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

Georgia Katsoula[1,2,3,*], Ana Luiza Arruda[2,*], Mauro Tutino[2], Ene Reimann[2,4], Peter Kreitmaier[1,2,3], Karan M Shah[5], Diane Swift[5], Lorraine Southam[2], Siim Suutre[6], Galadriel Luzia Velazquez Silva[4], Kaspar Tootsi[7,8], Aare Märtson[7,8], Reedik Mägi[4], J Mark Wilkinson[5,×], Eleftheria Zeggini[2,3,×]

[1] Technical University of Munich (TUM) Klinikum, TUM School of Medicine and Health, 81675 Munich, 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] Estonian Genome Centre, Institute of Genomics, University of Tartu, 51010 Tartu, Estonia

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

[6]Department of Anatomy, University of Tartu, Ravila 19, 50411 Tartu, Estonia

[7]Department of Orthopaedics, University of Tartu, Puusepa 8, 50406 Tartu, Estonia

[8]Orthopaedics Clinic, Tartu University Hospital, Puusepa 8, 50406 Tartu, Estonia.

* equal contribution

× equal contribution

Abstract

Osteoarthritis, a whole-joint degenerative disorder, is a major public health burden that affects nearly 600 million individuals worldwide, but no disease-modifying treatment exists. Molecular profiling of relevant tissues is crucial for understanding the biological mechanisms underlying disease development. Here, we generate a comprehensive map of transcriptional regulation in 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. Using the largest osteoarthritis genome-wide association study (GWAS) to date, we find colocalization evidence with 117 genes, 67 of which had not been identified as effector genes before. We prioritise 38 high-confidence effector genes for osteoarthritis, based on multiple lines of molecular and functional genomics evidence, including *MUSTN1*, which is implicated in cartilage integrity. Our study provides insights into the molecular mechanisms underpinning osteoarthritis and offers much-needed drug repurposing opportunities.

Introduction

Osteoarthritis, the most prevalent form of arthritis, is a chronic degenerative complex disease that is projected to affect over one 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 960 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 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.

Expression quantitative trait loci (eQTLs) 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 Han

Chinese 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 macroscopically intact (low-grade) and degraded (high-grade) cartilage, synovium, and fat pad. This is the first eQTL study for fat pad tissue and the sample size for the other three tissues is increased substantially compared to previous studies (low-grade cartilage: 3.02 times, high-grade cartilage: 2.25, synovium: 1.38 times). We identify shared and disease 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

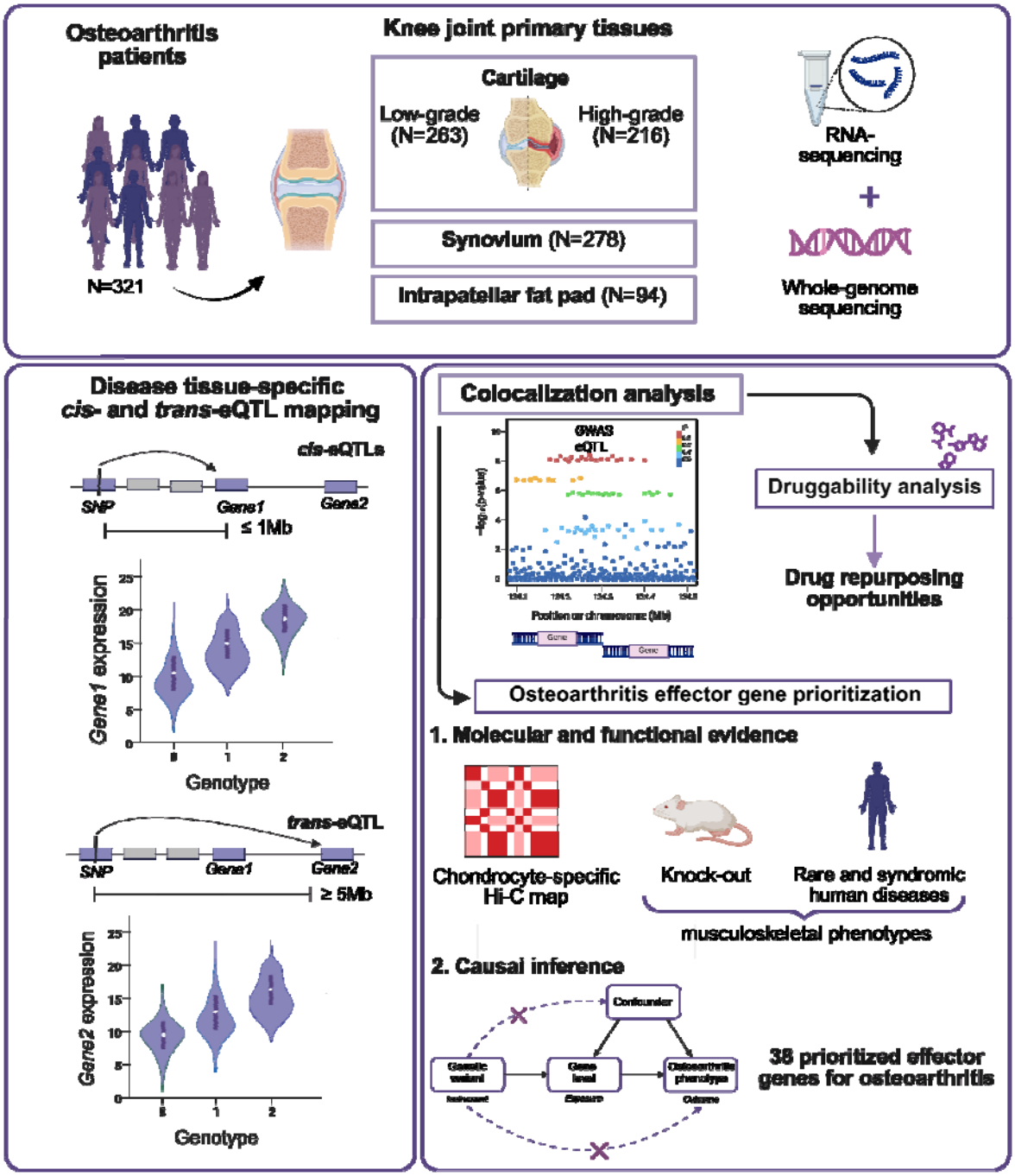


Figure 1: Study design.

Osteoarthritis primary tissues *cis*-eQTL discovery

We performed *cis*-eQTL analyses in 321 knee osteoarthritis patients (Figure 1, 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 2, Table S2). In synovium we find 6,411 *cis*-eGenes targeted by 8,043 conditionally independent *cis*-eQTLs (Figure 2, Table S2), and in fat pad osteoarthritis tissue, we identify 592 *cis*-eGenes regulated by 606 conditionally independent *cis*-eQTLs (Figure 2, Table S2).

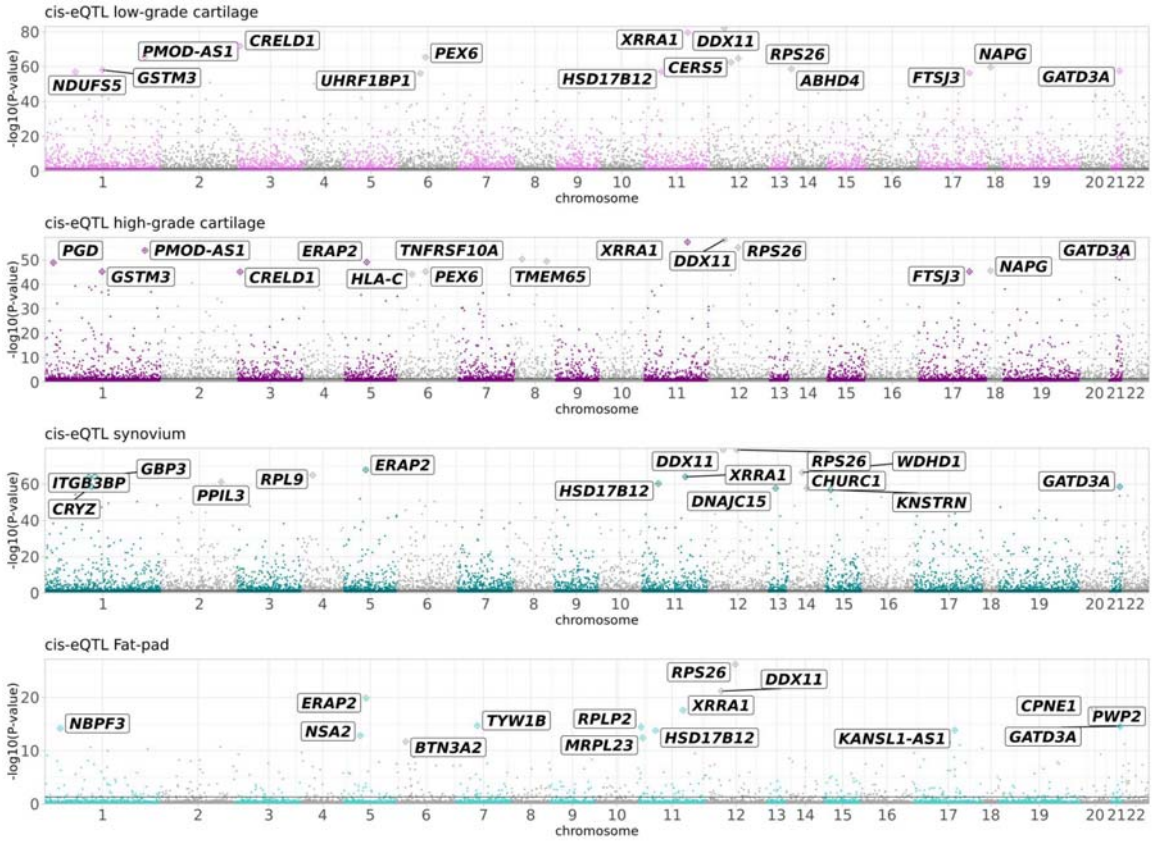


Figure 2: Genomic distribution of *cis*-eQTLs. Manhattan plots depicting the negative log of the permutation p-value (q-value) of the most significant association per gene across all variants within 1 Mb of the gene's transcription start site for low-grade cartilage, high-grade cartilage, synovium and fat pad tissues. Top 15 *cis*-eGenes per disease tissue are labeled in black.

Distinct eQTL effects in osteoarthritis tissues

Understanding how genetic regulation varies across osteoarthritis-affected tissues is important for deciphering the specificity of disease mechanisms. We sought to identify distinct genetic regulatory effects present in individual disease tissues using two complementary approaches. Firstly, we investigated eQTLs regulating genes expressed exclusively in a single disease tissue and not in others, identifying 597 unique eGenes in cartilage, 70 in synovium, and 14 in fat pad (Table S3). Secondly, to formally assess distinct genetic regulatory patterns across osteoarthritis tissues for commonly expressed genes, we performed a meta-analysis across all examined disease tissues (low-grade cartilage, high-grade cartilage, synovium and fat pad), and identify 53 additional eGenes with eQTL effects exclusive to low-grade cartilage, 24 to high-

grade cartilage, 267 to synovium, and 54 to fat pad, with no effect in other tissues (Figure S3, Tables S4-S5).

Several genes showed large eQTL effects in a single disease tissue and are related to osteoarthritis biology. *CXCL14*, a chemokine associated with osteoarthritis pain[17] and regulating connective tissue during limb development[18], shows distinct regulation of its expression in low-grade cartilage (*CXCL14:*rs35039539) (Figure S3, Table S4). *CXCL14* also acts as a natural inhibitor of the *CXCL12/CXCR4* axis, which is known to regulate chondrocyte hypertrophy[19] and is a potential therapeutic target for osteoarthritis[20].

SNAI1 (*SNAI1*:rs437276) and *CXCL5 (CXCL5:*rs17230459*)* show distinct expression regulation in high-grade cartilage (Figure S3, Table S4). *SNAI1* is a transcription factor involved in chondrocyte proliferation through repression of *COL2AI* and *ACAN* transcription *in vitro*[21] and has also been shown to act downstream of FGFR3 signaling in mouse chondrocytes, regulating both Stat and MAPK and ERK1/2 pathways. The latter are critical for chondrocyte terminal differentiation[22–24].

In the synovium, we find distinct regulation for *C1QTNF1*, *DMRT2* and *RAMP3*, among other genes (Figure S3, Tables S4-S5). *C1QTNF1* codes for an adipokine that enables collagen binding, has a role in inflammation, and is involved in chondrocyte maturation with activation of the ERK1/2 signaling pathway[25]. *DMRT2* codes for a transcription factor involved in skeletal development and endochondral bone formation[26], while *RAMP3* is involved in angiogenesis, which also takes place in osteoarthritis and is linked to pain and synovitis[27]. Several other genes with proposed roles in osteoarthritis show specific regulatory patterns in synovium including *TIMP3*, a critical inhibitor of matrix metalloproteinases (MMPs), *GSK3B* inducing osteoarthritis through activation of Wnt pathway in cartilage[28] and *SOX6* a transcription factor with established roles in synovial joint development[29].

AMPH, coding for a protein involved in synaptic vesicle endocytosis[30], shows the most distinct regulation in osteoarthritis fat pad compared to other tissues (Figure S3, Tables S4-S5), pointing to a possible role in local neuroinflammatory processes. Several further genes with fat pad–specific regulation, have been connected to osteoarthritis. For instance, *IL34* is involved in macrophage differentiation and is associated with synovitis, supporting its role in local immune activation[31]. *PER1*, a core circadian clock gene, with a proposed role in suppressing chondrocyte differentiation[32], has been linked to the temporal regulation of inflammatory mediators[33], which may contribute to daily variation in osteoarthritis symptoms. *ZEB2* is the mediator of the differentiation of age-associated B cells[34], which may contribute to the low-grade inflammation observed in osteoarthritis Together, these genes underscore the unique immunoregulatory and neuroimmune environment of the osteoarthritis fat pad.

Shared eQTL effects among osteoarthritis primary tissues

In line with GTEx findings[35], we observe widespread sharing of eQTL effects across osteoarthritis-relevant tissues, reflecting common regulatory architecture. Specifically, 72.3% of *cis*-eQTLs are shared across all four tissues when considering both effect direction and magnitude (see Methods). Tissue pairs with closer biological similarity show even greater

overlap, with 85.9% of *cis*-eQTLs shared between low- and high-grade cartilage, and 73.6% between synovium and fat pad (Figure 3A).

