Molecular genetic and epigenetic analysis of osteoarthritis risk at the *WWP2* locus

Thesis submitted to Newcastle University for the degree of Doctor of Philosophy

by

Jack Brian Roberts

December 2024

Biosciences Institute, Faculty of Medical Sciences

Word count: 36,353

Abstract

Osteoarthritis is a common multifactorial disease characterised by pathological changes to multiple joint tissues, primarily cartilage. Millions are affected yet no disease-modifying osteoarthritis drugs are available, highlighting a major clinically unmet need. Osteoarthritis is polygenic, with genome-wide association studies reporting over 100 associated variants. Most osteoarthritis-associated variants reside within noncoding regions of the genome, suggesting they confer pathogenicity via changes to gene expression. A quarter of osteoarthritis-associated variants co-localise with DNA methylation (DNAm) at CpG dinucleotides, forming methylation quantitative trait loci (mQTLs). mQTLs often act as functional intermediaries between risk variant and effector gene.

This thesis investigates the osteoarthritis association signal rs34195470, mapping to the E3 ubiquitin ligase gene *WWP2*. Joint tissues from osteoarthritis patients were analysed to detect mQTLs at a putative enhancer. Epigenetic editing of the enhancer was used to establish causality between DNAm and *WWP2* expression in a chondrocyte cell line. *In-silico* prediction of transcription factors (TFs) informed downstream TF-DNA interaction experiments. To establish the causal variant, reporter assays and CRISPR/Cas9 editing were employed.

The osteoarthritis risk allele of rs34195470 correlated with increased DNAm levels of the enhancer in cartilage, marking a differentially methylated region (DMR). Epigenetic editing to increase DNAm at the DMR resulted in higher *WWP2* expression. HIF-1α was predicted to bind the DMR, with downstream studies demonstrating this TF exclusively binds unmethylated DNA. rs34195470 exhibited allele-dependent regulatory function in the reporter assay. CRISPR/Cas9 deletion of rs111837947, a variant in high linkage disequilibrium with rs34195470, resulted in decreased *WWP2* expression.

Specific isoforms of *WWP2* are regulatory targets of the rs34195470 association signal, modulated by the functional intermediary of DNAm at the DMR. Together, the work presented in this thesis provides a framework for the molecular characterisation of osteoarthritis genetic risk and the transition from associated variant to effector gene, revealing targets for future clinical intervention.

[300 words]

Presentations and Publications

Some of the data presented in this thesis has also been presented at several scientific conferences and seminars. The details of these presentations are listed in the table below:

Date	Conference / Seminar	Location	Presentation Type
April 2022	OARSI World Congress	Berlin, Germany	Poster
July 2022	Cutting Edge OA	Oxford, UK	Oral and Poster
August 2022	Leiden University Medical Centre seminar	Leiden, Netherlands	Oral
September 2022	BSMB	Liverpool, UK	Poster*
October 2022	CIMA	York, UK	Oral
March 2023	GRS: Cartilage Biology and Pathology	Lucca, Italy	Oral and Poster*
March 2023	GRC: Cartilage Biology and Pathology	Lucca, Italy	Poster
May 2023	CIMA/CMAR	Birmingham, UK	Poster*
September 2023	Centre for Life seminar	Newcastle upon Tyne, UK	Oral
September 2023	NUBI Live!	Newcastle upon Tyne, UK	Poster
April 2024	OARSI World Congress	Vienna, Austria	Poster
September 2024	3rd International Workshop on the Epigenetics of OA	Toronto, Canada	Oral*

*Designates presentations that received awards.

In addition to these presentations, the following peer-reviewed publications for which I am the first author feature from work presented in this thesis:

Roberts, J. B. & Rice, S. J. Osteoarthritis as an Enhanceropathy: Gene Regulation in Complex Musculoskeletal Disease. *Curr Rheumatol Rep* **26**, 222–234 (2024).

Roberts, J. B. *et al*. Specific isoforms of the ubiquitin ligase gene *WWP2* are targets of osteoarthritis genetic risk via a differentially methylated DNA sequence. *Arthritis Res Ther* **26**, 78 (2024).

Acknowledgements

The collection and processing of human osteoarthritis arthroplasty samples was performed historically by several researchers within Prof John Loughlin's research group. Collection and processing of foetal donor cartilage samples was performed by the Human Developmental Biology Resource at Newcastle University. Pyrosequencing assays were designed with the assistance of Dr Guillaume Aubourg. DNA methylation quantification and rs34195470 genotyping data for foetal cartilage samples was collected in part by Ms Olivia Boldvig, and by Mr Sai Tanishq Kanchenapally for some of the osteoarthritis synovium samples. Collection of allelic expression imbalance data from osteoarthritis cartilage and osteoarthritis synovium patient samples was performed alongside Mr Sai Tanishq Kanchenapally. This data was presented in both students' dissertations. All data was re-analysed for presentation in this thesis.

I would like to thank my supervisory team, Prof John Loughlin and Dr Sarah Rice, for their guidance and support throughout the entirety of this project. Both have contributed greatly to my development as a researcher, and I have the utmost gratitude for their mentorship during my PhD studies. I would also like to thank the Ruth and Lionel Jacobson Charitable Trust for funding this research project, without which I would not have been able to continue studying in the field of research I enjoy most.

To the wider Skeletal Research Group, thank you for creating an enjoyable work environment and providing emotional support during my dramatic episodes when results weren't what I expected! A special shout-out to Dr Matt Barter for providing plenty of alt-pop bangers and *club classics* during lengthy pipetting sessions, and of course, to my work sibling and desk neighbour Abby Brumwell who has been there for me since day one (literally). Here's one for the memory jar! It turns out you really can have the moon on a stick.

I'd like to thank my family and friends - for everything. To my wonderful partner Nathan, thank you for your unconditional support and building me up in times of doubt. I really do appreciate everything you have done for me over the last few years. And finally, to my dear cat Charlie, thank you for being oblivious to the struggles of academia and meowing for food incessantly whilst I wrote this thesis.

Table of Contents

Abstract ... **i**

Presentations and Publications ... **ii**

Acknowledgements ... **iii**

Table of Contents ... **iv**

List of Figures ... **viii**

List of Tables ... **x**

Abbreviations ... **xi**

Chapter 1: Introduction ... **1**

- 1.1 Synovial joint formation ... 2
  - *1.1.1 Skeletogenesis ... 2*
  - *1.1.2 Synovial joint formation ... 2*
  - *1.1.3 Endochondral ossification ... 3*
- 1.2 Synovial joint tissues and function ... 5
  - *1.2.1 Articular cartilage ... 5*
  - *1.2.2 Subchondral bone ... 6*
  - *1.2.3 Synovium ... 7*
  - *1.2.4 Other tissues of the articular joint ... 7*
- 1.3 Osteoarthritis (OA) ... 8
  - *1.3.1 Pathophysiology and molecular mechanisms of OA ... 8*
  - *1.3.2 Epidemiology, risk factors, socioeconomic burden and treatment ... 11*
- 1.4 OA as a polygenic disease ... 12
  - *1.4.1 Genome-wide association studies (GWAS) and polygenic disease ... 12*
  - *1.4.2 Genetics of OA ... 14*
  - *1.4.3 Developmental origins of OA ... 15*
- 1.5 Epigenetics and OA ... 15
  - *1.5.1 Epigenetics ... 15*
  - *1.5.2 DNAm and OA ... 17*
  - *1.5.3 Post-translational histone modifications and OA ... 20*
  - *1.5.4 Noncoding RNAs and OA ... 20*
- 1.6 Functional follow-up studies ... 21
  - *1.6.1 From GWAS signals to effector genes ... 21*
  - *1.6.2 Utilisation of functional genomic datasets ... 22*
  - *1.6.3 QTL analyses ... 22*

1.6.4 Functional laboratory experiments 23
1.7 *WWP2* and miR-140 24
1.7.1 WWP2 24
1.7.2 miR-140 27
1.7.3 OA risk variants mapping to the WWP2/miR-140 locus 28
1.7.4 OA cartilage mQTLs at the WWP2/miR-140 locus 30
1.8 Aims and hypotheses 30
Chapter 2: Materials and Methods 32
2.1 Reagents 33
2.2 Methods 34
2.2.1 In silico analyses 34
2.2.2 OA patient arthroplasty and foetal donor samples 36
2.2.3 Nucleic acid extraction from OA patient arthroplasty and foetal donor samples 36
2.2.4 Bisulphite conversion of DNA 37
2.2.5 Genotyping 37
2.2.6 DNAm quantification 38
2.2.7 Complementary DNA (cDNA) synthesis 39
2.2.8 AEI analysis 40
2.2.9 Gene expression analysis 41
2.2.10 TC28a2 cell culture and passaging 41
2.2.11 Construction of Lucia reporter vectors 42
2.2.12 Methylation and mock-methylation of Lucia reporter constructs containing the region of interest 43
2.2.13 Lucia reporter assays in TC28a2 cells 43
2.2.14 Construction of dCas9-DNMT3a plasmids for epigenetic editing 44
2.2.15 dCas9-DNMT3a epigenetic editing in TC28a2 cells 45
2.2.16 Stabilisation of HIF-1α in TC28a2 cells 46
2.2.17 Nuclear protein extraction 46
2.2.18 Western blotting 47
2.2.19 Electrophoretic mobility shift assays (EMSAs) 48
2.2.20 Construction of CRISPR/Cas9 vectors 49
2.2.21 CRISPR/Cas9 deletion of SNVs in TC28a2 cells 49
2.2.22 Construction of luciferase reporter vectors and site-directed mutagenesis 51
2.2.23 Luciferase reporter gene assays in TC28a2 cells 52
2.2.24 Statistical analyses 53

Chapter 3: *In-silico* analysis of the *WWP2* locus 54
3.1 Introduction 55
3.2 Results 56
3.2.1 Identification of rs34195470-eQTLs 56
3.2.2 Chromatin accessibility and regulatory state at the region of interest 57
3.2.3 Chromatin looping at the region of interest 58
3.3 Discussion 59
Chapter 4: mQTL analysis and *WWP2* expression analysis in synovial joint tissues 62
4.1 Introduction 63
4.2 Results 64
4.2.1 mQTL analysis in foetal cartilage and OA cartilage 64
4.2.2 AEI analysis 70
4.2.3 Multi-tissue meQTL analysis 72
4.2.4 Multi-tissue WWP2 expression profiling and eQTL analysis 76
4.3 Discussion 79
Chapter 5: Functional studies of the DMR in an immortalised chondrocyte cell line 82
5.1 Introduction 83
5.2 Results 84
5.2.1 Lucia reporter gene assay 84
5.2.2 dCas9-DNMT3A epigenetic editing of the DMR 85
5.2.3 In-silico prediction of TF binding sites across the DMR 87
5.3 Discussion 88
Chapter 6: The role of HIF-1α as a transcriptional regulator of *WWP2* 92
6.1 Introduction 93
6.2 Results 94
6.2.1 HIF-1α is expressed in TC28a2 cells and can bind the WWP2 DMR in-vitro 94
6.2.2 Lucia reporter gene assays in $CoCl_2$-treated TC28a2 chondrocytes 97
6.2.3 dCas9-DNMT3A epigenetic editing in $CoCl_2$-treated TC28a2 chondrocytes 98
6.2.4 Effect of $CoCl_2$ on WWP2 expression 100
6.2.5 In-silico TF prediction following updates to JASPAR Core 2024 101
6.3 Discussion 103
Chapter 7: Determination of the causal variant marked by rs34195470 107
7.1 Introduction 108
7.2 Results 109
7.2.1 In-silico analysis of the region 109

7.2.2 Luciferase reporter gene assay 111
7.2.3 CRISPR/Cas9 deletion of the genomic regions housing rs34195470, rs9746247 and rs111837947 112
7.2.4 In-silico TF binding prediction at the three regions housing variants 115
7.3 Discussion 118
Chapter 8: General Discussion 122
8.1 Introduction 123
8.2 Summary of results 124
8.3 Clinical utility of findings 126
8.4 Limitations, unanswered questions and future directions 129
8.5 Conclusions 131
Chapter 9: Appendices 132
References 154

List of Figures

Figure 1.1: The process of endochondral ossification. 4
Figure 1.2: The healthy articular joint and the zones of articular cartilage. 6
Figure 1.3: Pathological changes to the tissues of the articular joint in OA. 11
Figure 1.4: The relationship between OA risk variants, their associated mQTLs and target gene expression. 19
Figure 1.5: Protein domains encoded by WWP2-FL, WWP2-N and WWP2-C. 25
Figure 1.6: WWP2 isoforms target SMAD signalling proteins for degradation. 27
Figure 1.7: The genomic region encompassing *WWP2* and miR-140. 29
Figure 2.1: Vector map for dCas9-DNMT3a. 45
Figure 3.1: Expression of genes identified as rs34195470-eQTLs using GTEx in OA cartilage. 56
Figure 3.2: Chromatin accessibility and regulatory elements in relevant cell types at the *WWP2* locus. 58
Figure 3.3: Chromatin looping at the *WWP2* locus. 59
Figure 4.1: mQTL analysis in OA cartilage and foetal cartilage. 66
Figure 4.2: mQTL analysis in OA bone and OA synovium. 68
Figure 4.3: Mean DNAm levels across the DMR. 69
Figure 4.4: Genotypic effect upon DNAm levels across the DMR. 69
Figure 4.5: AEI assay validation. 70
Figure 4.6: AEI analysis in synovial joint tissues. 72
Figure 4.7: meQTL analysis in OA cartilage and foetal cartilage. 73
Figure 4.8: meQTL analysis in OA bone and OA synovium. 75
Figure 4.9: *WWP2* expression in synovial joint tissues. 77
Figure 4.10: No eQTLs were detected for any of the three *WWP2* transcript isoforms in articular joint tissues. 78
Figure 5.1: Investigation of the transcriptional regulatory function of the DMR in TC28a2 chondrocytes. 84
Figure 5.2: Epigenetic modulation of the DMR in TC28a2 chondrocytes. 86
Figure 5.3: TFs predicted to bind at the DMR. 87
Figure 5.4: Proposed molecular mechanism. 90

Figure 6.1: HIF-1α is expressed in TC28a2 chondrocytes and can bind the *WWP2* DMR. 96
Figure 6.2: Investigation of the transcriptional regulatory function of the DMR in $CoCl_2$-treated TC28a2 chondrocytes. 98
Figure 6.3: Epigenetic modulation of the DMR in $CoCl_2$-treated TC28a2 chondrocytes. 99
Figure 6.4: Effect of $CoCl_2$ treatment on *WWP2* expression in TC28a2 chondrocytes. 101
Figure 6.5: TFs predicted to bind at the DMR by JASPAR Core 2024. 102
Figure 7.1: In-silico analysis of the region encompassing the three variants associated with OA genetic risk marked by rs34195470. 110
Figure 7.2: Luciferase reporter assay in TC28a2 chondrocytes. 111
Figure 7.3: CRISPR/Cas9 deletion of the three variants associated with OA genetic risk marked by rs34195470. 114
Figure 7.4: *WWP2* splicing unaffected by deletion of rs34195470, rs9746247 and rs111837947. 115
Figure 7.5: TFs predicted to bind rs34195470, rs9746247 and rs111837947. 117

List of Tables

Table 2.1. Details of reagents used in the methods. 33
Table 2.2. List of online databases and tools used. 35
Table 2.3. List of published datasets used. 35
Table 2.4. Summary details for OA patient arthroplasty and foetal donor samples. 36
Table 2.5. Summary of statistical analyses used in this thesis. 53
Table 5.1. Summary of the effects of dCas9-DNMT3A epigenetic editing upon DNAm levels at the DMR. 86
Table 5.2: List of TFs predicted to bind the DMR. 88
Table 6.1. Summary of the effects of dCas9-DNMT3A epigenetic editing in $CoCl_2$-treated TC28a2 cells upon DNAm levels at the DMR. 100
Table 6.2: List of TFs predicted to bind the DMR by JASPAR Core 2024. 103
Table 7.1: Summary information for the lead OA risk variant rs34195470 and the variants in high LD. 109
Table 7.2. Summary of the effects of CRISPR/Cas9 deletions upon DNAm levels at the DMR. 115
Table 7.3: List of TFs predicted to bind rs34195470 and rs111837947. 118

Abbreviations

5mC	5-methylcytosine
AEI	Allelic expression imbalance
ATAC-Seq	Assay for transposase accessible chromatin with high-throughput sequencing
BMI	Body mass index
cDNA	Complementary DNA
ChIP-Seq	Chromatin immunoprecipitation with high-throughput sequencing
CpG	Cytosine-guanine dinucleotide
$CoCl_2$	Cobalt chloride
CRISPR	Clustered regularly interspaced short palindromic repeats
dCas9	Catalytically inactive ‘dead’ Cas9
DMOADs	Disease-modifying osteoarthritis drugs
DNAm	DNA methylation
DNMT	DNA methyltransferase
DMR	Differentially methylated region
ECM	Extracellular matrix
EMSA	Electrophoretic mobility shift assay
ER	Endoplasmic reticulum
eQTL	Expression quantitative trait locus
FLS	Fibroblast-like synoviocytes
gRNA	Guide RNA
GWAS	Genome-wide association study
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HIF	Hypoxia inducible factor
HPAC	Human primary articular chondrocytes
HRE	Hypoxia response element
JSW	Joint space width
LD	Linkage disequilibrium
lncRNA	Long noncoding RNA
MBD	Methyl-CpG binding domain proteins

miRNA	Micro-RNA
meQTL	Methylation-expression quantitative trait locus
MSC	Mesenchymal stem cell
mQTL	Methylation quantitative trait locus
OA	Osteoarthritis
OR	Odds ratio
PCR	Polymerase chain reaction
PHD	Prolyl hydroxylase
pri-miRNA	Primary microRNA
NSAIDs	Nonsteroidal anti-inflammatory drugs
pcw	Post-conception weeks
RISC	RNA-induced silencing complex
RNA-Seq	RNA sequencing
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
SNV	Single nucleotide variant
TF	Transcription factor
TFBS	Transcription factor binding site
TPM	Transcripts per million
TSS	Transcription start site
UTR	Untranslated region

Chapter 1: Introduction

1.1 Synovial joint formation

1.1.1 Skeletogenesis

The development of the human skeletal system, comprising of bone, cartilage and joints is a process known as skeletogenesis. This process begins when multipotent mesenchymal stem cells (MSCs) arise from the mesoderm and organise themselves to form the template of the human skeleton[1]. These mesenchymal condensates are densely populated and begin to differentiate into chondrocytes to form cartilage anlage of future bones (via endochondral ossification) or into osteoblasts to directly form bone (via intramembranous bone formation)[2]. The former is responsible for the formation of most of the human skeleton, including the appendicular skeleton, whilst the latter occurs in the membranous neuro- and viscerocranium and in part of the clavicle[2].

Skeletogenesis is controlled by several factors. Skeletal patterning is regulated by many transcription factors (TFs) in the Homeobox and Paired-box families[2], with mutations in these factors resulting in skeletal abnormalities[3,4]. Bone shape is regulated by the Transforming Growth Factor Beta (TGFβ) superfamily, including Growth Differentiation Factors (GDFs) and Bone Morphogenic Proteins (BMPs)[2]. Differentiation of MSCs into chondrocytes that form the cartilage anlagen is driven by the master chondrogenesis TF SRY (Sex-Determining Region Y)-Box 9 (SOX9) following activation by BMP signalling[5,6]. Other factors important for skeletogenesis include Indian Hedgehog (IHH) and Parathyroid Hormone-related Peptide (PTHrP), which control chondrocyte hypertrophy and bone formation by osteoblasts[7], and the Fibroblast Growth Factor (FGF) and Wnt/β-catenin signalling pathways, which are required for endochondral ossification and the development of long bones[8].

1.1.2 Synovial joint formation

In humans, the development of the synovial joint occurs between four- and eight-weeks post-conception (pcw)[9]. Dense MSC condensations within the uninterrupted cartilage anlagen become identifiable by their flattened morphology, marking the future joint site known as the interzone[10]. Loss of this region prevents the formation of limb synovial joints, emphasising its pivotal role[11]. In the early stages, the interzone is characterised by high expression of *GDF5* and low expression of the chondrocyte markers type II collagen (*COL2A1*) and Matrillin 1 (*MATN1*)[10,12]. *GDF5* is critically important to joint formation, with mutations resulting in skeletal abnormalities[12]. These *GDF5*-expressing progenitor cells are responsible for the development of all mature joint structure lineages, including the menisci, ligaments, and synovium, whilst the outer regions surrounding the interzone differentiate into chondrocytes and begin synthesising the extracellular matrix (ECM) that forms articular cartilage[10,13].

Around 8pcw, joint cavitation takes place[9]. Historically, this process is thought to have been driven by interzone cell apoptosis, however, studies suggest that activation of the MEK-ERK pathway may also play a role[13]. Activation of this pathway by mechanical stimulation results in hyaluronan synthesis and loss of tissue cohesion[14]. Other factors such as SOX5 and SOX6 are also necessary for cavitation, with their knockout in mice leading to incomplete joint formation[15]. As the interzone diminishes, the surrounding MSCs form major components of the synovial capsule. The synovial capsule becomes filled with synovial fluid containing high levels of lubricin and hyaluronic acid[16,17]. Other tissues such as the infrapatellar fat pad develop later (18pcw)[9].

1.1.3 Endochondral ossification

The replacement of the temporary cartilaginous anlagen into permanent bone is a process known as endochondral ossification. The primary ossification site forms in the centre of the bone, characterised by chondrocyte hypertrophy and vascular invasion. Secondary sites of ossification develop, segmenting cartilage responsible for growth into two sites: the growth plate (between the primary and secondary sites of ossification) and the articular cartilage lining the end of the bone. The growth plate drives longitudinal extension of bones whilst the articular cartilage shapes the epiphysis and continues until early adulthood. At skeletal maturity, the growth plate is converted into bone as the primary and secondary sites of ossification merge. Only articular cartilage remains, providing normal joint function[18].

Endochondral ossification is a complex and tightly regulated process requiring several systemic factors, transcriptional regulators and secreted factors that work in synchronicity to control the growth rate and morphology of long bones (Fig. 1.1). The expression and secretion of these factors is driven by growth plate chondrocytes that are arranged in morphologically distinct zones, reflecting their role in endochondral ossification[18]. The ‘resting’ zone furthest from the ossification front contains chondrocytes expressing the ECM components aggrecan (encoded by *ACAN*) and type II collagen[13]. These cells mature into columnar, rapidly proliferating chondrocytes that continue to express ECM components. Proliferative chondrocytes undergo transition to a pre-hypertrophic phenotype, secreting IHH and PTHrP. These control the rate of proliferation and hypertrophy (and therefore control the rate of endochondral ossification), which determines bone dimension and length[13,18]. These chondrocytes eventually become fully hypertrophic as they reach the ossification front, expressing high levels of RUNX Family Transcription Factor 2 (*RUNX2*), an osteogenic factor and master regulator of chondrocyte hypertrophy driving terminal differentiation[19]. Hypertrophic chondrocytes also express the ECM-degrading enzymes Matrix Metallopeptidase 13 (MMP13) and ADAM Metallopeptidase With Thrombospondin Type 1 Motif 5 (ADAMTS5), which break down the cartilage matrix to allow for vascular invasion and bone remodelling by osteoclasts and osteoblasts[18,20]. This complex, interconnected process is also regulated by members of the TGFβ/BMP and the FGF signalling families as well as the Hypoxia Inducible Factors (HIFs) HIF-1α and HIF-2α, which have opposing anabolic and catabolic roles, respectively[18,21].

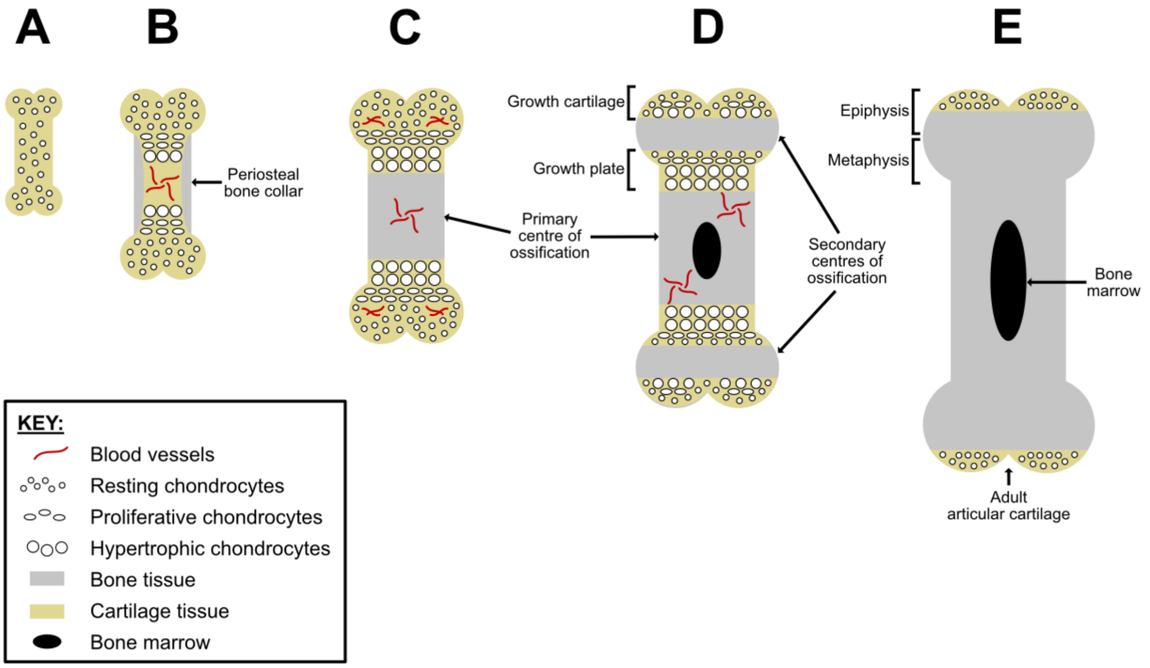


Figure 1.1: The process of endochondral ossification. A) Endochondral ossification begins with the cartilage model. **B)** Initiation of the formation of the primary centre of ossification forms, creating a periosteal bone collar. This is accompanied by vascular invasion and chondrocyte hypertrophy closer to the centre of the cartilage model. **C)** The primary centre of ossification becomes well established, with some vasculature within the cartilage tissue at the ends of developing long bones. **D)** Formation of the secondary centres of ossification, dividing

cartilage into growth cartilage (lining the developing articular joint surfaces) which shapes the epiphysis; and the growth plate (between the primary and secondary centres of ossification). The growth plate is responsible for longitudinal growth. **E)** In the adult bone, the metaphyseal and epiphyseal bone have fused together, leading to the disappearance of the growth plate cartilage. Permanent adult articular cartilage lines the ends of the bone. Adapted from Mackie *et al.*, 2008[18].

1.2 Synovial joint tissues and function

1.2.1 Articular cartilage

Articular cartilage is an avascular, aneural tissue lining the load-bearing surfaces of the articular joint and is composed of a dense ECM secreted by chondrocytes, the sole cell type found in cartilage[22]. Although they are responsible for depositing the ECM, healthy adult articular chondrocytes remain in a non-proliferative, postmitotic state across life[23]. Up to 80% of articular cartilage wet weight is water, which facilitates diffusion of molecules necessary for the survival of chondrocytes and confers its ability to withstand significant loading[24]. Collagens form the most abundant structural macromolecule present in the ECM, comprising of up to 60% of the dry weight. Of this, type II collagen represents nearly 95%, forming fibrils and bundles with proteoglycan aggregates. Other collagens, including types I, IV, V, VI, IX, and XI are also present. Together, the collagen fibril network provides articular cartilage with tensile strength to resist shear stress[24,25]. Aggrecan, the most abundant proteoglycan of the ECM, provides cartilage with its ability to resist compressive loading via its osmotic properties. Other proteoglycans such as decorin, biglycan and fibromodulin contribute to the structural network of collagen fibrils[22,24].

The ECM of articular cartilage can be segmented into four distinct zones: (i) superficial; (ii) transitional; (iii) radial; and (iv) calcified. The superficial zone is densely packed with collagen fibrils arranged tangentially to the joint, surrounding flattened, ellipsoid-shaped chondrocytes[22]. Superficial zone chondrocytes secrete lubricin, a proteoglycan responsible for maintaining low-friction movement of the articular joint surfaces[26,27]. Moving deeper, the transitional zone features spherical chondrocytes surrounded by bundles of thicker collagen fibrils. The radial zone represents the deepest layer of non-mineralised cartilage, comprising of the thickest collagen bundles[22]. Chondrocytes in this zone become hypertrophic as depth increases until they reach the final zone, underneath the tidemark, and are characterised by their expression of *IHH*, *RUNX2* and *COL10A1*[27]. This final zone is comprised of calcified cartilage and acts as the final barrier between the cartilage and the underlying subchondral bone[28,29]. The morphological differences between the zone subpopulations of chondrocytes are largely driven by the mechanical environment, whilst the arrangement of collagen fibres between these layers confers tensile strength that protects the subchondral bone from mechanical stress[22]. The avascular nature of articular cartilage means that chondrocytes exist in a hypoxic environment, with the O_2 tension ranging from 10% at the surface to less than 1% in the deep zones[22]. Adaptation to hypoxia is provided by the HIFs, which mediate changes to metabolism and transcription, contributing to the chondrocyte phenotype[21]. An overview of the articular joint and the zones of articular cartilage is depicted in Fig. 1.2.

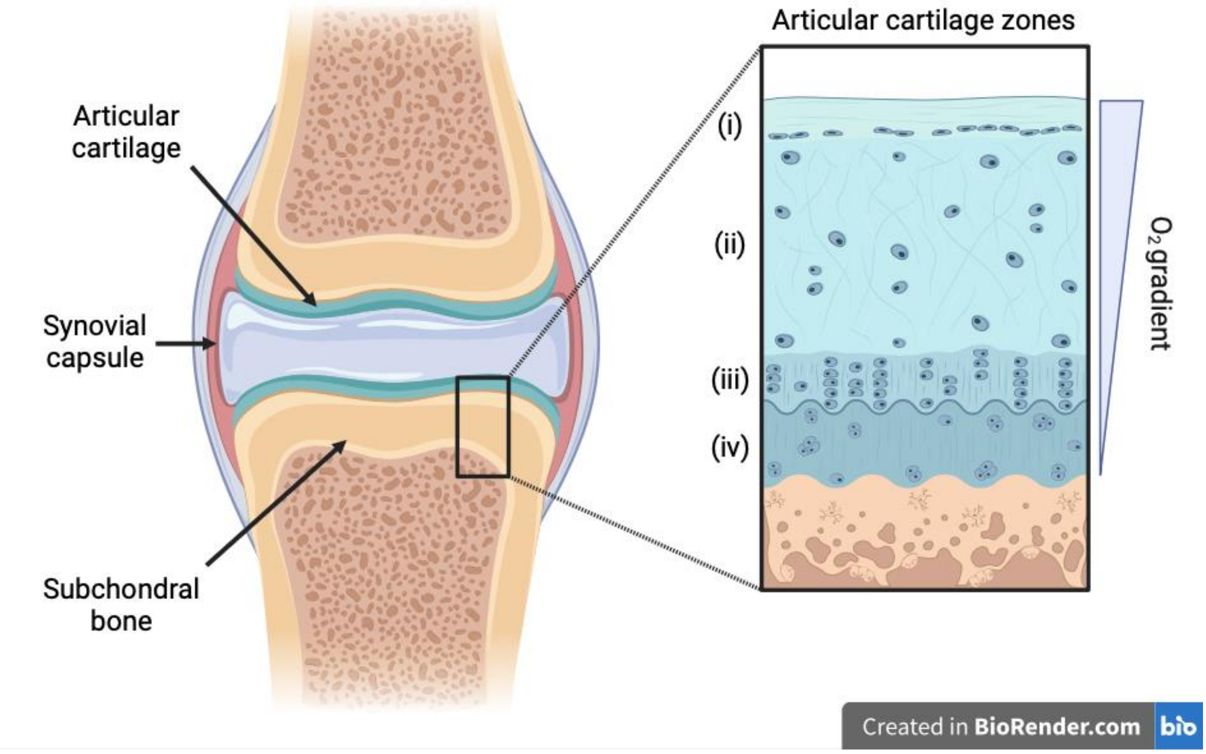


Figure 1.2: The healthy articular joint and the zones of articular cartilage. Left, schematic representation of the articular joint and its core tissues: articular cartilage, the synovial capsule and underlying subchondral bone. Right, schematic representation of the four zones of articular cartilage: (i) superficial zone, (ii) transitional zone, (iii) radial zone, (iv) calcified zone. The O_2 gradient moving from the superficial zone to the calcified zone is also shown. Created in BioRender (available at https://biorender.com).

1.2.2 Subchondral bone

Subchondral bone refers to the bone found underneath the calcified cartilage layer and can be split into two morphological compartments: a cortical plate directly beneath the calcified cartilage, and a subchondral trabecular bone that is porous and less dense than the cortical plate[30]. Vasculature and nerves traverse the cortical plate from the subchondral bone across to the calcified layer of cartilage, underlining the interactions between these tissue types[31,32]. The subchondral bone acts as a shock absorber and provides joint stability[31,32]. The maintenance and turnover of bone tissue is mediated by osteoblasts, which synthesise bone matrix, and osteoclasts, which resorb bone tissue[33]. In addition, osteocytes reside within the synthesised bone matrix and act as mechanosensors, signalling to osteoblasts and osteoclasts to orchestrate bone remodelling[33,34]. Several pathways are implicated in subchondral bone homeostasis, including Wnt/β-catenin and TGFβ signalling[35,36].

1.2.3 Synovium

The joint capsule is lined with synovial tissue that can be categorised into the intima, a layer of joint-facing cells (macrophages and fibroblast-like synoviocytes (FLS)) that synthesise synovial fluid, and the outer subintima, a dense fibrous tissue consisting of collagen fibres, nerves, lymph and blood vessels[37]. The subintima is made up of various cell types, including FLS, macrophages, monocytes, CD4$^+$ T cells and mast cells[37]. FLS are the most abundant cell type found in the synovium and are responsible for synthesising lubricin and hyaluronic acid that is deposited into the synovial fluid, facilitating frictionless movement[38]. Vasculature of the synovium facilitates the delivery of nutrients and cellular O_2 throughout the synovial tissue[37]. The synovium also nourishes chondrocytes via diffusion of nutrients from the synovial fluid into the cartilage ECM[37]. Innervation of the synovium promotes joint proprioception and pain sensing, aiding in the detection of potential joint damage[39].

1.2.4 Other tissues of the articular joint

In addition to the articular cartilage, synovium and subchondral bone, the articular joint comprises of menisci, tendons and ligaments. The menisci form fibrocartilaginous discs that act to stabilise the joint and facilitate mechanical loading[40]. Tendons and ligaments also contribute to joint stability, forming strong fibrous bands predominantly composed of type I collagen that connect the bones spanning the joint site[41]. These tissues, in addition to the infrapatellar fat pad and the surrounding skeletal muscle, all contribute to articular joint function and facilitate smooth movement of the human skeleton[42].

1.3 Osteoarthritis (OA)

1.3.1 Pathophysiology and molecular mechanisms of OA

OA is a disease of the whole joint characterised by complex pathophysiology, manifesting primarily through molecular dysregulation and subsequent anatomical and physiological changes that lead to cartilage degradation, bone remodelling, osteophyte formation, synovial inflammation and eventually loss of normal joint function[43,44]. Although OA can affect any joint site, it primarily affects the knee, the hand and the hip[45]. These changes are illustrated in Fig. 1.3.

Articular cartilage degradation is widely considered to be the core hallmark of OA, beginning primarily at sites of greater mechanical stress[42,46]. Chondrocytes transition from their previous resting state into an active state of proliferation with increased production of the ECM-degrading enzymes MMP3 and ADAMTS5[20,42]. This transition drives matrix remodelling and terminal hypertrophic differentiation, which in turn leads to calcification of the cartilage[22]. Matrix degradation continues with the activation of MMP13, which degrades type II collagen[42]. These changes appear to be irreversible as the once-intact collagen network and interspersed proteoglycans become disorganised and degraded[42]. Loss of articular cartilage is accompanied by joint space width (JSW) narrowing. Magnetic resonance imaging in OA patients indicates that higher JSW narrowing grades are linked with larger subchondral bone areas, suggesting that an increase in subchondral bone area occurs in advanced OA[47]. Aberrant TGFβ/SMAD signalling is associated with OA, with the SMAD2/3 pathway, which prevents chondrocyte hypertrophy, downregulated. As a result, the SMAD1/5/8 pathway becomes dominant, contributing to chondrocyte hypertrophy and calcification of the ECM[48–50]. Mutations in *SMAD3* result in early-onset OA[51,52], whilst decreased Smad3 promotes cartilage degradation in mice[53,54]. Changes to TGFβ/SMAD signalling exhibit pleiotropic effects in early adult versus aged cartilage, with abundant TGFβ preventing chondrocyte hypertrophy in the young, whereas increased TGFβ confers deleterious effects upon cartilage integrity in the aged or OA joint[49]. This is primarily driven by a decrease in the expression of Activin Receptor-Like Kinase 5 (ALK5), which promotes the SMAD2/3 pathway, resulting in a switch towards activation of the ALK1-mediated SMAD1/5/8 pathway[49]. Three TGFβ ligands are present in mammals: TGFβ1, TGFβ2 and TGFβ3, each exhibiting a high degree of sequence homology and overlapping functional roles in cartilage development and homeostasis[55].

However, distinct functional roles and expression patterns also exist between the three TGFβ isoforms. TGFβ1 is primarily responsible for cartilage ECM homeostasis, positively regulating SMAD2/3 signalling and inhibiting chondrocyte hypertrophy[56,57]. TGFβ2 is required during chondrogenesis and endochondral ossification, with *Tgfb2* KO mice exhibiting chondrodysplasia[58]. TGFβ3 also promotes chondrogenesis and inhibits terminal differentiation *in-vitro*[59]. Age-related changes in TGFβ expression have been observed in mice, with a decrease in Tgfβ2 and Tgfβ3, but not Tgfβ1[60].

Remodelling of subchondral bone in OA occurs following the formation of cracks and subchondral lesions caused by excessive loading[42]. Endochondral ossification is activated at the site of the tidemark, resulting in the formation of multiple tidemarks and increased mineralisation of cartilage as the subchondral bone encroaches, resulting in cartilage thinning[30,42]. Angiogenesis and innervation at the osteochondral junction are accompanied by Vascular Endothelial Growth Factor (*VEGF*) expression, which replaces bone marrow with fibrovascular tissue and is also associated with the development of nerve fibres expressing Nerve Growth Factor (*NGF*) that may be a driver of pain in OA patients[61–63]. Remodelling is stimulated by TGFβ and the inflammatory cytokines Interleukin-1 (IL-1) and IL-6, which are also expressed in degraded OA cartilage[64]. The rate of remodelling leads to changes in subchondral bone composition, with regions of high remodelling having reduced mineralisation and decreased elastic modulus, and the inverse for regions of low remodelling, resulting in brittle bones with increased elastic modulus[65]. Increased rates of bone remodelling in early-stage OA may result in changes to joint morphology and the distribution of loading that can predispose to cartilage degradation[30]. This creates a feedback loop in which overloading of the joint increases the rate of bone remodelling and the cycle continues[30]. The formation of osteophytes in OA accompanies changes in bone remodelling. Osteophytes develop following recapitulation of the developmental process of endochondral ossification, driven by TGFβ and BMP2[66]. However, the functional role of osteophytes is not well understood and may contribute to joint stabilisation rather than OA progression[42,66].

Synovial inflammation (synovitis) has been reported in both early- and late-stage OA, suggesting it is involved in disease progression[67,68], as well as following joint injury[69]. Injury to the joint induces the release of many proinflammatory cytokines and MMPs into the synovial fluid, which impact upon cartilage integrity[70]. Synovitis is characterised by infiltration of macrophages and lymphocytes, hyperplasia of the synovial membrane, fibrosis and increased vasculature[42,68]. Cellular changes include FLS proliferation and higher numbers of macrophages that induce hyperplasia of the synovial lining[38]. Abundance of proinflammatory macrophages leads to higher MMP levels as well as ADAMTS4 and ADAMTS5, which can lead to cartilage degradation[71], whereas depletion of synovial macrophages can reduce cartilage catabolism[72]. OA synovial fluid has higher levels of proinflammatory cytokines including IL-1β and Tumour Necrosis Factor-α (TNF-α), which signal chondrocytes within the articular cartilage towards catabolic mechanisms[22], and reduced levels of lubricin and hyaluronic acid[73]. Synovial fluid may also contain fragments of degraded bone and cartilage, which stimulate FLS towards a chronic inflammatory state[74]. VEGF is also increased in the synovial fluid and may contribute to vascular changes observed at the osteochondral junction[75]. Early-stage OA has markedly higher levels of IL-15 and is associated with progression of radiographic OA[76]. IL-15 induces synthesis of MMP1 and MMP9, which are involved in cartilage degradation and OA progression[77,78]. IL-17 produced by synovial fibroblasts and articular chondrocytes can contribute to synovial hyperplasia[38,79]. Synovitis is one of the main drivers of pain in OA, with upregulation of nociceptive pain signalling pathways and NGF expression[80].

Changes to the menisci and ligaments are common in OA, and injury to these tissues is a major risk factor for OA development[42,81]. Pathological changes include matrix synthesis disruption and collagen fibre disorganisation as well as increased vascularity and neuronal activity[82–84]. Degenerative changes to these tissues have been observed with ageing and following joint replacement surgery[85].

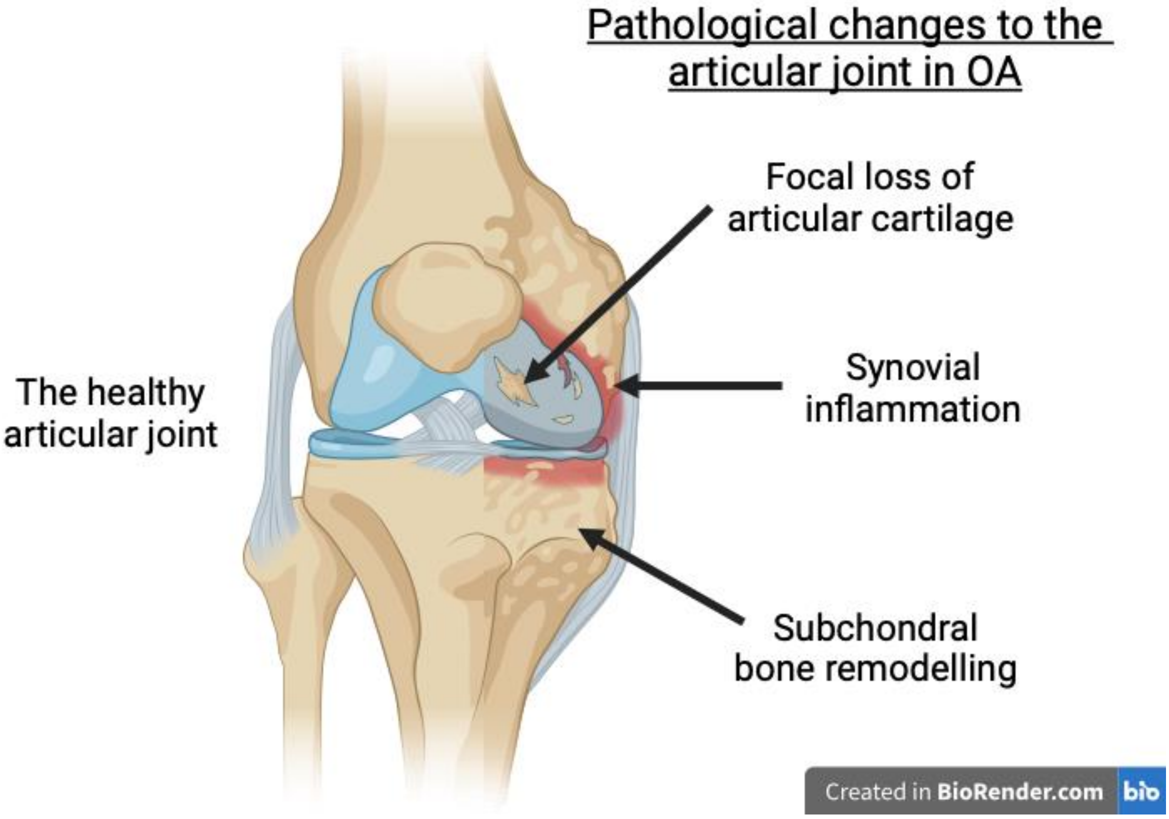


Figure 1.3: Pathological changes to the tissues of the articular joint in OA. Left, the healthy articular joint. Right, schematic representation of three pathological changes to the articular joint, including the focal loss of cartilage, synovial inflammation, and subchondral bone remodelling. Created in BioRender (available at https://biorender.com).

1.3.2 Epidemiology, risk factors, socioeconomic burden and treatment

More than 500 million people are affected by OA worldwide, with incidence rising by 48% between 1990 and 2019[45,86]. By 2050, the number of individuals with OA is estimated to exceed one billion. Incidence of OA increases with age, peaking in individuals aged 70 and above, and is a key factor driving the projected increase in OA incidence globally as the worldwide population ages[45]. OA is more common in females than in males[45]. Other risk factors include high body mass index (BMI $\geq$ 30)[45], previous joint injury such as anterior cruciate ligament rupture[81], and genetics[87–89].

OA confers great burden upon patients, causing pain, restricted movement and disability, cumulatively reducing their quality of life[45,86,90]. These factors contribute negatively to fatigue, mood and sleep[90]. Because of the disabling nature of OA, many affected individuals require assistance with their personal needs[90]. Mortality rates are higher in patients with OA in comparison to the wider population, though a direct mechanism is unclear[91]. It has been suggested that lack of physical activity contributes to the association of OA with other comorbidities[44], including cardiovascular disease and dementia[91]. The economic burden of OA is substantial, costing between 1- and 2.5% of gross national product in several major market economies, including the UK[92]. These costs manifest from direct medical costs (e.g. pain relief, physical therapy and surgery) as well as indirect costs from lost economic productivity[93]. Loss of economic productivity is driven by absenteeism from sickness, premature death (compounded by comorbidities) and early retirement by the individual, but also from carers who compensate upon daily household duties on behalf of OA sufferers[90,93].

Despite widespread prevalence and the significant burden conferred upon the individual and the economy, OA is largely ignored in global health strategic planning[86]. There are widespread misconceptions that OA is an inevitable part of ageing, with many patients reporting that their concerns are dismissed by medical professionals[86,94]. Management of OA is limited to nonpharmacological interventions such as exercise and weight management, and pharmacological treatments such as nonsteroidal anti-inflammatory drugs (NSAIDs) and intra-articular corticosteroid injections for pain relief[95,96]. However, NSAIDs carry cardiovascular and gastrointestinal risks whilst repeated intra-articular injections have been reported to be associated with cartilage thinning and degradation[97–99]. Therefore, there is a major clinically unmet need in developing disease-modifying OA drugs (DMOADs) for therapeutic intervention[44].

End-stage OA, characterised by persistent pain and functional loss of joint mobility, is treated by total joint replacement (arthroplasty)[44]. Arthroplasty has low rates of serious complications resulting in patient mortality[100] and low rates of revision surgery[101,102]. Typically, patients who undergo arthroplasty surgery over the age of 70 are more likely to die with their original implants than undergo revision[103]. Patient satisfaction following arthroplasty of the hip or knee is high, with most reporting little to no pain[104]. However, arthroplasty incurs great economic burden, costing the UK over £850 million in 2010[105]. These costs are expected to rise as global populations age and already high BMI rates continue to increase[45,86].

1.4 OA as a polygenic disease

1.4.1 Genome-wide association studies (GWAS) and polygenic disease

Genetic variation can influence phenotype in two ways: protein-coding variants that introduce nonsynonymous mutations, leading to altered protein function; or variants that alter the regulation of gene expression leading to changes in mRNA levels and thus translated protein abundance. Understanding of Mendelian (monogenic) diseases has benefitted greatly from genetic studies investigating nonsynonymous mutations. However, for many common, non-Mendelian (polygenic) diseases including cancer, Alzheimer's disease, type 2 diabetes and OA, risk is conferred by the inheritance of multiple genetic variants that each require detailed investigation[106,107].

Over the last two decades, genome-wide association studies (GWAS) have revealed an enormous number of genetic variants associated with a plethora of traits. These studies typically utilise very large cohorts of individuals with a particular trait (e.g. knee OA) and controls (e.g. healthy individuals without OA) to determine which variants, commonly single nucleotide variants (SNVs), are associated with the trait[108–110]. Variants associated with polygenic traits tend to have small odds ratios (OR < 1.5), conferring small effects upon phenotype[111]. Stringent *P*-values are essential for the identification of truly associated variants, accounting for the high number of statistical associations genome-wide identified by multiple testing and therefore negating the high false discovery rate[112,113]. GWAS are biased towards European populations, limiting understanding of genetic architecture across all humans, though in recent years East Asian cohorts and others have contributed to GWAS participant diversity[106,114]. The number of disease-associated variants discovered has tripled over the last five years, emphasising the highly polygenic nature of many common, complex diseases[106].

Most GWAS variants reside within noncoding regions of the genome, suggesting they confer their pathogenic effects by modulating gene expression[115]. However, variants do not always modulate the expression of the most proximal gene, interacting with physically proximal but genomically distant target loci through the formation of enhancer-promoter loops[116]. Variants reported by GWAS typically reside within large linkage disequilibrium (LD) blocks with other co-inherited SNVs and act as a proxy for the causal variant[107]. Causal variants with an OR < 1.5 are less likely to be reported as the lead signal, impeding their identification[111]. Similarly, SNVs in LD blocks may exert haplotypic effects, acting in concert rather than individually[117]. In LD blocks where more than one variant is associated with a trait, conditional analyses are performed to determine whether association signals are independent of one another[118]. Biological interpretation of GWAS variants is further impeded by tissue-specificity and determining whether variants exert additive effects[88,107].

To increase confidence in the likelihood of reported GWAS variants being causal rather than proxies of genetic association signals, fine-mapping is routinely applied post-GWAS[119]. The general strategy of fine-mapping partitions the genome into subregions that confer independent effects upon the trait. These subregions are then subject to statistical fine-mapping to determine a credible set of SNVs with a high probability of being causal[120]. Genomic regions can then be annotated with functional genomic datasets to determine SNVs likely to be functional[111,119].

All reported GWAS signals require functional follow-up laboratory studies to provide evidence of a causal role and to elucidate their gene targets[107]. However, these are often time-consuming and expensive to perform. These studies therefore rely upon informative fine-mapping and functional annotation to allow for prioritisation of likely causal variants. Functional follow-up studies can then be intelligently applied to elucidate the biological mechanisms through which variants exert their trait-associated effects, without which translation of GWAS discoveries into therapeutic candidates is limited[87]. At present, the number of functional studies lags far behind the number of SNV-trait associations, leading to a bottleneck in their translation[107].

1.4.2 Genetics of OA

The heritability of OA is approximately 23% at any joint site, 15% for knee OA, and 52% for hip OA[45]. To date, more than 100 independent OA association signals have been reported[89,121–124]. As observed in other polygenic traits, most OA risk variants reside within noncoding regions of the genome[87]. Some variants are associated with all forms of OA, whilst others are specific to a particular joint site or even one of the sexes. A signal associated with hip arthroplasty (rs10282983, *C8orf34* intronic variant) confers opposing OR directional effects between females and males, highlighting sex-specific pleiotropy in OA[89]. Most OA risk variants confer modest effect sizes ($OR < 1.5$). It is therefore hypothesised that the inheritance of multiple risk alleles leads to an accumulation of gene dysregulation, resulting in OA pathogenesis[87]. This theory is known as the “liability threshold of polygenic disease”[125].

Fine-mapping of OA GWAS data and integration with human and mouse musculoskeletal data has identified many ‘high-confidence’ putative effector genes prior to functional analysis, providing candidates for functional prioritisation. Many of these candidate gene targets are associated with joint development and homeostasis, including several collagens (*COL11A1*, *COL27A1*, *COL2A1*); the osteoblastic TF *RUNX2*; members of the TGFβ signalling pathway (*SMAD3*, *TGFB1*); growth factors (*IGF1R*, *GDF5*); the inflammatory cytokine *IL11*; and proteins implicated in other important joint signalling pathways (*FGF18*; *FGFR3*; *WNT1*)[89]. Causal variants have been reported for *GDF5* (rs143383 and rs6060369), with the risk alleles impacting upon transcriptional activity and joint development, respectively[126,127].

1.4.3 Developmental origins of OA

An increasing body of evidence suggests that OA genetic risk is active during development. Variants selectively conserved during evolution may promote joint shape that is beneficial in those of reproductive age but confers detrimental effects during ageing[128]. Richard *et al* demonstrated that the OA risk variant rs6060369 alters knee morphology via changes to *Gdf5* expression in mice[127]. However, the modification is subtle enough to be tolerated during development and early postnatal locomotion, suggesting that alterations in joint shape in mice and humans may lead to OA as we age, accumulating deleterious effects such as abnormal joint loading and resulting in excessive cartilage degradation[127]. In this scenario, it is probable that these variants have accumulated in the population due to no selective pressure acting upon them during reproductive age, achieving noticeable frequency via genetic drift[127], or antagonistic pleiotropy, where phenotypic effects that are beneficial during early adulthood have negative consequences as we age[129]. The deleterious effects of genetic drift and pleiotropic variants therefore has significant consequences as populations age and the incidence of OA increases[87]. In addition to the *GDF5* locus, other OA association signals have been reported to map to genes implicated in joint development and skeletogenesis, including *CHST3* (rs3740129), *SOX5* (rs10842226) and *WWP2* (rs34195470 and rs6499244)[89].

1.5 Epigenetics and OA

1.5.1 Epigenetics

Epigenetics describes the mechanism through which changes to gene expression are caused without changing the underlying genomic DNA sequence[130]. The three core epigenetic mechanisms are DNA methylation (DNAm), post-translational histone modifications and noncoding regulatory RNAs. These epigenetic mechanisms exert their effects upon gene expression directly or indirectly, via changes to chromatin accessibility and TF binding, and are susceptible to change across life and in disease[88,131]. Epigenetic alterations are considered one of the key hallmarks of ageing, manifesting in multiple age-associated diseases including OA[88,131,132].

DNAm is the presence of a methyl group covalently attached to cytosine residues (forming 5-methylcytosine, 5mC) most commonly at cytosine-guanine dinucleotides (CpGs) and represents the most widely studied epigenetic mark[133]. The establishment of CpG methylation is performed by DNA methyltransferase (DNMT) enzymes that perform *de novo* addition of methyl groups (DNMT3A and DNMT3B) whilst the maintenance of DNAm levels is conducted by DNMT1 and DNMT3L[134]. De-methylation of CpGs can be transient following functional loss of DNAm machinery and during cell division, or as an active process mediated by Ten-Eleven Translocation (TET) enzymes which oxidise 5mC, returning CpGs to their unmodified, unmethylated state[135]. This process, known as DNA hydroxymethylation, is implicated in both mammalian development and disease, acting as an intermediary step during DNA de-methylation[136,137]. CpGs are underrepresented across the genome but often appear in high-density clusters known as CpG islands[138,139]. CpG islands are often found in gene regulatory regions including promoters and enhancers[139]. DNAm in gene promoters is associated with epigenetic silencing, whilst at enhancer regions the impact upon target gene expression appears to be locus specific[88,139,140].

Post-translational histone modifications are diverse and play a key role in chromatin structure and accessibility with functional consequences upon gene expression. Compact chromatin with strong histone-DNA interactions (heterochromatin) prevents access for transcriptional machinery whilst open, accessible chromatin (euchromatin) facilitates gene expression mediated by TF binding[141]. Histone acetyltransferases (HATs) and deacetylases (HDACs and Sirtuins, SIRTs) modulate acetylation of lysine residues, altering the charge affinity between histones and DNA, therefore modulating TF binding potential[141]. Other modifications, including histone phosphorylation, methylation and ubiquitylation, are also present[141]. Specific modifications correlate with regulatory elements and can therefore be used to determine chromatin state of genomic regions. For example, Histone 3 Lysine 4 mono-methylation (H3K4me1) and H3K27 acetylation (H3K27ac) are typically associated with enhancer activity, whilst H3K4 tri-methylation (H3K4me3) modifications signal promoter activity, and H3K27me3 indicates transcriptionally repressed regions[88,142].

Noncoding regulatory RNAs are functional molecules of several classes, including long noncoding RNAs (lncRNAs) and microRNAs (miRNAs). lncRNAs are greater than 200 nucleotides in length, poorly conserved and have many diverse roles including gene regulation[143]. miRNAs are short (~22 nucleotides in length), double-stranded RNA molecules that are evolutionarily conserved and act to negatively regulate gene expression at the post-transcriptional level by binding to specific target sequences within the mature mRNA[144]. miRNAs require processing from their longer primary miRNA transcript (pri-miRNA), initially via Drosha in the nucleus, which removes the hairpin base of the pri-miRNA, before being exported into the cytoplasm where it is cleaved by Dicer, producing a miRNA duplex that can interact with Argonaute proteins. The strand is separated within the Argonaute complex, retaining a single miRNA strand that represses mRNA targets complementary to the seed sequence and forms the RNA-induced silencing complex (RISC)[145,146]. The retention of one of the strands, referred to as -5p and -3p, is governed by sequence and structural features of the miRNA duplex[147].

1.5.2 DNAm and OA

In MSCs and chondrocytes, changes in DNAm levels have been associated with changes in the expression of ECM components (*COL2A1*, *COL9A1*, *COL10A1*)[148,149], ECM-degrading enzymes (*MMP13*, *ADAMTS4*)[150,151], inflammatory markers (*IL1B*)[152], chondrogenic and osteogenic TFs (*SOX9*, *RUNX2*)[153] and growth factor signalling proteins (*BMP7*, *GDF5*)[154,155]. Differentially methylated genes are also reported in OA subchondral bone, including *RUNX2* and *TGFB1*[156]. In cartilage, distinct methylome signatures are present between joint sites[157], disease state[158] and in development[159], suggesting these changes may have functional consequences.

