



This is a repository copy of *Genetic and epigenetic interplay within a COLGALT2 enhancer associated with osteoarthritis*.

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/177870/>

Version: Published Version

---

**Article:**

Kehayova, Y.S., Watson, E., Wilkinson, J.M. et al. (2 more authors) (2021) Genetic and epigenetic interplay within a COLGALT2 enhancer associated with osteoarthritis. *Arthritis & Rheumatology*, 73 (10). pp. 1856-1865. ISSN 2326-5191

<https://doi.org/10.1002/art.41738>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Genetic and Epigenetic Interplay Within a *COLGALT2* Enhancer Associated With Osteoarthritis

Yulia S. Kehayova,<sup>1</sup>  Emily Watson,<sup>1</sup> J. Mark Wilkinson,<sup>2</sup>  John Loughlin,<sup>1</sup>  and Sarah J. Rice<sup>1</sup> 

**Objective.** The osteoarthritis (OA)-associated single-nucleotide polymorphism (SNP) rs11583641 is located in *COLGALT2*, encoding a posttranslational modifier of collagen. In cartilage, the SNP genotype correlates with DNA methylation in a putative enhancer. This study was undertaken to characterize the mechanistic relationship between rs11583641, the putative enhancer, and *COLGALT2* expression using cartilage samples from human patients and a chondrocyte cell model.

**Methods.** Nucleic acids were extracted from articular cartilage samples obtained from patients with OA (n = 137). Samples were genotyped, and DNA methylation was quantified at 12 CpGs using pyrosequencing. The putative enhancer was deleted in Tc28a2 chondrocytes using clustered regularly interspaced short palindromic repeat/Cas9, and the impact on nearby gene expression was determined using real-time quantitative polymerase chain reaction. Targeted modulation of the epigenome using catalytically dead Cas9 (dCas9) constructs fused to DNA methyltransferase 3a or ten-eleven translocase 1 allowed for the investigation of a causal relationship between DNA methylation and enhancer activity.

**Results.** The genotype at rs11583641 correlated with DNA methylation at 3 CpGs, and the presence of the OA risk allele, C, corresponded to reduced levels of methylation. Deletion of the enhancer resulted in a 2.7-fold reduction in *COLGALT2* expression. Targeted methylation and demethylation of the CpGs had antagonistic effects on *COLGALT2* expression. An allelic imbalance in the expression of *COLGALT2* was identified in the cartilage from patients with OA, with relative overexpression of the OA risk allele. Allelic expression ratios correlated with DNA methylation at 4 CpGs.

**Conclusion.** *COLGALT2* is a target of OA genetic risk at this locus. The genotype at rs11583641 impacts DNA methylation in a gene enhancer, which, in turn, modulates *COLGALT2* expression. *COLGALT2* encodes an enzyme that initiates posttranslational glycosylation of collagens and is therefore a compelling OA susceptibility target.

## INTRODUCTION

Osteoarthritis (OA) is a multifactorial musculoskeletal condition that most commonly affects the hips, knees, and base of the thumbs (1). The disease is characterized by focal thinning and loss of articular cartilage, which leads to pain and stiffness in the affected joint. This results in reduced mobility and comorbidities, such as cardiovascular disease, ultimately leading to an increase in the rate of all-cause mortality (2,3). Current treatments are limited to symptomatic pain relief and physical therapy, and surgical interventions involving joint replacement are common in end-stage disease.

Genetic predisposition accounts for >50% of total OA susceptibility (4). Approximately 90 independent genetic loci have been significantly associated with OA, according to the findings of genome-wide association studies (GWAS) (5–15). Risk loci are marked by single-nucleotide polymorphisms (SNPs), of which the majority reside within noncoding regions of the genome. Functional SNPs are thought to contribute to OA pathogenesis by altering the expression of target genes, disrupting cartilage homeostasis. Due to the genetic complexity of the reported regions, the identification of functional SNPs and effector genes at the majority of OA risk loci remains elusive (16,17). In recent years, post-GWAS

Supported by Versus Arthritis (grant 20771), the MRC and Arthritis Research UK (MRC–Arthritis Research UK Centre for Integrated Research into Musculoskeletal Ageing grant JXR 10641), and the EU Seventh Framework Program (D-BOARD grant 305815). Funds to purchase PyroMark Q24 Advanced were provided by The Ruth and Lionel Jacobson Charitable Trust.

<sup>1</sup>Yulia S. Kehayova, MRes, Emily Watson, MRes, John Loughlin, PhD, Sarah J. Rice, PhD: International Centre for Life and Newcastle University, Newcastle-upon-Tyne, UK, and MRC–Arthritis Research UK Centre for Integrated Research into Musculoskeletal Ageing and University of Liverpool,

Liverpool, UK; <sup>2</sup>J. Mark Wilkinson, PhD, FRCS: University of Sheffield, Sheffield, UK, and MRC–Arthritis Research UK Centre for Integrated Research into Musculoskeletal Ageing and University of Liverpool, Liverpool, UK.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Sarah J. Rice, PhD, Newcastle University, Skeletal Research Group, Biosciences Institute, International Centre for Life, Newcastle-upon-Tyne NE1 3BZ, UK. Email: sarah.rice@ncl.ac.uk.

Submitted for publication November 30, 2020; accepted in revised form March 11, 2021.

integration of epigenetic data sets has increasingly been conducted to colocalize genetic risk loci with changes to the cartilage epigenome (18–21). Correlations between genotypes at OA association SNPs and DNA methylation have been identified in cartilage, marking methylation quantitative trait loci (QTLs) (19–22). Identification of methylation QTLs prioritizes SNPs, genes, and regulatory elements for downstream analysis and provides insight into the potential molecular mechanisms through which risk variants influence disease susceptibility (23).

A GWAS using the UK Biobank data set identified the SNP rs11583641 at chromosome 1q25.3 to be significantly associated with hip OA ( $P = 5.6 \times 10^{-10}$ ) (11). Subsequently, a cartilage methylation analysis revealed that rs11583641 acts as a methylation QTL (20). Presence of the major, and OA effect, allele of rs11583641 was found to correlate with reduced DNA methylation at a single intronic CpG (cg18131582) (20). Both the SNP and CpG are located in the gene body of *COLGALT2*. Furthermore, *COLGALT2* expression was shown to be significantly increased in OA hip cartilage compared to non-OA cartilage (20). Based on these findings, the *COLGALT2* gene and its putative regulatory element have been prioritized for further investigation.

*COLGALT2* encodes the enzyme Colgalt2, which is one of two known galactosyltransferases (the other encoded by *COLGALT1*) that initiate collagen glycosylation, a posttranslational modification (24). Glycosylation has been functionally linked to oligomerization and stabilization of the collagen triple helix (25,26). Collagen constitutes up to 60% of the dry weight of articular cartilage; therefore, anomalies in folding and stability could result in a loss of tissue integrity and breakdown of cartilage. This highlights *COLGALT2* as a compelling target of the OA genetic risk at chromosome 1q25.3.

