

Title: Allelic expression imbalance in articular cartilage and subchondral bone refined genome-wide association signals in osteoarthritis.

Rodrigo Coutinho de Almeida¹, Margo Tuerlings¹, Yolande Ramos¹, Wouter Den Hollander¹, Eka Suchiman¹, Nico Lakenberg¹, Rob (RGHH) Nelissen², Hailiang Mei³, Ingrid Meulenbelt^{1*}.

1 Department of Biomedical Data Sciences, Section Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands.

2 Department of Orthopaedics Leiden University Medical Center, Leiden, The Netherlands.

3 Sequencing Analysis Support Core, Dept. of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands.

*To whom correspondence should be addressed:

Ingrid Meulenbelt

Address: Department of Biomedical Data Sciences, Section of Molecular Epidemiology, Leiden University Medical Center, LUMC Post-Zone S-05-P, PO Box 9600, 2300 RC Leiden, The Netherlands.

Telephone: +31 (0) 71 526 9734

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
E-mail: i.meulenbelt@lumc.nl

Abstract

Osteoarthritis (OA) is an age-related joint disease with a strong and complex genetic component. Genome-wide association studies (GWAS) discovered a large number of genomic regions associated with OA. Nevertheless, to link associated genetic variants affecting the expression of OA-risk genes in relevant tissues remains a challenge. Here, we showed an unbiased approach to identify transcript single nucleotide polymorphisms (SNPs) of OA risk loci by allelic expression imbalance (AEI). We used RNA sequencing of articular cartilage (N = 65) and subchondral bone (N= 24) from OA patients. AEI was determined for all genes present in the 100 regions reported by GWAS catalog. The count fraction of the alternative allele (ϕ) was calculated for each heterozygous individual with the risk-SNP or with the SNP in linkage disequilibrium (LD) with it. Furthermore, a meta-analysis was performed to generate a meta- ϕ (null hypothesis median $\phi=0.49$) and P-value for each SNP. We identified 30 transcript SNPs subject to AEI (28 in cartilage and 2 in subchondral bone). Notably, 10 transcript SNPs were located in genes not previously reported in the GWAS catalogue, including two long intergenic non-coding RNAs (lincRNAs), *MALAT1* (meta- $\phi=0.54$, $FDR=1.7 \times 10^{-4}$) and *ILF3-DT* (meta- $\phi=0.6$, $FDR=1.75 \times 10^{-5}$). Moreover, 14 drugs were interacting with 7 genes displaying AEI, of which 7 drugs has been already approved. By prioritizing proxy transcript SNPs that mark AEI in cartilage and/or subchondral bone at loci harboring GWAS signals, we present an unbiased approach to identify the most likely functional OA risk-SNP and gene. We identified 10 new potential OA risk genes ready for further, translation towards underlying biological mechanisms.

Introduction

Osteoarthritis (OA) is an age-related, joint disease, characterized by progressive heterogeneous changes in articular cartilage and subchondral bone. Over 80% of OA patients have limitations in movement and 25% inhibition in major daily activities of life (1,2). Up until now, there is no disease modifying treatment except for costly total joint replacement surgery at end-stage disease. As a result, OA puts a high social and economic burden on society (3). OA has a considerable and complex genetic component (4) and well-powered genetic studies are accumulating at high rate, highlighting strong OA risk genes and associated underlying disease pathways. By addressing functionality of identified OA risk alleles, such as for example for *COLGAT2* (5) and *MGP* (6), it was confirmed that OA risk single nucleotide polymorphisms (SNPs) frequently modulated OA pathology due to subtle altered transcription of a positional gene *in cis*, as such imposing a persistent negative influence on joint tissue homeostasis throughout life (6). Targeted testing of allele imbalanced expression of OA risk genes is, however, cumbersome (7–9). To address the full potential of biologically relevant DNA variants in cartilage, we previously captured allelic expression imbalance (AEI) at a genome wide scale, by virtue of a relatively small transcriptomics (RNA sequencing) dataset of paired preserved and lesioned OA cartilage (10). This study presented, however, a database of AEI SNPs in cartilage for researcher to probe their gene of interest without providing the linkage disequilibrium (LD) structure relative to the OA risk SNPs. Moreover, genome wide association studies (GWAS) in OA with ever increasing sample sizes have resulted in increasing number of OA risk SNPs with 100 currently mapped loci associating to OA related phenotypes (7). Finally, given the pathways in which the genes located in these loci act, the AEI analyses should be extended to other OA relevant tissues and particularly subchondral bone as was recognized by recently studies (5,8,9)

In the current study we used a relatively large RNA-sequencing dataset of OA cartilage and subchondral bone to capture AEI of transcript SNPs that tag OA risk SNPs recently reported by GWAS, as such ready for further study towards underlying biological mechanisms e.g. in human 3D *in vitro* model systems incorporating bone and/or cartilage tissue units.

Results

Identification of Allelic Expression Imbalance of transcript SNPs tagged by the OA risk SNPs

We initially performed a look-up on the expression levels in our previously published cartilage and subchondral bone datasets (11,12) for genes located at all 100 OA risk loci reported by GWAS catalog (7,13) residing within a 1Mb window. To obtain relevant information on the functional effect of the identified OA risk alleles on expression of positional genes, we next prioritized on positional transcript SNPs that were in LD ($r^2 > 0.6$) to the OA risk SNP and subject to AEI.

We identified 28 transcript SNPs subject to AEI in cartilage of positional genes in LD with the most associated risk SNP reported on OA GWAS (**Figure 1; Supplementary Table 1**). Notable among cartilage, we identified 10 AEI SNPs located in genes not previously highlighted as possible OA genes candidates according to GWAS catalogue (**Table 2**), including two long intergenic non-coding RNAs (lincRNAs) such as, *MALAT1* (AEI meta- $\phi=0.54$, FDR= 1.7×10^{-4}) and *ILF3-DT* (meta- $\phi=0.6$, FDR= 1.7×10^{-5}) (**Figure 2A-B**). Moreover, we identified two SNPs subject to AEI in subchondral bone of OA patients affecting the expression of three GWAS reported genes, the HLA genes *HLA-DPA1* and *HLA-DPB1* (meta- $\phi=0.76$, FDR=0.0033; **Figure 2C**) and MGP (meta- $\phi=0.33$, FDR = 1.1×10^{-19}) (**Supplementary Table 1**).

