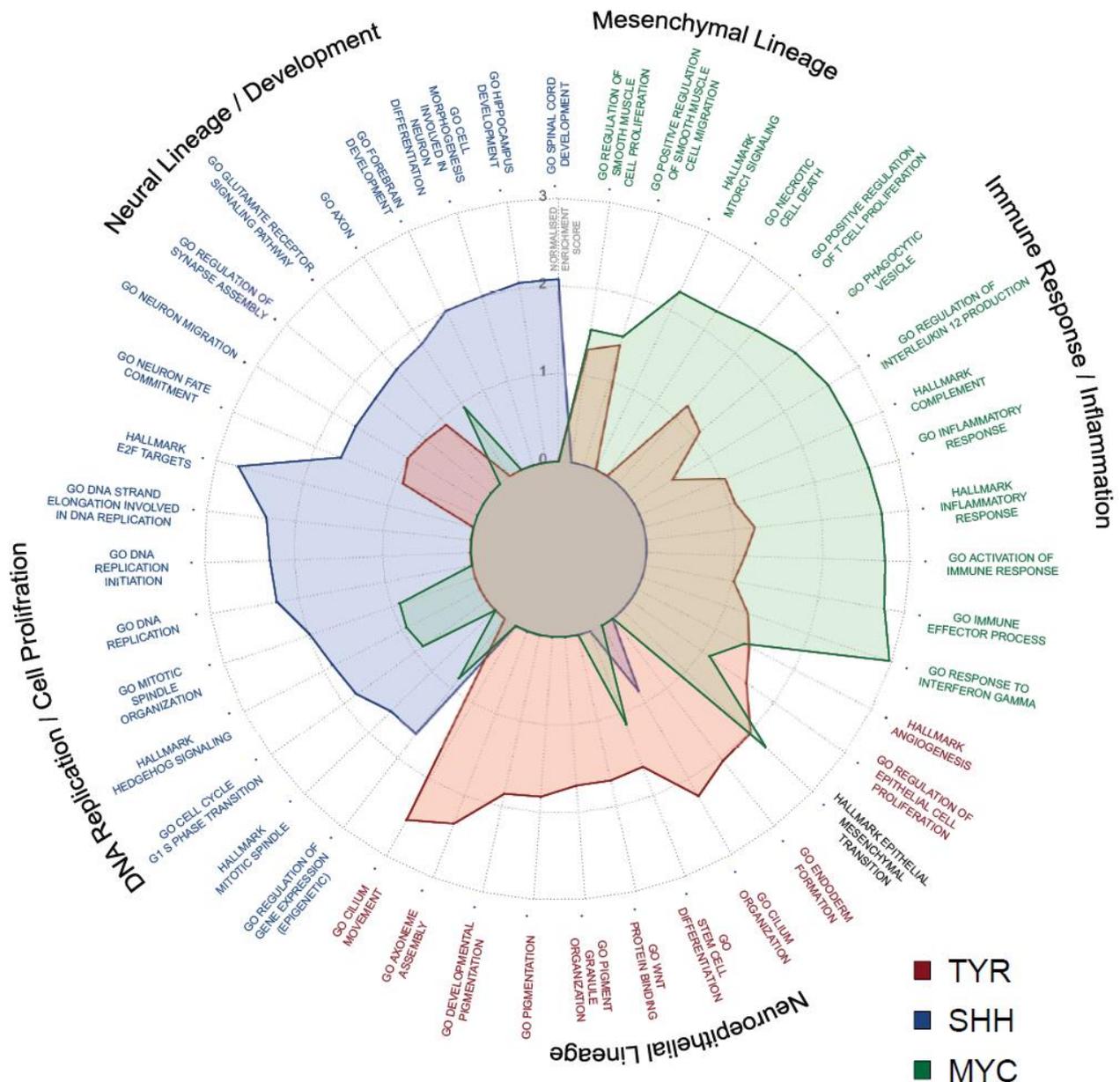


Figure 14. Radar plot showing the genesets identified to be significantly enriched across MRT sub-groups in HGU133Plus2.0. The Normalised Enrichment Score (NES) is plotted. Genesets are coloured by their associated significantly enriched sub-group. Red = TYR, blue = SHH, green = MYC, black = significant in both TYR and MYC differential expression analyses with different genes enriched.

4.3.2 Meta-analysis of methylation array data

Clustering of DNA methylation array data comprising 263 primary tumours (ATRT n = 213, ECRT n = 21, RTK n = 19, MRT n = 10, see Methods 2.1.3) was subject to consensus NMF clustering. This identified 3 stable clusters (Figure 15) with 258/263 samples being assigned a consensus NMF call. 5 samples were not assigned a sub-group due to lack of robust cluster assignment.

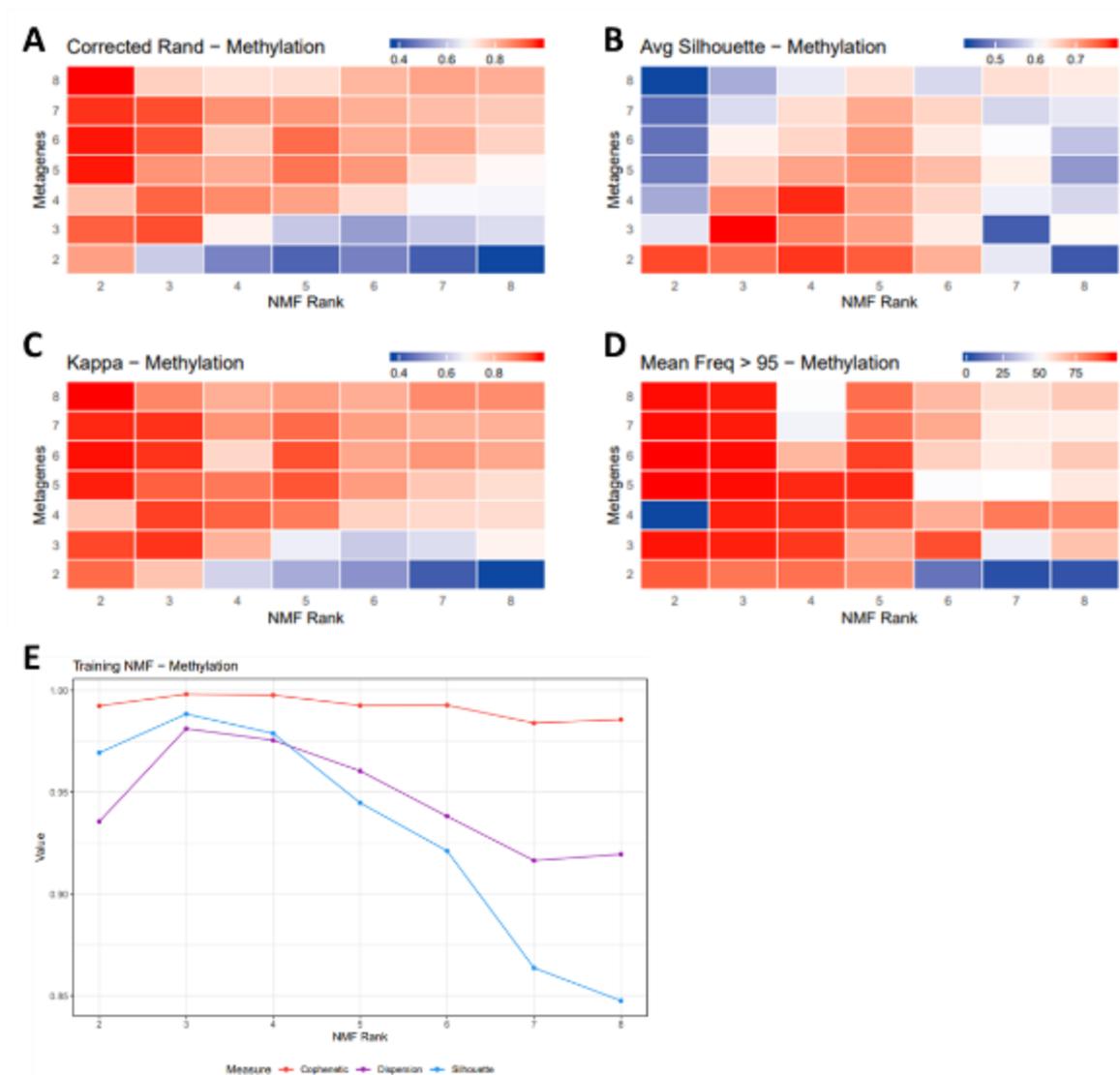


Figure 15. Consensus NMF cluster metrics from clustering DNA methylation array data comparisons are made across every NMF rank and every combination of metagenes A) Corrected Rand index calculated from *k*-means clustering of projected NMF metagenes, red denotes higher similarity between clustering iterations B) Average silhouette from *k*-means clustering of projected NMF metagenes red denotes higher silhouette score C) Cohen's kappa calculated from *k*-means clustering of projected NMF metagenes, red denotes higher level of agreement across iterations D) Percentage of samples in dataset which were assigned a group call with greater than 95% frequency following resampling, red denotes more samples receiving robust call E) Training NMF cluster metrics red denotes cophenetic correlation, purple denotes dispersion index and blue denotes the silhouette score

Comparison of the combined MRT consensus NMF sub-group assignment with the published sub-group calls from Johann *et al.* (2016) showed high concordance. 86% of ATRT-SHH (51/59) samples clustered to the SHH group, 98% of ATRT-TYR (45/46) clustered to the TYR consensus group and 97% of ATRT-MYC (31/32) to the MYC sub-group.

The TYR group, again, shows the lowest age (mean = 20.8 months, maximum = 132.4 months) with SHH showing an intermediate age distribution (mean = 26.6 months, maximum = 127.6 months) and MYC containing older cases (mean = 45.8, maximum

= 181.4). Chi-square testing did not show a significant difference between age categories.

Unlike the gene expression array results, not all ECRT cases were segregated to the MYC sub-group TYR also contains ECRT and RTK (4/39 and 6/39, respectively). Supratentorial tumours are significantly enriched in SHH (38/66, $p = 0.0023$), while infratentorial tumours are significantly enriched in TYR (46/66, $p < 0.001$). Although not statistically significant, the MYC sub-group features the only example of a spinal cord tumour, while TYR features 3 cases where tumours span across the tentorial boundary (termed “transtentorial”). These cases were derived from the UK MRT cohort based on more detailed CNS localisation information, and while it is possible that other transtentorial cases exist in the dataset, location information for published data is limited and it was not possible to explore this further at this time.

Analysis of available mutation data showed significant enrichment of partial loss of chromosome 22 (as defined by loss of a region spanning multiple genes, $p = 0.025$) and *SMARCB1* point mutations in the TYR subgroup ($p = 0.012$). Although 8/15 reported cases of *SMARCB1* deletion was in the MYC sub-group this was not found to be significant in contrast to the gene expression consensus NMF cohort.

4.3.3 MRT methylation analysis reveals additional heterogeneity

Having shown that combined MRT clustering can recapitulate multiple previously sub-grouping strategies including the existing 3 sub-group definitions previously proposed in ATRT, analysis was carried out to investigate whether the expanded cohort could yield any additional information and further explain the heterogeneity seen in the disease. To this end, a consensus NMF results beyond $k=3$ were considered for their clustering robustness, and the rank of $k=5$ was chosen as after this result, there was a significant drop-off in all NMF cluster quality metrics (Figure 15).

The $k = 5$ NMF solution was able to assign 256/263 samples a sub-group call with over 95% consistency. SHH-Infratentorial (SHH.Inf) and SHH-Supratentorial (SHH.Sup) sub-groups clusters mapped very closely to the consensus NMF $k=3$ SHH cluster ($n = 33/33$, $n = 46/50$, respectively). The names for these two sub-groups originating from the $k = 3$ SHH cluster is due to the significant differential enrichment of the two broad CNS localisations and is discussed in more detail later in this chapter. The $k = 5$ TYR sub-group mapped almost completely to its $k = 3$ TYR (66/68) counterpart but comprises ATRT only, unlike the $k=3$. A novel cluster, termed MRT-Hypomethylated

(HYPO, nomenclature discussed below) mapped partially to $k = 3$ TYR and MYC ($n = 23/42$, $n = 18/42$, respectively), and cluster $k = 5$ MYC mapped fully to MYC (63/63). $K = 5$ HYPO and MYC were the only clusters to contain ECRT and RTK cases as well as ATRT. In addition, group assignments were obtained from the MNP2.0 for a proportion of the total cohort, kindly provided by Dr Pascal Johann (DKFZ, Heidelberg) as part of the ATRT consensus analysis. 'ATRT, SHH' mapped to clusters SHH.Inf/SHH.Sup, with only 8/72 being reclassified (HYPO $n = 4/72$, MYC = 1/72, NC = 3/72), 'ATRT, TYR' mapped wholly to its $k = 5$ counterpart and 'ATRT, MYC' to MYC with one sample not receiving a consensus NMF call (Figure 16).

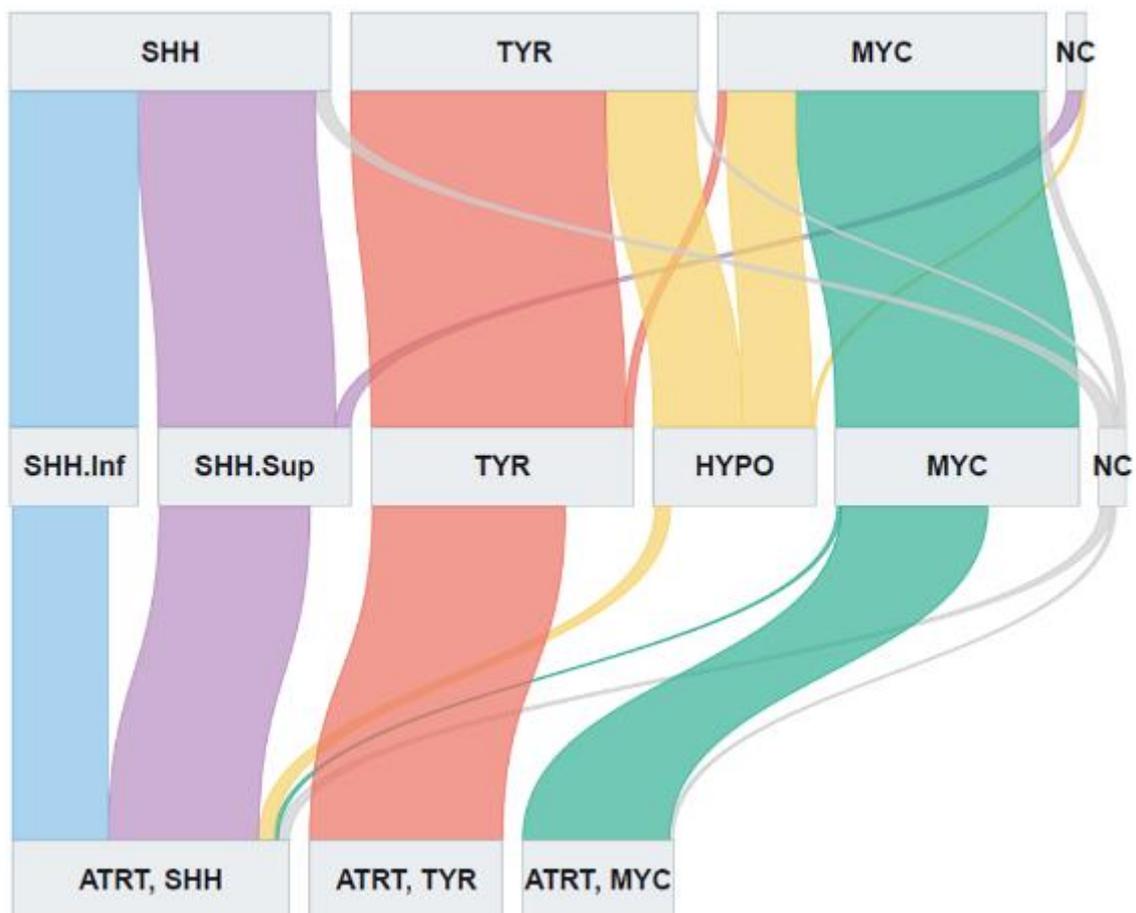


Figure 16. Sankey diagram of consensus NMF sub-group assignments from $k=3$ (top) and $k=5$ (middle) solutions, with additional comparison to the MNP2.0 classifier (bottom); NC = no consensus

Analysis of differences between the two groups SHH.Inf and SHH.Sup showed they significantly vary by the CNS localisation and patient age. SHH.Inf is significantly enriched for infratentorial tumours (23/28, $p < 0.001$) and contained the youngest patients (mean = 11.1 months, maximum = 44.6) while SHH.Sup was significantly enriched for supratentorial tumours (35/38, $p < 0.001$) and older patients (mean 39.3 months, maximum 127.6) (Figure 17A,C).

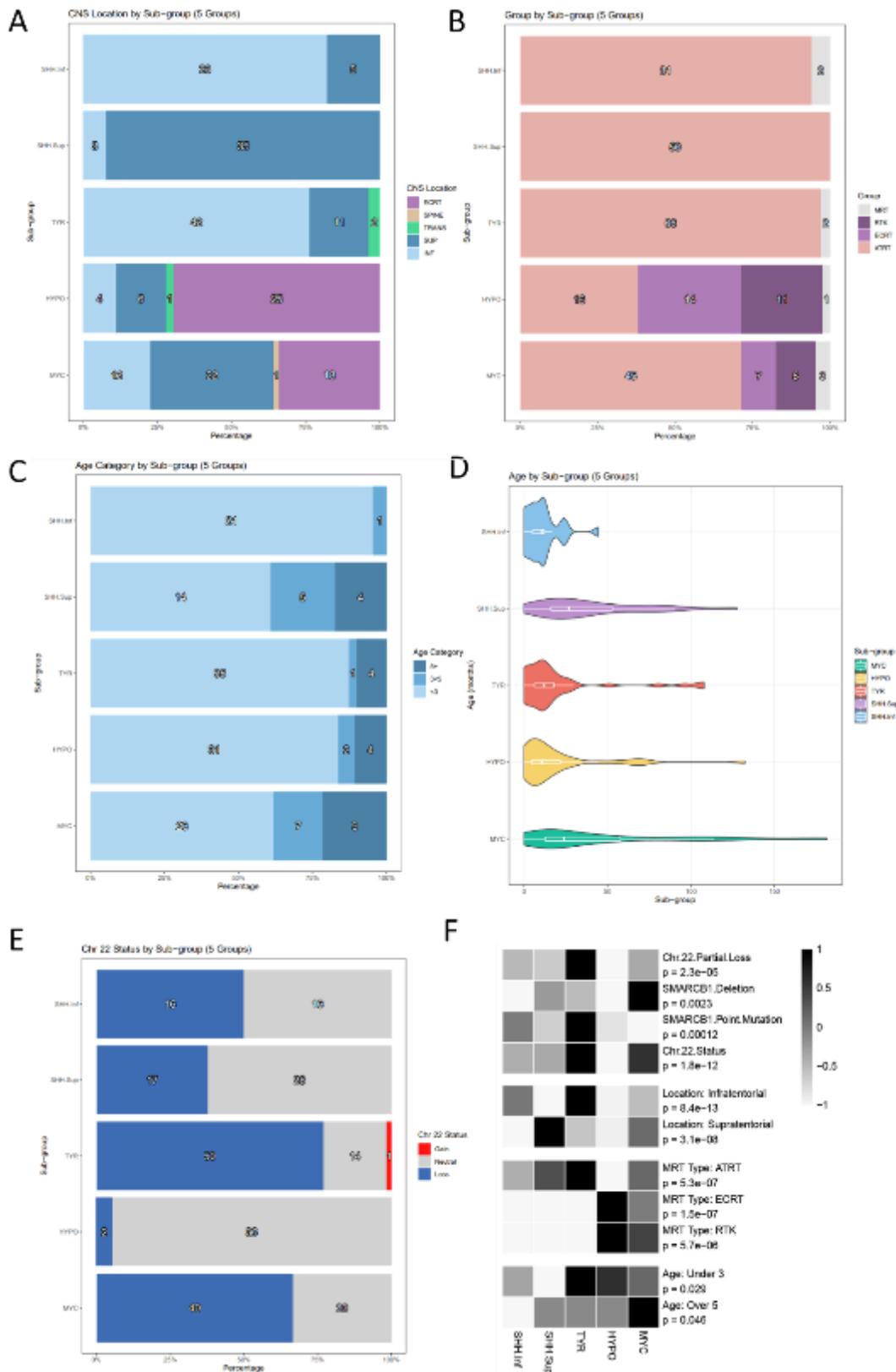


Figure 17. Sub-group characteristics as defined by consensus NMF sub-group assignment from clustering methylation array data; A) CNS location by sub-group INF = infratentorial, SUP = supratentorial, SPINE = spinal cord tumours, TRANS = transtentorial; B) Tumour type by sub-group MRT = cases where clinical data on MRT locale is not available; C) Age distribution by sub-group; A,B,C) Data is shown as a proportion of the total, missing values are removed; D) Violin plot of age distribution across sub-groups; E) Chr 22 copy-number estimation as generated by conumee F) Chi-square test residuals for significant comparisons.

Next, the HYPO sub-group was analysed in order to understand the features that differentiate this novel cluster from the $k = 3$ solution. This group showed a unique profile of CpG methylation when comparing the top 8675 most differentially methylated probes used for NMF clustering (Figure 18). The average beta-values in this group of patients were lower than any other cluster, suggesting a general hypomethylation across all tested CpGs. This effect persisted even when comparing different probe types based on their understood relationship to CpG islands (Shelf/Shore/Island/OpenSea). This consistent hypomethylation across all CpGs led to this group being termed as the Hypomethylated “HYPO” group.

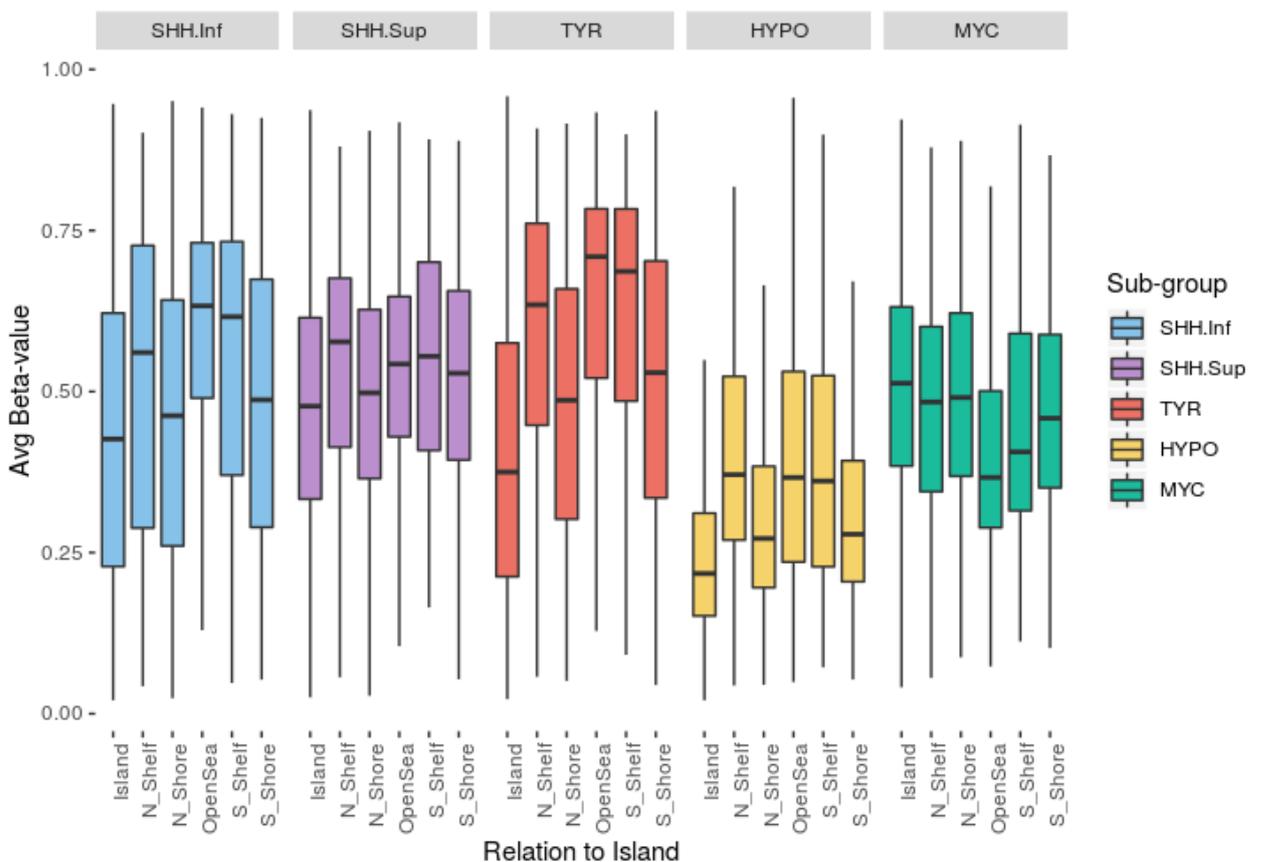


Figure 18. Distribution of average beta-value across sub-groups defined from the $k = 5$ consensus NMF solution. CpGs are annotated based on the Illumina 450K/EPIC manifest for relation to CpG island.

This group contains younger patients (mean = 23.3 months, maximum = 132.4) compared to the MYC $k = 5$ group (mean 45 months, maximum 181.44). Notably, this group contains the only case of reported *SMARCA4* mutation which was assigned a sub-group call by consensus NMF at $k = 5$. This cluster comprises all 3 types of MRT. Estimation of chromosomal arm copy number using conumee (R/Bioconductor), identified a significant enrichment of chromosome 22q arm copy-number neutral events (36/38, $p < 0.001$) compared to other sub-groups (Figure 17A,C,E).

4.4 Discussion

This chapter contained multiple meta-analyses comparing existing sub-grouping strategies for MRT, examining both gene expression and methylation array data. The combined ATRT and ECRT cohort recapitulated previous analyses focusing on only one tumour type and allowed parallels to be drawn between different sub-grouping methods, highlighting the effectiveness of consensus NMF in generating robust sub-group assignments by utilising resampling approaches.

Based on the findings discussed in this chapter, MRT can be defined as a heterogeneous disease displaying at least 3 molecular sub-groups with differences in patient age, *SMARCB1* mutation, and localisation, based on gene expression and methylation profiling. Although different methods may be used in order to cluster MRT, the resulting sub-groups are largely recapitulated regardless of algorithm used and reinforce the existence of these sub-groups as more than just the result of clustering high-dimensional biological data. In addition, the use of consensus NMF allows assessment of cluster robustness by testing the frequency a sample is assigned a particular sub-group call and allowing low-confidence samples to be identified

As well as a meta-analysis of the current proposed sub-grouping strategies, a proposed 5 group sub-grouping based on methylation array is presented and can provide a more comprehensive definition of MRT heterogeneity by capturing specific differences in tumour localisation, patient age and the nature of *SMARCB1*-inactivating mutations.

The analysis presented here also suggests avenues for further expansion of the sub-grouping, in particular the need for an expanded gene expression cohort to match the power of the available methylation data. Features of individual sub-groups such as the immune reactivity of the MYC sub-group, the involvement of DNA and cell replication pathways in SHH and the differences between infratentorial and supratentorial localisation of these tumours should be investigated further. The newly defined HYPO sub-group as identified by $k = 5$ consensus NMF remains to be fully characterised, although a number of features of this group have been presented. Lack of clinicopathological annotation is a significant limiting factor to further defining sub-group characteristics in MRT. Expanding sub-grouping strategies to a larger number of sub-groups requires more statistical power to identify sub-group specific changes, which is currently not available for a number of features including *SMARCB1* mutation and other gene alterations as well as detailed tumour localisation.

Having defined a consensus sub-grouping strategy for MRT, it is necessary to assess whether it is possible sub-grouping can improve our understanding of MRT survival and be useful in patient prognostication. Only limited survival analysis has been carried out in rhabdoid tumours, and never on a combined MRT cohort. As such it is a significant gap in the current understanding of the disease.

5 Survival analysis of MRT using sub-group specific and hypothesis-free testing

5.1 Aims

The aim of this chapter is to carry out a survival analysis using clinical factors and derived subgrouping information of a cohort of 113 MRT samples collected from UK Children's Cancer and Leukaemia Group (CCLG) cancer centres. The analysis will explore common survival features as well as examine any clinical relevance of subgrouping MRT using consensus NMF subgroups derived in Chapter 4.

5.2 Summary of MRT cohort

Data was collected alongside primary tumour material and was received either fully anonymised or was anonymised by either Dr Stephen Crosier, Prof Simon Bailey or Dr Claire Keeling (Newcastle upon Tyne Hospitals Trust). Additional anonymised clinical annotation was provided by Dr Patricia O'Hare and Dr Jessica C Pickles (Great Ormond Street Hospital for Children Trust) as part of the INSTINCT high risk paediatric brain tumour collaboration programme.

The raw clinical data was reduced to relevant fields where information was relevant to this study and amenable to statistical analysis. For tumour type, various classification schemes were collapsed to the following: ATRT for any tumour occurring within the CNS including the brainstem and spinal cord; ECRT for tumours occurring outside of the central nervous system but not within the kidney; RTK for tumours occurring specifically within the kidney. Where the tumour type classification was not available, it was inferred from pathological information or broadly inferred from other clinical data such as therapeutic approach. Samples were received with clear evidence of *SMARCB1* loss, as well as compatible histology but site of tumour was not available and not possible to infer based on current data (n = 7). Those samples were still profiled by DNA methylation array and are currently being investigated further with the data providers. For the purpose of this study there were classified as simply MRT, but not included in survival analysis.

CNS location for ATRT was derived from the localisation information of the tumour. The boundary of the tentorium cerebelli was used as a distinguishing factor, and tumours were classified as either infratentorial when occurring in structures below this boundary (cerebellum, pons, medulla, and brainstem, infratentorial ventricles) or supratentorial (lobes, central brain structures, supratentorial ventricles) when occurring beyond this boundary. Spinal ATRT was defined as any tumours classified as arising in the spinal cord of the CNS. Transtentorial tumours are defined as those that cannot be confidently said to have emerged either from the infratentorial or the supratentorial

space. Both spinal and transtentorial tumours were excluded from survival analysis due to low numbers (n = 1, 2 respectively).

Consensus NMF subgroups derived from previous work outlined in Chapter 4 were included as additional annotation. Where confident subgroup assignment could not be given following resampling, subgroup has not been assigned and samples removed from analysis. Sex was classified as male or female based on reported information. Age was defined as either the age of the patient on the date of sample tumour material being obtained by surgical intervention, or if surgery was not attempted, data was taken from the data of the pathological report detailing a diagnosis of MRT. Metastasis status was recoded from a number of compatible fields and collapsed to the common Yes/No factor. Where tumour stage was given, a stage of M0 or M0/1 was classified as No evidence, the latter was further defined from available clinical information, M1+ was classified as evidence of metastasis.

Overall survival was calculated from age at surgery/diagnosis to recorded age of death or last patient review. Progression-free survival was calculated from age of surgery/diagnosis until the first record of an event – classified as progressive or recurring disease, second malignancy or metastasis.

Table 5. A summary table of the MRT cohort assembled as part of this thesis. Summary statistics and numbers of samples that are annotated for various clinical features are provided.

Variable	Category	Number	Percent (%)
Primary Cases	Total	113	100
Group	ATRT	63	56
	RTK	19	17
	ECRT	24	21
	MRT	7	6
CNS Location	Infratentorial	28	44
	Supratentorial	24	38
	Transtentorial	2	3
	Spinal	1	2
	N/A	8	13
Subgroup k = 3	SHH	27	24
	TYR	36	32
	MYC	47	42
	N/A	3	3
Subgroup k = 5	SHH.Inf	10	9
	SHH.Sup	15	13
	TYR	17	15

	HYPO	28	25
	MYC	37	33
	N/A	6	5
Sex	F	46	41
	M	60	53
	N/A	7	6
Age Summary (months)	Mean	30.02	N/A
	Median	14.25	N/A
	Min	0.00	N/A
	Max	181.44	N/A
Age Category (years)	Under 1	41	36
	1 to 3	36	32
	Over 3	27	24
	N/A	9	8
Metastasis	Yes	35	31
	No	49	43
	N/A	29	26
PFS	Known	76	67
	Unknown	37	33
OS	Known	78	69
	Unknown	35	31

MRT primary tumour samples used in survival analysis were collected from UK CCLG cancer centres, or obtained from Brain UK. Loss of *SMARCB1* was confirmed by IHC during diagnosis and prior to being submitted to the respective biobanks, where *SMARCB1* status was not available it was confirmed prior to inclusion in this study. A full summary of the cohort is provided in Methods 2.1.1.

Of the total cohort of 113 primary cases, 56% (n = 63) were ATRT, 17% (n = 19) were RTK, 21% (n = 24) were ECRT, and 7 samples held a diagnosis of MRT but location information was not available for these tumours. As such, while these 7 cases were included in sub-group discovery, they were not included in the survival analysis. The most common tumour site in ATRT was Posterior Fossa (41%, n = 26/63), and for ECRT Liver and Thorax were joint most common (17%, n = 4/24 in both). Of the total 113 cases, OS was available for 69% (n = 78) and PFS for 67% (n = 76). Overall mean OS was 32.9 months (median = 13.5) and mean PFS was 28.9 months (median = 9.6 months). 5-year OS was 21%, PFS was 14% and longest OS/PFS was 205.6 months

(17.1 years). Information on whether evidence for metastases was present was available for 74% of the cohort (n = 84), information on either recurrence or progression 70% (n = 79), extent of resection was known for 51% (n = 58) after excluding cases with only biopsies recorded as the only surgical intervention, and 73% (n = 83) of cases had information on whether radiotherapy was received

Age data was available for 92% of cases (n = 104). Average age at diagnosis was 30 months (2.5 years) while the median was 14.25 months (1.2 years), the oldest patient at diagnosis was 181.44 months (15.1 years). Separated by MRT tumour type, ATRT median age was 15 months (1.25 years) and maximum was 181.44 (15.1 years), ECRT median age was 13 months (1.1 years) and maximum 104.1 months (8.7 years), RTK had a median age of 15.35 months (1.3 years) and a maximum of 74.6 months (6.2 years). Of the total cohort, 36% were under the age of 12 months when diagnosed (n = 41). Sex information was available for 94% of the cohort and the F:M ratio was 0.77. Sub-group assignment using consensus NMF was obtained for 110/113 cases for k = 3, and 107/113 cases for k = 5, based on procedures outlined in Chapter 4. Estimation of chromosome 22q loss from DNA methylation array profiling was available for 84% (n = 95) after excluding cases with a noise score ≥ 1.8 as defined by the conumee R analysis.

5.3 Survival analysis MRT using clinicopathological association

Survival analysis was carried out by constructing survival curves and comparing groups using logrank testing and likelihood ratio testing. All comparisons were carried out on the whole cohort, and then split into either ATRT or ECRT where it was reasonable to do so.

OS and PFS did not significantly differ by MRT localisation. For ATRT, infratentorial or supratentorial location was also found not to be significantly associated with survival differences. Comparisons of survival between consensus NMF k = 3 and k = 5 sub-groups showed no significant difference. There was also no difference associated with specific loss of chromosome 22q as estimated by DNA methylation array analysis.

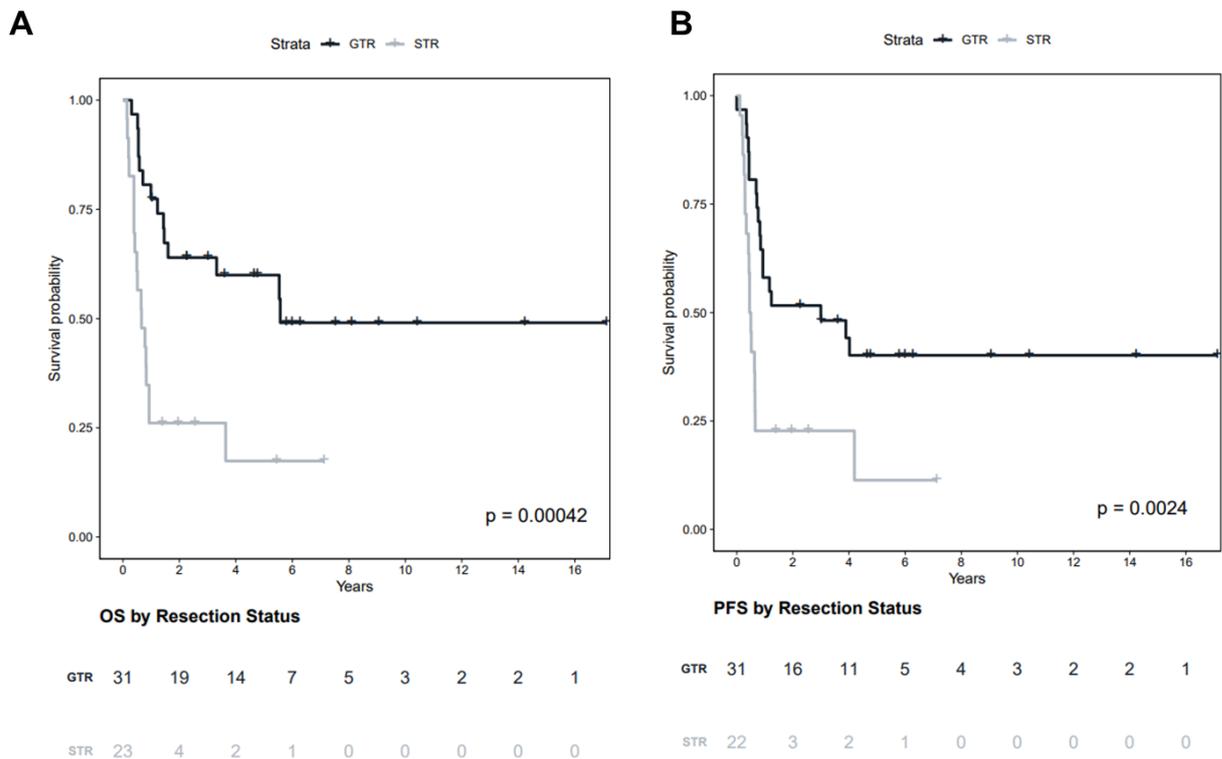


Figure 19. Kaplan-Meier curves showing survival differences for extent of resection in MRT A) Overall survival B) Progression Free Survival GTR = gross total resection STR = subtotal resection; p-value is provided from log-rank analysis, number of patients in each category is shown below the graphs.

Likelihood ratio and logrank testing was carried out to test for significance in clinical data, p values for both are provided unless the value is the same, or a different test used. Significant differences in OS and PFS were identified comparing extent of resection in the full cohort (OS $p < 0.001$, $n = 54$; PFS $p = 0.004/p = 0.002$ $n = 53$) also in ATRT (OS $p = 0.002$, $n = 39$; PFS $p = 0.03$, $n = 38$)(Figure 19). Receipt of radiotherapy was significant in all comparisons and in both OS and PFS (MRT OS $p < 0.001$, $n = 73$; PFS $p < 0.001$, $n = 71$). Evidence of metastases is also significant for all comparisons in both OS and PFS (MRT OS $p < 0.001$, $n = 74$; PFS $p = 0.003/0.002$, $n = 72$).

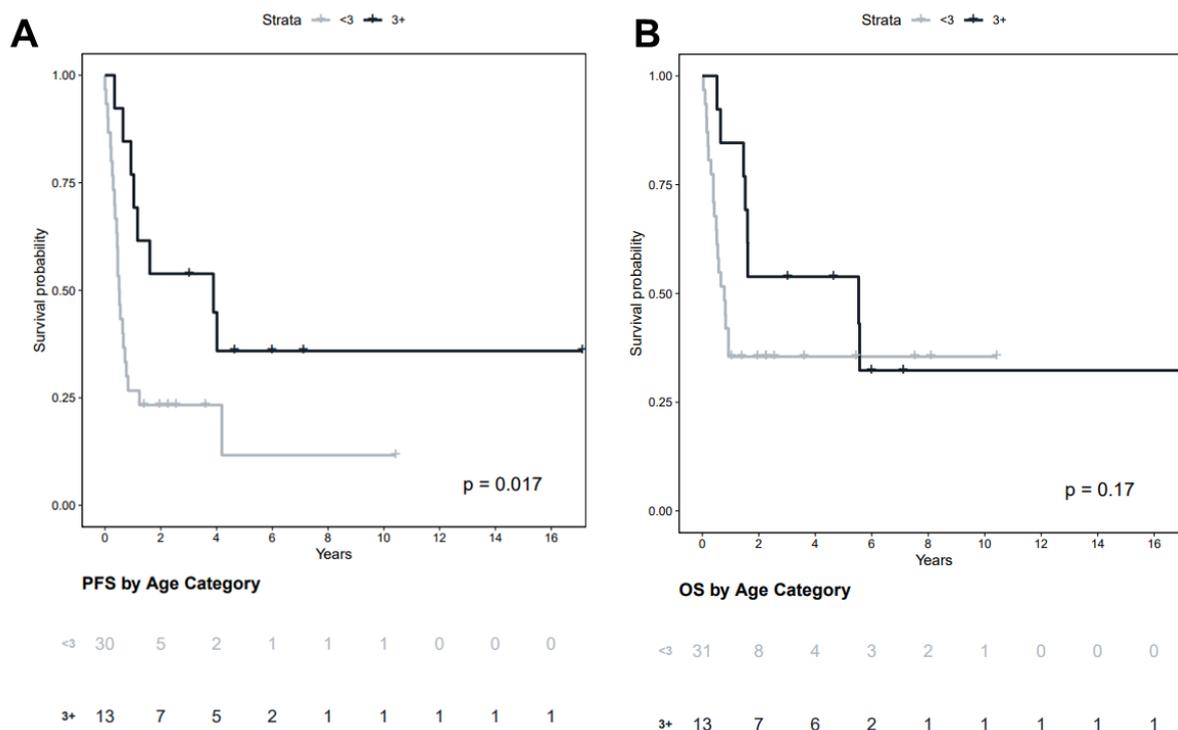


Figure 20. Kaplan-Meier curves showing survival differences for age categories in ATRT between patients under 3 years of age and over 3 years of age A) Overall survival B) Progression Free Survival <3 = younger than 3 years >=3 = 3 years or older at diagnosis; p-value is provided from log-rank analysis, number of patients in each category is shown below the graphs.