Few functional analyses of OA risk loci have successfully demonstrated causal relationships between the SNP genotype, DNA methylation, and the expression of target genes. In this study, we set out to test the prioritized region for enhancer activity, to identify the gene targets of the SNP and enhancer, and to investigate a functional effect of changes in DNA methylation associated with OA risk on gene expression.

## PATIENTS AND METHODS

**Cartilage samples and ethics approval.** Cartilage samples were obtained from the joints of patients undergoing knee or hip replacement surgery due to primary OA or femoral neck fracture. The Newcastle and North Tyneside Research Ethics Committee granted ethics approval for sample collection. Verbal and written informed consent was obtained from each patient prior to surgery (Research Ethics Committee reference no. 14/NE/1212). Nucleic acids were extracted from cartilage as previously described (27). Patient details can be found in Supplementary Table 1 (available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>).

**Genotyping.** Following polymerase chain reaction (PCR) amplification of the SNP region, samples were genotyped using a PyroMark Q24 system (Qiagen), according to the manufacturer's instructions. Primer sequences were generated using PyroMark assay design software (Qiagen) and purchased from IDT (Supplementary Table 2, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>).

**Methylation quantification.** Genomic DNA (500 ng) was treated with sodium bisulfite using EZ DNA Methylation Kits (Zymo Research). Assays were designed to capture the 12 investigated CpGs (PyroMark assay design software; Qiagen) (Supplementary Table 3, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>). Bisulfite-treated DNA was amplified by PCR, and DNA methylation was quantified using a PyroMark Q24 system. Each measurement was performed in duplicate, and replicate values that differed by >5% were excluded from the analysis.

**Allelic expression imbalance analysis.** The relative ratio of C:T at rs11583641 was quantified using pyrosequencing in DNA and complementary DNA (cDNA) from heterozygous patients. Assay details can be found in Supplementary Table 2 (available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>). Cartilage samples were analyzed in triplicate. Replicate values with >5% differences were excluded. Allelic expression in cDNA was normalized to that in genomic DNA for each patient.

**Reporter gene assay.** The investigated region was cloned into the Lucia CpG-free promoter vector (InvivoGen). The putative enhancer was amplified from pooled genomic DNA samples using primers containing the required restriction enzyme sequences for downstream cloning (Supplementary Table 2, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>). The PCR product was first cloned into the pCR2.1-TOPO vector (ThermoFisher) and transformed into chemically competent *Escherichia coli* (*E. coli*). Colonies were grown overnight, and extracted plasmid DNA was sequenced using the Sanger method (Source BioScience). Plasmids containing each of the 3 haplotypes for rs943409 and rs734657 were selected, and DNA was digested with *Avr* II and *Spe* I (New England Biolabs). The inserts were subcloned into a CpG-free-promoter vector containing a Lucia reporter gene (InvivoGen). Plasmids were methylated or mock-methylated in vitro using CpG Methyltransferase (New England Biolabs), which was confirmed using *Hga* I restriction digest (New England Biolabs). Tc28a2 cells, an immortalized human chondrocyte cell line, were seeded onto a 96-well plate (5,000 cells/well) and transfected with 100 ng of pCpG-free promoter and 10 ng of

pGL3-promoter (Promega) using Lipofectamine 2000 (Invitrogen). Cells were lysed after 24 hours. Luminescence was read and analyzed, as previously described (28).

**Genome and epigenome modulation using Cas9.** Two guide RNA (gRNA) sequences (gRNAs 1 and 2) were designed (using the IDT gRNA Design Tool) to target upstream and downstream of the CpGs (Supplementary Table 3, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>). Sequences were ordered as single-stranded DNA oligonucleotides, annealed, and cloned into the PX462 Cas9 plasmid, as previously described in detail (27). Constructs were nucleofected into Tc28a2 cells, and deletions were confirmed as previously described (27). Complementary DNA was synthesized from 1  $\mu$ g of RNA in a reverse transcription reaction with SuperScript II Reverse Transcriptase (Invitrogen). Gene expression was measured by real-time quantitative PCR (QuantStudio 3), using TaqMan primers and probes. The expression of target genes, relative to that of housekeeping genes *18S*, *GAPDH*, and *HPRT1*, was calculated using the formula  $2^{-\Delta\Delta C_t}$  (28). The predesigned TaqMan assays used in this study were purchased from IDT (Supplementary Table 4, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>).

For epigenome modulation, 6 gRNAs targeting the region were designed as described above (Supplementary Table 3, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>). For DNA methyltransferase 3a (DNMT3a)-mediated methylation of the CpGs, the gRNA sequences were synthesized as single-stranded DNA oligonucleotides, annealed, and ligated to the catalytically dead Cas9 (dCas9)-DNMT3a-enhanced green fluorescent protein (EGFP) plasmid (Addgene plasmid 71666), as previously described (27). Plasmid DNA (5  $\mu$ g) was nucleofected into Tc28a2 cells, and successful transfection was confirmed after 24 hours using GFP visualization (AxioVision; Zeiss). Supplementary Figure 1 contains representative images of Tc28a2 cells following transfection with dCas9-DNMT3a-EGFP plasmid (available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>).

For demethylation of the enhancer using dCas9-ten-eleven translocation 1 (TET1), Tc28a2 cells that had been previously stably integrated with a gene expressing an inducible dCas9-TET1 construct were created and cultured as described by Parker et al (29). The 6 sequences targeting the enhancer (Supplementary Table 3, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>) were ordered from IDT as gRNAs. The gRNAs and *trans*-activating CRISPR RNA (tracrRNA; IDT) were diluted to 100  $\mu$ M with Duplex Buffer (IDT). Each gRNA was mixed with tracrRNA (1:1) in a 4  $\mu$ l reaction. The gRNA and the tracrRNA were annealed at 95°C for 5 minutes

and cooled to room temperature to form duplexes. The duplexes were then transfected into the Tc28a2/dCas9-TET1 cells 24 hours after fusion protein induction. The cells were grown in a standard culture medium containing doxycycline (2  $\mu$ g/ml) for another 48 hours, after which they were left to recover in an antibiotic-free medium for 24 hours.

For both experiments, cells were harvested 72 hours post-transfection. DNA was extracted from harvested cell pellets using the PureLink Genomic DNA Mini kit (ThermoFisher), and RNA was extracted using a NucleoSpin TriPrep kit (Macherey-Nagel).

**Statistical analysis.** Genotype and methylation correlations were calculated using the Kruskal-Wallis test. For Lucia reporter assays, we corrected for multiple comparisons using the Holm-Sidak or Dunn's test. Changes in gene expression following Cas9 modulation were calculated using paired *t*-tests. Allelic expression imbalance and DNA methylation relationships were determined using linear regression analysis. All tests were performed in GraphPad Prism 8.3.1.