Nevertheless, a subset of 26 eQTL–eGene pairs, though significant across all disease-relevant tissues, exhibit discordant effects (Tables S4-S5), showing opposite directions of association in at least one tissue (Figure 3B). In 14 out of 26 cases, the direction of effect was concordant between low- and high-grade cartilage, as well as between synovium and fat pad, reflecting greater similarity within these tissue pairs. However, we also identified cases with opposite effects in a single tissue. For example, the insertion allele (TGCAC) of the rs70956447 variant (reference allele T), associated with *ENPEP* expression, shows a positive effect on *ENPEP* expression in cartilage (both low- and high-grade) and synovium, but a negative effect in the expression of the same gene in the fat pad. *ENPEP* encodes glutamyl aminopeptidase, an enzyme involved in angiogenesis and also a marker of common skeletal progenitors in the postnatal limb bone[36]. Among the genes with opposite effects specifically in fat pad, we observed several involved in ECM remodeling and previously linked to osteoarthritis, including *MATN3* and *MIA3* (Figure 3B). These discordant eQTLs may reflect tissue-specific regulatory mechanisms or context-dependent gene regulation relevant to osteoarthritis pathophysiology.

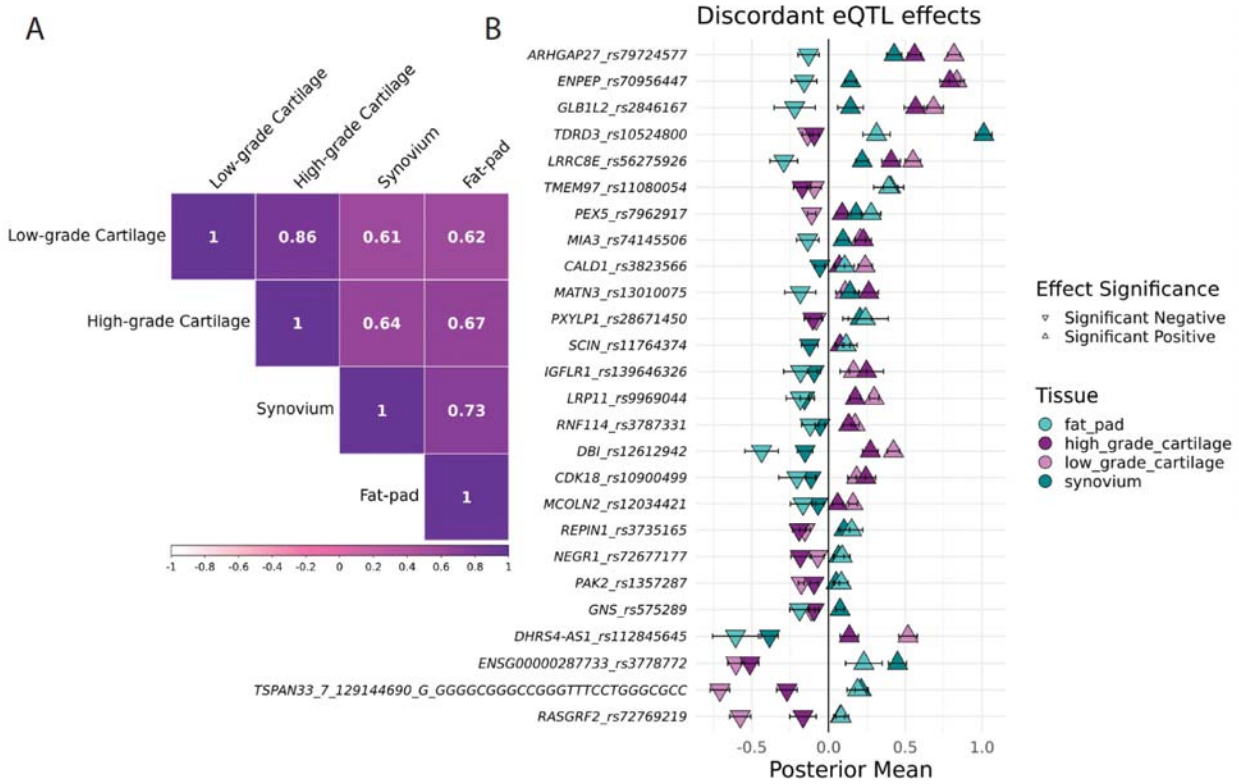


Figure 3: **Patterns of shared and discordant eQTL effects among osteoarthritis primary tissues.** (A) Heatmap showing pairwise sharing by magnitude of *cis*-eQTLs among tissues. (B) 26 eQTL–eGene pairs showing opposite directions of effect between osteoarthritis tissues.

Disease progression eQTLs in osteoarthritis cartilage

Cartilage was the only tissue that included both low- and high-grade osteoarthritis samples, enabling the investigation of stage-specific regulatory changes that may reflect disease progression. We identified 117 eQTL–eGene pairs with effects exclusive to low-grade cartilage and 38 pairs with effects specific to high-grade cartilage (Figure S4, Tables S5-S6).

Among the genes with low-grade-specific regulation, several are implicated in pathways central to chondrocyte function and cartilage integrity. These include *PIK3R1*, *MTOR*, *INSR*, and *PREX1*, key components of the PI3K/AKT/mTOR signaling axis, which regulates chondrocyte survival, metabolism, and autophagy[37]. We also observed regulation of *PTGER2*, a prostaglandin receptor linked to cartilage inflammation[38], and *ADAMTS5*, a central aggrecanase in osteoarthritis pathophysiology[39]. Additional low-grade-specific effects were found for *AHR*, which plays an important role in the immune system within the skeletal niche and the differentiation of mesenchymal stem cells into other cell lineages including chondrocytes and adipocytes[40] and *FOXJ1*, a ciliogenesis regulator with emerging roles in chondrocyte mechanotransduction[41]. These findings highlight that regulatory variation in early-stage osteoarthritis preferentially affects genes involved in signaling, inflammation, and matrix turnover in cartilage.

In high-grade cartilage, several eQTLs were associated with genes involved in oxidative stress (*MT1F*, *OSGIN2*), inflammation (*CXCL5*, *SELE*), and DNA repair (*RAD50*, *NUP98*). *TMEM165*, important for glycosaminoglycan chain elongation and chondrocyte maturation[42], also showed high-grade-specific regulation. Additionally, we identified lncRNAs *ALDH1A3-AS1* and *LINC01614* as targets of high-grade-specific eQTLs. *ALDH1A3-AS1* may influence retinoic acid signaling via its sense gene *ALDH1A3*, which is associated with higher expression of chondrocyte-specific genes such as *COL2A1* and *SOX9*[43]. *LINC01614*, previously reported to be upregulated in lesioned versus preserved osteoarthritis cartilage[44,45], may contribute to disease-related transcriptional changes. Together, these findings suggest that regulatory variation in advanced-stage osteoarthritis increasingly impacts genes involved in cellular stress responses, matrix regulation and transcriptional reprogramming.

Notably, we identified four eQTL–eGene pairs that exhibited significant associations in both low- and high-grade cartilage, but with opposite directions of effect (Figure 4A, Table S7). These included *TTPAL*, *MAF1*, *SPATA13*, and *TMEM220* (Figure 4B). Such directionally discordant effects may reflect dynamic, stage-specific regulatory mechanisms, potentially altering gene function or cellular context as osteoarthritis advances.

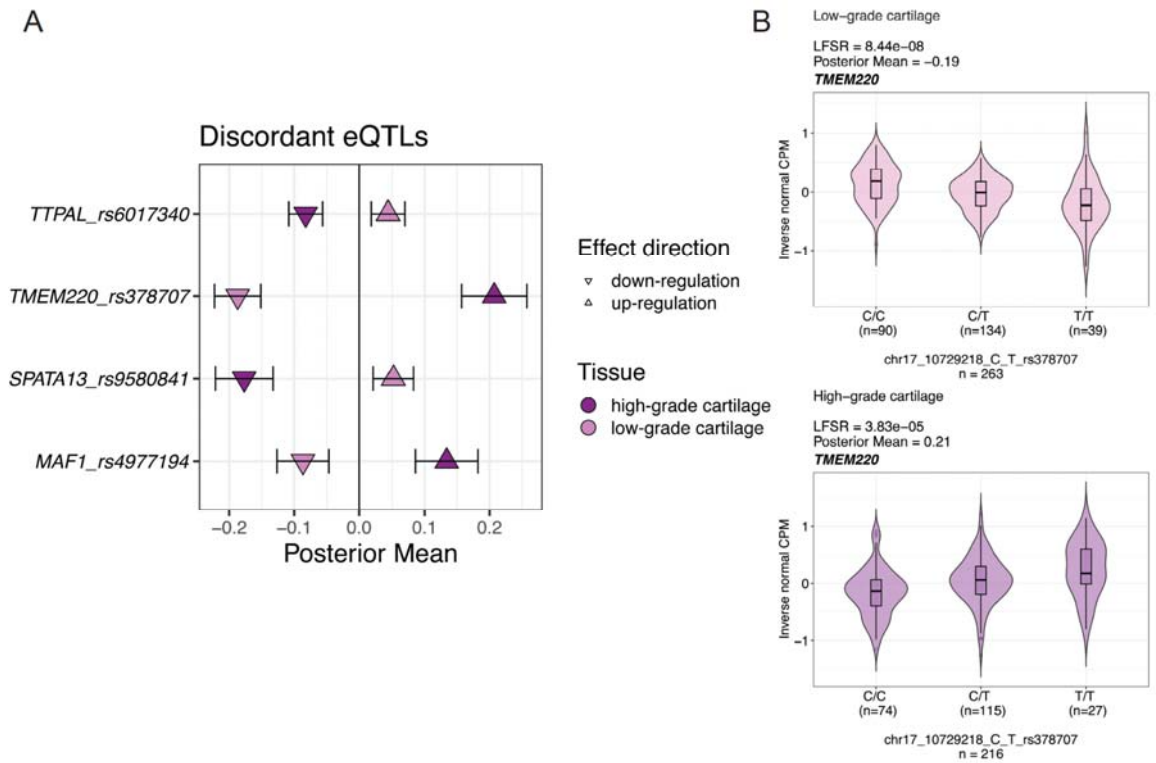


Figure 4: Disease progression eQTLs. (A) Four eQTL–eGene pairs exhibit opposite directions of effect between low- and high-grade osteoarthritis cartilage, indicating stage-specific regulatory changes.
(B) Violin plot illustrating the rs378707 C allele, which is associated with decreased *TMEM220* expression in low-grade cartilage but increased expression in high-grade cartilage.

Trans-eQTLs is osteoarthritis primary tissues

We investigated distal gene regulation (≥5 Mb from transcription start sites) in osteoarthritis primary tissues. In low-grade cartilage, we identified 363 *trans*-eQTLs (12 independent signals) regulating 12 *trans*-eGenes; in high-grade cartilage, 180 *trans*-eQTLs (5 independent) for 5 genes; in synovium, 163 *trans*-eQTLs (6 independent) for 5 genes; and a single *trans*-eQTL–eGene pair in fat pad. Four *trans*-eGenes - *UQCRHL*, *ORC6*, *MYLK3*, and *KANSL1-AS1* - were shared between low- and high-grade cartilage, with *KANSL1-AS1* also appearing in fat pad and *UQCRHL* in synovium (Table S8, Figures S5-S6).

Most *trans*-eGenes were also *cis*-regulated: 9/12 in low-grade cartilage, 4/5 in high-grade, and 2/5 in synovium. These included genes linked to mitochondrial function and metabolism (*UQCRHL*, *TEFM*, *COQ8A*), cell cycle and DNA replication (*ORC6*, *CDK10*), and protein degradation (*DCUN1D5*, *TRIM69*). Notably, *CRHR1* and *KANSL1-AS1*, both located in the osteoarthritis-associated 17q21.31 locus, show elevated expression in knee osteoarthritis cartilage compared to non-osteoarthritis controls[46].

To explore potential mechanisms underlying these distal regulatory effects, we next asked whether *trans*-eQTL variants might influence gene expression by disrupting transcription factor binding sites[47]. Across all tissues, we identified 46 unique variants predicted to affect motifs for 57 transcription factors (Table S9), with several linked to retinoic acid signaling, including

RXRA, RXRG, RARA::RXRA, RXR::RAR_DR5, RORA, and RORC. In synovium, multiple variants associated with *TRIM69* expression are predicted to disrupt retinoic acid-related binding sites, suggesting that retinoic acid-responsive regulatory elements may mediate these effects. Although direct regulation of *TRIM69* by these factors has not been shown, other TRIM family members, such as *TRIM32*, interact with retinoic acid receptors[48]. As retinoic acid signaling is a pathway previously implicated and proposed as a therapeutic target in osteoarthritis[6,52] its *trans*-regulation may be of particular relevance for the disease.

Resolution of osteoarthritis GWAS signals

We conducted statistical colocalization analysis between knee-related osteoarthritis GWAS risk loci[6] and primary knee osteoarthritis tissues *cis*-eQTLs (Table S10, Table S11). Among the 211 osteoarthritis GWAS risk loci examined, 74 (34.1%) showed evidence of a shared association signal with *cis*-eQTLs from primary osteoarthritis tissues, spanning 117 unique genes. 67 out of the 117 colocalizing genes have not been previously prioritized as effector genes for osteoarthritis[6].

To glean biological insights from the 117 colocalizing genes, we performed biological pathway over-representation analysis (Table S12). Compared to all eGenes across all four primary osteoarthritis tissues, the colocalizing genes are enriched for osteoblast differentiation (q-value=0.037) as well as signaling pathways including transforming growth factor beta (TGF-β) (q-value=0.037), mitogen-activated protein kinase (MAPK) (q-value=0.042) and phosphoinositide 3-kinase/protein kinase B (PI3K/AKT) (q-value=0.048) (Figure 5). These three signaling pathways have been previously implicated in osteoarthritis[37,49,50], as well as osteoblast differentiation[51]. The 117 colocalizing genes are also enriched for gene ontology pathways from broad biological mechanisms including cell fate and morphogenesis, as well as for 9-cis-retinoic acid (9cRA) metabolic process (q-value=0.032). While all-trans retinoic acid has been more extensively studied in osteoarthritis[52], 9cRA may have similar effects due to its ability to bind to retinoic acid receptors and influence retinoid signaling.