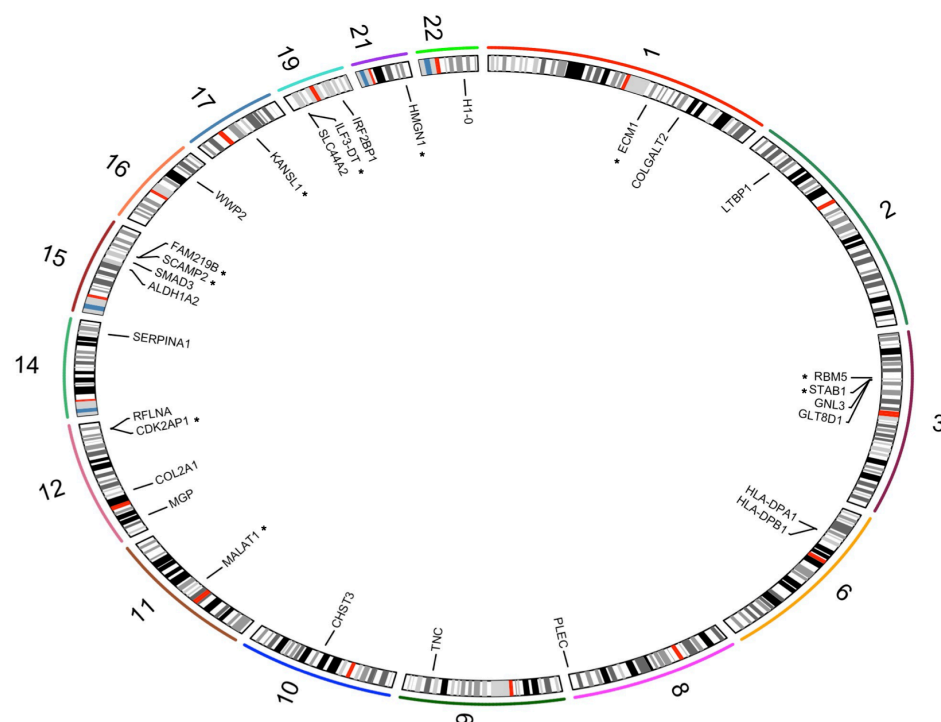


Figure 1: Circos plot showing all genes with allelic expression imbalance in OA across the genome. Stars represents genes that were not previously reported by GWAS catalog.

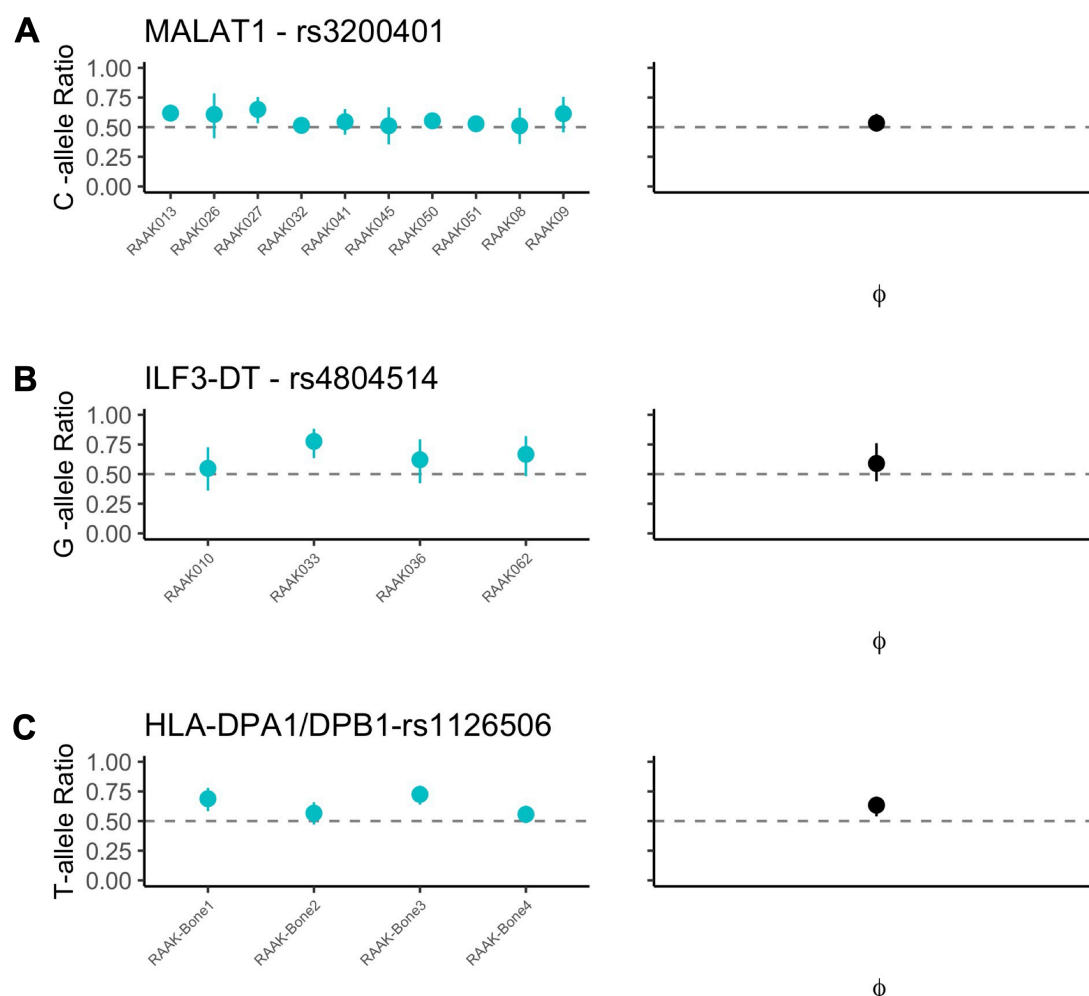


Figure 2: Allelic expression imbalance of two lincRNAs in articular cartilage and HLA genes in subchondral bone. (A) SNP rs3200401 showing increasing expression of lincRNA MALAT1 of the risk allele C in OA cartilage (meta- ϕ = 0.54). (C) LincRNA ILF3-DT with risk allele G displaying increasing expression in OA cartilage (meta- ϕ = 0.6). (D) HLA-DPA1/DPB1 with risk allele T increasing expression in OA subchondral bone (meta- ϕ = 0.76).

