
Article:
Kehayova, Y.S., Wilkinson, J.M. orcid.org/0000-0001-5577-3674, Rice, S.J. et al. (1 more author) (2022) Independent osteoarthritis risk-conferring alleles mediate the same epigenetic and transcriptional effect on a shared target gene, COLGALT2. Arthritis & Rheumatology. ISSN 2326-5191

https://doi.org/10.1002/art.42427

Arthritis & Rheumatology

AN OFFICIAL JOURNAL OF THE AMERICAN COLLEGE OF RHEUMATOLOGY

Independent osteoarthritis risk-conferring alleles mediate the same epigenetic and transcriptional effect on a shared target gene, COLGALT2

Journal:	*Arthritis & Rheumatology*
Manuscript ID	Draft
Wiley - Manuscript type:	Full Length
Date Submitted by the Author:	n/a
Complete List of Authors:	Kehayova, Yulia; Newcastle University, Biosciences Institute Wilkinson, J. Mark; University of Sheffield, Department of Human Metabolism; Rice, Sarah; Newcastle University, Biosciences Institute Loughlin, John; Newcastle University, ;
Keywords:	Osteoarthritis, COLGALT2, Epigenetics, Genetics, Risk Factors
<B>Disease Category: Please select the category from the list below that best describes the content of your manuscript.:	Osteoarthritis

Independent osteoarthritis risk-conferring alleles mediate the same epigenetic and transcriptional effect on a shared target gene, *COLGALT2*

Yulia S. Kehayova[1], J. Mark Wilkinson[2], Sarah J. Rice[1*], John Loughlin[1*]

[1]Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE1 3BZ, UK; [2]Department of Oncology and Metabolism, University of Sheffield, Sheffield, S10 2RX, UK.

*Address correspondence to:

John Loughlin, PhD, Biosciences Institute, International Centre for Life, Newcastle University, Newcastle upon Tyne, NE1 3BZ, UK.

E-mail: john.loughlin@ncl.ac.uk

Telephone: (+44) 191 241 8988

Sarah J. Rice, PhD, Biosciences Institute, International Centre for Life, Newcastle University, Newcastle upon Tyne, NE1 3BZ, UK.

E-mail: sarah.rice@ncl.ac.uk

Telephone: (+44) 191 241 8850

Objective. Over 100 DNA variants have been associated with osteoarthritis (OA), including rs1046934, located within a linkage disequilibrium block encompassing part of *COLGALT2* and *TSEN15*. Here, we used human foetal cartilage, cartilage from arthroplasty patients, and a chondrocyte cell model to determine the target gene(s) at the locus and the mechanism of action.

Methods. Cartilage array data (n=87) were used to determine if rs1046934 genotype correlated with differential DNA methylation at proximal CpGs. Results were replicated in arthroplasty (n=132) and foetal (n=77) cartilage DNA using pyrosequencing. Allelic expression imbalance (AEI) measured the effect of genotype upon *COLGALT2* and *TSEN15* expression. Reporter gene assays and epigenetic editing determined the functional role of regions harbouring differentially methylated CpGs. *In silico* analyses complemented these experiments.

Results. Three differentially methylated CpGs residing within regulatory regions were detected, two of which, cg15204595 and cg21606956, replicated. AEI was detected for *COLGALT2* and *TSEN15*, with correlations between expression and methylation for *COLGALT2*. Reporter assays confirmed that the CpGs are in chondrocyte enhancers with epigenetic editing directly linking methylation with *COLGALT2* expression.

Conclusion. *COLGALT2* is a target of this OA locus. We previously characterised another OA locus, marked by rs11583641, that independently targets *COLGALT2*. rs1046934, like rs11583641, mediates its effect by modulating the expression of *COLGALT2* via methylation changes to CpGs located in enhancers. The SNPs, CpGs and enhancers are distinct between the loci but the effect on *COLGALT2* is the same. *COLGALT2* is the target of independent OA risk loci sharing a common mechanism of action.

INTRODUCTION

Genome-wide association studies (GWAS) have identified over 100 DNA variants that associate with osteoarthritis (OA) risk (reviewed in 1). The samples sizes used are impressive, with recent investigations analysing the genomes of hundreds of thousands of individuals (2-4). Biological comprehension of GWAS signals requires elucidation of the molecular effects of the risk-conferring alleles on their target genes (5-8). Since the individual contribution of most variants to disease risk is small, assessing these effects is challenging (5-8). Furthermore, determining the causal variants underpinning an association signal is not straightforward, as genetic variants commonly occur within linkage disequilibrium (LD) blocks (5-8). Despite these difficulties, the application of statistical fine-mapping combined with laboratory-based studies of primary cells and cellular models has started to generate functional insight into the molecular basis of OA genetic risk (3,4,9-15).

As with other polygenic diseases, most OA associated variants reside within the non-coding genome and contribute to disease by altering expression of genes within the same topologically associated domain (TAD), thereby acting as expression-quantitative trait loci (eQTLs) (1). We have reported that DNA methylation at CpG dinucleotides also often correlates with genotype at OA associated variants, forming methylation-QTLs (mQTLs), and that this epigenetic effect may act as an intermediate between risk allele and gene expression change (16-21). One recent example was our investigation of the OA association signal marked by single nucleotide polymorphism (SNP) rs11583641 (22). This common variant resides within the 3′ untranslated region (3′UTR) of *COLGALT2*, a gene that encodes a galactosyltransferase that post-translationally modifies collagen (22). Using cartilage from patients and a chondrocyte cell line, we discovered that the OA risk allele of rs11583641 correlated with lower methylation levels of CpGs within an intronic enhancer of *COLGALT2*

and that this reduced methylation increased enhancer activity and expression of the gene (22). Increased glycosylation of collagen molecules reduces the amount of inter-molecule cross-linking, leading to collagen fibrils with reduced diameters and lower tensile strength (23). We concluded that increased *COLGALT2* expression, and therefore increased galactosyltransferase activity, is detrimental to cartilage health via impacts on collagen biosynthesis (22). We subsequently reported that for some OA risk loci, including rs11583641, genotype correlations with gene expression and CpG methylation observed in arthroplasty cartilage are also observed in foetal cartilage (24). This implies that OA genetic risk may be programmed in during development.

In the most recent OA GWAS, a second association signal was reported that maps close to *COLGALT2* (4). Three SNPs were highlighted; rs12047271 and rs1327123, which reside in-between *COLGALT2* and *TSEN15*, and rs1046934, which resides within *TSEN15*. The TSEN15 protein is a subunit of tRNA splicing endonuclease. The splicing of introns from pre-tRNAs is performed by a heterotetrameric endonuclease comprised of TSEN15, TSEN34, TSEN2, and TSEN54 (25,26). TSEN15-34 are the structural subunits of the endonuclease, whilst TSEN2-54 form the catalytic domains (26). TSEN15 adopts a compact α-α-β-β-β-β-α-β-β fold, preceded by a disordered N-terminal region, which has not been structurally resolved (25,26).

rs12047271, rs1327123 and rs1046934 are in very high LD with each other (r^2 values ≥ 0.95 in European ancestry cohorts) and are part of an LD block containing 21 SNPs (r^2 values ≥ 0.8) spanning a 30kb region. Furthermore, they are in near perfect linkage equilibrium with rs11583641 (r^2 values of zero, D' values ≤ 0.08). This second *COLGALT2* signal, which we will henceforth refer to as the rs1046934 locus, is therefore genetically independent of the first *COLGALT2* signal. In this study, we set out to investigate the gene targets of this new locus using a range of molecular, cellular and *in silico* techniques.

PATIENTS AND METHODS

Protein modelling. TSEN15 crystal structures were downloaded from the Protein Data Bank (Supplementary Table 1) and visualised in complex with TSEN34 (6Z9U) and as a monomeric structure (2GW6) using PyMOL Molecular Graphics System, version 2.1.1 (Schrödinger; https://pymol.org). The PyMOL Mutagenesis Wizard was used to perform *in silico* mutagenesis to model the missense variant Gln59His introduced by rs1046934. The gnomAD database (27) (Supplementary Table 1) was used to predict the effect of this variant and of Gly19Asp, introduced by rs2274432, on TSEN15 function.

Cartilage samples and ethics approval. Cartilage samples were obtained from 132 patients undergoing joint arthroplasty at the Newcastle upon Tyne NHS Foundation Trust hospitals for primary hip OA (n = 43), primary knee OA (n = 63), or for a neck-of-femur (NOF) fracture (n = 26). Ethical approval was granted by the NHS Health Research Authority with each donor providing written consent (REC reference number 19/LO/0389; patient details in Supplementary Table 2). Samples were processed, and the nucleic acids extracted, as previously described (20-22). Seventy-seven matched foetal DNA and RNA samples (Supplementary Table 3) were provided by the Human Developmental Biology Resource (HDBR; https://hdbr.org; project 200363) (24). The nucleic acids were extracted by the HDBR from human foetal cartilage, as previously described (24).

Genotyping. Allelic quantification pyrosequencing assays were designed using PyroMark Assay Design 2.0 software (Qiagen) with oligonucleotide primers ordered from Integrated DNA Technologies (IDT). Genomic DNA encompassing the SNP of interest was PCR

amplified using the PyroMark PCR kit (Qiagen) with genotype determined using the PyroMark Q24 Advanced system (Qiagen). Supplementary Table 4 has oligonucleotide sequences.

Allelic expression imbalance. Transcript SNPs were used to investigate allelic expression imbalance. For *COLGALT2*, we used the 5'UTR SNP rs114661926, for *TSEN15*, we used the missense SNP rs2274432 (Supplementary Table 5). For both genes, patients who were compound heterozygote at rs1046934 and the respective transcript SNP were investigated. cDNA was reverse transcribed from 500ng RNA using SuperScript IV (Invitrogen). The relative ratio of risk/non-risk allele at the SNPs was quantified by pyrosequencing in DNA and cDNA, as previously described (17,20,22). Oligonucleotides were ordered from IDT (Supplementary Table 4). Patient samples were analysed in triplicate and replicate values with >5% difference were excluded. Allelic expression in cDNA was normalised to that in genomic DNA for each patient.

mQTL discovery. We used genotype and cartilage DNA methylation data that we had generated previously using the Human Omni Express array and Infinium Human-Methylation450 array (28). Both datasets were generated from 87 patients who had undergone hip or knee arthroplasty (28). We covered a 0.4Mb region, 200kb upstream and 200kb downstream of rs1046934.

mQTL replication. CpGs with nominal *P* value < 0.05 in the mQTL discovery were taken forward for replication in an independent cohort of cartilage arthroplasty samples and in foetal cartilage samples. The samples were genotyped at rs1046934 by pyrosequencing. For methylation quantification, 500ng of genomic DNA was bisulphite converted using EZ DNA

methylation kits (Zymo Research). The regions of the CpG sites were PCR amplified in bisulphite converted DNA and methylation levels were quantified using the PyroMark Q24 Platform (Qiagen). Measurements were performed in duplicate and replicate values with >5% difference were excluded. Oligonucleotide sequences were generated by PyroMark Assay Design Software (Qiagen) and ordered from IDT (Supplementary Table 4).

***In silico* analysis.** Genomic databases (Supplementary Table 1) were searched to identify regulatory functions of the regions encompassing the associated SNPs and the mQTL CpGs. We focussed on data generated in human cells of the musculoskeletal system: primary mesenchymal stem cells (MSCs), MSC derived chondrocytes, MSC derived adipocytes, adipose derived MSCs, and primary osteoblasts. To assess if the SNPs or CpGs were in open or closed chromatin, we investigated ATAC-sequencing data generated from the cartilage chondrocytes of five OA knee patients, five OA hip patients, and from six foetal knee and six foetal hip samples (24). To assess if transcription factors predicted to bind at or close to the CpGs were expressed in cartilage, we investigated RNA-sequencing data generated from the hip cartilage of ten OA and six NOF patients (29; Supplementary Table 1, GEO accession number GSE111358).

Reporter gene assay. The investigated regions surrounding cg15204595 (290bp) and cg21606956 (260bp) were cloned into the Lucia CpG-free-promoter vector (InvivoGen). The putative enhancers were amplified from pooled genomic DNA samples using oligonucleotides containing the required restriction enzyme sequences for downstream cloning (Supplementary Table 6). The PCR products were cloned into the vector as previously described (21,22). The plasmids were methylated or mock-methylated *in vitro* using *M.SssI*

(New England BioLabs). Cells from the human chondrocyte cell line Tc28a2 (30) were seeded at 5000 cells/well in a 96-well plate and transfected with 100ng pCpG-free-promoter constructs, and 10ng pGL3-promoter control vector (Promega) using Lipofectamine 2000 (Invitrogen). Cells were lysed after 24h and luminescence read using the Dual-Luciferase Reporter Assay System (Promega) and analysed as previously described (21).