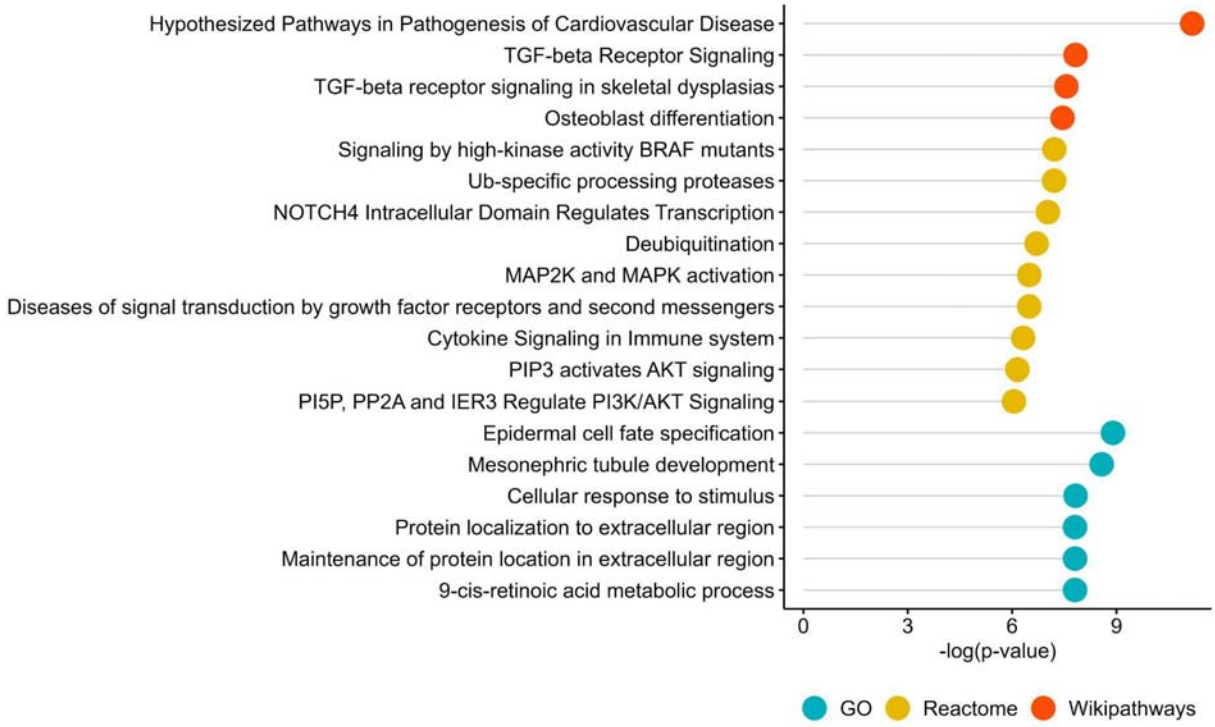


Figure 5: **Pathway enrichment analysis of the colocalizing genes.** Results (q-value<0.05) of pathway-based over-representation analysis using ConsensusPathDB[53] on the 117 genes colocalizing with knee-related osteoarthritis phenotypes. All eGenes across all four primary osteoarthritis tissues studied here were used as the background.

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 these variants by overlaying them with active enhancer-promoter loops from primary knee osteoarthritis chondrocytes[54]. Given that the high-throughput chromosome conformation capture (HiC) data were derived from cartilage tissue, we focused primarily on colocalizing regions in high-grade and low-grade cartilage. Two eQTLs prioritized by the colocalization analysis are located within enhancers of active enhancer-promoter loops for the same respective eGene: rs10948 (*SLC44A2*) and rs12932078 (*WWP2*). This suggests that these eQTL-eGene pairs may exert their effect on knee osteoarthritis by altering enhancer-promoter interactions. *WWP2* and *SLC44A2* have been previously identified as osteoarthritis effector genes[6] (Figure 6, Figures S7-8).

Three cartilage eQTLs colocalizing with osteoarthritis GWAS signals reside within active promoters: rs3742566 (*LGALS3*), rs2102066 (*WWP2*) and rs35346340 (*FES*) (Figures S9-11). 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. Additionally, three colocalizing synovium eQTLs reside within active promoters: rs12129705 (*ZNF697*), rs11539637 (*FES*) and rs117827273 (*MSL1*) (Figures S11-13). *MSL1* and *ZNF697* have not been prioritized as effector genes in the latest osteoarthritis GWAS meta-analysis[6].

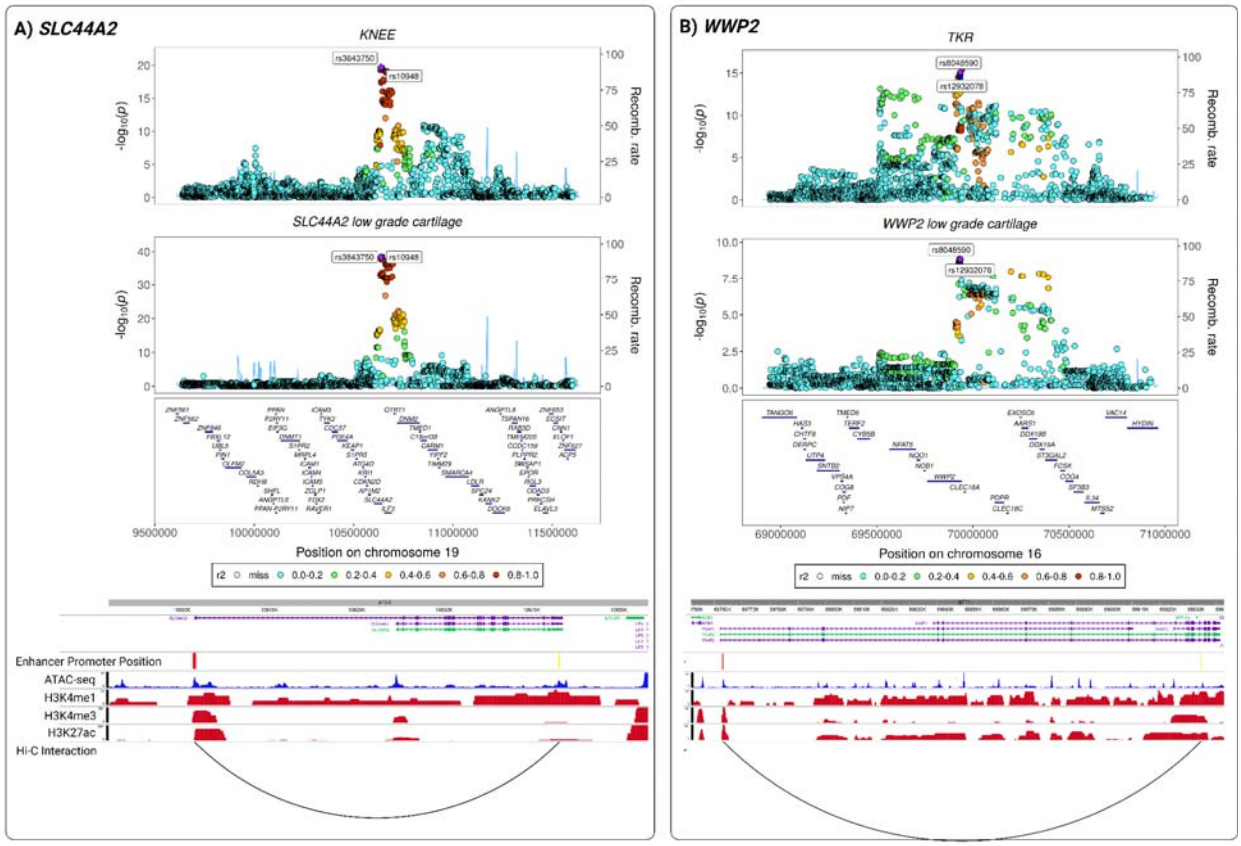


Figure 6: **Examples of colocalizing *cis*-eQTLs in low-grade cartilage overlapping with chondrocyte active enhancer-promoter loops derived from Hi-C data.**

(A) 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.

(B) Example of rs12932078, a colocalizing *cis*-eQTL of *WWP2* in low-grade cartilage and total knee replacement 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[55], showing annotated gene transcripts, SNP positions (yellow), enhancer and promoter locations (red), chromatin states, and histone modifications derived from chondrocyte ATAC-seq[56] and ChIP-seq[57] data (TKR= total knee replacement, KNEE = knee osteoarthritis).

Prioritization of 38 high-confidence effector genes

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)[58] (Table S13). Out of the 117 colocalizing genes, 87 show evidence of a causal effect on knee-related osteoarthritis for matching disease phenotype and tissue[59]. In addition to causal inference evidence, we have scored the genes based on orthogonal lines of evidence including knockout mouse phenotypes[60], rare and syndromic human disease phenotypes[61] and HiC information in cartilage[54] (Table S14). Out of the 117 colocalizing genes 40 are associated with musculoskeletal-related phenotypes (see Methods) in knockout mouse models and 10 in rare and syndromic human diseases. We prioritize 38 genes

that showed at least two lines of evidence pointing to their involvement in osteoarthritis, in addition to colocalizing with GWAS signals (Figure 7). Of these, 12 have not been defined as osteoarthritis effector genes in the latest GWAS meta-analysis[6].

We further evaluated the spatial expression patterns of the colocalizing genes using spatial transcriptomics data from osteoarthritis degraded cartilage (see Methods). We find that 37 out of the 38 prioritized colocalizing genes are expressed in at least one of the analyzed cartilage zones (Table S15). All of these genes are detected in the superficial layer of osteoarthritis cartilage. Nine genes (*ALDH1A2, FBN2, LMX1B*, *LTBP1, MSL1*, *POLB*, *SCYL1*, *SMAD7* and *UXS1*) are also expressed in the middle zone, while three genes (*ALDH1A2, HLA-DPB1* and *RBPJ*) are expressed in the deep cartilage zone. This shows that the majority of prioritized genes are active in the superficial layer of osteoarthritis cartilage, potentially reflecting its increased involvement in osteoarthritis pathology.

Among the prioritized effector genes, *WWP2*, an established effector gene for osteoarthritis[6], showed the strongest evidence of involvement, with four lines of evidence (Figure 7). 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=1.23, q-value=6.34×10^{-38}) and knee replacement surgery (OR=1.37, q-value=2.84×10^{-42}), whereas increased expression in synovium has a protective effect on the risk of knee osteoarthritis Our findings point to contrary mechanisms of action for *WWP2* in cartilage and synovium. We find evidence of expression of *WWP2* in the superficial zone of osteoarthritis cartilage.

NDNF accrued three lines of evidence in support of its role in osteoarthritis and has not been prioritized as an effector gene before. This gene colocalizes with knee osteoarthritis in low-grade (PP4=0.79, PP4/PP3=3.86) and high-grade cartilage (PP4=0.75, PP4/PP3=3.03) as well as in fat pad tissue (PP4=0.76, PP4/PP3=3.31). We find evidence of a risk-increasing effect of *NDNF* expression on knee osteoarthritis risk in high-grade cartilage (OR=1.1, q-value=6.69×10^{-17}) and fat pad (OR=1.05, q-value=5.54×10^{-13}). *NDNF* also showed evidence of expression in the superficial layer of osteoarthritis cartilage. Knockout mice for *NDNF* show skeletal muscle atrophy. Mutations in this gene are linked to the rare human disease hypogonadotropic hypogonadism 25 with anosmia, which is characterized by delayed or absent puberty and is associated with delayed bone age. *NDNF* promotes neuron migration, growth and survival, and may also be involved in regulating growth factor activity[62]. Neurotrophic factors can modulate inflammation and pain perception, an important symptom for osteoarthritis[63,64].

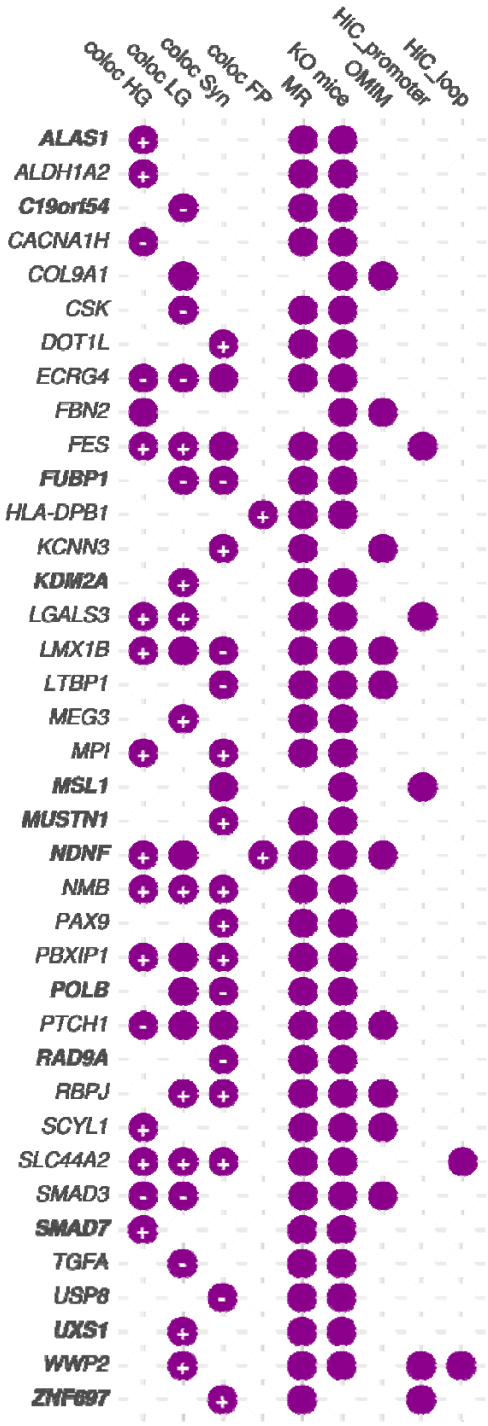


Figure 7: **Prioritized colocalizing genes based on additional evidence of involvement in knee-related osteoarthritis**. In addition to showing the osteoarthritis primary tissue of colocalization (HG=high-grade cartilage, LG=low-grade cartilage, Syn=synovium, FP=fat pad), the heat map visualizes the lines of evidence used to assess involvement in disease, including evidence of knockout mouse models (KO mice), rare and syndromic diseases (OMIM), causal inference analysis (MR), overlap with active promoter regions from chondrocytes (HiC_promoter) and overlap with enhancer-promoter loops from chondrocytes (Hi_loop). The signs (+/-) in the colocalization columns indicate the direction of putative causal effect

identified by MR. The genes highlighted in bold indicate genes that were not previously prioritized as effector genes for osteoarthritis in the latest disease GWAS meta-analysis[6].

Prioritized genes are enriched for chondrocyte-related pathways

The 38 prioritized colocalizing genes are enriched for gene ontology pathways related to morphogenesis, cell fate, cell signaling, cell regulation and metabolism (Table S16). Some genes are involved in osteoarthritis-relevant pathways specifically associated with cartilage, chondrocyte phenotypic stability and bone (Table S16). For instance, *MUSTN1* and *SMAD7*, two genes that were not previously prioritized as effector genes for osteoarthritis[6], are involved in the chondrocyte proliferation (q-value=0.012) and cartilage development pathways (q-value=0.035). *SMAD7* is further involved in regulation of ossification (q-value=0.003).