AEI genes differentially expressed between preserved and lesioned in cartilage and subchondral bone

To further strengthen that these genes are likely involved in OA pathophysiology, we lookup for differential expression between preserved and lesioned articular cartilage (N=35 paired) and subchondral bone (N=24) in our previously published datasets (11,12). We found nine AEI genes differentially expressed in cartilage and two in subchondral bone (**Table 3**), including *TNC* gene (meta- ϕ =0.47, FDR=2.08x10⁻⁷,

FC=1.4) which showed a downregulation AEI for the risk allele but upregulated in lesioned cartilage, suggesting that its expression in lesioned cartilage is merely a beneficial response to the occurring OA pathophysiological process (**Figure 3A**). On the other hand, the *COLGALT2* gene was differentially expressed between preserved and lesioned cartilage and with an upregulated AEI effect in cartilage (meta- ϕ =0.54, FDR = 0.00132, FC=1.4; **Figure 3B**), suggesting that a lower expression in lesioned cartilage could be potentially beneficial. Additionally, *SLC44A2* gene was differentially expressed in subchondral bone and in cartilage (**Table 3**), however with AEI was exclusively in cartilage (meta- ϕ =0.65, FDR=1.13x10⁻¹¹, **Figure 3C**), indicating that lower expression of SLC44A2 in cartilage is conferring risk to OA likely by an aberrant cartilage specific transcription factor.

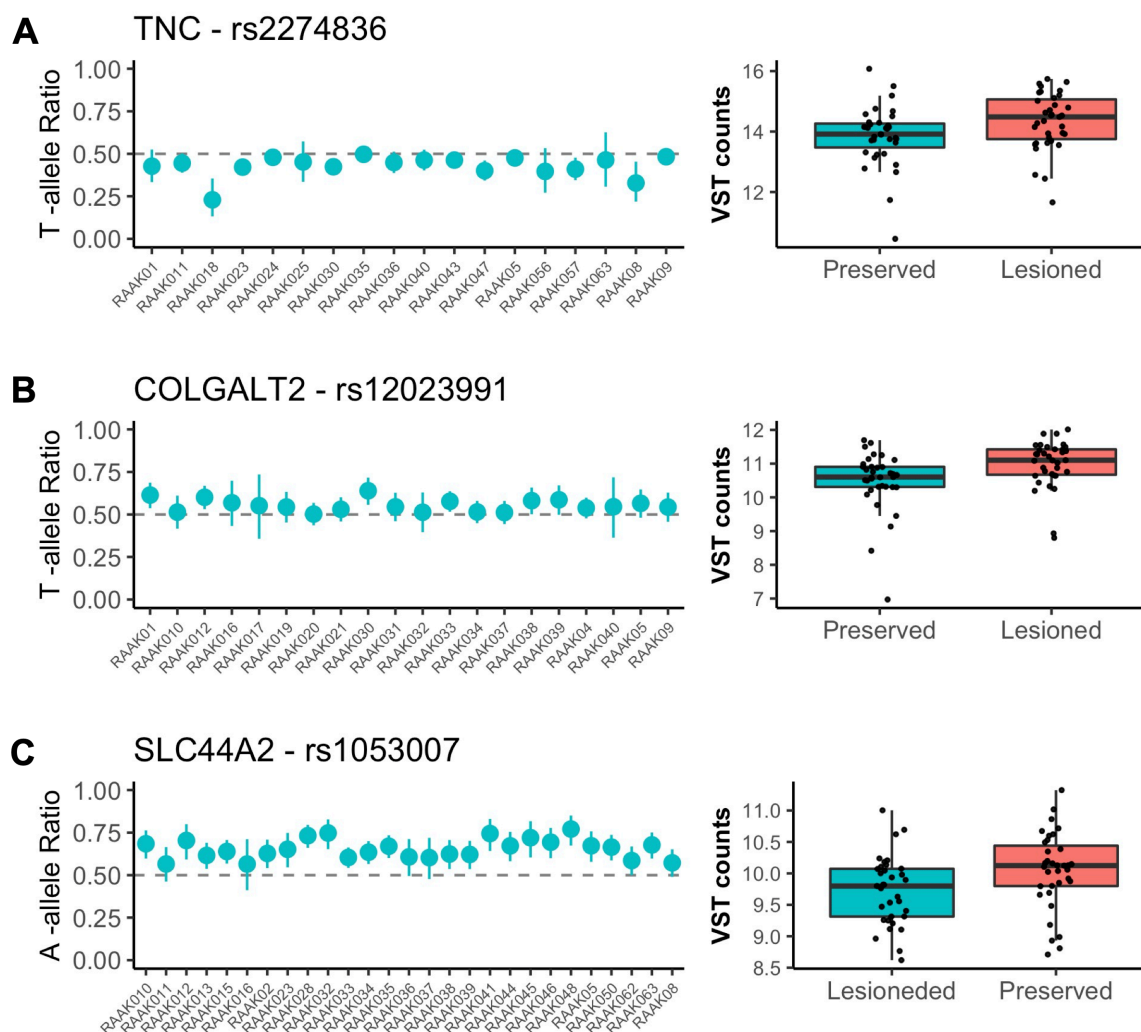


Figure 3: Genes with allelic expression imbalance (AEI) also showing differentially expressed between preserved and lesioned cartilage. (A) TNC upregulated in lesioned cartilage and with AEI downregulated effect (meta- ϕ = 0.47); (B) COLGALT2. (meta- ϕ = 0.54); (C) SLC44A2 (meta- ϕ = 0.65).

AEI genes interacting with 14 drugs

By assessing effects of identified OA risk alleles on expression of positional genes, a reliable inference can be made on potential druggable targets that could counteract this aberrant gene expression. Therefore, we next combining OA risk alleles subject to AEI to known drugs and their targets on all 30 AEI genes. In total 7 genes showed interaction with at least one drug (DGBI IG score > 0.5). Out of 15,000 drugs present on the DGBI database, 14 drugs showed 2.1-fold significant enrichment (Fisher's exact test p-value = 0.044) interaction with an AEI gene with a mode of action consistent with the direction of the risk allele expression (**Figure 4; Supp Table 2**). In addition, seven drugs have been already approved by federal agencies.