Next it was examined age was associated with survival. Two sets of age categories were employed. First, patients were divided into either younger than 1 or 1 year old and older, second, they were divided into under 3 or 3 years and older. These categories were chosen to coincide with the typical approach to defer radiotherapy before age 3 in ATRT, and the higher risk previously described in patients under 1 year old. In the under 3 / older than 3 category OS was not significantly different in the whole MRT cohort, or ATRT and ECRT examined separately (Figure 20). Additionally, PFS was not found to be significantly different in ECRT when comparing across the two age categories. ATRT PFS was significantly lower for patients under the age of 3 ($p = 0.01/p = 0.02$, $n = 43$) with 83% patients in that group succumbing to disease after the first year. While PFS analysis in the whole cohort was also significant, it is likely due to it also comprising the ATRT patients. Conversely, in the under 1 / older than 1 OS in ATRT was not found to significantly vary, while PFS for patients under 1 was significantly poorer ($p = 0.04/p = 0.03$, $n = 43$). In MRT, and ECRT, both OS and PFS varied significantly (ECRT OS $p = 0.008/p = 0.003$, $n = 28$; PFS $p = 0.008/p = 0.003$, $n = 27$) (Figure 21).

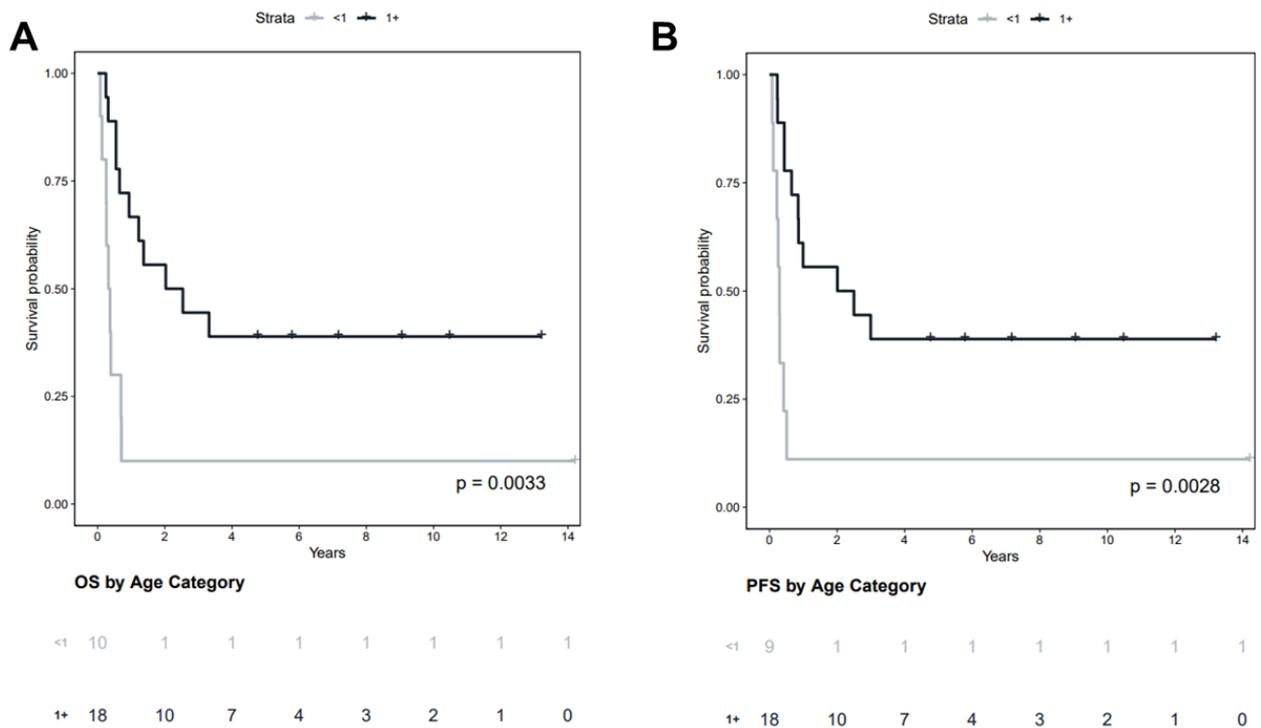


Figure 21. Kaplan-Meier curves showing survival differences for age categories in ATRT between patients under 1 years of age and over 3 years of age A) Overall survival B) Progression Free Survival <1 = younger than 1 years >=1 = 1 years or older at diagnosis; p-value is provided from log-rank analysis, number of patients in each category is shown below the graphs.

In a univariate Cox proportional hazard regression analysis, significant factors identified for MRT were radiotherapy, evidence of metastasis, extent of tumour resection, and patients being either older or younger than 3 at diagnosis. In ATRT evidence of progression or recurrence did not pass significance, while in ECRT subtotal resection status was only recorded in 1 case, and no evidence of progression was only found in 4 annotated cases making both of these variables not suitable for the analysis. A summary of the associated Hazard Ratios and p-values is available in Table 6.

Table 6. Summary of Cox proportional hazard regression univariate analyses

	HR (95% CI for HR)	z	Wald Test	p value
Combined MRT				
Radiotherapy (N)	0.21 (0.11-0.4)	-4.7	22	2.10E-06
Metastasis (N)	3.1 (1.7-5.5)	3.8	15	0.00012
Reccurence/Progression (N)	4.9 (2.1-12)	3.6	13	0.00032
Extent of Resection(GTR)	3.4 (1.7-7)	3.3	11	0.00081
Chr 22q (Loss)	0.97 (0.53-1.8)	-0.097	0.01	0.92
Sex M	0.68 (0.4-1.2)	-1.4	2	0.16
Age (Under 3)	0.57 (0.3-1.1)	-1.8	3.1	0.078

Age (Under 1)	0.46 (0.26-0.8)	-2.7	7.4	0.0066
Subgroup SHH	1 (0.55-1.8)	0.031	0	0.98
Subgroup TYR	0.77 (0.42-1.4)	-0.85	0.72	0.39
Subgroup MYC	1.3 (0.72-2.2)	0.8	0.64	0.42
ATRT Only				
Radiotherapy (N)	0.21 (0.086-0.53)	-3.4	11	0.00079
Metastasis (N)	3.4 (1.5-7.6)	2.9	8.5	0.0036
Reccurence/Progression (N)	2.3 (0.91-5.7)	1.8	3.1	0.077
Extent of Resection (GTR)	3.9 (1.5-9.7)	2.9	8.2	0.0041
Chr 22q (Loss)	0.55 (0.24-1.3)	-1.4	1.9	0.17
Sex M	0.6 (0.28-1.3)	-1.3	1.8	0.18
Age (Under 3)	0.56 (0.24-1.3)	-1.4	1.8	0.17
Age (Under 1)	0.62 (0.29-1.3)	-1.2	1.5	0.22
Subgroup (SHH)	1.1 (0.52-2.3)	0.23	0.05	0.82
Subgroup (TYR)	0.83 (0.38-1.8)	-0.47	0.22	0.64
Subgroup (MYC)	1.2 (0.44-3.2)	0.33	0.11	0.74
ECRT Only				
Radiotherapy (N)	0.2 (0.074-0.56)	-3.1	9.4	0.0021
Metastasis (N)	2.9 (1-8.1)	2	4	0.046
Reccurence/Progression (N)				
Extent of Resection(GTR)				
Chr 22q (Loss)	1.6 (0.52-4.8)	0.8	0.64	0.42
Sex M	0.87 (0.35-2.1)	-0.3	0.09	0.76
Age (Under 3)	0.63 (0.23-1.7)	-0.89	0.8	0.37
Age (Under 1)	0.27 (0.11-0.69)	-2.8	7.6	0.0057
Subgroup (SHH)				
Subgroup (TYR)	0.67 (0.22-2)	-0.71	0.51	0.48
Subgroup (MYC)	1.5 (0.49-4.5)	0.71	0.51	0.48

Variables found to be significant in univariate analysis were analysed as covariates in a multivariate Cox proportional analysis. In a multivariate Cox proportional hazard analysis of MRT OS, radiotherapy and age category over/under 3 years was found to be independently significant (receipt of radiotherapy HR 0.09 (0.018 – 0.41 95% CI), $p = 0.002$, Age ≥ 3 HR 5.4 (1.069 – 27.22 95% CI) $p = 0.041$ respectively) and for PFS only radiotherapy was significant (HR 0.14 (0.04 – 0.52 95% CI) $p = 0.003$) (Figure 22). OS was analysed in the ATRT cohort using a multivariate Cox proportional hazard analysis showing, receipt of radiotherapy (HR 0.054 (0.008 – 0.36 95% CI), $p = 0.003$), extent of resection (HR 3.68 (1.04 – 13.05 95% CI), $p = 0.044$) and age over/under 3

years (HR 8.17 (1.2 – 55.4 95% CI), p = 0.032) were significant. Cox proportional hazard testing in ATRT PFS, and both ECRT OS and PFS, found only radiotherapy remained a significant variable in the multivariate analysis.

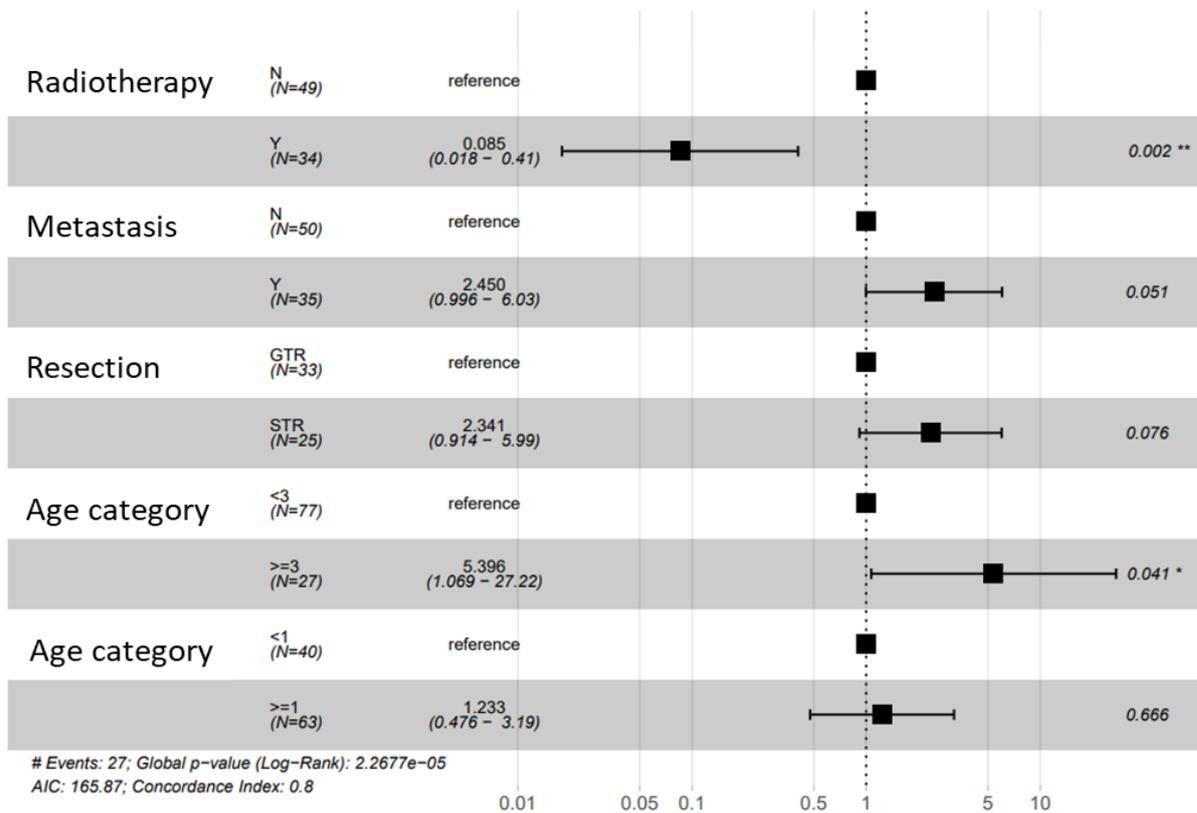


Figure 22. Forest plot from multivariate cox proportional hazard analysis for MRT overall survival

Despite the lack of significance of any subgroup-specific survival analysis at univariate level, further analysis was carried out to see whether combinations of variables involving sub-group could be used to stratify patients. In 2018, Michael Fruhwald, presented a poster abstract analysing a European ATRT cohort and identified a sub-group specific risk stratification comprising patient age and membership of the TYR sub-group (Fruehwald *et al.*, 2018).

Receiver operating characteristic (ROC) curves were generated for combinations of variables from Cox proportional analysis models. These types of visualisations are a useful way of displaying the sensitivity (true positive rate) and 1-specificity (false positive rate) of a diagnostic marker. Here, time-dependent ROC curves are used in order to compare the effectiveness of different combinations of variables at predicting 2-year survival from MRT OS and PFS. Variables tested were either consensus NMF k = 3 subgroup and age under/over 3 years, or age under/over 1 year. Area under the

curve (AUC) was compared to select the most effective combination of variables that predicts survival.

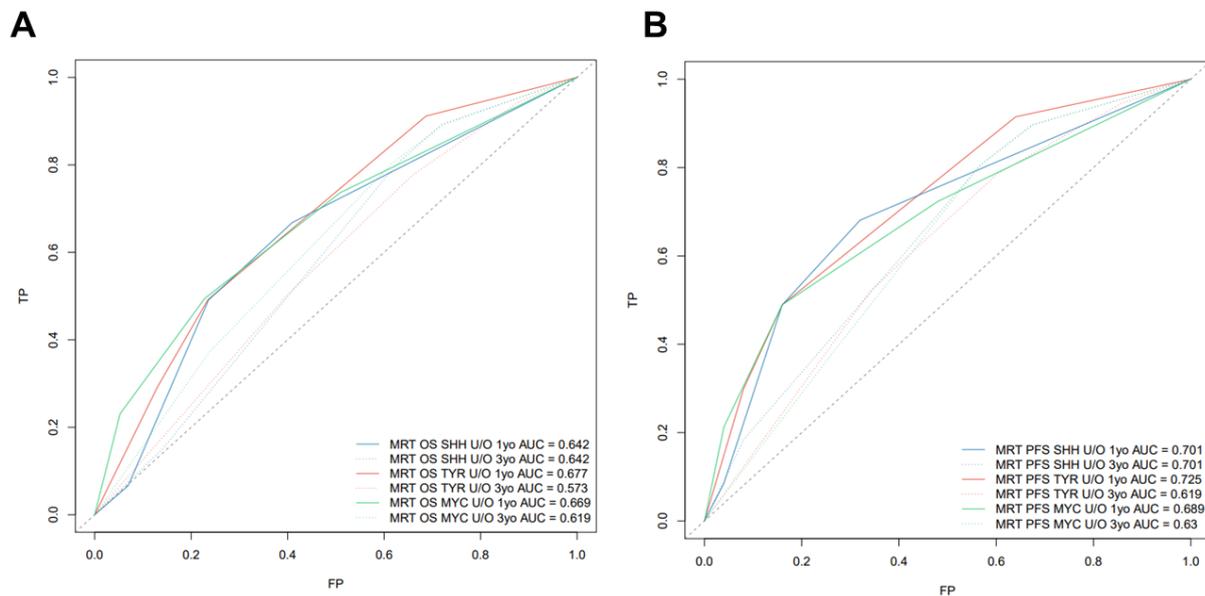


Figure 23. ROC curves for Cox proportional hazard ratio 2-year survival predictions for multiple combinations of age and subgroup categories A) overall survival B) progression free survival; U/O 1yo = Under/Over 1 year old; U/O 3yo = Under/Over 3 year old; TP = true positive, FP = false positive

For both OS and PFS, the combination of TYR subgroup and the category of under/over 1 year generated the highest AUC (OS 0.677, PFS 0.725). (Figure 23) although there was only a small difference overall between the predictive power of the combinations used. Survival curves of the resulting stratification were then generated. A category of moderate risk (MR) was assigned to any patient with a subgroup of TYR and under the age of 1 at diagnosis, and a high risk (HR) category was assigned for any patients in another subgroup, or above the age of 1. This scheme differs somewhat to the scheme originally presented in the abstract where TYR subgroup and age <1 year was classified as low risk, either TYR or age >1 year as intermediate risk and high risk was assigned to cases > 1 year not in TYR subgroup. Due to the limitations of the cohort size the classification was reduced to two categories by combining the intermediate and high risk categories. MRT and ATRT cohorts both showed significant survival differences between MR and HR groups (MRT OS $p = 0.02$, log rank; PFS $p = 0.005$, log rank; ATRT OS $p = 0.046$, log rank; PFS $p = 0.0052$, log rank). ECRT only did not show significant difference between risk, however as there were only 4 patients that were identified as having moderate risk, it is not clear whether this could be a significant survival difference in a larger cohort.

5.4 Discussion

This chapter presented the survival analysis for the UK MRT cohort and highlighted a number of significant clinical factors which impact survival. In addition, to presenting a survival overview of the cohort generated as part of this thesis, the analysis highlighted clinical factors that were significant in both ATRT and ECRT tumours and showed a combined risk associated across tumour types. Location specific survival features were also identified and likely correspond to both the nature of the tumours and the general approach and practicalities of therapy that are available and utilised in treating ATRT and ECRT.

The main variable identified to be highly significant across all groups was radiotherapy received. In the cohort, of the 13 ATRT patients where clinical data was available, only 1 patient did not receive radiotherapy, while in the ECRT cohort over half 16/30 patients did not receive radiotherapy. While the reason for this high number of patients in ECRT not receiving radiotherapy is not clear, it highlights the need for all patients with ECRT to undergo radiotherapy if possible given the survival for non-receivers was extremely poor (2-year survival 17%) much in the same way as other survival analyses have done for MRT. The results also highlighted the importance of complete surgical resection in this cohort. While there are many reasons why complete resection is not always possible, the impact on survival in both ATRT and ECRT is clear and is consistent with findings from other studies.

An interesting finding of the analysis was the antagonistic relationship between age and PFS. When comparing patients under and over the age of 3, PFS in ATRT was significantly different between the two categories suggesting patients under the age of 3 were much more likely to have progressive disease, but did not significantly vary in ECRT. Patients under the age of 1 were significantly less likely to survive in with ECRT as well as showed more progressive disease, while survival in ATRT was not affected. This could perhaps reflect the type of therapy the patients received, although unfortunately there is a lack of chemotherapy information for this cohort, which is a significant limitation. Additional efforts are currently ongoing in order to collect this data, but it is currently available for only a small fraction of the cohort.

This analysis also partially validated a stratification approach combining both subgroup and patient age in order to improve stratification over the predictive ability that the two presented as separate single predictors. The initial concept from the stratification was presented at the 18th International Symposium on Pediatric Neuro-Oncology (ISPNO

2018) June 30 – July 3. It is currently expected that the expanded version of that analysis will be submitted to peer review and publication shortly. Despite this, the use of the subgrouping information in MRT in order to generate a novel prognostic scheme is an encouraging first step in the efforts to characterise differences between MRT subgroups and define clinically-relevant molecular differences between tumours. While currently limited by low numbers of tumours in this study, the combination and meta-analysis of multiple survival cohorts in future could identify additional novel prognostic features of subgroups. It should therefore be a primary focus in future studies given how historically it has been difficult to obtain large, multicentre survival analyses. This also underlines the importance of generating a consensus subgrouping scheme which enables the comparison of cohorts generated in different studies to undergo a common classification strategy and therefore allow any subgroup associations to be applied to a much wider cohort.

Finally, as well as working towards expanding current cohort size, efforts should also be made to continue to explore MRT biology and further develop current understanding of differences between MRT localisations and subgroups. As previously discussed in Chapter 3 and Chapter 4 the MYC subgroup in MRT shows overexpression of a large set of genes associated in immune development, and function. Any immune interaction of this group, and others is a potential target of future research and may be able to allow for novel prognostic and therapeutic approaches.

6 Investigating the immunological landscape of CNS tumours

6.1 Introduction

There is evidence to suggest that the immune system may have some involvement in the biology of MRT, although the extent and specific association remains to be fully elucidated. Overexpression of genes associated with the immune system was reported when analysing sub-groups of ECRT (Chun *et al.*, 2016). In this thesis, sub-group specific expression differences in MRT highlighted the enrichment of immune-specific genes in the MYC sub-group (Chapter 4). Although immune involvement in MRT has not previously been examined, the role and nature of the tumour immune micro-environment (TIME) has been interrogated for a number of other CNS tumours with a view to investigate suitability for immune-therapy.

Immune-therapies are an attractive alternative anti-cancer strategy alongside the conventional approaches of surgery, chemotherapy and radiotherapy that may be particularly well suited to targeting diffuse infiltrative growing tumours. The field of cancer immunotherapy has grown expansively in recent years to include the therapeutic use of cancer vaccinations, chimeric antigen receptor (CAR) T-cell therapy and agents which block immune-checkpoint receptors and/or ligand interactions such as CTLA-4 and PD-1. Each can provoke a significant anti-tumour response in patients within varied tumour types (Prins *et al.*, 2011; Quail and Joyce, 2013; Hinrichs and Rosenberg, 2014; Topalian *et al.*, 2015; Butowski *et al.*, 2016; Voena and Chiarle, 2016; Quail and Joyce, 2017). However, for each patient who derives clinical benefit from a particular immunotherapeutic agent there are many whom do not (Bockmayr *et al.*, 2019). The composition of the TIME is a critical determinant of tumour-immune interactions and can direct response to treatment (Hirata and Sahai, 2017). Therefore, to take full advantage of the potential of immunotherapy - or combinations with targeted agents - treatment approaches need to be tailored to the specific TIME.

Detailed studies of the TIME are being conducted to predict response to immunotherapy and uncover mechanisms of treatment resistance. While anti-PD-1 antibodies nivolumab and pembrolizumab and an anti-CTLA-4 antibody Ipilimumab are FDA approved for, and can produce durable responses in, patients with metastatic melanoma (Robert *et al.*, 2015a; Robert *et al.*, 2015b; Weber *et al.*, 2015), non-small cell lung cancer (Rizvi *et al.*, 2015) and renal cell carcinoma (Tomita *et al.*, 2019), the majority of patients do not respond. Comparative studies between responders and non-responders indicate that multiple factors, including pre-existing T-cell infiltration, checkpoint molecule expression within the tumour and mutational burden with

consequent production of neo-antigens correlate with response to immune-therapy. For instance, colorectal cancer of the molecular subtype CMS1 are characterised by DNA mismatch-repair defects, microsatellite instability and hypermutation with accompanying infiltration of CD8+ T cells (Mlecnik *et al.*, 2016) and expression of immune-checkpoint proteins CTLA-4, PD-1, PD-L1 and IDO-1 (Gatalica *et al.*, 2014; Angelova *et al.*, 2015; Becht *et al.*, 2016). CMS1 patients show significant responses to anti-PD-1 therapies (Boland and Ma, 2017).

Tumours are frequently described as being immunologically “hot” or “cold” with a presumed implication for the effectiveness of particular tumour immune therapies. “Hot” tumour TIMEs are broadly characterised by high expression of the PD-1 ligand (PD-L1) and by infiltration of cytotoxic lymphocytes (CTLs) expressing PD-1. “Cold” tumours being relatively sparsely infiltrated with CTLs, at least within the tumour core (Binnewies *et al.*, 2018). Childhood brain tumours are thought to be relatively immunologically “cold” due to paucity of mutations (i.e. generally lacking neoantigens (Grobner *et al.*, 2018)). To date, quite limited information on TIME in childhood brain tumours has been published and in piecemeal fashion. In adult brain tumours, several immune cell types have identified roles in, and associations with, tumour development. For instance, TAMs (Tumour Associated Macrophages) are believed to make up a large proportion of immune cells in gliomas (Graeber *et al.*, 2002), and to be generally pro-tumourigenic and associated with a higher tumour grade (Komohara *et al.*, 2008; Hambardzumyan *et al.*, 2016). Furthermore, the number of neutrophils appears to have prognostic value (Fossati *et al.*, 1999; Bertaut *et al.*, 2016) and immunosuppressive Regulator T-cells (Treg) are significantly increased in patients with Glioma as a proportion of the peripheral CD4+ cell pool; they also account for a substantial proportion of the TIME (Fecci *et al.*, 2006; Hussain *et al.*, 2006). Simple extrapolation from adult brain tumours is unlikely to be informative given the underlying differences in tumour biology.

A number of methods exist to characterize and quantify TIME directly e.g. IHC, Fluorescence-assisted cytometry (FACS), Cy-TOF, single cell RNA-sequencing. These may be costly, laborious and/or difficult to multiplex. Indirect techniques have been developed to estimate TIME *in silico* by deconvoluting complex mixtures of cell types from profiles of bulk populations using pure populations of cell types as a reference (Gentles *et al.*, 2015; Newman *et al.*, 2015; Teschendorff and Zheng, 2017). CIBERSORT is a notable algorithm which uses support vector regression modelling to

deconvolute cell types and has been applied to several cancer datasets (Newman *et al.*, 2015). CNS tumours have been extensively DNA methylation profiled using arrays, most prominently by Capper *et al.* (2018) who published a cohort of 3,764 CNS-tumours (including 1403 patients < 18 years old) representing 80 tumour DNA methylation types and sub-types closely related to WHO histopathological entities. The Paediatric Brain Tumour Group, Newcastle University and others have published further large series of some of the major paediatric CNS types i.e. MB (Cavalli *et al.*, 2017; Northcott *et al.*, 2017; Schwalbe *et al.*, 2017a; Sharma *et al.*, 2019a), ATRT (Johann *et al.*, 2016; Torchia *et al.*, 2016) and paediatric High Grade Gliomas (pHGG) (Mackay *et al.*, 2017; Mackay *et al.*, 2018) with extensive clinical annotation and parallel multiomic data (RNA-seq, copy-number profiles, Exome/Whole-genome Sequencing). This chapter discusses the implementation of methylCIBERSORT - a recent adaptation of the CIBERSORT algorithm which uses genome-wide DNA methylation data (Chakravarthy *et al.*, 2018) - to characterize the TIME of >6000 CNS tumours, assessing variation and the relationship with clinico-pathology or outcome.

6.2 Aims

This chapter aimed to develop a DNA methylation based approach to estimation tumour infiltration on CNS tumours in order to carry out a primary investigation of the extent of immune infiltration in CNS malignancies. First a signature matrix needed to be generated which could accurately estimate immune infiltration on CNS tumours. Secondly clinicopathological factors would be compared with resulting estimations of immune infiltration to identify significant associations.

6.3 Results

6.3.1 Validation and benchmarking of signature matrix

A signature matrix was constructed from reference DNA methylation profiles of pure flow-sorted populations of cells. This signature matrix represents a set of differentially methylated CpGs selected and weighted to reflect specificity for a given cell type and is used as the basis of cell deconvolution by methylCIBERSORT. The final signature matrix consisted of 2215 differentially methylated CpGs distinguishing between 12 broad cell types: regulatory T-cells (Treg), CD4+ T-cells (CD4T), CD8+ T-cells (CD8T), B-cells (B-cell), Natural Killer (NK) cells, eosinophils, neutrophils, monocytes, endothelial cells, glial cells, neurons and cancer. The matrix was verified such that (i) specific differentially methylated CpGs were captured for each cell type (ii) the absence

of batch effects following processing (iii) the CpGs selected were not confounded by being specific to any particular CNS cancer type (Figure 24A,B,E).

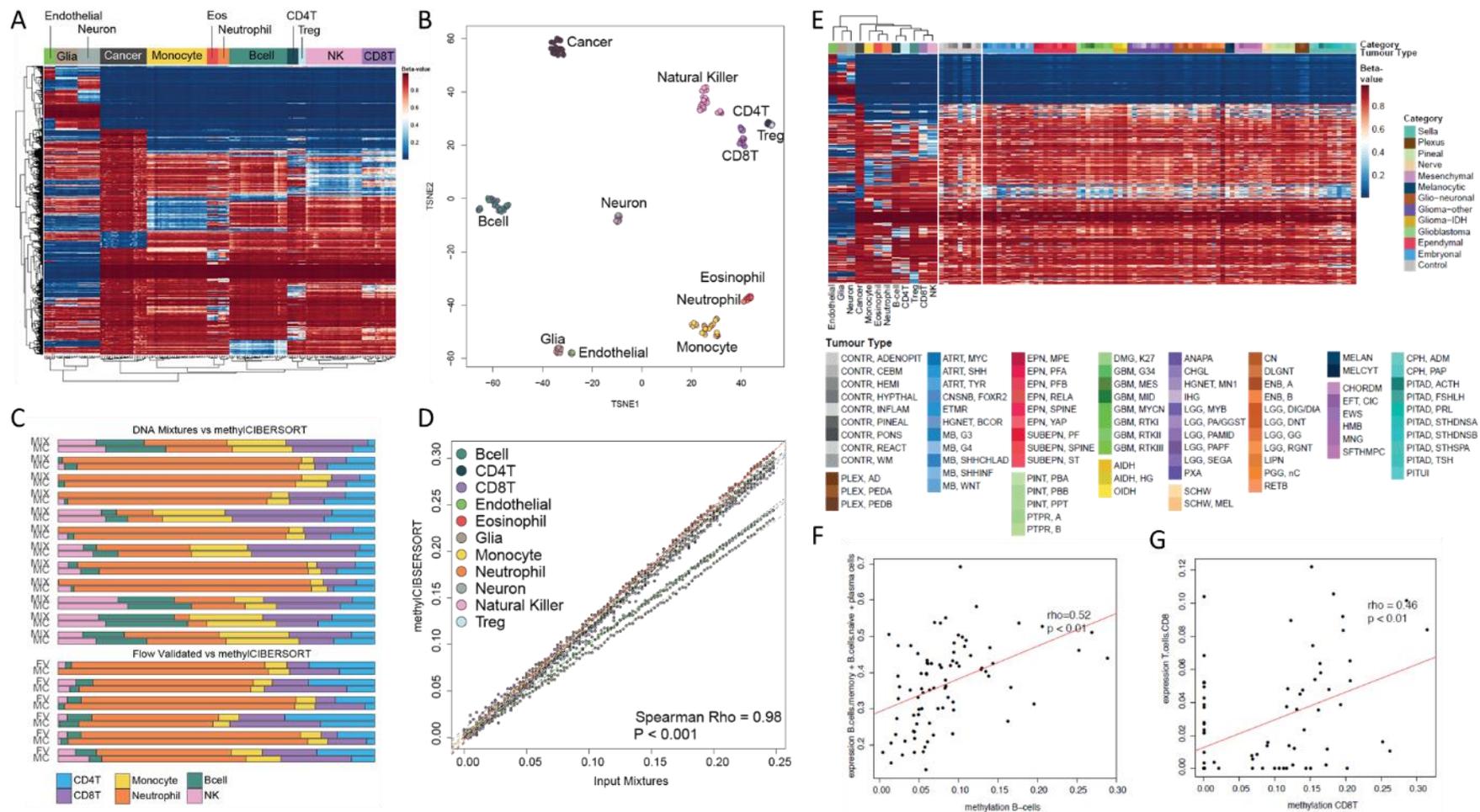


Figure 24. A) Heatmap of the resulting 2215 CpG probe beta-value signature matrix as generated by the methylCIBERSORT package. Hierarchical clustering was carried out on columns and rows as denoted by dendrograms. B) t-SNE embedding of resulting signature matrix beta-values C) Relative proportion comparisons between DNA-mixture and Flow Validated test data and the resulting methylation-based CIBERSORT estimate D) Scatter-plot showing the comparison between initial proportions of modelled mixture data and resulting methylation-based CIBERSORT estimate E) Pan-CNS cohort showing only 2215 CpG probes selected in the signature matrix F) Scatter-plot showing comparison between methylation-based CIBERSORT fraction of B-cells and the expression-based estimates from parallel data G) Scatter-plot showing comparison between methylation-based CIBERSORT fraction of CD8T and the expression-based estimates from parallel data

The new signature matrix was benchmarked using publicly available methylation profiles of PBMCs with known cell composition as determined by flow-cytometry or constituted from mixtures of reference DNAs of known proportions. A significant level of correlation was found between the methylCIBERSORT estimates and the flow cytometry measurements and known DNA mixtures (Rho = 0.84 $p < 0.001$, $n = 36$ and Rho = 0.91 $p < 0.001$, $n = 72$ respectively, Figure 24C). 100 synthetic mixtures for each cell type generated *in silico* were tested using methylation profiles of random pure cell populations mixed 1:4 with a mixture of cancer cell line profiles (Figure 24D). Again, there was a highly significant correlation between estimated and actual cell composition (Rho = 0.98 $p < 0.001$, $n = 1100$). A dataset comprising Breast Cancer samples for which parallel 450K Methylation and Affymetrix U133Plus2.0 expression profiles were available was analysed using both methylCIBERSORT and standard expression CIBERSORT (LM22 signature matrix). Where reference cell populations were comparable (i.e. had been flow sorted using the same antibodies) directly, or by aggregation and where tumour infiltration was present, there was a significant correlation (e.g. B-cells, T-cells) between methyl and expression CIBERSORT (Figure 24E,F).

6.3.2 Tumour Immune Microenvironment in Malignant Rhabdoid Tumours is associated with subtype and prognosis in a Tumour location dependent manner.

MethylCIBERSORT analysis was ran on a set of 229 MRT, made up of 192 ATRT samples and 37 ECRT. MRT are on average infiltrated predominantly by Tregs (19% of non-cancer cells), monocytes (18%), B-cells (15%) and CD8T (13%) (Figure 25A). Taking the three previously published molecular subgroups of ATRT (ATRT-TYR, ATRT-SHH, ATRT-MYC [23]) and ECRT the distribution of each estimated immune cell type is significantly different with respect to ATRT subgroup (all $p < 0.05$) (Figure 25A). Post-hoc testing shows the most significant are NK, Treg, B-cells (each greater in ATRT-TYR) and CD8T (significantly greater in ATRT-MYC and ATRT-SHH) (Figure 25A-B). Surprisingly, no immune cell types were found to be significantly different overall between ATRT (all subtypes) and ECRT (Figure 25A).

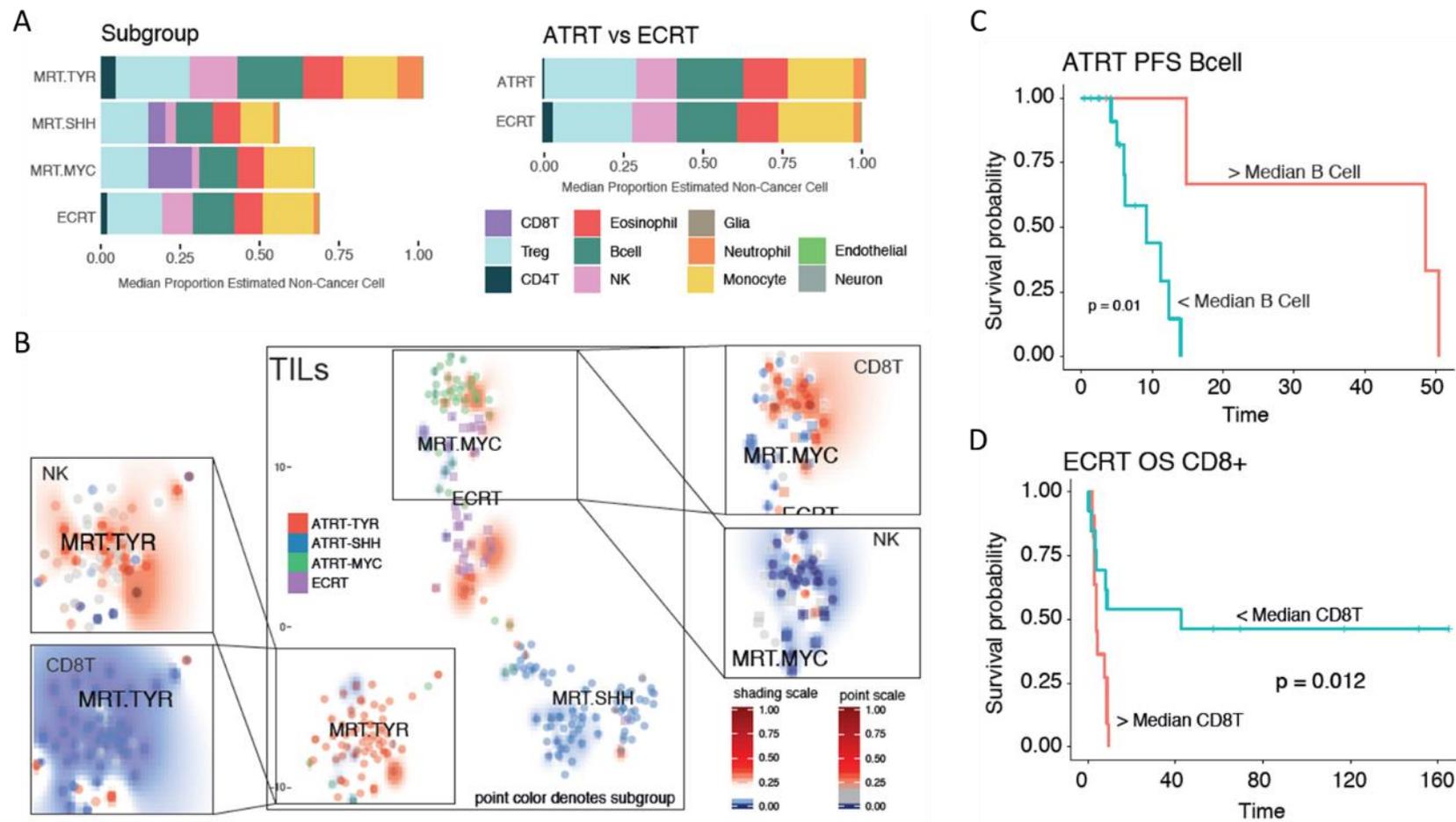


Figure 25. A) Comparison of estimated proportions and type of non-cancer cells in MRT by sub-group and by location B) t-SNE plot representing the methylation profiles of 229 MRT. The colours of dots in the central panel map to published molecular subgroups. Text represents centroids of individual subtypes. Background shading represents the 2D spatial density estimation of the amount of Tumour Infiltrating Lymphocytes (TILs); red shading equals relatively greater than average infiltration and blue less than average. Exploded side panels represent enlarged areas of interest wherein both dot colour and background shading represent the relative amount of the particular immune cell infiltration denoted. Red denotes relatively greater than average infiltration and blue less than average. C) Kaplan-Meier plot showing significantly different overall survival (OS) in ATRT with > or < median numbers of B-cells. D) Kaplan-Meier plot showing significantly different overall survival (OS) in ECRT with > or < median numbers of CD8+T cells.

Consensus clustering of MRT immune cell infiltration estimates identifies four robust immune subgroups which cut across the tumour subgroups and named here MRT IC1-4. MRTIC1 and MRTIC3 constitute minor clusters, only 2% (4/229) and 6% (14/229) of all MRT, and have relatively high proportion of neutrophils and monocytes respectively. Both clusters contain a disproportionate number of ECRT and ATRT-TYR (Chi-Square=48.218, $p<0.001$) (Figure 26A). MRTIC4 constitutes 32% (74/229) of all MRT and is characterised by a relatively high proportion of CD8T and relatively low infiltration of other immune cell types. MRTIC2 constitutes 60% (137/229) of all MRT and is characterised by a relative lack of CD8T and relatively moderate infiltration of other immune cell types; 83% (59/71) of ATRT-TYR are of this type.

Examining association with outcome in ATRT, a greater than median level of B-cells was associated with a significantly improved PFS (Log-rank, $p=0.01$, $n=21$) (Figure 25C). In ECRT a greater than median level of CD8T was associated with a significantly poorer overall survival (Log-Rank $p=0.0023$, $n=30$) (Figure 25D). It should be noted that molecular subgroup alone was non-significant with respect to overall survival in both ATRT and ECRT.