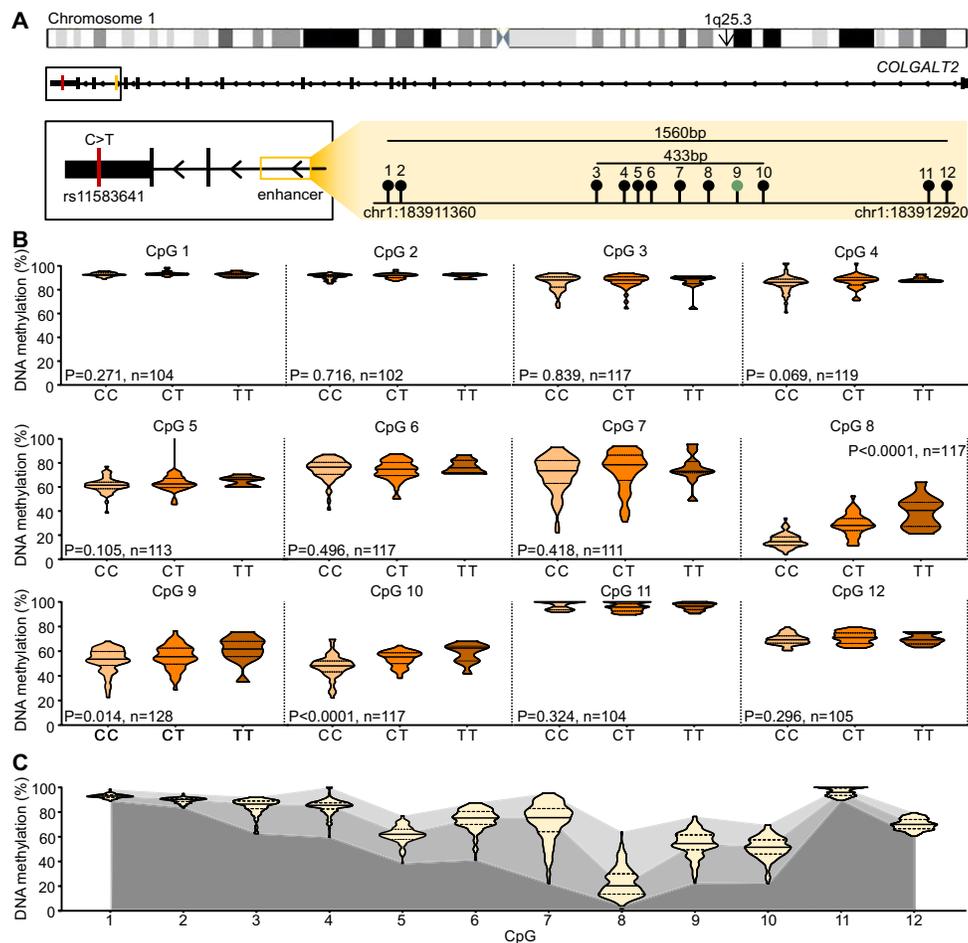
## RESULTS

### Investigation of OA-associated methylation QTLs.

The region surrounding cg18131582, at which an OA methylation QTL has previously been identified, is a putative intronic enhancer, located 6 kb upstream of the association SNP, rs11583641 (Figure 1A). We have identified that cg18131582 resides in a cluster of 8 CpGs spanning 433 bp (CpGs 3–10) (Figure 1A). To determine the physical limits of the differentially methylated region, we initially focused our attention on this cluster, along with the 4 most proximal CpGs flanking the region (CpG1–2 and CpG11–12) (Figure 1A).

We quantified DNA methylation at the 12 CpGs in 3 cartilage sample types: OA knee, OA hip, and femoral neck fracture (non-OA controls). In all samples, DNA methylation was stratified by rs11583641 genotype (Figure 1B). Significant methylation QTLs were identified at 3 of 12 CpGs: CpG8 ( $P < 0.0001$ ), CpG9 (cg18131582) ( $P = 0.014$ ), and CpG10 ( $P < 0.001$ ). This defined the limits of the differentially methylated region in arthroplasty cartilage to a 210-bp region. At the 3 CpGs, the effect allele, C, of rs11583641 was associated with lower levels of DNA methylation as compared to that in the presence of the non-risk allele, T (Figure 1B).

We analyzed median DNA methylation levels across the region (Figure 1C). At 8 of 12 CpGs, DNA was hypermethylated (median >75%). At CpGs 8–10 the median DNA methylation level was lower (20.3–54.5%), while interindividual variability was high (64% DNA methylation range at CpG8). This provides further evidence for the regulatory function of the CpG cluster, specifically the 210-bp differentially methylated region.



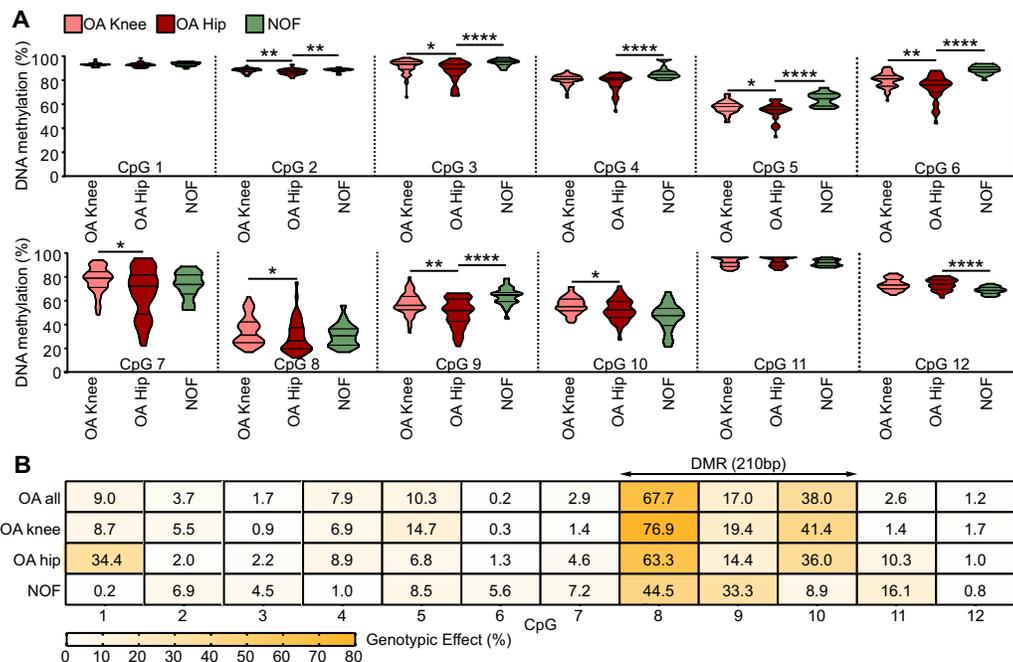
**Figure 1.** Cartilage methylation quantitative trait locus analysis of 12 CpG sites within the putative *COLGALT2* enhancer. **A**, A schematic representation of chromosome 1 shows the location of the *COLGALT2* gene at 1q25.3 (arrow). The osteoarthritis-associated single-nucleotide polymorphism rs11583641 within the 3'-untranslated region of *COLGALT2* and the putative intronic enhancer region containing cg18131582 are shown. The locations of the 12 investigated CpGs are indicated. Circles represent CpGs (CpGs 1–12); cg18131582 is CpG9 (green circle). **B**, Violin plots show DNA methylation values in arthroplasty cartilage samples ( $n = 102$ – $128$ ) at the 12 investigated CpGs, stratified by genotype at rs11583641.  $P$  values were calculated by Kruskal-Wallis nonparametric test. **C**, DNA methylation levels were measured in all arthroplasty cartilage samples ( $n = 102$ – $128$ ) at each of the 12 CpGs across the enhancer, plotted irrespective of genotype. In **B** and **C**, solid and dashed horizontal lines represent the median and interquartile range. Shaded gradients in **C** represent the upper and lower range of data points at each CpG.