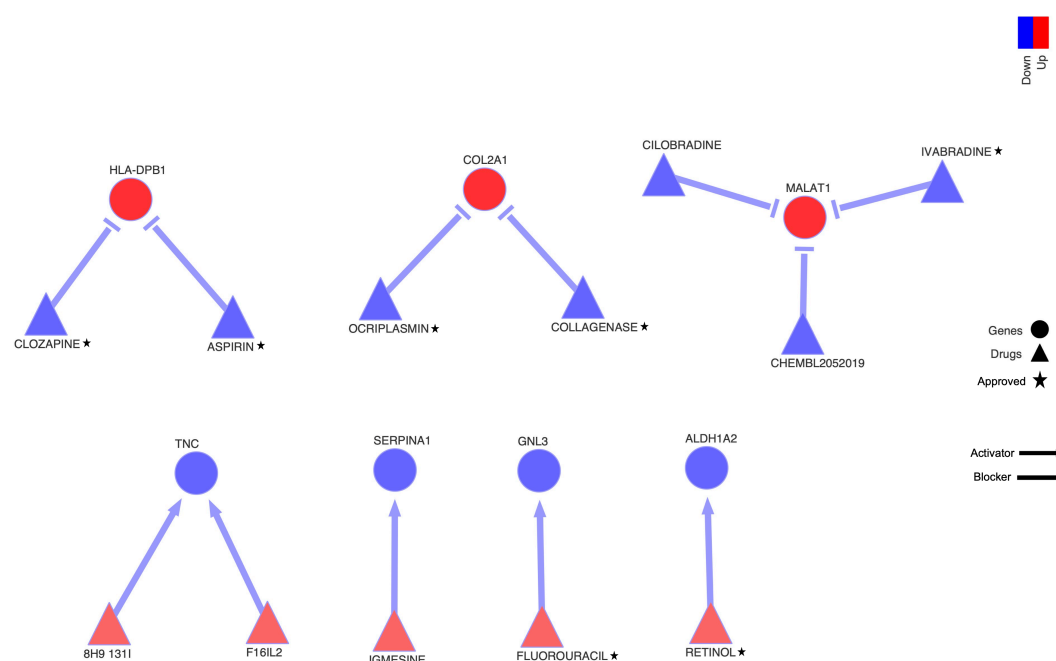


Figure 4: Seven AEI genes interacting with 14 drugs. Genes are represented by circles and drugs are triangles. The start shows drug approved federal agencies according to the database. Blue are genes with AEI downregulated effect and for drugs represents the blockers interaction. Red are genes with AEI upregulated effect and drugs with an activator interaction.

Discussion

In the current manuscript, we outlined an unbiased approach to identify positional proxy transcript SNPs of currently strong OA risk loci by AEI analyses in transcriptome-wide datasets of the OA relevant tissues articular cartilage and subchondral bone from OA patients. Furthermore, we integrated drug-gene interaction database to identify potential drugs that could counteract effects the OA risk alleles in the relevant tissue. Subsequently, we identified 10 genes subject to AEI in cartilage (**Table 2**), that were not previously recognized as the main gene by GWAS, including two lincRNAs *MALAT1* and *ILF3-DT*. Moreover, we identified AEI in HLA-DPA1/DPB1 and MGP genes also in subchondral bone. We advocate that these genes should be prioritized for further, translation towards underlying biological mechanisms e.g. in human 3D in vitro model systems incorporating bone and/or cartilage tissue units.

To our knowledge, we are the first to identify AEI of lincRNAs *MALAT1* and *ILF3-DT* in articular cartilage, highlighting that the OA risk is associated to upregulated expression (meta- ϕ =0.54 and meta- ϕ =0.6, respectively, **Table 2**). Although these two lincRNAs were also expressed in subchondral bone (data not shown) we could not find significant effect due to the low number of heterozygous in this dataset. The *Metastasis Associated Lung Adenocarcinoma Transcript 1* (*MALAT1*) is a well-conserved lncRNA involved in several diseases, and it is known to play a role in transcription, epigenetic regulation and splicing(14). *MALAT1* regulate inflammation, cellular responses to oxidative stress and cellular senescence(15). Noteworthy, *MALAT1* is decreased in senescent cells(16) and in bleomycin-induced murine fibrosis where myeloid deletion of *MALAT1* increases susceptibility to fibrosis and the number of profibrotic M2 macrophages(17). *MALAT1* was identified as sponges for miR-150-5p and it was found up-regulated in OA chondrocytes compared with normal chondrocytes(18). Nevertheless, *MALAT1* was not significant differentially expressed between lesioned and preserved cartilage(19) or in subchondral bone(20) in our dataset. We found allelic imbalance of the allele C (in LD with the risk allele G) with an upregulated effect (**Figure 2B**). It is tempting to speculate that *MALAT1* risk allele could be partially regulating inflammation, or even senescence pathways in OA. Being highly tissue specific and condition-specific expression pattern lincRNAs have been

proposed as new tool for a particular and personalized therapeutic approach(21). In addition, cartilage is a postmitotic tissue, highly dependent on adequate epigenetic mechanisms to establish dynamic changes in gene expression, which make both lincRNAs (*MALAT1* and *ILF3-DT*) promising candidates to be use as therapeutical tools. Nevertheless, more function studies (e.g CRISPR-cas9) on these regulatory variants will be necessary to clarify this issue.

Even though we took a different approach than our previously allelic transcriptomic-wide study (10), we replicated previously reported AEI in the important OA risk genes, *MGP*, *TNC* and *WWP2* in cartilage. Additionally, we showed AEI for *MGP* in both cartilage and subchondral bone tissues with the same direction of effect but different SNPs in high LD with the top GWAS SNP (downregulation effect of the rs1800801-T risk allele, meta- ϕ =0.37 in cartilage and rs4236-C risk allele meta- ϕ =0.33 in subchondral bone, **Table 3**). Previously *MGP* was identified with a low expression level for the risk allele in cartilage, fat pad and synovium, however in blood *MGP* expression was increased (8).Here we confirmed similar effect for AEI in *MGP* in the two main OA related tissues, cartilage and subchondral bone. In this regard, *TNC* and *WWP2* appeared to be showing AEI specifically in articular cartilage. Despite the fact that they are robustly expressed in subchondral bone and had a reasonable number of heterozygote samples (seven and three, respectively) they did not show significant AEI in this tissue. Henceforth, these OA risk alleles may likely confer risk to OA particularly via aberrant function in articular cartilage. Nevertheless, further confirmation in a larger transcriptomic dataset of subchondral bone is necessary to further corroborate this statement.