No significant differences in immune infiltration are seen with respect to age category (<2 vs >2 years), presence of metastases at diagnosis and type of SMARCB1 mutation. The only significant clinico-pathological association is a lower proportion of monocytes and a higher proportion of NK cells in infratentorial compared to supratentorial ATRT ($W=1469.5$ & $W=2726.5$ respectively, both $p<0.001$) (Figure 26C,D).

CYT score, PDL1 expression, calculated in samples for which parallel RNA-seq data was available, was significantly correlated with methylCIBERSORT estimates of TILs in MRT (all $p<0.01$, $n = 23$) (Figure 26E,F). Taken as a whole this underlines the significant relationships between subgroup, prognosis and immune cell infiltration in MRT.

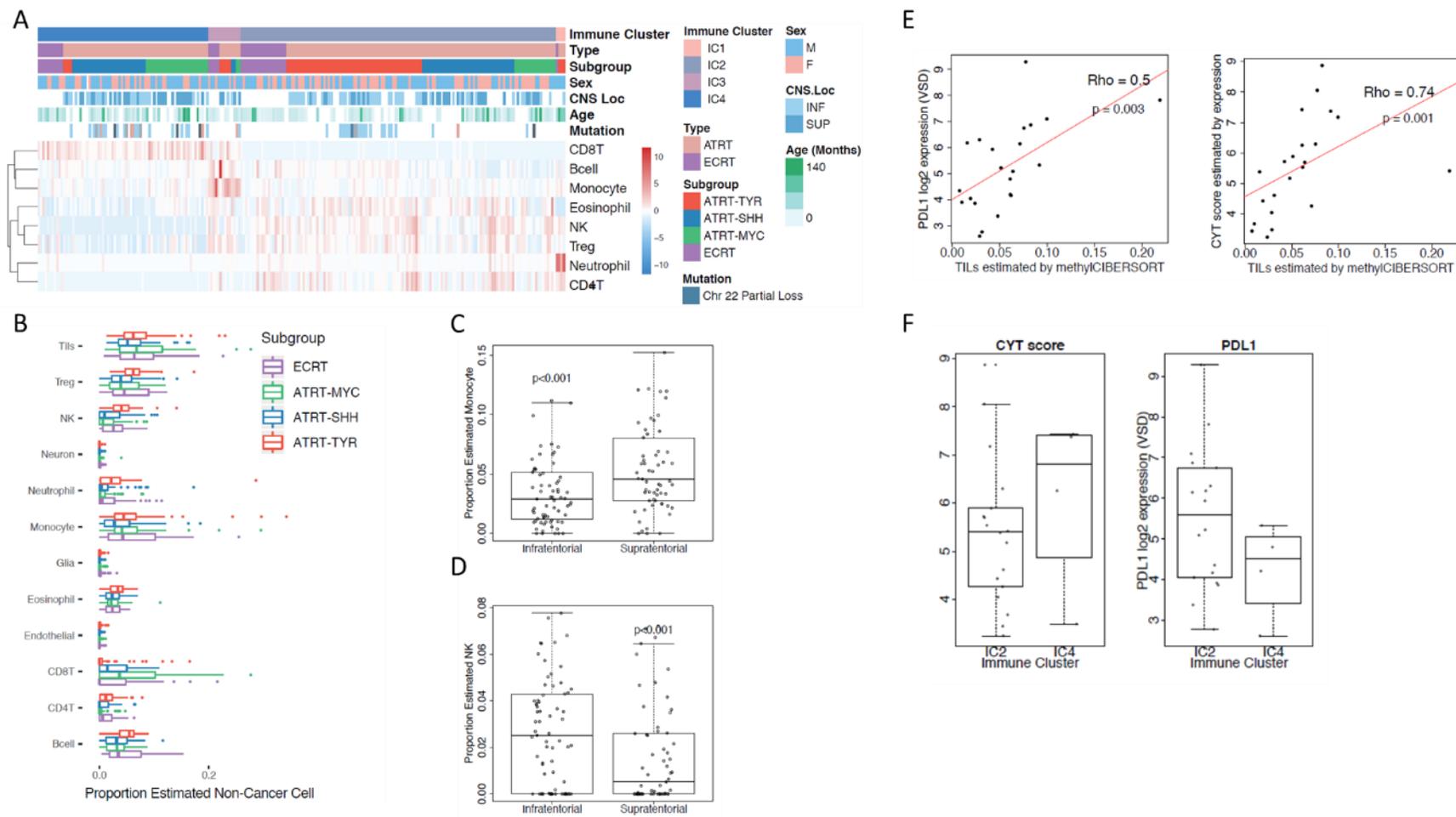


Figure 26. A) Heatmap showing row-scaled relative levels of immune cell infiltration in 229 Malignant Rhabdoid Tumors (MRT) ordered by immune cluster MRTIC1-4. B) Barplot showing estimated proportion of cell infiltration by molecular subgroup. C) Boxplot showing estimated monocyte infiltration and D) NK infiltration in ATRT by CNS location (infratentorial/supratentorial) E) scatterplot showing PDL1 expression and F. CYT score correlation with proportion of TILs as estimated by methylCIBERSORT. G) Boxplot showing CYT score and PDL1 expression by MRT immune cluster.

6.3.3 methylCIBERSORT analysis of a pan-CNS Tumour methylation cohort shows significant differences in Tumour immune microenvironment related to Tumour type and grade

MethylCIBERSORT was applied to a set of 3,764 pan-CNS tumour methylation profiles (plus an additional 141 control/hematopoietic samples) published by Capper *et al.* (2018). This reference set is the training resource of the Molecular Neuro-Pathology 2.0 (MNP2.0) classifier and represents 80 methylation tumour types/sub-types closely related to WHO histopathological entities and divided into 13 broad histological categories. A relative proportion of the 12 cell types were estimated and indicators of deconvolution performance examined. As further validation, the relative proportion of cancer cells estimated by methylCIBERSORT was significantly correlated with the estimate of tumour purity provided by Capper *et al.* (2018) (based on machine learning estimates trained on a set of known glioma positives) ($Rho = 0.71$, $p < 0.01$, $n = 3784$, Figure 27B). The sum of the estimated proportions of all Tumour Infiltrating Lymphocytes (TILs) (i.e. Treg, CD4T, CD8T and NK), correlates significantly with the meTIL score (an independent measure of T Lymphocyte infiltration based upon methylation status of 5 CPGs) defined by Jeschke *et al.* (2017) ($Rho = 0.29$, $p < 0.001$, $n = 3764$, Figure 27B). As expected, control samples having a known inflammatory or reactive tumour microenvironment were associated with a large increase in the estimated median proportion of neutrophils (86% vs 0%, $W=0$, $p < 0.001$) and monocytes (50% vs 17%, $W=17$ $p < 0.001$) respectively compared to the average of other CNS control tissues (Figure 27E).

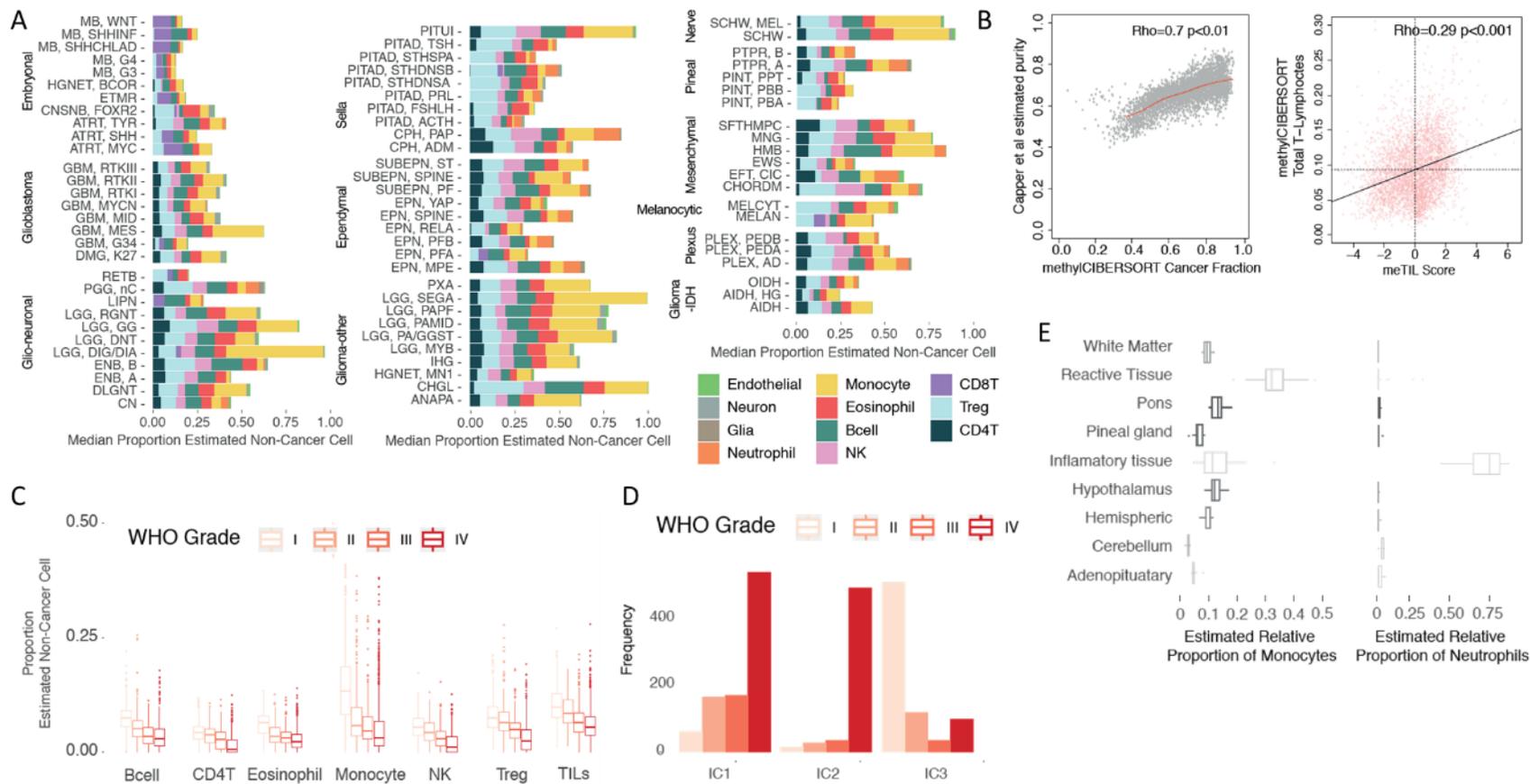


Figure 27. Barplots of the estimated median infiltration of specific cell types as a proportion of all non-cancer cell types (range scaled from 0-1) in 3,763 CNS tumour samples from the panCNS tumor cohort. Data shown by tumor type/subtype highlighting the range and variation of immune cell infiltration in different CNS tumor types. B) Scatterplot showing the estimated methylCIBERSORT cancer fraction correlates significantly with published estimates of tumor purity; Scatterplot showing the estimated methylCIBERSORT estimate of total T-lymphocyte infiltration correlates significantly with an independent meTIL score C) Boxplot showing a negative association between proportion of estimated cell types and WHO-grade. D) Bar chart showing differences in frequency of patients of different WHO grade by immune cluster. E) Boxplot showing methylCIBERSORT estimates of monocyte and neutrophil infiltration in control samples included within the pan-CNS cohort. As expected significantly greater proportions of monocytes and neutrophils were observed in reactive and inflammatory tissues respectively.

Calculating the median estimated relative proportions of non-cancer cell types showed that on average across all CNS tumour types the largest fractions of non-cancer cells proportionally were Tregs (20% of all non-cancer cells) and monocytes (20%) followed by B-cells (16%), CD8T (14%), eosinophils (12%), NK cells (12%), CD4T (9%), and neutrophils (8%). Relatively modest proportions of neuronal (3%), endothelial cells (2%), and glia (1%) were estimated.

Individual tumour types/subtypes varied significantly in the relative proportions of infiltrating cell types; each cell type was significantly non-randomly distributed with respect to tumour type/subtype (as calculated by Kruskal-Wallis one way analysis of variance (KW), each $p < 0.001$) (Figure 27A). Post-hoc testing (Dunn test) reveals the relative number of TILs, and indeed the total amount of infiltrating cells, was significantly less in high grade tumour types such as embryonal tumours (i.e. MB, ATRT, ETMR) than in Low Grade Gliomas (LGG) ($p < 0.001$). Examining the median relative proportions of the 11 non-cancer cell types across CNS tumours those with the greatest variance are monocytes, Treg and CD8T. Notably, LGG subtypes have a proportionally greater number of monocytes making up an estimated 35% of all infiltrating cells compared to 13% in embryonal tumours. CD8T, for example, is proportionally greater in MB_{Grp3} and MB_{SHHCHLD}, making up an estimated 48% and 40% of all infiltrating cells respectively compared to 6% in LGG. Tregs are relatively greater proportionally in the Sellar tumours (specifically pituitary adenomas) constituting an estimated 36% of all infiltrating cells compared to 14% in glioblastoma and 17% in embryonal tumours (Figure 27A).

Consensus clustering of immune cell estimates identifies an optimal 3 immune clusters referred to as panCNS_{IC1-3}. Members of panCNS_{IC1} have a relatively high proportion of Tregs and a relative lack of CD8T cells. panCNS_{IC2} have a relatively high proportion of CD8T and low proportions of CD4T/Tregs and NK cells. panCNS_{IC3} has a relatively high proportion of monocytes and relative lack of CD8T (Figure 28). Membership of an immune cluster was related to but by no means exclusively dictated by tumour type. Whilst immune cluster is significantly non-random with respect to tumour subgroup/subtype (Chi-square=3303, $p < 0.001$, most tumour subgroups cut across multiple immune clusters to some extent.

Immune Clusters

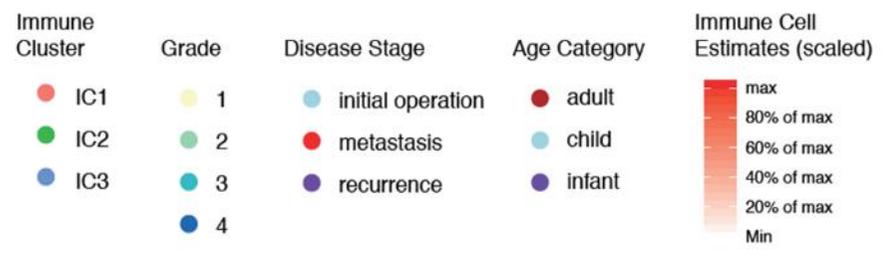
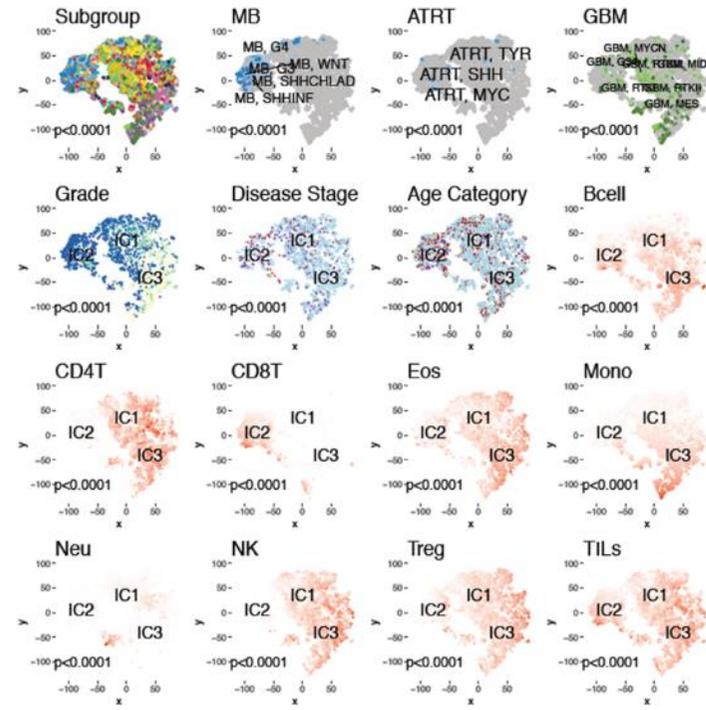
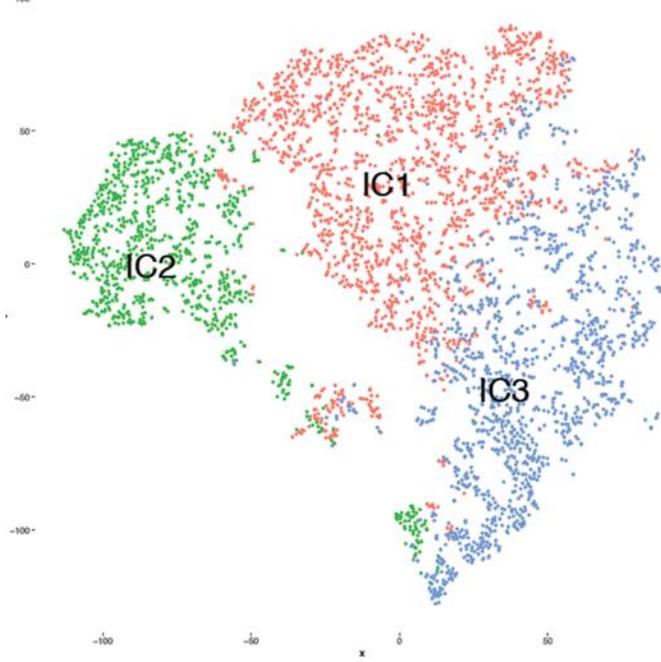


Figure 28. t-SNE plot showing clustering of the panCNS cohort by immune cell estimates. Large panel shows three immune clusters (IC1-3), smaller panels show the location and distribution of tumours of particular subgroup, grade, stage and age, immune cell estimates are represented as a red-white colour scale. P-values represent statistical test for non-random association of a given characteristic with immune-cluster

The significance of association with available clinico-pathological characteristics (WHO grade, disease stage i.e. metastases/relapse/diagnosis, age category i.e. <3 years / 3-16 years / >16 years, tumour location, gender) was assessed and the proportion of immune cell types was significantly associated with each of the clinico-pathological characteristics with the exception of gender. The strongest association was with WHO grade for which the average infiltration of certain immune cell types (eosinophils, CD4T, B-cell, Treg, NK, monocytes and TILs) decreases proportionally with increasing WHO Grade (I-IV) (Figure 27C, Figure 27E). Immune cluster membership is significantly associated with WHO grade (Chi-square = 1249.3, $p < 0.01$). 87% (509/587) of all WHO Grade I tumours belong to panCNS_{IC3} and panCNS_{IC2} consists of 86% (492/571) Grade IV tumours (Figure 27D). Such associations are unsurprising given the strong interdependence of clinico-pathological factors with tumour subtype. However, a regression analysis using only tumour types for which grade, age category and tumour location was variable showed a number of clinico-pathological associations significant independently of tumour subgroup. B-cells, CD4T, eosinophils and Tregs, were each significantly negatively associated with tumour stage (each $p < 0.01$) independently of subgroup. Monocytes were also significantly positively associated with spinal location independent of subgroup. In summary, this analysis reveals the existence of at least three distinct TIME classes across CNS tumours strongly related to but not exclusively dictated by tumour subgroup and grade.

6.3.4 Tumour Immune Microenvironment in Medulloblastoma is related to molecular subtype but provides independent prognostic information

Having estimated TIME in a panCNS cohort, more specific analyses were applied to the single tumour entity medulloblastoma; applying methylCIBERSORT to a set of 2325 MB methylation profiles, published by The Paediatric Brain Tumour Group and others, for which more detailed clinico-pathological and parallel multiomics data was available. Each of these studies elaborated upon the 4 classic subgroups of MB (MBWNT, MBSHH, MBGrp3 & MBGrp4) to describe further derivative subtypes including high-risk or low-risk subtypes of MBGrp3/Grp4. The most abundantly estimated infiltrated non-cancer cell types on average across all MB subgroups were CD8T (27% of all non-cancer cells), B-cells (16%) and eosinophils (15%). The proportion of each cell type was significantly different with respect to the 4 classic subgroups (all $p < 0.001$) and post-hoc testing shows significantly greater CD8T in MBGrp3 vs MBGrp4 (7.3-fold, $p < 0.001$), greater NK in MBGrp4 vs other subgroups

(9.7-fold, all comparisons $p < 0.001$) and greater B-cells in MBSHH vs other subgroups (3-fold, all comparisons $p < 0.001$) (Figure 29A,B).

A meta-analysis was recently published describing a further refinement of the MBGrp3/Grp4 subgroups into eight subtypes I-VIII. These subtypes are also associated with differences in estimated levels of each cell type with the exception of monocytes (each $p < 0.002$). Post-hoc analysis shows the most significant differences to be CD8T (greater in subtype II), Tregs (less in subtype II), NK (greater in subtypes VIII), B-cells (less in subtype III) (all comparisons $p < 0.01$) (Figure 4A,B). Significant differences were apparent between MBSHH subtypes. Both the infant SHH subtype and the SHH gamma subtype show significantly greater proportions of B-cells than other MBSHH subtypes (2.6 and 2.5 fold respectively, both $p < 0.001$) (Figure 29A,B).

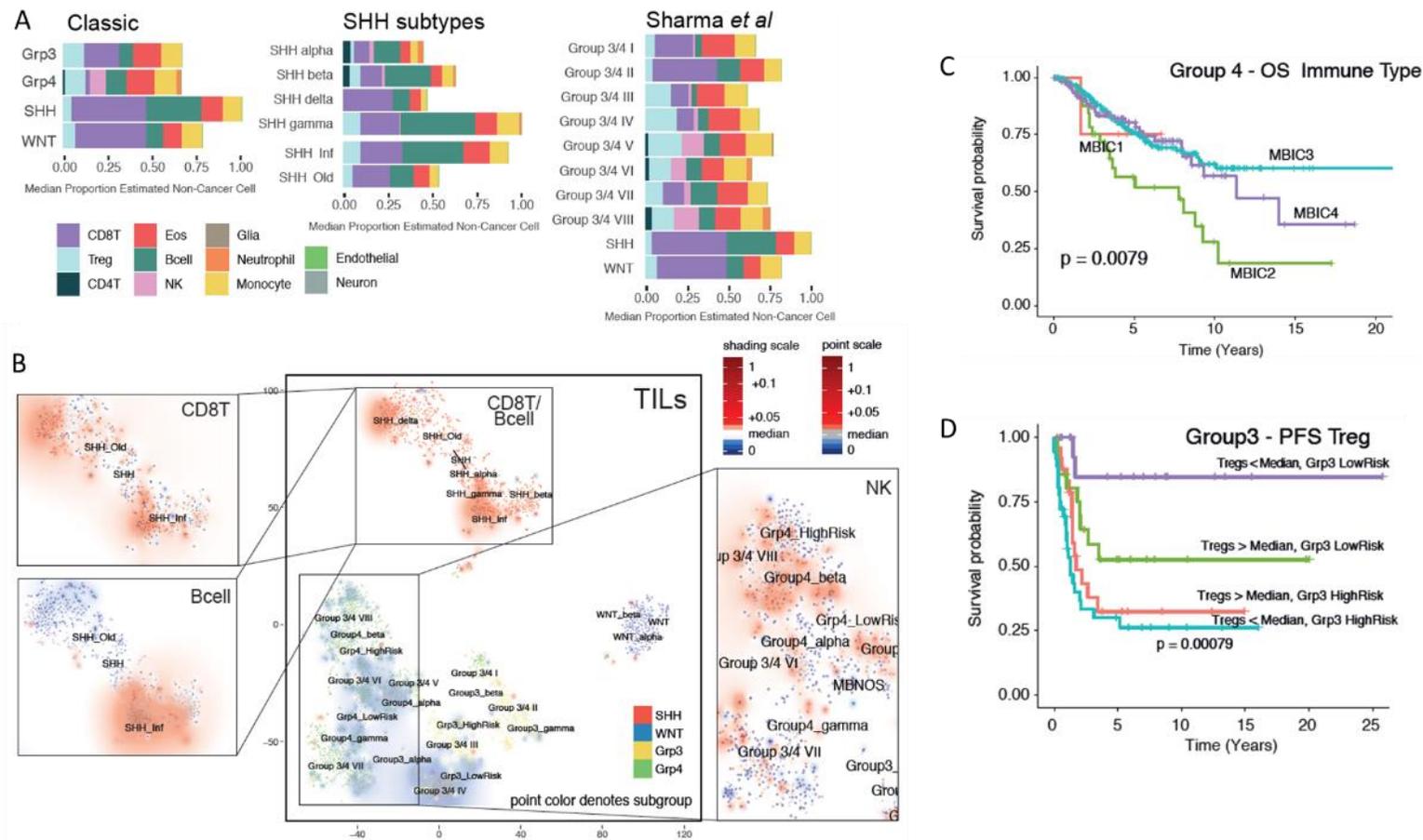


Figure 29. A) Barplots of the estimated median infiltration of specific cell types as a proportion of all non-cancer cell types (range scaled from 0-1) in 2,325 Medulloblastoma by subgroup (classic 4 medulloblastoma consensus subgroups) by SHH subtype and by 10 group consensus. B) t-SNE plot representing the methylation profiles of 2,325 Medulloblastoma. The colours of dots in the central panel map to the classic 4 molecular subgroups; red = SHH, blue = WNT, yellow = Grp3, green = Grp4. Text represents centroids of individual subtypes as reported variously. Background shading represents the 2D spatial density estimation of the amount of Tumour Infiltrating Lymphocytes (TILs); red shading equals relatively greater than average infiltration and blue less than average. Exploded side panels represent enlarged areas of interest wherein both dot colour and background shading represent the relative amount of the particular immune cell infiltration denoted. Red denotes relatively greater than average infiltration and blue less than average. C) Kaplan-Meier plot showing significant difference in overall survival in MBGrp4 by immune cluster. D) Kaplan-Meier plot showing significantly different progression free survival (PFS) within the MBGrp3 subtypes by low (< median) or high (> median) levels of Treg infiltration

Consensus clustering of MB immune cell estimates identifies an optimal 4 immune clusters referred to here as MBIC1-4 which cut across each of the MB subgroups/subtypes (Figure 30A). MBIC1 constitutes 7% (167/2325) of all MB and is characterised by relatively high proportions of B-cells and CD8T and a disproportionately high number of MBSHH patients; 83% of MBIC1 are also MBSHH (Chi-Square = 425.59, $p < 0.001$). MBIC2 constitutes 7% (162/2325) of all MB and is characterised by relatively high proportions of Treg, eosinophils, NK and low proportions of CD8T. MBIC3 constitutes 42% (981/2325) of all MB has relatively low proportions of CD8T, relatively moderate levels of all other infiltrating immune types and a disproportionately high proportion of MBGrp3/Grp4 (78% of MBIC3). MBIC4 constitutes 44% (1015/2325) of all MB and is characterised by a relatively high proportion of CD8T cells and relatively low-moderate levels of other infiltrating immune cell types (Figure 30A,B).

For a subset of MB samples, both methylation and RNA-seq data were available. It was therefore possible to calculate the expression based metric "Cytolytic score" (CYT = the mean expression of GZMA and PRF1) as described by Rooney et al [37] and this was significantly correlated with methylCIBERSORT estimates of TILs ($Rho = 0.18$, $p = 0.015$, $n = 185$) and differed significantly by immune cluster ($F = 4.1$, $p = 0.008$, $n = 185$) being greatest in MBIC1 and poorest in MBIC3 (Figure 30B). Expression of immune checkpoint gene PDL1 was also significantly different with respect to immune clusters (both $p < 0.01$); MBIC1 in particular showed high expression of PDL1.

Further associations between infiltrating cell estimations and clinico-pathological variables (within the four classic subgroups) were examined including: MYC/MYCN amplification, TP53 mutation and metastatic stage. MYC amplification in MBGrp3 was associated with a significantly higher proportion of TILs, CD8T and B-cells (KW=8.7, 16.7, 18.9 respectively, each $p < 0.01$, $n = 408$), and a lower infiltration of Tregs (KW=11, $p = 0.012$, $n = 408$) (Figure 30C).

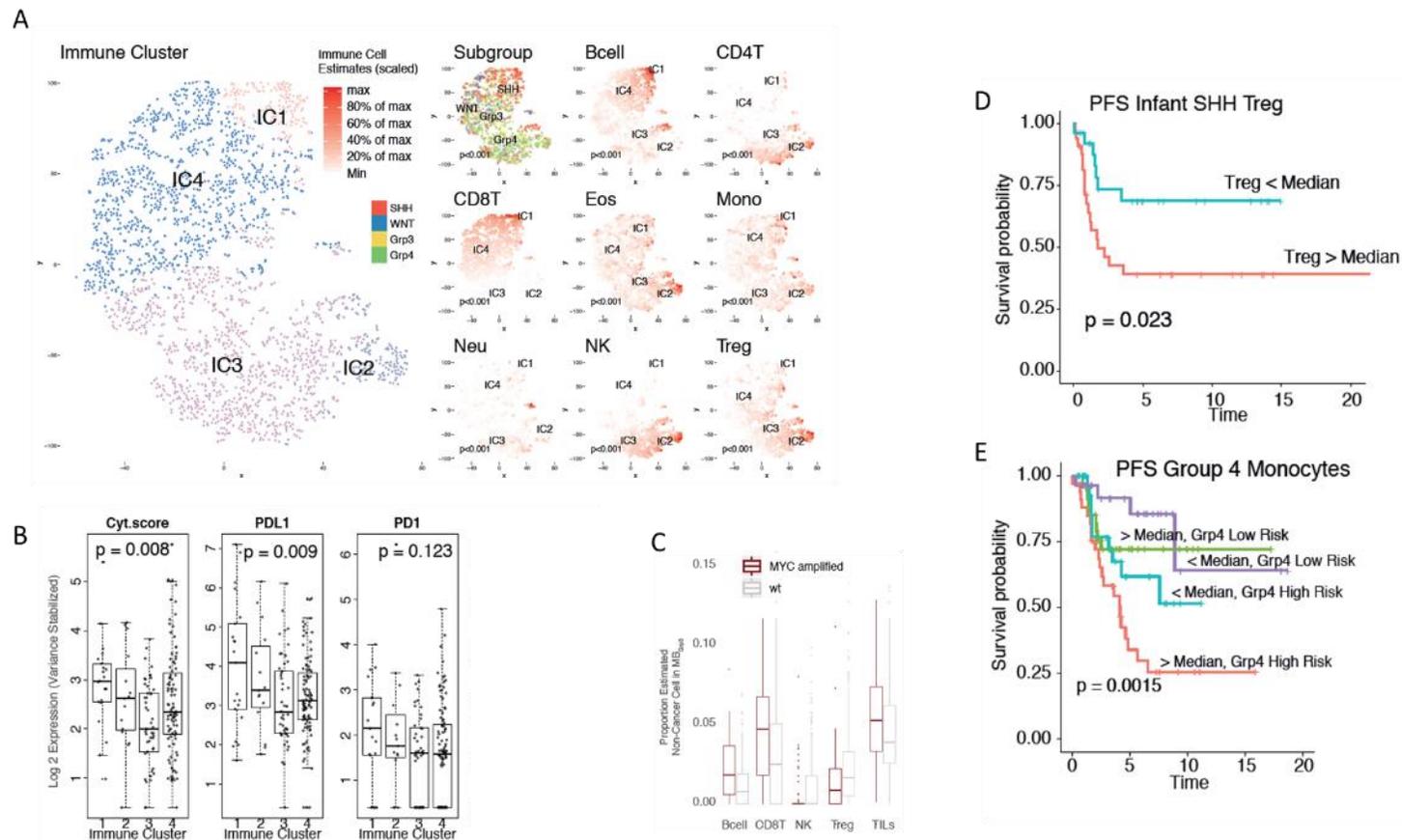


Figure 30. A) *t*-SNE plot showing clustering of the MB cohort by immune cell estimates. Large panel shows four immune clusters (MBIC1-4), smaller panels show the location and distribution of tumour of particular subgroup, immune cell estimates are represented as a red-white colour scale. P-values represent statistical test for non-random association of a given characteristic with immune-cluster. B) Boxplot showing expression of PDL1, PD1 and CYT score by MB immune cluster. C) Boxplot showing proportion of non-cancer cells by presence of MYC amplification in MBGrp3. D) Kaplan-Meier plot showing significantly different progression free survival (PFS) in infant MBSHH by low (< median) or high (> median) levels of Treg infiltration. E) Kaplan-Meier plot showing significantly different progression free survival (PFS) in MBGrp4 by low (< median) or high (> median) levels of monocyte infiltration.

Estimated immune cell infiltration was examined for association with survival in each subtype (excluding MBWNT). Membership of MBIC2 was associated with poorer overall survival (OS) in MBGrp4 (Log-Rank $p=0.0079$, $n = 399$) (Figure 29C). Cox-regression shows several individual cell types are significantly associated with outcome. For Infant MBSHH a greater than median proportion of Tregs was significantly associated with a poor progression free survival (PFS: $z=-2.187$ $p=0.029$ $n=59$) (Figure 30D). In some instances, immune cell estimates provide prognostic information independent of previously established survival associated methylation subtypes. For instance, a greater than median proportion of monocytes in MBGrp4 is associated with a poor prognosis (OS: $z=-2.742$, $p=0.006$, $n=399$, PFS: $z=-2.06$ $p=0.039$ $n=133$). Multivariate analysis shows that this association is significantly prognostic independent of the MBGrp4 High-risk/Low-risk subgrouping (OS: $z=-2.742$, $p=0.006$, $n=399$, PFS: $z=-2.06$ $p=0.039$ $n=133$) (Figure 30E). Likewise, the proportion of Tregs distinguishes two groups within the previously described MBGrp3 Low Risk subtype with significantly different survival (Log-Rank $p<0.001$, 5yrEFS 88% vs 52%) (Figure 29D). This demonstrates that immune infiltration estimates are able to add additional prognostic information not readily available from previous methylation-based analysis.

6.3.5 Differences in proportion of immune cell infiltration in HGG are associated with

subtype, Histone/MAPK mutation, clinicopathological characteristics and prognosis

A cohort of 401 primarily paediatric High-Grade Glioma (pHGG) samples were analysed. pHGG were on average infiltrated predominantly by monocytes (26% of non-cancer cells), Tregs (15%) and eosinophils (13%) (Figure 31A). CD8T infiltration in pHGG was generally less than MB and MRT. Several cell types varied significantly with respect to tumour subgroup i.e. WT-A, WT-B, WT-C, IDH, GBM G34 & GBM K27. These include monocytes, CD8T, TILs and eosinophils (each $p<0.001$), Figure 31A,B). Post-hoc testing shows significantly greater monocytes in WT-A vs other subgroups, (2.9-fold, all comparisons $p<0.001$), significantly greater CD8T in GBM with G34 mutations, (1.7-fold, all comparisons $p<0.05$) and significantly less eosinophils in GBM with G34 mutations (2.3-fold less, all comparisons $p<0.001$). Furthermore, the number of TILs and indeed the overall level of immune cell infiltration is significantly higher in the WT-A subgroup (1.6-fold greater, and 1.9-fold greater respectively, all comparisons $p<0.01$) and significantly lower in GBM-G34 than other pHGGs (1.8-fold and 2.0-fold

respectively, all comparisons $p < 0.01$). The WT-A subgroup generally contains pHGG otherwise referred to as PXA or LGG-like, they are also enriched for MAPK mutations.

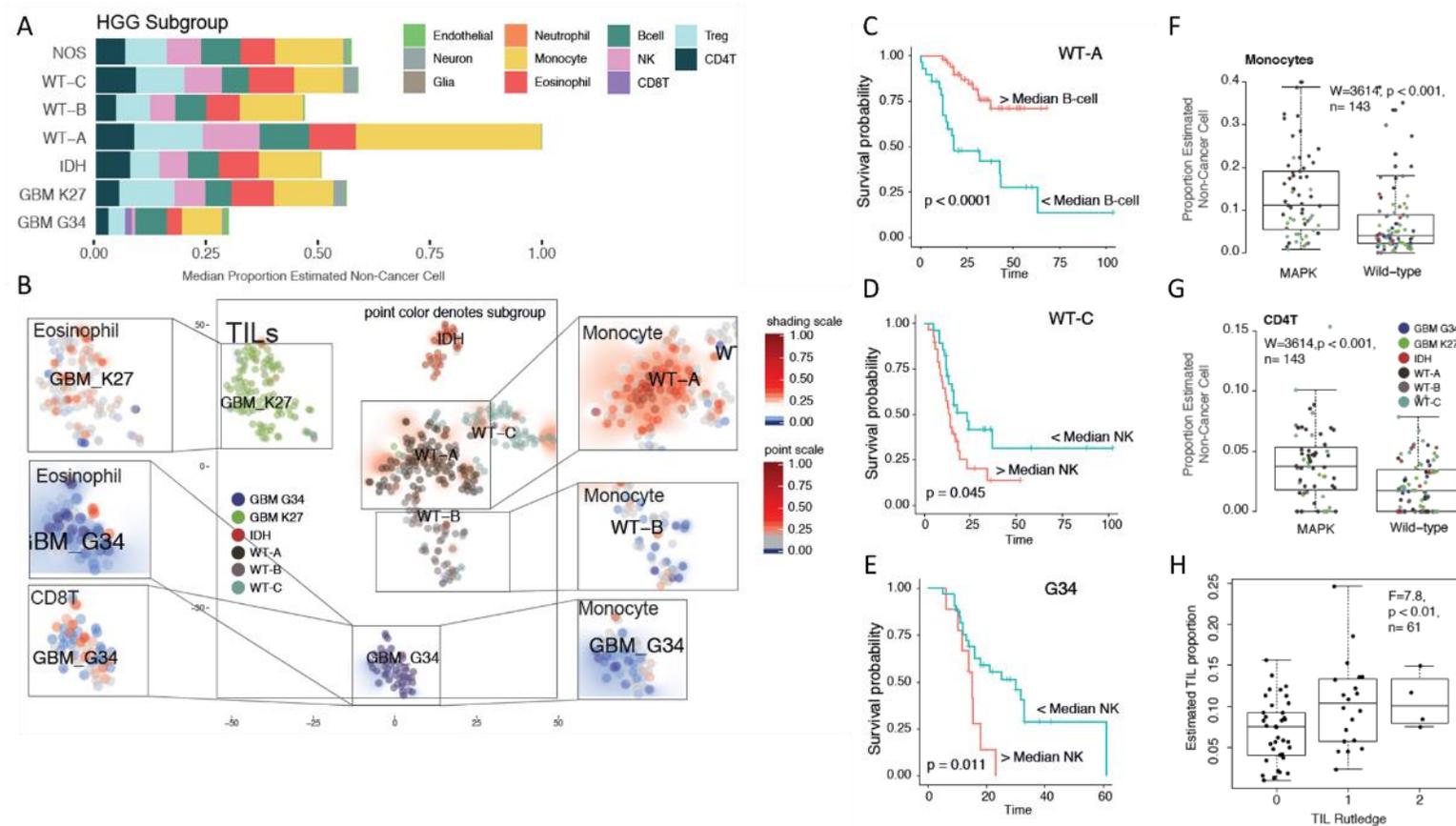


Figure 31. A) Barplots of the estimated median infiltration of specific cell types as a proportion of all non-cancer cell types (range scaled from 0-1) in 401 pHGG. B) t-SNE plot representing the methylation profiles of 401 pHGG. The colours of dots in the central panel map to subgroup. Background shading represents the 2D spatial density estimation of the amount of Tumour Infiltrating Lymphocytes (TILs); red shading equals relatively greater than average infiltration and blue less than average. Exploded side panels represent enlarged areas of interest wherein both dot colour and background shading represent the relative amount of the particular immune cell infiltration denoted. Red denotes relatively greater than average infiltration and blue less than average. C, D, E) Kaplan-Meier plot showing significant difference in overall survival in WT-A, WT-C and G34 subgroups by low (< median) or high (> median) levels of B-cell and NK infiltration. F) Boxplot showing the proportion of monocytes and CD4T cells in pHGG by presence/absence of a MAPK mutation. G) Boxplot showing TIL proportion as estimated by methylCIBERSORT for a subset of pHGG samples for which histopathology-based estimates of lymphocyte infiltration were available. H) Estimates of TILs were significantly greater in patients classified as Categories 1 (present) or 2 (abundant) than Category 0 (absent).

Consensus clustering of pHGG immune cell estimates identifies an optimal 3 immune clusters referred to here as pHGGIC1-3 which cut across each of the pHGG subtypes (Figure 6SA). pHGGIC3 overlaps primarily with panCNSIC1/panCNSIC2 and pHGGIC2 overlaps with panCNSIC3 (Figure 3SC). pHGGIC1 constitutes 31% (126/401) of all pHGG and is characterised by high proportions of Tregs, eosinophils, NK and CD4T. pHGGIC2 constitutes 17% (71/401) of all pHGG and is characterised by high proportions of monocytes and a disproportionately high frequency of WT-A subtypes; 77% (55/73) of all pHGGIC2 are also WT-A. pHGGIC3 constitutes 51% (204/401) of all pHGG and tumours show intermittently moderate levels of CD8T and relatively low levels of other infiltrating immune cell types. 87% (43/49) of all GBM G34 belong to this cluster (Figure 32A,B).