### Impact of joint site and disease state on DNA methylation.

We investigated the effect of the joint site on DNA methylation in OA. At 8 of 12 CpGs, DNA methylation levels were significantly lower ( $P < 0.05$ ) in hip cartilage ( $n = 33$ – $43$ ) than in knee cartilage ( $n = 44$ – $55$ ) (Figure 2A).

Next, we tested for a disease-specific effect on DNA methylation in hip samples. At 6 of 12 CpGs, higher median DNA methylation levels were measured in femoral neck fracture samples ( $n = 20$ – $27$ ) compared to OA hip samples ( $P < 0.01$ ) (Figure 2A). The greatest differences between median values (12.0% and 12.5%, respectively) were observed at CpG6 ( $P = 4.5 \times 10^{-13}$ ) and CpG9 ( $P = 8.5 \times 10^{-9}$ ). The exception to this trend was CpG12, the most distal from the cluster, at which femoral neck fracture DNA methylation was lower ( $P = 9.2 \times 10^{-5}$ ) than in OA hip samples (Figure 2A).

We stratified DNA methylation data from sample type groups by rs11586341 genotype (Figure 2B). We used a linear regression analysis to calculate the percentage contribution of genotype to differences in DNA methylation in each of the 4 groups: all OA (hip and knee), OA knee, OA hip, and femoral neck fracture. Within the 210-bp differentially methylated region, the strongest effect was measured at CpG8 in all sample types, and the greatest genotype effect was seen in OA knee samples (76.9%) (Figure 2B). Indeed, the effect was stronger in knee cartilage than hip cartilage at all 3 differentially methylated region CpGs (Figure 2B). At CpG8 and CpG10, the genotype effects were significantly greater in all OA disease groups compared to femoral neck fracture controls (for CpG8, 63.3–76.9% in OA groups versus 44.5% in controls; for CpG10, 36.0–41.4% in OA groups versus 8.9% in controls).



**Figure 2.** Effects of joint site and disease status on DNA methylation. **A**, DNA methylation levels at the 12 investigated CpGs in cartilage samples from patients with knee osteoarthritis (OA) ( $n = 47-60$ ), hip OA ( $n = 33-44$ ), or femoral neck fracture (NOF) ( $n = 20-27$ ). Solid and dashed horizontal lines represent the median and interquartile range.  $* = P < 0.05$ ;  $** = P < 0.01$ ;  $**** = P < 0.0001$ , calculated by Mann-Whitney test. **B**, Heatmap showing the influence of the rs11583641 genotype on DNA methylation levels 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 femoral neck fracture ( $n = 20-27$ ). Values ( $r^2$ ) were determined by linear regression analysis. DMR = differentially methylated region.

Interestingly, at CpG9, the genotype effect was greatest in femoral neck fracture samples (33.3%) (Figure 2B).

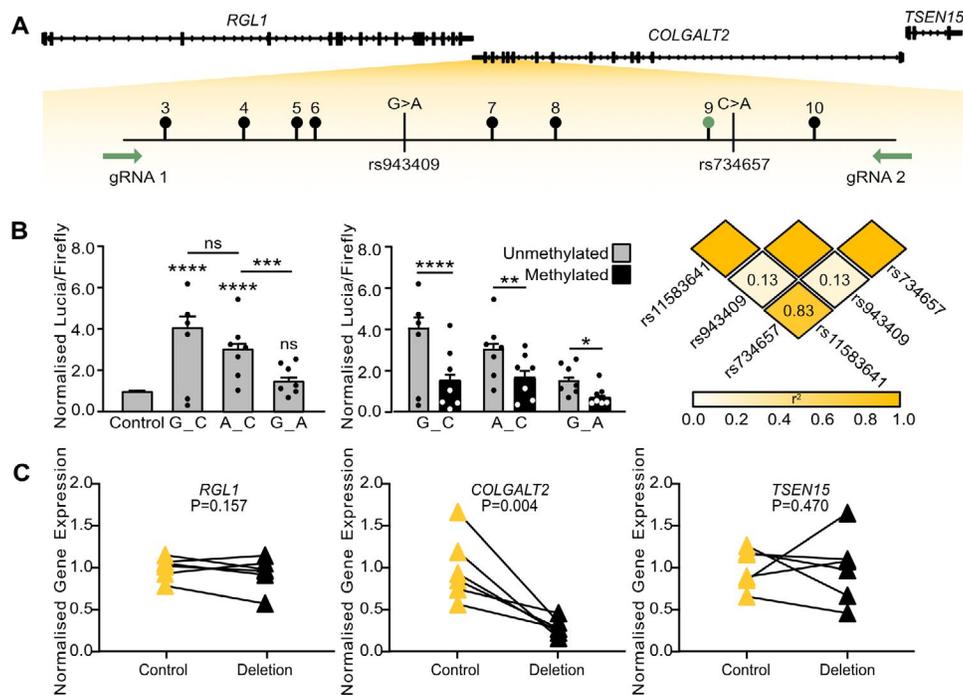
**Role of the region as a COLGALT2 enhancer.** We next tested the region for regulatory activity using a reporter gene assay. For this assay, and for subsequent investigations reported in this study, we chose to focus on the 8 CpGs within the cg18131582 cluster. A 503-bp region encompassing CpGs 3–10 was cloned into a Lucia reporter gene vector (Figure 3A). The cloned region also contained 2 SNPs, rs943409 (G>A) and rs734657 (C>A), the latter of which is in high linkage disequilibrium with rs11583641 ( $r^2 = 0.83$  for the British in England and Scotland population) (Figure 3B). Genotypes at rs943409 and rs734657 naturally occur in 3 haplotypes: G\_C, A\_C, and G\_A. All 3 haplotypes were tested for their impact on enhancer activity.

Two of the three constructs had increased enhancer activity ( $P < 0.001$ ) in Tc28a2 chondrocytes (Figure 3B). Both of these haplotypes contained the rs734657 C allele (G\_C and A\_C) and both showed a level of enhancer activity that was 3.1–4.1-fold above that of the control construct. No significant difference ( $P > 0.05$ ) was observed between the activity of these 2 constructs (Figure 3B). The G\_A haplotype conferred significantly lower activity ( $P < 0.001$ ) (Figure 3B), indicating that the rs734757 genotype impacts enhancer function, whereby the C allele (corresponding to the OA effect allele, C, at rs11583641) increases the enhancer activity of the region.