Epigenetic modulation. Two guide RNAs, gRNA1 and gRNA2, targeting cg15204595 and cg21606956, respectively, were designed using the CRISPR-Cas9 guide RNA design tool (IDT). The gRNA sequences (Supplementary Table 6) were synthesised as single-stranded complementary DNA oligonucleotides (IDT) with overhangs to facilitate cloning. For methylation, oligonucleotides were annealed and ligated into pdCas9-DNMT3a-EGFP plasmid (31) (Addgene, 71666) and the catalytically inactivated control plasmid pdCas9-DNMT3a-EGFP (ANV) (31) (Addgene, 71685) as previously described (21,22). For demethylation, the pdCas9-DNMT3a-EGFP plasmids containing the two gRNAs were digested with *PvuI* and *XbaI* (New England BioLabs) and scaffold regions were subcloned into pSpdCas9-huTET1CD-T2A-mCherry plasmid (Addgene, 129027) and the catalytically inactivated control plasmid pSpdCas9-hudTET1CD-T2A-mCherry (Addgene, 129028), as previously described (31). Each construct (5µg) was nucleofected into $1x10^6$ Tc28a2 cells using the 4D-Nucleofector kit (Lonza), with successful transfection confirmed after 24h by GFP (for DNMT3a plasmids) or mCherry (for TET1 plasmids) visualisation (Zeiss AxioVision).

Cells were harvested 72h after transfection. Nucleic acids were extracted using a DNA/RNA Purification Kit (Norgen Bioteck Corp). DNA methylation levels at cg15204595 and cg21606956 were measured using pyrosequencing. RNA (500ng) was reverse transcribed using SuperScript IV Reverse Transcriptase (Invitrogen) and gene expression measured by

reverse transcription quantitative PCR (RT-qPCR) using a Quant Studio 3 (Applied Biosystems). The expression of *COLGALT2* and *TSEN15*, normalised to that of housekeeping genes *18S, GAPDH* and *HPRT1*, was calculated using the $2^{-\Delta ct}$ method (32). TaqMan assays were purchased from IDT (Supplementary Table 7).

Statistical analysis. Wilcoxon matched pairs signed rank test was used to calculate *P* values in AEI analysis. For graphical representations of DNA methylation data, methylation status was plotted in the form of β-values, ranging from 0 (no methylation) to 1 (100% methylation). For statistical analysis of methylation data, β-values were converted to M-values (33). In mQTL analysis, linear regression was used to assess the relationship between CpG methylation and genotype (0, 1 or 2 copies of the minor allele) at rs1046934. For mQTL discovery, these calculations were performed using the Matrix eQTL package (34) in R, with age, sex and joint site (hip or knee) used as covariates. Correlations between AEI and DNA methylation were also determined using linear regression. Mann-Whitney U test was used to calculate *P* values when comparing methylation levels irrespective of genotype. For Lucia reporter gene assays, *P* values were calculated by paired and unpaired t-tests. Paired t-tests were used to calculate *P* values for changes in gene expression following DNMT3a or TET1 epigenetic modulation. Unless stated otherwise, statistical tests were performed in GraphPad Prism.

RESULTS

Missense variants not predicted to affect TSEN15 protein. The rs1046934 locus encompasses transcript SNPs that introduce amino acid (missense) substitutions into TSEN15;

rs1046934 itself (A>C, p.Gln59His), and rs2274432 (G>A, p.Gly19Asp). These SNPs are in perfect LD. The Gln59 residue falls within the α2 helix of TSEN15 (Figure 1A and 1B). *In silico* mutagenesis of the residue predicts an outward facing position of the histidine side chain, away from the coiled-coil interactions between the α1 and α2 helices (Figure 1C). This indicates that the missense variant is unlikely to affect TSEN15 structure or stability. Furthermore, a search of the gnomAD database (27; Supplementary Table 1) reported the variant as benign. We could not undertake *in silico* mutagenesis of the Gly19 residue since the Gly19Asp variant resides within the structurally unresolved N-terminal region of TSEN15. However, gnomAD also predicts this variant as benign. We conclude that the risk of OA residing at the rs1046934 locus is not driven by changes to TSEN15 protein function.

Correlation of the genotype at rs1046934 with *COLGALT2* and *TSEN15* expression in arthroplasty cartilage. The rs1046934 OA association signal was reported along with the observation that genotype at the SNP correlated with expression of *COLGALT2* and *TSEN15* in a range of tissues in the Genotype-Tissue Expression (GTEx) portal, forming eQTLs (4). However, none of the tissues comprising GTEx originate from the articulating joint. We therefore undertook an allelic expression imbalance (AEI) analysis in OA patient cartilage samples to assess whether rs1046934 genotype correlated with expression of either gene in this disease relevant tissue.

Both genes demonstrated AEI (Figure 2), with the OA risk allele C of *COLGALT2* transcript SNP rs114661926 showing an average 1.21-fold increase in *COLGALT2* expression ($P = 0.003$), and the OA risk allele G of *TSEN15* transcript SNP rs2274432 showing an average 1.09-fold increase in *TSEN15* expression ($P = 0.02$).

rs1046934 mQTLs operate within putative enhancers in human arthroplasty cartilage. We next analysed an arthroplasty cartilage epigenome wide DNA methylation dataset (28) to assess whether rs1046934 genotype correlated with proximal DNA methylation levels. We analysed 58 CpGs in a 400kb interval surrounding rs1046934 (Supplementary Table 8) and identified three CpGs whose methylation status nominally ($P < 0.05$) correlated with genotype, forming mQTLs: cg15204595 ($P = 0.005$), cg01436608 ($P = 0.04$), and cg21606956 ($P = 0.002$). At all three, the OA risk-conferring allele A of rs1046934 associated with reduced methylation (Figure 3A).

rs1046934 and the 20 SNPs in high pairwise LD ($r^2 > 0.8$) 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 3B, panels 1 and 2). Two of the mQTL CpGs, cg15204595 and cg01436608, are 2.35kb apart and located within intron 1 of *COLGALT2* (Figure 3B, panels 1,3). They are close to the LD block, with cg01436608 being 595bp from rs74767794, the most upstream variant in the block. cg15204595 and cg01436608 reside within a region that is marked as an enhancer and a transcriptionally active site in musculoskeletal cells (Figure 3B, panel 4), and as an open chromatin region in OA and foetal chondrocytes (Figure 3B, panel 5). Conversely, cg21606956 is distal to the LD block and over 200kb from cg15204595 and cg01436608 (Figure 3B, panels 1-3). It falls within an intergenic enhancer (Figure 3B, panels 3 and 4) that is marked as an open chromatin region in OA and foetal chondrocytes (Figure 3B, panel 5). MSC capture Hi-C data showed physical interactions between a broad region encompassing rs1046934 and the enhancer containing cg15204595 (Figure 3B, panel 6). Additional interactions were observed between the *COLGALT2* promoter and the enhancer containing cg21606956 (Figure 3B, panel 6).

Replication of the mQTLs. We set out to replicate the three mQTLs in an independent cohort of arthroplasty cartilage DNAs. We were able to design pyrosequencing assays for cg15204595 and cg21606956 but not for cg01436608, due to a long run of thymine bases following bisulphite conversion and subsequent PCR amplification. The cg15204595 and cg21606956 mQTLs replicated and confirmed the correlation of the OA risk-conferring allele A of rs1046934 with reduced methylation (Figure 4A).

The arthroplasty cartilage DNAs used for replication were derived from OA (hip and knee) and NOF patients. When the data were stratified by disease state (OA or NOF; Supplementary Figure 1A) mQTLs were detectable in both groups of patients, indicating that the differential methylation is not a consequence of the OA disease state in cartilage. When the methylation data was stratified by disease state irrespective of rs1046934 genotype (Supplementary Figure 1B), methylation at cg15204595 was significantly higher in NOF relative to OA ($P = 0.003$). The NOF patients were on average older than the OA patients at surgery (77.35 years versus 65.32 years for OA knee and 66.51 years for OA hip; Supplementary Table 2). No significant contribution of age to DNA methylation was identified ($P > 0.05$; Supplementary Figure 2).

CpG methylation correlates with *COLGALT2* expression. We subsequently assessed whether there were correlations between DNA methylation and gene expression in samples with matched data (Figure 4B). For *COLGALT2*, significant correlations were observed at cg15204595 ($r^2 = 0.51$, $P = 0.004$) and cg21606956 ($r^2 = 0.57$, $P = 0.005$), marking methylation-expression QTLs (meQTLs). Neither CpG showed significant correlations for *TSEN15*.

OA genetic risk mechanisms at the rs1046934 locus operate in human foetal development. We next determined whether the AEI, mQTL and meQTL effects that we had observed in arthroplasty cartilage were detectable in foetal cartilage. To do this, we analysed matched foetal DNA and RNA samples.

AEI was detected for both genes (Figure 5A), in the same direction as that observed in arthroplasty cartilage (Figure 2), with the OA risk allele C at rs114661926 showing an average 1.35-fold increase in *COLGALT2* expression ($P < 0.0001$), and the OA risk allele G at rs2274432 showing an average 1.03-fold increase in *TSEN15* expression ($P = 0.04$).

cg15204595 and cg21606956 both displayed mQTL effects in the foetal DNA (Figure 5B) in the same direction as that observed in the arthroplasty DNA (Figure 4A), with the OA risk-conferring allele A of rs1046934 correlating with reduced methylation. Mean DNA methylation levels at cg15204595 were higher in foetal cartilage (66.7%) compared to arthroplasty cartilage (62.8%) (P=0.0002), with the opposite observed at cg21606956, with mean values of 40.0% in foetal versus 61.1% in arthroplasty ($P < 0.0001$) (Supplementary Figure 3).

In the foetal cartilage samples, meQTLs were observed for *COLGALT2* but not *TSEN15* (Figure 5C), consistent with our observations in the aged cartilage samples (Figure 4B).

In both arthroplasty and foetal cartilage, the slopes of the *COLGALT2* meQTLs at cg15204595 and cg21606956 were in opposite directions. At cg15204595, high M-values correlated with low AEI ratios, whereas for cg21606956, high M-values correlated with high AEI ratios (Figure 4B and Figure 5C). In Supplementary Text and Supplementary Figure 4, we propose a model to account for this.

306 **cg15204595 and cg21606956 reside in enhancers and their demethylation increases**
307 ***COLGALT2* expression.** We next undertook an *in vitro* investigation of the genomic regions
308 harbouring cg15204595 and cg21606956, and of the CpGs themselves, using the chondrocyte
309 cell line Tc28a2.

310 The regions surrounding cg15204595 and cg21606956 were cloned into the CpG-free
311 Lucia reporter gene vector and tested for enhancer activity in either a methylated or
312 unmethylated state. No other CpGs were captured within the cloned regions. For
313 cg15204595, both the unmethylated and the methylated constructs showed increased Lucia
314 readings compared to the empty control vectors, with average increase in activity of 1.36-fold
315 ($P < 0.01$) and 1.35-fold ($P < 0.01$) respectively (Figure 6A, left). The region encompassing
316 cg21606956 also acted as an enhancer, with an average 1.41-fold ($P < 0.01$) and 1.32-fold (P
317 < 0.001) increase in Lucia activity in the unmethylated and methylated constructs,
318 respectively (Figure 6A, right). *In vitro* methylation status had no significant effect on the
319 function of the enhancers.

320 Targeted demethylation and methylation of cg15204595 and cg21606956 was
321 performed to investigate the impact of DNA methylation on *COLGALT2* and *TSEN15*
322 expression using catalytically dead Cas9 (dCas9) protein coupled with catalytically active TET1
323 (to demethylate) or DNMT3a (to methylate). Control cells were transfected with the same
324 gRNAs coupled with dCas9 and dead TET1 (dTET1) or dead DNMT3a (dDNMT3a). A mean
325 reduction in methylation at cg15204595 and cg21606956 of 12.8% and 17.3%, respectively,
326 was achieved using TET1 (Figure 6B, left). This resulted in 1.3-fold ($P = 0.0009$) and 1.2-fold (P
327 $= 0.01$) increases in *COLGALT2* expression but no significant change in *TSEN15* expression
328 (Figure 6B, right). A mean increase in methylation at cg15204595 and cg21606956 of 10.5%

and 10.7%, respectively, was achieved using DNMT3a (Figure 6C, left). This did not significantly alter the expression of either gene (Figure 6C, right).