Differences in DNAm levels epigenome-wide mediated by changes in DNMT/TET expression may contribute to OA risk. Cartilage-specific knockout of *Dnmt3b* in mice induces early-onset OA whilst overexpression protected mice from surgically induced OA[160]. The same study showed *DNMT3B* is downregulated in human OA chondrocytes in comparison to healthy donors[160]. *TET1* expression has also been reported to be downregulated in OA chondrocytes[137]. Together, these findings suggest an important role for DNAm in the maintenance of cartilage homeostasis. Most tissues exhibit global hypomethylation during ageing driven by a loss of maintenance during cell division, but some loci become hypermethylated with age[161]. Interestingly, most of these changes affect noncoding regions of the genome[162].

Approximately 20% of DNAm is attributed to genetics[163,164], with co-localisation analyses revealing approximately one quarter of OA risk SNVs associate with DNAm at proximal CpG sites[87]. Correlations between DNAm and SNV genotype are known as methylation quantitative trait loci (mQTLs). mQTLs have been reported in developmental and OA cartilage, synovium, subchondral bone, infrapatellar fat pad, and blood[158,165–170]. Cartilage mQTLs are enriched in enhancer regions[88], with several functional studies demonstrating that these OA mQTL regions act as functional intermediaries through which OA risk variants exert their regulatory effects (Fig. 1.4)[166,171–173]. More recently, Mendelian randomisation has been used to estimate the causal effect of DNAm upon OA at mQTL sites in cartilage, synovium and infrapatellar fat pad, revealing nearly 100 CpGs where changes in DNAm levels are likely to be implicated in OA pathogenesis[158,169].

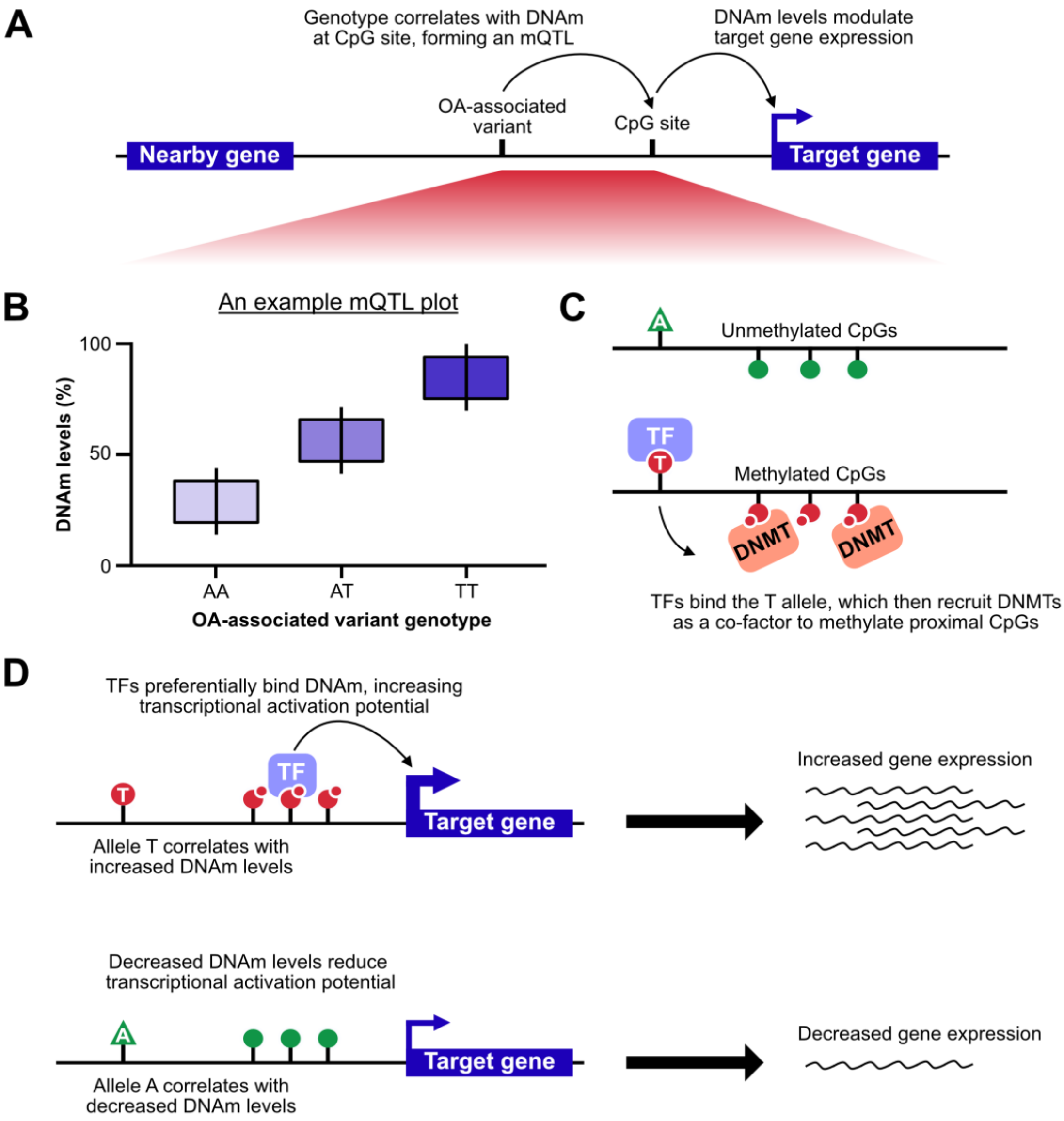


Figure 1.4: The relationship between OA risk variants, their associated mQTLs and target gene expression. A) Schematic representation of DNAm as a functional intermediary between an OA association signal, a proximal CpG site where DNAm levels correlate with the OA-associated variants genotype (an mQTL), and a target gene whose expression is modulated by changes to DNAm levels at the CpG site. Genes are represented as blue boxes, intergenic space is represented by black lines. **B)** An example of an mQTL box-plot, where DNAm levels at a particular CpG site are stratified by an OA-associated variants genotype. **C)** The proposed mechanism for how variants drive changes to DNAm levels at proximal CpGs. Top, the A allele leads to no changes in DNAm levels at the unmethylated CpGs. Bottom, the T allele facilitates TF binding. The TF recruits DNMTs as co-factors, increasing the levels of DNAm at proximal CpGs. **D)** The proposed mechanism for mQTLs as functional intermediaries, modulating target gene expression. Top, the T allele correlates with increased DNAm levels at the CpG region. TFs preferentially bind methylated CpGs (in this example), increasing transcriptional activation potential and increased target gene expression. Bottom, the A allele correlates with decreased

DNAm levels. TF binding is not facilitated, resulting in decreased transcriptional activation potential and decreased target gene expression. Wavy lines represent mRNA transcripts.

1.5.3 Post-translational histone modifications and OA

Histone modifications are required for normal cartilage development, with *in-vivo* murine studies demonstrating that knockout of *Hdac3*, *Hdac4*, *Hdac5* and *Hdac7* impairs endochondral ossification[174–176]. Expression of *HDAC4* is decreased in OA patients compared to healthy cartilage samples[177]. Mechanistically, HDAC4 is required for inhibition of *RUNX2*-mediated activation of *MMP13*, increased expression of *COL2A1* and *ACAN*, and suppression of the inflammatory cytokine IL-1β, highlighting an important role of HDACs in cartilage homeostasis[177]. SIRTs are also required for cartilage homeostasis. In chondrocytes with mitochondrial dysfunction (a hallmark of ageing and senescence[132]), increased *SIRT1* expression can reverse loss of mitochondrial biogenesis capacity[178]. *Sirt6* knockout mice exhibit skeletal defects whilst *Sirt6* overexpression in mice protects against surgically induced OA[179,180]. In humans, *SIRT6* depletion is associated with cellular senescence and increased *MMP1* and *MMP13* expression[181].

OA risk variants map to post-translational histone modification proteins, including the histone acetylation/de-ubiquitination enzyme *SUPT3H* and the histone methyltransferase *DOT1L*[124,182]. Upregulation of *SUPT3H* is observed with the OA risk allele in the cartilage, synovium and trabecular bone of OA patients[170]. Loss of H3K79me1 mediated by *DOT1L* reduced the expression of *COL2A1* and *ACAN* in human chondrocytes, whilst intra-articular injection of a *DOT1L* inhibitor into the knees of mice resulted in cartilage degradation[88,183]. Dysregulation of post-translational histone modifications, which in turn alter chromatin state and accessibility, therefore contributes to OA development and progression[88].

1.5.4 Noncoding RNAs and OA

miRNAs are crucial for skeletogenesis and cartilage homeostasis, with their dysregulation associated with OA and disease progression[184]. Conditional knockout of the miRNA processing machinery Dicer and Drosha in limb mesenchyme and the growth plate reduces chondrocyte proliferation and increases hypertrophy, resulting in skeletal growth defects and smaller limbs[185–187]. miRNAs are important in chondrogenesis, with miR-455-3p inhibiting *RUNX2* expression *in-vitro* during early chondrogenesis and miR-140 knockout mice exhibiting skeletal defects[188,189]. Interactions between the TGFβ signalling pathway and miRNAs have been reported. miR-29 and TGFβ downregulate one another, whilst miR-455 expression can be induced by TGFβ in chondrocytes which in turn inhibits TGFβ signalling by SMAD2, highlighting feedback loops and suggests miRNAs aid in regulating steady-state levels of TGFβ components[184,190,191]. miRNAs are also implicated in chondrocyte senescence and may be used as a biomarker for OA due to their stability in the circulatory system[192–194]. Many miRNAs are reported to be differentially expressed between young adult and aged cartilage[195], and between OA and non-OA cartilage samples from aged donors[196]. In the synovium, differential expression of miRNAs has been reported between early- and late-stage OA[197], whilst intra-articular injection of miR-26a into the knee joints of rats has been shown to reduce synovitis[198]. Several OA risk SNVs map within or close to the host genes of miRNAs and to their processing machinery, including miR-455 (rs1078301, rs919642; host gene *COL27A1*), miR-140 (rs34195470, rs6499244; host gene *WWP2*), miR-8068 (rs17659798, intergenic) and *DICER1* (rs28929474)[89,121,123].

Many lncRNAs are differentially expressed in cartilage samples taken from healthy adults and OA patients[199]. Deletion of the cartilage-specific lncRNA ROCR in MSCs prevents chondrogenesis via inhibition of *SOX9* expression[200]. lncRNA-HIT promotes histone acetylation for normal skeletal development in mice[201]. Intra-articular injection of the lncRNA HOTAIR into the knees of rats induced ECM degradation via modulation of the Wnt/β-Catenin signalling pathway[202]. A single OA risk SNV has been mapped to a lncRNA (rs11105466, LINC02399)[121], though this was not replicated in the largest OA GWAS to date[89].

1.6 Functional follow-up studies

1.6.1 From GWAS signals to effector genes

Whilst variants within coding regions of the genome can easily be linked to effector genes due to changes in the amino acid sequence, variants residing within noncoding regions of the genome are harder to interpret. Noncoding variants are thought to confer their pathogenicity via changes to target gene expression[87]. However, their target gene is often not the one genomically closest to the variant due to the 3D topography of the genome. Variants may in fact alter the expression of genes that are physically proximal yet genomically distal due to chromatin looping[88]. To further complicate the identification of effector genes, some variants only exert their pathogenic effects in particular tissues[87,88]. This is problematic for OA risk signals, where multiple tissues are affected[42]. Furthermore, tissues such as synovium and subchondral bone consist of multiple cell types, where some may exert the observed pathogenic effect upon OA risk whilst others are benign. Single-cell approaches provide an emerging strategy to demarcate tissues and cell types that play an active role in OA pathogenesis[203,204]. In addition, some variants confer their pathogenicity during development rather than in later life, for example during joint development[87,127], meaning that the variants functional timepoint must be considered[87,165]. To alleviate these issues, researchers should utilise three tools: functional genomic datasets, QTL analyses and downstream functional follow-up laboratory studies[87].

1.6.2 Utilisation of functional genomic datasets

Integration of GWAS variants with functional genomic datasets (*in-silico* analysis) can aid in characterising their putative regulatory role in multiple cell types relevant to OA. For noncoding variants that likely exert their functional effects via modulation of gene expression, understanding the local chromatin context in OA-relevant tissues is crucial[87,88]. Segmentation of the genome into regulatory states depending on post-translational histone modifications data derived from Chromatin Immunoprecipitation with high-throughput Sequencing (ChIP-Seq), provided by the epigenomics projects ENCODE and FANTOM, can be used to identify regions associated with promoter or enhancer activity[205–207]. Integration with Assay for Transposase Accessible Chromatin with high-throughput Sequencing (ATAC-Seq) identifies regions of open chromatin that are accessible to transcriptional machinery, suggestive of a functional role[165,208,209]. Utilising chromatin conformation capture technologies such as Capture Hi-C can aid in understanding the 3D spatial organisation of the genome and determine whether OA-associated variants are within close physical proximity to a gene promoter[210–212]. Together, these data can inform downstream *in-vitro* laboratory studies.

1.6.3 QTL analyses

To determine tissues where OA risk variants are likely to be functional, gene expression data from available patient samples or healthy donors can be stratified by OA risk signal genotype[87–89]. Significant correlations represent expression QTLs (eQTLs) and demonstrate a relationship between OA risk variants and putative effector genes, although causality must still be established by functional laboratory experiments. Unfortunately, there is a dearth of large-scale eQTL datasets available for OA-relevant tissue types. Furthermore, detecting eQTLs using gene expression data can be limited by interindividual variability that masks correlations in smaller sample cohorts[87]. Databases like GTEx therefore require very large sample cohorts to detect eQTLs[213]. Allelic expression imbalance (AEI) analysis, which determines the relative ratio of mRNA transcripts produced by each allele of a SNV, is a complementary and highly sensitive approach that overcomes this problem[87]. Where variants reside within intronic or intergenic DNA regions, proxies in high LD ($r^2 > 0.8$) and within the mRNA transcript sequence of the gene can be used to determine the effects of OA risk variants upon gene expression. AEI analyses have been performed in multiple OA-relevant tissue types, including cartilage, subchondral bone, fat pad and synovium, aiding in the identification of several OA effector genes[166,171–173,214,215]. mQTL analyses also provide an informative strategy for determining effectors of OA genetic risk, where mQTL sites act as functional intermediaries (see Section 1.5.2)[166,168,170–173]. DNAm data can also be correlated against expression data, revealing methylation-expression QTLs (meQTLs)[166,171,216]. The presence of an meQTL is suggestive of a functional mechanism for DNAm in driving changes to gene expression mediated by an OA risk variant[87].

1.6.4 Functional laboratory experiments

Following *in-silico* and QTL analyses of an OA risk locus, a credible set of putative effector genes can be identified. Functional follow-up studies can then be performed to establish a causal relationship between the risk variant and its target gene[87]. These experiments are routinely performed using cartilage cell lines, although studies are now expanding to include cell lines of other OA-relevant tissues such as synovium[167]. To determine the regulatory function of DNA regions housing OA risk variants or their associated mQTLs, reporter gene assays can be performed. These have been used to demonstrate the transcriptional effects of OA risk alleles as well as the effects of DNAm at mQTL sites, providing evidence that these regions are capable of modulating gene expression in an allele- or DNAm-dependent manner[127,171–173,217,218].

Targeted editing of the genome and epigenome has been revolutionised in the last decade following the advent of the CRISPR/Cas9 toolbox[219]. CRISPR/Cas9 deletion of putative regulatory elements associated with OA has led to the identification of several effectors of OA genetic risk, including *COLGALT2*[173], *PLEC*[168], *RUNX2*[170] and *TGFB1*[171]. More recently, the development of epigenetic editors which fuse catalytically inactive 'dead' Cas9 (dCas9) to enzymes that can modulate DNAm levels at proximal CpGs, namely dCas9-DNMT3A and dCas9-TET1, have been utilised to identify effectors of OA genetic risk[216,220]. Targeted epigenetic editing to methylate or de-methylate CpGs previously identified as OA-associated mQTLs has been used to identify several effector genes in which DNAm levels act as a functional intermediary of OA genetic risk[166,171–173].

Once target genes are established, researchers may choose to focus upon the mediators that facilitate changes to gene expression, such as TFs. The presence of a SNV can either create or destroy a TF binding site, or modulate binding affinity, resulting in changes to transcriptional potential[88]. Similarly, the addition of a methyl group to CpG sites can affect TF binding via steric hindrance, highlighting a possible mechanism through which mQTLs exert their effects (Fig. 1.4D)[221]. Computational prediction of TF binding sites (TFBS) using tools such as JASPAR and SNP2TFBS is therefore beneficial[222,223]. TFs abundantly expressed in relevant tissue types may point towards a functional role[107]. Following identification of TFs predicted to bind OA risk variants or their associated mQTL sites using available databases, their binding affinity can be validated using focussed electrophoretic mobility shift assays (EMSAs) or genome-wide TF ChIP-Seq[127,224]. EMSAs, which study interactions between nuclear protein and DNA probes matching sequences of interest, have been used to demonstrate preferential allelic binding of TFs expressed in chondrocytes for several OA risk variants[171,225,226].

1.7 *WWP2* and miR-140

1.7.1 WWP2

WWP2 encodes WW Domain Containing E3 Ubiquitin Protein Ligase 2, an enzyme active in protein ubiquitination cascades that result in the post-translational addition of ubiquitin to target proteins[227]. Ubiquitin ligase (E3) enzymes work in conjunction with ubiquitin-activating (E1) and ubiquitin-conjugating (E2) enzymes to mediate the transfer of ubiquitin to substrates, altering their cellular localisation or marking them for proteasomal degradation[228]. By regulating steady-state levels of target proteins, the ubiquitin-proteasomal pathway controls many cellular processes, including the cell cycle, apoptosis and the immune response[227,229–231]. E3 ligases can be categorised into two main classes: RING domain E3 ligases, which represent the majority of this group of enzyme and act as substrate scaffolds, and HECT domain E3 ligases[229]. The HECT domain possesses intrinsic ligase activity and directly mediates the covalent attachment of ubiquitin to substrate proteins[229]. *WWP2* is a member of the NEDD4 superfamily subdivision of HECT E3 ligases, which are characterised by an N-terminal C2 domain which facilitates E2 binding, WW domains that confer substrate specificity, and a C-terminal HECT domain[229]. The WW domains bind proteins containing PPxY (double proline-X-tyrosine) or similar motifs[232]. The specificity of HECT E3 ligases is further refined by variant protein isoforms produced by alternative splicing or distinct promoter sequences[233,234].

Three common isoforms of WWP2 protein have been characterised of varying amino acid (aa) length and encoding differing combinations of protein domains (Fig. 1.5). Full-length WWP2 (WWP2-FL, 870aa, RefSeq ID: NM_007014) contains three functional components: the C2 domain, four WW domains (WW1-WW4), and the catalytic HECT domain. The N-terminal isoform (WWP2-N, 336aa, RefSeq ID: NM_001270455) encodes the C2 and WW1 domains, whilst the C-terminal isoform (WWP2-C, 440aa, RefSeq ID: NM_199424) encodes the WW4 and HECT domains[231]. The generation of WWP2-N is thought to be the result of a splicing failure at intron 9-10 resulting in an alternate polyA site, whilst WWP2-C is generated from a distinct internal promoter within intron 10-11[234].

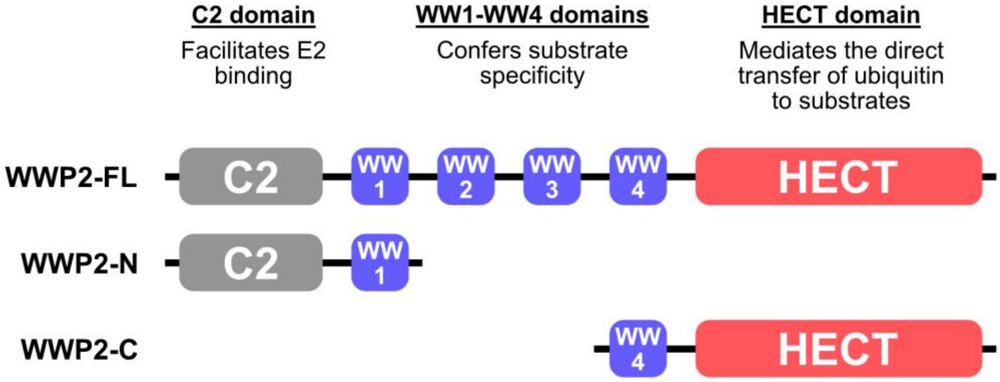


Figure 1.5: Protein domains encoded by WWP2-FL, WWP2-N and WWP2-C. Protein domains (C2, WW1-WW4, HECT) are represented by coloured boxes.

WWP2 is a regulator of TGFβ signalling. In the absence of TGFβ, WWP2-N and WWP2-FL complex to facilitate the rapid degradation of unstimulated SMAD2 and SMAD3. In the presence of TGFβ, SMAD2 and SMAD3 are phosphorylated and become the preferred binding partner of WWP2-N. This blocks WWP2-FL mediated degradation of SMAD2/3 and re-directs

WWP2-FL to act in concert with WWP2-C to rapidly degrade the inhibitory SMAD7, subsequently maintaining TGFβ signalling [231,234,235]. The levels of each isoform are therefore required to fine-tune TGFβ signalling (Fig. 1.6), influencing cellular phenotype[230,231,234–236]. *Wwp2* knockout in mice induces an OA-like phenotype with articular cartilage degradation[237]. WWP2 is required for polyubiquitination and subsequent degradation of RUNX2 in chondrocytes, targeting the PY motif of RUNX2[237]. Interestingly, WWP2 is capable of mono-ubiquitinating RUNX2 in osteoblasts but this does not lead to protein degradation, instead leading to augmentation of its activity and positively regulating osteogenesis[238]. Other E3 ligases, including the NEDD4 family members SMURF1, SMURF2 and WWP1 are also capable of polyubiquitination and subsequent degradation of RUNX2[239,240]. Another target of WWP2 is Goosecoid (GSC)[241], a paired-like homeobox protein required for craniofacial development, though the impact of *WWP2* upon craniofacial development has since been rebutted in favour of a role for miR-140[242]. In humans, upregulation of the WWP2-FL isoform in 3D pellet cultures of OA chondrocytes induces downregulation of *COL2A1* and *ACAN*, suggesting a loss of cartilage integrity[243].

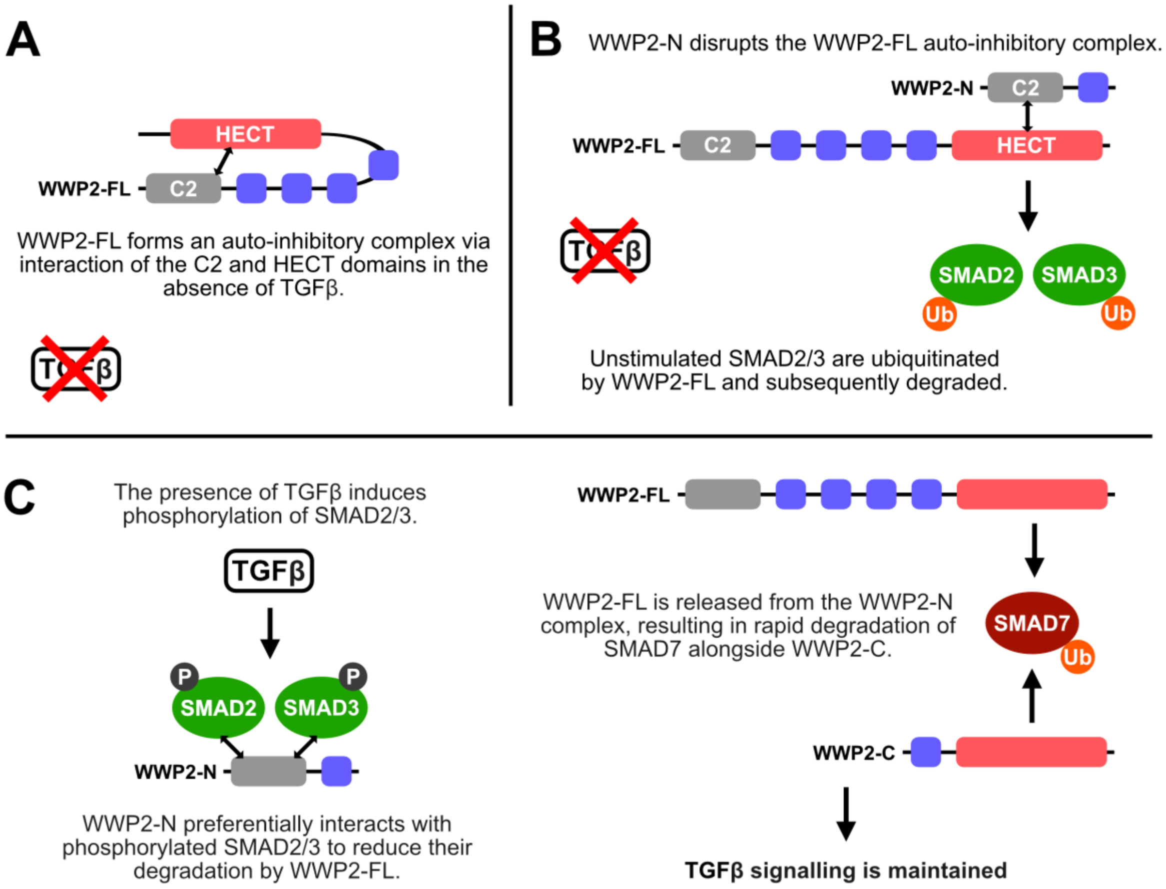


Figure 1.6: WWP2 isoforms target SMAD signalling proteins for degradation. A) In the absence of TGFβ, WWP2-FL forms an auto-inhibitory complex via interaction between the N-terminal C2 domain and the C-terminal HECT domain. **B)** WWP2-N disrupts the WWP2-FL auto-inhibitory complex, facilitating WWP2-FL-mediated ubiquitination of unstimulated SMAD2/3 and subsequent degradation. **C)** In the presence of TGFβ, SMAD2/3 are phosphorylated. WWP2-N preferentially binds phosphorylated SMAD2/3 instead of WWP2-FL, reducing their degradation. WWP2-FL is released by WWP2-N to target its preferred substrate, the inhibitory SMAD7. WWP2-FL and WWP2-C work in concert to rapidly degrade SMAD7, maintaining downstream TGFβ signalling. P, phosphorylation. Ub, ubiquitin. Adapted from Soond *et al.*, 2011[234].

1.7.2 miR-140

WWP2 is the host gene of miR-140, residing within an intron shared between *WWP2*-FL and *WWP2*-C (Fig. 1.7). miR-140 is co-expressed with *WWP2*-C and widely recognised to be a cartilage-specific marker, suggesting important roles for cartilage homeostasis[184,244]. Processing of the pri-miRNA-140 transcript by Drosha and Dicer produces two mature strands (miR-140-5p and miR-140-3p), which are differentially expressed and have differing targets in cartilage[245,246]. These targets include regulators of TGFβ and Wnt signalling (*SMAD3*, *FZD6*)[247,248], the histone deacetylase *HDAC4*[249], and the ECM-degrading enzyme *ADAMTS5*[189]. miR-140-5p is required for chondrogenesis, indicating an important role during development[248,250]. miR-140 expression is directly regulated by SOX9, in combination with SOX5 and SOX6, during chondrogenesis[248,251]. miR-140 has also been reported to be regulated by DNAm, with higher levels of DNAm at CpGs in the miR-140 upstream regulatory region decreasing SMAD3-mediated transcriptional activation[252].

Small animal studies have shown that miR-140 knockout in mice results in craniofacial truncation due to impaired chondrogenesis[189,253], whilst in zebrafish aberrant miR-140 expression results in palatogenesis defects[254]. In the articular joint, loss of miR-140 in mice leads to cartilage degradation[237]. Double knockout of miR-140 and *Wwp2* exacerbates loss of cartilage, suggesting they act cooperatively to maintain cartilage integrity[237]. A gain-of-function mutation within miR-140-5p has been reported for a human skeletal dysplasia[255]. miR-140 has also been highlighted as a potential biomarker of OA, with circulating miR-140-3p downregulated in OA patients compared to healthy individuals as observed in articular cartilage[193]. miR-140 may be protective against OA, with intra-articular injection of miR-140-5p decreasing incidence of OA in rats[256].

1.7.3 OA risk variants mapping to the WWP2/miR-140 locus

In 2018, a GWAS meta-analysis of Icelandic and UK datasets reported a knee OA association signal mapping to the *WWP2*/miR-140 locus[123]. The lead variant, rs34195470 (A > G), is located within an intron of *WWP2*, with the G allele conferring increased OA risk (OR = 1.07)[123]. In 2019, a GWAS of the UK Biobank and arcOGEN datasets reported another OA association signal close to the *WWP2*/miR-140 locus, this time with knee or hip OA[121]. The lead variant, rs6499244 (T > A), resides upstream of *WWP2* within the 3' untranslated region (UTR) of *NFAT5*. The A allele of rs6499244 confers increased OA risk (OR = 1.06)[121]. The pairwise LD between the two variants is modest in European ancestry cohorts but sufficiently high to suggest they are marking the same association signal ($r^2 = 0.22$, D' = 0.51). In 2021, the largest OA GWAS also reported rs34195470 and rs6499244 as associated variants[89]. rs34195470 was highlighted as the lead one of the two, and a potentially causal variant following conditional analyses, with an association *P*-value of 3.13×10^{-13}. Several GWAS signals map to genes encoding members of the TGFβ signalling pathway, suggesting that this pathway, and regulators of it (including *WWP2* and miR-140) are targets of OA genetic risk[89].

rs34195470 is in high LD ($r^2 > 0.8$) with two other variants: rs9746247 and rs111837947. rs9746247 (C > G) has never been identified to associate with any trait, whereas the indel rs111837947 (insertion of GT, insGT) has previously been reported to associate with surgical and nonsurgical (radiographic) knee OA[122]. The insGT allele of rs111837947 confers increased OA risk (OR = 1.05-1.09)[122]. Both variants are housed within the same intron as rs34195470 and therefore do not alter the amino acid coding sequence of *WWP2* or the sequence of miR-140, implying that the genetic signal mapping to this locus mediates its effect via changes to gene expression.

In 2019, AEI analysis reported several *WWP2* transcript SNVs that associated with AEI of the gene in OA cartilage[214]. The most significant imbalance was at rs1052429 (G > A), with a relative increased expression of allele A (56% of this allele versus 44% for allele G), equivalent to an A/G ratio of 1.27[214]. rs1052429 has a pairwise r^2 and D' of 0.41 and 0.91 with rs34195470 respectively, which means that the OA risk-conferring allele G of rs34195470 nearly always occurs on a haplotype containing allele A of rs1052429 in European ancestry cohorts. The OA risk allele of rs34195470 therefore correlates with increased *WWP2* expression in cartilage. Schematic representation of the genomic region housing *WWP2* and miR-140 and their proximity to rs34195470, rs6499244 and rs1052429 is shown in Fig. 1.7.

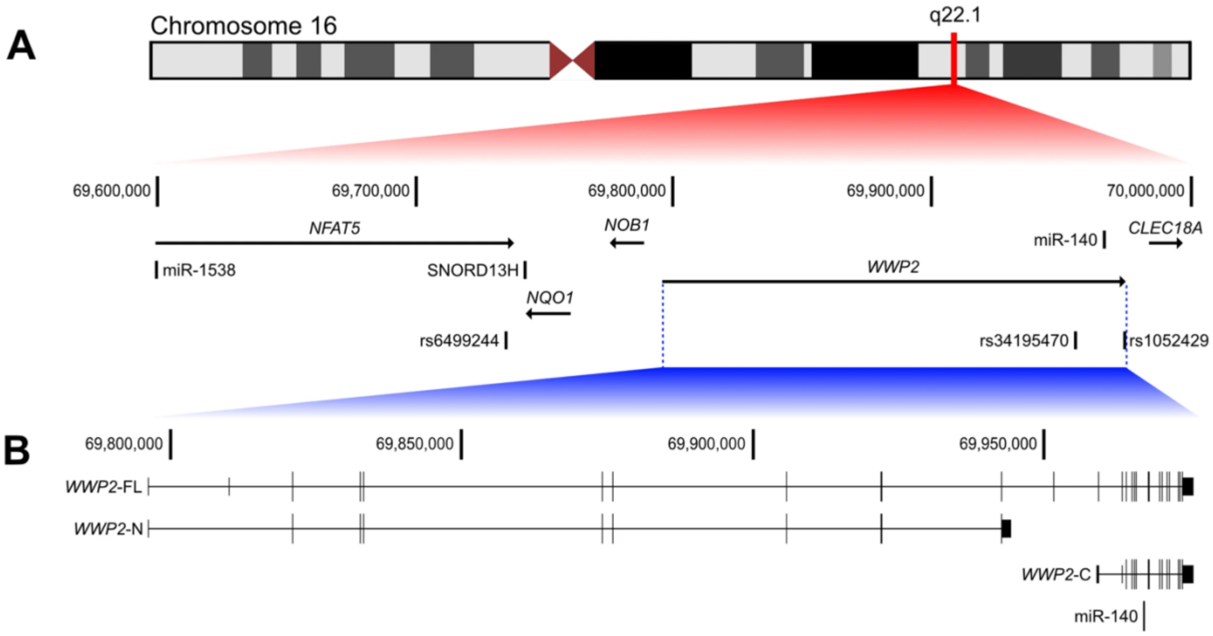


Figure 1.7: The genomic region encompassing *WWP2* and miR-140. A) Schematic representing Chromosome 16. The region q22.1, which houses *WWP2* and miR-140, is expanded to show all neighbouring genes across 400Kb. Protein-coding genes are represented

by arrows. Arrowheads denote orientation of transcription (pointing left, antisense; pointing right, sense). The miRNAs miR-1538 and miR-140 and the small nucleolar RNA SNORD13H are represented by vertical bars. Two OA risk SNVs mapped to the *WWP2* locus (rs6499244, 3' UTR of *NFAT5*; rs34195470, distal end of *WWP2*) and the proxy transcript SNV used previously in AEI analysis (rs1052429, 3' UTR of *WWP2*) are also shown, represented by vertical bars. **B)** The three core *WWP2* transcript isoforms and miR-140 are shown. Half-height vertical bars = 5' and 3' UTRs; full-height vertical bars = exons; horizontal bars = introns.

1.7.4 OA cartilage mQTLs at the WWP2/miR-140 locus

Using genome-wide DNAm array data created from OA arthroplasty patient cartilage samples, my principal investigator's research group has previously reported that genotype at rs34195470 and rs6499244 correlates with DNAm levels of CpGs located within the gene body of *WWP2*, forming mQTLs[87,257]. Both variants associate with DNAm levels of the same CpGs, cg26736200 and cg26661922, and for each CpG the OA risk-conferring alleles of rs34195470 and rs6499244 associate with increased DNAm levels[87,257]. cg26736200 and cg26661922 are 115bp apart, less than 4Kb from rs34195470, and reside within a region marked as transcriptionally active[257]. This implies that DNAm may be a functional intermediary between the rs34195470 association signal and changes in target gene expression, with the causal variant altering DNAm levels of the transcriptional regulator which then alters expression of the target gene.

1.8 Aims and hypotheses

In this thesis, I conducted a detailed molecular genetic and epigenetic analysis of OA risk marked by the rs34195470 association signal. The risk-conferring G allele of rs34195470 has previously been reported to correlate with increased *WWP2* expression in OA cartilage[214]. In addition, the risk-conferring G allele of rs34195470 has previously been reported to correlate with increased DNAm levels at two CpGs within the gene body of *WWP2*[87]. The biological relevance of *WWP2* and miR-140 in cartilage development and homeostasis make the rs34195470 association signal a particularly compelling target for further investigation. Based on the previous reports and the available literature, the following hypothesis is proposed:

> "Increased DNAm levels at the two OA cartilage mQTL sites results in higher levels of *WWP2* and/or miR-140 expression in chondrocytes"

To test this hypothesis, human joint tissues from OA patients who had undergone arthroplasty and from foetal donor samples were subject to molecular genetic and epigenetic analyses to replicate and expand on the previous findings. Then, *in-vitro* functional follow-up studies and *in-silico* data analyses were performed to further characterise the rs34195470 association signal. Finally, experiments were undertaken to decipher the TFs responsible for mediating the effects of the rs34195470 association signal on gene expression and to determine the causal variant.

Chapter 2: Materials and Methods

2.1 Reagents

Table 2.1. Details of reagents used in the methods.

Reagent	Supplier	Product code
Agar	BD	213000
Agarose	Sigma-Aldrich	A9539
AllPrep DNA/RNA/miRNA Universal kit	Qiagen	80224
Ammonium persulphate (APS)	Sigma-Aldrich	A3678
Ampicillin	Sigma-Aldrich	A9393
Anti GAPDH monoclonal antibody (mAb)	Merck	MAB374
Antarctic Phosphatase	NEB	M0289S
AvrII	NEB	R0174S
BbsI	NEB	R0539S
Bis-acrylamide (40%)	Invitrogen	HC2040
Casein hydrolysate	Sigma-Aldrich	22090-100G
Chloroform	Sigma-Aldrich	650498
Cobalt(II) chloride hexahydrate ($CoCl_2$)	Merck	255599-5G
CpG methyltransferase (*M. SssI*)	NEB	M0226S
dCas9-DNMT3a-EGFP vector	Addgene	71666
Dulbecco's Modified Eagle Medium/Nutrient Mixture F12 medium (DMEM:F12)	Gibco	11320033
DNase I	Invitrogen	18068015
DNA/RNA kit	Norgen Bio-Tek	48700
dNTPs	Invitrogen	18427089
Dual-Luciferase Reporter Assay System	Promega	E1910
Ethylenediaminetetraacetic acid (EDTA)	Merck	819040
Ethanol	VWR	20821.330
EPIXTRACT Nuclear Protein Isolation kit	Enzo Life Sciences	ENZ-45016-0100
EZ DNA Methylation kit	Zymo Research	D5002
E.Z.N.A. DNA/RNA Isolation kit	Omega Bio-Tek	R6731
E.Z.N.A. Tissue DNA kit	Omega Bio-Tek	D3396
Foetal bovine serum (FBS)	Merck	F7524
Glycine	Sigma-Aldrich	G7126
GT115 competent *E. coli*	Invivogen	GT115-11
HaeII	NEB	R0107S
Hank's Balanced Salt Solution (HBSS)	Gibco	14170112
HIF-1α XP Rabbit mAb	Cell Signaling Technology	36169
Hydrochloric acid (HCl), 30%	VWR	20251.365
IgG from rabbit serum	Sigma-Aldrich	I5006
Immobilon Western Chemiluminescent HRP Substrate	Merck	WBKLS
Isopropanol	Sigma-Aldrich	34863
Lipofectamine^{TM} 2000	Invitrogen	1168500
LB Broth (Lennox)	Sigma-Aldrich	L9234
Methanol	VWR	6712-25
MluI	NEB	R3198S
MultiScribe^{TM} Reverse Transcriptase kit	Invitrogen	4311235
NuPAGE^{TM} Bis-Tris Mini Protein Gels, 4-12%	Invitrogen	NP0321BOX
NuPAGE^{TM} MOPS SDS Running Buffer (20X)	Invitrogen	NP0001
Nystatin	Sigma-Aldrich	N1638
Odyssey EMSA kit	LI-COR Biosciences	82907910
One Shot^{TM} TOP10 Chemically Competent *E. Coli*	Invitrogen	C404010
PageRuler^{TM} Plus Prestained Protein Ladder	Thermo Scientific	26619
pCpG-free Lucia Basic vector	Invivogen	pcpgf-bas
pCpG-free Lucia Promoter vector	Invivogen	pcpgf-prom

pCR-Blunt-TOPO-II vector	Invitrogen	K280002
Penicillin-Streptomycin	Gibco	15140122
pGL3-Promoter (Firefly luciferase) vector	Addgene	212939
Phire II Hot Start DNA polymerase	Thermo Scientific	F122L
Polyclonal Goat Anti-Mouse Immunoglobulins/HRP	Dako	P0447
Polyclonal Goat Anti-Rabbit Immunoglobulins/HRP	Dako	P0448
pRL-TK (Renilla luciferase) vector	Promega	E2241
Protein Quantification Assay kit	Macherey-Nagel	740967
Protein Solving Buffer - tris(2-carboxyethyl)phosphine (PSB-TCEP) solution	Macherey-Nagel	740941
PureYield™ Plasmid Miniprep kit	Promega	A1222
Puromycin	Sigma-Aldrich	P7255
PVDF Transfer Membrane	Thermo Scientific	88518
pX462 Cas9 vector	Addgene	62987
PyroMark PCR kit	Qiagen	978705
Pyromark Q24 Advanced kit	Qiagen	970902
PyroMark Wash Buffer	Qiagen	979008
QIAquick Gel Extraction kit	Qiagen	28506
QuikChange Lightning Site-Directed Mutagenesis kit	Agilent	210518
Random primers	Invitrogen	48190011
RNase H	NEB	M0297S
RNaseOUT™	Invitrogen	10777019
RNeasy kit	Qiagen	74104
Roswell Park Memorial Institute (RPMI) 1640 medium	Gibco	11875093
S-adenosylmethionine (SAM)	NEB	B9003S
Semi-Skimmed Powdered Milk	Tesco	N/A
SF Cell Line 4D X Kit L	Lonza	V4XC-2012
Sodium chloride (NaCl)	Fisher Scientific	447302500
Sodium dodecyl sulphate (SDS)	Sigma-Aldrich	L3771
Sodium hydroxide (NaOH)	Fisher Scientific	S/4880/60
SpeI	NEB	R3133S
Streptavidin Sepharose™ High Performance beads	Sigma-Aldrich	GE17-5113-01
SuperScript™ IV Reverse Transcriptase	Invitrogen	18090010
TaqMan™ Fast Advanced PCR Master Mix for qPCR	Applied Biosystems	4444558
TBE Buffer (Tris-borate-EDTA, 10X)	Thermo Scientific	B52
Tetramethylethylenediamine (TEMED)	Thermo Scientific	17919
Trizma base	Sigma-Aldrich	T1503
TRIzol™ Reagent	Invitrogen	15596026
Trypsin-EDTA (0.05%), phenol red	Gibco	25300104
Tween 20	Sigma-Aldrich	P2287
T4 DNA Ligase	NEB	M0202S
XhoI	NEB	R0146S
Zeocin	Invivogen	ant-zn-05
Zero Blunt™ TOPO™ Cloning kit	Invitrogen	451245

2.2 Methods

2.2.1 In silico analyses

This project utilised several databases, tools, and published datasets to characterise the regulatory capacity of the *WWP2* locus and inform downstream experimental studies. Details of publicly available databases and tools used are listed in Table 2.2. Four datasets providing ATAC-Seq and RNA-Seq data in OA-relevant tissues were utilised and are listed in Table 2.3. Functional data was mapped to the human genome (hg19) assembly using UCSC Genome Browser[258].

Table 2.2. List of online databases and tools used.

Database / Tool	Utility	Source	Reference
Agilent QuikChange Primer Design	Design of primers for site-directed mutagenesis.	https://www.agilent.com/store/primerDesignProgram.jsp	N/A
CRISPR Targets	Design of gRNAs for CRISPR/Cas9 deletions.	Available on UCSC Genome Browser (hg19).	259
GTEx Portal	Identification of eQTLs associated with rs34195470.	https://www.gtexportal.org/home/	213
IDT gRNA Design	Design of gRNAs for dCas9-DNMT3a epigenetic editing.	https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE	N/A
JASPAR Core 2022	Transcription factor (TF) motif prediction.	https://jaspar.elixir.no	223
JASPAR Core 2024	TF motif prediction.	https://jaspar.elixir.no	260
ReMap Atlas of Regulatory Regions	Identification of TF ChIP-Seq data in the following biotypes (all TFs): cartilage, chondrosarcoma, fibroblast, hESC, hMSC, osteoblast, SW1353, U2OS. Identification of TF ChIP-Seq data for HIF-1α and HIF-2α (all biotypes).	Available on UCSC Genome Browser (hg19).	261
ROADMAP	Identification of regulatory elements using histone ChIP-Seq data (Primary ChromHMM) in the following cell types: E006 H1-derived MSCs (MSCs); E049 MSC-derived chondrocyte cultured cells (chondrocytes); E129 osteoblast primary cells (osteoblasts).	https://egg2.wustl.edu/roadmap/web_portal/	262
SNP2TFBS	TF motif prediction.	https://epd.expasy.org/snp2tfbs/	222
UCSC Genome Browser (hg19)	Visualisation of genomic loci and mapping functional elements.	https://genome.ucsc.edu	258
3D Genome Browser	Identification of long-range chromatin interactions (looping) using Capture Hi-C data.	http://3dgenome.fsm.northwestern.edu/chic.php	263

Table 2.3. List of published datasets used.

Dataset	Data Type	GEO Accession	Reference
Chromatin accessibility (foetal chondrocytes and OA chondrocytes)	ATAC-Seq	GSE214394	165
Chromatin accessibility (Saos-2 cell line)	ATAC-Seq	GSE120755	264
Chromatin accessibility (OA fibroblast-like synoviocytes, FLS)	ATAC-Seq	GSE112658	265
Gene expression (OA chondrocytes)	RNA-Seq	GSE111358	199

2.2.2 OA patient arthroplasty and foetal donor samples

Cartilage, synovium and subchondral bone tissue samples were collected from patients undergoing total joint arthroplasty of the knee and hip for advanced primary OA by the Newcastle upon Tyne NHS Foundation Trust. Ethical approval was granted by the NHS Health Research Authority with each donor providing written consent (REC reference number 19/LO/0389). Foetal cartilage taken from the developing ends of the femur and tibia was provided by the Human Developmental Biology Resource (HDBR) at Newcastle University (project number 200363). Samples were provided with full maternal written consent and approval from the Newcastle and North Tyneside NHS Health Authority Joint Ethics Committee. Summary data is shown in Table 2.4. Full patient sample details and their use in molecular genetic and epigenetic analyses are available in Appendices A, B, C and D.

Table 2.4. Summary details for OA patient arthroplasty and foetal donor samples.

	OA cartilage	**Foetal cartilage**	**OA bone**	**OA synovium**
n	55	50	61	71
Female	30	23	33	41
Male	25	27	28	30
Hip	25	N/A	61	4
Knee	30	N/A	0	67
Femur	N/A	3	N/A	N/A
Tibia	N/A	25	N/A	N/A
Femur and Tibia	N/A	22	N/A	N/A
Age range	41-93 years	8-17 pcw	51-86 years	45-91 years
Appendix	A	B	C	D

2.2.3 Nucleic acid extraction from OA patient arthroplasty and foetal donor samples

Following arthroplasty surgery, patient samples were transferred to pots containing HBSS (Gibco) supplemented with 10,000U/mL penicillin-streptomycin (Gibco) and 10,000U/mL nystatin (Sigma-Aldrich) and stored at 4°C. Tissues were isolated from samples in a sterile laminar flow hood, then frozen on dry ice and stored at -80°C. Samples were ground using the MM400 Tissue Grinder (Retsch) with grinding receptacles cooled in liquid N_2 to inhibit DNase and RNase activity.

Nucleic acids from synovium samples (300-400mg) were extracted using the E.Z.N.A. DNA/RNA Isolation kit (Omega Bio-Tek) according to manufacturer protocol. DNA from cartilage (200-500mg) was extracted using the E.Z.N.A. Tissue DNA kit (Omega Bio-Tek) according to manufacturer protocol. DNA from bone (50mg) was extracted using the E.Z.N.A. DNA/RNA Isolation kit (Omega Bio-Tek) according to manufacturer protocol. RNA from cartilage (250mg) was extracted using phenol-chloroform phase separation and the RNeasy kit (Qiagen). Briefly, 1mL TRIzol^{TM} Reagent (Invitrogen) was added to cartilage and homogenised by vortexing. Samples were incubated at room temperature for 15 minutes and then centrifuged. The supernatant was homogenised with 200µL chloroform (Sigma-Aldrich) and centrifuged. The RNA fraction (upper aqueous phase) was then transferred to an RNeasy spin column (Qiagen) and the manufacturer protocol was followed to purify RNA. RNA from bone (200-220mg) was extracted using phenol-chloroform phase separation as previously described. Then, the RNA fraction was homogenised with 500µL isopropanol (Sigma-Aldrich) to precipitate RNA and centrifuged. The supernatant was aspirated and washed with 1mL ice-cold 75% ethanol (VWR) before final resuspension in RNase-free water. Nucleic acids from foetal cartilage samples were extracted by the HDBR using the QIAcube Automated DNA/RNA isolation instrument (Qiagen) and the AllPrep DNA/RNA/miRNA Universal kit (Qiagen) according to manufacturer protocol.

Quantification of nucleic acids from OA arthroplasty samples was performed using the Nanodrop 1000 (Thermo Scientific). Quantification of nucleic acids from foetal cartilage samples was performed using the 2100 Bioanalyzer (Agilent). DNA was stored at -20°C and RNA was stored at -80°C.

2.2.4 Bisulphite conversion of DNA

To quantify DNAm levels at CpGs of interest, DNA from human samples or cultured cells was bisulphite-converted using the EZ DNA Methylation kit (Zymo Research) according to manufacturer protocol. Briefly, DNA (200-500ng) was incubated at 50°C for 16 hours in the presence of sodium bisulphite. This converts unmethylated cytosine nucleotides to uracil, whilst methylated cytosines remain unchanged. Bisulphite-converted DNA was then desulphonated and purified. Downstream analyses can therefore quantify the ratio of unmethylated to methylated cytosine residues.

2.2.5 Genotyping

Samples were genotyped at rs34195470 and rs1052429 using the PyroMark Q24 Advanced pyrosequencer (Qiagen). Genotyping assays were designed using the PyroMark Assay Design 2.0 software (Qiagen) and primers were ordered from Integrated DNA Technologies (IDT) as single-stranded DNA oligonucleotides. The complementary strand for the sequencing primer to bind was ordered with 5’ biotin modifications for Streptavidin Sepharose^{TM} High-Performance bead (Sigma-Aldrich) binding. Primer sequences are available in Appendix E.

For rs1052429 genotyping, 1μL DNA (50ng/μL) was amplified using the PyroMark PCR kit (Qiagen) in a 20μL reaction consisting of 10μL PyroMark Master Mix (containing HotStart Taq polymerase, dNTPs and 1.5mM $MgCl_2$), 2μL CoralLoad dye, 1μL forward primer and 1μL reverse primer (10mM each), and 5μL nuclease-free water. For rs34195470 genotyping, Q-solution (Qiagen) was added to increase amplification specificity: 1μL DNA (50ng/μL), 10μL PyroMark Master Mix, 2μL CoralLoad dye, 1μL forward primer and 1μL reverse primer (10mM each), 4μL Q-Solution and 1μL nuclease-free water. The following PCR parameters were used: 95°C, 15 mins; 45 cycles of 94°C, 30 seconds, 60°C, 30 seconds, 72°C, 30 seconds; and a final extension of 72°C, 10 minutes. PCR products were run on a 1% agarose-TBE gel and visualised using the GelDoc-It Imaging System (UVP).

The PCR product (10μL) was mixed with 70μL binding mix (containing 40μL Binding Buffer (Qiagen), 1.5μL Streptavidin Sepharose^{TM} High-Performance beads (Sigma-Aldrich) and 28.5μL nuclease-free water) in the wells of a 0.2mL 24-well PCR plate (Starlab). The mixture was agitated for 10 minutes at room temperature. The PyroMark Q24 Vacuum Workstation (Qiagen) was used to prepare DNA samples bound to streptavidin sepharose beads. Briefly, the binding solution was aspirated by vacuum with beads bound to filter probes. Then, beads were washed in 70% ethanol, 5 seconds; 1M NaOH, 5 seconds; PyroMark Wash Buffer (Qiagen), 10 seconds. The vacuum was released, and beads were transferred to Q24 plates containing 25μL sequencing mix (24.25μL annealing buffer (Qiagen) and 0.75μL 10mM sequencing primer). Plates were incubated at 80°C for 5 minutes and then transferred to the PyroMark Q24 Advanced (Qiagen) for pyrosequencing. The PyroMark Q24 Cartridge (Qiagen) was pre-loaded with reagents from the PyroMark Q24 Advanced kit (Qiagen), including reaction substrate, reaction enzyme and dNTPs. Analysis was performed using the PyroMark Q24 software (Qiagen).

2.2.6 DNAm quantification

To quantify DNAm levels at CpGs of interest, CpG assays were designed using the PyroMark Assay Design 2.0 software and primers ordered from IDT as single-stranded DNA oligonucleotides. Primers captured 16 CpGs in total, including cg26736200 (CpG8) and cg26661922 (CpG13). Primer sequences are available in Appendix E.

Bisulphite-converted DNA was amplified using the PyroMark PCR kit in duplicate. The 20μL PCR reaction consisted of 1μL bisulphite-converted DNA, 10μL PyroMark Master Mix, 2μL CoralLoad dye, 1μL forward primer and 1μL reverse primer (10mM each), and 5μL nuclease-free water. The following PCR parameters were used: 95°C, 15 mins; 45 cycles of 94°C, 30 seconds, 56°C, 30 seconds, 72°C, 30 seconds; and a final extension of 72°C, 10 minutes. PCR products were run on a 1% agarose-TBE gel and visualised using the GelDoc-It Imaging System.

Samples were prepared for analysis on the PyroMark Q24 Advanced as described in Section 2.2.5, with each sample run in duplicate. The DNAm level quantified at each CpG analysed for duplicates was subject to a 5% quality control threshold. Sample duplicates that were outside of the 5% quality control threshold were excluded from downstream analyses.

2.2.7 Complementary DNA (cDNA) synthesis

To quantify AEI and gene expression in human samples and cultured cells, total RNA was reverse transcribed using two methods to capture standard-length RNAs (>200 nucleotides) and small RNAs (<200 nucleotides in length) including miRNAs and small nuclear RNAs (snRNAs).

For standard-length RNAs, RNA (1μg) was reverse transcribed using SuperScript^{TM} IV Reverse Transcriptase (Invitrogen). RNA was incubated with DNase I (Invitrogen) for 15 minutes at room temperature. 1μL EDTA (25mM) was added, then DNase I was denatured by incubating samples for 10 minutes, 65°C. 1μL random primers (Invitrogen) and 1μL 10mM dNTPs (Invitrogen) were added, then incubated for 5 minutes, 65°C, then 1 minute, 4°C. Next, 1μL DTT (0.1M), 1μL RNaseOUT^{TM} (Invitrogen), 1μL SuperScript^{TM} IV, and 4μL SuperScript^{TM} IV buffer were added to each sample and reverse transcription was performed using the following thermocycler parameters: 23°C, 10 minutes; 55°C, 50 minutes; 80°C, 10 minutes. 1μL RNase H (NEB) (diluted 2:3) was added and incubated at 37°C, 20 minutes. cDNA samples were stored at -20°C.

To detect DNA contamination and confirm that cDNA synthesis was successful, cDNA was amplified using the PyroMark PCR kit and primers spanning exons 4 and 5 of the *HBP1* gene. Primer sequences are available in Appendix F. Briefly, 1µL cDNA was amplified in a 20µL reaction consisting of 10µL PyroMark Master Mix, 2µL CoralLoad dye, 1µL forward primer and 1µL reverse primer (10mM each), and 5µL nuclease-free water. The following PCR parameters were used: 95°C, 15 mins; 45 cycles of 94°C, 30 seconds, 60°C, 30 seconds, 72°C, 30 seconds; and a final extension of 72°C, 10 minutes. PCR products were run on a 1% agarose-TBE gel and visualised using the GelDoc-It Imaging System. In the presence of DNA contamination, a PCR product of 570bp would be amplified, incorporating the intron between exons 4 and 5 of *HBP1*. If the cDNA remained uncontaminated, a single band 152bp in length would be visible on the gel.

For small RNAs including miRNAs and snRNAs, reverse transcription was performed using MultiScribe^{TM} Reverse Transcriptase (Invitrogen) to manufacturer protocol using primers specific to miR-140-5p, miR-140-3p and the snRNA *U6* (Appendix F). Briefly, 0.15µL dNTPs (100mM), 1µL MultiScribe^{TM} (50U/µL), 1.5µL RT Buffer, 0.188µL RNase Inhibitor, 4.162µL RNase-free water and 3µL RT Primer were combined and mixed with 5µL RNA (20ng/µL). Reverse transcription was performed using the following parameters: 16°C, 30 minutes; 42°C, 30 minutes; 85°C, 5 minutes. cDNA samples were stored at -20°C.

2.2.8 AEI analysis

To determine the ratio that heterozygotes expressed each allele of the transcript SNV rs1052429, samples were first genotyped at rs34195470 and rs1052429 as described in Section 2.2.5. Only patient arthroplasty and foetal donor samples that were compound heterozygotes at rs34195470 and rs1052429 were used in the AEI analysis.

Matched DNA and cDNA samples were PCR amplified using the PyroMark PCR kit using primers targeting rs1052429 in triplicate and allelic ratios quantified by pyrosequencing as previously described (Section 2.2.5). Allelic quantification ratios for triplicates were subjected to a 5% quality control threshold. Sample triplicates that were outside of the 5% quality control threshold were excluded from downstream analyses. The mean cDNA allelic ratio of each sample was normalised to the mean DNA allelic ratio of the same sample. The results were plotted as a ratio of the risk allele to non-risk allele. The presence of AEI was determined by a ratio significantly different to 1:1.

2.2.9 Gene expression analysis

For standard mRNA, cDNA was diluted 1:20. 2.5µL diluted cDNA and 7.5µL master mix (containing 5µL TaqMan™ Advanced Master Mix for qPCR (Applied Biosystems), 1.5µL nuclease-free water and 1µL primer-probe mix (either pre-designed by IDT (PrimeTime qPCR Assays) or designed in-house, see Appendix F for details)) was combined in the wells of a MicroAmp Fast 96-Well Reaction Plate (Applied Biosystems). Individual samples were run in triplicate.