We found no evidence that the rs943409 genotype influenced activity. Methylation of the enhancer region significantly reduced the enhancer activity of all 3 haplotype constructs ( $P < 0.05$ ) (Figure 3B, middle panel).

Next, we investigated the gene target of the enhancer in cartilage. Using CRISPR-Cas9 and paired gRNAs targeting upstream and downstream of the region, we deleted 483 bp of the Tc28a2 genome, encompassing CpGs 3–10 (Figure 3A). The deletion resulted in a 2.7-fold reduction in COLGALT2 expression ( $P = 0.004$ ) (Figure 3C). No significant difference ( $P > 0.16$ ) in the expression of RGL1 or TSEN15, which flank COLGALT2, was detected (Figure 3C). To ensure that the intronic deletion did not affect COLGALT2 splicing, PCR amplification of cDNA from control and deletion cells was performed using primers spanning the deletion. No aberrant splicing was observed (Supplementary Figure 2A, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>). These results confirm that COLGALT2 is a gene target of the enhancer in chondrocytes.

Previous studies have demonstrated compensatory expression of COLGALT1 following COLGALT2 knockdown in U2OS cells (23). In Tc28a2 cells, no subsequent change in COLGALT1 expression was observed following deletion of the COLGALT2 enhancer (Supplementary Figure 2B, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>).



**Figure 3.** Effect of DNA methylation at the CpG cluster and investigation of enhancer activity in the region. **A**, A schematic representation of the location of *COLGALT2*, *RGL1*, and *TSEN15*. Circles represent the 8 CpGs (CpGs 3–10) that were captured; cg18131582 is CpG9 (green circle). Vertical lines represent the genomic positions of the 2 enhancer single-nucleotide polymorphisms (SNPs). Positions of the guide RNAs (1 and 2) used for the deletion of the region are indicated by **arrows**. **B**, Lucia reporter assays assessing enhancer activity in chondrocytes in the presence of constructs containing the 3 haplotypes of rs943409 and rs734657 (left) and the 3 haplotypes in a methylated or unmethylated state (middle); values were normalized to those in an empty vector control. Symbols represent individual samples ( $n = 7$  per group); bars show the mean  $\pm$  SEM. \* =  $P < 0.05$ ; \*\* =  $P < 0.01$ ; \*\*\* =  $P < 0.001$ ; \*\*\*\* =  $P < 0.0001$ , by Kruskal-Wallis test with Dunn's test for correction (left panel) or by multiple  $t$ -tests with the Holm-Sidak test for correction (middle panel). Linkage disequilibrium ( $r^2$ ) values between rs11583641 and the enhancer SNPs rs943409 and rs734657 in the British in England and Scotland population are shown (right). **C**, *RGL1*, *COLGALT2*, and *TSEN15* gene expression following deletion of the enhancer in chondrocytes; values were normalized to those measured in control (unedited) cells.  $P$  values were calculated by paired  $t$ -test ( $n = 6$ ).

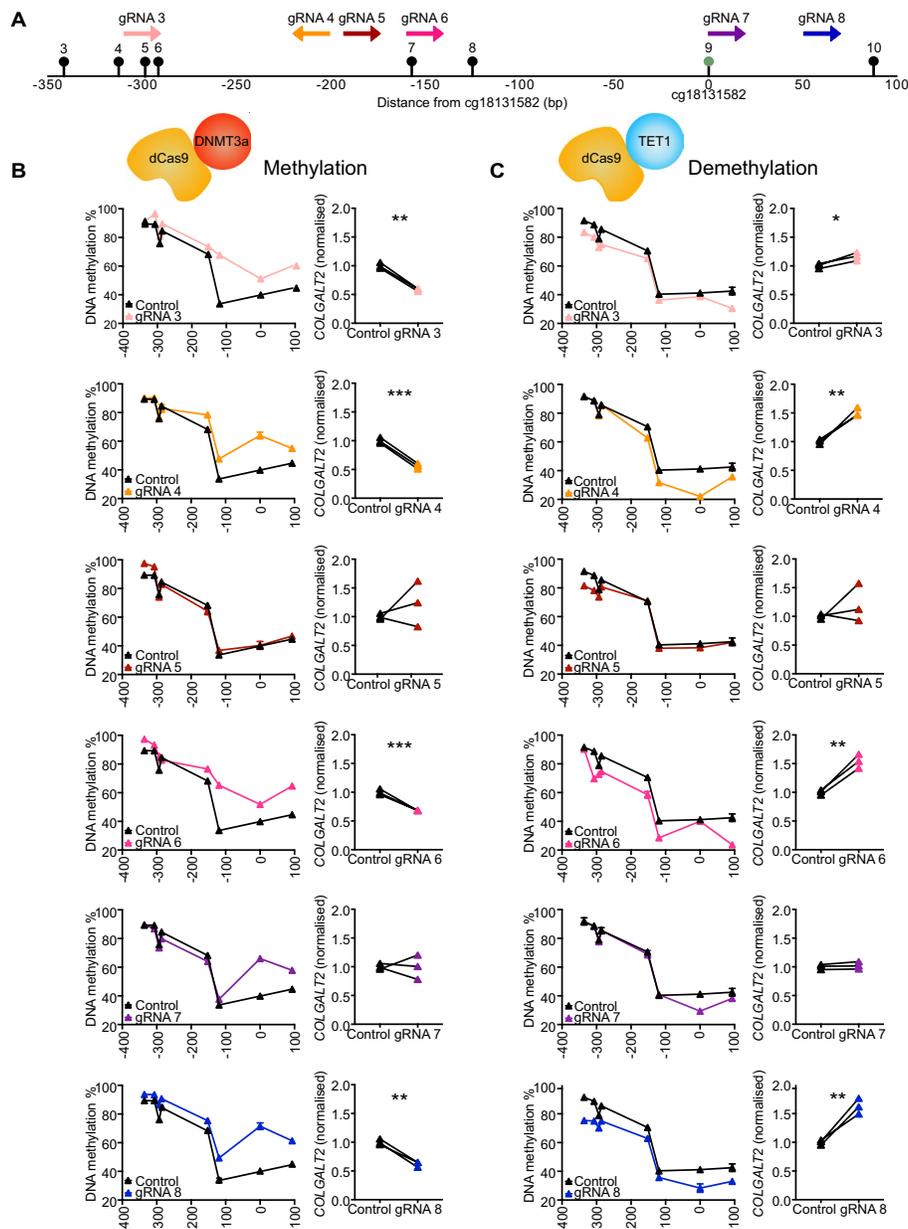
### Regulation of *COLGALT2* expression by enhancer methylation.

Next, we investigated whether enhancer DNA methylation impacts *COLGALT2* expression. We used dCas9 modulators of CpG methylation in Tc28a2 chondrocytes: dCas9-DNMT3a (to methylate the CpGs) or dCas9-TET1 (to demethylate the CpGs). Six gRNAs (gRNA3–8) were designed, which tiled across the region to target the enhancer CpGs (Figure 4A). We expressed individual gRNAs in cells along with the dCas9 constructs.