Targeted epigenetic modulation indicated that demethylation of cg15204595 and cg21606956 has direct effects on the function of their respective enhancer regions. Furthermore, methylation at CpGs has the potential to alter the binding efficiency of transcription factors to DNA (35,36). We therefore hypothesised that these CpGs fall within transcription factor binding sites. To assess this, we searched JASPAR (37; Supplementary Table 1) and identified multiple transcription factors predicted to bind at or near the CpGs (Supplementary Figure 5A and 5B), many of which are expressed in cartilage (Supplementary Figure 5C).

DISCUSSION

Functional investigation of OA genetic risk loci requires a combination of statistical fine-mapping, *in silico* analyses, and laboratory-based experiments (3,4,9-21). In this report, we studied a novel OA association locus marked by rs1046934. This signal maps close to *COLGALT2*, a gene that we had previously highlighted as a target of a completely independent OA risk locus, marked by rs11583641 (22). We discovered that the rs1046934 locus, like the rs11583641 locus, mediates its effect by modulating the expression of *COLGALT2* via methylation changes to CpGs located in enhancers. The associated SNPs, the CpGs and the enhancers are entirely distinct between the two loci but the ultimate effect on *COLGALT2* is the same. To our knowledge, this is the first time that a gene has been demonstrated through functional investigations to be the target of two independent OA association signals.

In our analysis of arthroplasty cartilage, the risk-conferring allele A of rs1046934 associated with increased *COLGALT2* expression and decreased methylation of CpGs

cg15204595, cg01436608 and cg21606956, with the methylation effects observed at cg15204595 and cg21606956 confirmed in an independent cohort. Importantly, we identified a correlation between methylation and *COLGALT2* expression. Epigenetic modulation demonstrated this to be a direct causal link, with demethylation increasing expression. Furthermore, reporter gene assays confirmed that the genomic regions harbouring cg15204595 and cg21606956 are chondrocyte enhancers. *In silico* data revealed that the CpGs reside in or close to transcription factor binding sites, and in open chromatin regions in chondrocytes, further supporting their functional role. MSC capture Hi-C highlighted physical interactions encompassing the associated SNPs, the cg15204595 and cg21606956 enhancers, and the *COLGALT2* promoter. We conclude therefore that these enhancers interact with *COLGALT2* to regulate its expression, with genotype at the association signal modulating the methylation status and consequently the function of the enhancers.

Although OA is a disease of older people, it has been reported that OA susceptibility has developmental origins, with many OA SNPs correlating with joint shape phenotypes (38-43). This implies that a proportion of OA genetic risk is functionally active during skeletogenesis and early post-natal life and becomes manifest as we age (44-46). We have previously investigated this by assessing AEI and mQTLs of OA genes in foetal cartilage samples (24). For a proportion of the studied genes, the AEI and mQTLs observed in arthroplasty cartilage were also observed in foetal cartilage (24). This included the rs11583641 *COLGALT2* locus (24), which prompted us to investigate foetal cartilage at the rs1046934 locus. The rs1046934 AEI, mQTL and meQTL effects detected in arthroplasty cartilage were also detected in foetal cartilage, implying that this locus is one in which molecular effects on a target gene are activated during development. Our cohort of arthroplasty patient cartilage samples included NOF patients, who lack OA cartilage lesions in

their hip joints. RNA from these patient samples was not available for analysis, however, using DNA we detected mQTLs at cg15204595 and cg21606956 in these samples. Combined, our foetal and NOF data imply that the molecular effects of the rs1046934 signal on *COLGALT2* are not dependent on age or on OA disease status yet contribute to this highly polygenic disease across the life course.

In our dCas9 experiment, demethylation of cg15204595 and cg21606956 had significant effects on *COLGALT2* expression. Demethylating cg15204595 and cg21606956 *in vitro* mimics the effect of the risk-conferring allele A of rs1046934 in cartilage, which associates with reduced methylation of the CpGs and with increased *COLGALT2* expression. We propose that the enhancers harbouring cg15204595 and cg21606956 are particularly sensitive to decreased methylation, accounting for the changes in *COLGALT2* expression, which were only measured when DNA methylation levels were reduced, and not increased.

Throughout our report, we investigated *TSEN15* alongside *COLGALT2* as both genes were highlighted in the discovery GWAS as potential targets of the association signal, primarily due to rs1046934 eQTLs at each gene in GTEx (4). We observed AEI at *TSEN15*, albeit the fold differences in expression between risk/non-risk alleles were not as large as those measured for *COLGALT2*. We did not however observe meQTLs for *TSEN15*, and the epigenetic modulation of cg15204595 and cg21606956 did not significantly alter *TSEN15* expression. Furthermore, our *in silico* analyses of the *TSEN15* missense variants did not indicate any impact of the changes upon protein structure or function. Despite these observations, we cannot definitively exclude *TSEN15* as an additional target of the rs1046934 association signal.

Clinical exploitation of OA genetic discoveries will require an understanding of the molecular mechanism by which risk-conferring alleles impact their target genes (1,46,47). In

this report, we undertook a detailed experimental analysis of the OA locus marked by rs1046934, highlighting its effect on the expression of *COLGALT2* via two distal enhancers that are epigenetically regulated. Our data points toward the important role of development in OA and, for the first time, provides compelling evidence of a target gene being impacted in a near identical manner by two genetically independent OA association signals and disease-relevant gene enhancers. Epigenetic effects on gene expression are an increasingly common observation in OA genetic studies (19,47,48) and may provide opportunities for therapeutic intervention through the application of epigenetic editing tools, such as those utilised in this report (49,50).

REFERENCES

1. Aubourg G, Rice SJ, Bruce-Wootton P, Loughlin J. Genetics of osteoarthritis. Osteoarthritis Cartilage 2022;30:636-49.
2. Styrkarsdottir U, Lund SH, Thorleifsson G, Zink F, Stefansson OA, Sigurdsson JK, 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 2018;50:1681-7.
3. Tachmazidou I, Hatzikotoulas K, Southam L, Esparza-Gordillo J, Haberland V, Zheng J, et al. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nat Genet 2019;51:230-6.
4. Boer CG, Hatzikotoulas K, Southam L, Stefánsdóttir L, Zhang Y, Coutinho de Almeida R, et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell 2021;184:4784-818.

5. Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. Am J Hum Genet 2018;102:717-30.
6. Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. Front Genet 2020;11:424.
7. Lichou F, Trynka G. Functional studies of GWAS variants are gaining momentum. Nat Commun 2020;11:6283.
8. Lappalainen T, MacArthur DG. From variant to function in human disease genetics. Science 2021;373:1464-8.
9. Steinberg J, Ritchie GRS, Roumeliotis TI, Jayasuriya RL, Clark MJ, Brooks RA, et al. Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. Sci Rep 2017;7:8935.
10. Liu Y, Chang J-C, Hon C-C, Fukui N, Tanaka N, Zhang Z, et al. Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation on osteoarthritis. Sci Rep 2018;8:15499.
11. Shepherd C, Zhu D, Skelton AJ, Combe J, Threadgold H, Zhu L, et al. Functional characterization of the osteoarthritis genetic risk residing at ALDH1A2 identifies rs12915901 as a key target variant. Arthritis Rheumatol 2018;70:1577-87.
12. den Hollander W, Pulyakhina I, Boer C, Bomer N, van der Breggen R, Arindrarto W, et al. Annotating transcriptional effects of genetic variants in disease-relevant tissue: transcriptome-wide allelic imbalance in osteoarthritic cartilage. Arthritis Rheumatol 2019;71:561-70.
13. Shepherd C, Reese AE, Reynard LN, Loughlin J. Expression analysis of the osteoarthritis genetic susceptibility mapping to the matrix Gla protein gene MGP. Arthritis Res Ther 2019;21:149.

14. Klein JC, Keith A, Rice SJ, Shepherd C, Agarwal V, Loughlin J, et al. Functional testing of thousands of osteoarthritis-associated variants for regulatory activity. Nat Commun 2019;10:2434.
15. Steinberg J, Southam L, Roumeliotis TI, Clark MJ, Jayasuriya RL, Swift D, et al. A molecular quantitative trait locus map for osteoarthritis. Nat Commun 2021;12:1309.
16. Rice S, Aubourg G, Sorial A, Almarza D, Tselepi M, Deehan D, et al. Identification of a novel, methylation-dependent, RUNX2 regulatory region associated with osteoarthritis risk. Hum Mol Genet 2018;27:3464-74.
17. Rice SJ, Tselepi M, Sorial AK, Aubourg G, Shepherd C, Almarza D, et al. Prioritization of PLEC and GRINA as osteoarthritis risk genes through the identification and characterization of novel methylation quantitative trait loci. Arthritis Rheumatol 2019;71:1285-96.
18. Sorial AK, Hofer IM, Tselepi M, Cheung K, Parker E, Deehan DJ, et al. Multi-tissue epigenetic analysis of the osteoarthritis susceptibility locus mapping to the plectin gene PLEC. Osteoarthritis Cartilage 2020;28:1448-58.
19. Rice SJ, Beier F, Young DA, Loughlin J. Interplay between genetics and epigenetics in osteoarthritis. Nat Rev Rheumatol 2020;16:268-81.
20. Parker E, Hofer IM, Rice SJ, Earl L, Anjum S, Deehan D, et al. Multi-tissue epigenetic and gene expression analysis combined with epigenome modulation identifies RWDD2B as a target of osteoarthritis susceptibility. Arthritis Rheumatol 2021;73:100-9.
21. Rice SJ, Roberts JB, Tselepi M, Brumwell A, Falk J, Steven C, et al. Genetic and epigenetic fine-tuning of TGFB1 expression within the human osteoarthritic joint. Arthritis Rheumatol 2021;73:1866-77.

22. Kehayova YS, Watson E, Wilkinson JM, Loughlin J, Rice SJ. Genetic and epigenetic interplay within a COLGALT2 enhancer associated with osteoarthritis. Arthritis Rheumatol 2021;73:1856-65.
23. Dominguez LJ, Barbagallo M, Moro L. Collagen overglycosylation: a biochemical feature that may contribute to bone quality. Biochem Biophys Res Commun 2005;330:1-4.
24. Rice SJ, Brumwell A, Falk J, Kehayova YS, Casement J, Parker E, et al. Genetic risk of osteoarthritis operates during human fetal development. Hum Mol Genet (under review). Accessible via Research Square, DOI: 10.21203/rs.3.rs-2056256/v1
25. Song J, Markley JL. Three-dimensional structure determined for a subunit of human tRNA splicing endonuclease (Sen15) reveals a novel dimeric fold. J Mol Biol 2007;366:155-64.
26. Sekulovski, S. Devant P, Panizza S, Gogakos T, Pitiriciu A, Heitmeier K, et al. Assembly defects of human tRNA splicing endonuclease contribute to impaired pre-tRNA processing in pontocerebellar hypoplasia. Nature Commun 2021;12:5610.
27. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434-43.
28. Rice SJ, Cheung K, Reynard LN, Loughlin J. Discovery and analysis of methylation quantitative trait loci (mQTLs) mapping to novel osteoarthritis genetic risk signals. Osteoarthritis Cartilage 2019;27:1545-56.
29. Ajekigbe B, Cheung K, Xu Y, Skelton AJ, Panagiotopoulos A, Soul J, et al. Identification of long non-coding RNAs expressed in knee and hip osteoarthritic cartilage. Osteoarthritis Cartilage 2019;27:694-702.

30. Kokenyesi R, Tan L, Robbins JR, Goldring MB. Proteoglycan production by immortalized human chondrocyte cell lines cultured under conditions that promote expression of the differentiated phenotype. Arch Biochem Biophys 2000;383:79-90.
31. Vojta A, Dobrinic P, Tadic V, Bockor L, Korac P, Julg B, et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res 2016;44:5615-28.
32. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 2001;25:402-8.
33. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 2010;11:587.
34. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 2012;28:1353-8.
35. Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet 2016;17:551-65.
36. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 2017;356:eaaj2239.
37. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res 2022;50:D165-73.
38. Slagboom E, Meulenbelt I (2008) Genetics of osteoarthritis: early developmental clues to an old disease. Nat Clin Pract Rheumatol 2008;4:563.
39. Pitsillides AA, Beier F. Cartilage biology in osteoarthritis - lessons from developmental biology. Nat Rev Rheumatol 2011;7:654-63.