For the miRNA strands miR-140-5p and miR-140-3p and the snRNA *U6*, cDNA was diluted 1:6. 4µL diluted cDNA and 6µL master mix (containing 5µL TaqMan™ Advanced Master Mix for qPCR (Applied Biosystems), 0.5µL nuclease-free water and 0.5µL primer-probe mix (pre-designed by IDT, see Appendix F for details)) was combined in the wells of a MicroAmp Fast 96-Well Reaction Plate (Applied Biosystems). Individual samples were run in duplicate.

Gene expression was quantified using reverse transcription quantitative PCR (RT-qPCR) and performed on the QuantStudio 3 (Applied Biosystems) using the following parameters: 95°C, 20 seconds; then 40 cycles of 95°C, 1 second and 60°C, 20 seconds. Analysis was performed using the QuantStudio Design and Analysis software (Thermo Fisher Scientific). Expression of genes of interest for standard mRNAs was normalised to the housekeeping genes *18S*, *HPRT1* and *GAPDH*. Expression of miR-140-5p and miR-140-3p were normalised to the snRNA *U6*. Normalisation was performed using the $2^{-\Delta Ct}$ method. Sample replicates were subject to a quality control threshold of 1 cycle (Ct), with any outside of the threshold excluded from downstream analyses.

2.2.10 TC28a2 cell culture and passaging

The TC28a2 human immortalised chondrocyte cell line[266] (Sigma-Aldrich) was cultured in DMEM:F12 medium (Gibco), supplemented with 10% (v/v) heat-inactivated FBS (Merck) and 1% (v/v) 10,000U/mL penicillin-streptomycin (Gibco) – hereafter referred to as 'complete medium'. Cells were handled in sterile conditions under a laminar flow hood and cultured in a HERAcell 150i incubator (Thermo Scientific) at 37°C, 5% CO_2.

Once cells reached 70-80% confluence, they were dissociated from culture flasks using 0.05% trypsin-EDTA (Gibco) for 10 minutes at 37°C. Following detachment using gentle tapping, cells were re-suspended in 10x (v/v) complete medium and transferred into centrifuge tubes. Cell suspensions were centrifuged at 500 x g, 5 minutes, the supernatant aspirated, and cells were re-suspended into complete medium. Cells were then seeded at a density of 1.6×10^4 for subculturing every 4-5 days and incubated at 37°C, 5% CO_2. The range of passage numbers for all TC28a2 chondrocytes used for functional laboratory experiments was between 7 and 20.

2.2.11 Construction of Lucia reporter vectors

The region encompassing CpGs 1-16 was PCR amplified from 50ng pooled blood DNA using Phire Hot Start II DNA Polymerase (Thermo Scientific) and primers that contained additional *AvrII* and *SpeI* restriction site sequences for downstream cloning (Appendix G). Briefly, 1μL DNA was combined with 19μL mastermix (4μL 5x Buffer, 1μL forward primer (10μM), 1μL reverse primer (10μM), 0.5μL dNTPs (10mM), 0.4μL Phire Hot Start II DNA Polymerase, 12.1μL nuclease-free water) and PCR amplified using the following parameters: 98°C, 30 seconds; then 25 cycles of 98°C, 5 seconds, 65°C, 5 seconds, 72°C, 10 seconds; and a final extension of 72°C for one minute. PCR products were run on a 1% agarose-TBE gel and visualised using the GelDoc-It Imaging System.

The amplified regions were cloned into the pCR-Blunt-TOPO II vector (Invitrogen) according to manufacturer protocol and transformed into One Shot™ TOP10 chemically competent *E. coli* (Invitrogen) using the heat-shock method. 100μL bacterial transformation solution was plated onto LB-agar plates supplemented with zeocin (25μg/mL) and cultured overnight at 37°C. Individual colonies were picked and cultured in LB broth supplemented with zeocin (25μg/mL) for 16 hours, 37°C, 225 rpm. Plasmid DNA was purified using the PureYield™ Plasmid Miniprep kit (Promega) according to manufacturer protocol. Successful ligation of the region of interest into pCR-Blunt-TOPO II vectors was confirmed by Sanger sequencing (Source Bioscience).

Following confirmation of successful ligation, 1μg plasmid construct was digested with *AvrII* (NEB) and *SpeI* (NEB) to manufacturer protocol. The pCpG-free-Lucia-Basic (Invivogen) and pCpG-free-Lucia-Promoter (Invivogen) vectors were linearised using *AvrII* and *SpeI* using the same parameters, then treated with Antarctic Phosphatase (NEB) to prevent self-ligation. All digested products were run on a 0.8% agarose-TBE gel. Gel bands of interest (linearised pCpG-free-Lucia vectors and the region of interest free from its TOPO backbone) were excised and purified using the QIAquick Gel Extraction kit (Qiagen). The region of interest encompassing the 16 CpGs was then ligated into the pCpG-free-Lucia-Basic and pCpG-free-Lucia-Promoter vectors using T4 DNA Ligase (NEB) for 16 hours, 16°C.

5μL ligation reaction was transformed into GT115 chemically competent *E. coli* (Invivogen) and cultured overnight at 37°C on LB-agar plates supplemented with zeocin (25μg/mL). Individual colonies were picked and cultured in LB broth supplemented with zeocin (25μg/mL) for 16 hours at 37°C, 225rpm. Plasmid DNA was purified using the PureYield^{TM} Plasmid Miniprep kit (Promega) according to manufacturer protocol. Successful ligation of the region of interest into pCpG-free-Lucia-Basic and pCpG-free-Lucia-Promoter vectors was confirmed by Sanger sequencing (Source Bioscience).

2.2.12 Methylation and mock-methylation of Lucia reporter constructs containing the region of interest

Lucia reporter constructs containing the region of interest (4μg) were methylated using the CpG methyltransferase *M. SssI* (NEB) by incubating at 37°C for 4 hours in the presence of 1,600μM S-adenosylmethionine (SAM). An unmethylated control (mock-methylation) was performed using the same parameters in the absence of *M. SssI*. Plasmid DNA was purified by ethanol precipitation. Confirmation of methylation status was performed by digesting Lucia reporter constructs with the methylation-sensitive restriction enzyme *HaeII* (NEB). Unmethylated plasmids are therefore linearised, whilst methylated plasmids remain intact. Digested products were run on a 1% agarose-TBE gel and visualised using the GelDoc-It Imaging System.

2.2.13 Lucia reporter assays in TC28a2 cells

TC28a2 cells were seeded in 96-well plates at a density of 5,000 cells/well and incubated at 37°C, 5% CO_2. Twelve biological replicates with six technical replicates per condition were seeded. After 24 hours, cells were transfected with 100ng Lucia reporter construct and 10ng pGL3-Promoter using 2μL Lipofectamine 2000 (Invitrogen) per well according to manufacturer protocol.

24 hours post-transfection, cells were lysed in 1X Lysis Buffer (Promega) and the levels of Lucia and Firefly luciferase were measured using the GloMax Luminometer (Promega) with the Dual-Luciferase Reporter Assay System (Promega). Lucia readings were normalised to Firefly luciferase to correct for differences in transfection efficiency between wells. For analysis of regulatory activity, readings were normalised to empty vector controls.

2.2.14 Construction of dCas9-DNMT3a plasmids for epigenetic editing

Guide RNAs (gRNAs) targeting up- and downstream of the 16 CpGs were designed using the IDT gRNA Design tool (IDT). For the control, a nontargeting scrambled gRNA was designed. gRNA sequences are available in Appendix G. gRNAs were ordered as single-stranded complementary DNA oligonucleotides and annealed using T4 DNA Ligase by incubating at 37°C, 30 minutes, then 95°C, 5 minutes and gradually cooled to 25°C at a rate of -6°C/min. The dCas9-DNMT3a plasmid (Addgene) was linearised using *BbsI* (NEB) to manufacturer protocol. Annealed gRNAs were ligated into the linearised dCas9-DNMT3a plasmid using T4 DNA Ligase for 16 hours, 16°C. Samples were cloned into One Shot^{TM} TOP10 *E. coli* using the heat-shock method. 100µL bacterial transformation solution was plated onto LB-agar plates supplemented with ampicillin (100µg/mL) and cultured overnight at 37°C. Individual colonies were picked and cultured in LB broth supplemented with ampicillin (100µg/mL) for 16 hours at 37°C, 225rpm. Plasmid DNA was purified using the PureYield^{TM} Plasmid Miniprep kit (Promega) according to manufacturer protocol. Successful ligation of gRNAs into the dCas9-DNMT3a plasmid was confirmed by Sanger sequencing (Source Bioscience). The vector map for dCas9-DNMT3a is shown in Fig. 2.1.

Figure 2.1: Vector map for dCas9-DNMT3a. Vector map taken from https://addgene.com and created using SnapGene^{TM} by the depositor.

2.2.15 dCas9-DNMT3a epigenetic editing in TC28a2 cells

TC28a2 cells were cultured in monolayer to 70-80% confluence, then treated with trypsin-EDTA (0.05%, 10 minutes, 37°C) and counted. 10^6 cells per biological replicate were incubated with 5μg plasmid DNA (gRNA-dCas9-DNMT3a constructs) in 1mL nucleofection cuvettes and nucleofected using the SF Cell Line kit (Lonza) and the 4D-Nucleofector X System (Lonza) using the pulse code EH-100. Post-nucleofection, cells were incubated at room temperature for 15 minutes, then 400μL RPMI medium supplemented with 10% (v/v) FBS was added, and cells were incubated for 10 minutes at 37°C, 5% CO_2. Cells were seeded in 6-well plates containing 2mL pre-warmed complete medium and incubated for 24 hours, 37°C, 5% CO_2.

Successful nucleofection of dCas9-DNMT3a plasmids into TC28a2 cells was confirmed by enhanced green fluorescent protein (EGFP) imaging using the Axiovert M200 fluorescent microscope (Zeiss). 72 hours post-nucleofection, TC28a2 cells were dissociated from 6-well plates and harvested as described in Section 2.2.10. Pellets were snap-frozen on dry ice and stored at -80°C. Nucleic acids were extracted using the Norgen DNA/RNA kit (Norgen Bio-Tek) according to manufacturer protocol. DNA was stored at -20°C and RNA at -80°C. The RNA collected included small RNAs including miRNAs and snRNAs. Samples were quantified using the Nanodrop 1000.

To determine whether dCas9-DNMT3a epigenetic editing successfully modulated DNAm levels at the CpG region, DNA was bisulphite-converted as described in Section 2.2.4. DNAm levels were then quantified as described in Section 2.2.6. RNA was reverse transcribed into cDNA as described in Section 2.2.7. The effect of epigenetic editing on *WWP2* and miR-140 expression was measured by RT-qPCR as described in Section 2.2.9. The expression of genes of interest in the dCas9-DNMT3a epigenetically edited samples were normalised to the control samples using the $2^{-\Delta\Delta Ct}$ method[267].

2.2.16 Stabilisation of HIF-1α in TC28a2 cells

HIF-1α protein is unstable in normoxic conditions due to the degrading activity of prolyl hydroxylases (PHDs)[21]. To inhibit the action of PHDs, 200μM $CoCl_2$ was added per 1mL complete media. HIF-1α stability was determined by Western blot 24 hours post-$CoCl_2$ treatment (see Section 2.2.18).

For functional studies, TC28a2 cells were transferred into complete medium supplemented with 200μM $CoCl_2$ 24 hours before seeding of cells (for Lucia reporter gene assays) or 24 hours before nucleofection (for dCas9-DNMT3a epigenetic editing). Cells were continuously cultured in $CoCl_2$-supplemented complete medium until the end of these experiments.

2.2.17 Nuclear protein extraction

Nuclear protein was extracted from TC28a2 cells using the EPIXTRACT Nuclear Protein Isolation kit (Enzo Life Sciences) according to manufacturer protocol. Briefly, up to 10^7 cells were cultured to 80% confluence and dissociated as described in Section 2.2.10. Cells were pelleted by centrifugation (1000rpm, 5 minutes) and the supernatant was aspirated. Cells were resuspended in ice-cold 1x Pre-Extraction Buffer (supplemented with DTT and protease inhibitor cocktail) to a concentration of 10^7 cells/mL. Cells were vortexed vigorously and centrifuged at 12,000rpm, 1 minute. The cytoplasmic extract was aspirated from the nuclear pellet. To lyse nuclei, Extraction Buffer (supplemented with DTT and protease inhibitor cocktail) was added at a concentration of 10^7 cells/mL and incubated on ice for 15 minutes with frequent vortexing. Cells were centrifuged for 10 minutes, maximum speed at 4°C. The supernatant containing isolated nuclear proteins was transferred to a fresh 1.5mL microtube and stored at -80°C.

Nuclear protein extracts were quantified using the Protein Quantification Assay kit (Macherey-Nagel) according to manufacturer protocol. Absorbances were measured at 595nm using the Varioscan LUX plate reader (Thermo Fisher). Protein concentrations were determined using a calibration curve created by serial dilution of bovine serum albumin (BSA, supplied with the Protein Quantification Assay kit).

2.2.18 Western blotting

10μg nuclear protein was diluted in 30μL Protein Solving Buffer-Tris-(2-carboxyethyl) Phosphine Hydrochloride (PSB-TCEP, Macherey-Nagel) and reduced for five minutes at 95°C. Protein samples were loaded on a NuPAGE^{TM} 4-12% Bis-Tris gel (Invitrogen) alongside 4μL PageRuler^{TM} Plus Prestained Protein Ladder (Invitrogen) and resolved by running in 1X NuPAGE^{TM} MOPS SDS Running Buffer (Invitrogen) for 90 minutes at 120v. A transfer stack containing two sheets of blotting paper saturated with transfer buffer (containing 39mM glycine, 48mM Tris base, 0.0325% (w/v) SDS and 20% (v/v) methanol), one sheet of methanol-activated PVDF membrane (Thermo Scientific), the pre-cast gel containing the protein samples, and two more sheets of transfer buffer-saturated blotting paper was assembled on a V20-SDB semi-dry blotter (Scie-Plas) and run at 80mA (1mA/cm^2) for 90 minutes.

Following protein transfer from gel to PVDF membrane, the blot was blocked for one hour to prevent nonspecific protein binding using Tris-buffered saline (10mM Trizma base (Sigma-Aldrich), 150mM NaCl (Fisher Scientific)) with 0.02% (v/v) Tween-20 (Sigma-Aldrich) (TBS-T) containing 5% (v/v) milk powder (Tesco). The blot was then incubated with 1:2,000 HIF-1α primary antibody (Cell Signaling Technology) in TBS-T containing 5% (v/v) milk powder overnight at 4°C in the dark on a rocking platform (Stuart Scientific). Post-incubation, the blot was washed in TBS-T (3 x 10 minutes). The polyclonal goat anti-rabbit horseradish peroxidase (HRP)-conjugated secondary antibody (Dako) was added at 1:5,000 in TBS-T containing 5% (v/v) milk powder and incubated for one hour at room temperature. Post-incubation, the blot was washed in TBS-T (3 x 10 minutes). HIF-1α protein chemiluminescence was detected by incubation of the blot with Immobilon Western Chemiluminescent HRP Substrate (Millipore) and the Azure c600 Gel Imaging System (Thermo Fisher Scientific).

As a loading control, blots were washed in TBS-T (3 x 10 minutes), then incubated for one hour at room temperature with 1:40,000 GAPDH primary antibody (Merck) in TBS-T containing 5% (v/v) milk powder for one hour at room temperature. Post-incubation, the blot was washed in TBS-T (3 x 10 minutes). The polyclonal goat anti-mouse horseradish peroxidase (HRP)-conjugated secondary antibody (Dako) was added at 1:5,000 in TBS-T containing 5% (v/v) milk powder and incubated for one hour at room temperature. Post-incubation, the blot was washed in TBS-T (3 x 10 minutes). GAPDH protein chemiluminescence was detected by incubation of the blot with Immobilon Western Chemiluminescent HRP Substrate and the Azure c600 Gel Imaging System.

2.2.19 Electrophoretic mobility shift assays (EMSAs)

Single-stranded complementary DNA oligonucleotides 31bp in length containing the DNA sequence up- and downstream of CpG7 and CpG8/cg26736200 were ordered with 5’ DY682 fluorescent modifications in an unmethylated or methylated state from Eurofins. Oligonucleotide sequences are available in Appendix G. Complementary oligonucleotides were annealed as EMSA probes by combining 10μL forward strand (100μM), 10μL reverse strand (100μM), 10μL annealing buffer (containing 100mM Tris HCl, 500mM NaCl, 10mM EDTA; pH 8.0), and 70μL nuclease-free water. The reaction was heated to 95°C for five minutes, then cooled gradually to room temperature to anneal the probes. Probes were diluted in nuclease-free water to a working concentration of 100fmol and stored at -20°C.

A 25mL gel mastermix containing 19.1mL distilled H_2O, 2.6mL 5X TBE, 3.2mL 40% bis-acrylamide (Invitrogen), 88μL 20% (w/v) APS (Fisher Scientific) and 25μL TEMED (Thermo Scientific) was poured between two 16cm x 16cm plates with a rubber seal and a 20-well comb to cast EMSA gels. The gel was left to polymerise for one hour at room temperature. Once set, the EMSA gel was transferred to a protein electrophoresis chamber (Atto) filled with 0.5X TBE and run for 30 minutes at 100v, 4°C to remove traces of APS.

An EMSA reaction containing 2μL 10X Binding Buffer (Tube 1 of Odyssey EMSA kit), 2μL 25mM DTT/2.5% Tween 20 (Tube 2 of Odyssey EMSA kit), 0.5μL 1μg/μL Poly(dI.dC) (Tube 3 of

Odyssey EMSA kit), 1μL NP-40 (Tube 6 of Odyssey EMSA kit), 2μL DY682-labelled probe (unmethylated or methylated), 12μg nuclear protein and nuclease-free water to a final volume of 20μL was incubated for 20 minutes in the dark. For EMSA super-shift reactions, the EMSA reaction volume was increased proportionally to 25μL to accommodate the addition of 5μL 100ng/μL antibody specific to HIF-1α or 5μL 100ng/μL IgG from rabbit serum (Sigma-Aldrich) for controls. Samples were then loaded with 1X Odyssey Loading Dye and run for four hours at 100v, 4°C in the dark. Once run, the gel was imaged using the Odyssey CLx infrared imager (LI-COR) using the following parameters: resolution, 169um; quality, medium; focus offset, 4mm; channel, 700; intensity, 8.0.

2.2.20 Construction of CRISPR/Cas9 vectors

gRNAs targeting up- and downstream of rs34195470, rs9746247 and rs111837947 were designed using the CRISPR Targets track[259] on UCSC Genome Browser (hg19)[258]. gRNA sequences are available in Appendix G. gRNAs were ordered from IDT as complementary single-stranded DNA oligonucleotides with *BbsI* (NEB) restriction site overhangs and annealed with the reverse complementary strand (95°C-25°C, Δ-6°C/min) using T4 DNA Ligase. The annealed gRNAs were ligated into the *BbsI*-linearised pSpCas9n(BB)-2A-Puro (pX462) vector (Addgene) using T4 DNA Ligase for 16 hours, 16°C.

gRNA-pX462 constructs were transformed into One Shot™ TOP10 chemically competent *E. coli* to manufacturer protocol and cultured overnight at 37°C on LB-agar plates supplemented with ampicillin (100μg/mL). Individual colonies were selected and cultured in 3mL LB broth supplemented with ampicillin (100μg/mL) overnight in a shaking incubator at 37°C, 225rpm. gRNA-pX462 constructs were purified from bacterial cultures using the PureYield™ Plasmid Miniprep kit (Promega) according to manufacturer protocol. Successful ligation of the gRNA into pX462 vectors was confirmed by Sanger sequencing (Source Bioscience).

2.2.21 CRISPR/Cas9 deletion of SNVs in TC28a2 cells

TC28a2 cells were cultured in monolayer to 70-80% confluence, then treated with trypsin-EDTA (0.05%, 10 minutes, 37°C) and counted. 10^6 cells per biological replicate were incubated with 5μg plasmid DNA (2.5μg of each gRNA for paired gRNA deletions) in 1mL nucleofection cuvettes and nucleofected using the SF Cell Line kit and the 4D-Nucleofector X System using the pulse code EH-100. Post-nucleofection, cells were incubated at room temperature for 15 minutes, then 400μL RPMI medium supplemented with 10% (v/v) FBS was added, and cells were incubated for 10 minutes, 37°C, 5% CO_2. Cells were seeded in 6-well plates containing 2mL pre-warmed complete medium and incubated for 24 hours, 37°C, 5% CO_2. Successfully nucleofected cells were selected using puromycin (1μg/mL) for 48 hours and then transferred into complete medium. Cells were expanded into $25cm^2$ flasks and cultured until confluent before harvesting and snap-frozen on dry ice. Nucleic acids were extracted using the DNA/RNA kit (Norgen Bio-Tek) according to manufacturer protocol.

Successful deletion of the target region was verified by using primers specific to each SNV region (Appendix G) and the PyroMark PCR kit. 1μL DNA (50ng/μL) was amplified in a 20μL reaction consisting of 10μL PyroMark Master Mix, 2μL CoralLoad dye, 1μL forward primer and 1μL reverse primer (10mM each), and 5μL nuclease-free water. The following PCR parameters were used: 95°C, 15 mins; 45 cycles of 94°C, 30 seconds, 60°C, 30 seconds, 72°C, 30 seconds; and a final extension of 72°C, 10 minutes. PCR products were run on a 1% agarose-TBE gel and visualised using the GelDoc-It Imaging System. For Cas9-targeting of rs34195470 and rs9746247, an observed deletion of 476bp was expected. For Cas9-targeting of rs111837947, an observed deletion of 66bp was expected.

RNA was reverse transcribed into cDNA as described in Section 2.2.7. To determine whether the intronic SNV deletion had affected splicing, primers spanning exons 10-13 of *WWP2*-FL were designed (Appendix G) and were used to amplify cDNA using the PyroMark PCR kit (Qiagen). Briefly, 1μL cDNA was amplified in a 20μL reaction consisting of 10μL PyroMark Master Mix, 2μL CoralLoad dye, 1μL forward primer and 1μL reverse primer (10mM each), and 5μL nuclease-free water. PCR products were run on a 1% agarose-TBE gel and visualised using the GelDoc-It Imaging System. The expected band size if splicing was unaffected is 291bp. The effect of SNV deletion on *WWP2* expression was measured by RT-qPCR as described in Section 2.2.9. The expression of each isoform in the deletion samples was normalised to the control samples using the $2^{-\Delta\Delta Ct}$ method[267].

To determine whether SNV deletion affected DNAm levels at the CpG region, DNA was bisulphite-converted as described in Section 2.2.4. DNAm levels were then quantified as described in Section 2.2.6.

2.2.22 Construction of luciferase reporter vectors and site-directed mutagenesis

The region encompassing rs34195470, rs9746247 and rs111837947 was PCR amplified from 50ng pooled blood genomic DNA using Phire Hot Start II DNA Polymerase. Briefly, 1μL DNA was combined with 19μL mastermix (4μL 5x Buffer, 1μL forward primer (10μM), 1μL reverse primer (10μM), 0.5μL dNTPs (10mM), 0.4μL Phire Hot Start II DNA Polymerase, 12.1μL nuclease-free water) and PCR amplified using the following parameters: 98°C, 30 seconds; then 25 cycles of 98°C, 5 seconds, 65°C, 5 seconds, 72°C, 10 seconds; and a final extension of 72°C for one minute. PCR products were run on a 1% agarose-TBE gel and visualised using the GelDoc-It Imaging System. Primer details are available in Appendix G.

The amplified region was cloned into the pCR-Blunt-TOPO II vector to manufacturer protocol and transformed into One Shot^{TM} TOP10 chemically competent *E. coli* (Invitrogen) using the heat-shock method. 100μL bacterial transformation solution was plated onto LB-agar plates supplemented with zeocin (25μg/mL) and cultured overnight at 37°C. Individual colonies were picked and cultured in LB broth supplemented with zeocin (25 μg/mL) for 16 hours, 37°C, 225 rpm. Plasmid DNA was purified using the PureYield^{TM} Plasmid Miniprep kit according to manufacturer protocol. Successful ligation of the region of interest into pCR-Blunt-TOPO II vectors was confirmed by whole plasmid sequencing using Oxford Nanopore technology (Plasmidsaurus).

To generate alternative alleles, site-directed mutagenesis was used to mutate each SNV individually using primers designed using the Agilent QuikChange Primer Design tool (Appendix G) and the QuikChange Lightning Site-Directed Mutagenesis kit (Agilent) to manufacturer protocol. Briefly, 5μL 10x QuikChange Lightning Buffer, 1μL plasmid template (50ng/μL), 1μL forward primer (100ng/μL), 1μL reverse primer (100ng/uL), 1μL dNTP mix, 1.5μL QuikSolution reagent, 38μL nuclease-free water and 1μL QuikChange Lightning enzyme were combined in a PCR tube. The mutagenesis thermocycler parameters were 95°C, 2 minutes, then 18 cycles of 95°C, 20 seconds, 60°C, 10 seconds, 68°C, 3.5 minutes, and a final extension of 68°C for 5 minutes. Template plasmid DNA was digested by the addition of 2μL *DpnI* and incubation at 37°C, 5 minutes. Amplification products were transformed into XL-Gold ultracompetent cells according to manufacturer protocol. 100μL bacterial transformation solution was plated onto LB-agar plates supplemented with zeocin (25μg/mL) and cultured overnight at 37°C. Individual colonies were picked and cultured in LB broth supplemented with zeocin (25μg/mL) for 16 hours, 37°C, 225 rpm. Plasmid DNA was purified using the PureYield™ Plasmid Miniprep kit according to manufacturer protocol. Successful site-directed mutagenesis of each SNV was confirmed by whole plasmid sequencing (Plasmidsaurus).

Each SNV was PCR-amplified individually using specific primer pairs (Appendix G) with *MluI* and *XhoI* restriction sites for downstream cloning using Phire Hot Start II DNA Polymerase. Briefly, 1μL plasmid DNA was combined with 19μL mastermix (4μL 5x Buffer, 1μL forward primer (10μM), 1μL reverse primer (10μM), 0.5μL dNTPs (10mM), 0.4μL Phire Hot Start II DNA Polymerase, 12.1μL nuclease-free water) and PCR amplified using the following parameters: 98°C, 30 seconds; then 25 cycles of 98°C, 5 seconds, 65°C, 5 seconds, 72°C, 10 seconds; and a final extension of 72°C for one minute. The PCR products were then digested using *MluI* (NEB) and *XhoI* (NEB) according to manufacturer protocol. The pGL3-Promoter vector (Addgene) was linearised using *MluI* and *XhoI* to manufacturer protocol, then treated with Antarctic Phosphatase to prevent self-ligation. All digested products were run on a 0.8% agarose-TBE gel. Gel bands were excised and purified using the QIAquick Gel Extraction kit according to manufacturer protocol. Each SNV region was then ligated into the linearised pGL3-Promoter vector using T4 DNA Ligase for 16 hours, 16°C. Ligation reactions were transformed into One Shot™ TOP10 chemically competent *E. coli* to manufacturer protocol and cultured overnight at 37°C on LB-agar plates supplemented with ampicillin (100μg/mL). Individual colonies were selected and cultured in 3mL LB broth supplemented with ampicillin (100μg/mL) overnight in a shaking incubator, 37°C, 225rpm. pGL3-Promoter constructs containing the SNV region were purified from bacterial cultures using the PureYield™ Plasmid Miniprep kit to manufacturer protocol. Successful ligation of each SNV was confirmed by whole plasmid sequencing (Plasmidsaurus).

2.2.23 Luciferase reporter gene assays in TC28a2 cells

TC28a2 cells were seeded in 96-well plates at a density of 5,000 cells/well and incubated at 37°C, 5% CO_2. Six biological replicates with six technical replicates per condition were seeded. After 24 hours, cells were transfected with 100ng pGL3-Promoter construct and 10ng pRL-TK (Renilla transfection efficiency control) using 2μL Lipofectamine 2000 (Invitrogen) per well according to manufacturer protocol.

24 hours post-transfection, cells were lysed in 1X Lysis Buffer (Promega) and the levels of Firefly and Renilla luciferase were measured using the GloMax Discovery luminometer (Promega) and the Dual-Luciferase Reporter Assay System (Promega). Firefly luciferase readings were normalised to Renilla luciferase to correct for differences in transfection efficiency between wells. For analysis of regulatory activity, readings were normalised to empty vector controls.

2.2.24 Statistical analyses

All statistical analyses were performed using Prism 10 (GraphPad). For statistical analysis of DNAm levels, β-values were converted to M-values[268]. Details of statistical analyses are listed in Table 2.5.

Table 2.5. Summary of statistical analyses used in this thesis.

Analysis	Chapter	Statistical Test Applied
mQTL analysis	4	Least squares linear regression
Sex and joint comparisons	4	Mann Whitney U
Age stratification	4	Simple linear regression
AEI analysis	4	Wilcoxon’s matched-pairs signed rank
meQTL analysis	4	Simple linear regression
eQTL analysis	4	Simple linear regression
Gene expression analysis	4	One-way ANOVA with Tukey’s multiple comparisons test
Lucia reporter assays	5 and 6	Mann Whitney U with Holm-Šídák multiple test correction
dCas9-DNMT3a epigenetic editing (gene expression analysis)	5 and 6	Paired *t*-test with Benjamini-Hochberg multiple test correction
Gene expression analysis	6	Paired *t*-test
Luciferase reporter assay	7	Mann Whitney U
CRISPR/Cas9 deletion of SNVs (gene expression analysis)	7	Paired *t*-test

Chapter 3: *In-silico* analysis of the *WWP2* locus

3.1 Introduction

Determining the effects of OA risk SNV genotype upon gene expression is one useful strategy towards identifying the targets of noncoding risk loci. Where differential gene expression correlates with genotype (eQTLs), this is indicative of a functional relationship[87]. Post-GWAS analyses often integrate eQTL data to prioritise candidates for functional follow-up analysis[89], though many of the larger scale eQTL data cohorts are not available in OA-relevant tissue types[213]. Nevertheless, the identification of an eQTL in one tissue is suggestive of a functional regulatory mechanism that may apply in other tissues too.

Using genome-wide DNAm array data from OA cartilage patient samples, our research group has previously reported that genotype at rs34195470 associates with DNAm levels of two CpGs at the *WWP2* locus (cg26736200 and cg26661922), forming mQTLs[87]. The risk-conferring G allele of rs34195470 correlates with increased DNAm at both CpGs. Several studies have shown that DNAm at CpGs marking OA cartilage mQTLs can act as functional intermediaries of OA risk SNVs, modulating target gene expression[166,171–173]. However, certain factors such as tissue specificity and the life stage that OA risk SNVs and mQTLs operate can impede downstream experimental studies[87,88]. For example, an OA risk SNV and/or mQTL that is active in cartilage may not be active in synovium (tissue specificity); whilst a variant and/or mQTL active in adulthood may not mediate its pathogenic effects during development (life-stage specificity).

OA risk loci should be integrated with (epi)genomic datasets to provide evidence of putative regulatory activity that can be used to prioritise and inform functional follow-up studies[87,88]. The GTEx Portal[213] provides a resource for eQTL identification in multiple tissues, albeit not in OA-relevant tissue types, providing a starting point for prioritisation of putative gene targets. Histone ChIP-Seq data that has been categorised into specific regulatory elements by the NIH Roadmap Epigenomics Project[262] can inform whether loci reside within enhancers or repressors, whilst ATAC-Seq data can determine chromatin accessibility and therefore regions accessible for TF binding[269]. Understanding chromatin conformation allows researchers to understand physically proximal interacting regions of DNA that are genomically distal, providing evidence of enhancer-promoter loops[88,270]. Additionally, because OA risk loci and their associated mQTLs may operate at specific time points across life (e.g. during development or in later life) or in different tissue types, other tissues of the synovial joint such as subchondral bone and synovium should be investigated in addition to cartilage[87,88,165].

Using functional genomic and epigenomic datasets and publicly available online tools, this chapter aims to:

1) Identify rs34195470-eQTLs using the GTEx Portal in all available tissues.
2) Identify whether rs34195470, cg26736200 and cg26661922 reside within regions of open, accessible chromatin.
3) Identify whether rs34195470, cg26736200 and cg26661922 reside within active regulatory elements.
4) Determine which joint tissues and life stages (foetal, adult) rs34195470, cg26736200 and cg26661922 are likely to exert functional effects upon gene regulation.
5) Characterise the 3D conformation of the *WWP2* locus and identify chromatin looping in relevant cell types.

3.2 Results

3.2.1 Identification of rs34195470-eQTLs

The GTEx Portal[213] was searched for rs34195470-eQTLs reported in all available tissues, identifying seven genes, two pseudogenes and three lncRNAs, all of which were within 1Mb of rs34195470 (Appendix H). In tibial artery, the risk-conferring G allele correlates with decreased *WWP2* expression ($P = 9.8\times10^{-6}$, n=584). The expression in OA cartilage of the 12 transcripts associating with rs34195470-eQTLs in GTEx was plotted, revealing *WWP2* as the most abundantly expressed gene (Fig. 3.1).

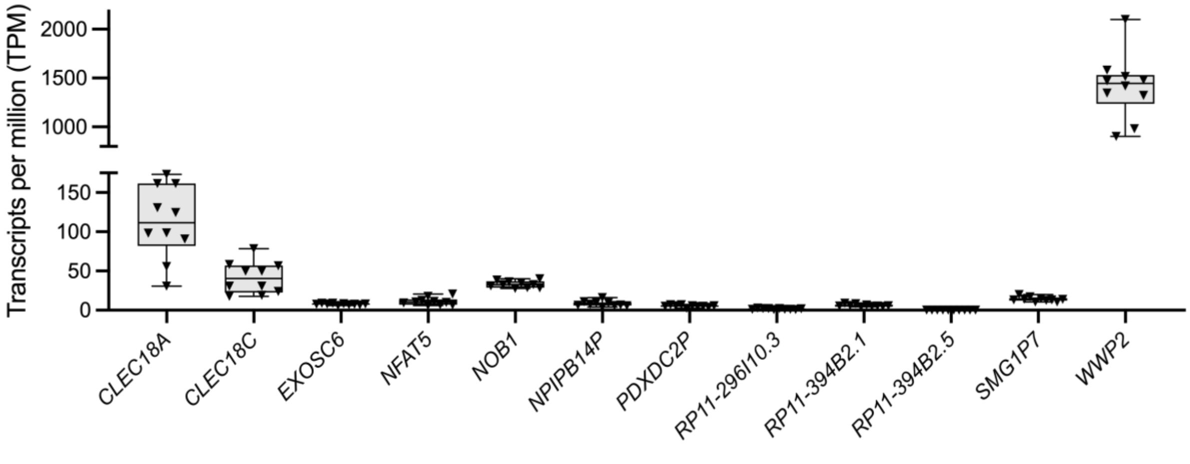


Figure 3.1: Expression of genes identified as rs34195470-eQTLs using GTEx in OA cartilage. Expression levels (TPM, transcripts per million) of the GTEx rs34195470-eQTL genes in ten OA

hip cartilage samples (available on GEO: GSE111358)[199]. Line inside the box represents the median, the box shows the interquartile range, the whiskers show the minimum and maximum values.

3.2.2 Chromatin accessibility and regulatory state at the region of interest

rs34195470 resides at the distal end of the locus within intron 11 of *WWP2*-FL and approximately 3.2Kb upstream of the *WWP2*-C TSS (Fig. 3.2A, in black). cg26736200 and cg26661922 are approximately 3.7Kb upstream of rs34195470 and span an exon-intron boundary: cg26736200 is within exon 11 of *WWP2*-FL and cg26661922 is within intron 11 of *WWP2*-FL (Fig. 3.2A, in red).

To determine chromatin accessibility at the region encompassing rs34195470 and the two CpGs, ATAC-Seq data was mapped to the region using the UCSC Genome Browser (hg19). Both cg26736200 and cg26661922 were identified to reside within ATAC-Seq peaks in foetal and OA (adult) chondrocytes of the hip and the knee as well as OA fibroblast-like synoviocytes (FLS, Fig. 3.2B). Additionally, cg26661922 was also found within an ATAC-Seq peak from the osteosarcoma cell line Saos-2. rs34195470 was not found within any ATAC-Seq peaks, indicating that it resides within closed, inaccessible chromatin.

To determine whether the region encompassing rs34195470 and the array CpGs were known to reside within putative regulatory elements, Roadmap regulatory elements were mapped to the region using the UCSC Genome Browser (hg19). rs34195470 is within a region marked as weak transcription in MSCs and as an enhancer in chondrocytes and osteoblasts (Fig. 3.2C). In MSCs, cg26736200 resides within an enhancer whilst cg26661922 is within a region of weak transcription. Both CpGs are marked as flanking an active TSS in chondrocytes and osteoblasts.

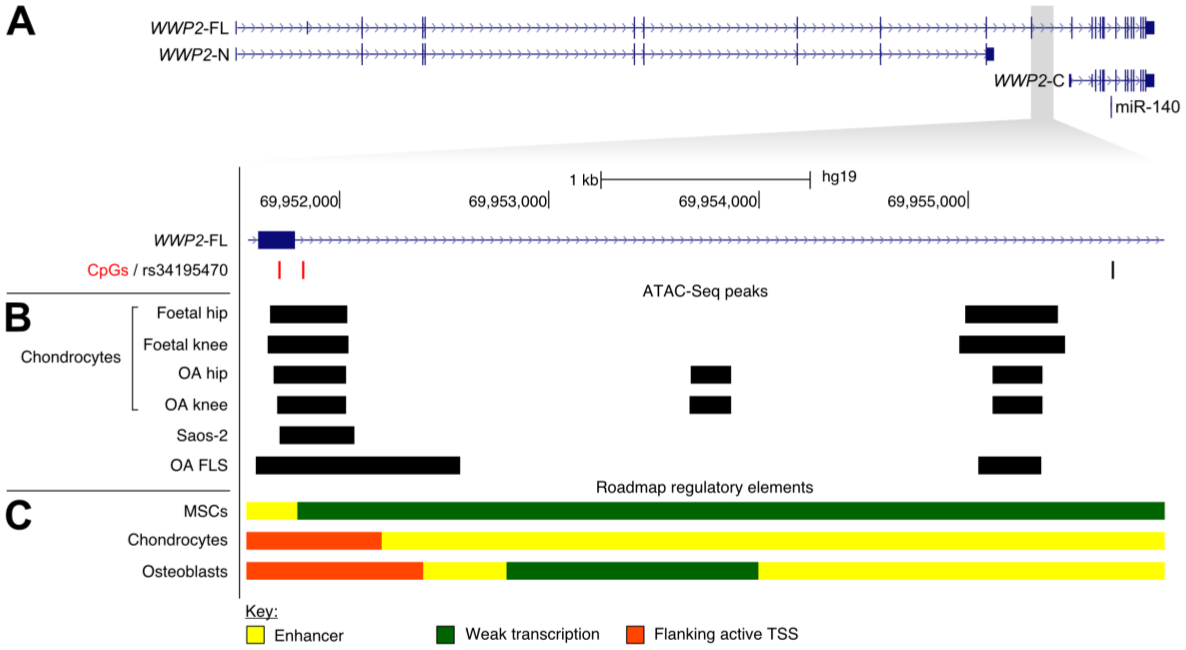


Figure 3.2: Chromatin accessibility and regulatory elements in relevant cell types at the *WWP2* locus. A) Schematic representation to show the gene structure of the *WWP2* isoforms *WWP2*-FL, *WWP2*-N and *WWP2*-C. Horizontal lines represent introns, full height vertical bars represent exons, and half-height vertical bars represent 5’ and 3’ UTRs. Arrows pointing right indicate direction of transcription (sense strand). miR-140 is also shown. The grey box region is expanded to show the proximity of rs34195470 (in black, right) and CpGs cg26736200 and cg26661922 (in red, left) to the *WWP2*-FL isoform. Coordinates from UCSC hg19. **B)** ATAC-Seq peaks generated from foetal hip and knee chondrocytes, OA hip and knee chondrocytes, Saos-2 cells, and OA FLS. Open chromatin regions are marked by black boxes. **C)** Chromatin regulatory state data from ROADMAP for MSCs (E006 H1-derived MSCs), chondrocytes (E049 MSC-derived chondrocyte cultured cells), and osteoblasts (E129 osteoblast primary cells). Colours corresponding to different regulatory elements are shown in the key at the bottom of the figure.

3.2.3 Chromatin looping at the region of interest

The 3D Genome Browser was searched for Capture Hi-C data in MSCs to identify long-range chromatin interactions at the region of interest[263]. A single chromatin loop was identified between the coordinates encompassing the *WWP2*-FL/*WWP2*-N shared TSS and the region harbouring rs34195470 and CpGs cg26736200 and cg26661922 (Fig. 3.3). The CpGs and rs34195470 are therefore brought to within close physical proximity to the *WWP2*-FL/*WWP2*-N TSS and may therefore be capable of regulating the expression of these transcript isoforms.

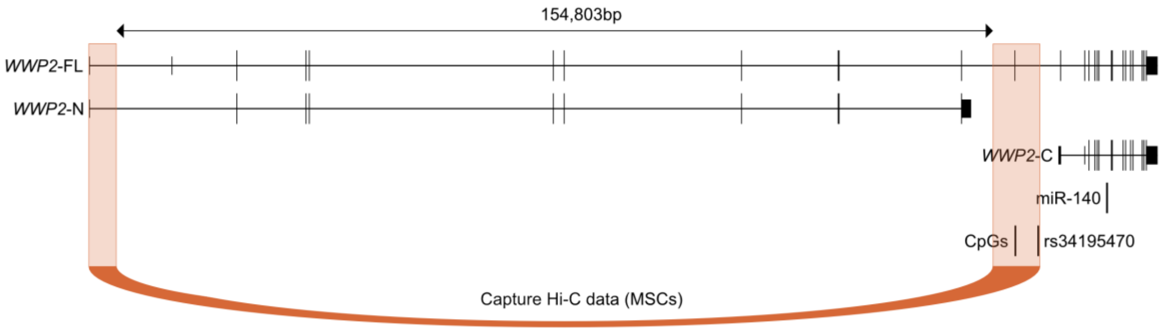


Figure 3.3: Chromatin looping at the *WWP2* locus. Schematic representation to show the gene structure of the *WWP2* isoforms *WWP2*-FL, *WWP2*-N and *WWP2*-C. Horizontal lines represent introns, full height vertical bars represent exons, and half-height vertical bars represent 5' and 3' UTRs. Arrows pointing right indicate direction of transcription (sense strand). The location of CpGs cg26736200 and cg26661922, rs34195470 and miR-140 are also shown. Capture Hi-C chromatin interactions identified at the CpG region in MSCs using the 3D Genome Browser, represented as an orange loop with anchors spanning the width of the interacting regions extending across the schematic of the *WWP2* locus. The genomic distance between the anchors of the chromatin loop (154,803bp) is represented by an arrow at the top of the figure. The interacting region coordinates in UCSC hg19 are Chr16:69,793,806-69,801,037 (*WWP2*-FL/*WWP2*-N TSS) and Chr16:69,948,024-69,955,840 (CpGs and rs34195470).

3.3 Discussion

Deciphering regulatory mechanisms and effector genes targeted by GWAS-reported OA risk SNVs can be aided through integration and fine-mapping of functional genomic and epigenomic datasets at regions of interest[87,89]. Several studies have demonstrated that understanding the local chromatin context of regions housing OA risk SNVs and their associated mQTLs can be a valuable tool informing downstream experiments to establish effector genes and putative molecular mechanisms[166,171–173,212].

Open, accessible chromatin facilitates the binding of TFs that can modulate target gene expression, whilst closed, inaccessible chromatin diminishes these interactions[269]. In the cell types investigated, rs34195470 was identified to reside within closed chromatin, suggesting TF binding would be impeded by steric hindrance and therefore unlikely to modulate target gene expression[130]. However, rs34195470 was also identified within a Roadmap enhancer in chondrocytes and osteoblasts, suggesting a potential role for transcriptional regulation in these cell types.

Using GTEx in all tissues, 12 genes were reported as rs34195470-eQTLs including *WWP2*, which exhibited tenfold greater expression than the next most abundant gene in OA cartilage. Using an alternative approach, *WWP2* AEI in OA cartilage has been reported using the transcript SNV rs1052429, where the A allele of rs1052429 correlates with increased *WWP2* expression[214]. The A allele of rs1052429 nearly always occurs on a haplotype with the OA risk-conferring G allele of rs34195470. Therefore, the OA risk allele of rs34195470 correlates with increased *WWP2* expression in OA cartilage and operates inversely to the observation made using tibial artery GTEx data. These findings suggest rs34195470 is capable of regulating *WWP2* in a genotype-dependent manner.

cg26736200 and cg26661922 reside within open chromatin in almost all the tissue types investigated, except for the former in the osteosarcoma Saos-2 cell line. In chondrocytes, this effect was not restricted to a particular joint site (hip, knee) or life stage (foetal, adult). It has been suggested that OA chondrocytes revert to a developmental phenotype, exhibiting increased expression of TFs associated with chondrocyte hypertrophy such as RUNX2 and HIF-2α[13]. However, it is equally plausible that this region may be accessible across the entirety of the life course. At present, there are no available datasets bridging these timepoints in cartilage to support either scenario. In chondrocytes and osteoblasts, both CpGs reside within a region marked as flanking an active TSS, indicating they are physically proximal to an active gene promoter. This is supported by chromatin looping observed in MSCs between the CpGs and the *WWP2*-FL/*WWP2*-N TSS. Together, these data imply cg26736200 and cg26661922 are within an active regulatory element and are accessible to TF binding in all cell types investigated in this chapter.

Enhancers typically reside within noncoding regions of the genome and are often distal to the target gene promoter they modulate. The most widely accepted theory on the mechanism driving enhancer activity is the formation of chromatin loops that bring enhancers and promoters into close physical proximity, facilitating transcriptional activity by RNA Pol II[88,270]. The region of DNA harbouring rs34195470 and CpGs cg26736200 and cg26661922 loops to interact with an upstream region containing the promoter of *WWP2*-FL and *WWP2*-N in MSCs, suggesting these transcript isoforms may be targets of OA risk marked by rs34195470, the CpGs, or both. However, the *WWP2*-C and miR-140 shared promoter also resides within intron 11 of *WWP2*-FL[244,251], indicating they are also valid candidate transcripts for further analysis.

The main limitation of this chapter is the lack of available OA-relevant tissue datasets. Regulatory elements in synovium were not characterised due to no available datasets on

Roadmap. There are also no available ATAC-Seq datasets for subchondral bone, therefore the Saos-2 cell line dataset was used. This cell line may not reflect the true state of chromatin accessibility at the locus in subchondral bone yet remained the most biologically relevant dataset available for this study.

In summary, all the investigated loci show indicators of regulatory capacity using ATAC-Seq and Roadmap data. Chromatin accessibility and regulatory state is not limited to a particular time point in chondrocytes, nor is it limited to one particular adult tissue type. These findings warrant expanding the OA cartilage mQTL study into other tissues to determine whether the mQTLs operate during development and whether the mQTL effect is cartilage-specific.

Chapter 4: mQTL analysis and *WWP2* expression analysis in synovial joint tissues

4.1 Introduction

cg26736200 and cg26661922 are the CpG sites whose methylation levels in OA cartilage DNA associate with genotype of the knee OA SNV rs34195470 (A > G), forming mQTLs[87]. The OA risk-conferring G allele of the SNV correlates with increased DNAm at the CpGs. *WWP2* AEI has also been reported using rs1052429 (G > A), a *WWP2* transcript SNV in moderate LD with rs34195470 in European populations (r^2 = 0.41, D' = 0.91), where the A allele of rs1052429 correlates with increased *WWP2* expression[214]. The A allele of rs1052429 nearly always occurs on a haplotype with the OA-risk conferring G allele of rs34195470. Together, these reports imply that the OA risk allele of rs34195470 associates with increased DNAm at the CpGs and with increased *WWP2* expression in cartilage. In the previous chapter, cg26736200 and cg26661922 were identified within a region of open chromatin marked as an active transcriptional regulator, suggesting they may be an intermediary between rs34195470 and its target gene. DNAm as a functional intermediate between a risk-conferring SNV and expression of the target gene is relatively common[271,272], with experimental perturbation of DNAm at OA cartilage mQTLs revealing several effector genes of OA genetic risk[166,171–173].

mQTLs that were first identified in OA cartilage have also been reported to be present during skeletogenesis[165]. In development, OA risk loci have been shown to demonstrably alter joint morphology[127], whilst OA chondrocytes exhibit the developmental phenotype of hypertrophy[13]. OA mQTLs may contribute to disease susceptibility during development, with their effects manifesting in later life[87]. Therefore, the utilisation of foetal donor samples as well as OA patient samples can aid in pinpointing when mQTLs exert their functional effects. mQTLs have also been reported in other articular joint tissues using epigenome-wide arrays[158,169]. These findings emphasise the need to investigate tissues other than cartilage, which has been the sole focus of many OA studies to date. Whilst mQTLs commonly present in the same direction across joint tissues[158,166], inverse relationships have also been identified with pleiotropic effects upon the expression of OA effector genes such as *COLGALT2*[167,171].

eQTLs can also be used to identify effector genes in complex disease. However, eQTL analyses require large sample cohorts to overcome interindividual variability in gene expression[87]. There is also a dearth of OA-relevant tissue eQTL datasets available. AEI analysis, which determines the relative ratio of allelic expression in heterozygotes using a much smaller sample size, has been proven to be a useful tool in elucidating effectors of OA risk SNVs[172,173]. Correlations between DNAm levels at CpGs and expression of genes of interest may also be used to detect tissues in which mQTLs may be functional and regulate gene expression, forming methylation-expression QTLs (meQTLs)[166,173].

Using a range of molecular genetic and epigenetic analyses in foetal donor and OA patient samples, this chapter aims to:

1) Replicate the previously reported *WWP2* mQTLs and AEI using an independent cohort of OA cartilage samples.
2) Determine whether the mQTLs and AEI are active during development using foetal cartilage samples.
3) Determine whether the mQTLs and AEI are cartilage-specific by expanding the study to include OA synovium and subchondral bone samples.
4) Perform an meQTL analysis using DNAm and AEI data in the four tissue types studied.
5) Quantify *WWP2* expression in these tissues.
6) Identify eQTLs associating with rs34195470.

4.2 Results

4.2.1 mQTL analysis in foetal cartilage and OA cartilage

Using genome-wide DNAm array data created from the cartilage of patients who had undergone total joint arthroplasty surgery, our research group has previously reported that rs34195470 genotype correlates with DNAm levels at cg26736200 and cg26661922, forming two mQTLs[87]. The risk-conferring allele (G) of rs34195470 correlates with increased DNAm at these CpGs. Here, I repeated the mQTL analysis of cg26736200 and cg26661922 in an independent cohort of OA cartilage DNA samples (Appendix A), simultaneously expanding the study to encompass nearby CpGs. A total of 16 CpGs were captured by pyrosequencing: cg26736200 (CpG8 in Fig. 4.1A) and cg26661922 (CpG13); the four CpGs between these two (CpG9-CpG12); the three CpGs immediately downstream of cg26661922 (CpG14-CpG16) and seven of the eight CpGs immediately upstream of cg26736200 (CpG1-CpG7; data could not be collected for a single CpG located between CpG1 and CpG2).

Fourteen of the 16 CpGs showed a significant association (P = 0.0010-0.0476) between their DNAm levels and rs34195470 genotype, with the risk-conferring G allele of rs34195470 correlating with increased DNAm (Fig. 4.1B). This replicated the original discovery[87] and highlighted the presence of a differentially methylated region (DMR) in OA cartilage spanning 228bp. The OA cartilage patient samples range in age from 41-93 years (Appendix I). The

DNAm levels for CpG11 and CpG13 showed a modest but significant reduction with age (CpG11, $r^2 = 0.0967$, $P = 0.0376$; CpG13, $r^2 = 0.0909$, $P = 0.0495$; Appendix J). Stratification by sex or joint site revealed that the mQTL was not restricted to a particular stratum (Appendix K and L). Stratification of the DNAm data by joint site or sex irrespective of rs34195470 genotype (Appendix L and M) revealed only one significant difference in DNAm levels, with CpG12 being slightly more methylated in hip compared to knee cartilage ($P = 0.0416$).

It has been reported that some OA cartilage mQTLs are also present in foetal cartilage, implying that a proportion of OA genetic risk may be functionally active during skeletogenesis[165,216]. The rs34195470 mQTL was therefore investigated in foetal cartilage DNA. Five of the 16 CpGs showed a significant association ($P = 0.0096$-0.0266) between DNAm and rs34195470 genotype, and as seen for the OA cartilage samples, the OA risk-conferring G allele correlated with increased DNAm (Fig. 4.1C). The cartilage DMR is therefore active during development and in OA. Foetal cartilage samples ranged in age from 8-17 pcw (Appendix I). No significant correlations between age and DNAm levels were observed ($P = 0.1479$-0.8827; Appendix J). None of the mQTLs were identified to be sex-specific (Appendix K, $P = 0.0671$-0.9724). There were also no differences in DNAm levels between females and males when foetal DNAm data was stratified irrespective of rs34195470 genotype (Appendix M, $P = 0.2799$-0.9699).

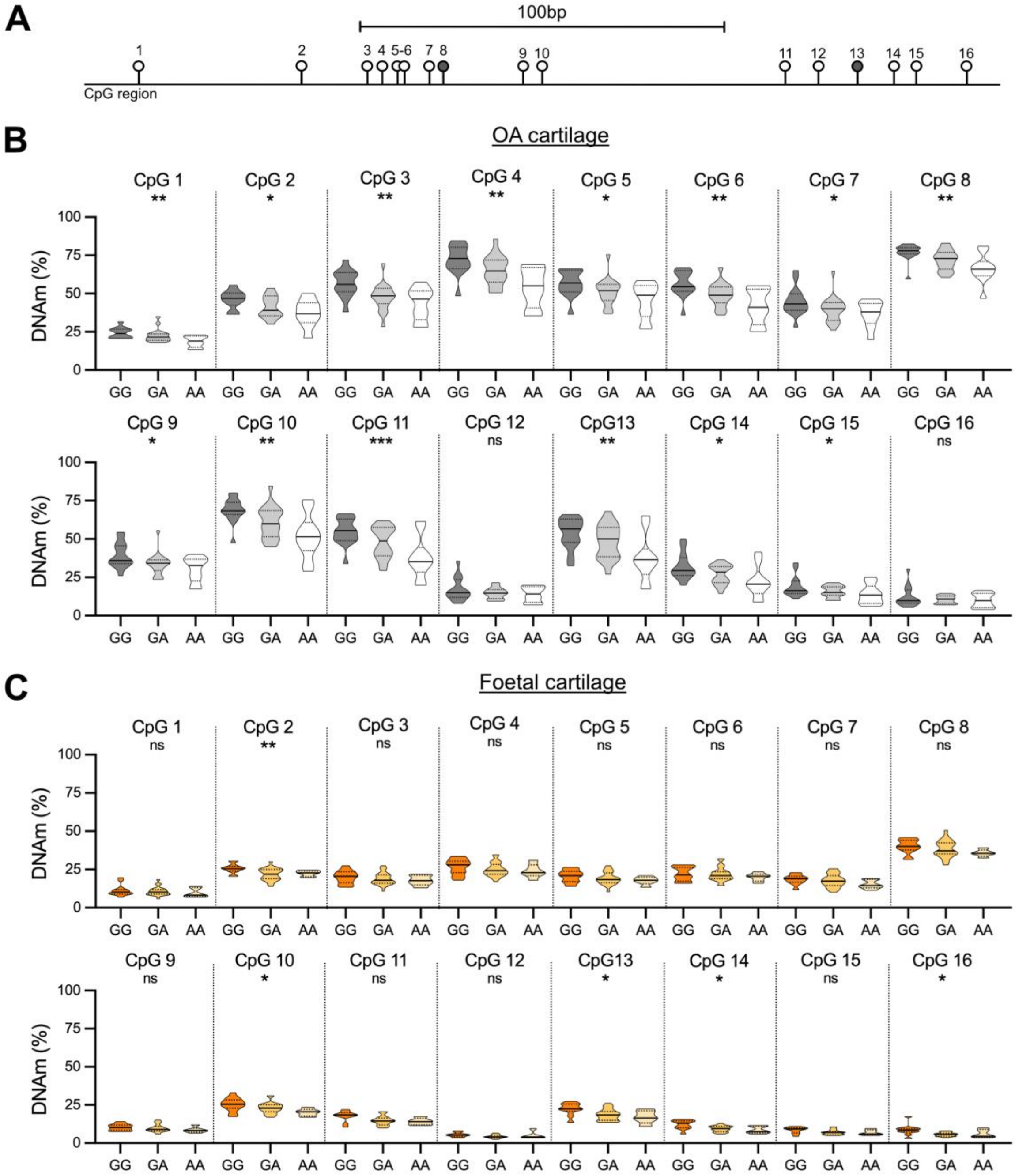


Figure 4.1: mQTL analysis in OA cartilage and foetal cartilage. A) Schematic representation of the 16 CpGs analysed in this study across a 228bp region. cg26736200 (CpG8) and cg26661922 (CpG13) are highlighted in grey. **B)** OA cartilage mQTL analysis (in grey, n=51). **C)** Foetal cartilage mQTL analysis (in orange, n=48). For panels B and C, DNAm levels at the 16 CpGs were stratified by rs34195470 genotype (GG, GA, AA). Methylation data is plotted in the form of β-values ranging from 0 (no methylation) to 1 (complete methylation) and expressed as a percentage. In the truncated violin plots, solid and dashed horizontal lines represent the median and interquartile range. *P*-values calculated by least squares linear regression. * = $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.001$; ns = not significant ($P > 0.05$).

mQTLs first reported in cartilage have also been found to be present in other synovial joint tissues[166,167,171]. The mQTL analysis was therefore expanded to include two other OA-relevant tissues: subchondral bone and synovium provided by OA patients undergoing arthroplasty. Eight of the 16 CpGs showed a significant correlation (P = 0.0178-0.0460) between DNAm levels and rs34195470 genotype in OA bone (Fig. 4.2B). The risk-conferring G allele of rs34195470 associated with the highest levels of DNAm at each CpG. However, heterozygotes (GA) had demonstrably lower DNAm levels at each CpG site than homozygotes of the non-risk allele (AA). The OA bone samples range in age from 51-86 years (Appendix I). One correlation between age and DNAm levels was identified at CpG2 (Appendix N, P = 0.0192, r^2 = 0.1160), with DNAm levels decreasing slightly with increasing age. Interestingly, stratification by sex revealed the mQTL effect observed was driven by males and absent in females (Appendix O), with all 16 CpGs significantly correlating with rs34195470 genotype in males ($P < 0.0001$-0.0183; the risk-conferring G allele of rs34195470 correlates with increased DNAm) whilst none of the CpGs were mQTLs in females (P = 0.0795-0.9641). DNAm levels at each CpG irrespective of genotype were comparable between sex (Appendix M, P = 0.0935-0.9352).