SMAD7 encodes an intracellular inhibitory SMAD protein that acts as a negative regulator of the TGF-β signaling pathway. TGF-β promotes chondrocyte proliferation and extracellular matrix synthesis, preserving cartilage integrity[65]. Variants associated with the expression of *SMAD7* in high-grade cartilage colocalize with total knee replacement (PP4=0.81) and show evidence of a putative risk-increasing causal effect on total knee replacement risk (OR=1.19, q-value=3.47×10^{-5}). *SMAD7* is the target of mongersen sodium, a SMAD-7 mRNA antisense inhibitor in clinical trials for Crohn's disease and ulcerative colitis. By inhibiting SMAD7, mongersen sodium aims to restore TGF-β activity, pointing to a potential repurposing opportunity of this drug for osteoarthritis. Knockout mouse models for *SMAD7* present abnormal bone marrow cell physiology[66], which can be associated with bone marrow lesions common at late stage osteoarthritis.

Variants associated with the expression of *MUSTN1* in synovium colocalize with knee osteoarthritis GWAS (PP4=0.86). In the same tissue, we find evidence of a risk-increasing effect of *MUSTN1* expression on knee osteoarthritis (OR=1.11, q-value=6.18×10^{-16}). *MUSTN1* has been implicated in critical processes for maintaining cartilage integrity and function, such as chondrogenesis, chondrocyte differentiation, and proliferation[67]. 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[68] . Knockout mouse models for this gene show skeletal muscle fibrosis[69].

Translational drug repurposing opportunities for the colocalizing genes

Considering the translational potential of drug targets with genetic evidence supporting their effect on disease[70], we conducted a druggability search on the 117 colocalizing genes using the Open Targets database[71]. A total of 14/117 genes are linked to drugs with at least phase II clinical trials completed (Table S17). *TGFA* and *CACNA1H* are the only colocalizing genes targeted by a drug with phase II clinical trials for osteoarthritis completed. *TGFA* is the target of fepixnebart, an antibody that inhibits TGFA interaction with its receptor, thereby hindering downstream signaling pathways. In addition to osteoarthritis, fepixnebart is under clinical trial with phase II completed for neuropathic pain and low back pain. Pregabalin is an anticonvulsant drug used to treat neuropathic pain conditions and fibromyalgia that targets *CACNA1H.* This

drug is used as chronic pain management in the treatment of osteoarthritis symptoms. *CACNA1H* is the target of several additional drugs, all of which act as an antagonist or blocker.

Two further genes targeted by drugs in completed phase II clinical trials are associated with osteoarthritis: *FGF1* and *MAPK3*. *FGF1* has been associated with osteoarthritis through its role in cartilage degradation and repair mechanisms. Studies suggest that *FGF1* may contribute to cartilage repair by promoting chondrocyte proliferation and extracellular matrix synthesis while also modulating inflammation[72,73]. Drugs targeting *FGF1* span inhibitory, agonistic and activator mechanisms of action. Variants associated with the expression of this gene colocalize with knee osteoarthritis, osteoarthritis at any site and total knee replacement in low-grade osteoarthritis cartilage (Table S10, Table S11).

MAPK3 is implicated in osteoarthritis by influencing chondrocyte function and inflammatory responses. Studies have shown that MAPK3 regulates chondrocyte proliferation, apoptosis, and the production of matrix metalloproteinases, enzymes that degrade cartilage matrix, thereby contributing to osteoarthritis progression[74]. It has been shown that MAPK inhibitors are protective against early stage osteoarthritis[75]. Both drugs targeting *MAPK3*, ulixertinib and temuterkib, are inhibitors of this gene. In addition to *MAPK3* eQTLs colocalizing with knee-related osteoarthritis in synovium (PP4=0.69, PP4/PP3=9.62), we find evidence of a protective effect of *MAPK3* expression on knee osteoarthritis in the same tissue (OR=0.87, q-value=7.21×10^{-5}).

Six further colocalizing genes targeted by drugs with at least phase II clinical trials completed have not been previously prioritized for osteoarthritis: *ALAS1*, *PSMA4, TNFRSF9, FRK, PIK3R3* and *KCNK3* (Table S17). *KCNK3* encodes a member of the two-pore domain potassium channel family, which is the target of sevoflurane. Sevoflurane, a widely used inhaled anesthetic, has been shown to modulate the activity of various ion channels and is indicated for pain treatment.

Discussion

Large efforts have been made to identify genetic risk signals for osteoarthritis. However, elucidating the function and tissue of action of these variants is challenging. 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 disease tissue- and 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 Han Chinese 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.

Colocalization between the *cis*-eQTL maps and knee-related osteoarthritis GWAS finds evidence of a shared signal for 34.1% of the tested loci, yielding 117 unique colocalizing genes. 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[76]. We prioritized 38 colocalizing genes supported by multiple lines of evidence of involvement in osteoarthritis. All were causally linked to knee-related osteoarthritis risk. Considering that genetic evidence support can accelerate drug discovery[77], the prioritized genes represent appealing candidates for clinical translation, including drug repurposing opportunities.

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 a well-recognized gap in the availability of molecular data from global populations[78]. Going forward, it is important that efforts focus on closing this gap. Recent single-cell RNA-seq (scRNA-seq) studies in osteoarthritis cartilage[79], synovium[80] and fat pad[81] have implicated specific cell types involved in osteoarthritis. Additionally, *trans*-eQTLs have been shown to capture cell-type specific effects[82]. Combining genomic data with scRNA-seq in larger sample sizes is needed to differentiate between real regulatory effects and different cell compositions.

Robust molecular data from primary tissues can be valuable in enhancing the translation of genetic findings for all complex diseases. Here, we have generated the largest molecular 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[83]. 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[84]. 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.[85] 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)[86] 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)[87]. We excluded samples from individuals that had bilateral knee replacement where low- and high-grade disease cartilage samples were pooled (7 individuals). 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 (Figure S14). 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[84]. 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).[88] 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)[89]. 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).

To characterize population structure and ensure consistency across datasets, we next performed principal component analysis (PCA) on the filtered genotype data using bcftools[90], PLINK[91], and the 1000 Genomes Project reference panel[89]. Genotype data in VCF format were filtered to retain only unique samples across all tissues using *bcftools view* and updating (MAF ≥ 0.05), call rate (CR ≥ 99%), and biallelic status as described above. The resulting VCF contained 7.285.082 variants across 321 samples. The resulting VCF was indexed and converted to PLINK binary format. To assess population structure, we downloaded phased genotype data from the 1000 Genomes Project (GRCh38), normalized and converted it to PLINK format, and LD-pruned each chromosome. We merged the pruned reference data across chromosomes and harmonized variant IDs with our dataset to identify common variants with 1000 Genomes Project reference panel (205.409 variants). PLINK was used to generate a merged dataset and compute principal components, which were used as covariates in downstream analyses. To generate PCA plots, we read eigenvectors and eigenvalues from PLINK output in R and calculated the proportion and cumulative variance explained by each principal component (PC). We annotated samples by merging with 1000 Genomes population metadata and labeled study samples accordingly. All eQTL cohort samples clustered with

European populations, primarily UK, thus genotype PCs were not included as covariates in downstream analyses (Figure S15).
Power analysis for the eQTL analysis was performed using the R package powerEQTL[92] (*powerEQTL.SLR*) across different MAF, slopes and sample sizes (Figure S1).

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[93] (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[94] representing hidden data variation.
We tested the association of each PEER factor with known covariates. For categorical covariates analysis of variance (ANOVA) was performed between each PEER factor and each covariate via the aov function in R. From the resulting ANOVA summary, the F-statistic and associated p-value (Pr(>F)) were extracted to assess statistical significance. For numeric covariates, a linear regression model was fitted using the lm function in R with PEER as the dependent variable and the covariate as the independent variable. From the regression summary, the t-statistic and p-value corresponding to the covariate were extracted to determine the strength and significance of the association (Figures S16-19).
The number of PEER factors included per tissue in the eQTL model was determined according to GTEx[35] 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[95]).

We conducted *cis*-eQTL analysis per tissue using tensorQTL (version 1.0.9)[96] 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[35].

Mapping *trans*-eQTL

We performed *trans*-eQTL analysis per tissue using MatrixeQTL[97]. 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. We performed *trans*-eQTL analysis on all variants whose distance to the genes's TSS was more than 5Mb. Given the large number of tests, only gene-SNP pairs with p-value < 10^{-6} were retained. Similar to the *cis*-eQTL analysis, and as previously reported[35], we calculated gene-level significance as follows:

- For each gene, we selected the most significant *trans*-eQTL (i.e., SNP-gene pair with the lowest nominal p-value) across chromosomes 1–22.
- The nominal p-values were multiplied by 10^6 to correct for the number of expected independent tests for a genome-wide association analysis at MAF 5% in a European population. Corrected p-values were capped at 1.
- FDR correction on the list of corrected p-values was performed using the qvalue package. *Trans*-eQTLs were defined as those with q-value < 0.05.
- To identify per-gene *trans*-eQTLs, a genome-wide significance threshold was calculated. We identified the largest p-value among significant *trans*-eQTLs (q-value ≤ 0.05) and the smallest p-value among non-significant *trans*-eQTLs (q-value > 0.05). We then took the midpoint between these two values to define a conservative genome-wide p-value threshold for significance at FDR 5%.

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[98]. We identified conditionally independent *trans*-eQTLs for each analyzed tissue by clumping the significant trans-eQTLs based on LD. We used PLINK[91] to perform clumping on a window of 1 Mb and a strict LD threshold $r^2 < 0.001$.

***Trans*-eQTL transcription factor binding motif disruption analysis**

To assess whether *trans*-eQTLs disrupt transcription factor binding motifs (TFBMs), we used the Bioconductor package motifbreakR[99] (version 2.20). We filtered the TF motif database to include only human motifs (organism == "Hsapiens") and restricted it to high-confidence data sources: HOCOMOCOv11-core-A, jaspar2022, JASPAR_CORE, and SwissRegulon. We ran motifbreakR() using the following parameters: filterp = TRUE, method = "ic" (information content), threshold = 1e-5, and a uniform background nucleotide distribution bkg = c(A=0.25, C=0.25, G=0.25, T=0.25). P-values for reference and alternative alleles were computed using calculatePvalue() with granularity = 1e-4. We retained only strong TFBM disruptions by filtering

for results with Refpvalue < 0.05, Altpvalue < 0.05, effect == "strong", and pvalueEffect == "strong". For each trans-eQTL, we recorded the disrupted TF motif and the associated target eGene to generate candidate TF–eGene regulatory links.

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[100] 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[101] to generate scatter plots with fitted regression lines and confidence intervals for each tissue type.

Identification of tissue-sharing and single-tissue eQTLs

To assess tissue-specific and shared eQTL patterns, we first explored eQTL effects detection in only one tissue due to tissue-specific gene expression (i.e., eQTLs in genes not common across tissues). To formally assess the presence of unique eQTL effects for genes that are commonly expressed in all disease tissues we applied multivariate adaptive shrinkage using MashR[102] 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. To identify tissue-specific *cis*-eQTLs, we implemented a filtering procedure that required: (1) LFSR < 0.05 in a single disease tissue; (2) LFSR ≥ 0.1 in all other tissues; (3) a ≥ two-fold larger absolute effect size in the single tissue compared to all others; and (4) a minimum absolute effect size of 0.1. This ensured eQTLs with unique effect in one tissue.

Variant annotation of eQTL

We annotated all variants tested in eQTL analyses for functional impact on known genes using the ANNOVAR tool (https://annovar.openbioinformatics.org/en/latest/)[103]. Regulatory

annotations were drawn from the Ensembl Regulatory Build[104], 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.[54]. 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[54]. We evaluated the distribution of eQTL annotations among both *cis*- and *trans*-eQTLs (Figure S20).

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 knee-related 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 the phenotype-specific osteoarthritis GWAS risk signals defined in Hatzikotoulas et al. [6].Colocalization analysis was run in the GWAS risk regions that had at least one study-wise significant eQTL. 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. 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[105].

In regions where no evidence of colocalization but slight evidence of shared causal association between both traits (PP4>0.25) 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[105]. We defined evidence of colocalization as for colo.abf(): PP4 >= 0.8 or PP4 > 0.6 and PP4/PP3 > 2. Given the lack of credible set information available in coloc.susie, we defined the lead signals from the GWAS and the eQTL as candidates for the lead colocalizing variant.

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

We defined candidates for the shared causal variants as the 95% credible set variants from coloc.abf() results as well as all lead signals colocalizing in coloc.susie(). For each locus

showing evidence of colocalization between cis-eQTL data and knee-related osteoarthritis, we mapped all the candidates for the shared causal variants to active enhancer-promoter loops from primary knee osteoarthritis chondrocytes[54]. To identify the target genes of the active promoters, the authors overlaid the extended promoter region with gene regions from ENSEMBL[106].

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

To evaluate putative causal effects of gene expression levels of colocalizing genes on knee-related osteoarthritis, we conducted Mendelian randomization analyses[58]. As exposures, we used the generated *cis*-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[89] 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[107] 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 ($beta^2/se^2$)[58]. 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/)[53] for both pathways and Gene Ontology terms using the biological process until level five. We further performed enrichment analysis for the genes showing evidence of colocalization 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. We selected pathways from the biological processes terms of Gene Ontology up to level 5. 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 colocalizing 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[61]) (https://omim.org/) via its API. We defined association with musculoskeletal phenotypes if the disease showed any skeletal or muscle soft tissue phenotype. Secondly, we used the batch query functionality of the Mouse Genome Informatics (MGI[60]) 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 colocalizing genes using the OpenTargets[71] 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[108]. 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.

Spatial transcriptomics study participants

The spatial transcriptomics project has been evaluated and approved by the Research Ethics Committee of the University of Tartu (379/M-10, 12th of June 2023). All study participants have voluntarily signed the informed consent form of the study. For spatial transcriptomics analyses three knee osteoarthritis patients from the Orthopaedics Clinic at the Tartu University Hospital were recruited. They were all female with average age 65 (standard deviation 6.78) years and with European ethnicity. They all have been diagnosed with knee osteoarthritis with Kellgren and Lawrence grade 4 (severe) and were hospitalised for the joint replacement surgery due to osteoarthritis. They did not have post-traumatic osteoarthritis , treatment with bisphosphonates, dysplasia, inflammatory arthritis (such as ankylosing spondylitis, rheumatoid arthritis, psoriatic arthritis), gout or lupus.