40. Sandell LJ. Etiology of osteoarthritis: genetics and synovial joint development. Nat Rev Rheumatol 2012;8:77-89.
41. Aspden RM, Saunders FR. Osteoarthritis as an organ disease: from the cradle to the grave. Eur Cell Mater 2019;37:74-87.
42. Tangredi BP, Lawler DF. Osteoarthritis from evolutionary and mechanistic perspectives. Anat Rec (Hoboken) 2020;303:2967-76.
43. Wilkinson JM, Zeggini E. The genetic epidemiology of joint shape and the development of osteoarthritis. Calcif Tissue Int 2021;109:257-76.
44. Richard D, Liu Z, Cao J, Kiapour AM, Willen J, Yarlagadda S, et al. Evolutionary selection and constraint on human knee chondrocyte regulation impacts osteoarthritis risk. Cell 2020;181:362-81.
45. Muthuirulan P, Zhao D, Young M, Richard D, Liu Z, Emami A, et al. Joint disease-specificity at the regulatory base-pair level. Nat Commun 2021;12:4161.
46. Loughlin J. Translating osteoarthritis genetics research: challenging times ahead. Trends Mol Med 2022;28:176-82.
47. Young DA, Barter MJ, Soul J. Osteoarthritis year in review: genetics, genomics, epigenetics. Osteoarthritis Cartilage 2022;30:216-25.
48. Izda V, Martin J, Sturdy C, Jeffries MA. DNA methylation and noncoding RNA in OA: recent findings and methodological advances. Osteoarthr Cartil Open 2021;3:100208.
49. Almarza D, Cucchiarini M, Loughlin J. Genome editing for human osteoarthritis - a perspective. Osteoarthritis Cartilage 2017;25:1195-8.
50. Tanikella AS, Hardy MJ, Frahs SM, Cormier AG, Gibbons KD, Fitzpatrick CK, et al. Emerging gene-editing modalities for osteoarthritis. Int J Mol Sci 2020;21:6046.

ACKNOWLEDGMENTS

Cartilage tissue was provided by the Newcastle Bone and Joint Biobank, supported by the NIHR Newcastle Biomedical Research Centre, awarded to the Newcastle-upon-Tyne NHS Foundation Trust and Newcastle University. We thank the surgeons and research nurses at the NHS Foundation Trust for providing us with access to these samples. The human embryonic and foetal material was provided by the Joint MRC/Wellcome Trust (grant# MR/R006237/1) Human Developmental Biology Resource (http://hdbr.org).

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Prof. Loughlin had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Kehayova, Wilkinson, Rice, Loughlin.

Acquisition of data. Kehayova, Rice.

Analysis and interpretation of data. Kehayova, Wilkinson, Rice, Loughlin.

FUNDING

Supported by Versus Arthritis (grants 20771 and 22615), by the Medical Research Council and Versus Arthritis Centre for Integrated Research into Musculoskeletal Ageing (CIMA, grant references MR/P020941/1 and MR/R502182/1), and by the Ruth & Lionel Jacobson Charitable Trust.

COMPETING INTEREST

The authors report no conflicts of interest.

FIGURE LEGENDS

Figure 1. TSEN15 protein structure. **A.** Crystal structure of TSEN15 (green)-TSEN34 (yellow/orange) heterodimer (Protein Data Bank, 6Z9U). The position of the Gln59 residue is highlighted and the amino acid side chain displayed. **B.** Monomeric crystal structure of TSEN15 (Protein Data Bank, 2GW6) with α1-3 and β1-6 of the α-α-β-β-β-α-β-β fold numbered. Gln59 is labelled, and side chains displayed (red, oxygen atom; blue, nitrogen atom). **C.** Structure shown as in B following *in silico* mutagenesis to predict the conformation of Gln59His (labelled). Structures were viewed and mutagenesis performed using the PyMOL Molecular Graphics System, version 2.1.1 (https://pymol.org).

Figure 2. Allelic expression imbalance (AEI) analysis of *COLGALT2* and *TSEN15* in arthroplasty cartilage samples. **A**. Allelic ratios for *COLGALT2* transcript SNP rs114661926 (C/G; C = OA risk allele). **B**. Allelic ratios for *TSEN15* transcript SNP rs2274432 (G/A; G = OA risk allele). Patient sample IDs on the x-axes. Each triangle represents the mean of three technical replicates. Boxplots represent the mean cDNA values measured across all samples, with the line inside the box representing the median, the box the interquartile range, and the whiskers the minimum and maximum values. The dashed line represents the allele ratios in genomic DNA. *P* values calculated using Wilcoxon matched-pairs signed rank test.

589 **Figure 3.** mQTL discovery and *in silico* analysis. **A.** Violin plots showing DNA methylation values
590 at cg15204595, cg01436608 and cg21606956 stratified by genotype at rs1046934. Solid and
591 dashed horizontal lines represent the median and interquartile range. *P* values calculated by
592 linear regression. **B.** Schematic overview of the rs1046934 locus. **Panel 1,** the relative genomic
593 position of the 5' end of *COLGALT2* and all of *TSEN15,* visualised in the UCSC Genome Browser
594 (hg19). **Panel 2,** the genomic position of rs1046934 (red line) and the SNPs in high LD with it
595 (pairwise r^2 values > 0.8; black lines). The SNPs comprise a 30kb block. **Panel 3,** the relative
596 genomic positions of cg15204595, cg01436608 and cg21606956 (black lines). **Panel 4,**
597 chromatin state data from ROADMAP for primary human MSCs (H1 MSC), MSC derived
598 chondrocytes (MSC.DR.CHON), MSC derived adipocytes (MSC.DR.ADIP), adipose derived
599 MSCs (ADIP.DR.MSC) and human osteoblasts (OSTEOBLASTS). The colours corresponding to
600 different chromatin states are shown at the bottom of the figure. **Panel 5,** ATAC-sequencing
601 peaks generated from OA hip and knee chondrocytes (open regions marked by orange blocks)
602 and from foetal hip and knee chondrocytes (open regions marked by blue blocks). **Panel 6,**
603 capture Hi-C chromatin interactions from the 3D Genome Browser in human MSCs,
604 represented as loops with the flat end of the loop spanning the width of the interacting
605 regions.
606
607 **Figure 4.** Replication of mQTLs and discovery of meQTLs in arthroplasty cartilage. **A.** Violin
608 plots showing DNA methylation values at cg15204595 and cg21606956 stratified by genotype
609 at rs1046934. Solid and dashed horizontal lines represent the median and interquartile range.
610 *P* values calculated by linear regression. **B.** AEI allelic ratios (log_2) for *COLGALT2* (rs114661926)
611 and *TSEN15* (rs2274432) plotted against matched DNA methylation levels (M-values) at

612 cg15204595 and cg21606956. Each dot is data from one individual. *P* values calculated by
613 linear regression.
614
615 **Figure 5.** AEI, mQTL and meQTL analysis in foetal cartilage. **A.** Allelic ratios for *COLGALT2*
616 transcript SNP rs114661926 (C/G; C = OA risk allele) and for *TSEN15* transcript SNP rs2274432
617 (G/A; G = OA risk allele). Patient sample IDs on the x-axes. Each triangle represents the mean
618 of three technical replicates. Boxplots represent the mean cDNA values measured across all
619 samples, with the line inside the box representing the median, the box the interquartile range,
620 and the whiskers the minimum and maximum values. The dashed line represents the allele
621 ratios in genomic DNA. *P* values calculated using Wilcoxon matched-pairs signed rank test. **B.**
622 Violin plots showing DNA methylation values at cg15204595 and cg21606956 stratified by
623 genotype at rs1046934. Solid and dashed horizontal lines represent the median and
624 interquartile range. *P* values calculated by linear regression. **C.** AEI allelic ratios (log_2)
625 for *COLGALT2* (rs114661926) and *TSEN15* (rs2274432) plotted against matched DNA
626 methylation levels (M-values) at cg15204595 and cg21606956. Each dot is data from one
627 individual. *P* values calculated by linear regression.
628
629 **Figure 6.** cg15204595 and cg21606956 reside in enhancers and increase *COLGALT2* expression
630 when demethylated. **A.** Normalised Lucia reporter gene luminescence readings measured in
631 Tc28a2 chondrocytes following transfection with a construct containing the region
632 surrounding cg15204595 (left) or cg21606956 (right) in an unmethylated or methylated state.
633 Dashed lines represent readings from cells transfected with empty control vectors. Individual
634 biological replicates (n = 8) are represented by black dots. **B.** Left, DNA methylation levels at
635 cg15204595 (top) and cg21606956 (bottom) in Tc28a2 chondrocytes following transfection

636 of gRNAs with dCas9 protein coupled with dTET1 in controls (black dots) or with active TET1
637 (orange dots). Six biological replicates per treatment. Right, effect of the methylation
638 decrease on *COLGALT2* and *TSEN15* expression. Values were normalized to the mean values
639 in control cells. **C.** Left, DNA methylation levels at cg15204595 (top) and cg21606956 (bottom)
640 in Tc28a2 chondrocytes following transfection of gRNAs with dCas9 protein coupled with
641 dDNMT3a in controls (black dots) or with active DNMT3a (orange dots). Six biological
642 replicates per treatment. Right, effect of the methylation increase on *COLGALT2* and *TSEN15*
643 expression. Values were normalized to the mean values in control cells. For A, B and C, bars
644 show the mean ± standard error of the mean (SEM). For A, *P* values calculated using a paired
645 t-test for empty control versus insert, and an unpaired t-test for unmethylated insert versus
646 methylated insert. For B and C, *P* values calculated using a paired *t*-test. **P* < 0.05; ***P* < 0.01;
647 ****P* < 0.001; ns = not significant (*P* > 0.05).
648
649 **Supplementary Figure 1.** Stratification of the arthroplasty methylation data. **A.** mQTL analysis
650 with the replication data stratified into OA and NOF. *P* values calculated by linear regression.
651 **B.** Stratification of the methylation data by OA knee, OA hip and NOF irrespective of
652 rs1046934 genotype. *P* values calculated by a Mann-Whitney U test. For A and B, the solid
653 and dashed horizontal lines of the violin plots represent the median and interquartile range.
654
655 **Supplementary Figure 2.** Age versus methylation. Linear regression was used to test for
656 correlation between age at surgery in years and DNA methylation levels at cg15204595 (top)
657 and cg21606956 (bottom). Patients were studied combined (OA knee, OA hip and NOF) and
658 following stratification. Each dot is data from one individual.
659

660 **Supplementary Figure 3.** CpG methylation levels in arthroplasty and foetal samples.
661 Stratification of the arthroplasty and foetal methylation data irrespective of rs1046934
662 genotype. *P* values calculated by a Mann-Whitney U test. Solid and dashed horizontal lines of
663 the violin plots represent the median and interquartile range.
664
665 **Supplementary Figure 4.** Theoretical model to account for the opposing slopes of the
666 *COLGALT2* meQTLs at cg15204595 and cg21606956. To be read in conjunction with
667 Supplementary Text. **A.** Schematic diagram showing the factors affecting the methylation
668 levels at CpG1 and CpG2 (marked by black circles). The SNP (M, major allele; m, minor allele)
669 is marked by a red line and Effect 1 and Effect 2 are marked with black lines. Arrows indicate
670 the effect the different factors have on the methylation levels at the two CpG sites with the
671 direction of the arrowhead indicating the direction of the effect (up = increased methylation;
672 down = decreased methylation) and the thickness of the arrow corresponding to the strength
673 of the effect (thin = weaker effect; thick = stronger effect). **B.** Table summarising the
674 cumulative effect of the SNP and Effect 1 and Effect 2 on the methylation levels at CpG1 (left)
675 and CpG2 (right) for individuals homozygous (MM or mm) and heterozygous (Mm) for the
676 SNP. The effects are represented by arrows with the direction of the arrowhead indicating
677 the direction of the effect (up = increased methylation; down = decreased methylation) and
678 the thickness of the arrow corresponding to the strength of the effect (thin = weaker effect;
679 thick = stronger effect). **C.** Predicted mQTL plots at CpG1 (left) and CpG2 (right) using the
680 expected overall methylation levels presented in B. Methylation levels are represented on the
681 y-axis by arrows with the direction of the arrowhead indicating the direction of the effect (up
682 = increased methylation; down = decreased methylation) and the thickness of the arrow
683 corresponding to the levels of methylation (thin = medium high/low levels; thick = very

high/low levels). Genotype at the SNP is on the x-axis (MM, Mm or mm). The overall methylation 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 methylation levels at CpG1 plotted against the expected overall methylation levels at CpG2 in all individuals (left) and in heterozygous individuals (right). Methylation levels are represented by arrows 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 levels of methylation (thin = medium high/low levels; thick = very high/low levels). The lines show the direction of the predicted correlations. **E.** Plots showing the measured methylation levels at cg15204595 plotted against the measured methylation levels at cg21606956 in all arthroplasty (top left) and all foetal (bottom left) samples, in arthroplasty (top middle) and foetal (bottom middle) samples homozygous (AA and CC) for rs1046934, and in arthroplasty (top right) and foetal (bottom right) samples heterozygous (AC) for rs1046934. 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 correlations.

