

The Molecular Genetics and Epigenetics of *COLGALT2*, a Risk Locus for Osteoarthritis

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

by

Yulia Sergeeva Kehayova



January 2023

Biosciences Institute, School of Medicine, Newcastle University

Word count: 47 974

Abstract

Osteoarthritis (OA) is a debilitating joint disorder affecting millions globally. No disease-modifying osteoarthritis drugs exist to date, highlighting the need for in-depth research into the molecular underpinnings of OA. The aetiology of OA is multifactorial, with genetics playing a crucial role in its development. Genome-wide association studies (GWAS) have to date linked over 125 genetic variants to OA. Functional follow-up studies on these variants are needed in order to gain profound insights into the biology of this complex trait and diseases and create therapeutic opportunities.

In this thesis I investigated two independent OA GWAS signals, rs11583641 and rs1046934, associated with a collagen galactosyltransferase gene *COLGALT2*. Cartilage and synovium tissues from arthroplasty patients were used to detect quantitative trait loci (QTLs) and link the genetic variants with gene expression and DNA methylation (DNAm) levels at enhancer regions. These effects were also investigated in developmental cartilage samples in order to gain insight into the developmental origins of OA. CRISPR/Cas9 tools were used to identify causal relationships between DNAm at the enhancers and gene expression *in vitro* cell models of cartilage and synovium. The role of the *COLGALT2* enzyme in chondrocyte biology was studied using a proteomic analysis of mesenchymal stem cells differentiated into chondrocytes

The OA associated variants were found to act as both expression and methylation -QTLs in arthroplasty and in developmental cartilage samples. The risk alleles of these two independent variants were found to correlate with DNAm levels at enhancer regions and with the expression of *COLGALT2*. Targeted epigenetic modulation of the enhancers revealed causal links between DNA methylation and *COLGALT2* expression. The rs11583641 locus was active in synovial tissue but the mechanism driving the effect was distinct from that observed in cartilage. Depletion of the *COLGALT2* enzyme revealed it to be crucial for the development of chondrocytes and for the secretion of healthy cartilage ECM.

The work presented in this thesis highlights the complex genetic landscape of OA, underscoring the importance of genetic and epigenetic factors in its pathogenesis. The identification of novel genetic targets opens avenues for the development of DMOADs, offering hope for improved management and treatment of OA.

Acknowledgements

The completion of this work is a result of continuous support from my supervisors, Prof John Loughlin, Dr Sarah Rice and Prof Mark Wilkinson. I would like to express my eternal gratitude to them for not giving up on me and motivating me to finish this work. I would also like to express my gratitude to all the members, past and present, of the Osteoarthritis research group. They have provided support, encouragement and valuable feedback to me throughout my MRes and PhD studies and created a positive and enjoyable work environment. I would also like to thank all the people who directly contributed to the work presented in this thesis.

I am grateful to the CIMA for providing the funding for my project and to all the patients who allowed their tissues to be used for science.

A special thanks to my partner, Thomas Ward, who was my lifeline through the past few years and kept me grounded, and to my friends who are the epitome of love and friendship.

Table of Contents

Chapter 1. Introduction	1
1.1.	1
1.1.1.	3
Skeletogenesis	3
Synovial joint determination and formation	4
Endochondral ossification	5
1.1.2.	5
1.2.	16
1.2.1.	17
1.2.2.	20
1.2.3.	21
1.2.4.	22
1.2.5.	23
1.2.6.	25
1.2.7.	33
1.3.	34
1.3.1.	34
1.3.2.	38
1.3.3.	39
1.3.4.	42
1.4.	43
1.5.	43
1.6.	44
Chapter 2. Methods and Materials	- 81 -
2.1. Tissue Samples	- 81 -
2.2. Extraction of Nucleic Acids from Human Samples	- 86 -
2.3. DNA Bisulfite Conversion	- 86 -

2.4. Complementary DNA (cDNA) Synthesis	- 87 -
2.5. Genotyping	- 88 -
2.6. Measuring DNAm Levels at CpG Sites	- 89 -
2.7. Allelic Expression Imbalance (AEI) analysis	- 90 -
2.8. Gene Expression Quantification	- 91 -
2.10. Lucia Reporter Gene Analysis	- 93 -
2.11. Activation and Repression of the Enhancer Containing CpGs 3-10 (rs11583641 locus)	- 95 -
2.12. Deletion of the Enhancer Region Containing CpGs 3-10 (rs11583641 locus)	- 97 -
2.13. Methylation and Demethylation of Regions of Interest in TC28a2 Cells	- 99 -
2.14. Assay for Transposase-Accessible Chromatin with Sequencing (ATAC-seq)	- 103 -
2.15. Differentiated ASC52teloSOX9 Cells as a New Chondrocyte Cell Model-	- 106 -
2.16. Knockout (KO) of <i>COLGALT1</i> and <i>COLGALT2</i> to Deplete their Respective Proteins	- 108 -
2.17. Graphs and Statistical Analyses	- 114 -
2.18. <i>In silico</i> Analyses	- 114 -
2.19. Contributions of Other People to the Work Presented in this Thesis	- 116 -
2.20. Overlap Between MRes and PhD	- 117 -
2.21. Presentations and publications	- 117 -
2.22. References	- 118 -
<i>Chapter 3. Genetic and Epigenetic Interplay Within a COLGALT2 Enhancer Associated with Osteoarthritis</i>	- 120 -
3.1. Introduction	- 120 -
3.2. Aims and Objectives	- 120 -

3.3. Results	- 122 -
3.4. Discussion	- 145 -
3.5. References	- 147 -
<i>Chapter 4. Independent Osteoarthritis Risk-Confering Alleles Mediate the Same Epigenetic and Transcriptional Effect on COLGALT2</i>	- 149 -
4.1. Introduction	- 149 -
4.2. Aims and Objectives	- 150 -
4.3. Results	- 151 -
4.4. Discussion	- 166 -
4.5. References	- 168 -
<i>Chapter 5. Investigation of the rs11583641 and rs1046934 OA association signals in synovium</i>	- 170 -
5.1. Introduction	- 170 -
5.2. Aims and objectives	- 170 -
5.3. Results	- 172 -
5.4. Discussion	- 185 -
5.5. References	- 190 -
<i>Chapter 6. Genetic and Epigenetic Regulation of COLGALT2 in Developmental Joint Tissues</i>	- 193 -
6.1. Introduction	- 193 -
6.2. Aims and objectives	- 195 -
6.3. Results	- 196 -
6.4. Discussion	- 208 -
6.5. References	- 212 -
<i>Chapter 7. Role of the COLGALT2 enzyme in chondrocyte biology</i>	- 215 -
7.1. Introduction	- 215 -

7.3. Results	- 218 -
7.4. Discussion	- 242 -
7.5. References	- 245 -
<i>Chapter 8. General Discussion</i>	- 248 -

List of Figures

Figure 1.1. Structure of the synovial joint (left) and articular cartilage (right).	-9-
Figure 1.2. Features of a healthy (left) versus an OA affected (right) knee joint.	-17-
Figure 1.3. Collagen synthesis and PTMs.	-34-
Figure 2.1. A representative image of the RFLP-PCR used to genotype rs734657.	-89-
Figure 3.1. Schematic overview of the region containing the GWAS signal and the investigated CpG sites.	-
122-	
Figure 3.2. Cartilage mQTL analysis of 12 CpG sites within the putative COLGALT2 enhancer.	-
124-	
Figure 3.3. mQTL analysis of the 12 investigated CpG sites in OA and non-OA cartilage samples.	-125-
Figure 3.4. Effects of joint site and disease status on DNAm.	-
127-	
Figure 3.5. mQTL analysis of the 12 investigated CpG sites in OA Knee, OA hip and non-OA cartilage samples (NOF).	-128-
Figure 3.6. Relationship between age of the patients at the time of surgery and DNAm levels at the investigated CpG sites.	-129-
Figure 3.7. Reporter gene analysis.	-131-
Figure 3.8. Schematic showing the position of the SNPs in high LD with rs11583641.	-133-
Figure 3.9. Genotypic contributions of the SNPs in high LD with rs11583641 to DNAm variability measured across the COLGALT2 enhancer.	-
134-	
Figure 3.10. Effect of activation and repression of the enhancer on gene expression.	-136-
Figure 3.11. Effect of deletion of the enhancer on gene expression.	-
138-	
Figure 3.12. Percentage change in DNAm following targeted epigenome modulation of the DMR.	-
140-	
Figure 3.13. Epigenetic modulation of the enhancer in TC28a2 cells.	-
141-	
Figure 3.14. Accuracy of the AEI assay used.	-
142-	
Figure 3.15. AEI analysis of COLGALT2 in arthroplasty cartilage samples.	-
144-	
Figure 4.1. mQTLs at the rs1046934 locus.	-150-

Figure 4.2. Protein structure of the structural unit of the eucaryotic tRNA splicing (TSEN) complex.	-152-
Figure 4.3. Accuracy of the AEI assays used.	-153-
Figure 4.4. AEI analysis of COLGALT2 and TSEN15 in arthroplasty cartilage samples.	-154-
Figure 4.5. Overview of the rs1046934 locus and mQTL replication in arthroplasty cartilage samples.	-155-
Figure 4.6. Stratification of DNAm levels at cg15204595 and cg21606956 against genotype at rs1046934, disease state, joint and age at surgery.	-157-
Figure 4.7. Relationship between DNAm at the enhancer and COLGALT2 AEI ratios.	-158-
Figure 4.8. Theoretical model to account for the opposing slopes of the COLGALT2 meQTLs at cg15204595 and cg21606956.	-160-
Figure 4.9. Reporter gene analyses.	-162-
Figure 4.10. Epigenetic modulation of the enhancers containing cg15204595 or cg21606956.	-164-
Figure 4.11. TFs predicted to bind at the enhancers.	-166-
Figure 5.1. Synovium mQTL analysis of 12 CpG sites within the COLGALT2 enhancer associated with rs11583641.	-172-
Figure 5.2. Comparison of the DNAm levels at the enhancer associated with rs11583641 between synovium and cartilage DNA samples from arthroplasty patients.	-174-
Figure 5.3. Synovium mQTL analysis for cg15204595 and cg21606956.	-174-
Figure 5.4. AEI analysis of COLGALT2 in arthroplasty synovium samples for the rs11583641 locus.	-176-
Figure 5.5. AEI analysis of COLGALT2 and TSEN15 in arthroplasty synovium samples for the rs1046934 locus.	-177-
Figure 5.6. Reporter gene analysis.	-178-
Figure 5.7. Percentage change in DNAm following targeted epigenome modulation of the enhancer associated with rs11583641 in SW982 cells.	-180-
Figure 5.8. Epigenetic modulation of the enhancer in SW982 cells.	-182-
Figure 5.9. TFs predicted to bind at the enhancer.	-184-
Figure 5.10. Model accounting for the opposite effects between synovium and cartilage of the risk allele on enhancer DNAm levels and COLGALT2 expression.	-188-

Figure 6.1. Cartilage mQTL analysis of 12 CpG sites within the COLGALT2 enhancer associated with rs11583641 in developmental samples.	-196-
Figure 6.2. Comparison of the DNAm levels at the enhancer associated with rs11583641 between foetal and arthroplasty cartilage samples.	-197-
Figure 6.3. Heatmap showing the contribution of the rs11583641 genotype to the observed DNAm levels.	-198-
Figure 6.4. mQTL analysis of 12 CpG sites within the COLGALT2 enhancer associated with rs11583641 in developmental mixed tissue limb bud samples.	-198-
Figure 6.5. mQTL analysis at cg15204595 and cg21606956 in developmental samples.	-199-
Figure 6.6. Comparison of the DNAm levels at cg15204595 and cg21606956 between foetal and arthroplasty cartilage samples.	-199-
Figure 6.7. AEI analysis of COLGALT2 for the rs11583641 locus in developmental cartilage samples.	-201-
Figure 6.8. Relationship between DNAm at the enhancer and COLGALT2 AEI ratios.	-202-
Figure 6.9. AEI analysis of COLGALT2 and TSEN15 in foetal cartilage samples for the rs1046934 locus.	-203-
Figure 6.10. Relationship between DNAm at the enhancer and COLGALT2 AEI ratios in foetal cartilage samples.	-204-
Figure 6.11. Differential chromatin accessibility between foetal and OA cartilage.	-205-
Figure 6.12. Prioritisation of SNPs at the COLGALT2 locus based on chromatin accessibility in OA and foetal chondrocytes and on histone marks in relevant cell types.	-206-
Figure 6.13. Chromatin accessibility at the COLGALT2 enhancer CpGs.	-208-
Figure 7.1. Chondrogenic potential of the ASC52teloSox9 cells.	-219-
Figure 7.2. Confirmation of COLAGLT protein depletion prior to differentiation.	-220-
Figure 7.3. Knockout efficiency in ASC52teloSox9 cells.	-221-
Figure 7.4. Phenotype of WT ASC52teloSox9 cells following 21-day differentiation into chondrocytes.	-222-
Figure 7.5. Protein intensities of chondrocyte markers following differentiation into chondrocytes.	-222-
Figure 7.6. Hierarchical clustering dendrogram of protein profiles.	-224-
Figure 7.7. Differential protein expression analysis of WT D0 and WT D21 cells.	-225-
Figure 7.8. Differential protein expression analysis of WT D0 and WT D7 cells.	-227-

Figure 7.9. Differential protein expression analysis of WT D7 and WT D14 cells.	-229-
Figure 7.10. Differential protein expression analysis of WT D14 and WT D21 cells.	-229-
Figure 7.11. Heatmap of Differentially Expressed Genes with Hierarchical Clustering.	-231-
Figure 7.12. Volcano plot of protein expression differences between the secretome of WT D0 and COLGALT2 KO D0 cells.	-233-
Figure 7.13. GO enrichment analysis of WT D0 vs COLGALT2 KO D0 cells.	-234-
Figure 7.14. Volcano plot of protein expression differences between the secretome of WT D21 and COLGALT2 KO D21 cells	-236-
Figure 7.15. GO enrichment analysis of WT D21 vs COLGALT2 KO D21 cells.	-236-
Figure 7.16. Differential protein expression analysis of COLGALT2 KO D0 and COLGALT2 KO D21 cells.	-238-
Figure 7.17. Differential protein expression analysis of the secretome of WT D21 and COLGALT2 KO D21 cells.	-240-

List of Tables

Table 1.1. The collagen superfamily.	-36-
Table 2.1. Patient details of the arthroplasty cartilage samples	-81-
Table 2.2. Patient details of the arthroplasty synovium samples.	-82-
Table 2.3. Details of the developmental cartilage samples.	-84-
Table 2.4. Details of the developmental and OA cartilage samples used for ATAC-seq.	-85-
Table 2.5. Details of the pre-designed RT-qPCR assays.	-92-
Table 2.6 Details of indexing primers used for ATAC-seq.	-105-

Table 2.7. Statistical analyses used.	-114-
Table 2.8. Details of databases and online tools used for in silico analyses.	-115-
Table 2.9. Details of -omics data.	-116-
Table 3.1. Details of the SNPs in high LD with rs11583641.	-
132-	
Table 3.2. Details of the patient samples used for the AEI analysis.	-145-
Table 7.1. Proteins relevant to cartilage biology differentially expressed in WT D0 and WT D21 cells	-226-
Table 7.2. Proteins relevant to cartilage biology differentially expressed in WT D0 and WT D7 cells.	-228-
Table 7.3. Proteins relevant to cartilage biology differentially expressed in WT D7 and WT D14 cells.	-230-
Table 7.4. Proteins relevant to cartilage biology differentially expressed in WT D14 and WT D21 cells.	-
231-	
Table 7.5. Proteins relevant to cartilage biology differentially expressed in WT D0 and COLGALT2 KO D0 cells.	-
235-	
Table 7.6. Proteins relevant to cartilage biology differentially expressed in WT D21 and COLGALT2 KO D21 cells.	-
238-	
Table 7.7. Proteins relevant to cartilage biology differentially expressed in COLGALT2 KO D0 and COLGALT2 KO D21 cells.	-
240-	
Table 7.8. Proteins relevant to cartilage biology differentially secreted by WT D21 and COLGALT2 KO D21 cells.	-
241-	

Chapter 1. Introduction

1.1. The musculoskeletal system and musculoskeletal health

The musculoskeletal system represents a complex and essential network within the human body, encompassing various interconnected components, including bones, muscles, tendons, ligaments, and articulating synovial joints. This intricate system fulfils a range of vital functions, including providing structural support, facilitating movement, protecting internal organs, and contributing to metabolic processes.

Bones serve as the foundational framework of the body, offering robust support and defining its overall shape. Additionally, bones play a pivotal role in metabolic functions, primarily by regulating calcium and phosphorus levels through storage and release. The bone marrow, situated within select bones, assumes the critical responsibility of producing blood cells, including red blood cells, white blood cells, and platelets¹.

Muscles, intimately linked to bones through tendons, act as the dynamic engines of the musculoskeletal system, orchestrating a wide array of movements crucial for everyday activities and mobility. The contraction and relaxation of muscle fibres generate forces transmitted to bones, resulting in joint motion. This intricate interplay enables a broad spectrum of movements, ranging from fine motor skills to powerful actions like running or lifting¹.

Ligaments, which are sturdy connective tissue bands, play a pivotal role in preserving the structural stability of joints. These fibrous structures connect bones to one another, defining the range of motion for each joint. By restraining excessive movement and guiding the articulation of bones, ligaments help prevent injuries and ensure the proper functioning of joints². Tendons are robust, flexible connective tissues crucial for movement. They link muscles to bones, enabling the transfer of muscular force to skeletal structures. This connection allows for the execution of all physical actions. Composed primarily of collagen, tendons are both strong and slightly elastic, adept at handling forces from muscle contractions while providing resilience against impacts. They are essential for maintaining proper joint mechanics and movement efficiency. However, tendons can be

prone to injuries, often due to overuse or intense stress, affecting movement and requiring careful management².

Articulating synovial joints, serving as the meeting points between bones, are integral components of the musculoskeletal system. They facilitate a diverse repertoire of movements. These specialized structures allow for flexibility and mobility while maintaining the critical aspects of stability and balance.

The musculoskeletal system's role extends beyond mere structural support and movement. It is integral to overall health and well-being. The system influences other body systems: it protects internal organs, stores minerals, and houses the marrow for blood cell production. Its health impacts the body's ability to engage in physical activity, which is essential for overall health, including cardiovascular and mental health, directly affecting the quality of life³.

Regarding the current state of musculoskeletal health, in 2020, musculoskeletal disorders were ranked as the second-highest cause of non-fatal disability, affecting over 1.63 billion people globally⁴. These disorders encompass a broad range of conditions affecting the locomotor and connective tissue systems, including bones, joints, ligaments, tendons, and muscles. Specifically, the study includes conditions such as rheumatoid arthritis (RA), osteoarthritis (OA), low back pain, neck pain, and gout, along with other musculoskeletal disorders comprising various acute and chronic conditions⁴.

The global burden of these disorders is substantial and growing. Between 1990 and 2020 the number of cases of musculoskeletal disorders had increased by 59.86%. This increase is projected to continue, with cases expected to rise by 115% from 2020 to 2050, reaching an estimated 1060 million prevalent cases⁴. The age-standardized prevalence of these disorders was notably higher in females than in males and increased with age, peaking at 65-69 years. The high prevalence of musculoskeletal disorders comes with a substantial impact on health and quality of life⁴.

The increasing prevalence and impact of musculoskeletal disorders highlight the need for focused public health efforts. These should include prevention strategies targeting modifiable risk factors, early detection and management of conditions, and research to better understand the underlying mechanisms and effective treatments. The growing

burden also underscores the need for healthcare systems to adapt to the increasing demand for musculoskeletal care, including rehabilitation services and long-term management strategies.

1.1.1. Musculoskeletal and Joint development

The processes of skeletogenesis, synovial joint determination and formation, and endochondral ossification are intricate and highly coordinated events in embryonic development.

Skeletogenesis

Skeletogenesis, the developmental process of skeletal formation during embryonic development, represents a complex orchestration of cellular differentiation and tissue patterning. This process encompasses the transformation of mesenchymal stem cells (MSCs) into the two primary cell types of the skeletal system: chondroprogenitor and osteoprogenitor cells⁵.

The journey of skeletogenesis begins with the condensation of MSCs in specific embryonic limb buds. Chondrogenesis and osteogenesis, the formation of cartilage and bone respectively, ensue following this initial condensation. Chondrocytes begin to form the cartilaginous templates of future bones. Simultaneously, osteoblasts embark on the creation of bone tissue. This bone formation occurs through two distinct pathways: directly via intramembranous ossification or indirectly through endochondral ossification, where the cartilage template is gradually replaced by bone^{5,6}.

Regulating skeletogenesis involves a sophisticated network of signalling pathways, transcription factors, and epigenetic mechanisms. Central to these regulatory pathways are Bone Morphogenic Proteins (BMPs), a subgroup of the Transforming Growth Factor beta (TGF- β) superfamily, which are potent inducers of osteoblast differentiation and play an instrumental role in the early stages of chondrogenesis⁷. Fibroblast Growth Factors (FGF) are also involved in the regulation of these processes, working in tandem with BMPs⁸. The Wnt/ β -catenin signalling pathway also plays a crucial role, promoting osteoblast

differentiation while simultaneously inhibiting chondrocyte maturation⁹. The Hedgehog signalling pathway, particularly the Indian Hedgehog (Ihh), is essential for both chondrocyte proliferation and maturation, as well as osteoblast differentiation^{5,6}. Feedback mechanisms involving these pathways are critical for the regulation of skeletogenesis. For instance, Ihh produced by chondrocytes stimulates the production of parathyroid hormone-related protein (PTHrP), which in turn modulates chondrocyte differentiation¹⁰.

Synovial joint determination and formation

The determination of synovial joints begins with the formation of the interzone at sites where two cartilaginous elements meet. Cells within the interzone undergo apoptosis, leading to the creation of the joint cavity. Surrounding mesenchymal cells differentiate into the articular cartilage and the synovial membrane. The joint cavity formed following apoptosis in the interzone is filled with synovial fluid, produced by the synovial membrane, and serves as a lubricant for the joint.

Central to the development of synovial joints is the role of Growth Differentiation Factor 5 (GDF5), a member of the TGF- β family. *GDF5* is expressed in the interzone of developing joints, playing a crucial role in establishing joint sites¹¹⁻¹³. Its importance is underscored by the skeletal abnormalities associated with *GDF5* mutations¹⁴⁻¹⁸. Wnt/ β -Catenin signalling, another crucial pathway in joint formation, regulates cell fate within the interzone. Activation of this pathway stabilizes β -catenin, which enters the nucleus to activate genes involved in joint formation¹⁹⁻²¹. Wnt signalling interacts with BMP signalling and Ihh signalling to orchestrate joint development^{20,22}.

The transcription factors SOX9 and RUNX2 are paramount in the regulation of chondrogenesis and endochondral ossification. SOX9 is essential for chondrogenesis, driving the expression of cartilage-specific genes^{23,24}, while RUNX2 is a master regulator of osteoblast differentiation²⁵, activating genes necessary for bone formation. The two factors work in synergy to regulate the formation of cartilage and bone²⁶.

Endochondral ossification

Endochondral ossification, a key mechanism in forming long bones, transforms cartilage into bone. Initially, chondrocytes create a cartilage template mirroring the future bone's shape. These cells mature, proliferate, and become hypertrophic chondrocytes, which secrete factors to calcify the cartilage matrix, preparing for bone formation. Vascular invasion brings blood vessels and osteoblasts into the calcified cartilage, beginning bone development while osteoclasts resorb the cartilage. This process forms the primary ossification centre in long bones.

Growth plate activity is crucial for bone elongation, with chondrocytes producing new cartilage replaced by bone. Endochondral ossification is regulated by systemic hormones and local factors like Ihh, PTHrP, and BMPs, influencing chondrocyte and osteoblast functions. Transcription factors including SOX9, RUNX2, and Osterix govern chondrogenesis and osteoblast differentiation²⁷⁻²⁹. This process is vital for growth, height, and limb length, with disruptions leading to disorders like achondroplasia.

1.1.2. Synovial joints

Synovial joints are the main functional joints in the human body characterized by opposing articulating bone ends that are each covered by a thin layer of cartilage and enclosed within a synovial cavity (Figure 1.1). These joints are surrounded by a fibrous joint capsule and contain synovial fluid, a lubricating fluid that reduces friction and provides essential nutrients to the articular cartilage. Synovial joints allow for a wide range of movements, making them crucial for mobility in various parts of the body, such as the knees, hips, hands, elbows, and shoulders.

Articular cartilage

Articular cartilage is a highly specialized connective tissue that covers the ends of bones in joints, facilitating smooth movement by reducing friction. Cartilage consists of chondrocytes, which make up about 1% to 5% of its volume, embedded in a well-organized extracellular matrix (ECM)³⁰. The ECM is composed of collagens, proteoglycans (PGs), and

glycoproteins, constituting approximately 65%, 10-20%, and 5-10% of the organic component of cartilage, respectively³⁰. Cartilage acts as a cushion, absorbing shock by efficiently transmitting forces through the joint, thereby protecting the underlying bone³¹.

Chondrocytes are only cellular component of articular cartilage. These specialized cells are responsible for the synthesis and maintenance of the ECM, which is pivotal for the tissue's overall function. Chondrocytes are not uniform throughout the tissue but exhibit distinct shapes and densities within different zones of cartilage (Figure 1.1), reflecting their roles in matrix production and maintenance³⁰. These cells are well adapted to their avascular environment, with nutrient and waste exchange facilitated by diffusion through the ECM³².

The cartilage ECM is a complex and highly specialized network that plays a crucial role in the structure and function of cartilage. Its composition and organization are tailored to meet the specific mechanical and biological needs of cartilage tissue. It is composed of several critical components. Collagens are the major components of the dry weight of the ECM, providing tensile strength and structural support^{30,33}. This collagen network gives cartilage its white, glossy appearance and contributes significantly to its ability to withstand mechanical stresses³¹.

Type II collagen is the most abundant collagen in articular cartilage^{30,33,34}. In conjunction with other proteins and proteoglycans, this collagen constructs intricate extracellular frameworks which are crucial for enduring mechanical stresses, preserving physiological balance, and providing secure attachment points for chondrocytes, ECM molecules, and growth factors.

Up to 90-95% of the collagens in cartilage ECM can be type II collagen^{30,33}. Type IX and type XI collagens constitute up to 5% and up to 3%, respectively, of the collagens found in cartilage³⁵. Type II and XI collagen, in combination with type IX collagen, interlink to create a mixed fibrillar framework^{36,37}. This structure is fundamental to imparting tensile strength to cartilage, ensuring its resilience and structural integrity. Collagen IX is co-expressed in human tissues with collagen II³⁸. While collagen XI is a quantitatively minor collagen in cartilage, it is essential for the regulation of collagen fibril diameter³⁹⁻⁴¹. This collagen is expressed early during chondrogenesis, highlighting its crucial role in collagen fibrillogenesis^{39,42}.

Other minor collagens in cartilage ECM include types III, IV, V, VI, X, XII, XIV, XVI, XXII, and XXVII³⁶. Collagen III can be found in healthy cartilage³⁴ but is upregulated in diseased tissue⁴³, suggesting it might interact with type II collagen in cartilage in an attempt of tissue repair⁴⁴ similar to the way it interacts with collagen type I in other tissues to initiate wound healing⁴⁵. Type IV collagen is a key component of basement membranes³³, which separate cartilage from surrounding tissues. It provides structural support and helps maintain the cartilage's boundaries. Type V collagen, a minor fibrillar collagen, has been found to assemble as type V/XI collagen in cartilage⁴⁶. It is essential in forming a fibrillar collagen mesh and regulates fibrogenesis and fibre size⁴⁷. It also aids cellular adhesion and matrix repair⁴⁷. Type VI collagen plays a role in cartilage ECM. It forms microfibrils that associate with Type II collagen fibrils⁴⁸ as well as proteoglycans and glycoproteins present in the ECM^{48,49}. As the main collagen in the pericellular matrix, collagen VI helps mediate chondrocyte-matrix interactions⁵⁰. Type X collagen is associated with hypertrophic chondrocytes and the calcified zone in cartilage endochondral ossification⁵¹. This type of collagen is increased in diseased cartilage⁵².

Types XII and XIV collagens are commonly present in regions subjected to intense mechanical stress^{53,54}. They play a significant role in fibrillogenesis and are essential for maintaining the structural integrity and mechanical properties of tissues⁵⁵. Type XVI collagen is found in the chondrocytes' territorial matrix and is associated with thin collagen fibrils containing type II and XI collagen⁵⁶. Type XXII collagen, on the other hand, is located at the interface between synovial fluid and the surface of articular cartilage, playing a role in the cartilage's extrafibrillar matrix⁵⁷. Type XXVII collagen, which forms fibrils, predominantly occurs where cartilage transitions to bone and around proliferative chondrocytes in the epiphyseal growth plate. It is thought to be crucial for the structural integrity of the extracellular matrix in the growth plate, particularly in organizing the proliferative zone⁵⁸.

Proteoglycans are the second most abundant organic component of the cartilage ECM³⁰. Aggrecan is the major proteoglycan in the ECM and plays a vital role in its ability to withstand compressive forces⁵⁹⁻⁶¹. Aggrecan molecules are decorated with glycosaminoglycan (GAG) chains, including chondroitin sulphate and keratan sulphate. These GAG chains exhibit high hydrophilicity, attracting water molecules. This property

gives cartilage its gel-like consistency and empowers it to resist compression effectively. Furthermore, the interaction between the negatively charged GAG chains and positively charged ions contributes to the tissue's osmotic properties, allowing it to absorb and release water in response to mechanical loading^{59,60,62}. Other important proteoglycans in cartilage ECM biology are decorin, biglycan, versican, and perlecan - small proteoglycans that bind to collagen fibrils or other proteoglycans, attract growth factors or interact with the pericellular matrix. They influence matrix organization and fibrillogenesis and play a role in cell adhesion and proliferation⁶³⁻⁶⁶.

Glycoproteins are the third most abundant component of cartilage ECM³⁰. These molecules often serve as adhesive proteins, facilitating cell-matrix interactions and helping anchor the cartilage matrix. Important glycoproteins for the proper function of the ECM are fibronectin, cartilage oligomeric matrix protein (COMP), matrilins, and hyaluronan and lubricin. Fibronectin interacts with collagen and proteoglycans, contributing to matrix assembly⁶⁷⁻⁶⁹. COMP participates in the assembly of the ECM, influencing collagen fibrillogenesis^{70,71}. Matrilins serve as adaptors in the ECM, linking different matrix components and contributing to mechanical stability⁷². Hyaluronan and lubricin ensure the lubrication of the cartilage surface^{68,73-75}.

Water is not merely a passive filler but an active participant in cartilage function. Water constitutes 65-80% of the wet weight of cartilage. It interacts with the hydrophilic GAG chains of proteoglycans, creating an osmotic pressure that resists compression. This hydration-dehydration mechanism enables cartilage to adapt to varying mechanical loads, maintaining its shock-absorbing properties and preventing tissue damage^{31,76}.

The organization of the ECM in articular cartilage exhibits a zonal structure (Figure 1.1). This zonal organization ensures that the cartilage ECM can effectively withstand and distribute the various mechanical stresses it encounters, such as compression, tension, and shear forces⁷⁶⁻⁷⁹. The superficial zone, located at the joint surface, contains high concentration of type II collagen fibres oriented parallel to the surface^{76,77,80}. This orientation provides resistance to shear forces, which can be generated during joint movement. This zone constitutes about 10% of the volume of articular cartilage and has lower proteoglycan content but a higher concentration of hyaluronan and lubricin⁷⁶, contributing to surface lubrication^{73,74}. The middle zone, beneath the superficial zone,

represents between 40 and 60% of the cartilage volume⁷⁶ and features randomly oriented collagen fibres^{77,78} which allows it to resist compressive forces effectively. The concentration of proteoglycans in this zone is higher than that in the superficial zone⁷⁶. The deep zone, 30% of the volume of cartilage, closest to the subchondral bone, contains collagen fibres oriented perpendicularly to the joint surface, providing resistance to both compression and shear forces, and has the highest proteoglycan content in the tissue⁷⁶⁻⁷⁸. The calcified cartilage, adjacent to the subchondral bone, acts as the interface between non-calcified cartilage and the underlying bone⁸¹.

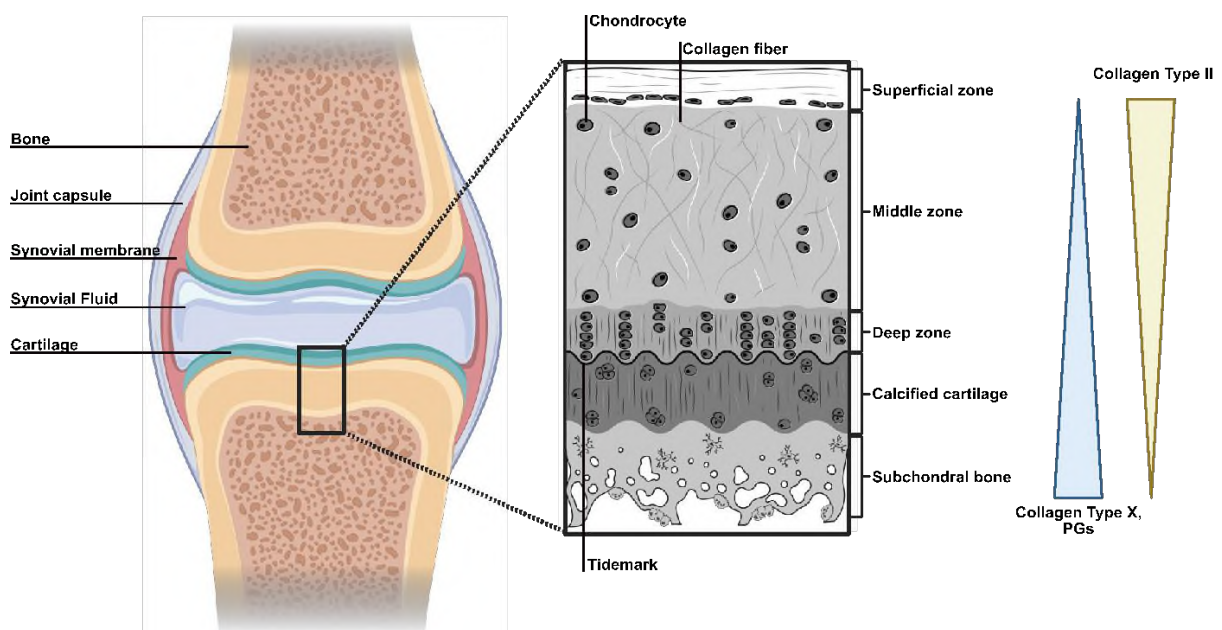


Figure 1.1. Structure of the synovial joint (left) and articular cartilage (right). The image was created using BioRender and adapted in Affinity Designer 2.1.

The ECM can also be divided into pericellular matrix (PCM), territorial matrix and interterritorial matrix. The PCM surrounds the chondrocytes directly. It is rich in proteoglycans, hyaluronic acid, and type VI collagen and functions as a protective cushion for chondrocytes, mediating their interaction with the surrounding ECM⁸²⁻⁸⁵. A chondrocyte with its PCM is called a chondron. The territorial matrix immediately surrounds the chondron. This region contains higher concentrations of proteoglycans and a dense network of collagen fibrils which are randomly organized and may serve to provide

structural support to the chondrocyte and the PCM^{76,85,86}. The interterritorial matrix is the largest region of the cartilage ECM, filling the space between the territorial matrices of adjacent chondrons. It is characterized by a lower concentration of cells and a less dense distribution of collagen fibrils and proteoglycans compared to the territorial matrix. This matrix contains primarily type II collagen and large aggregating proteoglycans like aggrecan and is responsible for the bulk of the mechanical properties of cartilage, such as its resistance to compression and tensile strength^{76,85}.

The functions of articular cartilage are fundamental for joint health and optimal function. Cartilage efficiently distributes mechanical loads across the joint surface, reducing stress on underlying bone structures and preventing damage. This even load distribution contributes to the long-term health and longevity of joints⁷⁶. Additionally, cartilage serves as an exceptional shock absorber, primarily due to the water-retaining properties of its proteoglycans. This capacity to absorb and dissipate impact forces protects the joint and prevents injury^{31,49,60,61}. The smooth, friction-reducing surface of articular cartilage, lubricated by synovial fluid, allows for seamless, painless joint movement^{68,74,75,87}. Cartilage adapts to joint congruity, conforming to the shapes of articulating bones. This feature ensures that joint surfaces fit together precisely, providing stability and uniform load distribution during movement. Moreover, it acts as a protective layer for the subchondral bone beneath, absorbing impacts that could otherwise damage the bone and maintaining its integrity³¹.

Cartilage is an avascular, aneural, and alymphatic tissue which leads to very limited intrinsic repair capacity⁸⁸. Despite the great importance of its many functions, the tissue relies exclusively on the ability of chondrocytes, the only resident cell type in cartilage, to maintain the ECM production and homeostasis. Chondrocytes play a pivotal role in the dynamic equilibrium of cartilage through their matrix synthesis and degradation activities. These processes are finely tuned to maintain the structural integrity of cartilage tissue. On the one hand, chondrocytes are engaged in anabolic activities, synthesising key ECM components. Growth factors like Insulin-like Growth Factor 1 (IGF-1) and TGF-® are instrumental in promoting ECM synthesis, ensuring the cartilage matrix's strength and resilience⁸⁹⁻⁹⁴. Conversely, catabolic processes are equally important for cartilage homeostasis. The matrix metalloproteinases (MMPs) MMP3 and MMP13, and the

Disintegrin and Metalloproteinase with Thrombospondin motifs (ADAMTS) enzymes ADAMTS4 and ADAMTS5, are responsible for the targeted degradation of collagens and proteoglycans, respectively^{30,95–97}. These catabolic enzymes are regulated by NF-κB signalling and cytokines including interleukin-1 (IL-1), interleukin-6 (IL-6) and Tumor Necrosis Factor alpha (TNF-α)^{98–100}. To prevent excessive matrix breakdown, these enzymatic activities are tightly regulated by tissue inhibitors of metalloproteinases (TIMPs), maintaining the delicate balance between synthesis and degradation within the cartilage matrix¹⁰¹.

Chondrocytes are remarkably sensitive to mechanical loading, and their metabolic activities are influenced by mechanical stimuli. Mechanical stress under physiological conditions can induce changes in chondrocyte gene expression, promote ECM synthesis, and modulate cell proliferation¹⁰². Abnormal mechanical loads, however, are detrimental to cartilage homeostasis shifting the balance towards catabolic processes.^{103–106}

Autophagy, a cellular process responsible for recycling damaged organelles and proteins, plays a critical role in chondrocyte maintenance. In chondrocytes, autophagy contributes to cell viability and function¹⁰⁷. Dysregulation of autophagy has been linked to cartilage degeneration and conditions like OA^{108,109}.

Chondrocytes operate in a unique microenvironment characterized by low oxygen levels, or hypoxia. In response to this hypoxic environment, chondrocytes adapt by relying on anaerobic glycolysis for energy production. Hypoxia-inducible factors (HIFs), particularly HIF-1α and HIF-2α, and their ratio within the cartilage tissue, play a pivotal role in orchestrating these metabolic adaptations^{110,111}. Additionally, due to cartilage's avascular nature, direct blood supply is limited. Consequently, chondrocytes rely on diffusion through the ECM for nutrient and waste exchange^{112,113}. Efficient nutrient uptake and waste removal are vital for chondrocyte survival and function, ensuring the maintenance of cartilage homeostasis in this specialized tissue.³⁰

Articular cartilage, with its intricate structure, multifaceted functions, and the finely tuned processes governing its development and maintenance, is a crucial component of the musculoskeletal system. Its unique composition and mechanical properties enable seamless joint movement, protect underlying bone, and maintain joint health. A deeper understanding of cartilage biology holds great promise for the development of

interventions and therapies to enhance joint function and alleviate the burden of joint diseases.

Synovial capsule

The synovial capsule, a complex, layered structure critical for the proper functioning of synovial joints, displays a highly organized composition essential for maintaining joint health and facilitating extensive mobility. Its outermost layer, known as the fibrous layer or membrane, is mainly composed of a dense connective tissue matrix, rich in collagen fibres, fibroblasts, and elastic fibres¹¹⁴. This arrangement endows the capsule with necessary tensile strength. The fibrous layer's role extends beyond providing structural integrity to the joint by limiting its movements within a physiological range and offering protection against external influences¹¹⁵. The fibrous layer of the synovial capsule is embedded with blood vessels and nerves. These blood vessels are crucial for delivering nutrients and oxygen, supporting the metabolic needs of both the capsule and synovium. The sensory nerve fibres within this layer transmit important information about joint positioning and pain, enabling the body to respond to changes in joint status or potential harm¹¹⁴.

Below this fibrous layer resides the synovial membrane, an intricate and crucial interface situated between the fibrous layer and the joint cavity. Characterized by a two-tiered structure, it comprises an outer layer of loosely arranged connective tissue, rich in blood vessels and lymphatics, and an inner layer of specialized cells called synoviocytes¹¹⁵. These synoviocytes, categorized into Type A (macrophage-like) and Type B (fibroblast-like), are responsible for producing the synovial fluid, a critical component for joint functionality^{112,116}. Macrophage-like synoviocytes, derived from monocytes, engage in phagocytosis and contribute to immune responses, producing inflammatory cytokines and enzymes in pathological conditions. Fibroblast-like synoviocytes resemble fibroblasts and are crucial for producing hyaluronic acid and lubricin, components of the synovial fluid. This viscoelastic fluid, rich in hyaluronic acid, lubricin, and proteins like albumin, chondroitin sulphate and globulins, encapsulates joint surfaces, reducing friction and facilitating smooth movement. It simultaneously nourishes the articular cartilage, which lacks its own blood supply, and aids in waste removal⁸⁷.

Synovial homeostasis, a balance between the constructive and destructive activities of synoviocytes, is influenced by cytokines and growth factors. Disruptions in this equilibrium can lead to synovitis, commonly seen in inflammatory joint diseases¹¹⁷. The synovium develops from mesenchymal stem cells during embryogenesis, with TGF- β and FGF pathways playing pivotal roles in this process. It is vital in maintaining a healthy joint environment, regulating fluid production and resorption to ensure normal joint function and pressure¹¹⁷.

Inflammatory cytokines like IL-1 β and TNF- α are crucial in modulating synovial function, particularly in joint diseases, by stimulating synoviocytes to produce inflammatory mediators and ECM-degrading enzymes^{118,119}. In contrast, anti-inflammatory cytokines such as IL-10 and IL-4 aid in preserving synovial homeostasis¹²⁰⁻¹²².

The synovium is involved in both innate and adaptive immune responses within the joint. Fibroblast-like synoviocytes can present antigens and activate T cells, playing a role in autoimmune responses in conditions like RA¹²³⁻¹²⁵.

Subchondral bone

In the dynamic architecture of synovial joints, the subchondral bone underpins the articular cartilage, serving as a critical structural component providing stability and it is integral to the mechanics of movement within the synovial joint system¹²⁶. The subchondral bone plate, a delicate cortical lamina situated beneath the calcified cartilage, exhibits significant porosity rather than being a solid barrier. This structure is permeated by channels that establish a direct connection between the articular cartilage and the underlying subchondral trabecular bone. These conduits facilitate the penetration of numerous arterial and venous vessels, along with nerves, into the calcified cartilage^{126,127}. The formation and distribution of these channels are influenced by factors such as aging and the magnitude of compressive forces experienced by the cartilage and subchondral bone in joints¹²⁷. Additionally, the channels' morphology and size vary according to the cortical plate's thickness, with different patterns observed in areas of different stress levels and plate thicknesses. The subchondral trabecular bone, arising from this plate, is integral for shock absorption and support in joints and, although not the main source of nutrients,

it plays a role in the nutrient supply and metabolic processes of cartilage¹¹³. The life cycle and sustainability of this bone component are orchestrated by the concerted efforts of osteoblasts, which lay down new bone, osteoclasts, which strategically resorb bone, and osteocytes, the sentinels embedded within the bone matrix¹²⁸. These osteocytes act as mechanosensors, responding to the ebb and flow of mechanical demands and coordinating the remodelling activities accordingly¹²⁹.

To maintain a state of homeostasis within the subchondral bone, a symphony of signalling pathways—comprising RANK/RANKL/OPG, Wnt/ β -catenin, and TGF- β —regulate the balance between the bone-forming and bone-resorbing cells^{129–132}. This meticulous control ensures that the bone remodelling aligns responsively with the needs imposed by mechanical stress and any injury to the joint.

Other joint tissues

Infrapatellar fat pad

In the intricate biomechanical landscape of synovial joints, fat pads, such as the infrapatellar fat pad (IFP) located in the knee, serve a multifaceted function. These specialized structures are comprised of adipose tissue and are distinguished by their rich innervation and vascularization, integral to their function and responsiveness to physiological as well as pathological stimuli¹³³. The anatomical positioning of fat pads within the joint capsule and synovium contributes to patellar stabilization during physical activity and shielding the knee from mechanical harm¹³⁴. IFP's pressure and volume dynamically alter with knee joint movement, providing stability at the extremities of motion¹³⁵. Additionally, it serves as a buffer between the patellar tendon and anterior tibial plateau. Removal of the IFP can lead to complications like patella baja, heightened patellofemoral joint pressure, anterior impingement, and knee pain¹³⁶.

Functioning like other adipose tissues, the IFP secretes various cytokines and adipokines, playing a role in joint disease pathogenesis^{137,138}. Its potential in tissue engineering, particularly as a source of reparative cells, is significant. Adipose-derived MSCs from IFP exhibit notable chondrogenic potential for inflammatory OA treatment^{139,140}.

Comparative studies of connective tissue progenitors from IFP, synovium, and periosteum reveal IFP as a preferred tissue source for cartilage repair in many patients¹⁴¹.

Meniscus

The meniscus is a crucial fibrocartilaginous element. Shaped like a crescent, it sits between the femur and tibia, made predominantly of collagen fibres intertwined with proteoglycans, facilitating its attachment to the tibial plateau and connections to surrounding ligaments¹⁴².

The meniscus plays several roles: distributing loads across the joint, acting as a shock absorber, and enhancing joint stability^{142,143}. It contains fibrochondrocytes, cells synthesizing its ECM^{144,145} which consists of collagens (predominantly type I and type II), proteoglycans and glycoproteins^{146,147}. Meniscal development involves mesenchymal cell condensation and differentiation¹⁴⁸, its growth and repair being regulated by growth factors like TGF- β and FGF¹⁴⁹. Meniscal integrity is crucial for joint health and functionality¹⁴³.

Tendons and ligaments

Tendons serve as robust yet flexible conduits of force between muscle and bone. Tendons display a hierarchical fibrillar structure, comprising collagen molecules, fibrils, fibres, fascicles, and the tendon unit, all aligned along the tendon's longitudinal axis for efficient tensile load transmission¹⁵⁰. Collagen, primarily type I, is crucial for tendon strength, complemented by other collagen types (III, V, IX, X, XI, XII)¹⁵¹ and proteins like proteoglycans and glycoproteins^{152,153}, which function together to enhance mechanical stability, facilitate tendon repair, and increase elasticity. The predominant cell type in tendons are tenocytes, fibroblast-like cells interspersed between the collagen fibrils, which maintain tendon homeostasis and repair¹⁵⁴. Tendon stem cells (TSCs) within tendons are key for self-renewal and differentiation, particularly under mechanical stress^{155,156}. Tendon properties, including stiffness and viscoelasticity, vary with tendon type, age, and strain rate, affecting their mechanical behaviour and energy absorption capacity¹⁵⁰.

Ligaments serve as stabilizers in joints by connecting bones. Predominantly composed of Type I collagen, known for its tensile strength, they also contain elastin fibres and proteoglycans for elasticity and resilience. Ligaments' structural hierarchy allows for load distribution, joint stability, and force absorption ^{157,158}. Ligaments guide joint movements, provide stability, and contribute to proprioception ¹⁵⁹. Fibroblasts within the ligaments maintain and repair the extracellular matrix. Ligament homeostasis balances matrix synthesis and degradation, influenced by mechanical stress and environmental factors ¹⁶⁰.

Due to their role in stabilising the joints, distributing load and absorbing mechanical force, injuries or dysregulation of the homeostatic environment of tendons and ligaments has a detrimental effect on the entire joint ¹⁶¹⁻¹⁶³.

1.2. Osteoarthritis

Joint homeostasis is maintained across all joint tissues through a delicate balance between anabolic and catabolic processes, regulated by a network of biochemical and mechanical signals. Disruption in this balance can lead to joint pathologies such as OA.

OA is a chronic degenerative joint disorder. The hallmark of OA is the progressive degeneration of articular cartilage. Although traditionally considered a non-inflammatory arthritis, synovitis (inflammation of the synovial membrane) is commonly observed in OA as well as remodelling and sclerosis of the subchondral bone and the formation of osteophytes at the joint margins. There can be secondary changes in the surrounding soft tissues, including thickening of the joint capsule, weakening or stiffening of ligaments, and potential involvement of periarticular muscles and tendons ¹⁶⁴ (Figure 1.2).

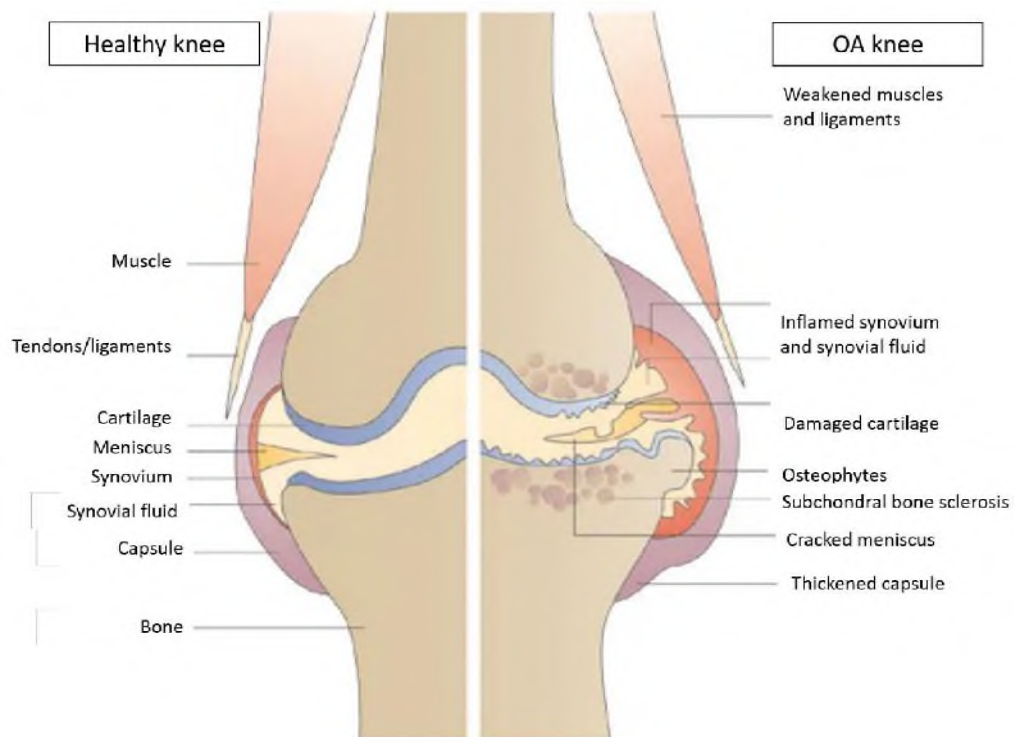


Figure 1.2. Features of a healthy (left) versus an OA affected (right) knee joint. [Adapted from: Wieland et al ¹⁶⁵]

1.2.1. The synovial joint and OA

Articular cartilage

The ECM of cartilage undergoes accelerated degradation in OA. While it is not yet fully understood whether metabolic dysregulation of chondrocytes precedes the onset of OA or is a response to this degradation, the cells' behaviour during the early stages of OA plays a crucial role in the disease's progression. At the initial stages of OA, chondrocytes appear to be activated in an attempt to repair and restore the damaged cartilage. This response involves an upregulation in the production of ECM macromolecules^{19,140,166}. This heightened activity is essentially a compensatory mechanism aimed at maintaining the structural integrity of the cartilage. However, in advanced age, the chondrocytes become less proliferative^{93,167–170} and have reduced abilities to synthesize the appropriate types and quantities of ECM macromolecules^{171,172}.

Furthermore, an increase in the accumulation of advanced glycation end products (AGEs) in the cartilage matrix is observed in old age and in OA. AGEs can crosslink with collagen fibres, making the matrix stiffer and more brittle, thus more susceptible to mechanical damage. AGEs also interact with receptors on chondrocytes (RAGE), inducing the production of pro-inflammatory cytokines such as IL-1 β and TNF- α and MMPs¹⁷³⁻¹⁷⁵.

The aged chondrocytes, with a limited regenerative capacity, cannot counteract the cartilage damage. As the disease progresses, the anabolic capabilities of chondrocytes are gradually overtaken by catabolic processes. These catabolic processes involve the activation of enzymes like MMPs and ADAMTSs, which break down the ECM components¹⁷⁶⁻¹⁷⁸. The net result is a homeostatic imbalance within the cartilage tissue. This imbalance tips the scales toward ECM degradation, leading to the weakening of the cartilage and ultimately its loss, especially in load-bearing parts of the joint.

Synovium

Changes in the synovium during OA can often even precede cartilage degradation¹⁷⁹.

In early stages of OA, synovial lining hypertrophy and hyperplasia, increased angiogenesis, a low level of inflammation and fibrosis are observed¹⁸⁰⁻¹⁸³. High levels of macrophage-like synoviocytes and synovitis can be found in over 50% of patients with OA, especially those in later stages of the disease^{184,185}. Synovitis is closely associated with the clinical symptoms of OA, particularly pain and joint swelling. The inflammatory mediators not only contribute to tissue damage but also sensitize nerve endings in the joint, leading to pain¹⁸⁶⁻¹⁸⁸.

The interplay between cartilage and synovium is a pivotal aspect of joint biology, especially in the context of joint diseases such as OA. This dynamic interaction involves a complex exchange of biochemical signals and mechanical feedback, significantly influencing the progression of the disease. The cartilage and synovium communicate via the synovial fluid, which facilitates the exchange of pro-inflammatory cytokines (e.g., IL-1 β , TNF- α) and catabolic enzymes from the inflamed synovium to the cartilage¹⁸⁹ and of cartilage matrix degradation products (e.g., fibronectin fragments, aggrecan breakdown products) into the synovium. These cartilage-derived degradation products can activate the

macrophage-like synoviocytes, leading to an increased production of pro-inflammatory cytokines, which further drive cartilage degradation, driving a regulatory loop¹⁹⁰.

Subchondral bone

The pathophysiology of OA involves significant changes beyond the articular cartilage and synovial membrane, notably in the subchondral bone. In OA, the subchondral bone undergoes increased turnover and remodelling¹⁹¹. This leads to alterations in bone architecture, including sclerosis and the formation of subchondral bone cysts. An imbalance between bone resorption and formation results in changes in the mechanical properties of the subchondral bone, which can exacerbate cartilage degradation as stiffening of the subchondral bone reduces its shock-absorbing capacity, increasing the load on the articular cartilage^{192,193}. Additionally, increased vascular penetration from subchondral bone into the calcified cartilage layer can occur^{194,195}. This can lead to the invasion of osteoblasts and osteoclasts, further disrupting the cartilage homeostasis.

Osteophytes at the joint margins are another hallmark of OA. They form as a result of the periosteal cell activation and endochondral ossification in response to mechanical stress and instability in the joint¹⁹⁶. In addition to osteophytes microfractures in subchondral bone can occur due to increased mechanical load and altered bone structure, and lead to bone marrow lesions associated with pain¹⁹⁷.

Other joint tissues

OA is a disease of the whole joint with all tissues being affected. The infrapatellar fat pad (in the knee) and other periarticular adipose tissues can become inflamed in OA, contributing to the inflammatory milieu of the joint. The adipose tissue can secrete adipokines (e.g., leptin, adiponectin) and cytokines that promote inflammation and pain. It may also impact the biomechanical environment of the joint¹⁹⁸.

Meniscal tears and degeneration are common in OA. The meniscus loses its elasticity and becomes more prone to injury. Damaged menisci can exacerbate cartilage wear by impairing the normal load distribution across the joint^{199,200}.

Tendons around an arthritic joint can become stressed or damaged due to altered joint mechanics and inflammation. Tendon pathologies can affect joint stability and alignment, further worsening OA symptoms by altering the normal biomechanical forces across the joint²⁰¹.

Ligaments also experience heightened stress due to changes in joint mechanics caused by OA-related cartilage degeneration. They might get overused as they compensate for joint instability, making them more susceptible to injury or strain^{201,202}.

1.2.2. Risk factors

OA is a complex condition influenced by a multitude of risk factors. Age, sex, genetic predisposition and various environmental factors such as lifestyle, occupation, and obesity play significant roles in the development and progression of OA. Understanding these factors is crucial for both prevention and management of the condition.

Age is the strongest predictor of OA. As individuals age, the prevalence and severity of OA increase, attributed to age-related changes in joint tissues^{203,204}. These include reduced cellularity and synthetic activity of chondrocytes, increased stiffness of the cartilage matrix, and a diminished capacity for repair^{31,167–171}. Additionally, the cumulative effect of mechanical stress on joints over time leads to wear and tear of cartilage, exacerbated by a decline in the quality of joint tissue repair mechanisms^{31,169–171,205}.

Biological sex has also been associated with the risk of developing OA. Women, particularly post-menopausal, are at a higher risk compared to men⁴. This disparity is thought to be linked to hormonal changes associated with menopause, as oestrogen is believed to play a protective role in cartilage health^{206,207}. Differences in joint anatomy, muscle strength, and biomechanics between men and women further contribute to this risk variance^{208,209}.

Environmental factors significantly impact the predisposition and progression of OA. Certain occupations involving repetitive joint motions, heavy lifting, or prolonged standing can put excessive stress on joints, accelerating cartilage wear^{210–213}. Lifestyle choices, including physical activity and diet, also play a role in OA risk. While moderate exercise is beneficial for joint health^{214,215}, excessive high-impact activities can increase the risk of

OA²¹⁶. A sedentary lifestyle can weaken joint-supporting structures, leading to weight gain and further stressing the joints^{217,218}. Poor nutrition may exacerbate inflammatory processes within the joint²¹⁸. Obesity is also a significant risk factor for OA, particularly in weight-bearing joints^{219,220}. The excess load from obesity can accelerate the breakdown of cartilage, leading to early onset and rapid progression of OA^{220,221}. Additionally, adipose tissue is an active endocrine organ that produces pro-inflammatory cytokines, contributing to a systemic inflammatory state that affects joint tissues²²². Obesity often coexists with metabolic syndrome, including conditions like diabetes mellitus, dyslipidaemia, and hypertension, which can influence systemic inflammation and have direct effects on joint tissues²²²⁻²²⁴.

The genetic contribution to the development of OA is also a significant factor, accounting for a substantial portion of the disease's risk. Research has shown that genetics plays a crucial role in the susceptibility to OA, with hereditary factors influencing up to 40-65% of OA cases depending on the joint site²²⁵.

1.2.3. *Diagnosis and treatment*

The diagnosis of OA begins with a clinical evaluation, including a patient's medical history and a physical examination. Key symptoms assessed are joint pain, morning stiffness, and functional impairment. The physical exam can reveal joint tenderness, crepitus, limited mobility, and swelling. For imaging, X-rays are common but have limitations such as not showing early cartilage loss. Magnetic Resonance Imaging (MRI) is more detailed, useful for measuring early cartilage changes, and for observing changes to other soft tissues. Blood tests, while not specific for OA, help exclude other inflammatory conditions²²⁶.

The treatment of OA is multifaceted, involving non-pharmacological interventions, pharmacological treatment, and in some cases, surgical interventions. Non-pharmacological approaches include regular exercise and physical therapy to improve joint mobility and muscle strength, weight management for overweight or obese patients to reduce joint load and pain, and the use of assistive devices for load distribution and joint stabilization²²⁷. Pharmacological treatment primarily focuses on pain management with

medications such as acetaminophen and Nonsteroidal Anti-Inflammatory Drugs (NSAIDs). Topical NSAIDs can be used for localized pain. Intra-articular injections, including corticosteroids and hyaluronic acid, are other options, though the efficacy of hyaluronic acid injections remains debated. Research is ongoing to develop Disease-Modifying OA Drugs (DMOADs) that can modify the disease process and prevent cartilage degradation²²⁸.

In advanced OA or when conservative management fails, surgical options such as arthroscopy, osteotomy, and arthroplasty may be considered. Arthroscopy is used for joint cleaning or addressing specific issues like meniscal tears, osteotomy involves cutting and realigning bones to relieve joint pressure, and arthroplasty involves replacing a severely damaged joint with an artificial one, common in knee and hip OA^{226,229}.

The approach to diagnosing and treating OA is comprehensive, integrating clinical evaluation with a combination of non-pharmacological, pharmacological, and surgical strategies, tailored to the individual needs of the patient. Emphasis on early diagnosis and comprehensive management aims to alleviate symptoms, improve function, and enhance quality of life. Ongoing research into OA's pathophysiology continues to inform and refine therapeutic options, with an increasing focus on disease-modifying treatments.

1.2.4. OA prevalence

The Global Burden of Disease Study 2019 categorizes musculoskeletal conditions into five specific categories: rheumatoid arthritis, OA, low back pain, neck pain, and gout, with an additional residual category encompassing various acute and chronic conditions affecting the locomotor system, including bones, joints, ligaments, tendons, and muscles²³⁰. OA, a significant constituent of these disorders, is a major cause of pain and disability worldwide. The burden of OA, along with other musculoskeletal disorders, has been on the rise globally. In 2020, there were an estimated 595 million people affected by OA (14.8% of the global population older than 30 years), a substantial increase from 221 million in 1990²³¹. The increasing prevalence of musculoskeletal disorders, including OA, underscores a growing public health challenge. Factors such as aging populations, the rising prevalence of obesity, and improved diagnostic capabilities contribute to this trend.

The epidemiology of OA varies, influenced by age, sex, geographic location, and specific joints affected. Knee OA is the most common form, followed by hip and hand OA. Spinal OA, particularly in the lumbar and cervical regions, is also prevalent in the older population²³¹.

The prevalence of OA is more pronounced in developed countries, attributed to longer life expectancy and higher obesity rates²³¹. In the UK, over 10 million people (one in six adults over the age of 45) are affected by the condition²³². Its prevalence rises sharply in individuals over 45, especially in populations aged 60 and older.

Significant geographical variations in OA prevalence exist, with knee OA more common in North America and Europe compared to Asia and Africa. These variations may be due to differences in genetic background, environmental exposures, lifestyle factors, and occupational activities. Additionally, the prevalence of OA varies among different ethnic groups, with African Americans having higher rates of knee OA but lower rates of hip OA compared to Caucasians²³¹.

By 2050, global estimates predict an increase in case numbers for OA of different joints of between 48.6% and 95.1%, with the increased rates varying between geographic regions of between 40% and over 200%²³¹.

The rising prevalence of OA, driven by demographic shifts and the obesity epidemic, poses significant challenges for healthcare systems. This includes increased demands for medical care, joint replacement surgeries, and long-term management of chronic pain and disability. Consequently, there is a pressing need for comprehensive public health strategies focusing on prevention, early intervention, and effective management to mitigate the impact of this chronic and debilitating disease.

1.2.5. Socioeconomic burden of OA

The socioeconomic impact of OA is substantial, reflecting its high prevalence and the chronic nature of the disease. The complexity of its socioeconomic burden is multi-dimensional, encompassing direct healthcare costs, indirect costs due to lost productivity and disability, and a considerable impact on quality of life.

Direct healthcare costs associated with OA include medical treatments and interventions such as pharmacological treatments for pain and inflammation, physical therapy, and surgical interventions ^{233,234}. These treatments contribute to a significant portion of healthcare expenditures, with joint replacement surgeries, in particular, being costly procedures increasingly in demand. Hospitalisations and regular physician visits for disease management also add to these costs^{233–235}.

The indirect costs of OA are substantial, largely due to lost work productivity. OA is a leading cause of work disability, and the limitations it imposes can reduce an individual's ability to work, leading to absenteeism, decreased productivity, and often early retirement. The economic impact of lost work productivity often surpasses the direct medical costs associated with OA. The cost of working days lost due to the two most prevalent forms of arthritis (OA and RA) are predicted to reach £3.8 billion in the UK by 2030²³⁵. Moreover, chronic pain and mobility limitations can lead to disability, necessitating assistance with daily activities and often requiring informal caregiving or formal caregiving services, adding further to the socioeconomic burden^{235,236}.

The quality of life for individuals with OA is significantly impacted due to chronic pain, reduced mobility, and functional limitations. These factors can lead to social isolation, depression, and reduced participation in community and family activities. Up to 20% of people suffering from OA report feelings of depression and anxiety^{237,238}. OA is also often associated with comorbid conditions such as obesity, cardiovascular disease, and diabetes, which adds complexity and cost to patient care^{224,239}.

Managing OA requires long-term strategies, including ongoing medical care and lifestyle modifications. The allocation of resources for OA treatment and management competes with other healthcare needs, posing challenges for healthcare policy and planning worldwide. The World Health Organisation (WHO) has designated 2021-2030 as a decade dedicated to healthy ageing and management and prevention of musculoskeletal conditions such as OA are crucial to achieving this goal ²⁴⁰.

1.2.6. Genetics of OA

As mentioned, OA has a large genetic component and studies into the molecular genetics of OA have revealed that it is a complex polygenic disease with a multitude of molecular mechanisms and cellular pathways being (dys)regulated by genetic variability.

Given the significant genetic influence on OA, extensive research has been conducted to pinpoint the specific genetic variations responsible for the development of OA pathology. Despite the examination of numerous single nucleotide polymorphisms (SNPs) in candidate gene studies for their potential association with OA, only a single SNP (rs143383), located within the *GDF5* gene²⁴¹, was eventually identified as significantly associated to the disease at a genome-wide significance level ($P < 5 \times 10^{-8}$) in subsequent investigations²⁴².

GWAS-era

The advent of array-based Genome-Wide Association Studies (GWAS) marked a significant advancement in identifying genetic loci associated with an increased risk of many complex polygenic diseases, including OA. These studies were instrumental in uncovering many disease-associated loci that were not previously known. However, these studies often found variants with small effect sizes and explained only a small fraction of the heritability of the traits. Improvements in genomic sequencing technologies and the reduction in costs have allowed for more comprehensive studies, including whole-genome sequencing²⁴³. The development of more sophisticated statistical methods has improved the power to detect associations and interpret the results²⁴⁴.

In the context of OA, GWAS have successfully identified both common and rare genetic variants that contribute to OA risk. Over 120 distinct genetic loci have been linked to OA through GWAS to date^{241,242,245–266}, encompassing a wide range of biological pathways, including those involved in joint health and homeostasis maintenance: *COL11A1* (rs4338381)²⁶⁵, *COL27A1* (rs1078301)²⁶², *TGFB* (rs75621460)²⁶⁵, *LTBP1* (rs2061026)²⁶⁵, *LTBP3* (rs10896015)²⁶⁵, *GDF5* (rs143383)²⁴², *BMP5* (rs80287694)²⁶⁵, *SMAD3* (rs12901071)²⁶², *RUNX2* (rs2396502)²⁶⁵, *SOX9* (rs8067763)²⁶⁵, *FGFR18* (rs3884606)²⁶⁵, *TGFA* (rs3771501)²⁶⁵, *TGFB1* (rs75621460)²⁶⁵, *IL11* (rs4252548)²⁶⁵, *HLA* (rs7775228)²⁶⁷, *HLA-DPA1*

(rs9277552)²⁶⁵. This multitude of genetic signals highlights the polygenic nature of OA, where multiple genetic variations each contribute a small effect to the overall risk of developing the disease²⁶⁸. The interaction of these genetic factors with environmental influences, like obesity, joint injury, and physical activity, further complicates the genetic landscape of OA.

Post GWAS-era

In what is now termed the “post-GWAS era”, the focus has shifted from identifying genetic associations to understanding their functional impact and clinical significance²⁶⁹.

While a significant number of genetic signals associated with OA have been identified, they account for less than half of the disease's heritability²⁶⁸. This reflects a common challenge in the post-GWAS era for many diseases²⁶⁹. The difficulties underlying this persistent issue include very small effect sizes, small minor allele frequencies which require large sample sizes to be investigated, the SNPs being part of large LD blocks making determining the causal variant difficult^{263,269}, and the fact that most of these signals reside within the non-coding genome with unknown gene targets²⁶³. The current prevalent hypothesis is that these non-coding SNPs mediate their effect by preferentially binding TFs and resulting in gene expression changes at their target gene(s). Such variants are known as expression Quantitative Trait Loci (eQTLs)²⁷⁰. Such eQTLs often overlap with functional regulatory elements like promoters and enhancers, open-chromatin regions and TF binding sites, and are *cis*-acting, affecting the expression of nearby genes^{263,271,272}.

The identification of eQTLs by correlating SNP genotypes with gene expression levels has been a widely used method of functional investigation into GWAS signals for many common diseases, including OA. Another approach, allelic expression imbalance (AEI), involves analysing transcript SNPs to assess if the risk allele correlates with altered gene expression²⁷⁰. Several OA-risk loci have been found to act as eQTLs using the AEI approach, including: rs10948172 (*RUNX2*)²⁷³, rs143383 (*GDF5*)²⁷⁴, rs12915901 (*ALDH1A2*)²⁷⁵, rs4764133 (*MGP*)²⁷⁶, rs75621460 (*TGFB1*)²⁷⁷, rs11780978 (*PLEC*)²⁷⁸, and rs6516886 (*RWDD2B*)²⁷⁹.

While a lot of these genes have a known OA-biology relevant function, many OA-associated SNPs are near genes with unknown disease-specific functions, suggesting the need for more in-depth studies of cell regulatory networks. It's also noted that there is considerable variation in eQTLs between tissues, cell types, and specific OA joint or sex associations²⁸⁰. Some signals may not directly influence OA in old age but instead predispose individuals early in life^{95,281–283}, potentially through processes like antagonistic pleiotropy, where a genotype beneficial in development may have adverse effects with aging²⁸⁴.

Steps towards the understanding of the interplay between genetic susceptibility (as identified by GWAS) and environmental/lifestyle factors are being made. This includes studying genome-epigenome interactions.

Epigenetics of OA

The development of our understanding of OA has been profoundly shaped by the increasing recognition of epigenetic mechanisms as pivotal elements in its pathology. These occur at the interface of genetics and the environment and represent heritable changes in gene expression without alterations to the underlying DNA sequence²⁸⁵. There are three major types of epigenetic mechanism: post-translational modifications (PTMs) of histones, non-coding regulatory RNAs, and chemical modifications to DNA, such as DNA methylation (DNAm). All three are crucial for the regulation of chondrocyte function, the maintenance of cartilage homeostasis, and the modulation of inflammatory responses, all of which are central to OA development and progression.

Histone modifications

Among the most studied types of histone PTMs are methylation and acetylation, with each having been extensively investigated in the context of OA.

Histone methylation-demethylation is carried out by histone methyltransferases (HMTs) and histone demethylases (HDMTs). The impact of histone methylation depends on the specific lysine (K) or arginine (R) residue that is modified in the histone tail, as well as the

number of methyl groups added. This modification can either activate or silence gene transcription²⁸⁶. While the precise histone PTMs implicated in OA have not yet been fully characterized, there are indications that they may serve as epigenetic regulators in the development of the disease²⁸⁷.

In the context of OA, elevated levels of trimethylation of H3K9 (H3K9me3) and H3K27 (H3K27me3) have been observed in the promoter region of the *SOX9* gene. The increased H3K9me3 and H3K27me3 levels contribute to the reduced expression of *SOX9* seen in OA²⁸⁸. Inhibition of a H3K9-specific methyltransferase in a chondrocyte cell line resulted in increased apoptosis, upregulation of ECM-degrading enzymes, and downregulation of anabolic factors, disrupting chondrocyte homeostasis²⁸⁹. Additionally, mechanical stimulation of chondrocytes was found to impact histone methylation, affecting chromatin structure and chondrocyte differentiation²⁹⁰. *DOT1L*, an H3K79 methyltransferase, has shown chondroprotective effects in both humans and mice by regulating the Wnt pathway²⁹¹⁻²⁹³.

HDMTs also contribute to OA pathogenesis. *KDM6A* (encoding an H3K27 demethylase) is involved in chondrogenic differentiation, and its depletion leads to transcriptional silencing and reduced chondrogenesis²⁹⁴. In contrast, *KDM6B* (also known as *JMJD3*) is highly expressed in OA and promotes inflammation and cartilage degradation. Inhibition of *KDM6B* can reverse its deleterious effects and promote cartilage synthesis²⁹⁵.

Histone acetylation modifies lysine residues located within the N-terminal tails of histones, leading to a reduction in their positive charge. This alteration weakens interactions between histones and between histones and DNA, ultimately resulting in the unwinding of the nucleosomal structure²⁸⁶. Histone acetylation is catalysed by histone acetyltransferases (HATs), while its reversal is catalysed by histone deacetylases (HDACs). Typically, HAT activity is associated with the activation of gene transcription, while HDACs are considered repressive marks²⁸⁶.

Although the role of HATs in OA remains relatively underexplored, there are promising examples that could serve as potential therapeutic targets. For instance, p300/CBP-associated factor (*PCAF*), a HAT, is upregulated in OA cartilage and in chondrocytes stimulated by TNF- α ²⁹⁶. *PCAF* is implicated in the NF- κ B pathway, suggesting its importance in regulating the inflammatory response in cartilage. Inhibition of *PCAF*

protein activity using salidroside, a natural compound found in *Rhodiola rosea*, has shown promise in suppressing inflammation and endoplasmic reticulum (ER) stress, both characteristic features of OA²⁹⁶.

Sirtuins, which are HDACs, have emerged as potential therapeutic targets. Notably, SIRT1 undergoes cleavage by cathepsin B when stimulated by TNF- α in human OA chondrocytes. This cleavage results in the generation of N-terminal and C-terminal fragments, leading to the loss of its deacetylase activity²⁹⁷. In the context of OA, researchers observed an elevated NT/CT SIRT1 ratio in early OA in human samples²⁹⁸. Silencing of SIRT1 resulted in the downregulation of chondrogenic markers, while activation of SIRT1 promoted autophagy in human chondrocytes²⁹⁹. SIRT6 has been associated with protection against aging-related pathologies and has demonstrated a role in OA by regulating chondrocyte redox balance³⁰⁰. SIRT7 appears to aggravate OA. Studies have shown that SIRT7 knockout mice are resistant to OA induced by aging or mechanical loading³⁰¹.

Non-coding RNAs

Non-coding RNAs such as microRNAs (miRNAs), circular RNAs (circRNAs), and long non-coding RNAs (lncRNAs) have also been investigated in the context of OA.

One of the noteworthy miRNAs in OA is miR-140, which is specifically expressed in cartilage and plays a critical role in chondrogenesis^{302–304}. It is located within an intronic region of the *WWP2* gene and generates two strands: miR-140-5p and miR-140-3p. Both of these are downregulated in OA chondrocytes³⁰⁵. MiR-140's regulatory sequence contains binding sites for transcription factors like SMAD3, and hypermethylation of CpG sites in that sequence in OA chondrocytes influences its expression and impacts the regulation of target genes such as *MMP13* and *ADAMTS5*^{304,306}.

Another important miRNA, miR-146a, is highly expressed in early OA cartilage³⁰⁷, while miR-146b is upregulated in OA cartilage³⁰⁸. In OA, the hypermethylation of its promoter in synoviocytes reduces miR-146a expression, resulting in increased expression of inflammatory factors³⁰⁹. MiR-146a expression can also be induced by IL-1 β stimulation in normal articular chondrocytes³⁰⁷. Interestingly, the modulation of miR-146a, along with

other miRNAs, has been shown to influence chondrocyte responses to mechanical loading³¹⁰.

Circular RNAs (circRNAs) represent a distinct class of non-coding RNAs known for their ubiquity, stability, and conservation across mammalian species³¹¹. Certain circRNAs function as competing endogenous RNAs that can naturally sequester miRNAs, effectively inhibiting their activity³¹².

In OA cartilage, 16 circRNAs were found to be upregulated, while 55 were downregulated³¹³. Notably, circRNA-100876 displayed increased expression following stimulation with IL-1 and TNF- α , paralleled by elevated *MMP13* expression³¹³. Further investigation revealed that circRNA-100876 contains binding sites for miR-636, miR-665, miR-217, miR-646, and miR-136. Intriguingly, miR-136 also targets the 3' untranslated region (UTR) of *MMP13*, indicating that circRNA-100876 is targeted by miRNAs that also target *MMP13*³¹³.

Circ-0136474 was found to be overexpressed in knee OA cartilage, leading to the downregulation of miR-127-5p. This miRNA, in turn, was also found to directly target the 3' UTR region of *MMP13*. Silencing miR-127-5p positively regulates *MMP13* expression at both the mRNA and protein levels, while miR-127-5p overexpression reduces *MMP13* levels³¹³.

Long non-coding RNAs (lncRNAs) have also been implicated in OA pathology. In an investigation of human hip cartilage, 198 lncRNAs exhibiting differential expression between OA and healthy individuals were discovered³¹⁴.

tRNA-derived fragments (tRFs) have also been reported to play a role in OA. In particular, tRF-50009A was reported to promote autophagy and protect against cartilage degradation in human knee samples from OA patients by inhibiting the mTOR pathway³¹⁵. Another tRF, tRF-3003a was induced in chondrocytes under inflammatory conditions and the data suggests that it prevents the pro-inflammatory signaling of the JAK/STAT pathway³¹⁶.

Small nucleolar RNAs (snoRNAs) are essential for ribosome functionality and play significant roles in cartilage homeostasis, ageing, and OA. A number of snoRNAs have showed changes in expression in cartilage ageing and OA, with SNORD26 elevated in OA³¹⁷

and SNORD44 and SNORD78 reduced in ageing³¹⁷. These changes were reported to influence chondrocyte differentiation, ECM production, and ribosome specialization.

Non-canonical snoRNAs, like SNORD33, SNORD35A, and SNORD116, are linked to oxidative stress^{318,319}, contributing to OA progression. These snoRNAs show differential expression patterns in minimally versus advanced OA cartilage degradation, suggesting snoRNAs are critical regulators of chondrocyte biology³¹⁷.

These findings on non-coding RNAs in the context of OA suggest a diverse and potentially therapeutically targetable role of these molecules in the progression of the disease.

DNA methylation (DNAm)

DNAm is extensively studied in the context of OA. It primarily involves the addition of a methyl group (-CH₃) to cytosine (C) nucleotides located 5' to guanine (G), resulting in the formation of 5-methylcytosine (5-mC). Key enzymes responsible for DNAm include DNA methyltransferases 1, 3A, and 3B (DNMT1, DNMT3A, DNMT3B). DNMT3A and DNMT3B establish *de novo* methylation patterns, while DNMT1 maintains methylation patterns³²⁰. In contrast, the process of DNA demethylation is carried out by ten-eleven translocation enzymes 1, 2, and 3 (TET1, TET2, TET3), converting 5-mC into 5-hydroxymethylcytosine (5-hmC). This epigenetic modification occurs in CpG dinucleotide sequences, particularly in CpG islands which are commonly found in gene promoter regions.

The interplay between DNAm and transcription factors (TFs) is a critical aspect of gene regulation and cellular function. This interaction involves both direct and indirect mechanisms. DNAm directly influences TF binding in several ways. Some TFs are sensitive to the methylation status of their binding sites, and when DNA is methylated at these sites, it can prevent the binding of these TFs, inhibiting gene transcription^{321,322}. For example, the CTCF transcription factor, crucial for organizing chromatin structure, has binding sites often inhibited by methylation^{323,324}. On the other hand, certain TFs preferentially bind to methylated DNA. These TFs, such as the MBD (methyl-CpG-binding domain) family, recognize and bind to methylated CpG sites. They can either repress transcription by recruiting co-repressors and HDACs or activate transcription in specific contexts^{325,326}. Indirect effects of DNAm on TF activity also play a significant role. DNAm can lead to a more

condensed chromatin state, especially when coupled with histone modifications like deacetylation. This chromatin compaction can limit TFs' access to their binding sites on DNA, contributing to gene silencing and the maintenance of cellular identity³²⁷. DNAm can also directly impact the shape of the DNA molecule and affect the chromatin landscape^{328,329}.

In the context of OA pathogenesis, DNAm plays a crucial role. For instance, the *COL9A1* gene exhibits hypermethylation at six CpG sites in the hips of OA patients. This hypermethylation leads to downregulated *COL9A1* expression, contributing to the loss of cartilage integrity³³⁰. DNMT1 and DNMT3A were found to hypermethylate the *PPARG* promoter, resulting in the suppression of PPAR γ expression in both human and murine knee OA, thus exacerbating the disease. However, the use of the pharmacological DNA demethylating agent 5-Aza-2'-deoxycytidine restored PPAR γ expression in the mice knee, providing chondroprotection by reducing oxidative stress and inflammation³³¹. Furthermore, inhibiting DNMT1 and DNMT3A in a murine model protected articular cartilage against OA progression, indicating their therapeutic potential³³¹. Knockdown of DNMT3B partially triggered early OA onset and progression in the knee joint, affecting the tricarboxylic acid cycle and mitochondrial respiration³³². Conversely, DNMT3B overexpression in knee chondrocytes enhanced chondroprotection by increasing *COL2A1* expression and reducing that of *RUNX2* and *MMP13*³³².

TET enzymes, particularly TET1, significantly contribute to OA pathogenesis. TET1 deficiency protects joints in a murine OA model, preventing knee cartilage surface damage and osteophyte formation³³³. Elevated 5-hmC levels were observed in *MMP3* and *MMP13* gene promoters in response to TET1 inhibition in knee chondrocytes from OA patients, resulting in cartilage-degrading MMP overexpression³³³.

Hypomethylation is another essential process in the altered synthesis of OA-related molecules. Hypomethylation of CpG sites in genes encoding MMPs and ADAMTS led to increased enzyme expression in OA cartilage^{334–336}. Furthermore, hypomethylation of a *NOS2* enhancer region in mice resulted in iNOS (inducible Nitric Oxide Synthase) overexpression in OA, promoting inflammation in the knee joints³³⁷. Demethylation of an *IL-8* promoter region increased proinflammatory chemokine expression in OA chondrocytes and in synovial tissue, which was associated with ECM proteoglycan loss,

MMP production, chondrocyte hypertrophy, apoptosis, immune cell infiltration into synovial tissue, and other detrimental effects^{338,339}. Similarly, hypomethylation of the *IL-6* promoter in OA was linked to IL-6 overexpression in synovial fibroblasts and synovial fluid, exacerbating the disease³⁴⁰. In rats with knee OA, hypomethylation occurred in CpG sites of genes related to cell differentiation, proliferation, and apoptosis, leading to increased expression of these genes³⁴¹. Hypomethylation also led to increased gene expression levels of *RUNX2*, promoting enhanced transcription of *MMP13*³⁴².

Epigenome-wide studies in OA-relevant tissues have revealed that DNAm is dynamic and highly tissue-, joint site-, age- and sex-specific^{343–351}. The interplay between genetic variation and DNAm has been extensively studied. Correlations between SNPs and DNAm levels at one or more CpGs have revealed multiple methylation QTLs (mQTLs) which can act as intermediates in the regulation of gene expression by SNPs, with correlations between gene expression and DNAm marking methylation-expression QTLs (meQTLs).

Over 20 OA GWAS signals have been shown to act as mQTLs in OA relevant tissues^{263,264,273,277,351–355}, including SNPs associated with genes with a known role in joint biology and homeostasis such as *GDF5* (rs143383)³⁵³, *RUNX2/SUPT3H* (rs10948155, rs1997995)²⁷³, and *TGFB1* (rs2820443)²⁷⁷.

1.2.7. OA susceptibility locus marked by the polymorphism rs1158364

In a GWAS using data from the UK Biobank, a significant association was discovered between the SNP rs11583641 (C>T) located on chromosome 1q25.3 and hip OA ($P = 5.6 \times 10^{-10}$)²⁶⁵. Subsequently, an analysis of the cartilage methylome revealed that rs11583641 functions as a mQTL. It was observed that the presence of the major C allele of rs11583641, associated with the development of OA, correlated with a reduction in DNAm at a CpG site, cg18131582³⁵¹. Both the SNP and the CpG site are located within the gene body of *COLGALT2*. Additionally, it was found that the expression of *COLGALT2* was significantly increased in hip cartilage affected by OA when compared to non-OA cartilage. These findings have led to the prioritization of further investigation into the *COLGALT2* gene and its potential regulatory element³⁵¹.