A notable finding presented here is the highly significant and consistent AEI of the OA risk gene *TNC* in articular cartilage. *TNC* encodes the glycoprotein Tenacin that is abundantly expressed (75th percentile) in the extracellular matrix (22). *TNC* is a key molecule in cartilage remodelling and participates in chondrogenesis and cartilage development (23). *TNC* may be involved in mechanotransduction in response to mechanical stress (22). The here reported AEI indicated that the OA risk allele T (rs2274836 SNP) is associated to downregulation of *TNC* in cartilage of OA patients, but *TNC* was upregulated in lesioned as compared to preserved OA cartilage. The latter suggesting a beneficial *TNC* response of chondrocytes in cartilage during the OA pathophysiology e.g. to restore tissue homeostasis. These findings are in line with *in vitro* cells models (24) showing that *TNC* expression levels was still normal at the

early OA phase, but with its expression disappearing at the late phase with cartilage maturation. In addition, it has been shown that administration of intra-articular TNC in rabbits using scaffolding matrices, promoted the repair of cartilage defects (25), suggesting, that *TNC* deficiency enhance cartilage degeneration. Together these data indicate that aberrant *TNC* signaling may confer important risk to OA, as such should be prioritized for functional follow up research towards underlying OA disease mechanisms.

COLGALT2 was among the nine genes with differential expression between preserved and lesioned cartilage (meta- ϕ =0.54 and FC=1.42, **Figure 3B**; **Table 2**). *COLGALT2* encodes procollagen galactosyltransferase 2, an enzyme that post-translationally glycosylates collagen. This gene has been also previously identified with methylation quantitative trait loci (mQTL) (26). Based on the mQTL it was speculated that the expression changes in this gene is mediated by differential enhancer methylation (26). The authors further suggested that a reparative response by *COLGAT2*. Recently, using several molecular biology tools, including CRISPR-Cas9, Kehayova et al.,(5) proposed that the OA risk allele from SNP rs11583641, mediates decreased levels of DNA methylation in cartilage at the *COLGALT2* enhancer, which ultimately increased expression of the gene (5). Our data support these findings, since we showed the SNP rs12023991 in high LD with the most associated GWAS SNP rs11583641, with upregulation effect of the risk allele (meta- ϕ =0.54, FDR = 0.00132), followed by an upregulation of *COLGALT2* in lesioned cartilage (FC=1.42, FDR =1.92x10⁻⁵), suggesting allele-specific co-regulation element that could explain these differences. The upregulation could require methylation however only via the risk allele.

The prioritization of druggable targets based on genetic support has been proven to highly increase efficacy of clinical trials. OA GWAS genes can result in candidate targets for drug discovery (7). We integrated allelic expression imbalance of OA risk variants (or variants in LD with it) with a large drug-gene target database (DGBIv4). We were able to identify 14 drugs interacting with at least one AEI gene (**Figure 4**). Of note, *SERPINA2* gene interacts with Igmesine drug, which could potentially reverse the allelic downregulation effect by enhancing *SERPINA2* expression. Nevertheless, this drug has not yet being approved by federal agencies to be use, therefore more studies to confirm this candidate should be performed. In the other hand, we found seven drugs which has been already approved to be used, including Retinol which interacts with *ALDH1A2* gene and Ivabradine which interacts with the lncRNA

MALAT1. The Retinol binding protein 4 (RBP4) acts as an immunomodulatory adipocytokine (a bioactive product produced by adipose tissue), and recently was found positively correlated with MMP-1 and MMP-3 in chondrocytes and highly expressed in synovial fluid and plasma from OA patients(27), suggesting a possible role of this molecule in the OA pathogenesis. Moreover, Ivabradine was found to induce up-regulation of matrix metalloproteinase-3 (MMP-3) and MMP-13 at gene and protein levels(28), which are key genes for OA pathophysiology. In addition, as we suggested above, the lncRNA MALAT1 appears to have an important role in OA. MALAT1 in chondrocytes promotes chondrocyte proliferation, suppress chondrocyte apoptosis and reduces extracellular matrix degradation (29,30). Recently, the use of Reversatrol to suppress the activation of MALAT1 showed a decreased of pro-inflammatory genes (NF- κ B1 and IL-6) by modulating a microRNA (miR-9) (31). Considering that Ivabradine is a blocker modulator, we can speculate that this drug could act in a similar way as Reversatrol with MALAT1 to reverse the effects of OA response. Nevertheless, more *in-vitro* and *in-vivo* studies will be essential to clarify this hypothesis.

Taken together, by applying AEI in disease-relevant tissues showed that we can improve the selection of possible causal genes associated to OA. Moreover, we were able to identify new candidate genes OA, including two lincRNAs. This was the first time that lincRNA involved in OA showed a direct interaction with approved drug. Thus, by pinpointing the risk allele which influence gene expression in disease-relevant tissues, we can pave the way for further post-GWAS *in-vitro* models in OA.

Methods

Samples

Ethical approval for the Research Arthritis and Articular Cartilage (RAAK) study(32) was obtained from the medical ethics committee of the Leiden University Medical Center (P08.239/P19.013), and informed consent was obtained from all patients included. The current study included 65 OA patients (39 knees and 26 hips) (**Table 1**) who underwent a joint replacement surgery, from which macroscopically unaffected (preserved N=56) cartilage entered the analyses complemented with independent samples for which we had available affected (lesioned N=9) OA articular cartilage only.

In addition, unaffected subchondral bone from 26 OA patients all overlapping with cartilage samples were also included.

RNA-Sequencing

Total RNA from articular cartilage and subchondral bone was isolated using Qiagen RNeasy Mini Kit (Qiagen, GmbH, Hilden, Germany). Paired-end 2×100 bp RNA-Sequencing (Illumina TruSeq RNA Library Prep Kit, Illumina HiSeq2000 and Illumina HiSeq4000) was performed. Strand specific RNA-seq libraries were generated which yielded a mean of 20 million reads per sample. Detailed description for quality control (QC) alignment, mapping and normalization are described earlier for cartilage (11) and for subchondral bone (12). In addition, common population SNP sites from the Genome of the Netherlands (33) were masked during read alignment to prevent potential reference alignment bias. AEI events were assessed on SNPs called using SNVMix2 with default settings (34) with minimum coverage of 25 and at least 10 reads per allele. Raw autosomal SNP-level data, for SNPs with ≥ 10 reads, was annotated by assigning heterozygous SNPs to genes using Ensembl v97 and SNPdb v151.