In OA synovium, none of the 16 CpGs were mQTLs (P = 0.1217-0.8587, Fig. 4.2C). The OA synovium samples range in age from 45-91 years (Appendix I). No correlations between age and DNAm levels were identified (Appendix N, P = 0.2502-0.9411). Stratification by sex did not reveal an mQTL effect specific to males or females (Appendix O, P = 0.2375-0.9856). Stratification of the DNAm data by sex irrespective of rs34195470 genotype (Appendix M) revealed slightly higher DNAm levels in females compared to males at three CpGs: CpG12 (P = 0.0120), CpG15 (P = 0.0431), and CpG16 (P = 0.0458). mQTL data was not stratified by joint site due to the low number of hip samples in the cohort (n = 4).

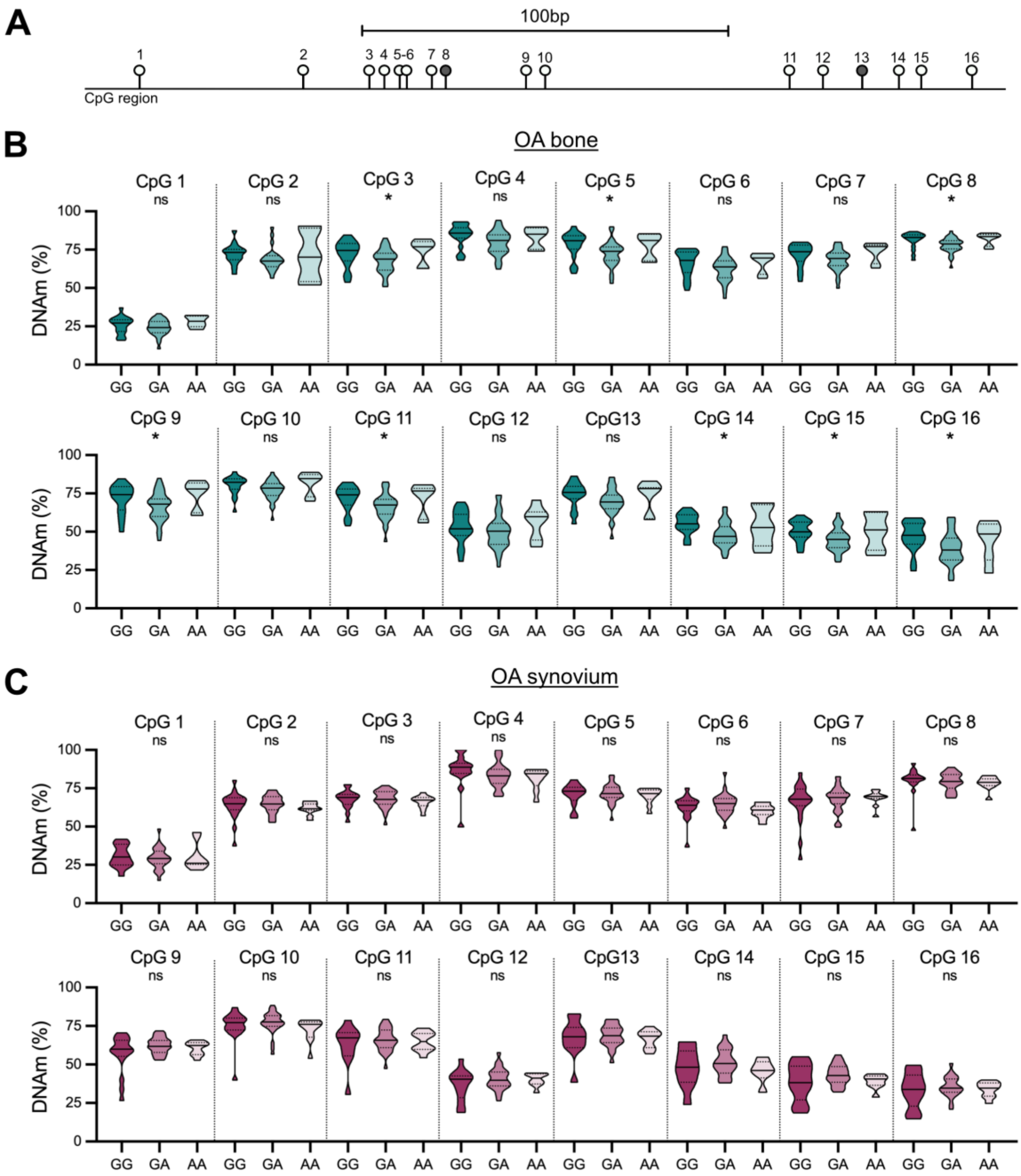


Figure 4.2: mQTL analysis in OA bone and OA synovium. A) Schematic representation of the 16 CpGs analysed in this study across a 228bp region. cg26736200 (CpG8) and cg26661922 (CpG13) are highlighted in grey. **B)** OA bone mQTL analysis (in teal, n=61). **C)** OA synovium mQTL analysis (in maroon, n=70). For panels B and C, DNAm levels at the 16 CpGs were stratified by rs34195470 genotype (GG, GA, AA). Methylation data is plotted in the form of β-values ranging from 0 (no methylation) to 1 (complete methylation) and expressed as a percentage. In the truncated violin plots, solid and dashed horizontal lines represent the median and interquartile range. *P*-values calculated by least squares linear regression. * = $P < 0.05$; ns = not significant ($P > 0.05$).

Mean DNAm levels at the 16 CpGs were compared between the four tissues investigated in this mQTL analysis (Fig. 4.3). Foetal cartilage was comparatively hypomethylated compared to the three OA tissues (mean DNAm range = 4.7-38.7%), whilst OA bone and OA synovium had comparably similar DNAm levels across the DMR and exhibited the highest levels of DNAm (mean DNAm range: OA bone = 25.2-81.6%; OA synovium = 29.8-83.6%). OA cartilage mean DNAm levels ranged from 11.0-72.5%.

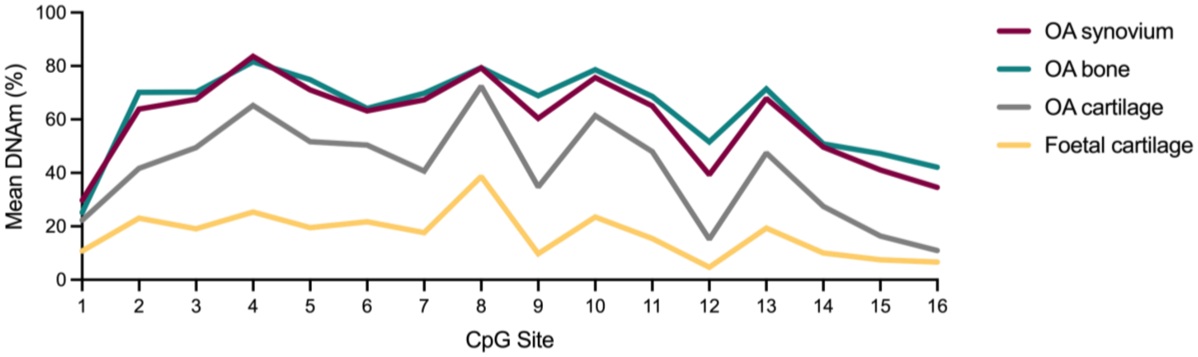
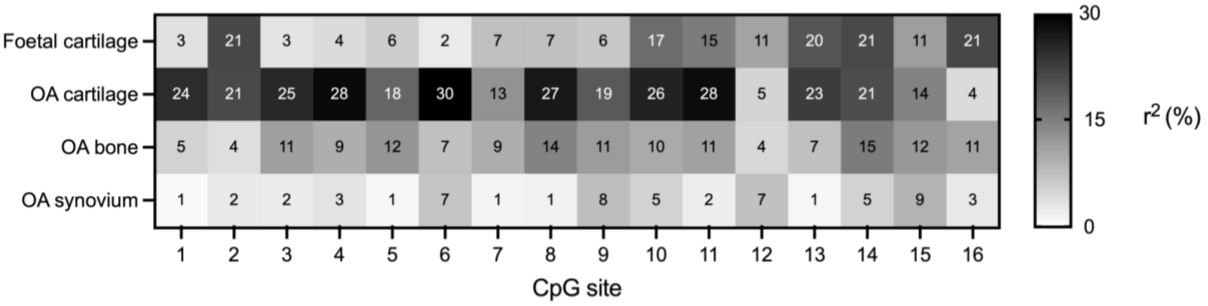


Figure 4.3: Mean DNAm levels across the DMR. Methylation data is plotted in the form of β-values ranging from 0 (no methylation) to 1 (complete methylation) and expressed as a percentage.

Linear regression was used to calculate the percentage contribution of rs34195470 genotype to differences in DNAm in all four tissues (Fig. 4.4). For 12 of the 16 CpGs investigated, the largest genotypic effect was observed in OA cartilage. Two of the four remaining CpGs had the same genotypic effect in OA cartilage and foetal cartilage (CpG2 and CpG14). At CpG12 and CpG16, the highest genotypic effect was in foetal cartilage. OA synovium showed the least genotypic contribution to DNAm levels at 14 of the 16 CpGs.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Foetal cartilage	3	21	3	4	6	2	7	7	6	17	15	11	20	21	11	21
OA cartilage	24	21	25	28	18	30	13	27	19	26	28	5	23	21	14	4
OA bone	5	4	11	9	12	7	9	14	11	10	11	4	7	15	12	11
OA synovium	1	2	2	3	1	7	1	1	8	5	2	7	1	5	9	3

Figure 4.4: Genotypic effect upon DNAm levels across the DMR. Heatmap showing the influence of the rs34195470 genotype on DNAm levels at the 16 CpGs in the four tissues

investigated. r^2 values determined by least squares linear regression and expressed as a percentage.

4.2.2 AEI analysis

Traditional eQTL analyses typically require hundreds of samples to detect correlations between gene expression and SNV genotype due to interindividual variability[87]. Therefore, AEI analysis was used to detect differences in the relative ratio of mRNA transcript produced by patients heterozygous at the OA risk SNV. As rs34195470 resides within an intron, the proxy SNV rs1052429 was used (r^2 = 0.41, D' = 0.91 in European populations) which resides within the 3' UTR of *WWP2*-FL and *WWP2*-C.

Before performing AEI analyses, assay validation was performed using a serial dilution of DNA with known ratios of A allele to G allele at rs1052429 (0-100%, in 10% increments). The expected and observed allelic ratios of A:G were highly congruent (r^2 = 0.98, $P < 0.0001$), validating the designed assay (Fig. 4.5).

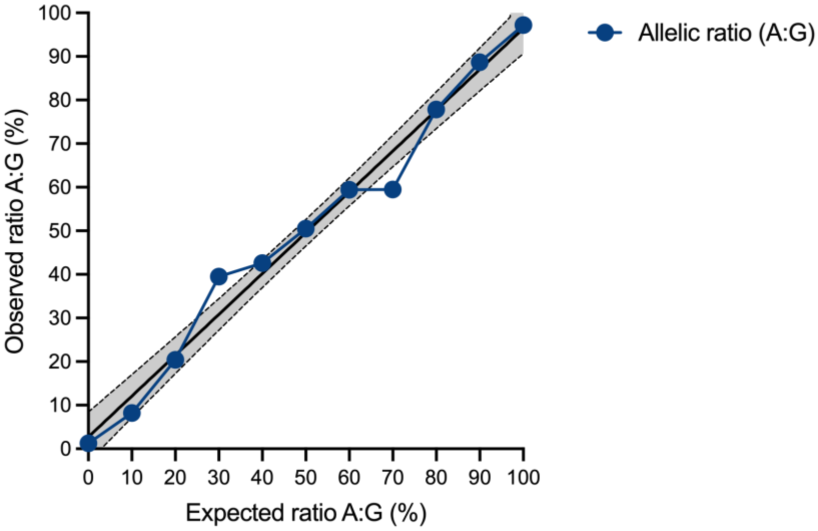


Figure 4.5: AEI assay validation. The designed AEI assay was validated using a serial dilution of known ratios of A allele and G allele at rs1052429. The measured ratios are shown, with the ratio of G allele plotted (blue dots and joining line). The expected ratio is plotted on the x-axis, and the observed ratio is plotted on the y-axis. *P*-value and r^2 value calculated using simple linear regression. The line of best fit (solid black line) and 95% confidence interval bands (dashed black lines) are shown, spanning the grey shaded area.

Significant AEI ($P = 0.0195$) was observed in nine OA cartilage samples heterozygous at rs1052429, with a mean A/G ratio of 1.15 (Fig. 4.6A). This is equivalent to the A/G ratio of 1.27

previously reported[214] and confirms that the OA risk-conferring G allele of rs34195470 associates with increased expression of *WWP2* in OA cartilage. No significant AEI was detected in foetal cartilage (*P* = 0.8438), OA bone (*P* = 0.0781) or OA synovium (*P* = 0.5333).

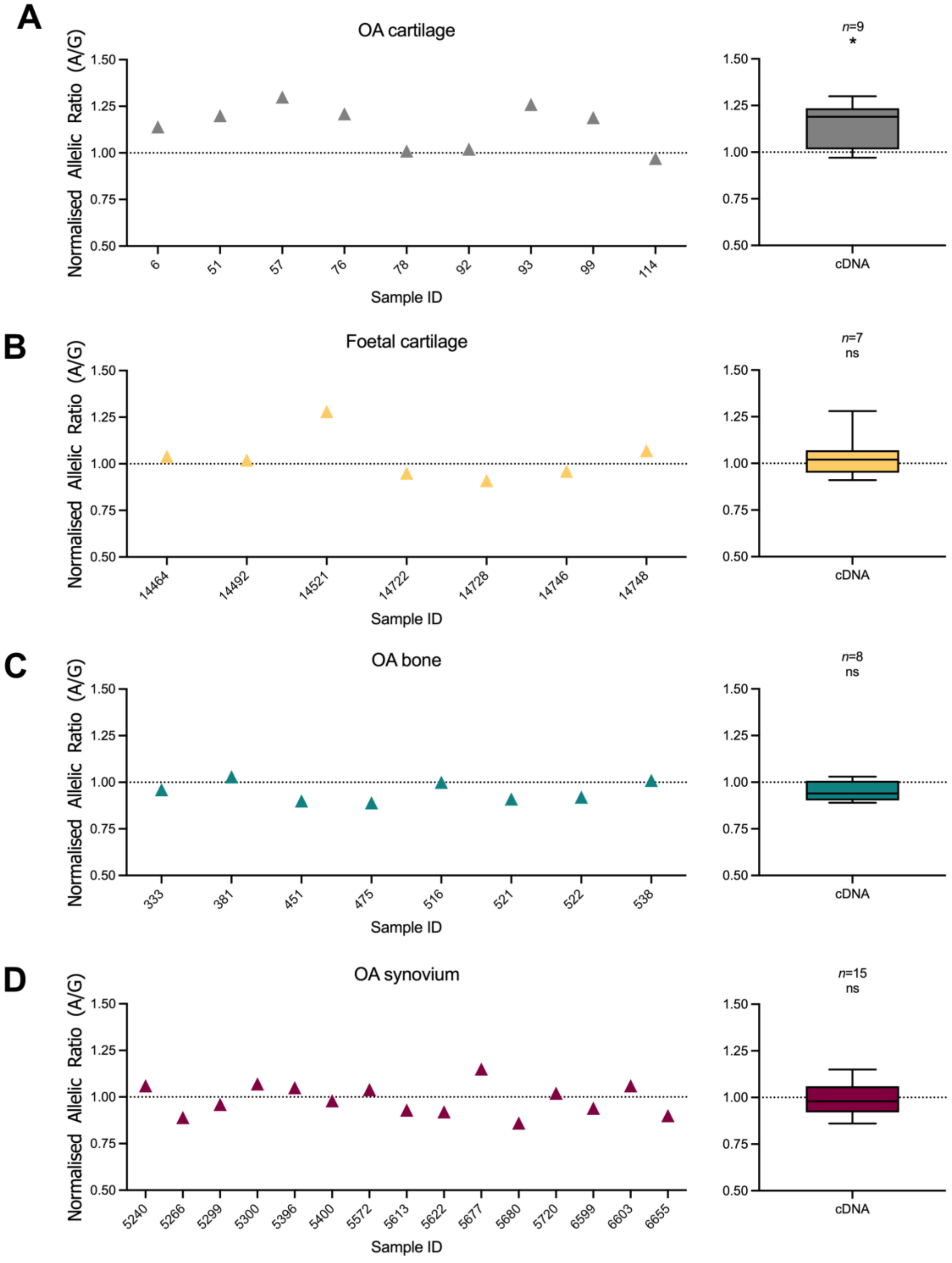


Figure 4.6: AEI analysis in synovial joint tissues. A) AEI analysis in OA cartilage samples (in grey, n=9). **B)** AEI analysis in foetal cartilage samples (in yellow, n=7). **C)** AEI analysis in OA bone samples (in teal, n=8). **D)** AEI analysis in OA synovium samples (in maroon, n=15). All left panels, allelic (A/G) ratios are plotted as triangles, representing the mean of three replicates normalised to DNA values (represented by the dotted line at y=1). All right panels, mean cDNA values for all samples represented as a box plot normalised to their corresponding DNA values, as above. Line inside the box represents the median, the box shows the interquartile range, the whiskers show the minimum and maximum values. *P*-values calculated by Wilcoxon matched-pairs signed rank test. * = $P < 0.05$; ns = not significant ($P > 0.05$).

4.2.3 Multi-tissue meQTL analysis

The OA risk-conferring allele (G) of rs34195470 correlates with increased DNAm and increased *WWP2* expression in OA cartilage, as previously reported[87,214] and replicated in Section 4.2.1 and Section 4.2.2. To identify correlations in OA patient and foetal donor samples between DNAm and *WWP2* expression, DNAm data for the 16 CpGs was plotted against AEI data for matched samples (Fig. 4.7 and 4.8). In OA cartilage, no meQTLs were detected (Fig. 4.7A, $P = 0.0511$-0.9634). In foetal cartilage, a single meQTL was identified at CpG1 (Fig. 4.7B, $P = 0.0397$) in which increased DNAm levels correlated with decreased *WWP2* expression. This correlation was observed despite CpG1 not being an mQTL and no observable AEI in this tissue. The remaining 15 CpGs were not meQTLs in foetal cartilage ($P = 0.0849$-0.9277). No meQTLs were identified in OA bone (Fig. 4.8A, $P = 0.3309$-0.9852) or in OA synovium (Fig. 4.8B, $P = 0.4759$-0.9114).

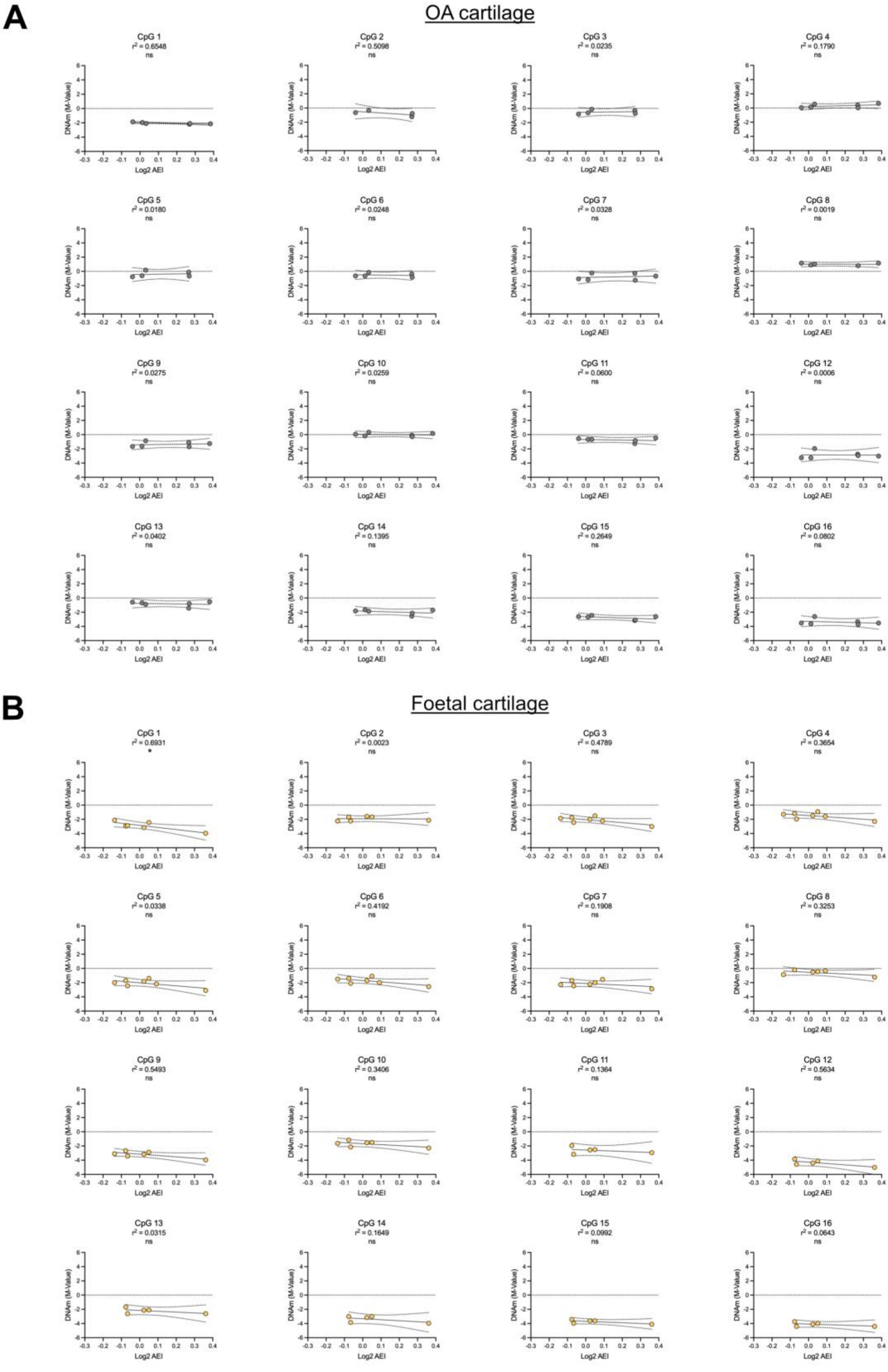


Figure 4.7: meQTL analysis in OA cartilage and foetal cartilage. A) meQTL analysis in OA cartilage (in grey, n=4-6). **B)** meQTL analysis in foetal cartilage (in yellow, n=5-7). For both panels, rs1052429 allelic ratios were plotted against matched DNAm data. Dotted line at *y*=0

represents 50% DNAm. DNAm levels were plotted as M-Values and AEI data was Log2 transformed. Coloured circles represent individual samples. Differences in numbers (n) due to variable number of patient samples per CpG with matched AEI and DNAm data. *P*-values and r^2 values calculated using simple linear regression. * = $P < 0.05$; ns = not significant, ($P > 0.05$). The solid line represents the regression line with 95% confidence intervals represented by dashed lines.

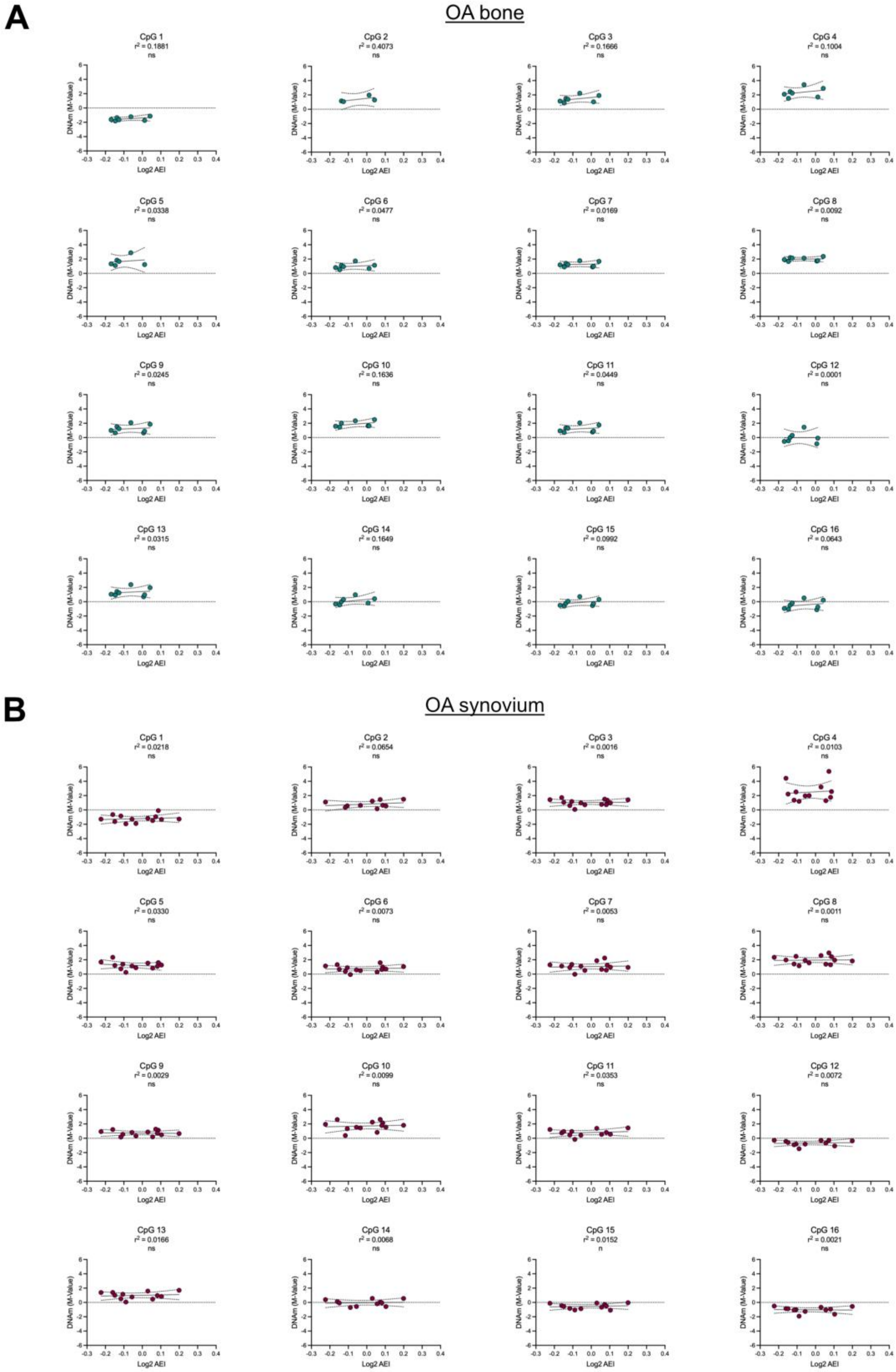


Figure 4.8: meQTL analysis in OA bone and OA synovium. A) meQTL analysis in OA bone (in teal, n=4-8). **B)** meQTL analysis in OA synovium (in maroon, n=10-14). For both panels, rs1052429 allelic ratios were plotted against matched DNAm data. Dotted line at *y*=0

represents 50% DNAm. DNAm levels were plotted as M-Values and AEI data was Log2 transformed. Coloured circles represent individual samples. Differences in numbers (n) due to variable number of patient samples per CpG with matched AEI and DNAm data. *P*-values and r^2 values calculated using simple linear regression. * = $P < 0.05$; ns = not significant, ($P > 0.05$). The solid line represents the regression line with 95% confidence intervals represented by dashed lines.

4.2.4 Multi-tissue WWP2 expression profiling and eQTL analysis

Expression of the three *WWP2* transcript isoforms was quantified by RT-qPCR in each of the tissue types investigated. Comparisons of each isoform across all tissues revealed that *WWP2*-FL and *WWP2*-N are more abundantly expressed in OA cartilage compared to all other tissues (Fig. 4.9A; *WWP2*-FL, $P < 0.0001$; *WWP2*-N, $P < 0.0001$–0.0424). No other significant differences in tissue expression were detected for *WWP2*-FL and *WWP2*-N ($P = 0.0510$-0.9988). *WWP2*-C is more abundantly expressed in foetal cartilage than OA bone and OA synovium ($P < 0.0001$, 5.9% increase), and more abundantly expressed in OA cartilage than OA bone and OA synovium ($P < 0.0001$, 7.6-7.7% increase in OA cartilage). No significant difference in *WWP2*-C expression was detected between foetal cartilage and OA cartilage ($P = 0.2316$) or between OA bone and OA synovium ($P = 0.9999$).

Comparisons between the three *WWP2* transcript isoforms per tissue revealed that *WWP2*-C was more abundantly expressed in foetal cartilage and OA cartilage in comparison to *WWP2*-FL and *WWP2*-N (Fig. 4.9B; $P < 0.0001$; 5.9% increase in foetal cartilage; 8.1% increase in OA cartilage). In OA bone, *WWP2*-C was also more abundantly expressed ($P < 0.0001$–0.0060) but the percentage change was much smaller (0.04-0.06%). No significant difference in *WWP2*-FL and *WWP2*-N expression was detected in these tissues ($P = 0.1777$-0.9999). In OA synovium, *WWP2*-N was more abundantly expressed than *WWP2*-FL or *WWP2*-C ($P < 0.0001$–0.0020). No significant difference in *WWP2*-FL and *WWP2*-C expression was detected ($P = 0.4583$).

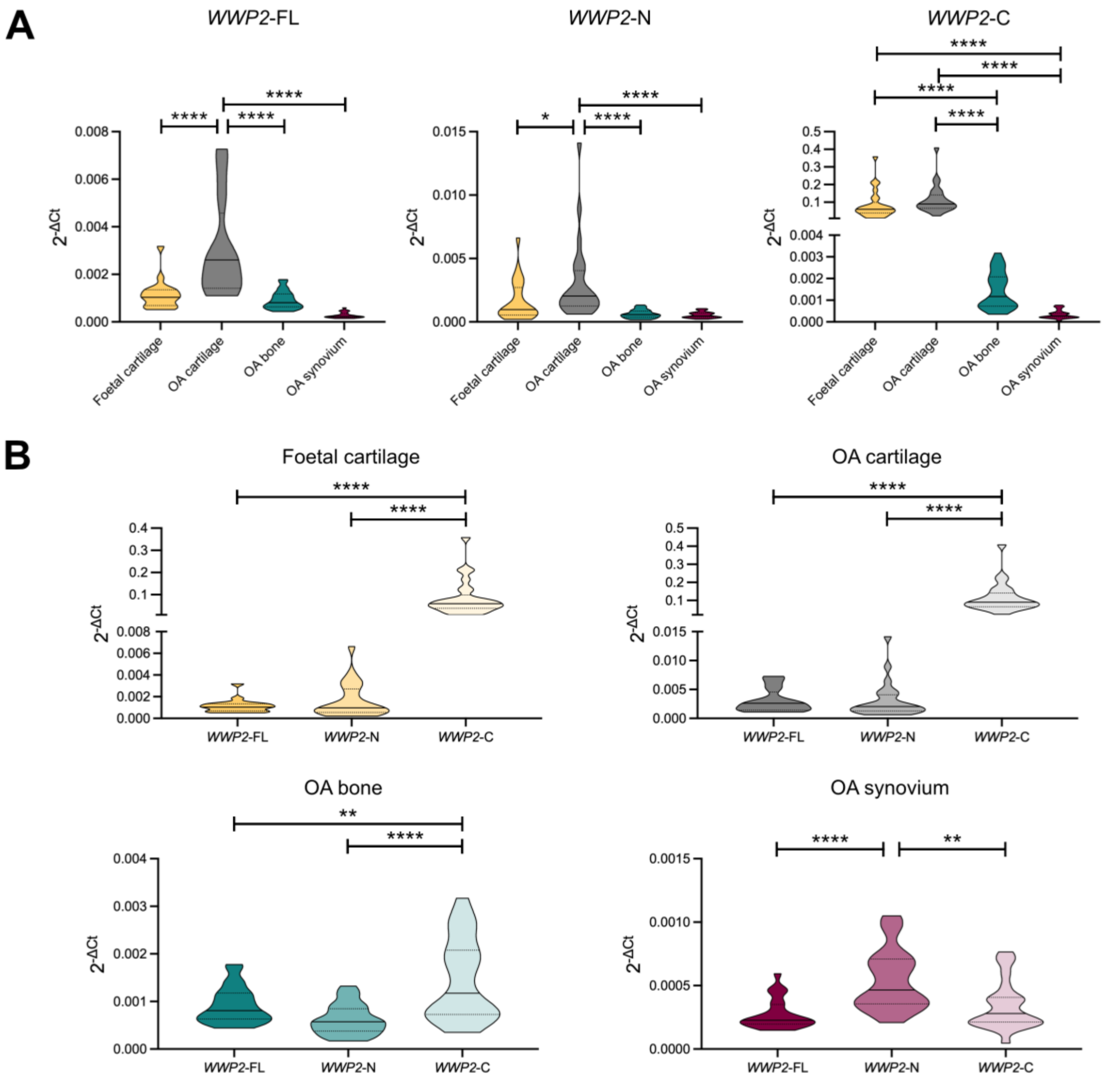


Figure 4.9: *WWP2* expression in synovial joint tissues. A) Expression of the *WWP2* transcript isoforms *WWP2*-FL, *WWP2*-N and *WWP2*-C compared between four tissue types: foetal cartilage (in yellow, n=25); OA cartilage (in grey, n=24); OA bone (in teal, n=24); and OA synovium (in maroon, n=23). **B)** Expression of the *WWP2* transcript isoforms *WWP2*-FL, *WWP2*-N and *WWP2*-C compared between each tissue. In the truncated violin plots, solid and dashed lines represent the median and interquartile range. *P*-values calculated by ordinary one-way ANOVA with Tukey's multiple comparisons test. * = $P < 0.05$; ** = $P < 0.01$; **** = $P < 0.0001$.

An eQTL analysis was performed using matched patient samples that were genotyped at rs34195470 in Section 4.2.1 (Fig. 4.10). No eQTLs were identified in any of the four tissues: OA cartilage (P = 0.5908-8270); foetal cartilage (P = 0.4941-9842); OA bone (P = 0.4693-8306); OA synovium (P = 0.5369-0.9500). This is likely due to the study being underpowered in comparison to other, larger scale eQTL analysis cohorts such as GTEx[87,213]. An meQTL analysis using RT-qPCR data was therefore not performed due to no correlations being present between *WWP2* expression and rs34195470 genotype.

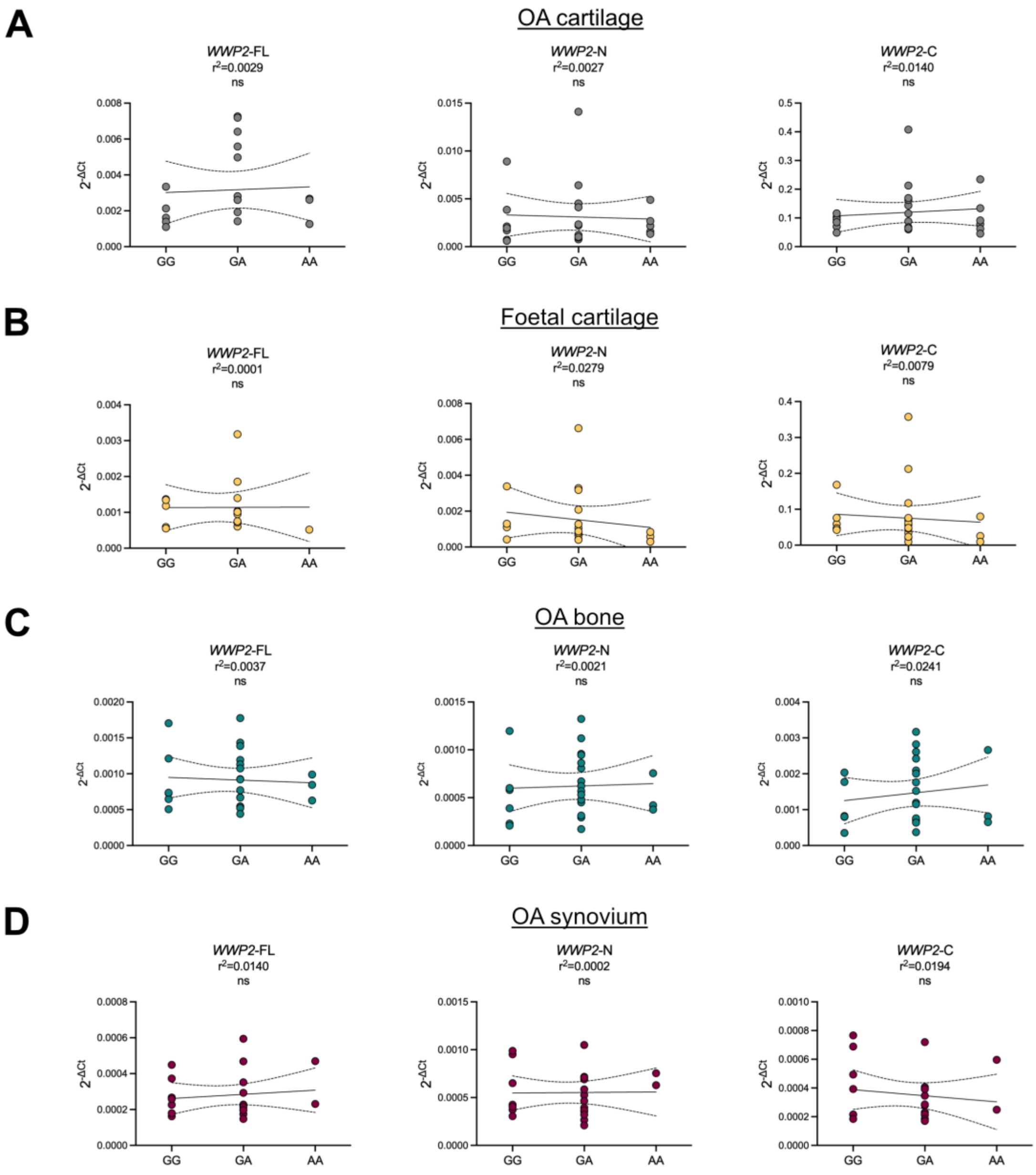


Figure 4.10: No eQTLs were detected for any of the three *WWP2* transcript isoforms in articular joint tissues. A) OA cartilage eQTL analysis (in grey, n=19-23). **B)** Foetal cartilage eQTL analysis (in yellow, n=15-23). **C)** OA bone eQTL analysis (in teal, n=24). **D)** OA synovium eQTL analysis (in maroon, n=22). For all panels, individual samples are represented by coloured circles. Differences in numbers (n) due to variable number of patient samples per transcript isoform not passing quality control. *P*-values and r^2 values calculated using simple linear regression. ns = not significant ($P > 0.05$). The solid line represents the regression line with 95% confidence intervals represented by dashed lines.

4.3 Discussion

At the 16 CpGs investigated in this chapter, mQTLs were identified in foetal cartilage, OA cartilage and OA bone. In OA cartilage, the previously reported mQTLs[87] at cg26736200 (CpG8) and cg26661922 (CpG13) were replicated with increased DNAm correlating with the rs34195470 risk allele (G). In 2022, Kreitmaier and colleagues also reported an OA cartilage mQTL at cg26736200 using the *WWP2* intronic SNV rs61611907, located approximately 8.6Kb upstream of rs34195470 (pairwise r^2 = 0.68, D' = 0.99 in European populations) and simultaneously identified it as a CpG with a potentially causal role in OA using Mendelian randomisation[158]. Although rs34195470 is reported as a knee OA risk signal, the DMR is detectable in both hip and knee cartilage. The risk-conferring allele was also confirmed by AEI analysis to be associated with increased *WWP2* expression.

The DMR was also active in foetal cartilage, exhibiting the same correlation observed in OA cartilage. However, both DNAm levels and the effect of rs34195470 genotype were lower in foetal compared to OA cartilage. An increase in CpG methylation over time is an active process mediated by DNMTs[221]. The DMR is therefore actively methylated after skeletogenesis and the effect of genotype upon DNAm levels increases in adulthood. The *WWP2* association signal may therefore be an example of antagonistic pleiotropy or genetic drift in OA, with the risk-conferring allele having a positive or neutral effect on joint formation but a negative effect on joint health as we age[127,273]. This is supported by the lack of AEI detected in foetal cartilage, suggesting allelic changes to *WWP2* expression occur after joint development. Age-related changes to DNAm levels in OA cartilage were limited to two CpGs which exhibited a modest but significant decline in CpG methylation. This may be due to a relatively common phenomenon towards global hypomethylation in ageing tissues known as epigenetic drift[161,274]. This is thought to be due to a loss of maintenance stringency, but it is often the case that specific loci are tightly regulated by SNVs to maintain their epigenetic state and thus contribute to OA pathogenesis[165], and this may apply to the remaining 14 CpGs.

In OA bone, the DMR was detectable and hypermethylated in comparison to foetal and OA cartilage. Of note, the mQTL effect was strikingly much stronger in males than in females despite rs34195470 not being marked as a sex-specific OA risk variant. This may be due to the SNV having roles in multiple joint tissues that increase OA susceptibility that are not sex-specific. Sex-specific mQTLs have been reported previously in OA cartilage[158,172,275] as well as other common polygenic diseases such as Alzheimer's[276]. Speculatively, this could be driven by sex hormones that may actively interact with an SNV in males, but not females, recruiting DNAm to the DMR in a genotype-dependent manner[164,277]. In this scenario, sex hormones may influence the cellular levels of TFs that bind rs34195470 in particular cell types[278]. These TFs may exhibit preferential binding for one of the alleles resulting in the observed mQTL effect. Additionally, sex hormones may influence chromatin state in subchondral bone at the DMR, in turn modulating DNAm status in a sex-specific manner[279]. However, by stratifying the mQTL data by sex the sample size is roughly halved and in turn, decreases statistical power whilst increasing the type I error rate. Replication of this analysis using an independent cohort of OA bone samples is therefore required before functional analysis. The AEI analysis revealed a trend towards decreased *WWP2* expression correlating with the risk-conferring allele, though this was not significant. No mQTLs or AEI were detected in OA synovium, suggesting that *WWP2* is not regulated by OA risk marked by rs34195470 in this tissue. Our methodology uses bulk synovial tissue rather than isolating single cells or performing single cell-type analyses. As synovium contains many cell types, it may be that the heterogeneity of this tissue masks the presence of a DMR[280]. cg26736200/CpG8 has been reported as an mQTL in OA synovium using the *WWP2* intronic SNV rs11647767, located approximately 1.7Kb upstream of rs34195470 (pairwise r^2 = 0.69, D' = 0.99 in European populations). As observed in cartilage, methylation of cg26736200 in synovium was identified potentially causal of OA using Mendelian randomisation[158]. However, these findings were not replicated in the present study.

A single meQTL was identified in foetal cartilage at CpG1, with *WWP2* expression decreasing with increased DNAm. Correlations between DNAm and target gene expression are difficult to identify due to the role of multiple *cis*-regulatory elements in fine-tuning gene expression[165]. Similarly, no eQTLs were detected for the three *WWP2* transcript isoforms in each tissue, a study by design that requires very large sample cohorts to overcome interindividual variability[87].

Using RT-qPCR, each of the *WWP2* isoforms was most abundantly expressed in OA cartilage. The three WWP2 isoforms regulate availability of SMAD2/3/7, which control TGFβ signalling activity[231,234]. Aberrant TGFβ signalling in cartilage is known to play a key role in OA, often resulting in cartilage degradation[49]. In addition, increased *WWP2*-FL expression in a 3D cartilage model has been shown to decrease expression of *ACAN* and *COL2A1*, suggesting this isoform confers detrimental effects upon ECM integrity and OA risk[243]. The expression of

WWP2-C was higher than the other two isoforms in both foetal cartilage and OA cartilage. *WWP2*-C is co-expressed with miR-140[244], which is required for cartilage development and homeostasis[189]. Expression of miR-140 was not measured by RT-qPCR due to small RNAs not being captured by the RNA isolation protocol that our group employs for joint tissues, which is a limitation of this chapter. However, miR-140 is known to be abundantly expressed in cartilage, with the miR-140-3p strand more abundantly expressed than the miR-140-5p strand[245]. The two miR-140 strands have also been reported to be expressed in OA synovium[281] and in bone[282].

In summary, these data demonstrate that the risk-conferring G allele of rs34195470 associates with increased DNAm at the DMR and increased *WWP2* expression in OA cartilage, validating previous reports. The DMR is also active during development and in OA bone, but the expression of *WWP2* is highest in OA cartilage. This implies that functional follow-up studies should be prioritised in chondrocytes. In the next chapter, I will aim to determine whether *WWP2* expression is regulated by the DMR identified in OA cartilage.

Chapter 5: Functional studies of the DMR in an immortalised chondrocyte cell line

5.1 Introduction

In Chapter 3, the originally reported rs34195470-mQTLs[87] (cg26736200/CpG8 and cg26661922/CpG13) were identified to reside within an active regulatory element in chondrocytes. In Chapter 4, these mQTLs were replicated and the study was expanded to identify an additional 12 mQTLs spanning a 228bp region in which the OA risk-conferring G allele of rs34195470 correlated with increased DNAm levels, revealing a DMR in OA cartilage. OA cartilage mQTLs are enriched in putative chondrocyte enhancers[88], with several studies demonstrating that experimental perturbation of DNAm levels at mQTL sites *in-vitro* can modulate gene expression[166,171–173]. Therefore, mQTLs can act as functional intermediaries between SNV and target gene. In the AEI analysis, the OA risk-conferring G allele of rs34195470 correlated with increased *WWP2* expression in OA cartilage, validating previous findings[214]. These data suggest increased DNAm at the DMR may drive increased *WWP2* expression in OA cartilage. Functional follow-up studies are now required to provide causative evidence of epigenetic regulation exerted by the DMR.

Epigenetic editing, which utilises a catalytically inactive 'dead' Cas9 protein fused to epigenetic modulators such as TET1 (for de-methylation) or DNMT3A (for methylation), is a valuable tool for experimental recapitulation of mQTLs *in-vitro*[166,171–173]. Understanding the regulatory capacity of mQTL regions can also be informed by reporter assays that assess promoter or enhancer function in an unmethylated or methylated state[171–173]. Once these mechanisms are characterised, the next step is to understand how these transcriptional effects are controlled. Transcription factors (TFs) are the key mediators of transcriptional regulation, binding to specific DNA sequences or 'motifs'. Methylation of CpGs within TF motifs can affect DNA binding affinity, altering transcriptional regulation and downstream expression of effector genes[133,221,283]. Therefore, the identification of TFs binding directly across CpGs marked as OA-associated mQTLs may point towards a molecular mechanism.

Combining functional experiments conducted using the immortalised chondrocyte cell line TC28a2[266] and an *in-silico* TF binding prediction tool, this chapter aims to:

1) Characterise the regulatory capacity of the DMR as a promoter and an enhancer.
2) Determine whether DMR regulatory function is methylation-sensitive.
3) Identify the target(s) genes of the DMR, if any, using epigenetic editing.
4) Identify TFs predicted to bind across the *WWP2* DMR and quantify their abundance in OA cartilage.

5.2 Results

5.2.1 Lucia reporter gene assay

The DMR was cloned into Lucia reporter gene vectors as a promoter and as an enhancer to investigate regulatory function in an unmethylated or methylated state and transfected into TC28a2 chondrocytes (Fig. 5.1A). When cloned as a promoter, the DMR repressed Lucia expression in comparison to an empty vector control for both the unmethylated DMR and methylated DMR constructs (Fig. 5.1B, left panel): unmethylated DMR versus control, $P = 0.0001$; methylated DMR versus control, $P < 0.0001$. The repressive effect was strengthened by methylation of the DMR in these constructs (unmethylated DMR versus methylated DMR, $P < 0.0001$).

When cloned as an enhancer, the DMR repressed Lucia expression in comparison to an empty vector control for both the unmethylated DMR and methylated DMR constructs (Fig. 5.1B, right panel): unmethylated DMR versus control, $P < 0.0255$; methylated DMR versus control, $P < 0.0001$. As observed in the promoter construct, the repressive effect upon enhancer activity was strengthened by methylation of the DMR in these constructs (unmethylated DMR versus methylated DMR, $P = 0.0173$). This data implies the DMR is a methylation-sensitive regulator of gene expression.

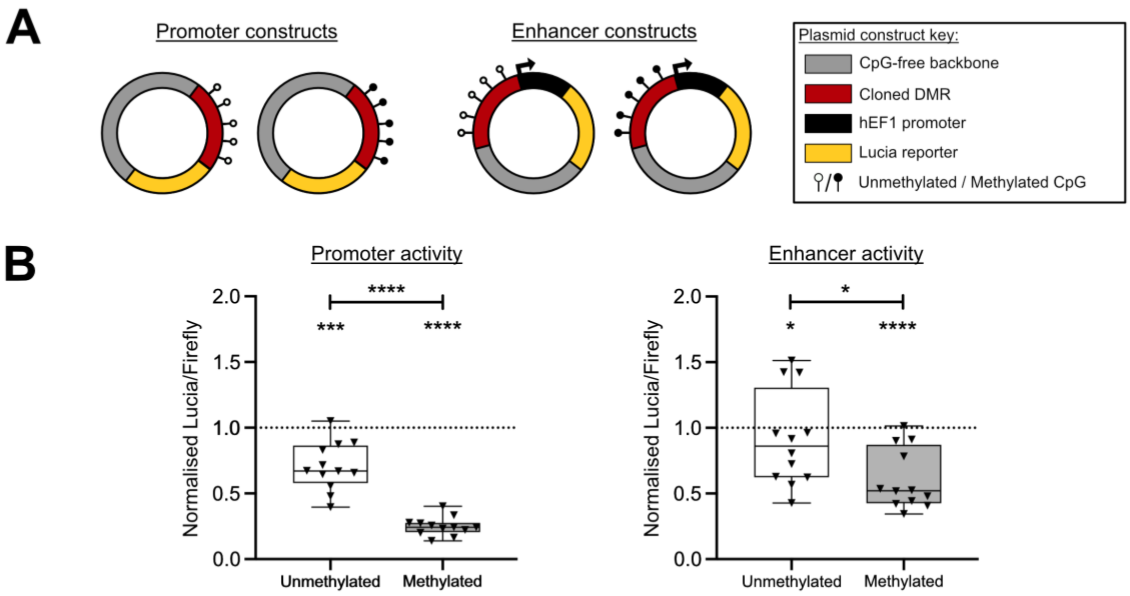


Figure 5.1: Investigation of the transcriptional regulatory function of the DMR in TC28a2 chondrocytes. A) Schematic representation of the promoter and enhancer plasmid vector constructs. **B)** Lucia reporter assays assessing promoter or enhancer activity in the presence

of construct containing the DMR in an unmethylated or methylated state. Values were normalised to those in empty vector control (dotted horizontal line). Black dots represent individual samples (12 biological replicates per group). Box plots show the median, 25th and 75th percentiles, and minimum and maximum values. *P*-values calculated by Mann Whitney U test with Holm-Šídák correction. * = P < 0.05; *** = P < 0.001; **** = P < 0.0001.

5.2.2 dCas9-DNMT3A epigenetic editing of the DMR

To observe the effects of differential DNAm at the DMR upon the expression of the three common isoforms of *WWP2* (*WWP2*-FL, *WWP2*-N and *WWP2*-C) and upon the two mature miR-140 strands (mir-140-5p and miR-140-3p), targeted epigenetic editing of the DMR was performed in TC28a2 chondrocytes. These cells are heterozygous (GA) at rs34195470 and moderately hypomethylated at the DMR, with DNAm levels ranging from 1.9% to 23.4% across the 16 CpGs (Fig. 5.2B, control data, in black). The DMR was therefore methylated using gRNA-dCas9-DNMT3A constructs, recapitulating the effect of the OA risk-conferring G allele of rs34195470.

To target the DMR in both orientations of the genome, five gRNAs were designed (gRNA 1-5, Fig. 5.2A). An increase in DNAm (> 5%) at a minimum of three CpGs was observed for each gRNA (Table 5.1), with gRNA5 increasing the levels of DNAm at 14 of 16 CpGs (Fig. 5.2B, in blue). The largest overall increase in DNAm was 37.7% at CpG7 by gRNA2 (Fig. 5.2B, in red). The mean percentage increase in DNAm levels across the DMR ranged from 4.0% (gRNA1) to 18.8% (gRNA5). Successful epigenetic editing of the DMR resulted in increased expression of *WWP2*-FL (*P* = 0.0100-0.0135) and *WWP2*-N (*P* = 0.0050-0.0340) for all five gRNAs used in comparison to a nontargeting control (Fig. 5C). No significant changes in expression for any gRNA were observed for *WWP2*-C (*P* = 0.9470), miR-140-5p (*P* = 0.0750-0.2340) or miR-140-3p (*P* = 0.1445-0.7100).

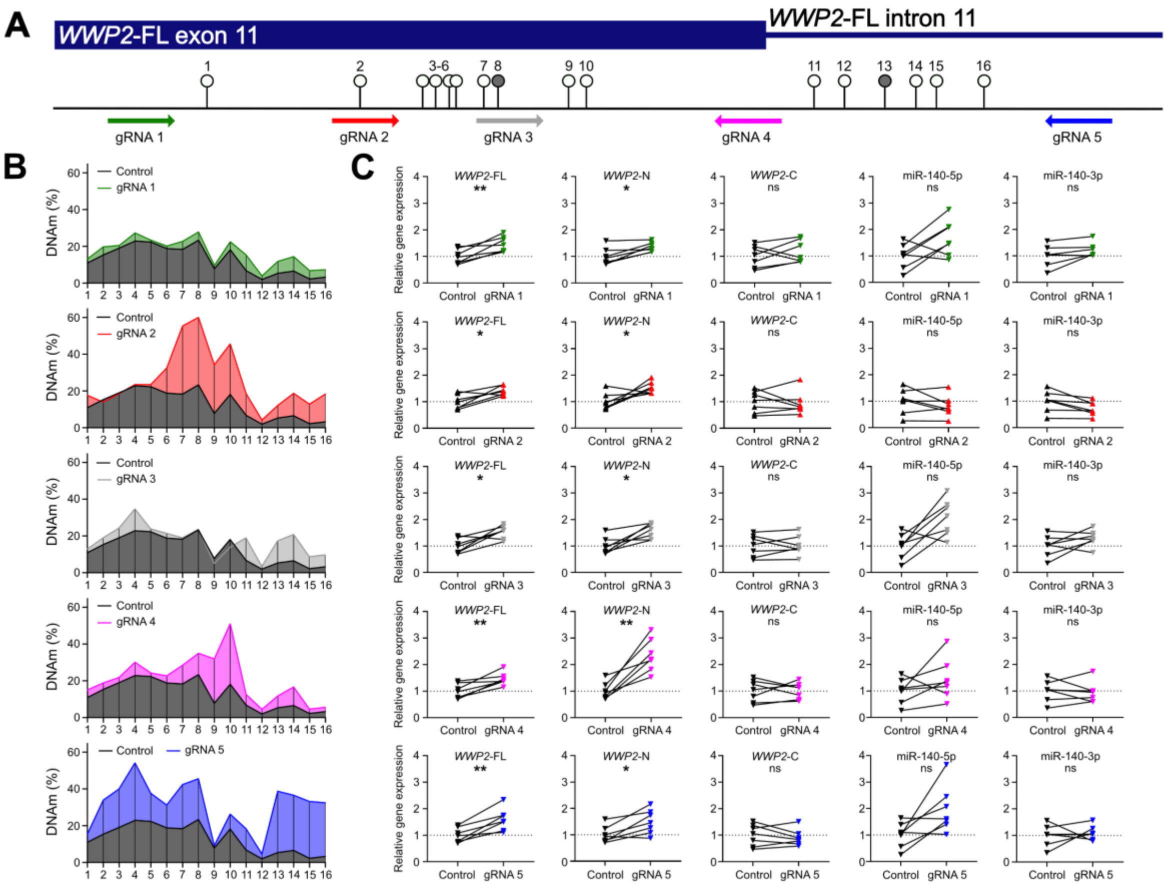


Figure 5.2: Epigenetic modulation of the DMR in TC28a2 chondrocytes. A) Schematic representation of the 16 CpGs and of the five gRNAs (arrow pointing left, antisense strand; arrow pointing right, sense strand), and their genomic positioning with regard to the *WWP2*-FL gene. cg26736200/CpG8 and cg26661922/CpG13 are highlighted. **B)** Mean DNAm levels (%) of the 16 CpGs following expression of dCas9-DNMT3A protein in control (black), with nontargeting gRNA, or in samples with a targeting gRNA (coloured). Seven replicates for control and for each targeting gRNA. **C)** Relative expression of the three *WWP2* transcripts (*WWP2*-FL, *WWP2*-N, *WWP2*-C) and of the two miR-140 strands (miR-140-5p, miR-140-3p) following epigenetic editing. Values were normalised to nontargeting gRNA control. *P*-values calculated using a paired t-test with Benjamini-Hochberg correction. * = $P < 0.05$; ** = $P < 0.01$; ns = not significant ($P > 0.05$).