COLGALT2 (also known as *GLT25D2*) encodes an enzyme called COLGALT2, which is one of the two galactosyltransferases known (the other being encoded by *COLGALT1* (also known as *GLT25D1*)) to initiate collagen glycosylation, a collagen PTM³⁵⁶. Given that collagen constitutes a significant portion of articular cartilage's dry weight, any abnormalities in its folding and stability could lead to a loss of tissue integrity and the breakdown of cartilage. This underscores the significance of *COLGALT2* and the rs11583641 locus as promising targets for further investigation into the genetics of OA predisposition.

1.3. Collagens

1.3.1. Biosynthesis, structure, and function

The biosynthesis of collagens is a multifaceted process that begins with the transcription of collagen genes into mRNA, followed by the translation into procollagen polypeptides^{96,357}. These polypeptides undergo critical PTMs³⁵⁸. The triple helix, a hallmark of collagen, is then formed and stabilized by hydrogen bonds. The helix is left-handed and stabilized by the repetition of the amino acid sequence Gly-X-Y, predominantly with proline and hydroxyproline in the X and Y position^{357,359}. This procollagen is transported to the Golgi apparatus, secreted into the extracellular space, and then transformed into tropocollagen through the cleavage of non-helical terminal regions. The cleaved collagens are then assembled into fibrils^{357,359} (Figure 1.3.).

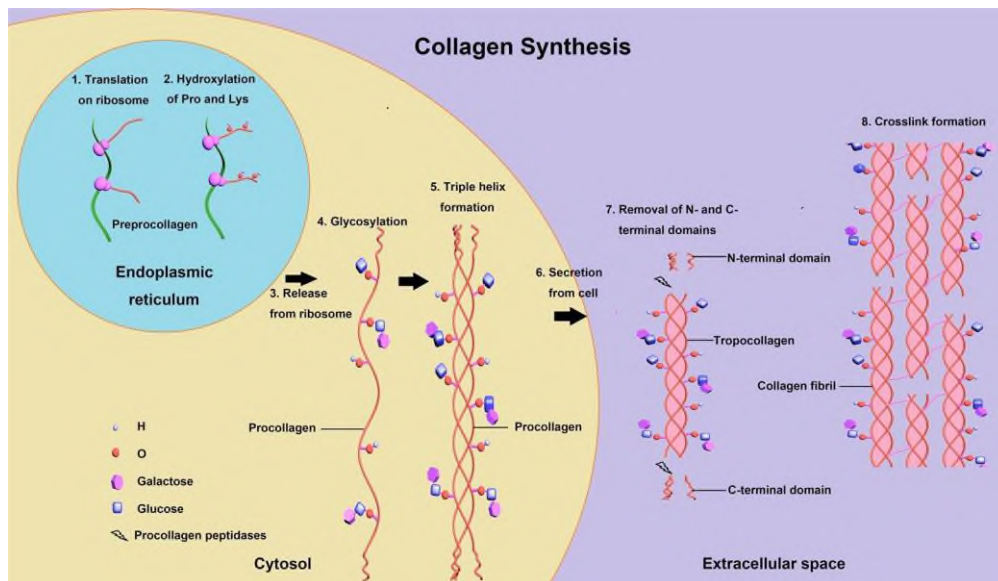


Figure 1.3. Collagen synthesis and PTMs. [Taken from <https://step1.medbullets.com/biochemistry/102078/collagen> , accessed on 25.09.2023]

The synthesis process is delicately regulated by cellular and environmental factors, ensuring synchronization with the body's needs³⁶⁰. The regulation of collagen biosynthesis at various levels – genetic, enzymatic, and environmental – is central to the process, as any disruptions can lead to a range of disorders^{361,362}. This regulation underscores the importance of collagen synthesis in maintaining tissue integrity and function.

The molecular diversity of collagen types, each with distinct structural attributes and tissue-specific distribution, speaks to the adaptability of this protein family. This variety, resulting in different physical and functional properties, allows collagens to perform specialized roles across a spectrum of tissues including skin, bone, cartilage, and blood vessels^{357,363,364}.

The collagen family encompasses 28 currently known types, each with unique structural features, functions, and tissue distributions (Table 1.1)³⁵⁷. The unique properties of these collagen types stem from variations in their amino acid sequences and post-translational modifications, which influence their assembly into fibrils and networks. These structural variations are encoded by a diverse set of genes, reflecting the genetic complexity underlying the collagen family (Table 1.1). The crucial role of this protein family in maintaining the structural integrity and functionality of various organs and tissues is

highlighted by the range of diseases associated with malformations and dysfunction of these collagens and their modifications^{361,365–379}.

Collagens can be subdivided into different categories based on their main functions and structure^{357,359}. Fibril-forming collagens include types I, II, III, V, and XI collagens. They form the long, thin fibrils seen in connective tissues like skin, tendons, and bone. The network-forming collagen group includes type IV collagen. Fibril-Associated Collagens with Interrupted Triple Helices (FACIT) collagens include types IX, XII, XIV, and others. They are typically associated with fibril-forming collagens and are found in tissues that require more flexibility, like cartilage. Collagen types XIII, XVII, XXIII, and XXV are transmembrane collagens and are part of the cell membrane, playing roles in cell adhesion and communication. Each type of collagen is adapted to its specific role in the body, whether it's providing structural support, elasticity, or facilitating cell interactions.

The diversity of this protein family allows collagens to fulfil a multitude of distinct functions within the body^{357,359,361,380}. Collagens provide structural support, essential tensile strength and resistance to intermittent pressures. Their functions in soft tissues contribute to elasticity and the tissue's ability to repair and remodel. Collagens are also actively involved in cell adhesion, cell-to-cell contact and signalling.

Table 1.1. The collagen superfamily [Based on Myllyharju, 2001³⁶¹ and Ricard-Blum, 2011³⁵⁷]

1.3.2. Collagen PTMs

Collagen PTMs are crucial biochemical processes that occur after the synthesis of the collagen polypeptide chain. These modifications are essential for the proper functioning, stability, and assembly of collagen fibrils. There are several types of PTMs, each mediated by specific enzymes and regulated by various factors: hydroxylation, glycosylation, cross-linking and telopeptide cleavage. These modifications are precisely regulated and are fundamental to the structural integrity of various tissues where collagen is a major component^{359,404,405}. Understanding these processes is vital not only for comprehending normal tissue physiology but also for recognizing the pathogenesis of a range of collagen-related disorders.

Collagen PTMs play an important role in the formation of collagen fibrils and their stability. Prior to secretion into the ECM, procollagen is subjected to modifications in the ER and Golgi apparatus. Proline and lysine hydroxylation are essential for stabilizing the triple helix structure of collagen^{405–408}. Glycosylation of hydroxylysine residues influences the assembly and diameter of collagen fibrils^{409–411}. Following secretion into the ECM, the tropocollagen undergoes C- and N-terminus cleavage, followed by enzymatic cross-linking. These processes provide mechanical strength and resistance to tensile forces^{380,412–414}.

Collagen PTMs exhibit tissue specificity. Different types of collagens exhibit varying degrees and patterns of PTMs, contributing to the diversity of collagen structures and functions in various tissues. Additionally, the same types of collagens can have different PTMs in different tissues, allowing them to perform slightly different roles. The extent and type of PTMs can dictate the biomechanical properties of the collagen fibres. PTMs can also affect collagen's interaction with other ECM components, such as proteoglycans and glycoproteins. This interaction is crucial for the structural organization and function of the ECM. Modified collagen serves as a binding site for cell receptors, such as integrins. This interaction is pivotal for cell adhesion, migration, and signalling, influencing processes like wound healing and tissue remodelling.

1.3.3. Types of collagen PTMs

Hydroxylation

Hydroxylation involves addition of hydroxyl groups to specific proline or lysine residues within the collagen molecule.

Proline hydroxylation is catalysed by prolyl hydroxylase enzymes: P4HA1, P4HA2, P4HA3 and P4HB. Hydroxyproline is critical for the stability of the collagen triple helix and for collagen interactions with other proteins^{406,407,415}. It increases the melting temperature of collagen, ensuring proper folding and stability at body temperature⁴⁰⁸. Vitamin C (ascorbic acid) is a necessary cofactor for this reaction. Its deficiency leads to impaired collagen synthesis, as seen in scurvy^{416,417}.

In the human genome, biallelic mutations in the *P4HA1* gene have been linked to a congenital connective tissue disorder. This disorder is characterized by joint hypermobility, contractures, mild skeletal dysplasia, and pronounced myopia³⁷⁹. Studies in genetically modified mice have demonstrated that complete homozygous inactivation of P4HA1 results in embryonic lethality at day E10.5⁴¹⁸. On the other hand, heterozygous mutations in the *P4HA2* gene are associated with the development of myopia⁴¹⁹. Mice with a P4ha2^{-/-} phenotype show no significant phenotypic abnormalities. However, mice with a compound genotype of P4ha1^{+/-} and P4ha2^{-/-} display moderate chondrodysplasia³⁷⁷. Furthermore, heterozygous mutations in the *P4HB* gene have been identified as the underlying genetic cause of Cole-Carpenter syndrome, a condition characterized by bone fragility³⁷⁶.

Lysine hydroxylation is performed by lysyl hydroxylase enzymes: PLOD1, PLOD2, PLOD3. Hydroxylysine is involved in the formation of bonds between collagen molecules. Hydroxylysine residues are also sites for glycosylation.

Lysine hydroxylation and its subsequent modifications are critically important. Genetic mutations in the *PLOD1* gene are linked to Ehlers–Danlos syndrome type VIA and Nevo syndrome^{375,420}. These syndromes are marked by symptoms such as hypermobility of joints, skin hyperextensibility, and kyphoscoliosis. Additionally, mutations in the *PLOD2* gene are associated with Bruck syndrome type 2^{378,421}. This rare autosomal-recessive disorder is characterized by bone fragility and congenital joint contractures. Moreover, recent findings

suggest that mutations in the *PLOD3* gene, along with reduced PLOD3 protein levels, are implicated in various congenital connective tissue malformations and epidermolysis bullosa simplex^{374,422}.

In osteoblastic cell cultures, alterations in levels of LH2b, a splice variant of *PLOD2*, have been observed to result in defective collagen cross-linking patterns, impaired collagen fibrillogenesis, and matrix mineralization⁴²³. Furthermore, in OA synovial fluid TGF- β induced increases in LH2b levels are observed³⁷³. The increase in LH2b levels is further associated with increased formation of pyridinoline cross-links by profibrotic cytokines and contributes to osteoarthritic synovial fibrosis⁴²⁴. Overexpression of the gene is linked with fibrosis in other tissues⁴²¹, as well as with a number of solid tumours³⁷².

Glycosylation

Collagen glycosylation involves the addition of carbohydrate moieties to hydroxylysine residues. Glycosylation occurs in two steps: first, galactosylation, where galactose is added by galactosyltransferases such as COLGALT1 and COLGALT2³⁵⁶, and then glucosylation. PLOD3, one of the lysine hydroxylases, has been shown to have glycosyltransferase properties and works in synergy with the COLGALT1 enzymes to produce a α -1,2-glucosylgalactosyl-O-hydroxylysine⁴²⁵.

Glycosylation enhances collagen's thermal stability, influences fibre formation, and facilitates cell-collagen interactions and interactions between collagens and other components of the ECM, impacting tissue integrity and function^{409–411,426}. The type and degree of glycosylation in collagen exhibit variability across different collagen types^{364,427–429}. Despite the differences, a certain level of this PTM is essential to life, as demonstrated by COLGALT1 knock-out (KO) resulting in lethality in mice at E11.5⁴³⁰. Further, mutations in the *COLGALT1* gene are associated with cerebral small vessel disease in both mice and humans. These mutations lead to decreased galactosyltransferase activity and abnormal intracellular accumulation of type IV collagen³⁶⁹. *COLGALT1* mutations are also linked to musculoskeletal defects in mice³⁶⁸.

COLGALT1 and COLGALT2, despite displaying 55% sequence identity, are differentially expressed in human tissues, with COLGALT1 found ubiquitously and COLGALT2 primarily in

the brain and skeletal muscle³⁵⁶. Neither enzyme is active on isolated hydroxylysyl (Hyl) residues and requires collagenous peptides as substrates. They do not show specificity for certain collagen sequences *in vitro*³⁵⁶. Additionally, the COLGALT1 enzyme also has activity on the collagenous domain of mannan-binding lectin and high molecular weight adiponectin, with glycosylation of these proteins facilitating their secretion⁴³¹. The COLGALT2 enzyme has also been associated with intercellular accumulation of high molecular weight adiponectin in the context of non-alcoholic fatty liver disease in mice³⁷⁰. The PLOD3 multi-functional enzyme also possesses collagen galactosyltransferase activity *in vitro*^{432,433}, but studies have demonstrated that *in vivo* it is not sufficient for physiologically relevant collagen glycosylation levels, as evident by the loss of function studies on COLGALT1⁴³⁰. Conversely, complete loss of PLOD3 function in mice resulted in embryonic death around E9.5 due to loss of secretion of collagen IV and its accumulation in the ER. The accumulation was shown to be a result of lack of PLOD3's glucosyltransferase activity, not its hydroxylase activity^{371,422,434}.

Changes in collagen glycosylation levels have been observed in several bone and skeletal disorders. These include conditions like osteogenesis imperfecta^{435,436}, osteosarcoma⁴³⁷, and OA²⁶⁵.

Glycosylation of procollagen peptides is an enzymatically controlled process that only contributes to pathological conditions if dysregulated. Non-enzymatic glycation of proteins, including collagens, on the other hand can lead to accumulation of AGEs which, as discussed earlier, can be detrimental to tissue health^{173–175,438,439}.

Telopeptide cleavage

Once released into the ECM, procollagen molecules undergo enzymatic cleavage by BMP1/tolloid-like proteinases and ADAMTS2, 3, and 14. These enzymes respectively remove the C- and N-terminal propeptides, resulting in the formation of a central triple helical structure with short non-collagenous sequences known as telopeptides at each end^{359,367}. This allows the collagen molecules to assemble correctly into fibrils. Mutations in the enzymes responsible for this cleavage can lead to various connective tissue disorders. For instance, recessive mutations in *ADAMTS2* are linked to dermatosparaxis Ehlers-Danlos syndrome type 7, characterized by severe skin fragility, short limbs, distinctive craniofacial features, and

osteopenia^{97,366}. On the other hand, recessive mutations in *ADAMTS3* cause Hennekam lymphangiectasia–lymphedema syndrome 3, marked by primary lymphedema primarily in the lower extremities^{97,366}. Similarly, mutations in *BMP1* can lead to a form of autosomal recessive osteogenesis imperfecta, stemming from defects in the extracellular processing of type I procollagen, particularly in the removal of the C-propeptide and possibly affecting subsequent collagen cross-linking^{365,440}. Additionally, BMP1 is also crucial for the processing of lysyl oxidase (LOX), a key enzyme in ECM collagen cross-linking, underscoring its vital role in maintaining collagen integrity⁴⁴¹.

The removal of the C-propeptides is essential for fibril formation as it impacts collagen solubility⁴¹² and facilitates the assembly of collagen fibrils by allowing tropocollagen molecules to pack tightly together. The removal of the N-propeptide aids in the correct alignment of collagen molecules, a prerequisite for the formation of strong and functional collagen fibrils⁴¹³.

Cross-linking

The formation of covalent cross-links in collagen fibrils, catalysed by LOX, is critical for the functional integrity of collagen in tissues. These cross-links, formed between aldehydes and lysine or hydroxylysine residues, are diverse and tissue- and collagen type-specific. They significantly enhance collagen's mechanical strength, influencing tissue rigidity and resilience^{380,414}. This cross-linking process is not only crucial for the structural integrity of tissues but also has implications in aging, where increased cross-linking can lead to decreased tissue elasticity, as seen in fibrosis⁴⁴². Inactivation of the mouse *Lox* gene leads to perinatal death due to aortic aneurysms and cardiovascular defects⁴⁴³, demonstrating that the PTM is crucial for health and development.

1.3.4. Regulation of PTMs

The enzymes involved in the PTMs of collagen are under strict regulatory controls to ensure precise and functional collagen synthesis^{360,425}. Factors influencing these enzymatic activities are diverse and multifaceted. The availability of substrates like proline and lysine is crucial for hydroxylation, a key step in collagen synthesis. Limited availability can restrict

enzyme activity, affecting collagen structure and function^{404,444}. Furthermore, essential cofactors like vitamin C are required for the hydroxylation of proline and lysine residues. Deficiency in cofactors can lead to impaired collagen synthesis, as seen in conditions like scurvy^{416,417}.

Environmental factors, such as oxygen levels significantly influence hydroxylation processes. Collagen synthesis can be affected under hypoxic conditions, which may alter the structure and function of the collagen produced³⁶⁰. Additionally, factors like pH and temperature can also impact the enzymatic activities involved in collagen PTMs⁴⁴⁵, as well as mechanochemical stimuli⁴⁴⁶.

Age-related changes in enzyme activity can also affect collagen cross-linking, leading to stiffer and less resilient tissues^{427,439}.

The expression of genes encoding the enzymes responsible for collagen modification is tightly regulated by signalling pathways responsive to both intracellular and extracellular cues. Genetic variation has been shown to play a role in the regulation of the expression of these enzymes and is a promising therapeutic target for diseases associated with loss of ECM integrity and structure, such as OA.

1.4. Hypotheses in this thesis

In this thesis, I conducted a molecular genetic analysis to study the risk of OA associated with specific genetic variations known as SNPs located within or near the *COLGALT2* gene locus. These SNPs are also linked to mQTLs at CpGs within the same locus. My main hypothesis was that *COLGALT2* is the gene of interest responsible for the OA association signals, and DNAm plays a functional role in mediating the relationship between these association signals and changes in *COLGALT2* gene expression.

1.5. Aims

The experiments conducted during my PhD research were focused on testing the hypothesis outlined above, utilizing primary human chondrocytes and synovial fibroblasts obtained from hip and knee arthroplasty patients, as well as foetal joint samples. These experiments were designed to address questions about the molecular mechanisms

underlying genetic susceptibility to OA, marked by two distinct GWAS signals and the epigenetically marked enhancer regions associated with them. My research explored these mechanisms across two joint tissues (cartilage and synovium) using a range of in silico and molecular techniques, including the CRISPR/Cas9 system to modify the epigenome. This approach aimed to establish directionality and causality between genetic and epigenetic marks in the context of disease and to enhance understanding of how the genome and epigenome collectively influence OA susceptibility in different joint tissues and during joint development by regulating the expression of target genes.

1.6. References

1. [review] DiGirolamo, D. J., Kiel, D. P. & Esser, K. A. Bone and Skeletal Muscle: Neighbors With Close Ties. *J. Bone Miner. Res.* **28**, 1509–1518 (2013).
2. [review] Bobzin, L., Roberts, R. R., Chen, H.-J., Crump, J. G. & Merrill, A. E. Development and maintenance of tendons and ligaments. *Development* **148**, (2021).
3. [review] Kell, R. T., Bell, G. & Quinney, A. Musculoskeletal Fitness, Health Outcomes and Quality of Life. *Sports Med.* **31**, 863–873 (2001).
4. Liu, S. *et al.* Global burden of musculoskeletal disorders and attributable factors in 204 countries and territories: a secondary analysis of the Global Burden of Disease 2019 study. *BMJ Open* **12**, e062183 (2022).
5. [review] Berendsen, A. D. & Olsen, B. R. Bone development. *Bone* **80**, 14–18 (2015).
6. [review] Ohba, S. Hedgehog Signaling in Skeletal Development: Roles of Indian Hedgehog and the Mode of Its Action. *Int. J. Mol. Sci.* **21**, 6665 (2020).
7. Norrie, J. L. *et al.* Dynamics of BMP signaling in limb bud mesenchyme and polydactyly. *Dev. Biol.* **393**, 270–281 (2014).
8. Verheyden, J. M. & Sun, X. An Fgf/Gremlin inhibitory feedback loop triggers termination of limb bud outgrowth. *Nature* **454**, 638–641 (2008).

9. Day, T. F., Guo, X., Garrett-Beal, L. & Yang, Y. Wnt/ β -Catenin Signaling in Mesenchymal Progenitors Controls Osteoblast and Chondrocyte Differentiation during Vertebrate Skeletogenesis. *Dev. Cell* **8**, 739–750 (2005).
10. Chen, X., Macica, C. M., Nasiri, A. & Broadus, A. E. Regulation of articular chondrocyte proliferation and differentiation by indian hedgehog and parathyroid hormone–related protein in mice. *Arthritis Rheum.* **58**, 3788–3797 (2008).
11. Chang, S. C. *et al.* Cartilage-derived morphogenetic proteins. New members of the transforming growth factor-beta superfamily predominantly expressed in long bones during human embryonic development. *J. Biol. Chem.* **269**, 28227–34 (1994).
12. [review] Sun, K., Guo, J., Yao, X., Guo, Z. & Guo, F. Growth differentiation factor 5 in cartilage and osteoarthritis: A possible therapeutic candidate. *Cell Prolif.* **54**, e12998 (2021).
13. Shwartz, Y., Viukov, S., Krief, S. & Zelzer, E. Joint Development Involves a Continuous Influx of Gdf5-Positive Cells. *Cell Rep.* **15**, 2577–2587 (2016).
14. Daans, M., Luyten, F. P. & Lories, R. J. U. GDF5 deficiency in mice is associated with instability-driven joint damage, gait and subchondral bone changes. *Ann. Rheum. Dis.* **70**, 208 (2011).
15. Pregizer, S. K. *et al.* Impact of broad regulatory regions on Gdf5 expression and function in knee development and susceptibility to osteoarthritis. *Ann. Rheum. Dis.* **77**, 450 (2018).
16. Li, J. *et al.* A missense GDF5 variant causes brachydactyly type A1 and multiple-synostoses syndrome 2. *JOR Spine* **7**, e1302 (2024).
17. Mang, T., Kleinschmidt-Dörr, K., Ploeger, F., Lindemann, S. & Gigout, A. The GDF-5 mutant M1673 exerts robust anabolic and anti-catabolic effects in chondrocytes. *J. Cell. Mol. Med.* **24**, 7141–7150 (2020).
18. Coveney, C. R. *et al.* Complex regulatory interactions at GDF5 shape joint morphology and osteoarthritis disease risk. *bioRxiv* 2024.11.01.621374 (2024) doi:10.1101/2024.11.01.621374.
19. [review] Li, X. *et al.* Role of Wnt signaling pathway in joint development and cartilage degeneration. *Front. Cell Dev. Biol.* **11**, 1181619 (2023).

20. Zhang, R. *et al.* Wnt/ β -catenin signaling activates bone morphogenetic protein 2 expression in osteoblasts. *Bone* **52**, 145–156 (2013).
21. Guo, X. *et al.* Wnt/ β -catenin signaling is sufficient and necessary for synovial joint formation. *Genes Dev.* **18**, 2404–2417 (2004).
22. Mak, K. K., Chen, M.-H., Day, T. F., Chuang, P.-T. & Yang, Y. Wnt/ β -catenin signaling interacts differentially with Ihh signaling in controlling endochondral bone and synovial joint formation. *Development* **133**, 3695–3707 (2006).
23. [review] Lefebvre, V. & Dvir-Ginzberg, M. SOX9 and the many facets of its regulation in the chondrocyte lineage. *Connect. Tissue Res.* **58**, 2–14 (2017).
24. Dy, P. *et al.* Sox9 Directs Hypertrophic Maturation and Blocks Osteoblast Differentiation of Growth Plate Chondrocytes. *Dev. Cell* **22**, 597–609 (2012).
25. [review] Komori, T. Regulation of skeletal development by the Runx family of transcription factors. *J. Cell. Biochem.* **95**, 445–453 (2005).
26. Eames, B. F., Sharpe, P. T. & Helms, J. A. Hierarchy revealed in the specification of three skeletal fates by Sox9 and Runx2. *Dev. Biol.* **274**, 188–200 (2004).
27. [review] Long, F. & Ornitz, D. M. Development of the Endochondral Skeleton. *Cold Spring Harb. Perspect. Biol.* **5**, a008334 (2013).
28. Prein, C. & Beier, F. ECM signaling in cartilage development and endochondral ossification. *Curr. Top. Dev. Biol.* **133**, 25–47 (2018).
29. [review] Lowery, J. W. & Rosen, V. The BMP Pathway and Its Inhibitors in the Skeleton. *Physiol. Rev.* **98**, 2431–2452 (2018).
30. [review] Carballo, C. B., Nakagawa, Y., Sekiya, I. & Rodeo, S. A. Basic Science of Articular Cartilage. *Clin. Sports Med.* **36**, 413–425 (2017).
31. [review] Sanchez-Adams, J., Leddy, H. A., McNulty, A. L., O’Conor, C. J. & Guilak, F. The Mechanobiology of Articular Cartilage: Bearing the Burden of Osteoarthritis. *Curr. Rheumatol. Rep.* **16**, 451 (2014).
32. [review] Komori, T. Molecular Processes in Chondrocyte Biology. *Int. J. Mol. Sci.* **21**, 4161 (2020).

33. [review] Eyre, D. R. Collagens and Cartilage Matrix Homeostasis. *Clin. Orthop. Relat. Res.* **427**, S118–S122 (2004).
34. Upadhyay, U., Kolla, S. & Chelluri, L. K. Extracellular matrix composition analysis of human articular cartilage for the development of organ-on-a-chip. *Biochem. Biophys. Res. Commun.* **667**, 81–88 (2023).
35. [review] Eyre, D. Articular cartilage and changes in Arthritis: Collagen of articular cartilage. *Arthritis Res. Ther.* **4**, 30 (2001).
36. Tanaka, N. *et al.* Expression of minor cartilage collagens and small leucine rich proteoglycans may be relatively reduced in osteoarthritic cartilage. *BMC Musculoskelet. Disord.* **20**, 232 (2019).
37. [review] Eyre, D. R., Wu, J.-J., Fernandes, R. J., Pietka, T. A. & Weis, M. A. Recent developments in cartilage research: matrix biology of the collagen II/IX/XI heterofibril network. *Biochem. Soc. Trans.* **30**, 893 (2001).
38. Eyre, D. R., Pietka, T., Weis, M. A. & Wu, J.-J. Covalent Cross-linking of the NC1 Domain of Collagen Type IX to Collagen Type II in Cartilage*. *J. Biol. Chem.* **279**, 2568–2574 (2004).
39. Li, Y. *et al.* A fibrillar collagen gene, Col11a1, is essential for skeletal morphogenesis. *Cell* **80**, 423–430 (1995).
40. Blaschke, U. K., Eikenberry, E. F., Hulmes, D. J. S., Galla, H.-J. & Bruckner, P. Collagen XI Nucleates Self-assembly and Limits Lateral Growth of Cartilage Fibrils*. *J. Biol. Chem.* **275**, 10370–10378 (2000).
41. Gregory, K. E. *et al.* Structural Organization of Distinct Domains within the Non-collagenous N-terminal Region of Collagen Type XI*. *J. Biol. Chem.* **275**, 11498–11506 (2000).
42. Thomas, J. T. *et al.* A human chondrodysplasia due to a mutation in a TGF- β superfamily member. *Nat. Genet.* **12**, 315–317 (1996).
43. Hosseininia, S. *et al.* Evidence for enhanced collagen type III deposition focally in the territorial matrix of osteoarthritic hip articular cartilage. *Osteoarthr. Cartil.* **24**, 1029–1035 (2016).
44. Wu, J.-J., Weis, M. A., Kim, L. S. & Eyre, D. R. Type III Collagen, a Fibril Network Modifier in Articular Cartilage*. *J. Biol. Chem.* **285**, 18537–18544 (2010).

45. [review] Kuivaniemi, H. & Tromp, G. Type III collagen (COL3A1): Gene and protein structure, tissue distribution, and associated diseases. *Gene* **707**, 151–171 (2019).
46. Wu, J.-J., Weis, M. A., Kim, L. S., Carter, B. G. & Eyre, D. R. Differences in Chain Usage and Cross-linking Specificities of Cartilage Type V/XI Collagen Isoforms with Age and Tissue*. *J. Biol. Chem.* **284**, 5539–5545 (2009).
47. [review] Mak, K. M., Png, C. Y. M. & Lee, D. J. Type V Collagen in Health, Disease, and Fibrosis. *Anat. Rec.* **299**, 613–629 (2016).
48. Bidanset, D. J. *et al.* Binding of the proteoglycan decorin to collagen type VI. *J. Biol. Chem.* **267**, 5250–6 (1992).
49. Wiberg, C. *et al.* Biglycan and Decorin Bind Close to the N-terminal Region of the Collagen VI Triple Helix*. *J. Biol. Chem.* **276**, 18947–18952 (2001).
50. Guilak, F. *et al.* The Pericellular Matrix as a Transducer of Biomechanical and Biochemical Signals in Articular Cartilage. *Ann. N. York Acad. Sci.* **1068**, 498–512 (2006).
51. Knuth, C. A. *et al.* Collagen type X is essential for successful mesenchymal stem cell-mediated cartilage formation and subsequent endochondral ossification. *Eur. cells Mater.* **38**, 106–122.
52. He, Y. *et al.* Type X collagen levels are elevated in serum from human osteoarthritis patients and associated with biomarkers of cartilage degradation and inflammation. *BMC Musculoskelet. Disord.* **15**, 309 (2014).
53. Taylor, D. W. *et al.* Collagen Type XII and Versican Are Present in the Early Stages of Cartilage Tissue Formation by Both Redifferentating Passaged and Primary Chondrocytes. *Tissue Eng. Part A* **21**, 683–693 (2015).
54. [review] Luo, Y. *et al.* The minor collagens in articular cartilage. *Protein Cell* **8**, 560–572 (2017).
55. Hemmavanh, C., Koch, M., Birk, D. E. & Espana, E. M. Abnormal Corneal Endothelial Maturation in Collagen XII and XIV Null Mice. *Investig. Ophthalmology Vis. Sci.* **54**, 3297 (2013).
56. Kassner, A. *et al.* Discrete integration of collagen XVI into tissue-specific collagen fibrils or beaded microfibrils. *Matrix Biol.* **22**, 131–143 (2003).

57. Koch, M. *et al.* A Novel Marker of Tissue Junctions, Collagen XXII*. *J. Biol. Chem.* **279**, 22514–22521 (2004).
58. Plumb, D. A. *et al.* Collagen XXVII Organises the Pericellular Matrix in the Growth Plate. *PLoS ONE* **6**, e29422 (2011).
59. [review] Roughley, P. J. & Mort, J. S. The role of aggrecan in normal and osteoarthritic cartilage. *J. Exp. Orthop.* **1**, 8 (2014).
60. Hayes, A. J. & Melrose, J. Aggrecan, the Primary Weight-Bearing Cartilage Proteoglycan, Has Context-Dependent, Cell-Directive Properties in Embryonic Development and Neurogenesis: Aggrecan Glycan Side Chain Modifications Convey Interactive Biodiversity. *Biomolecules* **10**, 1244 (2020).
61. Hascall, V. C. & Heinegård, D. Aggregation of cartilage proteoglycans: Oligosaccharides act as competitors of the proteoglycan-hyaluronic acid interaction. *J. Biol. Chem.* **249**, 4242–9 (1974).
62. Prakash, D. & Learmonth, D. Natural progression of osteo-chondral defect in the femoral condyle. *Knee* **9**, 7–10 (2002).
63. Wiberg, C. *et al.* Complexes of Matrilin-1 and Biglycan or Decorin Connect Collagen VI Microfibrils to Both Collagen II and Aggrecan*. *J. Biol. Chem.* **278**, 37698–37704 (2003).
64. Twomey, J., Thakore, P., Hartman, D., Myers, E. & Hsieh, A. Roles of type VI collagen and decorin in human mesenchymal stem cell biophysics during chondrogenic differentiation. *Eur. Cells Mater.* **27**, 237–250 (2014).
65. Hayes, A. J. *et al.* The CS Sulfation Motifs 4C3, 7D4, 3B3[–]; and Perlecan Identify Stem Cell Populations and Their Niches, Activated Progenitor Cells and Transitional Areas of Tissue Development in the Fetal Human Elbow. *Stem Cells Dev.* **25**, 836–847 (2016).
66. Choocheep, K. *et al.* Versican Facilitates Chondrocyte Differentiation and Regulates Joint Morphogenesis*. *J. Biol. Chem.* **285**, 21114–21125 (2010).
67. [review] Chevalier, X. Fibronectin, cartilage, and osteoarthritis. *Semin. Arthritis Rheum.* **22**, 307–318 (2009).
68. Flowers, S. A. *et al.* Lubricin binds cartilage proteins, cartilage oligomeric matrix protein, fibronectin and collagen II at the cartilage surface. *Sci. Rep.* **7**, 13149 (2017).

69. Tao, T. *et al.* Fibronectin Enhances Cartilage Repair by Activating Progenitor Cells Through Integrin $\alpha 5 \beta 1$ Receptor. *Tissue Eng. Part A* **24**, 1112–1124 (2018).
70. [review] Tseng, S., Reddi, A. H. & Cesare, P. E. D. Cartilage Oligomeric Matrix Protein (COMP): A Biomarker of Arthritis. *Biomark. Insights* **4**, BMI.S645 (2009).
71. Halász, K., Kassner, A., Mörgelin, M. & Heinegård, D. COMP Acts as a Catalyst in Collagen Fibrillogenesis*. *J. Biol. Chem.* **282**, 31166–31173 (2007).
72. Li, P. *et al.* Mice Lacking the Matrilin Family of Extracellular Matrix Proteins Develop Mild Skeletal Abnormalities and Are Susceptible to Age-Associated Osteoarthritis. *Int. J. Mol. Sci.* **21**, 666 (2020).
73. Majd, S. E. *et al.* Both Hyaluronan and Collagen Type II Keep Proteoglycan 4 (Lubricin) at the Cartilage Surface in a Condition That Provides Low Friction during Boundary Lubrication. *Langmuir* **30**, 14566–14572 (2014).
74. Ludwig, T. E., Hunter, M. M. & Schmidt, T. A. Cartilage boundary lubrication synergism is mediated by hyaluronan concentration and PRG4 concentration and structure. *BMC Musculoskelet. Disord.* **16**, 386 (2015).
75. [review] Jay, G. D. & Waller, K. A. The biology of Lubricin: Near frictionless joint motion. *Matrix Biol.* **39**, 17–24 (2014).
76. [review] Carballo, C. B., Nakagawa, Y., Sekiya, I. & Rodeo, S. A. Basic Science of Articular Cartilage. *Clin. Sports Med.* **36**, 413–425 (2017).
77. Korhonen, R. K., Julkunen, P., Wilson, W. & Herzog, W. Importance of Collagen Orientation and Depth-Dependent Fixed Charge Densities of Cartilage on Mechanical Behavior of Chondrocytes. *J. Biomech. Eng.* **130**, 021003 (2008).
78. Moger, C. J. *et al.* Cartilage Collagen Matrix Reorientation and Displacement in Response to Surface Loading. *J. Biomech. Eng.* **131**, 031008 (2009).
79. Yu, J. & Urban, J. P. G. The elastic network of articular cartilage: an immunohistochemical study of elastin fibres and microfibrils. *J. Anat.* **216**, 533–541 (2010).
80. Mansfield, J. C., Bell, J. S. & Winlove, C. P. The micromechanics of the superficial zone of articular cartilage. *Osteoarthr. Cartil.* **23**, 1806–1816 (2015).

81. [review] Evans, L. A. E. & Pitsillides, A. A. Structural clues to articular calcified cartilage function: A descriptive review of this crucial interface tissue. *J. Anat.* **241**, 875–895 (2022).
82. [review] Wilusz, R. E., Sanchez-Adams, J. & Guilak, F. The structure and function of the pericellular matrix of articular cartilage. *Matrix Biol.* **39**, 25–32 (2014).
83. GUILAK, F. *et al.* The Pericellular Matrix as a Transducer of Biomechanical and Biochemical Signals in Articular Cartilage. *Ann. N. York Acad. Sci.* **1068**, 498–512 (2006).
84. Liu, C. *et al.* Protective effects of the pericellular matrix components of chondrocyte on articular cartilage against the development of osteoarthritis. *Histol. Histopathol.* **33**, 757–764 (2018).
85. Youn, I., Choi, J. B., Cao, L., Setton, L. A. & Guilak, F. Zonal variations in the three-dimensional morphology of the chondron measured in situ using confocal microscopy. *Osteoarthr. Cartil.* **14**, 889–897 (2006).
86. Vincent, T. L., McClurg, O. & Troeberg, L. The Extracellular Matrix of Articular Cartilage Controls the Bioavailability of Pericellular Matrix-Bound Growth Factors to Drive Tissue Homeostasis and Repair. *Int. J. Mol. Sci.* **23**, 6003 (2022).
87. Hui, A. Y., McCarty, W. J., Masuda, K., Firestein, G. S. & Sah, R. L. A systems biology approach to synovial joint lubrication in health, injury, and disease. *Wiley Interdiscip. Rev.: Syst. Biol. Med.* **4**, 15–37 (2012).
88. [review] Liu, Y., Shah, K. M. & Luo, J. Strategies for Articular Cartilage Repair and Regeneration. *Front. Bioeng. Biotechnol.* **9**, 770655 (2021).
89. Kobayashi, T., Lyons, K. M., McMahon, A. P. & Kronenberg, H. M. BMP signaling stimulates cellular differentiation at multiple steps during cartilage development. *Proc. Natl. Acad. Sci.* **102**, 18023–18027 (2005).
90. Lui, J. C. *et al.* Cartilage-Targeted IGF-1 Treatment to Promote Longitudinal Bone Growth. *Mol. Ther.* **27**, 673–680 (2019).
91. Wang, W., Rigueur, D. & Lyons, K. M. TGF β signaling in cartilage development and maintenance. *Birth Defects Res. Part C: Embryo Today: Rev.* **102**, 37–51 (2014).
92. Yang, X. *et al.* TGF- β /Smad3 Signals Repress Chondrocyte Hypertrophic Differentiation and Are Required for Maintaining Articular Cartilage. *J. Cell Biol.* **153**, 35–46 (2001).

93. Bobacz, K. *et al.* Expression of bone morphogenetic protein 6 in healthy and osteoarthritic human articular chondrocytes and stimulation of matrix synthesis in vitro. *Arthritis Rheum.* **48**, 2501–2508 (2003).
94. Bonassar, L. J. *et al.* The effect of dynamic compression on the response of articular cartilage to insulin-like growth factor-I. *J. Orthop. Res.* **19**, 11–17 (2001).
95. [review] Pitsillides, A. A. & Beier, F. Cartilage biology in osteoarthritis—lessons from developmental biology. *Nat. Rev. Rheumatol.* **7**, 654–663 (2011).
96. [review] Onursal, C., Dick, E., Angelidis, I., Schiller, H. B. & Staab-Weijnitz, C. A. Collagen Biosynthesis, Processing, and Maturation in Lung Ageing. *Front. Med.* **8**, 593874 (2021).
97. [review] Mead, T. J. & Apte, S. S. ADAMTS proteins in human disorders. *Matrix Biol.* **71**, 225–239 (2018).
98. [review] Marcu, K. B., Otero, M., Olivotto, E., Borzi, R. M. & Goldring, M. B. NF- κ B Signaling: Multiple Angles to Target OA. *Curr. Drug Targets* **11**, 599–613 (2010).
99. [review] Wang, M. *et al.* Recent progress in understanding molecular mechanisms of cartilage degeneration during osteoarthritis. *Ann. N. York Acad. Sci.* **1240**, 61–69 (2011).
100. Kapoor, M., Martel-Pelletier, J., Lajeunesse, D., Pelletier, J.-P. & Fahmi, H. Role of proinflammatory cytokines IL-1 β , TNF- α and IL-6 in the pathophysiology of osteoarthritis. *Nat. Rev. Rheumatol.* **7**, 33–42 (2011).
101. Arpino, V., Brock, M. & Gill, S. E. The role of TIMP2 and TIMP3 in regulation of extracellular matrix proteolysis. *Matrix Biol.* **44**, 247–254 (2015).
102. Chowdhury, T. T., Salter, D. M., Bader, D. L. & Lee, D. A. Signal transduction pathways involving p38 MAPK, JNK, NF κ B and AP-1 influences the response of chondrocytes cultured in agarose constructs to IL-1 β and dynamic compression. *Inflamm. Res.* **57**, 306–313 (2008).
103. Stolberg-Stolberg, J. A. *et al.* Effects of cartilage impact with and without fracture on chondrocyte viability and the release of inflammatory markers. *J. Orthop. Res.* **31**, 1283–1292 (2013).
104. Quinn, T. M., Grodzinsky, A. J., Hunziker, E. B. & Sandy, J. D. Effects of injurious compression on matrix turnover around individual cells in calf articular cartilage explants. *J. Orthop. Res.* **16**, 490–499 (1998).

105. Chan, P. S. *et al.* Gene expression profile of mechanically impacted bovine articular cartilage explants. *J. Orthop. Res.* **23**, 1146–1151 (2005).
106. Natoli, R. M., Scott, C. C. & Athanasiou, K. A. Temporal Effects of Impact on Articular Cartilage Cell Death, Gene Expression, Matrix Biochemistry, and Biomechanics. *Ann. Biomed. Eng.* **36**, 780–792 (2008).
107. Zhang, Y. *et al.* Cartilage-specific deletion of mTOR upregulates autophagy and protects mice from osteoarthritis. *Ann. Rheum. Dis.* **74**, 1432 (2015).
108. Lin, Z. *et al.* JUNB-FBXO21-ERK axis promotes cartilage degeneration in osteoarthritis by inhibiting autophagy. *Aging Cell* **20**, e13306 (2021).
109. Caramés, B., Taniguchi, N., Otsuki, S., Blanco, F. J. & Lotz, M. Autophagy is a protective mechanism in normal cartilage, and its aging-related loss is linked with cell death and osteoarthritis. *Arthritis Rheum.* **62**, 791–801 (2010).
110. Ryu, J.-H. *et al.* Hypoxia-inducible factor-2 α regulates Fas-mediated chondrocyte apoptosis during osteoarthritic cartilage destruction. *Cell Death Differ.* **19**, 440–450 (2012).
111. Hu, S. *et al.* Stabilization of HIF-1 α alleviates osteoarthritis via enhancing mitophagy. *Cell Death Dis.* **11**, 481 (2020).
112. [review] Smith, M. D. The Normal Synovium. *Open Rheumatol. J.* **5**, 100–106 (2011).
113. Wang, Y., Wei, L., Zeng, L., He, D. & Wei, X. Nutrition and degeneration of articular cartilage upon IL-6 stimulation. *Knee Surg., Sports Traumatol., Arthrosc.* **21**, 1751–1762 (2013).
114. [review] Flandry, F. & Hommel, G. Normal Anatomy and Biomechanics of the Knee. *Sports Med. Arthrosc. Rev.* **19**, 82–92 (2011).
115. [review] Caola, G. The anatomy and physiology of diarthroses. *Vet. Res. Commun.* **27 Suppl 1**, 69–73 (2003).
116. Revell, P. A. Synovial lining cells. *Rheumatol. Int.* **9**, 49–51 (1989).
117. [review] Wei, Q. *et al.* Extracellular matrix in synovium development, homeostasis and arthritis disease. *Int. Immunopharmacol.* **121**, 110453 (2023).

118. Perlman, H. *et al.* IL-6 and Matrix Metalloproteinase-1 Are Regulated by the Cyclin-Dependent Kinase Inhibitor p21 in Synovial Fibroblasts. *J. Immunol.* **170**, 838–845 (2003).
119. Donlin, L. T., Jayatilleke, A., Giannopoulou, E. G., Kalliolias, G. D. & Ivashkiv, L. B. Modulation of TNF-Induced Macrophage Polarization by Synovial Fibroblasts. *J. Immunol.* **193**, 2373–2383 (2014).
120. Kaeppler, E. P. von *et al.* Interleukin 4 promotes anti-inflammatory macrophages that clear cartilage debris and inhibits osteoclast development to protect against osteoarthritis. *Clin. Immunol.* **229**, 108784 (2021).
121. Broeren, M. G. A. *et al.* Suppression of the inflammatory response by disease-inducible interleukin-10 gene therapy in a three-dimensional micromass model of the human synovial membrane. *Arthritis Res. Ther.* **18**, 186 (2016).
122. Helvoort, E. M. van, Heijden, E. van der, Roon, J. A. G. van, Eijkelkamp, N. & Mastbergen, S. C. The Role of Interleukin-4 and Interleukin-10 in Osteoarthritic Joint Disease: A Systematic Narrative Review. *CARTILAGE* **13**, 19476035221098170 (2022).
123. Yamamura, Y. *et al.* Effector Function of Resting T Cells in the activation of Synovial Fibroblasts. *J. Immunol.* **166**, 2270–2275 (2001).
124. Tran, C. N. *et al.* Presentation of arthritogenic peptide to antigen-specific T cells by fibroblast-like synoviocytes. *Arthritis Rheum.* **56**, 1497–1506 (2007).
125. Carmona-Rivera, C. *et al.* Synovial fibroblast-neutrophil interactions promote pathogenic adaptive immunity in rheumatoid arthritis. *Sci. Immunol.* **2**, (2017).
126. Milz, S. & Putz, R. Quantitative morphology of the subchondral plate of the tibial plateau. *J. Anat.* **185 (Pt 1)**, 103–110 (1994).
127. Holmdahl, D. E. & Ingelmark, B. E. The Contact Between the Articular Cartilage and the Medullary Cavities of the Bone. *Acta Orthop. Scand.* **20**, 156–165 (2009).
128. [review] Feng, X. & McDonald, J. M. Disorders of Bone Remodeling. *Annu. Rev. Pathol.: Mech. Dis.* **6**, 121–145 (2011).
129. [review] Palumbo, C. & Ferretti, M. The Osteocyte: From “Prisoner” to “Orchestrator.” *J. Funct. Morphol. Kinesiol.* **6**, 28 (2021).

130. Funck-Brentano, T. *et al.* Dkk-1–Mediated Inhibition of Wnt Signaling in Bone Ameliorates Osteoarthritis in Mice. *Arthritis Rheumatol.* **66**, 3028–3039 (2014).
131. Zhen, G. *et al.* Inhibition of TGF- β signaling in mesenchymal stem cells of subchondral bone attenuates osteoarthritis. *Nat. Med.* **19**, 704–712 (2013).
132. Tu, X. *et al.* Sost downregulation and local Wnt signaling are required for the osteogenic response to mechanical loading. *Bone* **50**, 209–217 (2012).
133. Macchi, V. *et al.* The Infrapatellar Adipose Body: A Histotopographic Study. *Cells Tissues Organs* **201**, 220–231 (2016).
134. [review] Gallagher, J., Tierney, P., Murray, P. & O’Brien, M. The infrapatellar fat pad: anatomy and clinical correlations. *Knee Surg., Sports Traumatol., Arthrosc.* **13**, 268–272 (2005).
135. Bohnsack, M. *et al.* Distribution of substance-P nerves inside the infrapatellar fat pad and the adjacent synovial tissue: a neurohistological approach to anterior knee pain syndrome. *Arch. Orthop. Trauma Surg.* **125**, 592–597 (2005).
136. Pinsornsak, P., Naratrikun, K. & Chumchuen, S. The Effect of Infrapatellar Fat Pad Excision on Complications After Minimally Invasive TKA: A Randomized Controlled Trial. *Clin. Orthop. Relat. Res.* **472**, 695–701 (2014).
137. Wang, K. *et al.* Serum levels of interleukin-17 and adiponectin are associated with infrapatellar fat pad volume and signal intensity alteration in patients with knee osteoarthritis. *Arthritis Res. Ther.* **18**, 193 (2016).
138. Ushiyama, T., Chano, T., Inoue, K. & Matsusue, Y. Cytokine production in the infrapatellar fat pad: another source of cytokines in knee synovial fluids. *Ann. Rheum. Dis.* **62**, 108 (2003).
139. Garcia, J. *et al.* Chondrogenic Potency Analyses of Donor-Matched Chondrocytes and Mesenchymal Stem Cells Derived from Bone Marrow, Infrapatellar Fat Pad, and Subcutaneous Fat. *Stem Cells Int.* **2016**, 6969726 (2016).
140. [review] Amaral, R. J. F. C. do, Almeida, H. V., Kelly, D. J., O’Brien, F. J. & Kearney, C. J. Infrapatellar Fat Pad Stem Cells: From Developmental Biology to Cell Therapy. *Stem Cells Int.* **2017**, 6843727 (2017).

141. Mantripragada, V. P. *et al.* Donor-matched comparison of chondrogenic progenitors resident in human infrapatellar fat pad, synovium, and periosteum - implications for cartilage repair. *Connect. Tissue Res.* **60**, 597–610 (2019).
142. [review] Renström, P. & Johnson, R. J. Anatomy and biomechanics of the menisci. *Clin. sports Med.* **9**, 523–38 (1990).
143. Karahan, M., Kocaoglu, B., Cabukoglu, C., Akgun, U. & Nuran, R. Effect of partial medial meniscectomy on the proprioceptive function of the knee. *Arch. Orthop. Trauma Surg.* **130**, 427–431 (2010).
144. Nakata, K. *et al.* Human meniscus cell: characterization of the primary culture and use for tissue engineering. *Clin. Orthop. Relat. Res.* S208-18 (2001).
145. Herwig, J., Egner, E. & Buddecke, E. Chemical changes of human knee joint menisci in various stages of degeneration. *Ann. Rheum. Dis.* **43**, 635 (1984).
146. Graverand, M.-P. *et al.* Comparative analysis of rabbit and human meniscus: their arrangement, interrelationship, morphological variations and cytoarchitecture. *J. Anat.* **198**, 525–535 (2000).
147. Melrose, J., Smith, S., Cake, M., Read, R. & Whitelock, J. Comparative spatial and temporal localisation of perlecan, aggrecan and type I, II and IV collagen in the ovine meniscus: an ageing study. *Histochem. Cell Biol.* **124**, 225–235 (2005).
148. [review] Clark, C. R. & Ogden, J. A. Development of the menisci of the human knee joint. Morphological changes and their potential role in childhood meniscal injury. *J. bone Jt. Surg. Am. Vol.* **65**, 538–47 (1983).
149. Ionescu, L. C., Lee, G. C., Huang, K. L. & Mauck, R. L. Growth factor supplementation improves native and engineered meniscus repair in vitro. *Acta Biomater.* **8**, 3687–3694 (2012).
150. [review] Wang, J. H.-C. Mechanobiology of tendon. *J. Biomech.* **39**, 1563–1582 (2006).
151. [review] Liu, S. H., Yang, R. S., al-Shaikh, R. & Lane, J. M. Collagen in tendon, ligament, and bone healing. A current review. *Clin. Orthop. Relat. Res.* 265–78 (1995).

152. Pins, G. D., Christiansen, D. L., Patel, R. & Silver, F. H. Self-assembly of collagen fibres. Influence of fibrillar alignment and decorin on mechanical properties. *Biophys. J.* **73**, 2164–2172 (1997).
153. Vogel, K. G. & Heinegård, D. Characterization of proteoglycans from adult bovine tendon. *J. Biol. Chem.* **260**, 9298–306 (1985).
154. Wall, M. E. *et al.* Metabolomic Study on Risk for Tendon Disorders. *Adv. Exp. Med. Biol.* **920**, 79–95 (2016).
155. Bi, Y. *et al.* Identification of tendon stem/progenitor cells and the role of the extracellular matrix in their niche. *Nat. Med.* **13**, 1219–1227 (2007).
156. Rui, Y.-F. *et al.* Isolation and Characterization of Multipotent Rat Tendon-Derived Stem Cells. *Tissue Eng. Part A* **16**, 1549–1558 (2010).
157. [review] Eisner, L. E., Rosario, R., Andarawis-Puri, N. & Arruda, E. M. The Role of the Non-Collagenous Extracellular Matrix in Tendon and Ligament Mechanical Behavior: A Review. *J. Biomech. Eng.* **144**, (2021).
158. Stender, C. J. *et al.* Modeling the effect of collagen fibril alignment on ligament mechanical behavior. *Biomech. Model. Mechanobiol.* **17**, 543–557 (2018).
159. Fleming, J. D., Ritzmann, R. & Centner, C. Effect of an Anterior Cruciate Ligament Rupture on Knee Proprioception Within 2 Years After Conservative and Operative Treatment: A Systematic Review with Meta-Analysis. *Sports Med.* **52**, 1091–1102 (2022).
160. [review] Frank, C. B. Ligament structure, physiology and function. *J. Musculoskelet. neuronal Interact.* **4**, 199–201 (2004).
161. [review] Leong, N. L. *et al.* Tendon and Ligament Healing and Current Approaches to Tendon and Ligament Regeneration. *J. Orthop. Res.* **38**, 7–12 (2020).
162. Meknas, K. *et al.* Induced tendonitis role in the progression of osteoarthritis *Scand. J. Med. Sci. Sports* **22**, 627–634 (2012).
163. Schulze-Tanzil, G. Intraarticular Ligament Degeneration Is Interrelated with Cartilage and Bone Destruction in Osteoarthritis. *Cells* **8**, 990 (2019).

164. [review] Iannone, F. & Lapadula, G. The pathophysiology of osteoarthritis. *Aging Clin. Exp. Res.* **15**, 364–372 (2003).
165. [review] Wieland, H. A., Michaelis, M., Kirschbaum, B. J. & Rudolphi, K. A. Osteoarthritis — an untreatable disease? *Nat. Rev. Drug Discov.* **4**, 331–344 (2005).
166. Saito, T. *et al.* Transcriptional regulation of endochondral ossification by HIF-2 α during skeletal growth and osteoarthritis development. *Nat. Med.* **16**, 678–686 (2010).
167. Scharstuhl, A., Beuningen, H. M. van, Vitters, E. L., Kraan, P. M. van der & Berg, W. B. van den. Loss of transforming growth factor counteraction on interleukin 1 mediated effects in cartilage of old mice. *Ann. Rheum. Dis.* **61**, 1095 (2002).
168. Loeser, R. F. *et al.* Reduction in the chondrocyte response to insulin-like growth factor 1 in aging and osteoarthritis: Studies in a non-human primate model of naturally occurring disease. *Arthritis Rheum.* **43**, 2110–2120 (2000).
169. Messai, H., Duchossoy, Y., Khatib, A.-M., Panasyuk, A. & Mitrovic, D. R. Articular chondrocytes from aging rats respond poorly to insulin-like growth factor-1: an altered signaling pathway. *Mech. Ageing Dev.* **115**, 21–37 (2000).
170. Iqbal, J., Dudhia, J., Bird, J. L. E. & Bayliss, M. T. Age-Related Effects of TGF- β on Proteoglycan Synthesis in Equine Articular Cartilage. *Biochem. Biophys. Res. Commun.* **274**, 467–471 (2000).
171. Tran-Khanh, N., Hoemann, C. D., McKee, M. D., Henderson, J. E. & Buschmann, M. D. Aged bovine chondrocytes display a diminished capacity to produce a collagen-rich, mechanically functional cartilage extracellular matrix. *J. Orthop. Res.* **23**, 1354–1362 (2005).
172. Lahm, A. *et al.* Changes in content and synthesis of collagen types and proteoglycans in osteoarthritis of the knee joint and comparison of quantitative analysis with Photoshop-based image analysis. *Arch. Orthop. Trauma Surg.* **130**, 557–564 (2010).
173. Sun, X., Zhang, J., Li, Y., Ren, W. & Wang, L. Etomidate ameliorated advanced glycation end-products (AGEs)-induced reduction of extracellular matrix genes expression in chondrocytes. *Bioengineered* **12**, 4191–4200 (2021).

174. Zeng, Y., Liu, Z., Tan, X. & Lei, L. The GPR55 antagonist CID16020046 mitigates advanced glycation end products (AGEs)- induced chondrocyte activation. *Chem.-Biol. Interact.* **325**, 109088 (2020).
175. Ma, C. *et al.* The Role of PPAR γ in Advanced Glycation End Products-Induced Inflammatory Response in Human Chondrocytes. *PLoS ONE* **10**, e0125776 (2015).
176. Freemont, A. J. *et al.* Gene expression of matrix metalloproteinases 1, 3, and 9 by chondrocytes in osteoarthritic human knee articular cartilage is zone and grade specific. *Ann. Rheum. Dis.* **56**, 542 (1997).
177. Neuhold, L. A. *et al.* Postnatal expression in hyaline cartilage of constitutively active human collagenase-3 (MMP-13) induces osteoarthritis in mice. *J. Clin. Investig.* **107**, 35–44 (2001).
178. Bau, B. *et al.* Relative messenger RNA expression profiling of collagenases and aggrecanases in human articular chondrocytes in vivo and in vitro. *Arthritis Rheum.* **46**, 2648–2657 (2002).
179. Roemer, F. W. *et al.* What Comes First? Multitissue Involvement Leading to Radiographic Osteoarthritis: Magnetic Resonance Imaging–Based Trajectory Analysis Over Four Years in the Osteoarthritis Initiative. *Arthritis Rheumatol.* **67**, 2085–2096 (2015).
180. 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).
181. Smith, M. D., Triantafillou, S., Parker, A., Youssef, P. P. & Coleman, M. Synovial membrane inflammation and cytokine production in patients with early osteoarthritis. *J. Rheumatol.* **24**, 365–71 (1997).
182. Nanus, D. E. *et al.* Synovial tissue from sites of joint pain in knee osteoarthritis patients exhibits a differential phenotype with distinct fibroblast subsets. *EBioMedicine* **72**, 103618 (2021).
183. Benito, M. J., Veale, D. J., FitzGerald, O., Berg, W. B. van den & Bresnihan, B. Synovial tissue inflammation in early and late osteoarthritis. *Ann. Rheum. Dis.* **64**, 1263 (2005).

184. Roemer, F. W. *et al.* Anatomical distribution of synovitis in knee osteoarthritis and its association with joint effusion assessed on non-enhanced and contrast-enhanced MRI. *Osteoarthr. Cartil.* **18**, 1269–1274 (2010).
185. Ayral, X., Pickering, E. H., Woodworth, T. G., Mackillop, N. & Dougados, M. Synovitis: a potential predictive factor of structural progression of medial tibiofemoral knee osteoarthritis – results of a 1 year longitudinal arthroscopic study in 422 patients. *Osteoarthr. Cartil.* **13**, 361–367 (2005).
186. Zhang, Y. *et al.* Fluctuation of knee pain and changes in bone marrow lesions, effusions, and synovitis on magnetic resonance imaging. *Arthritis Rheum.* **63**, 691–699 (2011).
187. Hill, C. L. *et al.* Knee effusions, popliteal cysts, and synovial thickening: association with knee pain in osteoarthritis. *J. Rheumatol.* **28**, 1330–7 (2001).
188. Hill, C. L. *et al.* Synovitis detected on magnetic resonance imaging and its relation to pain and cartilage loss in knee osteoarthritis. *Ann. Rheum. Dis.* **66**, 1599 (2007).
189. Bolander, J. *et al.* The synovial environment steers cartilage deterioration and regeneration. *Sci. Adv.* **9**, eade4645 (2023).
190. [review] Sellam, J. & Berenbaum, F. The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. *Nat. Rev. Rheumatol.* **6**, 625–635 (2010).
191. MacKay, J. W. *et al.* Subchondral bone in osteoarthritis: association between MRI texture analysis and histomorphometry. *Osteoarthr. Cartil.* **25**, 700–707 (2017).
192. Bolbos, R. I. *et al.* Relationship between trabecular bone structure and articular cartilage morphology and relaxation times in early OA of the knee joint using parallel MRI at 3T. *Osteoarthr. Cartil.* **16**, 1150–1159 (2008).
193. Strassle, B. W. *et al.* Inhibition of osteoclasts prevents cartilage loss and pain in a rat model of degenerative joint disease. *Osteoarthr. Cartil.* **18**, 1319–1328 (2010).
194. Walsh, D. A. *et al.* Angiogenesis in the synovium and at the osteochondral junction in osteoarthritis. *Osteoarthr. Cartil.* **15**, 743–751 (2007).
195. Fransès, R. E., McWilliams, D. F., Mapp, P. I. & Walsh, D. A. Osteochondral angiogenesis and increased protease inhibitor expression in OA. *Osteoarthr. Cartil.* **18**, 563–571 (2010).

196. [review] Kraan, P. M. van der & Berg, W. B. van den. Osteophytes: relevance and biology. *Osteoarthr. Cartil.* **15**, 237–244 (2007).
197. Bowes, M. A. *et al.* Osteoarthritic bone marrow lesions almost exclusively collocate with denuded cartilage: a 3D study using data from the Osteoarthritis Initiative. *Ann. Rheum. Dis.* **75**, 1852–1857 (2015).
198. Klein-Wieringa, I. R. *et al.* The infrapatellar fat pad of patients with osteoarthritis has an inflammatory phenotype. *Ann. Rheum. Dis.* **70**, 851 (2011).
199. Ding, C. *et al.* Meniscal tear as an osteoarthritis risk factor in a largely non-osteoarthritic cohort: a cross-sectional study. *J. Rheumatol.* **34**, 776–84 (2007).
200. Hunter, D. J. *et al.* The association of meniscal pathologic changes with cartilage loss in symptomatic knee osteoarthritis. *Arthritis Rheum.* **54**, 795–801 (2006).
201. Tan, A. L. *et al.* Combined high-resolution magnetic resonance imaging and histological examination to explore the role of ligaments and tendons in the phenotypic expression of early hand osteoarthritis. *Ann. Rheum. Dis.* **65**, 1267 (2006).
202. Douglas, M. J. M., Hutchison, J. D. & Sutherland, A. G. Anterior cruciate ligament integrity in osteoarthritis of the knee in patients undergoing total knee replacement. *J. Orthop. Traumatol.* **11**, 149–154 (2010).
203. [review] Sacitharan, P. K. Biochemistry and Cell Biology of Ageing. *Subcell. Biochem.* **91**, 123–159 (2019).
204. [review] Anderson, A. S. & Loeser, R. F. Why is osteoarthritis an age-related disease? *Best Pr. Res. Clin. Rheumatol.* **24**, 15–26 (2010).
205. Xiao, S.-Q. *et al.* The role of apoptosis in the pathogenesis of osteoarthritis. *Int. Orthop.* **47**, 1895–1919 (2023).
206. Yang, X., Yan, K., Zhang, Q., Yusufu, A. & Ran, J. Meta-Analysis of Estrogen in Osteoarthritis: Clinical Status and Protective Effects. *Altern. Ther. Heal. Med.* **29**, 224–230 (2022).
207. [review] Linn, S., Murtaugh, B. & Casey, E. Role of Sex Hormones in the Development of Osteoarthritis. *PMR* **4**, S169–S173 (2012).

208. [review] Tschon, M., Contartese, D., Pagani, S., Borsari, V. & Fini, M. Gender and Sex Are Key Determinants in Osteoarthritis Not Only Confounding Variables. A Systematic Review of Clinical Data. *J. Clin. Med.* **10**, 3178 (2021).
209. Jeon, H. *et al.* Low skeletal muscle mass and radiographic osteoarthritis in knee, hip, and lumbar spine: a cross-sectional study. *Aging Clin. Exp. Res.* **31**, 1557–1562 (2019).
210. Franklin, J., Ingvarsson, T., Englund, M. & Lohmander, S. Association between occupation and knee and hip replacement due to osteoarthritis: a case-control study. *Arthritis Res. Ther.* **12**, R102 (2010).
211. Andersen, S., Thygesen, L. C., Davidsen, M. & Helweg-Larsen, K. Cumulative years in occupation and the risk of hip or knee osteoarthritis in men and women: a register-based follow-up study. *Occup. Environ. Med.* **69**, 325 (2012).
212. Momsen, W. E. & Brockman, H. L. Inhibition of pancreatic lipase B activity by taurodeoxycholate and its reversal by colipase. *J. Biol. Chem.* **251**, 384–8 (1976).
213. Bernard, T. E., Wilder, F. V., Aluoch, M. & Leaverton, P. E. Job-Related Osteoarthritis of the Knee, Foot, Hand, and Cervical Spine. *J. Occup. Environ. Med.* **52**, 33–38 (2010).
214. He, C. *et al.* Exercise-induced BCL2-regulated autophagy is required for muscle glucose homeostasis. *Nature* **481**, 511–515 (2012).
215. [review] Sun, H. B. Mechanical loading, cartilage degradation, and arthritis. *Ann. N. York Acad. Sci.* **1211**, 37–50 (2010).
216. Regnaud, J. *et al.* High-intensity versus low-intensity physical activity or exercise in people with hip or knee osteoarthritis. *Cochrane Database Syst. Rev.* **2015**, CD010203 (2015).
217. Cao, Z., Li, Q., Li, Y. & Wu, J. Causal association of leisure sedentary behavior with arthritis: A Mendelian randomization analysis. *Semin. Arthritis Rheum.* **59**, 152171 (2023).
218. Berenbaum, F., Wallace, I. J., Lieberman, D. E. & Felson, D. T. Identification of causal relationships between modern-day environmental factors in the pathogenesis of osteoarthritis. *Nat. Rev. Rheumatol.* **14**, 674–681 (2018).
219. Wang, C. *et al.* Causal associations of obesity related anthropometric indicators and body compositions with knee and hip arthritis: A large-scale genetic correlation study. *Front. Endocrinol.* **13**, 1011896 (2022).

220. [review] Francisco, V. *et al.* Biomechanics, obesity, and osteoarthritis. The role of adipokines: When the levee breaks. *J. Orthop. Res.* **36**, 594–604 (2018).
221. Cao, Z., Wu, Y., Li, Q., Li, Y. & Wu, J. A causal relationship between childhood obesity and risk of osteoarthritis: results from a two-sample Mendelian randomization analysis. *Ann. Med.* **54**, 1636–1645 (2022).
222. [review] Wang, T. & He, C. Pro-inflammatory cytokines: The link between obesity and osteoarthritis. *Cytokine Growth Factor Rev.* **44**, 38–50 (2018).
223. Guss, J. D. *et al.* The effects of metabolic syndrome, obesity, and the gut microbiome on load-induced osteoarthritis. *Osteoarthr. Cartil.* **27**, 129–139 (2019).
224. [review] Sampath, S. J. P., Venkatesan, V., Ghosh, S. & Kotikalapudi, N. Obesity, Metabolic Syndrome, and Osteoarthritis—An Updated Review. *Curr. Obes. Rep.* **12**, 308–331 (2023).
225. Norden, C. W. & Dickens, D. R. Experimental Osteomyelitis: Treatment with Cephaloridine. *J. Infect. Dis.* **127**, 525–528 (1973).
226. [review] Abramoff, B. & Caldera, F. E. Osteoarthritis Pathology, Diagnosis, and Treatment Options. *Méd. Clin. North Am.* **104**, 293–311 (2020).
227. [review] Gwinnutt, J. M. *et al.* 2021 EULAR recommendations regarding lifestyle behaviours and work participation to prevent progression of rheumatic and musculoskeletal diseases. *Ann. Rheum. Dis.* **82**, 48–56 (2023).
228. Richard, M. J., Driban, J. B. & McAlindon, T. E. Meta-analysis of pharmaceutical treatment of osteoarthritis. *Osteoarthr. Cartil.* **31**, 458–466 (2023).
229. [review] Buchanan, W. W., Hogan, M. G., Kean, C. A., Kean, W. F. & Rainsford, K. D. Surgery of joints. *Inflammopharmacology* 1–6 (2023) doi:10.1007/s10787-023-01224-x.
230. Vos, T. *et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1204–1222 (2020).
231. 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).

232. Jordan, K. P. *et al.* Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. *BMC Musculoskelet. Disord.* **11**, 144 (2010).
233. Coates, G. *et al.* Health economic impact of moderate-to-severe chronic pain associated with osteoarthritis in England: a retrospective analysis of linked primary and secondary care data. *BMJ Open* **13**, e067545 (2023).
234. Lohan, C. *et al.* Estimating the cost and epidemiology of mild to severe chronic pain associated with osteoarthritis in England: a retrospective analysis of linked primary and secondary care data. *BMJ Open* **13**, e073096 (2023).
235. Economics, Y. H. *The Cost of Arthritis: Calculation Conducted on Behalf of Arthritis Research UK.* (2017).
236. Versus, A. U. *The State of Musculoskeletal Health 2023.*
<https://versusarthritis.org/media/duybjusg/versus-arthritis-state-msk-musculoskeletal-health-2023pdf.pdf> (2023).
237. [review] Wang, S.-T. & Ni, G.-X. Depression in Osteoarthritis: Current Understanding. *Neuropsychiatr. Dis. Treat.* **18**, 375–389 (2022).
238. Stubbs, B., Aluko, Y., Myint, P. K. & Smith, T. O. Prevalence of depressive symptoms and anxiety in osteoarthritis: a systematic review and meta-analysis. *Age Ageing* **45**, 228–235 (2016).
239. Campos, G. C. de *et al.* Osteoarthritis, mobility-related comorbidities and mortality: meta-analysis. *Ther. Adv. Musculoskelet. Dis.* **12**, 1759720X20981219 (2020).
240. [review] Rudnicka, E. *et al.* The World Health Organization (WHO) approach to healthy ageing. *Maturitas* **139**, 6–11 (2020).
241. 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).
242. Valdes, A. M. *et al.* The GDF5 rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. *Ann. Rheum. Dis.* **70**, 873 (2011).
243. [review] Kumar, K., Cowley, M. & Davis, R. Next-Generation Sequencing and Emerging Technologies. *Semin. Thromb. Hemost.* **45**, 661–673 (2019).

244. [review] Wang, M. H., Cordell, H. J. & Steen, K. V. Statistical methods for genome-wide association studies. *Semin. Cancer Biol.* **55**, 53–60 (2019).
245. Kerkhof, H. J. M. *et al.* A genome-wide association study identifies an osteoarthritis susceptibility locus on chromosome 7q22. *Arthritis Rheum.* **62**, 499–510 (2010).
246. Evangelou, E. *et al.* A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. *Ann. Rheum. Dis.* **73**, 2130 (2014).
247. Casalone, E. *et al.* A novel variant in GLIS3 is associated with osteoarthritis. *Ann. Rheum. Dis.* **77**, 620 (2018).
248. Day-Williams, A. G. *et al.* A Variant in MCF2L Is Associated with Osteoarthritis. *Am. J. Hum. Genet.* **89**, 446–450 (2011).
249. Miyamoto, Y. *et al.* Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. *Nat. Genet.* **40**, 994–998 (2008).
250. Hackinger, S. *et al.* Evaluation of shared genetic aetiology between osteoarthritis and bone mineral density identifies SMAD3 as a novel osteoarthritis risk locus. *Hum. Mol. Genet.* **26**, 3850–3858 (2017).
251. Liu, Y. *et al.* Genetic Determinants of Radiographic Knee Osteoarthritis in African Americans. *J. Rheumatol.* **44**, 1652–1658 (2017).
252. Zengini, E. *et al.* Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. *Nat. Genet.* **50**, 549–558 (2018).
253. Evans, D. S. *et al.* Genome-wide association and functional studies identify a role for IGFBP3 in hip osteoarthritis. *Ann. Rheum. Dis.* **74**, 1861 (2015).
254. Hollander, W. den *et al.* Genome-wide association and functional studies identify a role for matrix Gla protein in osteoarthritis of the hand. *Ann. Rheum. Dis.* **76**, 2046 (2017).
255. Betancourt, M. C. C. *et al.* Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. *Proc. Natl. Acad. Sci.* **109**, 8218–8223 (2012).
256. Yau, M. S. *et al.* Genome-Wide Association Study of Radiographic Knee Osteoarthritis in North American Caucasians. *Arthritis Rheumatol.* **69**, 343–351 (2017).

257. arcOGEN, C. *et al.* Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* **380**, 815–823 (2012).
258. Castaño-Betancourt, M. C. *et al.* Novel Genetic Variants for Cartilage Thickness and Hip Osteoarthritis. *PLoS Genet.* **12**, e1006260 (2016).
259. Styrkarsdottir, U. *et al.* Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat. Genet.* **46**, 498–502 (2014).
260. 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 (2013).
261. Styrkarsdottir, U. *et al.* Whole-genome sequencing identifies rare genotypes in COMP and CHADL associated with high risk of hip osteoarthritis. *Nat. Genet.* **49**, 801–805 (2017).
262. 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).
263. Aubourg, G., Rice, S. J., Bruce-Wootton, P. & Loughlin, J. Genetics of osteoarthritis. *Osteoarthr. Cartil.* **30**, 636–649 (2022).
264. 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).
265. 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).
266. McDonald, M.-L. N. *et al.* Novel genetic loci associated with osteoarthritis in multi-ancestry analyses in the Million Veteran Program and UK Biobank. *Nat. Genet.* **54**, 1816–1826 (2022).
267. Nakajima, M. *et al.* New Sequence Variants in HLA Class II/III Region Associated with Susceptibility to Knee Osteoarthritis Identified by Genome-Wide Association Study. *PLoS ONE* **5**, e9723 (2010).

268. Gao, F., Yao, Y., Zhang, Y. & Tian, J. Integrating Genome-Wide Association Studies With Pathway Analysis and Gene Expression Analysis Highlights Novel Osteoarthritis Risk Pathways and Genes. *Front. Genet.* **10**, 827 (2019).
269. [review] Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *Am. J. Hum. Genet.* **102**, 717–730 (2018).
270. Nicolae, D. L. *et al.* Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
271. Liu, Y. *et al.* Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis. *Sci. Rep.* **8**, 15499 (2018).
272. Arensbergen, J. van *et al.* High-throughput identification of human SNPs affecting regulatory element activity. *Nat. Genet.* **51**, 1160–1169 (2019).
273. 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).
274. 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).
275. Shepherd, C. *et al.* Functional Characterization of the Osteoarthritis Genetic Risk Residing at ALDH1A2 Identifies rs12915901 as a Key Target Variant. *Arthritis Rheumatol.* **70**, 1577–1587 (2018).
276. Shepherd, C., Reese, A. E., Reynard, L. N. & Loughlin, J. Expression analysis of the osteoarthritis genetic susceptibility mapping to the matrix Gla protein gene MGP. *Arthritis Res. Ther.* **21**, 149 (2019).
277. Rice, S. J. *et al.* Genetic and Epigenetic Fine-Tuning of TGFB1 Expression Within the Human Osteoarthritic Joint. *Arthritis Rheumatol.* **73**, 1866–1877 (2021).
278. Sorial, A. K. *et al.* Multi-tissue epigenetic analysis of the osteoarthritis susceptibility locus mapping to the plectin gene PLEC. *Osteoarthr. Cartil.* **28**, 1448–1458 (2020).
279. Parker, E. *et al.* Multi-Tissue Epigenetic and Gene Expression Analysis Combined With Epigenome Modulation Identifies RWDD2B as a Target of Osteoarthritis Susceptibility. *Arthritis Rheumatol.* **73**, 100–109 (2021).

280. Peters, J. E. *et al.* Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease. *PLoS Genet.* **12**, e1005908 (2016).
281. [review] Sandell, L. J. Etiology of osteoarthritis: genetics and synovial joint development. *Nat. Rev. Rheumatol.* **8**, 77–89 (2012).
282. Richard, D. *et al.* Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk. *Cell* **181**, 362–381.e28 (2020).
283. Bos, S. D., Slagboom, P. E. & Meulenbelt, I. New insights into osteoarthritis: early developmental features of an ageing-related disease. *Curr. Opin. Rheumatol.* **20**, 553–559 (2008).
284. Carter, A. J. & Nguyen, A. Q. Antagonistic pleiotropy as a widespread mechanism for the maintenance of polymorphic disease alleles. *BMC Méd. Genet.* **12**, 160 (2011).
285. [review] Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016).
286. [review] Gibney, E. R. & Nolan, C. M. Epigenetics and gene expression. *Heredity* **105**, 4–13 (2010).
287. [review] Wan, C., Zhang, F., Yao, H., Li, H. & Tuan, R. S. Histone Modifications and Chondrocyte Fate: Regulation and Therapeutic Implications. *Front. Cell Dev. Biol.* **9**, 626708 (2021).
288. Kim, K., Park, Y. & Im, G. Changes in the epigenetic status of the SOX-9 promoter in human osteoarthritic cartilage. *J. Bone Miner. Res.* **28**, 1050–1060 (2013).
289. Ukita, M., Matsushita, K., Tamura, M. & Yamaguchi, T. Histone H3K9 methylation is involved in temporomandibular joint osteoarthritis. *Int. J. Mol. Med.* **45**, 607–614 (2020).
290. Maki, K. *et al.* Hydrostatic pressure prevents chondrocyte differentiation through heterochromatin remodeling. *J. cell Sci.* **134**, (2020).
291. Monteagudo, S. *et al.* DOT1L safeguards cartilage homeostasis and protects against osteoarthritis. *Nat. Commun.* **8**, 15889 (2017).

292. Roover, A. D. *et al.* Hypoxia induces DOT1L in articular cartilage to protect against osteoarthritis. *JCI Insight* **6**, e150451 (2021).
293. Cornelis, F. M. F. *et al.* Increased susceptibility to develop spontaneous and post-traumatic osteoarthritis in Dot1l-deficient mice. *Osteoarthr. Cartil.* **27**, 513–525 (2019).
294. Wang, P. *et al.* KDM6A promotes chondrogenic differentiation of periodontal ligament stem cells by demethylation of SOX9. *Cell Prolif.* **51**, e12413 (2018).
295. Jun, Z. *et al.* Jumonji domain containing-3 (JMJD3) inhibition attenuates IL-1 β -induced chondrocytes damage in vitro and protects osteoarthritis cartilage in vivo. *Inflamm. Res.* **69**, 657–666 (2020).
296. Chen, D. *et al.* Pharmacological blockade of PCAF ameliorates osteoarthritis development via dual inhibition of TNF- α -driven inflammation and ER stress. *EBioMedicine* **50**, 395–407 (2019).
297. Dvir-Ginzberg, M. *et al.* Tumor necrosis factor α -mediated cleavage and inactivation of SirT1 in human osteoarthritic chondrocytes. *Arthritis Rheum.* **63**, 2363–2373 (2011).
298. Batshon, G. *et al.* Serum NT/CT SIRT1 ratio reflects early osteoarthritis and chondrosenescence. *Ann. Rheum. Dis.* **79**, 1370–1380 (2020).
299. Sacitharan, P. K., Bou-Gharios, G. & Edwards, J. R. SIRT1 directly activates autophagy in human chondrocytes. *Cell death Discov.* **6**, 41 (2020).
300. Collins, J. A. *et al.* Sirtuin 6 (SIRT6) regulates redox homeostasis and signaling events in human articular chondrocytes. *Free Radic. Biol. Med.* **166**, 90–103 (2021).
301. Korogi, W. *et al.* SIRT7 is an important regulator of cartilage homeostasis and osteoarthritis development. *Biochem. Biophys. Res. Commun.* **496**, 891–897 (2018).
302. 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).
303. Tuddenham, L. *et al.* The cartilage specific microRNA-140 targets histone deacetylase 4 in mouse cells. *FEBS Lett.* **580**, 4214–4217 (2006).

304. Miyaki, S. *et al.* MicroRNA-140 plays dual roles in both cartilage development and homeostasis. *Genes Dev.* **24**, 1173–1185 (2010).
305. Tardif, G. *et al.* NFAT3 and TGF- β /SMAD3 regulate the expression of miR-140 in osteoarthritis. *Arthritis Res. Ther.* **15**, R197 (2013).
306. Si, H. -b. *et al.* Intra-articular injection of microRNA-140 (miRNA-140) alleviates osteoarthritis (OA) progression by modulating extracellular matrix (ECM) homeostasis in rats. *Osteoarthr. Cartil.* **25**, 1698–1707 (2017).
307. Yamasaki, K. *et al.* Expression of MicroRNA-146a in osteoarthritis cartilage. *Arthritis Rheum.* **60**, 1035–1041 (2009).
308. Budd, E., Andrés, M. C. de, Sanchez-Elsner, T. & Oreffo, R. O. C. MiR-146b is down-regulated during the chondrogenic differentiation of human bone marrow derived skeletal stem cells and up-regulated in osteoarthritis. *Sci. Rep.* **7**, 46704 (2017).
309. 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).
310. Cheleschi, S. *et al.* Hydrostatic Pressure Regulates Oxidative Stress through microRNA in Human Osteoarthritic Chondrocytes. *Int. J. Mol. Sci.* **21**, 3653 (2020).
311. [review] Barrett, S. P. & Salzman, J. Circular RNAs: analysis, expression and potential functions. *Development* **143**, 1838–1847 (2016).
312. Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).
313. Liu, Q. *et al.* Circular RNA Related to the Chondrocyte ECM Regulates MMP13 Expression by Functioning as a MiR-136 ‘Sponge’ in Human Cartilage Degradation. *Sci. Rep.* **6**, 22572 (2016).
314. Ajekigbe, B. *et al.* Identification of long non-coding RNAs expressed in knee and hip osteoarthritic cartilage. *Osteoarthr. Cartil.* **27**, 694–702 (2019).
315. Deng, Z. *et al.* tRNA-Derived Fragment tRF-5009A Regulates Autophagy and Degeneration of Cartilage in Osteoarthritis via Targeting mTOR. *Oxidative Med. Cell. Longev.* **2022**, 5781660 (2022).

316. Green, J. A., Ansari, M. Y., Ball, H. C. & Haqqi, T. M. tRNA-derived fragments (tRFs) regulate post-transcriptional gene expression via AGO-dependent mechanism in IL-1 β stimulated chondrocytes. *Osteoarthr. Cartil.* **28**, 1102–1110 (2020).
317. Peffers, M. J. *et al.* SnoRNA signatures in cartilage ageing and osteoarthritis. *Sci. Rep.* **10**, 10641 (2020).
318. [review] Scott, M. S. & Ono, M. From snoRNA to miRNA: Dual function regulatory non-coding RNAs. *Biochimie* **93**, 1987–1992 (2011).
319. Steinbusch, M. M. F. *et al.* Serum snoRNAs as biomarkers for joint ageing and post traumatic osteoarthritis. *Sci. Rep.* **7**, 43558 (2017).
320. [review] Golbabapour, S., Abdulla, M. A. & Hajrezaei, M. A Concise Review on Epigenetic Regulation: Insight into Molecular Mechanisms. *Int. J. Mol. Sci.* **12**, 8661–8694 (2011).
321. Kaluscha, S. *et al.* Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat. Genet.* **54**, 1895–1906 (2022).
322. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, (2017).
323. Wiehle, L. *et al.* DNA (de)methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries. *Genome Res.* **29**, 750–761 (2019).
324. Hashimoto, H. *et al.* Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. *Mol. Cell* **66**, 711-720.e3 (2017).
325. Du, Q., Luu, P.-L., Stirzaker, C. & Clark, S. J. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* **7**, 1051–1073 (2015).
326. Lee, H.-T., Oh, S., Ro, D. H., Yoo, H. & Kwon, Y.-W. The Key Role of DNA Methylation upon Histone Acetylation (H3K27ac) in Atherosclerosis plaques. *J. Lipid Atheroscler.* **9**, 419–434 (2020).
327. Li, S., Peng, Y. & Panchenko, A. R. DNA methylation Scanning: Computational method for accessing the precise modulation of chromatin structure and dynamics. *Curr. Opin. Struct. Biol.* **75**, 102430 (2022).

328. [review] Machado, A. C. D. *et al.* Evolving insights on how cytosine methylation affects protein–DNA binding. *Brief. Funct. Genom.* **14**, 61–73 (2015).
329. Rao, S. *et al.* Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein–DNA binding. *Epigenetics Chromatin* **11**, 6 (2018).
330. 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).
331. Zhu, X. *et al.* PPAR γ preservation via promoter demethylation alleviates osteoarthritis in mice. *Ann. Rheum. Dis.* **78**, 1420 (2019).
332. Shen, J. *et al.* DNA methyltransferase 3b regulates articular cartilage homeostasis by altering metabolism. *JCI Insight* **2**, e93612 (2017).
333. Smeriglio, P. *et al.* Inhibition of TET1 prevents the development of osteoarthritis and reveals the 5hmC landscape that orchestrates pathogenesis. *Sci. Transl. Med.* **12**, (2020).
334. Roach, H. I. *et al.* Association between the abnormal expression of matrix-degrading enzymes by human osteoarthritic chondrocytes and demethylation of specific CpG sites in the promoter regions. *Arthritis Rheum.* **52**, 3110–3124 (2005).
335. 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).
336. 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).
337. Andrés, M. C. de *et al.* Loss of methylation in CpG sites in the NF- κ B enhancer elements of inducible nitric oxide synthase is responsible for gene induction in human articular chondrocytes. *Arthritis Rheum.* **65**, 732–742 (2013).
338. Takahashi, A., Andrés, M. C. de, Hashimoto, K., Itoi, E. & Oreffo, R. O. C. Epigenetic regulation of interleukin-8, an inflammatory chemokine, in osteoarthritis. *Osteoarthr. Cartil.* **23**, 1946–1954 (2015).

