Transcriptome regulation analysis in osteoarthritis primary tissues identifies high-confidence effector genes

Authors:

Georgia Katsoula[1,2,3,*], Ana Luiza Arruda[2,*], Mauro Tutino[2], Peter Kreitmaier[1,2,3], Karan M Shah[4], Diane Swift[4], Lorraine Southam[2], J Mark Wilkinson[4,×], Eleftheria Zeggini[2,3,×]

[1] Technical University of Munich (TUM), School of Medicine and Health, Graduate School of Experimental Medicine, Munich, 81675, Germany

[2] Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, 85764 Neuherberg, Germany

[3] Technical University of Munich (TUM) and Klinikum Rechts der Isar, TUM School of Medicine and Health, 81675 Munich, Germany

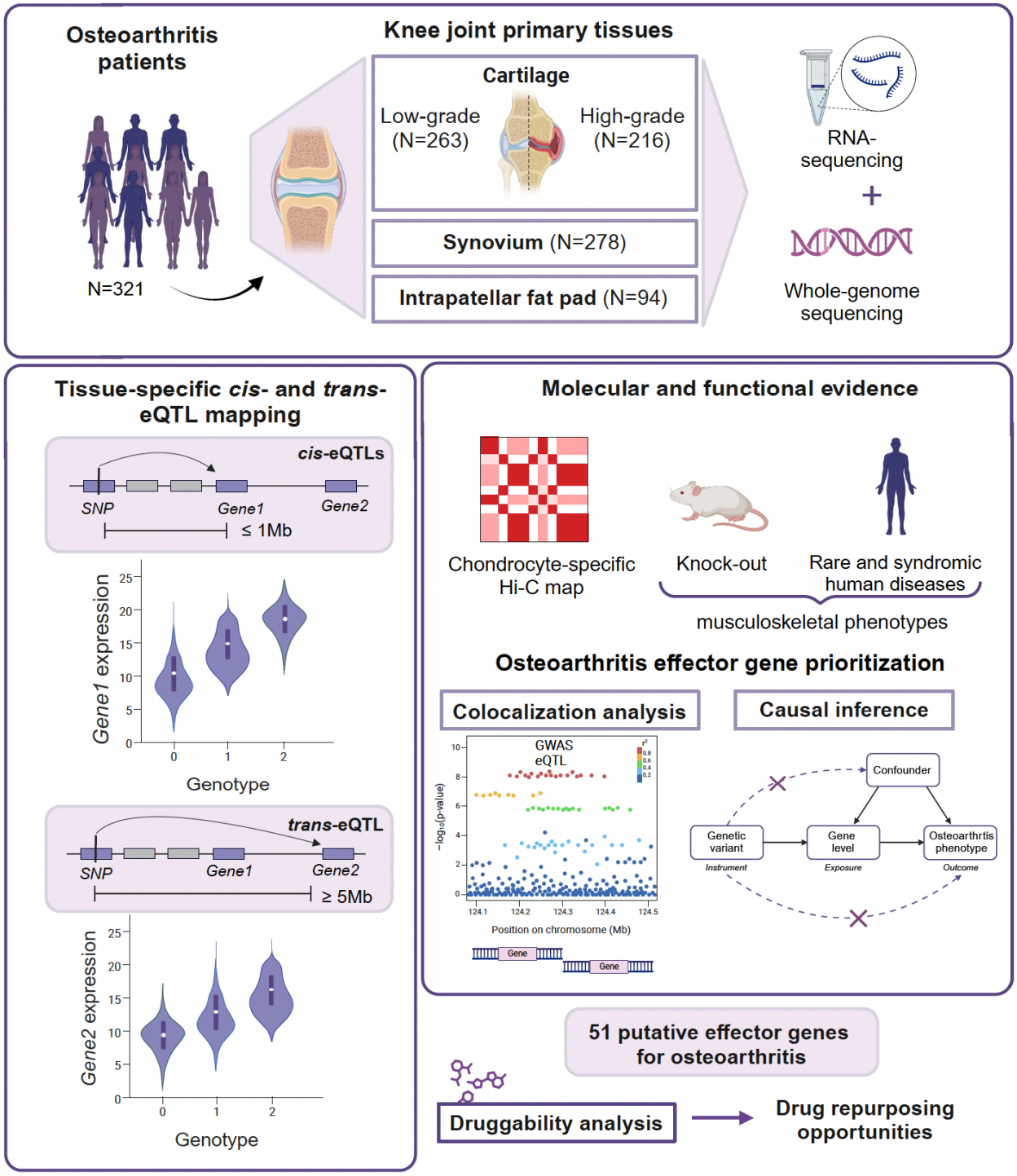
[4] School of Medicine and Population Health, University of Sheffield, Beech Hill Road, Sheffield, S10 2RX, United Kingdom

* equal contribution
× equal contribution

Summary

Osteoarthritis, a whole-joint degenerative disorder, is a major public health burden that affects nearly 600 million individuals worldwide. Molecular profiling of relevant tissues is crucial for understanding the biological mechanisms underlying disease development. We present the first comprehensive map of transcriptional regulation of disease-relevant primary tissues from knee osteoarthritis patients: macroscopically intact (low-grade, N=263) and degenerated (high-grade, N=216) cartilage, synovium (N=278), and fat pad (N=94). Of 9,738 unique expression quantitative trait loci (eQTL)-associated genes, 60.6% have not been reported previously. We identify 51 genes with directional evidence of tissue-specific colocalization and causal association with knee osteoarthritis, 25 of which are new high-confidence effector genes for osteoarthritis (including *MUSTN1*, which is implicated in cartilage integrity). Our study provides valuable insights into the tissue-specific and shared molecular mechanisms underlying osteoarthritis and offers drug repurposing opportunities.

Graphical abstract



Introduction

Osteoarthritis, the most prevalent form of arthritis, is a chronic degenerative complex disease that is projected to affect over 1 billion individuals worldwide by 2050[1]. Osteoarthritis is characterized by progressive breakdown of articular cartilage and low-grade inflammation of the

surrounding tissues[2,3], leading to pain, stiffness, and reduced mobility[4]. It most commonly affects the knee, hand, and hip joints[5]. In addition to risk factors such as age, obesity, and joint injury, osteoarthritis also has a genetic risk component. Currently, there is no effective disease-modifying treatment for osteoarthritis, and strategies focus on pain management and joint replacement surgery. Therefore, there is an urgent need for an in-depth understanding of the processes underpinning osteoarthritis to foster the development of new therapies. Genome-wide association studies (GWAS) have identified over 100 independent risk variants for osteoarthritis[6], the majority of which reside in non-coding sequences. The key challenge now is to identify the effector genes through which these associated variants affect osteoarthritis risk.

Osteoarthritis is a whole-joint disease, driven by a complex interplay of both shared and tissue-specific mechanisms[7]. Shared mechanisms involve inflammatory pathways, biomechanical stress, and metabolic imbalances that collectively contribute to the degeneration of articular cartilage, the primary hallmark of osteoarthritis[4]. This includes extensive changes in the extracellular matrix (ECM), including loss of proteoglycans and collagen through enzymatic cleavage, mediated by metalloproteinases and aggrecanases, and/or loss of chondrocyte ECM maintenance properties through senescence and apoptosis. In addition to cartilage degradation, osteoarthritis affects the synovium, subchondral bone, and fat pad tissues. The synovium often exhibits inflammation (synovitis) characterized by increased production of pro-inflammatory cytokines, contributing to cartilage degradation and joint inflammation[2]. Pro-inflammatory cytokines are also secreted by the adipose tissue located in the infrapatellar fat pad[8]. There is evidence that crosstalk between the synovium and fat pad could stimulate the infiltration of immune cells and the secretion of catabolic molecules across both tissues[9]. Subchondral bone undergoes abnormal remodeling and sclerosis, influenced by altered bone turnover and microdamage repair processes[10]. These interconnected pathophysiological changes underscore the importance of studying multiple primary tissues for dissecting the pathophysiology of osteoarthritis.

Expression quantitative trait loci (eQTL) capture the association between genetic variation and tissue-specific gene expression (proximal or distal)[11]. The regulatory relationships captured by eQTLs can reveal the function of disease-associated variants by identifying the genes and pathways they affect[12]. Given the tissue- and context-specificity underlying gene regulatory patterns[13], eQTL regulatory relationships are best captured in the context of the studied phenotype in the primary affected tissue(s).

There have been two *cis*-eQTL studies in osteoarthritis primary tissues to date. Steinberg et al.[14] identified *cis*-eQTLs in low-grade cartilage, high-grade cartilage, and synovium from up to 95 osteoarthritis patients of European ancestry. This study identified colocalization of eQTLs with osteoarthritis GWAS risk loci spanning five effector genes (*ALDH1A2*, *NPC1*, *SMAD3*, *FAM53A*, *SLC44A2*)[15]. A second study identified *cis*-eQTLs in the synovium of 202 Chinese Han individuals with osteoarthritis. The authors found evidence of colocalization with osteoarthritis GWAS loci for 34 genes with two, *FAM53A* and *SLC44A2*, overlapping with the European ancestry eQTL study[16]. These studies provide valuable insights into the proximal

gene regulation mechanisms in osteoarthritis primary tissue. However, a well-powered eQTL map across multiple human joint tissues, including *trans*-acting variants, is still missing.

Here, we have generated *cis*- and *trans*-eQTL maps from knee joint primary tissues of 321 osteoarthritis patients of European ancestry, including low-grade and high-grade cartilage, synovium, and fat pad. This is the first eQTL study for fat pad tissue and, for the other three tissues, the sample size here is increased by more than two-fold compared to previous studies. We identify shared and tissue-specific biological mechanisms involved in osteoarthritis and integrate this resource of gene expression molecular data with knee-related osteoarthritis GWAS to gain insights into loci associated with disease risk.

Results

Osteoarthritis primary cartilage eQTL discovery

We performed *cis*- and *trans*-eQTL analyses in 321 knee osteoarthritis patients (Table S1), achieving a significant increase in statistical power across the allele frequency spectrum (Figure S1) while also replicating previous findings (Figure S2). We identify 6,646 *cis*-eGenes (genes targeted by at least one eQTL) in low-grade cartilage and 4,639 in high-grade cartilage, including 8,378 and 5,355 conditionally independent *cis*-eQTLs, respectively (Figure 1, Table S2). We further find 363 *trans*-eQTL variants (representing 12 independent signals) for twelve genes in low-grade cartilage and 180 *trans*-eQTL (five independent) for five genes in high-grade cartilage (Supplemental notes, Table S2, Figure S3).

Tissue-specific eQTLs in low-grade cartilage regulate 1,783 genes enriched in terms related to anatomical structure development (q-value=2.11×10^{-3}) (Figure 2, Tables S3-4). Among these genes, several play distinct roles in limb development (e.g., *CTNNB1*, *GLI3*, *GNAS*, *MYH3*, *PITX1*, *GDF5*, *ALDH1A2*, *TBX4*, *LEF1*). Additionally, we find multiple genes involved in CXCR4 signaling, androgen receptor, VEGFA-VEGFR2, EGFR1, HIF-1α, TCF response to Wnt, PI3K/Akt, and Smad2/3 signaling pathways. These pathways are interconnected through signaling cascades that predominantly regulate cellular processes such as growth and differentiation. For instance, CXCR4, a G-protein-coupled chemokine receptor expressed in chondrocytes, activates several downstream pathways (PI3K/Akt, TGF-β/Smad3, NF-κB, MAPK, Wnt/β-catenin, and VEGF) upon binding its ligand SDF-1. SDF-1/CXCR4 is known to regulate chondrocyte hypertrophy[17] and is a potential therapeutic target for osteoarthritis[18]. These findings indicate that gene regulation in the early stages of knee osteoarthritis degeneration affects processes crucial for chondrocyte phenotypic stability and the initiation of chondrocyte hypertrophy.

In high-grade cartilage, tissue-specific eGenes (n=580) are enriched in actin monomer binding (q-value=6.57×10^{-3}) (Figure 2, Tables S3-4). The cytoskeleton plays a crucial role in the physical interactions between chondrocytes and their extracellular matrix, also contributing to the chondrocyte hypertrophic phenotype and biomechanical properties[19]. Cultured osteoarthritis chondrocytes have been shown to exhibit a disorganized actin cytoskeleton compared to healthy articular chondrocytes and differential expression of actin-binding proteins that are

critical to actin polymerization and hypertrophy[19–21]. This finding suggests that disruptions in actin cytoskeleton organization and actin-binding protein expression contribute to the altered chondrocyte phenotype and biomechanical dysfunctions in late-stage osteoarthritis degeneration. Among high-grade cartilage tissue-specific eGenes, we further find several involved in unfolded protein response (*SERPINH1, FICD, ATF6B, ERN1, HSPB1, DNAJC4, LMNA, PTPN2, VCP, MBTPS1, BCL2L11, DERL1, ARFGAP1, DERL2, RNF121, SEC31A, UBXN4*), highlighting the impact of endoplasmic reticulum stress in cartilage degeneration.

Osteoarthritis primary synovium eQTL discovery

In synovium, we find 6,411 *cis*-eGenes targeted by 8,043 conditionally independent *cis*-eQTLs (Figure 1, Table S2). We also find 163 *trans*-eQTL (six independent) for five genes (Supplemental notes, Table S2, Figures S3-4), two out of which are also subjected to *cis*-regulation of their expression.

Synovium-specific eGenes (n=2,308) (Figure 2) included multiple genes coding for calcium-activated cation channels (*KCNMB1*, *KCNN2*, *KCNN3*, *TMEM63A*, *TMEM63B, CATSPER2*, *ANO10*, *ANO6)* (Table S3-4). Several of these have been implicated in inflammation (*KCNMB1*, *KCNN2*, *KCNN3*)[22] and response to mechanical stress (*TMEM63A*, *TMEM63B*)[23]. Calcium-activated potassium channels have been associated with pro-invasive properties of synovial fibroblasts in rheumatoid arthritis models[24] and pain-associated synovial fibroblasts in osteoarthritis[25]. This finding suggests that these channels are involved in osteoarthritis synovial inflammation and pain perception. Among synovium-specific eGenes we further find several genes involved in the remodeling of the extracellular matrix (*MMP11, MMP19, MMP17, TIMP3, MMP24, TIMP2, MMP2, MMP9, MMP25, TIMP3, TIMP2, FURIN*). This finding underscores the importance of synovium in mediating collagen degradation in osteoarthritis cartilage and the complex interaction between these tissues. Synovium-specific *cis*-eGenes also included integral components of the JAK-STAT signaling pathway (*IL12RB1*, *JAK2*, *STAT4*, *TYK2*). The active JAK/STAT pathway contributes to osteoarthritis by regulating inflammation, immune response, mechanical loading, and apoptosis, leading to cartilage and subchondral bone degradation and synovial inflammation[26]. Therefore, the genetic regulation of JAK/STAT pathway in the osteoarthritis synovium highlights it as a putative key inflammatory mediator in disease pathogenesis.

Osteoarthritis primary fat pad eQTL discovery

In fat pad osteoarthritis tissue, we identify 592 *cis*-eGenes regulated by 606 conditionally independent *cis*-eQTLs (Figure 1, Table S2), along with one *trans*-eQTL gene pair (Supplemental notes, Table S2, Figure S3).

Among the *cis*-eGenes, 109 are specific to the fat pad (Figure 2), with significant enrichment in immune processes such as interferon gamma signaling (q-value=9.65×10^{-4}) (*HLA-DQA2, HLA-DQA1*, *HLA-DQB2*, *IRF6*, *HLA-DRB5*, *IRF9*), cytokine signaling (q-value=3.16×10^{-2}) (*HLA-DRB5*, *HLA-DQA1*, *IRF9*, *HLA-DQA2*, *KRAS*, *IRF6*, *HLA-DQB2*, *IL34*), and MHC class II antigen presentation (q-value=2.11×10^{-3}) (*HLA-DQA2*, *HLA-DQA1*, *ARF1*, *HLA-DQB2*, *HLA-DRB5*) (Table S3-4). This emphasizes the involvement of the fat pad in modulating immune

response within the joint environment. Additionally, we find enrichment for genes involved in the vesicle cellular compartment (q-value=6.58×10^{-3}) (*ARF1*, *RAB2A*, *ATP8A1*, *PRKACB*, *CLIC3*, *ARPC5*, *NME1*, *TMC4*, *ELAPOR1*, *PBLD*) (Table S3). The fat pad secretes small extracellular vesicles (sEVs) that promote chondrocyte extracellular matrix (ECM) catabolism, induce cellular senescence, and lead to osteoarthritis progression in murine models[27]. Given the crucial role of sEVs in intercellular communication, our findings suggest that sEVs propagate inflammatory signals from the fat pad to other joint tissues.