Table 5.1. Summary of the effects of dCas9-DNMT3A epigenetic editing upon DNAm levels at the DMR. Percentage change at each CpG site per gRNA in comparison to a nontargeting gRNA control.

gRNA	Number of CpGs edited (DNAm change >5%)	Mean percentage change in DNAm across DMR	Largest overall percentage change in DNAm
1	3	4.0%	CpG11 (8.5%)
2	11	13.2%	CpG7 (37.7%)
3	7	4.6%	CpG14 (14.4%)
4	8	8.3%	CpG10 (33.0%)
5	14	18.8%	CpG13 (33.7%)

5.2.3 In-silico prediction of TF binding sites across the DMR

Targeted epigenetic editing in TC28a2 chondrocytes using dCas9-DNMT3A revealed *WWP2*-FL and *WWP2*-N are regulated by DNAm at the DMR. If this were the case, the DMR would be expected to bind TFs[221]. To identify TF binding motifs at the DMR, the JASPAR Core 2022 dataset[223] was mapped to the DMR using the UCSC Genome Browser. Ten TFs were predicted to bind at or near to the DMR CpGs (Fig. 5.3A, $P < 1x10^{-5}$). To determine whether these TFs were abundantly expressed in OA cartilage, RNA-Seq data was plotted (Fig. 5.3B). Of the ten TFs predicted to bind the DMR, five were abundantly expressed in cartilage (transcripts per million, TPM > 10): *HIF1A*, encoding Hypoxia Inducible Factor 1 Subunit Alpha (HIF-1α); *RXRA*, encoding Retinoid X Receptor Alpha; *NR2C2*, encoding Nuclear Receptor Subfamily 2 Group C Member 2; *CREB3L2*, encoding CAMP Responsive Element Binding Protein 3 Like 2; and *MAX*, encoding MYC Associated Factor X. Of these five TFs, only HIF-1α was predicted to bind CpGs marked as mQTLs in OA cartilage (Table 5.2).

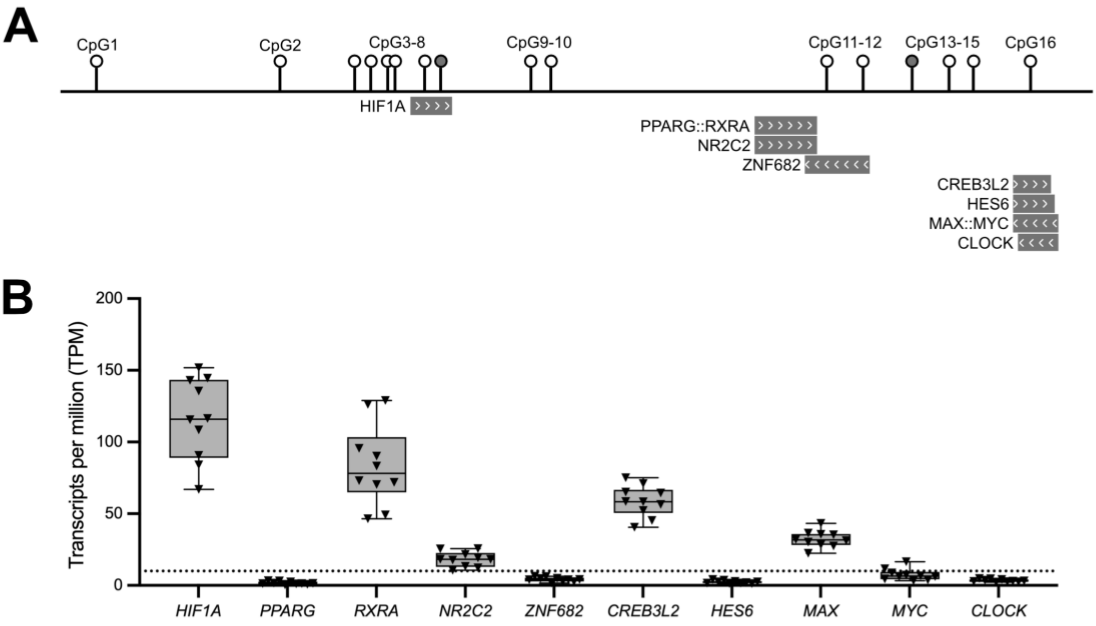


Figure 5.3: TFs predicted to bind at the DMR. A) Schematic representation of the 16 CpGs and of 10 TFs predicted to bind at or close to the CpGs. cg26736200 (CpG8) and cg26661922 (CpG13) are highlighted. The TFs are marked by grey rectangles with the direction of the arrows within the rectangles indicating the DNA strand the TF is predicted to bind to (arrows pointing to the left = antisense strand, arrows pointing to the right = sense strand). TF heterodimers are denoted by a double colon (::) between two TFs. **B)** Expression levels (TPM, transcripts per million) of the TFs in cartilage chondrocytes from 10 OA patients. Triangles

represent individual samples. Five TFs were abundantly expressed (median TPM > 10, represented by dotted line at y=10). Box plots show the median, 25th and 75th percentiles, and minimum and maximum values.

Table 5.2: List of TFs predicted to bind the DMR. The table lists the gene symbol, the encoded TF name, the CpGs predicted to be bound by TFs, whether the CpG was a significant mQTL in OA cartilage, and whether the TF was abundantly expressed in OA cartilage (median TPM > 10). N/A, not applicable.

Gene symbol	TF name	CpGs bound	OA cartilage mQTL?	Expression > 10 TPM?
HIF1A	Hypoxia Inducible Factor 1 Subunit Alpha	7 and 8	Yes	Yes
PPARG	Peroxisome Proliferator Activated Receptor Gamma	None	N/A	No
RXRA	Retinoid X Receptor Alpha	None	N/A	Yes
NR2C2	Nuclear Receptor Subfamily 2 Group C Member 2	None	N/A	Yes
ZNF682	Zinc Finger Protein 682	11 and 12	CpG 11 – yes CpG 12 – no	No
CREB3L2	CAMP Responsive Element Binding Protein 3 Like 2	16	No	Yes
HES6	Hes Family BHLH Transcription Factor 6	16	No	No
MAX	MYC Associated Factor X	16	No	Yes
MYC	MYC Proto-Oncogene BHLH Transcription Factor	16	No	No
CLOCK	Clock Circadian Regulator	16	No	No

5.3 Discussion

In this chapter, the *WWP2* DMR was studied *in-vitro* using reporter assays and epigenetic editing in the immortalised chondrocyte cell line TC28a2 to identify effectors of OA risk marked by rs34195470. Lucia reporter assays showed that the DMR has transcriptional effects, acting as a methylation-sensitive repressor of gene expression in both promoter and enhancer constructs. Epigenetic editing mediated by dCas9-DNMT3A demonstrated that increased DNAm levels at the DMR resulted in higher *WWP2* expression. The discordancy between the reporter assays and the epigenetic editing experiments likely reflects the relatively artificial nature of the reporter assay which, unlike epigenome editing, investigates a piece of DNA isolated from its normal genomic and chromatin context. *WWP2*-FL and *WWP2*-N were revealed as the target transcripts of the DMR, with no significant effects on the expression of *WWP2*-C or miR-140.

Whilst the use of a nontargeting gRNA control means that the dCas9-DNMT3a protein would not be targeted to the DMR, the observed changes in *WWP2*-FL and *WWP2*-N expression mediated by gRNAs 1-5 may not be solely explained by epigenetic editing. The binding of gRNAs to the region and the subsequent proximity of dCas9-DNMT3a protein to the DMR may block DNA-binding proteins (including TFs) by steric hindrance, potentially driving the observed effects upon gene expression independently of changes to DNAm levels[88,221]. A more appropriate control may be the use of dCas9 fused to catalytically inactive 'dead' DNMT3a (dDNMT3a), where the active site of DNMT3a is mutated by an E155A* substitution, thus inactivating methyltransferase activity[220]. This technology has since been used in two functional studies of OA-associated mQTLs at the *COLGALT2* locus[167,216].

The small relative fold changes observed in the expression of *WWP2*-FL and *WWP2*-N following dCas9-DNMT3A epigenetic editing match the fold differences in allelic expression measured in the AEI analysis using OA cartilage from patient samples (mean rs1052429 A/G ratio of 1.15; Chapter 4, Figure 4.6A) and reflect the small effect sizes observed for most polygenic risk loci[87]. In line with the proposed "liability threshold" model of polygenic disease, OA genetic risk conferred by rs34195470 and the associated DMR likely contributes to OA only when inherited with multiple other risk loci[125]. Both the reporter assays and dCas9-DNMT3A editing experiments were conducted using the immortalised chondrocyte cell line TC28a2, which does not reflect the true biological nature of articular chondrocytes[284]. Future studies undertaken on primary cells, including OA patient chondrocytes, may provide further insight into the role of the DMR as a regulator of *WWP2*.

In Chapter 3, the DMR was identified to physically interact with the promoter region of *WWP2*-FL and *WWP2*-N via the formation of a chromatin loop in MSCs. Epigenetic editing revealed these isoforms as targets of the DMR, which is predicted to bind TFs expressed in OA cartilage. Overall, these results suggest that the rs34195470 association signal regulates the expression of *WWP2*-FL and *WWP2*-N via the epigenetic intermediate of DNAm. Based on these findings, the following molecular mechanism is proposed: (1) the DMR is part of a transcriptional activator; (2) it loops to the shared promoter of *WWP2*-FL and *WWP2*-N; (3) in the presence of OA-risk conferring allele G of rs34195470, the DMR is more highly methylated; (4) this modulates the binding of TFs to the DMR and increases its transcriptional activation potential; (5) this results in greater relative expression of *WWP2*-FL and *WWP2*-N (Fig. 5.4).

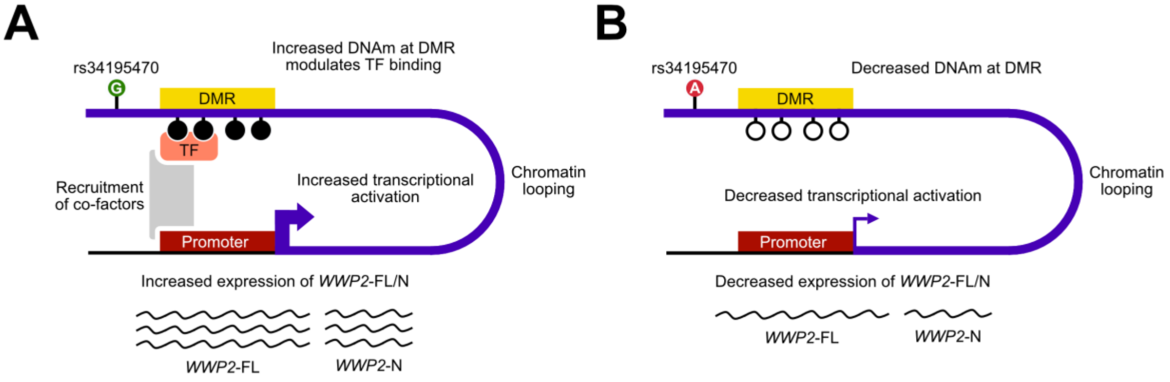


Figure 5.4: Proposed molecular mechanism. A) The DMR is part of a transcriptional activator that loops to the shared promoter of *WWP2*-FL and *WWP2*-N. In the presence of the OA-risk conferring allele G of rs34195470, the DMR is more highly methylated. This modulates binding of TFs to the DMR and increases its transcriptional activation potential via the recruitment of co-factors. This results in greater relative expression of *WWP2*-FL and *WWP2*-N. **B)** In the presence of the OA non-risk conferring allele A of rs34195470, the DMR is less methylated. This results in decreased transcriptional activation potential and decreased relative expression of *WWP2*-FL and *WWP2*-N. Methylated CpGs are represented by filled circles; unmethylated CpGs are represented as unfilled circles. *WWP2*-FL mRNA is represented by longer wavy lines; *WWP2*-N mRNA is represented by shorter wavy lines.

Ten TFs are predicted to bind the *WWP2* DMR, with five abundantly expressed in OA cartilage. Of these five, only HIF-1α bound CpGs marked as OA cartilage mQTLs. Increased DNAm levels at the CpGs constituting the DMR drives increased expression of *WWP2*-FL and *WWP2*-N, suggesting that a methylation-sensitive TF may mediate the observed effects. The HIF-1α binding motif is known as the hypoxia response element (HRE) and is well-characterised (5'-RCGTG-3', where R represents purine bases A or G)[285]. This sequence complements the surrounding sequence of CpG7 (5'-ACGTG-3'). Methylation of the CpG site within the HRE repels DNA binding of HIF-1α in studies of erythropoietin enhancer activity and in genome-wide studies of tumour immunotolerance via steric hindrance[286,287]. Conversely, other TFs have been demonstrated to preferentially bind specific methylated CpG sites, including the basic leucine zipper (bZIP) CCAAT/enhancer-binding protein alpha (CEBPα), the zinc-finger protein ZFP57, and the ZFP57-cofactor KRAB-associated protein 1 (KAP1)[288]. Interestingly, two other TFs identified by JASPAR, RXRα and PPARγ, have also been reported to bind methylated CpG sites in certain genomic contexts[289], although neither were predicted to bind directly across an OA cartilage mQTL site.

Other mechanisms influenced by changes to DNAm levels at the DMR could also produce the observed effects upon *WWP2* expression. For example, DNAm can alter local chromatin accessibility preventing TF binding of the DMR or encourage the binding of methyl-CpG binding domain proteins (MBDs) which block TF binding to proximal methylation-independent motifs[283,290]. However, *HIF1A* was the most abundantly expressed TF predicted to bind the DMR and is a known regulator of chondrogenesis and cartilage homeostasis[21,291]. Tuerlings and colleagues also highlighted hypoxia as an important pathway in the activity of the rs34195470 association signal following upregulation of *WWP2*-FL expression in a 3D cartilage model[243]. In addition to CpG7, HIF-1α is predicted to bind across CpG8/cg26736200. This CpG has been highlighted as potentially causal of OA by Mendelian randomisation using the *WWP2* intronic SNV rs61611907 (pairwise r^2 = 0.68, D' = 0.99 with rs34195470 in European populations) in OA cartilage[158]. These data imply that CpG8/cg26736200, HIF-1α and hypoxia are key drivers of the association signal. In the next chapter, their interactions will be investigated.

Chapter 6: The role of HIF-1α as a transcriptional regulator of *WWP2*

6.1 Introduction

In the previous chapters, epigenetic editing and *in-silico* analyses indicated that the DMR is a regulator of gene expression via interaction with the promoter region of *WWP2*-FL and *WWP2*-N. Modulation of DNAm levels at the DMR has direct effects on the expression of these isoforms, a mechanism that could be driven by altered TF binding affinity to CpGs constituting the DMR depending on their methylation status. The hypoxia-inducible factor HIF-1α was predicted to bind CpG7 and CpG8/cg26736200, suggesting it may be capable of mediating the observed effect upon *WWP2* expression.

HIF1A encodes the alpha subunit of the heterodimeric TF HIF-1, with the beta subunit (HIF-1β) encoded by *ARNT* (Aryl Hydrocarbon Receptor Nuclear Translocator)[292]. HIF-1α is continuously produced yet rapidly degraded under normal cellular O_2 levels (normoxia), whilst HIF-1β remains stable regardless of O_2 concentration[293]. Degradation of HIF-1α is mediated by prolyl hydroxylase (PHD) enzymes which catalyse the hydroxylation of two prolyl residues (Pro^{402} and Pro^{564}) within the O_2-dependent degradation domain of HIF-1α, a mechanism that requires cellular O_2 and Fe^{2+} as substrates for catalytic activity[21]. The hydroxylated HIF-1α is recognised by the Von Hippel-Lindau Tumour Suppressor protein (VHL), which is part of an E3 ligase complex and facilitates polyubiquitination of HIF-1α. HIF-1α is subsequently degraded by the ubiquitin-proteasomal pathway[294]. In reduced O_2 conditions (hypoxia), PHD hydroxylation is inhibited, thereby preventing O_2-dependent degradation of HIF-1α[21]. HIF-1α can therefore accumulate and dimerise with HIF-1β, interact with its DNA binding motif and exert the transcriptional response to hypoxia[293].

Due to the instability of HIF-1α in the presence of O_2, *in-vitro* investigations of its role typically require the use of a hypoxic chamber to ensure its long-term stability over the course of experiments. Where this equipment is unavailable, researchers often utilise chemical hypoxia mimetics such as cobalt chloride ($CoCl_2$) or dimethyloxyalylglycine (DMOG) to interrupt the mechanism of O_2-dependent degradation of HIF-1α[295,296]. The use of $CoCl_2$ in cell culture induces a similar transcriptional profile to true hypoxia and is thought to stabilise HIF-1α protein via displacement of Fe^{2+} in the catalytic site of PHDs, inhibiting their hydroxylation activity and preventing subsequent degradation of HIF-1α[297].

Two other HIF-α subunits have been identified: HIF-2α (encoded by *HIF2A*) and HIF-3α (encoded by *HIF3A*). These subunits are also susceptible to degradation by PHDs in the presence of O_2 and form heterodimeric TFs with HIF-1β. HIF-1α and HIF-2α exhibit high levels of structural identity and sequence similarity, both binding the same consensus DNA sequence (5'-RCGTG-3', where R represents purine bases A or G)[285,298]. Their DNA binding activity is influenced by epigenetic factors such as histone modifications and DNAm (due to the presence of a CpG site within the HIF-α binding motif)[298]. Less is known about HIF-3α, which is structurally dissimilar to HIF-1α and HIF-2α[299].

Articular cartilage is avascular and hypoxic, with its supply of O_2 and other nutrients provided via diffusion from vascularised subchondral bone and the synovial fluid which bathes the joint surface[300]. This hypoxic environment is maintained throughout life; therefore, chondrocytes must be adapted to these conditions to ensure their survival[21]. HIF-1α and HIF-2α mediate the transcriptional response to hypoxia, regulating changes to cellular metabolism as well as playing key roles in cartilage development and homeostasis[21,301]. HIF-1α activates *SOX9* expression during early skeletogenesis and maintains the expression of pro-anabolic markers *COL2A1* and *ACAN* whilst suppressing proliferation of adult chondrocytes[302–304]. HIF-2α is upregulated in OA cartilage and induces the expression of several genes associated with cartilage catabolism and chondrocyte hypertrophy, including *MMP13* and *RUNX2*[305,306]. Together, the HIFs mediate anabolic (via HIF-1α) and catabolic (via HIF-2α) processes within the ECM, with their dysregulation contributing to cartilage loss and OA[21,307,308].

Employing the previously used functional tools conducted using TC28a2 chondrocytes in combination with gel-shift assays, this chapter aims to:

1) Validate that HIF-1α can bind the DMR in chondrocytes.
2) Determine whether HIF-1α binding is methylation-sensitive.
3) Characterise the regulatory capacity of the DMR as a promoter and as an enhancer in the presence of HIF-1α.
4) Identify changes to *WWP2* and miR-140 expression following epigenetic editing of the DMR in the presence of HIF-1α.

6.2 Results

6.2.1 HIF-1α is expressed in TC28a2 cells and can bind the WWP2 DMR in-vitro

HIF-1α is rapidly degraded under normoxic conditions by PHDs. To inhibit their action, 200μM $CoCl_2$ was added to cell culture medium for 24 hours. HIF-1α protein expression was measured by Western blot using nuclear protein lysates derived from TC28a2 cells cultured in normal cell culture medium (control) or in cell culture medium supplemented with 200μM $CoCl_2$ ($CoCl_2$-treated, Fig. 6.1A). The intensity of HIF-1α protein abundance in the $CoCl_2$-treated condition was markedly stronger than the control, indicating that $CoCl_2$ successfully stabilised HIF-1α protein expression.

HIF-1α is predicted to bind the DNA sequence harbouring CpG7 and CpG8/cg26736200. To determine whether this DNA sequence is capable of binding TFs in an unmethylated or methylated state, an electrophoretic mobility shift assay (EMSA) was performed using control or $CoCl_2$-treated TC28a2 nuclear protein (Fig. 6.1B). Protein-DNA complexes were observed in all conditions. Some protein-DNA complexes were common across all conditions (Fig. 6.1B, green triangle), whilst others exhibited preferential binding to methylated DNA probes (Fig. 6.1B, pink triangle). One protein-DNA complex was identified to be unique to $CoCl_2$-treated TC28a2 nuclear protein incubated with unmethylated DNA probes (Fig. 6.1B, yellow triangle), suggesting this protein is only stably expressed when PHD activity is inhibited and only capable of binding unmethylated DNA.

As HIF-1α is predicted to bind these CpGs, an antibody specific to HIF-1α (anti-HIF-1α) was added to the EMSA reaction to observe a super-shift, where the molecular weight of the antibody-protein-DNA complex would be higher than the protein-DNA complex alone, therefore retarding its movement through the gel. A super-shift was observed for $CoCl_2$-treated TC28a2 nuclear protein incubated with unmethylated probes and anti-HIF-1α, confirming that this unique interaction was indeed HIF-1α (Fig. 6.1C). HIF-1α therefore binds CpG7 and CpG8/cg26736200 in a methylation-sensitive manner, where the presence of DNAm at these CpGs blocks DNA binding by HIF-1α. No super-shifts were observed in any of the other conditions.

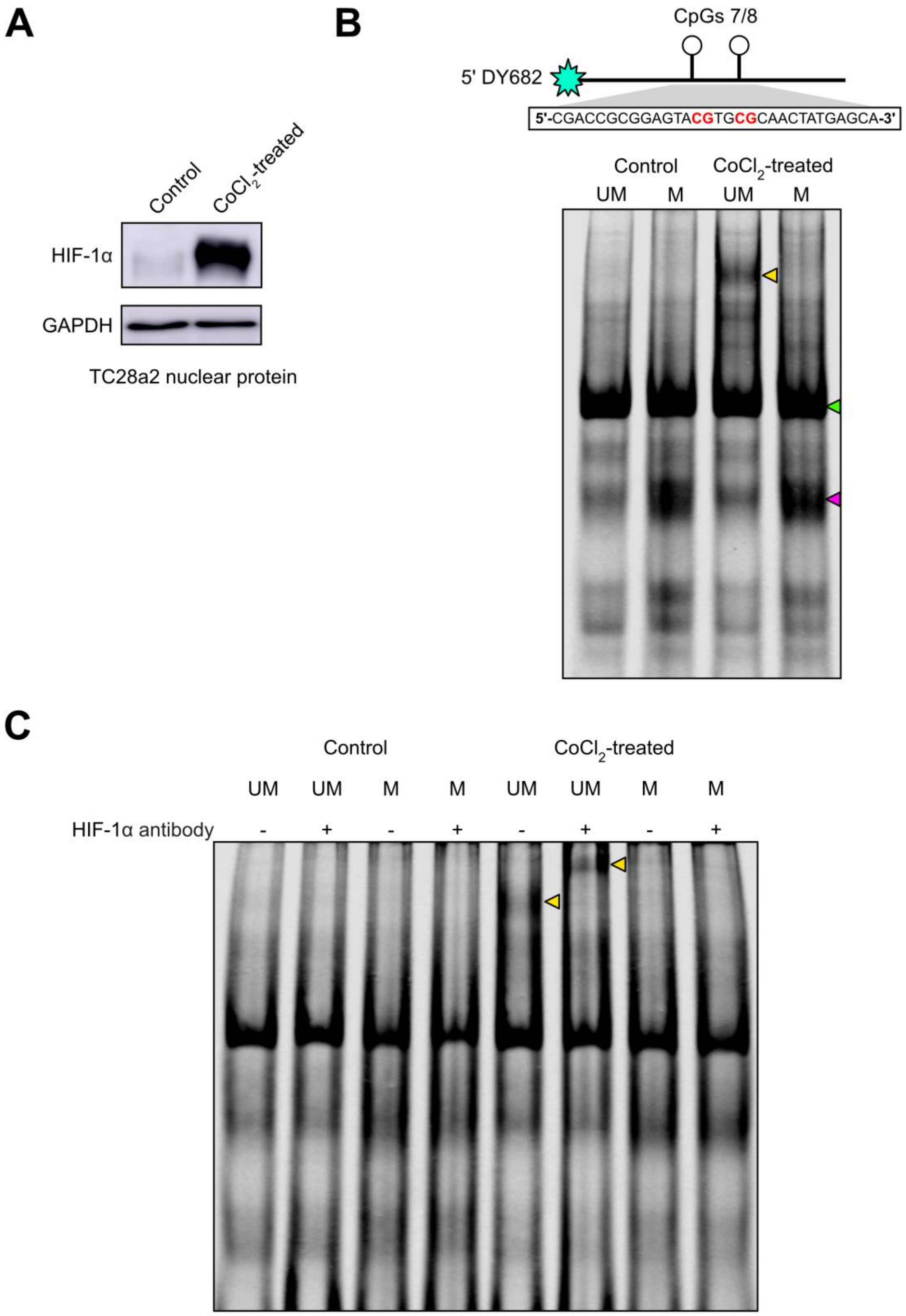


Figure 6.1: HIF-1α is expressed in TC28a2 chondrocytes and can bind the *WWP2* DMR. A) Western blot using TC28a2 nuclear protein derived from cells cultured in normal cell culture medium (control) or in cell culture medium supplemented with 200μM $CoCl_2$ ($CoCl_2$-treated).

GAPDH was used as a loading control. **B)** EMSA reaction demonstrating the interaction between 5' DY682-tagged fluorescent DNA probes encompassing the sequence surrounding CpG7 and CpG8/cg26736200 in an unmethylated (UM) or methylated (M) state, and TC28a2 nuclear protein (control or $CoCl_2$-treated). The DNA probe sequence is shown with CpGs 7 and 8 highlighted in bold, red text. Triangles indicate protein-DNA binding complexes. Yellow triangle, protein-DNA complex present only in lane with $CoCl_2$-treated nuclear protein with UM DNA probes. Green triangle, DNA-protein complex present in all conditions. Pink triangle, DNA-protein complex preferentially binding M probes. **C)** EMSA super-shift reaction demonstrating the interaction between DNA probes, TC28a2 nuclear protein, and the presence (+) or absence (-) of HIF-1α antibody. Yellow triangles indicate the presence of a super-shift for lanes with UM probes and $CoCl_2$-treated TC28a2 nuclear protein.

6.2.2 Lucia reporter gene assays in $CoCl_2$-treated TC28a2 chondrocytes

HIF-1α binds the *WWP2* DMR in a methylation-dependent manner, suggesting it can mediate transcriptional activity of the DMR observed in Chapter 5. However, the functional studies performed in the previous chapter were conducted in cell culture medium that would facilitate the rapid degradation of HIF-1α. These experiments were therefore repeated using $CoCl_2$-supplemented cell culture medium to maintain HIF-1α protein expression.

The same Lucia reporter constructs were utilised, with the DMR cloned as a promoter and as an enhancer to investigate regulatory function in an unmethylated or methylated state (Fig. 6.2A). The Lucia reporter constructs were then transfected into $CoCl_2$-treated TC28a2 chondrocytes. When cloned as a promoter, no transcriptional activity in comparison to an empty vector control was detected for the unmethylated DMR construct (Fig. 6.2B, left panel, $P = 0.4797$). However, the methylated DMR construct acted as a transcriptional repressor in comparison to the unmethylated DMR construct ($P < 0.0001$) and to an empty vector control ($P < 0.0001$).

When cloned as an enhancer, the unmethylated DMR construct acted as a weak enhancer of Lucia expression in comparison to an empty vector control (Fig. 6.2B, right panel), whilst the methylated DMR construct repressed enhancer activity: unmethylated DMR versus control, $P = 0.0002$; methylated DMR versus control, $P < 0.0001$. The repressive effect upon enhancer activity was strengthened by methylation of the DMR (unmethylated DMR versus methylated DMR, $P = 0.0017$). This data implies the DMR is a methylation-sensitive regulator of gene expression, with opposing transcriptional effects depending on methylation status.

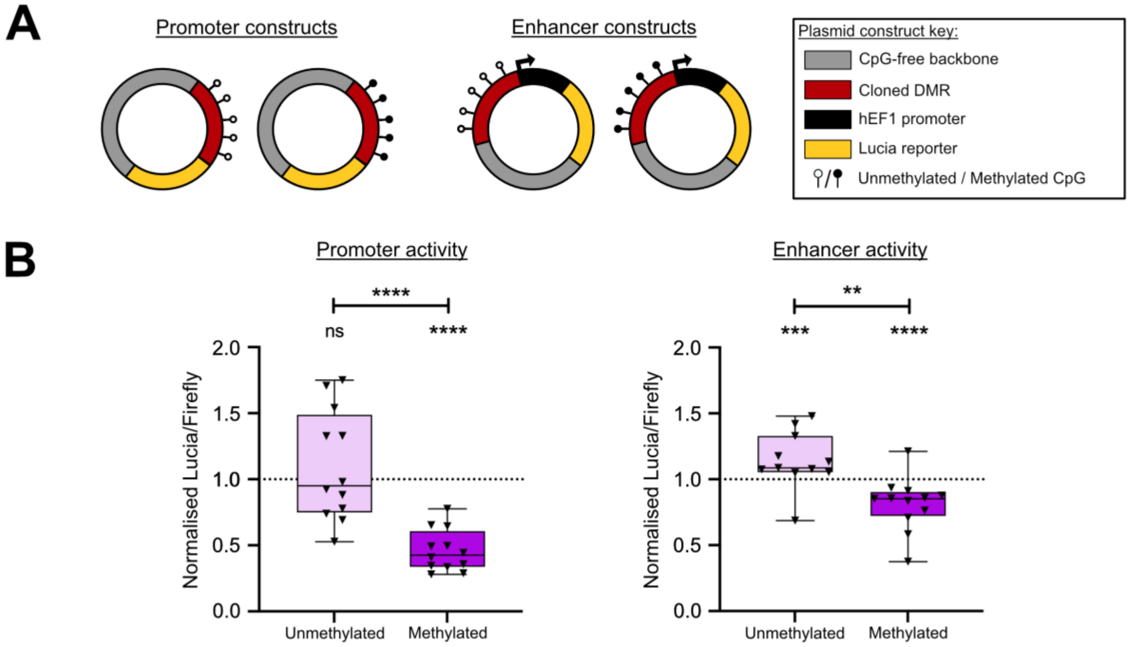


Figure 6.2: Investigation of the transcriptional regulatory function of the DMR in $CoCl_2$-treated TC28a2 chondrocytes. A) Schematic representation of the promoter and enhancer plasmid vector constructs. **B)** Lucia reporter assays assessing promoter or enhancer activity in the presence of construct containing the DMR in an unmethylated or methylated state. Values were normalised to those in empty vector control (dotted horizontal line). Black dots represent individual samples (12 biological replicates per group). For the unmethylated enhancer construct, one outlier was removed using the ROUT method. Box plots show the median, 25th and 75th percentiles, and minimum and maximum values. *P*-values calculated by Mann Whitney U test with Holm-Šídák correction. ** = P < 0.01; *** = P < 0.001; **** = P < 0.0001, ns = not significant (*P* > 0.05).

6.2.3 dCas9-DNMT3A epigenetic editing in $CoCl_2$-treated TC28a2 chondrocytes

Targeted epigenetic editing of the *WWP2* DMR by dCas9-DNMT3A was repeated using $CoCl_2$-treated TC28a2 chondrocytes. As observed previously, the DMR was moderately hypomethylated, with DNAm levels ranging from 3.4% to 25.7% across the 16 CpGs (Fig. 6.3B, control data, in black). The same five gRNAs were used to target the *WWP2* DMR in both orientations of the genome (Fig. 6.3A). Cells nucleofected with gRNA3-dCas9-DNMT3A constructs were not viable. As a result, nucleic acids were unable to be harvested and data could not be generated for this condition. TC28a2 cells targeted using the remaining four gRNAs (gRNA1, gRNA2, gRNA4, gRNA5) were viable and therefore used in the downstream analysis.

An increase in DNAm (> 5%) at a minimum of three CpGs was observed for two of the four gRNAs (gRNA2 and gRNA5, Table 6.1), with gRNA5 increasing the levels of DNAm at 11 of

16 CpGs (Fig. 6.3B, in blue). The largest overall increase in DNAm was 27.9% at CpG7 by gRNA2 (Fig. 6.3B, in red). The mean percentage increase in DNAm levels across the DMR was 6.6% for gRNA2 and 9.6% for gRNA5. These observations match the trend observed in Chapter 5 (Fig. 5.2), but the overall percentage changes in DNAm were reduced. Epigenetic editing of the DMR using gRNA1 (Fig. 6.3B, in green) and gRNA4 (Fig. 6.3B, in pink) was unsuccessful, with the mean percentage increase in DNAm levels across the DMR ranging from 1.2% (gRNA1) to 1.8% (gRNA4).

The effect of dCas9-DNMT3A epigenetic editing upon the expression of the three *WWP2* isoforms (*WWP2*-FL, *WWP2*-N, *WWP2*-C) and the two mature strands of miR-140 (miR-140-5p, miR-140-3p) was measured by RT-qPCR (Fig. 6.3C). No significant change in the expression of *WWP2* or miR-140 was observed for any of the four gRNAs in comparison to a nontargeting control: *WWP2*-FL, *P* = 0.5176-0.8325; *WWP2*-N, *P* = 0.1188-0.4213; *WWP2*-C, *P* = 0.6269-0.9739; miR-140-5p, *P* = 0.4396-6467; miR-140-3p, *P* = 0.0626-0.6946.

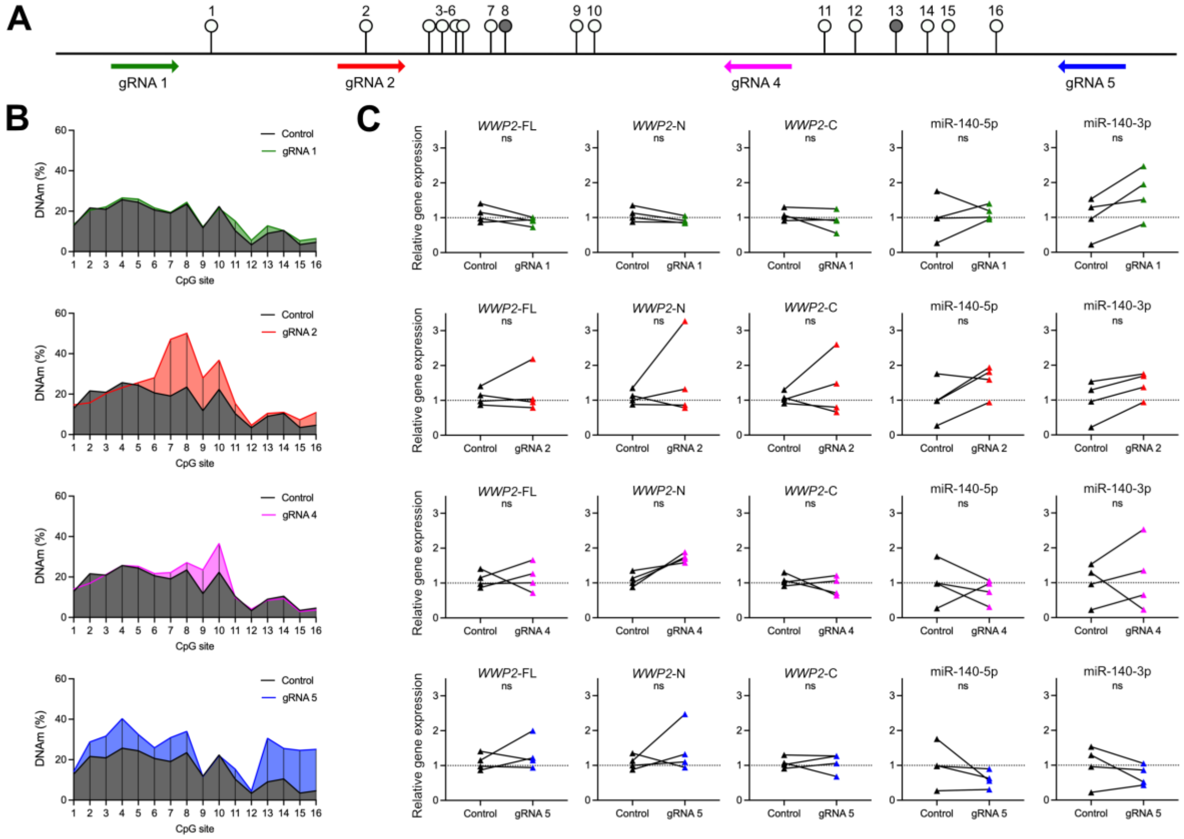


Figure 6.3: Epigenetic modulation of the DMR in $CoCl_2$-treated TC28a2 chondrocytes. A) Schematic representation of the 16 CpGs and of the four gRNAs (arrow pointing left, antisense strand; arrow pointing right, sense strand). cg26736200/CpG8 and cg26661922/CpG13 are

highlighted. **B)** Mean DNAm levels (%) of the 16 CpGs following expression of dCas9-DNMT3A protein in control (black), with nontargeting gRNA, or in samples with a targeting gRNA (coloured). Four biological replicates for control and for each targeting gRNA. **C)** Relative expression of the three *WWP2* transcripts (*WWP2*-FL, *WWP2*-N, *WWP2*-C) and of the two miR-140 strands (miR-140-5p, miR-140-3p) following epigenetic editing. Values were normalised to nontargeting gRNA control. *P*-values calculated using a paired *t*-test with Benjamini-Hochberg correction. ns = not significant (*P* > 0.05).

Table 6.1. Summary of the effects of dCas9-DNMT3A epigenetic editing in $CoCl_2$-treated TC28a2 cells upon DNAm levels at the DMR. Percentage change at each CpG site per gRNA in comparison to a nontargeting gRNA control.

gRNA	Number of CpGs edited (DNAm change >5%)	Mean percentage change in DNAm across DMR	Largest overall percentage change in DNAm
1	0	1.2%	CpG11 (4.7%)
2	7	6.6%	CpG7 (27.9%)
4	2	1.8%	CpG10 (14.2%)
5	11	9.6%	CpG13 (21.6%)

6.2.4 Effect of $CoCl_2$ on WWP2 expression

HIF-1α only binds the DMR in an unmethylated state. Therefore, targeted epigenetic editing using dCas9-DNMT3A to increase DNAm across the DMR should reduce binding efficiency of HIF-1α, attenuating its transcriptional effects. Targeted de-methylation of the DMR using dCas9-TET1 would facilitate HIF-1α, providing a better model to study the effects of HIF-1α transcriptional activity. However, the *WWP2* DMR is moderately hypomethylated. Overall changes in DNAm levels across the DMR would therefore be modest following dCas9-TET1 de-methylation and unlikely to lead to observable changes in gene expression. Instead, gene expression of $CoCl_2$-treated TC28a2 chondrocytes was compared to TC28a2 cells grown in normal cell culture medium to study the effects of HIF-1α stabilisation alone upon *WWP2* expression (Fig. 6.4). No significant change in expression was detected for any of the three *WWP2* isoforms: *WWP2*-FL (*P* = 0.1241); *WWP2*-N (*P* = 0.3275); *WWP2*-C (*P* = 0.5802).

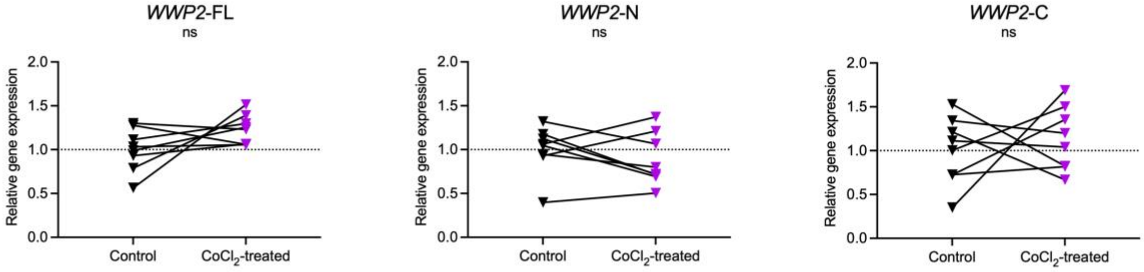


Figure 6.4: Effect of $CoCl_2$ treatment on *WWP2* expression in TC28a2 chondrocytes. Relative expression of the three *WWP2* isoforms (*WWP2*-FL, *WWP2*-N, *WWP2*-C) following treatment with $CoCl_2$ to stabilise HIF-1α protein. Seven biological replicates were performed. Values were normalised to TC28a2 cells cultured in normal cell culture medium (control). *P*-values calculated using paired *t*-test. ns = not significant ($P > 0.05$).

6.2.5 In-silico TF prediction following updates to JASPAR Core 2024

In March 2024, JASPAR released an updated version of its curated TF dataset[260]. In this update, HIF-1α was no longer predicted to bind the *WWP2* DMR. At the same genomic position, HIF-2α was predicted to bind CpG7 and CpG8/cg26736200 in the antisense direction (Fig. 6.5A). Six other TFs were also predicted to bind across the DMR, but not at any CpGs marked as OA cartilage mQTLs (Table 6.2). As HIF-1α and HIF-2α are predicted to bind the same DNA binding motif[285], the ReMap ChIP-Seq track[261] on UCSC Genome Browser was used to validate HIF-1α and HIF-2α binding predicted by JASPAR in 2022 and in 2024, respectively. Evidence of TF binding by HIF-1α and HIF-2α was identified in five ChIP-Seq datasets (Fig. 6.5B). In the cancer cell lines PC-3 and RCC10 and the epithelial cell line BEAS-2B, the ChIP-Seq peak summits for HIF-1α are within close proximity to CpG7 and CpG8/cg26736200. In the U2OS osteosarcoma cell line, which is more biologically relevant in the context of OA, HIF-1α bound the *WWP2* DMR downstream of the JASPAR predicted binding site. To determine the expression of the JASPAR-predicted TFs in OA cartilage, the same RNA-Seq dataset used in Chapters 3 and 5 was plotted (Fig. 6.5C). Of the seven TFs predicted to bind the DMR, five were abundantly expressed (transcripts per million, TPM > 10): *HIF2A*; *NR2C2*; *RXRA*; *CREB3L2*; and *THRA*, encoding Thyroid Hormone Receptor Alpha.

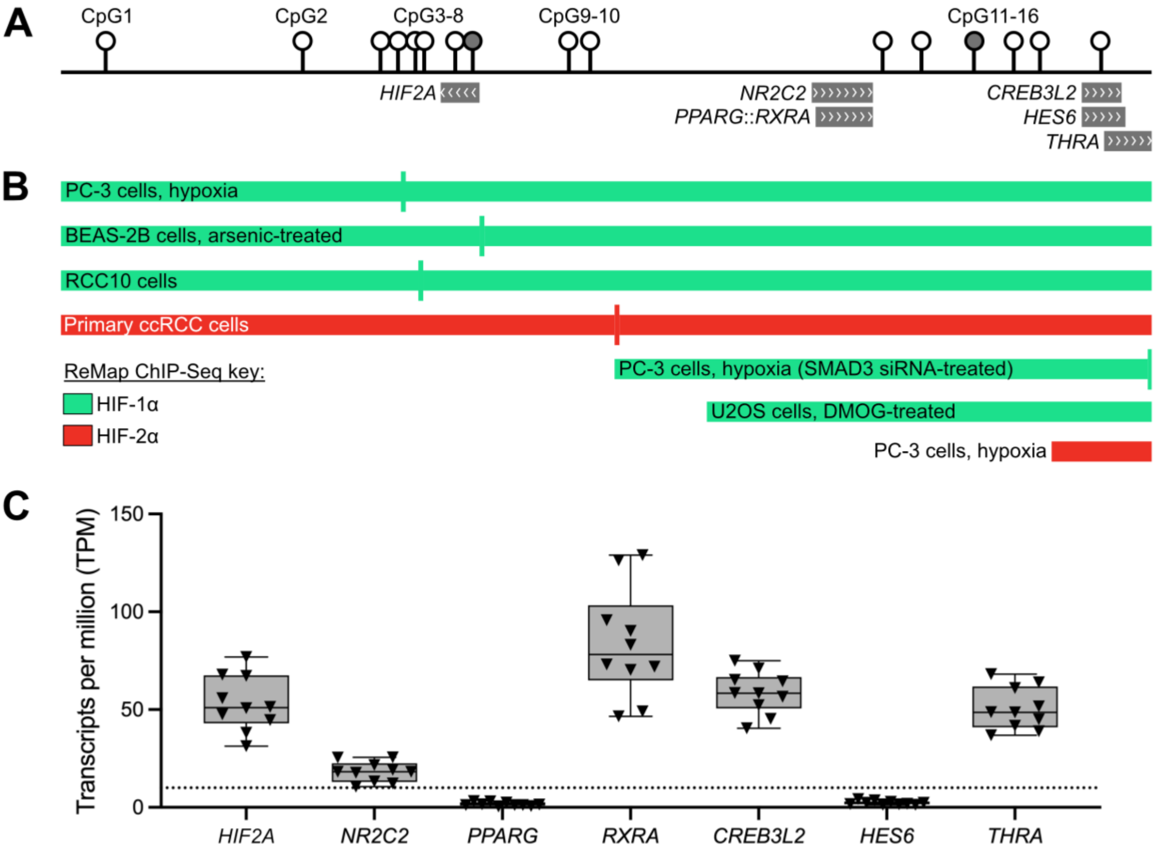


Figure 6.5: TFs predicted to bind at the DMR by JASPAR Core 2024. A) Schematic representation of the 16 CpGs and of seven TFs predicted to bind at or close to the CpGs. cg26736200 (CpG8) and cg26661922 (CpG13) are highlighted. The TFs are marked by grey rectangles with the direction of the arrows within the rectangles indicating the DNA strand the TF is predicted to bind to (arrows pointing to the left = antisense strand, arrows pointing to the right = sense strand). TF heterodimers are denoted by a double colon (::) between two TFs. **B)** ReMap ChIP-Seq data for HIF-1α and HIF-2α mapped to the DMR. Green tracks represent HIF-1α, red tracks represent HIF-2α DNA binding. Vertical bars represent ChIP-Seq peak summits. PC-3, human prostate cancer cell line; BEAS-2B, human bronchial epithelium cell line; RCC10, renal cell carcinoma cell line; primary ccRCC, primary clear cell renal carcinoma cells; U2OS, human osteosarcoma cell line. **C)** Expression levels (TPM, transcripts per million) of the TFs in cartilage chondrocytes from 10 OA patients. Triangles represent individual samples. Five TFs were abundantly expressed (median TPM > 10, represented by dotted line at y=10). Box plots show the median, 25th and 75th percentiles, and minimum and maximum values.

Table 6.2: List of TFs predicted to bind the DMR by JASPAR Core 2024. The table lists the gene symbol, the encoded TF name, the CpGs predicted to be bound by TFs, whether the CpG was a significant mQTL in OA cartilage, whether the TF was abundantly expressed in OA cartilage (median TPM > 10), and whether the prediction was novel to JASPAR 2024. N/A, not applicable.

Gene symbol	TF name	CpGs bound	OA cartilage mQTL?	Expression > 10 TPM?	Novel?
HIF2A	Hypoxia Inducible Factor 2 Subunit Alpha	7 and 8	Yes	Yes	Yes
NR2C2	Nuclear Receptor Subfamily 2 Group C Member 2	None	N/A	Yes	No
PPARG	Peroxisome Proliferator Activated Receptor Gamma	None	N/A	No	No
RXRA	Retinoid X Receptor Alpha	None	N/A	Yes	No
CREB3L2	CAMP Responsive Element Binding Protein 3 Like 2	16	No	Yes	No
HES6	Hes Family BHLH Transcription Factor 6	16	No	No	No
THRA	Thyroid Hormone Receptor Alpha	16	No	Yes	Yes

6.3 Discussion

In this chapter, HIF-1α was shown to bind the *WWP2* DMR in a methylation-dependent manner. These findings corroborate previous studies that demonstrate methylation of the CpG site within the HIF-1α binding motif prevents its interaction with the DNA[285,309]. Because the functional studies performed in Chapter 5 were conducted in conditions where HIF-1α would be rapidly degraded, the Lucia reporter assay and dCas9-DNMT3A epigenetic editing experiments were repeated using the hypoxia mimetic $CoCl_2$ to stabilise HIF-1α protein and allow its translocation to the nucleus for transcriptional activity. When cloned as an enhancer, the presence of HIF-1α reversed the repressive effect of the unmethylated DMR observed in Chapter 5 (Fig. 5.1B), enhancing Lucia expression. *HIF1A* is abundantly expressed in cartilage, a tissue that is naturally hypoxic and therefore facilitates transcriptional activity of HIF-1α protein[21]. The DMR may therefore act as an enhancer of *WWP2* when bound by HIF-1α, with this effect attenuated by higher DNAm levels exerted by the OA risk-conferring G allele of rs34195470. However, this scenario contradicts the previous findings that increased DNAm across the DMR drives increased *WWP2* expression in chondrocytes (Chapter 5, Fig. 5.2) and the observations of the mQTL and AEI studies using OA patient cartilage (Chapter 4, Fig. 4.1B and Fig. 4.6A)[87,214]. Due to the long-range chromatin interaction identified in Chapter 3 (Fig. 3.3), concatenation of the DMR to the *WWP2*-FL/*WWP2*-N promoter (effectively replacing the hEF1 promoter of the Lucia reporter vector) could provide insight into specific enhancer-promoter interactions and may address the discordancy between the findings of the reporter assays and the dCas9-DNMT3a epigenetic editing experiments performed in Chapters 5 and 6.

Reporter assays isolate a piece of DNA from their normal genomic and chromatin context. Epigenetic editing of the DMR was therefore undertaken, again in the presence of HIF-1α, but overall changes in DNAm levels were reduced in comparison to the original experiment. Speculatively, this could be due to HIF-1α binding the DMR, blocking access for dCas9-DNMT3A and preventing epigenetic editing. Treatment of cells with $CoCl_2$ can also lead to changes in histone modifications and, as observed in true hypoxia, may induce changes to chromatin accessibility[310–312]. It is plausible that the DMR is less accessible following $CoCl_2$ treatment in TC28a2 cells, reducing the capacity of dCas9-DNMT3A to produce large changes in DNAm levels. Future experiments should also investigate the effect of $CoCl_2$ on steady-state DNAm levels by comparing TC28a2 cells cultured in $CoCl_2$-supplemented culture medium to those cultured without $CoCl_2$ to ensure that this hypoxia mimetic does not affect DNAm levels.

As stated previously, HIF-1α only binds the DMR in an unmethylated state. Therefore, targeted epigenetic editing using dCas9-DNMT3A to increase DNAm across the DMR should reduce binding efficiency of HIF-1α, attenuating its transcriptional effects. The complementary targeted de-methylation tool dCas9-TET1 could not be performed due to TC28a2 chondrocytes being moderately hypomethylated. The effect of $CoCl_2$ treatment alone on *WWP2* expression was therefore measured, resulting in a slight increase in *WWP2*-FL expression, though this was not significant. Editing the DNAm levels in other cell lines or primary chondrocytes, with differing patterns of DNAm across the DMR that facilitates dCas9-TET1 epigenetic editing, may provide further insight into the interaction between HIF-1α, DNAm levels at the DMR and *WWP2* expression.

Three biological replicates for the dCas9-DNMT3A epigenetic editing experiment were not viable post-nucleofection, resulting in a smaller sample size (n=4) than in the original experiment (n=7). No biological replicates were viable for gRNA3, resulting in exclusion of this condition from the analysis. Whilst nucleofection can reduce cell viability[313], the prolonged exposure of TC28a2 chondrocytes to $CoCl_2$ may be responsible. $CoCl_2$ generates reactive oxygen species (ROS) that can produce oxidative stress and decrease cell proliferation[314]. At higher concentrations of 200μM and above, $CoCl_2$ has been shown to induce cell death due to oxidative DNA damage by ROS combined with inhibition of DNA repair mechanisms[297,315,316]. In addition, $CoCl_2$ has been shown to activate necrotic and apoptotic pathways, decreasing viability in multiple cell lines[317,318]. Although $CoCl_2$ mimics the transcriptional response observed in true hypoxia, differences remain, including molecular signalling pathways that are O_2-dependent[297]. The use of a hypoxic chamber may therefore provide better experimental design to study the true nature of HIF-1α transcriptional regulation.

The recent update to JASPAR predicts that HIF-2α binds the *WWP2* DMR, not HIF-1α. The shared DNA binding motif of these TFs and ChIP-Seq data suggests both can bind the DMR in other biological contexts. Like HIF-1α, methylation of the CpG site within the HIF binding site inhibits DNA binding of HIF-2α via steric hindrance[287]. However, HIF-2α appears to preferentially bind motifs within enhancers whilst HIF-1α preferentially binds motifs within promoter sequences, suggesting other factors govern their binding affinity to specific regions of the genome[287]. $CoCl_2$ can stabilise HIF-2α as well as HIF-1α[297]. It is therefore possible that both proteins were stabilised in the experiments presented in this chapter. Replication of the EMSAs to determine whether HIF-2α binds the DMR and further functional studies are therefore required, with the role of HIF-1α and HIF-2α in *WWP2* transcriptional regulation studied independently of one another and in combination. If HIF-1α or HIF-2α enhance transcriptional activity of the cloned DMR, this could be modelled *in-vitro* via co-transfection of the TFs with the Lucia reporter constructs to observe methylation-sensitive TF-mediated regulatory activity[319].

One of the key questions unanswered by this chapter is how *WWP2*-FL and *WWP2*-N are upregulated in response to increased DNAm at the DMR mediated by dCas9-DNMT3A in the absence of $CoCl_2$. In the original experiment (Chapter 5, Fig. 5.2), HIF-1α and HIF-2α protein would be rapidly degraded, implying these TFs would be unavailable to regulate *WWP2* expression. Identification of other TFs binding the *WWP2* DMR in an unmethylated and methylated state is therefore essential for pinpointing this mechanism. This could be achieved using dCas9-APEX to label proteins (including TFs) proximal to the DMR, facilitating targeted profiling of the DMR protein microenvironment[320].

In this chapter, HIF-1α was demonstrated to bind the *WWP2* DMR in a methylation-sensitive manner. Stabilisation of HIF-1α and HIF-2α using $CoCl_2$ mediated transcriptional changes to the unmethylated DMR when cloned as an enhancer. However, no functional role was established for these TFs in the regulation of *WWP2* or miR-140. Despite this, HIFs are known to play important roles in cartilage development and homeostasis[21]. The OA risk-conferring G allele of rs34195470 correlates with increased *WWP2* expression[214]. Tuerlings and colleagues recapitulated this effect via lentiviral transduction of the *WWP2*-FL isoform in primary chondrocytes, observing an increase in *HIF2A* expression[243]. *WWP2* is therefore upstream of *HIF2A*. HIF-2α is a key regulator of ECM catabolism and chondrocyte terminal differentiation[321,322], suggesting this may be one mechanism through which OA genetic risk is conferred by rs34195470. However, as transcriptional regulators of *WWP2*, the role of HIF-1α and HIF-2α is unclear. Future work that accounts for the limitations in the study design, particularly the use of $CoCl_2$ to mimic hypoxia, may provide better insight.

Chapter 7: Determination of the causal variant marked by rs34195470

7.1 Introduction

In the previous chapters, the OA risk-conferring G allele of rs34195470 was identified to correlate with increased *WWP2* expression and increased DNAm levels at 14 CpGs spanning a 228bp regulatory element, marking a DMR in OA cartilage. Functional studies in chondrocytes demonstrated that higher DNAm levels at the DMR results in increased expression of *WWP2*-FL and *WWP2*-N, identifying these transcripts as targets of OA risk marked by rs34195470. The DMR is therefore a functional intermediary between this OA association signal and its regulatory targets, *WWP2*-FL and *WWP2*-N.

rs34195470 was first reported in a 2018 meta-analysis of Icelandic and UK datasets, associating with knee OA as the most significant variant in its LD class[123]. This finding was replicated in the largest OA GWAS to date, with conditional analyses identifying rs34195470 as the lead variant[89]. However, reported variants are not always causal. Arrays typically employ a single index or 'tag' SNV to represent large LD blocks, most often acting as proxies for the true causal variant[88]. To overcome this, post-GWAS analyses such as conditional analyses and statistical fine-mapping are applied to focus in on likely causal variants[89,118,119]. Nevertheless, functional laboratory studies are essential to validate these findings. In addition, researchers must also consider the tissue-specific effects of causal variants[88]. For example, a risk variant may be beneficial or benign in one tissue, but produce adverse effects in another, contributing to disease risk. Integration of GWAS data with chromatin state data and identification of eQTLs are valuable *in-silico* analyses that can be used to inform downstream studies[87].

rs34195470 resides within an LD block ($r^2 > 0.8$) with two other variants in European populations: the SNV rs9746247 ($r^2 = 0.92$, $D' = 1.00$), and the indel rs111837947 ($r^2 = 0.85$, $D' = 0.99$). All three variants reside within intron 11 of *WWP2*-FL. Whilst rs9746247 has never been reported as an association signal in any OA GWAS, rs111837947 was recently reported as the most significant variant in its LD class to associate with surgical and nonsurgical (radiographic) knee OA in a meta-analysis of Danish, Icelandic and UK datasets[122]. The risk allele of rs111837947 is a 2-base insertion (insGT), which correlates with the risk allele G of rs34195470. In all three GWASs, *WWP2* was highlighted as the putative effector gene[89,122,123].