Allelic Expression Imbalance

To perform AEI, we used transcriptome-wide sequencing (RNA-seq) data after QC from articular cartilage (N=65) and subchondral bone (N=24) from OA patients. While the cartilage samples were not fully paired (56 preserved and 9 lesioned cartilage), subchondral bone samples were paired (24 pairs) and overlapping with cartilage. First, we calculated the LD of each 100 associated loci reported on GWAS catalog database(13) (**Supplementary Table 3**), taking a window of 1Mb from the most associated variant reported by each study. To identify AEI events, we used our previously transcriptome-wide approach(10). First, the count fraction of the alternative alleles among the alternative and reference alleles together (ϕ) was determined for each heterozygous individual. Finally, a meta-analysis was performed to generate a meta- ϕ and P-value for each genetic variant with a false discovery rate (FDR) correction for multiple tests. Additionally, differential expression analysis between paired samples of preserved and lesioned cartilage (N=35 pairs) and subchondral bone (N=24) was verified in our previous published datasets (11,12), and further intersected with all genes reported to be associated to OA in the GWAS catalog

database. We assess the average normalized expression in cartilage and subchondral bone using the DESeq2 normalization approach (35) for all OA associated genes.

Data integration with Drug–Gene Interaction Database (DGIdb 4.0)

The Drug Gene Interaction Database (DGIdb v4.0) (36) was used to determine potentially druggable targets for all genes that show allelic expression imbalance. Currently, the DGIdb database contains more than 40,000 genes (including protein-coding genes and non-coding RNAs) and 10,000 drugs involved in over 15,000 drug-gene interactions. To identify drugs known to act directly on the specific gene we limited the DGIdb interaction group (IG) score >0.50 . This score is based on: (1) the amount of evidence from different sources supporting the drug-gene (i.e. the number of publications and different databases); (2) the number of genes set on each search that interact with the given drug; (3) the degree to which the drug has known interactions with other gene (36). Moreover, to identify enrichment of the overlapping AEI genes with genes interacting with drugs in the DGI database, we applied a Fisher's exact test using all genes expressed in articular cartilage and subchondral bone ($N=21,666$) as background. Finally, to prioritize drug-gene targets in OA context we removed drugs showing a mode of action inconsistent with the AEI direction.

Acknowledgements

We thank all study participants of the RAAK study. The Leiden University Medical Centre have and are supporting the RAAK. We thank Evelyn Houtman, Marcella Hoolwerff, Enrike van der Linden, Robert van der Wal, Peter van Schie, Shaho Hasan, Maartje Meijer, Daisy Latijnhouwers, Anika Rabeling-Hoogenstraaten, and Geert Spierenburg for their contribution to the collection of the joint tissue. The research leading to these results has received funding from:

- TreatOA which is funded by the European Commission framework 7 programme grant 200800.
- BBMRI Metabolomics Consortium funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO, grant nr 184.021.007 and 184033111).
- Dutch Scientific Research council NWO /ZonMW VICI scheme (91816631/528)

Conflict of Interest Statement

All authors declare no competing interests.

Data availability statement

RNA-seq data are deposited at the European Genome-Phenome Archive (accession number: EGAS00001004476) and ArrayExpress (E-MTAB-7313). Data are available upon reasonable request by any qualified researchers who engage in rigorous, independent scientific research, and will be provided following review and approval of a research proposal and Statistical Analysis Plan (SAP) and execution of a Data Sharing Agreement (DSA). All other data relevant to the study are included in the article

References

1. Neogi, T. (2013) The epidemiology and impact of pain in osteoarthritis. *Osteoarthritis and Cartilage*, **21**,1145-53.
2. Mandl, L.A. (2019) Osteoarthritis year in review 2018: clinical. *Osteoarthritis and Cartilage*, **27**,359-364.
3. Woolf, A.D., Erwin, J. and March, L. (2012) The need to address the burden of musculoskeletal conditions. *Best. Pract. Res. Clin. Rheumatol*, **26**, 183–224.
4. MacGregor, A. (1999) Twins and the genetic architecture of osteoarthritis. *Rheumatology*, **38**, 583-8.
5. Kehayova, Y.S., Watson, E., Wilkinson, J.M., Loughlin, J. and Rice, S.J. (2021) Genetic and epigenetic interplay regulates *COLGALT2*, contributing to osteoarthritis genetic risk. *Arthritis & Rheumatology* **73**,1856-1865.
6. Loughlin, J. (2015) Genetic contribution to osteoarthritis development. *Current Opinion in Rheumatology*, **27**, 284-8.
7. Boer, C.G., Hatzikotoulas, K., Southam, L., Stefánsdóttir, L., Zhang, Y., Coutinho de Almeida, R., Wu, T.T., Zheng, J., Hartley, A., Teder-Laving, M., *et al.* (2021) Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell*, **184**, 4784-4818.e17.
8. Shepherd, C., Reese, A.E., Reynard, L.N. and Loughlin, J. (2019) Expression analysis of the osteoarthritis genetic susceptibility mapping to the matrix Gla protein gene MGP. *Arthritis Research & Therapy*, **21**, 149.
9. Shepherd, C., Zhu, D., Skelton, A.J., Combe, J., Threadgold, H., Zhu, L., Vincent, T.L., Stuart, P., Reynard, L.N. and Loughlin, J. (2018) Functional Characterization of the Osteoarthritis Genetic Risk Residing at *ALDH1A2* Identifies rs12915901 as a Key Target Variant. *Arthritis & Rheumatology*, **70**, 1577-1587.
10. Hollander, W., Pulyakhina, I., Boer, C., Bommer, N., Breggen, R., Arindrarto, W., Coutinho de Almeida, R., Lakenberg, N., Sentner, T., Laros, J.F.J., *et al.* (2019) Annotating Transcriptional Effects of Genetic Variants in Disease-Relevant Tissue: Transcriptome-Wide Allelic Imbalance in Osteoarthritic Cartilage. *Arthritis & Rheumatology*, **71**, 561-570.
11. Coutinho de Almeida, R., Ramos, Y.F.M., Mahfouz, A., den Hollander, W., Lakenberg, N., Houtman, E., van Hooijwerff, M., Suchiman, H.E.D., Rodriguez Ruiz,