Coexpression of dCas9-DNMT3a and gRNAs successfully increased DNA methylation at the CpGs (Figure 4B). The greatest increases in DNA methylation levels were reached at CpG8, CpG9, and CpG10, at which distinct gRNAs increased methylation by up to 38.0%, 37.5%, and 23.0%, respectively (Supplementary Figure 3, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>). No further increase in DNA methylation level was reached by expression of multiple gRNAs (Supplementary Figure 4A, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>). A significant reduction in *COLGALT2* ( $P < 0.01$ ) was measured

following coexpression of dCas9-DNMT3a with gRNA3 (0.58-fold decrease), gRNA4 (0.56-fold decrease), gRNA6 (0.68-fold decrease), or gRNA8 (0.62-fold decrease) (Figure 4B). These 4 gRNAs all edited DNA methylation at CpG8–10 (Supplementary Figure 3, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>). Of note, increasing DNA methylation at CpG9 and CpG10 alone (gRNA7) did not significantly decrease expression of *COLGALT2* (Figure 4B). In all instances, there were no significant changes ( $P > 0.05$ ) in the expression of *RGL1* and *TSEN15* (Supplementary Figure 4B).

A reduction in enhancer DNA methylation was achieved using dCas9-TET1 (Figure 4C). Decreasing DNA methylation resulted in an increase in *COLGALT2* expression (Figure 4C). As above, significant changes in gene expression resulted when gRNAs 3, 4, 6, and 8 were expressed in the cells. The mean DNA methylation levels at CpG8, CpG9, and CpG10 were reduced by up to 11.8% (with gRNA6), 19.2% (with gRNA4), and 18.7% (with gRNA6), respectively (Supplementary Figure 3, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41738/abstract>). As previously

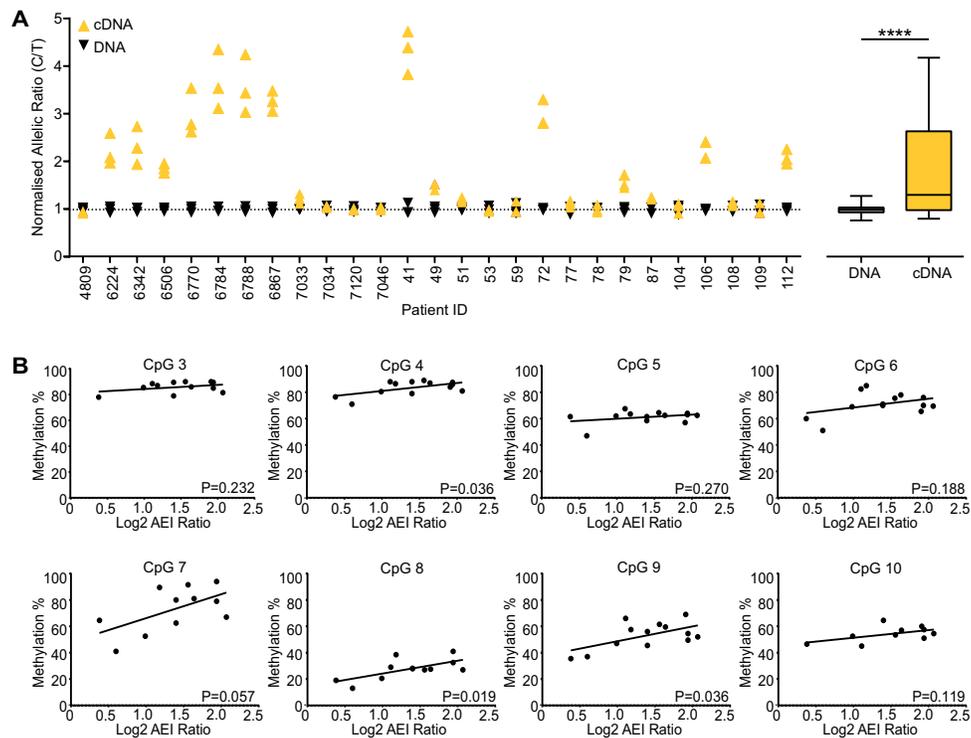


**Figure 4.** Epigenetic modulation of the enhancer in Tc28a2 cells. **A**, Schematic diagram showing the genomic position of the 6 guide RNAs (gRNAs 3–8), relative to the 8 CpGs; cg18131582 is CpG9 (green circle). **B**, Left, DNA methylation levels at the 8 CpGs in Tc28a2 chondrocytes following expression of DNA methyltransferase 3a (DNMT3a)–dead Cas9 (dCas9) protein in controls (no gRNA) or in samples with a targeting gRNA ( $n = 3$ ). Right, *COLGALT2* expression in chondrocytes following editing of DNA methylation with gRNAs. Values were normalized to the mean values in no-gRNA control cells (each  $n = 3$ ). **C**, Left, DNA methylation levels at the 8 CpGs in Tc28a2 chondrocytes following expression of dCas9-TET1 protein in controls or in samples with a targeting gRNA ( $n = 3$ ). Right, *COLGALT2* expression in control cells following editing of DNA methylation with gRNAs. Values were normalized to the mean values in no-gRNA control cells (each  $n = 3$ ). In **B** and **C**, \* =  $P < 0.05$ ; \*\* =  $P < 0.01$ ; \*\*\* =  $P < 0.001$ , by a 1-tailed paired *t*-test.

observed, modulation of DNA methylation levels did not lead to significant changes in *RGL1* or *TSEN15* expression (Supplementary Figure 4B).

In both experiments, targeted changes in DNA methylation at CpG8 and CpG10 resulted in changes in *COLGALT2* expression. These results complement the observations made using patient samples and confirm the importance of the methylation status of CpG8 and CpG10 in the regulatory function of the enhancer.

**Correlation of the genotype at rs11583641 with *COLGALT2* expression.** Finally, we returned our attention to the hip or knee cartilage samples from human patients in order to investigate the effect of the genotype at rs11583641 on allelic expression of *COLGALT2*. Allelic expression imbalance analysis revealed an imbalance between the C and T transcripts of *COLGALT2* in heterozygous patients (Figure 5A). A 1.94-fold mean increase ( $P < 0.0001$ ) in expression of the OA effect allele, C,



**Figure 5.** Allelic expression imbalance (AEI) analysis of rs11583641 and *COLGALT2*. **A**, Left, Allelic (C/T) ratios in cartilage samples from osteoarthritis (OA) patients heterozygous for rs11583641 ( $n = 27$ ). In each sample, the ratio of values for cDNA and DNA between the C (OA effect) allele and 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 box plots, with the line inside the box representing the median, boxes showing the interquartile range, and whiskers showing the minimum and maximum values. \*\*\*\* =  $P < 0.0001$  by Wilcoxon's matched pairs signed rank test (1-tailed). **B**, Allelic ratios ( $\log_2$ ) of *COLGALT2* plotted against matched DNA methylation level values at CpGs 3–10 ( $n = 11$ –13).  $P$  values were determined by linear regression analysis.

was observed. Interestingly, an allelic expression imbalance was detected in 13 of 27 samples used for the analysis, yet in the remaining 14 samples no imbalance was found (Figure 5A). This observation could not be explained by sex, age, disease, or joint differences between the patients.