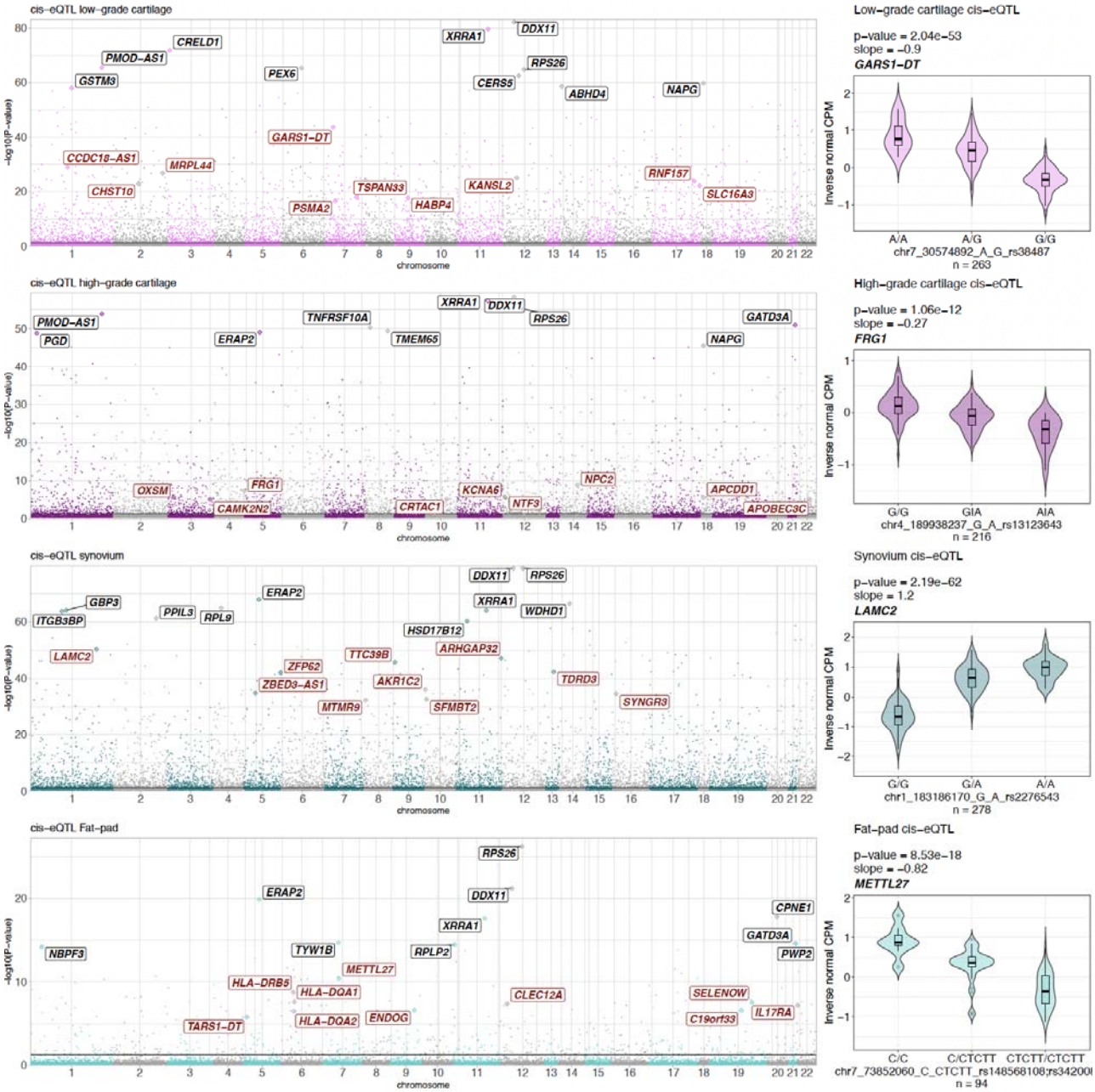


Figure 1: **Genomic distribution of *cis*-eQTLs.** Manhattan plots depicting the negative log of the permutation q-value of the most significant association per gene across all variants within 1 Mb of the gene's transcription start site for low-grade cartilage, high-grade cartilage, synovium

and fat-pad tissues. Top 10 *cis*-eGenes are labeled in black and tissue-specific top *cis*-eGenes are labeled in darkred. The violin plots show the effect of the most significant tissue-specific *cis*-eQTL on gene expression in each tissue respectively. The boxplots within the violins represent 25th, 50th, and 75th percentiles, and whiskers extend to 1.5 times the interquartile range. Nominal p-values and effect sizes (slopes) derived from *cis*-eQTL are displayed above the violin plots.

Shared *cis*-eQTL patterns across tissues reveal common disease mechanisms

To elucidate common gene regulatory mechanisms in osteoarthritis primary tissues, we evaluated *cis*-eQTL sharing patterns (Figure 2, Table S3). We find that 72.3% of eQTLs are shared by sign and magnitude (see Methods) between the four tissues. Shared *cis*-eGenes (n=3,392) are enriched for constituents of the mitochondrial matrix (q-value=3.1×10^{-2}) and endoplasmic reticulum (ER) (q-value=3.1×10^{-2}), supporting the role of mitochondrial dysfunction[28] and ER-stress[29] in knee osteoarthritis.

Low- and high-grade cartilage exhibit the largest overlap of *cis*-eQTL effects (85.9% of *cis*-eQTLs shared), suggesting that key aspects of *cis*- gene regulation are preserved across cartilage degradation stages. Shared eGenes between low- and high-grade cartilage (n=5,725) are enriched for ER constituents (q-value=2.67×10^{-4}), transport vesicle terms (q-value=7.03×10^{-4}), as well as for general metabolism pathway (q-value=2.8×10^{-3}) (Table S4). This is particularly relevant given the growing evidence that chondrocytes undergo metabolic shifts, such as mitochondrial dysfunction, enhanced anaerobic glycolysis, and alterations in lipid and amino acid metabolism, in response to environmental stressors like inflammatory microenvironments, mechanical overload, and oxidative damage[30]. The genetic regulation of these pathways in both low- and high-grade cartilage strengthens the connection between disrupted chondrocyte metabolism and osteoarthritis.

Considering the evidence that synovium and fat pad contribute to knee joint inflammation in osteoarthritis, we evaluated the presence of common regulatory processes in these two tissues. We find that synovium and fat pad share 73.6% of *cis*-eQTLs by both sign and magnitude. Shared eGenes (n=3,499) are enriched for glutathione-mediated detoxification (q-value=1.92×10^{-2}) (*MGST1*, *MGST3*, *GSTM3*, *GSTM4*, *GSTM5*, *GSTA4*, *GSTT2B*, *GSTK1*, *GSTM2*, *GSTZ1*, *ANPEP*, *GGH*), in addition to mitochondrial and ER processes (Table S4). Glutathione is an intracellular antioxidant, important for maintaining redox balance. Dysregulation of the glutathione antioxidant system has been associated with increased oxidative stress and oxidant-mediated chondrocyte death in osteoarthritis[31]. Given the beneficial effects of glutathione administration for osteoarthritis (reestablishment of the redox equilibrium; amelioration of oxidative stress and inflammation), targeting the glutathione antioxidant system in both osteoarthritis synovium and fat pad could be a promising treatment target for knee osteoarthritis.

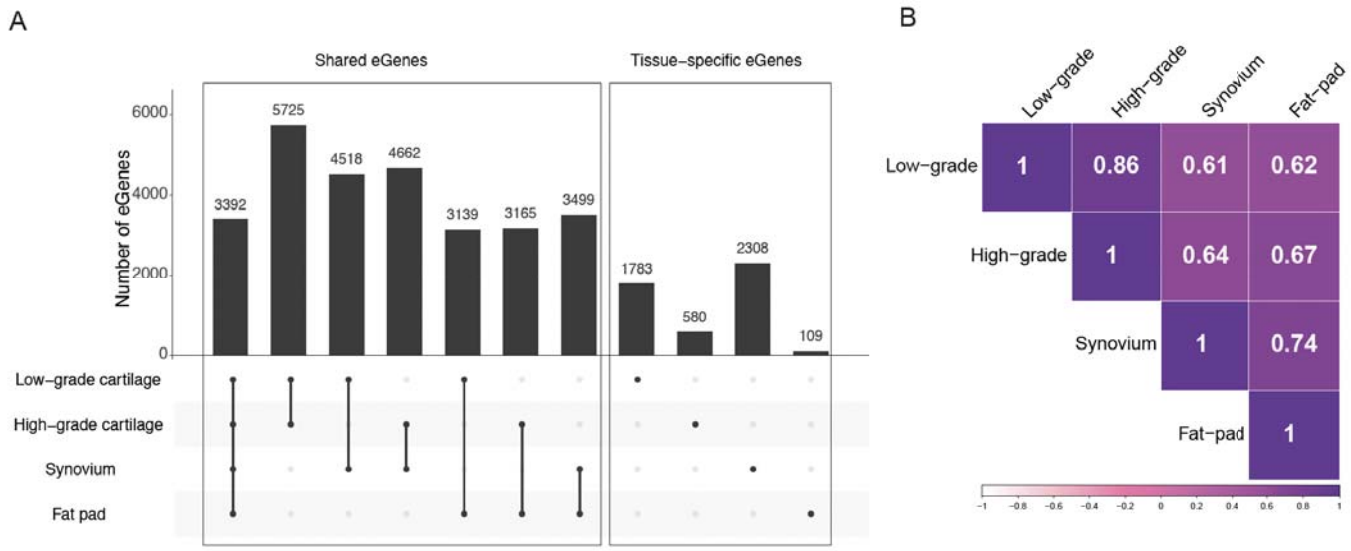


Figure 2: **Patterns of sharing and tissue-specificity among osteoarthritis primary tissues.** (A). Upset plot depicting the number of shared and tissue-specific *cis*-eGenes among tissues. (B) Heatmap showing pairwise sharing by magnitude of *cis*-eQTLs among tissues.

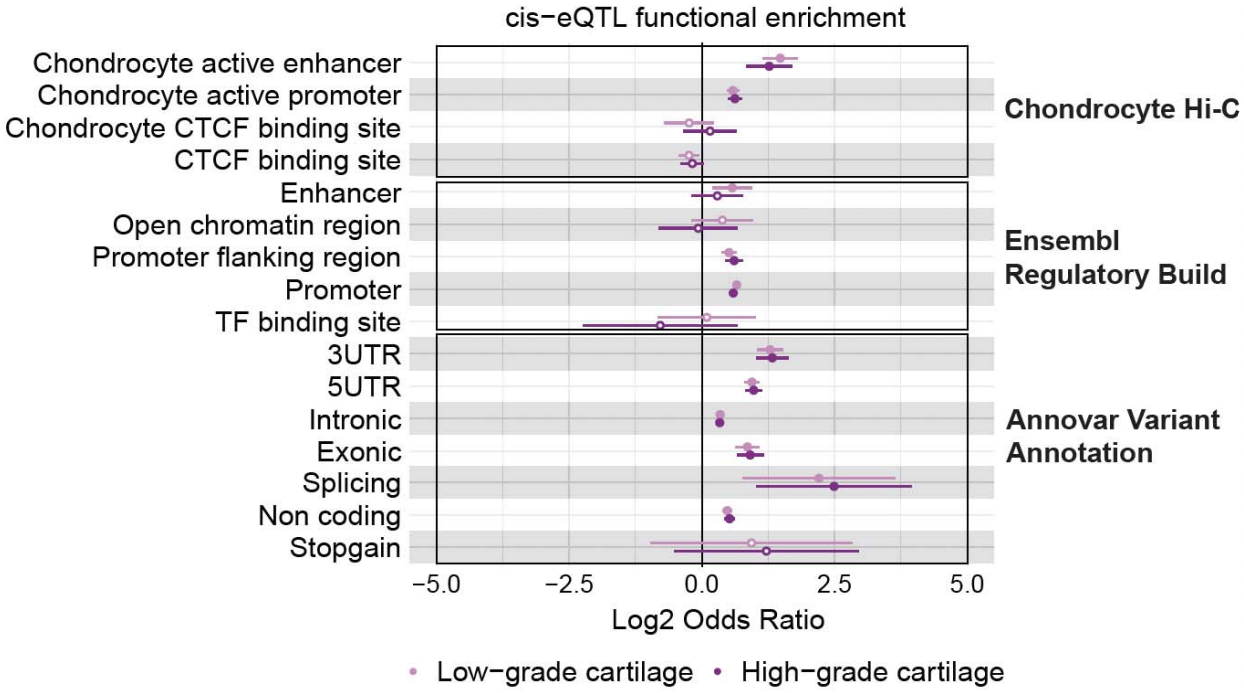
Distribution and functional enrichment of *cis*- and *trans*-eQTLs across regulatory regions

To understand the putative consequences of eQTLs in gene expression regulation, we explored their distribution among gene-regulatory regions. To this end, we utilized annotations from Ensembl Regulatory build[32], functional variant/gene annotations[33], and the first chondrocyte map of chromatin conformation in osteoarthritis patients using high-throughput chromosome conformation capture (HiC) technology[34] (Table S5).

Both *cis*- and *trans*-eQTLs are primarily located in intronic and intergenic regions, respectively. *Cis*-eQTLs are most abundant in intronic regions across all tissues, while *trans*-eQTLs are predominantly found in intergenic regions, with the exception of the synovium, where the majority of *trans*-eQTLs overlap with intronic regions. The majority of *cis*-eQTLs across all tissues overlap primarily with promoter-flanking or promoter regions, suggesting their role in regulating nearby gene expression. In contrast, *trans*-eQTLs (for both low- and high-grade cartilage) are mainly distributed across CTCF-binding sites, indicating a role in long-range regulatory interactions. In the synovium, *trans*-eQTLs are primarily found in enhancer regions reflecting putative tissue-specific regulatory dynamics (Figure S5, Table S5).

We further investigated whether *cis*-eQTLs display specific enrichment patterns within functional and regulatory elements (Figure 3). In low-grade cartilage, we observed the highest enrichment in active chondrocyte enhancers (odds ratio [OR] = 1.48, FDR < 10^{-1}), followed by 3' UTR (OR = 1.29, FDR < 10^{-1}) and 5' UTR regions (OR = 0.94, FDR < 10^{-1}). Similar enrichment patterns were found in high-grade cartilage. Both low- and high-grade cartilage eQTLs also exhibited enrichment in active chondrocyte promoters (low-grade cartilage: OR = 0.59, FDR < 10^{-1}) high-grade cartilage: OR = 0.62, FDR < 10^{-1}) (Table S5). These findings suggest that *cis*-eQTLs in both low- and high-grade cartilage are enriched in regions essential for chondrocyte function and regulation.

Synovium *cis*-eQTLs are most significantly enriched in 3' UTRs (OR = 1.52, FDR < 10^{-1}) and 5' UTRs (OR = 0.97, FDR < 10^{-1}) and, from a regulatory perspective, are enriched in promoters (OR = 0.62, FDR < 10^{-1}) and their flanking regions (OR = 0.55, FDR = 1.3 x 10^{-13}). Fat pad *cis*-eQTLs show similar enrichment patterns in functional gene elements, including 3' and 5' UTRs, as well as promoters (OR = 0.69, FDR = 1.3 x 10^{-13}). Both synovium and fat pad *cis*-eQTLs also show enrichment for splicing variants, though with lower significance than other categories but higher odds ratios (Figure 3, Table S4). Notably, fat pad *cis*-eQTLs display greater enrichment for splicing variants (OR = 3.51, FDR = 4.8 x 10^{-}) compared to synovium (OR = 2.14, FDR = 1.8 x 10^{-3}) (Table S5). These differences point to tissue-specific regulatory mechanisms, potentially linked to the unique functional requirements of each tissue.



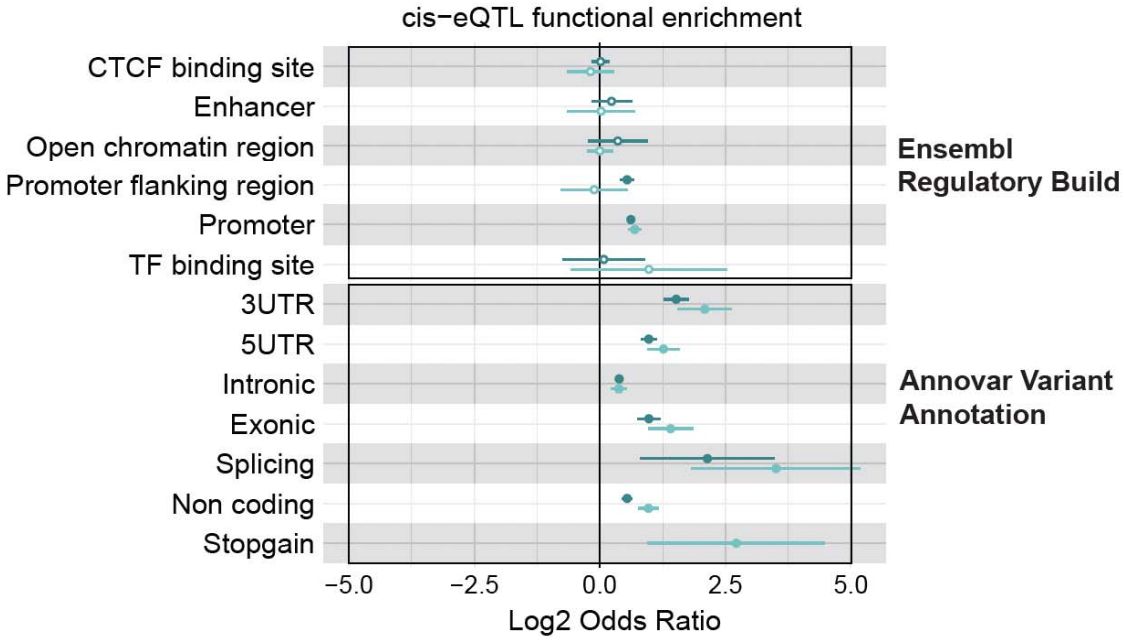


Figure 3: Functional annotation of *cis*-eQTLs across osteoarthritis primary tissues. Enrichment of *cis*-eQTLs in functional categories across tissues (upper panel: low- and high-grade cartilage, lower panel: synovium and fat-pad). Only enrichments with a standard error (SE) of less than one (SE < 1) are displayed. Filled graph dots indicate significant enrichment at a 5% false discovery rate (FDR).