Table 7.1: Summary information for the lead OA risk variant rs34195470 and the variants in high LD. N/A, not applicable; ins, insertion.

rsID	Co-ordinates (hg19)	Distance to rs34195470 (bp)	LD	Correlating alleles with rs34195470 (A > G)	Reported in OA GWAS?
rs34195470	Chr16:69,955,690	N/A	N/A	N/A	Yes[89,123]
rs9746247	Chr16:69,955,960	270bp	r^2 = 0.92 D' = 1.00	A = C (non-risk) G = G (risk)	No
rs111837947	Chr16:69,957,231	1,541bp	r^2 = 0.85 D' = 0.99	A = - (non-risk) G = insGT (risk)	Yes[122]

Using *in-silico* analyses in combination with a reporter gene assay and the CRISPR/Cas9 toolbox in TC28a2 chondrocytes, this chapter aims to:

1) Identify whether rs34195470, rs9746247 and rs111837947 are within active regulatory elements in relevant cell types using chromatin state data.
2) Identify allelic differences in regulatory function for rs34195470, rs9746247 and rs111837947.
3) Determine whether deletion of rs34195470, rs9746247 or rs111837947 affects *WWP2* expression.
4) Determine whether deletion of rs34195470, rs9746247 or rs111837947 affects DNAm levels at the previously identified cartilage DMR.
5) Identify TFs predicted to bind across rs34195470, rs9746247 and rs111837947 and quantify their abundance in OA cartilage.
6) Determine whether TFs predicted to bind across rs34195470, rs9746247 and rs111837947 exhibit preferential binding affinity to a particular allele.

7.2 Results

7.2.1 In-silico analysis of the region

rs34195470, rs9746247 and rs111837947 reside within intron 11 of *WWP2*-FL (Fig. 7.1A, in black). The previously identified OA cartilage DMR (Fig. 7.1A, in red) is approximately 3.7Kb upstream of rs34195470 and the TSS of *WWP2*-C is approximately 1.7Kb downstream of rs111837947.

To determine chromatin accessibility at the region encompassing the three variants, ATAC-Seq data was mapped to the region using the UCSC Genome Browser (hg19). The three variants were not identified to reside within ATAC-Seq peaks in any of the cell types investigated, suggesting that they reside within closed, inaccessible chromatin in these tissues (Fig. 7.1B). In contrast and as previously described (Chapter 3, Fig. 3.2), the DMR overlaps with ATAC-Seq peaks in foetal and OA (adult chondrocytes), in the osteosarcoma cell line Saos-2, and in OA fibroblast-like synoviocytes (FLS). To garner evidence of regulatory function, Roadmap regulatory elements were also mapped to the region using the UCSC Genome Browser (hg19). In MSCs, the three variants reside within a region of weak transcription, whilst in chondrocytes and osteoblasts they reside within an enhancer (Fig. 7.1C).

Finally, to determine whether rs34195470, rs9746247 or rs111837947 reside within regions known to bind TFs, the ReMap TF ChIP-Seq track[261] on UCSC Genome Browser was used to create a TF density plot at the region of interest using ChIP-Seq data generated in relevant cell types, including chondrocytes (Fig. 7.1D). The DMR and the TSS of *WWP2*-C exhibited the highest TF density peaks, indicating these regions are known to bind TFs in relevant cell types. No TF binding has been reported in relevant cell types at rs34195470, rs9746247 and rs111837947.

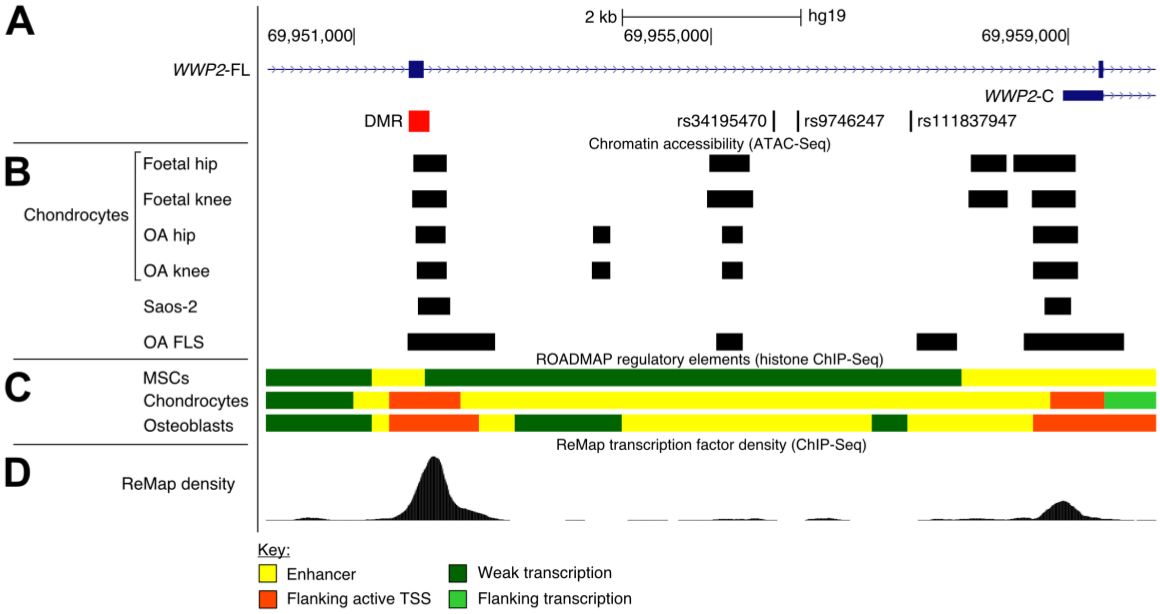


Figure 7.1: In-silico analysis of the region encompassing the three variants associated with OA genetic risk marked by rs34195470. A) Schematic representation of the *WWP2*-FL and *WWP2*-C transcript isoforms (in blue) in proximity to the *WWP2* DMR (in red). Introns are represented by horizontal lines with arrows pointing right indicating the encoded orientation (sense), half-height bars representing the 5' UTR of *WWP2*-C, and full-height bars representing the exons of *WWP2*-FL. The OA GWAS lead signal rs34195470 and the two variants in high LD (rs9746247 and rs111837947) are marked in black. **B)** ATAC-Seq peaks (denoted by black

boxes) in foetal and OA chondrocytes, the osteosarcoma cell line Saos-2, and OA fibroblast-like synoviocytes (FLS). **C)** ROADMAP regulatory elements in MSCs (E006), chondrocytes (E049) and osteoblasts (E129). A key denoting each chromatin state is found at the bottom of the figure. **(D)** ReMap TF density plot in relevant cell types. The peak height is equivalent to the number of TFs known to bind regions; i.e. higher peaks denote regions known to bind larger numbers of TFs.

7.2.2 Luciferase reporter gene assay

In-silico analyses suggest rs34195470, rs9746247 and rs111837947 reside within an enhancer in chondrocytes and osteoblasts. Differential allelic expression of *WWP2* has been observed in OA cartilage using AEI analysis (Chapter 4, Fig. 4.6A), but not in OA bone. Functional experiments were therefore prioritised in chondrocytes.

Each variant was cloned individually into Luciferase reporter vectors as an enhancer to assess the regulatory effects of the risk allele and the non-risk allele (Fig. 7.2). These constructs were then transfected into TC28a2 chondrocytes. Allelic differences in regulatory function were detected for rs34195470, with the non-risk allele (A) conferring greater repressive effects upon Luciferase activity than the OA risk-conferring G allele (P = 0.0216). The relative greater expression of allele G versus allele A matches the OA cartilage AEI data (Chapter 4, Fig. 4.6A). No allelic differences upon regulatory function were observed for rs9746247 (P = 0.6147) or rs111837947 (P = 0.7944). These data indicate that rs34195470 is the only variant capable of modulating regulatory activity in an allele-dependent manner.

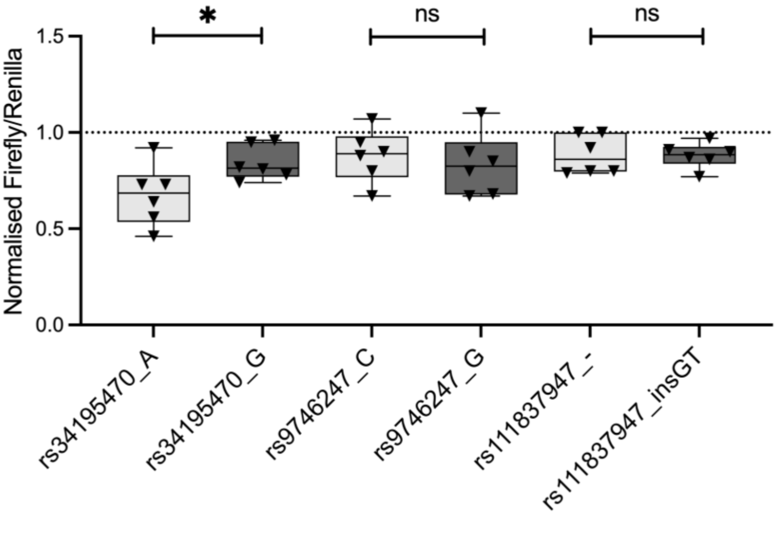


Figure 7.2: Luciferase reporter assay in TC28a2 chondrocytes. Luciferase reporter assays assessing enhancer activity in the presence of rs34195470, rs9746247 or rs111837947. Values were normalised to those in empty vector control (dotted horizontal line). Black triangles

represent individual samples (six biological replicates per group). Box plots show the median, 25th and 75th percentiles, and minimum and maximum values. Allelic differences in regulatory activity were measured between the risk allele (light grey boxes) and the non-risk allele (dark grey boxes). *P*-values calculated by Mann Whitney U test. * = $P < 0.05$, ns = not significant ($P > 0.05$).

7.2.3 CRISPR/Cas9 deletion of the genomic regions housing rs34195470, rs9746247 and rs111837947

To determine whether deletion of the variants and their surrounding DNA sequence could affect *WWP2* expression, gRNA primers targeting up- and downstream of the variants were designed and cloned into CRISPR/Cas9 expression vectors. Unique gRNAs targeting between rs34195470 and rs9746247 were unable to be designed due to high genomic complexity marked by nucleotide repeats frequently occurring across the region. rs34195470 and rs9746247 were therefore targeted for deletion together. Paired gRNA targeting the region encompassing these two variants produced an incomplete deletion (Fig. 7.3A, left panel). The third biological replicate was less successful, marked by an intense PCR product for the unedited band of 594bp compared to the deleted band of 118bp (Fig. 7.3A, left panel, DEL3). Paired gRNA targeting rs111837947 produced a near complete deletion for all three biological replicates, indicated by the presence of intense PCR products of the deleted band of 287bp compared to the unedited band of 353bp (Fig. 7.3A, right panel). The 16 CpGs constituting the DMR, located approximately 3.7Kb upstream of rs34195470, were not targeted by CRISPR/Cas9 deletion.

The effect of the deletions upon the expression of the three common isoforms of *WWP2* (*WWP2*-FL, *WWP2*-N and *WWP2*-C) was then measured by RT-qPCR (Fig. 7.3B). The expression of the two miR-140 strands (miR-140-5p and miR-140-3p) was not measured as these were not identified as targets of OA genetic risk marked by rs34195470 in Chapter 5 (Fig. 5.2). Deletion of rs34195470 and rs9746247 resulted in no significant changes in *WWP2* expression compared to nontargeting control (Fig. 7.3B, in blue): *WWP2*-FL ($P = 0.0533$), *WWP2*-N ($P = 0.8291$), or *WWP2*-C ($P = 0.0518$). This may reflect the incomplete deletion observed for this experiment or no effect of a partial deletion of the region harbouring the variants on *WWP2* expression. Deletion of rs111837947 resulted in decreased expression of *WWP2*-FL in comparison to nontargeting control ($P = 0.0324$, Fig. 7.3B, in orange). No significant changes in expression were observed for *WWP2*-N ($P = 0.0522$) or *WWP2*-C ($P = 0.0845$). The deletion experiments targeted intron 11 of *WWP2*-FL. To assess whether splicing was affected by the deletions, cDNA from all conditions were PCR-amplified with primers spanning exons 10-13 of *WWP2*-FL (Fig. 7.4). No differences were observed in comparison to cDNA from nontargeting control, confirming that splicing was unaffected.

The OA risk-conferring G allele of rs34195470 is associated with increased DNAm at the DMR and increased *WWP2* expression in OA cartilage (Chapter 4, Fig. 4.1B and Fig. 4.6A). Epigenetic editing confirmed that higher DNAm levels at the DMR result in increased *WWP2*-FL and *WWP2*-N expression (Chapter 5, Fig. 5.2). These data imply that increased DNAm at the DMR is mediated by rs34915470, rs9746247 or rs111837947, where their respective risk alleles contribute to methylation of the DMR, which in turn acts as a functional intermediary to modulate expression of *WWP2*-FL and *WWP2*-N. The effect of the deletions upon DNAm levels across the DMR were therefore measured (Fig. 7.3C). The mean percentage change in DNAm levels at the DMR was small (< 1%). The largest change in DNAm levels was a 2.8% decrease at CpG5 following deletion of rs34195470 and rs9746247 (Table 7.2). In this deletion experiment, the changes to *WWP2*-FL expression following deletion of rs111837947 are therefore mediated independently of DNAm levels at the DMR.

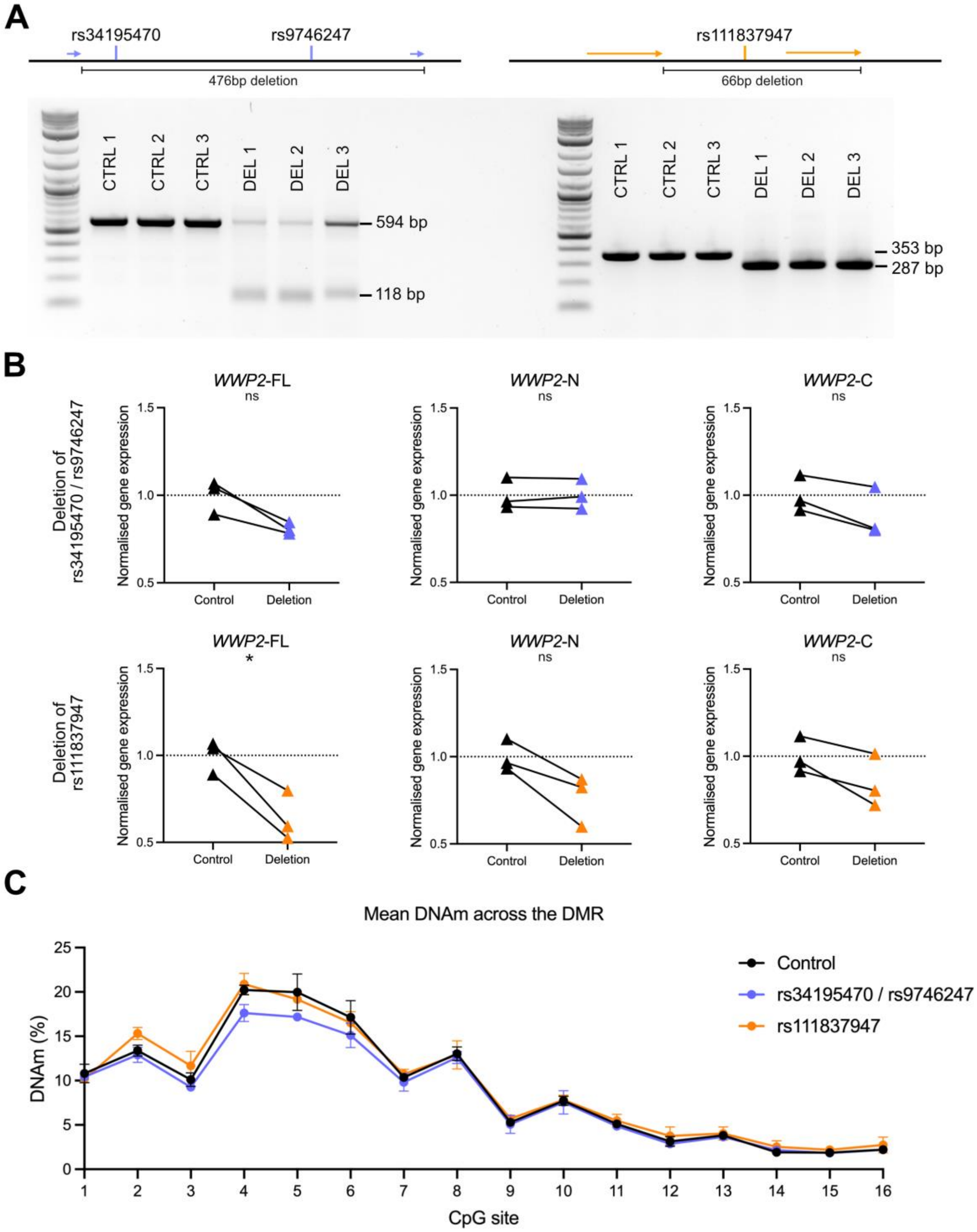


Figure 7.3: CRISPR/Cas9 deletion of the three variants associated with OA genetic risk marked by rs34195470. A) Schematic denoting genomic location of the paired gRNAs used to delete rs34195470 and rs9746247 (left, in blue) and rs111837947 (right, in orange). Arrows indicate the strand targeted by gRNAs (arrows pointing right = sense strand). The agarose-TBE gel below the schematic visualises the efficiency of CRISPR/Cas9 targeting (DEL) in comparison to a nontargeting control (CTRL). Three biological replicates were performed. A 1Kb+ ladder was run alongside samples. **B)** Normalised expression of the three *WWP2* transcripts (*WWP2*-FL, *WWP2*-N, *WWP2*-C) following CRISPR/Cas9 deletions. Values were normalised to a

nontargeting gRNA control. *P*-values calculated using paired *t*-test. * = $P < 0.05$, ns = not significant ($P > 0.05$). **C)** Mean DNAm levels (%) of the 16 CpGs of the DMR following expression of CRISPR/Cas9 protein in the nontargeting control (black), or in samples targeting rs34195470 and rs9746247 (blue) or rs111837947 (orange). Error bars represent the standard deviation. No error bars = range of standard deviation is less than the size of the point.

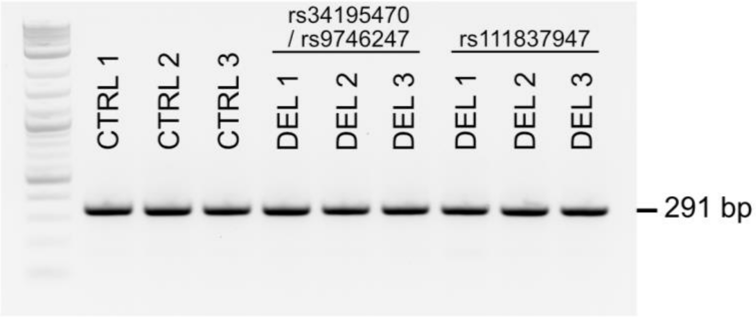


Figure 7.4: *WWP2* splicing unaffected by deletion of rs34195470, rs9746247 and rs111837947. Agarose-TBE gel visualising PCR products from amplification of cDNA derived from samples targeted by CRISPR/Cas9 (DEL) or by a nontargeting control (CTRL). The expected band size of 291bp indicated that splicing of *WWP2*-FL was consistent across all samples. A 1Kb+ ladder was run alongside samples.

Table 7.2. Summary of the effects of CRISPR/Cas9 deletions upon DNAm levels at the DMR. Percentage change at each CpG site per gRNA in comparison to a nontargeting gRNA control.

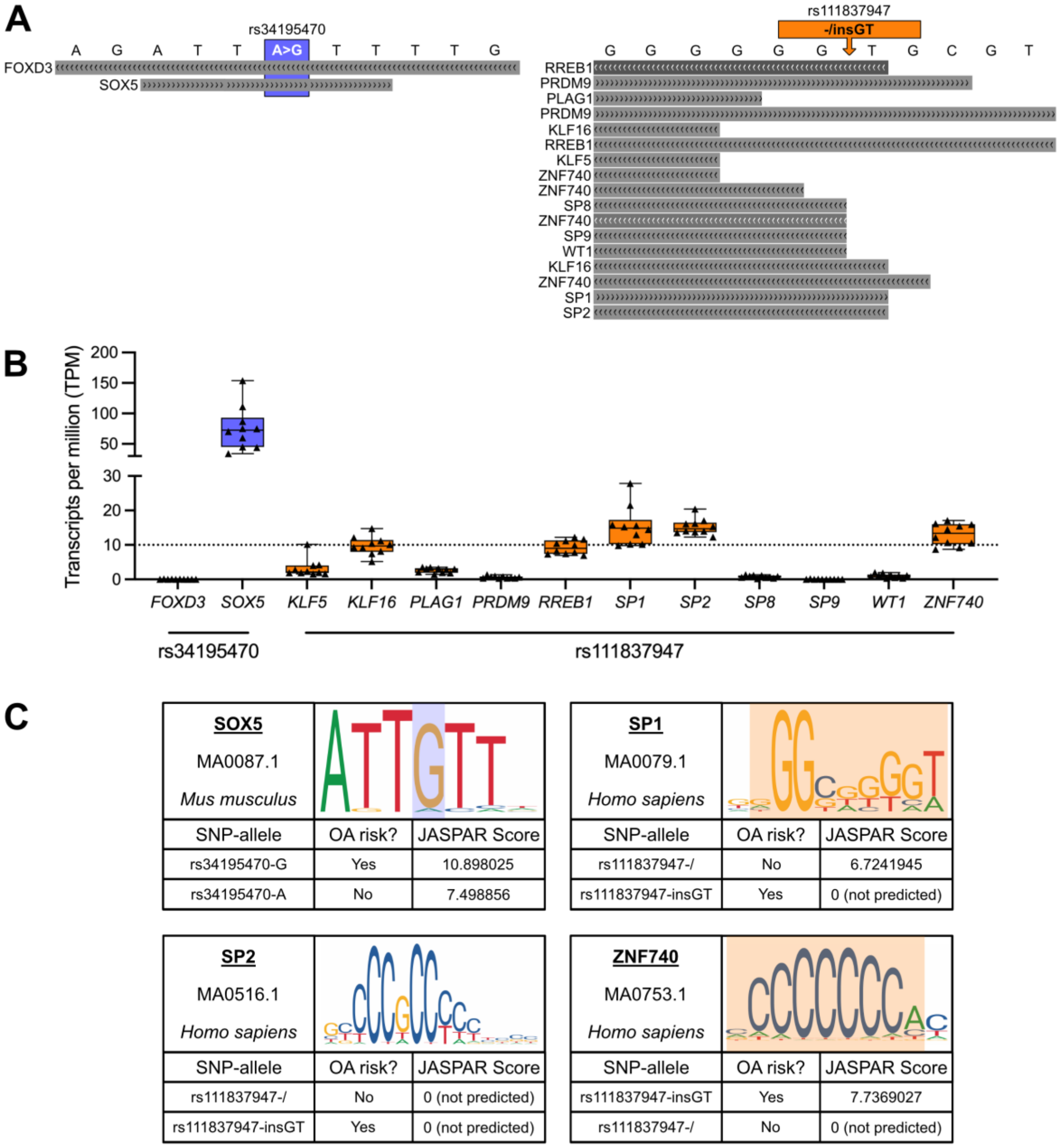
Target variant	Number of CpGs where DNAm levels changed (>5%)	Mean percentage change in DNAm across DMR	Largest overall percentage change in DNAm
rs34195470, rs9746247	0	0.7%	CpG5 (2.8%)
rs111873947	0	0.3%	CpG2 (1.9%)

7.2.4 In-silico TF binding prediction at the three regions housing variants

To determine TFs predicted to bind rs34195470, rs9746247 and rs111837947, the JASPAR Core 2024 dataset[260] was mapped to the region using the UCSC Genome Browser (Fig. 7.5A). As the UCSC Genome Browser sequence only includes the reference allele of each variant, the Expasy SNP2TFBS tool[323] was also used to detect TFs that bound the alternative allele. Two TFs were predicted to bind rs34195470: Forkhead Box D3 (FOXD3) and SRY-Box Transcription Factor 5 (SOX5) (Fig. 7.5A, left panel and Table 7.3). Eleven TFs were predicted to bind rs111837947 (Fig. 7.5A, right panel and Table 7.3). No TFs were predicted to bind rs9746247.

To determine whether these TFs were abundantly expressed in OA cartilage, RNA-Seq data was plotted (Fig. 7.5B). Of the 13 TFs predicted to bind rs34195470 or rs111837947, four were abundantly expressed in cartilage (transcripts per million, TPM > 10): *SOX5*; *SP1*, encoding Sp1 Transcription Factor; *SP2*, encoding Sp2 Transcription Factor; and *ZNF740*, encoding Zinc Finger Protein 740. *SOX5* was the most abundantly expressed TF in OA cartilage (median TPM = 72.6).

The JASPAR motif scan tool[260] was utilised to predict differential TF binding affinity to each allele of rs34195470 and rs111837947 (±5bp) for the four abundantly expressed TFs (Fig. 7.5C). All DNA binding motifs were validated in humans except for SOX5 where the DNA binding motif was validated in mouse. SOX5 is predicted to preferentially bind the risk allele (G) of rs34195470. SP1 was predicted to exclusively bind the non-risk allele (-) of rs111837947, whilst the risk allele (insGT) of rs111837947 was predicted to be exclusively bound by ZNF740. SP2 was not predicted to bind either the risk or non-risk alleles of rs111837947 using the JASPAR motif scan tool.



SOX5 MA0087.1 *Mus musculus*		
SNP-allele	OA risk?	JASPAR Score
rs34195470-G	Yes	10.898025
rs34195470-A	No	7.498856

SP1 MA0079.1 *Homo sapiens*		
SNP-allele	OA risk?	JASPAR Score
rs111837947-/	No	6.7241945
rs111837947-insGT	Yes	0 (not predicted)

SP2 MA0516.1 *Homo sapiens*		
SNP-allele	OA risk?	JASPAR Score
rs111837947-/	No	0 (not predicted)
rs111837947-insGT	Yes	0 (not predicted)

ZNF740 MA0753.1 *Homo sapiens*		
SNP-allele	OA risk?	JASPAR Score
rs111837947-insGT	Yes	7.7369027
rs111837947-/	No	0 (not predicted)

Figure 7.5: TFs predicted to bind rs34195470, rs9746247 and rs111837947. (A) Schematic representation of the DNA sequence surrounding rs34195470 (left, in blue), rs111837947 (right, in orange) and the TFs predicted to bind each variant. The TFs are marked by grey rectangles with the direction of the arrows within the rectangles indicating the DNA strand the TF is predicted to bind to (arrows pointing left = antisense strand, arrows pointing right = sense strand). **(B)** Expression levels (TPM, transcripts per million) of the TFs in cartilage chondrocytes from 10 OA patients. Triangles represent individual samples. Four TFs were abundantly expressed (median TPM > 10, represented by dotted line at y=10). Box plots show the median, 25th and 75th percentiles, and minimum and maximum values. **(C)** Results of JASPAR motif scan analysis. Boxes show TF name, JASPAR ID and the organism in which the TF motif was validated (*Mus musculus* or *Homo sapiens*). The motif is represented as a position weight matrix (PWM) and the variant location within each motif is represented by the blue shaded box (rs34195470) or the orange shaded box (rs111837947). No shaded box was

included for SP2 as the JASPAR motif scan tool was unable to predict TF binding for the input DNA sequence. The JASPAR score for each allele is shown underneath the motif, ranked by highest score. Higher scores indicate preferential binding.

Table 7.3: List of TFs predicted to bind rs34195470 and rs111837947. The table lists the gene symbol, the encoded TF name, the variant predicted to be bound by the TF, the database reporting the predicted TF, and whether the TF was abundantly expressed in OA cartilage (median TPM > 10).

Gene symbol	TF name	Variant bound	Predicted TF database	Expression > 10TPM?
FOXD3	Forkhead Box D3	rs34195470	JASPAR Core 2024	No
SOX5	SRY-Box Transcription Factor 5	rs34195470	Expasy SNP2TFBS	Yes
KLF5	KLF Transcription Factor 5	rs111837947	JASPAR Core 2024	No
KLF16	KLF Transcription Factor 16	rs111837947	JASPAR Core 2024	No
PLAG1	PLAG1 Zinc Finger	rs111837947	JASPAR Core 2024	No
PDRM9	PR/SET Domain 9	rs111837947	JASPAR Core 2024	No
RREB1	Ras Responsive Element Binding Protein 1	rs111837947	JASPAR Core 2024	No
SP1	Sp1 Transcription Factor	rs111837947	Expasy SNP2TFBS	Yes
SP2	Sp2 Transcription Factor	rs111837947	Expasy SNP2TFBS	Yes
SP8	Sp8 Transcription Factor	rs111837947	JASPAR Core 2024	No
SP9	Sp9 Transcription Factor	rs111837947	JASPAR Core 2024	No
WT1	WT1 Transcription Factor	rs111837947	JASPAR Core 2024	No
ZNF740	Zinc Finger Protein 740	rs111837947	JASPAR Core 2024	Yes

7.3 Discussion

Progressing from a GWAS-reported association signal to a causal variant is impeded by several factors: (1) common, complex trait-associated variants tend to reside within noncoding regions of the genome, complicating identification of their gene target(s); (2) most variants are within LD blocks, where the index variant often acts as a proxy for the true causal variant; (3) variants may exert tissue-specific effects, resulting in gene expression changes in certain tissues but not others[88]. The latter is particularly complex in OA where multiple tissues of the articular joint are affected. To overcome this, functional studies should utilise *in-silico* datasets to prioritise tissues and cell types that variants are likely to be functional within[87]. Using ATAC-Seq data, the three investigated variants (rs34195470, rs9746247 and rs111837947) were identified to reside within closed, inaccessible chromatin in OA relevant cell types, suggesting they are unable to bind TFs. This finding was supported by a lack of TF binding identified at the region using the ReMap ChIP-Seq dataset[261]. In OA-relevant cell types, TF binding was abundant at the DMR, supporting previous findings that this region acts as a transcriptional regulator in chondrocytes (Chapter 3, Fig. 3.2 and Chapter 5, Fig. 5.1 and Fig. 5.2).

Despite rs34195470, rs9746247 and rs111837947 residing within a putative enhancer in chondrocytes, all three variants exhibited reduced transcriptional activity in comparison to an empty vector control in the Luciferase reporter gene assay. No allelic differences in regulatory activity were observed for rs9746247 or rs111837947, suggesting they do not mediate expression changes exerted by the OA risk signal. The non-risk allele (A) of rs34195470 significantly repressed transcriptional activity in comparison to the risk allele (G), suggesting rs34195470 is capable of mediating changes in gene expression exerted by the OA risk signal. However, deletion of rs34195470 did not result in changes to *WWP2* expression. The targeted deletion of rs34195470 (in combination with rs9746247) was inefficient, with a proportion of cells in these samples not edited by CRISPR/Cas9. The incomplete deletion may therefore have attenuated any observable changes in *WWP2* expression. Both *WWP2*-FL and *WWP2*-C were downregulated, but this effect was not significant. The effect size of OA risk variants is small, thus leading to modest changes in gene expression[87]. Increasing the small sample size (n=3) may provide further insight into whether deletion of rs34195470 is capable of modulating *WWP2* expression *in-vitro*.

Targeted deletion of rs111837947 significantly downregulated *WWP2*-FL expression, indicating the genomic region housing rs111837947 is involved in the regulation of one of the two transcript isoforms identified as targets of OA genetic risk in Chapter 5 (Fig. 5.2). Whilst not significant, both *WWP2*-N and *WWP2*-C were also downregulated in comparison to the nontargeting control. Again, the modest sample size may have obscured detections of true changes in gene expression. However, no allelic differences in transcriptional activity for rs111837947 were detected by the Luciferase reporter assay. The region encompassing rs111837947 may therefore act as an important regulatory element required for increasing *WWP2*-FL expression regardless of whether the risk or non-risk allele is present. In this context, the utility of CRISPR/Cas9 deletion is limited by removing the DNA sequence surrounding the variant of interest with potentially wider implications upon chromatin state and disruption to other important TF binding sites. To better investigate the functional role of each variant upon *WWP2* expression, base editing should be used to specifically modify nucleotides of interest, limiting effects upon any nearby regulatory elements[324]. Furthermore, during DNA double-strand break repair post-CRISPR/Cas9 editing of the three variants, it is plausible that other altered sequences have been introduced (such as indel formation) that may impact upon the regulatory activity of these regions independently of each risk variant[325].

The use of endpoint PCR to confirm successful CRISPR/Cas9 targeting of the region provides no insight into the possibility of altered DNA sequences. Future work should utilise technologies such as Sanger sequencing to better characterise changes to the targeted DNA sequence, including indels that may modulate regulatory activity[326].

The findings of the previous chapters suggest that for this OA genetic risk signal, the risk allele confers increased DNAm levels at the DMR in cartilage which results in the upregulation of *WWP2*-FL and *WWP2*-N. However, no changes in DNAm at the DMR were observed following deletion of the three variants. Changes to DNAm levels can be actively mediated by DNMTs and TETs, or passively following multiple cell replications or loss of epigenetic maintenance, but these effects are not always immediate[135,221]. Nucleic acids were harvested 72 hours post-nucleofection, which may have been too short of a window to observe any changes to DNAm levels at the DMR. Culturing of cells over a prolonged period may therefore be required to detect changes in DNAm levels. In addition, if the risk allele is required to increase DNAm levels at the DMR, deletion of each variant would not show this. Again, base editing provides a promising tool for investigating the relationship between each variant and DNAm levels at the DMR in future studies[324]. Regardless, these data demonstrate that deletion of rs111837947 results in decreased *WWP2*-FL expression independently of DNAm levels at the DMR. The deleted region encompassing rs111837947 is marked as an enhancer in chondrocytes using Roadmap histone ChIP-Seq data. It is plausible that removal of this DNA sequence induces changes to the local chromatin landscape, attenuating its capability to enhance expression of *WWP2*-FL.

Of the 13 TFs predicted to bind rs34195470 and rs111837947, only four were abundantly expressed in OA cartilage. SOX5, which was predicted to preferentially bind the OA risk-conferring G allele of rs34195470, is required for normal joint cavitation during development and is upregulated in OA cartilage in comparison to non-OA cartilage[13,15,327]. The A allele of rs34195470 may therefore attenuate SOX5 binding potential and decrease transcriptional activity in comparison to the OA-risk conferring G allele, in concordance with the findings of the Luciferase reporter assay, with potential implications for cartilage development and homeostasis. SP1 and SP2 are members of the Specificity Protein TF family required for many cellular processes including growth, as well as critical roles in embryonic development[328]. SP1 is predicted to exclusively bind the non-risk allele of rs111837947, whilst SP2 was not predicted to bind either allele using the JASPAR motif scan tool. ZNF740 is a member of the Zinc Finger Protein (ZNF) family that is predicted to exclusively bind the risk allele (insGT) of rs111837947. Whilst little is known about the specific function of ZNF740, other ZNFs have been identified to play roles in chromatin remodelling as well as transcriptional regulation[329]. To validate the findings of the *in-silico* TF analyses, EMSAs should be performed using chondrocyte nuclear protein and DNA probes constituting the risk and non-risk alleles of rs34195470, rs9746247 and rs111837947. Additionally, co-transfection of the four abundantly expressed TFs (SOX5, SP1, SP2, ZNF740) alongside the luciferase reporter vectors containing the three variants in their risk and non-risk conferring allele constructs could provide further evidence of a functional role for these TFs in mediating the observed regulatory effects.

In this chapter, the Luciferase reporter assay suggests a regulatory role for rs34195470 whilst the CRISPR/Cas9 experiments indicate that rs111837947 is capable of regulating *WWP2*-FL expression. Both rs34195470 and rs111837947 are predicted to bind TFs abundantly expressed in cartilage, and both have been highlighted as lead variants in previous OA GWASs[89,122,123]. rs34195470 and rs111837947 may therefore act in unison to modulate *WWP2* expression. Future work should aim to delineate the effects of differing combinations of risk and non-risk alleles at rs34195470, rs9746247 and rs111837947 upon *WWP2* expression using reporter assays and base editing technologies. These investigations may aid in deciphering whether there is a single causal variant or a haplotypic effect. Deletion of rs9746247 produced no observable effect upon *WWP2* expression, and no allelic difference in regulatory activity was observed in the Luciferase reporter assay. In addition, no TFs are predicted to bind rs9746247. Hence, rs9746247 is unlikely to be the causal variant. Further investigations including *in-vitro* base editing and expansion of the current CRISPR/Cas9 sample size are required to determine the true causal variant(s)[118,324].

Chapter 8: General Discussion

8.1 Introduction

The GWAS era has revealed over 100 independent OA risk variants to date[88,89,330], and as cohorts increase and diversify to include underrepresented ethnic groups and subpopulations, it is inevitable that more GWAS hits will be reported. Despite the success in identifying OA-associated risk variants, the translation of these signals into functional candidates for therapeutic intervention has not materialised[107]. As observed in most polygenic traits, the associated variants tend to reside within noncoding regions of the genome suggesting they confer their pathogenicity via gene regulation, consequently complicating the identification of causal variants and their gene targets[115,330]. Furthermore, identification of a genetic signal alone provides no information on the tissue or the life stage that variants mediate their functional effects, creating additional layers of complexity that mask the biological interpretation of each signal[88,165,330]. It is therefore vital that laboratory studies are performed to elucidate targets of OA genetic risk with the overarching goal of closing the gap between reporting of GWAS hits and their functional candidates, thus accelerating their translation into the clinic.

The last decade of OA research has demonstrated that epigenetics plays an important role in OA development and progression, with changes to DNAm, chromatin state and miRNA expression all associating with disease phenotype[88,131,158,184]. Approximately a quarter of all OA risk variants co-localise with cartilage mQTLs[330], with the associated CpGs overwhelmingly enriched in putative chondrocyte enhancers[88]. Research published by my principal supervisor's research group has demonstrated that experimental perturbation of OA cartilage mQTLs *in-vitro* can modulate the expression of target genes, suggesting they act as functional intermediaries of OA genetic risk. These studies have revealed genes associated with TGFβ signalling (*TGFB1*)[171], proteostasis (*RWDD2B* and *TMEM129*)[166,172], collagen glycosylation (*COLGALT2*)[173], and osteoblastic differentiation (*RUNX2*)[170] as effectors of OA genetic risk, prioritising these genes and their associated cellular pathways as avenues for therapeutic targeting.

In this thesis, I have aimed to determine the effector gene(s) of OA genetic risk marked by the rs34195470 association signal via the integration of *in-silico* datasets, molecular genetic and epigenetic analyses of patient samples, and functional studies performed using the immortalised chondrocyte cell line TC28a2[284]. In addition, this thesis aimed to decipher the causal variant at this OA association signal, and the TFs capable of mediating the observed transcriptional effects.

8.2 Summary of results

Prior to performing costly functional laboratory experiments, OA association signals should be interrogated using available *in-silico* datasets to garner information about the locus[330]. Findings of *in-silico* analyses allow for prioritisation of candidates likely to possess regulatory function for downstream functional studies. In Chapter 3, the GTEx Portal[213] was searched for eQTLs in any tissue associating with rs34195470, whilst chromatin regulatory state data at the locus was visualised using the UCSC Genome Browser[258]. Twelve genes were identified as rs34195470-eQTLs, with *WWP2* identified as the most abundantly expressed gene associating with rs34195470 genotype in OA cartilage (Fig. 3.1). The risk-conferring G allele of rs34195470 was previously identified to correlate with increased DNAm at two CpGs (cg26736200 and cg26661922) in OA cartilage, forming mQTLs[330]. cg26736200 and cg26661922 are located within *WWP2* and were identified to reside within an open, accessible chromatin region with histone marks indicative of regulatory function in OA-relevant cell types, as well as during development (Fig. 3.2). In addition, a single chromatin loop at the locus was identified in MSCs between the *WWP2*-FL/*WWP2*-N shared TSS and the DNA region housing the two CpGs and rs34195470 (Fig. 3.3). *WWP2*-FL and *WWP2*-N were therefore identified as possible targets of the rs34195470 risk signal and the associated mQTLs.

In Chapter 4, molecular genetic and epigenetic analyses were performed using OA patient cartilage samples (Fig. 4.1B and 4.6A), validating previous reports that the OA risk-conferring G allele of rs34195470 correlated with increased *WWP2* expression via AEI analysis[214], and increased DNAm levels at cg26736200 and cg26661922[330]. The cartilage mQTL study was expanded across a 228bp region to capture DNAm levels at a total of 16 CpGs. Fourteen of the 16 captured CpGs were mQTLs, with the risk allele (G) of rs34195470 always correlating with increased DNAm levels, marking the identification of a DMR in OA cartilage. The mQTL and AEI analyses were expanded to include other OA tissues (subchondral bone and synovium) as well as foetal cartilage to determine whether the observed effects were specific to a particular tissue or life stage. The mQTL analysis revealed DMRs were present in foetal cartilage and OA bone (Fig. 4.1C and 4.2B), with the latter particularly striking in males versus females (Appendix O). The AEI analysis revealed changes to *WWP2* expression correlating with rs34195470 genotype were restricted to OA cartilage (Fig. 4.6). Downstream experiments were therefore prioritised using chondrocytes as these were most likely to exhibit the functional effects of the OA association signal.

The findings of Chapters 3 and 4 suggested *WWP2* was the likely target of the rs34195470 association signal in cartilage, with the risk allele (G) correlating with increased *WWP2* expression and increased DNAm levels at the DMR. Previous studies have demonstrated that *in-vitro* epigenetic editing of OA cartilage mQTLs can mediate changes to target gene expression[166,171–173]. It was therefore hypothesised that higher DNAm levels at the DMR would increase *WWP2* expression in chondrocytes. The DMR exhibited methylation-sensitive regulatory activity using a reporter gene assay system (Fig. 5.1), whilst targeted epigenetic editing of the DMR to recapitulate the mQTL effect revealed that increased DNAm levels at the DMR did indeed result in higher expression of two *WWP2* isoforms: *WWP2*-FL and *WWP2*-N (Fig. 5.2). As a result, these isoforms were confirmed as the targets of OA genetic risk marked by rs34195470, regulated by the functional intermediary of DNAm. From these findings, a molecular mechanism was proposed (Fig. 5.4).

Having identified *WWP2*-FL and *WWP2*-N as targets of OA genetic risk, the focus of this thesis moved onto determining the TFs responsible for mediating the observed effects. Among nine other TFs predicted to bind the DMR using the JASPAR 2022 dataset[223] was HIF-1α (Fig. 5.3), a key mediator of the transcriptional response to hypoxia[293]. The abundant expression of *HIF1A* in cartilage, alongside its known roles in skeletogenesis[13] and a previous report by Tuerlings *et al*[243] highlighting hypoxia as an important pathway linked to the rs34195470 association signal suggested that functional experiments focussing on HIF-1α should therefore be prioritised. However, HIF-1α protein is rapidly degraded in the presence of cellular O_2[293]. As the experiments performed in Chapter 5 were performed under normoxic conditions, HIF-1α was unlikely to mediate the observed effects. To address this, the reporter assay and epigenetic editing experiments were repeated in Chapter 6 with the addition of $CoCl_2$ to cell culture medium, thus stabilising HIF-1α protein[297]. The binding of HIF-1α to the DMR was validated using EMSAs (Fig. 6.1), revealing that the addition of methyl groups to CpG7 and CpG8/cg26736200 ablated HIF-1α DNA binding, corroborating previous reports that HIF-1α only binds unmethylated DNA sequences[285,286]. The reporter assay demonstrated that the presence of HIF-1α could modulate the transcriptional activity of the unmethylated DMR, producing a weak enhancer effect that had not been observed in previous experiments where HIF-1α would have been degraded (Fig. 5.1 and 6.2). However, targeted epigenetic editing of the DMR performed using chondrocytes treated with $CoCl_2$ was unable to verify whether HIF-1α acts as a transcriptional mediator of the rs34195470 association signal (Fig. 6.3). It was concluded that this experiment was limited by potential off-target effects incurred using $CoCl_2$, including decreased cell viability and changes to chromatin state[310,312,317,318]. Further challenges to the study design arose following the 2024 update of the JASPAR database[260] which predicted HIF-2α to bind the DMR instead of HIF-1α (Fig. 6.5). The experiments presented in this chapter were therefore unable to delineate the TF responsible for mediating the rs34195470 association signal.

In Chapter 7, *in-silico* analyses and *in-vitro* functional experiments were employed to determine the causal variant. rs34195470 and the two variants in high LD with it (rs9746247 and rs111837947, $r^2 > 0.8$ in European populations) were determined to reside within closed, inaccessible chromatin defined as an enhancer in chondrocytes (Fig. 7.1). Reporter assays in TC28a2 chondrocytes revealed rs34195470 as the only variant exhibiting allele-dependent transcriptional changes to regulatory activity (Fig. 7.2), whilst CRISPR/Cas9 deletion of rs111837947 decreased *WWP2*-FL expression in comparison to a nontargeting control (Fig. 7.3). rs34195470 and rs111837947 were also predicted to bind TFs abundantly expressed in cartilage, including *SOX5* (Fig. 7.5). Together, the data implies that rs34195470 and rs111837947, but not rs9746247, are likely to be the causal variant(s) of this OA association signal. However, more research is required to confirm the causal variant(s), including the use of base editing[324].

8.3 Clinical utility of findings

The overarching goal of molecular genetic studies of OA is to translate discoveries into clinical interventions for the benefit of patients. There are currently no licensed disease-modifying OA drugs (DMOADs) available to halt disease progression, underlining this major clinically unmet need[331,332]. However, many OA risk loci co-localise with genes that encode components of targetable signalling pathways, including the TGFβ signalling pathway, providing an opportunity for drug re-purposing and a faster route towards clinical utility[89]. The development of DMOADs must also consider the timing of treatment. OA risk variants are thought to confer their pathogenic effects in three ways: (i) during development, (ii) across the entirety of the life course, or (iii) late in life[127,165,330]. Targeting of OA risk loci in the early stages of life raises questions for the likelihood of treatment for a disease that has not yet manifested, whereas variants active in later life and at the onset of OA are more likely to be amenable to treatment[331]. In the given scenarios, the rs34195470 association signal appears to fall into category (iii), altering expression of *WWP2* in adulthood but not development, suggesting it is amenable to therapeutic intervention.

WWP2-FL, WWP2-N and WWP2-C control the availability of the SMAD signal transducers: WWP2-FL controls SMAD2/3 steady-state levels in conjunction with WWP2-N, whilst WWP2-C also interacts with WWP2-FL to control availability of SMAD7, a TGFβ signalling inhibitor[231,234–236]. The levels of WWP2-FL, WWP2-N and WWP2-C therefore fine-tune TGFβ signalling. In one study of cardiac fibrosis, it was reported that WWP2-FL and WWP2-N, but not WWP2-C, contribute to disease by interacting with and regulating the transcriptional activity of SMAD2[236]. A similar scenario may be taking place in OA, with carriers of the OA risk-conferring G allele of rs34195470 exhibiting higher levels of WWP2-FL and WWP2-N protein, altering interaction dynamics with substrate molecules and conferring detrimental effects upon cartilage homeostasis. Clinical exploitation of this association signal should therefore focus on the target substrates of WWP2-FL and WWP2-N, and on inhibitors targeting these two isoforms.

Small molecule WWP2 inhibitors have been identified, complexing with the HECT domain shared by WWP2-FL and WWP2-C[333]. The WWP2-C isoform is abundantly expressed in cartilage and targets SMAD7 for degradation *in-vitro*[234], suggesting it is required for maintaining TGFβ signalling. In the context of this association signal, other small molecule inhibitors are required that target the domains exclusively shared by WWP2-FL and WWP2-N: the C2 domain and WW1 domain. The WW domains are responsible for the recruitment of target substrates[230]. Therefore, disruption of the WW1 domain may be an avenue for therapeutic intervention. WWP2-FL rapidly degrades SMAD2/3 in the absence of TGFβ via disruption of an auto-inhibitory complex mediated by WWP2-N[234]. This provides two alternative mechanisms for therapeutic intervention: preventing WWP2-N-dependent disruption of the WWP2-FL auto-inhibitory conformation; or increasing the levels of TGFβ in cartilage. In these scenarios, the former may be preferred as increased TGFβ in the aged joint has been associated with multiple deleterious effects including chondrocyte hypertrophy, synovial fibrosis and osteophyte formation[49].

CRISPR/Cas9 gene editing has recently been used to edit the *BCL11A* erythroid-specific enhancer with no observed off-target effects, increasing foetal haemoglobin and eliminating vaso-occlusive episodes, and therefore demonstrates a targeted approach to treating β-thalassemia and sickle cell disease[334]. This treatment has since been approved by the FDA, marking the world's first CRISPR therapy. This study demonstrates a path for targeted genome editing that may be applicable to OA. Whilst further research is required to determine the causal variant(s) of this OA association signal, swapping risk alleles for non-risk alleles may be protective against disease. However, it is important to consider that this association signal confers modest effects upon disease risk (OR < 1.1)[89,123] and therefore may only confer pathogenic effects in combination with multiple other inherited risk loci[125].

In addition to the underlying genetics of OA, several studies have highlighted DNAm as a functional mediator of OA genetic risk[166,171–173]. This presents the epigenome as another potentially exploitable pharmacological target. In the context of this OA association signal, targeted de-methylation of the DMR in cartilage would likely decrease transcriptional activation potential at the *WWP2*-FL/*WWP2*-N TSS, attenuating detrimental effects upon cartilage health. Development of any therapeutic targeting the epigenome must consider target specificity. For example, nonspecific epigenetic regulators and small molecules targeting their activity such as the DNMT inhibitor 5-aza-2'-deoxycytidine may induce desirable effects upon the expression of certain genes whilst inadvertently modulating the expression of other genes that produce deleterious effects in the joints, exacerbating OA[335–337]. Translation of epigenetic editors fused to CRISPR/Cas9 machinery (e.g. dCas9-DNMT3a and dCas9-TET1) is therefore an appealing prospect[220,335]. In addition, treatments must be appropriately targeted to the affected tissues to prevent undesirable effects in other joint tissues. In 2023, Kehayova *et al* reported changes to DNAm levels associated with OA genetic risk at the *COLGALT2* locus conferred opposing effects upon gene expression between cartilage and synovium[167]. Consequently, epigenetic editing may decrease the effects of an OA association signal in one joint tissue but inadvertently increase the detrimental effects observed in another. These findings also highlight the necessity for a comprehensive understanding of (epi)genetic risk across all tissues affected by OA.

8.4 Limitations, unanswered questions and future directions

Whilst the body of work presented in this thesis has identified the effectors of OA genetic risk marked by rs34195470 and provided evidence that can be built upon for the identification of the causal variant, there are limitations in the study design and unanswered questions. In addition to the limitations discussed in each results chapter (Chapters 3-7), this section highlights four areas for consideration in future studies.

The reporter assays and (epi)genome editing experiments presented in this thesis were performed using an immortalised chondrocyte cell line routinely studied in cartilage research. TC28a2 chondrocytes, derived from the costal cartilage of a 15-year-old Caucasian female, express high levels of *COL2A1* and *SOX9*, representing a chondrogenic phenotype[266,284]. However, serial subculturing of TC28a2 cells in monolayer can lead to a loss of phenotypic stability, inducing the expression of osteogenic markers such as *COL1A1*[338]. Future experiments should therefore use human primary articular chondrocytes (HPACs) from healthy adult donors and OA arthroplasty patients, providing greater insight into the true biological mechanisms of OA genetic risk. The use of HPACs is not without its own limitations however, with HPACs rapidly de-differentiating following isolation and subculturing in monolayer[339,340]. Furthermore, the inter-individual variability between patient arthroplasty samples harvested for HPACs may limit the capability of the functional experiments employed here to observe modest changes in gene expression conferred by most OA risk loci[89,330]. The TC28a2 cell line therefore remains a biologically relevant model for studying OA association signals that future studies utilising HPACs can build upon.

Patient analyses demonstrate that the OA risk-conferring G allele of rs34195470 correlates with increased DNAm levels at the DMR and increased *WWP2* expression in cartilage. Functional studies established that increased DNAm levels at the DMR resulted in higher expression of *WWP2*-FL and *WWP2*-N. The DMR is therefore a functional intermediary through which OA genetic risk is exerted by the rs34195470 association signal. This raises an important question: what mechanism mediates the changes to DNAm levels at the DMR in the presence of the OA risk allele? Experiments such as base editing must be performed that establish a causal link between the alleles present and the DNAm levels at the DMR[324]. This limitation is present in all previous functional studies of OA risk loci and their associated mQTLs, where the mediators of epigenetic change have not yet been identified[166,171–173].

Having identified *WWP2*-FL and *WWP2*-N as effectors of OA genetic risk, research must now focus on how these isoforms confer their pathogenic effects in cartilage. As discussed in Section 8.3, the effect of the rs34195470 association signal upon *WWP2*-FL and *WWP2*-N expression likely leads to changes in TGFβ signalling as observed in cardiac fibroblasts and the human embryonic kidney cell line HEK293T[234,236]. This observation must be validated in a suitable cartilage model, such as the TC28a2 cell line or HPACs. Tuerlings *et al* demonstrated that upregulation of *WWP2*-FL in a 3D pellet culture of OA HPACs resulted in decreased expression of the anabolic markers *ACAN* and *COL2A1*[243]. However, proteomic analysis revealed no observable changes in the protein expression of these markers[243]. Meanwhile, Mokuda *et al* used mouse models to demonstrate that the catalytic HECT domain of WWP2 mediates degradation of RUNX2, in turn regulating expression of the matrix-degrading ADAMTS5, suggesting that WWP2 is protective against OA[237]. Whilst these results are inconsistent with the findings of the OA association signal, where OA risk is conferred by increased *WWP2*-FL and *WWP2*-N expression, it is important to point out that the findings of the Mokuda *et al* study simultaneously investigated the activity of WWP2-C in addition to the WWP2-FL[237]. Therefore, this model does not account specifically for OA risk conferred by the rs34195470 association signal. Nevertheless, the discordancy between the mouse knockout model and molecular studies using human OA patient samples suggest WWP2 activity in disease is inconsistent across species, therefore future studies should focus on human models of cartilage phenotyping for maximum translational potential, building upon the gene/protein expression analyses and histological approaches taken by Tuerlings *et al*[243].

The identification of WWP2 as an OA effector gene reveals the third E3 ubiquitin ligase or ubiquitin ligase-like gene identified as a target of OA risk via the epigenetic intermediary of DNAm. *TMEM129* encodes a RING-type E3 ubiquitin ligase known as Transmembrane Protein 129 and was identified to be downregulated in the presence of the OA risk-conferring T allele of rs11732213[172]. TMEM129 protein localises to the endoplasmic reticulum (ER) where it acts to facilitate protein degradation via proteolysis and mediates the unfolded protein response (UPR)[341]. ER stress, driven by an accumulation of protein aggregates in the ER, initiates the UPR[342]. However, chronic ER stress can contribute to OA and other diseases of the skeletal system, including chondrodysplasias[343,344]. *RWDD2B*, encoding RWD Domain Containing 2B, was identified by Parker *et al* as another target of OA genetic risk, with the risk allele of rs6516886 correlating with decreased expression of this gene[166]. Whilst little is known about

RWDD2B, proteins encoding RWD domains (including ubiquitin ligases) have the capacity to bind other proteins[345]. Other E3 ubiquitin ligases have also been reported to contribute to OA susceptibility independently of genetics. SMURF2 negatively regulates TGFβ signalling via targeting of SMAD3, which contributes to an OA phenotype in mice[54]. Loss of proteostasis has recently been described as a hallmark of ageing alongside genome instability and epigenetic alterations[132]. OA chondrocytes exhibit impaired proteasomal function, conferring detrimental effects upon cellular phenotype including decreased SOX9 and ACAN expression[346]. Changes to the steady-state levels of cellular proteins mediated by the targets of OA genetic risk may therefore contribute to the loss of proteostasis observed in OA and highlights an important area for further research.

8.5 Conclusions

The data presented in this thesis reports for the first time that functional fine-mapping tools have been used to reveal an OA association signal that regulates the expression of specific transcript isoforms of a gene (*WWP2*-FL and *WWP2*-N). In addition, the integration of *in-silico* datasets with molecular analyses of patient samples and *in-vitro* functional experiments provides a roadmap that future studies may replicate to elucidate effector genes of OA risk signals. Future studies should build upon the employed techniques to close the gap between GWAS discoveries and functional validation of causal variants and their effector genes, providing a comprehensive understanding of OA genetic risk and ultimately contributing to the development of therapeutic interventions that improve patient outcomes.

Chapter 9: Appendices

Appendix A. Sample details for OA cartilage samples.

Sample ID	Age (years)	Sex	Joint	mQTL analysis	AEI analysis	RTqPCR analysis
6	71	M	Knee	No	Yes	No
19	66	M	Hip	Yes	No	No
22	48	M	Hip	Yes	No	Yes
39	72	F	Knee	Yes	No	No
40	84	F	Hip	Yes	No	Yes
41	51	M	Knee	Yes	No	Yes
42	71	F	Hip	Yes	No	Yes
45	93	M	Knee	Yes	No	Yes
49	68	M	Hip	Yes	No	No
51	74	F	Hip	Yes	Yes	Yes
52	74	F	Hip	Yes	No	No
53	55	F	Hip	Yes	No	Yes
54	52	M	Hip	Yes	No	No
57	63	M	Knee	Yes	Yes	Yes
59	60	F	Knee	Yes	No	Yes
61	67	M	Knee	Yes	No	No
66	56	F	Hip	Yes	No	Yes
69	57	M	Hip	Yes	No	No
70	75	M	Hip	Yes	No	No
72	76	M	Knee	Yes	No	No
73	46	F	Hip	Yes	No	No
76	65	F	Knee	Yes	Yes	Yes
77	71	F	Knee	Yes	No	Yes
78	58	M	Knee	Yes	Yes	No
79	65	M	Knee	Yes	No	Yes
82	67	F	Knee	Yes	No	No
86	56	F	Knee	Yes	No	Yes
87	51	F	Hip	Yes	No	No
89	67	F	Knee	Yes	No	Yes
90	60	F	Hip	Yes	No	No
92	55	F	Hip	Yes	No	Yes
93	64	F	Knee	No	Yes	No
96	61	F	Hip	Yes	Yes	No
97	65	M	Hip	Yes	No	No
98	49	M	Hip	Yes	No	Yes
99	82	F	Knee	No	Yes	No
100	74	F	Hip	Yes	No	No
103	70	F	Knee	Yes	No	Yes
104	70	F	Knee	Yes	No	Yes

106	79	M	Knee	Yes	No	Yes
107	41	F	Knee	Yes	No	Yes
108	72	M	Knee	Yes	No	Yes
109	65	F	Knee	Yes	No	No
112	61	M	Hip	Yes	No	No
114	57	F	Knee	Yes	Yes	No
115	69	M	Knee	Yes	No	No
116	89	M	Hip	Yes	No	No
117	82	M	Hip	Yes	No	No
126	66	F	Hip	Yes	No	No
127	61	M	Knee	Yes	No	Yes
128	64	M	Knee	Yes	No	No
132	65	M	Knee	Yes	No	No
135	69	F	Hip	Yes	No	No
136	68	F	Knee	Yes	No	No
166	62	F	Knee	No	No	Yes

Appendix B. Sample details for foetal cartilage samples.