Examining the association of cell infiltration with survival within each of the pHGG subgroups using cox-regression reveals the following significant associations Lower than median concentrations of B-cell and CD8T in WT-A patients are associated with a poor OS ($z=3.735$, $p<0.001$, $n=80$ & $z=1.991$, $p=0.047$, $n=80$ respectively). Higher than median concentrations of CD4T and NK in GBM G34 patients is associated with a poor OS ($z=-2.193$, $p=0.028$, $n=42$ & $z=-2.417$, $p=0.016$, $n=42$ respectively) (Figure 31C & Figure 32C,D).

Clinico-pathological/biological features examined for association with estimated cell types include WHO stage, gender, age <1 or age < 3, presence of BRAF and/or other MAPK mutation. Several immune types were significantly associated with these clinicopathological criteria. As previously noted (Mackay *et al.*, 2018) the presence of MAPK mutations was associated with higher immune cell infiltration; specifically of monocytes and CD4T cells (W=3614 & W=3453 respectively, both $p < 0.001$). For a subset of samples histopathology-based estimates of lymphocyte infiltration were available which categorised patients as per Rutledge *et al.* (2013). Estimates of TILs were significantly greater in patients classified as Categories 1 (present) or 2 (abundant) than Category 0 (absent) ($F=7.839$, $p < 0.01$, $n = 61$). Again, taken as a whole, the significant relationships between molecular subgroup, prognosis, mutation and immune infiltration in pHGG are clear.

6.4 Discussion

Using a methylation-based deconvolution analysis the TIME of >6000 individual (primarily paediatric) CNS tumours was estimated. Diversity in TIME composition across these CNS tumours was demonstrated as well as significant associations variously with tumour type, subtype, stage, grade, location, mutation and survival. The notion of the CNS, and by association CNS tumours, as immune privileged and inaccessible to immune cells is increasingly outdated (Quail and Joyce, 2013), nevertheless the analysis lends weight to the idea of a diverse TIME across a wide range of CNS tumours.

The implications of the results are as follows. First, that the nature of immune cell content is associated with - but not exclusively dictated by - a particular tumour type or subtype. Second, that at least three broad CNS TIME subgroups strongly associated with tumour type and grade can be identified by clustering immune cell types and that within individual tumour types (MB, ATRT, pHGG) further immune subgroups may be described. Immune subgroups cut across the conventional CNS molecular tumour subgroups such that a patient may simultaneously belong to a given molecular subgroup and also independently a particular immune subgroup. Furthermore, these immune subgroups have different immunophenotypic characteristics (different CYT scores, expression of PDL1, etc) and are associated with WHO Grade. Third, that key molecular features recognised as molecular drivers, such as MYC amplification in MB or H3.3G34 mutations in HGG, are associated with distinct TIMEs and particular infiltrating cell types raising the possibility that these mutations are directly influencing

the tumour microenvironment; perhaps as adjunct to their intrinsic oncogenic mechanism. Fourth, that by extracting molecular information about TIME it is possible to access significant prognostic information independent of conventional molecular subgroups raising the possibility of their future incorporation into existing prognostic biomarker schemes. It should be noted that prognostic associations with immune cell infiltration appear to be context dependent; increased CD8T infiltration, for instance, does not universally denote a poor outcome.

The results are broadly in accordance with the small number of recent investigations into immune infiltration in paediatric CNS tumours. Mackay *et al.* (2018) identified a relative lack of TILs in histone mutant pHGG compared to hypermutator and PXA-like (WT-A) subgroups and this is borne out by the analysis here. Bockmayr *et al.* (2019) expression analysis of a mixed cohort of adult and paediatric gliomas identifies 4 immune clusters (including monocyte and T-cell dominated clusters) not wholly inconsistent with the results here. They show some associations with overall survival, however these are mainly within the older (>40 years) and IDH mutated subgroup. Bockmayr *et al.* (2018) also analysed expression (by microarray) of immune markers in 763 medulloblastomas and concluded, similarly, that MBSHH tumours had larger numbers of T-cells overall than other subgroups. In contrast to findings here, they did not identify associations with MB survival as was the case for Vermeulen *et al.* (2018) study of 26 MB patients.

methyCIBERSORT is a method of convenience especially given the prevalent use of methylation profiling within paediatric CNS tumours. Limitations of tumour biopsies and representative sampling notwithstanding, the analysis provides much breadth but clearly not the depth that may be achieved by single cell RNA-seq analysis. The analysis is further limited by its reliance on pure cell populations with no guarantee that the methylation signatures of these processed cells are identical to those within the tumour stroma. It should also be noted that there is likely “dark-matter” i.e. immune infiltration for which the reference population are absent or incomplete. Nevertheless, there have been several efforts to validate and benchmark the estimates using simulations and parallel expression/protein-based methods and provide justification for the broad accuracy of this approach in CNS tumours.

Finally, the results and the immune clusters developed here indicate important differences in TIME across paediatric brain tumour types. Immune clusters are clearly related to the expression of conventional immune targets such as PDL1 in MB and

ATRT and in a broad sense indicate which immune subgroups are “hot” or “cold”. The immune clusters identified break down, broadly speaking, into the Monocyte dominated (i.e. panCNSIC3, and pHGGIC1/2) the balanced or CD4+T type (i.e. panCNSIC1, MBIC2/3 and MRTIC2) and the CD8+T type (i.e. panCNSIC2, MBIC1/4 and pHGGIC3). With such information one may in future begin to match individuals or groups of individuals TIMEs to immunotherapy responses or lack thereof. Even in the most simplistic terms it seems to follow that an a priori paucity of infiltrating Cytotoxic T Lymphocytes and the lack of a supportive TIME may be uncondusive to immune checkpoint blockade as a therapeutic strategy, but instead may be amenable to approaches which alter the TIME or genetically redirect T-cell immunity.

In conclusion, this analysis gives first indications of the potential future therapeutic and prognostic possibilities of immuno-methylomic profiling as an adjunct to methylation/expression-based sub-classification. A future in-depth high-resolution approach incorporating spatial information is now required and *in silico* deconvolution approaches may ultimately be used to triage and to inform selection of immunotherapy approaches in paediatric CNS tumour patients.

7 Summary and Discussion

7.1 Malignant Rhabdoid Tumours

MRT are a rare and aggressive paediatric malignancy with unmet clinical need owing to the lack of effective therapies, and poor response to current therapy approaches. MRT can occur throughout the body and present different challenges when diagnosed in the CNS as ATRT and in other parts of the body as ECRT, limiting therapeutic options.

The main and typically only molecular feature of these tumours, SMARCB1 has been shown to play a key role in tumorigenesis in MRT and is present in a majority of cases (Versteeg *et al.*, 1998; Biegel *et al.*, 1999). SMARCB1 knock-out experiments in mouse models show rapid development of tumours including sarcomas and tumours resembling MRT transcriptional profiling (Klochender-Yeivin *et al.*, 2000; Guidi *et al.*, 2001; Han *et al.*, 2016). In addition, SMARCB1 loss has been linked to deregulation in a number of key developmental pathways such as WNT and SHH signalling, as well as effectors of chromatin remodelling such as *EZH2*. (Jagani *et al.*, 2010; Wilson *et al.*, 2010; Mora-Blanco *et al.*, 2014).

Despite its stable genome, MRT shows considerable heterogeneity in tumour appearance (Fanburg-Smith *et al.*, 1998), response to therapy and ultimately survival (Modena *et al.*, 2013; Abu Arja *et al.*, 2018). In an effort to characterise this heterogeneity, a number of subgrouping studies have been carried out since 2015 seeking to identify the number and biological character of putative subgroups. A number of subgrouping schemes have been proposed, most focusing on subgrouping only one type of MRT. The numbers of subgroups differ between publications and while features identified hint at common biology such as the difference between a neural lineage in SHH/Group1 to mesenchymal differentiation in MYC/Group2B as well as the common overexpression of HOX cluster genes in both ATRT and ECRT (Chun *et al.*, 2016; Han *et al.*, 2016; Johann *et al.*, 2016; Torchia *et al.*, 2016), currently consensus definition encompassing all MRT types exists.

This study was developed in order to interrogate current subgrouping strategies, develop a consensus and make recommendations on the direction of future MRT clinical and molecular studies. In order to do this, the relationship between MRT occurring at different locations in the body were compared to identify whether a common MRT-specific molecular signature can be identified. Subsequently, a meta-analysis of current MRT subgroups was carried out to identify optimum subgroups in a combined cohort of MRT and to explore additional subgroups as potential novel targets

for further research. Additional cases from CCLG cancer centres were collected as well as detailed clinical data and survival information. Subgroup specific survival analyses were carried out as well as more traditional comparisons between previously identified factors. Finally, prompted by association of MRT subgroups with immune signalling, an analysis to estimate immune cell content in a multiple CNS tumours, including MRT was carried out.

Parallel to this study, an international collaboration between Germany, Canada, France and the United Kingdom was launched in 2017 in order to generate a consensus ATRT subgrouping scheme. Data from the study is not presented in the main thesis body and instead a more comprehensive subgrouping analysis was carried out to examine subgroups in MRT as a whole. The ATRT consensus study text has been accepted for publication and is currently available as a pre-publication document ([doi: 10.1093/neuonc/noz235](https://doi.org/10.1093/neuonc/noz235)).

7.2 Investigating the biological relationship between ATRT and ECRT

To date, this thesis is one of the only large-scale studies examining ATRT and ECRT in a combined cohort. Despite a general agreement across the field that ATRT and ECRT share many biological features, clinical and profiling studies have typically restricted their focus on one type of tumour only. There are many good reasons for this, clinically approaches to treating ATRT and ECRT vary with radiotherapy being more common in young patients with ECRT while being typically deferred in patients under 3 years old in ATRT. Secondly, any chemotherapeutic agents identified for ATRT therapy would require access to the CNS via the blood-brain barrier.

Despite this, the analysis demonstrated a high degree of genetic and epigenetic overlap between ATRT and ECRT especially when compared against other embryonal tumours and tumours which originate from a single tissue or location which display more variability. An MRT-specific signature was identified highlighting common gene expression across ATRT and ECRT. As well as this, it was shown that subgrouping strategies which were developed in ATRT only can be recapitulated in a combined MRT cohort.

Based on the findings in this thesis, ATRT and ECRT are highly compatible with combined analyses with both transcriptomic and epigenetic approaches. Shared gene expression and high degree of overlap between ATRT and ECRT when compared to other embryonal tumours points to common biology that could be exploited

therapeutically. One conclusion of this study is that future analysis of MRT should encompass all tumour types and that any potential therapeutic strategies used in ATRT be considered in ECRT and *vice-versa*.

7.3 Generating a molecular signature of MRT subgroups

Subgrouping studies have proposed a number of different subgrouping schemes and their defining characteristics such as differential expression of neural and mesenchymal lineage genes, differences in *SMARCB1* mutation type, differences in age and location for CNS tumours. (Chun *et al.*, 2016; Han *et al.*, 2016; Johann *et al.*, 2016; Torchia *et al.*, 2016). In addition to this, a DNA methylation based classifier MNP2.0 has been developed which includes ATRT (Capper *et al.*, 2018).

In order to evaluate currently utilised subgrouping schemes in ATRT and ECRT and compare and contrast optimum subgroup number and content, a meta-analysis using a consensus NMF approach was carried out. This method utilised resampling to robustly call subgroups and identify poorly clustering samples. Using published data as well as a newly profiled cohort defined in this thesis, 450K/EPIC methylation analysis was carried out on a combined MRT cohort, and a gene expression array analysis on combined MRT HGU133Plus2.0 arrays.

The resulting subgrouping shows a high degree of concordance between previous studies and identified similar gene expression and methylation features between previous subgrouping strategies and the consensus subgrouping approach utilised here. This thesis presents a 3 group consensus subgrouping strategy which aligns a number of previous schemes and provides a robust method for analysis of additional profiles.

In addition, this thesis expanded on current subgrouping by exploring greater numbers of subgroups. The resulting subgrouping scheme identified novel clinicopathological differences between subgroups, such as localisation and age differences between SHH.Infratentorial and SHH.Supratentorial, or identified a novel subgroup with a unique methylation landscape and characterised by lack of broad chromosome 22q changes. These novel subgroups were previously only described as heterogeneity in the existing subgroup scheme (Johann *et al.*, 2016).

The clinical relevance of subgrouping MRT has been previously examined in a single study which found significant survival differences using subgroups the authors defined (Torchia *et al.*, 2015). Whether the consensus subgrouping scheme proposed here will

further expand current stratification in MRT remains to be seen. The generation of a large, well annotated clinical dataset will be required to rigorously test different clinical outcomes based on subgroup membership

7.4 Survival analysis in MRT

A cohort of 113 MRT cases was generated as part of this thesis and collected from CCLG centres and the Brain UK registry. Clinical data was obtained for cases where possible and survival information was collected to test the association of different clinicopathological features with MRT survival.

The most significant factors that affected survival in this cohort were radiotherapy, evidence of metastasis, surgical resection extent, and age of patients at diagnosis. Other factors such as consensus NMF subgroup, chromosomal arm loss in 22q, sex were not significantly associated with survival differences.

However, despite not initially being a significant predictor of survival in a univariate approach, consensus NMF $k = 3$ subgroup contributed to a significant survival stratification scheme when combined with a newly proposed (Fruehwald *et al.*, 2018) age stratification of < 1 year old and assignment to the TYR subgroup. Significant survival differences were shown in the MRT and ATRT cohorts with both the 5-year OS and PFS being $>50\%$ in patients allocated to the moderate risk group.

In addition to this, the survival analysis recapitulated results showed previously in ATRT only for factors such as radiotherapy and surgical resection (Torchia *et al.*, 2015) extent validating them in this cohort.

7.5 Estimating immune infiltration in CNS tumours

Expression of genesets associated with immune function and development was noted in this study in the MYC subgroup of MRT. Previous publications have also identified immune system genes as being significantly differentially expressed across MRT subgroups (Chun *et al.*, 2016)

Immune infiltration has been identified as an important hallmark of cancer as a whole (Hanahan and Weinberg, 2011) and more recently identified to be therapeutically targetable with a number of novel therapeutic agents (Rizvi *et al.*, 2015; Robert *et al.*, 2015b; Herbst *et al.*, 2016; Shen and Zhao, 2018; Tomita *et al.*, 2019). However, these studies have shown that response to therapy appears to be heavily influenced by the nature of the tumour immune infiltration and that understanding of the nature of the

TIME in a particular tumour can shed light on whether it would be a suitable target for immune therapeutics.

Carrying out a large-scale IHC immune cell analysis on the cohort developed as part of this thesis was not possible. However, a number of *in-silico* approaches for estimating TIME from genomic data have been developed (Newman *et al.*, 2015; Chakravarthy *et al.*, 2018). Based on the CIBERSORT method, adapted for DNA methylation array it was possible to carry out an initial estimation of tumour immune infiltration on large cohort of pan-CNS tumours including ATRT and to compare and contrast the types of immune differences seen across different types of CNS tumours originally published as part of the MNP2.0, as well as data provided by collaborators..

The method required the generation of a methylation signature matrix specific to immune cells of interest which was generated from publicly available immune cell methylation array profiling data for 11 immune and normal cell populations (CD8T, CD4T, Treg, Bcell, natural killer, monocytes, neutrophils, eosinophils, neuronal and glial tissue and endothelial tissue). Benchmarking and validation were carried out to measure the accuracy of the estimation which showed significant, high correlation to known flow-cytometry validated profiles and the estimation.

Analysis of MRT immune infiltration estimation identified a number of immune clusters based on the level of estimated infiltration. Novel associations with survival for specific cell types such Bcells in ATRT and CD8T in ECRT. Comparative analysis in other CNS tumour types identified a large number of different significant associations with immune cluster, such as tumour grade, patient age, and subgroups if tumours. The developed analysis pipeline provides a convenient and powerful method estimation of immune cell infiltration in CNS tumours.

7.6 Limitations

During the development of this thesis a number of limitations were identified as part of the study and may present potential caveats on results provided here.

Clinical data collected as part of this thesis cohort was not well annotated for therapeutic intent. Evidence of therapy was taken as a positive indicator of an attempt to treat the MRT but lack of therapeutic information was not excluded as this would heavily reduce the effective cohort size. In addition, chemotherapy information was highly variable and it was not always possible to ascertain whether HDCT was utilised. Germline SMARCB1 mutation testing in the cohort was available for a small number

of patients but was generally missing for the majority of cases. As this is a largely retrospective cohort, it is not possible to capture this data at this time.

In DNA methylation array profiling of samples, array quality was analysed and a decision was made for what cut-off threshold for a number of array quality metrics to use. As a large proportion of the dataset was FFPE, array quality scores were typically below recommended thresholds for fresh and high quality tissue arrays. Extra measures were taken in order to avoid any technical error effects from poor quality profiling and any samples which showed spurious clustering or an exceedingly noisy and poor quality array profile were excluded from the study.

SMARCB1 mutation information is only available as part of the published cohorts utilised in this thesis. Statistical analyses of enrichment of specific *SMARCB1* mutation types were therefore based on a smaller proportion of the overall cohort used in Chapter 3 & 4. In order to overcome this limitation, chromosome copy-number estimation in methylation array was used as a surrogate measure for chromosomal arm loss in chromosome 22q, however this method is not full proof and can be subject to noise in the array due to the overall array quality.

CIBERSORT analysis was carried out using a signature matrix of pure cell populations derived from publicly available data. The pure cell populations isolated as part of the datasets were not isolated with the same method and employed a number of different markers for cell-sorting and positive population enrichment. As there was no control over the methodology of the sorting and purification process, there is no guarantee that the populations included in the signature matrix are completely pure and represent a wholly pure population of the immune cells of interest. Due to this inherent uncertainty, despite high correlations with validation datasets, the results of the immune infiltration estimation were treated as an estimation only and that any subsequent findings would require validation before they can be treated as biologically real.

7.7 Future work

7.7.1 Expanding survival analysis cohort

A number of independently published survival cohorts are available for MRT however the data for these has largely not been made publically available. A collaboration between a number of large clinical data studies combined with the current subgroup consensus analysis would allow for a much more powerful survival analysis of clinicopathological factors. Work is currently ongoing to complete clinical annotation

for the MRT cohort developed as part of this study. Where available clinical annotation should be sought to be as complete as possible.

7.7.2 Further identification of subgroup-specific features

This study has highlighted an MRT subgrouping scheme that incorporates a number of previously proposed subgroupings and combines both ATRT and ECRT. Currently available datasets can be utilised in novel ways in order to further define the differences between subgroups. Differentially methylated DNA regions can be inferred from constituted CpG probe regions of known association and significant correlation. In addition to this, methylation array data can be correlated with currently available transcriptomic analysis to show direct correlation between gene expression DNA methylation in significantly differentially methylated genes.

7.7.3 Novel 'omics profiling

A number of new and established platforms are now optimised for use in low quality material such as FFPE. It would be possible to profile FFPE material using low-depth RNA-sequencing or bisulphite sequencing as an alternative to the more dated DNA methylation array and gene expression array. Generating a large cohort of detailed transcriptomic data for MRT is a highly important future goal.

7.8 Final summary

This study has applied a wide array of bioinformatic methods in order to interrogate heterogeneity in MRT biology. Differences between ATRT and ECRT were analysed in order to justify a future direction of combined MRT, subgroup-based research. Data presented in this thesis highlights a large degree of biological overlap between MRT tumours as a whole and provides evidence for a combined approach of targeting all MRT for future study and clinical analysis.

Subgroup consensus analysis identified robust, well annotated and characterised subgroups in MRT providing a methodology for future consensus efforts as well as further characterising novel aspects of subgrouping by identifying novel subgroup features such as tumour location, age and DNA methylation levels. In addition, novel approaches using already generated data, allow for new biological features to be explored such as the immune landscape of MRT. This thesis makes provides a number of significant foundations for future directed study in MRT.

Work carried out as part of this thesis has also informed an international ATRT subgroup consensus, which is currently in the process of being published.

Bibliography

- Abdelmoula, W.M., Balluff, B., Englert, S., Dijkstra, J., Reinders, M.J.T., Walch, A., McDonnell, L.A. and Lelieveldt, B.P.F. (2016) 'Data-driven identification of prognostic tumor subpopulations using spatially mapped t-SNE of mass spectrometry imaging data', *113*(43), pp. 12244-12249.
- Abu Arja, M.H., Patel, P., Shah, S.H., Auletta, J.J., Meyer, E.K., Conley, S.E., Aldrink, J.H., Pindrik, J.A. and AbdelBaki, M.S. (2018) 'Synchronous Central Nervous System Atypical Teratoid/Rhabdoid Tumor and Malignant Rhabdoid Tumor of the Kidney: Case Report of a Long-Term Survivor and Review of the Literature', *World Neurosurg*, *111*, pp. 6-15.
- Aloisi, F. (2001) 'Immune function of microglia', *Glia*, *36*(2), pp. 165-79.
- Angelova, M., Charoentong, P., Hackl, H., Fischer, M.L., Snajder, R., Krogsdam, A.M., Waldner, M.J., Bindea, G., Mlecnik, B., Galon, J. and Trajanoski, Z. (2015) 'Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy', *Genome Biol*, *16*, p. 64.
- Ann Zimmerman, M., Goumnerova, L.C., Proctor, M., Michael Scott, R., Marcus, K., Pomeroy, S.L., Turner, C.D., Chi, S.N., Chordas, C. and Kieran, M.W. (2005) 'Continuous remission of newly diagnosed and relapsed central nervous system atypical teratoid/rhabdoid tumor', *Journal of Neuro-Oncology*, *72*(1), pp. 77-84.
- Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A. and Campilho, A.J.P.o. (2017) 'Classification of breast cancer histology images using convolutional neural networks', *12*(6), p. e0177544.
- Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D. and Irizarry, R.A.J.B. (2014) 'Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays', *30*(10), pp. 1363-1369.
- Athale, U.H., Duckworth, J., Odame, I. and Barr, R. (2009) 'Childhood atypical teratoid rhabdoid tumor of the central nervous system: a meta-analysis of observational studies', *J Pediatr Hematol Oncol*, *31*(9), pp. 651-63.
- Bambakidis, N.C., Robinson, S., Cohen, M. and Cohen, A.R. (2002) 'Atypical teratoid/rhabdoid tumors of the central nervous system: clinical, radiographic and pathologic features', *Pediatr Neurosurg*, *37*(2), pp. 64-70.
- Becht, E., de Reynies, A., Giraldo, N.A., Pilati, C., Buttard, B., Lacroix, L., Selves, J., Sautes-Fridman, C., Laurent-Puig, P. and Fridman, W.H. (2016) 'Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy', *Clin Cancer Res*, *22*(16), pp. 4057-66.
- Beckwith, J.B. and Palmer, N.F. (1978) 'Histopathology and prognosis of Wilms tumors: results from the First National Wilms' Tumor Study', *Cancer*, *41*(5), pp. 1937-48.
- Bertaut, A., Truntzer, C., Madkouri, R., Kaderbhai, C.G., Derangère, V., Vincent, J., Chauffert, B., Aubriot-Lorton, M.H., Farah, W., Mourier, K.L., Boidot, R. and Ghiringhelli, F. (2016) 'Blood baseline neutrophil count predicts bevacizumab efficacy in glioblastoma', *Oncotarget*, *7*(43), pp. 70948-70958.
- Biegel, J.A., Rorke, L.B. and Emanuel, B.S. (1989) 'Monosomy 22 in rhabdoid or atypical teratoid tumors of the brain', *N Engl J Med*, *321*(13), p. 906.

- Biegel, J.A., Tan, L., Zhang, F., Wainwright, L., Russo, P. and Rorke, L.B. (2002) 'Alterations of the hSNF5/INI1 gene in central nervous system atypical teratoid/rhabdoid tumors and renal and extrarenal rhabdoid tumors', *Clin Cancer Res*, 8(11), pp. 3461-7.
- Biegel, J.A., Zhou, J.Y., Rorke, L.B., Stenstrom, C., Wainwright, L.M. and Fogelgren, B. (1999) 'Germline and acquired mutations of INI1 in atypical teratoid and rhabdoid tumors', *Cancer Res*, 59(1), pp. 74-9.
- Binnewies, M., Roberts, E.W., Kersten, K., Chan, V., Fearon, D.F., Merad, M., Coussens, L.M., Gabilovich, D.I., Ostrand-Rosenberg, S., Hedrick, C.C., Vonderheide, R.H., Pittet, M.J., Jain, R.K., Zou, W., Howcroft, T.K., Woodhouse, E.C., Weinberg, R.A. and Krummel, M.F. (2018) 'Understanding the tumor immune microenvironment (TIME) for effective therapy', *Nature Medicine*, 24(5), pp. 541-550.
- Birks, D.K., Donson, A.M., Patel, P.R., Sufit, A., Algar, E.M., Dunham, C., Kleinschmidt-DeMasters, B.K., Handler, M.H., Vibhakar, R. and Foreman, N.K. (2013) 'Pediatric rhabdoid tumors of kidney and brain show many differences in gene expression but share dysregulation of cell cycle and epigenetic effector genes', *Pediatr Blood Cancer*, 60(7), pp. 1095-102.
- Biswas, A., Julka, P.K., Bakhshi, S., Suri, A. and Rath, G.K. (2015) 'Intracranial atypical teratoid rhabdoid tumor: current management and a single institute experience of 15 patients from north India', *Acta Neurochir (Wien)*, 157(4), pp. 589-96.
- Bockmayr, M., Klauschen, F., Maire, C.L., Rutkowski, S., Westphal, M., Lamszus, K., Schüller, U. and Mohme, M. (2019) 'Immunologic Profiling of Mutational and Transcriptional Subgroups in Pediatric and Adult High-Grade Gliomas', *Cancer Immunology Research*.
- Bockmayr, M., Mohme, M., Klauschen, F., Winkler, B., Budczies, J., Rutkowski, S. and Schuller, U. (2018) 'Subgroup-specific immune and stromal microenvironment in medulloblastoma', *Oncoimmunology*, 7(9), p. e1462430.
- Boland, P.M. and Ma, W.W. (2017) 'Immunotherapy for Colorectal Cancer', *Cancers (Basel)*, 9(5).
- Boon, T., Cerottini, J.C., Van den Eynde, B., van der Bruggen, P. and Van Pel, A. (1994) 'Tumor antigens recognized by T lymphocytes', *Annu Rev Immunol*, 12, pp. 337-65.
- Bourdeaut, F., Freneau, P., Thuille, B., Lellouch-Tubiana, A., Nicolas, A., Couturier, J., Pierron, G., Sainte-Rose, C., Bergeron, C., Bouvier, R., Rialland, X., Laurence, V., Michon, J., Sastre-Garau, X. and Delattre, O. (2007) 'hSNF5/INI1-deficient tumours and rhabdoid tumours are convergent but not fully overlapping entities', *J Pathol*, 211(3), pp. 323-30.
- Brennan, B., De Salvo, G.L., Orbach, D., De Paoli, A., Kelsey, A., Mudry, P., Francotte, N., Van Noesel, M., Bisogno, G., Casanova, M. and Ferrari, A. (2016) 'Outcome of extracranial malignant rhabdoid tumours in children registered in the European Paediatric Soft Tissue Sarcoma Study Group Non-Rhabdomyosarcoma Soft Tissue Sarcoma 2005 Study—EPCSSG NRSTS 2005', *European Journal of Cancer*, 60, pp. 69-82.
- Brennan, B., Stiller, C. and Bourdeaut, F. (2013) 'Extracranial rhabdoid tumours: what we have learned so far and future directions', *The Lancet Oncology*, 14(8), pp. e329-e336.
- Brunet, J.-P., Tamayo, P., Golub, T.R. and Mesirov, J.P. (2004) 'Metagenes and molecular pattern discovery using matrix factorization', 101(12), pp. 4164-4169.

Buscariollo, D.L., Park, H.S., Roberts, K.B. and Yu, J.B. (2012) 'Survival outcomes in atypical teratoid rhabdoid tumor for patients undergoing radiotherapy in a Surveillance, Epidemiology, and End Results analysis', *Cancer*, 118(17), pp. 4212-9.

Butowski, N., Colman, H., De Groot, J.F., Omuro, A.M., Nayak, L., Wen, P.Y., Cloughesy, T.F., Marimuthu, A., Haidar, S., Perry, A., Huse, J., Phillips, J., West, B.L., Nolop, K.B., Hsu, H.H., Ligon, K.L., Molinaro, A.M. and Prados, M. (2016) 'Orally administered colony stimulating factor 1 receptor inhibitor PLX3397 in recurrent glioblastoma: an Ivy Foundation Early Phase Clinical Trials Consortium phase II study', *Neuro Oncol*, 18(4), pp. 557-64.

Capper, D., Jones, D.T.W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E., Kratz, A., Wefers, A.K., Huang, K., Pajtler, K.W., Schweizer, L., Stichel, D., Olar, A., Engel, N.W., Lindenberg, K., Harter, P.N., Braczynski, A.K., Plate, K.H., Dohmen, H., Garvalov, B.K., Coras, R., Hölsken, A., Hewer, E., Bewerunge-Hudler, M., Schick, M., Fischer, R., Beschorner, R., Schittenhelm, J., Staszewski, O., Wani, K., Varlet, P., Pages, M., Temming, P., Lohmann, D., Selt, F., Witt, H., Milde, T., Witt, O., Aronica, E., Giangaspero, F., Rushing, E., Scheurlen, W., Geisenberger, C., Rodriguez, F.J., Becker, A., Preusser, M., Haberler, C., Bjerkvig, R., Cryan, J., Farrell, M., Deckert, M., Hench, J., Frank, S., Serrano, J., Kannan, K., Tzirigos, A., Brück, W., Hofer, S., Brehmer, S., Seiz-Rosenhagen, M., Hänggi, D., Hans, V., Rozsnoki, S., Hansford, J.R., Kohlhof, P., Kristensen, B.W., Lechner, M., Lopes, B., Mawrin, C., Ketter, R., Kulozik, A., Khatib, Z., Heppner, F., Koch, A., Jouvét, A., Keohane, C., Mühleisen, H., Mueller, W., Pohl, U., Prinz, M., Benner, A., Zapatka, M., Gottardo, N.G., Driever, P.H., Kramm, C.M., Müller, H.L., Rutkowski, S., von Hoff, K., Frühwald, M.C., Gnekow, A., Fleischhack, G., Tippelt, S., Calaminus, G., Monoranu, C.-M., Perry, A., Jones, C., et al. (2018) 'DNA methylation-based classification of central nervous system tumours', *Nature*, 555, p. 469.

Cavalli, F.M.G., Remke, M., Rampasek, L., Peacock, J., Shih, D.J.H., Luu, B., Garzia, L., Torchia, J., Nor, C., Morrissy, A.S., Agnihotri, S., Thompson, Y.Y., Kuzan-Fischer, C.M., Farooq, H., Isaev, K., Daniels, C., Cho, B.K., Kim, S.K., Wang, K.C., Lee, J.Y., Grajkowska, W.A., Perek-Polnik, M., Vasiljevic, A., Faure-Contier, C., Jouvét, A., Giannini, C., Nageswara Rao, A.A., Li, K.K.W., Ng, H.K., Eberhart, C.G., Pollack, I.F., Hamilton, R.L., Gillespie, G.Y., Olson, J.M., Leary, S., Weiss, W.A., Lach, B., Chambless, L.B., Thompson, R.C., Cooper, M.K., Vibhakar, R., Hauser, P., van Veelen, M.C., Kros, J.M., French, P.J., Ra, Y.S., Kumabe, T., Lopez-Aguilar, E., Zitterbart, K., Sterba, J., Finocchiaro, G., Massimino, M., Van Meir, E.G., Osuka, S., Shofuda, T., Klekner, A., Zollo, M., Leonard, J.R., Rubin, J.B., Jabado, N., Albrecht, S., Mora, J., Van Meter, T.E., Jung, S., Moore, A.S., Hallahan, A.R., Chan, J.A., Tirapelli, D.P.C., Carlotti, C.G., Fouladi, M., Pimentel, J., Faria, C.C., Saad, A.G., Massimi, L., Liau, L.M., Wheeler, H., Nakamura, H., Elbabaa, S.K., Perezpena-Diazconti, M., Chico Ponce de Leon, F., Robinson, S., Zapotocky, M., Lassaletta, A., Huang, A., Hawkins, C.E., Tabori, U., Bouffet, E., Bartels, U., Dirks, P.B., Rutka, J.T., Bader, G.D., Reimand, J., Goldenberg, A., Ramaswamy, V. and Taylor, M.D. (2017) 'Intertumoral Heterogeneity within Medulloblastoma Subgroups', *Cancer Cell*, 31(6), pp. 737-754.e6.

Chakravarthy, A., Furness, A., Joshi, K., Ghorani, E., Ford, K., Ward, M.J., King, E.V., Lechner, M., Marafioti, T., Quezada, S.A., Thomas, G.J., Feber, A. and Fenton, T.R. (2018) 'Pan-cancer deconvolution of tumour composition using DNA methylation', *Nat Commun*, 9(1), p. 3220.

Chatterjee, S.S., Biswas, M., Boila, L.D., Banerjee, D. and Sengupta, A. (2018) 'SMARCB1 Deficiency Integrates Epigenetic Signals to Oncogenic Gene Expression Program Maintenance in Human Acute Myeloid Leukemia', *Mol Cancer Res*, 16(5), pp. 791-804.

Chen, Y.A., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J. and Weksberg, R. (2013) 'Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray', *Epigenetics*, 8(2), pp. 203-9.

Chen, Y.W., Wong, T.T., Ho, D.M., Huang, P.I., Chang, K.P., Shiau, C.Y. and Yen, S.H. (2006) 'Impact of radiotherapy for pediatric CNS atypical teratoid/rhabdoid tumor (single institute experience)', *Int J Radiat Oncol Biol Phys*, 64(4), pp. 1038-43.

Chi, S.N., Zimmerman, M.A., Yao, X., Cohen, K.J., Burger, P., Biegel, J.A., Rorke-Adams, L.B., Fisher, M.J., Janss, A., Mazewski, C., Goldman, S., Manley, P.E., Bowers, D.C., Bendel, A., Rubin, J., Turner, C.D., Marcus, K.J., Goumnerova, L., Ullrich, N.J. and Kieran, M.W. (2009) 'Intensive Multimodality Treatment for Children With Newly Diagnosed CNS Atypical Teratoid Rhabdoid Tumor', *Journal of Clinical Oncology*, 27(3), pp. 385-389.

Chun, H.E., Lim, E.L., Heravi-Moussavi, A., Saberi, S., Mungall, K.L., Bilenky, M., Carles, A., Tse, K., Shlafman, I., Zhu, K., Qian, J.Q., Palmquist, D.L., He, A., Long, W., Goya, R., Ng, M., LeBlanc, V.G., Pleasance, E., Thiessen, N., Wong, T., Chuah, E., Zhao, Y.J., Schein, J.E., Gerhard, D.S., Taylor, M.D., Mungall, A.J., Moore, R.A., Ma, Y., Jones, S.J.M., Perlman, E.J., Hirst, M. and Marra, M.A. (2016) 'Genome-Wide Profiles of Extra-cranial Malignant Rhabdoid Tumors Reveal Heterogeneity and Dysregulated Developmental Pathways', *Cancer Cell*, 29(3), pp. 394-406.

Cieřlik, M. and Chinnaiyan, A.M. (2017) 'Cancer transcriptome profiling at the juncture of clinical translation', *Nature Reviews Genetics*, 19, p. 93.

Clasie, B., Depauw, N., Fransen, M., Goma, C., Panahandeh, H.R., Seco, J., Flanz, J.B. and Kooy, H.M. (2012) 'Golden beam data for proton pencil-beam scanning', *Phys Med Biol*, 57(5), pp. 1147-58.

Coulie, P.G., Van den Eynde, B.J., van der Bruggen, P. and Boon, T. (2014) 'Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy', *Nat Rev Cancer*, 14(2), pp. 135-46.

D'Agostino, P.M., Gottfried-Blackmore, A., Anandasabapathy, N. and Bulloch, K. (2012) 'Brain dendritic cells: biology and pathology', *Acta Neuropathol*, 124(5), pp. 599-614.

De Amorim Bernstein, K., Sethi, R., Trofimov, A., Zeng, C., Fullerton, B., Yeap, B.Y., Ebb, D., Tarbell, N.J., Yock, T.I. and MacDonald, S.M. (2013) 'Early Clinical Outcomes Using Proton Radiation for Children With Central Nervous System Atypical Teratoid Rhabdoid Tumors', *International Journal of Radiation Oncology*Biophysics*, 86(1), pp. 114-120.

Delacher, M., Imbusch, C.D., Weichenhan, D., Breiling, A., Hotz-Wagenblatt, A., Trager, U., Hofer, A.C., Kagebein, D., Wang, Q., Frauhammer, F., Mallm, J.P., Bauer, K., Herrmann, C., Lang, P.A., Brors, B., Plass, C. and Feuerer, M. (2017) 'Genome-wide DNA-methylation landscape defines specialization of regulatory T cells in tissues', *Nat Immunol*, 18(10), pp. 1160-1172.

Devarajan, K. (2008) 'Nonnegative matrix factorization: an analytical and interpretive tool in computational biology', *PLoS Comput Biol*, 4(7), p. e1000029.

Dong, H., Strome, S.E., Salomao, D.R., Tamura, H., Hirano, F., Flies, D.B., Roche, P.C., Lu, J., Zhu, G., Tamada, K., Lennon, V.A., Celis, E. and Chen, L. (2002) 'Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion', *Nat Med*, 8(8), pp. 793-800.

Douglass, E.C., Valentine, M., Rowe, S.T., Parham, D.M., Wilimas, J.A., Sanders, J.M. and Houghton, P.J. (1990) 'Malignant rhabdoid tumor: a highly malignant childhood tumor with minimal karyotypic changes', *Genes Chromosomes Cancer*, 2(3), pp. 210-6.

Drake, C.G., Jaffee, E. and Pardoll, D.M. (2006) 'Mechanisms of immune evasion by tumors', *Advances in immunology*, 90, pp. 51-81.

Dunn, G.P., Bruce, A.T., Ikeda, H., Old, L.J. and Schreiber, R.D. (2002) 'Cancer immunoediting: from immunosurveillance to tumor escape', *Nat Immunol*, 3(11), pp. 991-8.

Dwass, M. (1957) 'Modified randomization tests for nonparametric hypotheses', *The Annals of Mathematical Statistics*, pp. 181-187.

Eaton, K.W., Tooke, L.S., Wainwright, L.M., Judkins, A.R. and Biegel, J.A. (2011) 'Spectrum of SMARCB1/INI1 mutations in familial and sporadic rhabdoid tumors', *Pediatr Blood Cancer*, 56(1), pp. 7-15.

Efron, B. (1992) 'Bootstrap methods: another look at the jackknife', in *Breakthroughs in statistics*. Springer, pp. 569-593.

Elliott, G., Hong, C., Xing, X., Zhou, X., Li, D., Coarfa, C., Bell, R.J., Maire, C.L., Ligon, K.L., Sigaroudinia, M., Gascard, P., Tlsty, T.D., Harris, R.A., Schalkwyk, L.C., Bilenky, M., Mill, J., Farnham, P.J., Kellis, M., Marra, M.A., Milosavljevic, A., Hirst, M., Stormo, G.D., Wang, T. and Costello, J.F. (2015) 'Intermediate DNA methylation is a conserved signature of genome regulation', *Nat Commun*, 6, p. 6363.

Fanburg-Smith, J.C., Hengge, M., Hengge, U.R., Smith, J.S.C. and Miettinen, M. (1998) 'Extrarenal rhabdoid tumors of soft tissue: A clinicopathologic and immunohistochemical study of 18 cases', *Annals of Diagnostic Pathology*, 2(6), pp. 351-362.

Fangusaro, J., Finlay, J., Sposto, R., Ji, L., Saly, M., Zacharoulis, S., Asgharzadeh, S., Abromowitch, M., Olshefski, R., Halpern, S., Dubowy, R., Comito, M., Diez, B., Kellie, S., Hukin, J., Rosenblum, M., Dunkel, I., Miller, D.C., Allen, J. and Gardner, S. (2008) 'Intensive chemotherapy followed by consolidative myeloablative chemotherapy with autologous hematopoietic cell rescue (AuHCR) in young children with newly diagnosed supratentorial primitive neuroectodermal tumors (sPNETs): report of the Head Start I and II experience', *Pediatr Blood Cancer*, 50(2), pp. 312-8.

Fecci, P.E., Mitchell, D.A., Whitesides, J.F., Xie, W., Friedman, A.H., Archer, G.E., Herndon, J.E., 2nd, Bigner, D.D., Dranoff, G. and Sampson, J.H. (2006) 'Increased regulatory T-cell fraction amidst a diminished CD4 compartment explains cellular immune defects in patients with malignant glioma', *Cancer Res*, 66(6), pp. 3294-302.

Finetti, M. (2014) *Next-Generation Sequencing Identifies Mechanisms Of Tumourigenesis Caused By Loss Of SMARCB1 In Malignant Rhabdoid Tumours*. Newcastle University.

Finkelstein-Shechter, T., Gassas, A., Mabbott, D., Huang, A., Bartels, U., Tabori, U., Janzen, L., Hawkins, C., Taylor, M. and Bouffet, E. (2010) 'Atypical teratoid or rhabdoid tumors: improved outcome with high-dose chemotherapy', *J Pediatr Hematol Oncol*, 32(5), pp. e182-6.