- A., Slagboom, P.E., *et al.* (2019) RNA sequencing data integration reveals an miRNA interactome of osteoarthritis cartilage. *Ann Rheum Dis*, **78**, 270–277.
12. Tuerlings, M., Hoolwerff, M., Houtman, E., Suchiman, E.H.E.D., Lakenberg, N., Mei, H., Linden, E.H.M.J., Nelissen, R.R.G.H.H., Ramos, Y.Y.F.M., Coutinho de Almeida, R., *et al.* (2021) RNA Sequencing Reveals Interacting Key Determinants of Osteoarthritis Acting in Subchondral Bone and Articular Cartilage: Identification of *IL11* and *CHADL* as Attractive Treatment Targets. *Arthritis & Rheumatology*, **73**, 789-799.
13. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., *et al.* (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, **47**, 47(D1):D1005-D1012.
14. Fei, J., Jadaliha, M., Harmon, T.S., Li, I.T.S., Hua, B., Hao, Q., Holehouse, A.S., Reyer, M., Sun, Q., Freier, S.M., *et al.* (2017) Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *Journal of Cell Science*, **130**, 4180-4192.
15. Puvvula, P.K. (2019) LncRNAs Regulatory Networks in Cellular Senescence. *International Journal of Molecular Sciences*, **20**, 2615.
16. Grammatikakis, I., Panda, A.C., Abdelmohsen, K. and Gorospe, M. (2014) Long noncoding RNAs (lncRNAs) and the molecular hallmarks of aging. *Aging*, **6**, 992-1009.
17. Omote, N. and Sauler, M. (2020) Non-coding RNAs as Regulators of Cellular Senescence in Idiopathic Pulmonary Fibrosis and Chronic Obstructive Pulmonary Disease. *Frontiers in Medicine*, **7**, 603047.
18. Zhang, Y., Wang, F., Chen, G., He, R. and Yang, L. (2019) LncRNA MALAT1 promotes osteoarthritis by modulating miR-150-5p/AKT3 axis. *Cell & Bioscience*, **9**.
19. Hoolwerff, M., 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.
20. Tuerlings, M., van Hoolwerff, M., van Bokkum, J.M., Suchiman, H.E.D., Lakenberg, N., Broekhuis, D., Nelissen, R.G.H.H., Ramos, Y.F.M., Mei, H., Cats, D., *et al.* (2021) Long non-coding RNA expression profiling of subchondral bone reveals *AC005165.1* modifying *FRZB* expression during osteoarthritis. *Rheumatology*. keab826. doi: 10.1093/rheumatology/keab826.
21. Arun, G., Diermeier, S.D. and Spector, D.L. (2018) Therapeutic Targeting of Long Non-Coding RNAs in Cancer. *Trends in Molecular Medicine*, **24**, 257-277.
22. Imanaka-Yoshida, K. and Aoki, H. (2014) Tenascin-C and mechanotransduction in the development and diseases of cardiovascular system. *Frontiers in Physiology*, **5**, 283.
23. Hasegawa, M., Yoshida, T. and Sudo, A. (2020) Tenascin-C in Osteoarthritis and Rheumatoid Arthritis. *Frontiers in Immunology*, **11**, 577015.
24. Nakoshi, Y., Hasegawa, M., Akeda, K., Iino, T., Sudo, A., Yoshida, T. and Uchida, A. (2010) Distribution and role of tenascin-C in human osteoarthritic cartilage. *Journal of Orthopaedic Science*, **15**, 666-73.
25. Ikemura, S., Hasegawa, M., Iino, T., Miyamoto, K., Imanaka-Yoshida, K., Yoshida, T. and Sudo, A. (2015) Effect of tenascin-C on the repair of full-thickness osteochondral defects of articular cartilage in rabbits. *Journal of Orthopaedic Research*, **33**, 563-71.

26. 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 and Cartilage*, **27**, 1545–1556.
27. Scotece, M., Koskinen-Kolasa, A., Pemmari, A., Leppänen, T., Hämäläinen, M., Moilanen, T., Moilanen, E. and Vuolteenaho, K. (2020) Novel adipokine associated with OA: retinol binding protein 4 (RBP4) is produced by cartilage and is correlated with MMPs in osteoarthritis patients. *Inflammation Research*, **69**, 415–421.
28. Xiang, X., Zhou, Y., Sun, H., Tan, S., Lu, Z., Huang, L. and Wang, W. (2019) Ivabradine abrogates TNF- α -induced degradation of articular cartilage matrix. *International Immunopharmacology*, **66**, 347–353.
29. Pan, C., Huang, W., Chen, Q., Xu, J., Yao, G., Li, B., Wu, T., Yin, C. and Cheng, X. (2021) LncRNA Malat-1 From MSCs-Derived Extracellular Vesicles Suppresses Inflammation and Cartilage Degradation in Osteoarthritis. *Frontiers in Bioengineering and Biotechnology*, **9**, 772002.
30. Kong, X., Wang, J., Cao, Y., Zhang, H., Lu, X., Wang, Y., Bo, C., Wang, T., Li, S., Tian, K., *et al.* (2019) The long noncoding RNA MALAT-1 functions as a competing endogenous RNA to regulate MSL2 expression by sponging miR-338-3p in myasthenia gravis. *Journal of Cellular Biochemistry*, **120**, 5542–5550.
31. Zhang, G., Zhang, H., You, W., Tang, X., Li, X. and Gong, Z. (2020) Therapeutic effect of Resveratrol in the treatment of osteoarthritis via the MALAT1/miR-9/NF- κ B signaling pathway. *Experimental and Therapeutic Medicine*.
32. Ramos, Y.F.M., den Hollander, W., Bovée, J.V.M.G., Bomer, N., van der Breggen, R., Lakenberg, N., Christiaan Keurentjes, J., Goeman, J.J., Eline Slagboom, P., Nelissen, R.G.H.H., *et al.* (2014) Genes involved in the osteoarthritis process identified through genome wide expression analysis in articular cartilage; the RAAK study. *PLoS ONE*, **9**, e103056.
33. The Genome of the Netherlands Consortium (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet*, **46**, 818–825.
34. Goya, R., Sun, M.G.F., Morin, R.D., Leung, G., Ha, G., Wiegand, K.C., Senz, J., Crisan, A., Marra, M.A., Hirst, M., *et al.* (2010) SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics*, **26**, 730–6.
35. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.
36. Freshour, S.L., Kiwala, S., Cotto, K.C., Coffman, A.C., McMichael, J.F., Song, J.J., Griffith, M., Griffith, O.L. and Wagner, A.H. (2021) Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowdsourcing efforts. *Nucleic Acids Research*, **49**, D1144–D1151.

List of Supplementary Tables

Supplementary Table 1: Overview of allelic expression imbalance SNPs in cartilage and bone.

Supplementary Table 2: Drug interaction with allelic imbalance expression genes.

Supplementary Table 3: GWAS catalog results for osteoarthritis used in this study.