Linking *cis*-eQTLs to knee-related osteoarthritis GWAS signals

To evaluate the contribution of the identified eQTLs in knee osteoarthritis genetic risk, we investigated *cis*-eQTL enrichment using knee osteoarthritis GWAS data[6]. We find significant enrichment of *cis*-eQTLs in low-grade, high-grade cartilage and synovium for knee osteoarthritis, osteoarthritis at any site, and knee replacement surgery. We do not find significant enrichment in fat pad tissue, which may be partly attributed to differences in sample size compared to the other tested tissues and, hence, power. Collectively, all tissues accounted for 10-16% of SNP-based heritability across osteoarthritis phenotypes (Table S6). This finding reflects the utility of eQTL maps in disentangling the genetic architecture of osteoarthritis.

We conducted statistical colocalization analysis between osteoarthritis GWAS signals[6] and primary tissue *cis*-eQTLs (Table S7). Among the 100 osteoarthritis GWAS risk loci examined, 40 show evidence of a shared association signal with *cis*-eQTLs from primary osteoarthritis tissues, spanning 65 unique genes. In low-grade cartilage, we find the largest percentage of significant colocalizations across all tested genes for osteoarthritis at any site (21%), followed by knee osteoarthritis (20%). Conversely, in high-grade cartilage and synovium, knee replacement surgery shows the greatest number of significant colocalizations across all tested genes (17% for high-grade cartilage and 16% for synovium). This aligns with the clinical understanding of total knee replacement as a last resort in osteoarthritis management, indicative of advanced cartilage degeneration.

For each colocalizing signal, we calculated a 95% credible set of variants, within which the shared causal variant lies with 95% probability. We studied the location of variants in the 95% credible sets by overlaying them with active promoter-enhancer loops from primary knee osteoarthritis chondrocytes[34]. Given that the HiC data were derived from cartilage tissue, we focused on colocalizing regions in high-grade and low-grade cartilage. Two eQTLs from the colocalization credible set are located within enhancers of active promoter-enhancer loops for the same respective eGene: rs10948 (*SLC44A2*) and rs655293 (*MPHOSPH9*). Both genes were previously identified as osteoarthritis effector genes (Figure 4, Figures S6-7). This suggests that these eQTL-eGene pairs may exert their effect on knee osteoarthritis risk by altering enhancer-promoter interactions. Seven eQTLs colocalizing with osteoarthritis GWAS signals reside within active promoters: rs13052882, rs36113648 (*BRWD1*), rs141002682 (*ANKRD54*), rs1800470 (*TMEM91*), rs2102066 (*WWP2*), rs6000899 (*GCAT*), rs798727 (*FAM53A*) (Figures S8-21). This suggests that these loci may play a critical role in the proximal regulation of gene expression and contribute to knee osteoarthritis risk by influencing promoter activity.

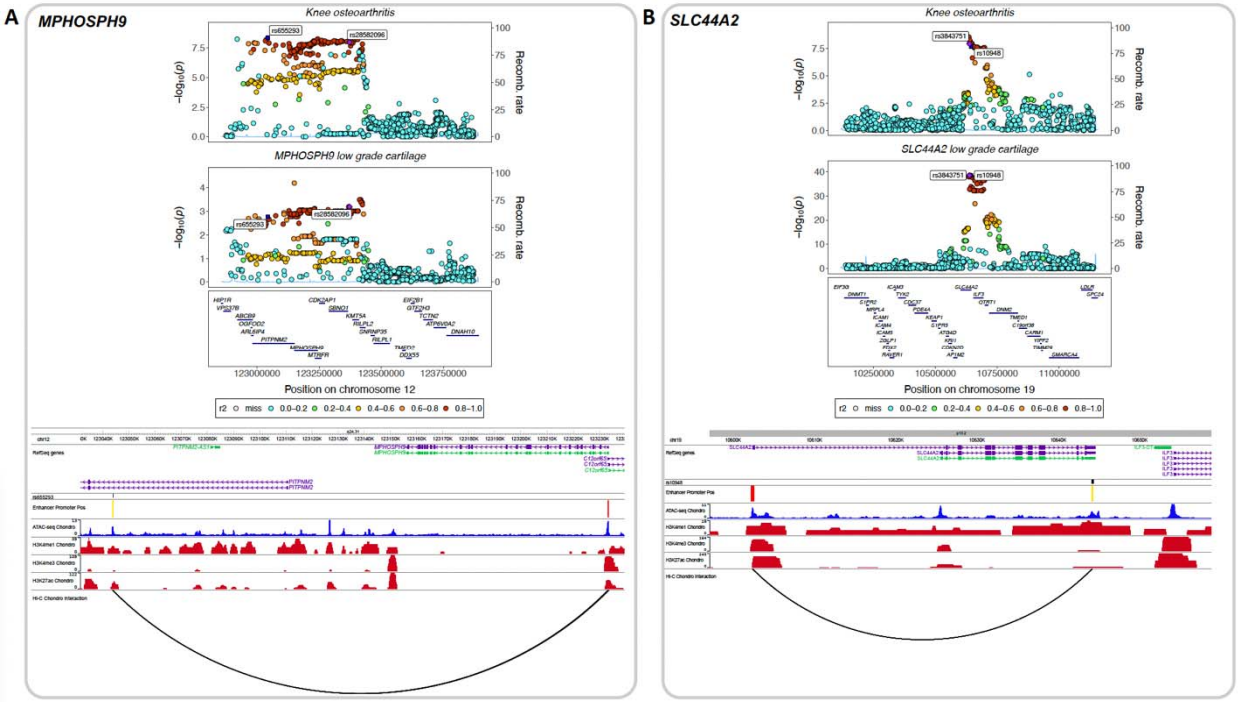


Figure 4: Examples of colocalizing *cis*-eQTLs in low-grade cartilage overlapping with chondrocyte active enhancer-promoter loops derived from Hi-C data.

(A) Example of rs655293, a colocalizing *cis*-eQTL of *MPHOSPH9* in low-grade cartilage and knee osteoarthritis found in the enhancer region of an active enhancer-promoter loop.

(B) Example of rs10948, a colocalizing *cis*-eQTL of *SLC44A2* in low-grade cartilage and knee osteoarthritis found in the enhancer region of an active enhancer-promoter loop.

For both examples: Upper panel: Regional association plot depicting the highlighted colocalizing region. The enhancer-overlapping cis-eQTL and lead colocalizing SNP are labeled. Lower panel: Track plot generated using WashU Epigenome Browser[35], showing annotated gene transcripts, SNP positions, enhancer and promoter locations, chromatin states, and histone modifications derived from chondrocyte ATAC-seq[36] and ChIP-seq[37] data.

Prioritization of 51 effector genes by combining colocalization and causal inference evidence

To infer putative causal effects of the identified colocalizing eGenes on knee-related osteoarthritis phenotypes, we conducted tissue-specific causal inference analyses using Mendelian randomization (MR)[38] (Table S8). We prioritize 51 effector genes showing evidence of colocalization and causal association for matching osteoarthritis phenotype and tissue[39] (Figure 5). Knockout mouse models for some of the prioritized genes show osteoarthritis-related phenotypes, for instance abnormal cartilage development and morphology (*GDF5*), decreased chondrocyte number (*GDF5*, *CREB3L2*), increased chondrocyte proliferation (*PTCH1*) and chondrodystrophy (*CREB3L2*). These genes have been previously defined as osteoarthritis effector genes[6]. Of the 51 prioritized genes, 25 have not been previously identified as effector genes for osteoarthritis[6] and 6 of these are associated with musculoskeletal-related phenotypes in knockout mouse models or rare and syndromic human diseases, including *ECRG4, MICU1, MUSTN1, NDNF, PBRM1* and *SNRPC* (Figure 5).

The prioritized effector gene showing the largest posterior probability of a shared causal variant (PP4=0.99) with osteoarthritis risk is *TMEM129*. This gene colocalizes with all tested knee-related osteoarthritis phenotypes in cartilage and synovium. We find evidence that increased levels of *TMEM129* in synovium are protective against knee-related osteoarthritis risk (knee replacement surgery OR=0.73, FDR=9.66×10^{-5}). *TMEM129* has been previously associated with osteoarthritis[6]. Among the 51 prioritized effector genes, *WWP2*, an established effector gene for osteoarthritis[6], showed the strongest evidence of causal association. Variants associated with the expression of this gene colocalize with knee osteoarthritis (PP4=0.98) and knee replacement surgery (PP4=0.89) in low-grade cartilage. One of the variants in the 95% credible set of the colocalization signal, rs2102066, resides in a chondrocyte active promoter-enhancer loop, suggesting a potential mechanism of action of this signal and *WWP2* on osteoarthritis. We find evidence that increased expression of *WWP2* in low-grade cartilage is causally associated with an increased risk of knee osteoarthritis (OR=1.29, FDR=2.06×10^{-35}) and knee replacement surgery (OR=1.38, FDR=1.93×10^{-19}), whereas increased expression in synovium has a protective effect on knee osteoarthritis risk (OR=0.82, FDR=0.003). Our findings point to contrary mechanisms of action for *WWP2* in cartilage and synovium.

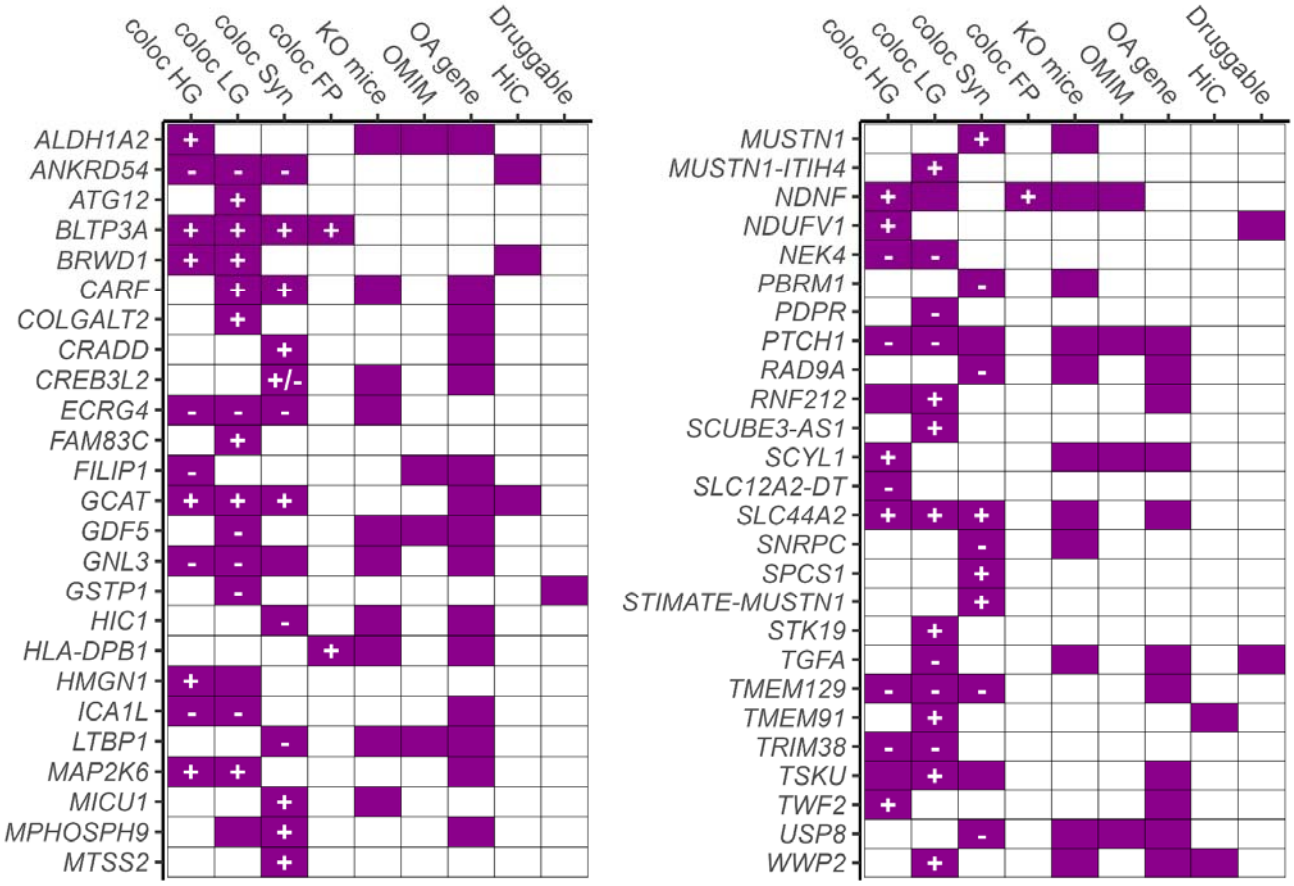


Figure 5: Prioritized effector genes showing evidence of colocalization and causal effect on osteoarthritis risk. In addition to showing the osteoarthritis primary tissue of colocalization (HG=high-grade cartilage, LG=low-grade cartilage, Syn=synovium, FP=fat pad), the heat map

visualizes additional evidence pointing to involvement in disease, including evidence of knockout mice (KO mice), rare and syndromic diseases (OMIM), previously defined effector genes (OA gene), overlap with open chromatin regions (HiC) and druggability evidence (drug).

Prioritized effector genes for osteoarthritis involved in chondrocyte differentiation and proliferation

Compared to all eGenes across the four primary osteoarthritis tissues analyzed here, the 51 prioritized effector genes were enriched for biological pathways related to cellular response to stimulus (q-value=0.035), histone deacetylase binding (q-value=0.035), transforming growth factor beta (TGF-β) binding (q-value=0.035), central nervous system neuron differentiation (q-value=0.052), and chondrocyte differentiation (q-value=0.053) (Figure 6, Table S9). Specifically, four genes were involved in the chondrocyte differentiation pathway: *GDF5*, *MUSTN1*, *TSKU* and *CREB3L2*. *MUSTN1*, which has not been previously associated with osteoarthritis[6], has been implicated in critical processes for maintaining cartilage integrity and function, such as chondrogenesis, chondrocyte differentiation, and proliferation[40]. The protein encoded by *MUSTN1* is predominantly expressed in skeletal muscle and bone and plays a role in myoblast differentiation, bone regeneration, chondrogenesis, and muscle hypertrophy[41]. Knockout mouse models for this gene show skeletal muscle fibrosis[42]. These findings further support the hypothesis that chondrocyte phenotypic transition from a non-proliferating articular phenotype to a development-like proliferating phenotype has a causal role in osteoarthritis.

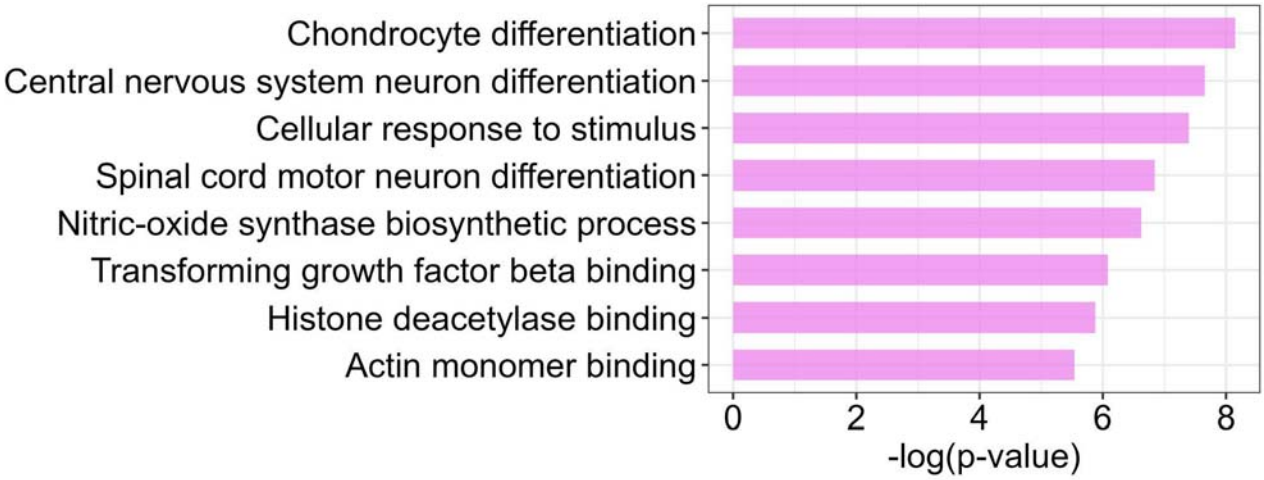


Figure 6: FDR-significant results of pathway-based over-enrichment analysis using ConsensusPathDB on the prioritized genes showing evidence of colocalization and causal association with knee-related osteoarthritis phenotypes.