Fossati, G., Ricevuti, G., Edwards, S.W., Walker, C., Dalton, A. and Rossi, M.L. (1999) 'Neutrophil infiltration into human gliomas', *Acta Neuropathologica*, 98(4), pp. 349-354.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. and Paul, C.L. (1992) 'A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands', *Proc Natl Acad Sci U S A*, 89(5), pp. 1827-31.

Fruehwald, M., Hasselblatt, M., Schneppenheim, R., Nemes, K., Bens, S., Johan, P., Hauser, P., Quiroga, E., Solano-Paez, P., Biassoni, V., Gil-da-Costa, M.J., Perek-Polnik, M., van de Wetering, M., Ebinger, M., Fleischhack, G., Furtwaengler, R., Hernaiz-Driever, P., Reinhard, H., Rukowski, S., Schlegel, P.-G., Schmid, I., Timmermann, B., Kordes, U., Gerss, J., Kerl, K., Nysom, K., Siebert, R., Kool,

- M. and Graf, N. (2018) 'ATRT-06. CLINICAL AND MOLECULAR RISK FACTORS IN CHILDREN WITH ATYPICAL TERATOID/RHABDOID TUMOUR (AT/RT) - EVIDENCE FROM THE EU-RHAB REGISTRY', *Neuro-Oncology*, 20(suppl_2), pp. i28-i28.
- Fruhwald, M.C., Hasselblatt, M., Wirth, S., Kohler, G., Schneppenheim, R., Subero, J.I., Siebert, R., Kordes, U., Jurgens, H. and Vormoor, J. (2006) 'Non-linkage of familial rhabdoid tumors to SMARCB1 implies a second locus for the rhabdoid tumor predisposition syndrome', *Pediatr Blood Cancer*, 47(3), pp. 273-8.
- Fujita, Y., Rooney, C.M. and Heslop, H.E. (2007) 'Adoptive cellular immunotherapy for viral diseases', *Bone Marrow Transplantation*, 41, p. 193.
- Gardiner-Garden, M. and Frommer, M. (1987) 'CpG islands in vertebrate genomes', *J Mol Biol*, 196(2), pp. 261-82.
- Gardner, S.L., Asgharzadeh, S., Green, A., Horn, B., McCowage, G. and Finlay, J. (2008) 'Intensive induction chemotherapy followed by high dose chemotherapy with autologous hematopoietic progenitor cell rescue in young children newly diagnosed with central nervous system atypical teratoid rhabdoid tumors', *Pediatric Blood & Cancer*, 51(2), pp. 235-240.
- Gatalica, Z., Snyder, C., Maney, T., Ghazalpour, A., Holterman, D.A., Xiao, N., Overberg, P., Rose, I., Basu, G.D., Vranic, S., Lynch, H.T., Von Hoff, D.D. and Hamid, O. (2014) 'Programmed cell death 1 (PD-1) and its ligand (PD-L1) in common cancers and their correlation with molecular cancer type', *Cancer Epidemiol Biomarkers Prev*, 23(12), pp. 2965-70.
- Gehrmann, J., Matsumoto, Y. and Kreutzberg, G.W. (1995) 'Microglia: intrinsic immune effector cell of the brain', *Brain Res Brain Res Rev*, 20(3), pp. 269-87.
- Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S.V., Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D., Diehn, M., West, R.B., Plevritis, S.K. and Alizadeh, A.A. (2015) 'The prognostic landscape of genes and infiltrating immune cells across human cancers', *Nat Med*, 21(8), pp. 938-945.
- Geyer, J.R., Sposto, R., Jennings, M., Boyett, J.M., Axtell, R.A., Breiger, D., Broxson, E., Donahue, B., Finlay, J.L., Goldwein, J.W., Heier, L.A., Johnson, D., Mazewski, C., Miller, D.C., Packer, R., Puccetti, D., Radcliffe, J., Tao, M.L. and Shiminski-Maher, T. (2005) 'Multiagent chemotherapy and deferred radiotherapy in infants with malignant brain tumors: a report from the Children's Cancer Group', *J Clin Oncol*, 23(30), pp. 7621-31.
- Giraud, M., Taubert, R., Vandiedonck, C., Ke, X., Levi-Strauss, M., Pagani, F., Baralle, F.E., Eymard, B., Tranchant, C., Gajdos, P., Vincent, A., Willcox, N., Beeson, D., Kyewski, B. and Garchon, H.J. (2007) 'An IRF8-binding promoter variant and AIRE control CHRNA1 promiscuous expression in thymus', *Nature*, 448(7156), pp. 934-7.
- Graeber, M.B., Scheithauer, B.W. and Kreutzberg, G.W. (2002) 'Microglia in brain tumors', *Glia*, 40(2), pp. 252-9.
- Grobner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V.A., Johann, P.D., Balasubramanian, G.P., Segura-Wang, M., Brabetz, S., Bender, S., Hutter, B., Sturm, D., Pfaff, E., Hubschmann, D., Zipprich, G., Heinold, M., Eils, J., Lawrenz, C., Erkek, S., Lambo, S., Waszak, S., Blattmann, C., Borkhardt, A., Kuhlen, M., Eggert, A., Fulda, S., Gessler, M., Wegert, J., Kappler, R., Baumhoer, D., Burdach, S., Kirschner-Schwabe, R., Kontny, U., Kulozik, A.E., Lohmann, D., Hettmer, S., Eckert, C., Bielack, S., Nathrath, M., Niemeyer, C., Richter, G.H., Schulte, J., Siebert, R., Westermann, F., Molenaar, J.J., Vassal, G., Witt, H., Burkhardt, B., Kratz, C.P., Witt, O., van Tilburg, C.M., Kramm,

- C.M., Fleischhack, G., Dirksen, U., Rutkowski, S., Fruhwald, M., von Hoff, K., Wolf, S., Klingebiel, T., Koscielniak, E., Landgraf, P., Koster, J., Resnick, A.C., Zhang, J., Liu, Y., Zhou, X., Waanders, A.J., Zwijnenburg, D.A., Raman, P., Brors, B., Weber, U.D., Northcott, P.A., Pajtler, K.W., Kool, M., Piro, R.M., Korbel, J.O., Schlesner, M., Eils, R., Jones, D.T.W., Lichter, P., Chavez, L., Zapatka, M. and Pfister, S.M. (2018) 'The landscape of genomic alterations across childhood cancers', *Nature*, 555(7696), pp. 321-327.
- Gruppenmacher, A.T., Halpern, A.L., Bonaldo Mde, F., Huang, C.C., Hamm, C.A., de Andrade, A., Tomita, T. and Sredni, S.T. (2013) 'Study of the gene expression and microRNA expression profiles of malignant rhabdoid tumors originated in the brain (AT/RT) and in the kidney (RTK)', *Childs Nerv Syst*, 29(11), pp. 1977-83.
- Guidi, C.J., Sands, A.T., Zambrowicz, B.P., Turner, T.K., Demers, D.A., Webster, W., Smith, T.W., Imbalzano, A.N. and Jones, S.N. (2001) 'Disruption of *Ini1* leads to peri-implantation lethality and tumorigenesis in mice', *Mol Cell Biol*, 21(10), pp. 3598-603.
- Gunderson, K.L., Kruglyak, S., Graige, M.S., Garcia, F., Kermani, B.G., Zhao, C., Che, D., Dickinson, T., Wickham, E., Bierle, J., Doucet, D., Milewski, M., Yang, R., Siegmund, C., Haas, J., Zhou, L., Oliphant, A., Fan, J.B., Barnard, S. and Chee, M.S. (2004) 'Decoding randomly ordered DNA arrays', *Genome Res*, 14(5), pp. 870-7.
- Haas, J.E., Palmer, N.F., Weinberg, A.G. and Beckwith, J.B. (1981) 'Ultrastructure of malignant rhabdoid tumor of the kidney. A distinctive renal tumor of children', *Hum Pathol*, 12(7), pp. 646-57.
- Haberler, C., Laggner, U., Slavic, I., Czech, T., Ambros, I.M., Ambros, P.F., Budka, H. and Hainfellner, J.A. (2006) 'Immunohistochemical analysis of INI1 protein in malignant pediatric CNS tumors: Lack of INI1 in atypical teratoid/rhabdoid tumors and in a fraction of primitive neuroectodermal tumors without rhabdoid phenotype', *Am J Surg Pathol*, 30(11), pp. 1462-8.
- Hadfield, K.D., Newman, W.G., Bowers, N.L., Wallace, A., Bolger, C., Colley, A., McCann, E., Trump, D., Prescott, T. and Evans, D.G. (2008) 'Molecular characterisation of SMARCB1 and NF2 in familial and sporadic schwannomatosis', *J Med Genet*, 45(6), pp. 332-9.
- Hambardzumyan, D., Gutmann, D.H. and Kettenmann, H. (2016) 'The role of microglia and macrophages in glioma maintenance and progression', *Nat Neurosci*, 19(1), pp. 20-7.
- Han, Z.Y., Richer, W., Freneaux, P., Chauvin, C., Lucchesi, C., Guillemot, D., Grison, C., Lequin, D., Pierron, G., Masliah-Planchon, J., Nicolas, A., Ranchere-Vince, D., Varlet, P., Puget, S., Janoueix-Lerosey, I., Ayrault, O., Surdez, D., Delattre, O. and Bourdeaut, F. (2016) 'The occurrence of intracranial rhabdoid tumours in mice depends on temporal control of *Smcarb1* inactivation', *Nat Commun*, 7, p. 10421.
- Hanahan, D. and Weinberg, R.A. (2000) 'The hallmarks of cancer', *Cell*, 100(1), pp. 57-70.
- Hanahan, D. and Weinberg, Robert A. (2011) 'Hallmarks of Cancer: The Next Generation', *Cell*, 144(5), pp. 646-674.
- Haskins, C.P., Jyoti, B., Hines, M., Simoneaux, V. and Buchsbaum, J.C. (2015) 'Single Center Results following Proton Beam Therapy in Children with Atypical Teratoid Rhabdoid Tumors of the Central Nervous System', *International Journal of Particle Therapy*, 2(1), pp. 1-10.
- Hasselblatt, M., Gesk, S., Oyen, F., Rossi, S., Viscardi, E., Giangaspero, F., Giannini, C., Judkins, A.R., Fruhwald, M.C., Obser, T., Schneppenheim, R., Siebert, R. and Paulus, W. (2011) 'Nonsense mutation and inactivation of SMARCA4 (BRG1) in an atypical teratoid/rhabdoid tumor showing retained SMARCB1 (INI1) expression', *Am J Surg Pathol*, 35(6), pp. 933-5.

- Hasselblatt, M., Nagel, I., Oyen, F., Bartelheim, K., Russell, R.B., Schuller, U., Junckerstorff, R., Rosenblum, M., Alassiri, A.H., Rossi, S., Schmid, I., Gottardo, N.G., Toledano, H., Viscardi, E., Balbin, M., Witkowski, L., Lu, Q., Betts, M.J., Foulkes, W.D., Siebert, R., Fruhwald, M.C. and Schneppenheim, R. (2014) 'SMARCA4-mutated atypical teratoid/rhabdoid tumors are associated with inherited germline alterations and poor prognosis', *Acta Neuropathol*, 128(3), pp. 453-6.
- Hayes, G.M., Woodroffe, M.N. and Cuzner, M.L. (1987) 'Microglia are the major cell type expressing MHC class II in human white matter', *J Neurol Sci*, 80(1), pp. 25-37.
- Herbst, R.S., Baas, P., Kim, D.W., Felip, E., Perez-Gracia, J.L., Han, J.Y., Molina, J., Kim, J.H., Arvis, C.D., Ahn, M.J., Majem, M., Fidler, M.J., de Castro, G., Jr., Garrido, M., Lubiniecki, G.M., Shentu, Y., Im, E., Dolled-Filhart, M. and Garon, E.B. (2016) 'Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial', *Lancet*, 387(10027), pp. 1540-1550.
- Herman, J.G., Latif, F., Weng, Y., Lerman, M.I., Zbar, B., Liu, S., Samid, D., Duan, D.S., Gnarr, J.R., Linehan, W.M. and et al. (1994) 'Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma', *Proc Natl Acad Sci U S A*, 91(21), pp. 9700-4.
- Hilden, J.M., Meerbaum, S., Burger, P., Finlay, J., Janss, A., Scheithauer, B.W., Walter, A.W., Rorke, L.B. and Biegel, J.A. (2004) 'Central Nervous System Atypical Teratoid/Rhabdoid Tumor: Results of Therapy in Children Enrolled in a Registry', *Journal of Clinical Oncology*, 22(14), pp. 2877-2884.
- Hinrichs, C.S. and Rosenberg, S.A. (2014) 'Exploiting the curative potential of adoptive T-cell therapy for cancer', *Immunol Rev*, 257(1), pp. 56-71.
- Hirata, E. and Sahai, E. (2017) 'Tumor Microenvironment and Differential Responses to Therapy', *Cold Spring Harb Perspect Med*, 7(7).
- Hodi, F.S., Mihm, M.C., Soiffer, R.J., Haluska, F.G., Butler, M., Seiden, M.V., Davis, T., Henry-Spires, R., MacRae, S., Willman, A., Padera, R., Jaklitsch, M.T., Shankar, S., Chen, T.C., Korman, A., Allison, J.P. and Dranoff, G. (2003) 'Biologic activity of cytotoxic T lymphocyte-associated antigen 4 antibody blockade in previously vaccinated metastatic melanoma and ovarian carcinoma patients', *Proc Natl Acad Sci U S A*, 100(8), pp. 4712-7.
- Hoel, P.G. (1943) 'On indices of dispersion', *The Annals of Mathematical Statistics*, 14(2), pp. 155-162.
- Hoot, A.C., Russo, P., Judkins, A.R., Perlman, E.J. and Biegel, J.A. (2004) 'Immunohistochemical analysis of hSNF5/INI1 distinguishes renal and extra-renal malignant rhabdoid tumors from other pediatric soft tissue tumors', *Am J Surg Pathol*, 28(11), pp. 1485-91.
- Hotelling, H. (1933) 'Analysis of a complex of statistical variables into principal components', *Journal of educational psychology*, 24(6), p. 417.
- Hulsebos, T.J., Plomp, A.S., Wolterman, R.A., Robanus-Maandag, E.C., Baas, F. and Wesseling, P. (2007) 'Germline mutation of INI1/SMARCB1 in familial schwannomatosis', *Am J Hum Genet*, 80(4), pp. 805-10.
- Hussain, S.F., Yang, D., Suki, D., Aldape, K., Grimm, E. and Heimberger, A.B. (2006) 'The role of human glioma-infiltrating microglia/macrophages in mediating antitumor immune responses', *Neuro-oncology*, 8(3), pp. 261-279.

- Hwang, B., Lee, J.H. and Bang, D. (2018) 'Single-cell RNA sequencing technologies and bioinformatics pipelines', *Experimental & Molecular Medicine*, 50(8), p. 96.
- Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. and Speed, T.P. (2003) 'Exploration, normalization, and summaries of high density oligonucleotide array probe level data', *Biostatistics*, 4(2), pp. 249-64.
- Isikay, I., Hanalioglu, S., Basar, I., Narin, F. and Bilginer, B. (2019) 'Survival Benefit with Gross Total Resection and Adjuvant Radiotherapy in Childhood Atypical Teratoid/Rhabdoid Tumors: Results of a Single-Center Cohort of 27 Cases', *Turk Neurosurg*.
- Iwai, Y., Ishida, M., Tanaka, Y., Okazaki, T., Honjo, T. and Minato, N. (2002) 'Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade', *Proc Natl Acad Sci U S A*, 99(19), pp. 12293-7.
- Jackson, E.M., Sievert, A.J., Gai, X., Hakonarson, H., Judkins, A.R., Tooke, L., Perin, J.C., Xie, H., Shaikh, T.H. and Biegel, J.A. (2009) 'Genomic analysis using high-density single nucleotide polymorphism-based oligonucleotide arrays and multiplex ligation-dependent probe amplification provides a comprehensive analysis of INI1/SMARCB1 in malignant rhabdoid tumors', *Clin Cancer Res*, 15(6), pp. 1923-30.
- Jagani, Z., Mora-Blanco, E.L., Sansam, C.G., McKenna, E.S., Wilson, B., Chen, D., Klekota, J., Tamayo, P., Nguyen, P.T., Tolstorukov, M., Park, P.J., Cho, Y.J., Hsiao, K., Buonamici, S., Pomeroy, S.L., Mesirov, J.P., Ruffner, H., Bouwmeester, T., Luchansky, S.J., Murtie, J., Kelleher, J.F., Warmuth, M., Sellers, W.R., Roberts, C.W. and Dorsch, M. (2010) 'Loss of the tumor suppressor Snf5 leads to aberrant activation of the Hedgehog-Gli pathway', *Nat Med*, 16(12), pp. 1429-33.
- Jeschke, J., Bizet, M., Desmedt, C., Calonne, E., Dedeurwaerder, S., Garaud, S., Koch, A., Larsimont, D., Salgado, R., Van den Eynden, G., Willard Gallo, K., Bontempi, G., Defrance, M., Sotiriou, C. and Fuks, F. (2017) 'DNA methylation-based immune response signature improves patient diagnosis in multiple cancers', *J Clin Invest*, 127(8), pp. 3090-3102.
- Jiang, D., Tang, C., Zhang, A.J.I.T.o.K. and Engineering, D. (2004) 'Cluster analysis for gene expression data: a survey', (11), pp. 1370-1386.
- Johann, P.D., Erkek, S., Zapatka, M., Kerl, K., Buchhalter, I., Hovestadt, V., Jones, D.T.W., Sturm, D., Hermann, C., Segura Wang, M., Korshunov, A., Rhyzova, M., Grobner, S., Brabetz, S., Chavez, L., Bens, S., Groschel, S., Kratochwil, F., Wittmann, A., Sieber, L., Georg, C., Wolf, S., Beck, K., Oyen, F., Capper, D., van Sluis, P., Volckmann, R., Koster, J., Versteeg, R., von Deimling, A., Milde, T., Witt, O., Kulozik, A.E., Ebinger, M., Shalaby, T., Grotzer, M., Sumerauer, D., Zamecnik, J., Mora, J., Jabado, N., Taylor, M.D., Huang, A., Aronica, E., Bertoni, A., Radlwimmer, B., Pietsch, T., Schuller, U., Schneppenheim, R., Northcott, P.A., Korbel, J.O., Siebert, R., Fruhwald, M.C., Lichter, P., Eils, R., Gajjar, A., Hasselblatt, M., Pfister, S.M. and Kool, M. (2016) 'Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups with Distinct Enhancer Landscapes', *Cancer Cell*, 29(3), pp. 379-393.
- Jones, P.A. and Baylin, S.B. (2002) 'The fundamental role of epigenetic events in cancer', *Nat Rev Genet*, 3(6), pp. 415-28.
- Kleihues, P. and Sobin, L.H. (2000) 'World Health Organization classification of tumors', *Cancer*, 88(12), pp. 2887-2887.
- Klochender-Yeivin, A., Fiette, L., Barra, J., Muchardt, C., Babinet, C. and Yaniv, M. (2000) 'The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression', *EMBO Rep*, 1(6), pp. 500-6.

- Komohara, Y., Ohnishi, K., Kuratsu, J. and Takeya, M. (2008) 'Possible involvement of the M2 anti-inflammatory macrophage phenotype in growth of human gliomas', *J Pathol*, 216(1), pp. 15-24.
- Kralik, S.F., Ho, C.Y., Finke, W., Buchsbaum, J.C., Haskins, C.P. and Shih, C.S. (2015) 'Radiation Necrosis in Pediatric Patients with Brain Tumors Treated with Proton Radiotherapy', *AJNR Am J Neuroradiol*, 36(8), pp. 1572-8.
- Kulis, M., Merkel, A., Heath, S., Queiros, A.C., Schuyler, R.P., Castellano, G., Beekman, R., Raineri, E., Esteve, A., Clot, G., Verdaguer-Dot, N., Duran-Ferrer, M., Russinol, N., Vilarrasa-Blasi, R., Ecker, S., Pancaldi, V., Rico, D., Agueda, L., Blanc, J., Richardson, D., Clarke, L., Datta, A., Pascual, M., Agirre, X., Prosper, F., Alignani, D., Paiva, B., Caron, G., Fest, T., Muench, M.O., Fomin, M.E., Lee, S.T., Wiemels, J.L., Valencia, A., Gut, M., Flicek, P., Stunnenberg, H.G., Siebert, R., Kuppers, R., Gut, I.G., Campo, E. and Martin-Subero, J.I. (2015) 'Whole-genome fingerprint of the DNA methylome during human B cell differentiation', *Nat Genet*, 47(7), pp. 746-56.
- Lafay-Cousin, L., Hawkins, C., Carret, A.S., Johnston, D., Zelcer, S., Wilson, B., Jabado, N., Scheinemann, K., Eisenstat, D., Fryer, C., Fleming, A., Mpofo, C., Larouche, V., Strother, D., Bouffet, E. and Huang, A. (2012) 'Central nervous system atypical teratoid rhabdoid tumours: The Canadian Paediatric Brain Tumour Consortium experience', *European Journal of Cancer*, 48(3), pp. 353-359.
- Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J.J., Cowey, C.L., Lao, C.D., Schadendorf, D., Dummer, R., Smylie, M., Rutkowski, P., Ferrucci, P.F., Hill, A., Wagstaff, J., Carlino, M.S., Haanen, J.B., Maio, M., Marquez-Rodas, I., McArthur, G.A., Ascierto, P.A., Long, G.V., Callahan, M.K., Postow, M.A., Grossmann, K., Sznol, M., Dreno, B., Bastholt, L., Yang, A., Rollin, L.M., Horak, C., Hodi, F.S. and Wolchok, J.D. (2015) 'Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma', *N Engl J Med*, 373(1), pp. 23-34.
- Lau, C.S., Mahendraraj, K. and Chamberlain, R.S. (2015) 'Atypical teratoid rhabdoid tumors: a population-based clinical outcomes study involving 174 patients from the Surveillance, Epidemiology, and End Results database (1973-2010)', *Cancer management and research*, 7, pp. 301-309.
- Lee, D.D. and Seung, H.S. (1999) 'Learning the parts of objects by non-negative matrix factorization', *Nature*, 401(6755), pp. 788-91.
- Lee, J., Kim, D.S., Han, J.W. and Suh, C.O. (2017) 'Atypical teratoid/rhabdoid tumors in children treated with multimodal therapies: The necessity of upfront radiotherapy after surgery', *Pediatr Blood Cancer*, 64(12).
- Lennerz, V., Fatho, M., Gentilini, C., Frye, R.A., Lifke, A., Ferel, D., Wolfel, C., Huber, C. and Wolfel, T. (2005) 'The response of autologous T cells to a human melanoma is dominated by mutated neoantigens', *Proc Natl Acad Sci U S A*, 102(44), pp. 16013-8.
- Li, J., Fu, C., Speed, T.P., Wang, W. and Symmans, W.F. (2018) 'Accurate RNA Sequencing From Formalin-Fixed Cancer Tissue To Represent High-Quality Transcriptome From Frozen Tissue', *JCO Precis Oncol*, 2018.
- Li, X., Liu, Y., Salz, T., Hansen, K.D. and Feinberg, A. (2016) 'Whole-genome analysis of the methylome and hydroxymethylome in normal and malignant lung and liver', *Genome Res*, 26(12), pp. 1730-1741.
- Maaten, L.v.d. and Hinton, G.J.J.o.m.l.r. (2008) 'Visualizing data using t-SNE', 9(Nov), pp. 2579-2605.

Mackay, A., Burford, A., Carvalho, D., Izquierdo, E., Fazal-Salom, J., Taylor, K.R., Bjerke, L., Clarke, M., Vinci, M., Nandhabalan, M., Temelso, S., Popov, S., Molinari, V., Raman, P., Waanders, A.J., Han, H.J., Gupta, S., Marshall, L., Zacharoulis, S., Vaidya, S., Mandeville, H.C., Bridges, L.R., Martin, A.J., Al-Sarraj, S., Chandler, C., Ng, H.K., Li, X., Mu, K., Trabelsi, S., Brahim, D.H., Kisljakov, A.N., Konovalov, D.M., Moore, A.S., Carcaboso, A.M., Sunol, M., de Torres, C., Cruz, O., Mora, J., Shats, L.I., Stavale, J.N., Bidinotto, L.T., Reis, R.M., Entz-Werle, N., Farrell, M., Cryan, J., Crimmins, D., Caird, J., Pears, J., Monje, M., Debily, M.A., Castel, D., Grill, J., Hawkins, C., Nikbakht, H., Jabado, N., Baker, S.J., Pfister, S.M., Jones, D.T.W., Fouladi, M., von Bueren, A.O., Baudis, M., Resnick, A. and Jones, C. (2017) 'Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma', *Cancer Cell*, 32(4), pp. 520-537.e5.

Mackay, A., Burford, A., Molinari, V., Jones, D.T.W., Izquierdo, E., Brouwer-Visser, J., Giangaspero, F., Haberler, C., Pietsch, T., Jacques, T.S., Figarella-Branger, D., Rodriguez, D., Morgan, P.S., Raman, P., Waanders, A.J., Resnick, A.C., Massimino, M., Garre, M.L., Smith, H., Capper, D., Pfister, S.M., Wurdinger, T., Tam, R., Garcia, J., Thakur, M.D., Vassal, G., Grill, J., Jaspan, T., Varlet, P. and Jones, C. (2018) 'Molecular, Pathological, Radiological, and Immune Profiling of Non-brainstem Pediatric High-Grade Glioma from the HERBY Phase II Randomized Trial', *Cancer Cell*, 33(5), pp. 829-842.e5.

McGovern, S.L., Okcu, M.F., Munsell, M.F., Kumbalasseriyl, N., Grosshans, D.R., McAleer, M.F., Chintagumpala, M., Khatua, S. and Mahajan, A. (2014) 'Outcomes and acute toxicities of proton therapy for pediatric atypical teratoid/rhabdoid tumor of the central nervous system', *International journal of radiation oncology, biology, physics*, 90(5), pp. 1143-1152.

McInnes, L., Healy, J. and Melville, J.J.a.p.a. (2018) 'Umap: Uniform manifold approximation and projection for dimension reduction'.

McKenna, E.S., Sansam, C.G., Cho, Y.J., Greulich, H., Evans, J.A., Thom, C.S., Moreau, L.A., Biegel, J.A., Pomeroy, S.L. and Roberts, C.W. (2008) 'Loss of the epigenetic tumor suppressor SNF5 leads to cancer without genomic instability', *Mol Cell Biol*, 28(20), pp. 6223-33.

Mlecnik, B., Bindea, G., Angell, H.K., Maby, P., Angelova, M., Tougeron, D., Church, S.E., Lafontaine, L., Fischer, M., Fredriksen, T., Sasso, M., Bilocq, A.M., Kirilovsky, A., Obenauf, A.C., Hamieh, M., Berger, A., Bruneval, P., Tuech, J.J., Sabourin, J.C., Le Pessot, F., Mauillon, J., Rafii, A., Laurent-Puig, P., Speicher, M.R., Trajanoski, Z., Michel, P., Sesboue, R., Frebourg, T., Pages, F., Valge-Archer, V., Latouche, J.B. and Galon, J. (2016) 'Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability', *Immunity*, 44(3), pp. 698-711.

Modena, P., Sardi, I., Brenca, M., Giunti, L., Buccoliero, A.M., Pollo, B., Biassoni, V., Genitori, L., Antonelli, M., Maestro, R., Giangaspero, F. and Massimino, M. (2013) 'Case report: long-term survival of an infant syndromic patient affected by atypical teratoid-rhabdoid tumor', *BMC Cancer*, 13(1), p. 100.

Mora-Blanco, E.L., Mishina, Y., Tillman, E.J., Cho, Y.J., Thom, C.S., Pomeroy, S.L., Shao, W. and Roberts, C.W. (2014) 'Activation of beta-catenin/TCF targets following loss of the tumor suppressor SNF5', *Oncogene*, 33(7), pp. 933-8.

Mueller, W., Eum, J.H., Lass, U., Paulus, W., Sarkar, C., Bruck, W. and von Deimling, A. (2004) 'No evidence of hSNF5/INI1 point mutations in choroid plexus papilloma', *Neuropathol Appl Neurobiol*, 30(3), pp. 304-7.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F. and Oliviero, S. (2017) 'Intragenic DNA methylation prevents spurious transcription initiation', *Nature*, 543(7643), pp. 72-77.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M. and Alizadeh, A.A. (2015) 'Robust enumeration of cell subsets from tissue expression profiles', *Nat Methods*, 12(5), pp. 453-7.

Nicolaides, T., Tihan, T., Horn, B., Biegel, J., Prados, M. and Banerjee, A. (2010) 'High-dose chemotherapy and autologous stem cell rescue for atypical teratoid/rhabdoid tumor of the central nervous system', *J Neurooncol*, 98(1), pp. 117-23.

Northcott, P.A., Buchhalter, I., Morrissy, A.S., Hovestadt, V., Weischenfeldt, J., Ehrenberger, T., Grobner, S., Segura-Wang, M., Zichner, T., Rudneva, V.A., Warnatz, H.J., Sidiropoulos, N., Phillips, A.H., Schumacher, S., Kleinheinz, K., Waszak, S.M., Erkek, S., Jones, D.T.W., Worst, B.C., Kool, M., Zapatka, M., Jager, N., Chavez, L., Hutter, B., Bieg, M., Paramasivam, N., Heinold, M., Gu, Z., Ishaque, N., Jager-Schmidt, C., Imbusch, C.D., Jugold, A., Hubschmann, D., Risch, T., Amstislavskiy, V., Gonzalez, F.G.R., Weber, U.D., Wolf, S., Robinson, G.W., Zhou, X., Wu, G., Finkelstein, D., Liu, Y., Cavalli, F.M.G., Luu, B., Ramaswamy, V., Wu, X., Koster, J., Ryzhova, M., Cho, Y.J., Pomeroy, S.L., Herold-Mende, C., Schuhmann, M., Ebinger, M., Liau, L.M., Mora, J., McLendon, R.E., Jabado, N., Kumabe, T., Chuah, E., Ma, Y., Moore, R.A., Mungall, A.J., Mungall, K.L., Thiessen, N., Tse, K., Wong, T., Jones, S.J.M., Witt, O., Milde, T., Von Deimling, A., Capper, D., Korshunov, A., Yaspo, M.L., Kriwacki, R., Gajjar, A., Zhang, J., Beroukhim, R., Fraenkel, E., Korbel, J.O., Brors, B., Schlesner, M., Eils, R., Marra, M.A., Pfister, S.M., Taylor, M.D. and Lichter, P. (2017) 'The whole-genome landscape of medulloblastoma subtypes', *Nature*, 547(7663), pp. 311-317.

Oka, H. and Scheithauer, B.W. (1999) 'Clinicopathological Characteristics of Atypical Teratoid/Rhabdoid Tumor', *Neurologia medico-chirurgica*, 39(7), pp. 510-518.

Ostrom, Q.T., Chen, Y., M de Blank, P., Ondracek, A., Farah, P., Gittleman, H., Wolinsky, Y., Kruchko, C., Cohen, M.L., Brat, D.J. and Barnholtz-Sloan, J.S. (2014) 'The descriptive epidemiology of atypical teratoid/rhabdoid tumors in the United States, 2001-2010', *Neuro-oncology*, 16(10), pp. 1392-1399.

Paatero, P. and Tapper, U. (1994) 'Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values', 5(2), pp. 111-126.

Park, E.S., Sung, K.W., Baek, H.J., Park, K.D., Park, H.J., Won, S.C., Lim, D.H. and Kim, H.S. (2012) 'Tandem high-dose chemotherapy and autologous stem cell transplantation in young children with atypical teratoid/rhabdoid tumor of the central nervous system', *J Korean Med Sci*, 27(2), pp. 135-40.

Pearson, K. (1901) 'LIII. On lines and planes of closest fit to systems of points in space', *The London, Edinburgh, Dublin Philosophical Magazine and Journal of Science*, 2(11), pp. 559-572.

Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P. and Fodor, S.P. (1994) 'Light-generated oligonucleotide arrays for rapid DNA sequence analysis', *Proc Natl Acad Sci U S A*, 91(11), pp. 5022-6.

Perry, A., Scheithauer, B.W., Stafford, S.L., Abell-Aleff, P.C. and Meyer, F.B. (1998) "'Rhabdoid" meningioma: an aggressive variant', *Am J Surg Pathol*, 22(12), pp. 1482-90.

Pidsley, R., Zotenko, E., Peters, T.J., Lawrence, M.G., Risbridger, G.P., Molloy, P., Van Dijk, S., Muhlhäusler, B., Stirzaker, C. and Clark, S.J. (2016) 'Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling', *Genome Biology*, 17(1), p. 208.

Pinto, E.M., Hamideh, D., Bahrami, A., Orr, B.A., Lin, T., Pounds, S., Zambetti, G.P., Pappo, A.S., Gajjar, A., Agnihotri, S. and Broniscer, A. (2018) 'Malignant rhabdoid tumors originating within and outside

- the central nervous system are clinically and molecularly heterogeneous', *Acta Neuropathol*, 136(2), pp. 315-326.
- Prins, R.M., Soto, H., Konkankit, V., Odesa, S.K., Eskin, A., Yong, W.H., Nelson, S.F. and Liau, L.M. (2011) 'Gene expression profile correlates with T-cell infiltration and relative survival in glioblastoma patients vaccinated with dendritic cell immunotherapy', *Clin Cancer Res*, 17(6), pp. 1603-15.
- Proust, F., Laquerriere, A., Constantin, B., Ruchoux, M.M., Vannier, J.P. and Freger, P. (1999) 'Simultaneous presentation of atypical teratoid/rhabdoid tumor in siblings', *J Neurooncol*, 43(1), pp. 63-70.
- Quail, D.F. and Joyce, J.A. (2013) 'Microenvironmental regulation of tumor progression and metastasis', *Nat Med*, 19(11), pp. 1423-37.
- Quail, D.F. and Joyce, J.A. (2017) 'The Microenvironmental Landscape of Brain Tumors', *Cancer Cell*, 31(3), pp. 326-341.
- Reardon, D.A., Kaley, T.J., Dietrich, J., Clarke, J.L., Dunn, G.P., Lim, M., Cloughesy, T.F., Gan, H.K., Park, A.J., Schwarzenberger, P., Ricciardi, T., Macri, M.J., Ryan, A. and Venhaus, R.R. (2017a) 'Phase 2 study to evaluate safety and efficacy of MEDI4736 (durvalumab [DUR]) in glioblastoma (GBM) patients: An update', *Journal of Clinical Oncology*, 35(15_suppl), pp. 2042-2042.
- Reardon, D.A., Omuro, A., Brandes, A.A., Rieger, J., Wick, A., Sepulveda, J., Phuphanich, S., de Souza, P., Ahluwalia, M.S., Lim, M., Vlahovic, G. and Sampson, J. (2017b) 'OS10.3 Randomized Phase 3 Study Evaluating the Efficacy and Safety of Nivolumab vs Bevacizumab in Patients With Recurrent Glioblastoma: CheckMate 143', *Neuro-Oncology*, 19(suppl_3), pp. iii21-iii21.
- Reddy, A., Strother, D., Judkins, A., Krailo, M., Gao, Y., Douglas, J., Mahajan, A., Lewis, V., Mazewski, C., Laningham, F., Pollack, I., Gajjar, A. and Biegel, J. (2016) 'AT-09TREATMENT OF ATYPICAL TERATOID RHABDOID TUMORS (ATRT) OF THE CENTRAL NERVOUS SYSTEM WITH SURGERY, INTENSIVE CHEMOTHERAPY, AND 3-D CONFORMAL RADIATION (ACNS0333). A REPORT FROM THE CHILDREN'S ONCOLOGY GROUP', *Neuro-Oncology*, 18(suppl_3), pp. iii2-iii2.
- Reinhard, H., Reinert, J., Beier, R., Furtwangler, R., Alkasser, M., Rutkowski, S., Fruhwald, M., Koscielniak, E., Leuschner, I., Kaatsch, P. and Graf, N. (2008) 'Rhabdoid tumors in children: prognostic factors in 70 patients diagnosed in Germany', *Oncol Rep*, 19(3), pp. 819-23.
- Rizvi, N.A., Mazieres, J., Planchard, D., Stinchcombe, T.E., Dy, G.K., Antonia, S.J., Horn, L., Lena, H., Minenza, E., Mennezier, B., Otterson, G.A., Campos, L.T., Gandara, D.R., Levy, B.P., Nair, S.G., Zalcman, G., Wolf, J., Souquet, P.J., Baldini, E., Cappuzzo, F., Chouaid, C., Dowlati, A., Sanborn, R., Lopez-Chavez, A., Grohe, C., Huber, R.M., Harbison, C.T., Baudelet, C., Lestini, B.J. and Ramalingam, S.S. (2015) 'Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial', *Lancet Oncol*, 16(3), pp. 257-65.
- Robert, C., Long, G.V., Brady, B., Dutriaux, C., Maio, M., Mortier, L., Hassel, J.C., Rutkowski, P., McNeil, C., Kalinka-Warzocho, E., Savage, K.J., Hernberg, M.M., Lebbe, C., Charles, J., Mihalciou, C., Chiarion-Sileni, V., Mauch, C., Cognetti, F., Arance, A., Schmidt, H., Schadendorf, D., Gogas, H., Lundgren-Eriksson, L., Horak, C., Sharkey, B., Waxman, I.M., Atkinson, V. and Ascierto, P.A. (2015a) 'Nivolumab in previously untreated melanoma without BRAF mutation', *N Engl J Med*, 372(4), pp. 320-30.
- Robert, C., Schachter, J., Long, G.V., Arance, A., Grob, J.J., Mortier, L., Daud, A., Carlino, M.S., McNeil, C., Lotem, M., Larkin, J., Lorigan, P., Neyns, B., Blank, C.U., Hamid, O., Mateus, C., Shapira-Frommer,

- R., Kosh, M., Zhou, H., Ibrahim, N., Ebbinghaus, S. and Ribas, A. (2015b) 'Pembrolizumab versus ipilimumab in Advanced Melanoma', *N Engl J Med*, 372(26), pp. 2521-32.
- Rorke, L.B., Packer, R.J. and Biegel, J.A. (1996) 'Central nervous system atypical teratoid/rhabdoid tumors of infancy and childhood: definition of an entity', *J Neurosurg*, 85(1), pp. 56-65.
- Rossi, M.L., Hughes, J.T., Esiri, M.M., Coakham, H.B. and Brownell, D.B. (1987) 'Immunohistological study of mononuclear cell infiltrate in malignant gliomas', *Acta Neuropathol*, 74(3), pp. 269-77.
- Rousseeuw, P. (1987) 'Silhouettes: a graphical aid to the interpretation and validation of cluster analysis', *Journal of computational applied mathematics*, 20, pp. 53-65.
- Rutledge, W.C., Kong, J., Gao, J., Gutman, D.A., Cooper, L.A., Appin, C., Park, Y., Scarpace, L., Mikkelsen, T., Cohen, M.L., Aldape, K.D., McLendon, R.E., Lehman, N.L., Miller, C.R., Schniederjan, M.J., Brennan, C.W., Saltz, J.H., Moreno, C.S. and Brat, D.J. (2013) 'Tumor-infiltrating lymphocytes in glioblastoma are associated with specific genomic alterations and related to transcriptional class', *Clin Cancer Res*, 19(18), pp. 4951-60.
- Schneppenheim, R., Fruhwald, M.C., Gesk, S., Hasselblatt, M., Jeibmann, A., Kordes, U., Kreuz, M., Leuschner, I., Martin Subero, J.I., Obser, T., Oyen, F., Vater, I. and Siebert, R. (2010) 'Germline nonsense mutation and somatic inactivation of SMARCA4/BRG1 in a family with rhabdoid tumor predisposition syndrome', *Am J Hum Genet*, 86(2), pp. 279-84.
- Schubeler, D. (2015) 'Function and information content of DNA methylation', *Nature*, 517(7534), pp. 321-6.
- Schwalbe, E.C., Lindsey, J.C., Nakjang, S., Crosier, S., Smith, A.J., Hicks, D., Rafiee, G., Hill, R.M., Iliasova, A., Stone, T., Pizer, B., Michalski, A., Joshi, A., Wharton, S.B., Jacques, T.S., Bailey, S., Williamson, D. and Clifford, S.C. (2017a) 'Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: a cohort study', *Lancet Oncol*, 18(7), pp. 958-971.
- Schwalbe, E.C., Lindsey, J.C., Nakjang, S., Crosier, S., Smith, A.J., Hicks, D., Rafiee, G., Hill, R.M., Iliasova, A., Stone, T., Pizer, B., Michalski, A., Joshi, A., Wharton, S.B., Jacques, T.S., Bailey, S., Williamson, D. and Clifford, S.C. (2017b) 'Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: a cohort study', *The Lancet Oncology*, 18(7), pp. 958-971.
- Sevenet, N., Lellouch-Tubiana, A., Schofield, D., Hoang-Xuan, K., Gessler, M., Birnbaum, D., Jeanpierre, C., Jouvett, A. and Delattre, O. (1999a) 'Spectrum of hSNF5/INI1 somatic mutations in human cancer and genotype-phenotype correlations', *Hum Mol Genet*, 8(13), pp. 2359-68.
- Sevenet, N., Sheridan, E., Amram, D., Schneider, P., Handgretinger, R. and Delattre, O. (1999b) 'Constitutional mutations of the hSNF5/INI1 gene predispose to a variety of cancers', *Am J Hum Genet*, 65(5), pp. 1342-8.
- Sharma, T., Schwalbe, E.C., Williamson, D., Sill, M., Hovestadt, V., Mynarek, M., Rutkowski, S., Robinson, G.W., Gajjar, A., Cavalli, F., Ramaswamy, V., Taylor, M.D., Lindsey, J.C., Hill, R.M., Jager, N., Korshunov, A., Hicks, D., Bailey, S., Kool, M., Chavez, L., Northcott, P.A., Pfister, S.M. and Clifford, S.C. (2019a) 'Second-generation molecular subgrouping of medulloblastoma: an international meta-analysis of Group 3 and Group 4 subtypes', *Acta Neuropathol*, 138(2), pp. 309-326.
- Sharma, T., Schwalbe, E.C., Williamson, D., Sill, M., Hovestadt, V., Mynarek, M., Rutkowski, S., Robinson, G.W., Gajjar, A., Cavalli, F., Ramaswamy, V., Taylor, M.D., Lindsey, J.C., Hill, R.M., Jäger, N., Korshunov, A., Hicks, D., Bailey, S., Kool, M., Chavez, L., Northcott, P.A., Pfister, S.M. and Clifford, S.C.

