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

Almeida, R.C. de; Tuerlings, M.; Ramos, Y.; Hollander, W. den; Suchiman, E.; Lakenberg, N.; ... ; Meulenbelt, I.

Rheumatology, 2022, **00**, 1–8
https://doi.org/10.1093/rheumatology/keac498
Advance access publication 30 August 2022
Original article

Basic science

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

Rodrigo Coutinho de Almeida [1], Margo Tuerlings [1], Yolande Ramos [1], Wouter Den Hollander[1], Eka Suchiman[1], Nico Lakenberg[1], Rob G. H. H. Nelissen[2], Hailiang Mei [3], 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, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

*Correspondence to: Ingrid Meulenbelt, 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. E-mail: i.meulenbelt@lumc.nl

Abstract

Objectives: To present an unbiased approach to identify positional transcript single nucleotide polymorphisms (SNPs) of osteoarthritis (OA) risk loci by allelic expression imbalance (AEI) analyses using RNA sequencing of articular cartilage and subchondral bone from OA patients.

Methods: RNA sequencing from 65 articular cartilage and 24 subchondral bone from OA patients was used for AEI analysis. AEI was determined for all genes present in the 100 regions reported by the genome-wide association studies (GWAS) catalog that were also expressed in cartilage or bone. 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 ($r^2 > 0.6$). Furthermore, a meta-analysis was performed to generate a meta-φ (null hypothesis median $\varphi = 0.49$) and *P*-value for each SNP.

Results: We identified 30 transcript SNPs (28 in cartilage and two in subchondral bone) subject to AEI in 29 genes. Notably, 10 transcript SNPs were located in genes not previously reported in the GWAS catalog, including two long intergenic non-coding RNAs (lincRNAs), *MALAT1* (meta-$\varphi = 0.54$, FDR $= 1.7\times10^{-4}$) and *ILF3-DT* (meta-$\varphi = 0.6$, FDR $= 1.75\times10^{-5}$). Moreover, 12 drugs were interacting with seven genes displaying AEI, of which seven drugs have been already approved.

Conclusions: By prioritizing proxy transcript SNPs that mark AEI in cartilage and/or subchondral bone at loci harbouring 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.

Keywords: allelic expression imbalance, long non-coding RNA, GWAS, osteoarthritis, cartilage, subchondral bone

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 for 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 a high rate, highlighting strong OA risk genes and associated underlying disease pathways. A major challenge for genome-wide association studies (GWAS) is to understand the functional consequences of associated single nucleotide polymorphisms (SNPs), more specifically, because most of the associated SNPs are in non-coding regions in the genome. By addressing the functionality of identified OA risk alleles, such as for *COLGALT2* [5] and *MGP* [6], it was confirmed that OA risk SNPs frequently modulated OA pathology due to subtly altered transcription of a positional gene *in cis*, as such imposing a persistent negative influence on joint tissue homeostasis throughout life [6].

Allelic expression imbalance (AEI) happens when an unequal expression of one of the two transcribed alleles in heterozygous individuals displays different levels. This phenomenon in the genome could be caused either by epigenetic inactivation of one allele, or by a genetic variation in regulatory regions (e.g. transcription factors). Targeted testing of AEI of OA risk genes is, however, cumbersome [7–9]. To address the full potential of biologically relevant DNA variants in cartilage, we previously captured AEI at a genome-wide scale, by 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 researchers to probe their gene of interest without providing the linkage disequilibrium (LD) structure relative to the OA risk SNPs. Moreover, GWAS

Received: 19 May 2022. **Accepted:** 18 August 2022

Rheumatology key messages

- Unbiased approach to identify functional OA risk SNP and gene by allelic expression imbalance (AEI) in cartilage and subchondral bone.
- AEI was found in long intergenic non-coding RNAs not previously identified by genome-wide association studies.
- Fourteen drugs showed significant enrichment interaction with OA risk genes displaying AEI.

in OA with ever-increasing sample sizes has resulted in an increasing number of OA risk SNPs with 100 currently mapped loci associated with 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 recent studies [5, 8, 9].

In the current study, we used an extended RNA-sequencing dataset of OA cartilage ($n = 65$) and of matching subchondral bone tissue ($n = 24$) 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.

Methods

Samples

Ethical approval for the Research Arthritis and Articular Cartilage (RAAK) study [11] was obtained from the medical ethics committee of the Leiden University Medical Center (P08.239/P19.013) and written 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 was 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 (San Francisco, California, USA)) was performed. Strand-specific RNA-seq libraries were generated which yielded a mean of 20 million reads per sample. Detailed descriptions for quality control (QC) alignment, mapping and normalization are described earlier for cartilage [12] and subchondral bone [13]. In addition, common population SNP sites from the Genome of the Netherlands [14] were masked during read alignment to prevent potential reference alignment bias. AEI events were assessed on SNPs called using SNVMix2 with default settings [15] 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 between preserved and lesioned (56 preserved and nine lesioned cartilage), all subchondral bone samples were paired (24 paired with preserved and lesioned) and also matched with their respectively cartilage sample. First, we calculated the LD of each 100 associated loci reported on the GWAS catalog database [16] (Supplementary Table S1, available at *Rheumatology* online), taking a window of 1 Mb 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 test using a threshold of <0.05 for significance. 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 [12, 13], and further intersected with all genes reported to be associated with OA in the GWAS catalog database. We assess the average normalized expression in cartilage and subchondral bone using the DESeq2 normalization approach [17] for all OA-associated genes.