Supplementary Figure 5. Transcription factors (TFs) predicted to bind at or close to cg15204595 and cg21606956. TF binding sites within 200bp of cg15204595 (**A**) or 200bp of cg21606956 (**B**) as predicted by JASPAR, visualised in the UCSC Genome Browser (hg19). Sections 1 of **A** and **B** highlight the CpGs (black lines), sections 2 the positions of the TFs. 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. **C.** Expression levels (TPM, transcripts per million) of those TFs predicted to bind within 30bp of cg15204595 and cg21606956 in hip cartilage

708 RNA-sequencing data from OA (n = 10, dark grey) and NOF (n = 6, light grey) patients. Bars
709 show the mean ±SEM. Y-axis is a linear segmented scale with three segments.
710
711
712
713
714
715

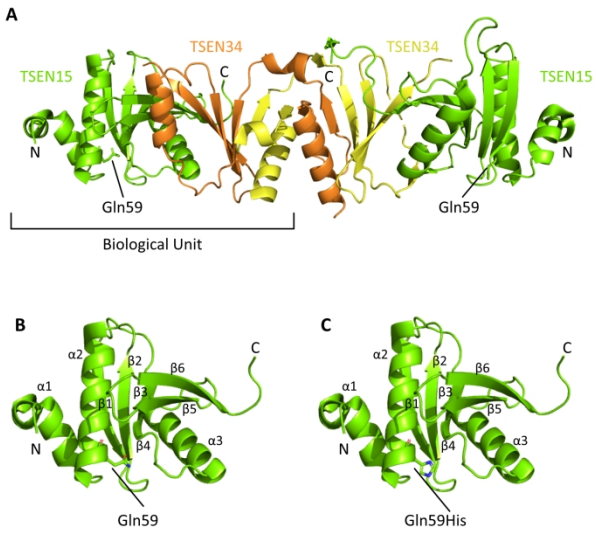


Figure 1. TSEN15 protein structure. A. Crystal structure of TSEN15 (green)-TSEN34 (yellow/orange) heterodimer (Protein Data Bank, 6Z9U). The position of the Gln59 residue is highlighted and the amino acid side chain displayed. B. Monomeric crystal structure of TSEN15 (Protein Data Bank, 2GW6) with α1-3 and β1-6 of the α-α-β-β-β-α-β-β fold numbered. Gln59 is labelled, and side chains displayed (red, oxygen atom; blue, nitrogen atom). C. Structure shown as in B following in silico mutagenesis to predict the conformation of Gln59His (labelled). Structures were viewed and mutagenesis performed using the PyMOL Molecular Graphics System, version 2.1.1 (https://pymol.org).

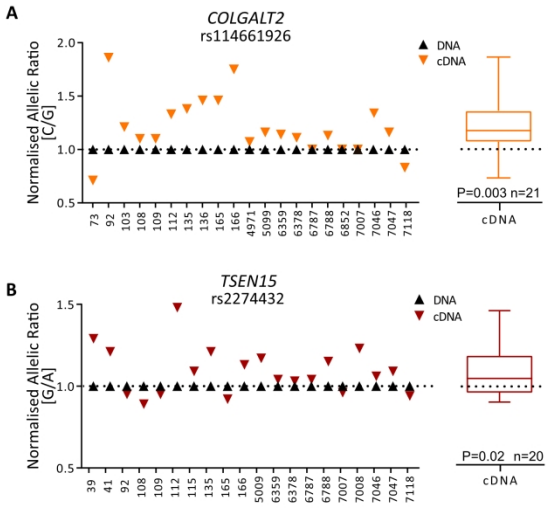


Figure 2. Allelic expression imbalance (AEI) analysis of COLGALT2 and TSEN15 in arthroplasty cartilage samples. A. Allelic ratios for COLGALT2 transcript SNP rs114661926 (C/G; C = OA risk allele). B. Allelic ratios for TSEN15 transcript SNP rs2274432 (G/A; G = OA risk allele). Patient sample IDs on the x-axes. Each triangle represents the mean of three technical replicates. Boxplots represent the mean cDNA values measured across all samples, with the line inside the box representing the median, the box the interquartile range, and the whiskers the minimum and maximum values. The dashed line represents the allele ratios in genomic DNA. P values calculated using Wilcoxon matched-pairs signed rank test.

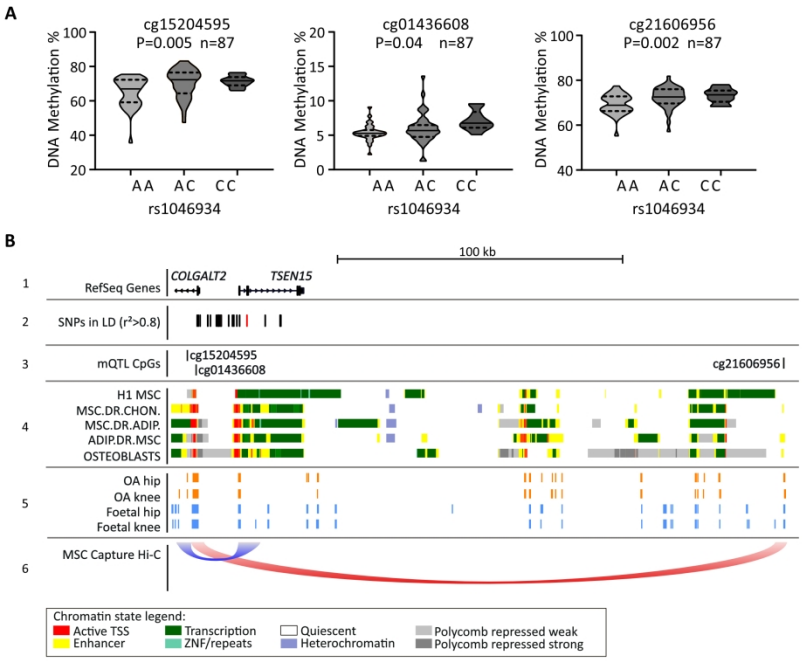


Figure 3. mQTL discovery and in silico analysis. A. Violin plots showing DNA methylation values at cg15204595, cg01436608 and cg21606956 stratified by genotype at rs1046934. Solid and dashed horizontal lines represent the median and interquartile range. P values calculated by linear regression. B. Schematic overview of the rs1046934 locus. Panel 1, the relative genomic position of the 5' end of COLGALT2 and all of TSEN15, visualised in the UCSC Genome Browser (hg19). Panel 2, the genomic position of rs1046934 (red line) and the SNPs in high LD with it (pairwise r2 values > 0.8; black lines). The SNPs comprise a 30kb block. Panel 3, the relative genomic positions of cg15204595, cg01436608 and cg21606956 (black lines). Panel 4, chromatin state data from ROADMAP for primary human MSCs (H1 MSC), MSC derived chondrocytes (MSC.DR.CHON), MSC derived adipocytes (MSC.DR.ADIP), adipose derived MSCs (ADIP.DR.MSC) and human osteoblasts (OSTEOBLASTS). The colours corresponding to different chromatin states are shown at the bottom of the figure. Panel 5, ATAC-sequencing peaks generated from OA hip and knee chondrocytes (open regions marked by orange blocks) and from foetal hip and knee chondrocytes (open regions marked by blue blocks). Panel 6, capture Hi-C chromatin interactions from the 3D Genome Browser in human MSCs, represented as loops with the flat end of the loop spanning the width of the

interacting regions.

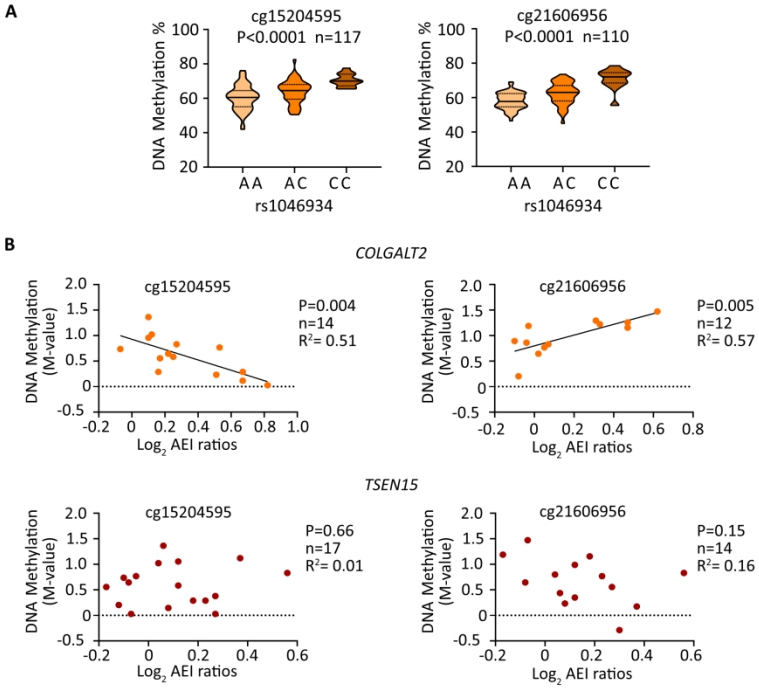


Figure 4. Replication of mQTLs and discovery of meQTLs in arthroplasty cartilage. A. Violin plots showing DNA methylation values at cg15204595 and cg21606956 stratified by genotype at rs1046934. Solid and dashed horizontal lines represent the median and interquartile range. P values calculated by linear regression. B. AEI allelic ratios (log2) for COLGALT2 (rs114661926) and TSEN15 (rs2274432) plotted against matched DNA methylation levels (M-values) at cg15204595 and cg21606956. Each dot is data from one individual. P values calculated by linear regression.

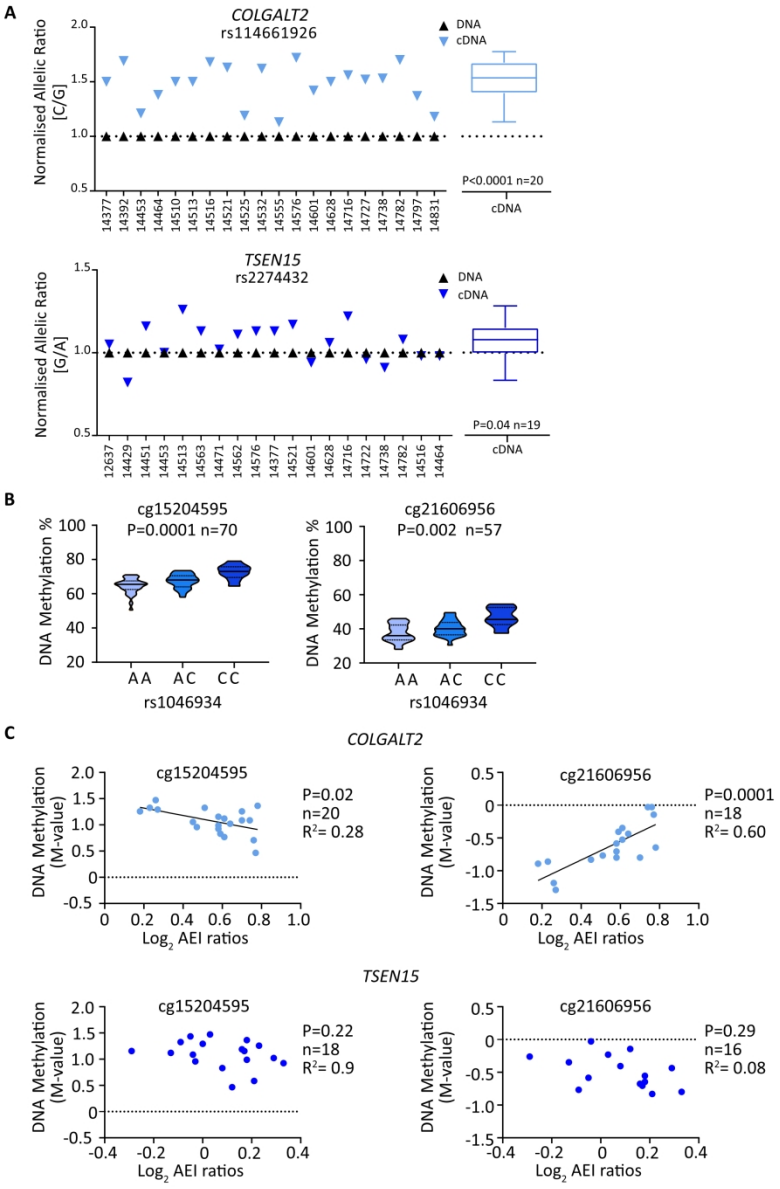


Figure 5. AEI, mQTL and meQTL analysis in foetal cartilage. A. Allelic ratios for COLGALT2 transcript SNP rs114661926 (C/G; C = OA risk allele) and for TSEN15 transcript SNP rs2274432 (G/A; G = OA risk allele). Patient sample IDs on the x-axes. Each triangle represents the mean of three technical replicates. Boxplots represent the mean cDNA values measured across all samples, with the line inside the box representing the median, the box the interquartile range, and the whiskers the minimum and maximum values. The dashed line represents the allele ratios in genomic DNA. P values calculated using Wilcoxon matched-pairs signed rank test. B. Violin plots showing DNA methylation values at cg15204595 and cg21606956 stratified by genotype at rs1046934. Solid and dashed horizontal lines represent the median and interquartile range. P values calculated by linear regression. C. AEI allelic ratios (log2) for COLGALT2 (rs114661926) and TSEN15 (rs2274432) plotted against matched DNA methylation levels (M-values) at cg15204595 and cg21606956. Each dot is data from one individual. P values calculated by linear regression.