(2019b) 'Second-generation molecular subgrouping of medulloblastoma: an international meta-analysis of Group 3 and Group 4 subtypes', *Acta Neuropathologica*, 138(2), pp. 309-326.

Shen, X. and Zhao, B. (2018) 'Efficacy of PD-1 or PD-L1 inhibitors and PD-L1 expression status in cancer: meta-analysis', *BMJ*, 362, p. k3529.

Shih, C.S., Hale, G.A., Gronewold, L., Tong, X., Laningham, F.H., Gilger, E.A., Srivastava, D.K., Kun, L.E., Gajjar, A. and Fouladi, M. (2008) 'High-dose chemotherapy with autologous stem cell rescue for children with recurrent malignant brain tumors', *Cancer*, 112(6), pp. 1345-53.

Slavc, I., Chocholous, M., Leiss, U., Haberler, C., Peyrl, A., Azizi, A.A., Dieckmann, K., Woehrer, A., Peters, C., Widhalm, G., Dorfer, C. and Czech, T. (2014) 'Atypical teratoid rhabdoid tumor: improved long-term survival with an intensive multimodal therapy and delayed radiotherapy. The Medical University of Vienna Experience 1992-2012', *Cancer Med*, 3(1), pp. 91-100.

Sokal, R.R. and Rohlf, F.J. (1962) 'The comparison of dendrograms by objective methods', *Taxon*, 11(2), pp. 33-40.

Streilein, J.W. (1993) 'Immune privilege as the result of local tissue barriers and immunosuppressive microenvironments', *Curr Opin Immunol*, 5(3), pp. 428-32.

Sullivan, L.M., Folpe, A.L., Pawel, B.R., Judkins, A.R. and Biegel, J.A. (2013) 'Epithelioid sarcoma is associated with a high percentage of SMARCB1 deletions', *Mod Pathol*, 26(3), pp. 385-92.

Sultan, I., Casanova, M., Al-Jumaily, U., Meazza, C., Rodriguez-Galindo, C. and Ferrari, A. (2010a) 'Soft tissue sarcomas in the first year of life', *European Journal of Cancer*, 46(13), pp. 2449-2456.

Sultan, I., Qaddoumi, I., Rodriguez-Galindo, C., Nassan, A.A., Ghandour, K. and Al-Hussaini, M. (2010b) 'Age, stage, and radiotherapy, but not primary tumor site, affects the outcome of patients with malignant rhabdoid tumors', *Pediatr Blood Cancer*, 54(1), pp. 35-40.

Szymanski, K.M., Tabib, C.H., Idrees, M.T. and Cain, M.P. (2013) 'Synchronous perivesical and renal malignant rhabdoid tumor in a 9-year-old boy: a case report and review of literature', *Urology*, 82(5), pp. 1158-60.

Tamayo, P., Scanfeld, D., Ebert, B.L., Gillette, M.A., Roberts, C.W.M. and Mesirov, J.P. (2007) 'Metagene projection for cross-platform, cross-species characterization of global transcriptional states', 104(14), pp. 5959-5964.

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S.C., Kok, C.Y., Noble, K., Ponting, L., Ramshaw, C.C., Rye, C.E., Speedy, H.E., Stefancsik, R., Thompson, S.L., Wang, S., Ward, S., Campbell, P.J. and Forbes, S.A. (2018) 'COSMIC: the Catalogue Of Somatic Mutations In Cancer', *Nucleic Acids Research*, 47(D1), pp. D941-D947.

Tekautz, T.M., Fuller, C.E., Blaney, S., Fouladi, M., Broniscer, A., Merchant, T.E., Krasin, M., Dalton, J., Hale, G., Kun, L.E., Wallace, D., Gilbertson, R.J. and Gajjar, A. (2005) 'Atypical Teratoid/Rhabdoid Tumors (ATRT): Improved Survival in Children 3 Years of Age and Older With Radiation Therapy and High-Dose Alkylator-Based Chemotherapy', *Journal of Clinical Oncology*, 23(7), pp. 1491-1499.

Teschendorff, A.E. and Zheng, S.C. (2017) 'Cell-type deconvolution in epigenome-wide association studies: a review and recommendations', *Epigenomics*, 9(5), pp. 757-768.

- Tian, T., Olson, S., Whitacre, J.M. and Harding, A. (2011) 'The origins of cancer robustness and evolvability', *Integr Biol (Camb)*, 3(1), pp. 17-30.
- Tomita, Y., Fukasawa, S., Shinohara, N., Kitamura, H., Oya, M., Eto, M., Tanabe, K., Saito, M., Kimura, G., Yonese, J., Yao, M. and Uemura, H. (2019) 'Nivolumab versus everolimus in advanced renal cell carcinoma: Japanese subgroup 3-year follow-up analysis from the Phase III CheckMate 025 study', *Jpn J Clin Oncol*, 49(6), pp. 506-514.
- Tomlinson, G.E., Breslow, N.E., Dome, J., Guthrie, K.A., Norkool, P., Li, S., Thomas, P.R., Perlman, E., Beckwith, J.B., D'Angio, G.J. and Green, D.M. (2005) 'Rhabdoid tumor of the kidney in the National Wilms' Tumor Study: age at diagnosis as a prognostic factor', *J Clin Oncol*, 23(30), pp. 7641-5.
- Topalian, S.L., Drake, C.G. and Pardoll, D.M. (2015) 'Immune checkpoint blockade: a common denominator approach to cancer therapy', *Cancer Cell*, 27(4), pp. 450-61.
- Torchia, J., Golbourn, B., Feng, S., Ho, K.C., Sin-Chan, P., Vasiljevic, A., Norman, J.D., Guilhamon, P., Garzia, L., Agamez, N.R., Lu, M., Chan, T.S., Picard, D., de Antonellis, P., Khuong-Quang, D.-A., Planello, A.C., Zeller, C., Barsyte-Lovejoy, D., Lafay-Cousin, L., Letourneau, L., Bourgey, M., Yu, M., Gendoo, D.M.A., Dzamba, M., Barszczyk, M., Medina, T., Riemenschneider, A.N., Morrissy, A.S., Ra, Y.-S., Ramaswamy, V., Remke, M., Dunham, C.P., Yip, S., Ng, H.-K., Lu, J.-Q., Mehta, V., Albrecht, S., Pimentel, J., Chan, J.A., Somers, G.R., Faria, C.C., Roque, L., Fouladi, M., Hoffman, L.M., Moore, A.S., Wang, Y., Choi, S.A., Hansford, J.R., Catchpoole, D., Birks, D.K., Foreman, N.K., Strother, D., Klekner, A., Bognár, L., Garami, M., Hauser, P., Hortobágyi, T., Wilson, B., Hukin, J., Carret, A.-S., Van Meter, T.E., Hwang, E.I., Gajjar, A., Chiou, S.-H., Nakamura, H., Toledano, H., Fried, I., Fults, D., Wataya, T., Fryer, C., Eisenstat, D.D., Scheinmann, K., Fleming, A.J., Johnston, D.L., Michaud, J., Zelcer, S., Hammond, R., Afzal, S., Ramsay, D.A., Sirachainan, N., Hongeng, S., Larbcharoensub, N., Grundy, R.G., Lulla, R.R., Fangusaro, J.R., Druker, H., Bartels, U., Grant, R., Malkin, D., McGlade, C.J., Nicolaidis, T., Tihan, T., Phillips, J., Majewski, J., Montpetit, A., Bourque, G., Bader, G.D., Reddy, A.T., Gillespie, G.Y., Warmuth-Metz, M., et al. (2016) 'Integrated (epi)-Genomic Analyses Identify Subgroup-Specific Therapeutic Targets in CNS Rhabdoid Tumors', *Cancer cell*, 30(6), pp. 891-908.
- Torchia, J., Picard, D., Lafay-Cousin, L., Hawkins, C.E., Kim, S.K., Letourneau, L., Ra, Y.S., Ho, K.C., Chan, T.S., Sin-Chan, P., Dunham, C.P., Yip, S., Ng, H.K., Lu, J.Q., Albrecht, S., Pimentel, J., Chan, J.A., Somers, G.R., Zielenska, M., Faria, C.C., Roque, L., Baskin, B., Birks, D., Foreman, N., Strother, D., Klekner, A., Garami, M., Hauser, P., Hortobágyi, T., Bognar, L., Wilson, B., Hukin, J., Carret, A.S., Van Meter, T.E., Nakamura, H., Toledano, H., Fried, I., Fults, D., Wataya, T., Fryer, C., Eisenstat, D.D., Scheineman, K., Johnston, D., Michaud, J., Zelcer, S., Hammond, R., Ramsay, D.A., Fleming, A.J., Lulla, R.R., Fangusaro, J.R., Sirachainan, N., Larbcharoensub, N., Hongeng, S., Barakzai, M.A., Montpetit, A., Stephens, D., Grundy, R.G., Schuller, U., Nicolaidis, T., Tihan, T., Phillips, J., Taylor, M.D., Rutka, J.T., Dirks, P., Bader, G.D., Warmuth-Metz, M., Rutkowski, S., Pietsch, T., Judkins, A.R., Jabado, N., Bouffet, E. and Huang, A. (2015) 'Molecular subgroups of atypical teratoid rhabdoid tumours in children: an integrated genomic and clinicopathological analysis', *Lancet Oncol*, 16(5), pp. 569-82.
- Triche, T.J., Jr., Weisenberger, D.J., Van Den Berg, D., Laird, P.W. and Siegmund, K.D. (2013) 'Low-level processing of Illumina Infinium DNA Methylation BeadArrays', *Nucleic Acids Res*, 41(7), p. e90.
- Tsuneyoshi, M., Daimaru, Y., Hashimoto, H. and Enjoji, M. (1987) 'The existence of rhabdoid cells in specified soft tissue sarcomas. Histopathological, ultrastructural and immunohistochemical evidence', *Virchows Arch A Pathol Anat Histopathol*, 411(6), pp. 509-14.
- Tumeh, P.C., Harview, C.L., Yearley, J.H., Shintaku, I.P., Taylor, E.J., Robert, L., Chmielowski, B., Spasic, M., Henry, G., Ciobanu, V., West, A.N., Carmona, M., Kivork, C., Seja, E., Cherry, G., Gutierrez, A.J., Grogan, T.R., Mateus, C., Tomicic, G., Glaspy, J.A., Emerson, R.O., Robins, H., Pierce, R.H., Elashoff,

- D.A., Robert, C. and Ribas, A. (2014) 'PD-1 blockade induces responses by inhibiting adaptive immune resistance', *Nature*, 515(7528), pp. 568-71.
- Ueyama, T., Nagai, E., Yao, T. and Tsuneyoshi, M. (1993) 'Vimentin-positive gastric carcinomas with rhabdoid features. A clinicopathologic and immunohistochemical study', *Am J Surg Pathol*, 17(8), pp. 813-9.
- Vermeulen, J.F., Van Hecke, W., Adriaansen, E.J.M., Jansen, M.K., Bouma, R.G., Villacorta Hidalgo, J., Fisch, P., Broekhuizen, R., Spliet, W.G.M., Kool, M. and Bovenschen, N. (2018) 'Prognostic relevance of tumor-infiltrating lymphocytes and immune checkpoints in pediatric medulloblastoma', *Oncoimmunology*, 7(3), p. e1398877.
- Versteeg, I., Sevenet, N., Lange, J., Rousseau-Merck, M.F., Ambros, P., Handgretinger, R., Aurias, A. and Delattre, O. (1998) 'Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer', *Nature*, 394(6689), pp. 203-6.
- Voena, C. and Chiarle, R. (2016) 'Advances in cancer immunology and cancer immunotherapy', *Discov Med*, 21(114), pp. 125-33.
- Vogel, J., Carmona, R., Ainsley, C.G. and Lustig, R.A. (2018) 'The Promise of Proton Therapy for Central Nervous System Malignancies', *Neurosurgery*, 84(5), pp. 1000-1010.
- Vujanic, G.M., Sandstedt, B., Harms, D., Boccon-Gibod, L. and Delemarre, J.F. (1996) 'Rhabdoid tumour of the kidney: a clinicopathological study of 22 patients from the International Society of Paediatric Oncology (SIOP) nephroblastoma file', *Histopathology*, 28(4), pp. 333-40.
- Wagner, L., Hill, D.A., Fuller, C., Pedrosa, M., Bhakta, M., Perry, A. and Dome, J.S. (2002) 'Treatment of metastatic rhabdoid tumor of the kidney', *J Pediatr Hematol Oncol*, 24(5), pp. 385-8.
- Waldron, P.E., Rodgers, B.M., Kelly, M.D. and Womer, R.B. (1999) 'Successful treatment of a patient with stage IV rhabdoid tumor of the kidney: case report and review', *J Pediatr Hematol Oncol*, 21(1), pp. 53-7.
- Warmuth-Metz, M., Bison, B., Dannemann-Stern, E., Kortmann, R., Rutkowski, S. and Pietsch, T. (2008) 'CT and MR imaging in atypical teratoid/rhabdoid tumors of the central nervous system', *Neuroradiology*, 50(5), pp. 447-52.
- Weber, J.S., D'Angelo, S.P., Minor, D., Hodi, F.S., Gutzmer, R., Neyns, B., Hoeller, C., Khushalani, N.I., Miller, W.H., Jr., Lao, C.D., Linette, G.P., Thomas, L., Lorigan, P., Grossmann, K.F., Hassel, J.C., Maio, M., Sznol, M., Ascierto, P.A., Mohr, P., Chmielowski, B., Bryce, A., Svane, I.M., Grob, J.J., Krackhardt, A.M., Horak, C., Lambert, A., Yang, A.S. and Larkin, J. (2015) 'Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial', *Lancet Oncol*, 16(4), pp. 375-84.
- Wick, M.R., Ritter, J.H. and Dehner, L.P. (1995) 'Malignant rhabdoid tumors: a clinicopathologic review and conceptual discussion', *Semin Diagn Pathol*, 12(3), pp. 233-48.
- Wilson, B.G., Wang, X., Shen, X., McKenna, E.S., Lemieux, M.E., Cho, Y.J., Koellhoffer, E.C., Pomeroy, S.L., Orkin, S.H. and Roberts, C.W. (2010) 'Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation', *Cancer Cell*, 18(4), pp. 316-28.

Woehrer, A., Slavic, I., Waldhoer, T., Heinzl, H., Zielonke, N., Czech, T., Benesch, M., Hainfellner, J.A. and Haberler, C. (2010) 'Incidence of atypical teratoid/rhabdoid tumors in children: a population-based study by the Austrian Brain Tumor Registry, 1996-2006', *Cancer*, 116(24), pp. 5725-32.

Wu, F.X. (2008) 'Genetic weighted k-means algorithm for clustering large-scale gene expression data', *BMC Bioinformatics*, 9 Suppl 6, p. S12.

Xiao, X., Cai, F., Niu, X., Shi, H. and Zhong, Y. (2016) 'Association between P16INK4a Promoter Methylation and Ovarian Cancer: A Meta-Analysis of 12 Published Studies', *PLoS One*, 11(9), p. e0163257.

Yang, I., Tihan, T., Han, S.J., Wrensch, M.R., Wiencke, J., Sughrue, M.E. and Parsa, A.T. (2010) 'CD8+ T-cell infiltrate in newly diagnosed glioblastoma is associated with long-term survival', *J Clin Neurosci*, 17(11), pp. 1381-5.

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A. and Liang, G. (2014) 'Gene body methylation can alter gene expression and is a therapeutic target in cancer', *Cancer Cell*, 26(4), pp. 577-90.

8 Appendix

8.1 A) Primary tumour cohort used as part of this study

Sample Name	Group	Subgroup NMF (k = 3)	Subgroup NMF (k = 5)	MNP2.0 Call	MNP2.0 Calib. Score	Sex	Age (months)	Location	PFS	OS	Metastasis
INSTINCT ATRT 1	ATRT	SHH	SHH.Inf	ATRT, SHH	0.962	M	4.3	Cerebral hemispheres; Posterior of Thalamus; 3rd Ventricle	Y	Y	FALSE
NMB 1040	ATRT	SHH	SHH.Inf	ATRT, SHH	0.972	M	17	Posterior fossa	Y	Y	FALSE
NMB 1081	ATRT	SHH	SHH.Inf	N/A	N/A	M	44.6	Pineal gland	Y	Y	FALSE
NMB 1083	ATRT	SHH	SHH.Inf	N/A	N/A	M	12.3	Pineal gland	Y	Y	FALSE
NMB 461	ATRT	SHH	SHH.Inf	N/A	N/A	F	8	Posterior fossa	Y	Y	FALSE
NMB 1074	ATRT	SHH	SHH.Inf	N/A	N/A	M	0.2	Posterior fossa	Y	Y	FALSE
NMB 1082	ATRT	SHH	SHH.Inf	N/A	N/A	F	10.5	Posterior fossa	Y	Y	FALSE
Chalker 108 0975	ATRT	SHH	SHH.Inf	ATRT, SHH		F	12	Posterior fossa	N	N	N/A
INSTINCT ATRT 2	ATRT	SHH	SHH.Sup	N/A	N/A	N/A	N/A	MRT NOS	N	N	N/A
Chalker 96 4458	ATRT	SHH	SHH.Sup	ATRT, SHH	0.301	F	90.84	Cerebral hemispheres	N	N	N/A
Chalker 80 2927	ATRT	SHH	SHH.Sup	ATRT, SHH	0.997	M	0	ATRT NOS	N	N	N/A
NMB 1076	ATRT	SHH	SHH.Sup	N/A	N/A	F	27	Cerebral hemispheres	Y	Y	TRUE
Chalker 98 4459	ATRT	SHH	SHH.Sup	ATRT, SHH	0.9	M	38.4	ATRT NOS	N	N	N/A
Chalker 78 2923	ATRT	SHH	SHH.Sup	ATRT, SHH	0.057	M	72	Cerebral hemispheres	N	N	N/A
NMB 1069	ATRT	SHH	SHH.Sup	N/A	N/A	M	7.2	Cerebral hemispheres	Y	Y	FALSE
NMB 1208	ATRT	SHH	SHH.Sup	N/A	N/A	F	127.6	Frontal lobe	Y	Y	FALSE
NMB 843	ATRT	SHH	SHH.Sup	N/A	N/A	F	74	Frontal lobe	Y	Y	FALSE
NMB 846	ATRT	SHH	SHH.Sup	N/A	N/A	F	79.2	Frontal lobe	Y	Y	FALSE
NMB 854	ATRT	SHH	SHH.Sup	ATRT, SHH		F	18.6	Frontal lobe; Parietal lobe	Y	Y	N/A
NMB 775	ATRT	SHH	SHH.Sup	N/A	N/A	F	36.9	Frontal lobe	Y	Y	FALSE
NMB 888	ATRT	SHH	SHH.Sup	ATRT, SHH	0.999	M	58	Posterior fossa	Y	Y	TRUE
NMB 1080	ATRT	SHH	N/A	N/A	N/A	M	14.8	Temporal lobe	Y	Y	TRUE
Chalker 98 4465	ATRT	SHH	N/A	ATRT, SHH	0.691	F	4.32	Cerebellum	N	N	N/A
NMB 1062	ATRT	SHH	N/A	N/A	N/A	M	64.6	Posterior fossa	Y	Y	TRUE
NMB 842	ATRT	SHH	N/A	ATRT, SHH	0.989	F	18.6	ATRT NOS	Y	Y	FALSE
NMB 1010	ATRT	TYR	TYR	N/A	N/A	M	9.9	Cerebropontine angle; Spine	Y	Y	N/A
NMB 1039	ATRT	TYR	TYR	ATRT, TYR	0.953	M	27	Posterior fossa	Y	Y	FALSE

Chalker 97 4961	ATR T	TYR	TYR	ATRT, TYR		1F	6.96	posterior fossa	N	N	N/A
NMB 16	ATR T	TYR	TYR	N/A	N/A	M	1.2	Posterior fossa	Y	Y	TRUE
Chalker 99 4464	ATR T	TYR	TYR	N/A	N/A	M	7.8	Cerebellum	N	N	N/A
NMB 1070	ATR T	TYR	TYR	N/A	N/A	M	0.8	Posterior fossa	Y	Y	FALSE
NMB 1072	ATR T	TYR	TYR	N/A	N/A	M	3.2	Posterior fossa	Y	Y	FALSE
NMB 1075	ATR T	TYR	TYR	N/A	N/A	F	15	Posterior fossa	Y	Y	TRUE
NMB 1211	ATR T	TYR	TYR	N/A	N/A	M	0.1	Posterior fossa	N	N	FALSE
NMB 488	ATR T	TYR	TYR	N/A	N/A	F	6	Posterior fossa	Y	Y	N/A
NMB 1215	ATR T	TYR	TYR	N/A	N/A	F	10.5	Posterior fossa; Midline other	Y	Y	TRUE
NMB 957	ATR T	TYR	TYR	N/A	N/A	M	N/A	Frontal lobe	N	N	FALSE
NMB 776	ATR T	TYR	TYR	ATRT, TYR		1M	80.5	Posterior fossa	Y	Y	FALSE
NMB 778	ATR T	TYR	TYR	N/A	N/A	M	106.5	Posterior fossa	Y	Y	FALSE
NMB 853	ATR T	TYR	TYR	ATRT, TYR		1M	7.8	Posterior fossa	N	Y	TRUE
NMB 919	ATR T	TYR	HYPO	N/A	N/A	M	3.2	Posterior fossa	Y	Y	FALSE
NMB 920	ATR T	TYR	HYPO	N/A	N/A	M	67.5	Posterior fossa; Pineal gland	Y	Y	FALSE
Chalker 83 3598	ATR T	TYR	HYPO	ATRT, SHH		0.765F	12	Posterior fossa	N	N	N/A
NMB 1073	ATR T	TYR	HYPO	N/A	N/A	M	67.5	Pineal gland	Y	Y	FALSE
NMB 779	ATR T	TYR	HYPO	N/A	N/A	M	132.4	ATRT NOS	Y	Y	TRUE
NMB 1212	ATR T	TYR	HYPO	N/A	N/A	F	14	Posterior fossa	Y	Y	FALSE
NMB 836	ATR T	TYR	HYPO	N/A	N/A	M	19.1	ATRT NOS	Y	Y	FALSE
NMB 878	ATR T	TYR	HYPO	N/A	N/A	M	7.8	ATRT NOS	Y	Y	FALSE
NMB 1063	ATR T	MYC	HYPO	N/A	N/A	F	22.6	Cerebral hemispheres; Temporal lobe; Optic chiasm; Thalamus	Y	Y	TRUE
NMB 1217	ATR T	MYC	HYPO	N/A	N/A	F	2.5	Cerebral hemispheres; Pineal gland; Other	Y	Y	TRUE
NMB 1213	ATR T	MYC	HYPO	N/A	N/A	M	3.7	Posterior fossa	Y	Y	FALSE
Chalker 43 3138	ATR T	MYC	MYC	ATRT, MYC		0.893F	24	Thalamus	N	N	N/A
NMB 885	ATR T	MYC	MYC	N/A	N/A	M	13.2	Posterior fossa	Y	Y	FALSE
Chalker 100 4453	ATR T	MYC	MYC	ATRT, MYC		0.736M	181.44	Cerebral hemispheres	N	N	N/A
NMB 1079	ATR T	MYC	MYC	N/A	N/A	F	16.5	Pineal gland	Y	Y	TRUE
Chalker 85 3792	ATR T	MYC	MYC	ATRT, MYC		1F	108	Brainstem	N	N	N/A
Chalker 101 4468	ATR T	MYC	MYC	ATRT, MYC		0.987M	171	Spinal cord	N	N	N/A
NMB 1214	ATR T	MYC	MYC	N/A	N/A	F	16.5	Pineal gland	Y	Y	TRUE
NMB 1216	ATR T	MYC	MYC	N/A	N/A	F	5	Posterior fossa	Y	Y	FALSE

NMB 834	ATR T	MYC	MYC	ATR, MYC		1	F	3	ATR NOS		N	N	FALSE
NMB 856	ATR T	MYC	MYC	ATR, MYC		0.992	F	135.3	Intramedullary		Y	Y	FALSE
NMB 876	ATR T	N/A	SHH.Sup	N/A	N/A		M	22.4	Frontal lobe		N	N	FALSE
NMB 478	ATR T	N/A	SHH.Sup	N/A	N/A		M	13.7	ATR NOS		N	N	FALSE
NMB 887	RTK	TYR	HYPO	N/A	N/A		M	44.4	Kidney		Y	Y	FALSE
NMB 997	RTK	TYR	HYPO	N/A	N/A		M	1.6	Kidney		Y	Y	TRUE
NMB 875	RTK	TYR	HYPO	N/A	N/A		F	8.3	Kidney		N	N	TRUE
NMB 1007	RTK	TYR	HYPO	N/A	N/A		M	18.2	Kidney		Y	Y	N/A
NMB 860	RTK	TYR	HYPO	N/A	N/A		F	6.7	Kidney		Y	Y	TRUE
NMB 886	RTK	TYR	HYPO	N/A	N/A		F	9.2	Kidney		N	N	TRUE
NMB 1006	RTK	MYC	HYPO	N/A	N/A		M	18.2	Kidney		Y	Y	FALSE
NMB 844	RTK	MYC	HYPO	N/A	N/A		M	4.4	Kidney		Y	Y	TRUE
NMB 847	RTK	MYC	HYPO	N/A	N/A		F	7.9	Kidney		N	Y	TRUE
NMB 879	RTK	MYC	HYPO	N/A	N/A		M	13.8	Kidney		N	N	N/A
NMB 1009	RTK	MYC	MYC	N/A	N/A		F	24	Kidney		Y	Y	TRUE
NMB 998	RTK	MYC	MYC	N/A	N/A		F	12	Kidney		Y	Y	TRUE
NMB 1261	RTK	MYC	MYC	N/A	N/A		N/A	N/A	Kidney		N	N	N/A
NMB 852	RTK	MYC	MYC	N/A	N/A		M	3.8	Kidney		Y	Y	TRUE
NMB 865	RTK	MYC	MYC	N/A	N/A		M	44.9	Kidney		Y	Y	FALSE
NMB 877	RTK	MYC	MYC	N/A	N/A		M	45.1	Kidney		Y	Y	TRUE
NMB 841	RTK	MYC	MYC	N/A	N/A		F	16.9	Kidney		N	N	N/A
NMB 881	RTK	MYC	MYC	N/A	N/A		M	29.1	Kidney		N	N	FALSE
NMB 848	RTK	N/A	HYPO	N/A	N/A		M	74.6	Kidney		N	N	N/A
NMB 1008	ECR T	TYR	HYPO	N/A	N/A		F	12	Lung		Y	Y	TRUE
NMB 1266	ECR T	TYR	HYPO	N/A	N/A		F	30	Bladder		Y	Y	FALSE
NMB 1273	ECR T	TYR	HYPO	N/A	N/A		N/A	0	Paraspine		N	N	N/A
NMB 840	ECR T	TYR	HYPO	N/A	N/A		M	5.4	Thorax		Y	Y	TRUE
NMB 1264	ECR T	MYC	HYPO	N/A	N/A		M	6	Thorax		Y	Y	TRUE
NMB 1265	ECR T	MYC	HYPO	N/A	N/A		N/A	N/A	Craniovertebral junction; Cervical spine		N	N	N/A
NMB 1269	ECR T	MYC	HYPO	N/A	N/A		F	9	Liver		Y	Y	FALSE
NMB 838	ECR T	MYC	HYPO	N/A	N/A		F	5.4	Abdomen		Y	Y	TRUE
NMB 845	ECR T	MYC	HYPO	N/A	N/A		M	14.5	Armpit		Y	Y	FALSE
NMB 839	ECR T	MYC	HYPO	N/A	N/A		M	2.2	Abdomen		Y	Y	TRUE
NMB 849	ECR T	MYC	HYPO	N/A	N/A		F	0.3	Spine		N	N	FALSE
NMB 862	ECR T	MYC	HYPO	N/A	N/A		M	13.4	Liver		Y	Y	TRUE
NMB 864	ECR T	MYC	HYPO	N/A	N/A		F	56.9	Buttock		Y	Y	TRUE
NMB 850	ECR T	MYC	HYPO	N/A	N/A		M	29	Spine		Y	Y	FALSE

NMB 1005	ECR T	MYC	MYC	N/A	N/A	F	21	Thorax; Mediastinum	Y	Y	FALSE
NMB 1262	ECR T	MYC	MYC	N/A	N/A	N/A	N/A	Paratracheal	N	N	N/A
NMB 1267	ECR T	MYC	MYC	N/A	N/A	F	51.6	Neck	Y	Y	FALSE
NMB 861	ECR T	MYC	MYC	N/A	N/A	M	93.3	Bladder	Y	Y	FALSE
NMB 896	ECR T	MYC	MYC	N/A	N/A	M	104.1	Thorax	Y	Y	TRUE
NMB 882	ECR T	MYC	MYC	N/A	N/A	M	12.6	Abdomen	Y	Y	TRUE
NMB 880	ECR T	MYC	MYC	N/A	N/A	M	15.6	Liver	N	N	TRUE
NMB 835	ECR T	MYC	MYC	N/A	N/A	M	17.1	Pelvis	N	N	N/A
NMB 863	ECR T	MYC	MYC	N/A	N/A	F	8.5	Pelvis	Y	Y	FALSE
NMB 883	ECR T	MYC	MYC	N/A	N/A	M	11.7	Liver	N	N	TRUE
NMB 959	MR T	SHH	SHH.Inf	ATRT, SHH		1F	N/A	MRT NOS	Y	Y	FALSE
NMB 958	MR T	SHH	SHH.Inf	N/A	N/A	F	N/A	MRT NOS	Y	Y	FALSE
NMB 1042	MR T	TYR	TYR	N/A	N/A	N/A	25.2	MRT NOS	N	N	N/A
NMB 914	MR T	TYR	TYR	N/A	N/A	N/A	N/A	MRT NOS	N	N	N/A
NMB 1041	MR T	TYR	HYPO	N/A	N/A	F	0	MRT NOS	Y	Y	N/A
NMB 960	MR T	MYC	N/A	ATRT, MYC		0.97M	N/A	MRT NOS	Y	Y	FALSE
NMB 921	MR T	MYC	N/A	N/A	N/A	M	7.3	MRT NOS	Y	Y	TRUE

8.2 B) HGU133Plus2 cohort used as part of this thesis

SampleID	GEOAccession	TumourType	GEOPlatform	GEODataset	SampleName	Exclude	Reason
TARGET-50-CAAAA-01A-01R		WT	GPL570	ftp://caftpdc.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-CAAAA	FALSE	NA
TARGET-50-CAAAAB-01A-01R		WT	GPL570	ftp://caftpdc.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-CAAAAB	FALSE	NA
TARGET-50-CAAAAC-01A-01R		WT	GPL570	ftp://caftpdc.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-CAAAAC	FALSE	NA
TARGET-50-CAAAAH-01A-01R		WT	GPL570	ftp://caftpdc.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-CAAAAH	FALSE	NA
TARGET-50-CAAAAJ-01A-01R		WT	GPL570	ftp://caftpdc.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-CAAAAJ	FALSE	NA
TARGET-50-CAAAAL-01A-01R		WT	GPL570	ftp://caftpdc.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-CAAAAL	FALSE	NA
TARGET-50-CAAAAM-01A-01R		WT	GPL570	ftp://caftpdc.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-CAAAAM	TRUE	Clustering
TARGET-50-CAAAAO-01A-01R		WT	GPL570	ftp://caftpdc.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-CAAAAO	FALSE	NA
TARGET-50-CAAAAP-01A-01R		WT	GPL570	ftp://caftpdc.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-CAAAAP	FALSE	NA
TARGET-50-CAAAAQ-01A-01R		WT	GPL570	ftp://caftpdc.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-CAAAAQ	FALSE	NA
TARGET-50-CAAAAR-01A-01R		WT	GPL570	ftp://caftpdc.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-CAAAAR	FALSE	NA

TARGET-50-PALLCK-01A-01R		WT	GPL570	ftp://caftpd.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-PALLCK	FAL SE	NA
TARGET-50-PALLFB-01A-01R		WT	GPL570	ftp://caftpd.nci.nih.gov/pub/OCG-DCC/TARGET/WT/gene_expression_array/L1/	TARGET-50-PALLFB	FAL SE	NA
OD005		RMS	GPL570	E-TABM-1202	RMS005	FAL SE	NA
OD006		RMS	GPL570	E-TABM-1202	RMS006	FAL SE	NA
OD009		RMS	GPL570	E-TABM-1202	RMS009	FAL SE	NA
OD010		RMS	GPL570	E-TABM-1202	RMS010	FAL SE	NA
OD012		RMS	GPL570	E-TABM-1202	RMS012	FAL SE	NA
OD015		RMS	GPL570	E-TABM-1202	RMS015	FAL SE	NA
OD017		RMS	GPL570	E-TABM-1202	RMS017	FAL SE	NA
OD026		RMS	GPL570	E-TABM-1202	RMS026	FAL SE	NA
OD027		RMS	GPL570	E-TABM-1202	RMS027	FAL SE	NA
OD029		RMS	GPL570	E-TABM-1202	RMS029	FAL SE	NA
OD032		RMS	GPL570	E-TABM-1202	RMS032	FAL SE	NA
OD033		RMS	GPL570	E-TABM-1202	RMS033	FAL SE	NA
OD034		RMS	GPL570	E-TABM-1202	RMS034	FAL SE	NA
OD035		RMS	GPL570	E-TABM-1202	RMS035	FAL SE	NA
OD038		RMS	GPL570	E-TABM-1202	RMS038	FAL SE	NA
OD039		RMS	GPL570	E-TABM-1202	RMS039	FAL SE	NA
OD041		RMS	GPL570	E-TABM-1202	RMS041	FAL SE	NA
OD042		RMS	GPL570	E-TABM-1202	RMS042	FAL SE	NA
OD043		RMS	GPL570	E-TABM-1202	RMS043	FAL SE	NA
OD046		RMS	GPL570	E-TABM-1202	RMS046	FAL SE	NA
OD049		RMS	GPL570	E-TABM-1202	RMS049	FAL SE	NA
OD050		RMS	GPL570	E-TABM-1202	RMS050	FAL SE	NA
OD051		RMS	GPL570	E-TABM-1202	RMS051	FAL SE	NA
OD052		RMS	GPL570	E-TABM-1202	RMS052	FAL SE	NA
OD054		RMS	GPL570	E-TABM-1202	RMS054	FAL SE	NA
OD055		RMS	GPL570	E-TABM-1202	RMS055	FAL SE	NA
OD057		RMS	GPL570	E-TABM-1202	RMS057	FAL SE	NA
OD058		RMS	GPL570	E-TABM-1202	RMS058	FAL SE	NA
OD060		RMS	GPL570	E-TABM-1202	RMS060	FAL SE	NA
OD064		RMS	GPL570	E-TABM-1202	RMS064	FAL SE	NA

OD067		RMS	GPL570	E-TABM-1202	RMS067	FAL	SE	NA
OD075		RMS	GPL570	E-TABM-1202	RMS075	FAL	SE	NA
OD076		RMS	GPL570	E-TABM-1202	RMS076	FAL	SE	NA
OD078		RMS	GPL570	E-TABM-1202	RMS078	FAL	SE	NA
OD080		RMS	GPL570	E-TABM-1202	RMS080	FAL	SE	NA
OD081		RMS	GPL570	E-TABM-1202	RMS081	FAL	SE	NA
OD082		RMS	GPL570	E-TABM-1202	RMS082	FAL	SE	NA
OD084		RMS	GPL570	E-TABM-1202	RMS084	FAL	SE	NA
OD085		RMS	GPL570	E-TABM-1202	RMS085	FAL	SE	NA
OD087		RMS	GPL570	E-TABM-1202	RMS087	FAL	SE	NA
OD090		RMS	GPL570	E-TABM-1202	RMS090	FAL	SE	NA
OD091		RMS	GPL570	E-TABM-1202	RMS091	FAL	SE	NA
OD092		RMS	GPL570	E-TABM-1202	RMS092	FAL	SE	NA
OD093		RMS	GPL570	E-TABM-1202	RMS093	FAL	SE	NA
OD094		RMS	GPL570	E-TABM-1202	RMS094	FAL	SE	NA
OD096		RMS	GPL570	E-TABM-1202	RMS096	FAL	SE	NA
OD100		RMS	GPL570	E-TABM-1202	RMS100	FAL	SE	NA
OD102		RMS	GPL570	E-TABM-1202	RMS102	FAL	SE	NA
OD104		RMS	GPL570	E-TABM-1202	RMS104	FAL	SE	NA
OD105		RMS	GPL570	E-TABM-1202	RMS105	FAL	SE	NA
OD106		RMS	GPL570	E-TABM-1202	RMS106	FAL	SE	NA
OD109		RMS	GPL570	E-TABM-1202	RMS109	FAL	SE	NA
OD110		RMS	GPL570	E-TABM-1202	RMS110	FAL	SE	NA
OD111		RMS	GPL570	E-TABM-1202	RMS111	FAL	SE	NA
OD112		RMS	GPL570	E-TABM-1202	RMS112	FAL	SE	NA
OD113		RMS	GPL570	E-TABM-1202	RMS113	FAL	SE	NA
OD114		RMS	GPL570	E-TABM-1202	RMS114	FAL	SE	NA
OD116		RMS	GPL570	E-TABM-1202	RMS116	FAL	SE	NA
OD117		RMS	GPL570	E-TABM-1202	RMS117	FAL	SE	NA
OD118		RMS	GPL570	E-TABM-1202	RMS118	FAL	SE	NA
OD120		RMS	GPL570	E-TABM-1202	RMS120	FAL	SE	NA
OD123		RMS	GPL570	E-TABM-1202	RMS123	FAL	SE	NA

OD130		RMS	GPL570	E-TABM-1202	RMS130	FAL SE	NA
OD131		RMS	GPL570	E-TABM-1202	RMS131	FAL SE	NA
OD132		RMS	GPL570	E-TABM-1202	RMS132	FAL SE	NA
OD133		RMS	GPL570	E-TABM-1202	RMS133	FAL SE	NA
OD134		RMS	GPL570	E-TABM-1202	RMS134	FAL SE	NA
OD136		RMS	GPL570	E-TABM-1202	RMS136	FAL SE	NA
OD138		RMS	GPL570	E-TABM-1202	RMS138	FAL SE	NA
OD139		RMS	GPL570	E-TABM-1202	RMS139	FAL SE	NA
OD141		RMS	GPL570	E-TABM-1202	RMS141	FAL SE	NA
OD142		RMS	GPL570	E-TABM-1202	RMS142	FAL SE	NA
OD143		RMS	GPL570	E-TABM-1202	RMS143	FAL SE	NA
OD144		RMS	GPL570	E-TABM-1202	RMS144	FAL SE	NA
OD148		RMS	GPL570	E-TABM-1202	RMS148	FAL SE	NA
OD149		RMS	GPL570	E-TABM-1202	RMS149	FAL SE	NA
OD151		RMS	GPL570	E-TABM-1202	RMS151	FAL SE	NA
OD152		RMS	GPL570	E-TABM-1202	RMS152	FAL SE	NA
OD153		RMS	GPL570	E-TABM-1202	RMS153	FAL SE	NA
OD155		RMS	GPL570	E-TABM-1202	RMS155	FAL SE	NA
OD156		RMS	GPL570	E-TABM-1202	RMS156	FAL SE	NA
OD157		RMS	GPL570	E-TABM-1202	RMS157	FAL SE	NA
OD158		RMS	GPL570	E-TABM-1202	RMS158	FAL SE	NA
OD159		RMS	GPL570	E-TABM-1202	RMS159	FAL SE	NA
OD160		RMS	GPL570	E-TABM-1202	RMS160	FAL SE	NA
OD307		RMS	GPL570	E-TABM-1202	RMS307	FAL SE	NA
OD308		RMS	GPL570	E-TABM-1202	RMS308	FAL SE	NA
OD311		RMS	GPL570	E-TABM-1202	RMS311	FAL SE	NA
OD316		RMS	GPL570	E-TABM-1202	RMS316	FAL SE	NA
OD317		RMS	GPL570	E-TABM-1202	RMS317	FAL SE	NA
OD321		RMS	GPL570	E-TABM-1202	RMS321	FAL SE	NA
OD323		RMS	GPL570	E-TABM-1202	RMS323	FAL SE	NA
OD324		RMS	GPL570	E-TABM-1202	RMS324	FAL SE	NA
OD326		RMS	GPL570	E-TABM-1202	RMS326	FAL SE	NA