Sample ID	Age (pcw)	Sex	Joint	mQTL analysis	AEI analysis	RTqPCR analysis
14377	14	M	Tibia	Yes	No	No
14378	14	M	Tibia	Yes	No	No
14392	9	M	Tibia	Yes	No	No
14393	14	M	Tibia	Yes	No	Yes
14397	15	F	Femur and tibia	No	No	Yes
14423	12	M	Tibia	Yes	No	Yes
14429	14	M	Tibia	Yes	No	No
14451	9	M	Femur	Yes	No	Yes
14453	12	F	Femur	Yes	No	Yes
14460	9	F	Tibia	Yes	No	No
14464	16	M	Tibia	Yes	Yes	Yes
14467	10	M	Tibia	Yes	No	Yes
14471	9	F	Femur and tibia	Yes	No	No
14475	16	M	Tibia	Yes	No	Yes
14492	12	F	Tibia	Yes	Yes	No
14501	16	M	Tibia	Yes	No	No
14510	14	F	Tibia	No	No	Yes
14512	10	M	Tibia	Yes	No	No
14513	12	F	Femur and tibia	Yes	No	Yes
14516	15	M	Tibia	Yes	No	No
14521	12	M	Tibia	Yes	Yes	No
14523	15	F	Tibia	Yes	No	No
14524	15	F	Tibia	Yes	No	Yes
14525	10	M	Tibia	Yes	No	Yes
14532	11	F	Tibia	Yes	No	No
14541	11	M	Tibia	Yes	No	No
14544	13	M	Tibia	Yes	No	No
14555	9	M	Femur and tibia	Yes	No	No
14562	9	F	Tibia	Yes	No	Yes
14576	8	M	Femur and tibia	Yes	No	No
14580	9	M	Femur and tibia	Yes	No	Yes
14586	13	F	Femur and tibia	Yes	No	Yes
14600	8	M	Tibia	Yes	No	Yes
14601	10	M	Femur and tibia	Yes	No	Yes
14604	14	F	Tibia	Yes	No	Yes
14617	13	F	Femur and tibia	Yes	No	No
14619	16	M	Femur and tibia	Yes	No	No
14628	17	F	Femur and tibia	Yes	No	No
14684	14	F	Femur	Yes	No	No

14703	12	F	Femur and tibia	Yes	No	Yes
14713	16	F	Femur and tibia	Yes	No	No
14715	12	M	Femur and tibia	Yes	No	Yes
14716	10	M	Femur and tibia	Yes	No	Yes
14720	15	F	Femur and tibia	Yes	No	Yes
14721	10	F	Femur and tibia	Yes	No	No
14722	17	M	Femur and tibia	Yes	Yes	Yes
14728	16	M	Femur and tibia	Yes	Yes	No
14746	16	F	Femur and tibia	Yes	Yes	Yes
14748	10	F	Femur and tibia	Yes	Yes	Yes
14831	9	F	Femur and tibia	Yes	No	No

Appendix C. Sample details for OA bone samples.

Sample ID	Patient Age	Sex	Joint	mQTL analysis	AEI analysis	RTqPCR analysis
310	69	M	Hip	Yes	No	No
325	81	F	Hip	Yes	No	Yes
331	56	F	Hip	Yes	No	No
333	60	F	Hip	Yes	Yes	Yes
337	68	M	Hip	Yes	No	Yes
350	81	F	Hip	Yes	No	No
353	75	M	Hip	Yes	No	No
354	74	F	Hip	Yes	No	Yes
359	80	F	Hip	Yes	No	No
361	59	M	Hip	Yes	No	No
372	83	F	Hip	Yes	No	Yes
374	86	F	Hip	Yes	No	No
377	64	F	Hip	Yes	No	Yes
380	77	M	Hip	Yes	No	Yes
381	59	F	Hip	Yes	Yes	Yes
387	59	F	Hip	Yes	No	No
390	72	M	Hip	Yes	No	No
393	75	M	Hip	Yes	No	No
394	71	F	Hip	Yes	No	No
399	67	M	Hip	Yes	No	No
432	65	F	Hip	Yes	No	No
435	77	F	Hip	Yes	No	No
440	62	M	Hip	Yes	No	Yes
442	60	F	Hip	Yes	No	No
447	82	F	Hip	Yes	No	Yes
449	77	F	Hip	Yes	No	Yes
451	74	F	Hip	Yes	Yes	Yes
456	75	M	Hip	Yes	No	No
459	64	M	Hip	Yes	No	No
460	75	M	Hip	Yes	No	No
469	74	M	Hip	Yes	No	No
475	69	M	Hip	Yes	Yes	Yes
477	63	M	Hip	Yes	No	Yes
481	85	F	Hip	Yes	No	No
482	67	M	Hip	Yes	No	No
483	60	M	Hip	Yes	No	No
488	61	M	Hip	Yes	No	Yes
489	73	M	Hip	Yes	No	No
490	74	F	Hip	Yes	No	No

491	62	F	Hip	Yes	No	Yes
492	75	M	Hip	Yes	No	No
495	51	F	Hip	Yes	No	No
508	74	F	Hip	Yes	No	No
510	69	F	Hip	Yes	No	No
514	58	F	Hip	Yes	No	No
516	75	M	Hip	Yes	Yes	Yes
518	80	F	Hip	Yes	No	Yes
521	77	M	Hip	Yes	Yes	Yes
522	73	M	Hip	Yes	Yes	Yes
525	75	M	Hip	Yes	No	No
529	52	F	Hip	Yes	No	No
530	81	F	Hip	Yes	No	No
534	54	M	Hip	Yes	No	No
537	65	F	Hip	Yes	No	Yes
538	53	F	Hip	Yes	Yes	Yes
544	66	F	Hip	Yes	No	No
548	67	M	Hip	Yes	No	No
562	73	M	Hip	Yes	No	No
566	70	M	Hip	Yes	No	Yes
569	53	F	Hip	Yes	No	No
577	58	F	Hip	Yes	No	Yes

Appendix D. Sample details for OA synovium samples.

Sample ID	Age (years)	Sex	Joint	mQTL analysis	AEI analysis	RTqPCR analysis
3975	60	M	Knee	Yes	No	No
4288	85	M	Knee	Yes	No	Yes
4429	65	M	Hip	Yes	No	Yes
4432	72	F	Knee	Yes	No	No
4512	55	F	Knee	Yes	No	No
4552	82	F	Knee	Yes	No	No
4983	75	M	Knee	Yes	No	No
4984	62	M	Knee	Yes	No	Yes
5014	55	M	Knee	Yes	No	Yes
5023	75	M	Knee	Yes	No	Yes
5240	62	F	Knee	Yes	Yes	Yes
5266	72	F	Knee	Yes	Yes	No
5281	59	F	Knee	Yes	No	No
5299	75	F	Knee	Yes	Yes	No
5300	87	M	Knee	Yes	Yes	No
5301	70	M	Knee	Yes	No	No
5351	66	F	Hip	Yes	No	No
5377	80	M	Knee	Yes	No	Yes
5378	65	M	Knee	Yes	No	No
5380	72	M	Knee	Yes	No	No
5396	52	F	Knee	Yes	Yes	Yes
5400	62	F	Knee	Yes	Yes	No
5410	67	M	Knee	Yes	No	No
5482	91	M	Knee	Yes	No	No
5506	71	F	Knee	Yes	No	No
5509	62	M	Knee	Yes	No	No
5536	53	F	Knee	Yes	No	Yes
5550	57	M	Knee	Yes	No	No
5551	49	M	Knee	Yes	No	No
5552	54	F	Knee	Yes	No	No
5558	75	F	Knee	Yes	No	Yes
5562	73	M	Knee	Yes	No	No
5564	73	F	Knee	Yes	No	No
5569	63	F	Knee	Yes	No	No
5572	72	F	Knee	Yes	Yes	No
5603	79	F	Knee	Yes	No	No
5613	75	M	Knee	Yes	Yes	Yes
5614	60	M	Knee	Yes	No	No
5622	61	F	Knee	Yes	Yes	Yes

5632	52	F	Knee	No	No	Yes
5643	76	M	Knee	Yes	No	Yes
5648	66	F	Knee	Yes	No	Yes
5652	82	F	Knee	Yes	No	No
5654	85	M	Knee	Yes	No	Yes
5665	61	M	Knee	Yes	No	No
5677	54	M	Knee	Yes	Yes	Yes
5680	80	F	Knee	Yes	Yes	Yes
5681	51	F	Knee	Yes	No	Yes
5712	67	F	Knee	Yes	No	Yes
5713	65	F	Knee	Yes	No	Yes
5720	68	M	Knee	Yes	Yes	Yes
5776	85	F	Knee	Yes	No	Yes
5798	52	F	Knee	Yes	No	No
6068	59	F	Knee	Yes	No	No
6094	54	F	Knee	Yes	No	No
6535	67	M	Knee	Yes	No	No
6547	50	F	Knee	Yes	No	No
6548	81	F	Knee	Yes	No	No
6549	45	M	Knee	Yes	No	No
6550	80	F	Knee	Yes	No	No
6593	72	M	Knee	Yes	No	No
6599	81	F	Knee	Yes	Yes	No
6603	54	F	Knee	Yes	Yes	No
6636	50	F	Knee	Yes	No	No
6655	58	M	Knee	Yes	Yes	No
6877	71	F	Knee	Yes	No	No
6940	72	F	Knee	Yes	No	No
1066778	69	F	Knee	Yes	No	No
8011503	55	F	Hip	Yes	No	No
8041192	72	M	Hip	Yes	No	No
8157410	82	F	Knee	Yes	No	No

Appendix E. Primers used for pyrosequencing. [btn], biotin tag at 5’ end of primer.

Application	Forward (5'-3')	Reverse (5'-3')	Sequencing (5'-3')	Captured Loci	Genomic Position (hg19)
Genotyping	CCTGCTGGGACCCACTGA	[btn]GCTGGGATTACAGGCGTGAG	TCTGCAGTATTGAAACAGA	rs34195470	Chr16:69,955,690
Genotyping / Allelic expression imbalance (AEI) analysis	TCCATGCTCCCAGATTCTCG	[btn]CCACAGGGAAGTACAGCAAATGAC	GCTTGCCACAGCGCA	rs1052429	Chr16:69,975,360
Methylation quantification	TGGGAGTGGGGTGGTTATT	[btn]CCAAACCCTATTTTCCCTACTCACTAAACA	GGTTTTTATTTTTAGTTGGGAA	CpG 1 (cg02147637)	Chr16:69,951,622
				CpG 2	Chr16:69,951,667
			AATATTAGGATTATTATTTGGTAG	CpG 3 (cg17669573)	Chr16:69,951,685
				CpG 4	Chr16:69,951,689
				CpG 5	Chr16:69,951,693
				CpG 6	Chr16:69,951,695
			TTAAAGATTTTTTTATTAGGTGAGA	CpG 11	Chr16:69,951,800
				CpG 12	Chr16:69,951,809
				CpG 13 (cg26661922)	Chr16:69,951,820
				CpG 14	Chr16:69,951,830
				CpG 15	Chr16:69,951,836
				CpG 16	Chr16:69,951,850
	[btn]TGGGAGTGGGGTGGTTATT	CCAAACCCTATTTTCCCTACTCACTAAACA	ACCCCCTAAAACTAATTC	CpG 9	Chr16:69,951,728
				CpG 10	Chr16:69,951,733
	[btn]TTTTGAGGTTTTTATTTTTAGTTGGGAAA	ACATAACCCCCTAAAACTAATTCC	TCCCCTACCACTACCACTACTCA	CpG 7	Chr16:69,951,702
				CpG 8 (cg26736200)	Chr16:69,951,706

Appendix F. cDNA synthesis and RT-qPCR primers. n/a, not applicable.

Product ID	Assay Type	Supplier	Primer/Probe Sequences (5'-3')	Gene	Exon Location	Captured Transcripts	RefSeq ID	Ensembl Transcript ID	Ensembl ID
n/a	cDNA synthesis PCR	Integrated DNA Technologies	Forward: TCGAAGAGTGAACCAGCCTT	*HBP1*	e4-e5	*HBP1*	NM_012257	ENST00000222574.9	HBP1-201
n/a			Reverse: GAAGGCCAGGAATTGCACCATCC						
Hs.PT.58.45621782	PrimeTime Std qPCR Assay	Integrated DNA Technologies	n/a	*WWP2*	e1 - e3	*WWP2*-FL	NM_007014	ENST00000359154.7	WWP2-202
Hs.PT.58.27020061.g	PrimeTime Std qPCR Assay	Integrated DNA Technologies	n/a		e13a	*WWP2*-C	NM_199424	ENST00000568684.1	WWP2-213
Universal Probe Library #13	IDT oligo + Universal Probe Library	Sigma Aldrich	n/a		e8 - e9	*WWP2*-N	NM_001270455	ENST00000569174.5	WWP2-216
n/a		Integrated DNA Technologies	Forward: GAAGGAGAGGAACCCAGCA						
			Reverse: AGAGTCTCAGGTCTTCTCAGT						
hsa-miR-140-5p	TaqMan Advanced miRNA Assay	Thermo Fisher Scientific	n/a	miR-140	n/a	miR-140-5p	NR_029681	ENST00000385282.3	MIR140-201
hsa-miR-140-3p	TaqMan Advanced miRNA Assay	Thermo Fisher Scientific	n/a		n/a	miR-140-3p	NR_029681		
Hs.PT.39a22214836	PrimeTime Std qPCR Assay	Integrated DNA Technologies	n/a	*GAPDH*	e2 - e3	*GAPDH*	NM_002046	ENST00000229239.10	GAPDH-201
Hs.PT.58v.45621572	PrimeTime Std qPCR Assay	Integrated DNA Technologies	n/a	*HPRT1*	e8 - e9	*HPRT1*	NM_000194	ENST00000298556.8	HPRT1-201
n/a	IDT oligo	Integrated DNA Technologies	Forward: CGAATGGCTCATTAAATCAGTTATGG	*18S*	e1	*18S*	NR_003286	n/a	n/a

			Reverse: TATTAGCTCTAGAATTACCACAGTTATCC						
			Probe: 56-FAM/TCCTTTGGTCGCTCGCTCCTCTCCC/36-TAMSp						
U6 snRNA	TaqMan microRNA Control Assay	Thermo Fisher Scientific	n/a	*U6*	n/a	*U6*	NR_004394	n/a	n/a

Appendix G. Oligonucleotide sequences for *in vitro* experiments, including cloning primers, gRNAs for genome and epigenome editing, site-directed mutagenesis, EMSAs, and primers for endpoint and splicing PCRs. Lower case sequences represent restriction enzyme motifs: cctagg, *AvrII*; actagt, *SpeI*; caccg, *BbsI* forward strand; aaac[gRNA]c, *BbsI* reverse strand; acgcgt, *MluI*; ctcgag, *XhoI*. n/a, not applicable. [5medC], methylated cytosine nucleotide.

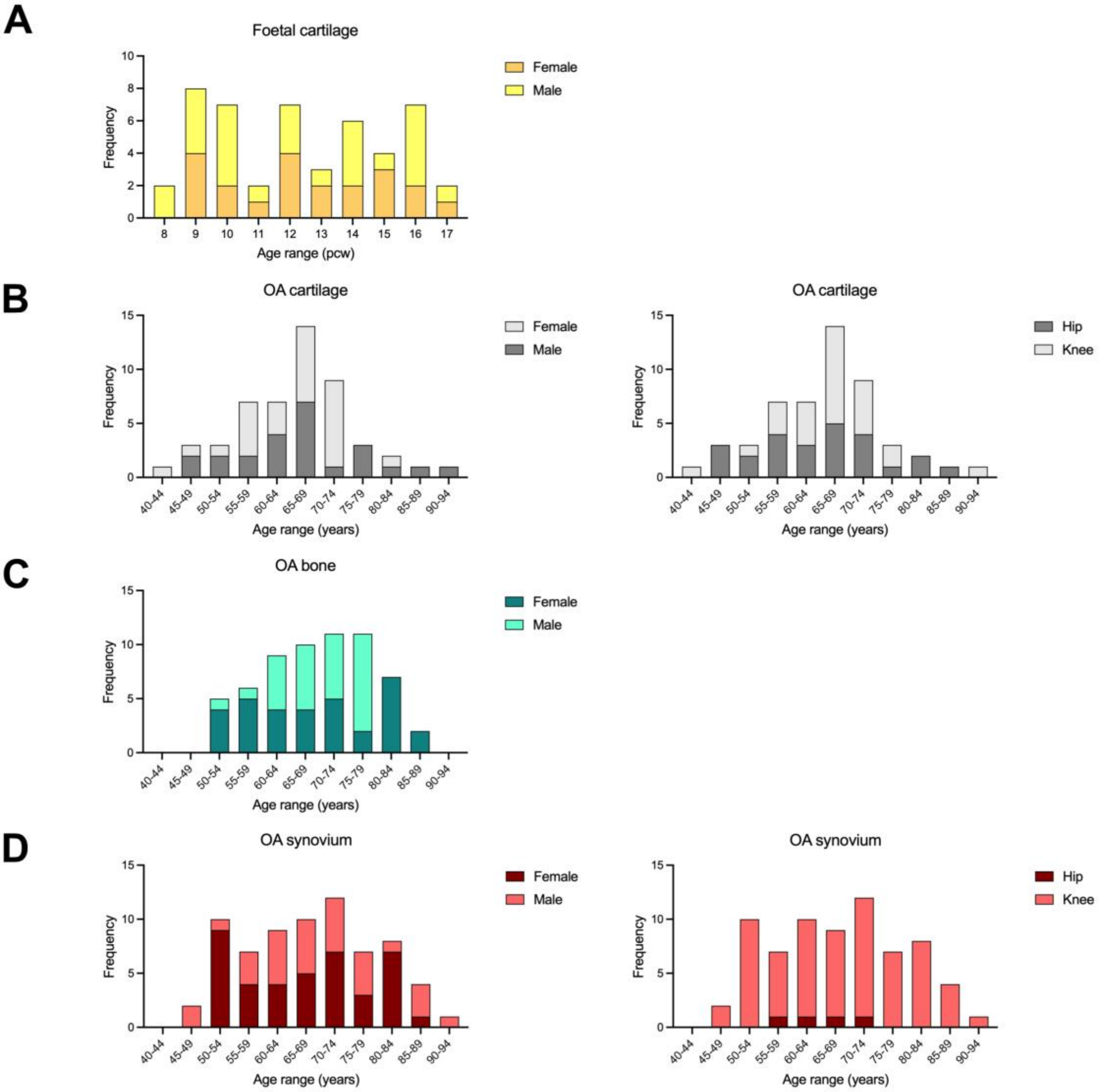
Application	Assay Name / Utility	SNV Target	Forward (5'-3')	Reverse (5'-3')
Lucia reporter gene assay	Cloning of DMR	n/a	cctaggGAGGTTCCTATTTCCAGCTGG	actagtTCACTAAGCAGGTCAACAGAGA
dCas9-DNMT3a epigenome modulation	gRNA 1	n/a	caccgCTGAGGTTCCTATTTCCAGC	aaacGCTGGAAATAGGAACCTCAGc
	gRNA 2	n/a	caccgACAATACTCGGACCACCACC	aaacGGTGGTGGTCCGAGTATTGTc
	gRNA 3	n/a	caccgACGTGCGCAACTATGAGCAG	aaacCTGCTCATAGTTGCGCACGTc
	gRNA 4	n/a	caccgTCACCTGGTAGAGGAATCTT	aaacAAGATTCCTCTACCAGGTGAc
	gRNA 5	n/a	caccgGTTTTCCCTGCTCACTAAGC	aaacGCTTAGTGAGCAGGGAAAACc
	Non-targeting gRNA (Control)	n/a	caccgCGTTAATCGCGTATAATACG	aaacCGTATTATACGCGATTAACGc
EMSA	Unmethylated probes	n/a	CGACCGCGGAGTACGTGCGCAACTATGAGCA	TGCTCATAGTTGCGCACGTACTCCGCGGTCG
	Methylated probes	n/a	CGACCGCGGAGTA[5medC]GTG[5medC]GCAACTATGAGCA	TGCTCATAGTTG[5medC]GCA[5medC]GTACTCCGCGGTCG
Luciferase reporter gene assay	Cloning of SNV region	rs34195470, rs9746247, rs111837947	GGGGacgcgtCTGCCATGTTTCCCTTTCCC	GGGGctcgagACAGGGGCATGGTCATCTTT
Site-directed mutagenesis	rs34195470 A>G	rs34195470	CCGCGCCCAGTCAAAACAATCTGTTTCAATACTGCA	TGCAGTATTGAAACAGATTGTTTTGACTGGGCGCGG
	rs9746247 C>G	rs9746247	CTCACTCTGTCTCTCTGGCTGGAGTGCAGT	ACTGCACTCCAGCCAGAGAGACAGAGTGAG

	rs111837947 insGT	rs111837947	GGGGTGGGGGGGGTTGCGTTCTGATAG	CTATCAGAACGCAACCCCCCCCACCCC
Luciferase reporter gene assay	SNV-specific PCR	rs34195470	GGGGacgcgtCTGCCATGTTTCCCTTTCCC	GGGGctcgagCTCAGCTCACCACAACCAAC
	SNV-specific PCR	rs9746247	GGGGacgcgtATTATTTTGACTGGGCGCGG	GGGGctcgagCATAAAGGTTGCAGCGTGGA
	SNV-specific PCR	rs111837947	GGGGacgcgtCCTCCTGGGTTTAAGCGATTA	GGGGctcgagACAGGGGCATGGTCATCTTT
CRISPR/Cas9 deletions	gRNA 1	rs34195470, rs9746247	caccgGTTGTGACGGGACTATCACA	aaacTGTGATAGTCCCGTCACAACc
	gRNA 2	rs34195470, rs9746247	caccgGCATTGAATTCTCAAGCACC	aaacGGTGCTTGAGAATTCAATGCc
	gRNA 1	rs111837947	caccgCTATAAGAAGCGAGGAGAAG	aaacCTTCTCCTCGCTTCTTATAGc
	gRNA 2	rs111837947	caccgTCTGATAGAATAACAAACAG	aaacCTGTTTGTTATTCTATCAGAc
Endpoint PCR (deletion confirmation)	n/a	rs34195470, rs9746247	TTTTGCCGTCTTGTGTCGT	CACCAGACGGATGCAGAAC
	n/a	rs111837947	TTGCACCAACCTAATAAATGTTGT	CACAGGGGCATGGTCATCTT
WWP2-FL splicing PCR	n/a	n/a	AGCGAGAGCTGCCCAACG	CCCGTCCATTGTCCTGTCTCT

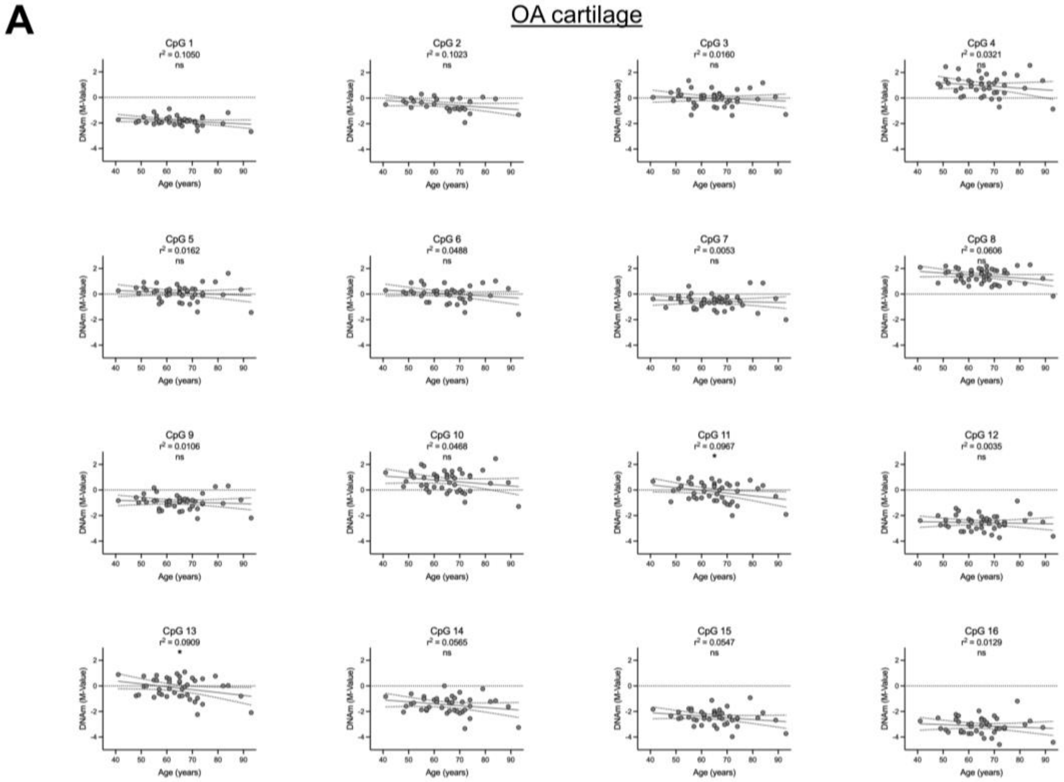
Appendix H. The 12 genes identified as rs34195470-eQTLs using GTEx Portal. Where more than three tissues were identified to host rs34195470-eQTLs, three representative tissues were listed. Gene coordinates are mapped to UCSC build hg19.

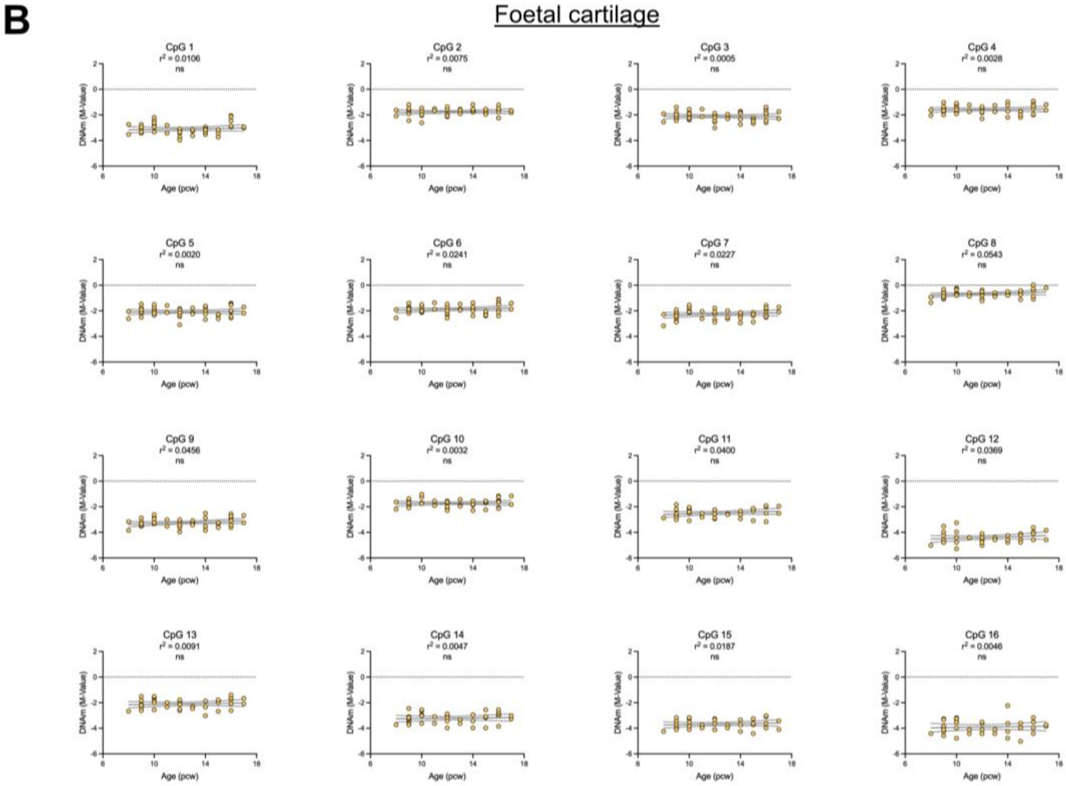
eQTL gene	Type	Tissue(s)	Gene Coordinates (hg19)
CLEC18A	Protein-coding	Adipose - Subcutaneous Nerve – Tibial Muscle – Skeletal	Chr16:69,984,608-69,998,250
CLEC18C	Protein-coding	Adrenal Gland Thyroid Whole Blood	Chr16:70,207,928-70,220,798
EXOSC6	Protein-coding	Adipose – Subcutaneous Muscle – Skeletal Whole Blood	Chr16:70,284,134-70,285,833
NFAT5	Protein-coding	Thyroid	Chr16:69,599,869-69,738,569
NOB1	Protein-coding	Oesophagus Thyroid	Chr16:69,775,757-69,788,871
NPIPB14P	Pseudogene	Adipose – Subcutaneous Artery – Tibial Muscle – Skeletal	Chr16:70,010,292-70,030,091
PDXDC2P	Pseudogene	Artery – Tibial Muscle – Skeletal Nerve – Tibial	Chr16:70,044,905-70,099,851
RP11-296I10.3	lncRNA	Testis	Chr16:70,192,860-70,207,351
RP11-394B2.1	lncRNA	Thyroid	Chr16:70,695,570-70,699,739
RP11-394B2.5	lncRNA	Oesophagus – Mucosa	Chr16:70,747,890-70,750,793
SMG1P7	Protein-coding	Adipose – Subcutaneous Muscle – Skeletal Thyroid	Chr16:70,253,484-70,259,936
WWP2	Protein-coding	Artery – Tibial	Chr16:69,796,235-69,975,642

Appendix I. Histograms showing the age distribution of patient arthroplasty and foetal donor samples used in Chapter 4. **(A)** Age distribution in post-conception weeks (pcw) of foetal cartilage samples stratified by sex. **(B)** Age distribution in years of OA cartilage samples stratified by sex (left panel) and joint site (right panel). **(C)** Age distribution in years of OA bone samples stratified by sex. **(D)** Age distribution in years of OA synovium samples stratified by sex (left panel) and joint site (right panel).

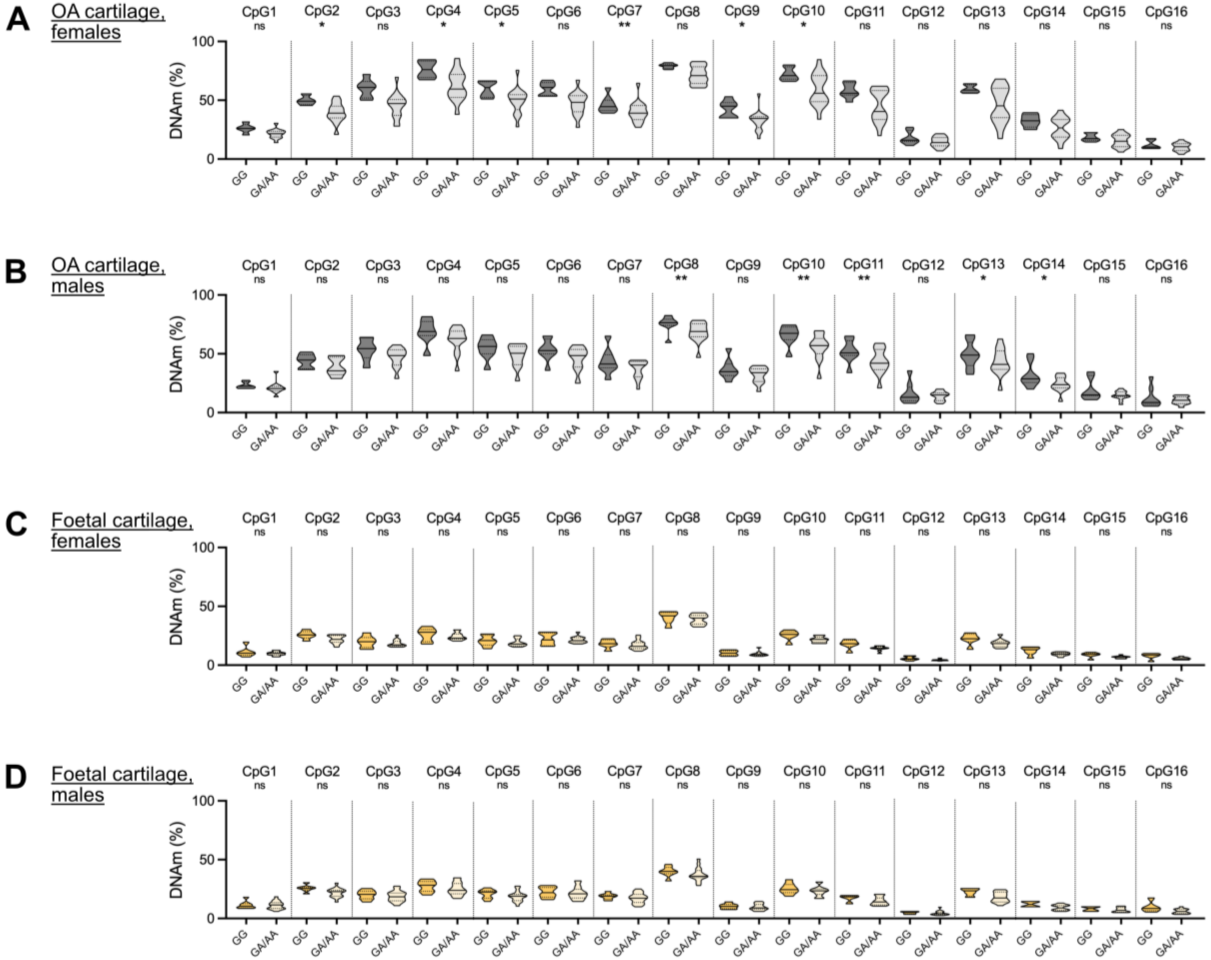


Appendix J. Stratification of DNAm levels at the 16 CpGs by age in OA cartilage (panel **A**, in grey) and foetal cartilage (panel **B**, in yellow). For both panels, the dotted line at *y*=0 represents 50% DNAm. Coloured circles represent individual samples. *P*-values and r^2 values calculated using simple linear regression. The solid line represents the regression line with 95% confidence intervals represented by dashed lines. * = $P < 0.05$; ns = not significant ($P > 0.05$).

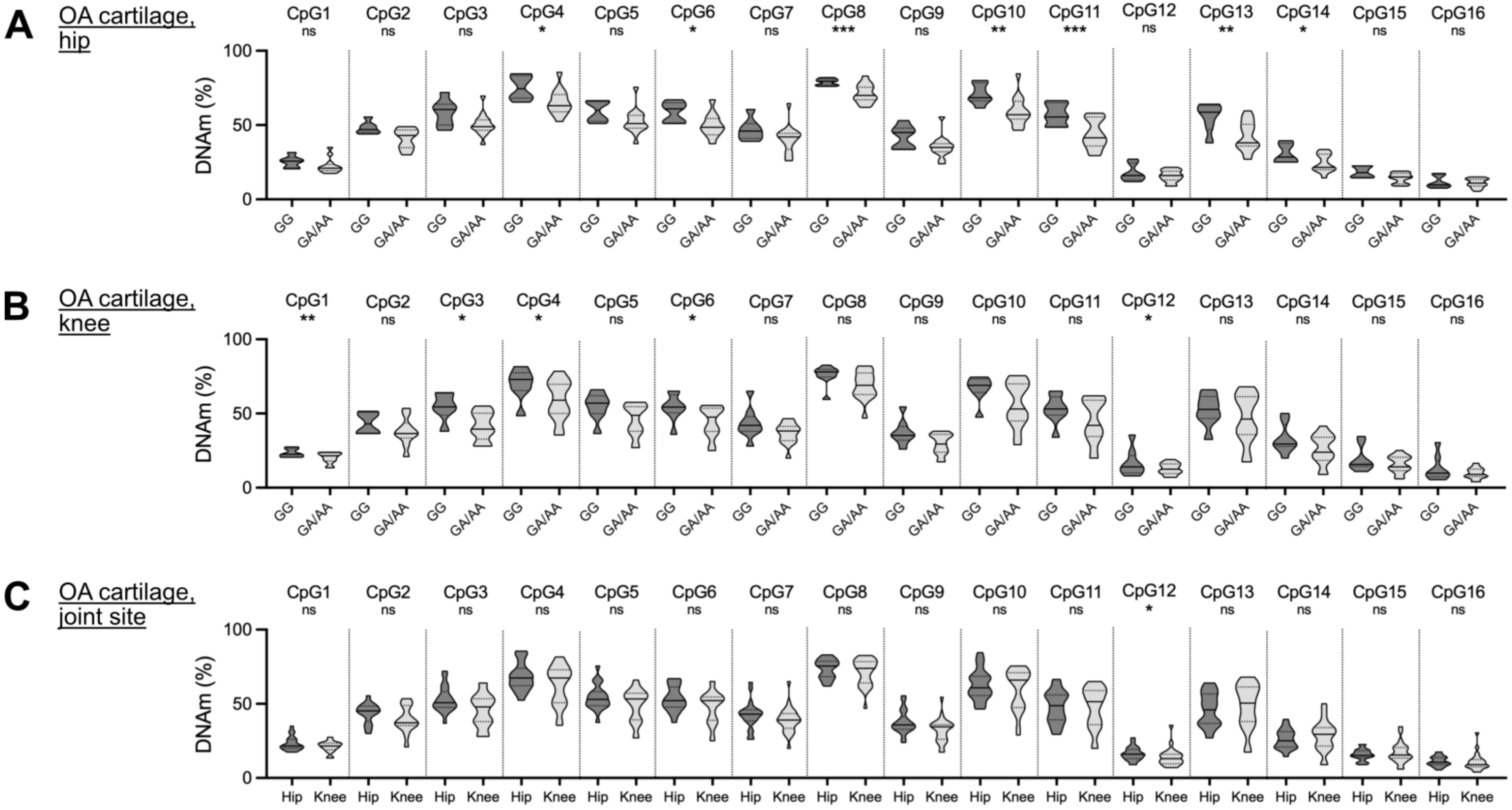




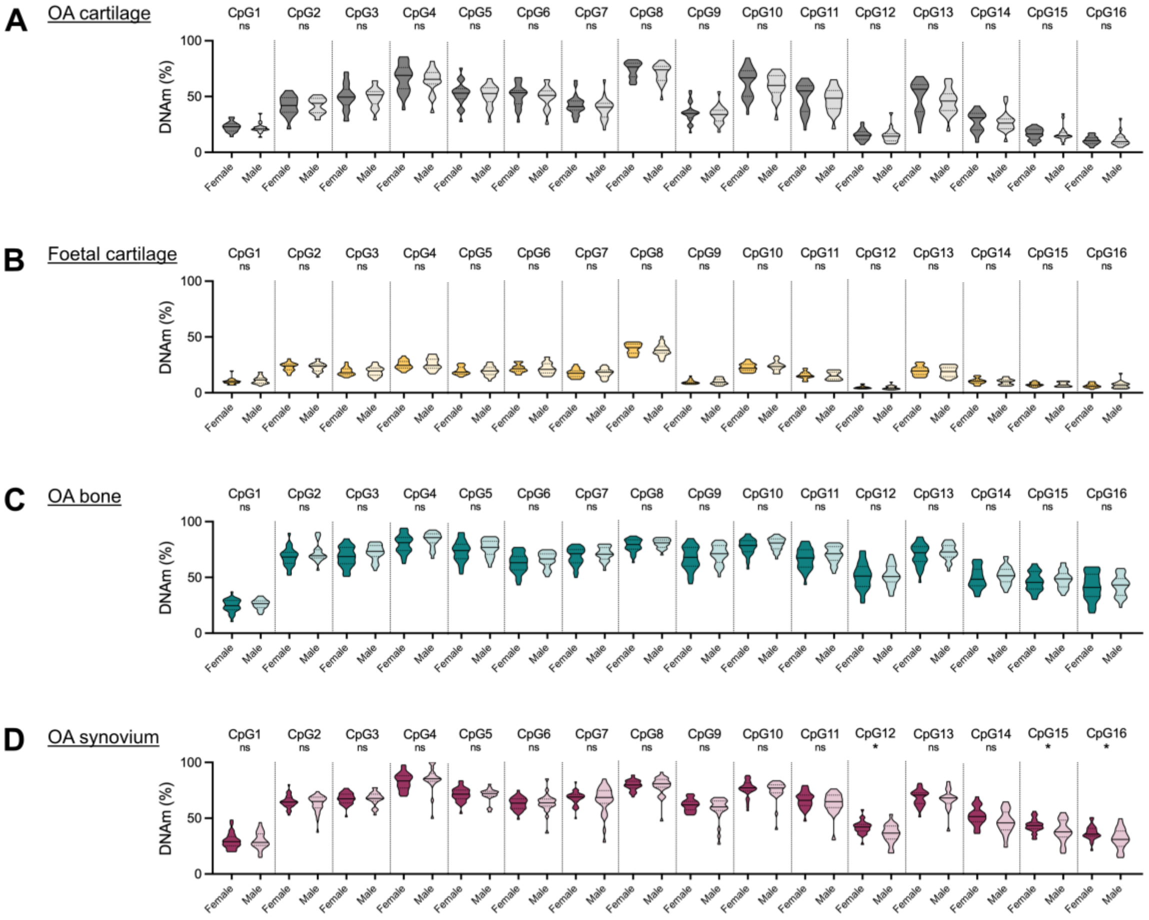
Appendix K. mQTL analysis stratified by sex in OA cartilage (panels **A** and **B**, in grey) and in foetal cartilage (panels **C** and **D**, in yellow). For all panels, DNAm levels at the 16 CpGs were stratified by rs34195470 genotype (GG, GA, AA). Heterozygotes (GA) and minor allele homozygotes (AA) were plotted together due to low number of AAs per CpG (n<3). Methylation data is plotted in the form of β-values ranging from 0 (no methylation) to 1 (complete methylation) and expressed as a percentage. In the truncated violin plots, solid and dashed horizontal lines represent the median and interquartile range. *P*-values calculated by least squares linear regression. * = $P < 0.05$; ** = $P < 0.01$; ns = not significant ($P > 0.05$).



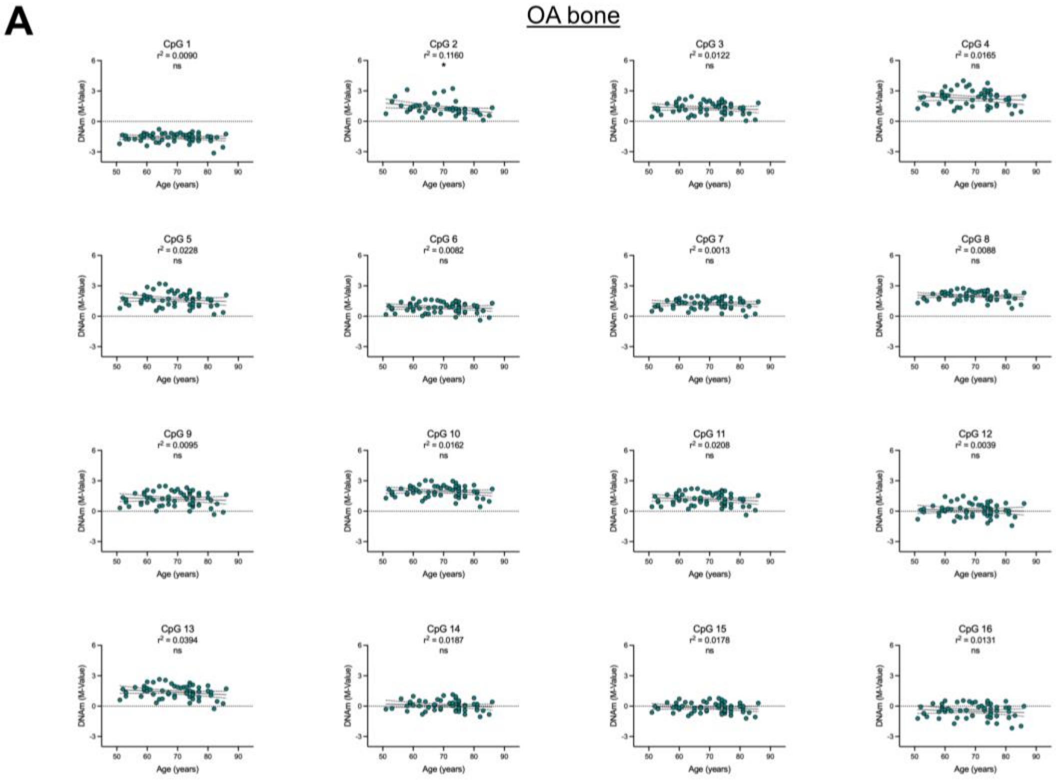
Appendix L. mQTL analysis stratified by joint site in OA cartilage (in hip, panel **A** and in knee, panel **B**). In both panels, DNAm levels at the 16 CpGs were stratified by rs34195470 genotype (GG, GA, AA). Heterozygotes (GA) and minor allele homozygotes (AA) were plotted together due to low number of AAs per CpG (n<3). *P*-values calculated by least squares linear regression. * = *P* < 0.05; ** = *P* < 0.01; *** = *P* < 0.001; ns = not significant (*P* > 0.05). **(C)** Stratification of DNAm levels at the 16 CpGs by joint site in OA cartilage irrespective of rs34195470 genotype. *P*-values calculated using Mann Whitney U test. * = *P* < 0.05; ns = not significant (*P* > 0.05). For all panels, methylation data is plotted in the form of β-values ranging from 0 (no methylation) to 1 (complete methylation) and expressed as a percentage. In the truncated violin plots, solid and dashed horizontal lines represent the median and interquartile range.

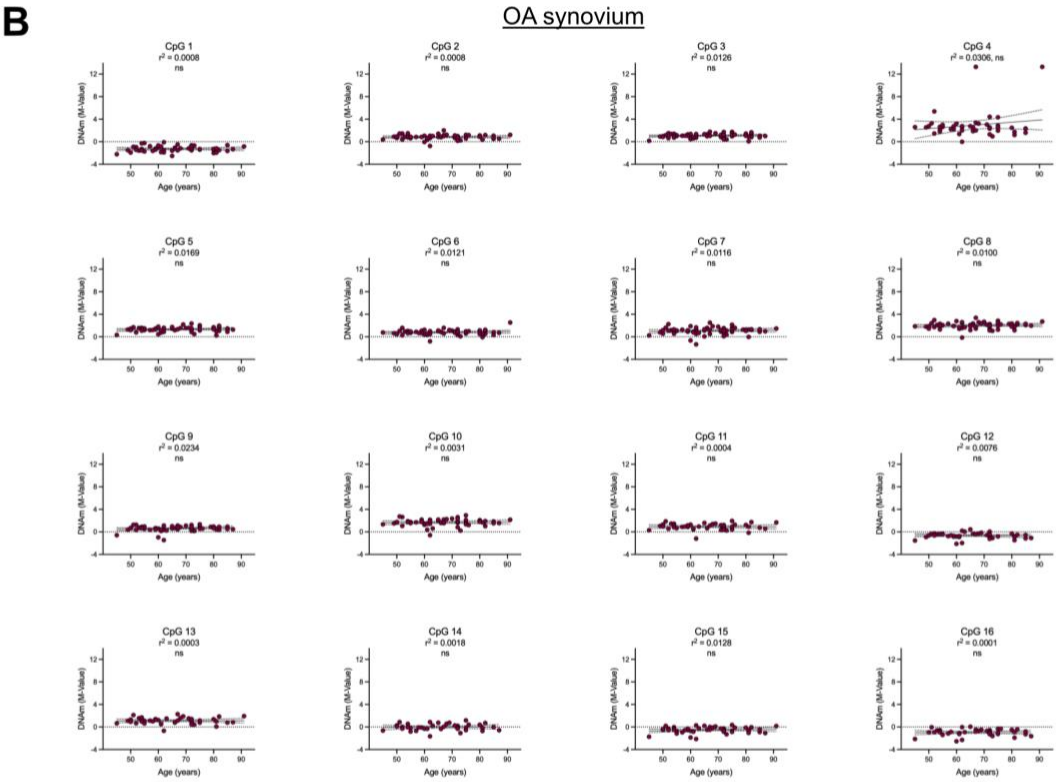


Appendix M. Stratification of DNAm levels at the 16 CpGs by sex in OA cartilage (in grey, panel **A**), foetal cartilage (in yellow, panel **B**), OA bone (in teal, panel **C**), and OA synovium (in maroon, panel **D**) irrespective of rs34195470 genotype. *P*-values calculated using Mann Whitney U test. * = *P* < 0.05; ns = not significant (*P* > 0.05). For all panels, methylation data is plotted in the form of β-values ranging from 0 (no methylation) to 1 (complete methylation) and expressed as a percentage. In the truncated violin plots, solid and dashed horizontal lines represent the median and interquartile range.

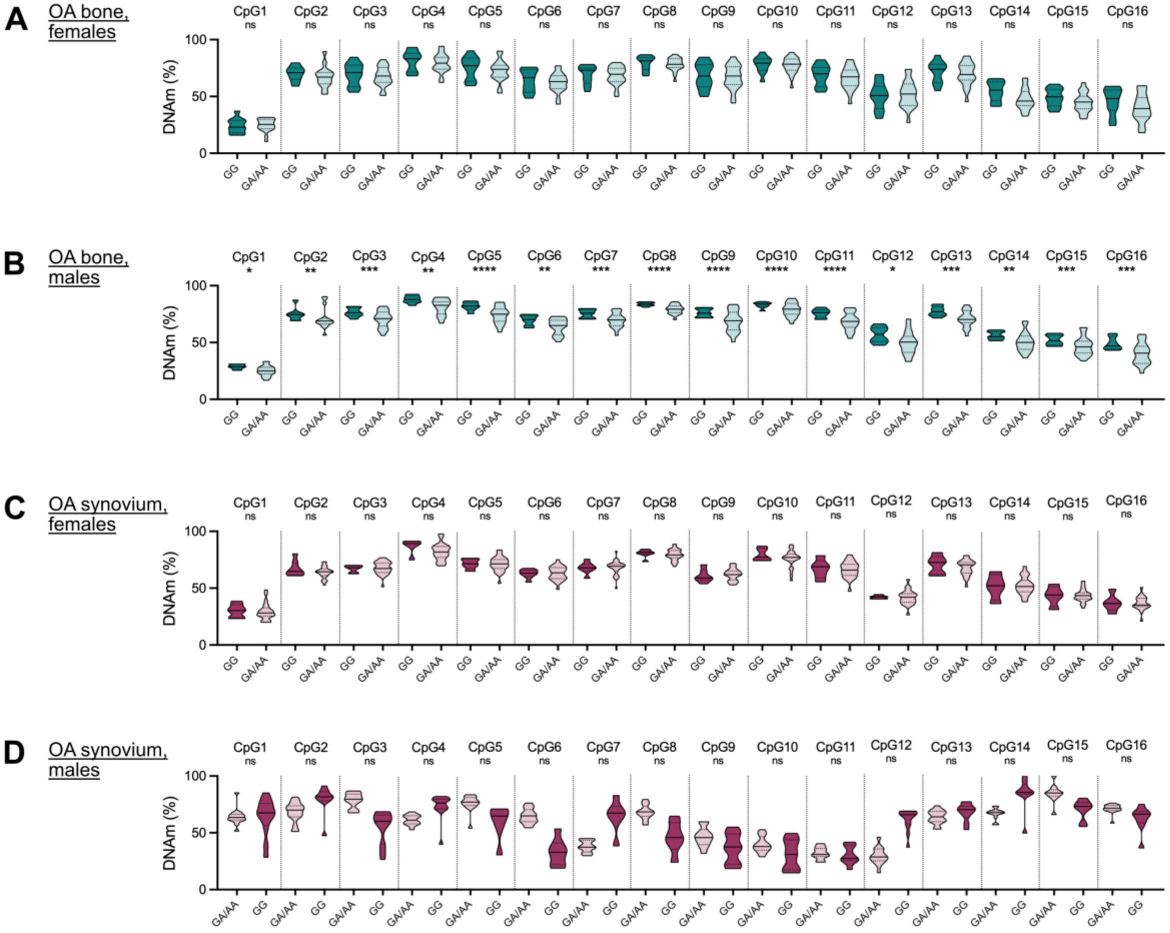


Appendix N. Stratification of DNAm levels at the 16 CpGs by age in OA bone (panel **A**, in teal) and OA synovium (panel **B**, in maroon). For both panels, the dotted line at *y*=0 represents 50% DNAm. Coloured circles represent individual samples. *P*-values and r^2 values calculated using simple linear regression. The solid line represents the regression line with 95% confidence intervals represented by dashed lines. * = $P < 0.05$; ns = not significant ($P > 0.05$).





Appendix O. mQTL analysis stratified by sex in OA bone (panels **A** and **B**, in teal) and in OA synovium (panels **C** and **D**, in maroon). For all panels, DNAm levels at the 16 CpGs were stratified by rs34195470 genotype (GG, GA, AA). Heterozygotes (GA) and minor allele homozygotes (AA) were plotted together due to low number of AAs per CpG (n<3). Methylation data is plotted in the form of β-values ranging from 0 (no methylation) to 1 (complete methylation) and expressed as a percentage. In the truncated violin plots, solid and dashed horizontal lines represent the median and interquartile range. *P*-values calculated by least squares linear regression. * = *P* < 0.05; ** = *P* < 0.01; *** = *P* < 0.001; **** = *P* < 0.0001; ns = not significant (*P* > 0.05).



References

1. Lefebvre, V. & Bhattaram, P. Vertebrate Skeletogenesis. *Curr Top Dev Biol* **90**, 291–317 (2010).
2. Berendsen, A. D. & Olsen, B. R. Bone development. *Bone* **80**, 14–18 (2015).
3. Muragaki, Y., Mundlos, S., Upton, J. & Olsen, B. R. Altered Growth and Branching Patterns in Synpolydactyly Caused by Mutations in HOXD13. *Science (1979)* **272**, 548–551 (1996).
4. Zlotogora, J., Lerer, I., Bar-David, S., Ergaz, Z. & Abeliovich, D. Homozygosity for Waardenburg syndrome. *Am J Hum Genet* **56**, 1173 (1995).
5. Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A. & De Crombrugghe, B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev* **16**, 2813–2828 (2002).
6. Kawakami, Y., Rodriguez-León, J. & Belmonte, J. C. I. The role of TGFβs and Sox9 during limb chondrogenesis. *Curr Opin Cell Biol* **18**, 723–729 (2006).
7. Mak, K. K., Kronenberg, H. M., Chuang, P. T., Mackem, S. & Yang, Y. Indian hedgehog signals independently of PTHrP to promote chondrocyte hypertrophy. *Development* **135**, 1947–1956 (2008).
8. Kronenberg, H. M. Developmental regulation of the growth plate. *Nature* **423**, 332–336 (2003).
9. Finnegan, M. A. & Uhthoff, H. K. The Development of the Knee. *The Embryology of the Human Locomotor System* **1**, 129–140 (1990).
10. Decker, R. S. Articular cartilage and joint development from embryogenesis to adulthood. *Semin Cell Dev Biol* **62**, 50 (2017).
11. Holder, N. An experimental investigation into the early development of the chick elbow joint. *J Embryol Exp Morphol* **39**, 115–127 (1977).
12. Thomas, J. T. *et al.* A human chondrodysplasia due to a mutation in a TGF-β superfamily member. *Nat Genet* **12**, 315–317 (1996).
13. Pitsillides, A. A. & Beier, F. Cartilage biology in osteoarthritis—lessons from developmental biology. *Nat Rev Rheumatol* **7**, 654–663 (2011).

14. Bastow, E. R. *et al.* Selective Activation of the MEK-ERK Pathway Is Regulated by Mechanical Stimuli in Forming Joints and Promotes Pericellular Matrix Formation. *Journal of Biological Chemistry* **280**, 11749–11758 (2005).
15. Dy, P. *et al.* Synovial joint morphogenesis requires the chondrogenic action of Sox5 and Sox6 in growth plate and articular cartilage. *Dev Biol* **341**, 346–359 (2010).
16. Pacifici, M., Koyama, E. & Iwamoto, M. Mechanisms of synovial joint and articular cartilage formation: Recent advances, but many lingering mysteries. *Birth Defects Res C Embryo Today* **75**, 237–248 (2005).
17. Chijimatsu, R. & Saito, T. Mechanisms of synovial joint and articular cartilage development. *Cellular and Molecular Life Sciences* **76**, 3939–3952 (2019).
18. Mackie, E. J., Ahmed, Y. A., Tatarczuch, L., Chen, K. S. & Mirams, M. Endochondral ossification: How cartilage is converted into bone in the developing skeleton. *Int J Biochem Cell Biol* **40**, 46–62 (2008).
19. Lefebvre, V. & Smits, P. Transcriptional control of chondrocyte fate and differentiation. *Birth Defects Res C Embryo Today* **75**, 200–212 (2005).
20. Cawston, T. E. & Wilson, A. J. Understanding the role of tissue degrading enzymes and their inhibitors in development and disease. *Best Pract Res Clin Rheumatol* **20**, 983–1002 (2006).
21. Zhang, F. J., Luo, W. & Lei, G. H. Role of HIF-1α and HIF-2α in osteoarthritis. *Joint Bone Spine* **82**, 144–147 (2015).
22. Goldring, M. B. & Marcu, K. B. Cartilage homeostasis in health and rheumatic diseases. *Arthritis Res Ther* **11**, 1–16 (2009).
23. Martin, J. A., Brown, T. D., Heiner, A. D. & Buckwalter, J. A. Chondrocyte senescence, joint loading and osteoarthritis. *Clin Orthop Relat Res* **427**, (2004).
24. Sophia Fox, A. J., Bedi, A. & Rodeo, S. A. The Basic Science of Articular Cartilage: Structure, Composition, and Function. *Sports Health* **1**, 461 (2009).
25. Maroudas A. Physiochemical properties of articular cartilage. in *Adult Articular Cartilage* (ed. Freeman MAR) 215–290 (Cambridge University Press, Kent, 1979).
26. Jay, G. D. & Waller, K. A. The biology of Lubricin: Near frictionless joint motion. *Matrix Biology* **39**, 17–24 (2014).