339. Chauffier, K. *et al.* Induction of the chemokine IL-8/Kc by the articular cartilage: Possible influence on osteoarthritis. *Jt. Bone Spine* **79**, 604–609 (2012).
340. Yang, F. *et al.* Epigenetic modifications of interleukin-6 in synovial fibroblasts from osteoarthritis patients. *Sci. Rep.* **7**, 43592 (2017).
341. Wang, X. *et al.* Analysis of DNA methylation in chondrocytes in rats with knee osteoarthritis. *BMC Musculoskelet. Disord.* **18**, 377 (2017).
342. Takahashi, A. *et al.* DNA methylation of the RUNX2 P1 promoter mediates MMP13 transcription in chondrocytes. *Sci. Rep.* **7**, 7771 (2017).
343. Izda, V. *et al.* A Pilot Analysis of Genome-Wide DNA Methylation Patterns in Mouse Cartilage Reveals Overlapping Epigenetic Signatures of Aging and Osteoarthritis. *ACR Open Rheumatol.* **4**, 1004–1012 (2022).
344. Housman, G., Havill, L. M., Quillen, E. E., Comuzzie, A. G. & Stone, A. C. Assessment of DNA Methylation Patterns in the Bone and Cartilage of a Nonhuman Primate Model of Osteoarthritis. *CARTILAGE* **10**, 335–345 (2019).
345. Lin, X. *et al.* Genome-wide analysis of aberrant methylation of enhancer DNA in human osteoarthritis. *BMC Méd. Genom.* **13**, 1 (2020).
346. Zhao, L., Wang, Q., Zhang, C. & Huang, C. Genome-wide DNA methylation analysis of articular chondrocytes identifies TRAF1, CTGF, and CX3CL1 genes as hypomethylated in osteoarthritis. *Clin. Rheumatol.* **36**, 2335–2342 (2017).
347. Fernández-Tajes, J. *et al.* Genome-wide DNA methylation analysis of articular chondrocytes reveals a cluster of osteoarthritic patients. *Ann. Rheum. Dis.* **73**, 668 (2014).
348. Zhang, Y. *et al.* Genome-wide DNA methylation profile implicates potential cartilage regeneration at the late stage of knee osteoarthritis. *Osteoarthr. Cartil.* **24**, 835–843 (2016).
349. Taylor, S. E. B., Li, Y. H., Wong, W. H. & Bhutani, N. Genome-Wide Mapping of DNA Hydroxymethylation in Osteoarthritic Chondrocytes. *Arthritis Rheumatol.* **67**, 2129–2140 (2015).
350. Delgado-Calle, J. *et al.* Genome-wide profiling of bone reveals differentially methylated regions in osteoporosis and osteoarthritis. *Arthritis Rheum.* **65**, 197–205 (2013).

351. 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. *Osteoarthr. Cartil.* **27**, 1545–1556 (2019).
352. 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).
353. Rushton, M. D. *et al.* Methylation quantitative trait locus analysis of osteoarthritis links epigenetics with genetic risk. *Hum. Mol. Genet.* **24**, 7432–7444 (2015).
354. 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 Rheumatol.* **71**, 1285–1296 (2019).
355. Kreitmaier, P. *et al.* An epigenome-wide view of osteoarthritis in primary tissues. *Am. J. Hum. Genet.* **109**, 1255–1271 (2022).
356. Schegg, B., Hülsmeier, A. J., Rutschmann, C., Maag, C. & Hennet, T. Core Glycosylation of Collagen Is Initiated by Two $\beta(1\text{-O})$ Galactosyltransferases. *Mol. Cell. Biol.* **29**, 943–952 (2009).
357. Ricard-Blum, S. The Collagen Family. *Cold Spring Harb. Perspect. Biol.* **3**, a004978 (2011).
358. [review] Myllyharju, J. Collagen, Primer in Structure, Processing and Assembly. *Top. Curr. Chem.* 115–147 (2005) doi:10.1007/b103821.
359. [review] Gelse, K., Pöschl, E. & Aigner, T. Collagens—structure, function, and biosynthesis. *Adv. Drug Deliv. Rev.* **55**, 1531–1546 (2003).
360. Kavitha, O. & Thampan, R. V. Factors influencing collagen biosynthesis under normal and fibrotic conditions. *J. Cell. Biochem.* **104**, 1150–1160 (2008).
361. [review] Myllyharju, J. & Kivirikko, K. I. Collagens and collagen-related diseases. *Ann. Med.* **33**, 7–21 (2001).
362. Mienaltowski, M. J., Gonzales, N. L., Beall, J. M. & Pechanec, M. Y. Identification of novel mechanisms governing heritable Soft Connective Tissue Diseases. *Adv. Exp. Med. Biol.* **1348**, 5–43 (2021).
363. [review] Bella, J. & Hulmes, D. J. S. Fibrous Proteins: Structures and Mechanisms. *Subcell. Biochem.* **82**, 457–490 (2017).

364. [review] Soroushanova, A. *et al.* The Collagen Suprafamily: From Biosynthesis to Advanced Biomaterial Development. *Adv. Mater.* **31**, e1801651 (2019).
365. Marini, J. C., Reich, A. & Smith, S. M. Osteogenesis imperfecta due to mutations in non-collagenous genes. *Curr. Opin. Pediatr.* **26**, 500–507 (2014).
366. Brady, A. F. *et al.* The Ehlers–Danlos syndromes, rare types. *Am. J. Méd. Genet. Part C: Semin. Méd. Genet.* **175**, 70–115 (2017).
367. [review] Mienaltowski, M. J. & Birk, D. E. Progress in Heritable Soft Connective Tissue Diseases. *Adv. Exp. Med. Biol.* **802**, 5–29 (2013).
368. Geister, K. A. *et al.* Loss of function of Colgalt1 disrupts collagen post-translational modification and causes musculoskeletal defects. *Dis. Model. Mech.* **12**, dmm037176 (2019).
369. Miyatake, S. *et al.* Biallelic COLGALT1 variants are associated with cerebral small vessel disease. *Ann. Neurol.* **84**, 843–853 (2018).
370. Yang, J. *et al.* Collagen $\beta(1-O)$ galactosyltransferase 2 deficiency contributes to lipodystrophy and aggravates NAFLD related to HMW adiponectin in mice. *Metabolism* **120**, 154777 (2021).
371. Rautavuoma, K. *et al.* Premature aggregation of type IV collagen and early lethality in lysyl hydroxylase 3 null mice. *Proc. Natl. Acad. Sci.* **101**, 14120–14125 (2004).
372. Chen, Y. *et al.* Lysyl Hydroxylase 2 Is Secreted by Tumor Cells and Can Modify Collagen in the Extracellular Space*. *J. Biol. Chem.* **291**, 25799–25808 (2016).
373. Remst, D. F. G. *et al.* Osteoarthritis-related fibrosis is associated with both elevated pyridinoline cross-link formation and lysyl hydroxylase 2b expression. *Osteoarthr. Cartil.* **21**, 157–164 (2013).
374. Salo, A. M. *et al.* A Connective Tissue Disorder Caused by Mutations of the Lysyl Hydroxylase 3 Gene. *Am. J. Hum. Genet.* **83**, 495–503 (2008).
375. Giunta, C. *et al.* Nevo syndrome is allelic to the kyphoscoliotic type of the Ehlers–Danlos syndrome (EDS VIA). *Am. J. Méd. Genet. Part A* **133A**, 158–164 (2005).
376. Rauch, F. *et al.* Cole-Carpenter Syndrome Is Caused by a Heterozygous Missense Mutation in P4HB. *Am. J. Hum. Genet.* **96**, 425–431 (2015).

377. Aro, E. *et al.* Severe Extracellular Matrix Abnormalities and Chondrodysplasia in Mice Lacking Collagen Prolyl 4-Hydroxylase Isoenzyme II in Combination with a Reduced Amount of Isoenzyme I*. *J. Biol. Chem.* **290**, 16964–16978 (2015).
378. Ha-Vinh, R. *et al.* Phenotypic and molecular characterization of Bruck syndrome (osteogenesis imperfecta with contractures of the large joints) caused by a recessive mutation in PLOD2. *Am. J. Méd. Genet. Part A* **131A**, 115–120 (2004).
379. Zou, Y. *et al.* P4HA1 mutations cause a unique congenital disorder of connective tissue involving tendon, bone, muscle and the eye. *Hum. Mol. Genet.* **26**, 2207–2217 (2017).
380. Wohlgemuth, R. P., Brashear, S. E. & Smith, L. R. Alignment, cross linking, and beyond: a collagen architect's guide to the skeletal muscle extracellular matrix. *Am. J. Physiol.-Cell Physiol.* **325**, C1017–C1030 (2023).
381. Gajko-Galicka, A. Mutations in type I collagen genes resulting in osteogenesis imperfecta in humans. *Acta Biochim. Pol.* **49**, 433–441 (2002).
382. Venable, E., Knight, D. R. T., Thoreson, E. K. & Baudhuin, L. M. COL1A1 and COL1A2 variants in Ehlers-Danlos syndrome phenotypes and COL1-related overlap disorder. *Am. J. Méd. Genet. Part C: Semin. Méd. Genet.* **193**, 147–159 (2023).
383. Kannu, P., Bateman, J. & Savarirayan, R. Clinical phenotypes associated with type II collagen mutations. *J. Paediatr. Child Heal.* **48**, E38–E43 (2012).
384. Handa, A., Grigelioniene, G. & Nishimura, G. Radiologic Features of Type II and Type XI Collagenopathies. *RadioGraphics* **41**, 192–209 (2020).
385. Hoornaert, K. P. *et al.* Stickler syndrome caused by COL2A1 mutations: genotype–phenotype correlation in a series of 100 patients. *Eur. J. Hum. Genet.* **18**, 872–880 (2010).
386. Guo, L. *et al.* Novel and recurrent COL11A1 and COL2A1 mutations in the Marshall–Stickler syndrome spectrum. *Hum. Genome Var.* **4**, 17040 (2017).
387. Frank, M. *et al.* The type of variants at the COL3A1 gene associates with the phenotype and severity of vascular Ehlers–Danlos syndrome. *Eur. J. Hum. Genet.* **23**, 1657–1664 (2015).
388. Storey, H., Savige, J., Sivakumar, V., Abbs, S. & Flinter, F. A. COL4A3/COL4A4 Mutations and Features in Individuals with Autosomal Recessive Alport Syndrome. *J. Am. Soc. Nephrol.* **24**, 1945–1954 (2013).

389. Daga, S. *et al.* Slowly progressive autosomal dominant Alport Syndrome due to COL4A3 splicing variant. *Eur. J. Hum. Genet.* 1–7 (2024) doi:10.1038/s41431-024-01706-8.
390. Pedchenko, V., Kitching, A. R. & Hudson, B. G. Goodpasture's autoimmune disease — A collagen IV disorder. *Matrix Biol.* **71**, 240–249 (2018).
391. Kiyohara, T. *et al.* COL5A1 gene mutation in three cases of classic type Ehlers–Danlos syndrome. *J. Dermatol. Sci.* **84**, e144 (2016).
392. Deconinck, N. *et al.* Bethlem myopathy: long-term follow-up identifies COL6 mutations predicting severe clinical evolution. *J. Neurol., Neurosurg. Psychiatry* **86**, 1337 (2015).
393. Bönnemann, C. G. Chapter 5 The collagen VI-related myopathies Ullrich congenital muscular dystrophy and Bethlem myopathy. *Handb. Clin. Neurol.* **101**, 81–96 (2011).
394. Dang, N. & Murrell, D. F. Mutation analysis and characterization of COL7A1 mutations in dystrophic epidermolysis bullosa. *Exp. Dermatol.* **17**, 553–568 (2008).
395. Biswas, S. *et al.* Missense mutations in COL8A2, the gene encoding the $\alpha 2$ chain of type VIII collagen, cause two forms of corneal endothelial dystrophy. *Hum. Mol. Genet.* **10**, 2415–2423 (2001).
396. Jackson, G. C. *et al.* Type IX collagen gene mutations can result in multiple epiphyseal dysplasia that is associated with osteochondritis dissecans and a mild myopathy. *Am. J. Méd. Genet. Part A* **152A**, 863–869 (2010).
397. Maghsoudi, D. *et al.* Retinal detachment in Type IX collagen recessive Stickler syndrome. *Eye* 1–6 (2024) doi:10.1038/s41433-024-03393-7.
398. Kong, L. *et al.* Identification of two novel COL10A1 heterozygous mutations in two Chinese pedigrees with Schmid-type metaphyseal chondrodysplasia. *BMC Méd. Genet.* **20**, 200 (2019).
399. Bateman, J. F., Wilson, R., Freddi, S., Lamandé, S. R. & Savarirayan, R. Mutations of COL10A1 in Schmid metaphyseal chondrodysplasia. *Hum. Mutat.* **25**, 525–534 (2005).
400. Micale, L. *et al.* Exon-Trapping Assay Improves Clinical Interpretation of COL11A1 and COL11A2 Intronic Variants in Stickler Syndrome Type 2 and Otospondylomegaepiphyseal Dysplasia. *Genes* **11**, 1513 (2020).

401. Zou, Y. *et al.* Recessive and dominant mutations in COL12A1 cause a novel EDS/myopathy overlap syndrome in humans and mice. *Hum. Mol. Genet.* **23**, 2339–2352 (2014).
402. NISHIE, W. Collagen XVII Processing and Blistering Skin Diseases. *Acta Derm.-Venereol.* **100**, 5662 (2020).
403. Patel, C. K. *et al.* Whole genome sequencing in a Knobloch syndrome family confirms the molecular diagnosis. *Ophthalmic Genet.* **43**, 201–209 (2022).
404. [review] Yamauchi, M. & Shiiba, M. Post-translational Modifications of Proteins, Tools for Functional Proteomics. *Methods Mol. Biol.* **446**, 95–108 (2008).
405. [review] Kivirikko, K. I. Posttranslational processing of collagens. *Princ. Méd. Biol.* **3**, 233–254 (1996).
406. Sipilä, K. H. *et al.* Proline hydroxylation in collagen supports integrin binding by two distinct mechanisms. *J. Biol. Chem.* **293**, 7645–7658 (2018).
407. Rappu, P., Salo, A. M., Myllyharju, J. & Heino, J. Role of prolyl hydroxylation in the molecular interactions of collagens. *Essays Biochem.* **63**, 325–335 (2019).
408. Fujii, K. K. *et al.* The Thermal Stability of the Collagen Triple Helix Is Tuned According to the Environmental Temperature. *Int. J. Mol. Sci.* **23**, 2040 (2022).
409. Holmgren, S. K., Taylor, K. M., Bretscher, L. E. & Raines, R. T. Code for collagen's stability deciphered. *Nature* **392**, 666–667 (1998).
410. Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. Crystal and Molecular Structure of a Collagen-Like Peptide at 1.9 Å Resolution. *Science* **266**, 75–81 (1994).
411. Terajima, M. *et al.* Glycosylation and Cross-linking in Bone Type I Collagen*. *J. Biol. Chem.* **289**, 22636–22647 (2014).
412. Kadler, K. E., Hojima, Y. & Prockop, D. J. Assembly of collagen fibrils de novo by cleavage of the type I pC-collagen with procollagen C-proteinase. Assay of critical concentration demonstrates that collagen self-assembly is a classical example of an entropy-driven process. *J. Biol. Chem.* **262**, 15696–701 (1987).
413. Silver, F. H., Freeman, J. W. & Seehra, G. P. Collagen self-assembly and the development of tendon mechanical properties. *J. Biomech.* **36**, 1529–1553 (2003).

414. Eekhoff, J. D., Fang, F. & Lake, S. P. Multiscale mechanical effects of native collagen cross-linking in tendon. *Connect. Tissue Res.* **59**, 410–422 (2018).
415. Beldowski, P. *et al.* Collagen type II–hyaluronan interactions – the effect of proline hydroxylation: a molecular dynamics study. *J. Mater. Chem. B* **10**, 9713–9723 (2022).
416. Bates, C. J., Prynne, C. J. & Levene, C. I. Ascorbate-dependent differences in the hydroxylation of proline and lysine in collagen synthesized by 3T6 fibroblasts in culture. *Biochim. Biophys. Acta (BBA) - Protein Struct.* **278**, 610–616 (1972).
417. Boyera, N., Galey, I. & Bernard, B. A. Effect of vitamin C and its derivatives on collagen synthesis and cross-linking by normal human fibroblasts. *Int. J. Cosmet. Sci.* **20**, 151–158 (1998).
418. Holster, T. *et al.* Loss of Assembly of the Main Basement Membrane Collagen, Type IV, but Not Fibril-forming Collagens and Embryonic Death in Collagen Prolyl 4-Hydroxylase I Null Mice*. *J. Biol. Chem.* **282**, 2512–2519 (2007).
419. Guo, H. *et al.* Mutations of P4HA2 encoding prolyl 4-hydroxylase 2 are associated with nonsyndromic high myopia. *Genet. Med.* **17**, 300–306 (2015).
420. Yeowell, H. N. & Walker, L. C. Mutations in the Lysyl Hydroxylase 1 Gene That Result in Enzyme Deficiency and the Clinical Phenotype of Ehlers–Danlos Syndrome Type VI. *Mol. Genet. Metab.* **71**, 212–224 (2000).
421. Slot, A. J. van der *et al.* Identification of PLOD2 as Telopeptide Lysyl Hydroxylase, an Important Enzyme in Fibrosis*. *J. Biol. Chem.* **278**, 40967–40972 (2003).
422. Risteli, M. *et al.* Reduction of Lysyl Hydroxylase 3 Causes Deleterious Changes in the Deposition and Organization of Extracellular Matrix*. *J. Biol. Chem.* **284**, 28204–28211 (2009).
423. Pornprasertsuk, S., Duarte, W. R., Mochida, Y. & Yamauchi, M. Overexpression of Lysyl Hydroxylase-2b Leads to Defective Collagen Fibrillogenesis and Matrix Mineralization. *J. Bone Miner. Res.* **20**, 81–87 (2005).
424. Slot, A. J. van der *et al.* Elevated formation of pyridinoline cross-links by profibrotic cytokines is associated with enhanced lysyl hydroxylase 2b levels. *Biochim. Biophys. Acta (BBA) - Mol. Basis Dis.* **1741**, 95–102 (2005).

425. Mattoteia, D. *et al.* Identification of Regulatory Molecular “Hot Spots” for LH/PLOD Collagen Glycosyltransferase Activity. *Int. J. Mol. Sci.* **24**, 11213 (2023).
426. Stawikowski, M. J., Aukszi, B., Stawikowska, R., Cudic, M. & Fields, G. B. Glycosylation Modulates Melanoma Cell $\alpha 2\beta 1$ and $\alpha 3\beta 1$ Integrin Interactions with Type IV Collagen*. *J. Biol. Chem.* **289**, 21591–21604 (2014).
427. Stammers, M. *et al.* Age-related changes in the physical properties, cross-linking, and glycation of collagen from mouse tail tendon. *J. Biol. Chem.* **295**, 10562–10571 (2020).
428. Murray, R. E., Barner, H. B. & Gantner, G. E. Diffuse Aneurysmal Degeneration of the Femoropopliteal Artery. *Arch. Surg.* **100**, 299–301 (1970).
429. Merl-Pham, J. *et al.* Quantitative proteomic profiling of extracellular matrix and site-specific collagen post-translational modifications in an in vitro model of lung fibrosis. *Matrix Biol. Plus* **1**, 100005 (2019).
430. He, L. *et al.* Down-regulation of GLT25D1 inhibited collagen secretion and involved in liver fibrogenesis. *Gene* **729**, 144233 (2020).
431. Liefhebber, J. M., Punt, S., Spaan, W. J. & Leeuwen, H. C. van. The human collagen beta(1-O)galactosyltransferase, GLT25D1, is a soluble endoplasmic reticulum localized protein. *BMC Cell Biol.* **11**, 33 (2010).
432. Scietti, L. *et al.* Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. *Nat. Commun.* **9**, 3163 (2018).
433. Wang, C. *et al.* The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagens in vitro. *Matrix Biol.* **21**, 559–566 (2002).
434. Ruotsalainen, H. *et al.* Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. *J. Cell Sci.* **119**, 625–635 (2006).
435. Tajima, S., Takehana, M. & Azuma, N. Production of Overmodified Type I Procollagen in a Case of Osteogenesis Imperfecta. *J. Dermatol.* **21**, 219–222 (1994).
436. Cabral, W. A. *et al.* Abnormal Type I Collagen Post-translational Modification and Crosslinking in a Cyclophilin B KO Mouse Model of Recessive Osteogenesis Imperfecta. *PLoS Genet.* **10**, e1004465 (2014).

437. Baumann, S. & Hennet, T. Collagen Accumulation in Osteosarcoma Cells lacking GLT25D1 Collagen Galactosyltransferase*. *J. Biol. Chem.* **291**, 18514–18524 (2016).
438. DeGroot, J. *et al.* Accumulation of advanced glycation end products as a molecular mechanism for aging as a risk factor in osteoarthritis. *Arthritis Rheum.* **50**, 1207–1215 (2004).
439. Reigle, K. L. *et al.* Non-enzymatic glycation of type I collagen diminishes collagen–proteoglycan binding and weakens cell adhesion. *J. Cell. Biochem.* **104**, 1684–1698 (2008).
440. Eyre, D. R. & Weis, M. A. Bone Collagen: New Clues to Its Mineralization Mechanism from Recessive Osteogenesis Imperfecta. *Calcif. Tissue Int.* **93**, 338–347 (2013).
441. Uzel, M. I. *et al.* Multiple Bone Morphogenetic Protein 1-related Mammalian Metalloproteinases Process Pro-lysyl Oxidase at the Correct Physiological Site and Control Lysyl Oxidase Activation in Mouse Embryo Fibroblast Cultures. *J. Biol. Chem.* **276**, 22537–22543 (2001).
442. Piersma, B. & Bank, R. A. Collagen cross-linking mediated by lysyl hydroxylase 2: an enzymatic battlefield to combat fibrosis. *Essays Biochem.* **63**, 377–387 (2019).
443. Mäki, J. M. *et al.* Inactivation of the Lysyl Oxidase Gene Lox Leads to Aortic Aneurysms, Cardiovascular Dysfunction, and Perinatal Death in Mice. *Circulation* **106**, 2503–2509 (2002).
444. [review] Kivirikko, K. I. & Pihlajaniemi, T. Advances in Enzymology and Related Areas of Molecular Biology. *Adv. Enzym. - Relat. Areas Mol. Biol.* **72**, 325–398 (2016).
445. Harris, J. R., Soliakov, A. & Lewis, R. J. In vitro fibrillogenesis of collagen type I in varying ionic and pH conditions. *Micron* **49**, 60–68 (2013).
446. [review] Siadat, S. M. & Ruberti, J. W. Mechanochemistry of collagen. *Acta Biomater.* **163**, 50–62 (2023).

Chapter 2. Methods and Materials

2.1. Tissue Samples

2.1.1. Arthroplasty samples for molecular analyses

The cartilage samples used in my studies were obtained from 145 patients undergoing joint arthroplasty at the Newcastle upon Tyne NHS Foundation Trust hospitals for primary hip OA (n = 50), primary knee OA (n = 61), or for a neck-of-femur (NOF) fracture (n = 29). Patient details are available in Table 2.1.

Table 2.1. Patient details of the arthroplasty cartilage samples used for molecular analyses

Sample	Sex	Joint	Age	Disease	Sample	Sex	Joint	Age	Disease	Sample	Sex	Joint	Age	Disease
T005	Male	Hip	69	NOF	117	Male	Hip	82	OA	5163	Male	Knee	55	OA
T007	Female	Hip	71	NOF	126	Female	Hip	66	OA	5164	Male	Knee	55	OA
T023	Female	Hip	68	NOF	5137	Female	Hip	52	OA	5178	Female	Knee	83	OA
T117	Male	Hip	85	NOF	5726	Female	Hip	82	OA	5188	Female	Knee	64	OA
T150	Female	Hip	80	NOF	5764	Male	Hip	64	OA	5214	Male	Knee	57	OA
T151	Male	Hip	62	NOF	6156	Female	Hip	67	OA	5999	Female	Knee	79	OA
T152	Female	Hip	95	NOF	6218	Female	Hip	79	OA	6135	Male	Knee	71	OA
T154	Male	Hip	75	NOF	7007	Male	Hip	77	OA	6184	Female	Knee	55	OA
T158	Male	Hip	79	NOF	7008	Female	Hip	75	OA	6224	Female	Knee	62	OA
T167	Female	Hip	62	NOF	7015	Female	Hip	64	OA	6342	Female	Knee	62	OA
T168	Female	Hip	81	NOF	7025	Female	Hip	76	OA	6359	Female	Knee	70	OA
T172	Female	Hip	77	NOF	7029	Female	Hip	67	OA	6363	Female	Knee	68	OA
T174	Male	Hip	86	NOF	7030	Male	Hip	63	OA	6378	Male	Knee	54	OA
T177	Female	Hip	72	NOF	7033	Female	Hip	81	OA	6506	Female	Knee	59	OA
T178	Female	Hip	62	NOF	7034	Female	Hip	60	OA	6770	Female	Knee	80	OA
T179	Female	Hip	81	NOF	7036	Female	Hip	88	OA	6772	Male	Knee	71	OA
T184	Female	Hip	62	NOF	7037	Female	Hip	80	OA	6778	Male	Knee	63	OA
T191	Female	Hip	87	NOF	7039	Female	Hip	68	OA	6783	Female	Knee	55	OA
T192	Female	Hip	71	NOF	7040	Male	Hip	74	OA	6784	Female	Knee	46	OA
T193	Female	Hip	86	NOF	7046	Female	Hip	56	OA	6786	Male	Knee	82	OA
T195	Female	Hip	80	NOF	7047	Female	Hip	78	OA	6787	Male	Knee	64	OA
T196	Female	Hip	83	NOF	7059	Male	Hip	81	OA	6788	Male	Knee	66	OA
T197	Female	Hip	92	NOF	39	Female	Knee	72	OA	6803	Female	Knee	65	OA
T199	Female	Hip	73	NOF	41	Male	Knee	51	OA	6818	Female	Knee	62	OA
T204	Female	Hip	76	NOF	45	Male	Knee	83	OA	6852	Female	Knee	83	OA
T244	Female	Hip	97	NOF	57	Male	Knee	63	OA	6867	Male	Knee	86	OA
T245	Female	Hip	81	NOF	59	Female	Knee	60	OA	7115	Female	Knee	83	OA
T246	Female	Hip	83	NOF	61	Male	Knee	67	OA	7116	Male	Knee	51	OA
T5026	Female	Hip	66	NOF	72	Male	Knee	76	OA	7118	Female	Knee	47	OA
19	Male	Hip	66	OA	76	Female	Knee	65	OA	7120	Female	Knee	61	OA
22	Male	Hip	48	OA	77	Female	Knee	71	OA	7124	Male	Knee	64	OA
40	Female	Hip	84	OA	78	Male	Knee	58	OA	6363	Female	Knee	68	OA
42	Female	Hip	71	OA	79	Male	Knee	65	OA	6378	Male	Knee	54	OA
49	Male	Hip	68	OA	82	Female	Knee	67	OA	6506	Female	Knee	59	OA
51	Female	Hip	74	OA	86	Female	Knee	56	OA	6770	Female	Knee	80	OA
52	Female	Hip	74	OA	89	Female	Knee	67	OA	6772	Male	Knee	71	OA
53	Female	Hip	55	OA	103	Female	Knee	70	OA	6778	Male	Knee	63	OA
54	Male	Hip	52	OA	104	Female	Knee	70	OA	6783	Female	Knee	55	OA
62	Male	Hip	91	OA	106	Male	Knee	79	OA	6784	Female	Knee	46	OA
66	Female	Hip	56	OA	107	Female	Knee	41	OA	6786	Male	Knee	82	OA
69	Male	Hip	57	OA	108	Male	Knee	72	OA	6787	Male	Knee	64	OA
70	Male	Hip	75	OA	109	Female	Knee	65	OA	6788	Male	Knee	66	OA
73	Female	Hip	46	OA	114	Female	Knee	57	OA	6803	Female	Knee	65	OA
87	Female	Hip	51	OA	115	Male	Knee	69	OA	6818	Female	Knee	62	OA
90	Female	Hip	60	OA	127	Male	Knee	61	OA	6852	Female	Knee	83	OA
92	Female	Hip	55	OA	128	Male	Knee	64	OA	6867	Male	Knee	86	OA
96	Female	Hip	61	OA	4809	Male	Knee	79	OA	7115	Female	Knee	83	OA
97	Male	Hip	65	OA	4883	Female	Knee	68	OA	7116	Male	Knee	51	OA
98	Male	Hip	49	OA	4971	Male	Knee	64	OA	7118	Female	Knee	47	OA
100	Female	Hip	74	OA	5099	Female	Knee	57	OA	7120	Female	Knee	61	OA
112	Male	Hip	61	OA	5103	Female	Knee	58	OA	7124	Male	Knee	64	OA
116	Male	Hip	89	OA	5104	Female	Knee	67	OA					

The synovium samples used in my studies were obtained from 88 patients undergoing joint arthroplasty for primary knee OA. Patient details are available in Table 2.2.

Table 2.2. Patient details of the arthroplasty synovium samples used for molecular analyses

Sample	Sex	Age	Sample	Sex	Age
3085	Female	55	5670	Female	73
3334	Female	73	5677	Male	54
3348	Male	57	5680	Female	70
3390	Female	79	5681	Female	51
3502	Male	67	5692	Female	66
3975	Male	60	5694	Female	81
4288	Male	85	5712	Female	67
4338	Female	79	5713	Female	65
4552	Female	82	5720	Male	68
4851	Male	58	5723	Male	79
4983	Male	75	5725	Male	81
4984	Male	62	5729	Female	58
5014	Male	55	5731	Female	58
5019	Male	50	5740	Female	87
5023	Male	75	5747	Female	88
5240	Female	62	5748	Male	67
5254	Female	62	5763	Female	54
5271	Female	68	5770	Female	52
5281	Female	59	5776	Female	85
5299	Female	75	5798	Female	52
5300	Male	87	5809	Male	79
5377	Male	80	5834	Male	71
5378	Male	65	5857	Male	75
5380	Male	72	6056	Female	68
5400	Female	62	6068	Female	59
5422	Female	62	6094	Female	54
5536	Female	53	6535	Male	65
5551	Male	49	6547	Male	72
5552	Female	54	6548	Female	72
5558	Female	75	6549	Male	63
5561	Female	71	6550	Female	51
5563	Male	64	6593	Female	50
5564	Female	73	6597	Female	78
5569	Female	63	6599	Male	72
5572	Female	72	6603	Female	67
5575	Male	61	6624	Female	66
5597	Male	71	6636	Male	53
5614	Male	60	6638	Female	84
5622	Female	61	6655	Female	70
5643	Male	76	6810	Female	69
5644	Male	76	6877	Female	63
5648	Female	66	6878	Male	58
5654	Male	85	6890	Female	60
5665	Male	61	6892	Female	66

Ethical approval was granted by the NHS Health Research Authority with each donor providing written consent (REC reference number 19/LO/0389, awarded to Professor Loughlin).

2.1.2. Developmental samples for molecular analyses

Human developmental tissues were obtained from the Human Developmental Biology Resource (HDBR) at Newcastle University (<http://www.hdbr.org>, project number 200363), with appropriate maternal written consent and approval from the Newcastle and North Tyneside NHS Health Authority Joint Ethics Committee. Details for the developmental samples used for my projects are listed in Table 2.3.

2.1.3. Arthroplasty and developmental samples for ATAC-seq

For the ATAC-seq experiment, ten hip and ten knee samples from OA patients undergoing arthroplasty were collected from Newcastle upon Tyne NHS Foundation Trust hospitals. Additionally, the cartilage from the proximal end (hip) and the distal end (knee) of six foetal (12pcw) femora was provided by the HDBR. The details are listed in Table 2.4.

Table 2.3. Details of the developmental cartilage samples used for molecular analyses. pcw, post conception weeks.

Sample	Sex	Tissue	Stage (pcw)	Sample	Sex	Tissue	Stage (pcw)
1193	Female	limb bud	10	14378	Male	Cartilage	14
1194	Female	limb bud	12	14392	Male	Cartilage	9
1820	Male	limb bud	CS23	14393	Male	Cartilage	14
1912	Female	limb bud	CS20	14408	Female	Cartilage	14
1955	Female	limb bud	10	14472	Male	Cartilage	12
10885	Female	limb bud	9	14492	Female	Cartilage	12
11993	Male	limb bud	CS13	14510	Female	Cartilage	14
11956	Female	limb bud	CS22	14511	Female	Cartilage	11
12023	Female	limb bud	10	14512	Male	Cartilage	10
12331	Female	limb bud	CS17	14516	Male	Cartilage	15
12388	Male	limb bud	CS23	14521	Male	Cartilage	12
12510	Male	limb bud	11	14522	Female	Cartilage	15
12630	Female	limb bud	CS23	14523	Female	Cartilage	15
12637	Male	limb bud	9	14524	Female	Cartilage	15
12692	Male	limb bud	CS17	14525	Male	Cartilage	10
14040	Female	limb bud	CS22	14532	Female	Cartilage	11
14090	Male	limb bud	CS19	14541	Female	Cartilage	11
14109	Male	limb bud	CS18	14577	Female	Cartilage	11
14339	Female	limb bud	CS19	14597	Male	Cartilage	11
14375	Male	Cartilage	12	14601	Male	Cartilage	10
14395	Female	Cartilage	10	14603	Female	Cartilage	11
14397	Female	Cartilage	15	14619	Male	Cartilage	16
14423	Male	Cartilage	12	14628	Female	Cartilage	17
14429	Female	Cartilage	14	14681	Female	Cartilage	11
14451	Male	Cartilage	9	14688	Male	Cartilage	11
14453	Female	Cartilage	12	14689	Male	Cartilage	10
14460	Female	Cartilage	9	14703	Female	Cartilage	12
14464	Male	Cartilage	16	14713	Female	Cartilage	16
14466	Male	Cartilage	12	14715	Male	Cartilage	12
14467	Male	Cartilage	10	14716	Male	Cartilage	10
14475	Male	Cartilage	16	14717	Female	Cartilage	14
14501	Male	Cartilage	16	14720	Female	Cartilage	15
14513	Female	Cartilage	12	14721	Female	Cartilage	10
14563	Male	Cartilage	14	14722	Male	Cartilage	17
14586	Female	Cartilage	13	14727	Male	Cartilage	12
14604	Female	Cartilage	14	14728	Male	Cartilage	16
14471	Female	Cartilage	9	14729	Male	Cartilage	12
14555	Male	Cartilage	9	14738	Male	Cartilage	15
14562	Female	Cartilage	9	14746	Female	Cartilage	16
14580	Male	Cartilage	9	14748	Female	Cartilage	10
14600	Male	Cartilage	8	14749	Male	Cartilage	13
14576	Male	Cartilage	8	14782	Male	Cartilage	10
14544	Male	Cartilage	13	14797	Male	Cartilage	12
14684	Female	Cartilage	14	14831	Female	Cartilage	9
14606	Female	Cartilage	9	14835	Female	Cartilage	12
14617	Female	Cartilage	13	14863	Male	Cartilage	10
14377	Male	Cartilage	14	14881	Female	Cartilage	10

CS - Carnegie Stage

Table 2.4. Details of the developmental and OA cartilage samples used for ATAC-seq. Red OA samples were digested with Type I collagenase overnight and the black OA samples were digested for 12h.

Sample	Sample type	Joint	Age
15192	Foetal	Proximal femur (hip)	12pcw
15192	Foetal	Distal femur (knee)	12pcw
15209	Foetal	Proximal femur (hip)	12pcw
15209	Foetal	Distal femur (knee)	12pcw
15210	Foetal	Proximal femur (hip)	12pcw
15210	Foetal	Distal femur (knee)	12pcw
15271	Foetal	Proximal femur (hip)	12pcw
15271	Foetal	Distal femur (knee)	12pcw
15285	Foetal	Proximal femur (hip)	12pcw
15285	Foetal	Distal femur (knee)	12pcw
15335	Foetal	Proximal femur (hip)	12pcw
15335	Foetal	Distal femur (knee)	12pcw
138	OA	Hip	63
158	OA	Hip	67
165	OA	Hip	64
167	OA	Hip	71
168	OA	Hip	62
150	OA	Knee	75
154	OA	Knee	57
157	OA	Knee	58
164	OA	Knee	73
166	OA	Knee	62
263	OA	Knee	70
274	OA	Knee	65
275	OA	Knee	71
296	OA	Knee	69
297	OA	Knee	82
293	OA	Hip	68
295	OA	Hip	68
310	OA	Hip	69
312	OA	Hip	73
316	OA	Hip	57

Pcw - post conception weeks.

2.2. Extraction of Nucleic Acids from Human Samples

Following arthroplasty surgery, patient samples were stored at 4° C in Hank's Balanced Salt Solution (HBSS) (Gibco, Life Technologies) supplemented with Penicillin-Streptomycin (10000U/ml) (Gibco, Life Technologies) and Nystatin (10000U/ml) (Sigma Aldrich). The tissues (synovium and cartilage) were isolated from the samples and frozen at -80°C. Prior to nucleic acid extraction, the tissues were ground in liquid nitrogen using a Retsch MM400 tissue Grinder.

DNA and RNA from the synovium samples were extracted from 250mg of ground tissue using the Omega E.Z.N.A. DNA/RNA Isolation Kit (Omega Bio-Tek).

DNA from the arthroplasty cartilage was extracted from 1g of ground cartilage tissue using the Omega E.Z.N.A. DNA/RNA Isolation Kit, according to manufacturer protocol.

RNA from the cartilage tissue was extracted using TRIzol Reagent (Invitrogen) and RNeasy kit (Qiagen) according to the manufacturer's instructions.

The nucleic acids from the synovium and the cartilage tissues were quantified by UV-Vis spectrophotometry. DNA was stored at -20° C and the RNA at -80° C.

Nucleic acids from the developmental pooled limb tissues and cartilage were extracted by the HDBR using a QIAGEN QIAcube Automated DNA, RNA isolation machine and AllPrep DNA/RNA/miRNA Universal Kit (Qiagen), according to the manufacturer's instructions. The samples were quantified on a 2100 Bioanalyzer (Agilent). The DNA was stored at -20° C and the RNA at -80° C.

2.3. DNA Bisulfite Conversion

Prior to measuring DNAm levels at CpGs of interest, genomic DNA from human samples or from cultured cells was bisulphite converted using the EZ DNA methylation kit (Zymo Research) according to manufacturer's protocol. Briefly, DNA (200-500ng) was incubated overnight (16h) at 37° C in the presence of sodium bisulphite which converted unmethylated cytosine bases to uracil but left the methylated ones unchanged.

2.4. Complementary DNA (cDNA) Synthesis

For studies of gene expression or allelic imbalance, RNA (500ng-1µg) was reverse-transcribed. For the experiments described in Chapters 3 and 5, RNA was reverse transcribed using SuperScript II Reverse Transcriptase (Invitrogen). DNA was removed by incubating the RNA in the presence of 0.5µl of Turbo DNase at 1h at 37° C (Thermo Fisher Scientific). The Reverse Transcription PCR parameters were as follows: 10min at 25°C, 50min at 42°C and 10min at 70° C. Removal of RNA hybridised to the DNA was achieved through incubation of the samples with RNase H (New England Biolabs) for 20min at 37° C.

For the experiments described in all other Chapters, RNA was reverse transcribed using SuperScript IV Reverse Transcriptase (Invitrogen). The RNA was incubated with DNase I (supplied with SuperScript IV Reverse Transcriptase, Invitrogen) for 15min at room temperature. The PCR parameters used following the addition of SuperScript IV Reverse Transcriptase were: 10min at 23° C , 50min at 55° C and 10min at 80° C . Samples were treated with RNase H as described above.

To determine if the reverse-transcription was successful and if the resulting cDNA still contains traces of contaminating genomic DNA, the cDNA was amplified using Pyromark PCR kit (Qiagen) and forward and reverse primers targeting the *HBP1* gene (Appendix A) ordered from Integrated DNA Technologies (IDT). Briefly, 1µl of cDNA was amplified in a 25µl reaction consisting of 12.5µl Pyromark Master Mix (HotStart Taq polymerase, dNTPs, and 1.5mM MgCl₂), 2.5µl CoralLoad dye, 0.5µl of each of the forward and reverse primers (10mM) and DEPC-treated Ambion™ Nuclease-Free water (Invitrogen). The following PCR cycling conditions were used: 15min at 95° C, followed by 45 cycles of 30sec at 94° C, 30sec at 60° C, 30sec at 72° C, and final extension at 72° C for 10min. The primers were designed to bind in exon four and exon five, respectively, of *HBP1*. Following amplification, the samples were run on 2% agarose-tris-borate-ETDA (TBE) (Agarose – Fisher Scientific; TBE – Invitrogen) and the results were visualised using GelDoc-It Imaging System (UVP). If there was still genomic DNA present in the samples following the reverse-transcription, amplicon sizes of 152bp (cDNA) and of 750bp (DNA) were expected.

The switch from SuperScript II Reverse Transcriptase to SuperScript IV Reverse Transcriptase was necessary as the new protocol proved to work better for RNA of low integrity (such as that extracted from cartilage).

2.5. Genotyping

Samples were genotyped at rs11583641, rs12083450, rs7522001, rs10752924 and rs1046934 using PyroMark Q24 Sequencer (Qiagen). The genotyping assays were designed using PyroMark Software 2.4 (Qiagen) and the primers were ordered from IDT as oligonucleotides. The primer sequences are listed in Appendix A.

DNA was PCR amplified using Pyromark PCR kits (Qiagen) as described in Section 2.4. The amplicons were run on agarose-TBE gels, and the successful amplification of the target region was visualised.

The amplified product (10µl) was mixed with 70µl of Binding mix (consisting of 40µl PyroMark binding buffer (Qiagen), 1.5µl streptavidin-coated Sepharose beads (Sigma Aldrich) and 28.5µl DEPC-treated Ambion™ Nuclease-Free water) in the wells of a 0.2ml 24-well PCR plate (StarLab). The mixture was agitated gently for 10min on a 96-well plate shaker at room temperature.

Using a PyroMark Q24 Vacuum Workstation (Qiagen), the Sepharose beads and DNA bound to them were captured on the filter probes (Qiagen). The beads were washed in 70% ethanol for 5sec, 0.20M sodium hydroxide for 5sec and washed with 1X PyroMark Wash buffer (Qiagen) for 10sec. The beads were then released into a PyroMark Q24 24-well plate (Qiagen) containing 25µl Sequencing mix (consisting of 24.75µl PyroMark annealing buffer (Qiagen) and 0.75µl sequencing primer(10mM)). The samples were then annealed to the primer by incubation for 5min at 80° C. A PyroMark Q24 Cartridge (Qiagen) was loaded with reagents from PyroMark Q24 Advanced kit (Qiagen), including reaction substrate, reaction enzyme and dNTPs. The pyrosequencing was performed in a PyroMark Q24 Sequencer. The allele quantification was performed by PyroMark Q24 Software and visualised as pyrograms.

Genotyping at rs734657 was performed using Restriction Fragment Length Polymorphism (RFLP)-PCR analysis and agarose gel electrophoresis. Forward and reverse primers for the amplification of the region surrounding the SNP were generated using Primer3 Input online software (Appendix A) and ordered as oligonucleotides from IDT.

The DNA samples (10-50ng) were amplified using AmpliTaq Gold Polymerase II (Thermo Fisher Scientific) in a 20µl reaction consisting of 2µl 10X Buffer II, 0.8µl MgCl₂ (25mM), 0.5µl of forward and reverse primers (10mM), 2.5µl dNTPs (100µM), 0.5U of AmpliTaq Gold Polymerase, made up to a total volume of 20µl with DEPC-treated Ambion™

Nuclease-Free water. The following parameters of the PCR program were used: 14min at 94° C, followed by 40 cycles of 30sec at 95° C, 30sec at 63° C, 30sec at 72° C, and final extension step at 72° C for 10min. The NEBcutter online tool was used to identify a restriction enzyme that recognises one of the alleles of the SNP. The chosen enzyme, *MluCI*, which recognises the sequence 5'...VAATT...3', was ordered from NEB (UK).

The PCR product was then digested with *MluCI* according to manufacturer's instructions. The amplified region contained two recognition sequences, one acting as a control for the digestion reaction, and a second one that is created or destroyed by genotype at rs734657. Following amplification of the region containing rs734657 and digestion with *MluCI*, the three genotypes at rs734657 would produce bands of the following sizes: CC - 390bp and 65bp; CA - 390bp, 237bp, 153bp and 65bp; and AA - 237bp, 153bp and 65bp (Figure 2.1).

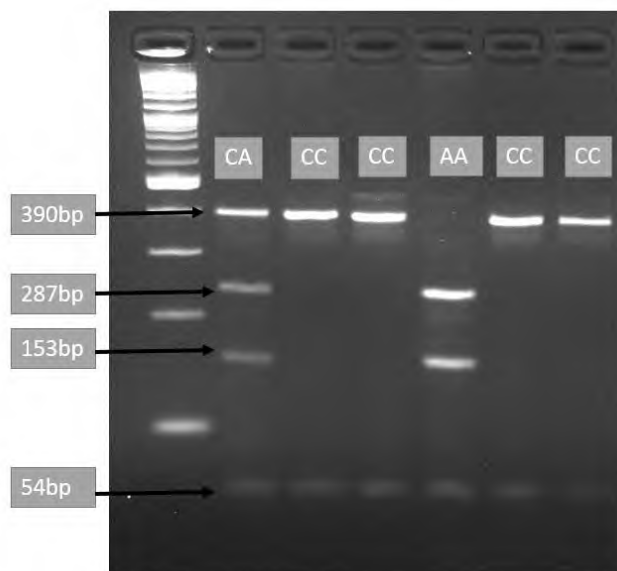


Figure 2.1. A representative image of the RFLP-PCR used to genotype rs734657. Left hand lane is a 1kb Plus DNA ladder (NEB). The size of each fragment is indicated in bp. For each sample the genotype is indicated.

2.6. Measuring DNAm Levels at CpG Sites

DNAm levels at the investigated CpG sites were measured using pyrosequencing. Assays capturing the CpGs of interest were designed using Pyromark Assay Design Software 2.0 and ordered from IDT. Primer sequences are listed in Appendix A.

Bisulphite converted DNA (bsDNA) was amplified using Pyromark PCR kit as described in Section 2.4. The PCR cycling conditions were the same as those used for genotyping except for a lower annealing temperature, 56° C, used for the bsDNA amplification. For the assay capturing CpG 3 and the assay capturing CpGs 4-6 the final concentration of MgCl₂ was optimised to 2mM to obtain a stronger PCR product.

Each sample was amplified and sequenced in duplicate. For quality control purposes, only replicate values within 5% of each other were used for downstream analyses. For the mQTL analyses, average DNAm levels of duplicate samples were taken and plotted against genotype at the SNP of interest.

2.7. Allelic Expression Imbalance (AEI) analysis

2.7.1. AEI at the *rs11583641* locus

Since *rs11583641* falls within a transcribed region of *COLGALT2*, the same assay as the one used for genotyping was used for the AEI analysis. Prior to the analysis, however, the assay had to be validated to ensure it provided accurate measurement of both alleles relative to each other. This was achieved by mixing DNA from patients homozygous at the SNP (CC and TT) to create samples with known ratios of the major and minor allele of the SNP. These samples were then amplified at the SNP and the alleles of the SNP were quantified by pyrosequencing (as described in Section 2.5). The expected values for each sample were compared to the actual measured values. The assay proved to be accurate in detecting the ratios between the two alleles and was, therefore, used for the AEI analysis.

For the AEI analysis, samples heterozygous at OA risk SNP *rs11583641* were used. For each patient, cDNA and its corresponding genomic DNA were amplified using PyroMark PCR Kits as described in Section 2.4 and the primers listed Appendix A. Each PCR reaction was repeated in triplicate for both the genomic DNA and the cDNA. The mean cDNA ratio for each individual patient was normalized to that measured in the genomic DNA of the same patient. The results were plotted as the ratio of the risk allele to the non-risk allele for each patient. A ratio different to 1:1 indicated differential expression of the risk-allele.

2.7.2. AEI at the rs1046934 locus

The transcript SNPs used for the analysis at this locus were rs114661926 for *COLGALT2* and rs2274432 for *TSEN15*. Assays for allele quantification at these SNPs were designed using Pyromark Assay Design Software 2.0 and ordered from IDT. The details of the assays are listed in Appendix A. The accuracy of the assays in detecting the ratios of the major and the minor allele was assessed using samples of mixed homozygous DNA at known ratios (as described in Section 2.7.1.). Both assays were successful at detecting differences in the ratios of the major and the minor allele.

For the AEI analyses at both *COLGALT2* and *TSEN15*, patients who were compound heterozygote at rs1046934 and the respective transcript SNP were used. The PCR reactions for both SNPs were optimised to obtain stronger PCR products by the addition of 5µl of Q-solution (supplied with PyroMark PCR kits). DNA and cDNA from these patients were amplified in triplicate using the optimised PCR reaction. The allelic ratios were measured, normalised and plotted as described in Section 2.7.1.

2.8. Gene Expression Quantification

Total gene expression levels were measured using reverse-transcription quantitative PCR (RT-qPCR). RNA was reverse transcribed as described in Section 2.4. cDNA was diluted 1:20 using DEPC-treated Ambion™ Nuclease-Free water. The diluted cDNA (2.5µl) was then used in 10µl total volume reactions with TaqMan® Fast Universal PCR Master Mix (Applied Biosystems) and PrimeTime qPCR Assays purchased from IDT (Table 2.5). The qPCR reaction was set up in MicroAmp® Fast Optical 96-well Reaction Plates (Applied Biosystems) and carried out using Quant Studio 3 (Applied Biosystems) with the following cycling conditions: 20sec at 95° C, followed by 40 cycles of 1sec at 95° C and 20sec at 60°C.

Gene expression was normalised to the house-keeping genes *18S*, *HPRT1* and *GAPDH* using the $2^{-\Delta Ct}$ method, $2^{-\Delta Ct(Gene\ of\ interest)} = 2^{-[Ct(Gene\ of\ interest) - Ct(mean\ of\ housekeeping\ genes)]}$, where Ct is the intersection point of the sample's reaction curve and the fluorescence threshold line. Each sample was analysed in triplicate.

Table 2.5. Details of the pre-designed RT-qPCR assays used to quantify gene expression

2.9.1. TC28a2 Human Chondrocyte Cell line

TC28a2 cells (ATCC) were cultured in Dulbecco's modified eagle's medium (DMEM): F12 Nutrient Mix (Ham) (1:1) (GIBCO, Life Technologies), containing 10% (v/v) heat-inactivated Foetal Bovine Serum (FBS) (Sigma-Aldrich), 1% (v/v) Penicillin/Streptomycin (10000U/ml). The cells were cultured at 37° C, 20% O₂, 5% CO₂ in a HERAccl 150i incubator (Thermo Scientific).

2.9.2. TC28a2-Cas9, TC28a2-dCas9/VP64, Tc28a2-dCas9/KRAB and TC28a2-dCas9/TET1 cell lines

TC28a2 cell lines stably expressing Cas9 protein, dCas9-VP64 transcriptional activation domain fusion protein, dCas9-KRAB transcriptional repression domain fusion protein, and dCas9-TET1 demethylating fusion protein had been engineered by Professor Loughlin's group prior to the start of my project. These cells were cultured under the same conditions as the TC28a2 cells with the addition of Puromycin (1µg/ml) to the culture medium.

2.9.3. SW982 Human Synovial Sarcoma Cell Line

SW982 cells (ATCC) were cultured under the same conditions as TC28a2 cells.

2.9.4. ASC52teloSox9 cells

Immortalised adipose derived MSC cells overexpressing SOX9, ASC52teloSox9¹, were grown in Mesenchymal Stem Cell Basal Medium (ATCC) supplemented with MSC supplement (2% FBS, 5ng/ml rh FGF basic, 5ng/ml rh FGF acidic, 5ng/ml rh EGF), L-Alanyl-L-Glutamine (2.4mM) and G41 disulphate salt solution (Sigma Aldrich) (0.2mg/ml) at 37°C, 20% O₂, 5% CO₂.

2.10. Lucia Reporter Gene Analysis

2.10.1. Cloning of the putative enhancer regions into the CpG-free-promoter-Lucia reporter gene plasmid

The region encompassing CpGs 3-10 (rs11583641 locus) contains two SNPs: rs943409 (G>A) and rs734657 (C>A), with three naturally occurring haplotypes in European populations: G/C, G/A and A/C. All three haplotypes were amplified from patient genomic DNA using primer sequences which were designed to include the required restriction enzyme sequences for downstream cloning (Table 2.9). The samples were amplified using AmpliTaq Gold Polymerase as described in Section 2.5. The PCR parameters used for the amplification are: 14min at 94° C, followed by 40 cycles of 30sec at 95° C, 30sec at 65° C, 30sec at 72° C, and final extension step at 72° C for 10min.

The regions containing cg15204595 and cg21606956 were amplified from patient genomic DNA using the Pyromark PCR kit as described in Section 2.4. The primer sequences, which included restriction enzyme sequences for downstream cloning, are listed in Appendix A.

The amplified regions were then cloned into a TOPO vector using a TOPO PCR cloning kit (Thermo Fisher Scientific) and transformed into One Shot™ TOP10 chemically competent *E. coli* cells (Invitrogen) using the heat-shock method. The bacteria were then plated onto LB-Ampicillin (100µg/ml) (Sigma Aldrich) agar (Sigma Aldrich) and X-Gal (Invitrogen) plates and grown overnight. White colonies (five from each plate) were picked and grown overnight in LB-Ampicillin (100µg/ml) broth. DNA was extracted using PureLink Quick Plasmid Miniprep kits (Invitrogen) and Sanger sequenced (Source BioScience, Cambridge). Plasmids were selected in which the regions of interest were successfully cloned (the region containing CpGs 3-10 with all three haplotypes for rs943409 and rs734657; the region containing cg15204595; the region containing cg21606956).

Following confirmation that the regions of interest were successfully cloned into the TOPO vector, the constructs (1µg) and the pCpG-free-promoter-Lucia vector (Invivogen) (1µg) were digested with appropriate restriction enzymes. For the constructs containing CpG 3-10 (with the three haplotypes of rs943409 and rs734657) and for the construct containing

cg21606956, *AvrII* and *SpeI* (NEB) enzymes were used; for the construct containing cg15204595 *BamHI* and *SpeI* (NEB) were used. The restriction digests were performed at 37° C for 3h. The products were resolved using gel electrophoresis on a 0.8% agarose-TBE gels to confirm successful linearization. The DNA of interest (the enhancer inserts from the TOPO plasmids and the CpG-free backbone of the pCpG-free-promoter-Lucia vector) were extracted using QIAquick Gel Extraction kit (Qiagen). The enhancers were then ligated into the CpG-free-promoter-Lucia vector using T4 DNA Ligase for 1h at room temperature (NEB).

5µl of the ligation reaction was transformed into chemically competent GT115 cells (Invivogen) and grown at 37° C overnight on LB-Zeocin (1000µg/ml) agar plates (Zeocin – InvivoGen). Colonies were isolated and cultured overnight in LB-Zeocin (1000µg/ml). DNA was extracted using PureLink Quick Plasmid Miniprep kits and Sanger sequenced (Source Bioscience, Cambridge) to confirm the integration of the putative enhancer regions into the CpG-free Lucia plasmids.

2.10.2. Methylation and mock-methylation of the CpG-free-promoter-Lucia plasmids containing the regions of interest

The plasmids containing the inserts of interest (4µg) were methylated *in vitro* using CpG Methyltransferase (*M.SssI*, NEB) by incubation at 37° C for 4h with *M.SssI* (4U/µl) in the presence of S-adenosylmethionine (SAM) (1600µM). The mock-methylated control plasmids were treated the same way but in the absence of *M. SssI*. The plasmid DNA was then purified by ethanol precipitation. To confirm successful methylation, both the methylated and the mock-methylated plasmids were digested with methylation-sensitive enzyme *HgaI* (NEB) (for the constructs containing CpGs 3-10 and all three haplotypes of rs943409 and rs734657), or *MspI* (NEB) (for the constructs containing cg21606956) and resolved by agarose gel electrophoresis. A methylation-sensitive restriction enzyme recognizing the sequence surrounding cg15204595 was not identified. The successful *in vitro* methylation of the reporter gene constructs containing the region surrounding cg15204595 was therefore confirmed by pyrosequencing.

2.10.3. Transfection of CpG-free-promoter-Lucia plasmids containing the regions of interest in TC28a2 cells and SW982 cells

The regulatory function of the region containing CpGs 3-10 (rs11583641 locus) was tested in TC28a2 cells and SW982 cells. The regulatory function of the regions containing cg15204595 and cg21606956 were only tested in TC28a2 cells.

The cells (TC28a2 or SW982) were seeded in 96-well plates at a density of 5000 cells/well 24h prior to the transfection. Each well was transfected with 100ng of the relevant CpG-free-promoter-Lucia construct, along with 10ng of the pGL3-promoter (Firefly luciferase reporter) using 2µl Lipofectamine 2000 (Invitrogen) per well according to manufacturer's protocol. Empty CpG-free-promoter-Lucia vectors were used as controls.

2.10.4. Lucia luminescence readings

24h post-transfection, the cells were lysed in 1X Lysis Buffer (Promega), and the levels of Lucia and Firefly Luciferase were measured using a GloMax multi-detection luminometer (Promega) with a Dual Luciferase Assay System (Promega). The Lucia readings for each methylated/mock-methylated condition were normalised to the corresponding Firefly readings to correct for differences in transfection efficiency between wells. The readings from the constructs containing the methylated or the mock-methylated DNA inserts of interest were then normalised to the readings obtained from the cells transfected with the empty CpG-free-promoter-Lucia plasmids (methylated and mock-methylated, respectively).

2.11. Activation and Repression of the Enhancer Containing CpGs 3-10 (rs11583641 locus)

2.11.1. Design of gRNAs targeting the region of interest

Six guide RNAs (gRNAs) targeting the region of the putative enhancer containing CpGs 3-10 (rs11583641 locus) were designed using IDT's CRISPR-Cas9 guide RNA design tool and ordered from the company. The sequences are listed in Appendix B.

2.11.2. Testing the targeting efficacy of the gRNAs

The efficiency of the gRNAs in targeting the enhancer was assessed in TC28a2 cells expressing Cas9. The cells were cultured as described in Section 2.9.2 and seeded in 6-well plates at a density of 100000 cells/well. The six gRNAs and trans-activating CRISPR RNA (tracrRNA, IDT) were diluted to 100µM with Duplex Buffer (IDT). Each gRNA was mixed with tracrRNA (1:1) in a 4µl reaction. The gRNA and the tracrRNA were annealed at 95° C for 5min and cooled to room temperature to form gRNA:tracrRNA duplexes. The duplexes (4µl) were transfected into TC28a2/Cas9 cells using 4µl DharmaFECT 1 transfection reagent (Horizon Discovery) per well. Once the cells reached 80% confluence, they were treated with 0.05% Trypsin-EDTA solution (GIBCO, Life Technologies) and incubated at 37° C for 5min. The cells were then collected and pelleted by centrifugation at 500xg for 5min and snap-frozen in dry ice.

DNA was extracted from the pellets using NucleoSpin TriPrep kit (Macherey-Nagel) following manufacturer instructions. The region targeted by the gRNAs was PCR amplified using AmpliTaq Gold Polymerase II in a 20µl reaction consisting of 2µl 10x BufferII, 1.6µl MgCl₂ (25mM), 0.1µl AmpliTaq Gold Polymerase II (0.5U), 0.5µl dNTPs (250µM) and 1µl of forward and reverse primers (10mM) (Appendix A) and DEPC-treated Ambion™ Nuclease-Free water. To estimate the editing efficiency, a T7 endonuclease I (T7EI) assay was used (Alt-R Genome editing Detection kit, IDT). The PCR amplicons were denatured and re-annealed forming heteroduplexes of wild-type DNA and CRISPR-mutated DNA. The samples were treated with T7EI which recognised and cleaved the mismatched heteroduplexes. The results were visualised by agarose gel electrophoresis.

2.11.3. Transfecting targeting gRNAs in TC28a2-dCas9/VP64 and TC28a2-dCas9/KRAB cells

Once the targeting efficiency of the gRNAs was confirmed, TC28a2-dCas9/VP64 and TC28a2-dCas9/KRAB were cultured as described in Section 2.9.2.

The gRNA:tracrRNA duplexes were formed and transfected into the TC28a2/dCas9-VP64 and TC28a2/dCas9-KRAB cells using DharmaFECT 1 as described in Section 2.11.2. The gRNAs were transfected alone or in combination (gRNAs 1, 3 and 5, and all gRNAs). Non-targeting gRNA was used as control (IDT).

72 hours post transfection, the cells were dissociated with Trypsin-EDTA and pelleted as described in Section 2.11.2 and RNA was extracted using NucleoSpin TriPrep kit following manufacturer's instructions. 1µg of RNA was reverse-transcribed using SuperScript II as described in Section 2.4 and the expression of *RGL1*, *COLGALT2* and *TSEN15* was measured by RT-qPCR as described in Section 2.5.

2.12. Deletion of the Enhancer Region Containing CpGs 3-10 (rs11583641 locus)

2.12.1. Design of gRNAs and cloning of the gRNAs into a PX462 plasmid containing the Cas9 gene

Two gRNAs (gRNA7 and gRNA8) located up- and downstream of the DMR containing cg18131582 were designed using IDT's CRISPR-Cas9 guide RNA design tool. The gRNAs were ordered from IDT as single-stranded complementary DNA oligonucleotides with overhangs for downstream cloning (Appendix B).

The PX462 plasmid containing the Cas9 gene (Addgene, 62987) was digested with *BbsI* (NEB) for 2h at 37° C followed by 20min at 65° C. The digested plasmid was then run on a 0.8% agarose-TBE gel. The linearized product was excised from the gel and extracted using QIAquick Gel Extraction kit.

The forward and reverse oligonucleotides of the two gRNAs were ligated to form a double-stranded product using T4 DNA Ligase by incubating for 30min at 37° C, followed by 5min at 95° C and gradually cooled to 25° C at a rate of -6° C /min.

The annealed gRNAs were then ligated overnight at 16° C into the linear PX462 plasmid using T4 DNA Ligase.

The plasmids were transformed into One Shot™ TOP10 Chemically Competent *E. coli* by heat shock. The bacteria were grown on LB-Ampicillin agar plates (100µg/ml). Colonies were picked and grown overnight in LB-Ampicillin (100µg/ml). The plasmid DNA was extracted and purified using PureLink Quick Plasmid Miniprep kits. Plasmids containing the correct gRNA sequences were selected after Sanger sequencing (Source Bioscience, Cambridge).

2.12.2. Transfection of PX462 plasmids containing the gRNA sequences into TC28a2 cells

TC28a2 cells were cultured as described in Section 2.9.1.

Once the cells were 80% confluent, they were dissociated with 0.05% Trypsin-EDTA and counted. One million cells were resuspended in 100µl SF nucleofection solution (Lonza) and 2.5µg of plasmid containing each of the guide sequences (5µg in total) were added to the cells. Empty PX462 plasmid was used as control. The reaction was transferred to 1ml cuvettes (Lonza). The plasmids were then transfected into the cells via electroporation using the 4D-NucleofectorX and the pulse code EH-100 (Lonza). The electroporated cells were incubated at room temperature for 5min. Warm 1X RPMI medium (GIBCO, Life Technologies) was added to the cuvettes, and they were incubated at 37° C for 15min. Following the incubation, the transfected cells were transferred to the wells of 6-well plates containing 2ml warm media. Cells were selected with puromycin (1µg/ml) 24h post transfection and expanded in puromycin-containing media until the wells of the plate reached 80% confluency (72h).

2.12.3. Confirmation of deletion

48h post transfection the cells were dissociated using Trypsin-EDTA and pelleted. DNA and RNA were extracted using the PureLink Genomic DNA Mini kit (Thermo Fisher Scientific) and NucleoSpin TriPrep kit, respectively. To confirm the successful deletion, the targeted region of the enhancer was PCR amplified from DNA from the control and the edited cells using AmpliTaq Gold Polymerase II in a 20µl reaction consisting of 2µl 10x Buffer 2.0, 1.6µl MgCl₂ (2mM), 0.1µl AmpliTaq Gold Polymerase II (0.5U), 0.5µl dNTPs (250µM) and 1µl of forward and reverse primers (10mM) (Appendix A). The PCR parameters used were: 14min at 94° C, followed by 40 cycles of 30sec at 95° C, 30sec at 65° C, 30sec at 72° C, and final extension step at 72° C for 10min. The results were visualised by agarose gel electrophoresis.

Once the deletion was confirmed, 1µg of RNA was reverse transcribed using SuperScript II (Thermo Fisher Scientific) as described in section 2.4. The gene expression of *RGL1*, *COLGALT2* and *TSEN15* was measured and normalised to the house-keeping genes as described in Section 2.8. The gene expression levels in the edited cells were then normalised to those measured in the control cells.

To confirm that the deletion had not affected *COLGALT2* splicing, the intron containing the enhancer (intron 10) was amplified using PyroMark PCR kit as described in Section 2.4 in cDNA from the control and the edited cells using primers binding in exon 11 and exon 10 of *COLGALT2* (main transcript COLGALT2-201, ENST00000361927.9). The primer sequences are listed in Appendix A. The results were visualised by gel electrophoresis.

2.13. Methylation and Demethylation of Regions of Interest in TC28a2 Cells

2.13.1. Demethylation of the enhancer containing CpGs 3-10 (rs11583461 locus) using TC28a2-dCas9/TET1 cells

The six gRNAs (gRNA1-6) used to target the region containing CpGs 3-10 were used to demethylate the CpG sites.

TC28a2-dCas9/TET1 cells were cultured as described in section 2.9.2. The cells were seeded in 6-well plates at a density of 100000 cells/well. 24h later, doxycycline (1µg/ml) was added to the cell culture media to induce the expression of the dCas9/TET1 protein complex.

gRNA:tracrRNA duplexes were formed as described in Section 2.11.2 and transfected into the TC28a2-dCas9/TET1 cells using DharmaFECT 1 as described in section 2.11.2. Similarly to the TC28a2-dCas9/VP64 and TC28a2-dCas9/KRAB experiments, gRNAs were transfected alone or in combination (gRNAs 1, 3 and 5, and all gRNAs) and non-targeting gRNA was used as control.

72h post transfection, the cells were trypsinised and pelleted. DNA was extracted using the PureLink Genomic DNA Mini kit and RNA was isolated using the NucleoSpin TriPrep kit.

DNA (200ng) was bisulphite converted as described in Section 2.3. DNA methylation at CpG 3-10 was measured by pyrosequencing as described in Section 2.6. RNA (500ng) was reverse-transcribed as described in Section 2.4. The expression of *RGL1*, *COLGALT2* and *TSEN15* was measured using RT-qPCR and normalised to the house-keeping genes as described in Section 2.8. The gene expression levels in the modulated cells were then normalised to that measured in the control cells.

2.13.2. Methylation of the enhancer containing CpGs 3-10 (*rs11583461* locus)

The same six gRNAs used for the demethylation of the enhancer and the non-targeting gRNA were ordered as single-stranded complementary DNA oligonucleotides with overhangs for subsequent cloning from IDT (Appendix B). The gRNA oligonucleotides were annealed to form double-stranded DNA fragments using T4 DNA Ligase as described in Section 2.12.1.

A pdCas9-DNMT3a-EGFP plasmid (Addgene, 71666) (1µg) was digested with *BbsI* enzyme for 1h at 37° C followed by 20min at 65° C. The linearized plasmid was run on 0.8% agarose-TBE gel and extracted using QIAquick Gel Extraction kit. The double-stranded gRNAs were ligated into the linearized pdCas9-DNMT3a-EGFP plasmid using the T4 DNA Ligase by overnight incubation at 16° C. The plasmids were transformed by heat shock into chemically competent OneShot™ TOP10 *E. coli* which were then cultured overnight in LB-Ampicillin (100µg/ml). Plasmid DNA from the bacteria was isolated and Sanger sequenced.

Once the successful integration of the six gRNA sequences and the non-targeting gRNA sequence into the pdCas9-DNMT3a-EGFP was confirmed, the gRNAs (alone or in combination, 5µg in total) were nucleofected into TC28a2 cells as described in section 2.12.2. The successful transfection of the plasmids into the cells was confirmed after 24h by GFP visualisation (Zeiss AxioVision).

Once the 6-well plates were confluent (72h post transfection) the cells were trypsinised and pelleted and DNA and RNA were extracted as described in section 2.12.3. DNA (300ng) was bisulphite converted (Section 2.3). DNAm levels at CpGs 3-10 was measured by pyrosequencing (Section 2.6). cDNA was synthesised from RNA (500ng) (Section 2.4). The expression of *RGL1*, *COLGALT2* and *TSEN15* was measured using RT-qPCR and normalised to the house-keeping genes as described in Section 2.8. The gene expression levels in the modulated cells were then normalised to those measured in the control cells.

2.13.4. Modulation of the enhancers containing *cg15204595* and *cg21606956* (*rs1046934* locus) using *dCas9/TET1* and *dCas9/dTET1*, and *dCas9/DNMT3a* and *dCas9/dDNMT3a*

TC28a2 cells were transfected with a plasmid containing dCas9 coupled with active TET1 or active DNMT3a (dCas9/TET1 or dCas9/DNMT3a) protein and a targeting gRNA or a

control plasmid containing dCas9 coupled with deactivated TET1 or deactivated DNMT3a (dCas9/dTET1 or dCas9/dDNMT3a) protein and the same targeting gRNA.

For the methylation of the two CpGs, the gRNA oligonucleotides (non-targeting, targeting cg15204595 or cg21606956) were annealed and ligated into the pdCas9-DNMT3a-EGFP plasmid (dCas9/DNMT3a) and the catalytically inactivated control plasmid pdCas9-dDNMT3a-EGFP (ANV) (dCas9/dDNMT3a) (Addgene, 71685) as described in section 2.13.2.

For demethylation, the pdCas9-DNMT3a-EGFP plasmids containing the two gRNAs targeting cg15204595 and cg21606956 and the non-targeting gRNA were digested with *PvuII* and *XbaI* (NEB) for 1h at 37° C followed by 20min at 65° C. This digestion separated the gRNA scaffold region of the plasmids, along with the U6 promoter, and part of the Ampicillin resistance gene from the backbone of the plasmid. The pSpdCas9-huTET1CD-T2A-mCherry plasmid (dCas9/TET1) (Addgene, 129027) and the catalytically inactivated control plasmid pSpdCas9-huTET1CD-T2A-mCherry (dCas9/dTET1) (Addgene, 129028) were also digested using the same restriction enzymes. The products of the digestion reactions were run on a 0.8% agarose-TBE gel. The backbones of the dCas9/TET1 and the dCas9/dTET1 plasmids, and the gRNA scaffolds of the pdCas9-DNMT3a-EGFP (containing either the gRNAs targeting cg15204595, cg21606956 or the non-targeting gRNA) were excised from the gel and the DNA was extracted using QIAquick Gel Extraction kit.

The scaffold regions containing the gRNAs targeting cg15204595 and cg21606956, and the non-targeting gRNA were ligated into the backbones of the dCas9/TET1 and the dCas9/dTET1 plasmids using T4 DNA Ligase by overnight incubation at 16° C. The plasmids were transformed into chemically competent OneShot™ TOP10 *E. coli* via heat shock. The bacteria were grown on LB-Ampicillin agar plates (100µg/ml), colonies were picked and grown overnight in LB-Ampicillin (100µg/ml). The plasmid DNA was extracted and purified using PureLink Quick Plasmid Miniprep kits and Sanger sequenced. Plasmids that successfully integrated the right gRNA sequences were used.

TC28a2 cells were cultured as described in Section 2.9.1. 5µg of dCas9/DNMT3a, dCas9/dDNMT3a, dCas9/TET1 or dCas9/dTET1 containing a non-targeting gRNA sequence or a sequence targeting either cg15204595 or cg21606956 were nucleofected into 1000000 TC28a2 cells. The efficiency of the transfection was determined after 24h by GFP (for dCas9/DNMT3a and dCas9/dDNMT3a plasmids) or mCherry (for dCas9/TET1 or dCas9/dTET1 plasmids) visualisation using a Zeiss AxioVision fluorescent microscope.

72h post transfection the cells were dissociated, pelleted and snap frozen. Nucleic acids were extracted as described in previous sections and the effects of the epigenetic modulation on DNAm levels and gene expression levels were measured as described in Sections 2.13.3 and 2.13.4.

2.13.5. Methylation and demethylation of the enhancer containing CpGs 3-10 (rs11583641 locus) in SW982 cells

The six gRNAs (gRNA1-6) used to target the enhancer containing CpG 3-10 were cloned into the pdCas9-DNMT3a-EGFP, pdCas9-dDNMT3a-EGFP (ANV), pSpdCas9-huTET1CD-T2A-mCherry and pSpdCas9-huTET1CD-T2A-mCherry plasmids as described in section 2.13.5. Plasmids confirmed to contain the correct gRNA sequences were used for the experiment.

To determine the optimal nucleofection conditions for the SW982 cell line, the cells were dissociated and counted. 1500000 cells were resuspended into 300µl of SF nucleofection solution or SE nucleofection solution (Lonza). 15µg of pmaxGFP™ Vector (1µg/µl in 10mM Tris pH 8.0) (Lonza) were added to the resuspended cells. 20µl of the cells (either in SE or SF solution) were pipetted into the wells of a 16-well Nucleocuvette™ Strip (Lonza). The Nucleocuvette™ Strip was placed in a 4D-NucleofectorX machine and the following pulse codes were applied to the wells containing the cells: CA-137, DS-150, CM-138, DS-120, CM-137, EH-100, CM-150, EO-100, DN-100, EN-138, DS-138, EN-150, DS-137, EW-113, DS-130. The optimal conditions were chosen following GFP visualisation 24h after the nucleofection. For the SW982 cells Lonza's SE nucleofection solution and the pulse code DN-100 resulted in an optimal balance between efficiency and cell viability.

Once the nucleofection conditions were optimised, 5µg of the gRNA containing plasmids (alone or in combination) were nucleofected in SW982 cells. The same protocol described in section 2.12.2 was followed with the optimised nucleofection parameters described above. The successful transfection of the plasmids into the cells was confirmed by fluorescent microscopy 24h after the nucleofection (Zeis AxioVert).

Once the cells reached 80% confluency (72h post nucleofection), they were pelleted and DNA and RNA were extracted using PureLink Genomic DNA Mini kit and NucleoSpin TriPrep kit, respectively. The effects of the epigenetic modulation on DNAm levels and gene

expression levels in the SW982 cell line were measured as described in Sections 2.13.3 and 2.13.4.

2.14. Assay for Transposase-Accessible Chromatin with Sequencing (ATAC-seq)

2.14.1. Digestion of developmental cartilage samples

The details of the developmental samples used for the ATAC-seq experiment are listed in Table 2.4. The cartilage was digested in 1% type I Collagenase solution (Sigma Aldrich) for 3h at 37° C on a rotating platform. The collagenase solution was removed, cells were washed with 1X PBS (GIBCO), re-suspended in 5% FBS-DMEM and counted.

2.14.2. Digestion of arthroplasty cartilage

For five of the OA knees and five of the OA hips used for the ATAC-seq experiment (Table 2.4, red OA samples) the cartilage was removed from the joint using a scalpel and 1g of it was digested in 4ml of Hyaluronidase (1mg/ml) (Sigma Aldrich) for 15min at 37° C, followed by a second digestion with 4ml of Trypsin (2.5mg/ml) (Sigma Aldrich) for 30min at 37° C. The samples were left to digest overnight in 3ml of 0.2% type I Collagenase at 35° C. The collagenase solution was then removed, cells were washed with 1X PBS, re-suspended in 10% FBS-DMEM and counted.

The other five OA knee and five OA hip samples (Table 2.4, black OA samples) were digested using 1% type I Collagenase solution (Sigma Aldrich) for 12h at 37° C. Following the digestion, the cells were washed with 1X PBS, re-suspended in 5% FBS-DMEM and counted.

2.14.3. Cell Lysis

Following isolation of the chondrocytes from either developmental or OA cartilage, the cell nuclei from 50 000 cells were isolated by 3min incubation on ice in lysis buffer consisting of 5M NaCl (Sigma Aldrich), 1M MgCl₂ (Sigma Aldrich), 1M Tris-HCl (pH 7.5), 10% (w/v) NP-40 (Roche), 10% (w/v) Tween 20 (Roche) and 1% (w/v) Digitonin (Promega). The successful

isolation of the nuclei was confirmed visually by observing the nuclei under a microscope at 20X magnification.

2.14.4. Tagmentation and DNA purification

The isolated nuclei were re-suspended in a transposase mix (25µl 2xTD Buffer, 2.5µl TDE1 enzyme (NexteraTn5 transposase, Illumina), 0.5µl 1% (w/v) Digitonin, 0.5µl 10% (w/v) Tween 20, 16.5µl 1x PBS and 5µl DEPC-treated Ambion™ Nuclease-Free water) and incubated on a thermomixer for 30min at 37° C at 1000rpm. DNA was purified using MinElute PCR Purification kit (Qiagen) according to manufacturer's instructions and eluted in 10µl of Elution Buffer.

2.14.5. Library amplification

The purified transposed DNA was partially PCR amplified in a 50µl reaction using 25µl NEB Next High-Fidelity PCR Master Mix (NEB) and 2.5µl of primer Ad1_noMX (25µM) (AATGATACGGCGACCAACGAGATCTACACTCGTCGGCAGCGTCAGATGTG) in combination with 2.5µl of one of 22 barcoded primers (25µM) (Table 2.6). The following PCR cycling conditions were used: 5min at 72° C, 30sec at 98° C, followed by five cycles of 10sec at 98° C, 30sec at 63° C and 60sec at 72° C.

An aliquot of the reaction was amplified by qPCR. The PCR reaction consisted of 5µl of partially amplified tagmented DNA, 0.5µl Ad1_noMX (25µM), 0.5µl of one of 22 barcoded primers (25µM) (Table 2.10), 25µl NEB Next High-Fidelity PCR Master Mix, 0.15µl SYBR Green I (100X) and DEPC-treated Ambion™ Nuclease-Free water to adjust to a final volume of 15µl.

Table 2.6. Details of indexing primers used for ATAC-seq. Red OA samples were digested with Type I collagenase overnight and the black OA samples were digested for 12h.

Patient ID	Sample origin	Indexing Primer
15192	Foetal	Ad2.1_TAAGGCGA CAAGCAGAAGACGGCATAACGAGATTCGCCCTAGTCTCGTGGGCTCGGAGATGT
15192	Foetal	Ad2.2_CGTACTAG CAAGCAGAAGACGGCATAACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
15209	Foetal	Ad2.3_AGGCAGAA CAAGCAGAAGACGGCATAACGAGATTTCTGCCCTGTCTCGTGGGCTCGGAGATGT
15209	Foetal	Ad2.4_TCTGAGC CAAGCAGAAGACGGCATAACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
15210	Foetal	Ad2.5_GGACTCCT CAAGCAGAAGACGGCATAACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
15210	Foetal	Ad2.6_TAGGCATG CAAGCAGAAGACGGCATAACGAGATCATGCCCTAGTCTCGTGGGCTCGGAGATGT
15271	Foetal	Ad2.7_CTCTCTAC CAAGCAGAAGACGGCATAACGAGATGTAGAGAGTCTCGTGGGCTCGGAGATGT
15271	Foetal	Ad2.8_CAGAGAGG CAAGCAGAAGACGGCATAACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
15285	Foetal	Ad2.9_GCTACGCT CAAGCAGAAGACGGCATAACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
15285	Foetal	Ad2.10_CGAGGCTG CAAGCAGAAGACGGCATAACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
15335	Foetal	Ad2.11_AAGAGGCA CAAGCAGAAGACGGCATAACGAGATTGCCCTTGTCTCGTGGGCTCGGAGATGT
15335	Foetal	Ad2.12_GTAGAGGA CAAGCAGAAGACGGCATAACGAGATTCCCTACGTCTCGTGGGCTCGGAGATGT
138 and 264	OA	Ad2.13_GTCGTGAT CAAGCAGAAGACGGCATAACGAGATATCACGAGTCTCGTGGGCTCGGAGATGT
150 and 274	OA	Ad2.14_ACCACTGT CAAGCAGAAGACGGCATAACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
154 and 275	OA	Ad2.15_TGGATCTG CAAGCAGAAGACGGCATAACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
157 and 296	OA	Ad2.16_CCGTTTGT CAAGCAGAAGACGGCATAACGAGATACAAACGGTCTCGTGGGCTCGGAGATGT
158 and 297	OA	Ad2.17_TGCTGGGT CAAGCAGAAGACGGCATAACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
164 and 293	OA	Ad2.18_GAGGGGTT CAAGCAGAAGACGGCATAACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
165 and 295	OA	Ad2.19_AGGTTGGG CAAGCAGAAGACGGCATAACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
166 and 310	OA	Ad2.20_GTGTGGTG CAAGCAGAAGACGGCATAACGAGATACCCACACGTCTCGTGGGCTCGGAGATGT
167 and 312	OA	Ad2.21_TGGGTTTC CAAGCAGAAGACGGCATAACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT
168 and 316	OA	Ad2.22_TGGTCACA CAAGCAGAAGACGGCATAACGAGATTGTGACCAAGTCTCGTGGGCTCGGAGATGT

The qPCR reaction was set up in MicroAmp® Fast Optical 96-well Reaction Plates and carried out using Quant Studio 3 with the following cycling conditions: 30sec at 98° C, followed by 20 cycles of 10sec at 98° C, 30sec at 63° C and 60sec at 72° C.

The number of additional PCR cycles needed for each sample was then calculated by determining the number of cycles needed to reach 1/3 of the maximum Rn, which is defined as the fluorescence emission intensity of the reporter dye normalised to the fluorescence emission intensity of the passive reference dye. For each sample, the Rn was determined based on the amplification plots generated by the qPCR analysis tool of ThermoCloud (Thermo Fisher Scientific).

The remaining 45µl of the partially amplified DNA libraries were amplified by PCR for an additional number of cycles, as calculated for each individual sample based on the qPCR readings, to avoid amplification to saturation.

2.14.6. Library purification

The amplified DNA libraries were purified using AMPure XP magnetic beads (Beckman Coulter). For each sample, the amplified PCR product (45µl) was mixed with 22.5µl (0.5X) of AMPure XP magnetic beads. The samples were incubated for 10min at room temperature and the beads were separated on magnetic racks for 5min. The supernatant was transferred to

clean 1.5ml microcentrifuge tubes and mixed thoroughly with 58.5µl (0.85X) of AMPure XP magnetic beads. Samples were incubated for 10min at room temperature and the beads were separated with a magnet for 5min. The supernatant was discarded, and the beads were washed twice with 200µl freshly prepared 80% ethanol. The ethanol was discarded, and the beads were left to air dry for 10min. The beads were resuspended in 20µl DEPC-treated Ambion™ Nuclease-Free water and incubated at room temperature for 5min. The beads were separated on magnetic racks for 5min and the supernatant containing the purified DNA was transferred to clean 1.5ml microcentrifuge tubes and stored at -20° C.

2.14.7. Sequencing and data analysis

The DNA libraries were sequenced by Newcastle University's Genomics Core Facility on a NextSeq S1 generating paired-end 100 bp reads.

The analysis of the data was performed by Mr John Casement (Experimental Scientific Officer) from Newcastle University's Bioinformatics Support Unit.

2.15. Differentiated ASC52teloSOX9 Cells as a New Chondrocyte Cell Model

2.15.1. Chondrogenesis protocol

ASC52teloSox9 cells were cultured as described in section 2.9.4.

For the chondrogenic differentiation, the cells were dissociated and counted. Cells were resuspended in differentiation medium consisting of DMEM-HG (4.5g/L Glucose) (GIBCO, Life Technologies) supplemented with TGFβ3 (10ng/ml) (Sigma Aldrich), Dexamethasone (100nM) (Sigma Aldrich), Ascorbic acid-2-phosphate (50µg/ml) (Sigma Aldrich), Proline (40µg/ml) (Sigma Aldrich), 1X ITS+L premix (Sigma Aldrich) and 1% Penicillin/Streptomycin and seeded in the wells of a 96-well round bottom plate at a density of 400000 cell/well. The plate was spun for 5min at 500xg to pellet the cells. The pellets were left at 37° C, 20% O₂, 5% CO₂. The remaining cells were pelleted (400000 cells/pellet) in 1.5-ml tubes and stored at -80° C to serve as 'Day 0' base line control.