Translational drug repurposing opportunities for the prioritized effector genes

Considering the translational potential of drug targets with genetic evidence supporting their effect on disease[43], we conducted a druggability search on the 51 prioritized effector genes using the Open Targets database[44]. Three genes were linked to drugs with at least phase II clinical trials completed: *TGFA, GSTP1*, and *NDUFV1*. *TGFA* is the target of fepixnebart, an

antibody that inhibits TGFA interaction with its receptor, thereby hindering downstream signaling pathways. This drug is under clinical trial for osteoarthritis, neuropathic pain, and low back pain. *GSTP1* is targeted by ezatiostat, a small molecule that inhibits GSTP1, leading to the activation of the JNK pathway[45]. Ezatiostat showed promising results for the treatment of myelodysplastic syndrome[46]. Both *GSTP1* and *TGFA* are well-established osteoarthritis effector genes. Metformin targets *NDUFV1*, a newly prioritized effector gene for osteoarthritis, that encodes a critical component of the electron transport chain in mitochondria. Metformin is a widely used medication primarily prescribed for the treatment of type 2 diabetes that modulates AMPK activity[47]. The beneficial effects of metformin treatment have also been observed for osteoarthritis risk[48,49].

Discussion

Osteoarthritis is the most common whole-joint disorder affecting nearly 600 million people worldwide, for which no curative therapy exists. Large efforts have been made to identify genetic risk signals for osteoarthritis. However, elucidating the function and tissue of action of these variants still remains a challenge. Here, we have generated the largest *cis*- and *trans*-eQTL maps in low-grade cartilage (n=263), high-grade cartilage (n=216), synovium (n=278), and fat-pad (n=94) knee osteoarthritis joint tissues. The availability of this resource aids in uncovering shared and tissue- and disease-stage specific biological mechanisms. By combining the newly-generated eQTL maps with osteoarthritis GWAS and HiC data of patient chondrocytes, we prioritize effector genes and highlight the mechanism of action of risk variants. By focusing on primary tissues directly affected by osteoarthritis, the identified *cis*-eQTLs provide a more precise and context-specific understanding of the genetic architecture underlying the disease highlighting the distinct tissue-specific genetic regulatory mechanisms operating in osteoarthritis-affected tissues.

Comparing our results with the first *cis*-eQTL study in cartilage and synovium[14], we recapitulate ~70% of *cis*-eQTL-gene pairs across tissues with largely concordant effect sizes. The comparison of our study with the synovium *cis*-eQTL map in 201 osteoarthritis patients of Chinese Han ancestry[16] yielded a 50% overlap for eGenes. This may be attributed to population-specific effects, including differences in allele frequencies and linkage disequilibrium patterns, and/or to analysis approach differences.

The enrichment of heritability within *cis*-eQTLs in cartilage and synovium for knee osteoarthritis GWAS signals underscores the significant role of gene expression regulation by nearby genetic variants to disease risk and clinical outcomes. Colocalization between the *cis*-eQTL maps and knee-related osteoarthritis GWAS finds evidence of a shared signal for 40% of the tested osteoarthritis risk signals. The partial elucidation of GWAS signals by colocalization with eQTL data underscores the increasing evidence that eQTLs and GWAS identify partially different sets of associated variants partly due to natural selection[50]. We prioritized 51 colocalizing genes showing evidence of a causal effect on knee-related osteoarthritis, which are enriched in chondrocyte differentiation and proliferation. Considering that genetic evidence support can accelerate drug discovery[51], our prioritized genes represent appealing candidates for clinical translation, including drug repurposing opportunities highlighted here.

Despite the gain in power to elucidate GWAS signals presented in this study, the generated data are restricted to samples from individuals of European ancestry. This reflects the gap in the availability of molecular data from underrepresented global populations[52]. Going forward, it is of utmost importance that efforts be put into closing this gap. Recent single-cell RNA-seq (scRNA-seq) studies in osteoarthritis cartilage[53], synovium[54], and fat-pad[55] have implicated specific cell types involved in osteoarthritis. Additionally, *trans*-eQTLs have been shown to capture cell-type specific effects[56]. Combining scRNA-seq with genomic data from larger sample sizes is needed to differentiate between real regulatory effects and different cell compositions.

The tissue-specific nature of gene regulation underscores the need for robust molecular data from primary tissues to enhance the translation of genetic findings for all complex diseases. Here, we have generated the largest molecular *cis* and *trans*-eQTL maps for four osteoarthritis-relevant tissues and two cartilage disease stages. Leveraging these powerful data, we detect shared and tissue-specific biological mechanisms underlying knee osteoarthritis. Our findings identify new candidate effector genes for this disease along with tissue of action and direction of effect, pointing to drug repurposing opportunities.

Methods

Sample selection

In this study, we included osteoarthritis patients who had undergone total knee replacement surgery for osteoarthritis with no history of significant knee surgery, infection, fracture, or malignancy within the previous 5 years. Additionally, we ensured that no patient had been treated with corticosteroids (systemic or intra-articular) within the previous 6 months, or any other drug associated with immune modulation. We isolated matched cartilage samples from weight-bearing regions of the joint along with synovium and fat-pad tissue from the same patients. Cartilage samples were also scored macroscopically using the International Cartilage Repair Society (ICRS) scoring system[57]. We collected from each patient one sample of ICRS grade 0 or 1, representing low-grade osteoarthritis degradation, and one sample of ICRS grade 3 or 4, representing high-grade osteoarthritis degradation. All study participants provided informed consent and samples were collected under Human Tissue Authority license 12182 and National Research Ethics Service approval 15/SC/0132 and 20/SC/0144, South Yorkshire and North Derbyshire Musculoskeletal Biobank, University of Sheffield, UK.

Library preparation

RNA-sequencing

We extracted RNA using Qiagen AllPrep RNA Mini Kit according to the manufacturer's instructions and as previously described in Steinberg et.al[14]. We isolated and sequenced samples in five batches that were later uniformly processed. In brief, we isolated Poly-A-tailed RNA from total RNA using Illumina's TruSeq RNA Sample Prep v2 for batches 1-4, while for the 5th batch, we used SMART-Seq® Ultra® Low Input RNA Kit. The libraries were then sequenced

on the Illumina HiSeq 2000 and HiSeq 4000 (75bp paired-ends) (batches 1-4) and Novaseq6000 (batch 5) recovering a median of 80.8 million reads per sample.

Whole-genome sequencing (WGS)

WGS data were generated from patients' DNA as previously described[58]. In brief, libraries were prepared from DNA samples using the standard Illumina paired-end DNA protocol as per manufacturers' instructions. Then, they were amplified and sequenced in two batches on the NovaSeq platform.

Preprocessing and quality control (QC) of RNA-sequencing data

We performed alignment of the fastq files to the reference genome Ensembl GRCh38 release 105 (cDNA and non-coding) using STAR version 2.7.9a.[59] We summarized the reads using FeatureCounts (Subread 2.0.3) software and called duplicates using PicardTools [https://broadinstitute.github.io/picard/]. We further used RSeqQC (version 4.0.0)[60] to summarize alignment metrics. We first performed sample-level quality control and excluded samples that did not pass the following criteria: RNA integrity number >5, percentage of uniquely mapped reads >70%, reads mapping to genomic features >40%, reads mapping to intronic regions or rRNA <30%, reads that failed RseQC strandness check <30% and including >10 million reads. Using the aforementioned criteria we excluded 42 low-grade cartilage samples, 35 high-grade cartilage samples, 14 synovium samples, and 13 fat-pad adipocyte samples. This way we ensured the inclusion of only high-quality samples.

We further excluded outlier samples in a tissue-specific manner. Outliers were identified using robust principal component analysis considering the first three principal components (PcaGrid method rrcov R package)[61]. We excluded samples from individuals that had bilateral knee replacement where low- and high-grade disease cartilage samples were pooled (7 indivuduals). For individuals with bilateral knee replacement, we excluded one pair of matched cartilage samples each (3 samples), keeping only the sample pair with the best quality. For samples with technical replicates we also kept to one with the best quality (2 samples). The final number of samples after the exclusions across tissue was 881, derived from 331 patients (low-grade cartilage:272, high-grade cartilage:223, synovium:288, fat-pad:98). After QC we performed principal component analysis (PCA) to find main drivers of variation and evaluate the presence of batch effects. The principal component 1 (PC1), explaining 29.89% of the variation in the data, corresponded to the different tissues, while PC2, explaining 13.22% of the variation, corresponded to the sequencing batches.

Preprocessing and QC of Whole Genome Sequencing (WGS) data

WGS preprocessing and QC was performed as previously described[58]. CRAM files were converted to BAM format using samtools (version samtools conda version 1.14). The later were further converted to fastq format using *bedtools bamtofastq* function (bedtools conda version 2.30.0). We performed variant calling per batch and using Sarek nf-core pipeline (version 2.7.1, https://nf-co.re/sarek/2.7.1) with the additional options "*--tools HaplotypeCaller, --generate_gvcf*". We further performed joint variant calling by adapting the GATK4 pipeline (https://github.com/IARCbioinfo/gatk4-GenotypeGVCFs-nf) to be used with GATK (docker

container broadinstitute/gatk:4.2.5.0).[62] For all variant calling purposes GRCh38 reference was specified.

For variant-level quality control, we used Variant Quality Score Recalibration (VQSR) method specifying a tranche threshold of 99.5% for SNPs and 99% for INDELS. The expected false positive rate for SNPs was 2.5% and the expected sensitivity was 97%.

We excluded samples with high heterozygosity rate (2 samples), non-reference allele concordance with the directly typed genotype data using variants MAF >0.01 (1 sample - also showing heterozygosity rate), outliers in sequencing depth (1 sample), sex mismatched (2 samples), relatedness > 0.2 (2 samples), ethnicity outliers (2 samples). Ethnicity outliers (3 individuals) were identified using the Ancestry and kinship toolkit (based on 1000G data from phase three; https://github.com/Illumina/akt/tree/master)[63]. Nine samples were excluded in total.

After sample-level QC, we performed additional filtering and excluded variants with MAF <0.01, Hardy-Weinberg equilibrium P<10^{-5} and call rate <=0.99. After selecting individuals matching WGS and RNA-seq data per tissue we performed further variant filtering to keep only bi-allelic and common variants (MAF > 0.05). The final WGS data sets per tissue after filtering were as follows: low-grade cartilage (263 individuals, 6,304,167 variants), high-grade cartilage (216 individuals, 6,323,956 variants), synovium: (278 individuals, 6,338,438 variants), fat-pad (94 individuals, 6,295,898 variants). Power analysis for the eQTL analysis was performed using the R package powerEQTL[64] (*powerEQTL.SLR*) across different MAF, slopes and sample sizes.

Mapping *cis*-eQTL

For the eQTL analysis, we included WGS data and matched RNA-seq tissue-specific expression data from the same patients (n=321 osteoarthritis patients across all tissues). For eQTL analyses the gene expression count matrix was normalized using the trimmed mean of M-values (TMM) normalization[65] (between samples normalization) and then inverse-normal transformed (across genes normalization). This procedure was performed separately per tissue (cartilage, synovium and fat-pad). The resulting matrices were then used to estimate PEER factors[66] representing hidden data variation. The number of PEER factors included per tissue in the eQTL model was determined according to GTEx[67] recommendations on the sample size (low-grade cartilage: 45, high-grade cartilage: 30, synovium: 45, fat-pad: 15) [https://github.com/broadinstitute/gtex-pipeline/tree/master/qtl]. For eQTL analyses we included protein-coding and long non-coding RNA (lncRNA) genes excluding other gene biotypes due to higher quantification uncertainty. We further included genes on the autosomes. The total number of genes per tissue used in eQTL testing was formed as follows: cartilage (both low-grade and high-grade): 16,009, synovium: 14,482, fat-pad:15,156. We further defined the transcription start-site (TSS) for each gene based on its most abundant transcript across the tested tissues (transcript abundance quantified using RSEM[68]).

We conducted *cis*-eQTL analysis per tissue using tensorQTL (version 1.0.9)[69] following the GTEx pipeline [https://github.com/broadinstitute/gtex-pipeline/tree/master/qtl]. In brief, this method uses linear regression to test for associations between variant–gene pairs in a specified window (1-Mb specified for *cis*-eQTL) within the transcription start site of a gene (nominal pass). The nominal p-values were estimated for every variant–gene pair with the following model:

$$Expression \sim genotype + age + sex + WGS_sequencing_batch + \#PEER_factors$$

Out of 321 WGS samples, 247 were extracted in the first batch and 74 in the second sequencing batch. The number (#PEER_factors) was determined per tissue.

To test for association between gene expression and the lead variant (top in *-cis*) we used tensorQTL with the implemented permutation scheme specifying 1,000 permutations per gene. This way, we calculated the nominal p value threshold per gene and the gene-level q values corrected for multiple testing (Storey-Tibshirani). To identify eGenes, we performed q value correction of the permutation p values for the lead associated variant per gene at a threshold of 0.05. To identify significant eQTL-gene pairs we filtered the nominal output for the pairs below the gene-specific p value threshold. We further identified conditionally independent eQTL by performing a stepwise regression procedure as described in GTEx[67].

Mapping *trans*-eQTL

We performed *trans*-eQTL analysis per tissue using MatrixeQTL[70]. Expression and genotype input data were prepared as described for the *cis*-eQTL analysis. The same covariates and number of PEER factors were included in the final model. For *trans*-eQTL, we tested variants' whose distance to the genes's TSS was more than 5Mb. We defined significant *trans*-eQTL-gene pairs as the ones that had a nominal p value smaller than the p value threshold that corresponded to 5% FDR. Finally, we removed potential false positive *trans*-eQTLs caused by read cross-mapping. To this end, we utilized publicly available cross-mappability resources for the GRCh38 reference genome and excluded all *trans*-eQTL-gene pairs that had a mappability score less than 0.8[71]. We identified conditionally independent *trans*-eQTLs for each analyzed tissue by clumping the significant trans-eQTLs based on LD. We used PLINK[72] to perform clumping on a window of 1 Mb and a strict LD threshold $r^2 < 0.001$.

Comparison with existing *cis*-eQTL maps

To find newly reported associations we compared the significant eQTL-gene pairs with published eQTL studies from Streinberg et al[14] and Jiang et al.[16]. Both of these studies report eQTL associations on reference genome build 37 while our eQTL maps are on build 38. To establish a fair comparison, we converted the variant positions to the corresponding Reference SNP cluster ID (rsID) as rsIDs do not depend on a reference genome. We mapped variants' genomic positions to rsIDs using dbSNP database[73] available through Bioconductor package SNPlocs.Hsapiens.dbSNP144.GRCh37.

We further evaluated the association of effect sizes (slopes) for *cis*-eQTLs by conducting a linear regression analysis of common eQTL effect sizes between our study and that of

Steinberg et al.[14], which was performed in a subset of individuals from our study, all of European ancestry. Specifically, we analyzed 46,433 common *cis*-eQTLs with rsIDs in synovium, 63,294 in low-grade cartilage, and 79,826 in high-grade cartilage. For each tissue type, we fit a linear regression model with the effect sizes from our study as the predictor variable and those from Steinberg et al.[14] as the outcome variable. From this model, we extracted the R-squared value to quantify the strength of association and the p-value to evaluate statistical significance. To visualize these associations, we used ggplot2[74] to generate scatter plots with fitted regression lines and confidence intervals for each tissue type.

Identification of tissue-sharing and tissue-specific eQTLs

To assess tissue-specific and shared eQTL patterns, we applied multivariate adaptive shrinkage using MashR[75] on the top *cis*-eQTLs for each gene across four tissues, focusing on genes expressed in all tissues (N = 13,181). We input the estimated effect sizes and their associated standard errors, along with 59,310 randomly selected SNP-gene pairs tested across all tissues, to fit the Mash model as previously done (https://github.com/stephenslab/gtex-eqtls). MashR's output of effect size estimates and local false sign rates (LFSR), were used as measures of *cis*-eQTL magnitude and significance, respectively. We set an LFSR threshold of < 0.05 to identify significant eQTLs within each tissue. Shared cis-eQTLs were defined as those with an LFSR < 0.05 and an effect size within a two-fold range of the strongest effect observed across tissues. Tissue-specific eQTLs were identified by either a minimum two-fold larger effect size in one tissue compared to all others or by detection in only one tissue (i.e., eQTLs in genes not common across tissues).