27. Cheng, C., Conte, E., Pleshko-Camacho, N. & Hidaka, C. Differences in matrix accumulation and hypertrophy in superficial and deep zone chondrocytes are controlled by bone morphogenetic protein. *Matrix Biology* **26**, 541–553 (2007).
28. Goldring MB. Chapter 3: cartilage and chondrocytes. in *Kelley's Textbook of Rheumatology* (eds. Firestein GS et al.) vol. 8 37–69 (WB Saunders, an imprint of Elsevier Inc, Philadelphia, 2008).
29. Poole AR. Cartilage in health and disease. in *Arthritis and Allied Conditions: A Textbook of Rheumatology* (ed. Koopman WS) vol. 15 223–269 (Lippincott, Williams, and Wilkins, Philadelphia, 2005).
30. Burr, D. B. & Gallant, M. A. Bone remodelling in osteoarthritis. *Nat Rev Rheumatol* **8**, 665–673 (2012).
31. Milz, S. & Putz, R. Quantitative morphology of the subchondral plate of the tibial plateau. *J Anat* **185**, 103 (1994).
32. Holmdahl, D. E. & Ingelmark, B. E. The Contact Between the Articular Cartilage and the Medullary Cavities of the Bone. *Acta Orthopaedica Scandinavica* 156–165 (2009).
33. Feng, X. & McDonald, J. M. Disorders of bone remodeling. *Annual Review of Pathology: Mechanisms of Disease* **6**, 121–145 (2011).
34. Palumbo, C. & Ferretti, M. The Osteocyte: From "Prisoner" to "Orchestrator". *J Funct Morphol Kinesiol* **6**, 28 (2021).
35. Tu, X. *et al.* Sost downregulation and local Wnt signaling are required for the osteogenic response to mechanical loading. *Bone* **50**, 209–217 (2012).
36. Zhen, G. *et al.* Inhibition of TGF-β signaling in mesenchymal stem cells of subchondral bone attenuates osteoarthritis. *Nat Med* **19**, 704–712 (2013).
37. D. Smith, M. Suppl 1: The Normal Synovium. *Open Rheumatol J* **5**, 100 (2011).
38. Sanchez-Lopez, E., Coras, R., Torres, A., Lane, N. E. & Guma, M. Synovial inflammation in osteoarthritis progression. *Nat Rev Rheumatol* **18**, 258–275 (2022).
39. Flandry, F. & Hommel, G. Normal anatomy and biomechanics of the knee. *Sports Med Arthrosc Rev* **19**, 82–92 (2011).
40. Markes, A. R., Hodax, J. D. & Ma, C. B. Meniscus Form and Function. *Clin Sports Med* **39**, 1–12 (2020).

41. Roessler, P. P., Schüttler, K. F., Heyse, T. J., Wirtz, D. C. & Efe, T. The anterolateral ligament (ALL) and its role in rotational extra-articular stability of the knee joint: a review of anatomy and surgical concepts. *Arch Orthop Trauma Surg* **136**, 305–313 (2016).

42. Loeser, R. F., Goldring, S. R., Scanzello, C. R. & Goldring, M. B. Osteoarthritis: A disease of the joint as an organ. *Arthritis Rheum* **64**, 1697–1707 (2012).

43. Kraus, V. B., Blanco, F. J., Englund, M., Karsdal, M. A. & Lohmander, L. S. Call for standardized definitions of osteoarthritis and risk stratification for clinical trials and clinical use. *Osteoarthritis Cartilage* **23**, 1233–1241 (2015).

44. Katz, J. N., Arant, K. R. & Loeser, R. F. Diagnosis and Treatment of Hip and Knee Osteoarthritis: A Review. *JAMA* **325**, 568–578 (2021).

45. Steinmetz, J. D. *et al.* Global, regional, and national burden of osteoarthritis, 1990–2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Rheumatol* **5**, e508–e522 (2023).

46. Andriacchi, T. P. *et al.* A framework for the in vivo pathomechanics of osteoarthritis at the knee. *Ann Biomed Eng* **32**, 447–457 (2004).

47. Eckstein, F. *et al.* Magnitude and regional distribution of cartilage loss associated with grades of joint space narrowing in radiographic osteoarthritis – data from the Osteoarthritis Initiative (OAI). *Osteoarthritis Cartilage* **18**, 760–768 (2010).

48. Hellingman, C. A. *et al.* Smad signaling determines chondrogenic differentiation of bone-marrow-derived mesenchymal stem cells: inhibition of Smad1/5/8P prevents terminal differentiation and calcification. *Tissue Eng Part A* **17**, 1157–1167 (2011).

49. van der Kraan, P. M. The changing role of TGFβ in healthy, ageing and osteoarthritic joints. *Nat Rev Rheumatol* **13**, 155–163 (2017).

50. De Kroon, L. M. G. *et al.* Activin Receptor-Like Kinase Receptors ALK5 and ALK1 Are Both Required for TGFβ-Induced Chondrogenic Differentiation of Human Bone Marrow-Derived Mesenchymal Stem Cells. *PLoS One* **10**, (2015).

51. Van De Laar, I. M. B. H. *et al.* Mutations in SMAD3 cause a syndromic form of aortic aneurysms and dissections with early-onset osteoarthritis. *Nat Genet* **43**, 121–126 (2011).

52. Yao, J. Y. *et al.* Mutation analysis of the Smad3 gene in human osteoarthritis. *Eur J Hum Genet* **11**, 714–717 (2003).

53. Yang, X. *et al.* TGF-beta/Smad3 signals repress chondrocyte hypertrophic differentiation and are required for maintaining articular cartilage. *J Cell Biol* **153**, 35–46 (2001).

54. Wu, Q. *et al.* Induction of an osteoarthritis-like phenotype and degradation of phosphorylated Smad3 by Smurf2 in transgenic mice. *Arthritis Rheum* **58**, 3132–3144 (2008).

55. Govinden, R. & Bhoola, K. D. Genealogy, expression, and cellular function of transforming growth factor-β. *Pharmacol Ther* **98**, 257–265 (2003).

56. Li, T. F. *et al.* Smad3-deficient chondrocytes have enhanced BMP signaling and accelerated differentiation. *Journal of bone and mineral research* **21**, 4–16 (2006).

57. Ferguson, C. M. *et al.* Smad2 and 3 mediate transforming growth factor-beta1-induced inhibition of chondrocyte maturation. *Endocrinology* **141**, 4728–4735 (2000).

58. Sanford, L. P. *et al.* TGFβ2 knockout mice have multiple developmental defects that are non-overlapping with other TGFβ knockout phenotypes. *Development* **124**, 2659–2670 (1997).

59. Mueller, M. B. *et al.* Hypertrophy in Mesenchymal Stem Cell Chondrogenesis: Effect of TGF-β Isoforms and Chondrogenic Conditioning. *Cells Tissues Organs* **192**, 158–166 (2010).

60. Davidson, E. N. B., Scharstuhl, A., Vitters, E. L., van der Kraan, P. M. & van den Berg, W. B. Reduced transforming growth factor-beta signaling in cartilage of old mice: role in impaired repair capacity. *Arthritis Res Ther* **7**, 1–10 (2005).

61. Ashraf, S., Mapp, P. I. & Walsh, D. A. Contributions of angiogenesis to inflammation, joint damage, and pain in a rat model of osteoarthritis. *Arthritis Rheum* **63**, 2700–2710 (2011).

62. Walsh, D. A. *et al.* Angiogenesis and nerve growth factor at the osteochondral junction in rheumatoid arthritis and osteoarthritis. *Rheumatology (Oxford)* **49**, 1852–1861 (2010).

63. Walsh, D. A. *et al.* Angiogenesis in the synovium and at the osteochondral junction in osteoarthritis. *Osteoarthritis Cartilage* **15**, 743–751 (2007).

64. Mansell, J. P., Collins, C. & Bailey, A. J. Bone, not cartilage, should be the major focus in osteoarthritis. *Nat Clin Pract Rheumatol* **3**, 306–307 (2007).

65. Day, J. S. *et al.* A decreased subchondral trabecular bone tissue elastic modulus is associated with pre-arthritic cartilage damage. *Journal of Orthopaedic Research* **19**, 914–918 (2001).
66. van der Kraan, P. M. & van den Berg, W. B. Osteophytes: relevance and biology. *Osteoarthritis Cartilage* **15**, 237–244 (2007).
67. Ene, R., Sinescu, R., Ene, P., Cirstoiu, M. & Cirstoiu, F. Synovial inflammation in patients with different stages of knee osteoarthritis. *Rom J Morphol Embryol* (2015).
68. Benito, M. J., Veale, D. J., FitzGerald, O., Van Den Berg, W. B. & Bresnihan, B. Synovial tissue inflammation in early and late osteoarthritis. *Ann Rheum Dis* **64**, 1263–1267 (2005).
69. Scanzello, C. R. *et al.* Synovial inflammation in patients undergoing arthroscopic meniscectomy: Molecular characterization and relationship to symptoms. *Arthritis Rheum* **63**, 391–400 (2011).
70. Swärd, P., Frobell, R., Englund, M., Roos, H. & Struglics, A. Cartilage and bone markers and inflammatory cytokines are increased in synovial fluid in the acute phase of knee injury (hemarthrosis)--a cross-sectional analysis. *Osteoarthritis Cartilage* **20**, 1302–1308 (2012).
71. Manferdini, C. *et al.* From osteoarthritic synovium to synovialderived cells characterization: Synovial macrophages are key effector cells. *Arthritis Res Ther* **18**, 1–13 (2016).
72. Bondeson, J. *et al.* The role of synovial macrophages and macrophage-produced mediators in driving inflammatory and destructive responses in osteoarthritis. *Arthritis Rheum* **62**, 647–657 (2010).
73. Temple-Wong, M. M. *et al.* Hyaluronan concentration and size distribution in human knee synovial fluid: variations with age and cartilage degeneration. *Arthritis Res Ther* **18**, (2016).
74. Hamasaki, M. *et al.* Transcriptional profiling of murine macrophages stimulated with cartilage fragments revealed a strategy for treatment of progressive osteoarthritis. *Sci Rep* **10**, 7558 (2020).
75. Brown, R. A. *et al.* Relationship of angiogenesis factor in synovial fluid to various joint diseases. *Ann Rheum Dis* **42**, 301–307 (1983).

76. Scanzello, C. R. *et al.* Local cytokine profiles in knee osteoarthritis: elevated synovial fluid interleukin-15 differentiates early from end-stage disease. *Osteoarthritis Cartilage* **17**, 1040–1048 (2009).

77. Constantinescu, C. S., Grygar, C., Kappos, L. & Leppert, D. Interleukin 15 stimulates production of matrix metalloproteinase-9 and tissue inhibitor of metalloproteinase-1 by human peripheral blood mononuclear cells. *Cytokine* **13**, 244–247 (2001).

78. Rose, B. J. & Kooyman, D. L. A Tale of Two Joints: The Role of Matrix Metalloproteases in Cartilage Biology. *Dis Markers* **2016**, 4895050 (2016).

79. Honorati, M. C., Bovara, M., Cattini, L., Piacentini, A. & Facchini, A. Contribution of interleukin 17 to human cartilage degradation and synovial inflammation in osteoarthritis. *Osteoarthritis Cartilage* **10**, 799–807 (2002).

80. Sousa-Valente, J. *et al.* Role of TrkA signalling and mast cells in the initiation of osteoarthritis pain in the monoiodoacetate model. *Osteoarthritis Cartilage* **26**, 84–94 (2018).

81. Brown, T. D., Johnston, R. C., Saltzman, C. L., Marsh, J. L. & Buckwalter, J. A. Posttraumatic osteoarthritis: A first estimate of incidence, prevalence, and burden of disease. *J Orthop Trauma* **20**, 739–744 (2006).

82. Ashraf, S. *et al.* Increased vascular penetration and nerve growth in the meniscus: a potential source of pain in osteoarthritis. *Ann Rheum Dis* **70**, 523–529 (2011).

83. Katsuragawa, Y. *et al.* Changes of human menisci in osteoarthritic knee joints. *Osteoarthritis Cartilage* **18**, 1133–1143 (2010).

84. Mullaji, A. B., Marawar, S. V., Simha, M. & Jindal, G. Cruciate ligaments in arthritic knees: a histologic study with radiologic correlation. *J Arthroplasty* **23**, 567–572 (2008).

85. Watanabe, A., Kanamori, A., Ikeda, K. & Ochiai, N. Histological evaluation and comparison of the anteromedial and posterolateral bundle of the human anterior cruciate ligament of the osteoarthritic knee joint. *Knee* **18**, 47–50 (2011).

86. Hunter, D. J., March, L. & Chew, M. Osteoarthritis in 2020 and beyond: a Lancet Commission. *The Lancet* **396**, 1711–1712 (2020).

87. Aubourg, G., Rice, S. J., Bruce-Wootton, P. & Loughlin, J. Genetics of osteoarthritis. *Osteoarthritis Cartilage* **30**, 636–649 (2022).

88. Roberts, J. B. & Rice, S. J. Osteoarthritis as an Enhanceropathy: Gene Regulation in Complex Musculoskeletal Disease. *Curr Rheumatol Rep* **26**, 1–13 (2024).

89. Boer, C. G. *et al.* Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell* **184**, 4784–4818 (2021).

90. Hunter, D. J., Schofield, D. & Callander, E. The individual and socioeconomic impact of osteoarthritis. *Nat Rev Rheumatol* **10**, 437–441 (2014).

91. Nüesch, E. *et al.* All cause and disease specific mortality in patients with knee or hip osteoarthritis: population based cohort study. *BMJ* **342**, 638 (2011).

92. March, L. M. & Bachmeier, C. J. M. Economics of osteoarthritis: a global perspective. *Baillieres Clin Rheumatol* **11**, 817–834 (1997).

93. Leifer, V. P., Katz, J. N. & Losina, E. The burden of OA-health services and economics. *Osteoarthritis Cartilage* **30**, 10–16 (2022).

94. Alami, S. *et al.* Patients' and Practitioners' Views of Knee Osteoarthritis and Its Management: A Qualitative Interview Study. *PLoS One* **6**, e19634 (2011).

95. Cheng, O. T., Souzdalnitski, D., Vrooman, B. & Cheng, J. Evidence-Based Knee Injections for the Management of Arthritis. *Pain Medicine (United States)* **13**, 740–753 (2012).

96. Bannuru, R. R. *et al.* OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. *Osteoarthritis Cartilage* **27**, 1578–1589 (2019).

97. Domper Arnal, M. J., Hijos-Mallada, G. & Lanas, A. Gastrointestinal and cardiovascular adverse events associated with NSAIDs. *Expert Opin Drug Saf* **21**, 373–384 (2022).

98. Zeng, C. *et al.* Intra-articular corticosteroids and the risk of knee osteoarthritis progression: results from the Osteoarthritis Initiative. *Osteoarthritis Cartilage* **27**, 855–862 (2019).

99. McAlindon, T. E. *et al.* Effect of Intra-articular Triamcinolone vs Saline on Knee Cartilage Volume and Pain in Patients With Knee Osteoarthritis: A Randomized Clinical Trial. *JAMA* **317**, 1967–1975 (2017).

100. Memtsoudis, S. G. *et al.* Risk Factors for Perioperative Mortality After Lower Extremity Arthroplasty: A Population-Based Study of 6,901,324 Patient Discharges. *J Arthroplasty* **25**, 19–26 (2010).

101. Evans, J. T. *et al.* How long does a hip replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. *The Lancet* **393**, 647–654 (2019).

102. Evans, J. T. *et al.* How long does a knee replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. *The Lancet* **393**, 655–663 (2019).

103. Katz, J. N. *et al.* Twelve-year risk of revision after primary total hip replacement in the U.S. medicare population. *Journal of Bone and Joint Surgery* **94**, 1825–1832 (2012).

104. Beswick, A. D., Wylde, V., Gooberman-Hill, R., Blom, A. & Dieppe, P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open* **2**, e000435 (2012).

105. Chen, A., Gupte, C., Akhtar, K., Smith, P. & Cobb, J. The Global Economic Cost of Osteoarthritis: How the UK Compares. *Arthritis* **2012**, 698709 (2012).

106. Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. REVIEW 15 years of GWAS discovery: Realizing the promise. *The American Journal of Human Genetics* **110**, 179–194 (2023).

107. Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *The American Journal of Human Genetics* **102**, 717–730 (2018).

108. Yengo, L. *et al.* A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).

109. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237–244 (2019).

110. Scott, R. A. *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).

111. van de Bunt, M., Cortes, A., Brown, M. A., Morris, A. P. & McCarthy, M. I. Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci. *PLoS Genet* **11**, (2015).

112. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* **32**, 381–385 (2008).

113. Manolio, T. A. Genomewide Association Studies and Assessment of the Risk of Disease. *New England Journal of Medicine* **363**, 166–176 (2010).

114. Fatumo, S. *et al.* A roadmap to increase diversity in genomic studies. *Nat Med* **28**, 243–250 (2022).

115. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science (1979)* **337**, 1190–1195 (2012).

116. Blackwood, E. M. & Kadonaga, J. T. Going the distance: a current view of enhancer action. *Science* **281**, 60–63 (1998).

117. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science (1979)* **296**, 2225–2229 (2002).

118. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369–375 (2012).

119. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491–504 (2018).

120. Carlin, B. P. & Louis, T. A. *Bayesian Methods for Data Analysis*. *Bayesian Methods for Data Analysis, Third Edition* (Taylor and Francis, NewYork, 2008).

121. Tachmazidou, I. *et al.* Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. *Nat Genet* **51**, 230–236 (2019).

122. Henkel, C. *et al.* Genome-wide association meta-analysis of knee and hip osteoarthritis uncovers genetic differences between patients treated with joint replacement and patients without joint replacement. *Ann Rheum Dis* **82**, 384–392 (2023).

123. Styrkarsdottir, U. *et al.* Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11, COL11A1 and 13 more new loci associated with osteoarthritis. *Nat Genet* **50**, 1681–1687 (2018).

124. Zeggini, E. *et al.* Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* **380**, 815–823 (2012).

125. Dempster, E. R. & Lerner, I. M. Heritability of Threshold Characters. *Genetics* **35**, 212 (1950).

126. Miyamoto, Y. *et al.* A functional polymorphism in the 5′ UTR of GDF5 is associated with susceptibility to osteoarthritis. *Nat Genet* **39**, 529–533 (2007).

127. Richard, D. *et al.* Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk. *Cell* **181**, 362–381 (2020).

128. Capellini, T. D. *et al.* Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. *Nat Genet* **49**, 1202–1210 (2017).

129. Byars, S. G. & Voskarides, K. Antagonistic Pleiotropy in Human Disease. *J Mol Evol* **88**, 12–25 (2020).

130. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487–500 (2016).

131. Rice, S. J., Beier, F., Young, D. A. & Loughlin, J. Interplay between genetics and epigenetics in osteoarthritis. *Nat Rev Rheumatol* **16**, 268–281 (2020).

132. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243–278 (2023).

133. Villicaña, S. & Bell, J. T. Genetic impacts on DNA methylation: research findings and future perspectives. *Genome Biol* **22**, 127 (2021).

134. Lyko, F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet* **19**, 81–92 (2017).

135. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet* **18**, 517–534 (2017).

136. Richa, R. & Sinha, R. P. Hydroxymethylation of DNA: an epigenetic marker. *EXCLI J* **13**, 592 (2014).

137. Taylor, S. E. B., Smeriglio, P., Dhulipala, L., Rath, M. & Bhutani, N. A global increase in 5-hydroxymethylcytosine levels marks osteoarthritic chondrocytes. *Arthritis & Rheumatology* **66**, 90–100 (2014).

138. Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* **8**, 1499 (1980).

139. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010–1022 (2011).

140. Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).

141. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381–395 (2011).

142. Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications - writers that read. *EMBO Rep* **16**, 1467–1481 (2015).

143. Mattick, J. S. *et al.* Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol* **24**, 430–447 (2023).

144. Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**, 281–297 (2004).

145. Czech, B. & Hannon, G. J. Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet* **12**, 19–31 (2010).

146. Bartel, D. P. Leading Edge Review Metazoan MicroRNAs. *Cell* **173**, 20–51 (2018).

147. Shang, R., Lee, S., Senavirathne, G. & Lai, E. C. microRNAs in action: biogenesis, function and regulation. *Nat Rev Genet* **24**, 816–833 (2023).

148. Imagawa, K. *et al.* Association of reduced type IX collagen gene expression in human osteoarthritic chondrocytes with epigenetic silencing by DNA hypermethylation. *Arthritis Rheumatol* **66**, 3040–3051 (2014).

149. Zimmermann, P. *et al.* Correlation of COL10A1 induction during chondrogenesis of mesenchymal stem cells with demethylation of two CpG sites in the COL10A1 promoter. *Arthritis Rheum* **58**, 2743–2753 (2008).

150. Bui, C. *et al.* cAMP response element-binding (CREB) recruitment following a specific CpG demethylation leads to the elevated expression of the matrix metalloproteinase 13 in human articular chondrocytes and osteoarthritis. *FASEB J* **26**, 3000–3011 (2012).

151. Cheung, K. S. C., Hashimoto, K., Yamada, N. & Roach, H. I. Expression of ADAMTS-4 by chondrocytes in the surface zone of human osteoarthritic cartilage is regulated by epigenetic DNA de-methylation. *Rheumatol Int* **29**, 525–534 (2009).

152. Hashimoto, K., Oreffo, R. O. C., Gibson, M. B., Goldring, M. B. & Roach, H. I. DNA De-methylation at Specific CpG Sites in the IL1B Promoter in Response to Inflammatory Cytokines in Human Articular Chondrocytes. *Arthritis Rheum* **60**, 3303 (2009).

153. Ezura, Y., Sekiya, I., Koga, H., Muneta, T. & Noda, M. Methylation status of CpG islands in the promoter regions of signature genes during chondrogenesis of human synovium-derived mesenchymal stem cells. *Arthritis Rheum* **60**, 1416–1426 (2009).

154. Reynard, L. N., Bui, C., Canty-laird, E. G., Young, D. A. & Loughlin, J. Expression of the osteoarthritis-associated gene GDF5 is modulated epigenetically by DNA methylation. *Hum Mol Genet* **20**, 3450–3460 (2011).

155. Loeser, R. F., Im, H. J., Richardson, B., Lu, Q. & Chubinskaya, S. Methylation of the OP-1 promoter: potential role in the age-related decline in OP-1 expression in cartilage. *Osteoarthritis Cartilage* **17**, 513–517 (2009).

156. Jeffries, M. A. *et al.* Genome-wide DNA methylation study identifies significant epigenomic changes in osteoarthritic cartilage. *Arthritis Rheumatol* **66**, 2804–2815 (2014).

157. Rushton, M. D. *et al.* Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. *Arthritis & Rheumatology* **66**, 2450–2460 (2014).

158. Kreitmaier, P. *et al.* An epigenome-wide view of osteoarthritis in primary tissues. *The American Journal of Human Genetics* **109**, 1255–1271 (2022).

159. McDonnell, E. *et al.* Epigenetic mechanisms of osteoarthritis risk in human skeletal development. *medRxiv* 2024.05.05.24306832 (2024) doi:10.1101/2024.05.05.24306832.

160. Shen, J. *et al.* DNA methyltransferase 3b regulates articular cartilage homeostasis by altering metabolism. *JCI Insight* **2**, (2017).

161. Jones, M. J., Goodman, S. J. & Kobor, M. S. DNA methylation and healthy human aging. *Aging Cell* **14**, 924–932 (2015).

162. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* **19**, 371–384 (2018).

163. Bell, J. T. *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* **8**, e1002629 (2012).

164. van Dongen, J. *et al.* Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun* **7**, 11115 (2016).

165. Rice, S. J. *et al.* Genetic risk of osteoarthritis operates during human skeletogenesis. *Hum Mol Genet* **32**, 2124–2138 (2023).

166. Parker, E. *et al.* Multi-Tissue Epigenetic and Gene Expression Analysis Combined With Epigenome Modulation Identifies RWDD2B as a Target of Osteoarthritis Susceptibility. *Arthritis & Rheumatology* **73**, 100–109 (2021).

167. Kehayova, Y. S., Wilkinson, J. M., Rice, S. J. & Loughlin, J. Osteoarthritis genetic risk acting on the galactosyltransferase gene COLGALT2 has opposing functional effects in articulating joint tissues. *Arthritis Res Ther* **25**, 83 (2023).

168. Sorial, A. K. *et al.* Multi-tissue epigenetic analysis of the osteoarthritis susceptibility locus mapping to the plectin gene PLEC. *Osteoarthritis Cartilage* **28**, 1448–1458 (2020).

169. Kreitmaier, P. *et al.* Epigenomic profiling of the infrapatellar fat pad in osteoarthritis. *Hum Mol Genet* **33**, 501–509 (2024).

170. Rice, S. J. *et al.* Identification of a novel, methylation-dependent, RUNX2 regulatory region associated with osteoarthritis risk. *Hum Mol Genet* **27**, 3464–3474 (2018).

171. Rice, S. J. *et al.* Genetic and Epigenetic Fine-Tuning of TGFB1 Expression Within the Human Osteoarthritic Joint. *Arthritis & Rheumatology* **73**, 1866–1877 (2021).

172. Brumwell, A. *et al.* Identification of TMEM129, encoding a ubiquitin-protein ligase, as an effector gene of osteoarthritis genetic risk. *Arthritis Res Ther* **24**, 189 (2022).

173. Kehayova, Y. S., Watson, E., Wilkinson, J. M., Loughlin, J. & Rice, S. J. Genetic and Epigenetic Interplay Within a COLGALT2 Enhancer Associated With Osteoarthritis. *Arthritis & Rheumatology* **73**, 1856–1865 (2021).

174. Feigenson, M. *et al.* Histone Deacetylase 3 Deletion in Mesenchymal Progenitor Cells Hinders Long Bone Development. *J Bone Miner Res* **32**, 2453–2465 (2017).

175. Nishimori, S. *et al.* PTHrP targets HDAC4 and HDAC5 to repress chondrocyte hypertrophy. *JCI Insight* **4**, (2019).

176. Bradley, E. W., Carpio, L. R., Olson, E. N. & Westendorf, J. J. Histone Deacetylase 7 (Hdac7) Suppresses Chondrocyte Proliferation and β-Catenin Activity during Endochondral Ossification. *Journal of Biological Chemistry* **290**, 118–126 (2015).

177. Cao, K. *et al.* Decreased histone deacetylase 4 is associated with human osteoarthritis cartilage degeneration by releasing histone deacetylase 4 inhibition of runt-related transcription factor-2 and increasing osteoarthritis-related genes: a novel mechanism of human osteoarthritis cartilage degeneration. *Arthritis Res Ther* **16**, 1–13 (2014).

178. Wang, Y., Zhao, X., Lotz, M., Terkeltaub, R. & Liu-Bryan, R. Mitochondrial biogenesis is impaired in osteoarthritis chondrocytes but reversible via peroxisome proliferator-activated receptor γ coactivator 1α. *Arthritis Rheumatol* **67**, 2141–2153 (2015).

179. Piao, J. *et al.* Sirt6 regulates postnatal growth plate differentiation and proliferation via Ihh signaling. *Sci Rep* **3**, 3022 (2013).

180. Wu, Y. *et al.* Overexpression of Sirtuin 6 suppresses cellular senescence and NF-κB mediated inflammatory responses in osteoarthritis development. *Sci Rep* **5**, 17602 (2015).

181. Nagai, K. *et al.* Depletion of SIRT6 causes cellular senescence, DNA damage, and telomere dysfunction in human chondrocytes. *Osteoarthritis Cartilage* **23**, 1412–1420 (2015).

182. Evangelou, E. *et al.* The DOT1L rs12982744 polymorphism is associated with osteoarthritis of the hip with genome-wide statistical significance in males. *Ann Rheum Dis* **72**, 1264–1265 (2013).

183. Monteagudo, S. *et al.* DOT1L safeguards cartilage homeostasis and protects against osteoarthritis. *Nat Commun* **8**, 15889 (2017).

184. Swingler, T. E. *et al.* The function of microRNAs in cartilage and osteoarthritis. *Clin Exp Rheumatol* **37**, 40–47 (2019).

185. Kobayashi, T. *et al.* Early postnatal ablation of the microRNA-processing enzyme, Drosha, causes chondrocyte death and impairs the structural integrity of the articular cartilage. *Osteoarthritis Cartilage* **23**, 1214–1220 (2015).

186. Kobayashi, T. *et al.* Dicer-dependent pathways regulate chondrocyte proliferation and differentiation. *Proc Natl Acad Sci U S A* **105**, 1949–1954 (2008).

187. Harfe, B. D., McManus, M. T., Mansfield, J. H., Hornstein, E. & Tabin, C. J. The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. *Proc Natl Acad Sci U S A* **102**, 10898–10903 (2005).

188. Zhang, Z. *et al.* MiR-455-3p regulates early chondrogenic differentiation via inhibiting Runx2. *FEBS Lett* **589**, 3671–3678 (2015).

189. Miyaki, S. *et al.* MicroRNA-140 plays dual roles in both cartilage development and homeostasis. *Genes Dev* **24**, 1173–1185 (2010).

190. Le, L. T. T. *et al.* The microRNA-29 family in cartilage homeostasis and osteoarthritis. *J Mol Med (Berl)* **94**, 583 (2016).

191. Swingler, T. E. *et al.* The expression and function of microRNAs in chondrogenesis and osteoarthritis. *Arthritis Rheum* **64**, 1909–1919 (2012).

192. Zhang, W. *et al.* MiR-34a Enhances Chondrocyte Apoptosis, Senescence and Facilitates Development of Osteoarthritis by Targeting DLL1 and Regulating PI3K/AKT Pathway. *Cell Physiol Biochem* **48**, 1304–1316 (2018).

193. Ntoumou, E. *et al.* Serum microRNA array analysis identifies miR-140-3p, miR-33b-3p and miR-671-3p as potential osteoarthritis biomarkers involved in metabolic processes. *Clin Epigenetics* **9**, 1–15 (2017).

194. Zhao, X. *et al.* MicroRNA-495 enhances chondrocyte apoptosis, senescence and promotes the progression of osteoarthritis by targeting AKT1. *Am J Transl Res* **11**, 2232 (2019).

195. Balaskas, P. *et al.* MicroRNA Profiling in Cartilage Ageing. *Int J Genomics* **2017**, 2713725 (2017).

196. Díaz-Prado, S. *et al.* Characterization of microRNA expression profiles in normal and osteoarthritic human chondrocytes. *BMC Musculoskelet Disord* **13**, 1–14 (2012).

197. Li, Y. H. *et al.* Identification of synovial fluid microRNA signature in knee osteoarthritis: differentiating early- and late-stage knee osteoarthritis. *Osteoarthritis Cartilage* **24**, 1577–1586 (2016).

198. Zhao, Z., Dai, X. S., Wang, Z. Y., Bao, Z. Q. & Guan, J. Z. MicroRNA-26a reduces synovial inflammation and cartilage injury in osteoarthritis of knee joints through impairing the NF-κB signaling pathway. *Biosci Rep* **29**, 20182025 (2019).

199. Ajekigbe, B. *et al.* Identification of long non-coding RNAs expressed in knee and hip osteoarthritic cartilage. *Osteoarthritis Cartilage* **27**, 694–702 (2019).

200. Barter, M. J. *et al.* The long non-coding RNA ROCR contributes to SOX9 expression and chondrogenic differentiation of human mesenchymal stem cells. *Development* **144**, 4510–4521 (2017).

201. Carlson, H. L. *et al.* LncRNA-HIT Functions as an Epigenetic Regulator of Chondrogenesis through Its Recruitment of p100/CBP Complexes. *PLoS Genet* **11**, e1005680 (2015).

202. Hu, J. *et al.* Long non-coding RNA HOTAIR promotes osteoarthritis progression via miR-17-5p/FUT2/β-catenin axis. *Cell Death Dis* **9**, 711 (2018).

203. Li, H. *et al.* Combining single-cell RNA sequencing and population-based studies reveals hand osteoarthritis-associated chondrocyte subpopulations and pathways. *Bone Res* **11**, 1–13 (2023).

204. Rai, M. F. *et al.* Single Cell Omics for Musculoskeletal Research. *Curr Osteoporos Rep* **19**, 131–140 (2021).

205. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045–1048 (2010).

206. Abascal, F. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).

207. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).

208. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21–29 (2015).

209. Liu, Y. *et al.* Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis. *Sci Rep* **8**, 15499 (2018).

210. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* **47**, 598–606 (2015).

211. Bittner, N. *et al.* Primary osteoarthritis chondrocyte map of chromatin conformation reveals novel candidate effector genes. *Ann Rheum Dis* **83**, 1048–1059 (2024).

212. Boer, C. G. *et al.* Genome-wide association of phenotypes based on clustering patterns of hand osteoarthritis identify WNT9A as novel osteoarthritis gene. *Ann Rheum Dis* **80**, 367–375 (2021).

213. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580–585 (2013).

214. den Hollander, W. *et al.* Annotating Transcriptional Effects of Genetic Variants in Disease-Relevant Tissue: Transcriptome-Wide Allelic Imbalance in Osteoarthritic Cartilage. *Arthritis & Rheumatology* **71**, 561–570 (2019).

215. Rice, S. J. *et al.* Prioritization of PLEC and GRINA as Osteoarthritis Risk Genes Through the Identification and Characterization of Novel Methylation Quantitative Trait Loci. *Arthritis & Rheumatology* **71**, 1285–1296 (2019).

216. Kehayova, Y. S., Wilkinson, J. M., Rice, S. J. & Loughlin, J. Mediation of the Same Epigenetic and Transcriptional Effect by Independent Osteoarthritis Risk–Conferring Alleles on a Shared Target Gene, COLGALT2. *Arthritis & Rheumatology* **75**, 910–922 (2023).

217. Klein, J. C. *et al.* Functional testing of thousands of osteoarthritis-associated variants for regulatory activity. *Nat Commun* **10**, 2434 (2019).

218. Shepherd, C. *et al.* Functional Characterization of the Osteoarthritis Genetic Risk Residing at ALDH1A2 Identifies rs12915901 as a Key Target Variant. *Arthritis & Rheumatology* **70**, 1577–1587 (2018).

219. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9**, 1–13 (2018).

220. Vojta, A. *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* **44**, 5615–5628 (2016).

221. Héberlé, É. & Bardet, A. F. Sensitivity of transcription factors to DNA methylation. *Essays Biochem* **63**, 727–741 (2019).

222. Kumar, S., Ambrosini, G. & Bucher, P. SNP2TFBS – a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res* **45**, D139 (2017).

223. Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50**, D165–D173 (2022).

224. Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat Protoc* **2**, 1849–1861 (2007).

225. Syddall, C. M., Reynard, L. N., Young, D. A. & Loughlin, J. The Identification of Trans-acting Factors That Regulate the Expression of GDF5 via the Osteoarthritis Susceptibility SNP rs143383. *PLoS Genet* **9**, e1003557 (2013).

226. Reynard, L. N., Bui, C., Syddall, C. M. & Loughlin, J. CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. *Hum Genet* **133**, 1059–1073 (2014).

227. Chen, W., Jiang, X. & Luo, Z. WWP2: A Multifunctional ubiquitin ligase Gene. *Pathology and Oncology Research* **20**, 799–803 (2014).

228. Kerscher, O., Felberbaum, R. & Hochstrasser, M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol* **22**, 159–180 (2006).

229. Scheffner, M. & Kumar, S. Mammalian HECT ubiquitin-protein ligases: Biological and pathophysiological aspects. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**, 61–74 (2014).

230. Chantry, A. WWP2 ubiquitin ligase and it's isoforms: new biological insight and promising disease targets. *Cell Cycle* **10**, 2437 (2011).

231. Soond, S. M. & Chantry, A. How ubiquitination regulates the TGF-β signalling pathway: new insights and new players: new isoforms of ubiquitin-activating enzymes in the E1-E3 families join the game. *Bioessays* **33**, 749–758 (2011).

232. Lu, P. J., Zhou, X. Z., Shen, M. & Lu, K. P. Function of WW domains as phosphoserine- or phosphothreonine-binding modules. *Science (1979)* **283**, 1325–1328 (1999).

233. Rotin, D. & Kumar, S. Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol* **10**, 398–409 (2009).

234. Soond, S. M. & Chantry, A. Selective targeting of activating and inhibitory Smads by distinct WWP2 ubiquitin ligase isoforms differentially modulates TGFβ signalling and EMT. *Oncogene* **30**, 2451–2462 (2011).

235. Wahl, L. C. *et al.* Smad7 Binds Differently to Individual and Tandem WW3 and WW4 Domains of WWP2 Ubiquitin Ligase Isoforms. *Int J Mol Sci* **20**, 4682 (2019).

236. Chen, H. *et al.* WWP2 regulates pathological cardiac fibrosis by modulating SMAD2 signaling. *Nat Commun* **10**, 3616 (2019).

237. Mokuda, S. *et al.* Wwp2 maintains cartilage homeostasis through regulation of Adamts5. *Nat Commun* **10**, (2019).

238. Zhu, W. *et al.* The E3 ubiquitin ligase WWP2 facilitates RUNX2 protein transactivation in a mono-ubiquitination manner during osteogenic differentiation. *Journal of Biological Chemistry* **292**, 11178–11188 (2017).

239. Shen, R. *et al.* Smad6 interacts with Runx2 and mediates Smad ubiquitin regulatory factor 1-induced Runx2 degradation. *Journal of Biological Chemistry* **281**, 3569–3576 (2006).

240. Zhao, M., Qiao, M., Oyajobi, B. O., Mundy, G. R. & Chen, D. E3 ubiquitin ligase Smurf1 mediates core-binding factor α1/Runx2 degradation and plays a specific role in osteoblast differentiation. *Journal of Biological Chemistry* **278**, 27939–27944 (2003).

241. Zou, W. *et al.* The E3 ubiquitin ligase Wwp2 regulates craniofacial development through mono-ubiquitylation of Goosecoid. *Nat Cell Biol* **13**, 59–65 (2010).

242. Inui, M. *et al.* Dissecting the roles of miR-140 and its host gene. *Nat Cell Biol* **20**, 516–518 (2018).

243. Tuerlings, M. *et al.* WWP2 confers risk to osteoarthritis by affecting cartilage matrix deposition via hypoxia associated genes. *Osteoarthritis Cartilage* **31**, 39–48 (2023).

244. Yang, J. *et al.* MiR-140 is co-expressed with Wwp2-C transcript and activated by Sox9 to target Sp1 in maintaining the chondrocyte proliferation. *FEBS Lett* **585**, 2992–2997 (2011).

245. Woods, S. *et al.* microRNA-seq of cartilage reveals an overabundance of miR-140-3p which contains functional isomiRs. *RNA* **26**, 1575–1588 (2020).

246. Crowe, N. *et al.* Detecting new microRNAs in human osteoarthritic chondrocytes identifies miR-3085 as a human, chondrocyte-selective, microRNA. *Osteoarthritis Cartilage* **24**, 534–543 (2016).

247. Pais, H. *et al.* Analyzing mRNA expression identifies Smad3 as a microRNA-140 target regulated only at protein level. *RNA* **16**, 489–494 (2010).

248. Barter, M. J. *et al.* Genome-Wide MicroRNA and Gene Analysis of Mesenchymal Stem Cell Chondrogenesis Identifies an Essential Role and Multiple Targets for miR-140-5p. *Stem Cells* **33**, 3266–3280 (2015).

249. Tuddenham, L. *et al.* The cartilage specific microRNA-140 targets histone deacetylase 4 in mouse cells. *FEBS Lett* **580**, 4214–4217 (2006).

250. Karlsen, T. A., Jakobsen, R. B., Mikkelsen, T. S. & Brinchmann, J. E. microRNA-140 targets RALA and regulates chondrogenic differentiation of human mesenchymal stem cells by translational enhancement of SOX9 and ACAN. *Stem Cells Dev* **23**, 290–304 (2014).

251. Yamashita, S. *et al.* L-Sox5 and Sox6 proteins enhance chondrogenic miR-140 MicroRNA expression by strengthening dimeric Sox9 activity. *Journal of Biological Chemistry* **287**, 22206–22215 (2012).

252. Papathanasiou, I., Trachana, V., Mourmoura, E. & Tsezou, A. DNA methylation regulates miR-140-5p and miR-146a expression in osteoarthritis. *Life Sci* **228**, 274–284 (2019).

253. Nakamura, Y., Inloes, J. B., Katagiri, T. & Kobayashi, T. Chondrocyte-Specific MicroRNA-140 Regulates Endochondral Bone Development and Targets Dnpep To Modulate Bone Morphogenetic Protein Signaling. *Mol Cell Biol* **31**, 3019–3028 (2011).

254. Eberhart, J. K. *et al.* MicroRNA Mirn140 modulates Pdgf signaling during palatogenesis. *Nat Genet* **40**, 290–298 (2008).

255. Grigelioniene, G. *et al.* Gain-of-function mutation of microRNA-140 in human skeletal dysplasia. *Nat Med* **25**, 583–590 (2019).

256. Tao, S. C. *et al.* Exosomes derived from miR-140-5p-overexpressing human synovial mesenchymal stem cells enhance cartilage tissue regeneration and prevent osteoarthritis of the knee in a rat model. *Theranostics* **7**, 180–195 (2017).

257. Rice, S. J., Cheung, K., Reynard, L. N. & Loughlin, J. Discovery and analysis of methylation quantitative trait loci (mQTLs) mapping to novel osteoarthritis genetic risk signals. *Osteoarthritis Cartilage* **27**, 1545–1556 (2019).

258. Nassar, L. R. *et al.* The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res* **51**, D1188–D1195 (2023).

259. Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**, 1–12 (2016).

260. Rauluseviciute, I. *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **52**, D174–D182 (2024).

261. Hammal, F., De Langen, P., Bergon, A., Lopez, F. & Ballester, B. ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res* **50**, D316–D325 (2022).

262. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

263. Wang, Y. *et al.* The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* **19**, 1–12 (2018).

264. Morris, J. A. *et al.* An atlas of genetic influences on osteoporosis in humans and mice. *Nat Genet* **51**, 258–266 (2018).

265. Ge, X. *et al.* Functional genomics atlas of synovial fibroblasts defining rheumatoid arthritis heritability. *Genome Biol* **22**, 1–39 (2021).

266. Goldring, M. B. *et al.* Interleukin-1 beta-modulated gene expression in immortalized human chondrocytes. *J Clin Invest* **94**, 2307–2316 (1994).

267. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**, 402–408 (2001).

268. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 1–9 (2010).

269. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**, 207–220 (2019).

270. Blackwood, E. M. & Kadonaga, J. T. Going the distance: a current view of enhancer action. *Science* **281**, 60–63 (1998).

271. Pierce, B. L. *et al.* Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms. *Nat Commun* **9**, 804 (2018).

272. Hannon, E. *et al.* Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits. *Am J Hum Genet* **103**, 654–665 (2018).

273. Richard, D. & Capellini, T. D. Shifting epigenetic contexts influence regulatory variation and disease risk. *Aging (Albany NY)* **13**, 15699–15749 (2021).

274. Seale, K., Horvath, S., Teschendorff, A., Eynon, N. & Voisin, S. Making sense of the ageing methylome. *Nat Rev Genet* **23**, 585–605 (2022).

275. Rushton, M. D. *et al.* Methylation quantitative trait locus analysis of osteoarthritis links epigenetics with genetic risk. *Hum Mol Genet* **24**, 7432–7444 (2015).

276. Zhang, L. *et al.* Sex-specific DNA methylation differences in Alzheimer's disease pathology. *Acta Neuropathol Commun* **9**, 77 (2021).

277. Ober, C., Loisel, D. A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nat Rev Genet* **9**, 911–922 (2008).

278. Lopes-Ramos, C. M. *et al.* Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues. *Cell Rep* **31**, 107795 (2020).

279. Sugathan, A. & Waxman, D. J. Genome-Wide Analysis of Chromatin States Reveals Distinct Mechanisms of Sex-Dependent Gene Regulation in Male and Female Mouse Liver. *Mol Cell Biol* **33**, 3594–3610 (2013).

280. Alajem, A. *et al.* DNA methylation patterns expose variations in enhancer-chromatin modifications during embryonic stem cell differentiation. *PLoS Genet* **17**, e1009498 (2021).

281. Peng, J. S. *et al.* Amelioration of Experimental Autoimmune Arthritis Through Targeting of Synovial Fibroblasts by Intraarticular Delivery of MicroRNAs 140-3p and 140-5p. *Arthritis & Rheumatology* **68**, 370–381 (2016).

282. Keller, A. *et al.* miRNATissueAtlas2: an update to the human miRNA tissue atlas. *Nucleic Acids Res* **50**, D211–D221 (2022).

283. Zhu, H., Wang, G. & Qian, J. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet* **17**, 551–565 (2016).

284. Finger, F. *et al.* Molecular phenotyping of human chondrocyte cell lines T/C-28a2, T/C-28a4, and C-28/I2. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* **48**, 3395–3403 (2003).

285. Mole, D. R. *et al.* Genome-wide association of hypoxia-inducible factor (HIF)-1α and HIF-2α DNA binding with expression profiling of hypoxia-inducible transcripts. *Journal of Biological Chemistry* **284**, 16767–16775 (2009).

286. Wenger, R. H., Kvietikova, I., Rolfs, A., Camenisch, G. & Gassmann, M. Oxygen-regulated erythropoietin gene expression is dependent on a CpG methylation-free hypoxia-inducible factor-1 DNA-binding site. *Eur J Biochem* **253**, 771–777 (1998).

287. D'Anna, F. *et al.* DNA methylation repels binding of hypoxia-inducible transcription factors to maintain tumor immunotolerance. *Genome Biol* **21**, 1–36 (2020).

288. Liu, Y., Toh, H., Sasaki, H., Zhang, X. & Cheng, X. An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. *Genes Dev* **26**, 2374–2379 (2012).

289. Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *Elife* **2013**, (2013).

290. Li, S., Peng, Y. & Panchenko, A. R. DNA methylation: Precise modulation of chromatin structure and dynamics. *Curr Opin Struct Biol* **75**, 102430 (2022).

291. Okada, K. *et al.* Hypoxia-inducible factor-1 alpha maintains mouse articular cartilage through suppression of NF-κB signaling. *Sci Rep* **10**, 5425 (2020).

292. Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proc Natl Acad Sci U S A* **92**, 5510–5514 (1995).

293. Semenza, G. L. Hypoxia-inducible factor 1 (HIF-1) pathway. *Science's STKE* **2007**, cm8 (2007).

294. Taylor, C. T. & Scholz, C. C. The effect of HIF on metabolism and immunity. *Nat Rev Nephrol* **18**, 573–587 (2022).

295. Hu, S. *et al.* Stabilization of HIF-1α alleviates osteoarthritis via enhancing mitophagy. *Cell Death Dis* **11**, 481 (2020).

296. Grimmer, C. *et al.* Hypoxia-inducible factor 1α is involved in the prostaglandin metabolism of osteoarthritic cartilage through up-regulation of microsomal prostaglandin E synthase 1 in articular chondrocytes. *Arthritis Rheum* **56**, 4084–4094 (2007).

297. Muñoz-Sánchez, J. & Chánez-Cárdenas, M. E. The use of cobalt chloride as a chemical hypoxia model. *Journal of Applied Toxicology* **39**, 556–570 (2019).

298. Choudhry, H. & Harris, A. L. Cell Metabolism Review Advances in Hypoxia-Inducible Factor Biology. *Cell Metab* **27**, 281–298 (2018).

299. Ravenna, L., Salvatori, L. & Russo, M. A. HIF3α: the little we know. *FEBS J* **283**, 993–1003 (2016).

300. Milner, P. I., Fairfax, T. P. A., Browning, J. A., Wilkins, R. J. & Gibson, J. S. The effect of O2 tension on pH homeostasis in equine articular chondrocytes. *Arthritis Rheum* **54**, 3523–3532 (2006).

301. Semenza, G. L. Hypoxia-Inducible Factors in Physiology and Medicine. *Cell* **148**, 399–408 (2012).

302. Amarilio, R. *et al.* HIF1alpha regulation of Sox9 is necessary to maintain differentiation of hypoxic prechondrogenic cells during early skeletogenesis. *Development* **134**, 3917–3928 (2007).

303. Duval, E. *et al.* Hypoxia-inducible factor 1alpha inhibits the fibroblast-like markers type I and type III collagen during hypoxia-induced chondrocyte redifferentiation: hypoxia not only induces type II collagen and aggrecan, but it also inhibits type I and type III collagen in the hypoxia-inducible factor 1alpha-dependent redifferentiation of chondrocytes. *Arthritis Rheum* **60**, 3038–3048 (2009).

304. Provot, S. & Schipani, E. Fetal growth plate: a developmental model of cellular adaptation to hypoxia. *Ann N Y Acad Sci* **1117**, 26–39 (2007).

305. Saito, T. *et al.* Transcriptional regulation of endochondral ossification by HIF-2α during skeletal growth and osteoarthritis development. *Nat Med* **16**, 678–686 (2010).

306. Hirata, M. *et al.* C/EBPβ and RUNX2 cooperate to degrade cartilage with MMP-13 as the target and HIF-2α as the inducer in chondrocytes. *Hum Mol Genet* **21**, 1111–1123 (2012).

307. Pfander, D., Cramer, T. & Swoboda, B. Hypoxia and HIF-1α in osteoarthritis. *Int Orthop* **29**, 6–9 (2005).

308. Ryu, J. H. *et al.* Interleukin-6 plays an essential role in hypoxia-inducible factor 2α-induced experimental osteoarthritic cartilage destruction in mice. *Arthritis Rheum* **63**, 2732–2743 (2011).

309. Wenger, R. H., Kvietikova, I., Rolfs, A., Camenisch, G. & Gassmann, M. Oxygen-regulated erythropoietin gene expression is dependent on a CpG methylation-free hypoxia-inducible factor-1 DNA-binding site. *Eur J Biochem* **253**, 771–777 (1998).

310. Li, Q., Ke, Q. & Costa, M. Alterations of histone modifications by cobalt compounds. *Carcinogenesis* **30**, 1243–1251 (2009).

311. Johnson, A. B., Denko, N. & Barton, M. C. Hypoxia induces a novel signature of chromatin modifications and global repression of transcription. *Mutation*

Research/Fundamental and Molecular Mechanisms of Mutagenesis **640**, 174–179 (2008).

312. Batie, M., Frost, J., Shakir, D. & Rocha, S. Regulation of chromatin accessibility by hypoxia and HIF. *Biochemical Journal* **479**, 767–786 (2022).
313. Søndergaard, J. N. *et al.* Successful delivery of large-size CRISPR/Cas9 vectors in hard-to-transfect human cells using small plasmids. *Commun Biol* **3**, 319 (2020).
314. Triantafyllou, A. *et al.* Cobalt induces hypoxia-inducible factor-1α (HIF-1α) in HeLa cells by an iron-independent, but ROS-, PI-3K- and MAPK-dependent mechanism. *Free Radic Res* **40**, 847–856 (2006).
315. Simonsen, L. O., Harbak, H. & Bennekou, P. Cobalt metabolism and toxicology—A brief update. *Science of The Total Environment* **432**, 210–215 (2012).
316. Mahey, S. *et al.* Effect of cobalt(II) chloride hexahydrate on some human cancer cell lines. *Springerplus* **5**, 1–14 (2016).
317. Kanaji, A. *et al.* Cytotoxic effects of cobalt and nickel ions on osteocytes in vitro. *J Orthop Surg Res* **9**, 1–8 (2014).
318. Huk, O. L. *et al.* Induction of apoptosis and necrosis by metal ions in vitro. *J Arthroplasty* **19**, 84–87 (2004).
319. Andrés, M. C. De *et al.* Loss of methylation in CpG sites in the NF-κB enhancer elements of inducible nitric oxide synthase is responsible for gene induction in human articular chondrocytes. *Arthritis Rheum* **65**, 732–742 (2013).
320. Gao, X. D., Rodríguez, T. C. & Sontheimer, E. J. Adapting dCas9-APEX2 for subnuclear proteomic profiling. *Methods Enzymol* **616**, 365–383 (2019).
321. Stewart, A. J., Houston, B. & Farquharson, C. Elevated expression of hypoxia inducible factor-2alpha in terminally differentiating growth plate chondrocytes. *J Cell Physiol* **206**, 435–440 (2006).
322. Yang, S. *et al.* Hypoxia-inducible factor-2alpha is a catabolic regulator of osteoarthritic cartilage destruction. *Nat Med* **16**, 687–693 (2010).
323. Kumar, S., Ambrosini, G. & Bucher, P. SNP2TFBS – a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res* **45**, D139–D144 (2017).
324. Adikusuma, F. *et al.* Optimized nickase- and nuclease-based prime editing in human and mouse cells. *Nucleic Acids Res* **49**, 10785–10795 (2021).

325. Bennett, E. P. *et al.* INDEL detection, the 'Achilles heel' of precise genome editing: a survey of methods for accurate profiling of gene editing induced indels. *Nucleic Acids Res* **48**, 11958–11981 (2020).

326. Dehairs, J., Talebi, A., Cherifi, Y. & Swinnen, J. V. CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. *Sci Rep* **6**, 1–5 (2016).

327. Liu, F. *et al.* NEAT1/miR-193a-3p/SOX5 axis regulates cartilage matrix degradation in human osteoarthritis. *Cell Biol Int* **44**, 947–957 (2020).

328. Zhao, C. & Meng, A. Sp1-like transcription factors are regulators of embryonic development in vertebrates. *Dev Growth Differ* **47**, 201–211 (2005).

329. Cassandri, M. *et al.* Zinc-finger proteins in health and disease. *Cell Death Discov* **3**, 17071 (2017).

330. Aubourg, G., Rice, S. J., Bruce-Wootton, P. & Loughlin, J. Genetics of osteoarthritis. *Osteoarthritis Cartilage* **30**, 636–649 (2022).

331. Loughlin, J. Translating osteoarthritis genetics research: challenging times ahead. *Trends Mol Med* **28**, 176–182 (2022).

332. Oo, W. M., Little, C., Duong, V. & Hunter, D. J. The Development of Disease-Modifying Therapies for Osteoarthritis (DMOADs): The Evidence to Date. *Drug Des Devel Ther* **15**, 2921–2945 (2021).

333. Watt, J. E. *et al.* Discovery of Small Molecule WWP2 Ubiquitin Ligase Inhibitors. *Chemistry – A European Journal* **24**, 17677–17680 (2018).

334. Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *New England Journal of Medicine* **384**, 252–260 (2021).

335. Grandi, F. C. & Bhutani, N. Epigenetic Therapies for Osteoarthritis. *Trends Pharmacol Sci* **41**, 557–569 (2020).

336. Kadler, S. *et al.* Effects of 5-aza-2´-deoxycytidine on primary human chondrocytes from osteoarthritic patients. *PLoS One* **15**, e0234641 (2020).

337. Zuscik, M. J. *et al.* 5-azacytidine alters TGF-β and BMP signaling and induces maturation in articular chondrocytes. *J Cell Biochem* **92**, 316–331 (2004).

338. Steimberg, N. *et al.* SV40 Large T Antigen Expression Driven by col2a1 Regulatory Sequences Immortalizes Articular Chondrocytes but Does Not Allow Stabilization of Type II Collagen Expression. *Exp Cell Res* **249**, 248–259 (1999).

339. Stokes, D. G. *et al.* Regulation of type-II collagen gene expression during human chondrocyte de-differentiation and recovery of chondrocyte-specific phenotype in culture involves Sry-type high-mobility-group box (SOX) transcription factors. *Biochemical Journal* **360**, 461–470 (2001).

340. Stokes, D. G. *et al.* Assessment of the gene expression profile of differentiated and dedifferentiated human fetal chondrocytes by microarray analysis. *Arthritis Rheum* **46**, 404–419 (2002).

341. Van De Weijer, M. L. *et al.* A high-coverage shRNA screen identifies TMEM129 as an E3 ligase involved in ER-associated protein degradation. *Nat Commun* **5**, 3832 (2014).

342. Oikonomou, C. & Hendershot, L. M. Disposing of misfolded ER proteins: A troubled substrate's way out of the ER. *Mol Cell Endocrinol* **500**, 110630 (2020).

343. Briggs, M. D., Dennis, E. P., Dietmar, H. F. & Pirog, K. A. New developments in chondrocyte ER stress and related diseases. *F1000Res* **9**, (2020).

344. Rellmann, Y., Eidhof, E. & Dreier, R. Review: ER stress-induced cell death in osteoarthritic cartilage. *Cell Signal* **78**, 109880 (2021).

345. Alontaga, A. Y. *et al.* RWD domain as an E2 (Ubc9)-interaction module. *Journal of Biological Chemistry* **290**, 16550–16559 (2015).

346. Serrano, R. L., Chen, L. Y., Lotz, M. K., Liu-Bryan, R. & Terkeltaub, R. Impaired Proteasomal Function in Human Osteoarthritic Chondrocytes Can Contribute to Decreased Levels of SOX9 and Aggrecan. *Arthritis Rheumatol* **70**, 1030–1041 (2018).