The differentiation media on the pellets in the 96-well plate was changed every 3-4 days. Pellets were collected at Day 3, 7, 14 and 21 (five pellets/time point) and stored at -80° C.

Nucleic acids were extracted from the pellets using RNA/DNA Purification Micro Kit (Norgen Biotek Corp.) following manufacturer's instructions. RNA (500ng) was reverse transcribed using SuperScript IV Reverse Transcriptase as described in section 2.4. Expression levels of genes relevant to cartilage biology (*ACAN*, *COL2A1*, *COL1A1*, *COL10A1*, *SOX9*, *RUNX2*) were measured by RT-qPCR as described in Section 2.8.

The chondrogenic differentiation experiment was repeated by Mr Jack Roberts (PhD student with Professor Loughlin) who tested if different number of cells per pellet (600000, 800000 and 1000000 cells/pellet) had an impact on the differentiation of the cells. By comparing expression level of the above genes, he determined that an increased number of cells per pellet did not have an impact on the chondrogenic potential of the cells. For future experiments I used 1000000 cells/pellet.

2.15.2. Optimisation of nucleofection conditions for the ASC52teloSox9

To determine the optimal transfection conditions for the ASC52teloSox9 cells, the cells were grown in monolayer as described in section 2.9.4. The cells were dissociated and counted. 1500000 cells were resuspended into 300µl of P1 4D-Nucleofector™ X Solution (Supplied with P1 4D-Nucleofector™ X kit, Lonza). 15µg of pmaxGFP™ Vector (1µg/µl in 10 mM Tris pH 8.0) (Supplied with P1 4D-Nucleofector™ X kit, Lonza) were added to the resuspended cells. 20µl of the cell solution were pipetted into the wells of a 16-well Nucleocuvette™ Strip (Supplied with P1 4D-Nucleofector™ X kit, Lonza). The Nucleocuvette™ Strip was placed in a 4D-NucleofectorX machine and the following pulse codes were applied to the wells containing the cells: CA-137, DS-150, CM-138, DS-120, CM-137, EH-100, CM-150, EO-100, DN-100, EN-138, DS-138, EN-150, DS-137, EW-113, DS-130.

24h post nucleofection the GFP was visualised using a Zeis AxioVert fluorescent microscope. The pulse CM-138 was chosen as the cells nucleofected using that code showed an optimal balance between transfection efficiency and cell viability.

2.16. Knockout (KO) of *COLGALT1* and *COLGALT2* to Deplete their Respective Proteins

2.16.1. gRNAs for CRISPR/Cas9 KO of *COLGALT1* and *COLGALT2*

A gRNA sequence targeting exon two of *COLGALT1* (gRNA-T1) and a gRNA targeting exon three of *COLGALT2* (gRNA-T2) were designed using IDT's gRNA design software. The gRNA sequences were ordered as complementary single-stranded oligonucleotide with restriction enzyme sequences added for downstream cloning (Appendix B).

The oligonucleotides were annealed and cloned into PX462 plasmid containing the Cas9 gene as described in Section 2.12.1. Following confirmation of the successful integration of the gRNA sequences into the plasmids, the plasmids were nucleofected into ASC52teloSOX9 cells as described previously (Section 2.15.2.). A non-targeting gRNA was used as a control. 24h after the nucleofection the cell media was changed to cell media containing puromycin (1mg/ml). Once the cells were confluent (96h post-nucleofection), they were dissociated with 0.05% Trypsin-EDTA. Half of the cells were frozen in liquid nitrogen for future use and the other half was pelleted. DNA, RNA and protein were extracted from the cell pellets using NucleoSpin TriPrep kit following manufacturer instructions. The region targeted by the gRNAs was PCR amplified using Pyromark PCR kits. The primers used are listed in Appendix A.

To estimate the editing efficiency, a T7 endonuclease I (T7EI) assay was used as described in section 2.11.2. Both gRNA-T1 and gRNA-T2 were confirmed to cut at their target regions.

2.16.2. Confirmation of the depletion of the *COLGALT1* and *COLGALT2* proteins

The protein extracted from the ASC52teloSOX9 cells was quantified in a microplate using the Protein Quantification Assay (Machery-Nagel), according to the manufacturer's protocol. Absorbances were read at 595nm using Varioscan LUX Plate reader (ThermoFisher). Protein concentrations were determined based on a calibration curve created via serial dilution of BSA stock protein (supplied with the Protein Quantification Assay, Macherey-Nagel).

10µg of the protein samples from the ASC52teloSOX9 cells were diluted to 30µl with Protein Solving Buffer - Tris - (2-carboxyethyl) phosphine hydrochloride reducing agent (PSB-

TCEP) (Supplied with NucleoSpin TriPrep kit, Macherey-Nagel). The samples were reduced for 5min at 95°C.

The protein samples from the control cells, the *COLGALT1* KO cells and the *COLGALT2* KO cells were resolved on a NUPAGE 10-12% Bis-Tris gel (Thermo Fisher Scientific). The resolved proteins were then transferred to a Immobilon-P polyvinylidene fluoride (PVDF) membrane (Merck Millipore) by electroblotting for 150min at 80mA using a V20-SDB semi-dry blotter (Scie-Plas). The PVDF membrane was cut in half (one half containing proteins from the control cells and the *COLGALT1* KO cells and the other half containing the proteins from the control cells and *COLGALT2* KO cells) and the two membranes were blocked for non-specific protein binding by incubation at room temperature for 1h in 1X TBS containing 5% (v/v) powdered skimmed milk (Tesco) and 0.02% (v/v) Tween-20 (Sigma Aldrich). The membrane was washed on a rocking platform (Stuart Scientific) for 15 min (three washes of 5min) with 1X TBS containing 0.02% (v/v) Tween-20.

The PVDF membrane containing proteins from the control cells and the *COLGALT1* KO cells was incubated overnight at 4°C with GLT25D1 (alternate name for the protein encoded by *COLGALT1*) rabbit polyclonal antibody (Proteintech, 16768-1-AP) diluted 1:1000 in 1X TBS containing 5% (w/v) powdered milk and 0.02% (v/v) Tween-20. The PVDF membrane containing proteins from the control cells and the *COLGALT2* KO cells was incubated overnight at 4° C with GLT25D2 (alternate name for the protein encoded by *COLGALT2*) rabbit polyclonal antibody (Proteintech, 25993-1-AP) diluted 1:500 in 1X TBS containing 5% (w/v) powdered milk and 0.02% (v/v) Tween-20.

The membranes were then washed on as above for 15min (three washes of 5min) with 1X TBS containing 0.02% (v/v) Tween-20 and incubated with the polyclonal horse anti-rabbit horseradish peroxidase (HRP) conjugated secondary antibody (DakoCytomation) (diluted 1:2000 in 1X TBS containing 5% (w/v) powdered milk and 0.02% (v/v) Tween-20) for 1.5h at room temperature. As a loading control, the blots were incubated for 1.5h at room temperature with mouse monoclonal GAPDH primary antibody (Chemicon) (diluted 1:10000 in 1X TBS containing 5% (w/v) powdered milk and 0.02% (v/v) Tween-20) with subsequent incubation in goat anti-mouse horseradish peroxidase (HRP) conjugated secondary antibody (DakoCytomation) (diluted 1:2000 in 1X TBS containing 5% (w/v) powdered milk and 0.02% (v/v) Tween-20) for 1.5h at room temperature. Detection was achieved with Immobilon

Western Chemiluminescent HRP Substrate (Millipore) and the Azure c600 Gel imaging system (Thermo Fisher Scientific).

RNA extracted from the cells was reverse-transcribed and the expression levels of *COLGALT1* and *COLGALT2* was measured by RT-qPCR as described in section 2.8.

2.16.3. Generation of ASC52teloSOX9 cells with COLGALT1 and COLGALT2 double KO

Once the successful depletion of the *COLGALT1* and the *COLGALT2* proteins achieved using single CRISPR gRNAs was confirmed, a disruption of both proteins simultaneously in ASC52teloSOX9 cells was attempted. To achieve this, the frozen cells in which *COLGALT1* was knocked out were grown. Once the cells were confluent, they were nucleofected as previously described with the PX462 plasmid containing the sequence of gRNA-T2 (targeting *COLGALT2*). Once the cells were confluent (96h post-transfection) they were dissociated. Half of the cells were frozen, as before, for future use and the other half was pelleted. Nucleic acids and protein were extracted using NucleoSpin TriPrep kit. Western blot for *COLGALT1* and *COLGALT2* was performed as described in section 2.17.2 to confirm the depletion of both proteins. Reduced expression of the *COLGALT1* and the *COLGALT2* genes was also confirmed at the mRNA level using RT-qPCR as described in section 2.8.

2.16.5. Chondrogenic differentiation of the ASC52teloSox9 – COLGALT1/ COLGALT2/ COLGALT1+COLGALT2 KO cells

Once the successful depletion of the *COLGALT1* and the *COLGALT2* proteins was confirmed in ASC52teloSox9 cells, the cell lines with the depleted proteins and the control cells were expanded. After 11 days of culture in ASC52teloSox9 culture media (as described in section 2.9.4) the cells were dissociated and counted. The control cells, *COLGALT1* KO cells, *COLGALT2* KO cells and *COLGALT1+COLGALT2* KO cells were resuspended in chondrogenic differentiation medium and pelleted in the wells of 96 deep-well plates at a concentration of 1000000 cells/ well in 1ml of differentiation medium (Section 2.16.1). The control cells were pelleted into 20 pellets (five pellets per each time point – Day 0, Day 7, Day 14 and Day 21). The *COLGALT1* KO, *COLGALT2* KO and *COLGALT1+COLGALT2* KO cells were each pelleted into 10 pellets (five pellets per time point – Day 0 and Day 21). The Day 0 pellets for each condition

were snap frozen in dry ice and stored at -80° C. The chondrogenic media was changed every 3-4 days and pellets were collected at the appropriate time points. On the day of the collection of the last pellets (Day 21), the chondrogenic medium of the cells was removed and 1ml of Reduced Serum Medium Opti-MEM (1x) (GIBCO, Life Technologies) was added to the pellets. The cells were kept in this medium for eight hours after which the cells were snap frozen and stored at -80° C. The media was collected, and the samples were centrifuged at 2000xg for 10min at 4°C. The supernatant was then transferred to 1.5ml LoBind tubes (Eppendorf) and stored at -80° C.

2.16.6. Protein extraction from chondrocyte pellets

The chondrocyte pellets were broken down to release the cells with the aid of a hand-operated grinder (KIMBLE, DWK Life Sciences) and pellet pestles (Bel-Art, Sigma Aldrich). The cells were then resuspended in Protein Solving Buffer (supplied with NucleoSpin TriPrep kit, Macherey-Nagel) and centrifuged for 5min at 2000xg. The supernatant was discarded, and the cells were lysed in 50µl of lysis buffer consisting of in 5% SDS, 50mM Tris (pH8), 1mM TCEP. The cells were spun at 140000xg for 10min. The protein contained in the supernatant was stored at -80° C.

2.16.7. Protein extraction from media samples

The media samples were transferred to 5ml Protein LoBind tubes (Eppendorf). Ice cold methanol (960µl) was added to the media samples and vortexed briefly. Ice cold chloroform (Sigma Aldrich) was added (160µl) and mixed thoroughly. Ice cold water (1.5ml) was then added to each tube. The samples were mixed by vortexing and then centrifuged at 4,000xg for 30min at 4° C. The samples separated into layers. The top one was carefully discarded and the bottom layer containing the protein was kept. Ice cold methanol (500µl) was added and mixed thoroughly. The samples were transferred to 1.5ml Protein LoBind tubes and centrifuged at 20,000xg for 30 min at 4° C. The supernatant was discarded, and the pellets were air dried. The pellets were then resuspended in 50µl of buffer consisting of 5% SDS, 50mM Tris (pH8), 1mM TCEP.

2.16.8. Protein Library preparation

For the proteomic sample preparation 20µg of protein from the cell pellets and 9.5µg of protein from the media samples were purified by suspension trapping (S-Trap) (ProtiFi). Samples were reduced with 5mM Tris (2-carboxyethyl) phosphine (Pierce) for 30min at 37° C and then alkylated with 5mM iodoacetamide for 30min at 37° C . The samples were then acidified with 2.5µl of 12% phosphoric acid. Subsequently, 165µl of S-Trap binding buffer (90% methanol in 100 mM triethylammonium bicarbonate buffer (pH 7.1)) was added to the samples. The acidified samples were loaded on S-trap microspin columns and centrifuged at 4000xg for 1min. Each column was washed with 150µl of S-trap binding buffer and centrifuged again at 4000xg for 1min. The wash step was repeated three times. 25µl of 50mM TEAB (pH 8.0) containing sequencing-grade trypsin (1:10 ratio of trypsin:protein) was added to the samples, followed by proteolysis for 2h at 47° C. Peptides were eluted with 50mM TEAB (pH 8.0) and centrifuged at 1000xg for 1min. Elution steps were repeated using 0.2% formic acid (FA) and 0.2% FA in 50% acetonitrile, respectively. The combined eluted peptides from each sample were dried and stored at -80° C.

2.16.9. Protein peptide injection and liquid chromatography

The protein libraries were dried and dissolved in 2% acetonitrile, 0.1% trifluoroacetic acid and injected with an UltiMate 3000 RSLCnano System (Thermo Scientific). Samples were first separated on an Acclaim PepMap 100 C18 LC trap column (inner diameter: 100µm × 20 mm, 3µm, 100 Å) and then on an EASY-Spray nanoLC-C18 column (inner diameter: 75µm × 500 mm, 2µm, 100 Å) at a flow rate of 250nL/min. For Buffer A water containing 0.1% FA was used and for buffer B - 80% acetonitrile containing 0.1% FA. The following gradient was used for the analysis: Buffer A and buffer B at 3% for 5 min, followed by an increase from 3% to 35% for Buffer B over 90 min, 35% to 90% for Buffer B over 0.5 min, maintained at 90% for Buffer B for 5 min, followed by a decrease to 3% over 0.5 min and equilibration at 3% for 10 min.

2.16.10. Data independent acquisition LC-MS/MS analysis

The protein samples were run on a Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Scientific) by Ms Francis Sidgwick, a PhD student in Professor Matthias Trost's proteomics group (Newcastle University). The spectrometer operated in positive-ion and data-independent acquisition (DIA) modes. During each DIA cycle, a survey scan was conducted to derive windows for isolating and fragmenting precursor ions. Full precursor ion scans were performed within a range of 390-1,500 m/z, with each window spanning 8 m/z. The analysis employed a maximum injection time of 100ms, a collision energy of 30%, and a fixed cycle time of 3s. The electrospray setup utilized a voltage of 2.0 kV and a capillary temperature of 275°C, with no sheath or auxiliary gas flow.

2.16.11. Proteomics analysis

Raw files were searched against a Uniprot *Human* reference database (containing 20323 entries, including isoforms) using DIA-NN version 1.8 with default settings to identify protein groups, as well as common contaminants. A false discovery rate (FDR) of <1% was employed. The databases were downloaded on 13/09/2021.

The list of proteins identified by DIA was filtered to remove contaminants and the proteins identified by <2 precursors per peptide. Downstream analysis of the protein group intensities was performed in R-Studio where the data was logarithmically transformed. Median imputation was used to deal with missing at random values. The *limma* R-Studio package was then used to run t-tests to determine which proteins show significant differential abundance (adj. *P-value* <0.05, Benjamin-Hochberg correction for multiple testing). This work was carried out by Ms Frances Sidgwick.

Hierarchical clustering of the biological replicate samples from distinct time points was used to identify anomalous data points and was carried out in R-Studio using the *dendextend* package.

A Gene Ontology (GO) enrichment for biological processes (BP) analysis was performed by Mr Victor Cheng (University of California Los Angeles) against a MSC chondrogenic differentiation proteome dataset. R-Studio and the *ClusterProfiler* package. Fischer's exact test with Benjamin-Hochberg adjustment was used.

2.17. Graphs and Statistical Analyses

All graphs were produced in GraphPad Prism 9.0 unless stated otherwise. Statistical analyses were performed in GraphPad Prism 9.0. A summary of the types of analyses used is presented in Table 2.7. Throughout the thesis the specific statistical analyses used are also listed within figure legends for the different data presented.

Table 2.7. Statistical analyses used throughout the thesis.

Experiment	Variables tested	Test used
mQTL analysis	DNAm levels at CpGs (M-values) and genotype at SNPs	Simple Linear Regression
DNAm comparisons	Mean DNAm levels at CpGs between different groups (e.g. OA knee vs OA hip)	Mann-Whitney test
AEI analysis	AEI ratios measured in DNA compared to those measured in cDNA	Wilcoxon's matched pairs signed rank test
AEI and DNAm	AEI ratios vs DNAm levels at CpGs	Simple Linear Regression
Reporter gene analysis	Construct containing insert of interest vs empty reporter gene vector	Paired t test
	Methylated vs non-methylated construct	Non-paired t test
	Multiple variables (e.g. haplotype at various SNPs and methylation status)	Multiple t-tests with Holm-Sidak correction for multiple testing

mQTL – methylation quantitative trait locus; CpG- cytosine-guanine dinucleotide, SNP – single nucleotide polymorphism; OA – osteoarthritis; AEI – allelic expression imbalance; DNAm – DNA methylation

2.18. In silico Analyses

Throughout my studies several commercially available databases and online tools were used for *in silico* characterisation of genetic loci of interest. Details of the databases and online tools used is provided in Table 2.8.

Some *in silico* analyses described in this thesis made use of publicly available -omics data generated by Professor Loughlin's group or by other research groups. The details of the datasets are listed in Table 2.9.

Table 2.8. *Details of databases and online tools used for in silico analyses.*

ChIP-seq – chromatin immunoprecipitation with sequencing; HiC – high throughput chromosome conformation capture; WGS – whole genome sequencing; WES – whole exome sequencing; RNA-seq – RNA sequencing

Table 2.9. *Details of -omics data utilised throughout the thesis.*

OA – osteoarthritis; RNA-seq – RNA sequencing; ATAC-seq - assay for transposase-accessible chromatin with sequencing

2.19. Contributions of Other People to the Work Presented in this Thesis

The collection of arthroplasty samples and the tissue processing (both cartilage and synovium) was done by other members of Professor John Loughlin's group.

The pyrosequencing assays used for the methylation quantification at CpGs 3-10 (rs11583641 locus), as well as the pyrosequencing assay used for the genotyping at rs11583641 were designed by Dr Sarah Rice prior to the start of my project.

The pyrosequencing assays used for the methylation quantification at cg15204595 (rs1046934 locus) was designed and tested by Mr Jack Roberts.

DNAm levels at CpGs 3-10 (rs11583641 locus) in 44 of the developmental cartilage samples were measured by Ms Julia Falk (research technician with Professors Loughlin and Young).

The chondrogenic differentiation of the ASC52teloSOX9 cells and the optimisation of the nucleofection parameters for these cells was performed alongside Mr Antony Sorial (PhD student with Professor Loughlin).

The protein modelling presented in Chapter 4 was performed by Dr Sarah Rice. The TSEN15 crystal structures were obtained from the Protein Data Bank and were examined both in combination with TSEN34 (PDB ID: 6Z9U) and as an individual monomeric structure (PDB ID: 2GW6) using the PyMOL Molecular Graphics System by Schrödinger. To simulate the missense variant Gln59-His introduced by rs1046934, she conducted *in silico* mutagenesis using the PyMOL Mutagenesis Wizard. To assess the potential impact of this variant, as well as the Gly19-Asp variant introduced by rs2274432, on the function of TSEN15, she utilized various tools including gnomAD, PolyPhen, and Mutation Taster.

Other contributions made to the acquisition or analysis of the data presented in this piece of work are listed in the text.

2.20. Overlap Between MRes and PhD

The work on the project described in this thesis started during my MRes. Genotyping at rs11583641 and rs743567, as well as DNAm quantification at CpGs 1-12 for 79 of the arthroplasty cartilage samples and 50 of the foetal cartilage samples was done as part of my MRes project. AEI analysis at the rs11583641 locus in 12 arthroplasty cartilage samples was also performed and these data were presented in my MRes dissertation.

Some plots contain the data generated during my MRes with additional samples added to the analyses as part of my PhD project.

2.21. Presentations and publications

Work described in this thesis has been presented in front of the following scientific forums:

April 2022 Allelic heterogeneity in *COLGALT2* expression contributes to osteoarthritis risk through decreased DNA methylation at multiple enhancers. Kehayova YS*, Wilkinson JM, Rice

SJ, Loughlin J. Osteoarthritis Research Society International (OARSI) Congress, Berlin, Germany. (Poster presentation)

April 2022 Epigenomic analysis of osteoarthritis genetic risk during human foetal development. Rice SJ*, Falk J, Brumwell A, Kehayova YS, Casement J, Parker E, Hofer IM, Shepherd C, Loughlin J. Osteoarthritis Research Society International (OARSI) Congress, Berlin, Germany.

March 2021 The molecular genetics and epigenetics of *COLGALT2*, a risk locus for osteoarthritis. Kehayova YS*, Wilkinson JM, Loughlin J, Rice SJ. Osteoarthritis Research Society International (OARSI) Congress, Vienna, Austria. Conference cancelled due to COVID-19. Talk delivered online at “OARSI Live”.

The following peer-reviewed publications also feature work from this thesis:

Kehayova YS, Watson E, Wilkinson JM, Loughlin J, Rice SJ (2021). Genetic and epigenetic interplay within a *COLGALT2* enhancer associated with osteoarthritis. *Arthritis and Rheumatology* 73(10):1856- 1865

Rice SJ, Falk J, Brumwell A, **Kehayova YS**, Casement J, Parker E, Hofer IM, Shepherd C, Loughlin J (2023). Genetic Risk of osteoarthritis operates during human skeletogenesis. *Human Molecular Genetics* 32(13):2124-2138

Kehayova YS, Wilkinson JM, Rice SJ, Loughlin J (2023). Mediation of the same epigenetic and transcriptional effect by independent osteoarthritis risk-conferring alleles on a shared target gene, *COLGALT2*. *Arthritis and Rheumatology* 75(6):910-922

Kehayova YS, Wilkinson JM, Rice SJ, Loughlin J (2023). Osteoarthritis genetic risk acting on the galactosyltransferase gene *COLGALT2* has opposing functional effects in articulating joint tissues. *Arthritis Research & Therapy* 25(83)

2.22. References

1. Katz, D. B., Huynh, N. P. T., Savadipour, A., Palte, I. & Guilak, F. An immortalized human adipose-derived stem cell line with highly enhanced chondrogenic properties. *Biochem. Biophys. Res. Commun.* **530**, 252–258 (2020).

2. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
3. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
4. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
5. 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 (2021).
6. Alt-RTMCRISPR-Cas9 crRNA design tool, IDT, Coralville, Iowa, USA
7. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
8. Vincze, T., Posfai, J. & Roberts, R. J. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res.* **31**, 3688–3691 (2003).
9. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
10. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
11. Rushton, M. D. *et al.* Characterization of the Cartilage DNA Methylome in Knee and Hip Osteoarthritis. *Arthritis Rheumatol.* **66**, 2450–2460 (2014).
12. 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).
13. Ge, X. *et al.* Functional genomics atlas of synovial fibroblasts defining rheumatoid arthritis heritability. *Genome Biol.* **22**, 247 (2021).
14. Rice, S. J. *et al.* Genetic risk of osteoarthritis operates during human skeletogenesis. *Hum. Mol. Genet.* **32**, 2124–2138 (2022).

Chapter 3. Genetic and Epigenetic Interplay Within a *COLGALT2* Enhancer Associated with Osteoarthritis

3.1. Introduction

Along with the number of susceptibility loci being uncovered by GWAS, the evidence that common disease risk is linked to alleles which influence gene expression via the modulation of epigenetic factors such as DNAm, is growing¹. Characterising the exact mechanisms driving common diseases, therefore, requires thorough investigation of the links between genetic variation, epigenetic marks, and gene expression.

A highly powered GWAS using the UK Biobank dataset identified the SNP rs11583641 as significantly associated with hip OA ($P=5.6 \times 10^{-10}$)². A subsequent cartilage methylome analysis revealed that rs11583641 acts as an mQTL³. The major and OA effect allele of rs11583641, C, correlates with reduced DNAm at an intronic CpG, cg18131582. Both the SNP and CpG are located within the gene body of *COLGALT2*. The expression of *COLGALT2* was shown to be significantly increased in OA hip cartilage when compared to control NOF cartilage³.

Considering the available information, I hypothesised that the OA-association signal marked by rs11583641, or another SNP in high LD with it, contributes to OA predisposition by affecting the expression levels of *COLGALT2* in cartilage. Furthermore, I hypothesised that the region harbouring cg18131582 acts as enhancer in cartilage and DNAm levels at one or more CpG sites within this enhancer act as an intermediate between the genetic predisposition marked by rs11583641 and *COLGALT2* expression.

3.2. Aims and Objectives

3.2.1. Aims

- Characterise the relationship between genotype at rs11583641 and DNAm levels at the putative enhancer region.

- Determine the regulatory activity of the putative enhancer in a human chondrocyte cell model.
- Determine if the activity of the putative enhancer is methylation sensitive.
- Identify the target gene of the activity of the putative enhancer.
- Investigate the role of DNAm at the putative enhancer in the regulation of gene expression of its target gene(s) in chondrocytes.
- Determine the relationship between genotype at rs11583641 and gene expression of the target gene(s) in arthroplasty cartilage patient samples.

3.2.2. Objectives

- Genotype cartilage arthroplasty patient samples at rs11583641.
- Replicate the mQTL signal at cg18131582 in the patient samples.
- Measure DNAm levels at an additional 11 CpG sites within the putative enhancer and determine the physical limits of the DMR.
- Clone the DMR in CpG-free Lucia reporter gene vector, methylate the constructs *in vitro*, transfect the constructs in TC28a2 cells and measure Lucia readings.
- Activate and/or repress the putative enhancer using dCas9 coupled with VP64 or KRAB proteins, respectively, and measure the expression of *COLGALT2* and its neighbouring genes, *RGL1* and *TSEN15*, following the modulation.
- Delete the enhancer using Cas9 and measure the effect of this on the expression of *COLGALT2*, *RGL1* and *TSEN15*.
- Methylate and demethylate the differentially methylated CpG sites within the putative enhancer using dCas9/DNMT3a or dCas9-TET1 fusion proteins, respectively, and measure the expression of *COLGALT2* and, *RGL1* and *TSEN15*, following the modulation.
- Perform AEI analysis in arthroplasty patient samples for the target gene of the rs11583641 signal.

3.3. Results

3.3.1. Characterisation of the region encompassing rs11583641 and cg118131582

The UCSC Genome Browser was used to visualise the region containing rs11583641 and cg18131582. As previously discussed, rs11583641 is located on chromosome 1q25.3 within the 3'UTR of *COLGALT2* (Figure 3.1, panels 1-2). The region surrounding cg18131582 is a putative intronic enhancer, located 6kb upstream of the SNP. cg18131582 resides in a cluster of 8 CpGs spanning 433bp (Figure 3.1, panels 3-4). There are two SNPs within this cluster of CpG sites, rs943409 (G>A) and rs734657 (C>A) (Figure 3.1, panel 2). The major allele, G, of rs943409 forms a CpG site which was not investigated due to its polymorphic nature. To determine the physical limits of the DMR, the attention was initially focussed on this cluster, along with the four most proximal CpGs flanking the region, covering a region of 1560bp (Figure 3.1, panel 3). The CpGs were numbered 1-12 based on their physical location with CpG 9 being cg18131582.

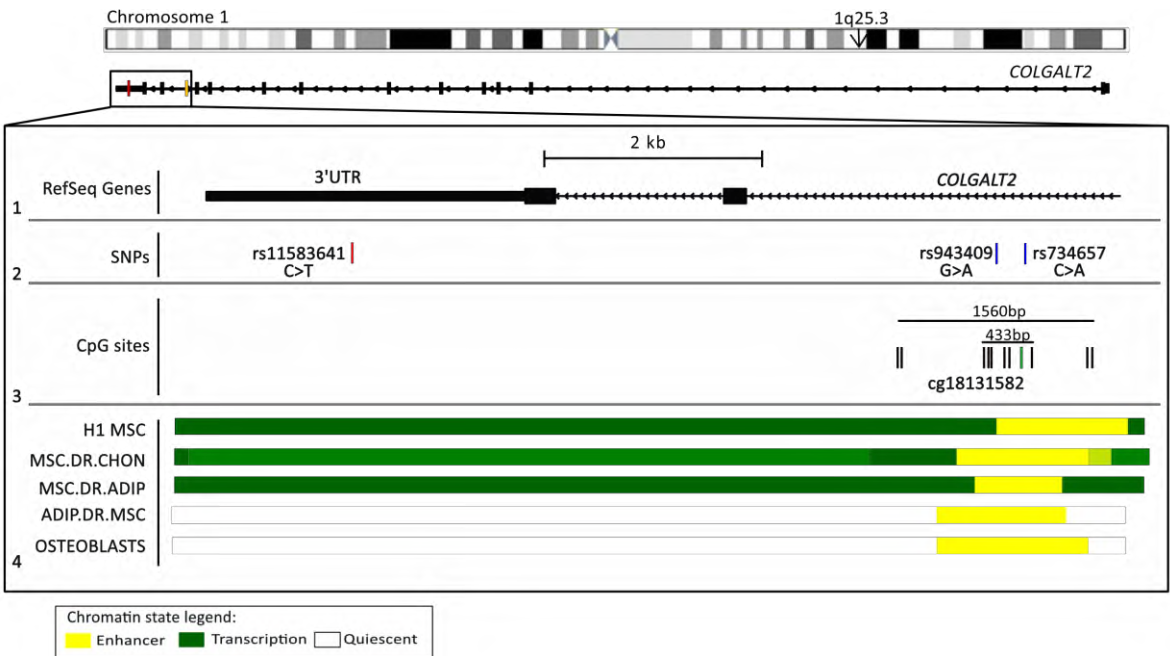


Figure 3.1 Schematic overview of the region containing the GWAS signal and the investigated CpG sites. **Panel 1**, The relative genomic position of *COLGALT2* (shown are only the 3'UTR, exon 11, and introns 11 and 10 (partially)) visualised in the UCSC browser (hg19). **Panel 2**, The genomic position of rs11583641 (red line), rs943409 and rs734657 (blue lines) **Panel 3**, The

genomic position of cg18131582 (green line) and the other investigated CpG sites (black lines) **Panel 4**, Chromatin state data, as determined from the Roadmap project, for primary human mesenchymal stem cells (H1 MSC), MSC-derived chondrocytes (MSC.DR.CHON) and adipocytes (MSC.DR.ADIP), adipose-derived MSCs (ADIP.DR.MSC), and human osteoblasts. Colours correspond to different chromatin states, as indicated in legend at bottom.

3.3.2. Replication of the mQTL at cg18131582 and determination of the limits of the DMR in arthroplasty cartilage

DNAm at the 12 CpGs in DNA was quantified in DNA from OA knee, OA hip, or control non-OA NOF patients. DNAm in all samples was stratified by genotype at rs11583641 (Figure 3.2A). Significant mQTLs were identified at 5/12 CpGs: CpG 4 ($P=0.05$), CpG 5 ($P=0.04$), CpG 8 ($P<0.0001$), CpG 9 (cg18131582, $P=0.01$), and CpG 10 ($P<0.001$).

The median DNAm levels across the region were compared (Figure 3.2B). At 8/12 CpGs DNA was hypermethylated (median DNAm >60%). Based on this data, the DMR was determined to span the 398bp between CpG 3 and CpG 10. At CpGs 8-10 median DNAm was lower (CpG 6- 20.3%, CpG 7 – 54.8, CpG 8 – 51.5%) with high interindividual variability (64% DNAm range at CpG 8, 54.5% at CpG 9, 47.5% at CpG 10). This provides evidence for a regulatory function of the region containing these CpG sites.

When DNAm was stratified by genotype at the SNP separately in OA and NOF samples, 5/12 CpGs sites were differentially methylated in OA samples (Figure 3.3A) and 2/12 in the control NOF samples (Figure 3.3B). These results showed that while rs11583641 affects DNAm levels at the putative enhancer in all aged cartilage samples, in non-OA samples the allele-dependent methylation effect is restricted to CpGs 8 and 9. At all differentially methylated CpG sites the OA-risk allele, C, of rs11583641 was associated with reduced DNAm compared to the non-risk, T, allele.

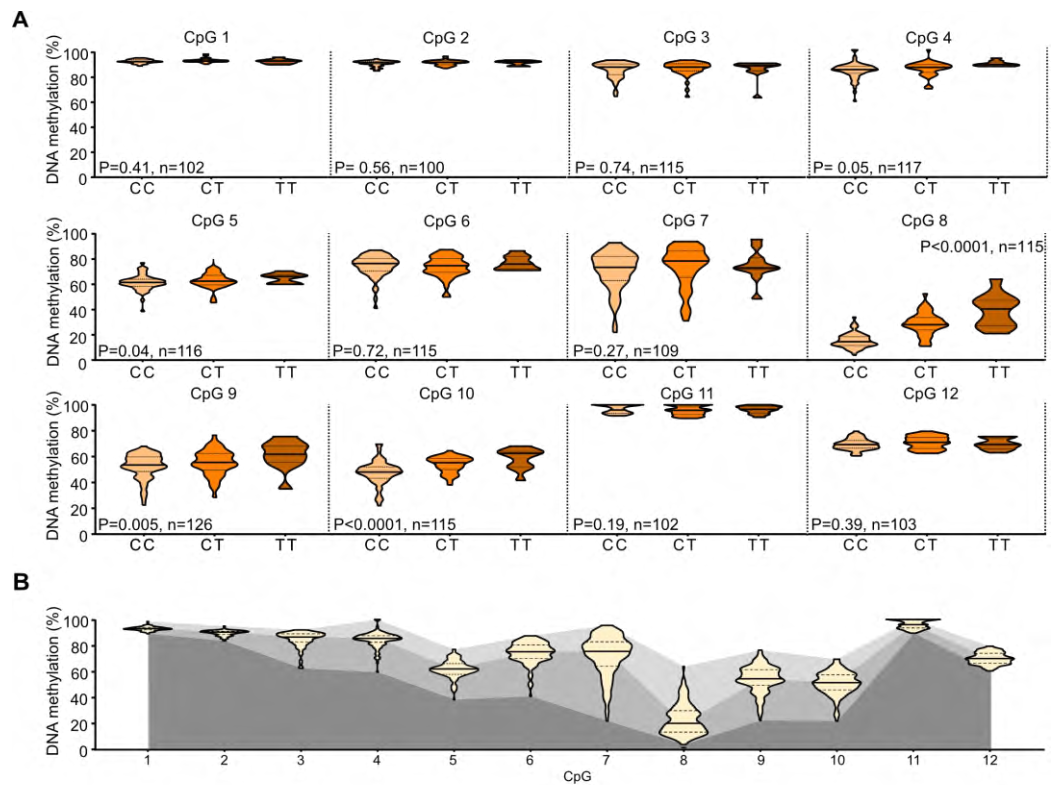


Figure 3.2. Cartilage mQTL analysis of 12 CpG sites within the putative COLGALT2 enhancer. **A.** Violin plots showing DNAm values in arthroplasty cartilage samples at the 12 investigated CpGs, stratified by rs11583641 genotype. Solid lines and dashed lines inside the plots represent median and interquartile range, respectively. P-values were calculated using simple linear regression. **B.** Violin plots showing median DNAm levels in all cartilage samples at each of the 12 investigated CpGs, plotted irrespective of genotype. Solid and dashed horizontal lines represent the median and interquartile range, respectively. Shaded gradients represent the upper and lower range of data points at each CpG. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A. and B.).

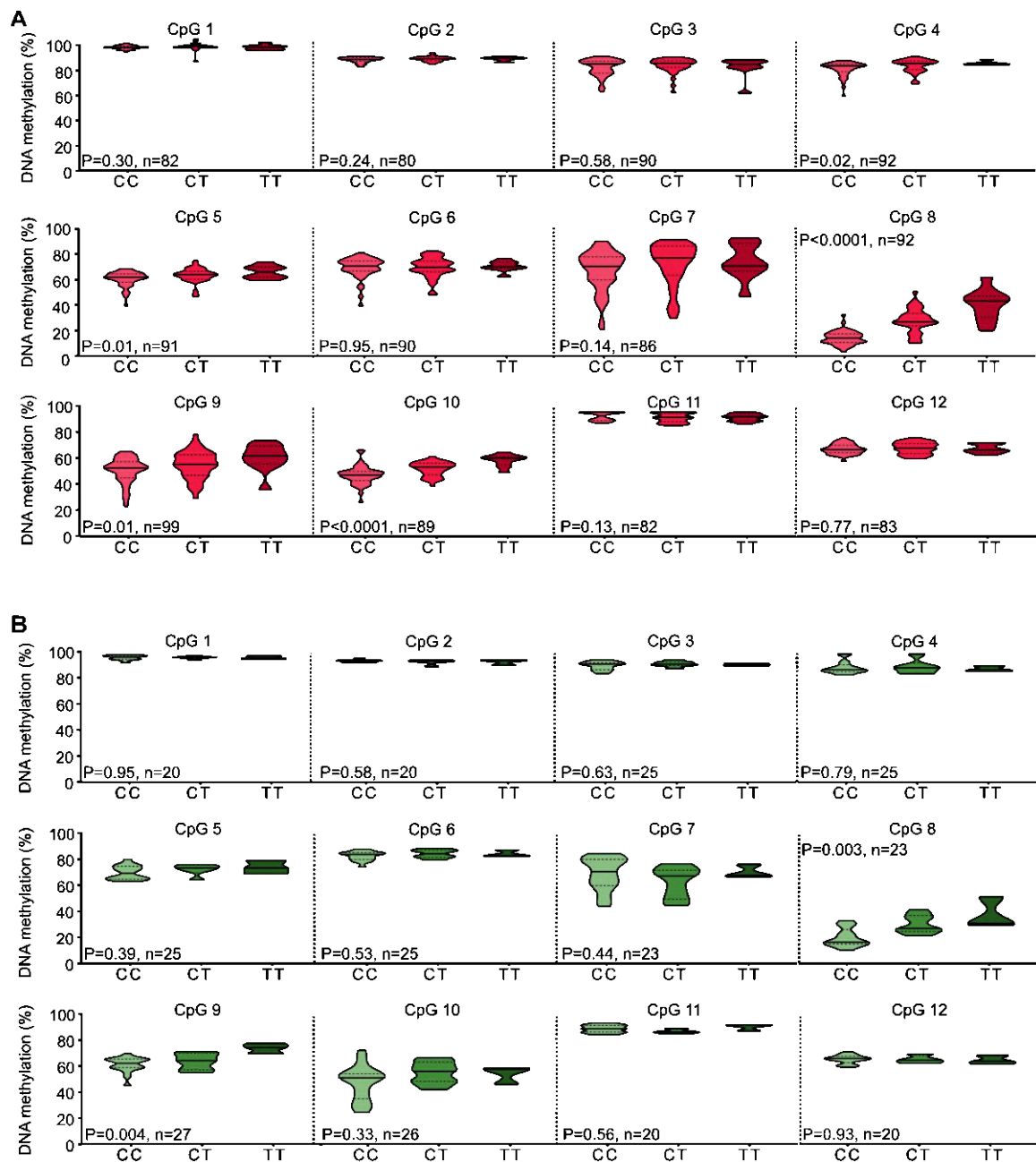


Figure 3.3. mQTL analysis of the 12 investigated CpG sites in OA and non-OA cartilage samples. **A.** Violin plots showing DNAm values in OA cartilage samples at the 12 CpGs, stratified by genotype at rs11583641. Solid and dashed horizontal lines represent the median and interquartile range, respectively. P-values were calculated using simple linear regression. **B.** Violin plots showing DNAm values in NOF cartilage samples at the 12 CpGs, stratified by genotype at rs11583641. Solid and dashed horizontal lines represent the median and interquartile range, respectively. P-values were calculated using Simple Linear Regression. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A. and B.).

3.3.3. Joint site and disease impact upon DNAm at the putative enhancer

We studied the effect of joint site (hip or knee) upon DNAm. At 8/12 CpGs, DNAm was significantly lower (CpG 2 - $P=0.003$, CpG 3 - $P=0.01$, CpG 5 - $P=0.03$, CpG 6 - $P=0.003$, CpG 7 - $P=0.02$, CpG 8 - $P=0.02$, CpG 9 - $P=0.001$, CpG 10 - $P=0.03$) in hip than in knee cartilage (Figure 3.4A).

Comparison of the DNAm between OA hip and NOF samples allowed investigation of disease status upon DNAm. At 7/12 CpGs, higher median DNAm was measured in NOF ($n=20-27$) compared to OA hip (CpG 2 - $P=0.003$, CpG 3 - $P<0.0001$, CpG 4 - $P<0.0001$, CpG 5 - $P<0.0001$, CpG 6 - $P<0.0001$, CpG 8 - $P=0.05$, CpG 9 - $P<0.001$) (Figure 3.4A). The greatest differences between median values (12% and 12.5%, respectively) were observed at CpG 6 ($P<0.001$), and CpG 9 ($P<0.001$). The exception to this trend was CpG 12, the most distal from the cluster, at which NOF DNAm was lower ($P<0.001$) than in OA hip samples.

The strength of the effect of genotype at rs11583641 upon DNAm across the 12 CpG sites between the sample groups (OA, OA hip, OA knee and NOF) was also measured (Figure 3.4B). Linear regression analysis was used to calculate the genotypic effect (GE), the percentage contribution of genotype to the observed differences in DNAm, within each of the groups. Within the DMR, OA knee samples showed stronger GE at 3/5 differentially methylated CpG sites compared to the other groups (CpG 5- 14.7%, CpG 8 – 76.9%, CpG 10 – 41.4%). At CpG 4 the effect was stronger in OA hip samples (8.9%) and, interestingly, at CpG 9 the GE was greatest in NOF samples (33.3%). The GE at CpG1 was very strong (34.4%) for OA hip samples but not for NOF nor for combined OA. This is due to an mQTL detected at the CpG in OA hip samples but not in OA knee nor NOF samples (Figure 3.5).

Although the NOF patients were on average older than the OA patients when they underwent arthroplasty (77.4 years compared to 65.3 years for OA knee and 66.5 years for OA hip), there were no correlations between age and DNAm at any CpG sites in the combined arthroplasty samples (Figure 3.6A), OA samples (Figure 3.6B) or NOF samples (Figure 3.6C).

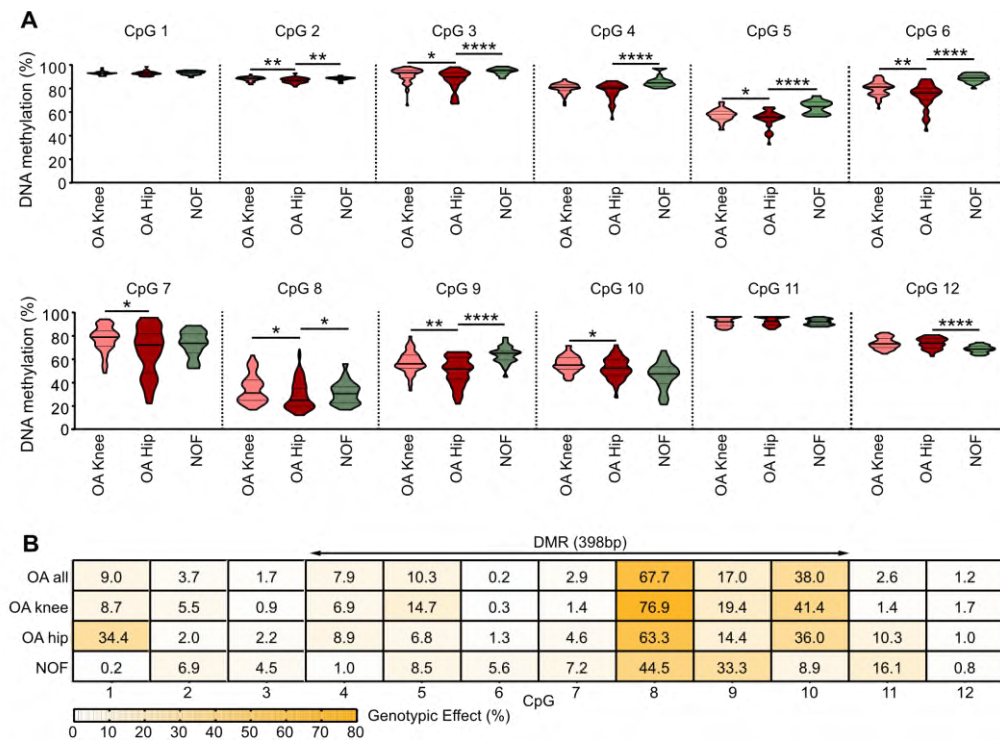


Figure 3.4. Effects of joint site and disease status on DNAm. **A.** Violin plots showing DNAm levels at the 12 investigated CpGs in cartilage samples from patients with knee OA ($n = 47-60$), hip OA ($n = 33-44$), or NOF ($n = 20-27$). Solid and dashed horizontal lines represent the median and interquartile range, respectively. P -values were calculated using a Mann-Whitney test. * = $P < 0.05$; ** = $P < 0.01$; **** = $P < 0.0001$. **B.** Heatmap showing the contribution of the rs11583641 genotype to the observed DNAm levels differences at the 12 CpGs in cartilage samples from patients with all OA (both hip and knee; $n = 79-92$), knee OA ($n = 44-55$), hip OA ($n = 33-43$), or NOF ($n = 20-27$). The heatmap displays r^2 values calculated by simple linear regression analysis, converted to percentages. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A. and B.).

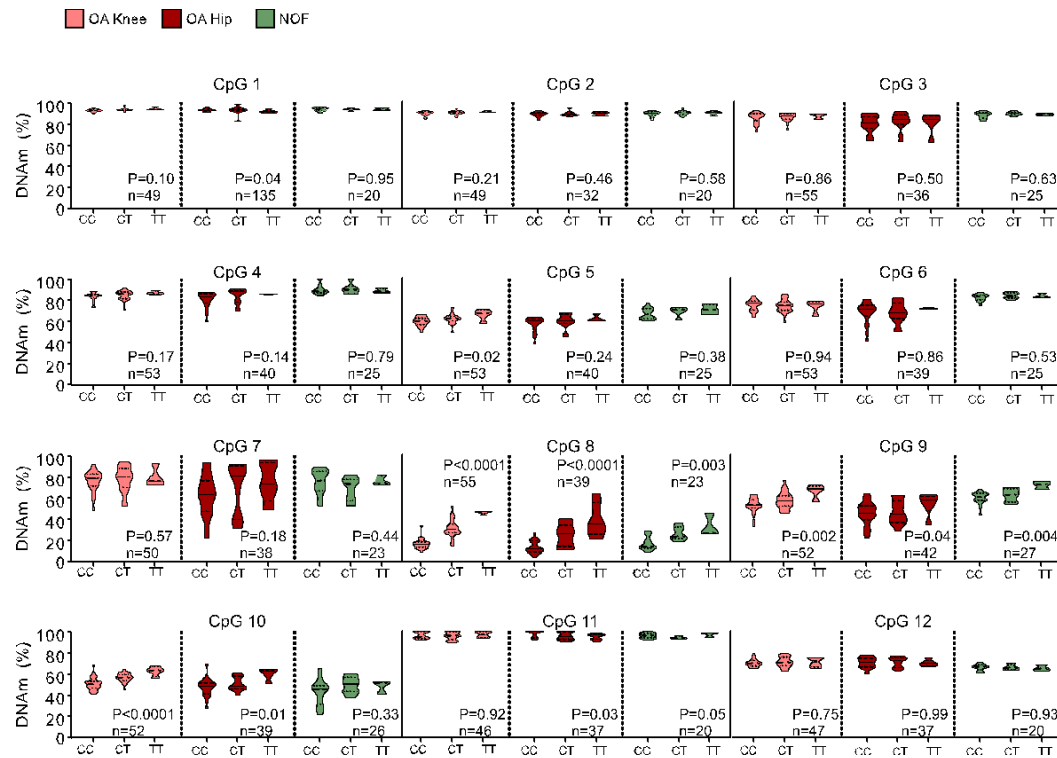


Figure 3.5. mQTL analysis of the 12 investigated CpG sites in OA Knee, OA hip and non-OA cartilage samples (NOF). Violin plots showing DNAm values in the samples (y-axis) at the 12 CpGs, stratified by genotype at rs11583641 (x-axis). Solid and dashed horizontal lines represent the median and interquartile range, respectively. P-values were calculated using simple linear regression. GraphPad Prism 9.0 was used for statistical analyses and data visualisation.

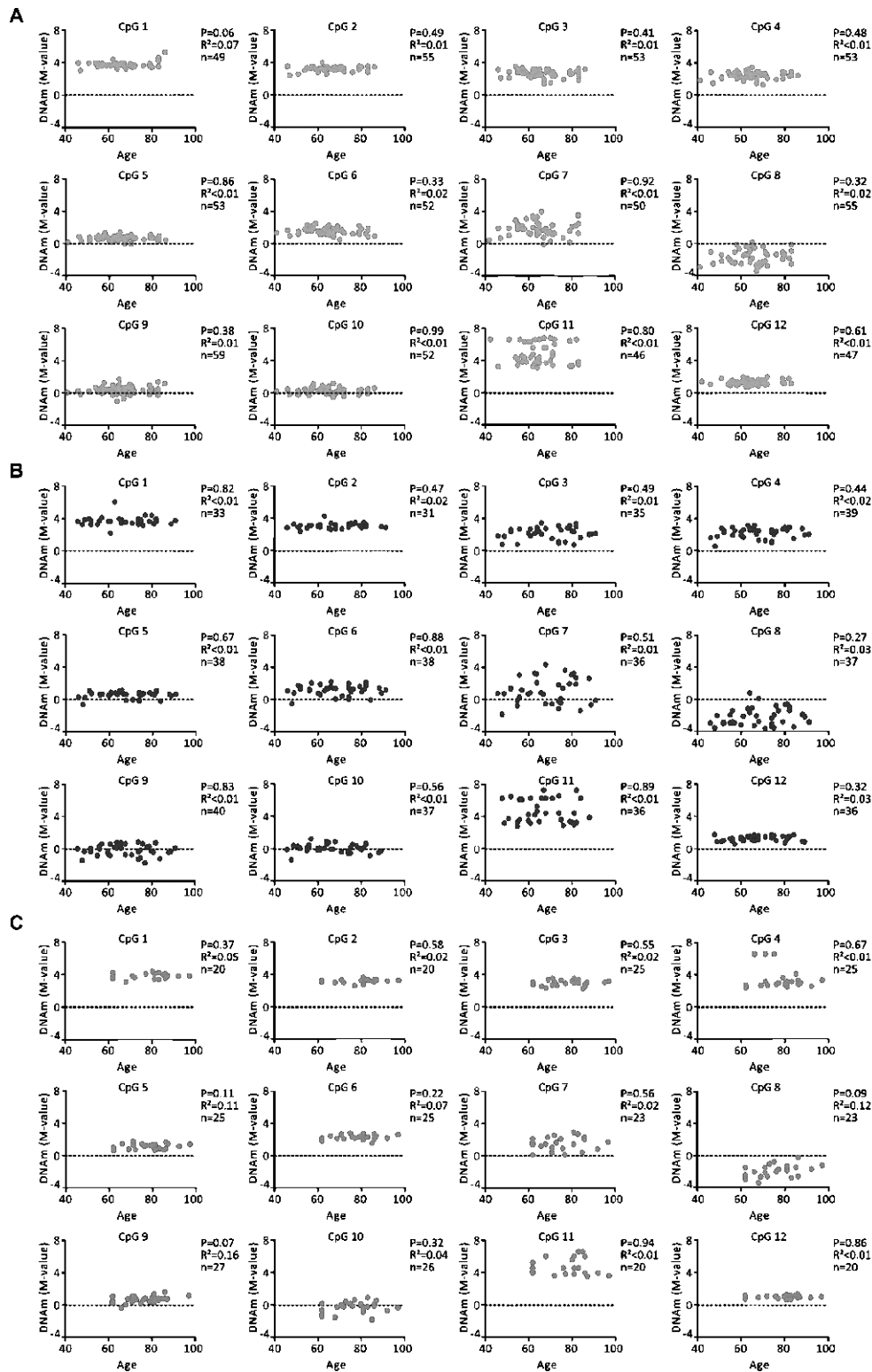


Figure 3.6. Relationship between age of the patients at the time of surgery and DNAm levels at the investigated CpG sites. **A.** Age at time of surgery of OA knee patients (n=47-59) (y-axis) plotted against DNAm levels at the 12 CpGs represented as β -values (x-axis). P-values were calculated using simple linear regression. **B.** Age at time of surgery of OA hip patients (n=31-43) (y-axis) plotted against DNAm levels at the 12 CpGs represented as β -values (x-axis). P-values were calculated using simple linear regression. **C.** Age at time of surgery of NOF

patients ($n=20-27$) (y-axis) plotted against DNAm levels at the 12 CpGs represented as β -values (x-axis). P-values were calculated using simple linear regression. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A., B., and C.).

3.3.4. Assessment of the regulatory function of the region containing the differentially methylated CpGs

The region containing the differentially methylated CpGs was tested for regulatory activity in the human chondrocyte cell line TC28a2 using a reporter gene assay. A 503bp region encompassing CpGs 3-10 was cloned into a Lucia reporter-gene vector. The cloned region also contained the two SNPs, rs943409 (G>A) and rs734657 (C>A) (Figure 3.7A), the latter of which is in high LD with rs11583641 ($r^2=0.83$, GBR (British in England and Scotland)) (Figure 3.7B). rs943409 and rs734657 have an r^2 of 0.13 and a D' of 1.0 in the GBR and occur in three haplotypes in that population: G_C, A_C, and G_A (Figure 3.7B). The impact of the three haplotypes upon enhancer activity was tested in both the methylated and the unmethylated constructs.

Two of the constructs showed increased enhancer activity in TC28a2 chondrocytes (Figure 3.7C). These two haplotypes both contained the rs734657 C-allele (G_C and A_C) and showed 3.1-fold ($P<0.0001$) and 4.1-fold ($P<0.0001$) increase in Lucia reporter gene activity, respectively, compared to the control. No significant difference was observed between the activity of these two haplotypes, suggesting that the genotype at rs943409 did not influence the activity of the enhancer. The G_A haplotype conferred significantly lower activity compared to the G_C ($P=0.006$) and the A_C ($P=0.0007$) haplotypes (Figure 3.7C), indicating that C allele of rs734657, which corresponds to the OA risk allele, C, at rs11583641, leads to increased enhancer activity.

In all three haplotype constructs, DNAm at the CpGs significantly reduced enhancer activity (Figure 3.7D). The differences in Lucia activity between the methylated and the unmethylated constructs are 2.5-fold ($P<0.0001$) for G_C, 1.4-fold ($P=0.001$) for A_C and 0.8-fold ($P=0.04$) for G_A constructs (Figure 3.7D).

Figure 3.7. Reporter gene analysis. **A.** A schematic representation of the DMR cloned into the reporter gene plasmid in relation to the location of COLGALT2 and its neighbouring genes RGL1 and TSEN15. Circles represent the 8 CpGs (CpGs 3–10) that were captured with cg18131582 being CpG 9 (green circle). The position of the two SNPs located within the DMR, rs943409 and rs734657, is indicated by two black vertical lines. The image was adapted from UCSC Genome browser. **B.** A schematic of the relationship between the two SNPs in the putative enhancer, rs943409 and rs734657, and rs11583641. The intensity of the colour represents the LD between the SNPs with darker colour corresponding to higher r^2 value. The numbers in the squares represent the exact r^2 values. The image was adapted from LDLink (NIH). **C.** Lucia reporter assays readings from TC28a2 cells transfected with the unmethylated constructs containing the 3 haplotypes of rs943409 and rs734657. **D.** Lucia reporter assays readings from TC28a2 cells transfected with the constructs containing the 3 haplotypes of rs943409 and rs734657 in a methylated (black bars) or unmethylated state (gray bars). For C. and D. values were normalized to those measured in an empty vector control. Dots represent individual samples ($n = 7$), and the error bars show the mean $\pm$ SEM. P-values were calculated using multiple t-tests with Holm-Sidak correction for multiple testing. * = $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.001$; **** = $P < 0.0001$. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (C. and D.).

3.3.5. Determination of the effect of genotype at rs734657 upon DNAm at the enhancer in arthroplasty samples in TC28a2 cells

The results from the reporter gene analysis confirmed that the region between CpG3-CpG10 acts as an enhancer in TC28a2 cells and suggested that genotype at rs734657 might have a functional role in the activity of the region. Since rs734657 is in high LD with rs11583641 it is possible that this is the causal SNP driving the observed mQTLs in arthroplasty cartilage. In the British population there are 24 SNPs in high LD ($r^2 > 0.8$) with rs11583641. They form 5 groups within which the SNPs are in perfect LD ($r^2 = 1.0$) with each other and in high LD

($r^2 > 0.8$) with rs11583641 (Table 3.1) (Figure 3.8). In order to determine if the genotype at any of those SNP explains the DNAm changes across the enhancer better than rs11583641, the arthroplasty samples used for the mQTL analysis were genotyped at SNPs from those groups (Group 1 – rs11583641, Group 2 – rs12083450, Group 3 – rs7522001, Group 4 – rs10752924 and Group 5 – rs734657) and the DNAm at the 12 CpG sites within the enhancer was stratified by the new genotypes. Linear regression was used to determine the GE in arthroplasty (OA and NOF combined) (Figure 3.9A), OA (Figure 3.9B) and NOF (Figure 3.9C) cartilage.

Table 3.1. Details of the SNPs in high LD with rs11583641.

Group	SNP	Alleles	MAF (GBR)	Position (hg19)	r^2 with rs11583641
1	rs11583641	C>T	0.324	chr1:183906245-183906245	1.0
	rs10911472	A>G	0.324	chr1:183908813-183908813	1.0
	rs3828029	C>A	0.324	chr1:183901357-183901357	1.0
	rs12023991	T>C	0.324	chr1:183905477-183905477	1.0
	rs71657827	AA> -	0.324	chr1:183905485-183905488	1.0
2	rs12083450	G>A	0.319	chr1:183880299-183880299	0.97
3	rs7522001	T>C	0.341	chr1:183898790-183898790	0.93
	rs10797923	T>C	0.341	chr1:183901966-183901966	0.93
4	rs11591089	C>T	0.335	chr1:183876182-183876182	0.90
	rs148826550	- > T	0.335	chr1:183876527-183876526	0.90
	rs115896819	A>G	0.335	chr1:183876527-183876527	0.90
	rs11586086	T>C	0.335	chr1:183876610-183876610	0.90
	rs12119315	C>G	0.335	chr1:183879084-183879084	0.90
	rs4400572	C>G	0.335	chr1:183881798-183881798	0.90
	rs10911467	G>A	0.335	chr1:184308761-184308761	0.90
	rs10752924	C>A	0.335	chr1:183887032-183887032	0.90
	rs10911468	T>C	0.335	chr1:184308766-184308766	0.90
	rs10911470	T>C	0.335	chr1:183889204-183889204	0.90
	rs7555062	T>C	0.335	chr1:183889674-183889674	0.90
	rs7519678	T>C	0.335	chr1:183892785-183892785	0.90
	rs7552871	G>T	0.335	chr1:183892786-183892786	0.90
	rs10797922	T>G	0.335	chr1:183894775-183894775	0.90
	rs71708186	- > A	0.335	chr1:183897303-183897302	0.90
5	rs734657	C>A	0.319	chr1:183912319-183912319	0.83
	rs730601	C>T	0.319	chr1:183911602-183911602	0.83

SNP- single nucleotide polymorphism; MAF- minor allele frequency

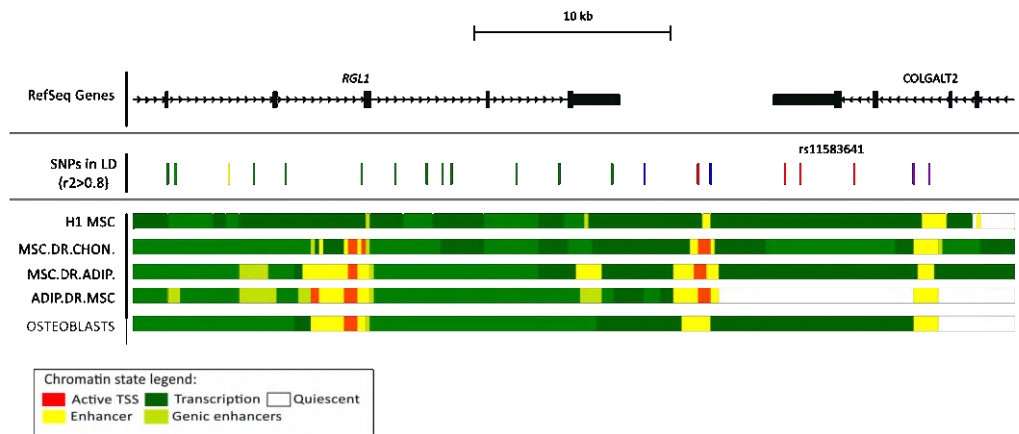


Figure 3.8. A schematic showing the position of the SNPs in high LD with rs11583641 in relation to the genes in the region (**top panel**) (hg19). Each line (**middle panel**) shows a SNP with rs11583641 being labelled. The colour of the line indicates the group the SNPs belongs to: red – Group 1 (r^2 with rs11583641 = 1.0, GBR); yellow – Group 2 (r^2 with rs11583641 = 0.97, GBR); blue – Group 3 (r^2 with rs11583641 = 0.93, GBR); green – Group 4 (r^2 with rs11583641 = 0.90, GBR); purple – Group 5 (r^2 with rs11583641 = 0.83, GBR). Chromatin state predictions by the ROADMAP database of histone modifications (**bottom panel**) for primary human mesenchymal stem cells (H1 MSC), MSC-derived chondrocytes (MSC.DR.CHON) and adipocytes (MSC.DR.ADIP), adipose-derived MSCs (ADIP.DR.MSC), and human osteoblasts. Colours correspond to different chromatin states, as indicated in legend at bottom. The image was adapted from UCSC genome browser (hg19).

In all arthroplasty cartilage samples, irrespective of disease status (Figure 3.8A), the GE of the SNPs from Group 1 was strongest at 3/5 differentially methylated CpG sites (CpG 4–5.4%, CpG 5 – 6.0%, CpG 8 – 52.0 %). At CpG 5 the GE of the SNPs from Group 3 and Group 4 was the highest (7.0%). The GE at CpG9 (cg19131582) was strongest for SNPs from Group 4 (12%). The GE of rs173657 (Group 5) was lower than that of rs11583641 (Group 1) for all differentially methylated CpG sites. In OA samples (Figure 3.8B) the GE of the SNPs from Group 1 was stronger than that of SNPs from other groups at 4/5 differentially methylated CpG sites with Group 5 SNPs having 0.3% higher GE at CpG4. In NOF samples, in which mQTLs were detected only at CpG 8 and CpG 9, the GE at CpG 8 was strongest for SNPs from Group 5 and for CpG 9 – for SNPs from Group 3 (Figure 3.8C).

The differences in the GE between the SNPs from the 5 group were small (<10%) for all samples, except for CpG 8 at which the greatest difference of GE was 12% for arthroplasty cartilage and 14% for NOF samples. Overall, no one group of SNPs appeared to explain the differences of DNAm observed across the enhancer better than rs11583641.

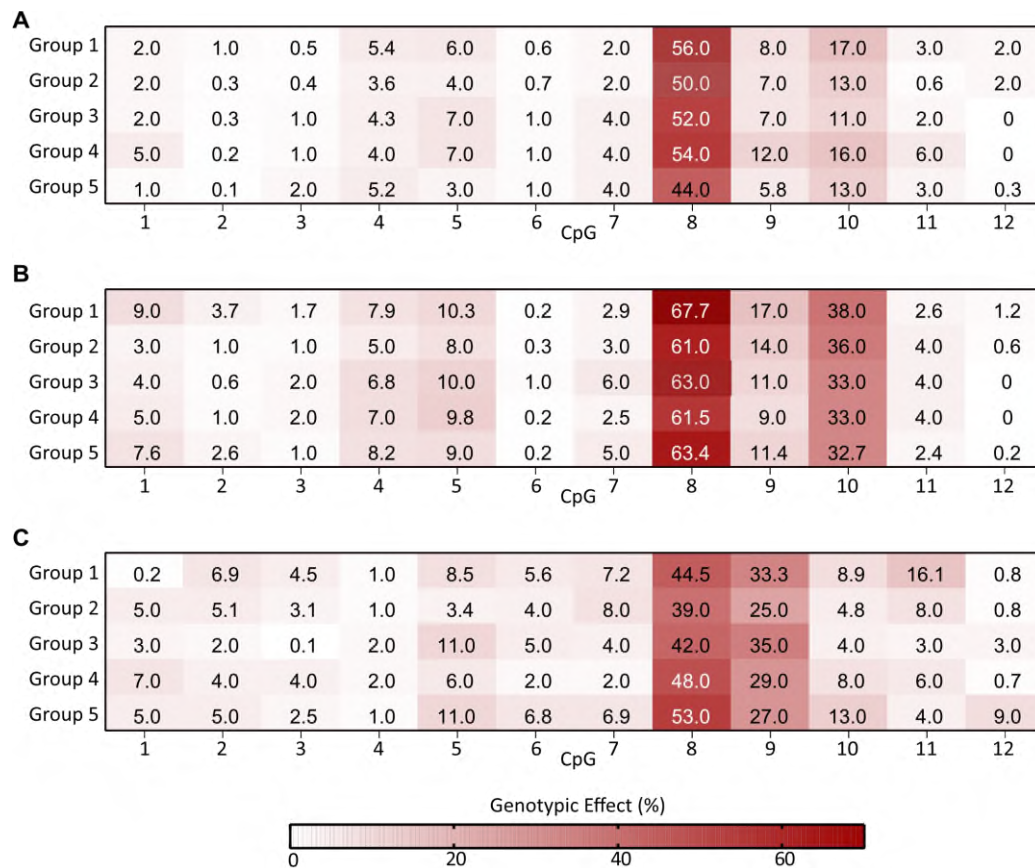


Figure 3.9. Genotypic contributions of the SNPs in high LD with rs11583641 to DNAm variability measured across the COLGALT2 enhancer. **A.** Heatmap showing the contribution of the genotype of SNPs from the LD block containing rs11583641 to the observed DNAm levels differences at the 12 CpGs in arthroplasty cartilage samples (n=100-126). **B.** Heatmap showing the contribution of the genotype of SNPs from the LD block containing rs11583641 to the observed DNAm levels differences at the 12 CpGs in OA cartilage samples (n=80-99). **C.** Heatmap showing the contribution of the genotype of SNPs from the LD block containing rs11583641 to the observed DNAm levels differences at the 12 CpGs in NOF cartilage samples (n=20-27). The heatmaps displays r^2 values calculated by linear regression analysis, converted to percentages. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A., B. and C.).

3.3.6. Identification of the target gene of the enhancer

The next step in the investigation of this locus was to determine which gene is the target of the enhancer. To achieve this goal, the enhancer was targeted using 6 gRNAs in TC28a2 cells engineered to express dCas9 fused with either VP64 transcriptional activator or KRAB transcriptional repressor. The 6 gRNAs (gRNA1-6) were designed to target the region between CpG 3 and CpG 10 containing the differentially methylated CpG sites (Figure 3.10A).

The gRNAs were tested in a TC28a2 cell line engineered to express active Cas9. Following transfection with the gRNAs, DNA was isolated from the cells and T7 endonuclease I (T7EI) assay confirmed that the gRNAs successfully target the region (Figure 3.10B. and C.). The gRNAs (alone or in combination) were coupled with a fluorescently labelled tracrRNA and transfected into the TC28a2-dCas9/VP64 and Tc28a2-dCas9/KRAB cells. The efficiency of the transfection was confirmed by fluorescent microscopy.

The expression of *COLGALT2* and its two neighbouring genes, *RGL1* and *TSEN15*, was measured in the cells following the transfections. No individual gRNA or a combination of gRNAs led to a significant increase (for TC28a2-dCas9/VP64 cells, Figure 3.10C) or decrease (for TC28a2-dCas9/KRAB cells, Figure 3.10D) of expression of any of the three genes.

Figure 3.10. Effect of activation and repression of the enhancer on gene expression. **A.** Schematic diagram showing the genomic position of the 6 gRNAs (arrows) used for the activation/repression of the enhancer (gRNAs 1– 2), relative to the 8 CpGs of the DMR represented by circles with cg18131582 being CpG 9 (green circle). The image was created in Affinity Designer 2.1. **B.** A graph showing the expected band sizes for the heteroduplexes produced by each guide (gRNA 1 – 6) following the T7EI digestion. **C.** Gel image showing results

from the T7EI assay. The left-hand lane is a 1kb Plus DNA Ladder and the samples treated with the T7EI in the remaining lanes, gRNA1 (G1) to gRNA6 (G6). **C.** qPCR results from the dCas9-VP64 transcriptional activation experiment. The graphs show the relative gene expression of *RGL1* (left), *COLGALT2* (middle) and *TSEN15* (right). The gene expression levels of each of the three genes are normalised to those measured in the non-targeting control cells (dashed line) and shown for all conditions: TC28a2/dCas9-VP64 cells transfected individually with gRNAs 1-6, cells transfected with all gRNAs, and cells transfected with gRNAs 1, 3 and 5. The bars show the mean of six biological repeats and the error bars represent the SEM. **D.** qPCR results from the dCas9-KRAB transcriptional repression experiment. The graphs show the relative gene expression of *RGL1* (left), *COLGALT2* (middle) and *TSEN15* (right). The gene expression levels of each of the three genes are normalised to those measured in the non-targeting control cells (dashed line) and shown for all conditions: TC28a2/dCas9-KRAB cells transfected individually with gRNAs 1-6, cells transfected with all gRNAs, and cells transfected with gRNAs 1, 3 and 5. The bars show the mean of six biologic replicates and the error bars represent the SEM. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (C. and D.).

Since the VP64 and KRAB experiments had no effects on the three genes, Cas9 and paired gRNAs (gRNA7 and gRNA8) (Figure 3.11A) were used to delete 483bp of the TC28a2 genome containing the DMR (Figure 3.11B). The deletion resulted in a 2.7-fold reduction in *COLGALT2* expression ($P=0.0004$, Figure 3.11C). No significant difference in expression of *RGL1* ($P=0.16$) or *TSEN15* ($P=0.47$) was detected (Figure 3.11C). To ensure that the intronic deletion did not affect *COLGALT2* splicing, PCR amplification of cDNA from control (C) and edited (D) cells was performed using primers spanning the deletion. No aberrant splicing was observed (Figure 3.11D).

To determine if reduction in *COLGALT2* results in compensatory increase in *COLGALT1*, the expression of *COLGALT1* in the TC28a2 edited cells was measured and no change in the expression of the gene following deletion of the *COLGALT2* enhancer ($P=0.45$) was observed (Figure 3.11E).

The results from the experiment confirmed *COLGALT2* as a target gene of the enhancer containing the DMR.

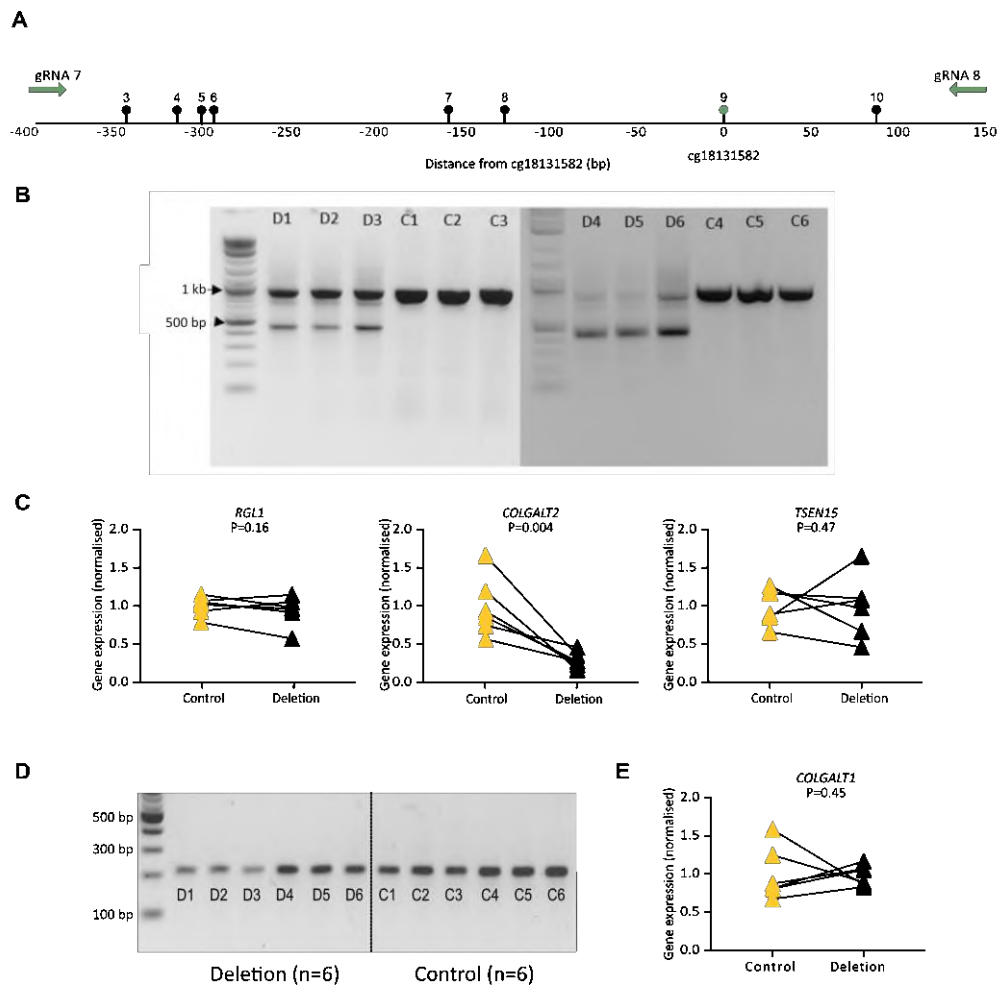


Figure 3.11. Effect of deletion of the enhancer on gene expression. **A.** Schematic diagram showing the genomic position of the two gRNAs (arrows) used in combination with Cas9 to edit the genome of TC28a2 cells (gRNAs 1– 2), relative to the 8 CpGs of the DMR represented by circles with cg18131582 being CpG 9 (green circle). The image was created in Affinity Designer 2.1. **B.** Gel images showing the successful deletion of the targeted region. Lane 1 is a 1kb Plus DNA Ladder. The control non-edited samples are marked with C (1-6) and the samples from the edited cells are marked with D (1-6). Samples 1-3 (D1-3 and C1-3) were run on one gel (left) and samples 4-6 (D4-6 and C4-6) were run on another gel (right). **C.** The relative gene expression of RGL1 (left), COLGALT2 (middle) and TSEN15 (right) following the deletion of the enhancer. The gene expression levels of each of the three genes following the deletion (black triangles) are normalised to those measured in the non-edited cells (yellow triangles). P-values were calculated using paired t-tests. **D.** PCR amplification of cDNA spanning the enhancer region showing that the deletion had not affected splicing. cDNA was reverse transcribed from RNA extracted from Tc28a2 cells with the enhancer deleted (D1-D6) (expected band size: 425bp if aberrant splicing, and 211bp if normal splicing) and non-edited controls (C1-C6) (expected band size: 908bp if aberrant splicing, and 211bp if normal splicing). **E.** Gene expression of COLGALT1 following deletion of the enhancer. The expression levels of the gene in cells harbouring the deletion (black triangles) have been normalised to those measured in the non-edited control cells (yellow triangles) (n=6). P-values were calculated using paired t-tests. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (C. and E.).

3.3.7. Modulation of the DNAm levels at the enhancer in TC28a2 cells

The impact of DNAm at the DMR upon *COLGALT2* expression was next investigated. To achieve that, dCas9 coupled with DNMT3a, to methylate the CpGs, or -TET1, to demethylate the CpGs, was used. The six gRNAs (gRNAs 1-6) used in the dCas9-VP64/KRAB experiment were expressed individually or in combination (all gRNAs, gRNAs 1,3,5) in cells along with the dCas9-DNMT3a/TET1 constructs.

Co-expression of dCas9-DNMT3a and gRNAs successfully increased DNAm at the CpGs (Figure 3.12 and Figure 3.13A, Panel 1). The greatest increases in DNAm were achieved at CpGs 8-10, at which gRNAs increased methylation by up to 38% (gRNA 1), 37.5% (gRNA 6), and 23% (gRNA 4), respectively (Figure 3.12). No further increase in DNAm was achieved by expression of multiple gRNAs (Figure 3.12). A significant reduction in *COLGALT2* expression was measured following co-expression of dCas9-DNMT3a with gRNA 1 (0.58-fold, $P=0.002$), gRNA 2 (0.56-fold, $P=0.0003$), gRNA 4 (0.68-fold, $P=0.009$), and gRNA 6 (0.62-fold, $P=0.008$) (Figure 3.13A, Panel 2). These 4 gRNAs all had an impact on DNAm at CpGs 8-10. Increasing DNAm at CpGs 9 and 10 alone (gRNA 5) did not impact the expression of *COLGALT2* (Figure 3.13, right panels). In all instances, no significant changes ($P>0.05$) in expression of *RGL1* and *TSEN15* were measured (Figure 3.13A, Panel 2).

Reducing enhancer DNAm in the chondrocytes was achieved using dCas9-TET1 (Figure 3.12 and Figure 3.13B, Panel 1). DNAm at CpGs 8-10 was reduced by up to 14%, 22.5% (both gRNA 2), and 24% (gRNA 4), respectively (Figure 3.12). Decreasing enhancer DNAm resulted in an increase in *COLGALT2* expression (Figure 3.13B, Panel 2). Significant increase in *COLGALT2* expression resulted when gRNAs 1 (1.16-fold, $P=0.03$), 2 (1.51-fold, $P=0.02$), 4 (1.54-fold, $P=0.01$), and 6 (1.62-fold, $P=0.02$) were used. The greatest increase in *COLGALT2* was achieved following expression of gRNA 6, which decreased mean DNAm at CpGs 8 and 10 by 11.8% and 18.8%, respectively (Figure 3.12). No significant changes ($P>0.5$) in *RGL1* or *TSEN15* expression were observed (Figure 3.13B, Panel 2).

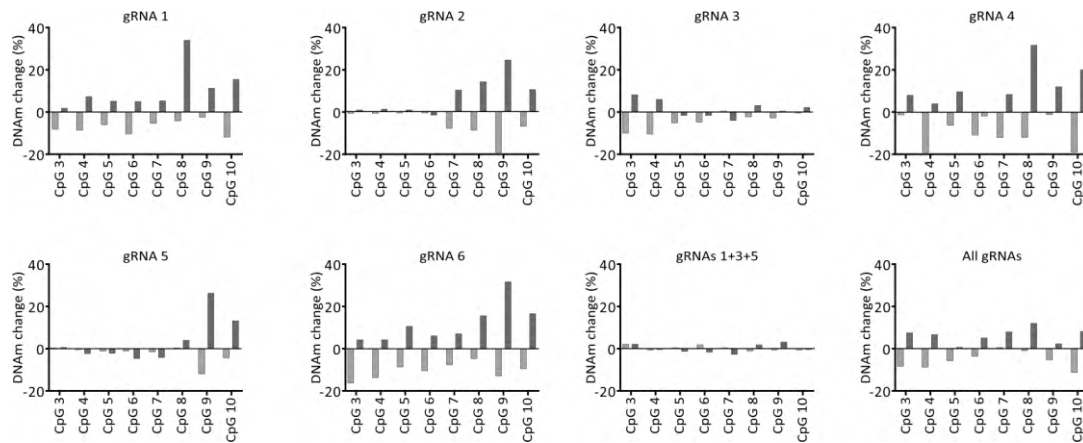


Figure 3.12. Percentage change in DNAm following targeted epigenome modulation of the DMR. Each plot shows the change in DNAm following epigenetic modulation with the gRNAs (gRNA1-6) or with combinations of the gRNAs (gRNA 1, 3, and 5 or with all gRNAs). The bars show the percentage change in DNAm (compared to the control cells) following modulation with dCas9-DNMT3a (red bars, n=3) or dCas9-TET1 (blue bars, n=3) at each of the 8 targeted CpGs (CpG 3- CpG 10). GraphPad Prism 9.0 was used for data visualisation.

Figure 3.13. Epigenetic modulation of the enhancer in TC28a2 cells. **A. Panel 1**, DNAm levels at the 8 CpGs in TC8a2 chondrocytes following expression of the dCas9-DNMT3a fusion protein in controls (non-targeting gRNA, black triangles) or in samples with a targeting gRNA (coloured triangles). Each dot represents a biologic replicate. **Panel 2**, Effect of modulation on gene expression of RGL1 (left), COLGALT2 (middle) and TSEN15 (right). Values measured in the modulated cells (coloured triangles) were normalized to those measured in the control cells (black triangles). **B. Panel 1**, DNAm levels at the 8 CpGs in TC8a2 chondrocytes following expression of the dCas9-DTET1 fusion protein in controls (non-targeting gRNA, black triangles) or in samples with a targeting gRNA (coloured triangles). Each dot represents a biologic replicate. **Panel 2**, Effect of modulation on gene expression of RGL1 (left), COLGALT2 (middle) and TSEN15 (right). Values measured in the modulated cells (coloured triangles) were normalized to those measured in the control cells (black triangles). P-values were calculated using paired t-tests. * = $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.001$. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A. and B.).

In both experiments, targeted changes in DNAm at CpGs 8 and 10 resulted in changes in *COLGALT2* expression, highlighting their importance in the regulatory function of the enhancer.

3.3.8. Identification of *COLGALT2* allelic imbalance in arthroplasty patient samples

In light of the findings in TC28a2 cells, the focus was returned to patient samples to investigate the effect of genotype at rs11583641 upon allelic expression of *COLGALT2*. The pyrosequencing assay used for the AEI analysis was the same one used for the genotyping at rs11583641. However, prior to the analysis it was necessary to confirm that the assay measures accurately the levels of the C and the T alleles of the SNP. An assay validation experiment was performed, and the assay proved to be correctly measuring the C and the T alleles in different ratios (Figure 3.14).

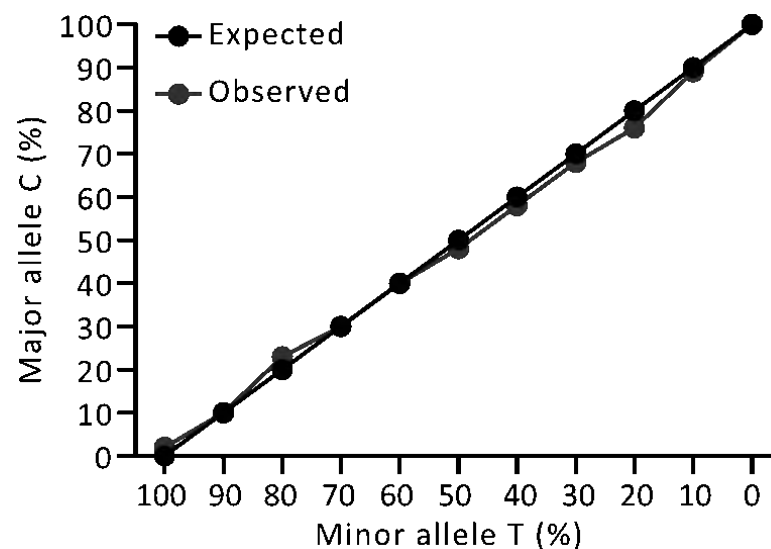


Figure 3.14. Accuracy of the AEI assay used. The measured levels of the major allele (C) (y-axis) and the minor allele (T) (x-axis) (red dots) compared to the expected values (black dots). GraphPad Prism 9.0 was used for data visualisation.

The AEI analysis was carried out in arthroplasty patient samples heterozygous for rs11583641 and an imbalance between the C and T transcripts of *COLAGLT2* was detected (Figure 3.15A). A 1.94-fold mean increase ($P < 0.0001$) in expression of the OA-risk allele, C,

was observed. Interestingly, AEI was detected in 13/27 samples used for analysis, yet in the remaining 14 patients no imbalance was measured (Figure 3.15B). This observation could not be explained by sex, age, joint (knee or hip), or genotype at any of the other SNPs forming part of rs11583641 LD block (Table 3.2).