In the cartilage samples that showed significant allelic expression imbalance, allelic ratios correlated with DNA methylation at the enhancer CpGs (Figure 5B). Significant correlations marking methylation-expression QTLs were discovered at CpG4, CpG8, and CpG9 ( $P = 0.036$ ,  $P = 0.019$ , and  $P = 0.036$ , respectively), and a trend was emerging at CpG7 ( $P = 0.057$ ) (Figure 5B), supporting OA-associated regulation of *COLGALT2*, mediated by genetic and epigenetic interplay in cartilage.

## DISCUSSION

The GWAS era has spanned more than a decade and has resulted in the identification of >90 independent OA genetic association signals. However, it has proven challenging to biologically interpret these results and translate genetic discoveries into effective therapies (22). Consequently, epigenetic data sets are increasingly being applied post-GWAS to prioritize causal genes and

their regulatory elements for focused analysis (10,11,18,20). In the current study, we performed a focused analysis of one such risk locus, on chromosome 1, at which intronic DNA methylation has previously been correlated with OA genetic risk (20). We applied a broad range of molecular biology tools in cartilage samples and in a chondrocyte cell model, to characterize the genetic and epigenetic factors regulating gene expression at the locus.

In patient arthroplasty samples, we analyzed DNA methylation at 12 CpGs and defined the differentially methylated region as spanning 210 bp and containing 3 CpGs (CpGs 8–10). The rs11583641 effect allele, C, corresponded to a reduction in DNA methylation levels compared to levels in the non-effect allele, T. At the 3 CpGs, the strongest effect size was observed in knee cartilage. Allelic expression imbalance analysis confirmed that the OA effect allele of rs11583641 correlated with increased expression of *COLGALT2*. Furthermore, correlations between allelic expression imbalance ratios and DNA methylation revealed a methylation-expression QTL.

The region was shown to have regulatory function in vitro, which was significantly hindered by DNA methylation. We identified a genetic influence on enhancer function, whereby the major allele of rs734657, located within the enhancer and corresponding to the effect allele at rs11583641, resulted in increased Lucia

expression. These in vitro findings demonstrate a synergistic relationship between genetic and epigenetic factors in the regulation of gene expression. The effect allele at OA-associated SNPs correlates with an increase in gene expression.

The advent of CRISPR-Cas9 and subsequent development of the Cas9 toolbox has revolutionized targeted editing of the genome and epigenome (29). CRISPR-Cas9 deletion of the enhancer confirmed that *COLGALT2* was a target gene, while gene expression of *RGL1* and *TSEN15*, which flank *COLGALT2*, remained unchanged. The development of a catalytically dead Cas9 fused to enzymes that either methylate (DNMT3a) or demethylate (TET1) CpGs has provided an elegant tool for precision editing of DNA methylation (30). Here, we have applied this technology in human immortalized chondrocytes and demonstrated a causal relationship between DNA methylation and *COLGALT2* expression. Concordant with the measurements of DNA methylation and gene expression in cartilage samples, a reduction in DNA methylation resulted in increased *COLGALT2* expression. Furthermore, the results of this investigation confirmed that the 3 differentially methylated region CpGs, particularly CpG8 and CpG10, are the functional mediators.

*COLGALT2* is a compelling candidate gene in the etiology of OA. Collagens are a major constituent of the extracellular matrix of articular cartilage, the central tissue in OA pathology. Two enzymes are required for collagen glycosylation: procollagen galactosyltransferase 1 and 2, which are encoded by *COLGALT1* and *COLGALT2*, respectively. *COLGALT1* mutations cause the collagen deficiency condition known as cerebral small vessel disease (31), and a recent genetic investigation identified a rare coding variant in *COLGALT1* as a potential cause of erosive hand OA (32). Polymorphisms in *COLGALT2* have been associated with height, body fat distribution, and schizophrenia (33–35), in addition to OA (11). The many associations between the disruption of collagen posttranslational modification and diseases, including osteogenesis imperfecta and Bruck syndrome, provide evidence that aberrant posttranslational modification can be detrimental to cartilage integrity (36–38).

Taking into consideration all our observations in both analyses of human tissue and in vitro cartilage models, we propose the following model: the OA effect allele marked by the SNP rs11583641 mediates decreased cartilage DNA methylation at the *COLGALT2* enhancer, resulting in increased expression of the gene and a subsequent increase in galactosyltransferase activity. We hypothesize that resulting over-modification of the collagen triple helix could detrimentally affect the structural or mechanotransductive properties of articular cartilage, contributing to OA. This requires further investigation.

Using novel epigenome modulating tools, we established a causal relationship between enhancer DNA methylation and *COLGALT2* expression. The absence of *COLGALT2* allelic expression imbalance in a proportion of the heterozygous patients in the study, along with the evidence that the genotype

at rs734657 contributes to regulatory function, indicates that the genotype at multiple SNPs in the region could function in concert to finely tune *COLGALT2* expression. This also requires further investigation.

Our understanding of the interplay between genetics and epigenetics in OA is rapidly developing (16). The discovery of a role for epigenetics in musculoskeletal diseases exposes the human epigenome as an exploitable pharmacologic target. To enable this, complete knowledge of the molecular mechanisms underpinning pathogenic subtypes is required. A precision medicine approach to treat musculoskeletal disorders requires a deeper understanding of the relationship between genetics and disease, as well as stratification of patients based on the underlying biology (39). In OA, there is increasing evidence for different molecular pathways underlying disease subtypes, ultimately presenting with the same endotype but requiring distinct therapeutic approaches (40,41). The identification of OA risk genes, such as *COLGALT2*, provides scope for novel pharmacologic interventions within patient subgroups.

## 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.

## 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. Dr. Rice had full access to all of 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, Loughlin, Rice.

**Acquisition of data.** Kehayova, Watson, Rice.

**Analysis and interpretation of data.** Kehayova, Rice.

## REFERENCES

1. Iannone F, Lapadula G. The pathophysiology of osteoarthritis [review]. *Aging Clin Exp Res* 2003;15:364–72.
2. Wang H, Bai J, He B, Hu X, Liu D. Osteoarthritis and the risk of cardiovascular disease: a meta-analysis of observational studies. *Sci Rep* 2016;6:39672.
3. Wilkie R, Parmar SS, Blagojevic-Bucknall M, Smith D, Thomas MJ, Seale BJ, et al. Reasons why osteoarthritis predicts mortality: path analysis within a Cox proportional hazards model. *RMD Open* 2019;5:e001048.
4. Palazzo C, Nguyen C, Lefevre-Colau MM, Rannou F, Poiraudou S. Risk factors and burden of osteoarthritis. *Ann Phys Rehabil Med* 2016;59:134–8.
5. Zeggini E, Panoutsopoulou K, Southam L, Rayner NW, Day-Williams AG, Lopes MC, et al, on behalf of the arcOGEN Consortium and arcOGEN Collaborators. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* 2012;380:815–23.
6. Evangelou E, Kerkhof HJ, Styrkarsdottir U, Ntzani EE, Bos SD, Esko T, et al. A meta-analysis of genome-wide association studies