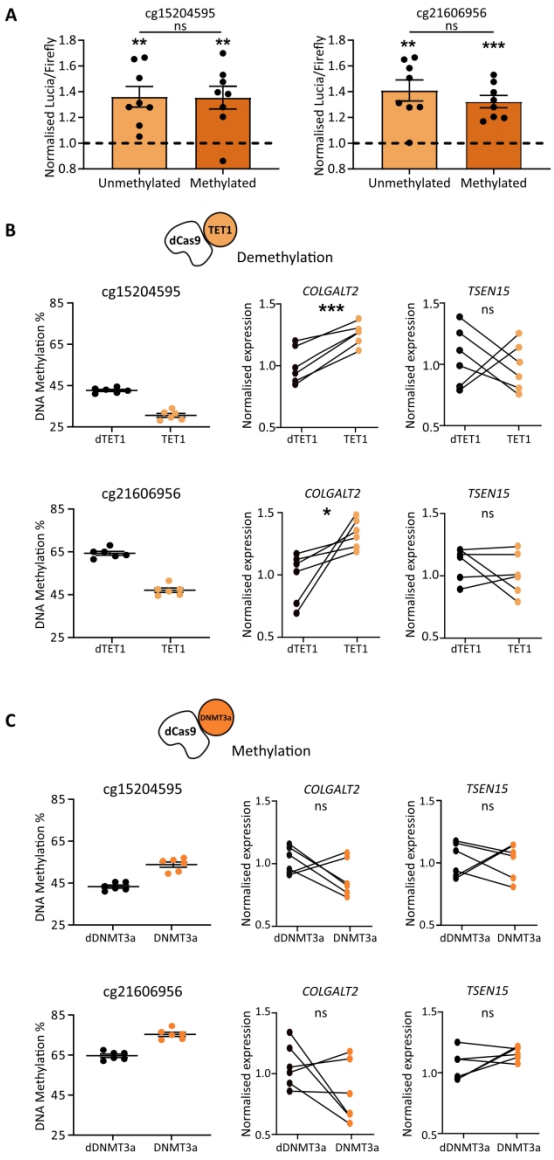


Figure 6. cg15204595 and cg21606956 reside in enhancers and increase COLGALT2 expression when demethylated. A. Normalised Lucia reporter gene luminescence readings measured in Tc28a2 chondrocytes following transfection with a construct containing the region surrounding cg15204595 (left) or cg21606956 (right) in an unmethylated or methylated state. Dashed lines represent readings from cells transfected with empty control vectors. Individual biological replicates (n = 8) are represented by black dots. B. Left, DNA methylation levels at cg15204595 (top) and cg21606956 (bottom) in Tc28a2 chondrocytes following transfection of gRNAs with dCas9 protein coupled with dTET1 in controls (black dots) or with active TET1 (orange dots). Six biological replicates per treatment. Right, effect of the methylation decrease on COLGALT2 and TSEN15 expression. Values were normalized to the mean values in control cells. C. Left, DNA methylation levels at cg15204595 (top) and cg21606956 (bottom) in Tc28a2 chondrocytes following transfection of gRNAs with dCas9 protein coupled with dDNMT3a in controls (black dots) or with active DNMT3a (orange dots). Six biological replicates per treatment. Right, effect of the methylation increase on COLGALT2 and TSEN15 expression. Values were normalized to the mean values in control cells. For A, B and C, bars show the mean ± standard error of the mean (SEM). For A, P values calculated using a paired t-test

for empty control versus insert, and an unpaired t-test for unmethylated insert versus methylated insert. For B and C, P values calculated using a paired t-test. *P < 0.05; **P < 0.01; ***P < 0.001; ns = not significant (P > 0.05).

Supplementary text

Accounting for the observation that the *COLGALT2* meQTLS have opposing slopes.

The OA risk-conferring allele A of rs1046934 is associated with lower methylation levels at cg15204595 and cg21606956 compared to the non-risk allele C in both arthroplasty (Figure 4A) and foetal (Figure 5B) samples. Allele A is also associated with higher *COLGALT2* expression, as evidenced by the AEI analyses, in both tissue types (Figure 2A and Figure 5A). Methylation-expression (meQTL) plots show correlations between methylation levels at both CpGs and *COLGALT2* AEI ratios (Figure 4B and Figure 5C). 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. This is true for both arthroplasty (Figure 4B) and foetal (Figure 5C) samples.

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

In 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 methylation and its minor allele (m) with high methylation at both CpG1 and CpG2, individuals homozygous for the major allele (MM) will have low methylation levels, individuals homozygous for the minor allele (mm) will have high methylation levels, and heterozygous individuals will have intermediate methylation levels at both CpGs (Supplementary Figure 4A and 4B). Assume two additional factors affect the methylation levels at CpG1 and CpG2 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 (Supplementary Figure 4A and 4B). If these two hypothetical effects are weaker than the effect the SNP has on the methylation at CpG1 and CpG2, they will be masked in individuals homozygous at the SNP (MM or mm). However, in heterozygous individuals (Mm), they 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 (Supplementary Figure 4B, 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 (Supplementary Figure 4B, right).

If both the effects of the SNP and Effect 1 or Effect 2 (for CpG1 and CpG2, respectively) are taken into consideration, the expected mQTL plots for CpG1 and CpG2 will show a similar trend (Supplementary Figure 4C). 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 (Supplementary Figure 4D, left). When the same plot is created for heterozygous individuals, however, the trend is reversed and low methylation levels at one CpG will correspond to higher levels at the other (Supplementary Figure 4D, right). This observation is expected since the weaker effects of Effect 1 and Effect 2 complement the effects of the minor and the major alleles (respectively) of the SNP, with the SNP alleles have opposing effects on methylation levels.

The theoretical scenarios described here are consistent with our actual observations in arthroplasty and foetal samples. When methylation levels at cg15204595 and cg21606956 are plotted against genotype at rs1046934 they show a similar trend in both arthroplasty (Figure 4A) and foetal (Figure 5B) samples. When methylation levels at cg15204595 are plotted against the levels at cg21606956 for all arthroplasty (Supplementary Figure 4E, top left) and all foetal (Supplementary Figure 4E, bottom left) samples, low methylation levels at one CpG correlate with low methylation levels at the other. This is particularly striking when only rs1046934 homozygotes (AA and CC) are plotted (Supplementary Figure 4E, top middle [arthroplasty] and bottom middle [foetal]). However, in heterozygous (AC) samples (Supplementary Figure 4E, top right [arthroplasty] and bottom right [foetal]), low methylation at one CpG correlates with higher methylation at the other.

The *COLGALT2* meQTL plots in Figure 4B and Figure 5C show methylation levels at cg15204595 and cg21606956 plotted against AEI ratios in arthroplasty and foetal samples. The AEI analysis is carried out in heterozygous individuals in which lower methylation levels at cg15204595 correspond to higher methylation levels at cg21606956. This leads to the meQTL plots having opposing slopes.

Database	Data type	URL
Protein Data Bank	Macromolecular structures	https://www.rcsb.org
gnomAD	Potential functional impact of genetic variants	https://gnomad.broadinstitute.org
ROADMAP	ChIP-Seq	http://www.roadmapepigenomics.org
3D Genome Browser	Chromatin capture Hi-C	http://3dgenome.fsm.northwestern.edu/chic.php
UCSC Genome Browser	Integrated genomic datasets	https://genome.ucsc.edu
LDlink	SNP linkage disequilibrium measures	https://ldlink.nci.nih.gov/?tab=home
JASPAR	Transcription factor binding profiles	https://jaspar.genereg.net
Gene Expression Omnibus (GEO)	Gene expression data	https://www.ncbi.nlm.nih.gov/geo/

Supplementary Table 1. The public databases used this study.

Sample ID	Sex	Joint	Age at sugery (years)	Disease
4809	Male	Knee	79	OA
4883	Female	Knee	68	OA
4971	Male	Knee	64	OA
5099	Female	Knee	57	OA
5103	Female	Knee	58	OA
5104	Female	Knee	67	OA
5137	Female	Hip	52	OA
5163	Male	Knee	55	OA
5164	Male	Knee	55	OA
5178	Female	Knee	83	OA
5188	Female	Knee	64	OA
5214	Male	Knee	57	OA
5764	Male	Hip	64	OA
5999	Female	Knee	79	OA
6135	Male	Knee	71	OA
6156	Female	Hip	67	OA
6184	Female	Knee	55	OA
6218	Female	Hip	79	OA
6224	Female	Knee	62	OA
6342	Female	Knee	62	OA
6359	Female	Knee	70	OA
6363	Female	Knee	68	OA
6378	Male	Knee	54	OA
6506	Female	Knee	59	OA
6770	Female	Knee	80	OA
6772	Male	Knee	71	OA
6778	Male	Knee	63	OA
6783	Female	Knee	55	OA
6784	Female	Knee	46	OA
6786	Male	Knee	82	OA
6787	Male	Knee	64	OA
6788	Male	Knee	66	OA
6803	Female	Knee	65	OA
6818	Female	Knee	62	OA
6852	Female	Knee	83	OA
6867	Male	Knee	86	OA
7007	Male	Hip	77	OA
7008	Female	Hip	75	OA
7015	Female	Hip	64	OA
7025	Female	Hip	76	OA
7029	Female	Hip	67	OA
7030	Male	Hip	63	OA
7034	Female	Hip	60	OA
7036	Female	Hip	88	OA
7037	Female	Hip	80	OA
7039	Female	Hip	68	OA
7040	Male	Hip	74	OA
7046	Female	Hip	56	OA
7047	Female	Hip	78	OA
7115	Female	Knee	83	OA
7116	Male	Knee	51	OA

7118	Female	Knee	47	OA
7120	Female	Knee	61	OA
7124	Male	Knee	64	OA
19	Male	Hip	66	OA
22	Male	Hip	48	OA
39	Female	Knee	72	OA
40	Female	Hip	84	OA
41	Male	Knee	51	OA
42	Female	Hip	71	OA
45	Male	Knee	83	OA
49	Male	Hip	68	OA
51	Female	Hip	74	OA
52	Female	Hip	74	OA
53	Female	Hip	55	OA
54	Male	Hip	52	OA
57	Male	Knee	63	OA
59	Female	Knee	60	OA
61	Male	Knee	67	OA
66	Female	Hip	56	OA
69	Male	Hip	57	OA
70	Male	Hip	75	OA
72	Male	Knee	76	OA
73	Female	Hip	46	OA
76	Female	Knee	65	OA
77	Female	Knee	71	OA
78	Male	Knee	58	OA
79	Male	Knee	65	OA
82	Female	Knee	67	OA
86	Female	Knee	56	OA
87	Female	Hip	51	OA
89	Female	Knee	67	OA
90	Female	Hip	60	OA
92	Female	Hip	55	OA
96	Female	Hip	61	OA
97	Male	Hip	65	OA
98	Male	Hip	49	OA
100	Female	Hip	74	OA
103	Female	Knee	70	OA
104	Female	Knee	70	OA
106	Male	Knee	79	OA
107	Female	Knee	41	OA
108	Male	Knee	72	OA
109	Female	Knee	65	OA
112	Male	Hip	61	OA
114	Female	Knee	57	OA
115	Male	Knee	69	OA
116	Male	Hip	89	OA
117	Male	Hip	82	OA
126	Female	Hip	66	OA
127	Male	Knee	61	OA
128	Male	Knee	64	OA
135	Female	Hip	69	OA
136	Female	Knee	68	OA
165	Female	Hip	64	OA

166	Female	Knee	62	OA
T023	Female	Hip	68	NOF
T117	Male	Hip	85	NOF
T151	Male	Hip	62	NOF
T167	Female	Hip	62	NOF
T168	Female	Hip	81	NOF
T172	Female	Hip	77	NOF
T174	Male	Hip	86	NOF
T177	Female	Hip	72	NOF
T178	Female	Hip	62	NOF
T179	Female	Hip	81	NOF
T184	Female	Hip	62	NOF
T191	Female	Hip	87	NOF
T192	Female	Hip	71	NOF
T193	Female	Hip	86	NOF
T195	Female	Hip	80	NOF
T196	Female	Hip	83	NOF
T204	Female	Hip	76	NOF
T244	Female	Hip	97	NOF
T245	Female	Hip	81	NOF
T246	Female	Hip	83	NOF
T005	Male	Hip	69	NOF
T007	Female	Hip	71	NOF
T150	Female	Hip	80	NOF
T152	Female	Hip	95	NOF
T154	Male	Hip	75	NOF
T158	Male	Hip	79	NOF

Supplementary Table 2. Details of the arthroplasty patient samples used in this study. OA, osteoarthritis; NOF, neck of femur fracture.