For the patient samples with significant AEI, allelic ratios were correlated with DNAm at the enhancer CpGs (Figure 3.15C). Significant correlations, marking methylation-expression QTLs (meQTLs) were discovered at CpGs 4, 8 and 9 ($P=0.04$, $P=0.03$ and $P=0.04$, respectively) and a trend was observed at CpG 7 ($P=0.07$). These data suggested that in OA patient cartilage the regulation of *COLGALT2* expression is subject to both genetic and epigenetic factors.

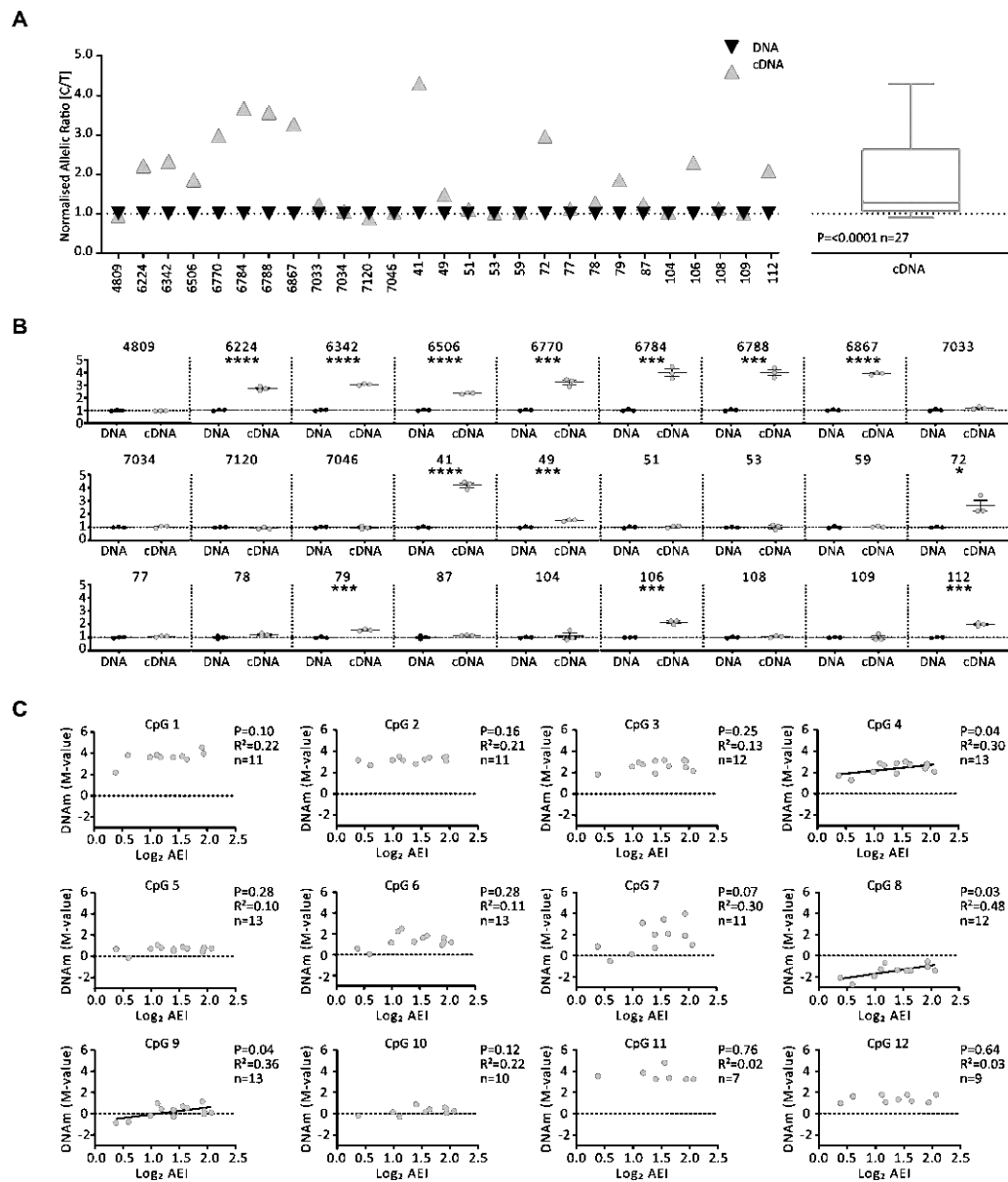


Figure 3.15. AEI analysis of COLGALT2 in arthroplasty cartilage samples. **A.** Left, Allelic (C/T) ratios in cartilage samples from OA patients heterozygous for rs11583641. In each sample, the ratio of values for cDNA and DNA between the OA-risk allele C and the non-risk T allele was plotted; each symbol represents the mean of 3 technical repeats. Right, Mean DNA and cDNA values in the presence of the C allele versus the T allele in heterozygous patients. Values are shown as a box plot, with the line inside the box representing the median, the box showing the interquartile range, and whiskers showing the minimum and maximum values. P-value was calculated using Wilcoxon's matched pairs signed rank test. **B.** Plots comparing the allelic ratios (C/T) measured in cDNA and DNA for each patient. P-values were calculated using paired t-tests. * = $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.001$; **** = $P < 0.0001$ **C.** Relationship between COLGALT2 AEI ratios and DNAm at the enhancer in cartilage. Allelic ratios ($\log_2$) of COLGALT2 (x-axis) are plotted against matched DNAm levels (M-values) (y-axis) at the investigated CpGs (CpG 1-12). Each dot represents data from 1 individual. P-values were calculated using simple linear regression (P-value, R^2 and number of samples used for the

analysis). GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A., B. and C.).

Table 3.2. Details of the patient samples used for the AEI analysis.

Sample	Significant AEI	Sex	Joint	Age	Genotype				
					Group 1	Group 2	Group 3	Group 4	Group 5
4809	NO	Male	Knee	79	1	1	1	1	1
6224	YES	Female	Knee	62	1	1	1	1	2
6342	YES	Female	Knee	62	1	1	1	1	1
6506	YES	Female	Knee	59	1	1	1	1	1
6770	YES	Female	Knee	80	1	1	1	1	1
6784	YES	Female	Knee	46	1	1	1	1	1
6788	YES	Male	Knee	66	1	1	1	1	2
6867	NO	Male	Knee	86	1	1	1	1	1
7033	NO	Female	Hip	81	1	1	2	1	1
7034	NO	Female	Hip	60	1	1	1	1	1
7046	NO	Female	Hip	56	1	1	1	1	1
7120	NO	Female	Knee	61	1	1	1	1	1
41	YES	Male	Knee	51	1	1	1	1	1
49	YES	Male	Hip	68	1	1	1	1	1
51	NO	Female	Hip	74	1	1	1	1	2
53	NO	Female	Hip	55	1	1	1	1	1
59	NO	Female	Knee	60	1	1	1	1	1
72	YES	Male	Knee	76	1	0	1	1	1
77	NO	Female	Knee	71	1	1	1	1	1
78	NO	Male	Knee	58	1	1	1	1	1
79	YES	Male	Knee	65	1	1	1	1	1
87	NO	Female	Hip	51	1	1	1	1	0
104	NO	Female	Knee	70	1	1	1	1	1
106	NO	Male	Knee	79	1	1	1	1	1
108	NO	Male	Knee	72	1	1	1	1	1
109	NO	Female	Knee	65	1	1	2	2	1
112	YES	Male	Hip	61	1	1	1	1	1

AEI – allelic expression imbalance

3.4. Discussion

The era of GWAS has extended over a decade and has led to the discovery of more than 100 independent genetic signals associated with OA⁴. However, it has been difficult to interpret these findings in a biological context and translate them into effective therapies. Consequently, researchers are increasingly utilizing epigenetic datasets to prioritize causal genes and their regulatory elements for focused analysis⁵⁻¹¹.

The work presented here has been focused on a specific risk locus on chromosome 1, where previous research had shown a correlation between intronic DNA methylation (DNAm) and genetic risk for OA³. In patient arthroplasty cartilage DNA samples, DNAm at 12 specific

sites (CpGs 1-12) was measured and a differentially methylated region (DMR) spanning 210 base pairs was identified, encompassing the originally reported mQTL at cg18131582³. The effect allele (C) of the genetic variant rs11583641 was associated with reduced DNAm compared to the non-effect allele (T) at all differentially methylated sites. The strongest effect was observed in knee cartilage between CpG 8 and 10. Allele-specific expression analysis confirmed that the effect allele of the SNP correlated with increased expression of the *COLGALT2* gene. Moreover, correlations between allele-specific expression ratios and DNAm revealed a meQTL.

Using a reporter gene assay, the DMR was found to have regulatory function, which is significantly affected by DNAm. Furthermore, a genetic influence on enhancer function was detected with rs734657. Alleles of this SNP, which is located within the enhancer and is in high LD with rs11583641, associated with different levels of expression of the reporter gene. While these findings highlight a potential relationship between genetic and epigenetic factors in the regulation of gene expression, it is important to acknowledge that a direct causal link between SNP genotype and DNAm in patient samples cannot be determined, which remains a limitation of this type of study. While the reporter gene analysis was informative, there are a number of limitations to this type of analysis, including the fact that regions of interest are not investigated in the context of the cells' chromatin environment and that the gene target of the regulatory function cannot be established. In order to understand the impact of the enhancer in the genomic environment within the TC28a2 chondrocytes, dCas9-VP64 and dCas9-KRAB editing complexes were used to activate and repress its function. This method does not cause and damage to the underlying DNA targeted in the form of DNA breaks and so the outcomes of the experiment could only be attributed to the activation/silencing of the enhancer. Furthermore, since the expression of the complexes is temporary, the process of activation/silencing is reversible. Versions of these constructs coupled with DNAm editing proteins (TET1 and DNMT3a) have been shown to regulate gene expression in a reversible manner, without affecting the DNA stability of the cells¹². This type of regulatory region activation/silencing did not work at this locus in the experiment described in section 3.3.6. In order to identify the gene target of the enhancer a genome editing approach was instead used.

Deletion of the enhancer using CRISPR-Cas9, confirmed *COLGALT2* as a target gene of the OA signal marked by rs11583641. Furthermore, the use of a catalytically dead Cas9 fused to enzymes that methylate (DNMT3a) or demethylate (TET1) CpGs demonstrated a causal relationship between DNAm and *COLGALT2* expression. Consistent with the measurements in patient samples, a reduction in DNAm led to increased *COLGALT2* expression and *vice versa*. Additionally, these investigations confirmed that the region surrounding CpGs 8, 9, and 10 in particular could have an important regulatory role. The signal marked by rs11583641 is one of not many instances where a causal gene has been identified at an OA risk locus via direct molecular investigation^{6,9-11}. It confirms the effectiveness of employing epigenome editing techniques to unravel functional risks associated with complex genetic diseases. This is important as it sheds light on a new potential therapeutic target for intervening in the disease, focusing on both the epigenome and the enzyme coded for by *COLGALT2*, which is a strong candidate in the development of OA due to its role in collagen PTM regulation.

3.5. References

1. [review] Do, C. et al. Genetic–epigenetic interactions in cis: a major focus in the post-GWAS era. *Genome Biol.* **18**, 120 (2017).
2. 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).
3. 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. *Osteoarthr. Cartil.* **27**, 1545–1556 (2019).
4. Boer, C. G. *et al.* Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell* **184**, 4784-4818.e17 (2021).
5. 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).

6. Parker, E. *et al.* Multi-Tissue Epigenetic and Gene Expression Analysis Combined with Epigenome Modulation Identifies RWDD2B as a Target of Osteoarthritis Susceptibility. *Arthritis Rheumatol.* **73**, 100–109 (2021).
7. Sorial, A. K. *et al.* Multi-tissue epigenetic analysis of the osteoarthritis susceptibility locus mapping to the plectin gene PLEC. *Osteoarthr. Cartil.* **28**, 1448–1458 (2020).
8. 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).
9. Rice, S. J. *et al.* Genetic and Epigenetic Fine-Tuning of TGFB1 Expression Within the Human Osteoarthritic Joint. *Arthritis Rheumatol.* **73**, 1866–1877 (2021).
10. Klein, J. C. *et al.* Functional testing of thousands of osteoarthritis-associated variants for regulatory activity. *Nat. Commun.* **10**, 2434 (2019).
11. 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).
12. Nuñez, J. K. *et al.* Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* **184**, 2503–2519.e17 (2021).

Chapter 4. Independent Osteoarthritis Risk-Confering Alleles Mediate the Same Epigenetic and Transcriptional Effect on COLGALT2

4.1. Introduction

In 2020 a second association signal that maps close to *COLGALT2* was reported in a GWAS including 177517 OA subjects and 649173 control subjects¹. SNPs highlighted by the study included rs12047271 ($P=3.96 \times 10^{-12}$) and rs1327123 ($P=2.44 \times 10^{-16}$), both residing within the intergenic region between *COLGALT2* and *TSEN15*, and rs1046934 ($P=3.81 \times 10^{-18}$), located within *TSEN15*. These SNPs were part of an LD block containing a total of 21 SNPs ($r^2 \geq 0.8$, GBR; British in England and Scotland). Two of these 21 SNPs coded for missense substitutions: rs1046934 (A>C, Gln59His) and rs2274432 (G>A, Gly19Asp), both located in *TSEN15*. This gene encodes a subunit of human tRNA splicing endonuclease (TSEN) complex responsible for the splicing of introns from eukaryotic pre-tRNAs. The complex is composed of TSEN15 and TSEN34, comprising the structural subunits, and TSEN2 and TSEN54 which form the catalytic domain^{2,3}.

A mQTL analysis performed by Dr Sarah Rice on genome wide CpG methylation data from arthroplasty cartilage samples ($n=87$) included an interval of 0.2Mbp surrounding rs1046934 (see Appendix C). mQTLs were detected at three CpG sites within this interval: cg21606956 ($P=0.002$), cg15204595 ($P=0.005$) and cg01436608 ($P=0.04$). At all three CpGs, the OA risk-conferring allele A of rs1046934 was associated with reduced methylation (Figure 4.1). Two of the CpG sites, cg15204595 and cg01436608, were located within the gene body of *COLGALT2*, whilst cg21606956 was in an intergenic region over 200kb from the other two CpGs.

In this chapter I describe the results of a functional analysis of this new OA signal (the rs1046934 signal). rs1046934 is in near perfect linkage equilibrium ($r^2 = 0.099$, GBR) with the rs11583641 locus discussed in Chapter 3, confirming that these two loci are genetically independent.

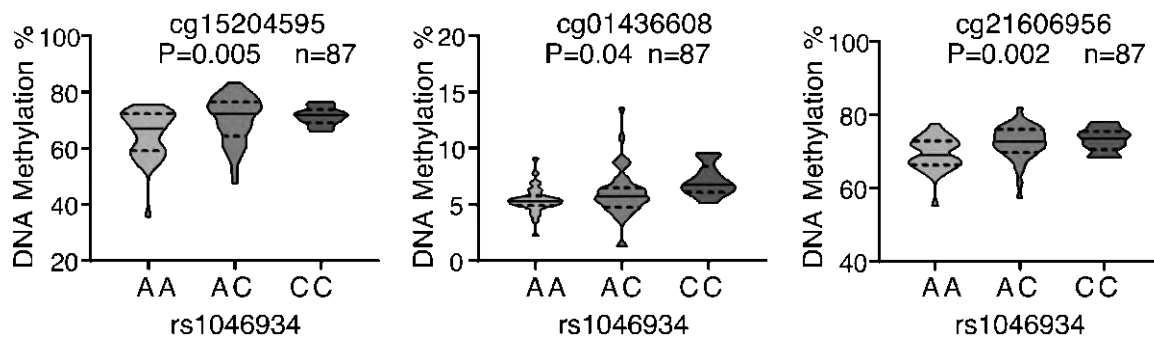


Figure 4.1. mQTLs at the rs1046934 locus. Violin plots showing DNAm values (y-axis) at cg15204595 (left plot), cg01436608 (middle plot), and cg21606956 (right plot) stratified by rs1046934 genotype (x-axis), using hip or knee arthroplasty cartilage DNA samples from OA patients ($n = 87$). Solid lines and dashed lines inside the plots represent median and interquartile range, respectively. P-values were calculated using linear regression. GraphPad Prism 9.0. was used for data visualisation and statistical analysis was performed in R-Studio.

Based on the position of the LD block containing rs1046934 and the results of the mQTL analysis, my hypothesis was that rs1046934 predisposed to the development of OA by regulating the expression of *COLGALT2* in cartilage through differential methylation at three CpG sites located in regulatory regions (cg21606956, cg15204595 and cg01436608), similarly to but independently of rs11583641 (see Chapter 3).

4.2. Aims and Objectives

4.2.1. Aims

- Determine if the missense variant rs1046934 and/or rs2274432 are likely to impact the function of TSEN15 protein.
- Determine if *COLGALT2* is the target of the OA susceptibility locus marked by rs1046934.
- Determine if the regions containing cg21606956, cg15204595 and cg01436608 have regulatory function in chondrocytes.

- Determine if the regulatory function (if any) of the regions containing cg21606956, cg15204595 and cg01436608 is dependent on the DNAm status of the CpGs.
- Identify the target gene(s) of rs1046943 and of the regions containing cg21606956, cg15204595 and cg01436608.
- Investigate the role of DNAm levels at cg21606956, cg15204595 and cg01436608 in regulating the expression of the target genes.

4.2.2. Objectives

- Perform *in silico* analyses to determine if the missense substitutions caused by rs1046934 and rs2274432 are likely to impact TSEN15 protein function.
- Perform AEI analysis of *COLGALT2* and *TSEN15* in arthroplasty patient cartilage samples heterozygote at rs1046934.
- Replicate the mQTLs at cg21606956, cg15204595 and cg01436608 in arthroplasty cartilage DNA.
- Clone the regions containing cg21606956 and cg15204595 in CpG-free Lucia reporter gene vectors, methylate the constructs *in vitro* and measure Lucia readings.
- Methylate and demethylate the regions containing cg21606956 and cg15204595 using dCas9/DNMT3a or dCas9-TET1 fusion proteins, respectively, and measure the expression of *COLGALT2* and *TSEN15*, following the modulation.

4.3. Results

4.3.1. Analysis of missense variants

The Protein Data Bank⁴ was used to investigate the likelihood of the missense variants rs1046934 (Gln59His) and rs2274432 (Gly19Asp) impacting the function of TSEN15 protein. As mentioned, this protein is a subunit of a heterotetrameric

endonuclease which is responsible for the splicing of introns from eukaryotic pre-tRNAs. A heterodimer formed by TSEN15 and TSEN34 comprises the structural unit of the TSEN complex (Figure 4.2A).

The Gln59 residue of TSEN15 falls within the $\alpha 2$ helix (Figure 4.2B). *In silico* mutagenesis of the residue carried out by Dr Sarah Rice predicts an outward facing position of the aromatic histidine side chain, away from the coiled-coil interactions between the $\alpha 1$ and $\alpha 2$ helices (Figure 4.2C), indicating that the missense variant is unlikely to affect the structure and/or stability of the protein or the TSEN complex. An *in silico* mutagenesis of the Gly19 residue could not be undertaken because the Gly19Asp variant resides within the structurally unresolved N-terminal region of *TSEN15*.

Genotype at both SNPs correlates with expression of *COLGALT2* and *TSEN15* in a range of non-joint related tissues from the Genotype-Tissue Expression (GTEx) portal, forming eQTLs³. Additionally, both rs1046934 and rs2274432 were reported as benign changes in the genomAD database⁵.

Based on these data, it is unlikely that the rs1046934 locus affects protein function and more likely that it effects gene expression.

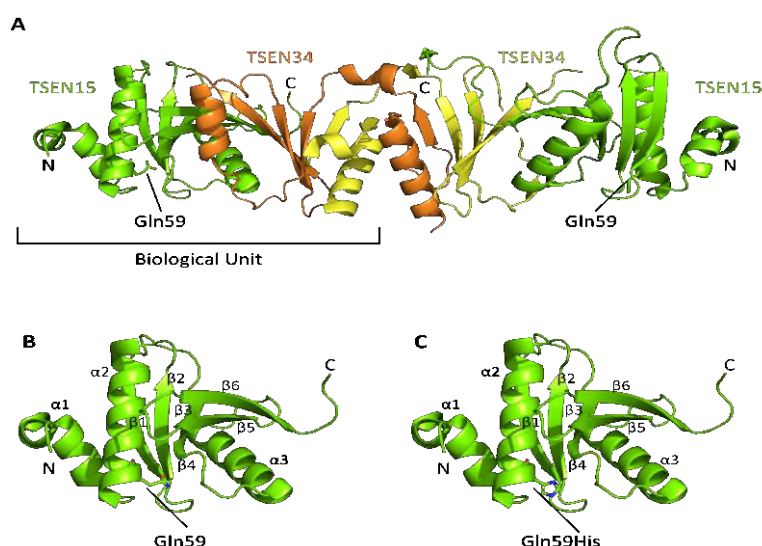


Figure 4.2. Protein structure of the structural unit of the eucaryotic tRNA splicing (TSEN) complex. **A.** Crystal structure of the TSEN15 (green)–TSEN34 (yellow and orange) heterodimer

(Protein Data Bank, identification no. 6Z9U). The position of the Gln59 residue is highlighted, and the amino acid side chain is displayed. **B.** Monomeric crystal structure of TSEN15 (Protein Data Bank, identification no. 2GW6), showing numbering of the α - β - β - β - α - β - β fold (α 1–3 and β 1–6). Gln59 is labelled, and the side chain is displayed. **C.** TSEN15 structure, shown as in B, following *in silico* mutagenesis to predict the conformation of Gln59-His (labelled). In B and C, red shows oxygen atoms and blue shows nitrogen atoms. The protein structures were examined using PyMOL Molecular Graphics System by Schrödinger (A., B., and C.). The *in silico* mutagenesis was carried out using the PyMOL Mutagenesis Wizard (C.).

4.3.2. Identification of COLGALT2 and TSEN15 allelic expression imbalance

AEI analysis was used to determine if rs1046934 forms eQTLs for *COLGALT2* and/or *TSEN15* in arthroplasty patient cartilage, an OA relevant tissue. The analysis was performed using coding SNPs in high LD with rs1046934: for *COLGALT2* the 5'UTR transcript SNP rs114661926 (C>G, $r^2= 0.86$, GBR) was used, for *TSEN15* the missense variant rs2274432 (G>A, $r^2= 1.0$, GBR) was used. The analysis was performed in DNA and matched cDNA from compound heterozygote patients at rs1046934 and the respective transcript SNPs using pyrosequencing. The two assays were validated prior to the experiment to show that they could reliably measure the alleles of the SNPs in any ratio (Figure 4.3).

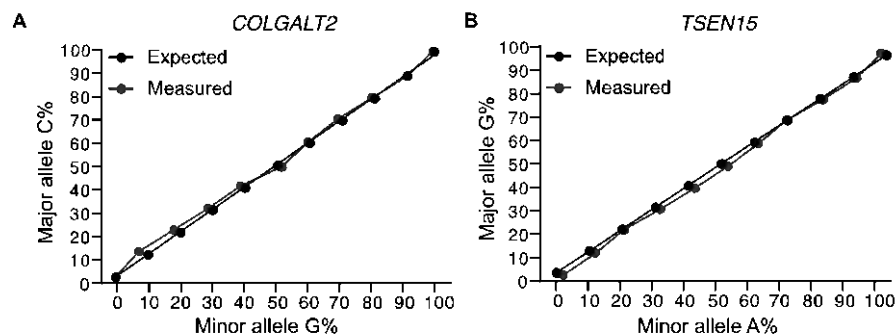


Figure 4.3. Accuracy of the AEI assays used. **A.** The measured levels of the major allele (C) of rs114661926 (y-axis) and the minor allele (G) (x-axis) (red dots) compared to the expected values (black dots). **B.** Same as A. for rs2274432. GraphPad Prism 9.0. was used for data visualisation.

AEI was identified for both genes with the OA risk allele C of rs114661926 showing an average 1.21-fold increase in *COLGALT2* expression compared to the non-risk G allele

($P=0.003$) (Figure 4.4A), and the OA risk allele G of rs2274432 showing an average 1.09-fold increase in *TSEN15* expression compared to the non-risk A allele ($P=0.02$) (Figure 4.4B).

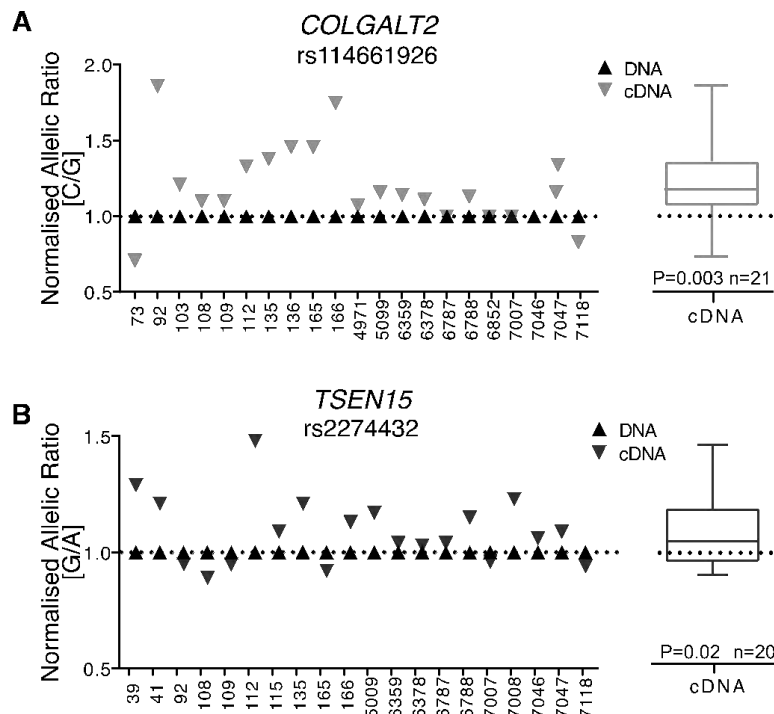


Figure 4.4. AEI analysis of *COLGALT2* and *TSEN15* in arthroplasty cartilage samples. **A.** Left, Allelic (C/G) ratios in patient samples heterozygous for *COLGALT2* transcript SNP rs11466192. In each sample, the ratio of values for cDNA and DNA between the OA-risk allele and the non-risk allele was plotted; each symbol represents the mean of 3 technical repeats. Right, Mean DNA and cDNA values in the presence of the risk allele versus the non-risk allele in heterozygous patients. Values are shown as a box plot, with the line inside the box representing the median, the box showing the interquartile range, and whiskers showing the minimum and maximum values. P -values were calculated using Wilcoxon's matched pairs signed rank test. **B.** Same as A. for *TSEN15* SNP rs2274432 (G/A; risk/non-risk). GraphPad Prism 9.0. was used for statistical analyses and data visualisation (A. and B.).

4.3.3. In silico analysis of the rs1046934 locus and replication of cg15204595 and cg21606956 cartilage DNA mQTLs

The region containing rs1046934 and the differentially methylated CpGs (cg15204595, cg01436608 and cg21606956) was investigated using publicly available genomic databases and online tools listed in Chapter 2, Table 2.8 (Figure 4.5A). rs1046934 and the 20 SNPs in LD with it are part of a 30kb block that encompasses the 5'UTR and promoter of *COLGALT2*, the promoter and part of the gene body of *TSEN15*, and the intergenic region between the two

genes (Figure 4.5A, panels 1 and 2). cg15204595 and cg01436608 are 2.35kb apart and located within intron 1 of *COLGALT2* (Figure 4.5A, panels 1-3). cg15204595 and cg01436608 reside within a region that is marked as an enhancer and a transcriptionally active site in OA-relevant cells (Figure 4.5A, panel 4). The region is also marked as open chromatin in OA hip and knee chondrocytes (Figure 4.5A, panel 5). cg21606956 is located at over 200kb from cg15204595 and cg01436608 in a distant intergenic enhancer which is also transcriptionally accessible in OA chondrocytes from the hip and the knee joint (Figure 4.5A, panel 1-5). MSC capture Hi-C data showed physical interactions between the region encompassing rs1046934 and the enhancer containing cg15204595 (Figure 4.5A, panel 6). Additional interactions were observed between the *COLGALT2* promoter and the enhancer containing cg21606956 (Figure 4.5A, panel 6).

The mQTLs at this locus were replicated in an independent cohort of arthroplasty cartilage DNA samples for cg15204595 ($P < 0.0001$) and cg21606956 ($P < 0.0001$) (Figure 4.5B). The pyrosequencing assay designed to capture cg15204595, measured also the DNAm levels at another CpG site, cg15204595-142. That is the only other CpG site within 1kb (500bp up- and down-stream) of cg15204595. No mQTL was detected at cg15204595-142 ($P = 0.66$) (Figure 4.5B). Due to the sequence surrounding cg01436608, a pyrosequencing assay that reliably measure the DNAm levels at the CpG site could not be designed and the mQTL was not replicated. Further analyses were, therefore, focused on cg15204595 and cg21606956.

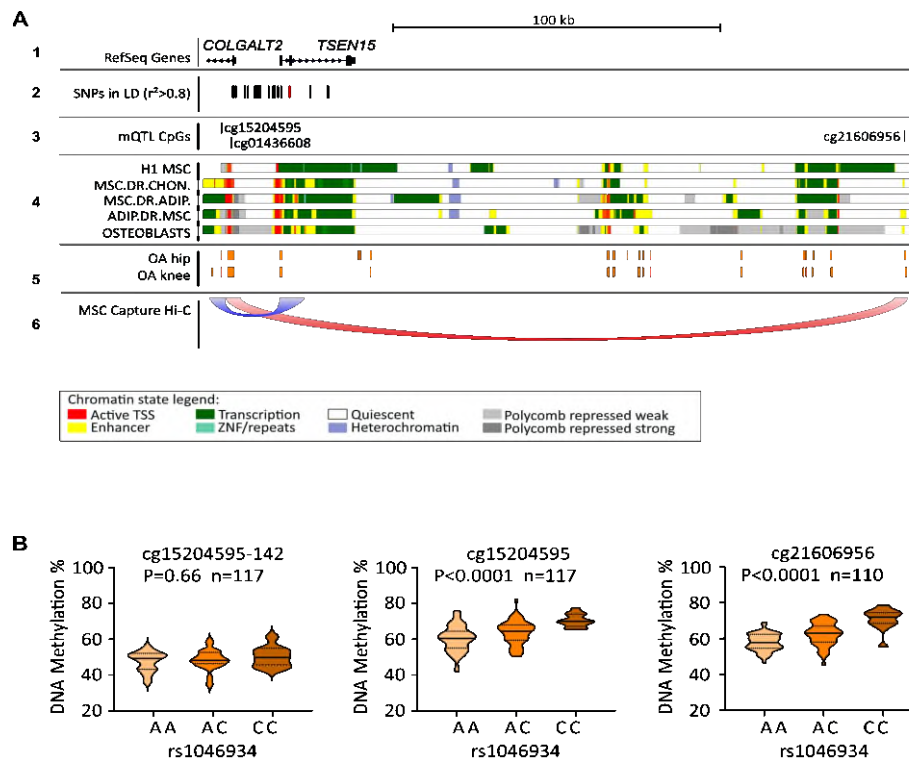


Figure 4.5. Overview of the rs1046934 locus and mQTL replication in arthroplasty cartilage samples. **A.** Schematic overview of the rs1046934 locus. **Panel 1**, Relative genomic position of the 5' end of COLGALT2 and all of TSEN15, visualized using the UCSC Genome Browser (hg19). **Panel 2**, The positions of rs1046934 (red line), and the SNPs in high LD (pairwise $r^2 > 0.8$) with rs1046934 (black lines), are indicated. **Panel 3**, The positions of cg15204595, cg01436608, and cg21606956 (black lines) are indicated. **Panel 4**, Chromatin state data, as determined from the Roadmap project, for primary human mesenchymal stem cells (H1 MSC), MSC-derived chondrocytes (MSC.DR.CHON) and adipocytes (MSC.DR.ADIP), adipose-derived MSCs (ADIP.DR.MSC), and human osteoblasts. Colours correspond to different chromatin states, as indicated in legend at bottom. **Panel 5**, Open regions marked by ATAC-sequencing peaks generated from OA hip and knee chondrocytes (orange blocks). **Panel 6**, Loops indicate capture Hi-C chromatin interactions from the 3D Genome Browser in human MSCs, with the flat ends of loops spanning the width of the interacting regions. **B.** Violin plots showing DNAm values in arthroplasty cartilage samples at cg15204595-142 (left), cg15204595 (middle), and cg21606956 (right), stratified by rs1046934 genotype. Solid lines and dashed lines inside the plots represent median and interquartile range, respectively. P-values were calculated using simple linear regression. GraphPad Prism 9.0. was used for statistical analyses and data visualisation (B.).

The arthroplasty cartilage DNAs used in the replication were derived from OA hip, OA knee and NOF patients. This allowed investigating the effects of joint site and disease status upon the DNAm at cg15204595 and cg21606956. The DNAm data were first compared between OA and NOF patients (Figure 4.6A). The mQTLs were observed in both OA (cg15204595 – $P < 0.001$; cg21606956 – $P < 0.0001$) and NOF (cg15204595 – $P = 0.002$;

cg21606956 – $P=0.001$), suggesting the effects are not disease-specific. DNAm levels at the two CpGs were then compared between OA hip, OA knee and NOF samples, irrespective of rs1046934 genotype (Figure 4.6B). Joint-specific differences were not observed for either CpG. In a comparison between the OA hip samples and the control NOF samples, cg15204595 showed a significantly higher level of methylation in NOF (mean DNAm - 66.6 %) relative to OA (mean DNAm - 61.9%) ($P = 0.004$). While the NOF patients were on average older than the OA patients at the time of surgery (NOF - 77.4 years (SD-9.7); OA knee - 65.3 years(SD-9.5); OA hip - 66.5 years(SD-13.1)), when the methylation data was stratified by age, no correlations with age for either CpG were observed (Figure 4.6C).

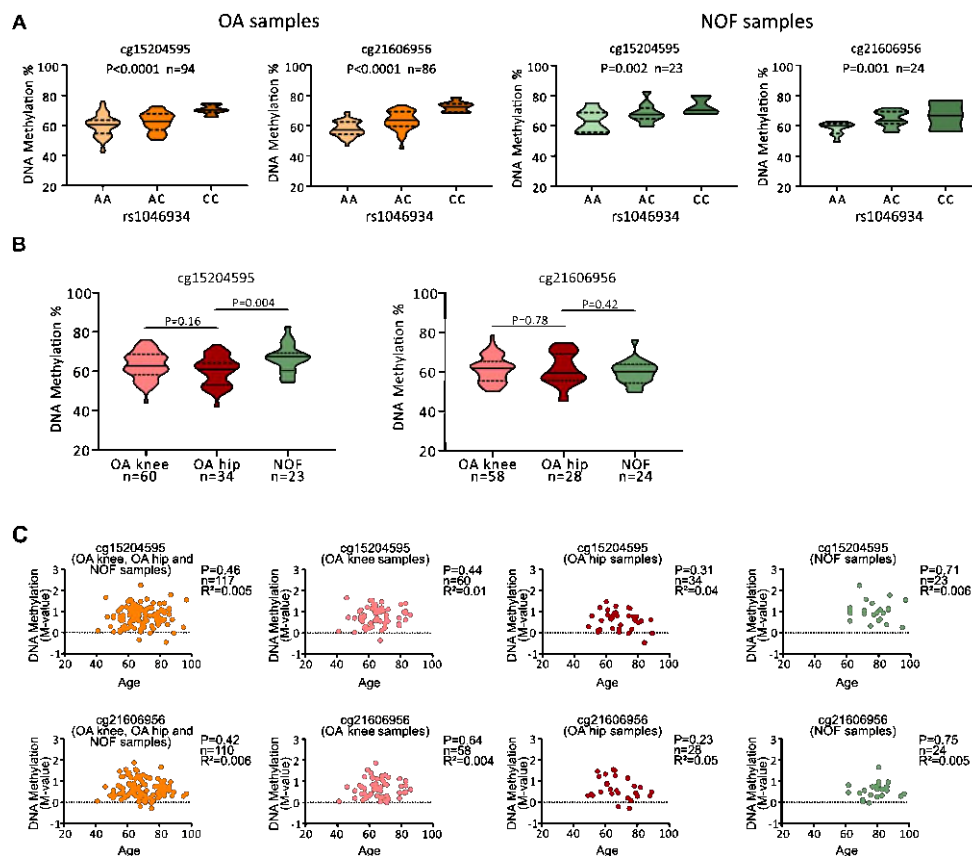


Figure 4.6. Stratification of DNAm levels at cg15204595 and cg21606956 against genotype at rs1046934, disease state (OA and NOF), joint (hip, knee) and age at surgery. **A.** DNAm values at cg15204595 and cg21606956 plotted against genotype at rs1046934, using DNA derived from cartilage from OA patients and NOF patients, are shown as violin plots. Solid lines and dashed lines inside the plots represent the median and interquartile range, respectively. P -values were calculated using simple linear regression. **B.** DNAm values at cg15204595 and cg21606956, presented as violin plots, measured in cartilage from OA knee, OA hip and NOF patients. P -values were calculated using a Mann-Whitney test. **C.** Graphs show DNAm levels at cg15204595 (top) and cg2160956 (bottom), measured in cartilage from OA knee, OA hip

and NOF patients, plotted against the age of the patients at the time of arthroplasty surgery in years. Each dot represents data from 1 individual. P-values were calculated using simple linear regression. GraphPad Prism 9.0. was used for statistical analyses and data visualisation (A., B., and C.).

4.3.4. Detection of methylation-expression QTLs (meQTLs) in arthroplasty cartilage

I moved forward by assessing whether methylation at cg15204595 and cg21606956 and expression of *COLGALT2* and/or *TSEN15* correlated in our patient samples. For both genes, the AEI allelic ratios were plotted against matched DNAm values. For *COLGALT2*, significant correlations were observed at cg15204595 ($P=0.004$) and cg21606956 ($P=0.005$), marking meQTLs (Figure 4.7A). Neither CpG showed significant correlations for *TSEN15* ($P=0.66$ and $P=0.15$, respectively) (Figure 4.7B).

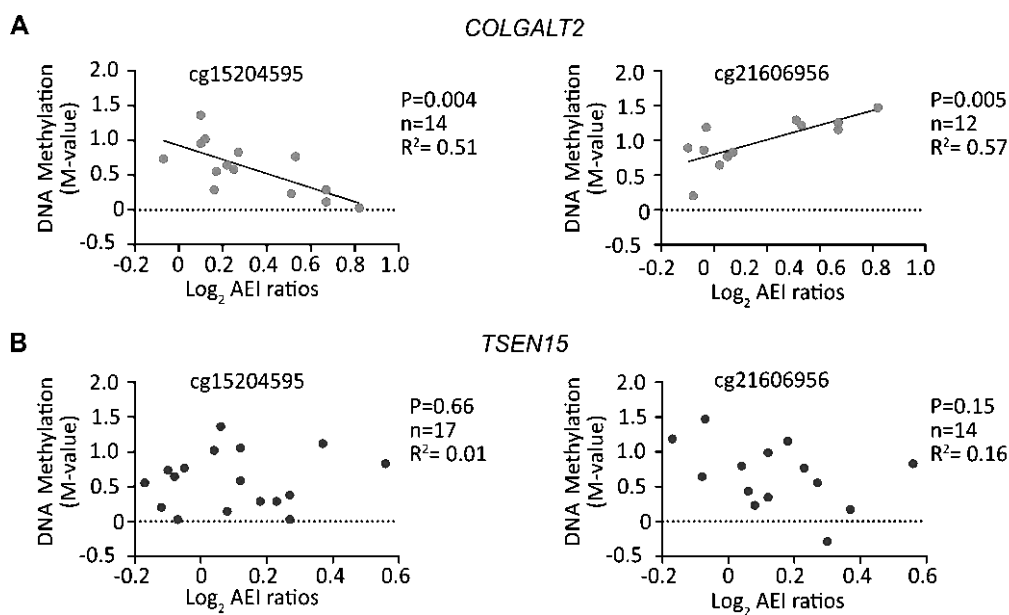


Figure 4.7. Relationship between DNAm at the enhancer and *COLGALT2* AEI ratios. **A.** Allelic ratios ($\log_2$) of *COLGALT2* (x-axis) are plotted against matched DNAm levels (M-values) (y-axis) at cg15204595 (left) and cg21606956 (right). Each dot represents data from 1 individual. P-values were calculated using simple linear regression and the test statistics are shown to the right of each plot (P-value, R^2 and number of samples used for the analysis). **B.** Same as A. for *TSEN15* (rs2274432). GraphPad Prism 9.0. was used for statistical analyses and data visualisation (A. and B.).

These results were particularly intriguing in light of the other observations in patient samples. The OA risk-conferring allele A of rs1046934 is associated with lower DNAm levels

at cg15204595 and cg21606956 compared to the non-risk allele C in arthroplasty samples (Figure 4.5B). Allele A is also associated with higher *COLGALT2* expression, as evidenced by the AEI analyses (Figure 4.4A). meQTL plots show correlations between methylation levels at both CpGs and *COLGALT2* AEI ratios (Figure 4.7A). However, while higher methylation levels at cg15204595 are associated with lower allelic ratios, the opposite is observed at cg21606956, with higher methylation levels corresponding to higher allelic ratios (Figure 4.7A).

This can be explained if the methylation levels at the two CpGs are influenced by factors other than rs1046934 genotype.

I propose a theoretical model where a regulatory SNP affects methylation levels at two CpG sites (CpG1 and CpG2), with its major allele (M) associated with low DNAm and its minor allele (m) with high DNAm at both CpG1 and CpG2. In this case, individuals homozygous for the major allele (MM) will have low DNAm levels, individuals homozygous for the minor allele (mm) will have high DNAm levels, and heterozygous individuals will have intermediate DNAm levels at both CpGs (Figure 4.8A and 4.8B). Two additional factors that affect the DNAm levels at CpG1 and CpG2 will be added in this theoretical model: Effect 1, which has a constant effect on the methylation at CpG1 leading to high levels of methylation at the CpG, and Effect 2, which has a constant effect on the methylation at CpG2 leading to low levels of methylation at the CpG (Figure 4.8A and 4.8B). If these two hypothetical effects are weaker than the effect the SNP has on the DNAm at CpG1 and CpG2, they will be masked in individuals homozygous at the SNP (MM or mm), individuals homozygous for the major allele (MM) will still have lower DNAm at the two CpGs compared to the DNAm levels measured in individuals homozygous for the minor allele (mm). However, in heterozygous individuals (Mm), the effects of Effect 1 and Effect 2 will not be masked: Effect 1 will complement the effect of the minor allele (m) of the SNP and the overall methylation levels at CpG1 will be higher than the expected mean but lower than (mm) homozygotes (Figure 4.8B, left); Effect 2 will complement the effect of the major allele (M) of the SNP and the overall methylation levels at CpG2 will be lower than the expected mean but higher than (MM) homozygotes (Figure 4.8B, right).

If both the effects of the SNP and Effect 1 or Effect 2 (for CpG1 and CpG2, respectively) are taken into consideration, similar trends will be expected for the mQTL plots for CpG1 and CpG2 (Figure 4.8C). If the methylation levels at CpG1 are plotted against the levels at CpG2

for all individuals irrespective of SNP genotype, low levels of methylation at one CpG will correspond to low levels at the other (Figure 4.8D, left). When the same plot is created for heterozygous individuals, however, the influence of Effect 1 and Effect 2, complementing the minor and the major alleles (respectively) of the SNP, will contribute to a reversed trend where low DNAm levels at one CpG will correspond to higher levels at the other (Figure 4.8D, right).

The theoretical scenarios described here are consistent with our actual observations in arthroplasty samples. When methylation levels at cg15204595 and cg21606956 are plotted against genotype at rs1046934 they show a similar trend in both arthroplasty samples (Figure 4.5B). When methylation levels at cg15204595 are plotted against the levels at cg21606956 for all arthroplasty samples (Figure 4.8E, left), low DNAm levels at one CpG correlate with low DNAm levels at the other. The trend is particularly strong when only rs1046934 homozygotes (AA and CC) are plotted (Figure 4.8E, middle). However, in heterozygous (AC) samples (Figure 4.8E, right) low DNAm at one CpG correlates with higher DNAm at the other.

The *COLGALT2* meQTL plots in Figure 4.7A show methylation levels at cg15204595 and cg21606956 plotted against AEI ratios in arthroplasty. Since the AEI analysis is carried out in heterozygous individuals in which lower DNAm levels at cg15204595 correspond to higher DNAm levels at cg21606956, the meQTL plots for the two CpGs have opposing slopes.

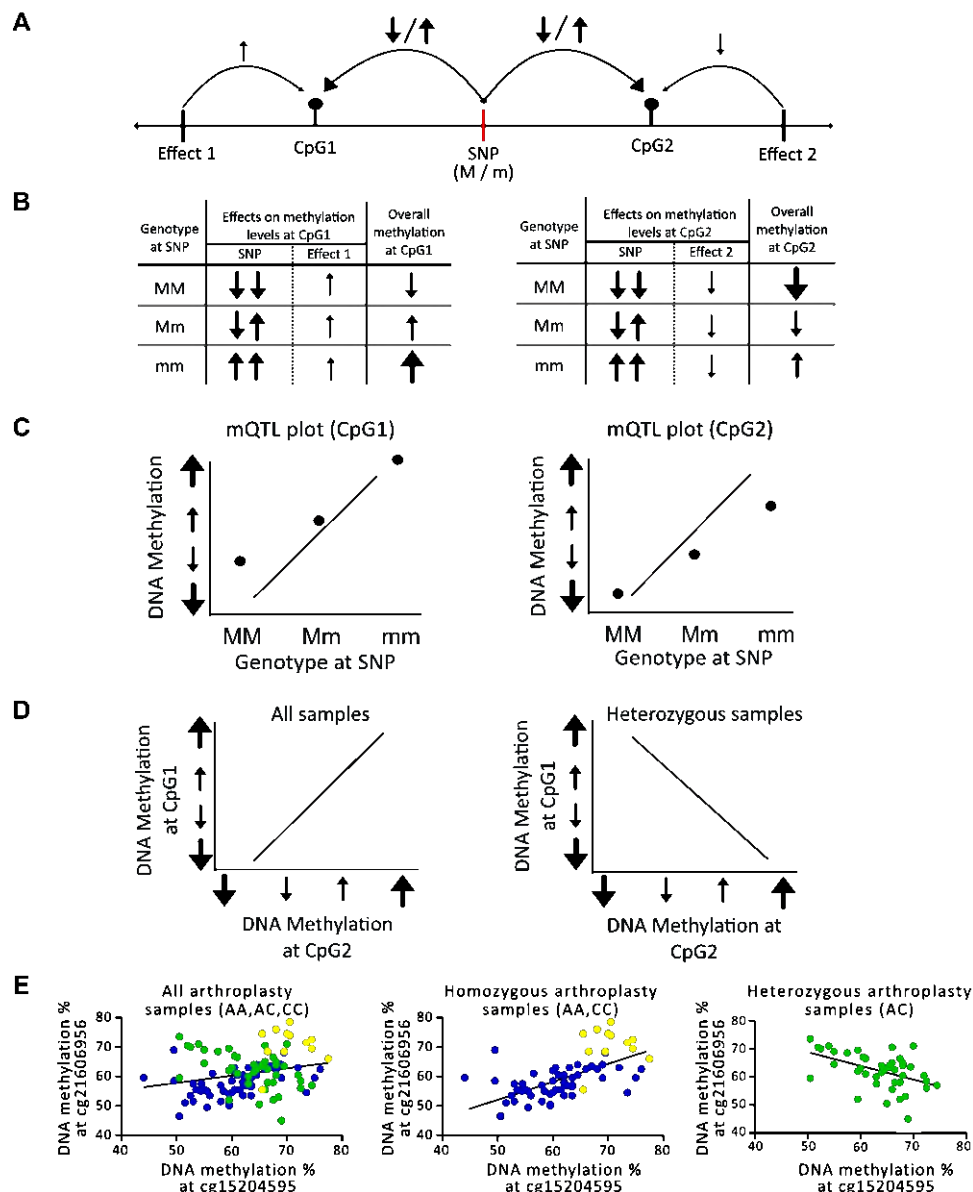


Figure 4.8. Theoretical model to account for the opposing slopes of the COLGALT2 meQTLs at cg15204595 and cg21606956. **A.** Schematic diagram showing the factors affecting the methylation levels at CpG1 and CpG2 (marked by black circles). The SNP (M/m) is marked by a red line and Effect 1 and Effect 2 are marked with black lines. Arrows indicate the effect the different factors have on the methylation levels at the two CpG sites with the direction of the arrowhead indicating the direction of the effect (up = increased methylation; down = decreased methylation) and the thickness of the arrow corresponding to the strength of the effect (thin = weaker effect; thick = stronger effect). **B.** Table summarising the cumulative effect of the SNP and Effect 1 and Effect 2 on the methylation levels at CpG1 (left) and CpG2 (right) for individuals homozygous (MM or mm) and heterozygous (Mm) for the SNP. The effects are represented by arrows with the direction of the arrowhead indicating the direction of the effect and the thickness of the arrow corresponding to the strength of the effect. **C.** Predicted mQTL plots at CpG1 (left) and CpG2 (right) using the expected overall methylation levels presented in B. DNAm levels are represented on the y-axis by arrows with the direction

of the arrowhead indicating the direction of the effect and the thickness of the arrow corresponding to the levels of methylation. Genotype at the SNP is on the x-axis (MM, Mm or mm). The overall DNAm levels expected for each genotype are represented by a dot and the mQTL direction is indicated by a line. **D.** Plots showing the expected overall DNAm levels at CpG1 plotted against the expected overall DNAm levels at CpG2 in all individuals (left) and in heterozygous individuals (right). DNAm levels are represented by arrows with the direction of the arrowhead indicating the direction of the effect and the thickness of the arrow corresponding to the levels of methylation. The lines show the direction of the predicted associations. **E.** Plots showing the measured DNAm levels at cg15204595 plotted against the measured DNAm levels at cg21606956 in all arthroplasty samples (left), in arthroplasty samples homozygous (AA and CC) for rs1046934 (middle), and in arthroplasty samples heterozygous (AC) for rs1046934 (right). Each sample is represented by a dot with the colour of the dot corresponding to rs1046934 genotype (blue = AA, green = AC, yellow = CC). The trend lines show the direction of the associations. The images were generated using Affinity Designer 2.1. (A. - D.) and GraphPad Prism 9.0 (E.).

4.3.5. Assessment of the regulatory function of the regions containing the differentially methylated cg15204595 and cg21606956 CpGs

The next step of this project was an *in vitro* investigation of the genomic regions harbouring cg15204595 and cg21606956 using the TC28a2 chondrocyte cells. The regions surrounding cg15204595 and cg21606956 were cloned into the CpG-free Lucia reporter gene plasmid. The constructs, which contained only the CpG sites of interest, were methylated *in vitro* and Lucia readings were taken from both the methylated and the unmethylated versions of the constructs allowing us to determine if the DNAm status of the CpGs affects the expression of the reporter gene.

The results suggested that both regions act as enhancers. The region surrounding cg15204595 showed increased Lucia readings compared to the empty control vectors with average increase in expression of the reporter gene of 1.36-fold ($P=0.003$) and 1.35-fold ($P=0.005$) for the unmethylated and the methylated constructs, respectively (Figure 4.9A). The cg21606956 region produced an average increase in Lucia activity of 1.41-fold ($P=0.002$) and 1.32-fold ($P=0.0003$) for the unmethylated and methylated constructs, respectively (Figure 4.9B). The methylation status of cg15204595 and cg21606956 had no significant effect on the regulatory function of the two enhancers ($P=0.96$ and $P=0.38$, respectively).

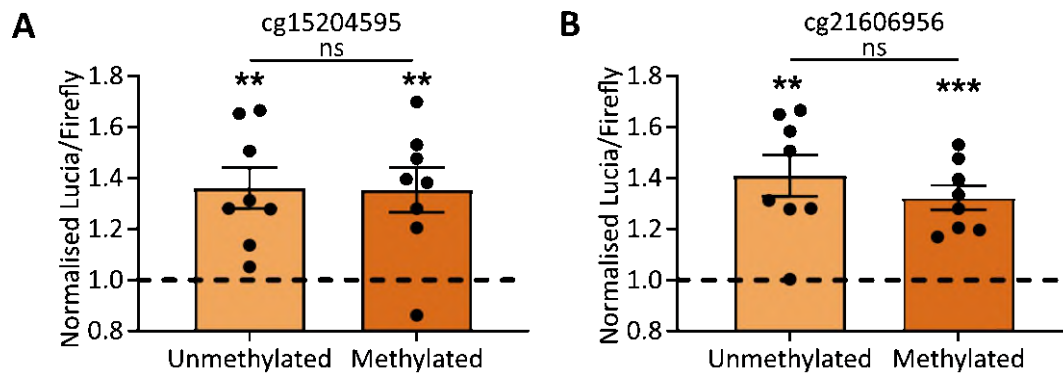


Figure 4.9. Reporter gene analyses. Lucia/Firefly luminescence ratios (y-axis) in TC28a2 chondrocytes following transfection with constructs containing cg15204595 (A.) or cg21606956 (B.) in unmethylated or methylated states, normalised to Lucia/Firefly luminescence ratios measured in cells transfected with empty vectors (dashed line). Each dot represents a biologic replicate. P-values were calculated using paired t-test (for comparisons between the empty vector controls and the vectors containing the inserts of interest) and non-paired t-tests (for comparisons between methylated and unmethylated constructs). ** = $P < 0.01$; *** = $P < 0.001$; ns = not significant. GraphPad Prism 9.0. was used for statistical analyses and data visualisation (A. and B.).

4.3.6. Modulation of the DNAm levels at the enhancers containing cg15204595 and cg21606956 in TC28a2 cells

Next, the role of DNAm at cg15204595 and cg21606956 in the gene expression regulation of *COLGALT2* and *TSEN15* was interrogated. DNAm levels at the two CpG sites and gene expression levels of *COLGALT2* and *TSEN15* were compared between TC28a2 chondrocytes transfected with gRNAs targeting the enhancers and dCas9 fused with catalytically active and inactive versions of TET1 and DNMT3A proteins (Figure 4.10).

The TC28a2 cells transfected with dCas9-TET1 plasmid containing the targeting gRNA sequences resulted in reduced methylation levels compared to the control at cg15204595 – 30.0% (Figure 4.13A, bottom left panel). and cg21606956 – 47.0% (Figure 4.10A, bottom left panel). When comparing the expression levels of *COLGALT2* and *TSEN15* in these cells, the 12.8% decrease in DNAm at cg15204595 lead to 1.3-fold ($P=0.0009$) increase in *COLGALT2* expression (Figure 4.10A, top right panels) and the 17.3% decrease in DNAm at cg21606956 resulted in 1.2-fold ($P=0.01$) increase in the gene expression (Figure 4.10A, bottom right

panels). No significant differences between the expression levels of *TSEN15* were observed ($P>0.05$) (Figure 4.10A, right panels).

Transfection with dCas9-DNMT3a plasmid containing the targeting gRNA sequences resulted the DNAm levels of 53.8% at cg15204595 (Figure 4.10B, top left panel) and 75.3% at cg21606956 (Figure 4.10B, bottom left panel) in the chondrocytes. This increase in DNAm levels at the two CpGs compared to the controls (10.5% increase for cg15204595 and 10.7% for cg21606956) did not significantly alter the expression of any of the three genes ($P>0.5$) (Figure 4.10B, right panels). The expression of *COLGALT2*, however, did show a small reduction of 0.2-fold ($P=0.09$) and 0.3-fold ($P=0.07$) following the increase of DNAm at cg15204595 and cg21606956, respectively (Figure 4.10B, right panels).

These results demonstrate that the two enhancers containing cg15204595 and cg21606956 regulate *COLGALT2* expression via DNAm at the two CpGs but not *TSEN15*.

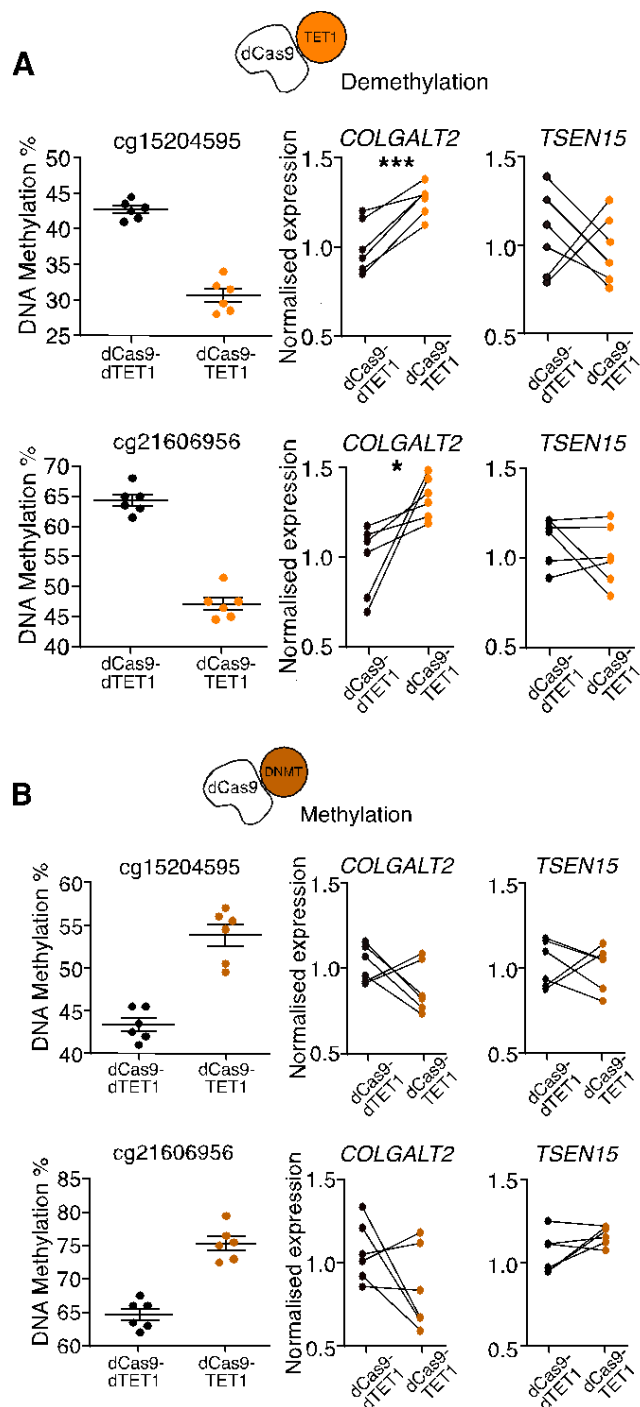


Figure 4.10. Epigenetic modulation of the enhancers containing cg15204595 or cg21606956. **A.** Left, DNAm levels at cg15204595 (top) and cg21606956 (bottom) in TC28a2 chondrocytes following transfection of targeting gRNAs with dCas9-dTET1 (black symbols) or with dCas9-TET1 (orange symbols). Each dot represents a biologic replicate. Bars show mean + SEM. Right, Effect of modulation on gene expression of COLGALT2 and TSEN15. Values were normalized to mean control values. *P*-values were calculated using paired *t*-tests. **B.** Left, DNAm levels at cg15204595 (top) and cg21606956 (bottom) in TC28a2 chondrocytes following transfection of targeting gRNAs with dCas9-dDNMT3a (black symbols) or with dCas9-DNMT3a (orange symbols). Each dot represents a biologic replicate. Bars show mean + SEM. Right, Effect of modulation on gene expression of COLGALT2 and TSEN15. Values

were normalized to mean control values. *P*-values were calculated using paired *t*-tests. *, *P* < 0.05; ***, *P* < 0.001. GraphPad Prism 9.0. was used for statistical analyses and data visualisation (A. and B.).

4.3.7. *In silico* prediction of the TF binding profile of the enhancer regions containing cg15204595 and cg21606956

The dCas9 experiment described in section 4.3.6 revealed that the enhancer regions containing cg15204595 and cg21606956 regulate the expression of *COLGALT2* in a DNAm dependent manner. This observation could be explained if TFs of *COLGALT2* bound to the enhancers and if their binding was modulated by DNAm. If the methylation status of cg15204595 and cg21606956 impacts TF binding affinity and thus regulates the expression of *COLGALT2*, these CpGs would be expected to be part of or in close physical proximity to TF binding sites. A search of the JASPAR TF database revealed several transcription factors predicted to bind at or near cg15204595 and cg21606956 (Figure 4.11A and 4.11B, respectively). Furthermore, many of these TF are expressed in cartilage from both OA and NOF patients⁶ (Figure 4.11C).

low likelihood of the substitutions having effects on the structure of the TSEN15 protein or on the larger tRNA splicing (TSEN) complex. rs1046934, similarly to but independently of the locus marked by rs11583641 (discussed in Chapter3), was shown to mediate its effect by modulating the expression of *COLGALT2* via methylation changes to two CpG sites, cg15204595 and cg21606956, located in enhancers. Using whole genome bisulphite DNAm sequencing data obtained from arthroplasty cartilage samples, as well as DNAm data obtained from pyrosequencing of DNA from an independent cohort of arthroplasty cartilage samples, the risk allele A of rs1046934 was found to correlate with decreased DNAm levels at the two enhancer CpGs - cg15204595 and cg21606956. The whole genome data highlighted another CpG, cg01436608, as potentially being affected by the SNP. Due to lack of pyrosequencing assays that could reliably capture this CpG, it was excluded from further studies. However, the location of the CpG, in close physical proximity to the TSS of *COLGALT2* and within a region that interacts with the enhancer containing cg21606956 in MSCs (Figure 4.6), makes this a very intriguing target for future studies.

In addition to correlating with low DNAm levels at cg15204595 and cg21606956, the OA risk allele of rs1046934 also associated with increased expression of both *COLGALT2* and *TSEN15* in arthroplasty patients, and a further correlation between methylation and *COLGALT2* expression was found.

Reporter gene assays confirmed that the genomic regions harbouring cg15204595 and cg21606956 are enhancers in TC28a2 chondrocytes. Modulation of the DNAm levels at cg15204595 and cg21606956 using catalytically active versions of the DNAm editing protein complexes resulted in increased expression of *COLGALT2* but not *TSEN15*, highlighting the importance of DNAm as a method of precise gene expression control and the complexity behind gene regulation across enhancer regions.

The *in vitro* data presented in this chapter supports the *ex vivo* data gathered from arthroplasty cartilage samples in that the enhancers regulate gene expression and genotype at rs1046934 is involved in this regulation to different degrees (AEI effects for *COLGALT2* were stronger than those observed for *TSEN15*) and via different mechanisms (epigenetic modulation did not affect the expression of *TSEN15* and no meQTLs were detected in patient samples for this gene). Both *COLGALT2* and *TSEN15* appear to be targets for the OA signal rs1046934, exerting its effects via activation/repression of the two regulatory elements

containing cg15204595 and cg21606956, respectively. The mechanisms through which the SNP affects expression of the two genes, however, are distinct: while the mechanism of regulation of *TSEN15* expression via the enhancers could not be determined, our data support a causal link between DNAm levels (correlated with genotype at rs1046934) at the two enhancer CpGs and *COLGALT2* expression, highlighting this as a potential regulatory mechanism. *In silico* data revealed that cg15204595 and cg21606956 reside in open chromatin regions in joint tissues and in or close to TF binding sites. Methylation of cytosines in CpG dinucleotides can influence steric and hydrophobic environment at the dinucleotide, rendering it similar to that of a thymine, a process termed thymine mimicry^{7,8}. This leads to a widening of the major groove and narrowing the minor groove at CpGs⁹ which results in an increased helix stiffness¹⁰. The effects of DNAm on TF binding have been detailed for a number of proteins^{11–13}. In many cases methylation is shown or predicted to prevent TF binding^{14,15}. This DNAm dependant TF binding affinity could be the mechanism via which *COLGALT2* expression is regulated in chondrocytes by the OA signals rs11583641 and rs1046934.

In this chapter and in Chapter 3, I describe the regulation of gene expression via DNAm at *cis*-acting enhancers as the method via which the risk alleles of two genetically independent risk loci, rs11583641 and rs1046934, contribute to altered *COLGALT2* expression in chondrocytes. In both cases, the OA risk-conferring alleles lead to increased *COLGALT2* expression in cartilage. Understanding the molecular mechanisms by which risk alleles impact their target genes is crucial for the clinical application of OA genetic discoveries and signals that impact epigenetic marks could lead to novel directions for drug development.

4.5. References

1. Boer, C. G. *et al.* Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell* **184**, 4784-4818.e17 (2021).
2. Sekulovski, S. *et al.* Assembly defects of human tRNA splicing endonuclease contribute to impaired pre-tRNA processing in pontocerebellar hypoplasia. *Nat. Commun.* **12**, 5610 (2021).
3. Song, J. & Markley, J. L. Three-dimensional Structure Determined for a Subunit of Human tRNA Splicing Endonuclease (Sen15) Reveals a Novel Dimeric Fold. *J. Mol. Biol.* **366**, 155–164 (2007).

4. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
5. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
6. 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. *Osteoarthr. Cartil.* **27**, 1545–1556 (2019).
7. Kribelbauer, J. F. *et al.* Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes. *Cell Rep.* **19**, 2383–2395 (2017).
8. Kribelbauer, J. F., Lu, X.-J., Rohs, R., Mann, R. S. & Bussemaker, H. J. Toward a Mechanistic Understanding of DNA Methylation Readout by Transcription Factors. *J. Mol. Biol.* **432**, 1801–1815 (2020).
9. [review] Machado, A. C. D. *et al.* Evolving insights on how cytosine methylation affects protein–DNA binding. *Brief. Funct. Genom.* **14**, 61–73 (2015).
10. Rao, S. *et al.* Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein–DNA binding. *Epigenetics Chromatin* **11**, 6 (2018).
11. Hashimoto, H., Zhang, X., Vertino, P. M. & Cheng, X. The Mechanisms of Generation, Recognition, and Erasure of DNA 5-Methylcytosine and Thymine Oxidations*. *J. Biol. Chem.* **290**, 20723–20733 (2015).
12. Wan, J. *et al.* Methylated cis-regulatory elements mediate KLF4-dependent gene transactivation and cell migration. *eLife* **6**, e20068 (2017).
13. [review] Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, (2017).
14. Kaluscha, S. *et al.* Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat. Genet.* **54**, 1895–1906 (2022).
15. [review] Héberlé, É. & Bardet, A. F. Sensitivity of transcription factors to DNA methylation. *Essays Biochem.* **63**, 727–741 (2019).

Chapter 5. Investigation of the rs11583641 and rs1046934 OA association signals in synovium

5.1. Introduction

As discussed in Chapter 1, OA affects multiple tissues in the articular joint. While a lot of research on OA is still focused on cartilage as a primary target tissue, in recent years there has been a shift in focus towards investigating the multi-tissue crosstalk underlying the complexity behind OA aetiology and progression. Targeted studies focused on the resolution of GWAS signals, as well as investigations utilizing different -omics technologies have been carried out in OA-relevant joint tissues including synovium¹⁻¹⁰, fat pad^{1-3,7} meniscus^{7,11,12} and subchondral bone^{8,13}, as well as peripheral tissues including blood^{1-3,14} and urine^{15,16}. A lot of these findings were integrated into the comprehensive database of OAomics and molecular biomarkers (OAOB)¹⁷.

Recent findings have suggested a role for the COLGALT2 enzyme in cancer progression^{18,19} and immune response^{20,21}. Based on these findings, we hypothesised that the COLGALT2 OA mQTLs and eQTLs operating in cartilage and described in Chapter 3 and Chapter 4 may also be active in non-cartilaginous joint tissues. To determine this, we investigated synovial tissue samples that our group had collected from OA patients undergoing arthroplasty.

5.2. Aims and objectives

5.2.1. Aims

- Determine if the mQTLs detected in cartilage and associated with the OA risk SNPs rs11583641 and rs1046934 are present in synovial samples from arthroplasty patients.
- Determine if the OA-associated SNPs rs11583641 and rs1046934 act as eQTLs in synovial samples from arthroplasty patients.

- Determine if the regions containing the mQTLs associated with the OA risk SNPs rs11583641 and rs1046934 act as enhancers in a human synovial cell line model and if the regulatory function of the regions (if any) is DNAm-dependent.

- Investigate the role of DNAm at the *COLGALT2* enhancers described in Chapter 3 and Chapter 4 in regulating the expression of their target gene(s) in a human synovial cell line model.

5.2.2. Objectives

- Genotype synovium arthroplasty patient DNA samples at rs11583641 and rs1046934.

- Measure DNAm levels at the 12 CpG sites associated with the OA signal rs11583641 in the synovium samples.

- Measure DNAm levels at the two CpG sites associated with the OA signal rs1046934 in the synovium samples.

- Perform AEI analysis in synovium patient samples on *COLGALT2* for rs11583641.

- Perform AEI analysis in synovium patient samples on *COLGALT2* and *TSEN15* for rs1046934.

- Clone the region containing the mQTLs associated with the OA signal rs11583641 in a CpG-free Lucia reporter gene vector, methylate the constructs *in vitro*, transfect the constructs in SW982 cells and measure Lucia readings.

- Methylate and demethylate the differentially methylated CpG sites associated with the OA signal rs11583641 using dCas9/DNMT3a or dCas9-TET1 fusion proteins, respectively, and measure the expression of *COLGALT2* and its neighbouring genes, *RGL1* and *TSEN15*, following the modulation.

5.3. Results

5.3.1. Detection of mQTLs associated with rs11583641 and rs1046934 in synovium tissue DNA from arthroplasty patients

DNAm levels at the 12 CpG sites located within the enhancer investigated in Chapter 3 were measured in synovial patient DNA samples and plotted against genotype at rs11583641 (Figure 5.1A). Significant mQTLs were detected at 8/12 CpG sites (CpG 3 – $P=0.002$, CpG 4 – $P<0.0001$, CpG 5 – $P<0.0001$, CpG 6 – $P=0.001$, CpG 7 – $P=0.01$, CpG 8 – $P<0.0001$, CpG 9 – $P<0.0001$, CpG 10 – $P<0.0001$).

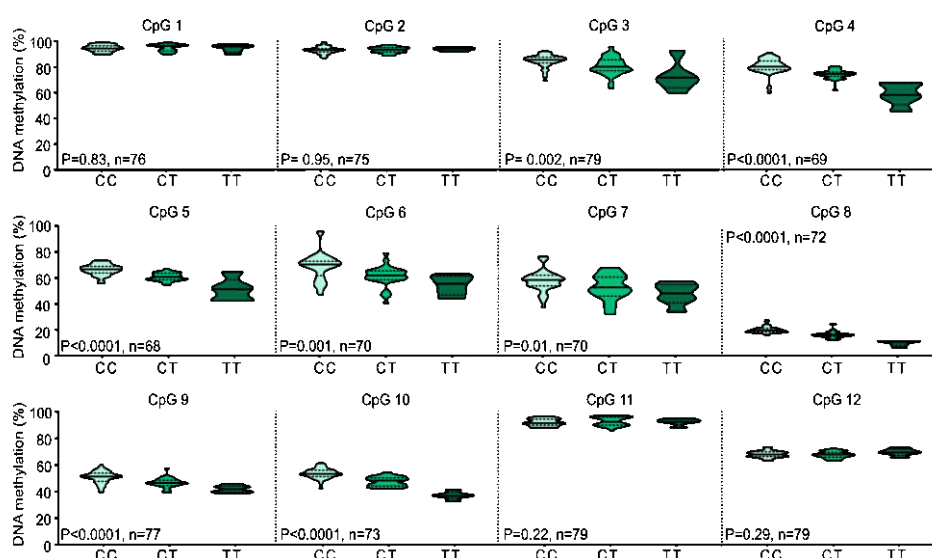


Figure 5.1. Synovium mQTL analysis of 12 CpG sites within the COLGALT2 enhancer associated with rs11583641. Violin plots showing DNAm values in arthroplasty synovium samples at the 12 investigated CpGs, stratified by rs11583641 genotype. Solid lines and dashed lines inside the plots represent median and interquartile range, respectively. P-values were calculated using simple linear regression. GraphPad Prism 9.0. was used for statistical analyses and data visualisation.

At all differentially methylated CpGs the OA-risk allele C of rs11583641 was associated with increased methylation compared to the non-risk allele, T (Figure 5.2A, left). This is in contrast with our observations in OA cartilage (Figure 5.2A, right) where at all differentially methylated CpG sites the risk allele, C, was associated with lower DNAm.

The strength of the effect of genotype at rs11583641 upon DNAm at the 12 CpG sites was compared between OA cartilage (n=80-99) and OA synovium (n=68-79) samples (Figure 5.2B). The genotypic effect % (GE%) was stronger in synovium at 9/12 CpG sites, including all differentially methylated ones (CpGs 3-10). The strongest effects in both tissues were observed at CpG 8 (69.8% in synovium and 67.7% in cartilage) and CpG 10 (62.5% in synovium and 38.0% in cartilage). In synovium, strong effects were also observed at CpGs 4 and 5 (53.7% and 51.7%, respectively). We also compared the mean DNAm levels at the 12 CpG sites, irrespective of genotype at rs11583641, between cartilage and synovium (Figure 5.2C). Significant differences were observed at 10/12 CpGs ($P < 0.05$). At 8 of these CpGs DNAm levels were higher in cartilage with the greatest difference in DNAm levels observed at CpG 6 (12.27%) and CpG 7 (19.49%). At most CpG sites the spread of the data was also greater in cartilage, suggesting higher variability between patients for this tissue.

The synovium samples were also genotyped at rs1046934 and DNAm was measured at cg15204595 and cg21606956 (discussed in Chapter 4). The mQTLs detected at these two CpGs in cartilage were not observed in synovium (Figure 5.3), suggesting that the mQTL at the rs1046934 locus is cartilage specific.

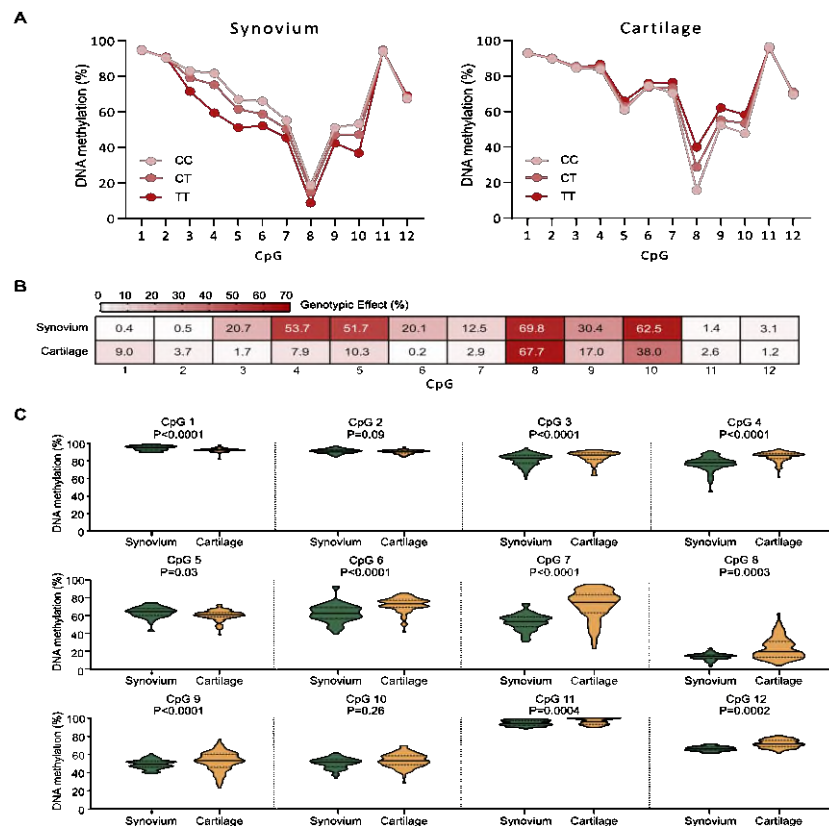


Figure 5.2. Comparison of the DNAm levels at the enhancer associated with rs11583641 between synovium (n= 49-79) and cartilage (n=81-98) DNA samples from arthroplasty patients. **A.** The median DNA methylation value for each of the 12 CpGs stratified by genotype at rs11583641 in synovium (left) and cartilage (right) from arthroplasty patients. **B.** Heatmap showing the contribution of the rs11583641 genotype to the observed DNAm levels differences at the 12 CpGs in synovium and cartilage samples. The heatmap displays r^2 values calculated by linear regression analysis, converted to percentages. **C.** Violin plots showing DNAm levels at the 12 investigated CpGs in synovium and cartilage samples. Solid and dashed horizontal lines represent the median and interquartile range, respectively. P-values were calculated using Mann-Whitney tests. GraphPad Prism 9.0. was used for statistical analyses and data visualisation (A., B., and C.).

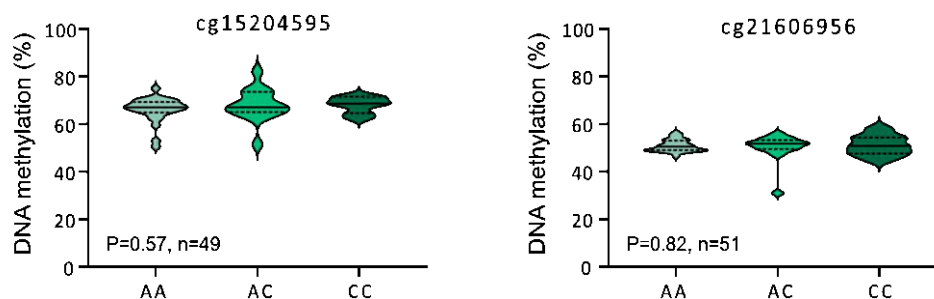


Figure 5.3. Synovium mQTL analysis for cg15204595 and cg21606956. Violin plots showing DNAm values in arthroplasty synovium DNA samples at cg15204595 (left) and cg21606956 (right).

(right), stratified by rs1046934 genotype. Solid and dashed horizontal lines represent the median and interquartile range, respectively. P-values were calculated using simple linear regression. GraphPad Prism 9.0. was used for statistical analyses and data visualisation.

5.3.2. Identification of COLGALT2 allelic imbalance in synovium tissue from arthroplasty patients

To determine if the COLGALT2 cartilage eQTL described in Chapter 3 is active in synovium, AEI analysis was performed in this tissue for patients heterozygous at rs11583641. A significant imbalance between the two alleles of the SNP was detected with the risk C allele showing a 0.81-fold mean decrease in expression compared to the non-risk T allele ($P < 0.0001$, Figure 5.4A). In cartilage, the risk allele showed an increased expression (1.94-fold mean increase, $P < 0.0001$, Chapter 3, Figure 3.14A). Like our observation following the mQTL analysis (section 5.3.1), the risk allele at this locus showed an opposing effect on COLGALT2 expression in synovium compared to cartilage. Furthermore, while the AEI effect was only present in 13/27 cartilage samples (Figure 3.15A), in synovium all 15 tested patients showed allelic imbalance (Figure 5.4A).

When the AEI ratios were plotted against DNAm levels at the 12 CpGs associated with the rs11583641 locus in synovium, no meQTLs were detected (Figure 5.4B).

Although no mQTLs were detected at the CpGs associated with the rs1046934 locus (Figure 5.3), it is possible that the SNP affects the expression of COLGALT2 and/or TSEN15 in synovium through a mechanism independent of DNAm at the enhancer regions harbouring cg15204595 and cg21606956. To determine if that was the case, we performed the AEI analysis using rs114661926 (COLGALT2) and rs2274432 (TSEN15), which are in high LD with rs1046934 ($r^2 > 0.8$, GBR), in synovium samples from patients who were compound heterozygotes. No allelic imbalance was observed for COLGALT2 or TSEN15 (Figure 5.5).

Since no mQTL nor eQTL effects were observed in synovium samples at the rs1046934 locus we did not proceed with further investigations of the locus in this tissue.

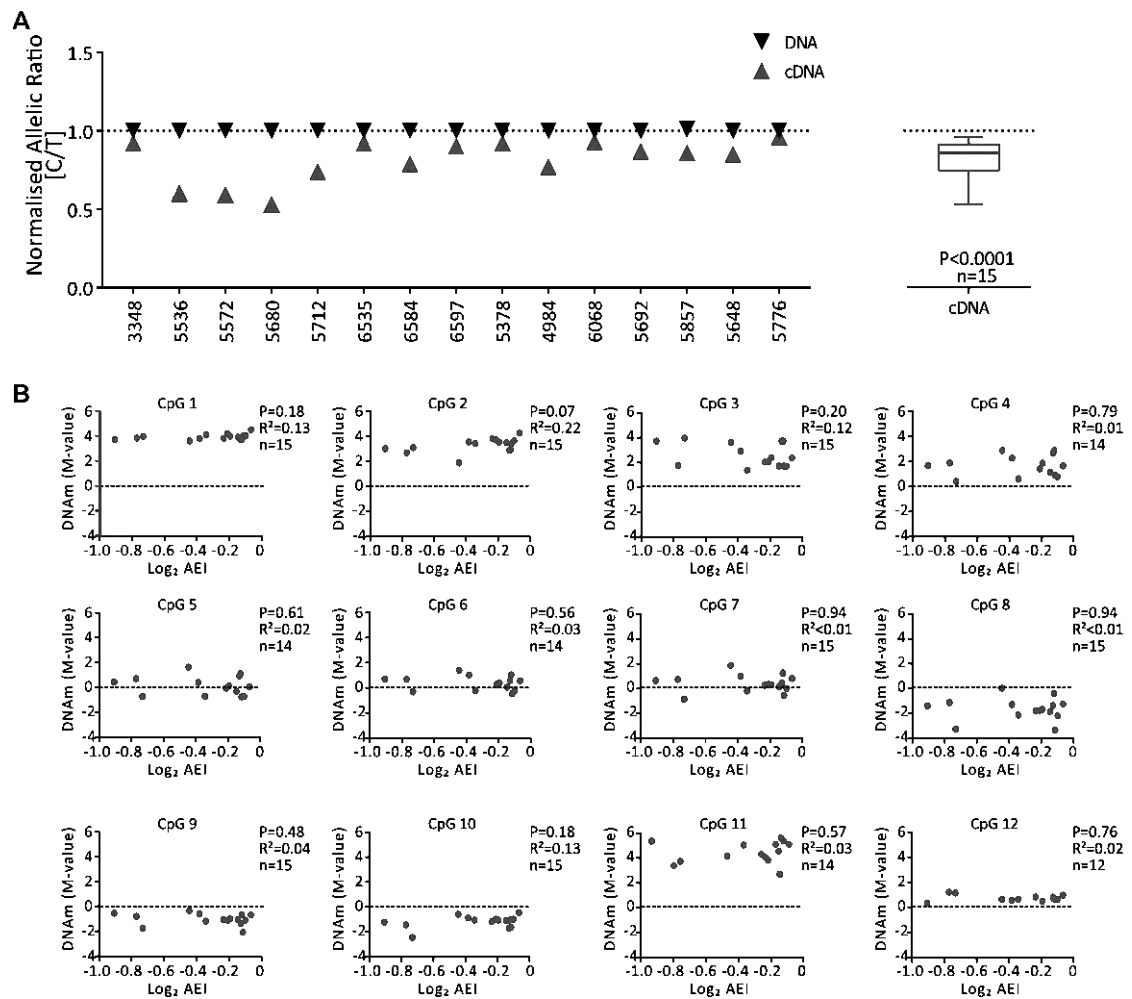


Figure 5.4. AEI analysis of *COLGALT2* in arthroplasty synovium samples for the *rs11583641* locus. **A.** Left, Allelic (C/T) ratios in synovium samples from OA patients heterozygous for *rs11583641*. In each sample, the ratio of values for cDNA and DNA between the OA-risk allele C and the non-risk T allele was plotted; each symbol represents the mean of 3 technical repeats. Right, Mean DNA and cDNA values in the presence of the C allele versus the T allele in heterozygous patients. Values are shown as a box plot, with the line inside the box representing the median, the box showing the interquartile range, and whiskers showing the minimum and maximum values. P-values were calculated using Wilcoxon's matched pairs signed rank test. **B.** Relationship between *COLGALT2* AEI ratios and DNAm at the enhancer in synovium. Allelic ratios ($\log_2$) of *COLGALT2* (x-axis) are plotted against matched DNAm levels (M-values) (y-axis) at the investigated CpGs (CpG 1-12). Each dot represents data from 1 individual. P-values were calculated using simple linear regression and the test statistics are shown to the right of each plot (P-value, R² and number of samples used for the analysis). GraphPad Prism 9.0. was used for statistical analyses and data visualisation (A. and B.).