Variant annotation of eQTL functional enrichment

We annotated all variants tested in eQTL analyses for functional impact on known genes using the ANNOVAR tool (https://annovar.openbioinformatics.org/en/latest/)[76]. Regulatory annotations were drawn from the Ensembl Regulatory Build[32], covering features such as enhancers, promoters, promoter-flanking regions, transcription factor binding sites, CTCF binding sites, and open chromatin regions. For low- and high-grade cartilage tissue, we further utilized enhancer-promoter loop coordinates specific to osteoarthritis cartilage, as mapped by chromatin accessibility Hi-C data from Bittner et al.[34]. From this Hi-C map (hg38 build), we extracted genomic coordinates for: (1) chondrocyte-specific CTCF binding sites, marking the most active regions within topologically associated domains (TADs); and (2) active chondrocyte promoters and enhancers within Hi-C loops. Active loops were identified using ATAC-seq open chromatin data combined with ChIP-seq peaks for H3K4me3 and H3K27ac signals[34]. We evaluated the distribution of eQTL annotations among both *cis*- and *trans*-eQTLs.

To determine whether *cis*-eQTLs, which were detected in significantly higher numbers than *trans*-eQTLs, were more prevalent in specific genomic regions, we conducted a functional enrichment analysis using the aforementioned annotations and the torus software[77], as previously described[67].

Partitioning heritability to eQTLs

We employed LD score regression to partition heritability[78]. Using this approach, we broke down the heritability of the tested osteoarthritis phenotypes (osteoarthritis, knee osteoarthritis, total knee replacement) across *cis*-eQTLs and *trans*-eQTLs for the different tissues. Leveraging data from the GWAS summary statistics a large GWAS meta-analysis on osteoarthritis[6] and ancestry-matched LD modeling from the 1000 Genomes reference panel[63], we quantified the proportion of the overall genome-wide heritability that can be attributed to eQTLs. EQTL categories were added to the full baseline model of 53 categories of broad regulatory functions described in Funicane et al.[78] to enhance model accuracy.

Genetic colocalization of *cis*-eQTL with GWAS signals

We conducted regional genetic colocalization analyses between eQTL from osteoarthritis primary tissues (low-grade and high-grade cartilage, synovium, and fat-pad) and three osteoarthritis GWAS (knee osteoarthritis, total knee replacement, and osteoarthritis at any site)[6]. Regions to run genetic colocalization were defined as 1 Mb windows on either side of osteoarthritis GWAS risk signals defined in Boer et al.[6]. To perform the analyses, we have used the *coloc.abf* function of the *coloc* R package (version 5.2.2). To overcome possible LD challenges due to for instance unmatched population structure, evidence of colocalization was set at a posterior probability of a shared causal variant (PP4) >= 0.8 or PP4 > 0.6 and PP4/PP3 > 2, where PP3 denotes the posterior probability of both traits having a distinct causal variant in the genomic region.

In regions where no evidence of colocalization but slight evidence of association for both traits (PP3>0.1) was found, we conducted additional analyses using the *coloc.susie* function which relaxes the assumption of a single shared causal variant by fine-mapping each trait and defining multiple credible sets to run colocalization. For this additional analysis, an LD reference panel is required for each investigated trait. For the eQTL dataset, we used the in-sample LD map and for the osteoarthritis GWAS, we used the UK Biobank genotype files to derive an LD reference map[79]. For each colocalized genomic locus, we calculated a 95% credible set for the shared causal variant by taking the cumulative sum of the variants' posterior probabilities to be causal conditional on H4 being true. LD between the single-nucleotide polymorphisms (SNPs) was calculated using plink (v.2.0 alpha) based on the UK biobank and was used for visualizing the results in regional association plots[79].

Overlap between variants in the 95% credible set of the colocalization and HiC data from chondrocytes

For each locus showing evidence of colocalization between cis-eQTL data and knee-related osteoarthritis phenotypes, we mapped all the variants in the 95% credible sets for the shared causal variants to active enhancer-promoter loops from primary knee osteoarthritis chondrocytes[34]. To identify the target genes of the active promoters, the authors overlaid the extended promoter region with gene regions from ENSEMBL[80].

Mendelian randomization analyses between *trans*- and *cis*-eQTL and knee-related osteoarthritis

To evaluate putative causal effects of gene expression levels on knee-related osteoarthritis, we conducted Mendelian randomization analyses[38]. As exposures, we used the generated *cis*- and *trans*-eQTL data from the colocalizing genes derived from osteoarthritis primary tissues (low-grade and high-grade cartilage, synovium, and fat-pad), and as the outcome, we employed GWAS summary statistics from knee osteoarthritis, osteoarthritis at any site, and total knee replacement[6]. Instrumental variables for the analyses were identified for each gene and tissue combination from the exposure summary statistics as independent variants significantly associated with gene expression. We defined independence through clumping of all significantly associated variants based on LD using the European reference LD panel from the 1000 Genomes project[63] within a 10Mb window on either side of the index variants, and with a strict threshold of R^2=0.001. In cases where an instrumental variable was not present in the outcome data, we conducted an LD-based proxy search using $R^2 > 0.7$ in the same LD reference panel as in the clumping step. For the proxy search, we used the *LDlinkR* R package (version 1.3.0). We used the *TwoSampleMR* R package (version 0.5.7) curated by MR-Base[81] to run the causal inference analysis based on the inverse variance weighted method (IVW) and the Wald ratio if only one instrumental variable was available.

Mendelian randomization relies on certain assumptions, which can partially be validated. Firstly, we ensured that the selected instrumental variables were strongly associated with the exposure by filtering out variants with an F-statistic < 10. The F-statistic was calculated as the ratio between the squared effect size estimates and its squared standard error ($beta^2/se^2$)[38]. In cases where multiple instrumental variables were available, we reanalyzed using the inverse variance weighted method with Steiger-filtered exposure data to assess reverse causation by comparing the direction of causal effect estimates. Additionally, if more than two instrumental variables were available, weighted median and MR-Egger analyses were conducted to assess the direction of effect concordance among different estimates of causal effects. To indicate a lack of evidence of pleiotropy, we required the MR-Egger intercept to be equal to zero (FDR adjusted p-value > 0.05). To assess heterogeneity, we examined the FDR adjusted p-value of the Cochran's Q-statistic, requiring it to be larger than 0.05 for no evidence of heterogeneity. Finally, we have adjusted the p-values of the causal estimates for multiple testing burdens using FDR correction. Hence, evidence for a causal effect was defined based on the following criteria:

1. FDR adjusted p-value of the Wald ratio or IVW method < 0.05
2. If applicable, concordant direction of effect across different tested methods
3. Lack of evidence of pleiotropy (FDR adjusted MR-Egger intercept p-value > 0.05)
4. Lack of evidence of heterogeneity (FDR adjusted Q-statistic p-value > 0.05)

Pathway enrichment analysis

We have conducted a pathway-based over-representation analysis for all tissue-specific and shared eGenes using all unique gene among tissues as a background (13,181 shared genes). All analyses were performed using ConsensusPathDB (http://cpdb.molgen.mpg.de/)[82] for both pathways and Gene Ontology terms using the default parameters. We further performed enrichment analysis for the genes showing evidence of colocalization and causal association with knee-related osteoarthritis phenotypes. For this analysis, we used as the background 9,758 genes that had an eQTL in at least one of the four analyzed osteoarthritis primary tissues, i.e.,

the genes that were tested in the MR and colocalization analyses. We selected pathways from the following databases: Reactome, KEGG, WikiPathways, and Gene Ontology. For the latter, we included all subcategories, namely molecular function, biological processes, and cellular components, up to level 4. We required a minimum overlap of two genes for enrichment. The significance threshold was set at false discovery rate (FDR) < 0.05.

Lookup of prioritized effector genes

To gain further insights into the biology and function of the prioritized genes, we searched these genes in knockout mice and rare and syndromic human diseases databases. Firstly, to look at rare and syndromic diseases linked to the prioritized genes, we extracted data from the Online Mendelian Inheritance in Man (OMIM[83]) (https://omim.org/) via its API. We defined association with musculoskeletal phenotypes if the disease showed any skeletal or muscular phenotype. Secondly, we used the batch query functionality of the Mouse Genome Informatics (MGI[84]) database (http://www.informatics.jax.org/) to retrieve knockout mice phenotypes for all prioritized genes. We then screened the output for the following musculoskeletal phenotypes: bone, muscle, skeleton, osteo, arthritis, muscular, joint, body size, growth, skeletal, stature, height, limb, appendage, cartilage, chondrocyte, body length, brachypodia, brachydactyly, brachyphalangia, femur, tibia, ulna, fibula, humerus, radial, spine curvature, posture, vertebral arch and syndactyly.

Drug targets analysis

We have queried the druggability of the prioritized effector genes using the OpenTargets[44] database (version 23.09), which includes information about 61,264 unique genes. Open Targets defines drugs as any bioactive molecule with drug-like properties, based on the EMBL-EBI ChEMBL database[85]. This definition excludes vaccines, blood products, cell therapies and multi-ingredient drugs. We considered drugs with information on the indication that had completed at least a Phase 2 trial.

Acknowledgements

The authors wish to thank Norbert Bittner for providing access to the HiC chondrocyte data and contributing to the discussion for their optimal use. We acknowledge the technical support of Core Facility Genomics at Helmholtz Munich. We thank Dr. Inti Alberto de la Rosa Velasquez, Dr. Peter Lichtner and Dr. Gertrud Eckstein and Dr. Jenny Hankinson for their help in sample handling and additional processing of RNA-seq and WGS data. We also wish to thank Iris Fischer for her helpful contributions during manuscript submission process, and all study participants for their willingness to donate study samples. The work on osteoarthritis cartilage samples was funded by Wellcome Trust (206194).

Author contributions

Study design: E.Z., J.M.W.; Clinical collection: J.M.W., K.M.S., D.S.; Data Analysis: G.K., A.A., M.T., P.K.; Interpretation of results: G.K., A.A., E.Z., J.M.W, L.S.; Data visualization: G.K., A.A., M.T; Manuscript drafting: G.K., A.A., E.Z.; Manuscript reviewing and editing: all authors.

Data and Code availability

Summary statistics of all analyses will be shared through the Musculoskeletal Knowledge Portal (mskkp.org). All software used in this study is available from free repositories or manufacturers as referenced throughout the methods section.

Declarations of interests

The authors declare no competing interests.

References

1. GBD 2021 Osteoarthritis Collaborators (2023). Global, regional, and national burden of osteoarthritis, 1990-2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Rheumatol *5*, e508–e522. https://doi.org/10.1016/S2665-9913(23)00163-7.

2. Sanchez-Lopez, E., Coras, R., Torres, A., Lane, N.E., and Guma, M. (2022). Synovial inflammation in osteoarthritis progression. Nat Rev Rheumatol *18*, 258–275. https://doi.org/10.1038/s41584-022-00749-9.

3. Clockaerts, S., Bastiaansen-Jenniskens, Y.M., Runhaar, J., Van Osch, G.J.V.M., Van Offel, J.F., Verhaar, J. a. N., De Clerck, L.S., and Somville, J. (2010). The infrapatellar fat pad should be considered as an active osteoarthritic joint tissue: a narrative review. Osteoarthritis Cartilage *18*, 876–882. https://doi.org/10.1016/j.joca.2010.03.014.

4. Martel-Pelletier, J., Barr, A.J., Cicuttini, F.M., Conaghan, P.G., Cooper, C., Goldring, M.B., Goldring, S.R., Jones, G., Teichtahl, A.J., and Pelletier, J.-P. (2016). Osteoarthritis. Nat Rev Dis Primers *2*, 16072. https://doi.org/10.1038/nrdp.2016.72.

5. Hunter, D.J., and Bierma-Zeinstra, S. (2019). Osteoarthritis. Lancet *393*, 1745–1759. https://doi.org/10.1016/S0140-6736(19)30417-9.

6. Boer, C.G., Hatzikotoulas, K., Southam, L., Stefánsdóttir, L., Zhang, Y., Coutinho de Almeida, R., Wu, T.T., Zheng, J., Hartley, A., Teder-Laving, M., et al. (2021). Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell *184*, 4784-4818.e17. https://doi.org/10.1016/j.cell.2021.07.038.

7. Chen, D., Shen, J., Zhao, W., Wang, T., Han, L., Hamilton, J.L., and Im, H.-J. (2017). Osteoarthritis: toward a comprehensive understanding of pathological mechanism. Bone Res *5*, 1–13. https://doi.org/10.1038/boneres.2016.44.

8. Zeng, N., Yan, Z.-P., Chen, X.-Y., and Ni, G.-X. (2020). Infrapatellar Fat Pad and Knee Osteoarthritis. Aging Dis *11*, 1317–1328. https://doi.org/10.14336/AD.2019.1116.

9. Zhou, S., Maleitzke, T., Geissler, S., Hildebrandt, A., Fleckenstein, F.N., Niemann, M., Fischer, H., Perka, C., Duda, G.N., and Winkler, T. (2022). Source and hub of inflammation: The infrapatellar fat pad and its interactions with articular tissues during knee osteoarthritis. J Orthop Res *40*, 1492–1504. https://doi.org/10.1002/jor.25347.

10. Hu, W., Chen, Y., Dou, C., and Dong, S. (2021). Microenvironment in subchondral bone: predominant regulator for the treatment of osteoarthritis. Ann Rheum Dis *80*, 413–422. https://doi.org/10.1136/annrheumdis-2020-218089.

11. Aguet, F., Alasoo, K., Li, Y.I., Battle, A., Im, H.K., Montgomery, S.B., and Lappalainen, T. (2023). Molecular quantitative trait loci. Nat Rev Methods Primers *3*, 1–22. https://doi.org/10.1038/s43586-022-00188-6.

12. Qi, T., Song, L., Guo, Y., Chen, C., and Yang, J. (2024). From genetic associations to genes: methods, applications, and challenges. Trends Genet *40*, 642–667. https://doi.org/10.1016/j.tig.2024.04.008.

13. Hekselman, I., and Yeger-Lotem, E. (2020). Mechanisms of tissue and cell-type specificity in heritable traits and diseases. Nat Rev Genet *21*, 137–150. https://doi.org/10.1038/s41576-019-0200-9.

14. Steinberg, J., Southam, L., Roumeliotis, T.I., Clark, M.J., Jayasuriya, R.L., Swift, D., Shah, K.M., Butterfield, N.C., Brooks, R.A., McCaskie, A.W., et al. (2021). A molecular quantitative trait locus map for osteoarthritis. Nat Commun *12*, 1309. https://doi.org/10.1038/s41467-021-21593-7.

15. Arruda, A.L., Katsoula, G., Chen, S., Reimann, E., Kreitmaier, P., and Zeggini, E. (2024). The Genetics and Functional Genomics of Osteoarthritis. Annual Review of Genomics and Human Genetics *25*, 239–257. https://doi.org/10.1146/annurev-genom-010423-095636.

16. Jiang, F., Hu, S.-Y., Tian, W., Wang, N.-N., Yang, N., Dong, S.-S., Song, H.-M., Zhang, D.-J., Gao, H.-W., Wang, C., et al. (2024). A landscape of gene expression regulation for synovium in arthritis. Nat Commun *15*, 1409. https://doi.org/10.1038/s41467-024-45652-x.

17. Wei, L., Kanbe, K., Lee, M., Wei, X., Pei, M., Sun, X., Terek, R., and Chen, Q. (2010). Stimulation of chondrocyte hypertrophy by chemokine stromal cell-derived factor 1 in the chondro-osseous junction during endochondral bone formation. Dev Biol *341*, 236–245. https://doi.org/10.1016/j.ydbio.2010.02.033.

18. Wei, F., Moore, D.C., Wei, L., Li, Y., Zhang, G., Wei, X., Lee, J.K., and Chen, Q. (2012). Attenuation of osteoarthritis via blockade of the SDF-1/CXCR4 signaling pathway. Arthritis Res Ther *14*, R177. https://doi.org/10.1186/ar3930.

19. Chan, B., Glogauer, M., Wang, Y., Wrana, J., Chan, K., Beier, F., Bali, S., Hinz, B., Parreno, J., Ashraf, S., et al. (2023). Adseverin, an actin-binding protein, modulates hypertrophic chondrocyte differentiation and osteoarthritis progression. Sci Adv *9*, eadf1130. https://doi.org/10.1126/sciadv.adf1130.

20. Fioravanti, A., Nerucci, F., Annefeld, M., Collodel, G., and Marcolongo, R. (2003). Morphological and cytoskeletal aspects of cultivated normal and osteoarthritic human articular chondrocytes after cyclical pressure: a pilot study. Clin Exp Rheumatol *21*, 739–746.

21. Lauer, J.C., Selig, M., Hart, M.L., Kurz, B., and Rolauffs, B. (2021). Articular Chondrocyte Phenotype Regulation through the Cytoskeleton and the Signaling Processes That Originate from or Converge on the Cytoskeleton: Towards a Novel Understanding of the Intersection between Actin Dynamics and Chondrogenic Function. Int J Mol Sci *22*, 3279. https://doi.org/10.3390/ijms22063279.