Sample ID	Sex	Limb/Joint	Developmental stage
14375	Male	Proximal and distal tibia	12pcw
14395	Female	Proximal and distal tibia	10pcw
14397	Female	Proximal and distal tibia and femur	15pcw
14423	Male	Proximal and distal tibia	12pcw
14429	Female	Proximal and distal tibia	14pcw
14451	Male	Proximal and distal femur	9pcw
14453	Female	Proximal and distal femur	12pcw
14460	Female	Proximal and distal tibia	9pcw
14464	Male	Proximal and distal tibia	16pcw
14466	Male	Proximal and distal tibia	12pcw
14467	Male	Proximal and distal tibia	10pcw
14475	Male	Proximal and distal tibia	16pcw
14408	Female	Proximal and distal tibia	14pcw
14501	Male	Proximal and distal tibia	16pcw
14513	Female	Proximal and distal tibia and femur	12pcw
14563	Male	Proximal and distal tibia	14pcw
14586	Female	Proximal and distal tibia and femur	13pcw
14604	Female	Proximal and distal tibia	14pcw
14471	Female	Proximal and distal tibia and femur	9pcw
14555	Male	Proximal and distal tibia and femur	9pcw
14453	Female	Proximal and distal femur	12pcw
14562	Female	Proximal and distal tibia	9pcw
14580	Male	Proximal and distal tibia and femur	9pcw
14600	Male	Proximal and distal tibia	8pcw
14576	Male	Proximal and distal tibia and femur	8pcw
14544	Male	Proximal and distal tibia	13pcw
14684	Female	Proximal and distal femur	14pcw
14606	Female	Proximal and distal tibia and femur	9pcw
14617	Female	Proximal and distal tibia and femur	13pcw
14377	Male	Proximal and distal tibia	14pcw
14378	Male	Proximal and distal tibia	14pcw
14392	Male	Proximal and distal tibia	9pcw
14393	Male	Proximal and distal tibia	14pcw
14408	Female	Proximal and distal tibia	14pcw
14472	Male	Proximal and distal tibia	12pcw
14492	Female	Proximal and distal tibia	12pcw
14510	Female	Proximal and distal tibia	14pcw
14511	Female	Proximal and distal tibia	11pcw
14512	Male	Proximal and distal tibia	10pcw
14516	Male	Proximal and distal tibia	15pcw
14521	Male	Proximal and distal tibia	12pcw
14522	Female	Proximal and distal tibia	15pcw
14523	Female	Proximal and distal tibia	15pcw
14524	Female	Proximal and distal tibia	15pcw
14525	Male	Proximal and distal tibia	10pcw
14532	Female	Proximal and distal tibia	11pcw
14541	Female	Proximal and distal tibia	11pcw
14577	Female	Proximal and distal tibia and femur	11pcw
14597	Male	Proximal and distal tibia	11pcw
14601	Male	Proximal and distal tibia and femur	10pcw
14603	Female	Proximal and distal tibia and femur	11pcw
14619	Male	Proximal and distal tibia and femur	16pcw
14628	Female	Proximal and distal tibia and femur	17pcw
14681	Female	Proximal and distal tibia and femur	11pcw

14688	Male	Proximal and distal tibia and femur	11pcw
14689	Male	Proximal and distal femur	10pcw
14703	Female	Proximal and distal tibia and femur	12pcw
14713	Female	Proximal and distal tibia and femur	16pcw
14715	Male	Proximal and distal tibia and femur	12pcw
14716	Male	Proximal and distal tibia and femur	10pcw
14717	Female	Proximal and distal tibia	14pcw
14720	Female	Proximal and distal tibia and femur	15pcw
14721	Female	Proximal and distal tibia and femur	10pcw
14722	Male	Proximal and distal tibia and femur	17pcw
14727	Male	Proximal and distal tibia	12pcw
14728	Male	Proximal and distal tibia and femur	16pcw
14729	Male	Proximal and distal tibia and femur	12pcw
14738	Male	Proximal and distal tibia and femur	15pcw
14746	Female	Proximal and distal tibia and femur	16pcw
14748	Female	Proximal and distal tibia and femur	10pcw
14749	Male	Proximal and distal tibia and femur	13pcw
14782	Male	Proximal and distal tibia and femur	10pcw
14797	Male	Proximal and distal tibia and femur	12pcw
14831	Female	Proximal and distal tibia and femur	9pcw
14835	Female	Proximal and distal tibia and femur	12pcw
14863	Male	Proximal and distal tibia and femur	10pcw
14881	Female	Proximal and distal tibia and femur	10pcw

Supplementary Table 3. Details of the foetal cartilage samples used in this study. PCW, post conceptional weeks.

Experiment	SNP/CpG	Position of SNP/CpG (hg19)	Oligonucleotide		
			Forward (5'-3')	Reverse (5'-3')	Sequencing (5'-3')
Genotyping	rs1046934	chr1:184023529	TTAGATATAGGAGATGCCACCCA	CAGGTAAACCAAGAACGCTACATA	AGGAGATGCCACCCA
Genotyping and AEI	rs114661926	chr1:184006549	GAGAGGAGCAGTAGCGACCAGG	AGCAAGGGGCTGCGAGGG	TGAGGGCGGCGGCGG
	rs2274432	chr1:184020945	GGGAAGCGGAACCCACAG	CCCTGAGGACGCCTGGAT	TGTTCGCGGCTTTGG
Methylation	cg15204595	chr1:184003184	TGTGGTTAGATTTAGAGGAAGTTGA	AACACTCTAATAAAAAATAATACCTCAATT	ATGAGAAATTAGAGGTAAGT
	cg21606956	chr1:184211935	TTTTGTATTTATTTTGTGGTTAATTTTTG	CAATAACTCTATACTATAATAAAATCCTCT	TGTGGTTAATTTTTGGTTT

Supplementary Table 4. Sequence of oligonucleotide primers used for genotyping, allelic expression imbalance (AEI) analysis, and methylation analysis.

Gene	SNP	Position (hg19)	Alleles (MAF in EUR)	OA-risk allele	Pairwise r2 with rs1046934 in EUR
COLGALT2	rs114661926	chr1:184006549	C>G (0.304)	C	0.79
TSEN15	rs2274432	chr1:184020945	G>A (0.327)	G	1

Supplementary Table 5. Transcript SNPs used for AEI analyses. MAF, minor allele frequency; EUR, European ancestry cohorts.

Experiment	Oligonucleotide/gRNA name	Sequence (5'-3')
Lucia - cg15204595	Lucia595 FWD	ATGCATTCAGTTTGGCATCTTGGAATGT
	Lucia595 REV	ACTAGTGATAATAGGGCATCAACCAGAGT
Lucia - cg21606956	Lucia956 FWD	ATGCATCAAAGTTCTGGGTAGCTCACA
	Lucia956 REV	ACTAGTGAATCTCTTTGGAAGTGCTGACT
Demethylation/Methylation of cg15204595	gRNA1 FWD	CACCGTGGAATGTTGCATAGAGCTT
	gRNA1 REV	AAACAAGCTCTATGCAACATTCCAC
Demethylation/Methylation of cg21606956	gRNA2 FWD	CACCGGTTATTGTGGTCAGATTCAG
	gRNA2 REV	AAACCTGAATCTGACCACAATAACC

Supplementary Table 6. Sequences of the oligonucleotide primers used for cloning the regions containing cg15204595 and cg21606959 into the Lucia CpG-free-promoter vector, and of the gRNAs used for the targeted demethylation/methylation of cg15204595 and cg21606959.

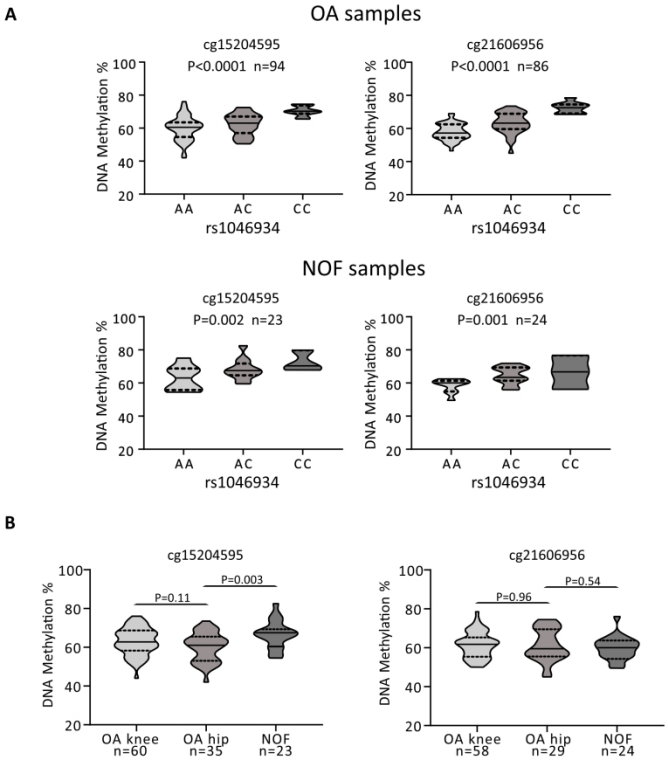
Sequences in red for the Lucia oligonucleotides are restriction enzyme sites used for cloning - ATGCAT, NsiI; ACTAGT, SpeI. Sequences in red for the gRNA oligonucleotides are to facilitate cloning.

Gene	Assay
COLGALT2	Hs.PT.58.19198739
TSEN15	Hs.PT.58.24776232

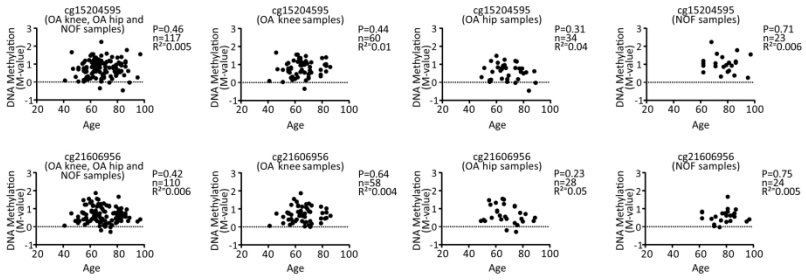
Supplementary Table 7. The pre-designed IDT RT-qPCR assays used to quantify gene expression.