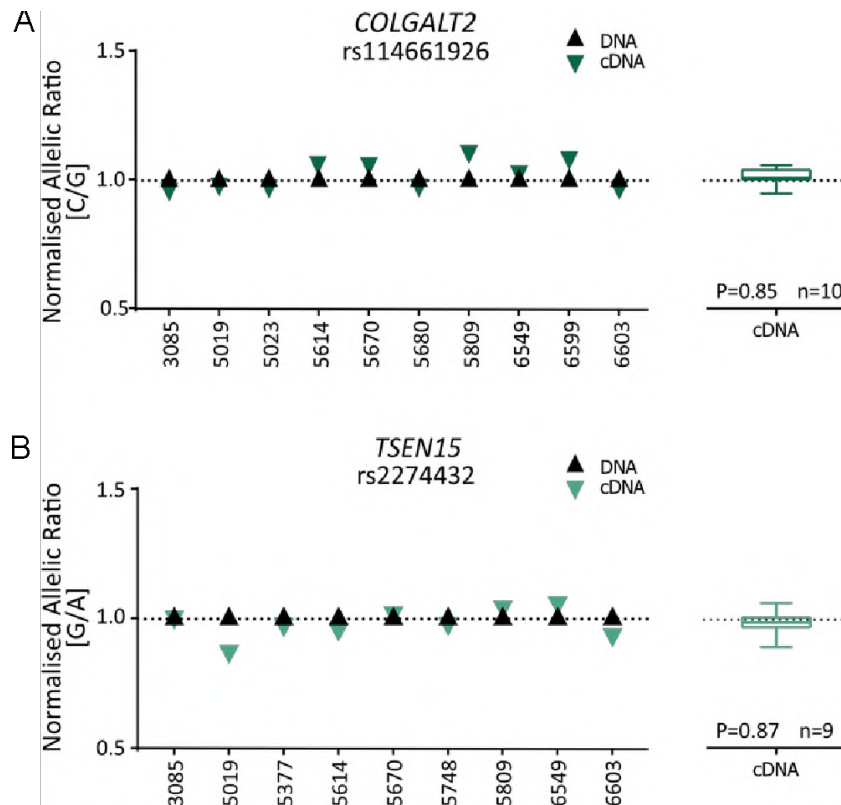


Figure 5.5. AEI analysis of *COLGALT2* and *TSEN15* in arthroplasty synovium samples for the *rs1046934* locus. **A.** Left, Allelic (C/G) ratios in patient samples heterozygous for *COLGALT2* transcript SNP rs11466192. In each sample, the ratio of values for cDNA and DNA between the OA-risk allele and the non-risk allele was plotted; each symbol represents the mean of 3 technical repeats. Right, Mean DNA and cDNA values in the presence of the risk allele versus the non-risk allele in heterozygous patients. Values are shown as a box plot, with the line inside the box representing the median, the box showing the interquartile range, and whiskers showing the minimum and maximum values. P-values were calculated using Wilcoxon's matched pairs signed rank test. **B.** Same as A. for *TSEN15* SNP rs2274432 (G/A; risk/non-risk). GraphPad Prism 9.0. was used for statistical analyses and data visualisation (A. and B.).

5.3.3. Reporter gene assessment in the SW982 synovial cell line of the regulatory function of the rs11583641 region containing the differentially methylated CpGs

Our previous experiments have confirmed that the region containing the rs11583641 mQTLs acts as an enhancer in TC28a2 chondrocyte cells (Chapter 3, Section 3.3.4). The regulatory effect of the region was shown to be DNAm-sensitive and dependent on genotype at rs734657, which is in high LD with rs11583641 ($r^2=0.83$, GBR). To test if the region has regulatory function in a synovial cell model, we transfected the *in vitro* methylated and mock-methylated Lucia reporter gene constructs containing the differentially methylated CpG sites

and the three haplotypes formed by the SNPs in the region (Section 3.3.4) into cells of the human synovial sarcoma cell line SW982.

The results of this experiment are presented in Figure 5.6. The constructs containing the C allele of rs734657, which corresponds to the OA-risk C allele of rs11583641 (G-C and A-C haplotypes) showed 5.5-fold ($P=0.004$) and 4.4-fold ($P=0.03$) increase in Lucia expression, respectively (Figure 5.6A). No differences were observed between the two constructs, methylated or unmethylated, indicating that the genotype at rs943409 had no influence on the reporter gene activity. The constructs containing the non-risk allele, A, of rs734657 (G_A haplotype) did not show significantly increased Lucia activity compared to the control ($P=0.71$) (Figure 5.6A). While the differences were not statistically significant, the cells transfected with the constructs containing the G_A haplotype showed lower mean Lucia readings compared to the ones containing the G_C and the A_C haplotypes, especially in their methylated state (5.7-fold mean difference with G_C ($P=0.23$) and 5.4-fold mean difference with A_C ($P=0.24$)). These results suggest that the differentially methylated region has a regulatory function in synovial sarcoma cells which is dependent on genotype at rs734657.

The enhancer function of the region also appeared to be affected by the methylation status of the constructs. The methylated constructs containing the G_C and the A_C haplotypes produced higher Lucia readings compared to the unmethylated constructs with mean differences of 3.6-fold ($P=0.002$) and 4.4-fold ($P=0.01$), respectively. There was no difference between the Lucia readings of the cells transfected with the methylated or unmethylated ($P=0.29$) construct containing the G_A haplotype (Figure 5.6B).

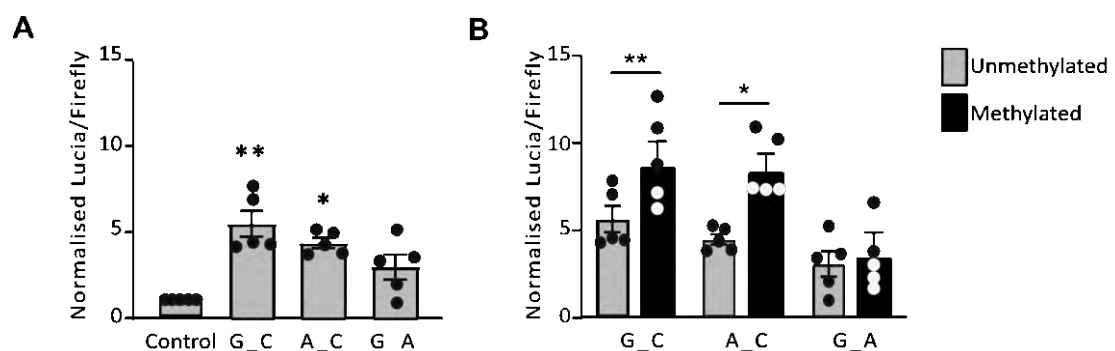


Figure 5.6. Reporter gene analysis. Lucia/Firefly readings in SW982 cells in the presence of constructs containing the 3 haplotypes of rs943409 and rs734657 (G_C, A_C, and G_A) (A.) and the 3 haplotypes in a methylated or unmethylated state (B.). Values were normalized to

those in an empty vector control. Dots represent individual samples ($n = 5$). Bars show the mean $\pm$ SEM. P-values were calculated using paired t-tests (A.) and non-paired t-tests (B.). * = $P < 0.05$; ** = $P < 0.01$. GraphPad Prism 9.0. was used for statistical analyses and data visualisation (A. and B.).

The results from the Lucia reporter gene assay in SW982 cells revealed that the region containing CpG3-10 acts as enhancer in this synovial cell model. The enhancer activity was dependent on genotype at rs734657 with the C allele of the SNP (corresponding to the C allele of rs11583641) associated with higher gene expression compared to the A allele. Methylation at the region resulted in lower enhancer activity. This agrees with the reporter gene data using the TC28a2 chondrocyte cells, although in those cells DNAm decreased enhancer activity of the G_C and A_C haplotype constructs (Figure 3.6B).

5.3.4. Modulation of the DNAm levels at the enhancer region in SW982 cells

We next tested the effect of methylation at CpGs 3-10 on the expression of *RGL1*, *COLGALT2* and *TSEN15* using dCas9 coupled with TET1 or DNMT3a and gRNAs targeting the region in SW982 cells. The same gRNAs used to target the region in the TC28a2 chondrocytes were used (Section 3.3.7, gRNA 1-6) alone or in combination. Our control cells were transfected with the same targeting gRNAs and dCas9-dTET1 or dCas9-dDNMT3a.

DNAm levels across the CpGs in the enhancer were successfully increased and decreased following expression of dCas9-DNMT3a or dCas9-TET1, respectively, and the six gRNAs targeting the region (alone or in combination) (Figure 5.7).

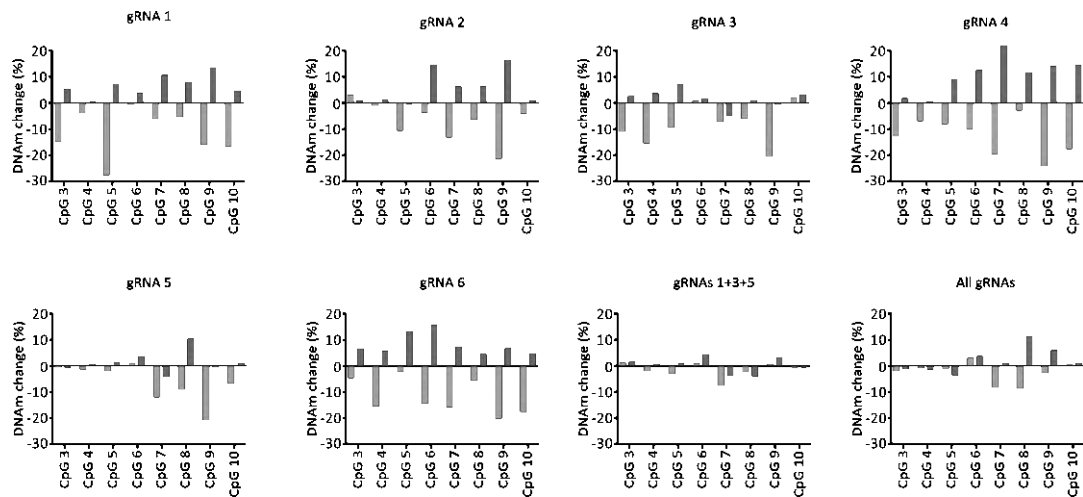


Figure 5.7. Percentage change in DNAm following targeted epigenome modulation of the enhancer associated with rs11583641 in SW982 cells. Each plot shows the change in DNAm following epigenetic modulation with the gRNAs (gRNA1-6) or with combinations of the gRNAs (gRNA 1, 3, and 5 or with all gRNAs). The bars show the percentage change in DNAm (compared to the control cells) following modulation with dCas9-DNMT3a (red bars, n=3) or dCas9-TET1 (blue bars, n=3) at each of the 8 targeted CpGs (CpG 3- CpG 10). GraphPad Prism 9.0 was used for data visualisation.

Transfection of dCas9-DNMT3a and gRNAs lead to increased DNAm at the CpGs (Figures 5.7 and 5.8A, Panel 1). The greatest increases in DNAm were achieved at CpGs 6, 7 and 9, at which distinct gRNAs increased methylation by 16.5% (gRNA 2), 20.5% (gRNA 4), and 14.9% (gRNA 1), respectively (Figure 5.7). Modest increase in DNAm at only a few CpGs was achieved by expression of gRNA 3 (up to 7.0%), gRNA 5 (up to 10.5%) or combinations of multiple gRNAs (gRNAs 1+3+5 – up to 4.4%; all gRNAs – up to 11.3%) (Figure 5.7).

Increased DNAm levels at the enhancer resulted in reduced *COLGALT2* expression in cells transfected with gRNA 1 (0.15-fold, $P=0.04$), gRNA 2 (0.29-fold, $P=0.04$), gRNA 4 (0.21-fold, $P=0.04$), and gRNA 6 (0.12-fold, $P=0.005$) (Figure 5.8A, Panel 2). These 4 gRNAs all had an impact on the DNAm at CpGs 6-9. While gRNA 5 had only a modest effect on DNAm (Figure 5.8B, Panel 1), the expression of *COLGALT2* measured in the cells transfected with that gRNA was also reduced (0.18-fold, $P=0.05$) (Figure 5.8B, Panel 2).

DNAm levels across the enhancer CpGs was successfully reduced using dCas9-TET1 (Figures 5.7 and 5.8B, Panel 1). The strongest effects were observed at CpGs 5, 7 and 9 at which the DNAm was reduced by up to 27.7% (gRNA 1), 19.5% (gRNA 4) and 24.1% (gRNA 4),

respectively (Figure 5.7). Combinations of gRNAs resulted in small reduction of DNAm across the enhancer.

Significant increase in *COLGALT2* expression was observed following reduced DNAm of the enhancer using gRNA 1 (1.1-fold, $P=0.02$), gRNA 2 (1.1-fold, $P=0.01$), gRNA 3 (1.2-fold, $P=0.02$), gRNA 4 (1.2-fold, $P=0.03$), and gRNA 6 (1.2-fold, $P=0.03$) (Figure 5.8B, Panel 2).

In the analysis of the enhancer in the TC28a2 chondrocyte cells, increased DNAm resulted in decreased *COLGALT2* expression, whilst decreased DNAm resulted in increased *COLGALT2* expression (Figure 3.12). *COLGALT2* expression therefore shows the same response in SW982 and TC28a2 cells to enhancer DNAm changes.

Changes to the DNAm levels of the CpGs at the enhancer had no significant effect (all P -values >0.05) on the expression of *RGL1* or *TSEN15* in the SW982 cells (Figure 5.8). This also matched what was observed for the TC28a2 cells (Figure 3.12).

Differences in *COLGALT2* expression between the control and modulated SW982 cells were observed following changes to CpGs 7-10 when the DNAm levels at one or more of these CpGs was increased or decreased (Figure 5.8). The DNAm levels at these 4 CpG sites was lower compared to the other CpGs in the SW982 cells ($<35\%$). Lower DNAm has been associated with higher levels of TF binding^{22,23}, marking these 4 CpG sites as potentially functionally active in SW982 cells.

Figure 5.8. Epigenetic modulation of the enhancer in SW982 cells. **A. Panel 1,** DNAm levels at the 8 CpGs in TC8a2 chondrocytes following expression of targeting gRNAs with dCas9-dDNMT3a (black symbols) or with dCas9-DNMT3a (coloured symbols). Each dot represents a biologic replicate. **Panel 2,** Effect of modulation on gene expression of RGL1 (left), COLGALT2 (middle) and TSEN15 (right). Values measured in the modulated cells (coloured triangles) were normalized to those measured in the control cells (black triangles). **B. Panel 1,** DNAm levels at the 8 CpGs in TC8a2 chondrocytes following expression of targeting gRNAs with dCas9-dTET1 (black symbols) or with dCas9-TET1 (coloured symbols). Each dot represents a biologic replicate. **Panel 2,** Effect of modulation on gene expression of RGL1 (left), COLGALT2 (middle) and TSEN15 (right). Values measured in the modulated cells (coloured triangles) were normalized to those measured in the control cells (black triangles). P-values were calculated using paired t-tests. * = $P < 0.05$; ** = $P < 0.01$. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A. and B.).

5.3.5. In silico prediction of the TF binding profile of the enhancer region associated with rs11583641 in synovium and cartilage

The results from the targeted methylation and demethylation using dCas9 at the enhancer in SW982 cells revealed that the regulation of *COLGALT2* expression through DNAm of the CpGs at the enhancer operates in a similar manner to what was observed to be the case in chondrocytes (Chapter 3, Figure 3.13): reducing the DNAm levels increased the expression of the gene and increasing DNAm resulted in decreased *COLGALT2* levels. We hypothesise that the enhancer regulates gene expression through DNAm dependent TF binding. Using publicly available datasets the region was investigated. The JASPAR TF database identified TFs predicted to bind at or near the differentially methylated CpG sites (CpG 3-10) (Figure 5.9A). Using RNA-seq data from chondrocytes from OA patients²⁴ and from synovial fibroblasts from OA patients²⁵ the expression of these TFs between the two cell types was compared. A total of 32 TFs and TF dimers were predicted to bind across the region encompassing CpGs 3-10 (+25bp either side) (Figure 5.9A, panel 3). Thirteen of these TFs are expressed in at least one of the two tissues at varying levels (Figure 5.9B).

The Lucia reporter gene assay results suggested a role for the SNP rs734657 in the regulation of gene expression (Figure 5.6). We looked at the region where the SNP resides to identify any TF that bind over the SNP and could be affected by the variant. The SNP is part of the consensus binding motif for RFX6, a TF almost exclusively expressed in pancreatic islets with a known role in the regulation of insulin sensing²⁶ and with a potential involvement in pain perception²⁷. Based on its genomic position, however, the different alleles of the rs734657 variant are not expected to affect the binding affinity of RFX6 (Figure 5.9C) and the transcription factor is not expressed in detectable amounts in synovial fibroblasts nor chondrocytes (Figure 5.9B and C).

These data confirm that the *COLGALT2* enhancer containing the differentially methylated CpGs is transcriptionally accessible in both cartilage and synovium, and it contains binding motifs for a range of TFs, some of which are expressed in both tissues but at different levels.

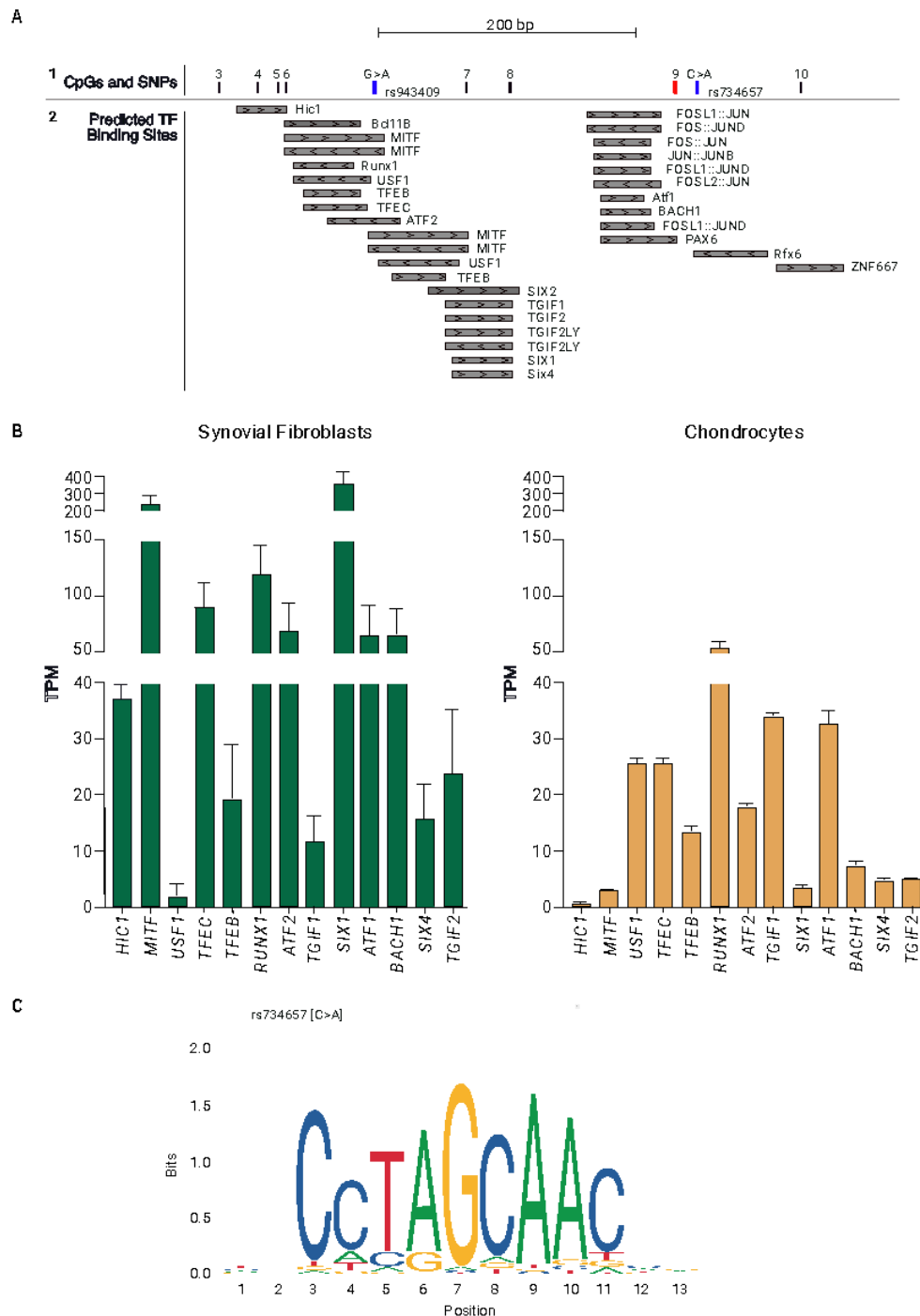


Figure 5.9. TFs predicted to bind at the enhancer. **A.** Top, location of the 8 DMR CpGs (black lines, CpGs 3-10); cg18131582 is CpG9 (red line). Also shown are the locations of the DNA variants (blue bars) rs943409 (G > A) and rs734657 (C > A). Bottom, TF-binding sites within 25bp of the DMR as predicted by JASPAR and visualised in the UCSC Genome Browser (hg19). The TFs are marked by grey bars with the direction of the arrows within the boxes 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). **B.** Expression levels (transcripts per million (TPM)) of TFs predicted to bind within the DMR in synovial fibroblasts (left) and chondrocytes (right). Bars show the mean +SEM. The y-axes are linear segmented scales with 3 segments. GraphPad Prism 9.0 was used for the data visualisation. **C.** RFX6 binding motif with the nucleotide

positions on the x-axis and the height of the letter corresponding to the binding affinity of the TF to that nucleotide. The position of the SNP rs734657 (position 2) is highlighted in pink. The image was adapted from the JASPAR database using Affinity Designer 2.1.

5.4. Discussion

Understanding the molecular mechanisms behind GWAS signals is an essential step in identifying potential therapeutic targets. Identification of QTLs can help identify the target genes of these signals. It is important to remember, however, that these QTLs can be cell type-specific and can occur under different physiological contexts or during different stages of development²⁸. The discovery of these cell type-specific QTLs is problematic for diseases where relevant tissues are not easily obtainable. In OA, however, arthroplasty surgery provides researchers with access to multiple joint relevant tissues. Despite this, cartilage-centric studies dominated the field for a long time with interest in the complexities of the multi-tissue crosstalk only gathering momentum in the past decade. In this chapter I undertook a functional analysis of the two OA signals operating on *COLGALT2* described in cartilage (Chapter 3 and Chapter 4) in another OA-relevant tissue, synovium. The effects of the signal marked by rs1046934 on DNAm at cg15204595 and cg21606956, and on the expression of *COLGALT2* and *TSEN15* were not detected in synovium, meaning the signal is not active in all joint-related tissues. The OA risk signal marked by rs11583641, however, was found to act as a QTL in both cartilage and synovium.

We started by investigating the relationship between the risk SNP rs11583641 and DNAm at the *COLGALT2* enhancer in synovium tissue from arthroplasty patients. Our investigation revealed that in synovium the risk allele of the SNP, C, is associated with higher DNAm compared to the non-risk allele, T, the opposite of what had been observed in cartilage. The mQTL effect is present at more CpG sites within the enhancer in synovium (7/12) compared to cartilage (5/12) and the effect is stronger at all differentially methylated CpGs. Average DNAm levels at the enhancer were also lower in synovium compared to cartilage. None of the samples used in these analyses were matched and direct comparisons between DNAm levels in the different tissues within the same patient was not possible.

AEI analysis revealed that the rs11583641 risk allele corresponds to higher *COLGALT2* expression relative to the non-risk allele. These findings, again, showed an opposite trend to what was observed in cartilage. The allelic imbalance was consistent across all samples.

Using a reporter gene assay, we tested the regulatory function of the region containing the rs11583641 differentially methylated CpGs and the impact DNAm at the CpGs has on that activity in the SW982 synovial fibroblast sarcoma cell line, a synovium cell model commonly used in studies of RA²⁹. Similar to our observations in the TC28a2 chondrocyte cell model (Chapter 3, Section 3.3.4), we confirmed that the region containing the differentially methylated CpG sites acts as an enhancer in synovial fibroblasts. Furthermore, as observed in chondrocytes, an allelic effect was observed. The Lucia-reporter gene constructs containing the C allele of rs734657, corresponding to the C allele of rs11583641, showed higher expression of the reporter gene compared to the constructs containing the non-risk allele. The methylation status of the region also influenced gene expression with lower DNAm resulting in higher Lucia readings. These results were not corroborated by our observations in patient samples where the risk allele of rs11583641 was associated with higher DNAm at the enhancer and lower expression of *COLGALT2*. While reporter gene analyses are an inexpensive, rapid, and sensitive tool for assessing the regulatory function of DNA elements, there are also several limitations to this *in vitro* model. The experiment investigates a piece of DNA isolated from its normal genomic and chromatin context, with a lack of the 3D conformation that would be observed *in vivo*; this can impact on the accessibility of the region to TFs. Furthermore, *in vitro* DNA methylation of the constructs using enzymes results in constructs that completely lack methylation and constructs that are completely methylated, whereas the DNAm status of a region between primary cells is more variable.

To test the true effect of DNAm at the enhancer upon gene expression in the SW982 cells we performed a targeted RNA-guided methylation and demethylation of the region using dCas9 coupled with DNMT3a and TET1, respectively. The results of the experiment showed an antagonistic relationship between DNAm at the enhancer and *COLGALT2* expression; increased DNAm led to decreased *COLGALT2* expression, and *vice versa*. The expression of the other two genes in the region, *RGL1* and *TSEN15*, was not affected following the epigenetic modulation of the enhancer. The DNAm status of CpGs 7-10 appeared to have the strongest effect on *COLGALT2* expression. DNAm levels at these CpGs is lower in SW982 cells and it is possible that TFs that bind in a DNAm-dependent fashion to these CpGs drive *COLGALT2* expression.

Based on this hypothesis we identified TFs predicted to bind at the region using the JASPAR TF database. Thirteen of the TF that were predicted to bind were expressed in cartilage and/or synovium, with many showing differences in their levels of expression between the two tissue types.

The enhancer region containing the differentially methylated CpGs was found to regulate the expression of *COLGALT2* in both cartilage and synovium from OA patients. Methylation and demethylation of the CpGs using dCas9-DNMT3a or dCas9-TET1 proteins and six gRNAs targeting the region showed a causal link between DNAm at the enhancer and *COLGALT2* expression in both TC28a2 chondrocytes and SW982 synovial fibroblasts, suggesting that the mechanism through which the enhancer drives *COLGALT2* expression is identical in cartilage and synovium.

The mechanism through which the OA GWAS SNP affects the DNAm at the enhancer and gene expression, however, appears to be different between the tissues. In patient samples we observed mQTL and eQTL effects showing opposite trends in cartilage compared to synovium. The risk allele of rs11583641 was associated with low DNAm at the enhancer and high expression of *COLGALT2* compared to the non-risk allele in cartilage, whereas in synovium the risk allele was associated with high DNAm and low *COLGALT2* expression.

We hypothesise that the observed differences in the directions of the mQTL and the eQTL are due to differences in TFs binding to the causal SNP between cartilage and synovium. These differences can be quantitative (specific TF(s) can be more expressed in one tissue compared to the other) and/or qualitative (a different repertoire of TF can be expressed in different cells). Since the causal SNP at the locus is not known, confirmation of what TFs bind at the SNP and if they are differentially expressed in cartilage and synovium is not possible.

Based on the available information, we propose a model to explain the differences we observe between cartilage and synovium in the regulation of *COLGALT2* expression by the enhancer containing the differentially methylated CpGs. The model is based on allelic differences in the TF binding affinity at the causal SNP (rs11583641 or another SNP in high LD with it), combined with different TF being expressed in the two tissue types.

In our theoretical model, the locus's causal SNP possesses a risk allele, denoted as “C” and a non-risk allele, denoted as “T”. Within cartilage, the risk allele “C” exhibits a higher

affinity for, or interacts with, a greater number of TFs in comparison to the non-risk allele “T”. The same applies to the alleles in synovium but with a different subset of TFs (Figure 5.10).

The TF(s) binding to the risk allele C in cartilage and the ones binding to the non-risk allele T in synovium result in lower DNAm levels at the enhancer which then recruits more of a TF that binds at the enhancer and leads to increased *COLGALT2* expression (Figure 5.10). The different set of TFs that bind to the non-risk T allele in cartilage and to the risk allele C in synovium lead to higher levels of DNAm at the enhancer, less of the TF that binds to the enhancer and lower levels of *COLGALT2* expression.

In our model the effect of the alleles of the disease-associated SNP on DNAm at the enhancer are different between the tissues but the mechanism driving *COLGALT2* expression through DNAm at the enhancer is the same (Figure 5.10).

This model accounts for the results from the targeted DNAm modulation of the enhancer in TC28a2 chondrocytes (Chapter 3, Figure 3.13) and in SW982 (Figure 5.8) which otherwise seem to contradict the mQTL and AEI data collected from patient cartilage (Chapter 3, Figure 3.2A and 3.15A, respectively) and synovium (Figure 5.1 and 5.4A) samples.

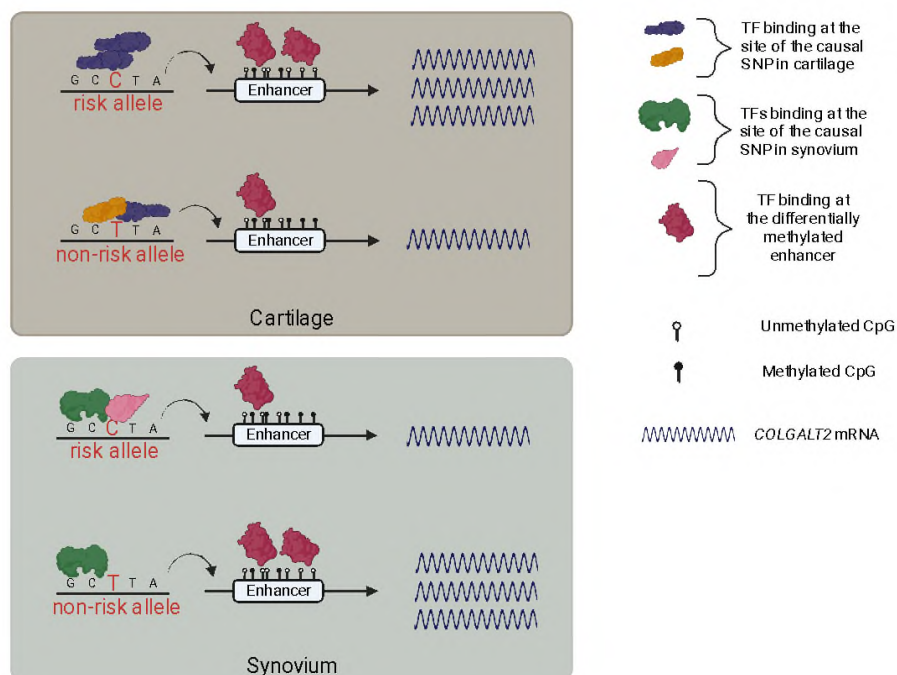


Figure 5.10. Model accounting for the opposite effects between synovium and cartilage of the risk allele on enhancer DNAm levels and *COLGALT2* expression. The causal variant has

hypothetical alleles C (risk) and T (non-risk). In cartilage (top box), the alleles differentially bind transcription factors (blue and orange); in synovium (bottom box), they differentially bind different transcription factors (green and pink). Differential transcription factor binding leads to allele-specific enhancer methylation, resulting in quantitative differences in the binding of common transcription factor (red). High levels of bound red TF to the enhancer lead to high levels of COLGALT2 transcription. In the cartilage (top box), the risk-allele C leads to higher DNAm levels at the enhancer than the non-risk allele T, resulting in less red TF binding to the enhancer, and therefore relatively low transcription of COLGALT2. The opposite occurs in the synovium (bottom box). The image was generated in BioRender and adapted in Affinity Designer 2.1.

Our investigation of the functional effects of the *COLGALT2* loci marked by rs11583641 and rs1046934 in both cartilage and synovium revealed that one of the signals, rs1046934, is not active across all joint tissues and does not contribute to the dysregulation of *COLGALT2* expression in synovium tissue from OA patients. The other signal, however, was found to affect gene expression in both tissue types. While the alleles of the SNP rs11583641 correlated with DNAm levels at CpGs at a *COLGALT2* enhancer in both tissues but in opposite directions, the effect of DNAm at the enhancer on *COLGALT2* expression was identical. These findings highlight the importance of performing targeted functional analyses of GWAS signals in all disease-relevant tissues and understanding the mechanisms driving any effects on gene expression before making decisions on potential therapeutic targets. In the case of the *COLGALT2* OA locus marked by rs11583641, a therapy targeted at increasing the DNAm at the enhancer in cartilage might decrease the expression of the gene in cartilage, preventing potential over-glycosylation of collagens leading to an increased formation of cross-links³⁰, which increases the stiffness of the ECM^{31,32} and preserves the structural stability of the tissue and its ability to withstand mechanical loading. However, if such an intervention is administered to the entire joint and not just the cartilage, the effect it might have in synovium could be deleterious and increase predisposition to OA. Based on the results presented in this chapter and in Chapter 3, I hypothesise that the OA risk allele C of the rs11583641 locus is responsible for exerting pleiotropic effects on various joint tissues. Specifically, in cartilage, it promotes heightened *COLGALT2* enzyme activity, leading to a compromise in collagen fibril integrity and cartilage resilience. Conversely, in the synovium, it results in reduced enzyme activity, impairing the tissue's ability to respond to injury and inflammation and to provide the underlying cartilage with adequate nutrition.

While risk variants having opposing effects in different tissues is rare, as evidenced by the epigenome-wide study of OA risk signals in disease-relevant tissues published in 2022¹⁰, it highlights the importance of targeted functional analyses in a multitude of tissues in order to obtain a full picture of the dynamics of gene regulation in health and in disease.

5.5. References

1. Sorial, A. K. *et al.* Multi-tissue epigenetic analysis of the osteoarthritis susceptibility locus mapping to the plectin gene PLEC. *Osteoarthr. Cartil.* **28**, 1448–1458 (2020).
2. Parker, E. *et al.* Multi-Tissue Epigenetic and Gene Expression Analysis Combined With Epigenome Modulation Identifies RWDD2B as a Target of Osteoarthritis Susceptibility. *Arthritis Rheumatol.* **73**, 100–109 (2021).
3. Rice, S. J. *et al.* Genetic and Epigenetic Fine-Tuning of TGFB1 Expression Within the Human Osteoarthritic Joint. *Arthritis Rheumatol.* **73**, 1866–1877 (2021).
4. Steinberg, J. *et al.* A molecular quantitative trait locus map for osteoarthritis. *Nat. Commun.* **12**, 1309 (2021).
5. Chou, C.-H. *et al.* Synovial cell cross-talk with cartilage plays a major role in the pathogenesis of osteoarthritis. *Sci. Rep.* **10**, 10868 (2020).
6. Zhou, Y. *et al.* Identification of differentially expressed miRNAs and mRNAs in synovial of osteoarthritis via RNA-sequencing. *BMC Méd. Genet.* **21**, 46 (2020).
7. Timur, U. T. *et al.* Identification of tissue-dependent proteins in knee OA synovial fluid. *Osteoarthr. Cartil.* **29**, 124–133 (2021).
8. Yuan, C. *et al.* Classification of four distinct osteoarthritis subtypes with a knee joint tissue transcriptome atlas. *Bone Res.* **8**, 38 (2020).
9. Steinberg, J. *et al.* Linking chondrocyte and synovial transcriptional profile to clinical phenotype in osteoarthritis. *Ann. Rheum. Dis.* **80**, 1070–1074 (2021).
10. Kreitmaier, P. *et al.* An epigenome-wide view of osteoarthritis in primary tissues. *Am. J. Hum. Genet.* **109**, 1255–1271 (2022).

11. Folkesson, E. *et al.* Proteomic comparison of osteoarthritic and reference human menisci using data-independent acquisition mass spectrometry. *Osteoarthr. Cartil.* **28**, 1092–1101 (2020).
12. Sun, H. *et al.* Single-cell RNA-seq analysis identifies meniscus progenitors and reveals the progression of meniscus degeneration. *Ann. Rheum. Dis.* **79**, annrheumdis-2019-215926 (2019).
13. Jeffries, M. A. *et al.* Genome-Wide DNA Methylation Study Identifies Significant Epigenomic Changes in Osteoarthritic Subchondral Bone and Similarity to Overlying Cartilage. *Arthritis Rheumatol.* **68**, 1403–1414 (2016).
14. Dunn, C. M., Nevitt, M. C., Lynch, J. A. & Jeffries, M. A. A pilot study of peripheral blood DNA methylation models as predictors of knee osteoarthritis radiographic progression: data from the Osteoarthritis Initiative (OAI). *Sci. Rep.* **9**, 16880 (2019).
15. Xiao, K. *et al.* Urine Proteomics Profiling and Functional Characterization of Knee Osteoarthritis Using iTRAQ Technology. *Horm. Metab. Res.* **51**, 735–740 (2019).
16. Abdelrazig, S. *et al.* Metabolic signatures of osteoarthritis in urine using liquid chromatography-high resolution tandem mass spectrometry. *Metabolomics* **17**, 29 (2021).
17. Li, J. *et al.* Multi-omics molecular biomarkers and database of osteoarthritis. *Database* **2022**, baac052 (2022).
18. Wang, Y. *et al.* Exosomes Secreted by Adipose-Derived Mesenchymal Stem Cells Foster Metastasis and Osteosarcoma Proliferation by Increasing COLGALT2 Expression. *Front. Cell Dev. Biol.* **8**, 353 (2020).
19. Guo, T. *et al.* COLGALT2 is overexpressed in ovarian cancer and interacts with PLOD3. *Clin. Transl. Med.* **11**, e370 (2021).
20. Zhang, X. *et al.* GLT25D2 Is Critical for Inflammatory Immune Response to Promote Acetaminophen-Induced Hepatotoxicity by Autophagy Pathway. *Front. Pharmacol.* **11**, 01187 (2020).
21. Hao, X. *et al.* Glt25d2 Knockout Directly Increases CD25+CD69– but Decreases CD25–CD69+ Subset Proliferation and is Involved in Concanavalin-Induced Hepatitis. *Cell. Physiol. Biochem.* **50**, 1186–1200 (2018).

22. Kaluscha, S. *et al.* Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat. Genet.* **54**, 1895–1906 (2022).
23. [review] Héberlé, É. & Bardet, A. F. Sensitivity of transcription factors to DNA methylation. *Essays Biochem.* **63**, 727–741 (2019).
24. 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).
25. Ge, X. *et al.* Functional genomics atlas of synovial fibroblasts defining rheumatoid arthritis heritability. *Genome Biol.* **22**, 247 (2021).
26. Smith, S. B. *et al.* Rfx6 directs islet formation and insulin production in mice and humans. *Nature* **463**, 775–780 (2010).
27. Kouraki, A. *et al.* Different genes may be involved in distal and local sensitization: A genome-wide gene-based association study and meta-analysis. *Eur. J. Pain* **26**, 740–753 (2022).
28. Arvanitis, M., Tayeb, K., Strober, B. J. & Battle, A. Redefining tissue specificity of genetic regulation of gene expression in the presence of allelic heterogeneity. *Am. J. Hum. Genet.* **109**, 223–239 (2022).
29. Chang, J.-H., Lee, K.-J., Kim, S.-K., Yoo, D.-H. & Kang, T.-Y. Validity of SW982 synovial cell line for studying the drugs against rheumatoid arthritis in fluvastatin-induced apoptosis signaling model. *Indian J. Méd. Res.* **139**, 117–24 (2014).
30. Dominguez, L. J., Barbagallo, M. & Moro, L. Collagen overglycosylation: A biochemical feature that may contribute to bone quality. *Biochem. Biophys. Res. Commun.* **330**, 1–4 (2005).
31. Pehrsson, M. *et al.* Enzymatic cross-linking of collagens in organ fibrosis – resolution and assessment. *Expert Rev. Mol. Diagn.* **21**, 1049–1064 (2021).
32. Mak, M. Impact of crosslink heterogeneity on extracellular matrix mechanics and remodeling. *Comput. Struct. Biotechnol. J.* **18**, 3969–3976 (2020).

Chapter 6. Genetic and Epigenetic Regulation of *COLGALT2* in Developmental Joint Tissues

6.1. Introduction

While OA is considered a disease of old age due to its late-life onset, the large genetic contribution to the predisposition to the disease and its strong correlation with joint morphology¹⁻³ has led to the question of whether OA has developmental origins. Further supporting this line of thinking, as discussed in Chapter 1, is the damaging impact of the attempt by chondrocytes from OA affected joints to repair articular cartilage, resulting in reversion of the chondrocytes to a more developmental-like phenotype⁴⁻⁶. A few studies so far have used developmental tissues/cells to understand OA predisposition^{3,7}.

Epigenetic effects associated with the genetic traits predisposing to disease development (e.g. QTLs) can also have links to development. Studies of neurological and metabolic disorders have shown that DNAm, as well as other epigenetic marks, are active during development and early life and lead to an increased predisposition to disease⁸⁻¹³. The phenotypic consequences of epigenetic modifications occurring during the developmental stages may only become apparent in subsequent stages in life triggered by environmental challenges. Such epigenetically driven developmental consequences could have a role to play in the initiation and progression of many complex diseases. However, to this day no studies have performed extensive detection of QTLs and a comparison between their effect in developmental and aged tissues relevant to musculoskeletal biology.

A large proportion of the genetic and epigenetic marks associated with common diseases such as OA, as discussed in Chapter 1, are located in non-coding regions of the genome and often reside within *cis*-acting regulatory elements such as transcriptional enhancers¹⁴⁻¹⁶. Enhancer features were originally identified in the context of developmental loci due to the dynamic nature inherent in developmental processes. Indeed, the comparative analysis of distinct developmental stages has the potential to unveil active enhancer features that may not be easily discernible within a more static, differentiated cell lineage. The functional redundancy, often attributed to developmental enhancers, serves to ensure

precise gene expression. Regulatory elements known as shadow enhancers exemplify this redundancy by driving similar expression patterns as primary enhancers. However, when functioning in concert, redundant enhancers contribute to a more faithful expression, particularly under suboptimal conditions. Redundant enhancers have been observed in mammals¹⁷ but the extent to which non-developmental genes rely on such enhancers remains to be established.

In Chapters 3 and 4 I described the genetic and epigenetic mechanisms driving *COLGALT2* regulation in articular cartilage chondrocytes via two independent OA GWAS signals marked by rs11583641 and rs1046934. Both signals were found to regulate the expression of the gene via differential DNAm at *cis*-acting enhancer elements. In the case of the rs1046934 locus, two separate enhancers were found to have the same regulatory effect on *COLGALT2*, demonstrating a potential regulatory redundancy similar to that observed in developmental gene regulation. Furthermore, in Chapter 5 I described that while the rs1046934 signal appears to not be active in all joint tissues, the rs11583641 signal was active in another joint tissue - synovium. The tissues used in these studies were derived from older individuals who had undergone arthroplasty.

To investigate if the QTLs described in Chapter 3 and 4 are active during development, I carried out the functional analyses that I had undertaken on arthroplasty patient samples on developmental foetal joint samples (cartilage and limb bud) provided by the Human Developmental Biology Resource (HDBR). The underlying hypothesis guiding these experiments is that the epigenetic processes governing the regulation of *COLGALT2* expression in aged human joint tissues are active during foetal joint development. If so, this would suggest a potential impact of OA genetic risk on the integrity of cartilage (and/or other joint tissues) from early life, which may contribute to the development of the disease with advancing age.

6.2. Aims and objectives

6.2.1. Aims

- Determine if the mQTLs that were detected in the arthroplasty cartilage, and which associated with the OA risk SNPs rs11583641 and rs1046934, are also present in developmental foetal joint samples.
- Determine if the OA-associated SNPs rs11583641 and rs1046934 act as eQTLs in the developmental samples.
- Compare the chromatin environment at the *COLGALT2* locus between OA arthroplasty and developmental cartilage chondrocytes.

6.2.2. Objectives

- Genotype developmental samples at rs11583641 and rs1046934.
- Measure DNAm levels at the 12 CpG sites associated with the OA signal rs11583641 in the developmental samples.
- Measure DNAm levels at the two CpG sites associated with the OA signal rs1046934 in the developmental samples.
- Perform AEI analysis in the developmental samples on *COLGALT2* for rs11583641.
- Perform AEI analysis in the developmental samples on *COLGALT2* and *TSEN15* for rs1046934.
- Generate whole-genome ATAC-seq data from OA and foetal cartilage chondrocytes. .
- Compare the chromatin accessibility at the *COLGALT2* locus between the OA and foetal cartilage chondrocytes.

6.3. Results

6.3.1. Detection of mQTLs associated with rs11583641 and rs1046934 in developmental joint tissue samples

For the investigation of the *COLGALT2* rs11583641 locus, the foetal cartilage samples (n=72) were genotyped at rs11583641 and DNAm levels were measured at the previously interrogated 12 CpGs located in an enhancer (Chapter 3). When the DNAm levels at the 12 CpGs were plotted against the genotype at the SNP, mQTLs were detected at 7/12 CpGs: CpG 4 (P=0.0001), CpG 5 (P<0.0001), CpG 6 (P<0.0001), CpG 7 (P<0.0001), CpG 8 (P<0.0001), CpG 9 (cg18131582, P<0.0001), and CpG 10 (P<0.001) (Figure 6.1). In comparison, in arthroplasty cartilage mQTLs were detected at 5/12 CpGs (Chapter 3, Figure 3.2A). In both developmental and aged cartilage, the risk allele C of rs11583641 is associated with lower DNAm levels compared to the non-risk allele T.

The DNAm levels at the enhancer were lower in foetal cartilage at 10/12 CpGs (Figure 6.2), with the mean differences at CpG 1, 3, 8 and 12 being <10% (0.06% (P>0.05), 1.55% (P>0.05), 4.55% (P=0.0005) and 8.9% (P<0.0001), respectively) while at the other CpGs the differences were larger: CpG 4 (13.76% (P<0.0001)), CpG 5 (16.05% (P<0.0001), CpG 6 (20.89% (P<0.0001)), CpG 7 (23.35% (P<0.0001)), CpG 9 (cg18131582, 23.09% (P<0.0001)) and CpG 10 (25.11% (P<0.0001).

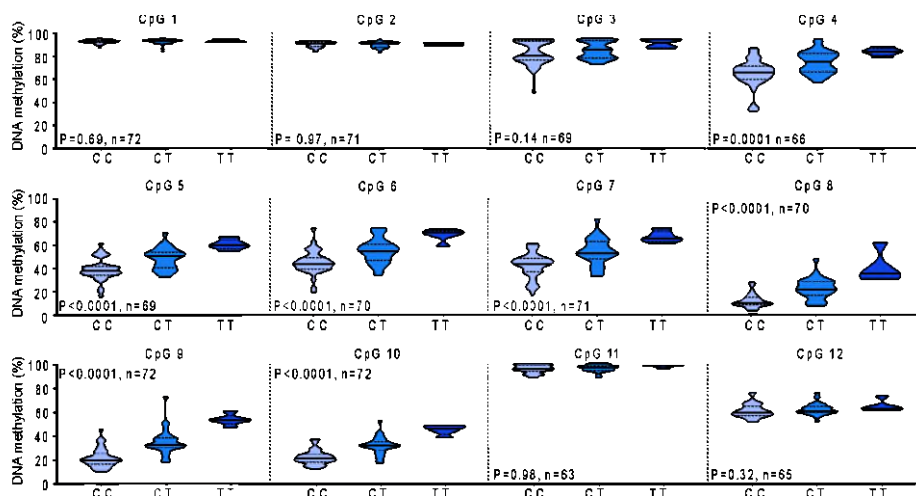


Figure 6.1. Cartilage mQTL analysis of 12 CpG sites within the COLGALT2 enhancer associated with rs11583641 in developmental samples. Violin plots showing DNAm values in foetal cartilage samples at the 12 investigated CpGs, stratified by rs11583641 genotype. Solid lines and dashed lines inside the plots represent median and interquartile range, respectively. P-values were calculated using simple linear regression. GraphPad Prism 9.0 was used for statistical analyses and data visualisation.

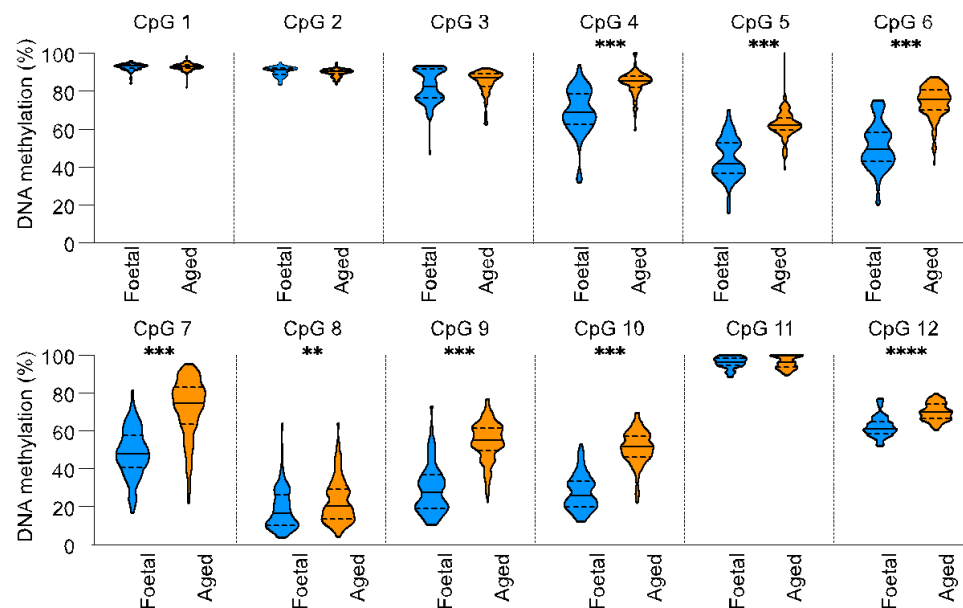


Figure 6.2. Comparison of the DNAm levels at the enhancer associated with rs11583641 between foetal and arthroplasty cartilage samples. Violin plots showing DNAm levels at the 12 investigated CpGs in developmental foetal (n= 65-72) and aged arthroplasty (n=102-126) cartilage samples. Solid and dashed horizontal lines represent the median and interquartile range, respectively. P-values were generated using Mann-Whitney tests. ** = $P < 0.01$; *** = $P < 0.001$, **** = $P < 0.0001$. GraphPad Prism 9.0 was used for statistical analyses and data visualisation.

The GE% was much stronger (>30% difference) in foetal cartilage compared to aged cartilage from arthroplasty samples at all differentially methylated CpG sites (Figure 6.3) except for CpG 8 (52% in arthroplasty samples and 51.4% in foetal cartilage).

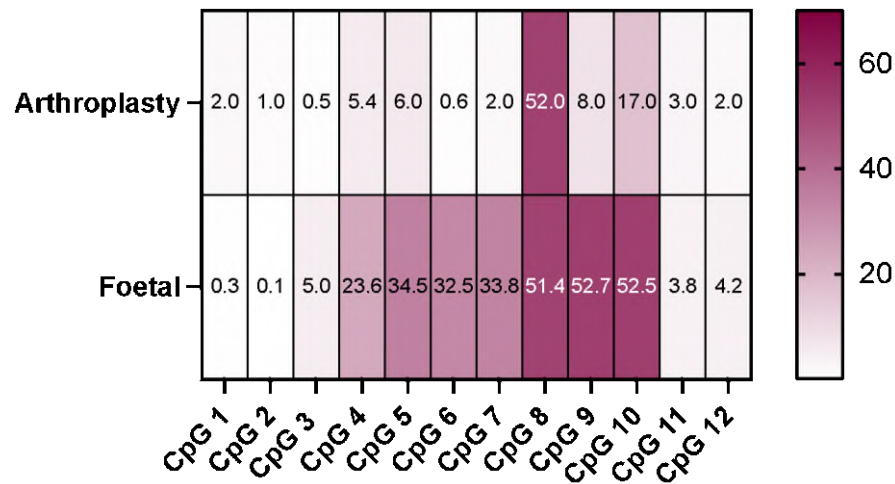


Figure 6.3. Heatmap showing the contribution of the rs11583641 genotype to the observed DNAm levels differences at the 12 CpGs in developmental foetal (n= 65-72) and aged arthroplasty (n=102-126) cartilage samples. The heatmap displays r^2 values calculated by linear regression analysis, converted to percentages. GraphPad Prism 9.0 was used for statistical analyses and data visualisation.

In the foetal limb bud samples (n=14-19) mQTLs were detected in 3/12 CpGs: CpG 8 (P=0.02), CpG 9 (cg18131582, P=0.0006) and CpG 10 (P=0.003) (Figure 6.4). These were the CpGs with the strongest GE% in foetal cartilage (Figure 6.3) and their detection in the much smaller number of limb bud samples is likely a reflection of the large contribution of rs11583641 genotype to DNAm at these three CpGs.

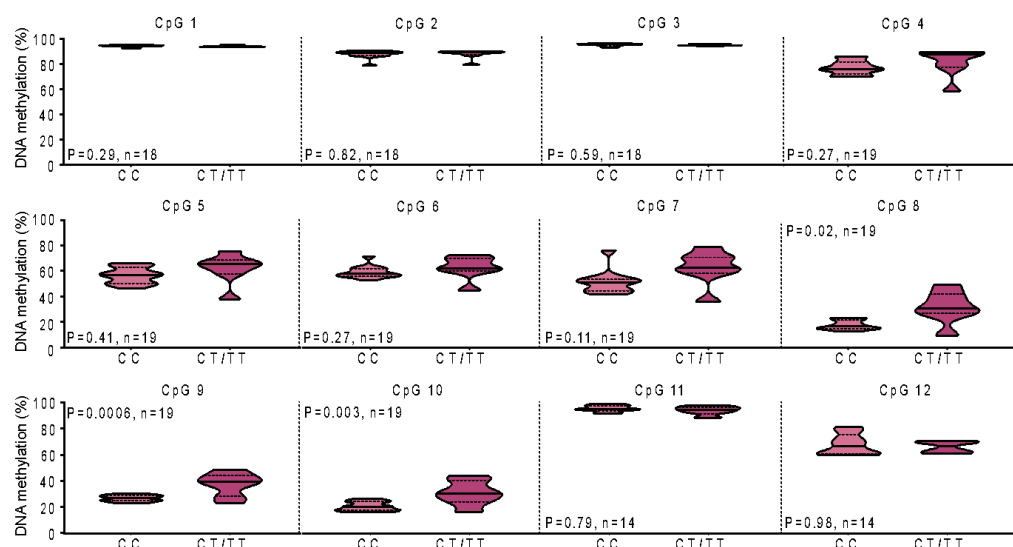


Figure 6.4. mQTL analysis of 12 CpG sites within the COLGALT2 enhancer associated with rs11583641 in developmental mixed tissue limb bud samples. Violin plots showing DNAm

values in foetal cartilage samples at the 12 investigated CpGs, stratified by rs11583641 genotype. Solid lines and dashed lines inside the plots represent median and interquartile range, respectively. P-values were calculated using simple linear regression. GraphPad Prism 9.0 was used for statistical analyses and data visualisation.

Since the joint origin (knee or hip) of the foetal cartilage and limb bud samples used in our DNAm analysis was undisclosed to us, a comparison was not undertaken of DNAm at the CpGs stratified by hip and knee.

For the rs1046934 locus, the arthroplasty mQTLs at cg15204595 and cg20606956 were detected in both foetal cartilage ($P=0.0001$ and $P=0.002$, respectively) (Figure 6.5A) and limb bud ($P=0.02$ and $P=0.009$, respectively) (Figure 6.5B). As observed in the aged cartilage (Figure 4.6B), the risk allele A at rs1046934 correlated with lower DNAm compared to the non-risk C allele at both differentially methylated CpGs in foetal samples. The mean DNAm levels at cg15204595 are higher in foetal (66.7%) compared to aged cartilage (62.83%) ($P=0.0002$) but are lower in the foetal compared to aged cartilage for cg20606956 (40.00% and 61.10%, respectively; $P<0.0001$) (Figure 6.6).

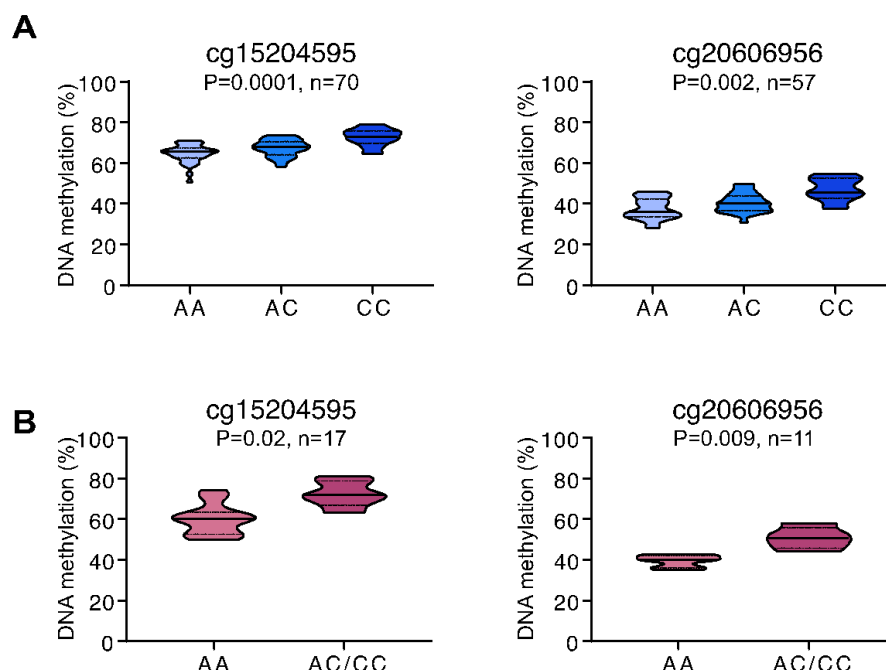


Figure 6.5. mQTL analysis at cg15204595 and cg21606956 in developmental samples. **A.** Violin plots showing DNAm values in foetal cartilage samples at cg15204595 (left) and cg21606956 (right), stratified by rs1046934 genotype. Solid lines and dashed lines inside the plots represent median and interquartile range, respectively. P-values were calculated using

simple linear regression. **B.** Same as A. for foetal limb bud samples. Due to the low number of CC homozygotes (<3), these samples were combined with AC heterozygotes for the limb bud analysis. P-values were calculated using simple linear regression. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A. and B.).

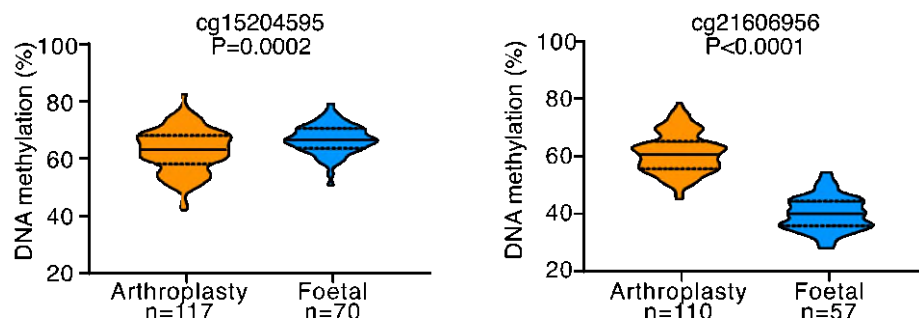


Figure 6.6. Comparison of the DNAm levels at cg15204595 and cg21606956 between foetal and arthroplasty cartilage samples. Violin plots showing DNAm levels at the 12 investigated CpGs in developmental foetal and aged arthroplasty cartilage samples. Solid and dashed horizontal lines represent the median and interquartile range, respectively. P-values were calculated using Mann-Whitney tests. GraphPad Prism 9.0 was used for statistical analyses and data visualisation.

6.3.2. Identification of COLGALT2 allelic expression imbalance in developmental joint tissue

Since the effects of rs11583641 and rs1046934 on DNAm at the COLGALT2 enhancers were detected in developmental samples, the presence of eQTLs was also investigated.

AEI analysis in foetal cartilage samples heterozygous for rs11583641 revealed that the risk allele C is associated with increased COLGALT2 expression (1.15-fold increase, $P=0.01$) (Figure 6.7A). The mean effect observed in arthroplasty cartilage was stronger (1.94-fold increase, Chapter 3, Figure 3.15A). As observed in the aged cartilage, some foetal cartilage samples do not show AEI (Figure 6.7A).

No AEI was detected in limb bud samples ($P=0.47$) (Figure 6.7B). The presence of an eQTL in these samples, however, cannot be ruled out, given the relatively small sample size (6 limb bud versus 20 cartilage) and the observed absence of AEI in many of the cartilage samples.

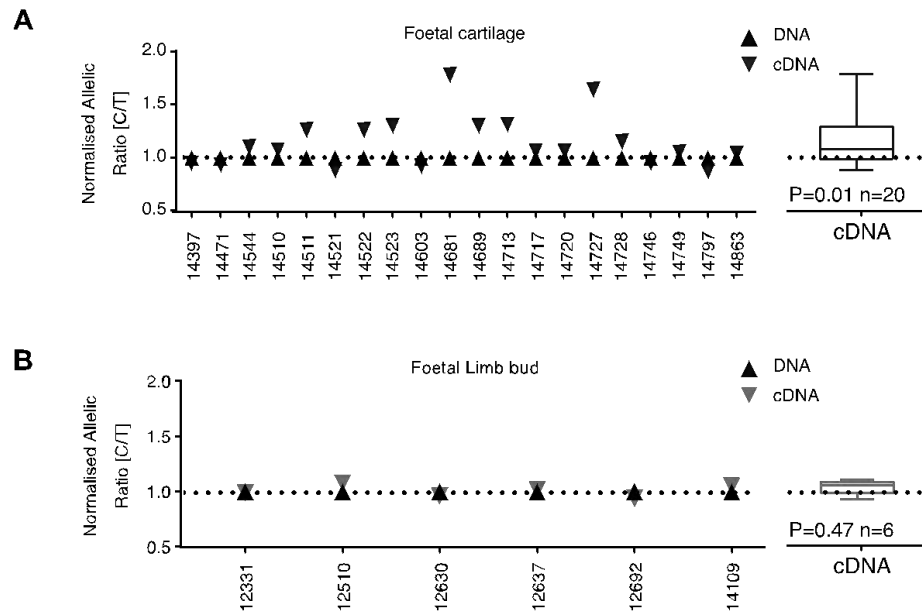


Figure 6.7. AEI analysis of *COLGALT2* for the rs11583641 locus in developmental cartilage samples. **A.** Left, Allelic (C/T) ratios in foetal cartilage samples heterozygous for rs11583641. In each sample, the ratio of values for cDNA and DNA between the OA-risk allele C and the non-risk T allele was plotted; each symbol represents 1 of 3 technical repeats. Right, Mean DNA and cDNA values in the presence of the C allele versus the T allele in heterozygous patients. Values are shown as a box plot, with the line inside the box representing the median, the box showing the interquartile range, and whiskers showing the minimum and maximum values. P-value was calculated using Wilcoxon's matched pairs signed rank test. **B.** Same as A. for foetal limb bud samples. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A. and B.).

No meQTLs for the 12 rs11583641 enhancer CpG sites were detected in foetal cartilage nor in limb bud samples (Figure 6.8A and B, respectively). In arthroplasty cartilage meQTLs were detected for CpGs 4, 8 and 9 (cg18131582) (Chapter 3, Figure 3.15C).

For the rs1046934 locus, AEI was identified for *COLGALT2* and *TSEN15* in foetal cartilage (Figure 6.9A and B, respectively), mirroring the pattern observed in human arthroplasty cartilage (Chapter 4, Figure 4.4). The OA risk allele C at rs114661926 (corresponding to the risk allele A of rs1046934) exhibited an average 1.35-fold increase in *COLGALT2* expression ($P < 0.0001$), while the OA risk allele G at rs2274432 (corresponding to the risk allele A of rs1046934) showed an average 1.03-fold increase in *TSEN15* expression ($P = 0.04$). Only one foetal limb bud sample passed AEI QC and AEI analysis of this tissue was not therefore performed.

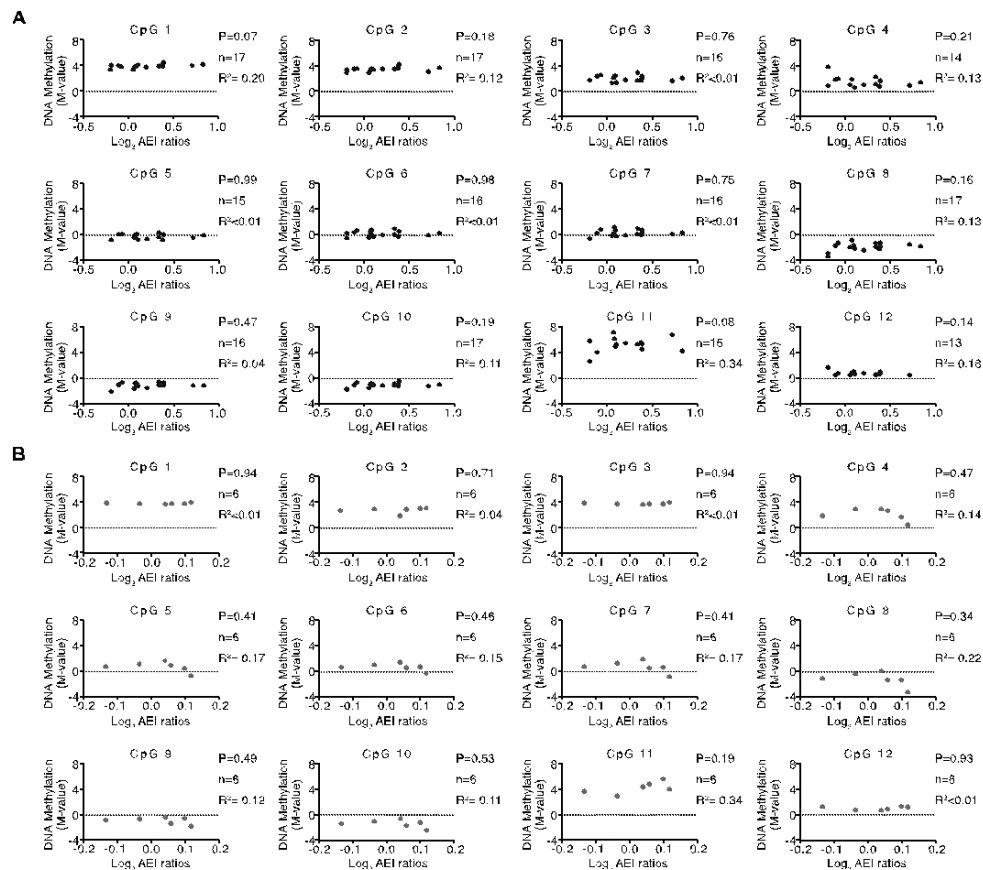


Figure 6.8. Relationship between DNAm at the enhancer and *COLGALT2* AEI ratios. **A.** Allelic ratios ($\log_2$) of *COLGALT2* (x-axis) are plotted against matched DNAm levels (M-values) (y-axis) at the investigated CpGs (CpG 1-12). Each dot represents data from 1 foetal cartilage sample. P-values were calculated using simple linear regression and the test statistics are shown to the right of each plot (P-value, R^2 and number of samples used for the analysis). **B.** Same as A. for foetal limb bud samples.

meQTLs for the rs1046934 CpGs cg15204595 and cg20606956 were observed in the foetal cartilage samples for *COLGALT2* ($P=0.02$ and $P=0.0001$, respectively) but not for *TSEN15* ($P=0.27$ and $P=0.29$, respectively) (Figure 6.10A and B), confirming that while the SNP rs1046934 may affect the expression of both genes, the expression of *COLGALT2* is specifically contingent upon the levels of DNAm at the enhancer regions containing cg15204595 and cg20606956. As observed in arthroplasty cartilage (Chapter 4, Figure 4.7A), the direction of the correlation between DNAm levels and *COLGALT2* expression at cg15204595 and cg20606956 is in opposing directions in foetal cartilage (Figure 6.10A) despite the individual effects of the SNP upon DNAm at the CpGs (Figure 6.5A) and upon the expression of the gene (Figure 6.9A) being in the same direction. These observations in foetal cartilage are in line with the theoretical model explaining this phenomenon described in Chapter 4 (Figure 4.9).

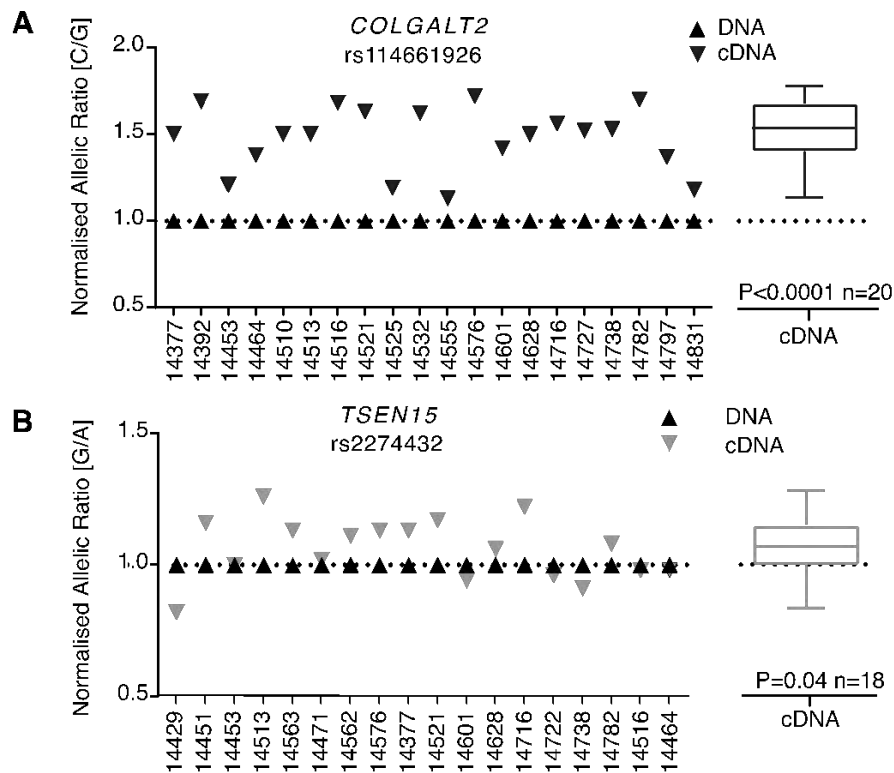


Figure 6.9. AEI analysis of *COLGALT2* and *TSEN15* in foetal cartilage samples for the rs1046934 locus. **A.** Left, Allelic (C/G) ratios in foetal cartilage samples heterozygous for *COLGALT2* transcript SNP rs11466192. In each sample, the ratio of values for cDNA and DNA between the OA-risk allele and the non-risk allele was plotted; each symbol represents the mean of 3 technical repeats. Right, Mean DNA and cDNA values in the presence of the risk allele versus the non-risk allele in heterozygous samples. Values are shown as a box plot, with the line inside the box representing the median, the box showing the interquartile range, and whiskers showing the minimum and maximum values. P-value was calculated using Wilcoxon's matched pairs signed rank test. **B.** Same as A. for *TSEN15* SNP rs2274432 (G/A; risk/non-risk). GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A. and B.).

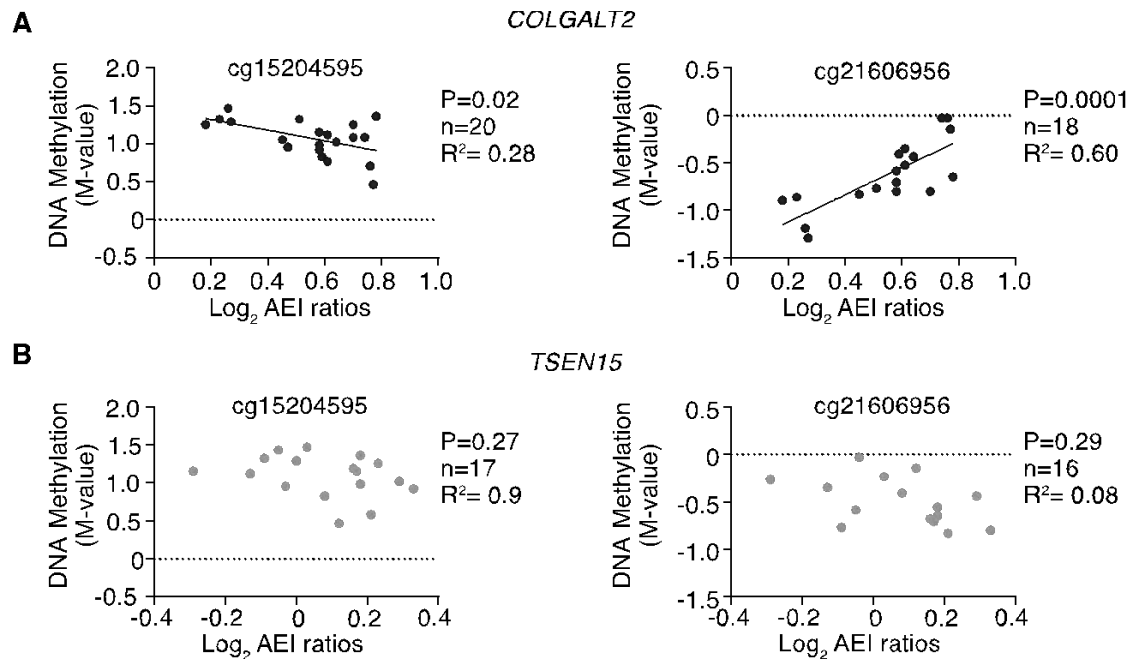


Figure 6.10. Relationship between DNAm at the enhancer and *COLGALT2* AEI ratios in foetal cartilage samples. **A.** Graphs show allelic ratios ($\log_2$) of *COLGALT2* (rs114661926) plotted against matched DNAm levels (M-values) at cg15204595 and cg21606956. Each dot represents data from 1 individual. P-values were calculated using simple linear regression. **B.** Same as A. for *TSEN15* (rs2274432). GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A. and B.).

6.3.3. Characterisation of chromatin accessibility at the *COLGALT2* locus in developmental and aged cartilage chondrocytes

While the mQTLs and eQTLs observed at the OA loci marked by rs11583641 and rs1046934 are present in both aged and developmental samples, the differences in the strength of the signals between these samples required further investigation. Data on genome-wide chromatin accessibility was generated (utilizing ATAC-seq) from OA patients' cartilage obtained during knee arthroplasty (n=5) and hip arthroplasty (n=5) procedures. Additionally, chromatin accessibility data was obtained from the cartilage of the proximal end (hip, n=6) and distal end (knee, n=6) of foetal femora (Chapter 2, Table 2.4).

The data revealed 78,219 open chromatin regions shared by all examined tissues (OA hip, OA knee, foetal hip, foetal knee). Among these, 49,922 open regions were exclusively found in foetal samples, while 63,058 were exclusive to aged samples. Furthermore, a comparison between foetal and aged cartilage revealed 113,887 differentially accessible regions at the hip joint and 121,050 such regions at the knee joint (Figure 6.11).

Comprehensive analysis and interpretation of the whole-genome data was carried out by Dr Sarah Rice and Mr John Casement and published as part of a larger project¹⁸. This chapter is focused on the chromatin environment at the *COLGALT2* locus and encompassing the two independent OA association signals: rs11583641 and rs1046934.

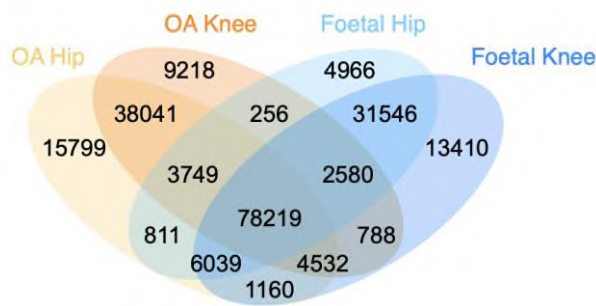


Figure 6.11. Differential chromatin accessibility between foetal and OA cartilage. Venn diagram displaying the number of differentially accessible regions between OA hip (light orange), OA knee (dark orange), Foetal hip (light blue) and Foetal knee (dark blue) cartilage. R-studio was used for data visualisation.

Since both rs11583641 and rs1046934 form part of larger LD blocks, I used the chromatin accessibility data to determine if any of the SNPs in the LD blocks fall within open chromatin regions in arthroplasty and/or foetal cartilage (Figure 6.11) and could therefore be prioritised as potential causal SNPs driving the observed effects.

Out of the 24 SNPs in high LD ($r^2 > 0.8$, GBR) with rs11583641, two fall within open chromatin regions (Figure 6.12A, panels b and d): rs10911467 ($r^2 = 1.00$, GBR) and rs734657 ($r^2 = 0.83$, GBR). rs10911467 resides within an intron of *RGL1*, in a region marked as an enhancer in MSC-derived adipocytes, adipose-derived MSCs and osteoblasts, and as a flanking TSS in MSC-derived chondrocytes (Figure 6.12A, panels a and c). The second SNP, rs734657, is located within the differentially methylated enhancer investigated in Chapters 3 and 5; the reporter gene assays from these chapters suggest that this SNP has allelic effects on gene transcription (Figures 3.6 and 5.6).

The LD block encompassing rs1046934 includes 20 SNPs (Figure 6.12B, panel b). Among these, three SNPs, rs114661926 ($r^2 = 0.86$, GBR), rs74767794 ($r^2 = 0.88$, GBR), and rs2274432 ($r^2 = 1.00$, GBR), are situated in open chromatin regions observed in both OA and

foetal samples (both hip and knee) (Figure 6.12B, panels b and d). The first two SNPs are positioned in intron 1 and the 5'UTR of *COLGALT2* (Figure 6.12B, panels a and b), respectively, within regions annotated as a TSS in joint-relevant cell types (Figure 6.12B, panel c). rs2274432 is also situated in a TSS, within the 5'UTR of *TSEN15* (Figure 6.12B, panels a, b, and c).

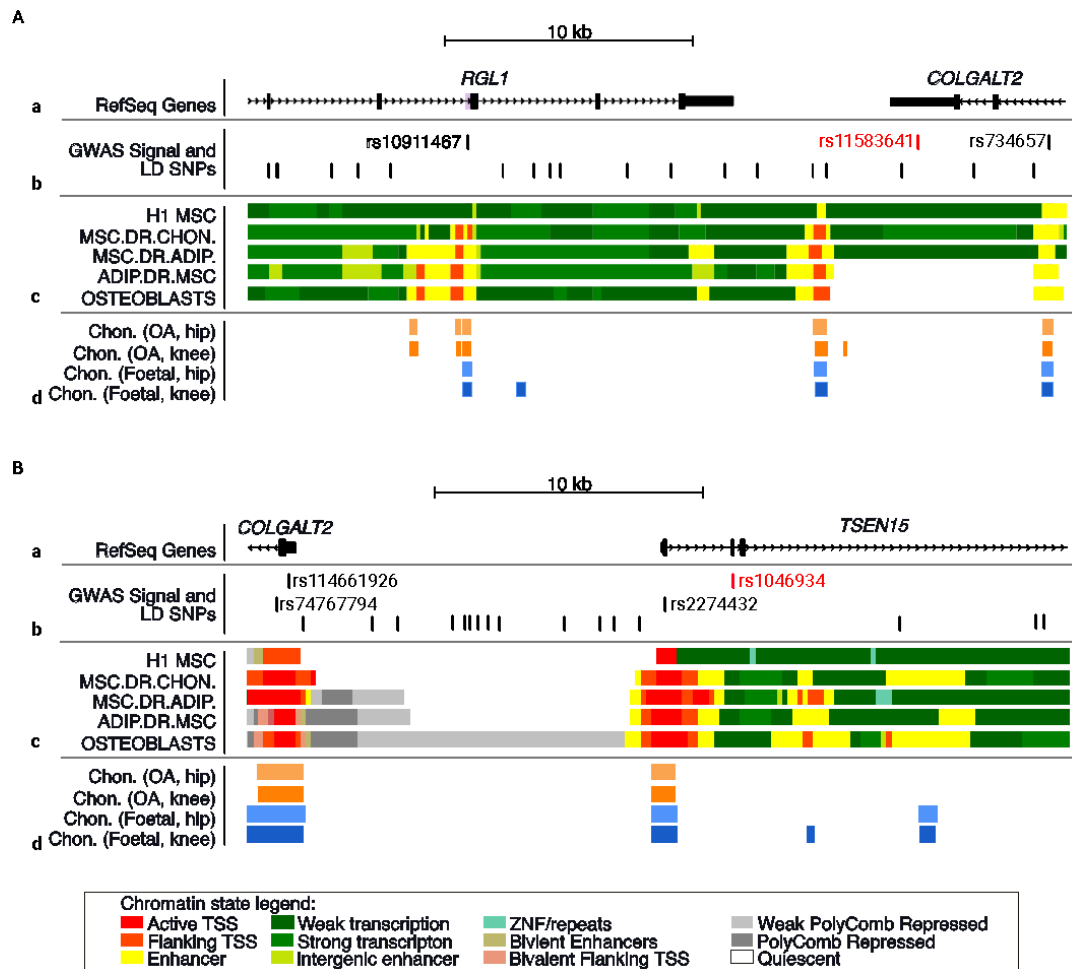


Figure 6.12. Prioritisation of SNPs at the *COLGALT2* locus based on chromatin accessibility in OA and foetal chondrocytes and on histone marks in relevant cell types. **A.** Schematic overview of the rs11583641 signal. **Panel a,** Relative genomic position of portions of the *COLGALT2* and *RGL1* genes, visualized using the UCSC Genome Browser (hg19). **Panel b,** The location of rs11583641 (red line) and the SNPs in high LD (pairwise $r^2 > 0.8$) with it (black lines). The two SNPs in open chromatin (rs10911467 and rs734657) are highlighted in purple. **Panel c,** Chromatin state data, as determined from the Roadmap project, for primary human mesenchymal stem cells (H1 MSC), MSC-derived chondrocytes (MSC.DR.CHON) and adipocytes (MSC.DR.ADIP), adipose-derived MSCs (ADIP.DR.MSC), and human osteoblasts. Colours correspond to different chromatin states, as indicated in legend at bottom. **Panel d,** Open regions marked by ATAC-sequencing peaks in hip and knee chondrocytes from OA (hip, light orange blocks; knee, dark orange blocks) and foetal (hip, light blue blocks; knee, dark blue

blocks) cartilage samples. **B.** Schematic overview of the rs1046934 signal. **Panel a**, Relative genomic position of portions of the COLGALT2 and TSEN15 genes, visualized using the UCSC Genome Browser (hg19). **Panel b**, The location of rs1046934 (red line) and the SNPs in high LD (pairwise $r^2 > 0.8$) with it (black lines). The three SNPs in open chromatin (rs114661926, rs74767794, and rs2274432) are highlighted in purple. **Panel c and Panel d**, same as for **A**.

I next examined the chromatin accessibility status of the regions containing the rs11583641 and rs1046934 mQTL enhancer CpGs (Figure 6.13).

For rs11583641, I focussed on the eight CpGs that had shown evidence of being mQTLs in the earlier studies; CpG 3-10. These are located within or just outside of ATAC-seq peaks in OA (hip and knee) and foetal (hip and knee) chondrocytes (Figure 6.13A). The level of accessibility of the chromatin corresponds to the height of the ATAC-seq peaks; the peaks in foetal chondrocytes are higher and broader than those in OA chondrocytes.

For rs1046934, cg15204595 does not fall within an open chromatin peak but is located close to one (Figure 6.13B). This peak is broader and higher in OA chondrocytes compared to foetal chondrocytes. The second rs1046934 CpG, cg21606956, is located within an open chromatin region in both OA and foetal chondrocytes, with the region being more accessible in knee chondrocytes (Figure 6.13C).