22. Haidar, O., O'Neill, N., Staunton, C.A., Bavan, S., O'Brien, F., Zouggari, S., Sharif, U., Mobasheri, A., Kumagai, K., and Barrett-Jolley, R. (2020). Pro-inflammatory Cytokines Drive Deregulation of Potassium Channel Expression in Primary Synovial Fibroblasts. Front Physiol *11*, 226. https://doi.org/10.3389/fphys.2020.00226.

23. Murthy, S.E., Dubin, A.E., Whitwam, T., Jojoa-Cruz, S., Cahalan, S.M., Mousavi, S.A.R., Ward, A.B., and Patapoutian, A. (2018). OSCA/TMEM63 are an Evolutionarily Conserved Family of Mechanically Activated Ion Channels. Elife *7*, e41844. https://doi.org/10.7554/eLife.41844.

24. Hu, X., Laragione, T., Sun, L., Koshy, S., Jones, K.R., Ismailov, I.I., Yotnda, P., Horrigan, F.T., Gulko, P.S., and Beeton, C. (2012). KCa1.1 potassium channels regulate key proinflammatory and invasive properties of fibroblast-like synoviocytes in rheumatoid arthritis. J Biol Chem *287*, 4014–4022. https://doi.org/10.1074/jbc.M111.312264.

25. Nanus, D.E., Badoume, A., Wijesinghe, S.N., Halsey, A.M., Hurley, P., Ahmed, Z., Botchu, R., Davis, E.T., Lindsay, M.A., and Jones, S.W. (2021). Synovial tissue from sites of joint pain in knee osteoarthritis patients exhibits a differential phenotype with distinct fibroblast subsets. EBioMedicine *72*, 103618. https://doi.org/10.1016/j.ebiom.2021.103618.

26. Zhou, Q., Ren, Q., Jiao, L., Huang, J., Yi, J., Chen, J., Lai, J., Ji, G., and Zheng, T. (2022). The potential roles of JAK/STAT signaling in the progression of osteoarthritis. Front Endocrinol (Lausanne) *13*, 1069057. https://doi.org/10.3389/fendo.2022.1069057.

27. Cao, Y., Ruan, J., Kang, J., Nie, X., Lan, W., Ruan, G., Li, J., Zhu, Z., Han, W., Tang, S., et al. (2024). Extracellular Vesicles in Infrapatellar Fat Pad from Osteoarthritis Patients Impair Cartilage Metabolism and Induce Senescence. Adv Sci (Weinh) *11*, e2303614. https://doi.org/10.1002/advs.202303614.

28. Blanco, F.J., Rego, I., and Ruiz-Romero, C. (2011). The role of mitochondria in osteoarthritis. Nat Rev Rheumatol *7*, 161–169. https://doi.org/10.1038/nrrheum.2010.213.

29. Rellmann, Y., Eidhof, E., and Dreier, R. (2021). Review: ER stress-induced cell death in osteoarthritic cartilage. Cell Signal *78*, 109880. https://doi.org/10.1016/j.cellsig.2020.109880.

30. Zheng, L., Zhang, Z., Sheng, P., and Mobasheri, A. (2021). The role of metabolism in chondrocyte dysfunction and the progression of osteoarthritis. Ageing Res Rev *66*, 101249. https://doi.org/10.1016/j.arr.2020.101249.

31. Carlo, M.D., and Loeser, R.F. (2003). Increased oxidative stress with aging reduces chondrocyte survival: correlation with intracellular glutathione levels. Arthritis Rheum *48*, 3419–3430. https://doi.org/10.1002/art.11338.

32. Zerbino, D.R., Wilder, S.P., Johnson, N., Juettemann, T., and Flicek, P.R. (2015). The ensembl regulatory build. Genome Biol *16*, 56. https://doi.org/10.1186/s13059-015-0621-5.

33. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res *38*, e164. https://doi.org/10.1093/nar/gkq603.

34. Bittner, N., Shi, C., Zhao, D., Ding, J., Southam, L., Swift, D., Kreitmaier, P., Tutino, M., Stergiou, O., Cheung, J.T.S., et al. (2024). Primary osteoarthritis chondrocyte map of chromatin conformation reveals novel candidate effector genes. Ann Rheum Dis *83*, 1048–1059. https://doi.org/10.1136/ard-2023-224945.

35. Li, D., Purushotham, D., Harrison, J.K., Hsu, S., Zhuo, X., Fan, C., Liu, S., Xu, V., Chen, S., Xu, J., et al. (2022). WashU Epigenome Browser update 2022. Nucleic Acids Research *50*, W774–W781. https://doi.org/10.1093/nar/gkac238.

36. Liu, Y., Chang, J.-C., Hon, C.-C., Fukui, N., Tanaka, N., Zhang, Z., Lee, M.T.M., and Minoda, A. (2018). Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis. Sci Rep *8*, 15499. https://doi.org/10.1038/s41598-018-33779-z.

37. Ferguson, G.B., Van Handel, B., Bay, M., Fiziev, P., Org, T., Lee, S., Shkhyan, R., Banks, N.W., Scheinberg, M., Wu, L., et al. (2018). Mapping molecular landmarks of human skeletal ontogeny and pluripotent stem cell-derived articular chondrocytes. Nat Commun *9*, 3634. https://doi.org/10.1038/s41467-018-05573-y.

38. Sanderson, E., Glymour, M.M., Holmes, M.V., Kang, H., Morrison, J., Munafò, M.R., Palmer, T., Schooling, C.M., Wallace, C., Zhao, Q., et al. (2022). Mendelian randomization. Nat Rev Methods Primers *2*, 6. https://doi.org/10.1038/s43586-021-00092-5.

39. Zuber, V., Grinberg, N.F., Gill, D., Manipur, I., Slob, E.A.W., Patel, A., Wallace, C., and Burgess, S. (2022). Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. Am J Hum Genet *109*, 767–782. https://doi.org/10.1016/j.ajhg.2022.04.001.

40. Hadjiargyrou, M. (2018). Mustn1: A Developmentally Regulated Pan-Musculoskeletal Cell Marker and Regulatory Gene. Int J Mol Sci *19*, 206. https://doi.org/10.3390/ijms19010206.

41. Gersch, R.P., and Hadjiargyrou, M. (2009). Mustn1 is expressed during chondrogenesis and is necessary for chondrocyte proliferation and differentiation in vitro. Bone *45*, 330–338. https://doi.org/10.1016/j.bone.2009.04.245.

42. Ducommun, S., Jannig, P.R., Cervenka, I., Murgia, M., Mittenbühler, M.J., Chernogubova, E., Dias, J.M., Jude, B., Correia, J.C., Van Vranken, J.G., et al. (2024). Mustn1 is a smooth muscle cell-secreted microprotein that modulates skeletal muscle extracellular matrix composition. Mol Metab *82*, 101912. https://doi.org/10.1016/j.molmet.2024.101912.

43. Rusina, P.V., Falaguera, M.J., Romero, J.M.R., McDonagh, E.M., Dunham, I., and Ochoa, D. (2023). Genetic support for FDA-approved drugs over the past decade. Nat Rev Drug Discov *22*, 864. https://doi.org/10.1038/d41573-023-00158-x.

44. Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Gonzalez-Uriarte, A., Malangone, C., Miranda, A., Fumis, L., Carvalho-Silva, D., Spitzer, M., et al. (2021). Open Targets Platform: supporting systematic drug-target identification and prioritisation. Nucleic Acids Res *49*, D1302–D1310. https://doi.org/10.1093/nar/gkaa1027.

45. Raza, A., Galili, N., Smith, S., Godwin, J., Lancet, J., Melchert, M., Jones, M., Keck, J.G., Meng, L., Brown, G.L., et al. (2009). Phase 1 multicenter dose-escalation study of ezatiostat hydrochloride (TLK199 tablets), a novel glutathione analog prodrug, in patients with myelodysplastic syndrome. Blood *113*, 6533–6540. https://doi.org/10.1182/blood-2009-01-176032.

46. Raza, A., Galili, N., Smith, S.E., Godwin, J., Boccia, R.V., Myint, H., Mahadevan, D., Mulford, D., Rarick, M., Brown, G.L., et al. (2012). A phase 2 randomized multicenter study of 2 extended dosing schedules of oral ezatiostat in low to intermediate-1 risk myelodysplastic syndrome. Cancer *118*, 2138–2147. https://doi.org/10.1002/cncr.26469.

47. Hasanvand, A. (2022). The role of AMPK-dependent pathways in cellular and molecular mechanisms of metformin: a new perspective for treatment and prevention of diseases. Inflammopharmacology *30*, 775–788. https://doi.org/10.1007/s10787-022-00980-6.

48. Baker, M.C., Sheth, K., Liu, Y., Lu, D., Lu, R., and Robinson, W.H. (2023). Development of Osteoarthritis in Adults With Type 2 Diabetes Treated With Metformin vs a Sulfonylurea. JAMA Netw Open *6*, e233646. https://doi.org/10.1001/jamanetworkopen.2023.3646.

49. Song, Y., Wu, Z., and Zhao, P. (2022). The effects of metformin in the treatment of osteoarthritis: Current perspectives. Front Pharmacol *13*, 952560. https://doi.org/10.3389/fphar.2022.952560.

50. Mostafavi, H., Spence, J.P., Naqvi, S., and Pritchard, J.K. (2023). Systematic differences in discovery of genetic effects on gene expression and complex traits. Nat Genet *55*, 1866–1875. https://doi.org/10.1038/s41588-023-01529-1.

51. Minikel, E.V., Painter, J.L., Dong, C.C., and Nelson, M.R. (2024). Refining the impact of genetic evidence on clinical success. Nature *629*, 624–629. https://doi.org/10.1038/s41586-024-07316-0.

52. Arruda, A.L., Morris, A.P., and Zeggini, E. (2024). Advancing equity in human genomics through tissue-specific multi-ancestry molecular data. Cell Genomics *4*. https://doi.org/10.1016/j.xgen.2023.100485.

53. Swahn, H., Li, K., Duffy, T., Olmer, M., D'Lima, D.D., Mondala, T.S., Natarajan, P., Head, S.R., and Lotz, M.K. (2023). Senescent cell population with ZEB1 transcription factor as its main regulator promotes osteoarthritis in cartilage and meniscus. Ann Rheum Dis *82*, 403–415. https://doi.org/10.1136/ard-2022-223227.

54. Chou, C.-H., Jain, V., Gibson, J., Attarian, D.E., Haraden, C.A., Yohn, C.B., Laberge, R.-M., Gregory, S., and Kraus, V.B. (2020). Synovial cell cross-talk with cartilage plays a major role in the pathogenesis of osteoarthritis. Sci Rep *10*, 10868. https://doi.org/10.1038/s41598-020-67730-y.

55. Tang, S., Yao, L., Ruan, J., Kang, J., Cao, Y., Nie, X., Lan, W., Zhu, Z., Han, W., Liu, Y., et al. (2024). Single-cell atlas of human infrapatellar fat pad and synovium implicates APOE signaling in osteoarthritis pathology. Sci Transl Med *16*, eadf4590. https://doi.org/10.1126/scitranslmed.adf4590.

56. Westra, H.-J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet *45*, 1238–1243. https://doi.org/10.1038/ng.2756.

57. van den Borne, M.P.J., Raijmakers, N.J.H., Vanlauwe, J., Victor, J., de Jong, S.N., Bellemans, J., and Saris, D.B.F. (2007). International Cartilage Repair Society (ICRS) and Oswestry macroscopic cartilage evaluation scores validated for use in Autologous Chondrocyte Implantation (ACI) and microfracture. Osteoarthritis and Cartilage *15*, 1397–1402. https://doi.org/10.1016/j.joca.2007.05.005.

58. Kreitmaier, P., Park, Y.-C., Swift, D., Gilly, A., Wilkinson, J.M., and Zeggini, E. (2023). Epigenomic profiling of the infrapatellar fat pad in osteoarthritis. Hum Mol Genet, ddad198. https://doi.org/10.1093/hmg/ddad198.

59. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

60. Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185. https://doi.org/10.1093/bioinformatics/bts356.

61. Chen, X., Zhang, B., Wang, T., Bonni, A., and Zhao, G. (2020). Robust principal component analysis for accurate outlier sample detection in RNA-Seq data. BMC Bioinformatics *21*, 269. https://doi.org/10.1186/s12859-020-03608-0.

62. Auwera, G. van der, and O'Connor, B.D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (O'Reilly Media, Incorporated).

63. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.

64. Dong, X., Li, X., Chang, T.-W., Scherzer, C.R., Weiss, S.T., and Qiu, W. (2021). powerEQTL: an R package and shiny application for sample size and power calculation of bulk tissue and single-cell eQTL analysis. Bioinformatics *37*, 4269–4271. https://doi.org/10.1093/bioinformatics/btab385.

65. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology *11*, R25. https://doi.org/10.1186/gb-2010-11-3-r25.

66. Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput Biol *6*, e1000770. https://doi.org/10.1371/journal.pcbi.1000770.

67. The GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.

68. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

69. Taylor-Weiner, A., Aguet, F., Haradhvala, N.J., Gosai, S., Anand, S., Kim, J., Ardlie, K., Van Allen, E.M., and Getz, G. (2019). Scaling computational genomics to millions of individuals with GPUs. Genome Biol *20*, 228. https://doi.org/10.1186/s13059-019-1836-7.

70. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358. https://doi.org/10.1093/bioinformatics/bts163.

71. Saha, A., and Battle, A. (2018). False positives in trans-eQTL and co-expression analyses arising from RNA-sequencing alignment errors. F1000Res *7*, 1860. https://doi.org/10.12688/f1000research.17145.2.

72. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559–575. https://doi.org/10.1086/519795.

73. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res *29*, 308–311. https://doi.org/10.1093/nar/29.1.308.

74. Wickham, H. (2009). ggplot2, Voice In Settings: Elegant Graphics for Data Analysis (Springer-Verlag) https://doi.org/10.1007/978-0-387-98141-3.

75. Urbut, S.M., Wang, G., Carbonetto, P., and Stephens, M. (2019). Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. Nat Genet *51*, 187–195. https://doi.org/10.1038/s41588-018-0268-8.

76. Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc *10*, 1556–1566. https://doi.org/10.1038/nprot.2015.105.

77. Wen, X. (2016). Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. The Annals of Applied Statistics *10*, 1619–1638. https://doi.org/10.1214/16-AOAS952.

78. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet *47*, 1228–1235. https://doi.org/10.1038/ng.3404.

79. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209. https://doi.org/10.1038/s41586-018-0579-z.

80. Cunningham, F., Allen, J.E., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Bennett, R., et al. (2022). Ensembl 2022. Nucleic Acids Res *50*, D988–D995. https://doi.org/10.1093/nar/gkab1049.

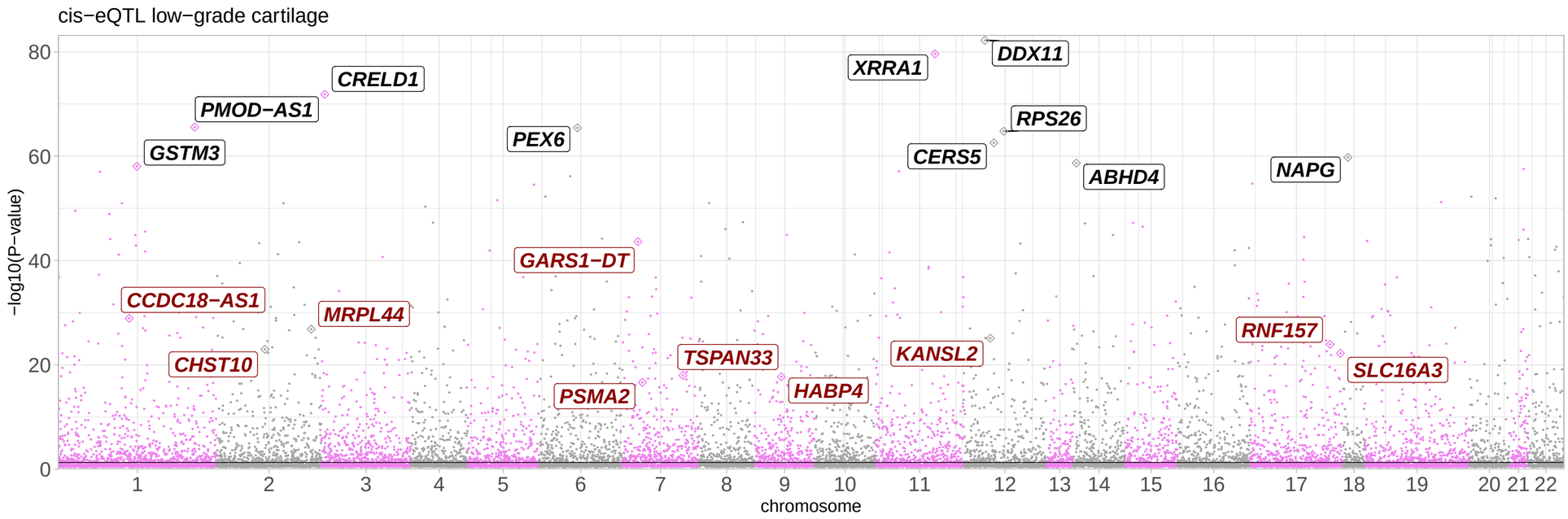
81. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. Elife *7*, e34408. https://doi.org/10.7554/eLife.34408.