Table 1: Samples characteristics

Cartilage	Total (N=65)	RNAseq - hip (N =26)	RNAseq - knee (N=39)
Age (SD)	64.5 (16.7)	64.2 (16)	64,6 (17.4)
Females	54	20	34
Bone	Total (N=26)	RNAseq - hip (N=6)	RNAseq – knee (N=18)
Age (SD)	68.1 (9.5)	67.8 (8.8)	65.7 (8.5)
Females	27	6	16

Table 2: AEI genes not reported in GWAS catalog.

AEI Gene	Biotype	GWAS SNP	GWAS associated gene	AEI SNP	R ²	chr	Position (hg38)	ALT	MAF	Meta - ϕ	FDR-AEI	Tissue
ECM1	PC	rs12040949	RPRD2	rs13294	0.93	chr1	150512511	A	0.26	0.53	3.5x10 ⁻³	Cartilage
CDK2AP1	PC	rs1060105	SBNO1	rs6633	0.94	chr12	123261262	T	0.12	0.44	2.2 x10 ⁻¹⁰	Cartilage
MALAT1	lincRNA	rs10896015	LTBP3	rs3200401	0.65	chr11	65504361	T	0.14	0.46	3.3x10 ⁻⁴	Cartilage
SCAMP2	PC	rs35206230	CSK	-								
			LMAN1L	rs12487	0.72	chr15	74844353	C	0.18	0.45	0.03	Cartilage
			CSK-									
FAM219B	PC	rs35206230	LMAN1L	rs1127796	0.63	chr15	74900663	C	0.38	0.48	0.01	Cartilage
ILF3-DT	lincRNA	rs1560707	SLC44A2	rs4804514	0.95	chr19	10653527	T	0.33	0.40	4.3x10 ⁻⁴	Cartilage
RBM5	PC	rs62262139	RBM6	rs1138536	0.66	chr3	50115923	C	0.25	0.48	7.5x10 ⁻³	Cartilage
KANSL1	PC	rs62063281	MAPT	rs35643216	0.90	chr17	46171471	C	0.19	0.41	0.04	Cartilage
STAB1	PC	rs11177	GNL3	rs13621	0.74	chr3	52524117	C	0.38	0.47	1.2x10 ⁻⁴	Cartilage
HMGN1	PC	rs9981884	BRWD1	rs3167757	0.92	chr21	39342552	T	0.47	0.41	1.1x10 ⁻⁵	Cartilage

AEI: Allelic Expression Imbalance; PC: protein coding; chr: chromosome; ALT: alternative allele; MAF: Minor Allele Frequency; FDR: False Discovery Rate.

Table 3: AEI genes differentially expressed in cartilage and bone.

AEI DEG	Average Expression (Normalized read counts)	Fold Change	FDR DEG	GWAS SNP	AEI ALLELE	r ²	Meta - ϕ	FDR AEI
<i>COLGALT2</i>	1918.57 (Upper quantile-75%) (cartilage)	1.42	1.92x10 ⁻⁵	rs11583641_C	rs12023991_C	1	0.46	0.021
<i>ILF3-DT</i>	142.32 (Upper quantile-75%) (cartilage)	0.69	3.41x10 ⁻⁴	rs1560707_T	rs4804514_T	0.95	0.4	4.3 x10 ⁻⁴
<i>MGP</i> (bone)	38663.18 (Upper quantile-75%) (cartilage) (3826.17- (Upper quantile-75%))	1.43 (1.53)	0.02 (NS)	rs4764133_T	rs1800801_T	0.95	0.37 (0.33)	1.06x10 ⁻⁶ (1.1x10 ⁻¹⁹)
<i>RFLNA</i>	90.98 (Down quantile-50%)	0.48	6.48x10 ⁻⁵	rs4765540_C	rs12551_C	0.74	0.48	0.006
<i>SLC44A2</i> (bone)	1014.102 (Upper 75%) (1383.285(Upper quantile-75%))	0.80 (0.84)	0.028 (0.019)	rs1560707_T	rs1053007_G	0.95	0.34 (NS)	1.6x10 ⁻⁹ (NS)
<i>SMAD3</i>	492.69 (Upper quantile-75%)	0.83	0.028	rs12901071_A	rs1061427_A	0.60	0.45	0.01
<i>TNC</i>	21349.84 (Upper quantile-75%)	1.41	0.010	rs2480930_A	rs2274836_T	1	0.47	2.08x10 ⁻⁷
<i>WWP2</i>	5381.64 (Upper quantile 75%)	0.79	0.034	rs34195470_A	rs2270841_C	0.65	0.54	0.0046
<i>CHTS3</i>	2509.1042 (Upper quantile 75%)	1.23	2.9x10 ⁻²	rs3740129_A	rs4148950_A	0.8	0.41	2.8x10 ⁻⁷

AEI: Allelic Expression Imbalance; DEG: Differentially Expression Gene; FDR: False Discovery Rate; NS: Non-significant.
Fold change direction related to the lesioned tissue.

Abbreviations

OA: Osteoarthritis

GWAS: Genome-wide association study

AEI: Allelic expression imbalance

LD: Linkage disequilibrium

COLGAT2: Collagen Beta(1-O)Galactosyltransferase 2

MGP: Matrix Gla Protein

SNPs: Single nucleotide polymorphisms

LincRNA: Long intergenic non-coding RNAs

MALAT1: Metastasis Associated Lung Adenocarcinoma Transcript 1

FDR: False Discovery Rate

ILF3-DT: Interleukin Enhancer Binding Factor 3 Divergent Transcript

HLA: Human Leukocyte Antigen

HLA-DPA1: Human Leukocyte Antigen-heterodimer consisting of an alpha

HLA-DPB1: Human Leukocyte Antigen-heterodimer consisting of an alpha

TNC: *Tenascin C*

FC: Fold Change

SLC44A2: Solute Carrier Family 44 Member 2

DGBI: Drug Gene Interaction Database

IG: Interaction group

CRISPR-cas9: Clustered Regularly Interspaced Short Palindromic Repeats
Associated Protein 9

WWP2: *WW Domain Containing E3 Ubiquitin Protein Ligase 2*

mQTL: methylation Quantitative Trait Loci

SERPINA2: Serpin Family A Member 2

ALDH1A2: *Aldehyde Dehydrogenase 1 Family Member A2*

RBP4: Retinol binding protein 4

MMP-3: matrix metalloproteinase-3

MMP-13: matrix metalloproteinase-13

NF-κB1: Nuclear Factor kappa B1

IL-6: Interleukin-6

miR-9: microRNA-9