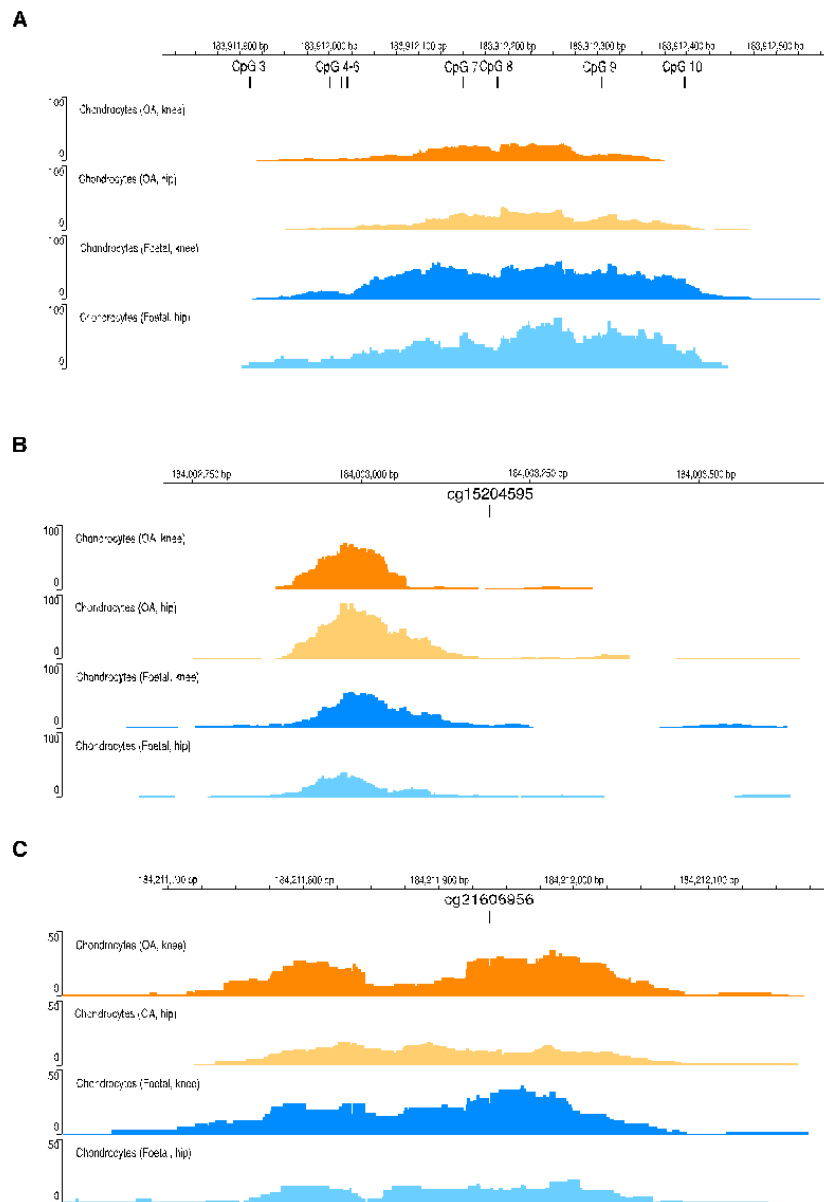


Figure 6.13. Chromatin accessibility at the *COLGALT2* enhancer CpGs. **A.** Genomic position of the differentially methylated CpGs associated with rs11583641 along Chromosome 1 (hg 19) visualised in the Integrative Genomics Viewer (hg19). The peaks show the average accessibility of the region in chondrocytes from OA knee (dark orange), OA hip (light orange), foetal knee (dark blue), and foetal hip (light blue) chondrocytes. The height of a peak corresponds to the level of accessibility with higher peaks marking more accessible areas. **B. and C.** Same as **A.** for the regions containing cg15204595 and cg21606956, respectively.

6.4. Discussion

In this chapter the genetic and epigenetic regulation of *COLGALT2* in developmental cartilage samples were investigated in to better understand the developmental origins of OA. The QTLs, associated with both expression and methylation, detected in OA-relevant tissues

from aged arthroplasty samples for the OA GWAS signals marked by rs11583641 and rs1046934, were detected in cartilage and mixed tissue limb bud samples of developmental origin.

The AEI and the mQTL effects detected in arthroplasty cartilage for the rs11583641 signal were also detected in foetal cartilage and the effects upon DNAm levels at the enhancer were stronger than those observed in aged cartilage. Based on the data presented in this chapter, I hypothesise that DNAm levels at the enhancer undergo precise regulation during development, requiring lower methylation levels to tightly regulate the expression of *COLGALT2* as the joint develops. As individuals age, methylation levels at the enhancer increase to ensure appropriate collagen glycosylation in aging cartilage. Conversely, in cartilage from joints affected by OA, there is a partial reversal of this process, evidenced by lowered DNAm levels compared to those in non-OA-affected NOF control samples (Chapter 3, Figure 3.4A), leading to dysregulated *COLGALT2* expression. This observed phenomenon may represent an attempt by chondrocytes in OA-affected joints to initiate cartilage repair, resulting in a more developmental-like phenotype.

Currently the prevalent hypothesis is that the crosstalk between genetic variants, epigenetic marks and gene expression regulation is carried out by TFs^{19,20}. As previously mentioned, TF binding can be methylation dependent, but the accessibility of the chromatin largely governs TF binding. The lower DNAm levels in foetal cartilage also suggest that TFs can bind more easily to the region^{21,22}. The ATAC-seq data generated suggest that the chromatin region harbouring the differentially methylated rs11583641 CpGs is more accessible in foetal samples compared to aged ones (Figure 6.13A). These data imply that the enhancer is more active in developmental cartilage and could be under tighter regulatory control by TFs.

The mQTLs at this enhancer were present in foetal limb bud tissue but only at CpGs 8, 9 and 10 (Figure 6.4). It is plausible that these effects primarily stem from the cartilage component within the more heterogeneous tissues of the limb bud, suggesting that the differentially methylated region is limited to these CpGs because the effect is not present in other tissues and is only detectable where it is strongest in cartilage. The significance of these three CpGs was underscored by CRISPR/Cas9 experiments conducted in Chapters 3 and 5, where targeted changes in DNA methylation at these specific CpGs led to the most substantial alterations in *COLGALT2* expression. Alternatively, it's possible that in other joint tissues, the

mQTL effect operates in the opposite direction to what was observed in foetal cartilage (as seen in aged synovium in Chapter 5, Figure 5.2A), and in a heterogenous tissue, it gets counterbalanced by the cartilage component of the mixture, with only the CpGs where the effect is most pronounced being detected. Additionally, it is important to acknowledge that the limb bud samples originate from earlier developmental stages (Chapter 2, Table 2.3), and the epigenetic influences on the regulatory elements may differ from those observed in later stages of development.

The eQTL detected in the developmental samples for the rs11583641 signal were not present in all foetal samples, a pattern that was observed in Chapter 3 for arthroplasty cartilage (Figure 3.15A). Additional but unrelated genetic variation could potentially explain these differences in expression between the samples. Further investigation into this could provide interesting and useful insights into the regulation of *COLGALT2* as the effect appears to be cancelling out the one caused by the OA-risk allele of rs11583641 which leads to increased *COLGALT2* expression in OA cartilage.

The second OA locus associated with *COLGALT2* expression in aged cartilage, rs1046934, was also studied in foetal samples and mQTLs were detected in both foetal cartilage and limb bud samples (Figure 6.5). The DNAm levels at cg15204595 were found to be higher in developmental cartilage samples than in aged ones (Figure 6.6) and the open chromatin peaks detected in very close proximity to the CpG were smaller and less broad (Figure 6.13B). On the other hand, the DNAm at cg21606956 is higher in developmental samples and the open chromatin peak which marks the region harbouring the CpG does not appear to be broader or higher in arthroplasty chondrocytes (Figure 6.13C). Differences in the accessibility at the region between knee and hip chondrocytes appear to be present (Figure 6.13C) but since no data on the joint origin of the foetal samples used for the mQTL analysis was available, the DNAm values measured at the CpG could not be stratified by joint site. Since the two enhancers were shown to regulate *COLGALT2* expression in arthroplasty cartilage via the same mechanism (differential DNAm levels) and in the same direction, it is possible that these enhancers regulate expression via the redundancy method briefly mentioned previously with one of the two enhancers serving as a “shadow” enhancer that is only used to fine-tune gene expression and ensure its faithful regulation¹⁷. In OA cartilage the co-operation of the two enhancers might be disrupted, with both enhancers being actively

involved in gene expression activation leading to levels of higher levels of *COLGALT2* expression compared to those measured in healthy aged tissue²³. CRISPR/Cas9 experiments of modulating the enhancers in developmental as well as in healthy aged chondrocytes could be used to test this hypothesis.

The rs1046934 eQTLs detected in arthroplasty cartilage were also detected in foetal cartilage, with the effects being observed in the same direction; risk A allele associating with higher expression of both *COLGALT2* and *TSEN15* (Figures 4.4 and 6.9). The *COLGALT2* meQTLs detected in arthroplasty cartilage (were also detected in foetal cartilage, with the slopes of the meQTLs at cg15204595 and cg20606956 in foetal operating in the same opposite directions as observed for arthroplasty (Figures 4.8A and 6.10A). The foetal data is therefore consistent with the explanation proposed earlier to explain the apparent discrepancy in the cg15204595 and cg20606956 *COLGALT2* meQTLs (Figure 4.9).

For the rs11583641 and rs1046934 OA signals, regional chromatin accessibility data combined with chromatin state annotations from the ROADMAP project²⁴ highlighted SNPs from the LD blocks that could potentially be causal and driving the observed effects, prioritizing them for further analyses. For the rs11583641 LD block containing, one of the two SNPs that fall within accessible chromatin regions is rs734657, located within the differentially methylated enhancer (Figure 6.12A). Lucia reporter gene assay results (Figures 3.6 and 5.6) highlight genotype at this SNP as potentially modulating gene expression and Lucia activity. However, in patient samples the correlation between the SNP and DNAm at the enhancer CpGs was in general weaker than that observed for the rs11583641 SNP (Section 3.3.5). While this highlights the SNP as a potential causal variant further *in vitro* investigations, including CRISPR/Cas9 directed base editing techniques in relevant patient derived primary cells, are needed to determine if it is driving the observed effects.

For the rs1046934 LD block, three SNPs were highlighted as potentially causal based on their location within open chromatin regulatory regions (Figure 6.12B). These SNPs reside in TSS associated with both *COLGALT2* and *TSEN15*, confirming their capacity to regulate the expression of both genes. While it is plausible that only one SNP per LD block is causal, it is possible that multiple SNPs work in unison (a haplotype effect) to exert an impact on gene expression and/or DNAm levels at CpGs and this might be the mechanism through which this

signal exerts effects on both *COLGALT2* (via DNAm levels at cg15204595 and cg20606956) and *TSEN15*.

Even though OA primarily affects older individuals, research has indicated that the susceptibility to OA may have its roots in early development³⁷. The rs11583641 and rs1046934 signals, that mediate their effect by regulating the expression of *COLGALT2*, are only two of a number of other OA genetic signals reported to be functionally active during joint development¹⁸.

Investigating the regulation of *COLGALT2* in developmental cartilage provides valuable insights into the origins of OA. Genetic and epigenetic mechanisms influencing certain OA-risk loci during early development may predispose individuals to the disease later in life. By targeting the developmental stages where the disease begins, innovative therapeutic strategies could be developed, offering more precise and effective interventions. Current therapy development strategies focus on restoring cartilage by enhancing chondrocyte homeostasis²⁵, preventing ROS accumulation²⁶, combating senescence²⁷ or even engineering artificial cartilage^{28,29}. However, given the significant genetic contribution to OA and the likelihood that some predispositional mechanisms are active during development, approaches such as gene therapies or using the CRISPR-Cas9 system to modulate the (epi)genome may provide more targeted and impactful treatments.

6.5. References

1. Richard, D. *et al.* Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk. *Cell* **181**, 362–381.e28 (2020).
2. [review] Loughlin, J. Genetic contribution to osteoarthritis development. *Curr. Opin. Rheumatol.* **27**, 284–288 (2015).
3. Muthuirulan, P. *et al.* Joint disease-specificity at the regulatory base-pair level. *Nat. Commun.* **12**, 4161 (2021).
4. [review] Pitsillides, A. A. & Beier, F. Cartilage biology in osteoarthritis—lessons from developmental biology. *Nat. Rev. Rheumatol.* **7**, 654–663 (2011).

5. Singh, P., Marcu, K. B., Goldring, M. B. & Otero, M. Phenotypic instability of chondrocytes in osteoarthritis: on a path to hypertrophy. *Ann. N. York Acad. Sci.* **1442**, 17–34 (2019).
6. Saito, T. *et al.* Transcriptional regulation of endochondral ossification by HIF-2 α during skeletal growth and osteoarthritis development. *Nat. Med.* **16**, 678–686 (2010).
7. 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).
8. Jaffe, A. E. *et al.* Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* **19**, 40–47 (2016).
9. [review] Perkovic, M. N. *et al.* Epigenetics of Alzheimer’s Disease. *Biomolecules* **11**, 195 (2021).
10. Jager, P. L. D. *et al.* Alzheimer’s disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).
11. [review] Starr, J. M. Ageing and epigenetics: linking neurodevelopmental and neurodegenerative disorders. *Dev. Med. Child Neurol.* **61**, 1134–1138 (2019).
12. Steegers-Theunissen, R. P. *et al.* Periconceptional Maternal Folic Acid Use of 400 μ g per Day Is Related to Increased Methylation of the IGF2 Gene in the Very Young Child. *PLoS ONE* **4**, e7845 (2009).
13. Godfrey, K. M. *et al.* Epigenetic Gene Promoter Methylation at Birth Is Associated With Child’s Later Adiposity. *Diabetes* **60**, 1528–1534 (2011).
14. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16**, 197–212 (2015).
15. Gaffney, D. J. *et al.* Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol.* **13**, R7 (2012).
16. [review] Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front. Genet.* **11**, 424 (2020).
17. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow Enhancers as a Source of Evolutionary Novelty. *Science* **321**, 1314–1314 (2008).

18. Rice, S. J. *et al.* Genetic risk of osteoarthritis operates during human skeletogenesis. *Hum. Mol. Genet.* **32**, 2124–2138 (2022).
19. Lempiäinen, J. K. & Garcia, B. A. Characterizing crosstalk in epigenetic signaling to understand disease physiology. *Biochem. J.* **480**, 57–85 (2023).
20. [review] Wang, H., Lou, D. & Wang, Z. Crosstalk of Genetic Variants, Allele-Specific DNA Methylation, and Environmental Factors for Complex Disease Risk. *Front. Genet.* **9**, 695 (2019).
21. Kaluscha, S. *et al.* Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat. Genet.* **54**, 1895–1906 (2022).
22. [review] Héberlé, É. & Bardet, A. F. Sensitivity of transcription factors to DNA methylation. *Essays Biochem.* **63**, 727–741 (2019).
23. 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. *Osteoarthr. Cartil.* **27**, 1545–1556 (2019).
24. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
25. [review] Kim, H. *et al.* The current state of the osteoarthritis drug development pipeline: a comprehensive narrative review of the present challenges and future opportunities. *Ther. Adv. Musculoskelet. Dis.* **14**, 1759720X221085952 (2022).
26. Wu, J. *et al.* ROS-responsive PPGF nanofibre membrane as a drug delivery system for long-term drug release in attenuation of osteoarthritis. *npj Regen. Med.* **7**, 66 (2022).
27. [review] Rizzo, M. G. *et al.* Therapeutic Perspectives for Inflammation and Senescence in Osteoarthritis Using Mesenchymal Stem Cells, Mesenchymal Stem Cell-Derived Extracellular Vesicles and Senolytic Agents. *Cells* **12**, 1421 (2023).
28. Rua, L. A. *et al.* Engineered nasal cartilage for the repair of osteoarthritic knee cartilage defects. *Sci. Transl. Med.* **13**, eaaz4499 (2021).
29. [review] Guilak, F., Estes, B. T. & Moutos, F. T. Functional tissue engineering of articular cartilage for biological joint resurfacing. *J. Orthop. Res.* **40**, 1721–1734 (2022).

Chapter 7. Role of the COLGALT2 enzyme in chondrocyte biology

7.1. Introduction

As discussed in Chapter 1, collagens are the most abundant proteins in the human body, performing a variety of functions from providing structural support to facilitating molecular processes such as cell adhesion, signal transduction, and maintaining tissue integrity¹. They contribute to the mechanical strength, stability, and functionality of collagen-rich tissues such as skin, bone, cartilage, tendons, and blood vessels. To produce functional collagen molecules, procollagen peptides undergo extensive PTMs which are fundamental for the proper assembly of fibrils, the formation of functional collagen networks, and the regulation of collagen turnover in tissues². Chapters 3 to 6 provide an in-depth examination of the functional characterization of GWAS signals associated with OA in two joint relevant tissues - synovium and cartilage. The analyses encompassed both healthy and diseased cartilage and included investigations in developmental and aged state. The target of these two signals in both tissues and across all conditions was found to be *COLGALT2*, a gene encoding one of the two enzymes responsible for carrying out the first of two steps modifying hydroxylysine to glucosylgalactosylhydroxylysine residues³. Changes in collagen glycosylation have been linked to various bone-related conditions, including osteogenesis imperfecta^{4,5}, bone diseases such as osteosarcoma⁶ osteofibrous dysplasia⁶, as well as melanoma⁷, cerebral small vessel abnormality and porencephaly⁸. Unlike its ortholog, *COLGALT1*, which is ubiquitously expressed, *COLGALT2* is expressed in fewer tissues, predominantly brain³. The results of Chapters 3-6 demonstrate that *COLGALT2* is also expressed in joint-related tissues – cartilage and synovium. While both COLGALT enzymes possess the same galactosyltransferase activity *in vitro*³, the tissue specific expression pattern of *COLGALT2* suggest that they play distinct roles, potentially mediating the glycosylation of specific collagens.

In order to study the biological function of this protein in the context of chondrocyte biology, a reliable *in vitro* model capturing the dynamics and phenotype of chondrocytes was required.

TC28a2 cells are the gold standard for investigating cartilage processes in an immortalised human chondrocyte model. Derived from human costal chondrocytes, their homogeneity and amenability to genetic manipulation render them invaluable for dissecting molecular pathways and gene functions in chondrocyte physiology. This cell line provides a consistent and reproducible platform, crucial for validating experimental outcomes. However, the relevance of TC28a2 cells in disease-specific contexts, particularly OA, is limited. Originating from healthy tissue and not from joints typically afflicted by OA (e.g., hip or knee), these cells may not accurately mirror the pathophysiological state of diseased cartilage. Furthermore, the phenotype of TC28a2 in monolayer culture raises concerns about limitations in recapitulating the physiological state, which is essential for the accurate modelling of cartilage diseases⁹.

In contrast, primary chondrocytes, often sourced from OA patients or aged donors, offer a more pathologically relevant model. These cells, however, are encumbered by significant drawbacks. The limited proliferation capacity of primary chondrocytes poses a challenge for their use in extensive research applications. Their unnameability to genetic modification further restricts the scope of mechanistic studies. Moreover, primary chondrocytes are susceptible to dedifferentiation when cultured in monolayer, potentially leading to an erosion of their native chondrocyte phenotype¹⁰.

The differentiation of MSCs into chondrocytes presents a developmental model for cartilage research, along with the ability to engineer *de novo* cartilage in the laboratory. Immortalised stem cells therefore offer a bridge between the use of primary chondrocytes and cell lines like TC28a2¹¹. A primary advantage of MSCs is their pluripotency. Unlike primary chondrocytes and cell lines, MSCs can differentiate into a variety of cell types, including chondrocytes, under specific culture conditions. This flexibility allows the interrogation of the complete process of chondrogenesis, from stem cell to mature chondrocyte, providing insights into the developmental pathways and factors influencing cartilage formation. Moreover, MSCs can be sourced from various tissues with a different degree of accessibility, such as bone marrow, adipose tissue, and umbilical cord blood, offering a diverse range of research opportunities. Their ability to proliferate extensively *in vitro* without significant loss of differentiation potential is another key advantage, addressing the limitation of limited proliferation capacity seen in primary chondrocytes.

The differentiation of MSCs into chondrocytes is not without challenges¹¹. The process can be complex and requires a well-defined and controlled environment, including specific growth factors and mechanical stimuli, to ensure successful and consistent differentiation. In the context of cartilage research, the use of MSCs as an intermediate model offers a complementary approach. While primary chondrocytes and cell lines like TC28a2 provide insights into specific aspects of chondrocyte biology and pathology, MSC-derived chondrocytes offer a dynamic system to study the complete spectrum of chondrogenesis. This model can significantly contribute to the understanding of cartilage development and regeneration, as well as to the development of new therapeutic strategies for cartilage repair and regeneration in degenerative diseases.

In this Chapter I describe a study into the role of the COLGALT2 protein in the biology of chondrocytes, using an immortalised adipose derived MSC line, which constitutively overexpress SOX9 to enhance their *in vitro* chondrogenic potential (ASC52teloSox9)¹².

7.2. Aims and objectives

7.2.1. Aims

- Identify the necessity of collagen glycosylation in chondrogenesis.
- Determine which collagen(s) are the target(s) of COLGALT2 in chondrocytes.
- Understand the changes in the proteome of cells during chondrogenic differentiation of MSCs.

7.2.2. Objectives

- Perform CRISPR knockout (KO) of *COLGALT1* and *COLGALT2*, along with a double KO in ASC52teloSox9 cells.
- Confirm successful KO of the proteins in the cell line.

- Perform a 21-day differentiation of the WT ASC52telSox9 lines, collecting cell pellets at D0, 7, 14 and 21 to investigate changes in the proteome through chondrogenesis.
- Perform a 21-day differentiation of KO ASC52telSox9 lines, collecting cell pellets at D0 and Day 21, as well as culture medium from Day 21 for proteomic analysis.

7.3. Results

7.3.1. Proteomic analysis of ASC52telSox9 throughout chondrogenesis

Prior to the genetic manipulation of the *COLGALT1/2* genes in the ASC52telSox9 cells line, I investigated how the proteome changes within this model throughout our protocol for chondrogenic differentiation.

The cells were cultured in differentiation medium in a 3D pellet for 21 days. Cell pellets (n=3) were collected at days 3, 7, 14 and 21 with cell pellets (n=3) collected at D0 as undifferentiated control. Throughout the differentiation the cartilage pellets (containing 1000000 cells/pellet) visibly increased in size, demonstrating an ability to secrete ECM (Figure 7.1A). The expression of four chondrocyte marker genes was measured at each time point (*COL2A1*, *COL10A1*, *SOX9* and *ACAN*; Figure 7.1B) as well as the expression of *COLGALT2* (Figure 7.1B). Throughout differentiation, a significant increase in the expression of *SOX9* (*D0 vs D21*, *FDR=0.026*), *COL2A1* (*D0 vs D21*, *FDR=0.0014*) and *ACAN* (*D0 vs D21*, *FDR=0.018*) was observed. *COL10A1*, encoding the alpha chain of collagen type 10, is a marker of chondrocyte hypertrophy. Across the 21 days of differentiation, the levels of this gene showed no significant change (*D0 vs D21*, *FDR=0.697*) (Figure 7.1B). These results confirmed successful chondrogenesis and were indicative of a healthy chondrocyte phenotype. The levels of *COLGALT2* throughout the differentiation were also measured. The expression of the gene was increased by Day 21 (*D0 vs D21*, *FDR=0.016*) (Figure 7.1B).

Figure 7.1. Chondrogenic potential of the ASC52teloSox9 cells. **A.** Neo cartilage pellets at Day 7, Day 14 and Day 21 of the differentiation protocol. **B.** Normalised relative gene expression of SOX9 (green), ACAN (purple), COL2A1 (orange), COL10A1 (red) and COLGALT2 (purple) at different time points during the differentiation protocol (x-axis). Each symbol represents the mean of three biological replicates. Bars show mean +SEM. GraphPad Prism 9.0 was used for statistical analyses and data visualisation.

7.3.2. CRISPR-Cas9 COLGALT1 and COLGALT2 knockout

In ASC52teloSox9 cells, CRISPR-Cas9 was used to introduce frame-shift mutations in *COLGALT1* and *COLGALT2* in ASC52teloSox9 cells, disrupting the encoded proteins both individually and in combination (to create a double KO). The successful targeting of the proteins was determined at the transcription level by qPCR and a conferred change in protein level was measured by Western blot.

COLGALT1 editing, resulted in a 5.34-fold decrease in mRNA levels ($P=0.0095$) confirming successful, yet incomplete KO of the protein. In the double (*COLGALT1* and *COLGALT2*) KO cells (DKO) a 9.93-fold decrease ($P=0.009$) was measured (Figure 7.2A). Similarly, *COLGALT2* editing resulted in a 12.13-fold ($P=0.04$) decrease in mRNA expression. However, in the DKO cells, only a 2.98-fold reduction was observed ($P=0.06$) (Figure 7.2B). Western blot analysis against the *COLGALT1* (Figure 7.2A, bottom) and *COLGALT2* (Figure 7.2B, bottom) revealed a decrease in the protein levels in the cells with the individual knockouts, as well as in the cells subjected to a double knockout.

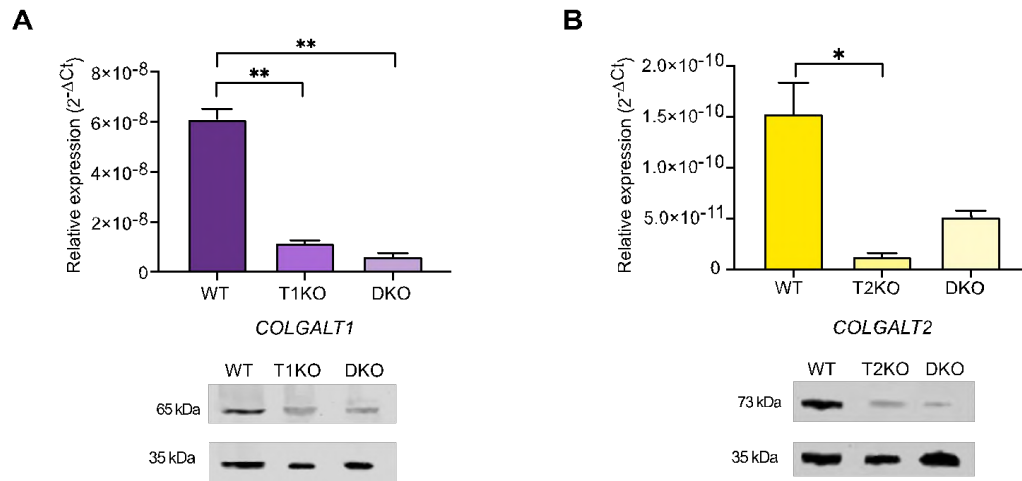


Figure 7.2. C Confirmation of COLGALT protein depletion prior to differentiation. **A.** COLGALT1 relative gene expression levels (top plot) were measured in control (WT) cells, COLGALT1 KO (T1KO) cells and in COLGALT1+COLGALT2 KO (DKO) cells. Bars show mean of three biologic replicates +SEM. P-values were calculated using non-paired t-tests. Western Blot against COLGALT1 (bottom) in control (WT) cells (n=1), COLGALT1 KO (T1KO) cells (n=1) and in COLGALT1+COLGALT2 KO (DKO) cells (n=1). COLGALT1 expected size band ~ 65KDa (top row). Loading control, GAPDH, expected size band is at 35KDa (bottom row). **B.** Same as A. for COLGALT2 (expected protein size ~ 73KDa). GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A. and B.).

7.3.3. Proteomic analysis of the differentiated chondrocytes

The proteomic analysis of the ASC52teloSOX9 differentiated cell pellets across the course of chondrogenesis and the analysis of the ASC52teloSOX9-COLGALT1 and -COLGALT2 KO cell pellets detected a total of 6624 proteins with >2 peptides (see Appendix D.).

Following chondrogenesis of WT and KO cell lines, proteomic analysis revealed no significant change in the expression of COLGALT1 compared to WT in the single (D0, FDR=0.56; D21, FDR=0.38) or double (D0, FDR=0.31; D21, FDR=0.49) KO cells. However, the COLGALT2 KO was significant at D0 (COLGALT2 KO, FDR<0.0001; double KO, FDR=0.0007) and D21 (COLGALT2 KO, FDR<0.0001; double KO, FDR<0.0001) (Figure 7.3). Based on this data, further investigations focused only on the COLGALT2 KO cells.

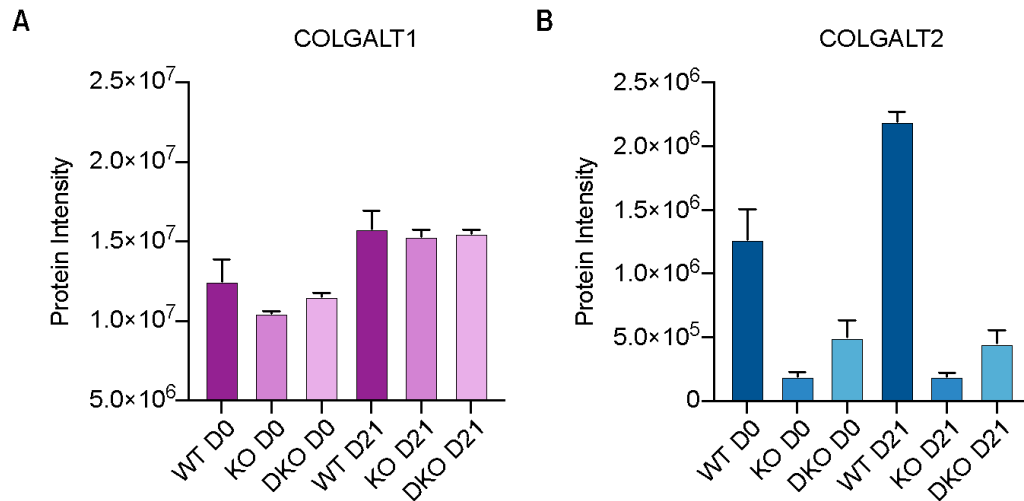


Figure 7.3. Knockout efficiency in ASC52teloSox9 cells. **A.** The measured protein abundance (protein intensity) (y-axis) of the COLGALT1 enzyme following differentiation into chondrocytes of WT (control, n=5), COLGALT1 KO (KO, n=5) and double COLGALT1+COLGALT2 KO (DKO, n=5) cells between D0 (D0) and Day 21 (D21). Bars show mean +SEM. **B.** Same as A. for COLGALT2. GraphPad Prism 9.0 was used for statistical analyses and data visualisation (A. and B.).

7.3.4. Phenotype of the differentiated chondrocytes

The expression of proteins encoded by the chondrocyte marker genes: SOX9, aggrecan, collagen type II, and collagen type X was assessed throughout chondrogenesis (Figure 7.4). The findings were in agreement with transcript level data (Figure 7.1B), revealing that during the 21-day differentiation period, there was a significant and sustained increase in the protein levels of aggrecan (D0 vs D21, FDR=0.018) and collagen type II (D0 vs D21, FDR=0.001), indicating the development of a mature chondrocyte phenotype. The protein expression of SOX9 also increased during the differentiation process (D0 vs D21, FDR=0.026). In contrast, collagen type X showed an initial increase in the early phase of differentiation (D0 vs D7, FDR=0.03), but its levels decreased towards the end, aligning with the levels observed at D0 (D0 vs D21, FDR=0.69).

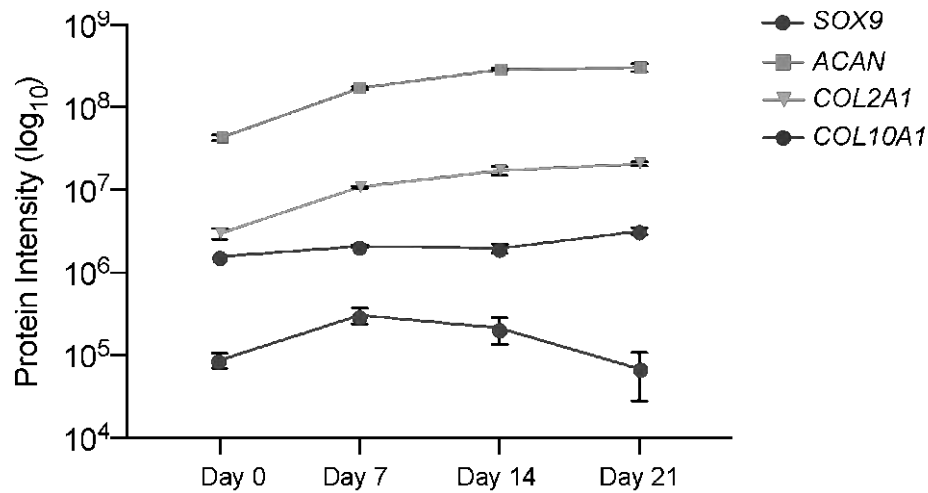


Figure 7.4. Phenotype of WT ASC52teloSox9 cells following 21-day differentiation into chondrocytes. The measured protein abundance (protein intensity) of SOX9 (green), ACAN (purple), COL2A1 (orange) and COL10A1 (red) in WT ASC52teloSox9 cells at different time points during the differentiation protocol (x-axis). Each symbol represents the mean of five biologic replicates. Bars show mean +SEM. GraphPad Prism 9.0 was used for the data visualisation.

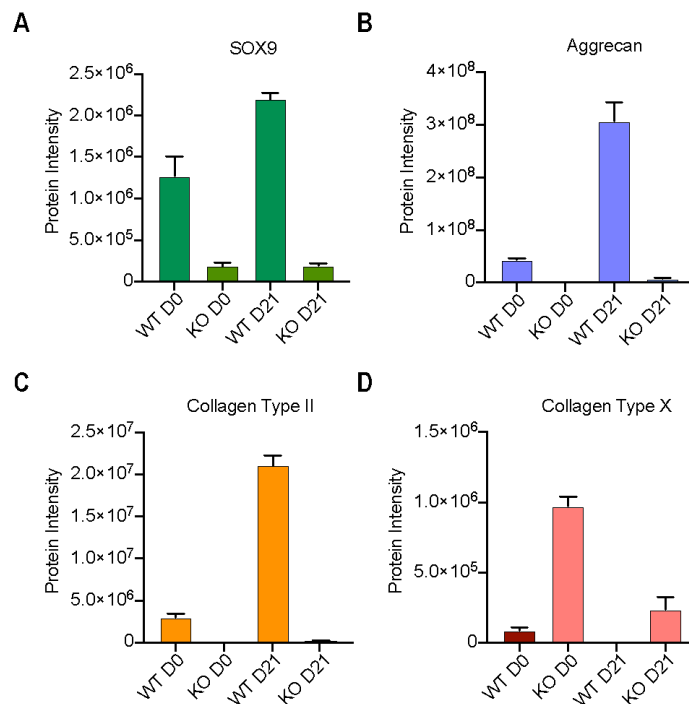


Figure 7.5. The measured protein abundance (protein intensity) of chondrocyte markers following differentiation into chondrocytes of WT (control, n=5), COLGALT2 KO (KO, n=5) cells at D0 (D0, n=5) and Day 21 (D21, n=5) for SOX9 (A), Aggrecan (B), Collagen Type II (C) and Collagen Type X (D). Bars show mean +/-SEM. GraphPad Prism 9.0 was used for data visualisation.

I next considered this panel of markers in the KO cells. Protein expression level of SOX9 was significantly lower compared to the control WT cells at both the start (D0, FDR=0.0002) and end (D21, FDR<0.0001) of the differentiation process (Figure 7.5A). This alteration in the expression pattern highlights the influence of COLGALT2 on the differentiation of chondrocytes and the expression of critical chondrogenesis genes.

Aggrecan levels in the COLGALT2 KO cells were consistently lower than those in the WT cells (D0, FDR=0.0178; D21, FDR<0.0001) (Figure 7.5B). Type-II collagen was undetectable in COLGALT2 KO cells at the outset (D0) and only present at minimal levels by D21, contrasting sharply with its robust presence in WT cells (D21, FDR<0.0001) (Figure 7.5C). The levels of collagen type X, associated with chondrocyte hypertrophy, were significantly higher compared to the control group at D21 (FDR=0.03) but comparable at D0 (FDR=0.31) since there was a significant reduction in the expression in the COLGALT2 KO cells between D0 and D21 (P=0.043) (Figure 7.5D).

These results collectively point to a distinct phenotype in the COLGALT2 KO ASC52teloSox9 cells compared to the WT controls. Notably, the COLGALT2 KO cells failed to develop characteristic features of mature chondrocytes following the differentiation protocol as demonstrated by the proteomics data.

7.3.4. Differential protein expression

Hierarchical clustering of the biological replicate samples from distinct time points in WT cell cohort indicated the presence of two anomalous data points collected at D21 (Figure 7.6A). The COLGALT2 KO samples collected at D0 and D21 segregated into two distinct clusters corresponding to the respective time points. Notably, the Day 21 samples displayed a higher degree of variability yet still formed a separate cluster distinguishable from the D0 samples (Figure 7.6B). For the secretome analysis involving WT and COLGALT2 KO cells at Day 21, the clustering result delineated three distinct groups. Two of these groups were primarily characterized by the cell type, while one sample from the COLGALT2 KO Day 21 exhibited marked dissimilarity compared to the other samples (Figure 7.6C).

Consequently, following these clustering outcomes, two replicates from the WT Day 21 cell pellets and one sample from the COLGALT2 KO Day 21 secretome were omitted from the differential analysis.

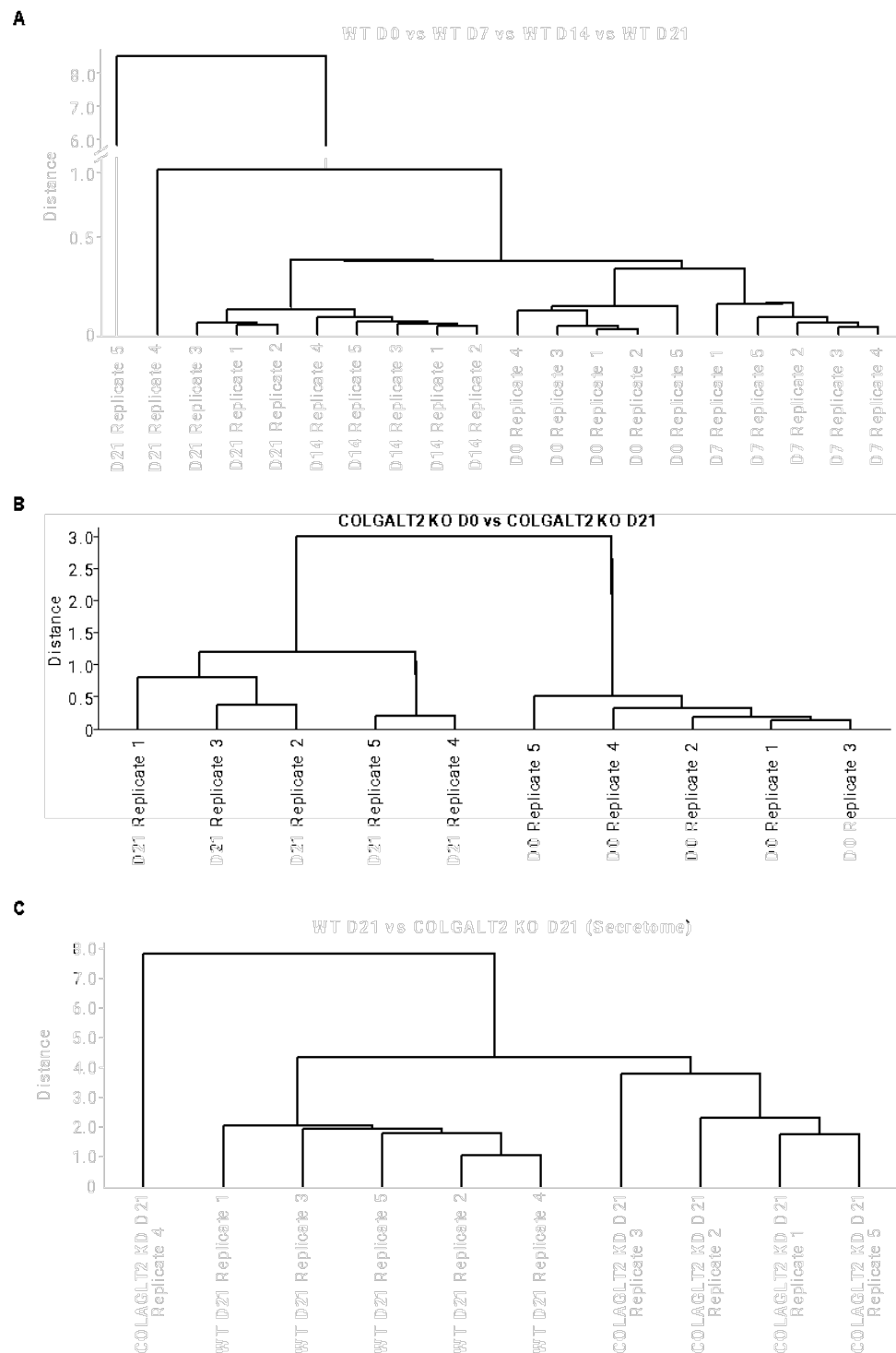


Figure 7.6. Hierarchical clustering dendrogram of protein profiles. The dendrogram illustrates the hierarchical clustering of experimental conditions (A. WT D0 vs D7, WT D0 vs D14, WT D0

vs D21, WT D7 vs D14, WT D14 vs D21; **B.** COLGALT2 KO D0 vs COLGALT2 KO D21; **C.** WT D21 Secretome vs COLGALT2 KO D21 Secretome) based on single linkage analysis. The vertical lines represent the individual replicates for each condition (x-axis), and the Manhattan distance (y-axis) indicates the similarity in protein profiles. Shorter distances between branches denote more similar profiles. The analysis and the data visualisation were performed in R-Studio.

First, the changes to the proteome between the stem cells and the differentiated chondrocytes were investigated prior to, and following, chondrogenesis (D0 vs D21). Of the 7929 measured proteins, 2918 showed differential expression ($FDR < 0.05$, $FC > [1.5]$) with 1322 (45%) increasing and 1596 decreasing (55%) (Figure 7.7A) (see Appendix E.). Gene ontology (GO) enrichment showed that the 20 most significantly overrepresented biological processes ($FDR < 0.05$) include metabolic processes, cellular structure organisation, as well as extracellular organisation and positive regulation of mesenchymal stem cell differentiation (Figure 7.7B). Table 7.1 lists the significantly differentially expressed proteins that play a known role in cartilage biology or are implicated indirectly in it. Among the proteins that increase in expression are collagens (type I-VI, X-XII, XIV, XVI and XVII), as well as proteins associated with collagen fibril formation (decorin, lumican, biglycan, matrilins), ECM components and (aggrecan, cartilage oligomeric matrix protein) and factors involved in chondrocyte homeostasis and development (growth differentiation factor 5, SRY-Box Transcription Factor 9, filamin B). The proteins downregulated during chondrogenesis include matrix metalloproteases (MMP1, MMP3), metalloprotease inhibitors (TIMP1, TIMP3) as well as ECM components (HAS1, PRG4) and components of the TGF β pathway (SMAD3, TGFBR1).

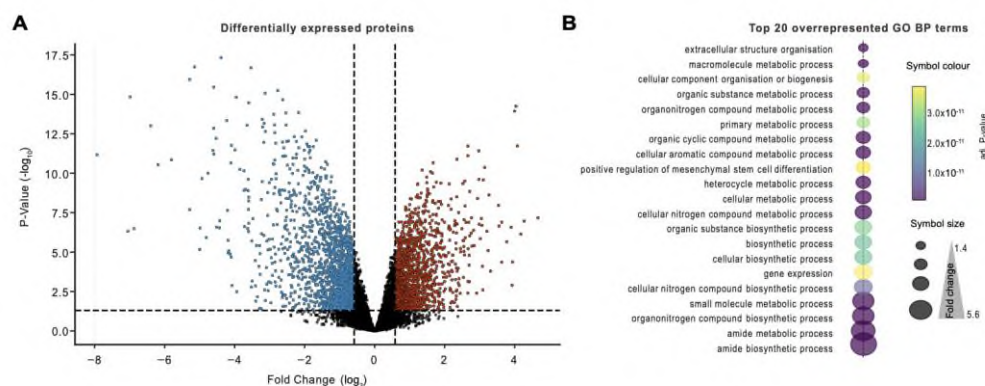


Figure 7.7. Differential protein expression analysis of WT D0 and WT D21 cells. **A.** Volcano plot of protein expression differences between WT D0 and WT D21 cells. The plot displays the

results of differential gene expression analysis. Log2 fold change (x-axis), indicates the magnitude of gene expression difference between the experimental conditions. Proteins to the right of the plot show upregulation in D21 cells, while those to the left show downregulation. The y-axis represents $-\log_{10}$ P-value. Higher values indicate greater significance. Each dot represents an individual protein. The blue dots are proteins showing a significant ($P < 0.05$) downregulation in D21 cells compared to D0 cells of more than 1.5FC. The red dots are proteins showing a significant ($P < 0.05$) upregulation in D21 cells compared to D0 cells of more than 1.5 FC. The threshold lines indicate cut-offs for significance levels ($P < 0.05$ (y-axis)), and fold change $FC > [1.5]$ (x-axis)). The P-values were calculated using t-tests with Benjamin-Hochberg correction for multiple testing. The volcano plots were produced in GraphPad Prism 9.0. **B.** A dot plot visualising the top 20 enriched GO biological processes for the differentially expressed proteins between WT D0 and WT D21 cells. Each dot represents a distinct GO term. The dot size is proportional to the fold change (from 1.4 to 5.6), indicating the magnitude of protein expression change. Larger dots represent higher fold changes. The colour represents FDR-adjusted P-value (Fischer's exact test with Benjamin-Hochberg correction), according to the indicated heatmap gradient where yellow-green colours represent less significant values and blue-purple colours show more significant values. R-Studio was used for the analysis and data visualisation.

Table 7.1. Proteins relevant to cartilage biology differentially expressed in WT D0 and WT D21 cells. The blue ones show increase between D0 and D21 and the orange ones – a decrease

Gene name	logFC	Levels at D0 compared to D21	Potential Role in Cartilage Biology
ACAN	-2.826235674	significantly lower	Aggrecan, critical for cartilage structure and resilience.
ASPN	-3.231288644	significantly lower	Asporin, involved in cartilage homeostasis, role in osteoarthritis.
BGN	-3.209521491	significantly lower	Biglycan, important in collagen fibrillogenesis.
COL10A1	-3.499038259	significantly lower	Type X collagen, involved in hypertrophic cartilage during bone formation.
COL11A2	-0.909081158	significantly lower	Part of collagen XI, role in cartilage biology not specified.
COL12A1	-1.989745196	significantly lower	Type XII collagen, supports structural integrity.
COL14A1	-3.209392513	significantly lower	Type XIV collagen, involved in assembly and strength.
COL16A1	-1.297704728	significantly lower	Type XVI collagen, associated with fibrillar collagens network.
COL18A1	-1.214200553	significantly lower	Type XVIII collagen, associated with vascular and connective tissues.
COL1A1	-3.605436269	significantly lower	Type I collagen, provides tensile strength.
COL1A2	-3.595788957	significantly lower	Forms type I collagen, important for structure.
COL2A1	-5.000241858	significantly lower	Essential for formation and maintenance, forms type II collagen.
COL3A1	-0.715126116	significantly lower	Type III collagen, maintains structure and strength.
COL4A1	-3.940804876	significantly lower	Forms type IV collagen, contributes to structure.
COL4A2	-0.892420262	significantly lower	Type IV collagen, part of basement membrane, influences development.
COL5A1	-0.750343328	significantly lower	Part of type V collagen, regulates fibrillar collagens assembly.
COL6A1	-2.008086569	significantly lower	Part of type VI collagen, contributes to structure.
COL6A2	-2.080100012	significantly lower	Part of type VI collagen, contributes to microfibrillar network.
COMP	-2.815590584	significantly lower	Maintains structure and involved in repair processes.
DCN	-3.168213365	significantly lower	Decorin, regulates collagen fibril formation, affects matrix organization.
FLNB	-0.979347476	significantly lower	Involved in skeletal development, influences cartilage structure.
GDF5	-2.063764156	significantly lower	Growth differentiation factor, important in development and repair.
LUM	-1.61055798	significantly lower	Lumican, regulates collagen fiber arrangement, affects matrix structure.
MATN3	-2.847848075	significantly lower	Matrilin-3, involved in matrix formation and stability.
SOX9	-1.03194325	significantly lower	Transcription factor critical for chondrogenesis and cartilage formation.
COL15A1	0.994229855	significantly higher	Component of collagen XV, contributes to structural integrity.
COL7A1	1.654508048	significantly higher	Encodes type VII collagen, important for anchoring fibrils in cartilage.
MMP1	1.838214447	significantly higher	Degrades matrix components, involved in cartilage repair.
MMP3	1.157339475	significantly higher	Degrades matrix components, involved in cartilage repair.
SMAD3	0.632994307	significantly higher	Part of the TGF-beta signaling pathway, affects cartilage formation.
TGFBR1	0.720904661	significantly higher	Part of TGF-beta receptor, signaling pathway important in cartilage development.
TIMP1	1.271583319	significantly higher	An MMP inhibitor, regulates matrix turnover and homeostasis.
TIMP3	1.1256641	significantly higher	An MMP inhibitor, regulates matrix turnover and homeostasis.

I next looked more closely at the different stages of chondrogenic differentiation, comparing protein expression at D7, and D14.

During the initial stages of differentiation (D0 vs D7) the largest number of protein changes were detected with 2339 differentially expressed proteins (see Appendix E.). The majority of these proteins showed increased expression at D7 compared to D0 (1596) (Figure 7.8A). The differentially expressed proteins at this early stage of chondrogenesis were primarily involved in biological processes associated with metabolism and cellular component organisation (Figure 7.9B). Among the differentially expressed proteins, many were related to cartilage biology, including collagens, ECM components and regulators of cartilage homeostasis and development (Table 7.2).

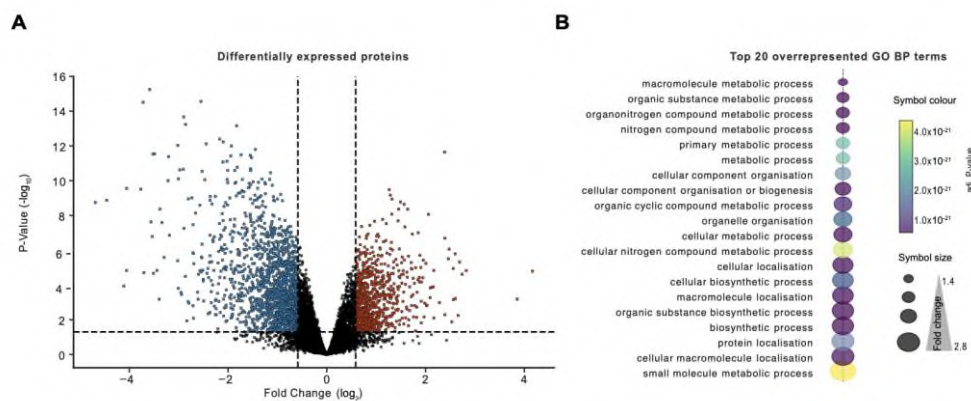


Figure 7.8. Differential protein expression analysis of WT D0 and WT D7 cells. **A.** Volcano plot of protein expression differences between WT D0 and WT D7 cells. The plot displays the results of differential gene expression analysis. Log₂ fold change (x-axis), indicates the magnitude of gene expression difference between the experimental conditions. Proteins to the right of the plot show upregulation in D7 cells, while those to the left show downregulation. The y-axis represents -Log₁₀ P-value. Higher values indicate greater significance. Each dot represents an individual protein. The blue dots are proteins showing a significant ($P < 0.05$) downregulation in D7 cells compared to D0 cells of more than 1.5FC. The red dots are proteins showing a significant ($P < 0.05$) upregulation in D7 cells compared to D0 cells of more than 1.5 FC. The threshold lines indicate cut-offs for significance levels ($P < 0.05$ (y-axis)), and fold change $FC > [1.5]$ (x-axis)). The P-values were calculated using t-tests with Benjamin-Hochberg correction for multiple testing. The volcano plots were produced in GraphPad Prism 9.0. **B.** A dot plot visualising the top 20 enriched GO biological processes for the differentially expressed proteins between WT D0 and WT D7 cells. Each dot represents a distinct GO term. The dot size is proportional to the fold change (from 1.4 to 2.8), indicating the magnitude of protein expression change. Larger dots represent higher fold changes. The colour represents FDR-adjusted P-value (Fischer's exact test with Benjamin-Hochberg correction), according to the indicated heatmap gradient where yellow-green colours represent less significant values and

blue-purple colures show more significant values. R-Studio was used for the analysis and data visualisation.

Table 7.2. Proteins relevant to cartilage biology differentially expressed in WT D0 and WT D7 cells. The blue ones show increase between D0 and D7 and the orange ones – a decrease.

Gene name	logFC	Levels at D0 compared to D7	Potential Role in Cartilage Biology
ACAN	-2.010208211	significantly lower	Aggrecan, critical for cartilage structure and resilience.
ASPN	-2.533929296	significantly lower	Asporin, involved in cartilage homeostasis, role in osteoarthritis.
BGN	-1.880210518	significantly lower	Biglycan, important in collagen fibrillogenesis.
CILP	-0.912683275	significantly lower	Involved in cartilage integrity, role in osteoarthritis.
COL10A1	-1.583375137	significantly lower	Type X collagen, involved in hypertrophic cartilage during bone formation.
COL11A2	-0.989128084	significantly lower	Part of collagen XI, role in cartilage biology not specified.
COL12A1	-1.059325936	significantly lower	Type XII collagen, supports structural integrity.
COL14A1	-1.571654479	significantly lower	Type XIV collagen, involved in assembly and strength.
COL16A1	-1.507302309	significantly lower	Type XVI collagen, associated with fibrillar collagens network.
COL1A1	-2.65760212	significantly lower	Type I collagen, provides tensile strength.
COL1A2	-3.561169057	significantly lower	Forms type I collagen, important for structure.
COL2A1	-1.918679776	significantly lower	Essential for formation and maintenance, forms type II collagen.
COL3A1	-1.042996589	significantly lower	Type III collagen, maintains structure and strength.
COL4A1	-1.689206078	significantly lower	Forms type IV collagen, contributes to structure.
COL4A2	-1.013166623	significantly lower	Type IV collagen, part of basement membrane, influences development.
COL5A1	-1.196210135	significantly lower	Part of type V collagen, regulates fibrillar collagens assembly.
COL6A1	-1.39540914	significantly lower	Part of type VI collagen, contributes to structure.
COL6A2	-1.677502021	significantly lower	Part of type VI collagen, contributes to microfibrillar network.
COMP	-2.470825265	significantly lower	Maintains structure and involved in repair processes.
DCN	-2.006644953	significantly lower	Decorin, regulates collagen fibril formation, affects matrix organization.
FLNB	-0.798033579	significantly lower	Involved in skeletal development, influences cartilage structure.
GDF5	-1.005969555	significantly lower	Growth differentiation factor, important in development and repair.
LUM	-1.407880554	significantly lower	Lumican, regulates collagen fiber arrangement, affects matrix structure.
MATN3	-1.279077752	significantly lower	Matrilin-3, involved in matrix formation and stability.
SLC26A2	-0.605826226	significantly lower	Involved in sulfate transport, important for proteoglycan synthesis.
COL15A1	0.851841657	significantly higher	Component of collagen XV, contributes to structural integrity.
HAS1	0.786521759	significantly higher	Synthesizes hyaluronan, important for cartilage structure and function.
MMP1	0.710051768	significantly higher	Degrades matrix components, involved in cartilage repair.
MMP3	1.189689913	significantly higher	Degrades matrix components, involved in cartilage repair.
PRG4	1.013060637	significantly higher	Lubricates cartilage surfaces, protects against wear.
TIMP3	1.195049584	significantly higher	Inhibits metalloproteinases, involved in extracellular matrix remodeling.

In later stages of differentiation, less protein expression changes occurred between the time points (825 differentially expressed proteins between D7 vs D14 and 795 D14 vs D21) (Figure 7.9A and 7.10A) (see Appendix E.). The topmost significant biological processes for which these proteins were overrepresented, however, included extracellular matrix assembly and organisation and collagen fibril organisation (Figure 7.9B and 7.10B), which are crucial for the proper function of healthy chondrocytes. Additionally, the overrepresented GO biological processes for later differentiation stage (D14 vs D21) also included regulation of apoptotic processes and programmed cell death (Figure 7.10B), suggesting that at those later stages in the differentiation, the chondrocytes have exhibited a phenotype somewhat consistent with that of aged mature chondrocytes.

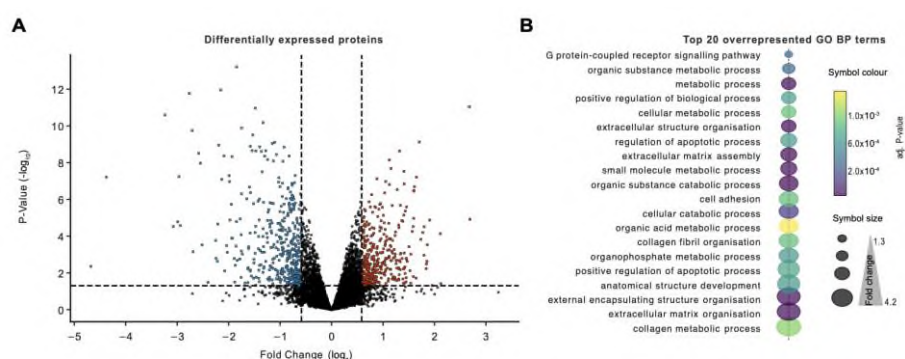


Figure 7.9. Differential protein expression analysis of WT D7 and WT D14 cells. **A.** Volcano plot of protein expression differences between WT D7 and WT D14 cells. The plot displays the results of differential gene expression analysis. Log2 fold change (x-axis), indicates the magnitude of gene expression difference between the experimental conditions. Proteins to the right of the plot show upregulation in D14 cells, while those to the left show downregulation. The y-axis represents $-\log_{10}$ P-value. Higher values indicate greater significance. Each dot represents an individual protein. The blue dots are proteins showing a significant ($P < 0.05$) downregulation in D14 cells compared to D7 cells of more than 1.5FC. The red dots are proteins showing a significant ($P < 0.05$) upregulation in D21 cells compared to D7 cells of more than 1.5 FC. The threshold lines indicate cut-offs for significance levels ($P < 0.05$ (y-axis)), and fold change $FC > [1.5]$ (x-axis)). The P-values were calculated using t-tests with Benjamin-Hochberg correction for multiple testing. The volcano plots were produced in GraphPad Prism 9.0. **B.** A dot plot visualising the top 20 enriched GO biological processes for the differentially expressed proteins between WT D7 and WT D14 cells. Each dot represents a distinct GO term. The dot size is proportional to the fold change (from 1.3 to 4.2), indicating the magnitude of protein expression change. Larger dots represent higher fold changes. The colour represents the FDR-adjusted P-value (Fischer's exact test with Benjamin-Hochberg correction), according to the indicated heatmap gradient where yellow-green colours represent less significant values and blue-purple colours show more significant values. R-Studio was used for the analysis and data visualisation.

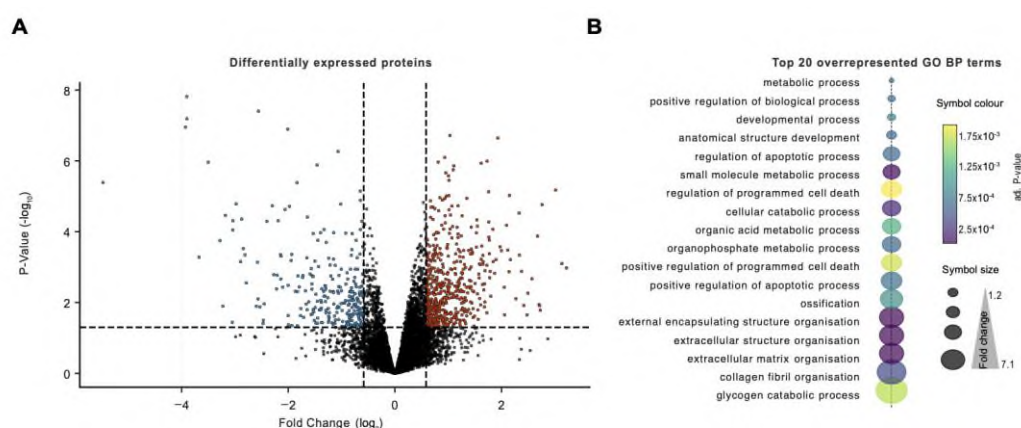


Figure 7.10. Differential protein expression analysis of WT D14 and WT D21 cells. **A.** Volcano plot of protein expression differences between WT D14 and WT D21 cells. The plot displays the

results of differential gene expression analysis. Log2 fold change (x-axis), indicates the magnitude of gene expression difference between the experimental conditions. Proteins to the right of the plot show upregulation in D21 cells, while those to the left show downregulation. The y-axis represents $-\log_{10}$ P-value. Higher values indicate greater significance. Each dot represents an individual protein. The blue dots are proteins showing a significant ($P < 0.05$) downregulation in D21 cells compared to D14 cells of more than 1.5FC. The red dots are proteins showing a significant ($P < 0.05$) upregulation in D21 cells compared to D14 cells of more than 1.5 FC. The threshold lines indicate cut-offs for significance levels ($P < 0.05$ (y-axis)), and fold change $FC > [1.5]$ (x-axis). The P-values were calculated using t-tests with Benjamin-Hochberg correction for multiple testing. The volcano plots were produced in GraphPad Prism 9.0. **B.** A dot plot visualising the top 20 enriched GO biological processes for the differentially expressed proteins between WT D14 and WT D21 cells. Each dot represents a distinct GO term. The dot size is proportional to the fold change (from 1.2 to 7.1), indicating the magnitude of protein expression change. Larger dots represent higher fold changes. The colour represents FDR-adjusted P-value (Fischer's exact test with Benjamin-Hochberg correction), according to the indicated heatmap gradient where yellow-green colours represent less significant values and blue-purple colours show more significant values. R-Studio was used for the analysis and data visualisation.

Among the cartilage biology-relevant proteins that increase in expression between D7 and D14 are still many collagens and ECM components (Table 7.3) but only FLNB and SOX show a significant increase between D14 and D21, with some collagens (type III, V and VII) showing a decrease alongside lumican (Table 7.4).

Table 7.3. Proteins relevant to cartilage biology differentially expressed in WT D7 and WT D14 cells. The blue ones show increase between D7 and D14 and the orange ones – a decrease.

Gene name	logFC	Levels at D7 compared to D14	Potential Role in Cartilage Biology
ACAN	-0.749480765	significantly lower	Aggrecan, critical for cartilage structure and resilience.
ASPN	-1.139036851	significantly lower	Asporin, implicated in cartilage homeostasis and osteoarthritis.
BGN	-1.072234323	significantly lower	Biglycan, involved in collagen fibrillogenesis and matrix assembly.
COL10A1	-1.628515937	significantly lower	Type X collagen, involved in hypertrophic cartilage during bone formation.
COL12A1	-0.811609531	significantly lower	Type XII collagen, interacts with other collagens for structural support.
COL14A1	-1.397873355	significantly lower	Involved in collagen network assembly, supports cartilage integrity.
COL1A1	-1.595151029	significantly lower	Forms type I collagen, key for tensile strength in connective tissues.
COL1A2	-0.657485912	significantly lower	Forms type I collagen, important for collagen structure.
COL2A1	-0.643825813	significantly lower	Essential for type II collagen, primary component of cartilage.
COL6A1	-0.662800093	significantly lower	Part of type VI collagen, contributes to cartilage microfibrillar network.
COL6A2	-0.770884082	significantly lower	Contributes to type VI collagen, important for cartilage matrix.
COL9A2	-0.846620725	significantly lower	Type IX collagen, involved in cartilage and intervertebral disc matrix.
DCN	-0.932712573	significantly lower	Decorin, regulates collagen fibril formation, affects matrix organization.
GDF5	-1.259423872	significantly lower	Growth differentiation factor 5, important in joint and cartilage development.
LUM	-0.995537012	significantly lower	Lumican, regulates collagen fiber arrangement, affects matrix structure.
MATN3	-1.575356893	significantly lower	Matrilin-3, contributes to cartilage matrix stability.
PRG4	-1.351630086	significantly lower	Lubricates cartilage surfaces, protects against wear.
TIMP2	-0.702319148	significantly lower	Inhibits MMPs, plays a role in extracellular matrix maintenance.
COL7A1	0.953208034	significantly higher	Encodes type VII collagen, important for anchoring fibrils in cartilage.
MMP1	1.859805553	significantly higher	Degrades matrix components, involved in cartilage repair.
TGFBR1	0.821277844	significantly higher	Part of TGF-beta receptor, signaling pathway important in cartilage development.
TIMP1	0.941446755	significantly higher	An MMP inhibitor, regulates matrix turnover and homeostasis.

Table 7.4. Proteins relevant to cartilage biology differentially expressed in WT D14 and WT D21 cells. The blue ones show increase between D14 and D21 and the orange ones – a decrease

Gene name	logFC	Levels at D14 compared to D21	Potential Role in Cartilage Biology
FLNB	-1.978605522	significantly lower	Involved in skeletal development, influences cartilage structure.
SOX9	-0.744181606	significantly lower	Transcription factor critical for chondrogenesis and cartilage formation.
CILP	0.653496807	significantly higher	Cartilage intermediate layer protein, involved in cartilage matrix regulation.
COL3A1	0.749189158	significantly higher	Type III collagen, supports connective tissue integrity.
COL5A2	0.726652441	significantly higher	Part of type V collagen, regulates fibrillar collagen assembly.
COL7A1	1.095615784	significantly higher	Encodes type VII collagen, important for anchoring fibrils in cartilage.
LUM	0.792859586	significantly higher	Lumican, regulates collagen fiber arrangement, affects matrix structure.

The proteome of the COLGALT2 KO cells was compared to that of the WT cells before the differentiation (at D0) and after the 21-day protocol (at D21) (Figure 7.11).

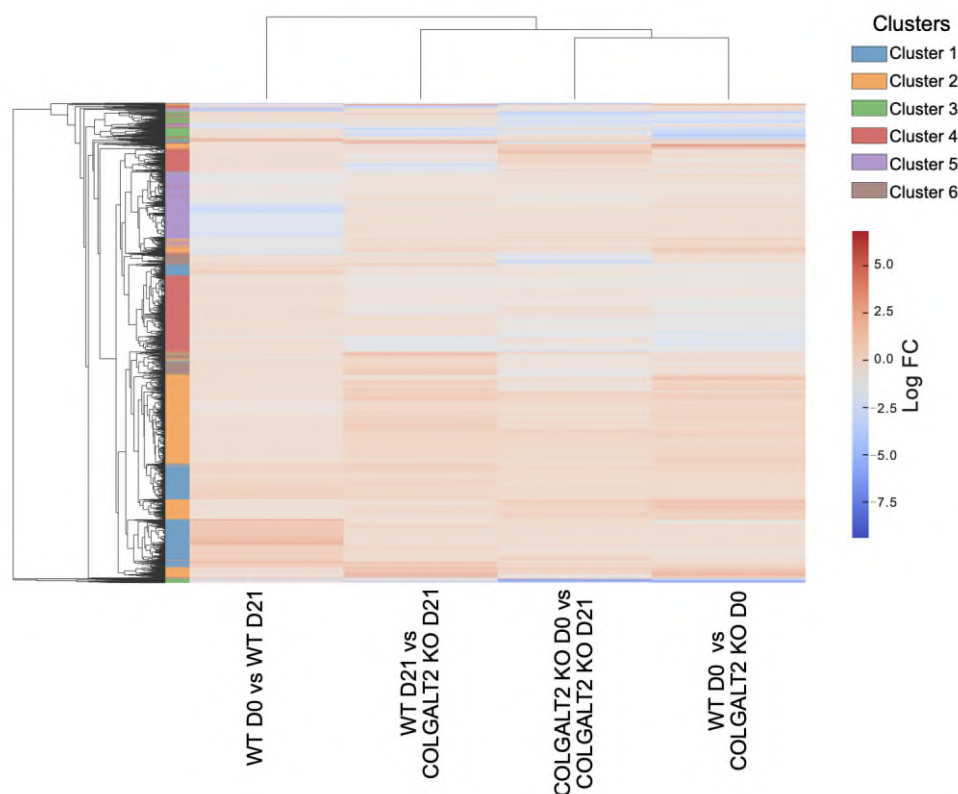


Figure 7.11. Heatmap of Differentially Expressed Genes with Hierarchical Clustering. The heatmap displays the log₂ fold-change (Log FC) values of the top 3000 most differentially expressed genes across four conditions: COLGALT2 KO D0 vs D21, WT D0 vs WT D21, WT D0 vs COLGALT2 KO D0, and WT D21 vs COLGALT2 KO D21. Hierarchical clustering was applied to group genes (rows) and conditions (columns) based on similarity in expression profiles. The rows are colour-coded by cluster membership, with six clusters identified using Ward's linkage method and a Manhattan distance metric. The dendrogram to the left of the heatmap represents the hierarchical clustering, reflecting the dissimilarity (linkage distance) between

merged clusters, providing insight into the relationships among gene expression patterns. Colour intensity indicates the magnitude of differential expression, with red representing upregulated genes (LogFC>1.5) and blue representing downregulated genes (LogFC<-1.5). Missing values are represented in white. Clustering and heatmap generation were performed in R-Studio, with the dendrogram height calculated directly from the linkage matrix.

The heatmap displays protein expression profiles for the top 3000 differentially expressed proteins across four experimental conditions, with hierarchical clustering identifying six distinct clusters (see Appendix F.). This analysis highlights regulatory trends and biological pathways affected by the COLGALT2 knockout. Proteins differentially expressed in the COLGALT2 knockout at D0 vs D21 show a closer association with other knockout comparisons (WT D0 vs COLGALT2 KO D0, WT D21 vs COLGALT2 KO D21) than with the WT D0 vs WT D21 comparison, suggesting that the largest fold changes are linked to pathways disrupted by COLGALT2 deficiency, rather than normal chondrogenic differentiation.

Cluster 1 includes proteins significantly upregulated in the knockout conditions, potentially regulated by COLGALT2, such as ECM components (COL1A1, COL3A1, COL5A1, MMP2, LOX) and stress response proteins (CCL20, CCL5, IL1RAP) (see Appendix F.). Cluster 2 proteins exhibit moderate expression differences, especially between WT D0 and WT D21, and are associated with normal chondrogenesis and cartilage formation, including ECM synthesis and regulation (collagens, aggrecan, BMP1, POSTN) (see Appendix F.). Cluster 3 proteins are strongly downregulated in the knockout, indicating COLGALT2's critical role in collagen biosynthesis and ECM stability, with key pathways involving TGF- β , Wnt/ β -catenin, MAPK, NF- κ B, and Notch (see Appendix F.). Cluster 4 proteins show mixed expression patterns, reflecting the involvement of dynamic signalling pathways influenced by both developmental stage and COLGALT2 activity (see Appendix F.). Cluster 5 proteins are downregulated in knockout conditions, suggesting impaired function, and participate in TGF- β , MAPK, and PI3K/Akt pathways critical for cartilage development (see Appendix F.). Cluster 6 proteins have context-dependent expression profiles, indicating complex regulatory mechanisms involving COLGALT2 and other factors (see Appendix F.).

Changes in collagen-related proteins, such as COL2A1 and COL3A1, in the COLGALT2 KO D0 vs D21 condition highlight disruptions in early ECM assembly and maintenance. Proteins with distinct patterns in the WT D0 vs WT D21 condition reflect normal

developmental progression. In contrast, the WT D0 vs COLGALT2 KO D0 condition shows clusters enriched with stress-response proteins, indicating an acute cellular response to COLGALT2 deficiency. The WT D21 vs COLGALT2 KO D21 condition reveals downregulation of structural and adhesion proteins, suggesting compromised cellular integrity and ECM remodelling during chondrogenesis due to COLGALT2 loss.

Delving deeper into the outcomes of the COLGALT2 depletion I observed a marked difference in the proteome makeup of the ASC52teloSox9 cells. At D0, 3286 proteins showed statistically significant differences in protein levels ($FDR < 0.05$) (1346 down- and 1940 upregulated in COLGALT2 KO compared to WT) (Figure 7.12) (see Appendix E.). The proteins significantly downregulated in the COLGALT2 KO D0 compared to WT D0 cells were overrepresented for GO terms including extracellular structure and matrix organisation, collagen fibril assembly and collagen biosynthesis, elastic fibre assembly, herpan sulphate and proteoglycan metabolism (Figure 7.13A). The proteins significantly upregulated in the edited cells at D0 were overrepresented for GO terms including cellular component and protein localisation, cellular metabolic processes, cellular components metabolism (Figure 7.13B).

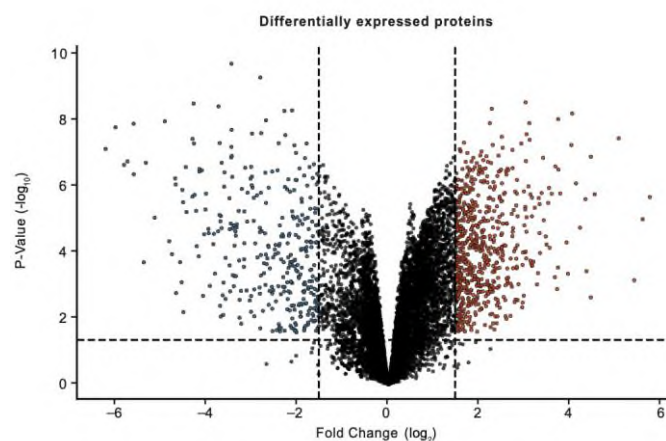


Figure 7.12. Volcano plot of protein expression differences between the secretome of WT D0 and COLGALT2 KO D0 cells. The plot displays the results of differential gene expression analysis. Log2 fold change (x-axis), indicates the magnitude of gene expression difference between the experimental conditions. Proteins to the right of the plot show upregulation in the proteome of COLGALT2 KO D0 cells compared to the WT D0 cells, while those to the left show downregulation. The y-axis represents $-\log_{10}$ P-value. Higher values indicate greater significance. Each dot represents an individual protein. The blue dots are proteins showing a significant ($P < 0.05$) downregulation in the proteome of COLGALT2 KO D0 cells compared to the proteome of WT D0 cells of more than 1.5FC. The red dots are proteins showing a

significant ($P < 0.05$) upregulation in the proteome of COLGALT2 KO D0 cells compared to the secretome of WT D0 cells of more than 1.5FC. The threshold lines indicate cut-offs for significance levels ($P < 0.05$ (y-axis)), and fold change $FC > [1.5]$ (x-axis)). The P-values were calculated using t-tests with Benjamin-Hochberg correction for multiple testing. The volcano plots were produced in GraphPad Prism 9.0.

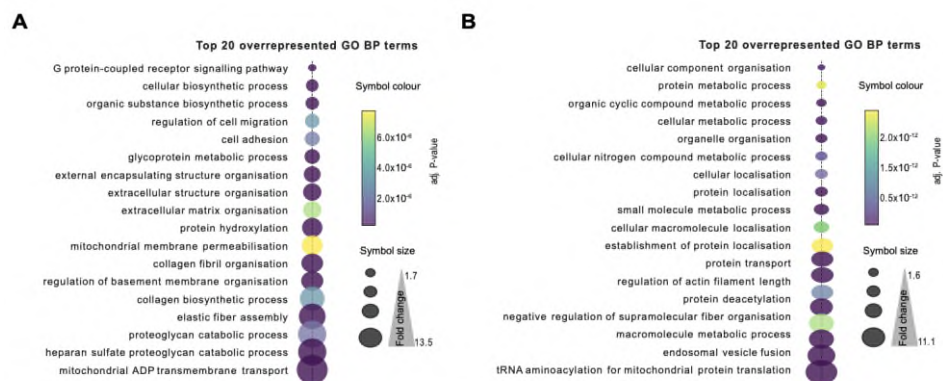


Figure 7.13. GO enrichment analysis of WT D0 vs COLGALT2 KO D0 cells. **A.** A dot plot visualising the top 20 enriched GO biological processes for the proteins showing a decrease in COLGALT2 KO D0 cells compared to WT D0 cells. Each dot represents a distinct GO Biological process (BP) term. The dot size is proportional to the fold change (from 1.7 to 13.5), indicating the magnitude of protein expression change. Larger dots represent higher fold changes. The colour represents the FDR-adjusted P-value (Fischer's exact test with Benjamin-Hochberg adjustment), according to the indicated heatmap gradient where yellow-green colours represent less significant values and blue-purple colours show more significant values. R-Studio was used for the analysis and data visualisation. **B.** A dot plot visualising the top 20 enriched GO biological processes for the proteins showing an increase in COLGALT2 KO D0 cells compared to WT D0 cells. Each dot represents a distinct GO Biological process (BP) term. The dot size is proportional to the fold change (from 1.6 to 11.1), indicating the magnitude of protein expression change. Larger dots represent higher fold changes. The colour represents the FDR-adjusted P-value (Fischer's exact test with Benjamin-Hochberg correction), according to the indicated heatmap gradient where yellow-green colours represent less significant values and blue-purple colours show more significant values. R-Studio was used for the analysis and data visualisation.

The cartilage biology-relevant proteins that show higher expression in the COLGALT2 KO compared to the WT cells before differentiation are COMP, SMAD3, and CD44 (Table 7.5). Among the proteins that are reduced in expression at D0 in the edited cells are many collagens (I, III, IV, V, VI, VII, XI, XII, XIV, XV, XVI, XVIII), aggrecan, SOX9 and many collagen PTM regulators (LOX, P3H2, P4HA2, PLOD1, PLOD2) (Table 7.5).

Table 7.5. Proteins relevant to cartilage biology differentially expressed in WT D0 and COLGALT2 KO D0 cells. The blue ones show increase between WT D0 and COLGALT2 KO D0 and the orange ones – a decrease.

Gene Name	logFC	Levels in COLGALT2 KO compared to WT cells	Potential Role in Cartilage Biology
ACAN	-8.370131347	significantly lower	Aggrecan, critical for cartilage structure and resilience.
COL11A1	-3.479133125	significantly lower	Type XI collagen, important for collagen fibril structure.
COL11A2	-4.524514526	significantly lower	Type XI collagen, important for collagen fibril structure.
COL12A1	-4.058899399	significantly lower	Type XII collagen, interacts with other collagens for structural support.
COL14A1	-2.657789285	significantly lower	Type XIV collagen, involved in assembly and strength.
COL15A1	-6.380164134	significantly lower	Component of collagen XV, contributes to structural integrity.
COL16A1	-2.910484851	significantly lower	Type XVI collagen, associated with fibrillar collagens network.
COL18A1	-6.95314307	significantly lower	Type XVIII collagen, associated with vascular and connective tissues.
COL1A1	-1.646751631	significantly lower	Type I collagen, provides tensile strength.
COL3A1	-2.593348542	significantly lower	Type III collagen, supports connective tissue integrity.
COL4A2	-3.011997005	significantly lower	Type IV collagen, part of basement membrane, influences development.
COL5A1	-1.680036207	significantly lower	Part of type V collagen, regulates fibrillar collagens assembly.
COL5A2	-3.739353043	significantly lower	Part of type V collagen, regulates fibrillar collagens assembly.
COL6A1	-2.283715834	significantly lower	Part of type VI collagen, contributes to structure.
COL6A2	-1.883829139	significantly lower	Part of type VI collagen, contributes to structure.
COL6A3	-3.308966313	significantly lower	Part of type VI collagen, contributes to structure.
COL7A1	-1.108192922	significantly lower	Part of type VII collagen, role in cartilage biology not known.
LOX	-1.396647786	significantly lower	Lysyl oxidase - involved in the crosslinking of collagen and elastin, important for cartilage strength.
LOXL1	-1.963256601	significantly lower	Lysyl oxidase-like 1 - similar function to LOX.
LOXL2	-4.514193681	significantly lower	Lysyl oxidase-like 2 - similar function to LOX.
LOXL3	-5.123413502	significantly lower	Lysyl oxidase-like 3 - similar function to LOX.
P3H1	-2.161535194	significantly lower	Proline hydroxylase - involved in collagen chain assembly, stability and cross-linking.
P4HA2	-1.091140851	significantly lower	Proline hydroxylase - involved in collagen chain assembly, stability and cross-linking.
PLD01	-1.009124165	significantly lower	Lysine hydroxylase - involved in collagen chain assembly, stability and cross-linking.
PLD02	-3.08880337	significantly lower	Lysine hydroxylase - involved in collagen chain assembly, stability and cross-linking.
SQX9	-1.44777279	significantly lower	Transcription factor critical for chondrogenesis and cartilage formation.
TGFB1	-1.543304075	significantly lower	Part of TGF-beta receptor, signaling pathway important in cartilage development.
CD44	0.620194605	significantly higher	Cell surface marker which helps chondrocytes interact with hyaluronic acid.
COMP	1.771091969	significantly higher	Cartilage oligomeric matrix protein - involved in the structural integrity of cartilage.
SMAD3	1.962528436	significantly higher	SMAD family member 3 - involved in TGF-beta signaling, relevant for cartilage development and homeostasis.

Between the cells that had undergone chondrogenic differentiation (D21), significant expression showed 3109 proteins (FDR<0.05) (894 down- and 2215 upregulated in COLGALT2 KO compared to WT) (Figure 7.14) (see Appendix E.). The proteins significantly downregulated in the COLGALT2 KO D21 compared to the WT D21 were overrepresented for GO terms including extracellular structure and matrix organisation and assembly, cartilage condensation, collagen activated signalling pathways, herpan sulphate and proteoglycan metabolism (Figure 7.15A). The proteins significantly upregulated in the edited cells at D21 were overrepresented for GO terms including metabolic and biosynthetic processes (Figure 7.15B).

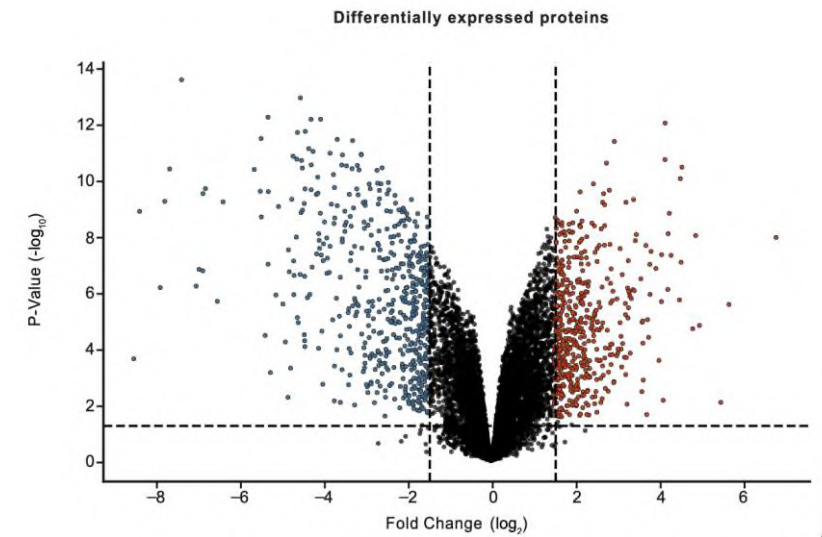


Figure 7.14. Volcano plot of protein expression differences between the WT D21 and COLGALT2 KO D21 cells. The plot displays the results of differential gene expression analysis. Log2 fold change (x-axis), indicates the magnitude of gene expression difference between the experimental conditions. Proteins to the right of the plot show upregulation in the proteome of COLGALT2 KO D21 cells compared to the WT D21 cells, while those to the left show downregulation. The y-axis represents $-\log_{10}$ P-value. Higher values indicate greater significance. Each dot represents an individual protein. The blue dots are proteins showing a significant ($P < 0.05$) downregulation in the proteome of COLGALT2 KO D21 cells compared to the proteome of WT D0 cells of more than 1.5FC. The red dots are proteins showing a significant ($P < 0.05$) upregulation in the proteome of COLGALT2 KO D21 cells compared to the secretome of WT D21 cells of more than 1.5FC. The threshold lines indicate cut-offs for significance levels ($P < 0.05$ (y-axis)), and fold change $FC > [1.5]$ (x-axis)). The P-values were calculated using t-tests with Benjamin-Hochberg correction for multiple testing. The volcano plots were produced in GraphPad Prism 9.0.