- identifies novel variants associated with osteoarthritis of the hip. *Ann Rheum Dis* 2014;73:2130–6.
7. Betancourt MC, Cailotto F, Kerkhof HJ, Cornelis FM, Doherty SA, Hart DJ, 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 U S A* 2012;109:8218–23.
  8. Styrkarsdottir U, Thorleifsson G, Helgadóttir HT, Bomer N, Metrustry S, Bierma-Zeinstra S, et al. Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat Genet* 2014;46:498–502.
  9. Styrkarsdottir U, Lund SH, Thorleifsson G, Zink F, Stafansson 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.
  10. Zengini E, Hatzikotoulas K, Tachmazidou I, Steinberg J, Hartwig FP, Southam L, et al. Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. *Nat Genet* 2018;50:549–58.
  11. 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 [letter]. *Nat Genet* 2019;51:230–6.
  12. Den Hollander W, Boer CG, Hart DJ, Yau MS, Ramos YF, Metrustry S, et al. Genome-wide association and functional studies identify a role for matrix Gla protein in osteoarthritis of the hand. *Ann Rheum Dis* 2017;76:2046–53.
  13. Casalone E, Tachmazidou I, Zengini E, Hatzikotoulas K, Hackinger S, Suveges D, et al. A novel variant in GLIS3 is associated with osteoarthritis. *Ann Rheum Dis* 2018;77:620–3.
  14. Styrkarsdottir U, Helgason H, Sigurdsson A, Norddahl GL, Agustsdóttir AB, Reynard LN, et al. Whole-genome sequencing identifies rare genotypes in COMP and CHADL associated with high risk of hip osteoarthritis. *Nat Genet* 2017;49:801–5.
  15. Yau MS, Yerges-Armstrong LM, Liu Y, Lewis CE, Duggan DJ, Renner JB, et al. Genome-wide association study of radiographic knee osteoarthritis in North American Caucasians. *Arthritis Rheumatol* 2017;69:343–51.
  16. Rice SJ, Beier F, Young DA, Loughlin J. Interplay between genetics and epigenetics in osteoarthritis [review]. *Nat Rev Rheumatol* 2020;16:268–81.
  17. Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function [review]. *Am J Hum Genet* 2018;102:717–30.
  18. Boer CG, Yau MS, Rice SJ, de Almeida RC, Cheung K, Styrkarsdottir U, et al. Genome-wide association of phenotypes based on clustering patterns of hand osteoarthritis identify WNT9A as novel osteoarthritis gene. *Ann Rheum Dis* 2020;80:367–75.
  19. 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.
  20. 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.
  21. Rushton MD, Reynard LN, Young DA, Shepherd C, Aubourg G, Gee F, et al. Methylation quantitative trait locus analysis of osteoarthritis links epigenetics with genetic risk. *Hum Mol Genet* 2015;24:7432–44.
  22. Hannon E, Gorrie-Stone TJ, Smart MC, Burrage J, Hughes A, Bao Y, et al. Leveraging DNA-methylation quantitative-trait loci to characterize the relationship between methylomic variation, gene expression, and complex traits. *Am J Hum Genet* 2018;103:654–65.
  23. Schegg B, Hülsmeier AJ, Rutschmann C, Maag C, Hennet T. Core glycosylation of collagen is initiated by two  $\beta(1-O)$ galactosyltransferases. *Mol Cell Biol* 2009;29:943–52.
  24. Webster JA, Yang Z, Kim YH, Loo D, Mosa RM, Li H, et al. Collagen  $\beta(1-O)$  galactosyltransferase 1 (GLT25D1) is required for the secretion of high molecular weight adiponectin and affects lipid accumulation. *Biosci Rep* 2017;37:20170105.
  25. Bann JG, Peyton DH, Bächinger HP. Sweet is stable: glycosylation stabilizes collagen. *FEBS Lett* 2000;473:237–40.
  26. Southam L, Rodriguez-Lopez J, Wilkins JM, Pombo-Suarez M, Snelling S, Gomez-Reino JJ, et al. An SNP in the 5'-UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage. *Hum Mol Genet* 2007;16:2226–32.
  27. Rice SJ, Aubourg G, Sorial AK, Almarza D, Tselepi M, Deehan DJ, et al. Identification of a novel, methylation-dependent, RUNX2 regulatory region associated with osteoarthritis risk. *Hum Mol Genet* 2018;27:3464–74.
  28. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  method. *Methods* 2001;25:402–8.
  29. Parker E, Hofer IM, Rice SJ, Earl L, Anjum SA, Deehan DJ, 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.
  30. Adli M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* 2018;9:1–13.
  31. Miyatake S, Schneeberger S, Koyama N, Yokochi K, Ohmura K, Shiina M, et al. Biallelic COLGALT1 variants are associated with cerebral small vessel disease. *Ann Neurol* 2018;84:843–53.
  32. Juryneć M, Grunwald D, Novak K, Honegger M, Hoshijima K, Kazmers N. Discovery and functional analysis of osteoarthritis susceptibility genes using pedigrees from a statewide population-based cohort [abstract]. *Osteoarthritis Cartilage* 2020;28:S344.
  33. Okada Y, Kamatani Y, Takahashi A, Matsuda K, Hosono N, Ohmiya H, et al. A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. *Hum Mol Genet* 2010;19:2303–12.
  34. Rask-Andersen M, Karlsson T, Ek WE, Johansson Å. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nat Commun* 2019;10:339.
  35. Drago A, Fischer EK. A molecular pathway analysis informs the genetic risk for arrhythmias during antipsychotic treatment. *Int Clin Psychopharmacol* 2018;33:1–14.
  36. Ha-Vinh R, Alanay Y, Bank RA, Campos-Xavier AB, Zankl A, Superti-Gurga A, 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 Med Genet A* 2004;131:115–20.
  37. Cabral WA, Chang W, Barnes AM, Weis M, Scott MA, Leikin S, et al. Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta [letter]. *Nat Genet* 2007;39:359–65.
  38. Cundy T, Dray M, Delahunt J, Hald JD, Langdahl B, Li C, et al. Mutations that alter the carboxy-terminal-propeptide cleavage site of the chains of type I procollagen are associated with a unique osteogenesis imperfecta phenotype. *J Bone Miner Res* 2018;33:1260–71.
  39. Delude CM. Deep phenotyping: the details of disease. *Nature* 2015;527:S14–5.
  40. Maumus M, Noel D, Ea HK, Moulin D, Ruiz M, Hay E, et al. Identification of TGF $\beta$  signatures in six murine models mimicking different osteoarthritis clinical phenotypes. *Osteoarthritis Cartilage* 2020;28:1373–84.
  41. Mobasher A, Saarakkala S, Finnilä M, Karsdal MA, Bay-Jensen AC, van Spil WE. Recent advances in understanding the phenotypes of osteoarthritis [review]. *F1000Res* 2019;8:2091.