Sample collection and preparation

Punch biopsies were taken from the tibial plateau (surgical waste from arthroplasty surgery) immediately after removal from the joint. Biopsies were collected from macroscopically degraded region. All biopsies contained cartilage and subchondral bone tissue layers. Tissues were transferred into 4 % formaldehyde solution at room temperature (RT). Within 2 hours the tissues were cut into smaller pieces (smallest dimension ~2 mm) and incubated in fresh 4 % formaldehyde solution at RT for at least 72 h with mild agitation. After this, tissues were washed with nuclease free water for 15 min at RT with mild agitation three times and then transferred into 14 % EDTA solution with pH at least 8 for the decalcification. Tissues were incubated in EDTA solution at RT with mild agitation and solution changed daily for altogether 11 days. After that the tissues were washed with nuclease free water for 15 min at RT with mild agitation three times and delivered to Tartu University Hospital Pathology Service laboratory for FFPE block preparation. The FFPE blocks were stored at 4 ○C until further use. For conducting the spatial transcriptomics sequencing the tissues were sectioned (5 µm) and placed onto Leica Bond plus slides (Leica Biosystems GmbH, Germany). In parallel the sections were graded based on OARSI grading system and all tissues were at OARSI grade 3.

Spatial Transcriptomics

Spatial transcriptomics analysis was conducted with Nanostring GeoMx Digital Spatial Profiler platform applying Human Whole Transcriptome Atlas panel (Burker Spatial Biology Inc, USA) and following sequencing on Illumina platform (Illumina Inc, USA). For selecting the areas of illumination (AOIs), the DNA dye signals were used. The AOIs included in the study were selected from the following tissue areas: (1) cartilage superficial zone, (2) cartilage middle zone, (3) cartilage deep zone, (4) calcified cartilage (below tidemark).

Evaluating the tissue-level specificity of the GO2 eGenes

To evaluate in which osteochondral tissue zones the colocalizing genes are expressed, the following workflow was applied. (1) We looked, which colocalizing genes were detected by Nanostring GeoMX platform in the samples. 97 out of the 117 colocalizing genes were expressed in at least one tissue. (2) We considered the genes not expressed if the probe count was below 6 in at least two samples out of three per zone group[109]. The probe counts can be found in Table S17.

Supplemental material

Supplemental tables

Table S1: Aggregate demographic data of individuals included in the study and summary technical covariate distributions across study samples.

Table S2: Permuted significant eQTL results and conditional independent eQTLs per osteoarthritis primary tissue.

Table S3: Unique significant eQTLs per osteoarthritis primary tissue due to unique tissue gene expression.

Table S4: Summary statistics of meta-analysis across osteoarthritis primary tissues.
Table S5: Summary statistics of disease tissue-specific and discordant eQTLs from meta-analysis across tissues.
Table S6: Summary statistics of meta-analysis across cartilage grades.
Table S7: Summary statistics of disease grade-specific and discordant eQTLs from meta-analysis across cartilage grades.
Table S8: Significant trans-eQTLs per osteoarthritis primary tissue using MatrixEQTL.
Table S9: Trans-eQTLs with evidence of transcription factor binding motif-disruption.
Table S10: Significant results of colocalization analysis using the coloc.abf() method between cis-eQTL data from osteoart0hritis primary tissues and osteoarthritis GWAS.
Table S11: Significant results of colocalization analysis using the coloc.susie() method between cis-eQTL data from osteoarthritis primary tissues and osteoarthritis GWAS.
Table S12: Results (q-value<0.05) of the pathway-based over-enrichment analysis using ConsensusPathDB on the 117 genes that colocalize with knee-related osteoarthritis phenotypes.
Table S13: Results of Mendelian randomization analyses between colocalizing cis-eGenes from osteoarthritis primary tissues eQTL data and knee-related osteoarthritis phenotypes.
Table S14: Scoring of colocalizing genes based on orthogonal lines of evidence pointing to involvement in osteoarthritis.
Table S15: Nanostring GeoMX probe counts of spatial transcriptomics data from three osteoarthritis patients.
Table S16: Results of the pathway-based over-enrichment analysis using ConsensusPathDB on the 38 prioritized genes involved in knee-related osteoarthritis phenotypes.
Table S17: Druggability information of colocalizing genes that are targets of drugs with at least phase II trial completed.

Supplemental figures

Figure S1: Power curves depicting the results of cis-eQTL power estimation across MAF and slope spectrum for the different osteoarthritis primary tissues. Colors reflect the sample sizes of Katsoula & Arruda et al. and Steinberg et al.[14].
Figure S2: Comparison of effect sizes between the commonly detected *cis*-eQTLs between our study and Steinberg et al.[14] using linear regression.
Figure S3: eQTL–eGene pairs exhibiting disease tissue-specific effects across primary osteoarthritis tissues identified through mashR[102]. The x-axis shows the posterior mean (effect size). Triangle orientation indicates the direction of effect (upward or downward). Significant associations (LFSR < 0.05) are shown as triangles, while non-significant pairs are shown as circles.
Figure S4: eQTL–eGene pairs associated with disease progression, identified through mashR[99] meta-analysis of low-grade and high-grade cartilage. The x-axis represents the posterior mean (effect size). Triangle orientation indicates the direction of effect (upward or downward). Significant associations (LFSR < 0.05) are shown as triangles, while non-significant pairs are represented as circles.

Figure S5: Violin plots depicting the effects of the most significant *trans*-eQTLs and their associated gene expression across osteoarthritis primary tissues with significant associations, including low-grade cartilage, high-grade cartilage, and synovium. Embedded boxplots within indicate the 25th, 50th (median), and 75th percentiles.

Figure S6: The rs6511291 lead variant exhibits a *cis* effect on the expression of the transcription factor coding gene *ZNF100* and a *trans* effect on the expression of *FHAD1* in synovial tissue. The violin plots display the distribution of expression levels, with the boxplots embedded within indicating the 25th, 50th (median), and 75th percentiles. Whiskers extend to 1.5 times the interquartile range. Nominal p-values and effect sizes (slopes) derived from cis- and trans-eQTL mapping analyses, respectively are displayed above the violin plots.

Figures S7-8: Regional association plots of the genetic loci where a colocalizing eQTL (variant included in the 95% credible set for the shared causal variant from the colocalization analysis with osteoarthritis GWAS) resides within an enhancer of an active promoter-enhancer loop derived from HiC data of primary knee osteoarthritis chondrocytes (Bittner et al.) from the same gene (KNEE = knee osteoarthritis; ALLOA = osteoarthritis at any site; TKR = total knee replacement surgery; PP4 = posterior probability of a shared causal variant).

Figures S9-S13: Regional association plots of the genetic loci where a colocalizing eQTL (variant included in the 95% credible set for the shared causal variant from the colocalization analysis with osteoarthritis GWAS) resides within a promoter of an active promoter-enhancer loop derived from HiC data of primary knee osteoarthritis chondrocytes (Bittner et al.) from the same gene (ALLOA = osteoarthritis at any site; TKR= total knee replacement surgery; KNEE = knee osteoarthritis; PP4 = posterior probability of a shared causal variant).

Figure S14: PCA plots of gene expression ($\log_2$(CPM+2) normalized) colored according to different sample information.

Figure S15: A. PCA plot of all populations from the 1000 Genomes Project Phase 3 dataset alongside our eQTL cohort. Each point represents an individual sample, colored by population. Major continental groups (AFR, AMR, EAS, EUR, SAS) form distinct clusters. Our eQTL cohort is overlaid in red, allowing comparison of its genetic structure to global reference populations. **B.** Zoomed-in PCA focusing on the European (EUR) subset of the 1000 Genomes Project and our eQTL cohort. The eQTL samples (red) are visualized relative to finer-scale European subpopulations.

Figures S16-S19: Association of PEER factors with known technical and biological covariates across four osteoarthritis-related tissues: low-grade cartilage (45 factors), high-grade cartilage (30 factors), synovium (45 factors), and fat pad (15 factors). Each heatmap shows $-\log_{10}$ (p-value) for associations between individual PEER factors and covariates such as age, sex, BMI, sequencing metrics, and tissue source. Significant associations (Bonferroni-adjusted $p<0.05$) are indicated, with darker blue reflecting stronger associations.

Figure S20: Distribution of *cis*- and *trans*- eQTLs across functional and regulatory region annotations for all tested osteoarthritis primary tissues. Regulatory region annotations are derived from ensemble regulatory build[104] and HiC map of primary osteoarthritis chondrocytes[54]. Variant-level annotations are derived from ANNOVAR[110].

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).
ER, RM and GLVS have received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 101095084 (ENDOTARGET). RM is supported by Estonian Research Council grant PUT (PRG1911) and by Estonian Ministry of Education and Research Funding (TK214). The work on KT and AM is supported by Estonian Research Council grant PSG610.

Author contributions

Study design: E.Z., J.M.W.; Clinical collection and library preparation: J.M.W., K.M.S., D.S., E.R;, S.S., G.L.V.S., K.T., A.M., R.M., Data Analysis: G.K., A.A., M.T., E.R, 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

The full summary statistics of all analyses will be shared through the Musculoskeletal Knowledge Portal[111] (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. Hatzikotoulas, K., Southam, L., Stefansdottir, L., Boer, C.G., McDonald, M.-L., Pett, J.P., Park, Y.-C., Tuerlings, M., Mulders, R., Barysenka, A., et al. (2025). Translational genomics of osteoarthritis in 1,962,069 individuals. Nature, 1–8. https://doi.org/10.1038/s41586-025-08771-z.

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. Thakur, M., Dawes, J.M., and McMahon, S.B. (2013). Genomics of pain in osteoarthritis. Osteoarthritis and Cartilage *21*, 1374–1382. https://doi.org/10.1016/j.joca.2013.06.010.

18. Nassari, S., Blavet, C., Bonnin, M.-A., Stricker, S., Duprez, D., and Fournier-Thibault, C. (2017). The chemokines CXCL12 and CXCL14 differentially regulate connective tissue markers during limb development. Sci Rep *7*, 17279. https://doi.org/10.1038/s41598-017-17490-z.

19. 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.

20. 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.

21. Seki, K., Fujimori, T., Savagner, P., Hata, A., Aikawa, T., Ogata, N., Nabeshima, Y., and Kaechoong, L. (2003). Mouse Snail family transcription repressors regulate chondrocyte, extracellular matrix, type II collagen, and aggrecan. J Biol Chem *278*, 41862–41870. https://doi.org/10.1074/jbc.M308336200.

22. Stanton, L.-A., Underhill, T.M., and Beier, F. (2003). MAP kinases in chondrocyte differentiation. Developmental Biology *263*, 165–175. https://doi.org/10.1016/S0012-1606(03)00321-X.

23. Chen, Z., Yue, S.X., Zhou, G., Greenfield, E.M., and Murakami, S. (2015). ERK1 and ERK2 regulate chondrocyte terminal differentiation during endochondral bone formation. J Bone Miner Res *30*, 765–774. https://doi.org/10.1002/jbmr.2409.

24. Kraan, P.M. van der, Davidson, E.N.B., Blom, A., and Berg, W.B. van den (2009). TGF-beta signaling in chondrocyte terminal differentiation and osteoarthritis: Modulation and integration of signaling pathways through receptor-Smads. Osteoarthritis and Cartilage *17*, 1539–1545. https://doi.org/10.1016/j.joca.2009.06.008.

25. Janowska, J.D. (2020). C1q/TNF-related Protein 1, a Multifunctional Adipokine: An Overview of Current Data. The American Journal of the Medical Sciences *360*, 222–228. https://doi.org/10.1016/j.amjms.2020.05.036.

26. Ono, K., Hata, K., Nakamura, E., Ishihara, S., Kobayashi, S., Nakanishi, M., Yoshida, M., Takahata, Y., Murakami, T., Takenoshita, S., et al. (2021). Dmrt2 promotes transition of endochondral bone formation by linking Sox9 and Runx2. Commun Biol *4*, 326. https://doi.org/10.1038/s42003-021-01848-1.

27. Bonnet, C.S., and Walsh, D.A. (2005). Osteoarthritis, angiogenesis and inflammation. Rheumatology *44*, 7–16. https://doi.org/10.1093/rheumatology/keh344.

28. Miclea, R.L., Siebelt, M., Finos, L., Goeman, J.J., Löwik, C.W.G.M., Oostdijk, W., Weinans, H., Wit, J.M., Robanus-Maandag, E.C., and Karperien, M. (2011). Inhibition of Gsk3β in cartilage induces osteoarthritic features through activation of the canonical Wnt signaling pathway. Osteoarthritis and Cartilage *19*, 1363–1372. https://doi.org/10.1016/j.joca.2011.07.014.

29. Dy, P., Smits, P., Silvester, A., Penzo-Méndez, A., Dumitriu, B., Han, Y., de la Motte, C.A., Kingsley, D.M., and Lefebvre, V. (2010). Synovial joint morphogenesis requires the chondrogenic action of Sox5 and Sox6 in growth plate and articular cartilage. Dev Biol *341*, 346–359. https://doi.org/10.1016/j.ydbio.2010.02.024.

30. Paolo, G.D., Sankaranarayanan, S., Wenk, M.R., Daniell, L., Perucco, E., Caldarone, B.J., Flavell, R., Picciotto, M.R., Ryan, T.A., Cremona, O., et al. (2002). Decreased Synaptic Vesicle Recycling Efficiency and Cognitive Deficits in Amphiphysin 1 Knockout Mice. Neuron *33*, 789–804. https://doi.org/10.1016/S0896-6273(02)00601-3.

31. Chemel, M., Le Goff, B., Brion, R., Cozic, C., Berreur, M., Amiaud, J., bougras, gwenola, Touchais, S., Blanchard, F., Heymann, M.-F., et al. (2012). Interleukin 34 expression is associated with synovitis severity in rheumatoid arthritis patients. Ann Rheum Dis *71*, 150–154. https://doi.org/10.1136/annrheumdis-2011-200096.

32. Rogers, N., and Meng, Q.-J. (2023). Tick tock, the cartilage clock. Osteoarthritis and Cartilage *31*, 1425–1436. https://doi.org/10.1016/j.joca.2023.05.010.

33. Fagiani, F., Di Marino, D., Romagnoli, A., Travelli, C., Voltan, D., Di Cesare Mannelli, L., Racchi, M., Govoni, S., and Lanni, C. (2022). Molecular regulations of circadian rhythm and implications for physiology and diseases. Sig Transduct Target Ther *7*, 1–20. https://doi.org/10.1038/s41392-022-00899-y.

34. Phillips, R. (2024). ZEB2 promotes formation of age-related B cells. Nat Rev Rheumatol *20*, 138–138. https://doi.org/10.1038/s41584-024-01089-6.

35. 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.

36. Chan, C.K.F., Lindau, P., Jiang, W., Chen, J.Y., Zhang, L.F., Chen, C.-C., Seita, J., Sahoo, D., Kim, J.-B., Lee, A., et al. (2013). Clonal precursor of bone, cartilage, and hematopoietic niche stromal cells. Proceedings of the National Academy of Sciences *110*, 12643–12648. https://doi.org/10.1073/pnas.1310212110.

37. Sun, K., Luo, J., Guo, J., Yao, X., Jing, X., and Guo, F. (2020). The PI3K/AKT/mTOR signaling pathway in osteoarthritis: a narrative review. Osteoarthritis Cartilage *28*, 400–409. https://doi.org/10.1016/j.joca.2020.02.027.