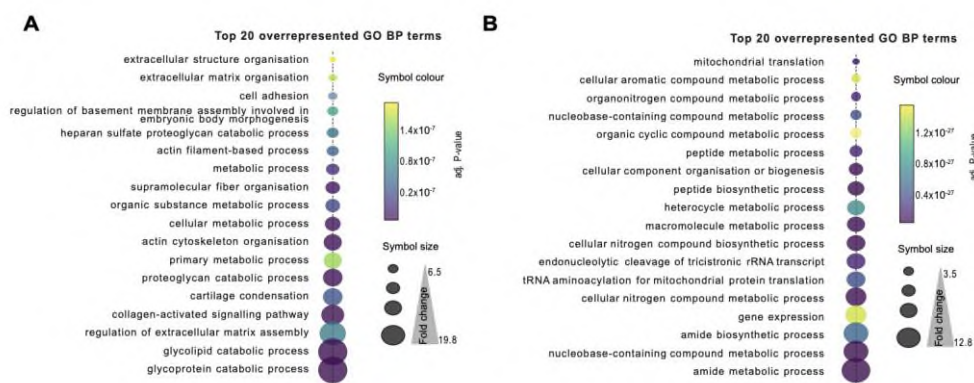


Figure 7.15. GO enrichment analysis of WT D21 vs COLGALT2 KO D21 cells. **A.** A dot plot visualising the top 20 enriched GO biological processes for the proteins showing a decrease in

COLGALT2 KO D21 cells compared to WT D21 cells. Each dot represents a distinct GO Biological process (BP) term. The dot size is proportional to the fold change (from 6.5 to 19.8), indicating the magnitude of protein expression change. Larger dots represent higher fold changes. The colour represents the FDR-adjusted P-value (Fischer's exact test with Benjamin-Hochberg adjustment), according to the indicated heatmap gradient where yellow-green colours represent less significant values and blue-purple colours show more significant values. R-Studio was used for the analysis and data visualisation. B. A dot plot visualising the top 20 enriched GO biological processes for the proteins showing an increase in COLGALT2 KO D21 cells compared to WT D21 cells. Each dot represents a distinct GO Biological process (BP) term. The dot size is proportional to the fold change (from 3.5 to 12.8), indicating the magnitude of protein expression change. Larger dots represent higher fold changes. The colour represents FDR-adjusted P-value (Fischer's exact test with Benjamin-Hochberg correction), according to the indicated heatmap gradient where yellow-green colours represent less significant values and blue-purple colours show more significant values. R-Studio was used for the analysis and data visualisation.

The cartilage biology-relevant proteins that show higher expression in the COLGALT2 KO compared to the WT cells before differentiation are collagens (I(2, III, V), lubricin, SMAD3, and CD44 (Table 7.6). Among the proteins that are reduced in expression at D0 in the edited cells are many collagens (I(1, II, IV, V, VI, VII, IX, XI, XII, XIV, XV, XVI, XVIII), ECM components (aggrecan, COMP, matrilin-3) and SOX9 (Table 7.6).

Table 7.6. Proteins relevant to cartilage biology differentially expressed in WT D21 and COLGALT2 KO D21 cells. The blue ones show increase between WT D21 and COLGALT2 KO D21 and the orange ones – a decrease.

Gene Name	logFC	Levels in COLGALT2 KO compared to WT cells	Potential Role in Cartilage Biology
ACAN	-5.339730978	significantly lower	Aggrecan, critical for cartilage structure and resilience.
COL11A1	-3.438662045	significantly lower	Type XI collagen, important for collagen fibril structure.
COL11A2	-3.71523082	significantly lower	Type XI collagen, important for collagen fibril structure.
COL12A1	-3.153346128	significantly lower	Type XII collagen, interacts with other collagens for structural support.
COL14A1	-1.49817908	significantly lower	Type XIV collagen, involved in assembly and strength.
COL15A1	-2.117842456	significantly lower	Component of collagen XV, contributes to structural integrity.
COL16A1	-0.892626978	significantly lower	Type XVI collagen, associated with fibrillar collagens network.
COL18A1	-4.161467556	significantly lower	Type XVIII collagen, associated with vascular and connective tissues.
COL1A1	-1.582561949	significantly lower	Type I collagen, provides tensile strength.
COL2A1	-6.009873928	significantly lower	Type II collagen, primary collagen in cartilage, provides tensile strength.
COL4A1	-2.617613024	significantly lower	Type IV collagen, part of basement membrane, influences development.
COL4A2	-3.461247862	significantly lower	Type IV collagen, part of basement membrane, influences development.
COL6A1	-0.950967051	significantly lower	Part of type VI collagen, contributes to structure.
COL6A2	-0.8688974	significantly lower	Part of type VI collagen, contributes to structure.
COL7A1	-2.392292329	significantly lower	Part of type VII collagen, role in cartilage biology not known.
COL9A3	-2.953737126	significantly lower	Type IX collagen, provides structural integrity.
COMP	-1.715269956	significantly lower	Cartilage oligomeric matrix protein - involved in the structural integrity of cartilage.
MATN3	-2.413655915	significantly lower	Matrilin-3, involved in matrix formation and stability.
SOX9	-2.807484502	significantly lower	Transcription factor critical for chondrogenesis and cartilage formation.
CD44	1.383286583	significantly higher	Cell surface marker which helps chondrocytes interact with hyaluronic acid.
COL1A2	0.940424163	significantly higher	Type I collagen, provides tensile strength.
COL3A1	1.24766578	significantly higher	Type III collagen, supports connective tissue integrity.
COL5A1	1.479862406	significantly higher	Part of type V collagen, regulates fibrillar collagen assembly.
LOX	3.12365417	significantly higher	Lysyl oxidase - involved in the crosslinking of collagen and elastin, important for cartilage strength.
PRG4	2.954128306	significantly higher	Lubricates cartilage surfaces, protects against wear.
SMAD3	2.623232978	significantly higher	Part of TGF-beta receptor, signaling pathway important in cartilage development.

Next, the influence of COLGALT2 KO upon chondrogenic differentiation was investigated. A comparative analysis was conducted on the proteomic profiles of

undifferentiated cells (COLGALT2 KO D0) and fully differentiated cells (COLGALT2 KO D21). Amongst the 8200 identified proteins, 1710 proteins exhibited statistically significant differences in expression levels (FDR<0.05) (1021 down- and 689 upregulated) (Figure 7.16A) (see Appendix E.).

The differentially expressed proteins were overrepresented for GO terms including cellular component organisation, extracellular matrix organisation, collagen fibril assembly and proteoglycan metabolism (Figure 7.16B).

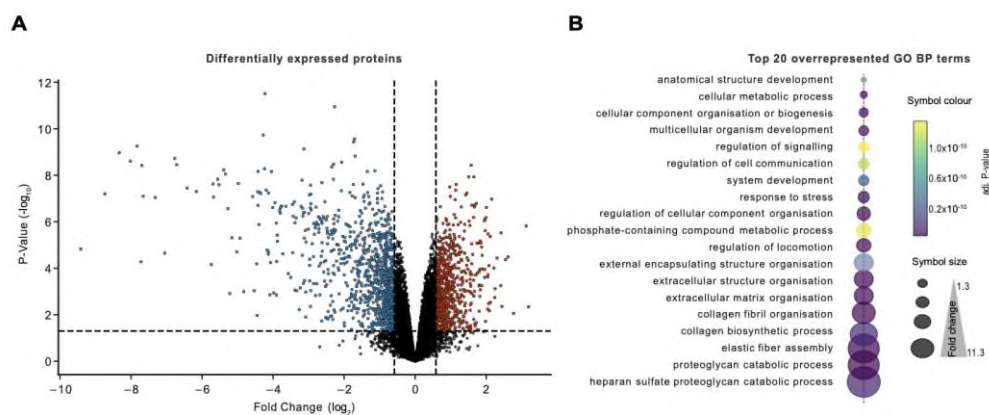


Figure 7.16. Differential protein expression analysis of COLGALT2 KO D0 and COLGALT2 KO D21 cells. **A.** Volcano plot of protein expression differences between COLGALT2 KO D0 and COLGALT2 KO D21 cells. The plot displays the results of differential gene expression analysis. Log₂ fold change (x-axis), indicates the magnitude of gene expression difference between the experimental conditions. Proteins to the right of the plot show upregulation in COLGALT2 KO D21 cells, while those to the left show downregulation. The y-axis represents -Log₁₀ P-value. Higher values indicate greater significance. Each dot represents an individual protein. The blue dots are proteins showing a significant ($P < 0.05$) downregulation in COLGALT2 KO D21 cells compared to COLGALT2 KO D0 cells of more than 1.5FC. The red dots are proteins showing a significant ($P < 0.05$) upregulation in COLGALT2 KO D21 cells compared to COLGALT2 KO D0 cells of more than 1.5 FC. The threshold lines indicate cut-offs for significance levels ($P < 0.05$ (y-axis)), and fold change $FC > [1.5]$ (x-axis)). The P-values were calculated using t-tests with Benjamin-Hochberg correction for multiple testing. The volcano plots were produced in GraphPad Prism 9.0. **B.** A dot plot visualising the top 20 enriched GO biological processes for the differentially expressed proteins between COLGALT2 KO D0 and COLGALT2 KO D21 cells. Each dot represents a distinct GO term. The dot size is proportional to the fold change (from 1.3 to 11.3), indicating the magnitude of protein expression change. Larger dots represent higher fold changes. The colour represents the FDR-adjusted P-value (Fischer's exact test with Benjamin-Hochberg correction), according to the indicated heatmap gradient where yellow-green colours represent less significant values and blue-purple colours show more significant values. R-Studio was used for the analysis and data visualisation.

The differentially expressed proteins with known relevance to cartilage biology were collagens (Type I, III, V, VI, VIII, XI, XII, XIV and XVI) and other ECM components (decorin, aggrecan, lumican, biglycan) (Table 7.7). Their expression increased following the differentiation.

Table 7.7. Proteins relevant to cartilage biology differentially expressed in COLGALT2 KO D0 and COLGALT2 KO D21 cells.

Gene name	logFC	Levels at D0 compared to D21	Potential Role in Cartilage Biology
ACAN	-5.676891119	significantly lower	Aggrecan, critical for cartilage structure and resilience.
BGN	-7.051375435	significantly lower	Biglycan, involved in collagen fibrillogenesis and matrix assembly.
CILP	-2.859163936	significantly lower	Cartilage intermediate layer protein, involved in cartilage matrix regulation.
COL11A2	-1.02699633	significantly lower	Type XI collagen, important for collagen fibril structure.
COL12A1	-2.629555445	significantly lower	Type XII collagen, interacts with other collagens for structural support.
COL14A1	-4.19292426	significantly lower	Involved in collagen network assembly, supports cartilage integrity.
COL16A1	-2.936210544	significantly lower	Type XVI collagen, associated with collagen network in cartilage.
COL1A1	-4.277364388	significantly lower	Forms type I collagen, key for tensile strength in connective tissues.
COL1A2	-2.297247092	significantly lower	Forms type I collagen, important for collagen structure.
COL1A2	-4.449748332	significantly lower	Forms type I collagen, important for collagen structure.
COL3A1	-4.228354679	significantly lower	Type III collagen, supports connective tissue integrity.
COL5A1	-3.131313898	significantly lower	Part of type V collagen, regulates fibrillar collagen assembly.
COL5A2	-4.243584045	significantly lower	Part of type V collagen, regulates fibrillar collagen assembly.
COL6A1	-3.18821206	significantly lower	Part of type VI collagen, contributes to cartilage microfibrillar network.
COL6A2	-2.945042782	significantly lower	Contributes to type VI collagen, important for cartilage matrix.
COL8A1	-4.429359206	significantly lower	Part of type VIII collagen, important for vascular and cartilage tissues.
DCN	-6.769540792	significantly lower	Decorin, regulates collagen fibril formation, affects matrix organization.
LUM	-7.657735168	significantly lower	Lumican, regulates collagen fiber arrangement, affects matrix structure.
PRG4	-3.079162785	significantly lower	Lubricates cartilage surfaces, protects against wear.
SLC26A2	-1.284159198	significantly lower	Sulphate transporter, important for cartilage development.
TGFBRI	-0.763991991	significantly lower	Part of TGF-beta receptor, signaling pathway important in cartilage development.
TIMP1	-0.692095198	significantly lower	Another MMP inhibitor, regulates matrix turnover and homeostasis.
TIMP2	-1.684230845	significantly lower	Inhibits MMPs, plays a role in extracellular matrix maintenance.
TIMP3	-0.603077054	significantly lower	Inhibits metalloproteinases, involved in extracellular matrix remodeling.

The effects of the knockout were further investigated through analysis of the secretome of the COLGALT2 KO D21 chondrocytes compared to the WT. Chondrocytes primarily serve the function of synthesizing and releasing ECM components. To assess the secretome composition of the COLGALT2 KO chondrocytes, a comparison was made with WT chondrocytes at D21. An examination of the expression levels of 572 secreted proteins revealed that 112 exhibited significantly elevated expression in the secretome of WT D21 chondrocytes compared to COLGALT2 KO D21 chondrocytes, whereas 104 displayed reduced expression (Figure 7.17A) (see Appendix G.). These differentially secreted proteins demonstrated an enrichment in biological processes related to cell differentiation, extracellular matrix organization, collagen fibril organization, and cartilage development (Figure 7.17B).

Among the proteins that are of potential relevance to cartilage biology and exhibited significant differences in secretion between WT D21 chondrocytes and COLGALT2 KO D21

chondrocytes, the collagen chains of type I and V, as well as MMP1 and MMP3, displayed notably increased secretion levels in COLGALT2 KO D21 chondrocytes. Conversely, collagen chains of type II, VI, XI, XII, and XVI, along with other crucial ECM components such as aggrecan, biglycan, lubricin, and filamin B, were less abundant in the secretome of the COLGALT2 KO D21 chondrocytes (Table 7.8).

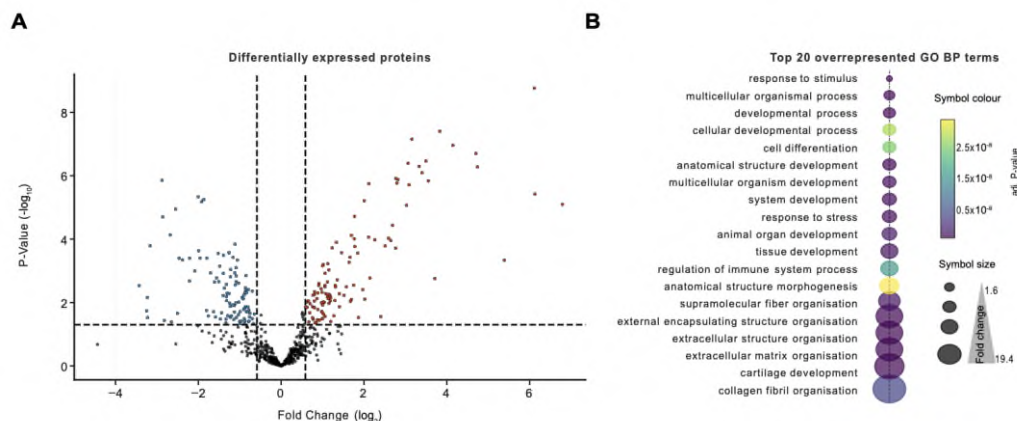


Figure 7.17. Differential protein expression analysis of the secretome of WT D21 and COLGALT2 KO D21 cells. **A.** Volcano plot of protein expression differences between the secretome of WT D21 and COLGALT2 KO D21 cells. The plot displays the results of differential gene expression analysis. Log2 fold change (x-axis), indicates the magnitude of gene expression difference between the experimental conditions. Proteins to the right of the plot show upregulation in the secretome of COLGALT2 KO D21 cells, while those to the left show downregulation. The y-axis represents $-\log_{10}$ P-value. Higher values indicate greater significance. Each dot represents an individual protein. The blue dots are proteins showing a significant ($P < 0.05$) downregulation in the secretome of COLGALT2 KO D21 cells compared to the secretome of WT D21 cells of more than 1.5FC. The red dots are proteins showing a significant ($P < 0.05$) upregulation in the secretome of COLGALT2 KO D21 cells compared to the secretome of WT D21 cells of more than 1.5FC. The threshold lines indicate cut-offs for significance levels ($P < 0.05$ (y-axis)), and fold change $FC > [1.5]$ (x-axis)). The P-values were calculated using t-tests with Benjamin-Hochberg correction for multiple testing. The volcano plots were produced in GraphPad Prism 9.0. **B.** A dot plot visualising the top 20 enriched GO biological processes for the differentially expressed proteins between WT D21 and COLGALT2 KO D21 cells. Each dot represents a distinct GO term. The dot size is proportional to the fold change (from 1.6 to 19.4), indicating the magnitude of protein expression change. Larger dots represent higher fold changes. The colour represents the FDR-adjusted P-value (Fischer's exact test with Benjamin-Hochberg correction), according to the indicated heatmap gradient where yellow-green colours represent less significant values and blue-purple colours show more significant values. R-Studio was used for the analysis and data visualisation.

Table 7.8. Proteins relevant to cartilage biology differentially secreted by WT D21 and COLGALT2 KO D21 cells. The blue ones show increase between WT D21 and COLGALT2 KO D21 and the orange ones – a decrease.

Gene name	logFC	Levels in WT D21 compared to KO D21	Potential Role in Cartilage Biology
COL1A2	-0.87244	significantly lower	Forms type I collagen, important for collagen structure.
COL5A1	-1.11621	significantly lower	Part of type V collagen, regulates fibrillar collagen assembly.
COL5A2	-0.90165	significantly lower	Part of type V collagen, regulates fibrillar collagen assembly.
MMP1	-1.37144	significantly lower	Breaks down extracellular matrix, plays a role in cartilage remodeling.
MMP3	-1.99335	significantly lower	Degrades matrix components, involved in cartilage repair.
ACAN	4.146729	significantly higher	Aggrecan, critical for cartilage structure and resilience.
BGN	1.331816	significantly higher	Biglycan, involved in collagen fibrillogenesis and matrix assembly.
COL11A1	3.830722	significantly higher	Type XI collagen, important for collagen fibril structure.
COL12A1	2.006236	significantly higher	Type XII collagen, interacts with other collagens for structural support.
COL16A1	1.772289	significantly higher	Type XVI collagen, associated with collagen network in cartilage.
COL2A1	5.385963	significantly higher	Essential for type II collagen, primary component of cartilage.
COL6A2	1.05767	significantly higher	Contributes to type VI collagen, important for cartilage matrix.
FLNB	0.81519	significantly higher	Influences skeletal development and cartilage structure.
PRG4	1.181903	significantly higher	Lubricates cartilage surfaces, protects against wear.

7.4. Discussion

This chapter presents the results of a proteomic investigation into the role of the COLGALT2 enzyme in chondrogenesis. To gain a comprehensive understanding of the influence of this collagen galactosyltransferase in a biologically relevant in vitro model, adipose-derived MSCs overexpressing SOX9 were differentiated into chondrocytes. These cells were reported to produce elevated levels GAG and type II collagen, and be responsive to both mechanical and inflammatory stimuli, resembling in that regard the behaviour of native human chondrocytes. The results presented in this chapter corroborate these observations. Examination of gene and protein expression levels associated with chondrocyte biology indicated that this cell model recapitulates the characteristics of healthy mature chondrocytes (Figure 7.1 and 7.4).

The experimental design enabled the exploration of protein changes during the chondrogenic differentiation of MSCs. Comparing the proteomic profiles of WT cells at different time points during chondrogenesis revealed that substantial phenotypic changes predominantly occur in the early stages of chondrogenesis. This phase is marked by significant shifts in metabolic activities and cellular structural processes, critical for the transition toward a chondrocyte-like phenotype (Figure 7.7 and Table 7.2). As differentiation progresses, there is a noticeable shift in the cellular proteome, aligning with the primary role of chondrocytes in ECM production and maintenance. This is evidenced by the increased expression of collagens and other key components of cartilage ECM (Figure 7.8 and Table 7.3). In the final stages of differentiation, chondrocytes exhibit mature/aged characteristics, with a notable decrease in ECM component production compared to earlier stages (Figure 7.9 and Table 7.4).

Other reported proteomic investigations into MSC chondrogenesis have revealed that the majority of altered proteins are associated with cellular metabolism, ECM production and

regulation, cytoskeleton organisation, and stress response^{13–16}. These results are in line with our observations (Figure 7.7 and Table 7.1). It is important, however, to highlight that all these studies differ in their experimental conditions by using different source of cells (bone marrow, human umbilical cord stroma), duration of differentiation (from 7 to 28 days) and time point collections. Such variations complicate the interpretation of comparisons between the results.

In stark contrast to the WT chondrocytes, the phenotype of COLGALT2 KO cells following the same 21-day differentiation period significantly differed from that of a healthy chondrocyte. COLGALT2 KO cells fail to manifest the distinct features associated with fully differentiated chondrocytes (Figure 7.5). Comparisons of the cells with depleted COLGALT2 levels and the unedited ASC52teloSox9 showed marked differences in the protein makeup both before and after chondrogenic differentiation (Figure 7.12 and 7.14, respectively). In particular the proteins that were downregulated in the knockout cells were overrepresented in GO terms associated with ECM organisation and assembly, as well as proteoglycan metabolism, cartilage condensation and collagen fibril assembly (Figure 7.13A and 7.13A). The levels of various collagen types (I – XII, XIV, XV, XVI, XVIII), collagen PTM enzymes (LOX, PLOD1, PLOD 2, P3H3, P4HA2), ECM components such as aggrecan, lubricin and COMP, as well as SOX9, the master-regulator of chondrogenesis, were differentially expressed between the COLGALT2 KO and the WT cells (Table 7.5 and 7.6). This observation suggests a role of the COLGALT2 enzyme in orchestrating the progression of chondrocyte-specific characteristics and in modulating the biosynthesis of essential ECM constituents vital for cartilage's structural integrity.

A comparative analysis of the proteomic profiles between undifferentiated COLGALT2 KO cells and those that had undergone full differentiation revealed that the affected proteins predominantly encompass various collagen types (I, III, V, VI, VIII, XI, XII, XIV, XVI) and other ECM constituents such as decorin, aggrecan, lumican, and biglycan (Table 7.7). Notably, during the differentiation process in WT cells, the expression of metalloproteases and their inhibitors decreases from D0 to Day 21 as the cells transitioned towards a more chondrocyte-like phenotype (Table 7.1). Conversely, COLGALT2 KO cells exhibit increased expression of metalloprotease inhibitors at Day 21 compared to D0 (Table 7.7), suggesting that these cells do not negatively regulate the function of metalloproteases. Metalloproteases play a crucial

role in maintaining ECM homeostasis by clearing damaged ECM components. Reduced COLGALT2 enzyme levels could lead to accumulation of collagen fibrils with aberrant glycosylation patterns in the ECM. These fibrils would necessitate continuous degradation, thus the need for unimpeded metalloprotease activity achieved through lower expression of TIMPs. This hypothesis gains further support from the increased secretion of metalloproteases (MMP1 and MMP3) from COLGALT2 KO D21 chondrocytes compared to WT D21 chondrocytes (Table 7.8).

The results presented in this chapter provide valuable insights into the role of COLGALT2 in chondrocyte biology and collagen biochemistry. While these findings do not definitively identify the specific targets of the COLGALT2 enzyme, the differential expression analyses of the secretome highlight collagens type II, VI, XI, XII, and XVI as potential targets. These collagens play crucial roles in cartilage biology, as detailed in Chapter I, where collagens II, IX, and XII form heteropolymers constituting a significant portion of the ECM collagen network^{17,18}. Collagen type VI has been implicated in regulating the pericellular matrix in cartilage^{19,20}. Collagens XI, XII and XVI primarily ensure the integrity and stability of the collagen fibril network²¹. Aberrant glycosylation of any or all of these collagens could disrupt the composition and structural integrity of the ECM secreted by chondrocytes, contributing to the observed weakened chondrocyte phenotype in COLGALT2 KO differentiated cells (Figure 7.4).

Given the pivotal roles of individual collagens and their post-translational modifications in cartilage, further comprehensive investigations into their roles are warranted and would contribute to the understanding of pathology of collagen rich tissues such as skin, bone, cartilage, and blood vessels. In this chapter, we focused on one of two enzymes involved in collagen glycosylation initiation, namely COLGALT2. While the expression pattern of *COLGALT2* suggests it is not the primary collagen galactosyltransferase in the body, our study was unable to assess the effects of complete glycosylation loss on chondrocyte biology due to incomplete depletion of the main enzyme, COLGALT1, in the ASC52teloSox9 cells. Despite reduced protein and gene expression of *COLGALT1* prior to differentiation in the ASC52teloSox9 cells (Figure 7.3), the prolonged expansion of the culture required to obtain necessary cell numbers might have given the unedited cells with a presumed selection advantage an opportunity to dominate the population, resulting in the cells that were

subjected to differentiation no longer having reduced levels of *COLGALT1*. Complete loss of *COLGALT1* activity results in embryonic lethality²². Due to the severe consequences of loss of this protein, it is highly likely that the cells with the depleted protein in my experiment could have had a growth disadvantage and been overcome by the unedited cells in the population following a number of population doublings. Despite being unable to study the effects of complete loss of galactosyltransferase activity or to compare the effects of *COLGALT1* loss to *COLGALT2* loss, the striking differences in phenotype between WT chondrocytes and *COLGALT2* KO chondrocytes suggest that *COLGALT2* is a key player in chondrocyte biology and collagen biochemistry.

Proteomics data offers valuable insights into protein expression and function but comes with several challenges. The complexity and dynamic nature of the proteome make it difficult to achieve comprehensive coverage, with low-abundance proteins and post-translational modifications often being overlooked. Data sensitivity to sample preparation, detection bias, and limited dynamic range can skew results, while quantification and reproducibility issues further complicate analysis. Advanced bioinformatics tools are needed to handle large datasets, but technical artifacts and variability between experiments can reduce reliability. Additionally, high costs and the need for specialized expertise limit accessibility. Proteomics data frequently lack direct functional insights and are often difficult to compare across studies due to methodological variability. These challenges underscore the necessity of validating findings through complementary techniques such as Western blotting for key targets, enzyme-linked immunosorbent assays (ELISA), immunohistochemistry, and flow cytometry for surface and intracellular proteins. Additionally, unbiased replication in independent proteomics experiments, incorporating a substantial number of biological and technical replicates, is crucial for ensuring data reliability. Integrating proteomics with other omics approaches further enhances the depth and breadth of biological understanding.

In previous chapters I described the effect common genetic variants associated with OA have on the regulation of the expression of *COLGALT2* in joint-relevant tissues from development to old age. In light of the observations made in this chapter regarding the role of the *COLGALT2* enzyme in chondrocyte biology and collagen biochemistry, a possible connection emerges between the genetic variants governing gene expression and the

development and progression of osteoarthritic pathology through aberrant glycosylation of ECM collagens.

7.5. References

1. [review] Ricard-Blum, S. The Collagen Family. *Cold Spring Harb. Perspect. Biol.* **3**, a004978 (2011).
2. [review] Soroushanova, A. *et al.* The Collagen Suprafamily: From Biosynthesis to Advanced Biomaterial Development. *Adv. Mater.* **31**, e1801651 (2019).
3. Schegg, B., Hülsmeier, A. J., Rutschmann, C., Maag, C. & Hennet, T. Core Glycosylation of Collagen Is Initiated by Two $\beta(1\text{-O})$ Galactosyltransferases. *Mol. Cell. Biol.* **29**, 943–952 (2009).
4. [review] Forlino, A., Cabral, W. A., Barnes, A. M. & Marini, J. C. New perspectives on osteogenesis imperfecta. *Nat. Rev. Endocrinol.* **7**, 540–557 (2011).
5. [review] Prockop, D. J. *et al.* Type I procollagen: The gene-protein system that harbors most of the mutations causing osteogenesis imperfecta and probably more common heritable disorders of connective tissue. *Am. J. Méd. Genet.* **34**, 60–67 (1989).
6. Lehmann, H. W. *et al.* Composition and posttranslational modification of individual collagen chains from osteosarcomas and osteofibrous dysplasias. *J. Cancer Res. Clin. Oncol.* **121**, 413–418 (1995).
7. Stawikowski, M. J., Aukshi, B., Stawikowska, R., Cudic, M. & Fields, G. B. Glycosylation Modulates Melanoma Cell $\alpha 2\beta 1$ and $\alpha 3\beta 1$ Integrin Interactions with Type IV Collagen*. *J. Biol. Chem.* **289**, 21591–21604 (2014).
8. Miyatake, S. *et al.* Biallelic COLGALT1 variants are associated with cerebral small vessel disease. *Ann. Neurol.* **84**, 843–853 (2018).
9. Finger, F. *et al.* Molecular phenotyping of human chondrocyte cell lines T/C-28a2, T/C-28a4, and C-28/I2. *Arthritis Rheum.* **48**, 3395–3403 (2003).
10. Schnabel, M. *et al.* Dedifferentiation-associated changes in morphology and gene expression in primary human articular chondrocytes in cell culture. *Osteoarthr. Cartil.* **10**, 62–70 (2002).
11. [review] Somoza, R. A., Welter, J. F., Correa, D. & Caplan, A. I. Chondrogenic Differentiation of Mesenchymal Stem Cells: Challenges and Unfulfilled Expectations. *Tissue Eng. Part B: Rev.* **20**, 596–608 (2014).
12. Katz, D. B., Huynh, N. P. T., Savadipour, A., Palte, I. & Guilak, F. An immortalized human adipose-derived stem cell line with highly enhanced chondrogenic properties. *Biochem. Biophys. Res. Commun.* **530**, 252–258 (2020).

13. Peffers, M. J., Collins, J., Loughlin, J., Proctor, C. & Clegg, P. D. A proteomic analysis of chondrogenic, osteogenic and tenogenic constructs from ageing mesenchymal stem cells. *Stem Cell Res. Ther.* **7**, 133 (2016).
14. Rocha, B. *et al.* Metabolic Labeling of Human Bone Marrow Mesenchymal Stem Cells for the Quantitative Analysis of their Chondrogenic Differentiation. *J. Proteome Res.* **11**, 5350–5361 (2012).
15. Fuente, A. D. la *et al.* Proteome Analysis During Chondrocyte Differentiation in a New Chondrogenesis Model Using Human Umbilical Cord Stroma Mesenchymal Stem Cells*. *Mol. Cell. Proteom.* **11**, M111.010496 (2012).
16. Wang, D. *et al.* Proteomic Profiling of Bone Marrow Mesenchymal Stem Cells upon Transforming Growth Factor β 1 Stimulation*. *J. Biol. Chem.* **279**, 43725–43734 (2004).
17. Mendler, M., Eich-Bender, S. G., Vaughan, L., Winterhalter, K. H. & Bruckner, P. Cartilage contains mixed fibrils of collagen types II, IX, and XI. *J. cell Biol.* **108**, 191–197 (1989).
18. Fernandes, R. J., Schmid, T. M. & Eyre, D. R. Assembly of collagen types II, IX and XI into nascent hetero-fibrils by a rat chondrocyte cell line. *Eur. J. Biochem.* **270**, 3243–3250 (2003).
19. [review] Wilusz, R. E., Sanchez-Adams, J. & Guilak, F. The structure and function of the pericellular matrix of articular cartilage. *Matrix Biol.* **39**, 25–32 (2014).
20. Zelenski, N. A. *et al.* Type VI Collagen Regulates Pericellular Matrix Properties, Chondrocyte Swelling, and Mechanotransduction in Mouse Articular Cartilage. *Arthritis Rheumatol.* **67**, 1286–1294 (2015).
21. [review] Luo, Y. *et al.* The minor collagens in articular cartilage. *Protein Cell* **8**, 560–572 (2017).
22. He, L. *et al.* Down-regulation of GLT25D1 inhibited collagen secretion and involved in liver fibrogenesis. *Gene* **729**, 144233 (2020).

Chapter 8. General Discussion

Exploring the intricacies involved in deciphering GWAS signals reveals a myriad of hurdles¹. The fundamental challenge resides in the essence of GWAS, which often pinpoints genetic variations linked to traits or diseases without providing immediate insights into the mechanisms through which these variations exert their effects. This enigma is compounded by the frequent occurrence of GWAS hits in non-coding regions of the genome, complicating the inference of their functional consequences^{2,3}. These regions might influence gene regulation, but comprehending the precise mechanisms necessitates intricate investigations. Furthermore, the polygenic character of many traits introduces an additional layer of complexity^{3,4}. Numerous genetic variations, each contributing minimally, collectively shape the trait, making it arduous to segregate the impact of individual variations. This polygenic nature implies that the causal mechanisms can be diffuse and intricately interconnected, involving complex gene networks rather than singular gene effects. In addition, the phenomenon of linkage disequilibrium further complicates the interpretation of GWAS signals⁵. Variants identified by GWAS are often in linkage disequilibrium with other variants, meaning they are inherited together. Disentangling which variant or variants within a linked group are genuinely causal represents a substantial challenge. Addressing this requires meticulous fine-mapping approaches to narrow down the potential causal variants. Lastly, the issue of population specificity presents itself⁶. GWAS outcomes can significantly vary across diverse populations due to disparities in genetic backgrounds and environmental exposures. This underscores the necessity for diverse population studies to ensure the broad applicability and comprehension of GWAS findings. These challenges highlight the complexity of translating GWAS findings into a profound understanding of genetic influence on intricate traits and diseases. Tackling these issues necessitates a blend of advanced genetic analysis techniques, functional investigations, and a nuanced interpretation of genetic data⁷.

Given these challenges, there exists a compelling imperative for functional validation studies to confirm and elucidate the biological significance of GWAS findings. Such studies involve meticulous scrutiny of the identified genetic variants to ascertain their influence on gene expression and function⁷. This encompasses assessing how these variants interact with regulatory elements, impact transcription factor binding, and alter gene expression patterns⁷⁻

[11](#). Functional validation is indispensable not only to corroborate the roles of identified variants but also to gain profound insights into the biological pathways and mechanisms underpinning complex traits and diseases. Functional validation studies explore how these variants affect gene expression, protein function, and ultimately, the phenotypical attributes of cells and tissues.

The intricacy of genetic regulation necessitates a multifaceted approach to functional validation studies. Distinct tissues and cell types can exhibit varying gene expression profiles and regulatory mechanisms. Consequently, it is imperative to investigate GWAS-identified variants across a spectrum of tissues and employ diverse methodologies. This comprehensive approach enables a more precise comprehension of how variants contribute to traits or diseases in a context-dependent manner. Techniques such as CRISPR-Cas9 gene editing, eQTL mapping, and chromatin conformation capture methods are employed to explore the regulatory landscapes across different tissues and decipher the intricate network of genetic regulation^{[12,13](#)}. Expanding upon the multifaceted approach for functional validation of GWAS signals, it is essential to underscore the amalgamation of cutting-edge technologies and interdisciplinary methods^{[14,15](#)}. Transcriptomic and proteomic analyses can provide insights into the impact of genetic variants at the RNA and protein levels, respectively. Epigenome interrogations help understand genetic regulation patterns and serve to connect the knowledge of the environmental context and the genetic landscape of any disease^{[16](#)}.

All of these investigations are vital for translating the wealth of data generated by GWAS into pragmatic applications, such as the development of novel therapeutic interventions and enhancements in disease diagnosis and prognosis.

Given the global prevalence of OA, its economic and social burden, the complex nature of the disease and the current lack of DMOADs, gaining further insight into any potentially targetable-by-therapeutics biological pathways is of paramount importance for healthcare systems worldwide. In this thesis I present the results from a detailed functional investigation of two independent OA risk GWAS hits, rs11583641 and rs1046934.

The first of these two signals, rs11583641, was found to function as a mQTL at a specific intronic CpG site, cg18131582, within the gene body of *COLGALT2*^{[17](#)}, an effect I replicated in an independent cohort of patient samples. It was also confirmed that the effect of the SNP on DNAm levels was not restricted to cg18131582 but affected 5 CpG sites within

a 398bp DMR marked as an enhancer in joint-relevant tissues (Chapter 3, Figure 3.2A). The effect allele (C) of the genetic variant rs11583641 associated with reduced DNAm at all differentially methylated sites, particularly pronounced in knee cartilage at cg18131582 and its neighbouring two CpG sites (Chapter 3, Figure 3.4B). Allele-specific expression analysis confirmed that the effect allele of the SNP correlated with increased expression of the *COLGALT2* gene (Chapter 3, Figure 3.15A). Additionally, correlations between allele-specific expression ratios and DNAm revealed an meQTL at 3 of the differentially methylated CpG sites (Chapter 3, Figure 3.15C). The AEI effect reported for this locus was replicated by a Netherlands research group using another SNP in LD with rs11583641, rs12023991 ($r^2=1.0$, GBR)¹⁸. The effect reported by the researchers was in the same direction as the effect described in Chapter 3. In the Netherlands study, differences between AEI ratios were reported between lesioned and preserved cartilage from the same patients with increase in *COLGALT2* being associated with lesioned cartilage¹⁸. In my studies, I detected strong AEI effects in some but not all cartilage samples from heterozygous patients (Chapter 3, Figure 3.15B). One way of interpreting this is by considering the potential effects of other variants that could be masking or counteracting the effect. Additionally, in light of the differences in AEI ratios between lesioned and preserved cartilage reported by the Netherlands group, it is possible that the effect was masked by the fact that the cartilage tissue samples used in my investigations were collected from the entire OA affected joint and the severity of cartilage degradation between OA patients is known to differ.

To assess the regulatory function of the DMR, I used a reporter gene assay, demonstrating its sensitivity to DNAm changes with lower DNAm at the enhancer leading to higher reporter gene activity (Chapter 3, Figure 3.7B). Furthermore, genetic influence on enhancer function was observed with rs734657, a SNP located within the enhancer and in high LD with rs11583641, associating with different levels of expression of the reporter gene. The SNP is a compelling target for further investigations as it is located in a regulatory region and in open chromatin in both hip and knee cartilage from OA patients (Chapter 6, Figure 6.13A). While informative, reporter gene analysis has its limitations and in order to comprehend the impact of the enhancer in its natural genomic context within TC28a2 chondrocytes, genome editing approaches were employed to delete the enhancer using CRISPR-Cas9, confirming *COLGALT2* as a target gene of the OA signal marked by rs11583641

(Chapter 3, Figure 3.11). Furthermore, the use of a catalytically inactive Cas9 fused to DNMT3a or TET1 enzymes demonstrated a causal relationship between DNAm and *COLGALT2* expression in this *in vitro* system (Chapter 3, Figure 3.13). These investigations consistently confirmed the significance of the region surrounding CpGs 8, 9, and 10, highlighting the rare instance where a causal gene has been identified at an OA risk locus through direct molecular inquiry. This underscores the effectiveness of employing epigenome editing techniques to unveil the functional roles of risk loci and their underlying mechanisms. However, it is essential to acknowledge that a direct causal link between SNP genotype and DNAm in primary cells from patient samples remains was not established. This remains a limitation of the studies described in this thesis and many other such functional investigations into the effects of GWAS signals.

A second genetically independent OA risk SNP, rs1046934, was also investigated¹⁹. This SNP resides within the gene *TSEN15*, in close proximity to *COLGALT2* (Chapter 4, Figure 4.5A). Although both rs1046934 and another SNP within the LD block are coding variants located within *TSEN15*, computational mutagenesis analyses indicated a low probability of these substitutions affecting the structure of the TSEN15 protein or the broader tRNA splicing (TSEN) complex (Chapter 4, Figure 4.2).

The variant rs1046934, independently of the locus marked by rs11583641, was observed to exert its influence by modulating the expression of *COLGALT2* through alterations in DNAm levels at two CpG sites, namely cg15204595 and cg21606956, which are located within regions marked as enhancers in joint relevant tissues. Using Illumina DNA methylation array data obtained from arthroplasty cartilage samples and pyrosequencing data from an independent cohort of arthroplasty cartilage samples, it was determined that the risk allele (A) of rs1046934 correlated with reduced DNAm levels at these two enhancer CpG sites (Chapter 4, Figure 4.1 and 4.5B). An additional CpG site, cg01436608, was identified as potentially impacted by the SNP, although it was excluded from further investigation due to the absence of reliable pyrosequencing assays (Chapter 4, Figure 4.1). Nevertheless, the proximity of cg01436608 to the transcription start site of *COLGALT2* and its association with the enhancer containing cg21606956 in MSCs (Chapter 4, Figure 4.5A) makes it a promising subject for future interrogations.

Moreover, apart from its association with reduced DNAm levels at cg15204595 and cg21606956, the OA risk allele of rs1046934 was also linked to increased expression of both *COLGALT2* and *TSEN15* in arthroplasty patients (Chapter 4, Figure 4.4). A further correlation between DNAm and *COLGALT2* expression was identified, marking meQTLs (Chapter 4, Figure 4.7). Reporter gene assays confirmed that the genomic regions housing cg15204595 and cg21606956 function as enhancers in TC28a2 chondrocytes (Chapter 4, Figure 4.9) and manipulation of the DNAm levels at these CpG sites using a dCas9 and TET1 fusion protein complex led to increased *COLGALT2* expression, confirming causal interactions between the DNAm levels at these sites and *COLGALT2* expression (Chapter 4, Figure 4.10A).

The *in vitro* findings on this locus presented in this chapter support the data collected from arthroplasty cartilage samples, indicating that the enhancers regulate gene expression and that the genotype at rs1046934 plays a role in this regulation. Both *COLGALT2* and *TSEN15* appear to be affected by the OA-associated signal rs1046934 but the mechanisms underlying the SNP's impact on the expression of these two genes are dissimilar. While the regulatory mechanism governing *TSEN15* expression through the enhancers remains undetermined, our data provide support for a causal connection between DNA methylation levels (correlated with the genotype at rs1046934) at the two enhancer CpG sites and *COLGALT2* expression, suggesting this as a potential regulatory mechanism.

In silico analyses revealed that cg15204595 and cg21606956 reside in open chromatin regions within joint tissues and are in or near TF binding sites (Chapter 4, Figure 4.11). Methylation of cytosines in CpG dinucleotides can alter the steric and hydrophobic properties of the dinucleotide which can lead to structural changes in the DNA helix, including an increased helix stiffness²⁰⁻²². The impact of DNA methylation on TF binding has been extensively documented for various proteins, often resulting in the prevention of TF binding²³⁻²⁵. This DNA methylation-dependent TF binding affinity may constitute the mechanism by which *COLGALT2* expression in chondrocytes is regulated by the OA-associated signals rs11583641 and rs1046934.

The findings described in this thesis highlight the complexity of gene expression regulation. Multiple SNPs from the same LD block or even from completely independent loci can govern the expression of a single gene with studies showing that allelic heterogeneity may be more common than originally anticipated²⁶⁻²⁸. Similarly, a single genetic signal can

have multiple targets and exert its effect in different ways²⁸. To further complicate matters, these effects are highly tissue specific, requiring scientists to consider multiple disease-relevant tissues when drawing conclusions based on functional analyses carried out on GWAS signals.

In my investigations I examined the effects of the two OA GWAS signals associated with *COLGALT2*, rs11583641 and rs1046934, in cartilage and then in synovium derived from arthroplasty patients. While the rs1046934 signal did not appear to be active in synovium (Chapter 5, Figure 5.3 and 5.5), the rs11583641 was (Chapter 5, Figure 5.1 and 5.4). The mQTL analysis for the CpGs associated with the genotype at rs11583641 revealed that in synovium, the risk allele (C) of the SNP is associated with increased DNAm compared to the non-risk allele (T), which is the opposite of the observations made in cartilage (Chapter 5, Figure 5.2A). This mQTL effect manifests at a greater number of CpG sites within the enhancer in synovium (7 out of 12 investigated) compared to cartilage (5 out of 12 investigated), and the effect is more pronounced at all differentially methylated CpGs. Furthermore, the average DNAm levels at the enhancer are lower in synovium in comparison to cartilage (Chapter 5, Figure 5.2B). It is essential to note that the samples employed in these analyses were not matched, and thus, direct comparisons between DNAm levels in different tissues within the same patient were not feasible. AEI analysis revealed that the risk allele corresponds to lower *COLGALT2* expression relative to the non-risk allele (Chapter 5, Figure 5.4A), unlike what is observed in cartilage (Chapter 3, Figure 3.15A). This allelic imbalance is less prominent in synovium when compared to cartilage.

To gain further insights, a reporter gene assay was carried out in the SW982 synovial fibroblast sarcoma cell line (an *in vitro* model for synovium) using the same methylated and unmethylated constructs with the different enhancer SNP haplotypes (rs943409 and rs734657, residing within the DMR) as the ones used in the TC28a2 cells. Similar to the findings in chondrocytes, the region housing the DMR was found to function as an enhancer in the synovial fibroblasts (Chapter 5, Figure 5.6). Moreover, akin to observations in chondrocytes, an allelic effect was evident. Constructs containing the C allele of rs734657, corresponding to the C allele of rs11583641, exhibited heightened expression of the reporter gene when compared to constructs containing the non-risk allele. The methylation status of the regions also exerted an influence on gene expression, with lower DNAm levels leading to

elevated reporter gene readings. These results were not corroborated by our observations in patient samples, where the risk allele of rs11583641 was linked to increased DNAm at the enhancer and reduced *COLGALT2* expression. It is imperative to recognize that while reporter gene analyses offer a cost-effective, expeditious, and sensitive means of assessing the regulatory function of DNA elements, there are inherent limitations to this *in vitro* model. The absence of three-dimensional conformation, as seen *in vivo*, can impact the accessibility of the region to TFs. Furthermore, *in vitro* DNA methylation of constructs using enzymes results in constructs that are uniformly unmethylated or methylated in all cells, whereas the DNAm levels of a region within living cells is more nuanced.

To elucidate more accurately the impact of DNAm at the enhancer on gene expression in SW982 cells, we executed a targeted RNA-guided methylation and demethylation of the region using dCas9 coupled with DNMT3a and TET1, respectively. The results of this experiment unveiled an antagonistic relationship between DNAm at the enhancer and *COLGALT2* expression, confirming the gene as a target of the enhancer (Chapter 5, Figure 5.8). As with chondrocytes, the DNAm status of cg18131582 and its neighbouring CpGs emerged as the most influential on *COLGALT2* expression in synovial fibroblasts. DNAm levels at these CpGs were lower in SW982 cells, and it is plausible that TFs that bind in a DNAm-dependent manner to these CpGs govern *COLGALT2* expression. Using the JASPAR TF database, TFs predicted to bind to this region were identified and were found to be expressed in either cartilage and/or synovium, with several exhibiting different levels of expression between the two tissue types (Chapter 5, Figure 5.9).

The rs11583641/cg18131582 enhancer region, housing the differentially methylated CpGs, was demonstrated to regulate *COLGALT2* expression in both cartilage and synovium from OA patients. Manipulation of CpG methylation using dCas9-DNMT3a or dCas9-TET1 proteins, in conjunction with six guide RNAs targeting the region, elucidated a causal connection between DNAm at the enhancer and *COLGALT2* expression in both TC28a2 chondrocytes and SW982 synovial fibroblasts. This suggests a common mechanism through which the enhancer modulates *COLGALT2* expression in cartilage and synovium. However, it is pertinent to note that the mechanism by which the OA GWAS SNP influences DNAm at the enhancer and gene expression appears to diverge between the two tissues. In patient samples, we observed mQTL and eQTL effects that displayed opposing trends in cartilage as

compared to synovium. In cartilage, the risk allele of rs11583641 correlated with decreased DNAm at the enhancer and heightened expression of *COLGALT2*, while in synovium, the risk allele was associated with increased methylation and diminished gene expression. I hypothesised that these contrasting directions of mQTL and eQTL effects are a consequence of disparities in TFs binding to the putative causal SNP between cartilage and synovium. These differences may be quantitative, with specific TFs being more abundantly expressed in one tissue relative to the other, or qualitative, with distinct repertoires of TFs expressed in different cell types. Since the exact causal SNP(s) at the locus remains undiscovered, it is currently not feasible to confirm the identity of TFs binding to the SNP and whether they are differentially expressed in cartilage and synovium. While the impact of the alleles of the disease-associated SNP on DNAm at the enhancer differs between the tissues, the mechanism through which the enhancer influences *COLGALT2* expression via DNAm remains consistent. This statement is corroborated by the results obtained from targeted DNAm modulation of the enhancer in TC28a2 chondrocytes and SW982 cells.

My comprehensive exploration of the functional repercussions of the *COLGALT2* loci marked by rs11583641 and rs1046934 in both cartilage and synovium has unveiled intriguing insights. While one of the signals, rs1046934, appears to not be active and to not contribute to changes in *COLGALT2* expression in synovium tissue from OA patients, the other signal exerts an influence on gene expression in both tissue types but via different mechanisms. These findings underscore the importance of conducting targeted functional analyses of GWAS signals across all pertinent disease-relevant tissues and comprehending the mechanisms that underlie any effects on gene expression before contemplating potential therapeutic targets.

Instances of risk variants exerting opposing effects in distinct tissues are infrequent, as evidenced by a 2022 epigenome-wide study of OA risk signals in the synovium and the intact cartilage from OA affected joints²⁹. The authors found that in only 0.02% of instances where mQTLs are detected in both tissues, they operate in opposing directions, making our findings highly uncommon and underscoring the importance of conducting targeted functional analyses across a spectrum of tissues to gain a holistic understanding of gene regulation dynamics in health and disease for different signals.

Since the effects of genetic variation on gene expression and/or epigenetic marks can be contingent upon specific cell types and can manifest under divergent physiological contexts or even during distinct developmental stages, the genetic and epigenetic regulation of *COLGALT2* was investigated in developmental cartilage samples. This inquiry aimed to enhance our comprehension of the developmental underpinnings of OA. The QTLs associated with both gene expression and DNAm at the rs11583641 and rs1046934 OA risk loci in aged cartilage were also observed in cartilage and mixed tissue limb bud samples originating from early developmental stages (Chapter 6, Figures 6.1, 6.5, 6.7 and 6.9). The AEI and mQTL effects associated with the rs11583641 signal in arthroplasty cartilage, had a more pronounced impact on DNAm levels at the enhancer containing cg18131582 in foetal cartilage compared to aged cartilage (Chapter 6, Figure 6.3 and 6.7). According to my interpretation of the results described in Chapter 6, during development DNAm levels at the rs11583641 associated enhancer region undergo precise regulation in developmental cartilage, necessitating lower methylation levels to tightly control *COLGALT2* expression as the joint forms. As individuals age, methylation levels at the enhancer increase to maintain appropriate collagen glycosylation in aging cartilage. Conversely, in cartilage affected by OA, there is a partial reversal of this process, with decreased DNA methylation levels compared to non-OA-affected control samples, potentially indicative of chondrocytes attempting cartilage repair and adopting a more developmental-like phenotype, a phenomenon reported in the literature³⁰⁻³².

The prevailing hypothesis suggests that the interplay between genetic variants, epigenetic marks, and gene expression regulation is mediated by TFs. TF binding can be influenced by DNA methylation, but it is chromatin accessibility that predominantly governs TF binding^{21,22,24,25,33-35}. Lower DNAm levels in foetal cartilage imply that TFs may bind more readily to the region. ATAC-seq data further supports this notion, indicating increased chromatin accessibility in the region harbouring the differentially methylated rs11583641 CpGs in foetal samples compared to aged ones (Chapter 6, Figure 6.13A). This suggests that the enhancer is more active in developmental cartilage and potentially subject to stricter regulatory control by TFs. This accessibility in the chromatin at the location of the enhancer in both aged and developmental cartilage also underscores the potential functional effect of the rs734657 SNP highlighted by the reporter gene analysis (Chapter 3, Figure 3.6B).

The mQTLs associated with this enhancer were found in foetal limb bud tissue but were limited to specific CpGs (Chapter 6, Figure 6.1 and 6.3). This implies that the differential methylation effect is primarily attributed to the cartilage component within the heterogeneous limb bud tissue. CRISPR/Cas9 experiments targeting DNAm changes at these specific CpGs in TC28a2 chondrocytes had the most substantial impact on *COLGALT2* expression, reinforcing their significance. In contrast, the mQTL effect in other joint tissues may operate differently, as we have seen in the synovium of aged OA patients, potentially counterbalanced by the cartilage component in the mixture, with only the most pronounced effects being detectable. It's important to note that the limb bud samples originated from earlier developmental stages, and epigenetic influences on regulatory elements may differ from those in later stages of development.

For the rs11583641 signal, the eQTL effect detected in developmental samples was not consistently present in all foetal samples, similar to the observations made in aged cartilage (Chapter 6, Figure 6.7), possibly due to additional unrelated genetic variation affecting expression or the nature of the cartilage sample. While it was possible that the sporadic presence of an effect in aged cartilage be attributed to the nature of the cartilage samples obtained from the patients (namely being intact), this theory is undermined by the observations made in developmental cartilage which does not have signs of OA- or age-related degradation. This leaves the possibility that the lack of AEI effect is due to undetected genetic variation in some patients.

The second OA locus associated with *COLGALT2* expression, rs1046934, was also examined in foetal samples, revealing mQTLs in both foetal cartilage and limb bud samples, with consistent effects on *COLGALT2* expression (Chapter 6, Figure 6.5). Differences in DNAm levels and chromatin accessibility were observed between developmental and aged cartilage samples at the two enhancers harbouring the differentially methylated CpGs, cg15204595 and cg21606956 (Chapter 6, Figure 6.13B and C). Based on these observations I hypothesise that the enhancers associated with rs1046934 function redundantly, fine-tuning gene regulation. In OA cartilage, however, dysregulation of epigenetic factors leads to both enhancers being actively involved and contributing to higher *COLGALT2* expression levels.

While the molecular studies carried out in cartilage and synovium samples from OA patients and in cartilage from developmental samples shed light on the complex regulatory

landscape in healthy and diseased joint tissues, the role of the protein encoded by the gene regulated by these GWAS signals remained unclear. In order to build a more comprehensive picture of how the GWAS signals associated with *COLGALT2* regulation contributed to the development of OA, I undertook a study exploring the role of the enzyme in chondrocyte development. To acquire a more complete understanding of the impact of this collagen galactosyltransferase in a biologically pertinent *in vitro* model, I induced the differentiation of adipose-derived MSCs overexpressing SOX9 (ASC52teloSox9) into chondrocytes. The evaluation of chondrocyte biology relevant genes and their protein expression levels indicated that this model faithfully recapitulates the characteristics observed in healthy mature chondrocytes (Chapter 7, Figure 7.1 and 7.4).

Our experimental design allowed for the investigation of protein alterations during the chondrogenic differentiation of MSCs. Comparative analysis of the proteomic profiles of wild-type (WT) cells at various time points during chondrogenesis unveiled significant phenotypic alterations, primarily occurring in the early stages of chondrogenesis (Chapter 7, Figure 7.7 and 7.8). This phase is distinguished by substantial shifts in metabolic activities and cellular structural processes, which are pivotal for the transition towards a chondrocyte-like phenotype. As the differentiation process advances, there is a discernible transformation in the cellular proteome, aligning with the primary role of chondrocytes in ECM production and maintenance, substantiated by a heightened expression of various types of collagens and other crucial constituents of cartilage ECM. In the later stages of differentiation, chondrocytes exhibit mature/aged characteristics, marked by a notable reduction in ECM component production compared to earlier stages (Chapter 7, Figure 7.10).

In stark contrast to WT chondrocytes, chondrocytes with depleted *COLGALT2* (*COLGALT2* KO) after the same 21-day differentiation period display a significant deviation from the attributes of healthy chondrocytes (Chapter 7, Figure 7.5). Comparative assessments of cells with reduced *COLGALT2* levels and unmodified (WT) ASC52teloSox9 cells revealed substantial differences in the protein composition both before (Chapter 7, Figure 7.11 and 7.12) and after chondrogenic differentiation (Chapter 7, Figure 7.13 and 7.14). Particularly, the proteins downregulated in the knockdown cells were overrepresented in gene ontology terms associated with ECM organization and assembly, proteoglycan metabolism, cartilage condensation, and collagen fibril assembly. Diverse collagen types (I - XII, XIV, XV, XVI, XVIII),

collagen post-translational modification enzymes (LOX, PLOD1, PLOD 2, P3H3, P4HA2), ECM components such as aggrecan, lubricin, COMP, and SOX9, the master regulator of chondrogenesis, exhibited differential expression between COLGALT2 KO and WT cells (Chapter 7, Table 7.5 and 7.6). These findings suggest that the COLGALT2 enzyme plays a pivotal role in orchestrating the progression of chondrocyte-specific characteristics and in regulating the biosynthesis of essential ECM constituents crucial for the structural integrity of cartilage.

A comparative analysis of proteomic profiles between undifferentiated COLGALT2 KO cells and fully differentiated ones revealed that the affected proteins primarily encompass various collagen types (I, III, V, VI, VIII, XI, XII, XIV, XVI) and other ECM constituents like decorin, aggrecan, lumican, and biglycan (Chapter 7, Figure 7.15 and Table 7.7). Notably, during the differentiation of WT cells, the expression of metalloproteases and their inhibitors decreases from Day 0 to Day 21 as the cells transition toward a more chondrocyte-like phenotype. Conversely, COLGALT2 KO cells exhibit increased expression of metalloprotease inhibitors at Day 21 compared to Day 0 (Chapter 7, Table 7.7). Metalloproteases play a critical role in maintaining ECM homeostasis by degrading damaged ECM components. I hypothesised that reduced levels of the COLGALT2 enzyme led to the accumulation of collagen fibrils with abnormal glycosylation patterns in the ECM. These fibrils would require continuous degradation, thus necessitating elevated levels of metalloproteases. Accumulation of collagens with aberrant glycosylation patterns in the ER and the ECM of cells has been reported following COLGALT1 depletion³⁶, corroborating my hypothesis.

These findings provided valuable insights into the role of the COLGALT2 enzyme in chondrocyte biology and collagen biochemistry. Although these results did not definitively identify the specific targets of COLGALT2, the differential expression analyses of the secretome highlight collagens type II, VI, XI, XII, and XVI as potential targets (Chapter 7, Figure 7.16 and Table 7.8). These collagens play crucial roles in cartilage biology, as elaborated in Chapter 1, where collagens II, IX, and XII form heteropolymers constituting a significant portion of the ECM collagen network³⁷. Collagen type VI has been implicated in regulating the pericellular matrix in cartilage³⁸⁻⁴⁰. Collagens XI, XII, and XVI primarily ensure the integrity and stability of the collagen fibril network⁴¹⁻⁴³. Aberrant glycosylation of any or all of these collagens could disrupt the composition and structural integrity of the ECM secreted by

chondrocytes, contributing to the observed weakened chondrocyte phenotype in COLGALT2 KO differentiated cells.

Considering the crucial roles played by individual collagens and their post-translational modifications in cartilage, further comprehensive investigations into their functions are warranted. Such investigations would contribute to a better understanding of the pathophysiology of collagen-rich tissues such as skin, bone, cartilage, and blood vessels. Our study was unable to evaluate the effects of complete glycosylation loss on chondrocyte biology due to incomplete depletion of the universal galactosyltransferase, COLGALT1, in the ASC52teloSox9 cells (Chapter 7, Figure 7.3). Despite the reduced protein and gene expression of COLGALT1 prior to differentiation in the ASC52teloSox9 (Chapter 7, Figure 7.2A), the extended culture expansion required to obtain sufficient cell numbers may have provided any unedited cells with a survival and growth advantage over COLGALT1 knock-out cells, resulting in the cells subjected to differentiation no longer having reduced levels of COLGALT1. The depletion of this protein has been shown to be embryonically lethal in mice⁴⁴ so a positive selection of the cells with intact COLGALT1 function is plausible.

In this thesis, my investigation primarily focused on assessing the impact of common genetic variants associated with OA on the regulation of COLGALT2 expression in tissues relevant to joint health across various stages of development and aging. Additionally, I delved into the role of the COLGALT2 enzyme within chondrocyte biology and collagen biochemistry by employing an *in vitro* chondrogenesis model. The outcomes of these different types of investigations underscore the significance of COLGALT2 as a promising target for the development of OA therapeutics. This is substantiated by the confirmation that both genetic and epigenetic factors differentially regulate COLGALT2 gene expression in patient samples. Furthermore, proteomics data provide compelling evidence implicating the enzyme encoded by the *COLGALT2* gene as a pivotal contributor to chondrocyte biology.

Furthermore, my research into the regulation of COLGALT2 in developmental cartilage samples has provided valuable insights into the origins of OA. These findings suggest that genetic and epigenetic mechanisms operating on OA-risk loci during early development may predispose individuals to OA later in life. Consequently, a focus on understanding the developmental stages of disease onset holds promise for the development of more precise and effective therapeutic strategies for managing OA.

In summary, while GWAS have proven invaluable in uncovering the genetic foundations of complex traits and diseases, the interpretation of GWAS signals presents a formidable challenge. It is crucial to implement a robust program of functional validation, encompassing diverse tissues and methodologies, in order to unravel the intricacies of genetic regulation. This effort is pivotal for advancing personalized medicine and formulating targeted therapeutic strategies for complex diseases including OA. With the prevalence and the burden of the disease predicted to only increase in the coming decades, the need for comprehensive functional investigations into GWAS signals which identify causal pathways as potential therapeutic targets is also on the rise.

References

1. [review] Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *Am. J. Hum. Genet.* **102**, 717–730 (2018).
2. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**, 1190–1195 (2012).
3. Aubourg, G., Rice, S. J., Bruce-Wootton, P. & Loughlin, J. Genetics of osteoarthritis. *Osteoarthr. Cartil.* **30**, 636–649 (2022).
4. [review] Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
5. [review] 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).
6. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
7. [review] Lichou, F. & Trynka, G. Functional studies of GWAS variants are gaining momentum. *Nat. Commun.* **11**, 6283 (2020).
8. Nicolae, D. L. *et al.* Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
9. [review] Li, S., Peng, Y. & Panchenko, A. R. DNA methylation: Precise modulation of chromatin structure and dynamics. *Curr. Opin. Struct. Biol.* **75**, 102430 (2022).
10. Liu, Y. *et al.* Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis. *Sci. Rep.* **8**, 15499 (2018).
11. Arensbergen, J. van *et al.* High-throughput identification of human SNPs affecting regulatory element activity. *Nat. Genet.* **51**, 1160–1169 (2019).
12. Alsheikh, A. J. *et al.* The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. *BMC Méd. Genom.* **15**, 74 (2022).

13. Horlbeck, M. A. *et al.* Mapping the Genetic Landscape of Human Cells using machine learning models. *Cell* **174**, 953–967.e22 (2018).
14. [review] Kumar, K., Cowley, M. & Davis, R. Next-Generation Sequencing and Emerging Technologies. *Semin. Thromb. Hemost.* **45**, 661–673 (2019).
15. [review] Wang, M. H., Cordell, H. J. & Steen, K. V. Statistical methods for genome-wide association studies. *Semin. Cancer Biol.* **55**, 53–60 (2019).
16. [review] Przybyla, L. & Gilbert, L. A. A new era in functional genomics screens. *Nat. Rev. Genet.* **23**, 89–103 (2022).
17. 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. *Osteoarthr. Cartil.* **27**, 1545–1556 (2019).
18. Almeida, R. C. de *et al.* Allelic expression imbalance in articular cartilage and subchondral bone refined genome-wide association signals in osteoarthritis. *Rheumatol. (Oxf., Engl.)* **62**, 1669–1676 (2022).
19. Boer, C. G. *et al.* Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell* **184**, 4784–4818.e17 (2021).
20. Kribelbauer, J. F., Lu, X.-J., Rohs, R., Mann, R. S. & Bussemaker, H. J. Toward a Mechanistic Understanding of DNA Methylation Readout by Transcription Factors. *J. Mol. Biol.* **432**, 1801–1815 (2020).
21. Hashimoto, H., Zhang, X., Vertino, P. M. & Cheng, X. The Mechanisms of Generation, Recognition, and Erasure of DNA 5-Methylcytosine and Thymine Oxidations*. *J. Biol. Chem.* **290**, 20723–20733 (2015).
22. Rao, S. *et al.* Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein–DNA binding. *Epigenetics Chromatin* **11**, 6 (2018).
23. Kribelbauer, J. F. *et al.* Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes. *Cell Rep.* **19**, 2383–2395 (2017).
24. Machado, A. C. D. *et al.* Meta-analysis of cytosine methylation affects upon protein–DNA binding. *Brief. Funct. Genom.* **14**, 61–73 (2015).
25. Yin, Y. *et al.* Chromosome conformation capture reveal the impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, (2017).
26. [review] Westra, H.-J. & Franke, L. From genome to function by studying eQTLs. *Biochim. Biophys. Acta (BBA) - Mol. Basis Dis.* **1842**, 1896–1902 (2014).
27. Abell, N. S. *et al.* Multiple causal variants underlie genetic associations in humans. *Science* **375**, 1247–1254 (2022).
28. Arvanitis, M., Tayeb, K., Strober, B. J. & Battle, A. Tissue specificity of genetic regulation of gene expression in the presence of allelic heterogeneity across multiple tissue types. *Am. J. Hum. Genet.* **109**, 223–239 (2022).
29. Kreitmaier, P. *et al.* An epigenome-wide view of osteoarthritis in primary tissues. *Am. J. Hum. Genet.* **109**, 1255–1271 (2022).

30. Singh, P., Marcu, K. B., Goldring, M. B. & Otero, M. Phenotypic instability of chondrocytes in osteoarthritis: on a path to hypertrophy. *Ann. N. York Acad. Sci.* **1442**, 17–34 (2019).
31. Saito, T. *et al.* Transcriptional regulation of endochondral ossification by HIF-2 α during skeletal growth and osteoarthritis development. *Nat. Med.* **16**, 678–686 (2010).
32. [review] Pitsillides, A. A. & Beier, F. Cartilage biology in osteoarthritis—lessons from developmental biology. *Nat. Rev. Rheumatol.* **7**, 654–663 (2011).
33. Wan, J. *et al.* Methylated cis-regulatory elements mediate KLF4-dependent gene transactivation and cell migration. *eLife* **6**, e20068 (2017).
34. Kaluscha, S. *et al.* Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat. Genet.* **54**, 1895–1906 (2022).
35. [review] Héberlé, É. & Bardet, A. F. Sensitivity of transcription factors to DNA methylation. *Essays Biochem.* **63**, 727–741 (2019).
36. Geister, K. A. *et al.* Loss of function of Colgalt1 disrupts collagen post-translational modification and causes musculoskeletal defects. *Dis. Model. Mech.* **12**, dmm037176 (2019).
37. Eyre, D. R., Wu, J.-J., Fernandes, R. J., Pietka, T. A. & Weis, M. A. Recent developments in cartilage research: matrix biology of the collagen II/IX/XI heterofibril network. *Biochem. Soc. Trans.* **30**, 893 (2001).
38. Twomey, J., Thakore, P., Hartman, D., Myers, E. & Hsieh, A. Roles of type VI collagen and decorin in human mesenchymal stem cell biophysics during chondrogenic differentiation. *Eur. Cells Mater.* **27**, 237–250 (2014).
39. Bidanset, D. J. *et al.* Binding of the proteoglycan decorin to collagen type VI. *J. Biol. Chem.* **267**, 5250–6 (1992).
40. Guilak, F. *et al.* The Pericellular Matrix as a Transducer of Biomechanical and Biochemical Signals in Articular Cartilage. *Ann. N. York Acad. Sci.* **1068**, 498–512 (2006).
41. Kassner, A. *et al.* Discrete integration of collagen XVI into tissue-specific collagen fibrils or beaded microfibrils. *Matrix Biol.* **22**, 131–143 (2003).
42. [review] Luo, Y. *et al.* The minor collagens in articular cartilage. *Protein Cell* **8**, 560–572 (2017).
43. Taylor, D. W. *et al.* Collagen Type XII and Versican Are Present in the Early Stages of Cartilage Tissue Formation by Both Redifferentating Passaged and Primary Chondrocytes. *Tissue Eng. Part A* **21**, 683–693 (2015).
44. He, L. *et al.* Down-regulation of GLT25D1 inhibited collagen secretion and involved in liver fibrogenesis. *Gene* **729**, 144233 (2020).

Appendix A. Primer sequences

Experiment	Locus	Target	Target genomic position (hg19)	Forward (5'→3')	Reverse (5'→3')	Sequencing (5'→3')
Genotyping and ABL	rs11583641	rs11583641	chr1:183909995	TGGG CAGTG AACTGAAG AATTAG	[Bnt]CGGGTGCTGCTACATACAA	TGGCTTAAAGTGATCTTAC
		rs12083450	chr1:183880289	[Bnt]GTATATCCCTCCGCAAGATT	GQCTCACTCCAACTTTGCTATTTA	ACTCTCAACCGGTGAC
		rs7523001	chr1:183887910	CTTGTCTCTTTGGAACTTTGT A	[Bnt]TGGAAATAACTTTGCTAGTGG	GCATACAACTGCTTTACG
		rs10752924	chr1:183887032	[Bnt]AGTGATGGGATTTGCTATTAAAG	TTCAAGGCTTATCTCTCTCTAC	CGGCTCCAGTCTACC
	rs1046934	rs734657	chr1:183912219	ACCTTAGATTCTCTGGGCTCA	GAAGCAGA AAGTTCCCAAGCA	N/A
		rs1046934	chr1:184029529	TTAGATATAGGAGATGCCACCCA	[Bnt]CAGGTAAACCAAGACGCTACATA	AGGAGATGCCACCCA
		rs114661926	chr1:184006549	GAGAGGAGCAGTAGCGACCAAG	[Bnt]AGCAAGGGGTCGAGGG	TGAGGGCGGGGCGGG
		rs2274432	chr1:184029445	[Bnt]GGGAAGCGGAACCCAAG	CCCTGAGGACGGCTGGAT	TGTTGCGGCTTGG
DNA quantification	rs11583641	CpG 1	chr1:183911362	[Bnt]GGTTTTTAAAGTTTGGATTATAG	TCCTCAATCCAAAACTAAATTTTCACTA	AGTGTGAGGATTATAGG
		CpG 2	chr1:183911374			
		CpG 3	chr1:183911971	[Bnt]AGGGGATATTTTGTATTAATTG	ACAACITCACTTAAACCTCAATAT	ATTCATAAATCATTTTCTCAAAC
		CpG 4	chr1:183912000			
		CpG 5	chr1:183912015	AGGGGATATTTTGTATTAATTG	[Bnt]ACAACTCCTAAACCTCAATATCA	GGTATTTTGTGGAGAAATG
		CpG 6	chr1:183912021			
		CpG 7	chr1:183912155		[Bnt]CCCTTAAATCACTATTTCTCTATA	AGGTTGTGAGGTTTAAATTATA
		CpG 8	chr1:183912188	GTA G T T T T T A A G T T T G T G A G G T T T T A A T T A	[Bnt]TCTAAATCACTAAATAACCTAACT	GGTATAGGTGTTTATATATAGTG
		CpG 9	chr1:183912306	AGGGATTTGTGAGATGTTAAATTAA	[Bnt]CCCAACTATATCAATAAACAATTCATC	CCCAACTATATCAATAAACAATTCATC
		CpG 10	chr1:183912398	TGGTTTATTTTGGAGGATAAGTAT	ACCAATTTATCAACTCTAAATACTCTAT	TTTGGTTGATTTTATATTTTAT
		CpG 11	chr1:183912892	[Bnt]TTTTAAAGAGGATTTTGGTTGATTT		
		CpG 12	chr1:183912920			
	rs1046934	cg15204595	chr1:184003184	TGTTGTTGGTTTGTAG AATGGATAT	[Bnt]AACACTCTATAATAAATAATCCCTCAAT	ATGAGAAATTAGAGGTAAAT
		cg21608956	chr1:184211934	TTTTGTATTTATTTTGTGTTAAATTTTG	[Bnt]CAATACCTCTATACATAATAAATCCCTCT	TGTGGTTAAATTTTGGTTT
Reporter gene analysis	rs11583641	region containing CpGs 3-10	N/A	CCTAGGTCGAAGACAACTGTTCTTCTCTCT	ACTAGTCTGTCCCTGCTTCTCCAC	N/A
	rs1046934	region containing cg15204595	N/A	ATGCACTAAAGTTCTGCGCTCA	ACTAGTGAAATCTCTTGGAGGTGCTGACT	N/A
		region containing cg21608956	N/A	CCTAGGCTGCTGTTTACACCAAGGCT	ACTAGTGAGTGGCTCATGTG TGGGG	N/A
Genome editing (confirmation of deletion)	rs11583641	region containing CpGs 3-10	N/A	CAACACTCCCTCAATGATTC	CTCTGCTCTCTCTCTCTCT	N/A
Genome editing (COL6ALT2 splicing)	rs11583641	intron 10 of COL6ALT2	N/A	GCAAGGAACCTACCTACGTC	AACATGACCAAGCTACGCT	N/A
Epigenome editing (testing gRNA targeting)	rs11583641	region containing CpGs 3-10	N/A	ACAACCTGTTCTTCTCTCTGCA	CTGTCCCTGCTTCTCCAC	N/A
COL6ALT1 and COL6ALT2 KO (testing gRNA targeting)	N/A	COL6ALT1 Exon 2	N/A	TCCTCTGAGTACCCCGAAG	CAGACCTGCTCTGTGATCAT	N/A
		COL6ALT2 Exon 3	N/A	ACACCTGAGATGACAGTTGTC	GGACAGAGCCCTGATACATG	N/A
cDNA synthesis	N/A	RP11 intron 4	N/A	TGG AAGAGTGACACCACTT	GAAGGCCAGGAAATGACCACTC	N/A

[Bnt] = biotinylated. Red sequences = overhangs for downstream cloning

Appendix B. gRNA sequences

Format	Experiment	Locus	gRNA name	Sequence (5'-3')
CRISPR gRNA sequence	Epigenetic modulation	rs11583641	gRNA 1 gRNA 2 gRNA 3 gRNA 4 gRNA 5 gRNA 6	CCACCGAACTACATGTGACC TGTCCACGTGACTTGATATAA GCCTGTGAGGCCCTTAACCAC GCGGAGCCATCCATTGAGAA GCATTAGTTGAGCCATTGCT AGGGAACAAGTATTCACCT
		rs1046934	gRNA-cg15204595 gRNA-cg21606956	TGGAATGTTGCATAGAGCTT GTTATTGTGGTCAGATTCAG
Complementary ssDNA oligonucleotides	Epigenetic modulation	rs11583641	gRNA 1 FWD gRNA 1 REV gRNA 2 FWD gRNA 2 REV gRNA 3 FWD gRNA 3 REV gRNA 4 FWD gRNA 4 REV gRNA 5 FWD gRNA 5 REV gRNA 6 FWD gRNA 6 REV	CACCGCCACCGAACTACATGTGACC AAACGGTCACATGTAGTTCGGTGGC CACCGTGCCACGTGACTTGATATAA AAACTTATACAAGTCACGTGGACAC CACCGCCTGTGAGGCCCTTAACCAC AAACGTGGTTAAGGCCTCACAGGCC CACCGCGGAGCCATCCATTGAGAA AAACTTCTCAATGGATGGCTCCGCC CACCGGCATTAGTTGAGCCATTGCT AAACAGCAATGGCTCAACTAATGCC CACCGAGGGAACAAGTATTCACCT AAACAGGTGGAATACTTGTCCCTC
			gRNA-cg15204595 FWD gRNA-cg15204595 REV gRNA-cg21606956 FWD gRNA-cg21606956 REV	CACCGTGAATGTTGCATAGAGCTT AAACAAGCTCTATGCAACATTCCAC CACCGGTTATTGTGGTCAGATTCAG AAACCTGAATCTGACCACAATAACC
	Genome editing	rs11583641	gRNA7 FWD gRNA7 REV gRNA 8 FWD gRNA 8 REV	CACCGCAACACTCCCTCAATGATTCCA AAACTGGAATCATTGAGGGAGTGTTC CACCGCCTCTGCCTCCTCTTCTC AAACGAGAAAGGAGGAGGCAGAGGC
			gRNA-T1 FWD gRNA-T1 REV gRNA-T2 FWD gRNA-T2 REV	CACCGGGTGGCTACGGACCACAACA AAACTGTTGTGGTCCGTAGCCACCC CACCGCCATGTGATGAACTACGAC AAACGTCGTAGTTTCATCACATGGC
	COLGALT1 and COLGALT2 KO	N/A		

ssDNA - single-stranded DNA; gRNA – guide RNA; Red sequences = overhangs for downstream cloning

Appendix C. mQTL analysis using Illumina 450K array data

SNP	CpG	pvalue	FDR	Slope	cpg.chr	cpg.pos	cpg.strand	cpg.kland	cpg.Group	cpg.GeneSymbol
rs1046934	cg21606956	0,002388789	0,138549819	0,158917452	chr1	184211934	-	OpenSea		
rs1046934	cg15204595	0,005077494	0,147247345	0,255571328	chr1	184003184	+	N_Shelf	Body	GLT25D2
rs1046934	cg01436608	0,041280749	0,627222855	0,178202058	chr1	184005534	-	Island	Body	GLT25D2
rs1046934	cg22340762	0,051489372	0,627222855	0,17776850	chr1	183823538	+	OpenSea	Body	RGL1
rs1046934	cg17002156	0,054070935	0,627222855	0,232924321	chr1	183995878	+	OpenSea	Body	GLT25D2
rs1046934	cg21404028	0,108366350	0,848349035	0,092775988	chr1	184020687	-	N_Shore	TSS200;TSS200;TSS200	TSEN15;TSEN15;TSEN15
rs1046934	cg12222949	0,113011900	0,848349035	0,160416090	chr1	184005360	+	Island	Body	GLT25D2
rs1046934	cg15111486	0,133341488	0,848349035	0,10419286	chr1	184195211	-	OpenSea		
rs1046934	cg05777716	0,144389698	0,848349035	0,08006347	chr1	184006285	-	Island	1stExon	GLT25D2
rs1046934	cg13951632	0,184928856	0,848349035	0,120100660	chr1	183897368	-	OpenSea	3'UTR	RGL1
rs1046934	cg15569801	0,187897515	0,848349035	0,066928549	chr1	184197096	-	OpenSea		
rs1046934	cg27177158	0,194193974	0,848349035	0,100123680	chr1	184121585	-	OpenSea		
rs1046934	cg02314308	0,194957334	0,848349035	0,110305009	chr1	184006972	-	S_Shore	TSS200	GLT25D2
rs1046934	cg27444414	0,230101786	0,848349035	0,092115213	chr1	184122874	-	OpenSea		
rs1046934	cg10312498	0,240591976	0,848349035	0,10911600	chr1	183957982	-	OpenSea	Body	GLT25D2
rs1046934	cg16655084	0,243080607	0,848349035	0,081155104	chr1	184005896	+	Island	Body	GLT25D2
rs1046934	cg12136256	0,248654027	0,848349035	0,13617857	chr1	183952796	+	OpenSea	Body	GLT25D2
rs1046934	cg14773259	0,292969985	0,928585405	0,085088592	chr1	184021034	-	Island	Body;Body;Body	TSEN15;TSEN15;TSEN15
rs1046934	cg22400420	0,304191770	0,928585405	0,088312671	chr1	183851242	-	OpenSea	Body	RGL1
rs1046934	cg18131582	0,363285827	0,965141685	0,104087734	chr1	183912305	-	OpenSea	Body	GLT25D2
rs1046934	cg00464025	0,381248494	0,965141685	0,04063193	chr1	183907542	-	OpenSea	3'UTR	GLT25D2
rs1046934	cg19101018	0,385337955	0,965141685	0,063402471	chr1	183925532	-	OpenSea	Body	GLT25D2
rs1046934	cg21417627	0,447531032	0,965141685	0,06737686	chr1	184008235	-	S_Shore	TSS1500	GLT25D2
rs1046934	cg14313310	0,485865458	0,965141685	0,055906006	chr1	184006816	-	S_Shore	5'UTR;1stExon	GLT25D2;GLT25D2
rs1046934	cg09940311	0,496887744	0,965141685	0,11007747	chr1	183901666	-	OpenSea		
rs1046934	cg06849459	0,514610056	0,965141685	0,04667025	chr1	183891439	-	OpenSea	Body	RGL1
rs1046934	cg01068808	0,530912365	0,965141685	0,07518952	chr1	184007101	-	S_Shore	TSS1500	GLT25D2
rs1046934	cg18073970	0,542309516	0,965141685	0,05760377	chr1	184006765	+	Island	5'UTR;1stExon	GLT25D2;GLT25D2
rs1046934	cg09684066	0,558049452	0,965141685	0,031355577	chr1	183891565	+	OpenSea	Body	RGL1
rs1046934	cg25605307	0,560179746	0,965141685	0,049859901	chr1	183965312	-	OpenSea	Body	GLT25D2
rs1046934	cg11678039	0,580064206	0,965141685	0,041394927	chr1	183841259	-	OpenSea	Body	RGL1
rs1046934	cg16716196	0,595667485	0,965141685	0,049572937	chr1	184195068	-	OpenSea		
rs1046934	cg03467001	0,604461198	0,965141685	0,06345798	chr1	183857649	-	OpenSea	Body	RGL1
rs1046934	cg03943177	0,611155154	0,965141685	0,052335755	chr1	184021360	+	S_Shore	Body;Body;Body	TSEN15;TSEN15;TSEN15
rs1046934	cg06915270	0,642327728	0,965141685	0,027943554	chr1	183891360	+	OpenSea	Body	RGL1
rs1046934	cg17357479	0,658357979	0,965141685	0,03704078	chr1	184005717	-	Island	Body	GLT25D2
rs1046934	cg10807943	0,689234253	0,965141685	0,02057012	chr1	184021131	+	S_Shore	Body;Body;Body	TSEN15;TSEN15;TSEN15
rs1046934	cg12702671	0,711635426	0,965141685	0,031402346	chr1	183885210	-	OpenSea	Body	RGL1
rs1046934	cg20942099	0,723326200	0,965141685	0,02379057	chr1	184007803	+	S_Shore	TSS1500	GLT25D2
rs1046934	cg00143655	0,754989126	0,965141685	0,03289700	chr1	184005063	-	N_Shore	Body	GLT25D2
rs1046934	cg12792264	0,765268544	0,965141685	0,03532027	chr1	183971636	+	OpenSea	Body	GLT25D2
rs1046934	cg00044463	0,766522433	0,965141685	0,023787640	chr1	184006933	+	S_Shore	TSS200	GLT25D2
rs1046934	cg24721630	0,769637771	0,965141685	0,027460413	chr1	184006216	+	Island	Body	GLT25D2
rs1046934	cg02890642	0,781538860	0,965141685	0,02156065	chr1	184017798	-	N_Shelf		
rs1046934	cg27291258	0,784525638	0,965141685	0,04348961	chr1	184173289	+	OpenSea		
rs1046934	cg12758231	0,787779993	0,965141685	0,019910918	chr1	184020711	+	N_Shore	TSS200;TSS200;TSS200	TSEN15;TSEN15;TSEN15
rs1046934	cg15359501	0,792639164	0,965141685	0,015547071	chr1	184006496	-	Island	5'UTR;1stExon	GLT25D2;GLT25D2
rs1046934	cg26429856	0,798737946	0,965141685	0,013574528	chr1	184005990	+	Island	Body	GLT25D2
rs1046934	cg18189236	0,821703244	0,966291552	0,02569492	chr1	184009497	-	S_Shelf		
rs1046934	cg05816006	0,839369641	0,966291552	0,018867471	chr1	184042579	+	OpenSea	3'UTR;3'UTR;Body	TSEN15;TSEN15;TSEN15
rs1046934	cg11387897	0,849670158	0,966291552	0,012868174	chr1	183914188	-	OpenSea	Body	GLT25D2
rs1046934	cg12631766	0,887366793	0,989755269	0,01550591	chr1	184025029	-	S_Shelf	Body;Body;Body	TSEN15;TSEN15;TSEN15
rs1046934	cg11346030	0,915841776	0,990278020	0,009511507	chr1	184090950	-	OpenSea		
rs1046934	cg06379531	0,921982984	0,990278020	0,005806731	chr1	184020410	-	N_Shore	TSS1500;TSS1500;TSS1500	TSEN15;TSEN15;TSEN15
rs1046934	cg23040305	0,969651660	0,995418040	0,004039352	chr1	183846243	-	OpenSea	Body	RGL1
rs1046934	cg15521745	0,983100264	0,995418040	0,001355846	chr1	184133404	-	OpenSea		
rs1046934	ch.1.1823595	0,987854651	0,995418040	0,001960897	chr1	184092903	+	OpenSea		
rs1046934	cg20271396	0,995418040	0,995418040	0,00047325	chr1	184020713	+	N_Shore	TSS200;TSS200;TSS200	TSEN15;TSEN15;TSEN15

Appendix D. A list of all proteins detected in the cell pellets and the secretome of the ASC52teloSOX9 differentiation and COLGALT2 KO experiments

Appendix E. A list of all differentially expressed proteins detected in the ASC52teloSOX9 differentiation and COLGALT2 KO experiments

Appendix F. List of proteins by cluster from the clustered heatmap of the top 3000 differentially expressed proteins in the *COLGALT2 KO D0 vs D21*, *WT D0 vs WT D21*, *WT D0 vs COLGALT2 KO D0*, and *WT D21 vs COLGALT2 KO D21*.

Appendix G. A list of all differentially expressed proteins in the secretome of WT D21 and COLGALT2 KD D21 cells