82. Kamburov, A., Pentchev, K., Galicka, H., Wierling, C., Lehrach, H., and Herwig, R. (2011). ConsensusPathDB: toward a more complete picture of cell biology. Nucleic Acids Res *39*, D712-717. https://doi.org/10.1093/nar/gkq1156.

83. Amberger, J.S., Bocchini, C.A., Scott, A.F., and Hamosh, A. (2019). OMIM.org: leveraging knowledge across phenotype–gene relationships. Nucleic Acids Research *47*, D1038–D1043. https://doi.org/10.1093/nar/gky1151.

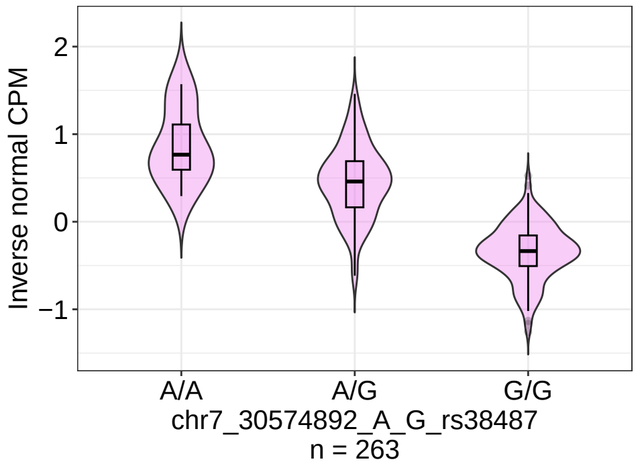
84. Baldarelli, R.M., Smith, C.L., Ringwald, M., Richardson, J.E., Bult, C.J., and Mouse Genome Informatics Group (2024). Mouse Genome Informatics: an integrated knowledgebase system for the laboratory mouse. Genetics, iyae031. https://doi.org/10.1093/genetics/iyae031.

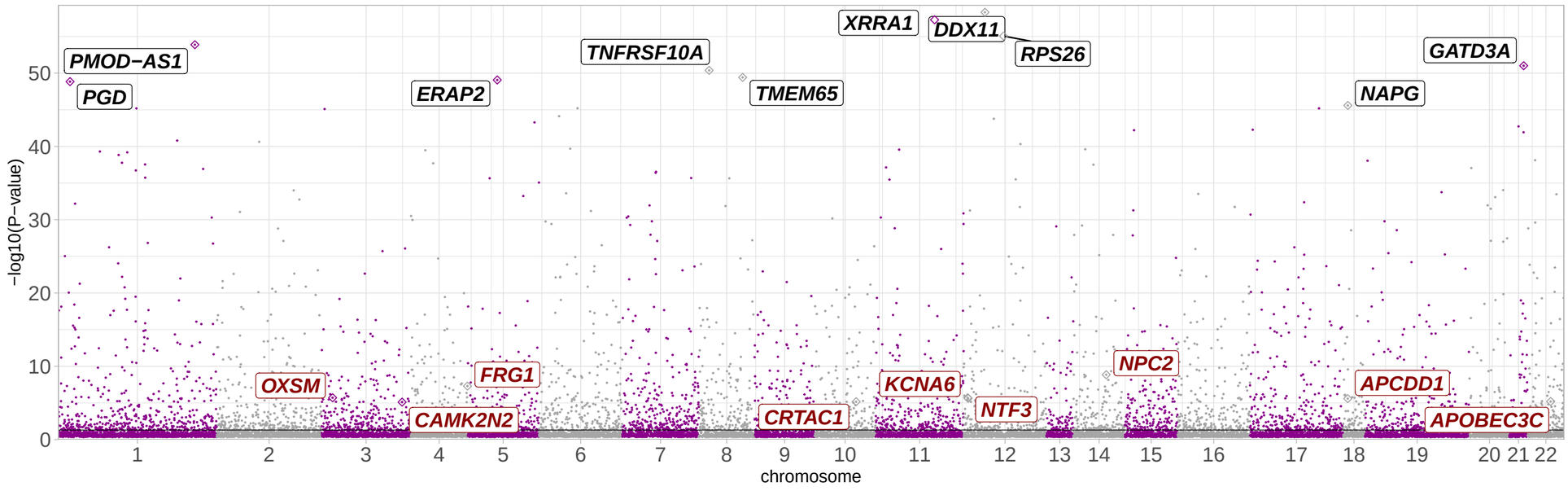
85. Zdrazil, B., Felix, E., Hunter, F., Manners, E.J., Blackshaw, J., Corbett, S., de Veij, M., Ioannidis, H., Lopez, D.M., Mosquera, J.F., et al. (2024). The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Research *52*, D1180–D1192. https://doi.org/10.1093/nar/gkad1004.



Low−grade cartilage cis−eQTL

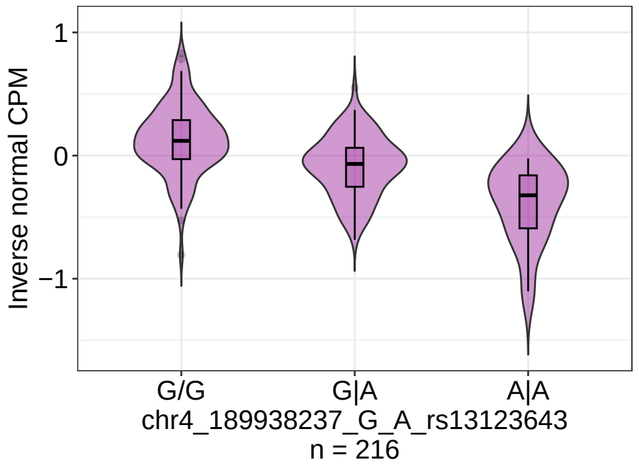
p−value = 2.04e−53
slope = −0.9
GARS1−DT