38. Yang, D., Xu, K., Xu, X., and Xu, P. (2024). Revisiting prostaglandin E2: A promising therapeutic target for osteoarthritis. Clin Immunol *260*, 109904. https://doi.org/10.1016/j.clim.2024.109904.

39. Jiang, L., Lin, J., Zhao, S., Wu, J., Jin, Y., Yu, L., Wu, N., Wu, Z., Wang, Y., and Lin, M. (2021). ADAMTS5 in Osteoarthritis: Biological Functions, Regulatory Network, and Potential Targeting Therapies. Front Mol Biosci *8*, 703110. https://doi.org/10.3389/fmolb.2021.703110.

40. Alhamad, D.W., Bensreti, H., Dorn, J., Hill, W.D., Hamrick, M.W., and McGee-Lawrence, M.E. (2022). Aryl hydrocarbon receptor (AhR)-mediated signaling as a critical regulator of skeletal cell biology. J Mol Endocrinol *69*, R109–R124. https://doi.org/10.1530/JME-22-0076.

41. Ruhlen, R., and Marberry, K. (2014). The chondrocyte primary cilium. Osteoarthritis and Cartilage *22*, 1071–1076. https://doi.org/10.1016/j.joca.2014.05.011.

42. Khan, S., Sbeity, M., Foulquier, F., Barré, L., and Ouzzine, M. (2021). TMEM165 a new player in proteoglycan synthesis: loss of TMEM165 impairs elongation of chondroitin- and heparan-sulfate glycosaminoglycan chains of proteoglycans and triggers early chondrocyte differentiation and hypertrophy. Cell Death Dis *13*, 1–14. https://doi.org/10.1038/s41419-021-04458-1.

43. Unguryte, A., Bernotiene, E., Bagdonas, E., Garberyte, S., Porvaneckas, N., and Jorgensen, C. (2016). Human articular chondrocytes with higher aldehyde dehydrogenase activity have stronger expression of *COL2A1* and *SOX9*. Osteoarthritis and Cartilage *24*, 873–882. https://doi.org/10.1016/j.joca.2015.11.019.

44. Hoolwerff, M. van, Metselaar, P.I., Tuerlings, M., Suchiman, H.E.D., Lakenberg, N., Ramos, Y.F.M., Cats, D., Nelissen, R.G.H.H., Broekhuis, D., Mei, H., et al. (2020). Elucidating Epigenetic Regulation by Identifying Functional cis-Acting Long Noncoding RNAs and Their Targets in Osteoarthritic Articular Cartilage. Arthritis & Rheumatology *72*, 1845–1854. https://doi.org/10.1002/art.41396.

45. Ajekigbe, B., Cheung, K., Xu, Y., Skelton, A.J., Panagiotopoulos, A., Soul, J., Hardingham, T.E., Deehan, D.J., Barter, M.J., and Young, D.A. (2019). Identification of long non-coding RNAs expressed in knee and hip osteoarthritic cartilage. Osteoarthritis and Cartilage *27*, 694–702. https://doi.org/10.1016/j.joca.2018.12.015.

46. Rice, S.J., Cheung, K., Reynard, L.N., and Loughlin, J. (2019). Discovery and analysis of methylation quantitative trait loci (mQTLs) mapping to novel osteoarthritis genetic risk signals. Osteoarthritis Cartilage *27*, 1545–1556. https://doi.org/10.1016/j.joca.2019.05.017.

47. Coetzee, S.G., Coetzee, G.A., and Hazelett, D.J. (2015). motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. Bioinformatics *31*, 3847–3849. https://doi.org/10.1093/bioinformatics/btv470.

48. Sato, T., Okumura, F., Kano, S., Kondo, T., Ariga, T., and Hatakeyama, S. (2011). TRIM32 promotes neural differentiation through retinoic acid receptor-mediated transcription. J Cell Sci *124*, 3492–3502. https://doi.org/10.1242/jcs.088799.

49. Shen, J., Li, S., and Chen, D. (2014). TGF-β signaling and the development of osteoarthritis. Bone Res *2*, 14002-. https://doi.org/10.1038/boneres.2014.2.

50. Li, Z., Dai, A., Yang, M., Chen, S., Deng, Z., and Li, L. (2022). p38MAPK Signaling Pathway in Osteoarthritis: Pathological and Therapeutic Aspects. J Inflamm Res *15*, 723–734. https://doi.org/10.2147/JIR.S348491.

51. Maruotti, N., Corrado, A., and Cantatore, F.P. (2017). Osteoblast role in osteoarthritis pathogenesis. J Cell Physiol *232*, 2957–2963. https://doi.org/10.1002/jcp.25969.

52. Davies, M.R., Ribeiro, L.R., Downey-Jones, M., Needham, M.R.C., Oakley, C., and Wardale, J. (2009). Ligands for retinoic acid receptors are elevated in osteoarthritis and may contribute to pathologic processes in the osteoarthritic joint. Arthritis Rheum *60*, 1722–1732. https://doi.org/10.1002/art.24550.

53. 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.

54. 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.

55. 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.

56. 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.

57. 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.

58. 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.

59. 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.

60. 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.

61. 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.

62. Messina, A., Pulli, K., Santini, S., Acierno, J., Känsäkoski, J., Cassatella, D., Xu, C., Casoni, F., Malone, S.A., Ternier, G., et al. (2020). Neuron-Derived Neurotrophic Factor Is Mutated in Congenital Hypogonadotropic Hypogonadism. Am J Hum Genet *106*, 58–70. https://doi.org/10.1016/j.ajhg.2019.12.003.

63. Morel, L., Domingues, O., Zimmer, J., and Michel, T. (2020). Revisiting the Role of Neurotrophic Factors in Inflammation. Cells *9*, 865. https://doi.org/10.3390/cells9040865.

64. Apfel, S.C. (2000). Neurotrophic factors and pain. Clin J Pain *16*, S7-11. https://doi.org/10.1097/00002508-200006001-00003.

65. Blaney Davidson, E.N., van der Kraan, P.M., and van den Berg, W.B. (2007). TGF-beta and osteoarthritis. Osteoarthritis Cartilage *15*, 597–604. https://doi.org/10.1016/j.joca.2007.02.005.

66. Gisterå, A., Robertson, A.-K.L., Andersson, J., Ketelhuth, D.F.J., Ovchinnikova, O., Nilsson, S.K., Lundberg, A.M., Li, M.O., Flavell, R.A., and Hansson, G.K. (2013). Transforming growth factor-β signaling in T cells promotes stabilization of atherosclerotic plaques through an interleukin-17-dependent pathway. Sci Transl Med *5*, 196ra100. https://doi.org/10.1126/scitranslmed.3006133.

67. 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.

68. 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.

69. 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.

70. 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.

71. 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.

72. Xie, Y., Zinkle, A., Chen, L., and Mohammadi, M. (2020). Fibroblast growth factor signalling in osteoarthritis and cartilage repair. Nat Rev Rheumatol *16*, 547–564. https://doi.org/10.1038/s41584-020-0469-2.

73. El-Seoudi, A., Abd El Kader, T., Nishida, T., Eguchi, T., Aoyama, E., Takigawa, M., and Kubota, S. (2017). Catabolic effects of FGF-1 on chondrocytes and its possible role in osteoarthritis. J Cell Commun Signal *11*, 255–263. https://doi.org/10.1007/s12079-017-0384-8.

74. Loeser, R.F., Erickson, E.A., and Long, D.L. (2008). Mitogen-activated protein kinases as therapeutic targets in osteoarthritis. Curr Opin Rheumatol *20*, 581–586. https://doi.org/10.1097/BOR.0b013e3283090463.

75. Lan, C.-N., Cai, W.-J., Shi, J., and Yi, Z.-J. (2021). MAPK inhibitors protect against early-stage osteoarthritis by activating autophagy. Mol Med Rep *24*, 829. https://doi.org/10.3892/mmr.2021.12469.

76. 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.

77. 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.

78. 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.

79. 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.

80. 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.

81. 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.

82. 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.

83. 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.

84. 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.

85. 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.

86. 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.

87. 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.

88. 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).

89. 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.

90. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

91. 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.

92. 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.

93. 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.

94. 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.

95. 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.

96. 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.

97. 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.

98. 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.

99. Coetzee, S.G., and Hazelett, D.J. (2024). motifbreakR v2: expanded variant analysis including indels and integrated evidence from transcription factor binding databases. Bioinformatics Advances *4*, vbae162. https://doi.org/10.1093/bioadv/vbae162.

100. 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.

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

102. 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.

103. 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.

104. 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.

105. 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.

106. 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.

107. 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.

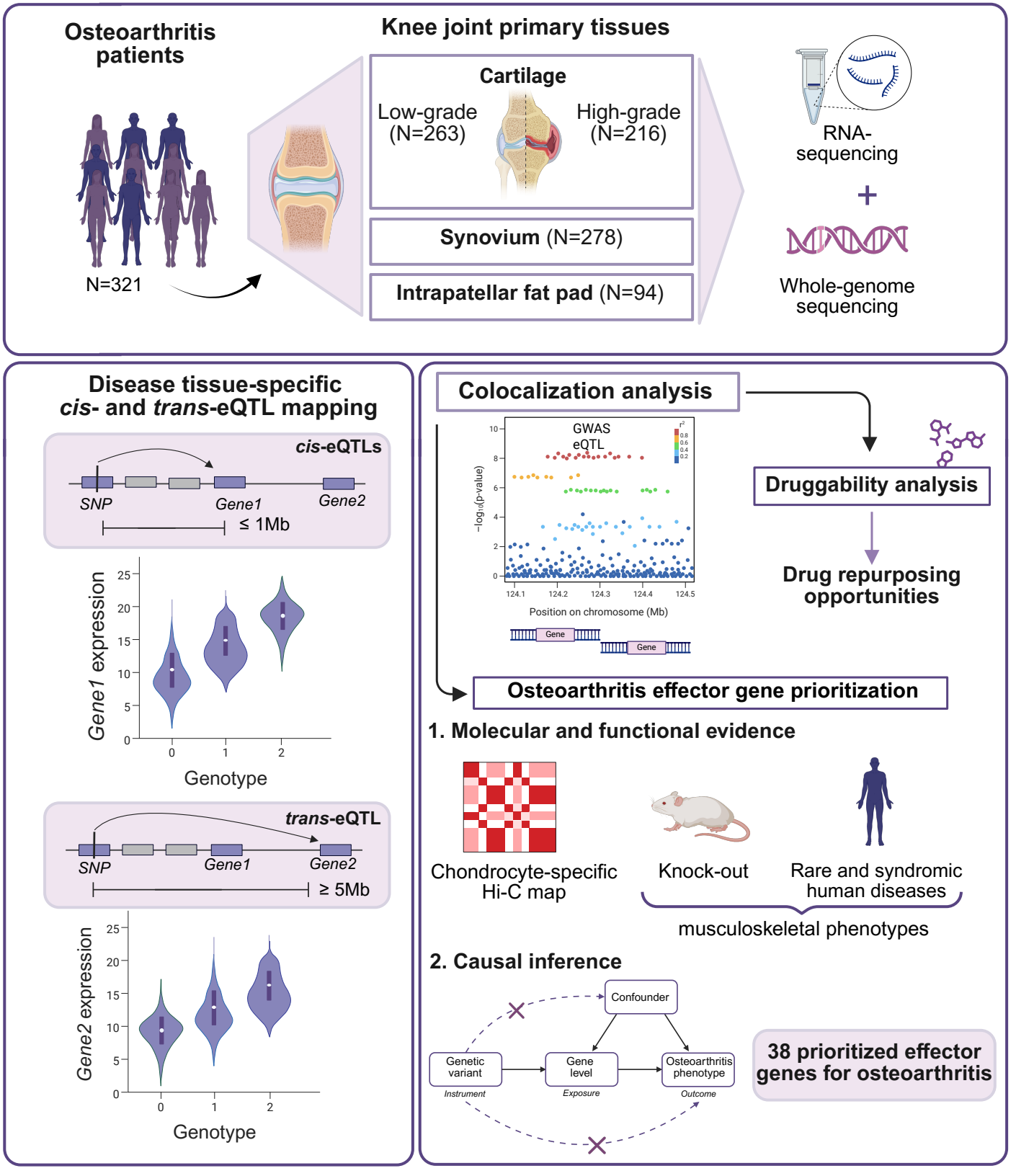
108. Zdrazil, B., Felix, E., Hunter, F., Manners, E.J., Blackshaw, J., Corbett, S., de Veij, M., Ioannidis, H., Lopez, D.M., Mosquera, J.F., et al. (2024). The ChEMBL Database in 2023: a

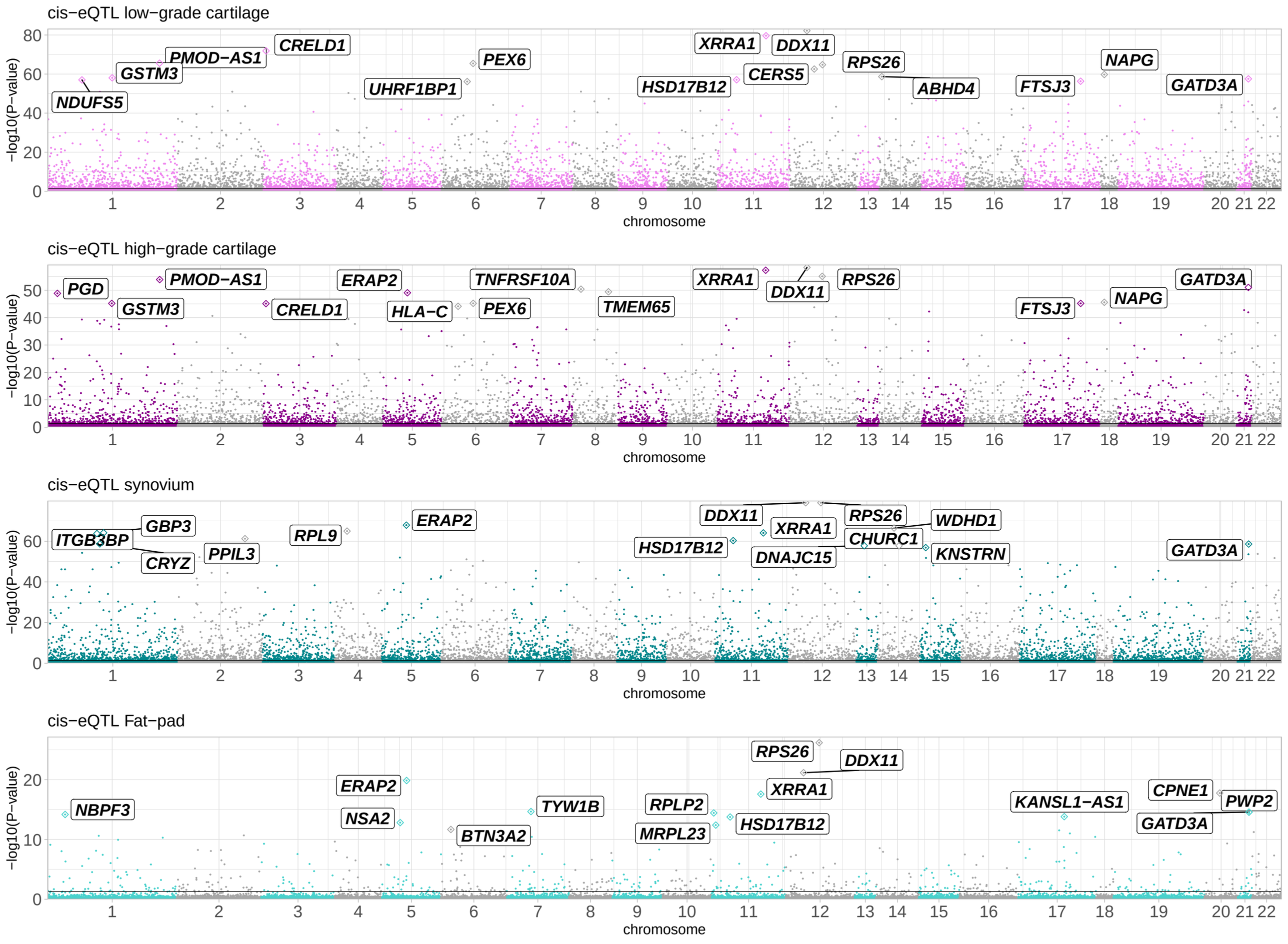
drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Research *52*, D1180–D1192. https://doi.org/10.1093/nar/gkad1004.