OD329		RMS	GPL570	E-TABM-1202	RMS329	FAL SE	NA
OD342		RMS	GPL570	E-TABM-1202	RMS342	FAL SE	NA
OD353		RMS	GPL570	E-TABM-1202	RMS353	FAL SE	NA
OD357		RMS	GPL570	E-TABM-1202	RMS357	FAL SE	NA
OD358		RMS	GPL570	E-TABM-1202	RMS358	FAL SE	NA
OD362		RMS	GPL570	E-TABM-1202	RMS362	FAL SE	NA
OD363		RMS	GPL570	E-TABM-1202	RMS363	FAL SE	NA
GSM260959	GSM26 0959	MB	GPL570	GSE10327	Medulloblastoma #255	FAL SE	NA
GSM260960	GSM26 0960	MB	GPL570	GSE10327	Medulloblastoma #256	FAL SE	NA
GSM260961	GSM26 0961	MB	GPL570	GSE10327	Medulloblastoma #258	FAL SE	NA
GSM260962	GSM26 0962	MB	GPL570	GSE10327	Medulloblastoma #259	FAL SE	NA
GSM260963	GSM26 0963	MB	GPL570	GSE10327	Medulloblastoma #260	FAL SE	NA
GSM260964	GSM26 0964	MB	GPL570	GSE10327	Medulloblastoma #261	FAL SE	NA
GSM260965	GSM26 0965	MB	GPL570	GSE10327	Medulloblastoma #262	FAL SE	NA
GSM260966	GSM26 0966	MB	GPL570	GSE10327	Medulloblastoma #264	FAL SE	NA
GSM260967	GSM26 0967	MB	GPL570	GSE10327	Medulloblastoma #265	FAL SE	NA
GSM260968	GSM26 0968	MB	GPL570	GSE10327	Medulloblastoma #267	FAL SE	NA
GSM260969	GSM26 0969	MB	GPL570	GSE10327	Medulloblastoma #268	FAL SE	NA
GSM260970	GSM26 0970	MB	GPL570	GSE10327	Medulloblastoma #269	FAL SE	NA
GSM260971	GSM26 0971	MB	GPL570	GSE10327	Medulloblastoma #270	FAL SE	NA
GSM260972	GSM26 0972	MB	GPL570	GSE10327	Medulloblastoma #272	FAL SE	NA
GSM260973	GSM26 0973	MB	GPL570	GSE10327	Medulloblastoma #273	FAL SE	NA
GSM260974	GSM26 0974	MB	GPL570	GSE10327	Medulloblastoma #274	FAL SE	NA
GSM260975	GSM26 0975	MB	GPL570	GSE10327	Medulloblastoma #275	FAL SE	NA
GSM260976	GSM26 0976	MB	GPL570	GSE10327	Medulloblastoma #311	FAL SE	NA
GSM260977	GSM26 0977	MB	GPL570	GSE10327	Medulloblastoma #312	FAL SE	NA
GSM260978	GSM26 0978	MB	GPL570	GSE10327	Medulloblastoma #313	TRU E	Cluste ring
GSM260979	GSM26 0979	MB	GPL570	GSE10327	Medulloblastoma #315	FAL SE	NA
GSM260980	GSM26 0980	MB	GPL570	GSE10327	Medulloblastoma #316	FAL SE	NA
GSM260981	GSM26 0981	MB	GPL570	GSE10327	Medulloblastoma #317	FAL SE	NA
GSM260982	GSM26 0982	MB	GPL570	GSE10327	Medulloblastoma #318	FAL SE	NA
GSM260983	GSM26 0983	MB	GPL570	GSE10327	Medulloblastoma #324	FAL SE	NA

GSM260984	GSM260984	MB	GPL570	GSE10327	Medulloblastoma #325	FAL SE	NA
GSM260985	GSM260985	MB	GPL570	GSE10327	Medulloblastoma #326	FAL SE	NA
GSM260986	GSM260986	MB	GPL570	GSE10327	Medulloblastoma #332	FAL SE	NA
GSM260987	GSM260987	MB	GPL570	GSE10327	Medulloblastoma #334	FAL SE	NA
GSM260988	GSM260988	MB	GPL570	GSE10327	Medulloblastoma #335	FAL SE	NA
GSM260989	GSM260989	MB	GPL570	GSE10327	Medulloblastoma #336	FAL SE	NA
GSM260990	GSM260990	MB	GPL570	GSE10327	Medulloblastoma #337	FAL SE	NA
GSM260991	GSM260991	MB	GPL570	GSE10327	Medulloblastoma #338	FAL SE	NA
GSM260992	GSM260992	MB	GPL570	GSE10327	Medulloblastoma #339	FAL SE	NA
GSM260993	GSM260993	MB	GPL570	GSE10327	Medulloblastoma #340	FAL SE	NA
GSM260994	GSM260994	MB	GPL570	GSE10327	Medulloblastoma #341	FAL SE	NA
GSM260995	GSM260995	MB	GPL570	GSE10327	Medulloblastoma #342	FAL SE	NA
GSM260996	GSM260996	MB	GPL570	GSE10327	Medulloblastoma #343	FAL SE	NA
GSM260997	GSM260997	MB	GPL570	GSE10327	Medulloblastoma #365	FAL SE	NA
GSM260998	GSM260998	MB	GPL570	GSE10327	Medulloblastoma #367	FAL SE	NA
GSM260999	GSM260999	MB	GPL570	GSE10327	Medulloblastoma #368	FAL SE	NA
GSM261000	GSM261000	MB	GPL570	GSE10327	Medulloblastoma #369	FAL SE	NA
GSM261001	GSM261001	MB	GPL570	GSE10327	Medulloblastoma #370	FAL SE	NA
GSM261002	GSM261002	MB	GPL570	GSE10327	Medulloblastoma #371	FAL SE	NA
GSM261003	GSM261003	MB	GPL570	GSE10327	Medulloblastoma #372	FAL SE	NA
GSM261004	GSM261004	MB	GPL570	GSE10327	Medulloblastoma #373	FAL SE	NA
GSM261005	GSM261005	MB	GPL570	GSE10327	Medulloblastoma #374	FAL SE	NA
GSM261006	GSM261006	MB	GPL570	GSE10327	Medulloblastoma #377	FAL SE	NA
GSM261007	GSM261007	MB	GPL570	GSE10327	Medulloblastoma #379	FAL SE	NA
GSM261008	GSM261008	MB	GPL570	GSE10327	Medulloblastoma #421	FAL SE	NA
GSM261009	GSM261009	MB	GPL570	GSE10327	Medulloblastoma #424	FAL SE	NA
GSM261010	GSM261010	MB	GPL570	GSE10327	Medulloblastoma #425	FAL SE	NA
GSM261011	GSM261011	MB	GPL570	GSE10327	Medulloblastoma #426	FAL SE	NA
GSM261012	GSM261012	MB	GPL570	GSE10327	Medulloblastoma #427	FAL SE	NA
GSM261013	GSM261013	MB	GPL570	GSE10327	Medulloblastoma #434	FAL SE	NA
GSM261014	GSM261014	MB	GPL570	GSE10327	Medulloblastoma #435	FAL SE	NA
GSM261015	GSM261015	MB	GPL570	GSE10327	Medulloblastoma #440	FAL SE	NA

GSM261016	GSM26 1016	MB	GPL570	GSE10327	Medulloblastoma #446	FAL SE	NA
GSM261017	GSM26 1017	MB	GPL570	GSE10327	Medulloblastoma #447	FAL SE	NA
GSM261018	GSM26 1018	MB	GPL570	GSE10327	Medulloblastoma #452	FAL SE	NA
GSM261019	GSM26 1019	MB	GPL570	GSE10327	Medulloblastoma #455	FAL SE	NA
GSM261020	GSM26 1020	MB	GPL570	GSE10327	Medulloblastoma #458	FAL SE	NA
GSM324062	GSM32 4062	MB	GPL570	GSE12992	Medulloblastoma tumor MB79	TRU E	Cluste ring
GSM324063	GSM32 4063	MB	GPL570	GSE12992	Medulloblastoma tumor MB80	FAL SE	NA
GSM324064	GSM32 4064	MB	GPL570	GSE12992	Medulloblastoma tumor MB81a	TRU E	Cluste ring
GSM324065	GSM32 4065	MB	GPL570	GSE12992	Medulloblastoma tumor MB82	FAL SE	NA
GSM324066	GSM32 4066	MB	GPL570	GSE12992	Medulloblastoma tumor MB87a	FAL SE	NA
GSM324067	GSM32 4067	MB	GPL570	GSE12992	Medulloblastoma tumor MB88	FAL SE	NA
GSM324068	GSM32 4068	MB	GPL570	GSE12992	Medulloblastoma tumor MB89	FAL SE	NA
GSM324069	GSM32 4069	MB	GPL570	GSE12992	Medulloblastoma tumor MB91	FAL SE	NA
GSM324082	GSM32 4082	MB	GPL570	GSE12992	Medulloblastoma tumor MB92	TRU E	Cluste ring
GSM324083	GSM32 4083	MB	GPL570	GSE12992	Medulloblastoma tumor MB93	TRU E	Cluste ring
GSM324084	GSM32 4084	MB	GPL570	GSE12992	Medulloblastoma tumor MB95	FAL SE	NA
GSM324085	GSM32 4085	MB	GPL570	GSE12992	Medulloblastoma tumor MB96	FAL SE	NA
GSM324090	GSM32 4090	MB	GPL570	GSE12992	Medulloblastoma tumor MB99	FAL SE	NA
GSM324091	GSM32 4091	MB	GPL570	GSE12992	Medulloblastoma tumor MB100	FAL SE	NA
GSM324092	GSM32 4092	MB	GPL570	GSE12992	Medulloblastoma tumor MB101	FAL SE	NA
GSM324093	GSM32 4093	MB	GPL570	GSE12992	Medulloblastoma tumor MB102	FAL SE	NA
GSM324104	GSM32 4104	MB	GPL570	GSE12992	Medulloblastoma tumor MB105	FAL SE	NA
GSM324111	GSM32 4111	MB	GPL570	GSE12992	Medulloblastoma tumor MB106	FAL SE	NA
GSM324112	GSM32 4112	MB	GPL570	GSE12992	Medulloblastoma tumor MB107	TRU E	Cluste ring
GSM324113	GSM32 4113	MB	GPL570	GSE12992	Medulloblastoma tumor MB108	FAL SE	NA
GSM324115	GSM32 4115	MB	GPL570	GSE12992	Medulloblastoma tumor MB109	FAL SE	NA
GSM324119	GSM32 4119	MB	GPL570	GSE12992	Medulloblastoma tumor MB112	FAL SE	NA
GSM324137	GSM32 4137	MB	GPL570	GSE12992	Medulloblastoma tumor MB116	FAL SE	NA
GSM324138	GSM32 4138	MB	GPL570	GSE12992	Medulloblastoma tumor MB117	FAL SE	NA
GSM324139	GSM32 4139	MB	GPL570	GSE12992	Medulloblastoma tumor MB118	FAL SE	NA
GSM324140	GSM32 4140	MB	GPL570	GSE12992	Medulloblastoma tumor MB119	FAL SE	NA
GSM324141	GSM32 4141	MB	GPL570	GSE12992	Medulloblastoma tumor MB120	FAL SE	NA

GSM324508	GSM32 4508	MB	GPL570	GSE12992	Medulloblastoma tumor MB121	TRU E	Cluste ring
GSM324512	GSM32 4512	MB	GPL570	GSE12992	Medulloblastoma tumor MB122	FAL SE	NA
GSM324513	GSM32 4513	MB	GPL570	GSE12992	Medulloblastoma tumor MB123	TRU E	Cluste ring
GSM324514	GSM32 4514	MB	GPL570	GSE12992	Medulloblastoma tumor MB124	FAL SE	NA
GSM324515	GSM32 4515	MB	GPL570	GSE12992	Medulloblastoma tumor MB125	FAL SE	NA
GSM324516	GSM32 4516	MB	GPL570	GSE12992	Medulloblastoma tumor MB127	FAL SE	NA
GSM324517	GSM32 4517	MB	GPL570	GSE12992	Medulloblastoma tumor MB128	FAL SE	NA
GSM324526	GSM32 4526	MB	GPL570	GSE12992	Medulloblastoma tumor MB130	FAL SE	NA
GSM325233	GSM32 5233	MB	GPL570	GSE12992	Medulloblastoma tumor MB131	FAL SE	NA
GSM325278	GSM32 5278	MB	GPL570	GSE12992	Medulloblastoma tumor MB133	FAL SE	NA
GSM325280	GSM32 5280	MB	GPL570	GSE12992	Medulloblastoma tumor MB134	FAL SE	NA
GSM325281	GSM32 5281	MB	GPL570	GSE12992	Medulloblastoma tumor MB135	FAL SE	NA
GSM325282	GSM32 5282	MB	GPL570	GSE12992	Medulloblastoma tumor MB136	FAL SE	NA
GSM408899	GSM40 8899	NB	GPL570	GSE16237	Neuroblastoma case 3	FAL SE	NA
GSM408900	GSM40 8900	NB	GPL570	GSE16237	Neuroblastoma case 4	FAL SE	NA
GSM408901	GSM40 8901	NB	GPL570	GSE16237	Neuroblastoma case 5	FAL SE	NA
GSM408902	GSM40 8902	NB	GPL570	GSE16237	Neuroblastoma case 9	FAL SE	NA
GSM408903	GSM40 8903	NB	GPL570	GSE16237	Neuroblastoma case 11	FAL SE	NA
GSM408904	GSM40 8904	NB	GPL570	GSE16237	Neuroblastoma case 18	FAL SE	NA
GSM408905	GSM40 8905	NB	GPL570	GSE16237	Neuroblastoma case 55	FAL SE	NA
GSM408906	GSM40 8906	NB	GPL570	GSE16237	Neuroblastoma case 59	FAL SE	NA
GSM408907	GSM40 8907	NB	GPL570	GSE16237	Neuroblastoma case 66	FAL SE	NA
GSM408908	GSM40 8908	NB	GPL570	GSE16237	Neuroblastoma case 142	FAL SE	NA
GSM408909	GSM40 8909	NB	GPL570	GSE16237	Neuroblastoma case 147	FAL SE	NA
GSM408910	GSM40 8910	NB	GPL570	GSE16237	Neuroblastoma case 151	FAL SE	NA
GSM408911	GSM40 8911	NB	GPL570	GSE16237	Neuroblastoma case 160	FAL SE	NA
GSM408912	GSM40 8912	NB	GPL570	GSE16237	Neuroblastoma case 163	FAL SE	NA
GSM408913	GSM40 8913	NB	GPL570	GSE16237	Neuroblastoma case 189	TRU E	Cluste ring
GSM408914	GSM40 8914	NB	GPL570	GSE16237	Neuroblastoma case 194	FAL SE	NA
GSM408915	GSM40 8915	NB	GPL570	GSE16237	Neuroblastoma case 209	FAL SE	NA
GSM408916	GSM40 8916	NB	GPL570	GSE16237	Neuroblastoma case 259	FAL SE	NA
GSM408917	GSM40 8917	NB	GPL570	GSE16237	Neuroblastoma case 280	FAL SE	NA

GSM408918	GSM408918	NB	GPL570	GSE16237	Neuroblastoma case 287	FAL SE	NA
GSM408919	GSM408919	NB	GPL570	GSE16237	Neuroblastoma case 288	FAL SE	NA
GSM408920	GSM408920	NB	GPL570	GSE16237	Neuroblastoma case 296	FAL SE	NA
GSM408921	GSM408921	NB	GPL570	GSE16237	Neuroblastoma case 311	FAL SE	NA
GSM408922	GSM408922	NB	GPL570	GSE16237	Neuroblastoma case 313	FAL SE	NA
GSM408923	GSM408923	NB	GPL570	GSE16237	Neuroblastoma case 314	FAL SE	NA
GSM408924	GSM408924	NB	GPL570	GSE16237	Neuroblastoma case 320	FAL SE	NA
GSM408925	GSM408925	NB	GPL570	GSE16237	Neuroblastoma case 338	FAL SE	NA
GSM408926	GSM408926	NB	GPL570	GSE16237	Neuroblastoma case 342	FAL SE	NA
GSM408927	GSM408927	NB	GPL570	GSE16237	Neuroblastoma case 346	FAL SE	NA
GSM408928	GSM408928	NB	GPL570	GSE16237	Neuroblastoma case 351	FAL SE	NA
GSM408929	GSM408929	NB	GPL570	GSE16237	Neuroblastoma case 360	FAL SE	NA
GSM408930	GSM408930	NB	GPL570	GSE16237	Neuroblastoma case 364	FAL SE	NA
GSM408931	GSM408931	NB	GPL570	GSE16237	Neuroblastoma case 365	FAL SE	NA
GSM408932	GSM408932	NB	GPL570	GSE16237	Neuroblastoma case 373	FAL SE	NA
GSM408933	GSM408933	NB	GPL570	GSE16237	Neuroblastoma case 379	FAL SE	NA
GSM408934	GSM408934	NB	GPL570	GSE16237	Neuroblastoma case 384	FAL SE	NA
GSM408935	GSM408935	NB	GPL570	GSE16237	Neuroblastoma case 400	FAL SE	NA
GSM408936	GSM408936	NB	GPL570	GSE16237	Neuroblastoma case 401	FAL SE	NA
GSM408937	GSM408937	NB	GPL570	GSE16237	Neuroblastoma case 402	FAL SE	NA
GSM408938	GSM408938	NB	GPL570	GSE16237	Neuroblastoma case 410	FAL SE	NA
GSM408939	GSM408939	NB	GPL570	GSE16237	Neuroblastoma case 413	FAL SE	NA
GSM408940	GSM408940	NB	GPL570	GSE16237	Neuroblastoma case 418	FAL SE	NA
GSM408941	GSM408941	NB	GPL570	GSE16237	Neuroblastoma case 419	FAL SE	NA
GSM408942	GSM408942	NB	GPL570	GSE16237	Neuroblastoma case 423	FAL SE	NA
GSM408943	GSM408943	NB	GPL570	GSE16237	Neuroblastoma case 424	FAL SE	NA
GSM408944	GSM408944	NB	GPL570	GSE16237	Neuroblastoma case 426	FAL SE	NA
GSM408945	GSM408945	NB	GPL570	GSE16237	Neuroblastoma case 428	FAL SE	NA
GSM408946	GSM408946	NB	GPL570	GSE16237	Neuroblastoma case 429	FAL SE	NA
GSM408947	GSM408947	NB	GPL570	GSE16237	Neuroblastoma case 430	FAL SE	NA
GSM408948	GSM408948	NB	GPL570	GSE16237	Neuroblastoma case 434	FAL SE	NA
GSM408949	GSM408949	NB	GPL570	GSE16237	Neuroblastoma case 452	TRU E	Clustering

GSM414000	GSM41 4000	NB	GPL570	GSE16476	Neuroblastoma. #1	FAL SE	NA
GSM414001	GSM41 4001	NB	GPL570	GSE16476	Neuroblastoma. #2	FAL SE	NA
GSM414002	GSM41 4002	NB	GPL570	GSE16476	Neuroblastoma. #3	FAL SE	NA
GSM414003	GSM41 4003	NB	GPL570	GSE16476	Neuroblastoma. #4	FAL SE	NA
GSM414004	GSM41 4004	NB	GPL570	GSE16476	Neuroblastoma. #5	FAL SE	NA
GSM414005	GSM41 4005	NB	GPL570	GSE16476	Neuroblastoma. #6	FAL SE	NA
GSM414006	GSM41 4006	NB	GPL570	GSE16476	Neuroblastoma. #7	FAL SE	NA
GSM414007	GSM41 4007	NB	GPL570	GSE16476	Neuroblastoma. #8	FAL SE	NA
GSM414008	GSM41 4008	NB	GPL570	GSE16476	Neuroblastoma. #9	FAL SE	NA
GSM414009	GSM41 4009	NB	GPL570	GSE16476	Neuroblastoma. #10	FAL SE	NA
GSM414010	GSM41 4010	NB	GPL570	GSE16476	Neuroblastoma. #11	FAL SE	NA
GSM414011	GSM41 4011	NB	GPL570	GSE16476	Neuroblastoma. #12	FAL SE	NA
GSM414012	GSM41 4012	NB	GPL570	GSE16476	Neuroblastoma. #13	FAL SE	NA
GSM414013	GSM41 4013	NB	GPL570	GSE16476	Neuroblastoma. #14	FAL SE	NA
GSM414014	GSM41 4014	NB	GPL570	GSE16476	Neuroblastoma. #15	FAL SE	NA
GSM414015	GSM41 4015	NB	GPL570	GSE16476	Neuroblastoma. #16	FAL SE	NA
GSM414016	GSM41 4016	NB	GPL570	GSE16476	Neuroblastoma. #17	FAL SE	NA
GSM414017	GSM41 4017	NB	GPL570	GSE16476	Neuroblastoma. #18	FAL SE	NA
GSM414018	GSM41 4018	NB	GPL570	GSE16476	Neuroblastoma. #19	FAL SE	NA
GSM414019	GSM41 4019	NB	GPL570	GSE16476	Neuroblastoma. #20	FAL SE	NA
GSM414020	GSM41 4020	NB	GPL570	GSE16476	Neuroblastoma. #21	FAL SE	NA
GSM414021	GSM41 4021	NB	GPL570	GSE16476	Neuroblastoma. #22	FAL SE	NA
GSM414022	GSM41 4022	NB	GPL570	GSE16476	Neuroblastoma. #23	FAL SE	NA
GSM414023	GSM41 4023	NB	GPL570	GSE16476	Neuroblastoma. #24	FAL SE	NA
GSM414024	GSM41 4024	NB	GPL570	GSE16476	Neuroblastoma. #25	FAL SE	NA
GSM414025	GSM41 4025	NB	GPL570	GSE16476	Neuroblastoma. #26	FAL SE	NA
GSM414026	GSM41 4026	NB	GPL570	GSE16476	Neuroblastoma. #27	FAL SE	NA
GSM414027	GSM41 4027	NB	GPL570	GSE16476	Neuroblastoma. #28	FAL SE	NA
GSM414028	GSM41 4028	NB	GPL570	GSE16476	Neuroblastoma. #29	FAL SE	NA
GSM414029	GSM41 4029	NB	GPL570	GSE16476	Neuroblastoma. #30	FAL SE	NA
GSM414030	GSM41 4030	NB	GPL570	GSE16476	Neuroblastoma. #31	FAL SE	NA
GSM414031	GSM41 4031	NB	GPL570	GSE16476	Neuroblastoma. #32	FAL SE	NA

GSM414032	GSM41 4032	NB	GPL570	GSE16476	Neuroblastoma. #33	FAL SE	NA
GSM414033	GSM41 4033	NB	GPL570	GSE16476	Neuroblastoma. #34	FAL SE	NA
GSM414034	GSM41 4034	NB	GPL570	GSE16476	Neuroblastoma. #35	FAL SE	NA
GSM414035	GSM41 4035	NB	GPL570	GSE16476	Neuroblastoma. #36	FAL SE	NA
GSM414036	GSM41 4036	NB	GPL570	GSE16476	Neuroblastoma. #37	FAL SE	NA
GSM414037	GSM41 4037	NB	GPL570	GSE16476	Neuroblastoma. #38	FAL SE	NA
GSM414038	GSM41 4038	NB	GPL570	GSE16476	Neuroblastoma. #39	FAL SE	NA
GSM414039	GSM41 4039	NB	GPL570	GSE16476	Neuroblastoma. #40	FAL SE	NA
GSM414040	GSM41 4040	NB	GPL570	GSE16476	Neuroblastoma. #41	FAL SE	NA
GSM414041	GSM41 4041	NB	GPL570	GSE16476	Neuroblastoma. #42	FAL SE	NA
GSM414042	GSM41 4042	NB	GPL570	GSE16476	Neuroblastoma. #43	FAL SE	NA
GSM414043	GSM41 4043	NB	GPL570	GSE16476	Neuroblastoma. #44	FAL SE	NA
GSM414044	GSM41 4044	NB	GPL570	GSE16476	Neuroblastoma. #45	FAL SE	NA
GSM414045	GSM41 4045	NB	GPL570	GSE16476	Neuroblastoma. #46	FAL SE	NA
GSM414046	GSM41 4046	NB	GPL570	GSE16476	Neuroblastoma. #47	FAL SE	NA
GSM414047	GSM41 4047	NB	GPL570	GSE16476	Neuroblastoma. #48	FAL SE	NA
GSM414048	GSM41 4048	NB	GPL570	GSE16476	Neuroblastoma. #49	FAL SE	NA
GSM414049	GSM41 4049	NB	GPL570	GSE16476	Neuroblastoma. #50	FAL SE	NA
GSM414050	GSM41 4050	NB	GPL570	GSE16476	Neuroblastoma. #51	FAL SE	NA
GSM414051	GSM41 4051	NB	GPL570	GSE16476	Neuroblastoma. #52	FAL SE	NA
GSM414052	GSM41 4052	NB	GPL570	GSE16476	Neuroblastoma. #53	FAL SE	NA
GSM414053	GSM41 4053	NB	GPL570	GSE16476	Neuroblastoma. #54	FAL SE	NA
GSM414054	GSM41 4054	NB	GPL570	GSE16476	Neuroblastoma. #55	FAL SE	NA
GSM414055	GSM41 4055	NB	GPL570	GSE16476	Neuroblastoma. #56	FAL SE	NA
GSM414056	GSM41 4056	NB	GPL570	GSE16476	Neuroblastoma. #57	FAL SE	NA
GSM414057	GSM41 4057	NB	GPL570	GSE16476	Neuroblastoma. #58	FAL SE	NA
GSM414058	GSM41 4058	NB	GPL570	GSE16476	Neuroblastoma. #59	FAL SE	NA
GSM414059	GSM41 4059	NB	GPL570	GSE16476	Neuroblastoma. #60	FAL SE	NA
GSM414060	GSM41 4060	NB	GPL570	GSE16476	Neuroblastoma. #61	FAL SE	NA
GSM414061	GSM41 4061	NB	GPL570	GSE16476	Neuroblastoma. #62	FAL SE	NA
GSM414062	GSM41 4062	NB	GPL570	GSE16476	Neuroblastoma. #63	FAL SE	NA
GSM414063	GSM41 4063	NB	GPL570	GSE16476	Neuroblastoma. #64	FAL SE	NA

GSM414064	GSM41 4064	NB	GPL570	GSE16476	Neuroblastoma. #65	FAL SE	NA
GSM414065	GSM41 4065	NB	GPL570	GSE16476	Neuroblastoma. #66	FAL SE	NA
GSM414066	GSM41 4066	NB	GPL570	GSE16476	Neuroblastoma. #67	FAL SE	NA
GSM414067	GSM41 4067	NB	GPL570	GSE16476	Neuroblastoma. #68	FAL SE	NA
GSM414068	GSM41 4068	NB	GPL570	GSE16476	Neuroblastoma. #69	FAL SE	NA
GSM414069	GSM41 4069	NB	GPL570	GSE16476	Neuroblastoma. #70	FAL SE	NA
GSM414070	GSM41 4070	NB	GPL570	GSE16476	Neuroblastoma. #71	FAL SE	NA
GSM414071	GSM41 4071	NB	GPL570	GSE16476	Neuroblastoma. #72	FAL SE	NA
GSM414072	GSM41 4072	NB	GPL570	GSE16476	Neuroblastoma. #73	FAL SE	NA
GSM414073	GSM41 4073	NB	GPL570	GSE16476	Neuroblastoma. #74	FAL SE	NA
GSM414074	GSM41 4074	NB	GPL570	GSE16476	Neuroblastoma. #75	FAL SE	NA
GSM414075	GSM41 4075	NB	GPL570	GSE16476	Neuroblastoma. #76	FAL SE	NA
GSM414076	GSM41 4076	NB	GPL570	GSE16476	Neuroblastoma. #77	FAL SE	NA
GSM414077	GSM41 4077	NB	GPL570	GSE16476	Neuroblastoma. #78	FAL SE	NA
GSM414078	GSM41 4078	NB	GPL570	GSE16476	Neuroblastoma. #79	FAL SE	NA
GSM414079	GSM41 4079	NB	GPL570	GSE16476	Neuroblastoma. #80	FAL SE	NA
GSM414080	GSM41 4080	NB	GPL570	GSE16476	Neuroblastoma. #81	FAL SE	NA
GSM414081	GSM41 4081	NB	GPL570	GSE16476	Neuroblastoma. #82	FAL SE	NA
GSM414082	GSM41 4082	NB	GPL570	GSE16476	Neuroblastoma. #83	FAL SE	NA
GSM414083	GSM41 4083	NB	GPL570	GSE16476	Neuroblastoma. #84	FAL SE	NA
GSM414084	GSM41 4084	NB	GPL570	GSE16476	Neuroblastoma. #85	FAL SE	NA
GSM414085	GSM41 4085	NB	GPL570	GSE16476	Neuroblastoma. #86	FAL SE	NA
GSM414086	GSM41 4086	NB	GPL570	GSE16476	Neuroblastoma. #87	FAL SE	NA
GSM414087	GSM41 4087	NB	GPL570	GSE16476	Neuroblastoma. #88	FAL SE	NA
GSM692982	GSM69 2982	ATRT	GPL570	GSE28026	AT/RT tumor sample ID03161	TRU E	Duplic ate
GSM692983	GSM69 2983	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00003	TRU E	Duplic ate
GSM692984	GSM69 2984	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00119	TRU E	Duplic ate
GSM692985	GSM69 2985	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00343	TRU E	Duplic ate
GSM692986	GSM69 2986	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00370	TRU E	Duplic ate
GSM692987	GSM69 2987	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00404	TRU E	Duplic ate
GSM692988	GSM69 2988	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00413	TRU E	Duplic ate
GSM692989	GSM69 2989	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00504	TRU E	Duplic ate

GSM692990	GSM692990	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00514	TRUE	Duplicate
GSM692991	GSM692991	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00515	TRUE	Duplicate
GSM692992	GSM692992	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00517	TRUE	Duplicate
GSM692993	GSM692993	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00605	TRUE	Duplicate
GSM692994	GSM692994	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00663	TRUE	Duplicate
GSM692995	GSM692995	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00687	TRUE	Duplicate
GSM692996	GSM692996	ATRT	GPL570	GSE28026	AT/RT tumor sample ID00737	TRUE	Duplicate
GSM692997	GSM692997	ATRT	GPL570	GSE28026	AT/RT tumor sample ID90004	TRUE	Duplicate
GSM692998	GSM692998	ATRT	GPL570	GSE28026	AT/RT tumor sample ID90005	TRUE	Duplicate
GSM692999	GSM692999	ATRT	GPL570	GSE28026	AT/RT tumor sample ID90007	TRUE	Duplicate
GSM852011	GSM852011	EWS	GPL570	GSE34620	Ewing sarcoma sample EW054	TRUE	Clustering
GSM852012	GSM852012	EWS	GPL570	GSE34620	Ewing sarcoma sample EW059	FALSE	NA
GSM852013	GSM852013	EWS	GPL570	GSE34620	Ewing sarcoma sample EW066	FALSE	NA
GSM852014	GSM852014	EWS	GPL570	GSE34620	Ewing sarcoma sample EW069	FALSE	NA
GSM852015	GSM852015	EWS	GPL570	GSE34620	Ewing sarcoma sample EW070	FALSE	NA
GSM852016	GSM852016	EWS	GPL570	GSE34620	Ewing sarcoma sample EW071	FALSE	NA
GSM852017	GSM852017	EWS	GPL570	GSE34620	Ewing sarcoma sample EW076	FALSE	NA
GSM852018	GSM852018	EWS	GPL570	GSE34620	Ewing sarcoma sample EW080	FALSE	NA
GSM852019	GSM852019	EWS	GPL570	GSE34620	Ewing sarcoma sample EW082	FALSE	NA
GSM852020	GSM852020	EWS	GPL570	GSE34620	Ewing sarcoma sample EW086	FALSE	NA
GSM852021	GSM852021	EWS	GPL570	GSE34620	Ewing sarcoma sample EW088	FALSE	NA
GSM852022	GSM852022	EWS	GPL570	GSE34620	Ewing sarcoma sample EW089	FALSE	NA
GSM852023	GSM852023	EWS	GPL570	GSE34620	Ewing sarcoma sample EW090	FALSE	NA
GSM852024	GSM852024	EWS	GPL570	GSE34620	Ewing sarcoma sample EW092	FALSE	NA
GSM852025	GSM852025	EWS	GPL570	GSE34620	Ewing sarcoma sample EW093	FALSE	NA
GSM852026	GSM852026	EWS	GPL570	GSE34620	Ewing sarcoma sample EW095	FALSE	NA
GSM852027	GSM852027	EWS	GPL570	GSE34620	Ewing sarcoma sample EW096	TRUE	Clustering
GSM852028	GSM852028	EWS	GPL570	GSE34620	Ewing sarcoma sample EW098	FALSE	NA
GSM852029	GSM852029	EWS	GPL570	GSE34620	Ewing sarcoma sample EW105	FALSE	NA
GSM852030	GSM852030	EWS	GPL570	GSE34620	Ewing sarcoma sample EW106	FALSE	NA
GSM852031	GSM852031	EWS	GPL570	GSE34620	Ewing sarcoma sample EW108	FALSE	NA
GSM852032	GSM852032	EWS	GPL570	GSE34620	Ewing sarcoma sample EW111	FALSE	NA

GSM852033	GSM85 2033	EWS	GPL570	GSE34620	Ewing sarcoma sample EW112	FAL SE	NA
GSM852034	GSM85 2034	EWS	GPL570	GSE34620	Ewing sarcoma sample EW116	FAL SE	NA
GSM852035	GSM85 2035	EWS	GPL570	GSE34620	Ewing sarcoma sample EW117b	FAL SE	NA
GSM852036	GSM85 2036	EWS	GPL570	GSE34620	Ewing sarcoma sample EW121	FAL SE	NA
GSM852037	GSM85 2037	EWS	GPL570	GSE34620	Ewing sarcoma sample EW123	FAL SE	NA
GSM852038	GSM85 2038	EWS	GPL570	GSE34620	Ewing sarcoma sample EW127	FAL SE	NA
GSM852039	GSM85 2039	EWS	GPL570	GSE34620	Ewing sarcoma sample EW128	FAL SE	NA
GSM852040	GSM85 2040	EWS	GPL570	GSE34620	Ewing sarcoma sample EW132	FAL SE	NA
GSM852041	GSM85 2041	EWS	GPL570	GSE34620	Ewing sarcoma sample EW135	TRU E	Cluste ring
GSM852042	GSM85 2042	EWS	GPL570	GSE34620	Ewing sarcoma sample EW139	FAL SE	NA
GSM852043	GSM85 2043	EWS	GPL570	GSE34620	Ewing sarcoma sample EW141	FAL SE	NA
GSM852044	GSM85 2044	EWS	GPL570	GSE34620	Ewing sarcoma sample EW143	FAL SE	NA
GSM852045	GSM85 2045	EWS	GPL570	GSE34620	Ewing sarcoma sample EW149	FAL SE	NA
GSM852046	GSM85 2046	EWS	GPL570	GSE34620	Ewing sarcoma sample EW153b	FAL SE	NA
GSM852047	GSM85 2047	EWS	GPL570	GSE34620	Ewing sarcoma sample EW155	FAL SE	NA
GSM852048	GSM85 2048	EWS	GPL570	GSE34620	Ewing sarcoma sample EW158	FAL SE	NA
GSM852049	GSM85 2049	EWS	GPL570	GSE34620	Ewing sarcoma sample EW159	FAL SE	NA
GSM852050	GSM85 2050	EWS	GPL570	GSE34620	Ewing sarcoma sample EW162	FAL SE	NA
GSM852051	GSM85 2051	EWS	GPL570	GSE34620	Ewing sarcoma sample EW165	FAL SE	NA
GSM852052	GSM85 2052	EWS	GPL570	GSE34620	Ewing sarcoma sample EW167	FAL SE	NA
GSM852053	GSM85 2053	EWS	GPL570	GSE34620	Ewing sarcoma sample EW168	FAL SE	NA
GSM852054	GSM85 2054	EWS	GPL570	GSE34620	Ewing sarcoma sample EW174	FAL SE	NA
GSM852055	GSM85 2055	EWS	GPL570	GSE34620	Ewing sarcoma sample EW195	FAL SE	NA
GSM852056	GSM85 2056	EWS	GPL570	GSE34620	Ewing sarcoma sample EW196	FAL SE	NA
GSM852057	GSM85 2057	EWS	GPL570	GSE34620	Ewing sarcoma sample EW197	FAL SE	NA
GSM852058	GSM85 2058	EWS	GPL570	GSE34620	Ewing sarcoma sample EW198	FAL SE	NA
GSM852059	GSM85 2059	EWS	GPL570	GSE34620	Ewing sarcoma sample EW200	FAL SE	NA
GSM852060	GSM85 2060	EWS	GPL570	GSE34620	Ewing sarcoma sample EW207	TRU E	Cluste ring
GSM852061	GSM85 2061	EWS	GPL570	GSE34620	Ewing sarcoma sample EW215	TRU E	Cluste ring
GSM852062	GSM85 2062	EWS	GPL570	GSE34620	Ewing sarcoma sample EW218	FAL SE	NA
GSM852063	GSM85 2063	EWS	GPL570	GSE34620	Ewing sarcoma sample EW220	FAL SE	NA
GSM852064	GSM85 2064	EWS	GPL570	GSE34620	Ewing sarcoma sample EW224b	FAL SE	NA

GSM852065	GSM852065	EWS	GPL570	GSE34620	Ewing sarcoma sample EW227b	FAL SE	NA
GSM852066	GSM852066	EWS	GPL570	GSE34620	Ewing sarcoma sample EW231	FAL SE	NA
GSM852067	GSM852067	EWS	GPL570	GSE34620	Ewing sarcoma sample EW236	TRU E	Clustering
GSM852068	GSM852068	EWS	GPL570	GSE34620	Ewing sarcoma sample EW237	FAL SE	NA
GSM852069	GSM852069	EWS	GPL570	GSE34620	Ewing sarcoma sample EW241	FAL SE	NA
GSM852070	GSM852070	EWS	GPL570	GSE34620	Ewing sarcoma sample EW242	FAL SE	NA
GSM852071	GSM852071	EWS	GPL570	GSE34620	Ewing sarcoma sample EW247	TRU E	Clustering
GSM852072	GSM852072	EWS	GPL570	GSE34620	Ewing sarcoma sample EW248	TRU E	Clustering
GSM852073	GSM852073	EWS	GPL570	GSE34620	Ewing sarcoma sample EW250	TRU E	Clustering
GSM852074	GSM852074	EWS	GPL570	GSE34620	Ewing sarcoma sample EW251	TRU E	Clustering
GSM852075	GSM852075	EWS	GPL570	GSE34620	Ewing sarcoma sample EW252b	FAL SE	NA
GSM852076	GSM852076	EWS	GPL570	GSE34620	Ewing sarcoma sample EW257b	FAL SE	NA
GSM852077	GSM852077	EWS	GPL570	GSE34620	Ewing sarcoma sample EW258	FAL SE	NA
GSM852078	GSM852078	EWS	GPL570	GSE34620	Ewing sarcoma sample EW278	FAL SE	NA
GSM852079	GSM852079	EWS	GPL570	GSE34620	Ewing sarcoma sample EW279	FAL SE	NA
GSM852080	GSM852080	EWS	GPL570	GSE34620	Ewing sarcoma sample EW284	FAL SE	NA
GSM852081	GSM852081	EWS	GPL570	GSE34620	Ewing sarcoma sample EW288	FAL SE	NA
GSM852082	GSM852082	EWS	GPL570	GSE34620	Ewing sarcoma sample EW294	FAL SE	NA
GSM852083	GSM852083	EWS	GPL570	GSE34620	Ewing sarcoma sample EW296	FAL SE	NA
GSM852084	GSM852084	EWS	GPL570	GSE34620	Ewing sarcoma sample EW297	FAL SE	NA
GSM852085	GSM852085	EWS	GPL570	GSE34620	Ewing sarcoma sample EW298	FAL SE	NA
GSM852086	GSM852086	EWS	GPL570	GSE34620	Ewing sarcoma sample EW299	FAL SE	NA
GSM852087	GSM852087	EWS	GPL570	GSE34620	Ewing sarcoma sample EW303	FAL SE	NA
GSM852088	GSM852088	EWS	GPL570	GSE34620	Ewing sarcoma sample EW306	FAL SE	NA
GSM852089	GSM852089	EWS	GPL570	GSE34620	Ewing sarcoma sample EW308	TRU E	Clustering
GSM852090	GSM852090	EWS	GPL570	GSE34620	Ewing sarcoma sample EW309	FAL SE	NA
GSM852091	GSM852091	EWS	GPL570	GSE34620	Ewing sarcoma sample EW338	TRU E	Clustering
GSM852092	GSM852092	EWS	GPL570	GSE34620	Ewing sarcoma sample EW340	TRU E	Clustering
GSM852093	GSM852093	EWS	GPL570	GSE34620	Ewing sarcoma sample EW341	FAL SE	NA
GSM852094	GSM852094	EWS	GPL570	GSE34620	Ewing sarcoma sample EW343	FAL SE	NA
GSM852095	GSM852095	EWS	GPL570	GSE34620	Ewing sarcoma sample EW369	FAL SE	NA
GSM852096	GSM852096	EWS	GPL570	GSE34620	Ewing sarcoma sample EW377	FAL SE	NA