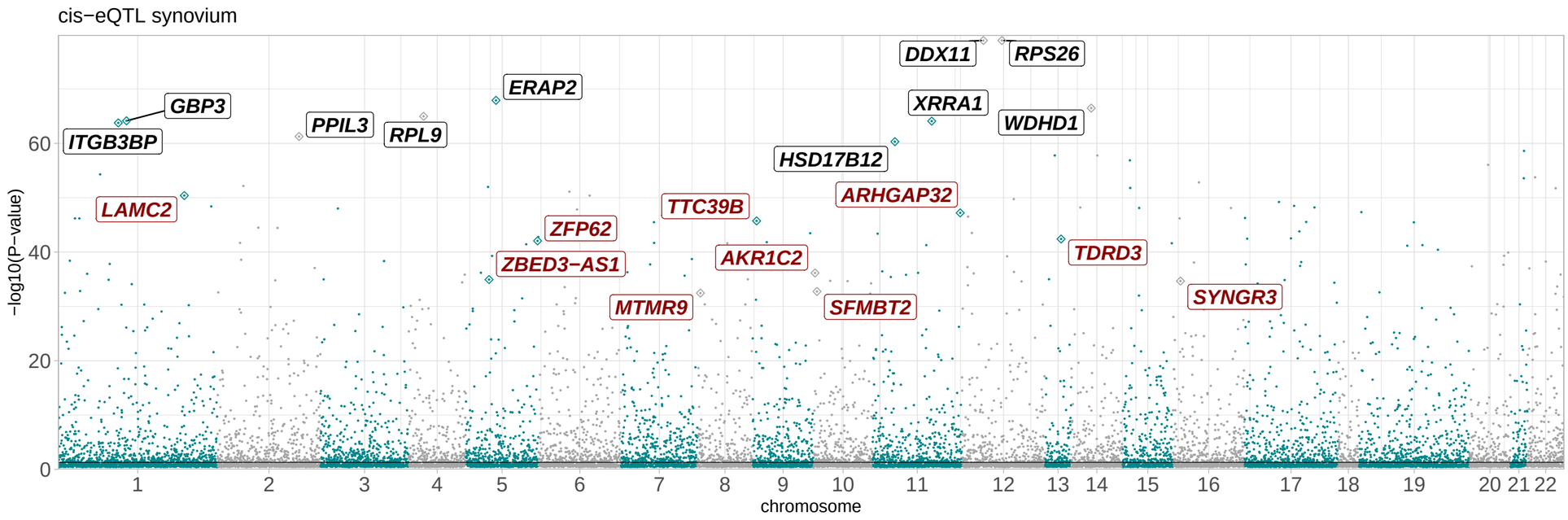




High−grade cartilage cis−eQTL

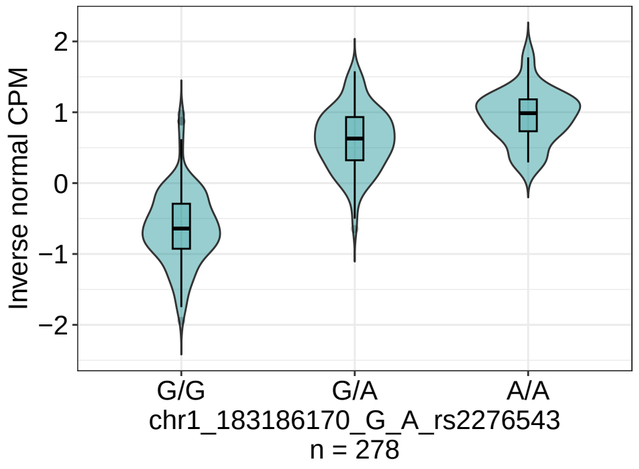
p−value = 1.06e−12
slope = −0.27
FRG1

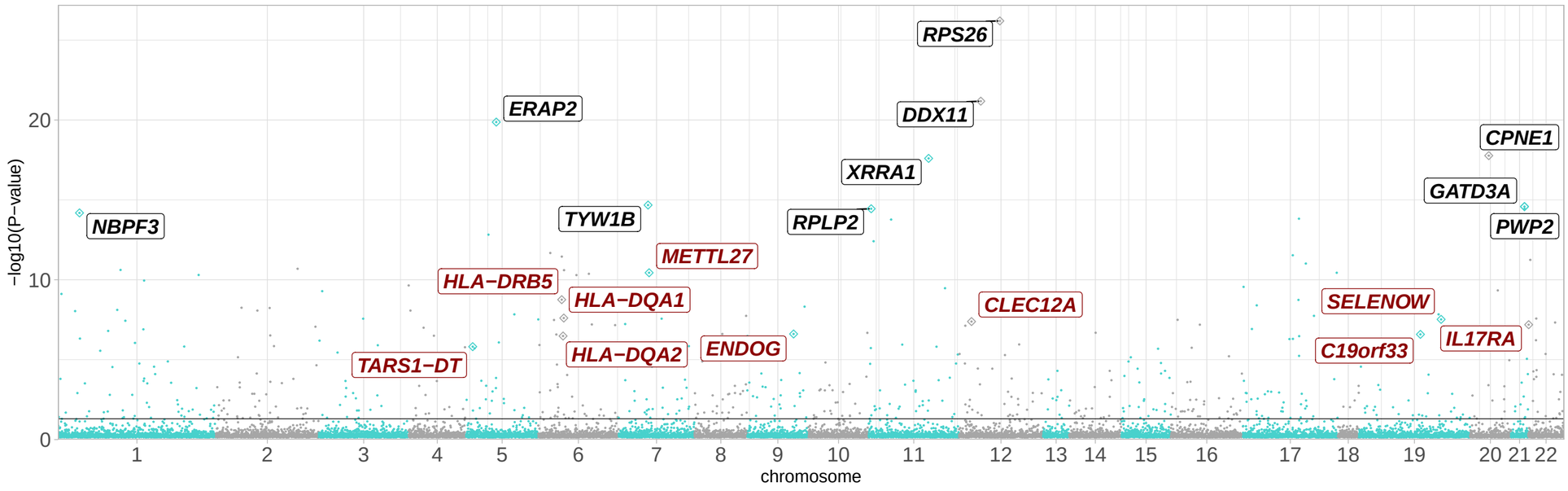




Synovium cis−eQTL

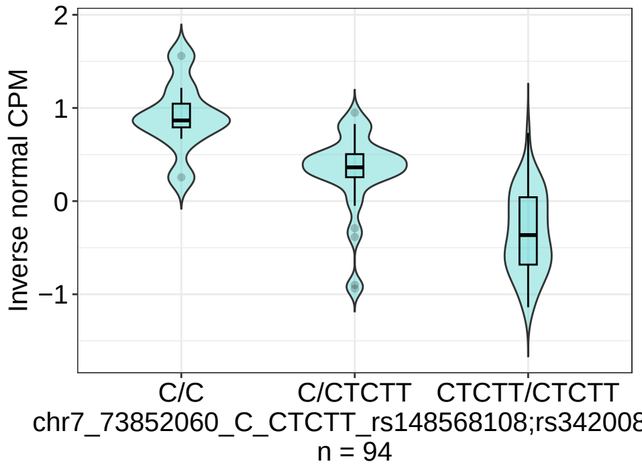
p−value = 2.19e−62
slope = 1.2
LAMC2

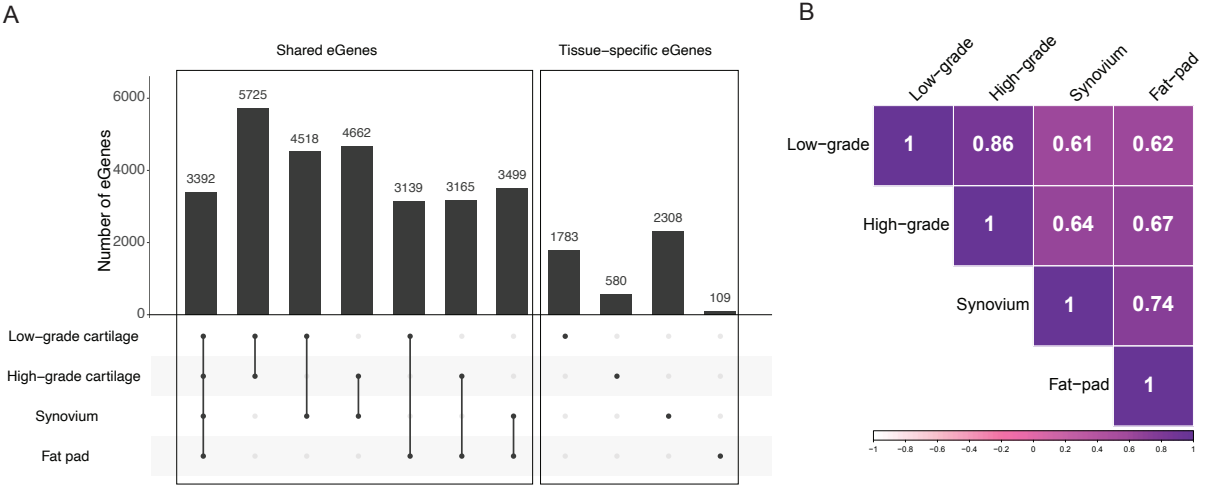


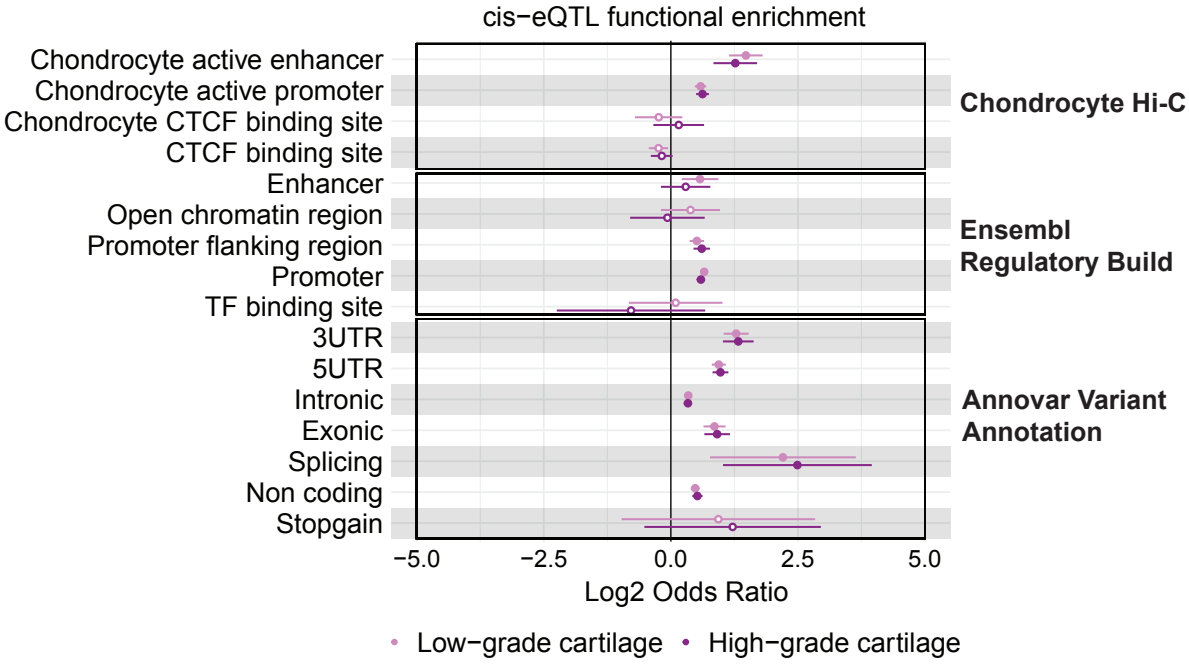


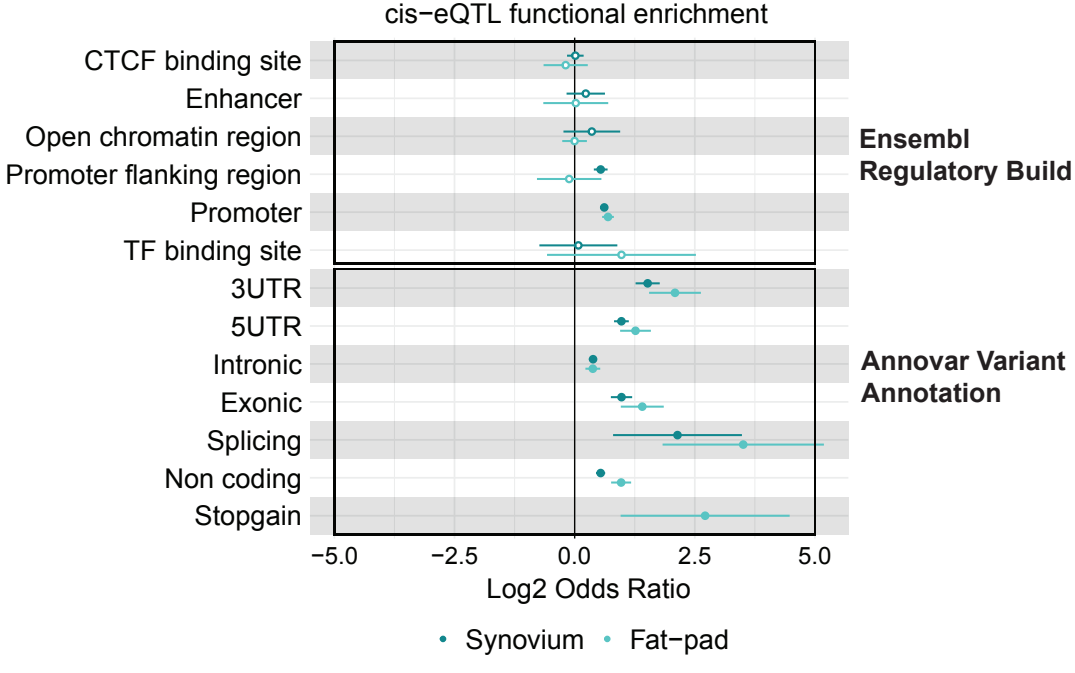
Fat−pad cis−eQTL

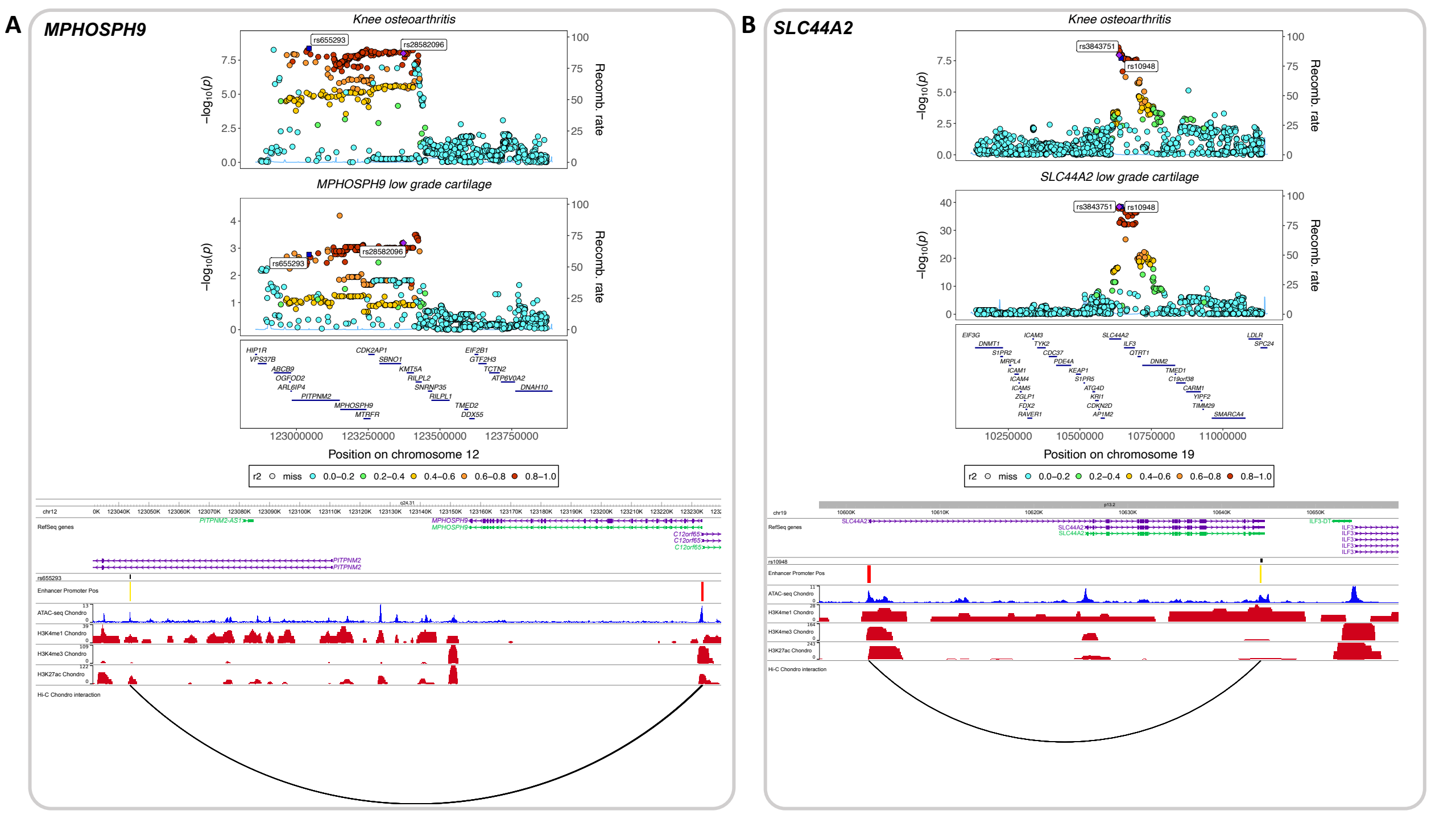
p−value = 8.53e−18
slope = −0.82
METTL27

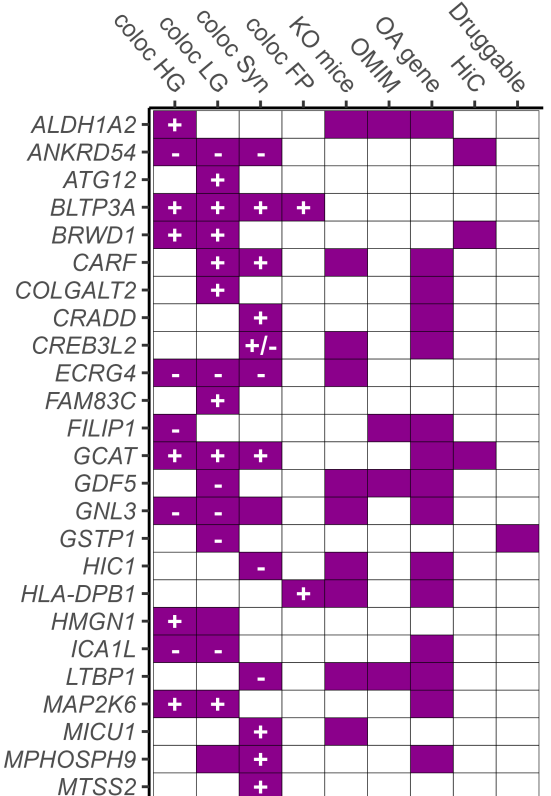


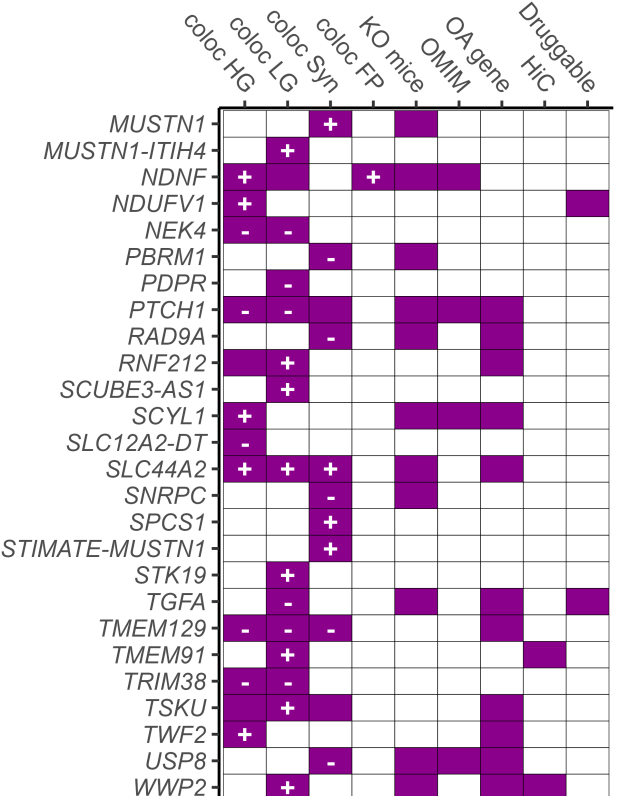

A
Shared eGenes
Tissue−specific eGenes
Number of eGenes
6000
4000
2000
0
3392
5725
4518
4662
3139
3165
3499
1783
580
2308
109
Low−grade cartilage
High−grade cartilage
Synovium
Fat pad
B
Low−grade
High−grade
Synovium
Fat−pad
1
0.86
0.61
0.62
1
0.64
0.67
1
0.74
1
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

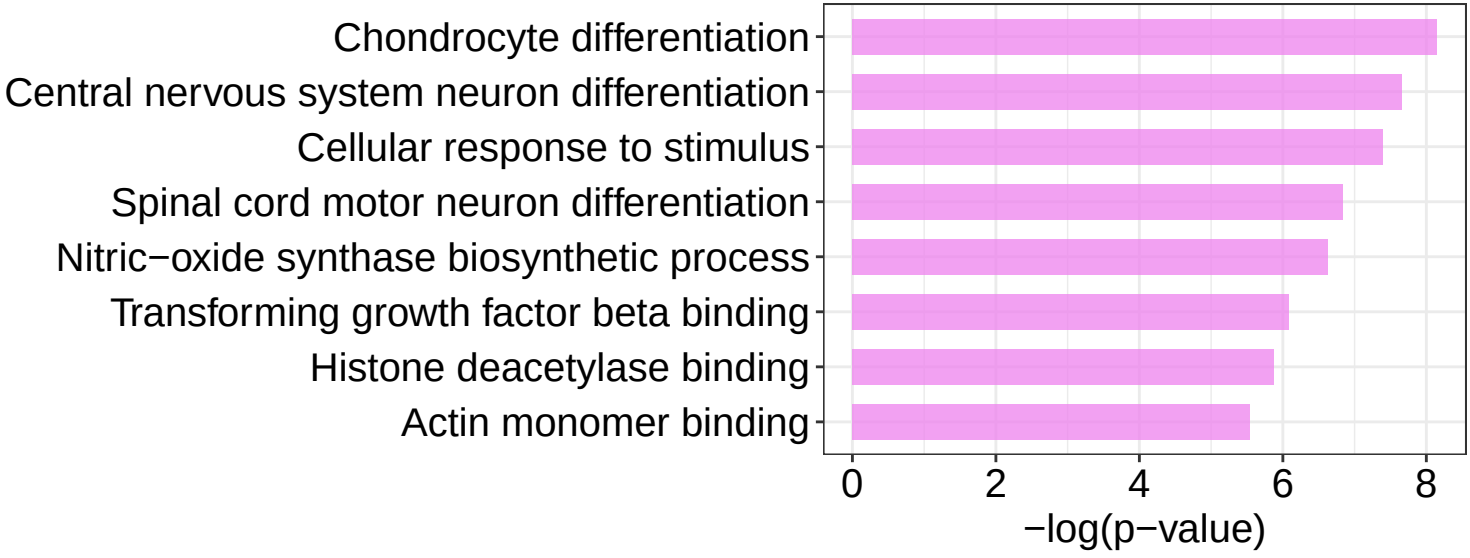

cis−eQTL functional enrichment
Chondrocyte active enhancer
Chondrocyte active promoter
Chondrocyte CTCF binding site
CTCF binding site
Enhancer
Open chromatin region
Promoter flanking region
Promoter
TF binding site
3UTR
5UTR
Intronic
Exonic
Splicing
Non coding
Stopgain
Chondrocyte Hi-C
Ensembl Regulatory Build
Annovar Variant Annotation
−5.0
−2.5
0.0
2.5
5.0
Log2 Odds Ratio
Low−grade cartilage
High−grade cartilage


cis−eQTL functional enrichment
CTCF binding site
Enhancer
Open chromatin region
Promoter flanking region
Promoter
TF binding site
3UTR
5UTR
Intronic
Exonic
Splicing
Non coding
Stopgain
Ensembl Regulatory Build
Annovar Variant Annotation
−5.0
−2.5
0.0
2.5
5.0
Log2 Odds Ratio
Synovium
Fat−pad


A
MPHOSPH9
Knee osteoarthritis
rs655293
rs28582096
−log10(p)
Recomb. rate
MPHOSPH9 low grade cartilage
HIP1R
VPS37B
ABCB9
OGFOD2
ARL6IP4
PITPNM2
MPHOSPH9
MTRFR
CDK2AP1
SBNO1
KMT5A
RILPL2
SNRNP35
RILPL1
TMED2
DDX55
EIF2B1
GTF2H3
TCTN2
ATP6V0A2
DNAH10
123000000
123250000
123500000
123750000
Position on chromosome 12
r2 miss 0.0–0.2 0.2–0.4 0.4–0.6 0.6–0.8 0.8–1.0
RefSeq genes
Enhancer Promoter Pos
ATAC-seq Chondro
H3K4me1 Chondro
H3K4me3 Chondro
H3K27ac Chondro
Hi-C Chondro interaction
B
SLC44A2
Knee osteoarthritis
rs3843751
rs10948
SLC44A2 low grade cartilage
EIF3G
DNMT1
S1PR2
MRPL4
ICAM1
ICAM4
ICAM5
ZGLP1
FDX2
RAVER1
ICAM3
TYK2
CDC37
PDE4A
KEAP1
S1PR5
ATG4D
KRI1
CDKN2D
AP1M2
SLC44A2
ILF3
QTRT1
DNM2
TMED1
C19orf38
CARM1
YIPF2
TIMM29
SMARCA4
LDLR
SPC24
10250000
10500000
10750000
11000000
Position on chromosome 19
r2 miss 0.0–0.2 0.2–0.4 0.4–0.6 0.6–0.8 0.8–1.0
RefSeq genes
Enhancer Promoter Pos
ATAC-seq Chondro
H3K4me1 Chondro
H3K4me3 Chondro
H3K27ac Chondro
Hi-C Chondro interaction


coloc HG
coloc LG
coloc Syn
coloc FP
KO mice
OMIM
OA gene
Druggable
HiC
ALDH1A2
ANKRD54
ATG12
BLTP3A
BRWD1
CARF
COLGALT2
CRADD
CREB3L2
ECRG4
FAM83C
FILIP1
GCAT
GDF5
GNL3
GSTP1
HIC1
HLA-DPB1
HMGN1
ICA1L
LTBP1
MAP2K6
MICU1
MPHOSPH9
MTSS2


coloc HG
coloc LG
coloc Syn
coloc FP
KO mice
OMIM
OA gene
Druggable
HiC
MUSTN1
MUSTN1-ITIH4
NDNF
NDUFV1
NEK4
PBRM1
PDPR
PTCH1
RAD9A
RNF212
SCUBE3-AS1
SCYL1
SLC12A2-DT
SLC44A2
SNRPC
SPCS1
STIMATE-MUSTN1
STK19
TGFA
TMEM129
TMEM91
TRIM38
TSKU
TWF2
USP8
WWP2


Chondrocyte differentiation
Central nervous system neuron differentiation
Cellular response to stimulus
Spinal cord motor neuron differentiation
Nitric−oxide synthase biosynthetic process
Transforming growth factor beta binding
Histone deacetylase binding
Actin monomer binding
0
2
4
6
8
−log(p−value)