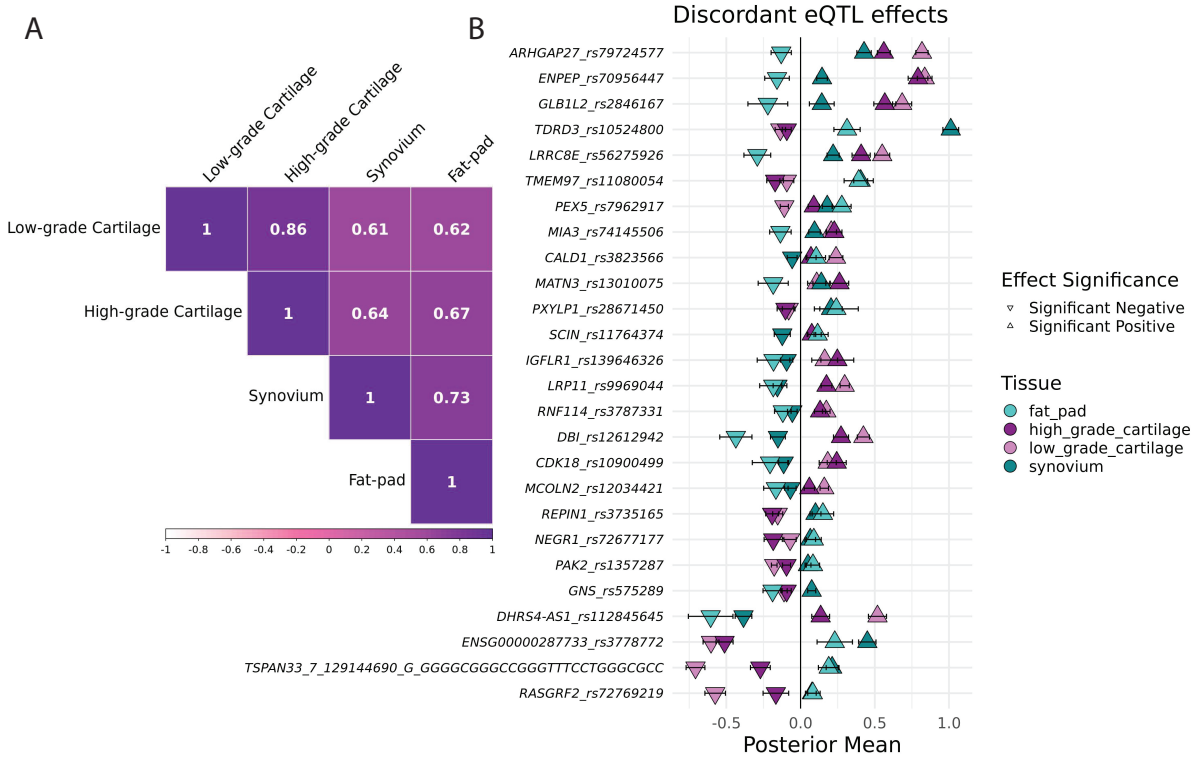
109. van Hijfte, L., Geurts, M., Vallentgoed, W.R., Eilers, P.H.C., Sillevis Smitt, P.A.E., Debets, R., and French, P.J. (2023). Alternative normalization and analysis pipeline to address systematic bias in NanoString GeoMx Digital Spatial Profiling data. iScience *26*, 105760. https://doi.org/10.1016/j.isci.2022.105760.

110. 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.

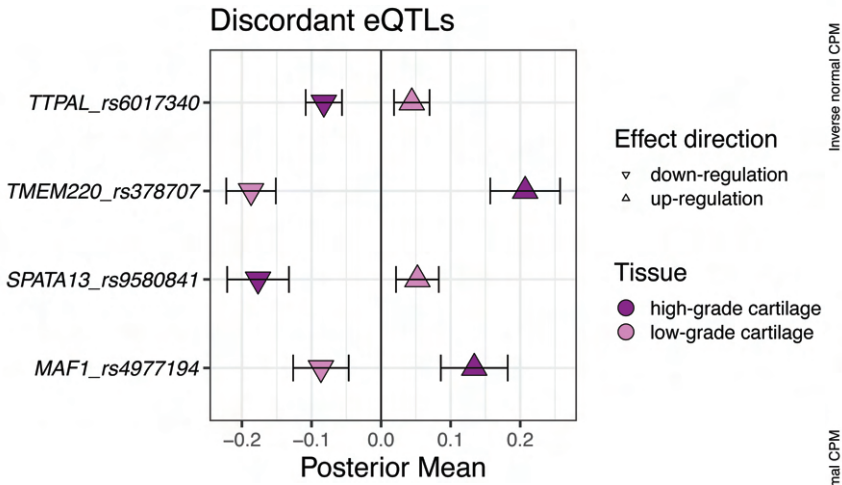
111. Westendorf, J.J., Bonewald, L.F., Kiel, D.P., and Burtt, N.P. (2022). The Musculoskeletal Knowledge Portal: improving access to multi-omics data. Nat Rev Rheumatol *18*, 1–2. https://doi.org/10.1038/s41584-021-00711-1.


Osteoarthritis patients
N=321
Knee joint primary tissues
Cartilage
Low-grade (N=263)
High-grade (N=216)
Synovium (N=278)
Intrapatellar fat pad (N=94)
RNA-sequencing
+
Whole-genome sequencing
Disease tissue-specific cis- and trans-eQTL mapping
cis-eQTLs
SNP
Gene1
Gene2
≤ 1Mb
Gene1 expression
Genotype
trans-eQTL
SNP
Gene1
Gene2
≥ 5Mb
Gene2 expression
Genotype
Colocalization analysis
GWAS
eQTL
−log10(p-value)
Position on chromosome (Mb)
Gene
Gene
Druggability analysis
Drug repurposing opportunities
Osteoarthritis effector gene prioritization
1. Molecular and functional evidence
Chondrocyte-specific Hi-C map
Knock-out
Rare and syndromic human diseases
musculoskeletal phenotypes
2. Causal inference
Confounder
Genetic variant
Instrument
Gene level
Exposure
Osteoarthritis phenotype
Outcome
38 prioritized effector genes for osteoarthritis

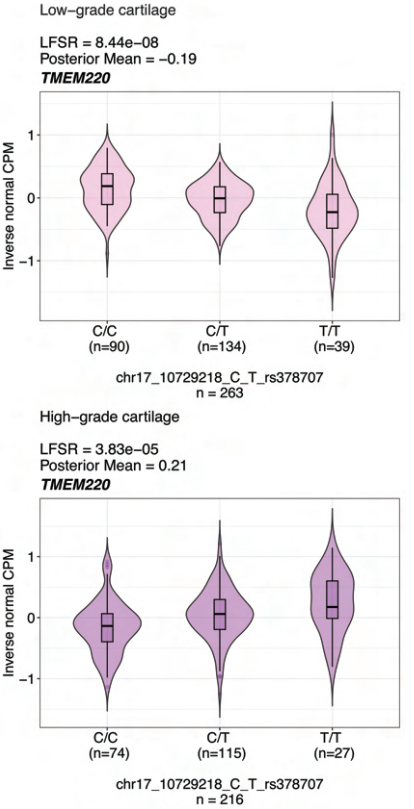

cis−eQTL low−grade cartilage
−log10(P−value)
chromosome
NDUFS5
GSTM3
PMOD−AS1
CRELD1
UHRF1BP1
PEX6
HSD17B12
XRRA1
DDX11
CERS5
RPS26
ABHD4
FTSJ3
NAPG
GATD3A
cis−eQTL high−grade cartilage
−log10(P−value)
chromosome
PGD
GSTM3
PMOD−AS1
CRELD1
ERAP2
HLA−C
TNFRSF10A
PEX6
TMEM65
XRRA1
DDX11
RPS26
FTSJ3
NAPG
GATD3A
cis−eQTL synovium
−log10(P−value)
chromosome
ITGB3BP
GBP3
CRYZ
PPIL3
RPL9
ERAP2
HSD17B12
DDX11
XRRA1
DNAJC15
RPS26
CHURC1
WDHD1
KNSTRN
GATD3A
cis−eQTL Fat−pad
−log10(P−value)
chromosome
NBPF3
ERAP2
NSA2
BTN3A2
TYW1B
RPLP2
MRPL23
HSD17B12
RPS26
DDX11
XRRA1
KANSL1−AS1
CPNE1
PWP2
GATD3A

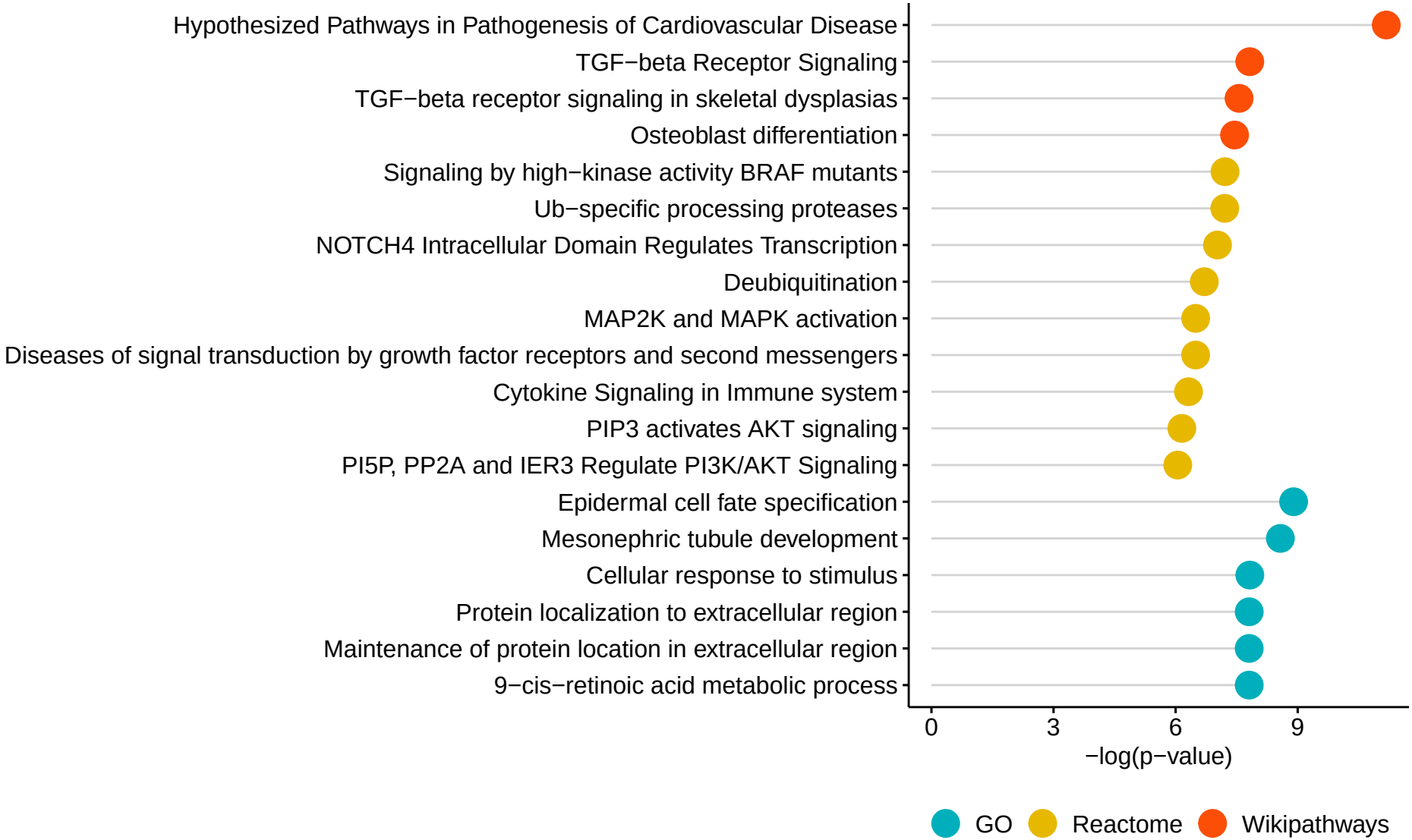

A
Low-grade Cartilage
High-grade Cartilage
Synovium
Fat-pad
1
0.86
0.61
0.62
1
0.64
0.67
1
0.73
1
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
B
Discordant eQTL effects
ARHGAP27_rs79724577
ENPEP_rs70956447
GLB1L2_rs2846167
TDRD3_rs10524800
LRRC8E_rs56275926
TMEM97_rs11080054
PEX5_rs7962917
MIA3_rs74145506
CALD1_rs3823566
MATN3_rs13010075
PXYLP1_rs28671450
SCIN_rs11764374
IGFLR1_rs139646326
LRP11_rs9969044
RNF114_rs3787331
DBI_rs12612942
CDK18_rs10900499
MCOLN2_rs12034421
REPIN1_rs3735165
NEGR1_rs72677177
PAK2_rs1357287
GNS_rs575289
DHRS4-AS1_rs112845645
ENSG00000287733_rs3778772
TSPAN33_7_129144690_G_GGGGCGGGCCGGGTTTCCTGGGCGCC
RASGRF2_rs72769219
-0.5
0.0
0.5
1.0
Posterior Mean
Effect Significance
Significant Negative
Significant Positive
Tissue
fat_pad
high_grade_cartilage
low_grade_cartilage
synovium