GSM852097	GSM852097	EWS	GPL570	GSE34620	Ewing sarcoma sample EW378	FALSE	NA
GSM852098	GSM852098	EWS	GPL570	GSE34620	Ewing sarcoma sample EW379	FALSE	NA
GSM852099	GSM852099	EWS	GPL570	GSE34620	Ewing sarcoma sample EW381	FALSE	NA
GSM852100	GSM852100	EWS	GPL570	GSE34620	Ewing sarcoma sample EW517	FALSE	NA
GSM852101	GSM852101	EWS	GPL570	GSE34620	Ewing sarcoma sample EW525b	FALSE	NA
GSM852102	GSM852102	EWS	GPL570	GSE34620	Ewing sarcoma sample EW531	FALSE	NA
GSM852103	GSM852103	EWS	GPL570	GSE34620	Ewing sarcoma sample EW532	FALSE	NA
GSM852104	GSM852104	EWS	GPL570	GSE34620	Ewing sarcoma sample EW554	FALSE	NA
GSM852105	GSM852105	EWS	GPL570	GSE34620	Ewing sarcoma sample EW556	FALSE	NA
GSM852106	GSM852106	EWS	GPL570	GSE34620	Ewing sarcoma sample EW557	FALSE	NA
GSM852107	GSM852107	EWS	GPL570	GSE34620	Ewing sarcoma sample EW563	FALSE	NA
GSM852108	GSM852108	EWS	GPL570	GSE34620	Ewing sarcoma sample EW576	FALSE	NA
GSM852109	GSM852109	EWS	GPL570	GSE34620	Ewing sarcoma sample EW577	TRUE	Clustering
GSM852110	GSM852110	EWS	GPL570	GSE34620	Ewing sarcoma sample EW579	FALSE	NA
GSM852111	GSM852111	EWS	GPL570	GSE34620	Ewing sarcoma sample EW582	FALSE	NA
GSM852112	GSM852112	EWS	GPL570	GSE34620	Ewing sarcoma sample EW585	FALSE	NA
GSM852113	GSM852113	EWS	GPL570	GSE34620	Ewing sarcoma sample EW587	FALSE	NA
GSM852114	GSM852114	EWS	GPL570	GSE34620	Ewing sarcoma sample EW604	FALSE	NA
GSM852115	GSM852115	EWS	GPL570	GSE34620	Ewing sarcoma sample EW608	FALSE	NA
GSM852116	GSM852116	EWS	GPL570	GSE34620	Ewing sarcoma sample EW612	FALSE	NA
GSM852117	GSM852117	EWS	GPL570	GSE34620	Ewing sarcoma sample EW613	FALSE	NA
GSM852118	GSM852118	EWS	GPL570	GSE34620	Ewing sarcoma sample EW616	FALSE	NA
GSM852119	GSM852119	EWS	GPL570	GSE34620	Ewing sarcoma sample EW652	FALSE	NA
GSM852120	GSM852120	EWS	GPL570	GSE34620	Ewing sarcoma sample EW658	FALSE	NA
GSM852121	GSM852121	EWS	GPL570	GSE34620	Ewing sarcoma sample EW661	FALSE	NA
GSM852122	GSM852122	EWS	GPL570	GSE34620	Ewing sarcoma sample EW662	FALSE	NA
GSM852123	GSM852123	EWS	GPL570	GSE34620	Ewing sarcoma sample EW665	FALSE	NA
GSM852124	GSM852124	EWS	GPL570	GSE34620	Ewing sarcoma sample EW666	FALSE	NA
GSM852125	GSM852125	EWS	GPL570	GSE34620	Ewing sarcoma sample EW667	FALSE	NA
GSM852126	GSM852126	EWS	GPL570	GSE34620	Ewing sarcoma sample EW668	FALSE	NA
GSM852127	GSM852127	EWS	GPL570	GSE34620	Ewing sarcoma sample EW669	FALSE	NA
GSM869617	GSM869617	ATRT	GPL570	GSE35493	AT/RT. ID03161	FALSE	NA

GSM869618	GSM86 9618	ATRT	GPL570	GSE35493	AT/RT. ID00003	FAL SE	NA
GSM869619	GSM86 9619	ATRT	GPL570	GSE35493	AT/RT. ID00119	FAL SE	NA
GSM869620	GSM86 9620	ATRT	GPL570	GSE35493	AT/RT. ID00343	FAL SE	NA
GSM869621	GSM86 9621	ATRT	GPL570	GSE35493	AT/RT. ID00370	FAL SE	NA
GSM869622	GSM86 9622	ATRT	GPL570	GSE35493	AT/RT. ID00404	FAL SE	NA
GSM869623	GSM86 9623	ATRT	GPL570	GSE35493	AT/RT. ID00413	FAL SE	NA
GSM869624	GSM86 9624	ATRT	GPL570	GSE35493	AT/RT. ID00504	FAL SE	NA
GSM869625	GSM86 9625	ATRT	GPL570	GSE35493	AT/RT. ID00514	FAL SE	NA
GSM869626	GSM86 9626	ATRT	GPL570	GSE35493	AT/RT. ID00515	FAL SE	NA
GSM869627	GSM86 9627	ATRT	GPL570	GSE35493	AT/RT. ID00517	FAL SE	NA
GSM869628	GSM86 9628	ATRT	GPL570	GSE35493	AT/RT. ID00605	FAL SE	NA
GSM869629	GSM86 9629	ATRT	GPL570	GSE35493	AT/RT. ID00663	FAL SE	NA
GSM869630	GSM86 9630	ATRT	GPL570	GSE35493	AT/RT. ID00687	FAL SE	NA
GSM869631	GSM86 9631	ATRT	GPL570	GSE35493	AT/RT. ID00737	FAL SE	NA
GSM869632	GSM86 9632	ATRT	GPL570	GSE35493	AT/RT. ID90004	FAL SE	NA
GSM869633	GSM86 9633	ATRT	GPL570	GSE35493	AT/RT. ID90005	FAL SE	NA
GSM869634	GSM86 9634	ATRT	GPL570	GSE35493	AT/RT. ID90007	FAL SE	NA
GSM869683	GSM86 9683	MB	GPL570	GSE35493	MED. ID00231	TRU E	Cluste ring
GSM869684	GSM86 9684	MB	GPL570	GSE35493	MED. ID00241	TRU E	Cluste ring
GSM869685	GSM86 9685	MB	GPL570	GSE35493	MED. ID00401	TRU E	Cluste ring
GSM869686	GSM86 9686	MB	GPL570	GSE35493	MED. ID00613	TRU E	Cluste ring
GSM869688	GSM86 9688	MB	GPL570	GSE35493	MED. ID00186	TRU E	Cluste ring
GSM869689	GSM86 9689	MB	GPL570	GSE35493	MED. ID00254	TRU E	Cluste ring
GSM869690	GSM86 9690	MB	GPL570	GSE35493	MED. ID00258	TRU E	Cluste ring
GSM869691	GSM86 9691	MB	GPL570	GSE35493	MED. ID00262	TRU E	Cluste ring
GSM869692	GSM86 9692	MB	GPL570	GSE35493	MED. ID00277	TRU E	Cluste ring
GSM869693	GSM86 9693	MB	GPL570	GSE35493	MED. ID00288	TRU E	Cluste ring
GSM869694	GSM86 9694	MB	GPL570	GSE35493	MED. ID00330	TRU E	Cluste ring
GSM869695	GSM86 9695	MB	GPL570	GSE35493	MED. ID00437	TRU E	Cluste ring
GSM869696	GSM86 9696	MB	GPL570	GSE35493	MED. ID00565	TRU E	Cluste ring
GSM869700	GSM86 9700	MB	GPL570	GSE35493	MED. ID00797	TRU E	Cluste ring
GSM982838	GSM98 2838	ATRT	GPL570	GSE35493	ATRT. ID00818	FAL SE	NA

GSM982839	GSM982839	ATRT	GPL570	GSE35493	ATRT. ID00880	FAL SE	NA
GSM982840	GSM982840	MB	GPL570	GSE35493	MED. ID00801	TRU E	Clustering
GSM982841	GSM982841	MB	GPL570	GSE35493	MED_ID00851	TRU E	Clustering
GSM982842	GSM982842	MB	GPL570	GSE35493	MED_ID00529	TRU E	Clustering
GSM982843	GSM982843	MB	GPL570	GSE35493	MED_ID00719	TRU E	Clustering
GSM982844	GSM982844	MB	GPL570	GSE35493	MED_ID00791	TRU E	Clustering
GSM982845	GSM982845	MB	GPL570	GSE35493	MED_ID00877	TRU E	Clustering
GSM982846	GSM982846	MB	GPL570	GSE35493	MED_ID00898	TRU E	Clustering
GSM918578	GSM918578	MB	GPL570	GSE37418	SJMB018	FAL SE	NA
GSM918579	GSM918579	MB	GPL570	GSE37418	SJMB043	FAL SE	NA
GSM918580	GSM918580	MB	GPL570	GSE37418	SJMB001	FAL SE	NA
GSM918581	GSM918581	MB	GPL570	GSE37418	SJMB014	FAL SE	NA
GSM918582	GSM918582	MB	GPL570	GSE37418	SJMB012	FAL SE	NA
GSM918583	GSM918583	MB	GPL570	GSE37418	SJMB015	FAL SE	NA
GSM918584	GSM918584	MB	GPL570	GSE37418	SJMB019	FAL SE	NA
GSM918585	GSM918585	MB	GPL570	GSE37418	SJMB045	FAL SE	NA
GSM918586	GSM918586	MB	GPL570	GSE37418	SJMB017	FAL SE	NA
GSM918587	GSM918587	MB	GPL570	GSE37418	SJMB048	FAL SE	NA
GSM918588	GSM918588	MB	GPL570	GSE37418	SJMB097	FAL SE	NA
GSM918589	GSM918589	MB	GPL570	GSE37418	SJMB050	FAL SE	NA
GSM918590	GSM918590	MB	GPL570	GSE37418	SJMB020	FAL SE	NA
GSM918591	GSM918591	MB	GPL570	GSE37418	SJMB009	FAL SE	NA
GSM918592	GSM918592	MB	GPL570	GSE37418	SJMB084	FAL SE	NA
GSM918593	GSM918593	MB	GPL570	GSE37418	SJMB051	FAL SE	NA
GSM918594	GSM918594	MB	GPL570	GSE37418	SJMB052	FAL SE	NA
GSM918595	GSM918595	MB	GPL570	GSE37418	SJMB016	FAL SE	NA
GSM918596	GSM918596	MB	GPL570	GSE37418	SJMB021	FAL SE	NA
GSM918597	GSM918597	MB	GPL570	GSE37418	SJMB023	FAL SE	NA
GSM918598	GSM918598	MB	GPL570	GSE37418	SJMB101	FAL SE	NA
GSM918599	GSM918599	MB	GPL570	GSE37418	SJMB024	FAL SE	NA
GSM918600	GSM918600	MB	GPL570	GSE37418	SJMB025	FAL SE	NA
GSM918601	GSM918601	MB	GPL570	GSE37418	SJMB006	FAL SE	NA

GSM918602	GSM91 8602	MB	GPL570	GSE37418	SJMB104	FAL SE	NA
GSM918603	GSM91 8603	MB	GPL570	GSE37418	SJMB028	FAL SE	NA
GSM918604	GSM91 8604	MB	GPL570	GSE37418	SJMB063	FAL SE	NA
GSM918605	GSM91 8605	MB	GPL570	GSE37418	SJMB064	FAL SE	NA
GSM918606	GSM91 8606	MB	GPL570	GSE37418	SJMB107	FAL SE	NA
GSM918607	GSM91 8607	MB	GPL570	GSE37418	SJMB067	FAL SE	NA
GSM918608	GSM91 8608	MB	GPL570	GSE37418	SJMB068	FAL SE	NA
GSM918609	GSM91 8609	MB	GPL570	GSE37418	SJMB055	FAL SE	NA
GSM918610	GSM91 8610	MB	GPL570	GSE37418	SJMB070	FAL SE	NA
GSM918611	GSM91 8611	MB	GPL570	GSE37418	SJMB030	FAL SE	NA
GSM918612	GSM91 8612	MB	GPL570	GSE37418	SJMB087	FAL SE	NA
GSM918613	GSM91 8613	MB	GPL570	GSE37418	SJMB088	FAL SE	NA
GSM918614	GSM91 8614	MB	GPL570	GSE37418	SJMB031	FAL SE	NA
GSM918615	GSM91 8615	MB	GPL570	GSE37418	SJMB091	FAL SE	NA
GSM918616	GSM91 8616	MB	GPL570	GSE37418	SJMB093	FAL SE	NA
GSM918617	GSM91 8617	MB	GPL570	GSE37418	SJMB032	FAL SE	NA
GSM918618	GSM91 8618	MB	GPL570	GSE37418	SJMB073	TRU E	Cluste ring
GSM918619	GSM91 8619	MB	GPL570	GSE37418	SJMB033	FAL SE	NA
GSM918620	GSM91 8620	MB	GPL570	GSE37418	SJMB003	FAL SE	NA
GSM918621	GSM91 8621	MB	GPL570	GSE37418	SJMB004	FAL SE	NA
GSM918622	GSM91 8622	MB	GPL570	GSE37418	SJMB010	FAL SE	NA
GSM918623	GSM91 8623	MB	GPL570	GSE37418	SJMB011	FAL SE	NA
GSM918624	GSM91 8624	MB	GPL570	GSE37418	SJMB094	FAL SE	NA
GSM918625	GSM91 8625	MB	GPL570	GSE37418	SJMB114	FAL SE	NA
GSM918626	GSM91 8626	MB	GPL570	GSE37418	SJMB035	FAL SE	NA
GSM918627	GSM91 8627	MB	GPL570	GSE37418	SJMB117	FAL SE	NA
GSM918628	GSM91 8628	MB	GPL570	GSE37418	SJMB118	TRU E	Cluste ring
GSM918629	GSM91 8629	MB	GPL570	GSE37418	SJMB008	FAL SE	NA
GSM918630	GSM91 8630	MB	GPL570	GSE37418	SJMB036	FAL SE	NA
GSM918631	GSM91 8631	MB	GPL570	GSE37418	SJMB079	FAL SE	NA
GSM918632	GSM91 8632	MB	GPL570	GSE37418	SJMB078	FAL SE	NA
GSM918633	GSM91 8633	MB	GPL570	GSE37418	SJMB038	FAL SE	NA

GSM918634	GSM91 8634	MB	GPL570	GSE37418	SJMB039	FAL SE	NA
GSM918635	GSM91 8635	MB	GPL570	GSE37418	SJMB080	FAL SE	NA
GSM918636	GSM91 8636	MB	GPL570	GSE37418	SJMB081	FAL SE	NA
GSM918637	GSM91 8637	MB	GPL570	GSE37418	SJMB082	FAL SE	NA
GSM918638	GSM91 8638	MB	GPL570	GSE37418	SJMB040	FAL SE	NA
GSM918639	GSM91 8639	MB	GPL570	GSE37418	SJMB122	FAL SE	NA
GSM918640	GSM91 8640	MB	GPL570	GSE37418	SJMB124	FAL SE	NA
GSM918641	GSM91 8641	MB	GPL570	GSE37418	SJMB137	FAL SE	NA
GSM918642	GSM91 8642	MB	GPL570	GSE37418	SJMB129	FAL SE	NA
GSM918643	GSM91 8643	MB	GPL570	GSE37418	SJMB139	FAL SE	NA
GSM918644	GSM91 8644	MB	GPL570	GSE37418	SJMB131	FAL SE	NA
GSM918645	GSM91 8645	MB	GPL570	GSE37418	SJMB085	FAL SE	NA
GSM918646	GSM91 8646	MB	GPL570	GSE37418	SJMB027	FAL SE	NA
GSM918647	GSM91 8647	MB	GPL570	GSE37418	SJMB069	FAL SE	NA
GSM918648	GSM91 8648	MB	GPL570	GSE37418	SJMB071	FAL SE	NA
GSM918649	GSM91 8649	MB	GPL570	GSE37418	SJMB089	FAL SE	NA
GSM918650	GSM91 8650	MB	GPL570	GSE37418	SJMB090	FAL SE	NA
GSM918651	GSM91 8651	MB	GPL570	GSE37418	SJMB013	FAL SE	NA
GSM918652	GSM91 8652	MB	GPL570	GSE37418	SJMB034	FAL SE	NA
GSM918653	GSM91 8653	MB	GPL570	GSE37418	SJMB037	FAL SE	NA
GSM1562889	GSM15 62889	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_._INI109)	FAL SE	NA
GSM1562890	GSM15 62890	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_._INI110)	FAL SE	NA
GSM1562891	GSM15 62891	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_._INI116)	FAL SE	NA
GSM1562892	GSM15 62892	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_._INI119)	FAL SE	NA
GSM1562893	GSM15 62893	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_._INI18)	FAL SE	NA
GSM1562894	GSM15 62894	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_._INI19)	FAL SE	NA
GSM1562895	GSM15 62895	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_._INI22)	FAL SE	NA
GSM1562896	GSM15 62896	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_._INI24)	FAL SE	NA
GSM1562897	GSM15 62897	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_._INI25)	FAL SE	NA
GSM1562898	GSM15 62898	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_._INI28)	FAL SE	NA
GSM1562899	GSM15 62899	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_._INI38)	FAL SE	NA
GSM1562900	GSM15 62900	ECRT	GPL570	GSE64019	Hs_CNS_MRT (hNA_._INI39)	FAL SE	NA

GSM1562901	GSM15 62901	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_.INI50)	FAL SE	NA
GSM1562902	GSM15 62902	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_.INI53b)	FAL SE	NA
GSM1562903	GSM15 62903	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_.INI56)	FAL SE	NA
GSM1562904	GSM15 62904	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_.INI59)	FAL SE	NA
GSM1562905	GSM15 62905	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_.INI90)	FAL SE	NA
GSM1562906	GSM15 62906	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_.INI91)	FAL SE	NA
GSM1562907	GSM15 62907	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_.INI93)	FAL SE	NA
GSM1562908	GSM15 62908	ECRT	GPL570	GSE64019	Hs_soft-tissue_MRT (hEC_.INI97)	FAL SE	NA
GSM1562909	GSM15 62909	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI103)	FAL SE	NA
GSM1562910	GSM15 62910	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI104)	FAL SE	NA
GSM1562911	GSM15 62911	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI159)	FAL SE	NA
GSM1562912	GSM15 62912	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI161)	FAL SE	NA
GSM1562913	GSM15 62913	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI162)	FAL SE	NA
GSM1562914	GSM15 62914	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI163)	FAL SE	NA
GSM1562915	GSM15 62915	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI168)	FAL SE	NA
GSM1562916	GSM15 62916	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI169)	FAL SE	NA
GSM1562917	GSM15 62917	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI84)	FAL SE	NA
GSM1562918	GSM15 62918	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI87)	FAL SE	NA
GSM1562919	GSM15 62919	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI88)	FAL SE	NA
GSM1562920	GSM15 62920	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI89)	FAL SE	NA
GSM1562921	GSM15 62921	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC1_.INI96)	FAL SE	NA
GSM1562922	GSM15 62922	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC2_.INI157)	FAL SE	NA
GSM1562923	GSM15 62923	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC2_.INI160)	FAL SE	NA
GSM1562924	GSM15 62924	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC2_.INI164)	FAL SE	NA
GSM1562925	GSM15 62925	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC2_.INI167)	FAL SE	NA
GSM1562926	GSM15 62926	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC2_.INI30)	FAL SE	NA
GSM1562927	GSM15 62927	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC2_.INI51)	FAL SE	NA
GSM1562928	GSM15 62928	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC2_.INI57)	FAL SE	NA
GSM1562929	GSM15 62929	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC2_.INI92)	FAL SE	NA
GSM1562930	GSM15 62930	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC2_.INI94)	FAL SE	NA
GSM1562931	GSM15 62931	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC2_.MB115)	FAL SE	NA
GSM1562932	GSM15 62932	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hiC3_.INI158)	FAL SE	NA

GSM1562933	GSM1562933	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hIC3_..INI165)	FALSE	NA
GSM1562934	GSM1562934	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hIC3_..INI166)	FALSE	NA
GSM1562935	GSM1562935	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hIC3_..INI170)	FALSE	NA
GSM1562936	GSM1562936	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hIC3_..INI99)	FALSE	NA
GSM1562937	GSM1562937	MB	GPL570	GSE64019	Hs_CNS_MB-SHH (hMB_..MB107)	TRUE	Clustering
GSM1562938	GSM1562938	MB	GPL570	GSE64019	Hs_CNS_MB-SHH (hMB_..MB141)	TRUE	Clustering
GSM1562939	GSM1562939	MB	GPL570	GSE64019	Hs_CNS_MB-SHH (hMB_..MB146)	TRUE	Clustering
GSM1562940	GSM1562940	MB	GPL570	GSE64019	Hs_CNS_MB-SHH (hMB_..MB147)	TRUE	Clustering
GSM1562941	GSM1562941	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hNA_..INI108)	FALSE	NA
GSM1562942	GSM1562942	ATRT	GPL570	GSE64019	Hs_CNS_MRT (hNA_..INI27)	FALSE	NA
GSM1587791	GSM1587791	ATRT	GPL570	GSE67851	AT/RT_INI1(-)_patient1	FALSE	NA
GSM1587792	GSM1587792	ATRT	GPL570	GSE67851	AT/RT_INI1(-)_patient2	FALSE	NA
GSM1587793	GSM1587793	ATRT	GPL570	GSE67851	AT/RT_INI1(-)_patient3	FALSE	NA
GSM1587794	GSM1587794	ATRT	GPL570	GSE67851	AT/RT_INI1(-)_patient4	FALSE	NA
GSM1587795	GSM1587795	ATRT	GPL570	GSE67851	AT/RT_INI1(-)_patient5	FALSE	NA
GSM1587796	GSM1587796	ATRT	GPL570	GSE67851	AT/RT_INI1(-)_patient6	FALSE	NA
GSM1587797	GSM1587797	ATRT	GPL570	GSE67851	AT/RT_INI1(-)_patient7	FALSE	NA
GSM1816332	GSM1816332	ATRT	GPL570	GSE70678	dkfz_ATRT_159	FALSE	NA
GSM1816333	GSM1816333	ATRT	GPL570	GSE70678	dkfz_ATRT_17	FALSE	NA
GSM1816334	GSM1816334	ATRT	GPL570	GSE70678	dkfz_ATRT_25	FALSE	NA
GSM1816335	GSM1816335	ATRT	GPL570	GSE70678	dkfz_ATRT_173	FALSE	NA
GSM1816336	GSM1816336	ATRT	GPL570	GSE70678	dkfz_ATRT_151	FALSE	NA
GSM1816337	GSM1816337	ATRT	GPL570	GSE70678	dkfz_ATRT_171	FALSE	NA
GSM1816338	GSM1816338	ATRT	GPL570	GSE70678	dkfz_ATRT_152	FALSE	NA
GSM1816339	GSM1816339	ATRT	GPL570	GSE70678	dkfz_ATRT_172	FALSE	NA
GSM1816340	GSM1816340	ATRT	GPL570	GSE70678	dkfz_ATRT_153	FALSE	NA
GSM1816341	GSM1816341	ATRT	GPL570	GSE70678	dkfz_ATRT_86	FALSE	NA
GSM1816342	GSM1816342	ATRT	GPL570	GSE70678	dkfz_ATRT_10	FALSE	NA
GSM1816343	GSM1816343	ATRT	GPL570	GSE70678	dkfz_ATRT_154	FALSE	NA
GSM1816344	GSM1816344	ATRT	GPL570	GSE70678	dkfz_ATRT_30	FALSE	NA
GSM1816345	GSM1816345	ATRT	GPL570	GSE70678	dkfz_ATRT_36	FALSE	NA
GSM1816346	GSM1816346	ATRT	GPL570	GSE70678	dkfz_ATRT_11	FALSE	NA

GSM1816347	GSM18 16347	ATRT	GPL570	GSE70678	dkfz_ATRT_12	FAL SE	NA
GSM1816348	GSM18 16348	ATRT	GPL570	GSE70678	dkfz_ATRT_41	FAL SE	NA
GSM1816349	GSM18 16349	ATRT	GPL570	GSE70678	dkfz_ATRT_43	FAL SE	NA
GSM1816350	GSM18 16350	ATRT	GPL570	GSE70678	dkfz_ATRT_164	FAL SE	NA
GSM1816351	GSM18 16351	ATRT	GPL570	GSE70678	dkfz_ATRT_16	FAL SE	NA
GSM1816352	GSM18 16352	ATRT	GPL570	GSE70678	dkfz_ATRT_6	FAL SE	NA
GSM1816353	GSM18 16353	ATRT	GPL570	GSE70678	dkfz_ATRT_160	FAL SE	NA
GSM1816354	GSM18 16354	ATRT	GPL570	GSE70678	dkfz_ATRT_7	FAL SE	NA
GSM1816355	GSM18 16355	ATRT	GPL570	GSE70678	dkfz_ATRT_48	FAL SE	NA
GSM1816356	GSM18 16356	ATRT	GPL570	GSE70678	dkfz_ATRT_49	FAL SE	NA
GSM1816357	GSM18 16357	ATRT	GPL570	GSE70678	dkfz_ATRT_51	FAL SE	NA
GSM1816358	GSM18 16358	ATRT	GPL570	GSE70678	dkfz_ATRT_47	FAL SE	NA
GSM1816359	GSM18 16359	ATRT	GPL570	GSE70678	dkfz_ATRT_46	FAL SE	NA
GSM1816360	GSM18 16360	ATRT	GPL570	GSE70678	dkfz_ATRT_101	FAL SE	NA
GSM1816361	GSM18 16361	ATRT	GPL570	GSE70678	dkfz_ATRT_157	FAL SE	NA
GSM1816362	GSM18 16362	ATRT	GPL570	GSE70678	dkfz_ATRT_155	FAL SE	NA
GSM1816363	GSM18 16363	ATRT	GPL570	GSE70678	dkfz_ATRT_168	FAL SE	NA
GSM1816364	GSM18 16364	ATRT	GPL570	GSE70678	dkfz_ATRT_156	FAL SE	NA
GSM1816365	GSM18 16365	ATRT	GPL570	GSE70678	dkfz_ATRT_163	FAL SE	NA
GSM1816366	GSM18 16366	ATRT	GPL570	GSE70678	dkfz_ATRT_162	FAL SE	NA
GSM1816367	GSM18 16367	ATRT	GPL570	GSE70678	dkfz_ATRT_158	FAL SE	NA
GSM1816368	GSM18 16368	ATRT	GPL570	GSE70678	dkfz_ATRT_166	FAL SE	NA
GSM1816369	GSM18 16369	ATRT	GPL570	GSE70678	dkfz_ATRT_169	FAL SE	NA
GSM1816370	GSM18 16370	ATRT	GPL570	GSE70678	dkfz_ATRT_170	FAL SE	NA
GSM1816371	GSM18 16371	ATRT	GPL570	GSE70678	dkfz_ATRT_88	FAL SE	NA
GSM1816372	GSM18 16372	ATRT	GPL570	GSE70678	dkfz_ATRT_72	FAL SE	NA
GSM1816373	GSM18 16373	ATRT	GPL570	GSE70678	dkfz_ATRT_167	FAL SE	NA
GSM1816374	GSM18 16374	ATRT	GPL570	GSE70678	dkfz_ATRT_85	FAL SE	NA
GSM1816375	GSM18 16375	ATRT	GPL570	GSE70678	dkfz_ATRT_84	FAL SE	NA
GSM1816376	GSM18 16376	ATRT	GPL570	GSE70678	dkfz_ATRT_165	FAL SE	NA
GSM1816377	GSM18 16377	ATRT	GPL570	GSE70678	dkfz_ATRT_14	FAL SE	NA
GSM1816378	GSM18 16378	ATRT	GPL570	GSE70678	dkfz_ATRT_32	FAL SE	NA

GSM1816379	GSM1816379	ATRT	GPL570	GSE70678	dkfz_ATRT_161	FAL SE	NA
GSM1816380	GSM1816380	ATRT	GPL570	GSE70678	dkfz_ATRT_90	FAL SE	NA
GSM1881102	GSM1881102	MB	GPL570	GSE73038	dkfz_MB_15-0040	FAL SE	NA
GSM1881103	GSM1881103	MB	GPL570	GSE73038	dkfz_MB_15-0041	FAL SE	NA
GSM1881104	GSM1881104	MB	GPL570	GSE73038	dkfz_MB_15-0043	FAL SE	NA
GSM1881105	GSM1881105	MB	GPL570	GSE73038	dkfz_MB_15-0026	FAL SE	NA
GSM1881106	GSM1881106	MB	GPL570	GSE73038	dkfz_MB_15-0011	FAL SE	NA
GSM1881107	GSM1881107	MB	GPL570	GSE73038	dkfz_MB_15-0012	FAL SE	NA
GSM1881108	GSM1881108	MB	GPL570	GSE73038	dkfz_MB_15-0013	FAL SE	NA
GSM1881118	GSM1881118	MB	GPL570	GSE73038	dkfz_MB_15-0015	FAL SE	NA
GSM1881119	GSM1881119	MB	GPL570	GSE73038	dkfz_MB_15-0004	FAL SE	NA
GSM1881120	GSM1881120	MB	GPL570	GSE73038	dkfz_MB_15-0006	FAL SE	NA
GSM1881121	GSM1881121	MB	GPL570	GSE73038	dkfz_MB_15-0008	FAL SE	NA
GSM1881122	GSM1881122	MB	GPL570	GSE73038	dkfz_MB_15-0002	FAL SE	NA
GSM1881123	GSM1881123	MB	GPL570	GSE73038	dkfz_MB_15-0033	FAL SE	NA
GSM1881124	GSM1881124	MB	GPL570	GSE73038	dkfz_MB_15-0023	FAL SE	NA
GSM1881125	GSM1881125	MB	GPL570	GSE73038	dkfz_MB_15-0034	FAL SE	NA
GSM1881126	GSM1881126	MB	GPL570	GSE73038	dkfz_MB_15-0025	FAL SE	NA
GSM1881127	GSM1881127	MB	GPL570	GSE73038	dkfz_MB_15-0027	FAL SE	NA
GSM1881128	GSM1881128	MB	GPL570	GSE73038	dkfz_MB_15-0038	FAL SE	NA
GSM1881129	GSM1881129	MB	GPL570	GSE73038	dkfz_MB_15-0037	FAL SE	NA
GSM1881144	GSM1881144	MB	GPL570	GSE73038	dkfz_MB_15-0021	FAL SE	NA
GSM1881145	GSM1881145	MB	GPL570	GSE73038	dkfz_MB_15-0029	FAL SE	NA
GSM1881146	GSM1881146	MB	GPL570	GSE73038	dkfz_MB_15-0030	FAL SE	NA
GSM1881147	GSM1881147	MB	GPL570	GSE73038	dkfz_MB_15-0031	FAL SE	NA
GSM1881148	GSM1881148	MB	GPL570	GSE73038	dkfz_MB_15-0024	FAL SE	NA
GSM1881159	GSM1881159	ATRT	GPL570	GSE73038	dkfz_ATRT_15-0002	FAL SE	NA
GSM1881167	GSM1881167	ATRT	GPL570	GSE73038	dkfz_ATRT_15-0006	TRU E	Lack of info
GSM1881168	GSM1881168	ATRT	GPL570	GSE73038	dkfz_ATRT_15-0003	FAL SE	NA
GSM1881187	GSM1881187	MB	GPL570	GSE73038	dkfz_MB_15-0028	FAL SE	NA
GSM1881188	GSM1881188	MB	GPL570	GSE73038	dkfz_MB_15-0010	FAL SE	NA
GSM1881189	GSM1881189	MB	GPL570	GSE73038	dkfz_MB_15-0009	FAL SE	NA

GSM1881190	GSM18 81190	MB	GPL570	GSE73038	dkfz_MB_15-0005	FAL SE	NA
GSM1881191	GSM18 81191	MB	GPL570	GSE73038	dkfz_MB_15-0035	FAL SE	NA
GSM1881195	GSM18 81195	ATRT	GPL570	GSE73038	dkfz_ATRT_15-0011	FAL SE	NA
GSM1881197	GSM18 81197	MB	GPL570	GSE73038	dkfz_MB_15-0032	FAL SE	NA
GSM1881198	GSM18 81198	MB	GPL570	GSE73038	dkfz_MB_15-0018	FAL SE	NA
GSM1881199	GSM18 81199	MB	GPL570	GSE73038	dkfz_MB_15-0044	FAL SE	NA
GSM1881200	GSM18 81200	MB	GPL570	GSE73038	dkfz_MB_15-0042	FAL SE	NA
GSM1881201	GSM18 81201	MB	GPL570	GSE73038	dkfz_MB_15-0016	FAL SE	NA
GSM1881202	GSM18 81202	MB	GPL570	GSE73038	dkfz_MB_15-0019	FAL SE	NA
GSM1881203	GSM18 81203	MB	GPL570	GSE73038	dkfz_MB_15-0020	FAL SE	NA
GSM1881204	GSM18 81204	MB	GPL570	GSE73038	dkfz_MB_15-0014	FAL SE	NA
GSM1881205	GSM18 81205	MB	GPL570	GSE73038	dkfz_MB_15-0017	FAL SE	NA
GSM1881206	GSM18 81206	MB	GPL570	GSE73038	dkfz_MB_15-0045	FAL SE	NA
GSM1881207	GSM18 81207	MB	GPL570	GSE73038	dkfz_MB_15-0046	FAL SE	NA
GSM1881208	GSM18 81208	MB	GPL570	GSE73038	dkfz_MB_15-0036	FAL SE	NA
GSM1881209	GSM18 81209	MB	GPL570	GSE73038	dkfz_MB_15-0039	FAL SE	NA
GSM1881237	GSM18 81237	ATRT	GPL570	GSE73038	dkfz_ATRT_15-0005	TRU E	Lack of info
GSM1881243	GSM18 81243	ATRT	GPL570	GSE73038	dkfz_ATRT_15-0009	FAL SE	NA
GSM1881250	GSM18 81250	ATRT	GPL570	GSE73038	dkfz_ATRT_15-0008	FAL SE	NA
GSM1881251	GSM18 81251	ATRT	GPL570	GSE73038	dkfz_ATRT_15-0007	TRU E	Lack of info
GSM1881252	GSM18 81252	MB	GPL570	GSE73038	dkfz_MB_15-0007	FAL SE	NA
GSM1881253	GSM18 81253	MB	GPL570	GSE73038	dkfz_MB_15-0003	FAL SE	NA
GSM1881254	GSM18 81254	MB	GPL570	GSE73038	dkfz_MB_15-0001	FAL SE	NA
GSM1881255	GSM18 81255	MB	GPL570	GSE73038	dkfz_MB_15-0047	TRU E	Lack of info
GSM1881256	GSM18 81256	ATRT	GPL570	GSE73038	dkfz_ATRT_15-0012	TRU E	Lack of info
GSM1881258	GSM18 81258	MB	GPL570	GSE73038	dkfz_MB_15-0022	FAL SE	NA

8.3 C) Cell populations used for the generation of the CIBERORT signature matrix

Dataset	Cell Type (number)	Isolation method
PMID: 27785870; GSE82234 (GEO/NCBI)	Endothelial/HUVEC (6)	Human umbilical vein endothelial cell donors; Supplemented Endopan 3 medium culture
Provided by Ankur Chakravarthy PMID: 24057217; GSE50798 (GEO/NCBI)	Glia (12)	Dissected Medial Orbito-Frontal Cortex tissue; NeuN- FACS (FACS Vantage with DiVa)

	Neuron (12)	Dissected Medial Orbito-Frontal Cortex tissue; NeuN+ FACS (FACS Vantage with DiVa)
Provided by Ankur Chakravarthy PMID: 23974203; GSE49667 (GEO/NCBI)	CD4+ T Effector (6)	Healthy donor blood buffy coat samples; CD45RA+ CD45RO- CD25- FACS
	CD4+ Treg (4)	Healthy donor blood buffy coat samples; CD45RA+ CD45RO- CD25+ FACS
FlowSorted.Blood.450k (Bioconductor) PMID: 22848472	CD19+ B cells (6)	Healthy donor PBMC; Ficoll-Paque Plus - PBMC (GE Healthcare, Sweden); CD19+ MACS (Miltenyi Biotech, Germany)
	CD8+ T cells (6)	Healthy donor PBMC; Ficoll-Paque Plus - PBMC (GE Healthcare, Sweden); CD8+ MACS (Miltenyi Biotech, Germany)
	Eosinophil (6)	Healthy donor PBMC; Ficoll-Paque Plus - Granulocyte (GE Healthcare, Sweden); Eosinophil Isolation kit II (Miltenyi Biotech, Germany)
	CD14+ Monocyte (6)	Healthy donor PBMC; Ficoll-Paque Plus - PBMC (GE Healthcare, Sweden); CD14+ MACS (Miltenyi Biotech, Germany)
	Neutrophil (6)	Healthy donor PBMC; Ficoll-Paque Plus - Granulocyte (GE Healthcare, Sweden); CD16+ MACS (Miltenyi Biotech, Germany)
	CD56+ NK cells (6)	Healthy donor PBMC; Ficoll-Paque Plus - PBMC (GE Healthcare, Sweden); CD56+ MACS (Miltenyi Biotech, Germany)
FlowSorted.CordBlood.450k (Bioconductor) PMID: 27019159	CD19+ B cells (14)	Healthy donor CBMC; Ficoll 1077; CD19+ MACS (Miltenyi Biotech, Germany)
	CD8+ T cells (12)	Healthy donor CBMC; Ficoll 1077; CD8+ MACS (Miltenyi Biotech, Germany)
	CD14+ Monocyte (15)	Healthy donor CBMC; Ficoll 1077; CD14+ MACS (Miltenyi Biotech, Germany)
	CD56+ NK cells (13)	Healthy donor CBMC; Ficoll 1077; CD56+ MACS (Miltenyi Biotech, Germany)
FlowSorted.CordBloodNorway.450k (Bioconductor) PMID: 27494297	CD19+ B cells (11)	Healthy donor CBMC; Lymphoprep (Stem Cell Technologies, Norway); CD19+ FACS (BD FACS Aria)
	CD14+ Monocyte (11)	Healthy donor CBMC; Lymphoprep (Stem Cell Technologies, Norway); CD14+ FACS (BD FACS Aria)
	CD56+ NK cells (11)	Healthy donor CBMC; Lymphoprep (Stem Cell Technologies, Norway); CD56+ FACS (BD FACS Aria)
unpublished	Cancer (25)	MRT/MB Cell lines (D283, D341, D384, D425 x2, D458 x2, D556, DAOY, HDMB03, MED1, MED8A, ONS76, UW228, A204, BT12 x2, BT16, CHLA259, CHLA266 x2, G401 x2, WT1 x2)

8.4 D) Cohorts of primary CNS tumours analysed by CIBERSORT

Dataset	Tumour Type (number)
GSE70460 (GEO/NCBI)	ATRT (150)
GSE109381 (GEO/NCBI)	Multiple CNS malignancies (3905)
GSE60274 (GEO/NCBI)	Glioblastoma (68)
E-MTAB-5528 (n = 99), E-MTAB-5552 (ArrayExpress)/ PMID:29763623 (n = 71), PMID:28966033 (n = 225). Additional unpublished (n = 6)	Paediatric HGG/DIPG (401)
Unpublished	MRT Primary (88)

PMID:28726821, GSE93646, GSE85212 (GEO/NCBI) with 232 sample exclusions to remove duplicates and poor QC	Medulloblastoma (2325)
GSE20713; GSE72308 (GEO/NCBI)	Breast cancer (87 HGU133Plus2; 87 HumanMethylation450K)

8.5 E) Validation cell populations used in benchmarking CIBERSORT

Dataset	Cell Type (number)
GSE112618 (GEO/NCBI)	WBC mix (6)
GSE110554 (GEO/NCBI)	Artificial DNA mix: CD4+ T, CD8+ T, CD14+ monocyte, neutrophil, CD19+ B, CD56+ NK (12)
GSE88824 (GEO/NCBI)	CD19+ B cells (8) CD4+ T (8) CD8+ T cells (8) CD14+ Monocyte (8) neutrophil (8) CD56+ NK cells (8)