CpG	Nominal P value	Slope	CpG.chr	CpG.pos	CpG.strand	CpG.Island	CpG.Group	CpG.GeneSymbol
cg21606956	0.00238879	0.158917453	chr1	184211934	-	OpenSea		
cg15204595	0.005077495	0.255571328	chr1	184003184	+	N_Shelf	Body	*COLGALT2*
cg01436608	0.04128075	0.178202059	chr1	184005534	-	Island	Body	*COLGALT2*
cg22340762	0.051489372	-0.177768504	chr1	183823538	+	OpenSea	Body	*RGL1*
cg17002156	0.054070936	0.232924321	chr1	183995878	+	OpenSea	Body	*COLGALT2*
cg21404028	0.10836635	0.092775989	chr1	184020687	-	N_Shore	TSS200;TSS200;TSS200	*TSEN15;TSEN15;TSEN15*
cg12222949	0.1130119	0.16041609	chr1	184005360	+	Island	Body	*COLGALT2*
cg15111486	0.133341488	-0.104192865	chr1	184195211	-	OpenSea		
cg05777716	0.144389699	-0.080063479	chr1	184006285	-	Island	1stExon	*COLGALT2*
cg13951632	0.184928856	0.120100661	chr1	183897368	-	OpenSea	3'UTR	*RGL1*
cg15569801	0.187897516	0.066928549	chr1	184197096	-	OpenSea		
cg27177158	0.194193974	0.100123681	chr1	184121585	-	OpenSea		
cg02314308	0.194957334	0.110305009	chr1	184006972	-	S_Shore	TSS200	*COLGALT2*
cg27444414	0.230101787	0.092115213	chr1	184122874	-	OpenSea		
cg10312498	0.240591977	-0.109116001	chr1	183957982	-	OpenSea	Body	*COLGALT2*
cg16655084	0.243080607	0.081155105	chr1	184005896	+	Island	Body	*COLGALT2*
cg12136256	0.248654028	-0.136178579	chr1	183952796	+	OpenSea	Body	*COLGALT2*
cg14773259	0.292969986	0.085088592	chr1	184021034	-	Island	Body;Body;Body	*TSEN15;TSEN15;TSEN15*
cg22400420	0.304191771	-0.088312676	chr1	183851242	-	OpenSea	Body	*RGL1*
cg18131582	0.363285827	0.104087735	chr1	183912305	-	OpenSea	Body	*COLGALT2*
cg00464025	0.381248495	-0.040631933	chr1	183907542	-	OpenSea	3'UTR	*COLGALT2*
cg19101018	0.385337956	-0.063402477	chr1	183925532	-	OpenSea	Body	*COLGALT2*
cg21417627	0.447531032	-0.067376864	chr1	184008235	-	S_Shore	TSS1500	*COLGALT2*
cg14313310	0.485865459	0.055906007	chr1	184006816	-	S_Shore	5'UTR;1stExon	*COLGALT2;COLGALT2*
cg09940311	0.496887744	-0.110077477	chr1	183901666	-	OpenSea		
cg06849459	0.514610056	-0.04667026	chr1	183891439	-	OpenSea	Body	*RGL1*
cg01068808	0.530912366	-0.075189524	chr1	184007101	-	S_Shore	TSS1500	*COLGALT2*
cg18073970	0.542309516	-0.057603771	chr1	184006765	+	Island	5'UTR;1stExon	*COLGALT2;COLGALT2*
cg09684066	0.558049453	0.031355578	chr1	183891565	+	OpenSea	Body	*RGL1*
cg25605307	0.560179747	-0.0498599	chr1	183965312	-	OpenSea	Body	*COLGALT2*
cg11678039	0.580064206	0.041394928	chr1	183841259	-	OpenSea	Body	*RGL1*
cg16716196	0.595667485	0.049572938	chr1	184195068	-	OpenSea		
cg03467001	0.604461199	-0.063457988	chr1	183857649	-	OpenSea	Body	*RGL1*
cg03943177	0.611155155	0.052335755	chr1	184021360	+	S_Shore	Body;Body;Body	*TSEN15;TSEN15;TSEN15*
cg06915270	0.642327728	0.027943555	chr1	183891360	+	OpenSea	Body	*RGL1*
cg17357479	0.658357979	-0.037040788	chr1	184005717	-	Island	Body	*COLGALT2*
cg10807943	0.689234253	-0.020570128	chr1	184021131	+	S_Shore	Body;Body;Body	*TSEN15;TSEN15;TSEN15*
cg12702671	0.711635427	0.031402346	chr1	183885210	-	OpenSea	Body	*RGL1*
cg20942099	0.7233262	-0.023790576	chr1	184007803	+	S_Shore	TSS1500	*COLGALT2*
cg09013655	0.754989126	-0.032897006	chr1	184005063	-	N_Shore	Body	*COLGALT2*
cg12792264	0.765268544	-0.035320275	chr1	183971636	+	OpenSea	Body	*COLGALT2*
cg00044463	0.766522434	0.02378764	chr1	184006933	+	S_Shore	TSS200	*COLGALT2*
cg24721630	0.769637772	0.027460413	chr1	184006216	+	Island	Body	*COLGALT2*
cg02890642	0.78153886	-0.021560653	chr1	184017798	-	N_Shelf		
cg27291258	0.784525638	-0.043489618	chr1	184173289	+	OpenSea		
cg12758231	0.787779993	0.019910919	chr1	184020711	+	N_Shore	TSS200;TSS200;TSS200	*TSEN15;TSEN15;TSEN15*
cg15359501	0.792639164	0.015547072	chr1	184006496	-	Island	5'UTR;1stExon	*COLGALT2;COLGALT2*
cg26429856	0.798737947	0.013574529	chr1	184005990	+	Island	Body	*COLGALT2*
cg18189236	0.821703244	-0.025694924	chr1	184009497	-	S_Shelf		
cg05816006	0.839369641	-0.018867477	chr1	184042579	+	OpenSea	3'UTR;3'UTR;Body	*TSEN15;TSEN15;TSEN15*
cg11387897	0.849670158	0.012868174	chr1	183914188	-	OpenSea	Body	*COLGALT2*
cg12631766	0.887366794	-0.015505912	chr1	184025029	-	S_Shelf	Body;Body;Body	*TSEN15;TSEN15;TSEN15*
cg11346030	0.915841776	0.009511507	chr1	184090950	-	OpenSea		
cg06379531	0.921982985	-0.00580673	chr1	184020410	-	N_Shore	TSS1500;TSS1500;TSS1500	*TSEN15;TSEN15;TSEN15*
cg23040305	0.96965166	0.004039352	chr1	183846243	-	OpenSea	Body	*RGL1*
cg15521745	0.983100265	0.001355847	chr1	184133404	-	OpenSea		
ch.1.182359526R	0.987854652	0.001960897	chr1	184092903	+	OpenSea		
cg20271396	0.99541804	-0.000473252	chr1	184020713	+	N_Shore	TSS200;TSS200;TSS200	*TSEN15;TSEN15;TSEN15*

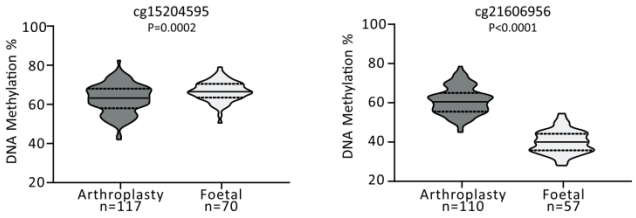
Supplementary Table 8. rs1046934 mQTL analysis of CpGs located 200kb upstream and downstream of the SNP. CpGs are ranked by *P* value. CpGs with $P < 0.05$ are highlighted.



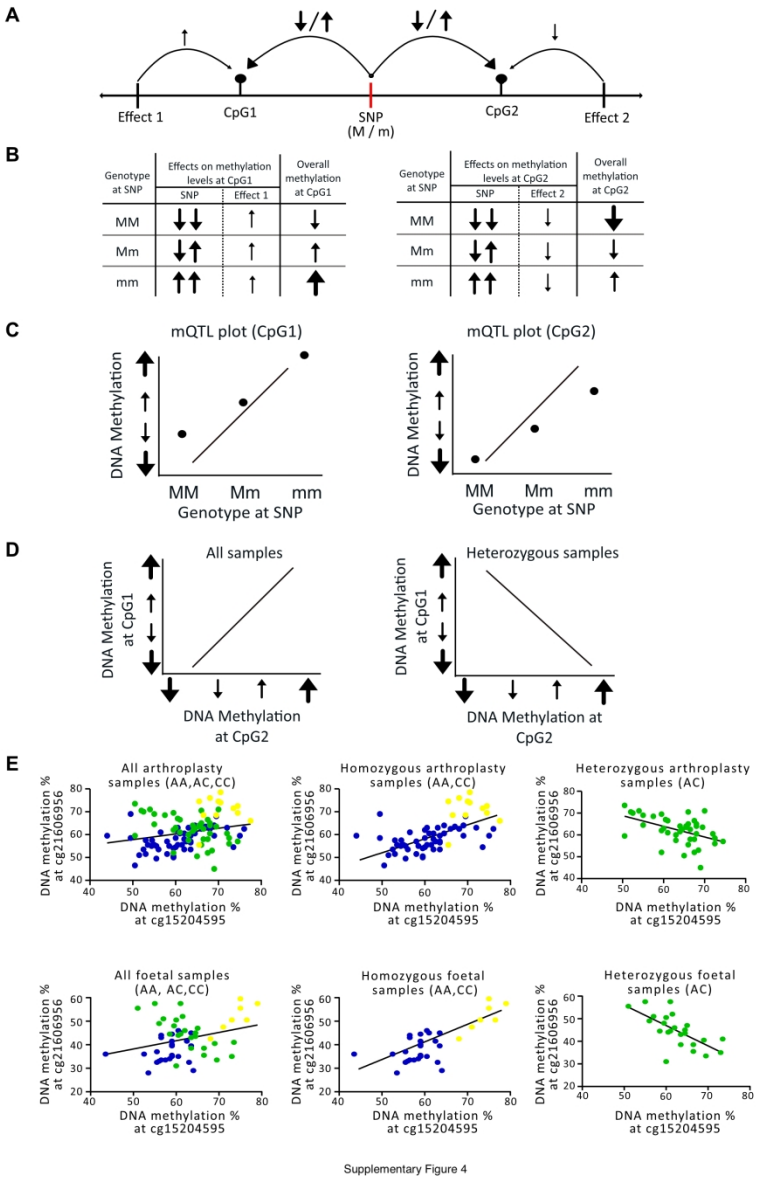
Supplementary Figure 1



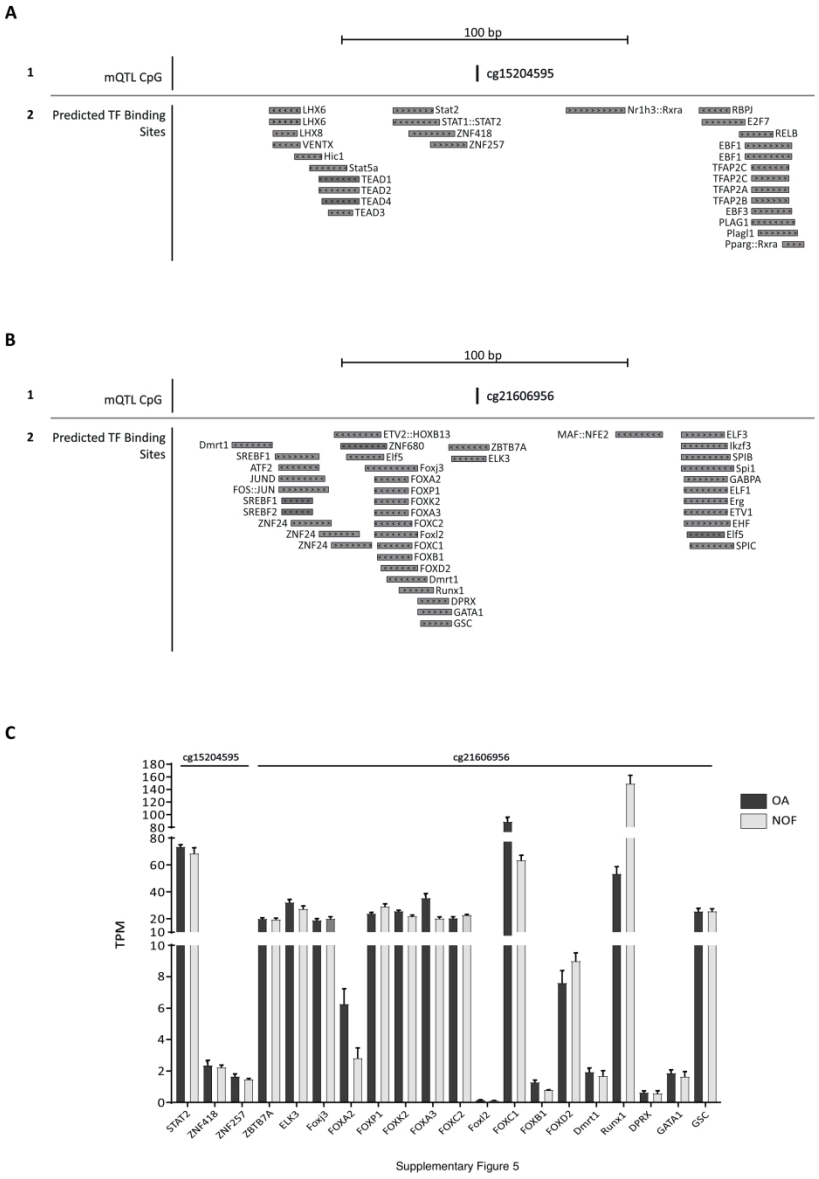
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5