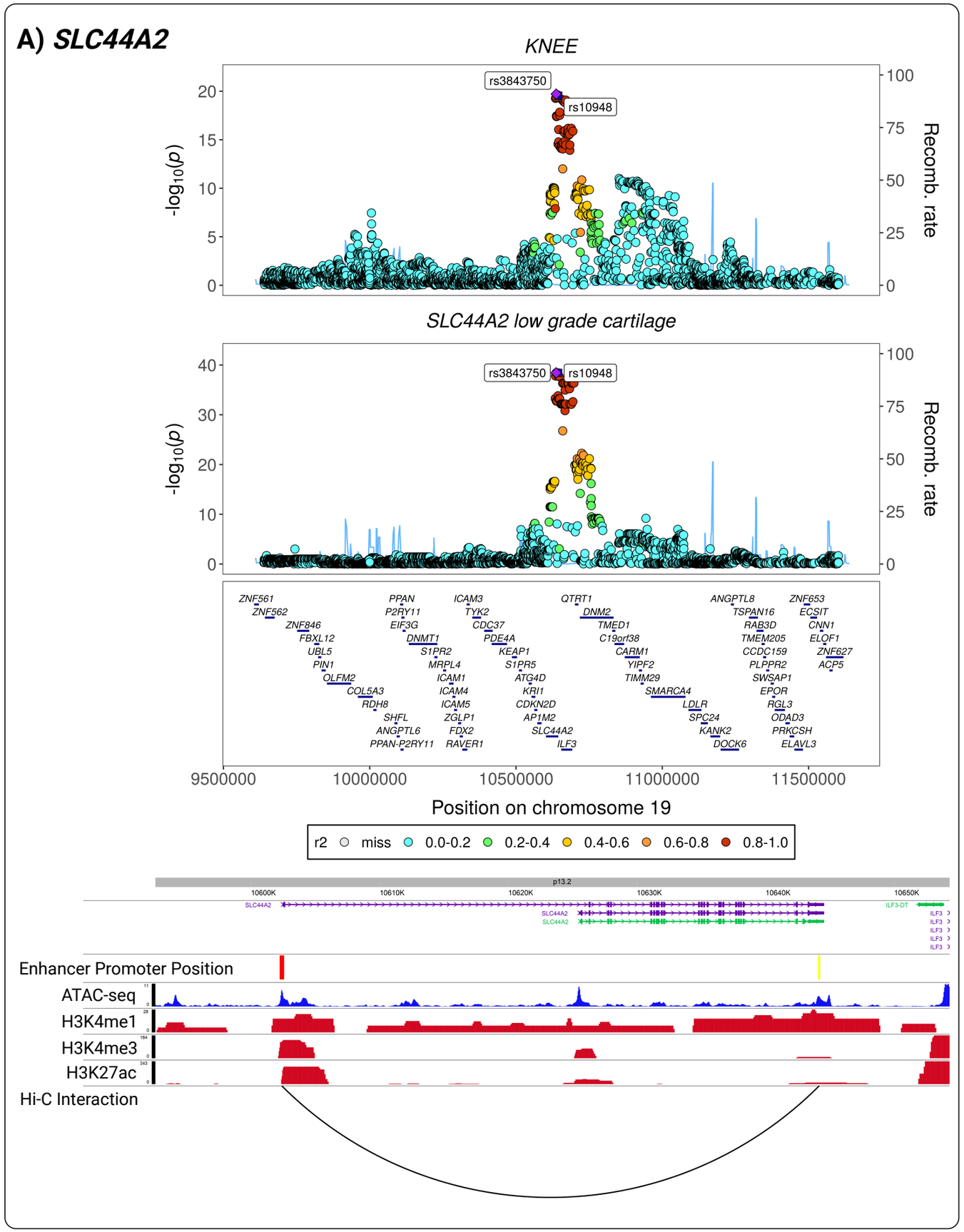
A

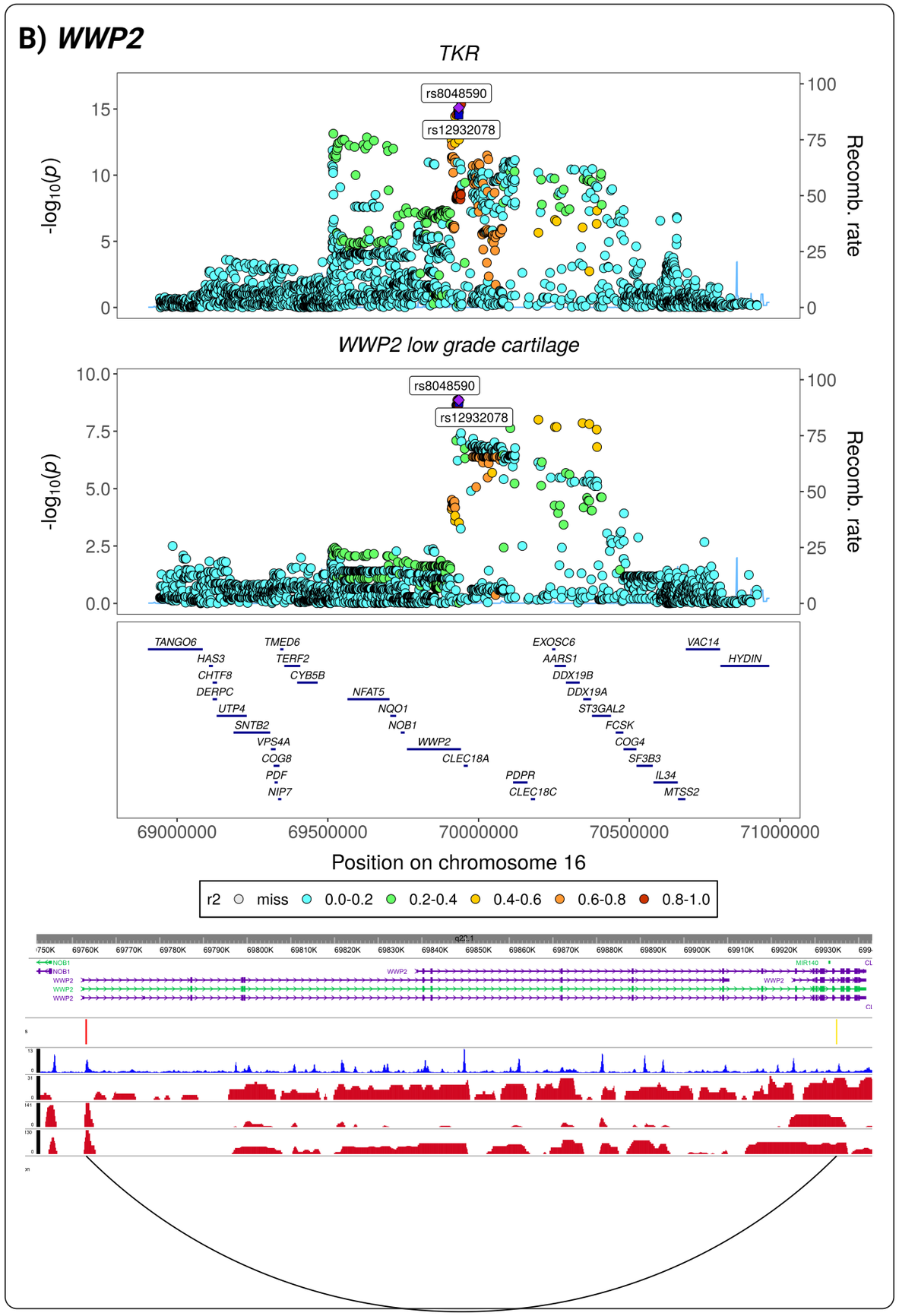

Discordant eQTLs
TTPAL_rs6017340
TMEM220_rs378707
SPATA13_rs9580841
MAF1_rs4977194
−0.2
−0.1
0.0
0.1
0.2
Posterior Mean
Effect direction
down-regulation
up-regulation
Tissue
high-grade cartilage
low-grade cartilage

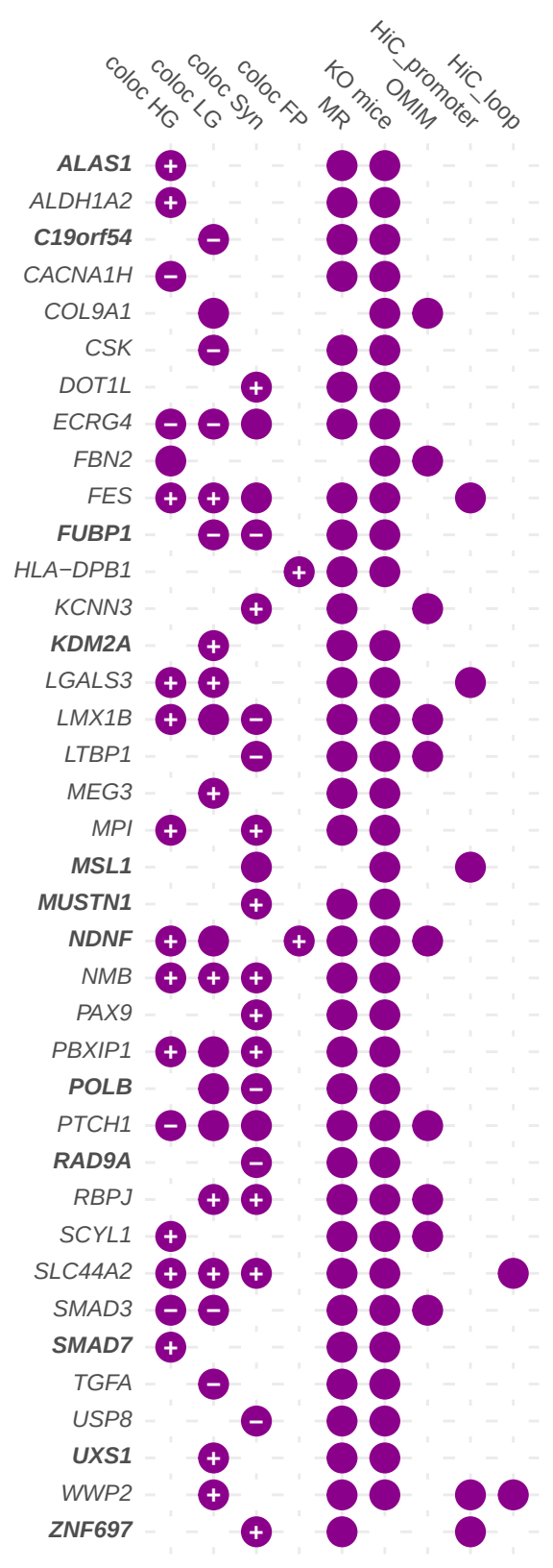
B


Low−grade cartilage
LFSR = 8.44e−08
Posterior Mean = −0.19
TMEM220
Inverse normal CPM
1
0
−1
C/C
(n=90)
C/T
(n=134)
T/T
(n=39)
chr17_10729218_C_T_rs378707
n = 263
High−grade cartilage
LFSR = 3.83e−05
Posterior Mean = 0.21
TMEM220
Inverse normal CPM
1
0
−1
C/C
(n=74)
C/T
(n=115)
T/T
(n=27)
chr17_10729218_C_T_rs378707
n = 216


Hypothesized Pathways in Pathogenesis of Cardiovascular Disease
TGF−beta Receptor Signaling
TGF−beta receptor signaling in skeletal dysplasias
Osteoblast differentiation
Signaling by high−kinase activity BRAF mutants
Ub−specific processing proteases
NOTCH4 Intracellular Domain Regulates Transcription
Deubiquitination
MAP2K and MAPK activation
Diseases of signal transduction by growth factor receptors and second messengers
Cytokine Signaling in Immune system
PIP3 activates AKT signaling
PI5P, PP2A and IER3 Regulate PI3K/AKT Signaling
Epidermal cell fate specification
Mesonephric tubule development
Cellular response to stimulus
Protein localization to extracellular region
Maintenance of protein location in extracellular region
9−cis−retinoic acid metabolic process
0
3
6
9
−log(p−value)
GO
Reactome
Wikipathways


A) SLC44A2
KNEE
rs3843750
rs10948
-log10(p)
Recomb. rate
SLC44A2 low grade cartilage
ZNF561
ZNF562
ZNF846
FBXL12
UBL5
PIN1
OLFM2
COL5A3
RDH8
SHFL
ANGPTL6
PPAN-P2RY11
PPAN
P2RY11
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
KANK2
DOCK6
ANGPTL8
TSPAN16
RAB3D
TMEM205
CCDC159
PLPPR2
SWSAP1
EPOR
RGL3
ODAD3
PRKCSH
ELAVL3
ZNF653
ECSIT
CNN1
ELOF1
ZNF627
ACP5
9500000
10000000
10500000
11000000
11500000
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
Enhancer Promoter Position
ATAC-seq
H3K4me1
H3K4me3
H3K27ac
Hi-C Interaction


B) WWP2
TKR
rs8048590
rs12932078
-log10(p)
Recomb. rate
WWP2 low grade cartilage
TANGO6
HAS3
CHTF8
DERPC
UTP4
SNTB2
VPS4A
COG8
PDF
NIP7
TMED6
TERF2
CYB5B
NFAT5
NQO1
NOB1
WWP2
CLEC18A
PDPR
CLEC18C
EXOSC6
AARS1
DDX19B
DDX19A
ST3GAL2
FCSK
COG4
SF3B3
IL34
MTSS2
VAC14
HYDIN
69000000
69500000
70000000
70500000
71000000
Position on chromosome 16
r2 miss 0.0-0.2 0.2-0.4 0.4-0.6 0.6-0.8 0.8-1.0


coloc HG
coloc LG
coloc Syn
coloc FP
KO mice
MR
HiC_promoter
OMIM
HiC_loop
ALAS1
ALDH1A2
C19orf54
CACNA1H
COL9A1
CSK
DOT1L
ECRG4
FBN2
FES
FUBP1
HLA−DPB1
KCNN3
KDM2A
LGALS3
LMX1B
LTBP1
MEG3
MPI
MSL1
MUSTN1
NDNF
NMB
PAX9
PBXIP1
POLB
PTCH1
RAD9A
RBPJ
SCYL1
SLC44A2
SMAD3
SMAD7
TGFA
USP8
UXS1
WWP2
ZNF697