Data integration with drug–gene interaction database (DGIdb 4.0)

The Drug Gene Interaction Database (DGIdb v4.0) [18] was used to determine potentially druggable targets for all genes that show allelic expression imbalance. Currently, the DGIdb database contains >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 only singular gene–drug interactions (i.e. one drug to target one specific gene) we set the DGIdb interaction group (IG) score to >0.50. The DGIdb IG score is based on: (i) the amount of evidence from different sources supporting the drug-gene (that is, the number of publications and different databases); (ii) the number of genes set on each search that interact with the given drug; and (iii) the degree to which the drug has known interactions with another gene [18]. 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 the OA context we removed drugs showing a mode of action inconsistent with the AEI direction.

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

Table 1. Samples characteristics

Cartilage	Total ($n = 65$)	RNAseq—hip ($n = 26$)	RNAseq—knee ($n = 39$)
Age (s.d.)	64.5 (16.7)	64.2 (16)	64,6 (17.4)
Females	54	20	34
Bone	**Total ($n = 24$)**	**RNAseq—hip ($n = 6$)**	**RNAseq—knee ($n = 18$)**
Age (s.d.)	68.1 (9.5)	67.8 (8.8)	65.7 (8.5)
Females	22	6	16

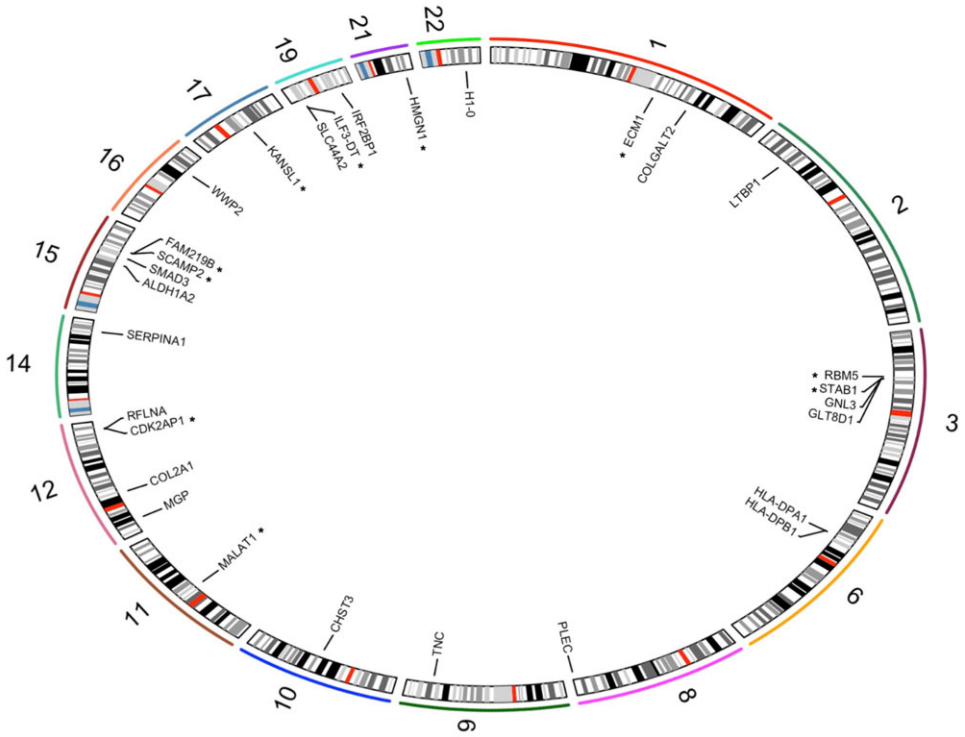

Figure 1. Circos plot showing all genes with allelic expression imbalance in OA across the genome. Stars represent genes that were not previously reported by the GWAS catalog

RNA-seq datasets [12, 13] for genes located at all 100 OA risk loci reported by the GWAS catalog [7, 16] residing within a 1 Mb window. To obtain relevant information on the functional effect of the identified OA risk alleles on the expression of positional genes, we next prioritized 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 (Fig. 1; Supplementary Table S2, available at *Rheumatology* online). Notable among cartilage, we identified 10 AEI SNPs located in genes not previously highlighted as possible OA genes candidates according to the GWAS catalog (Table 2), including two long intergenic non-coding RNAs (lincRNAs) such as *MALAT1* (AEI meta-$\varphi = 0.54$, FDR $= 1.7 \times 10^{-4}$) and *ILF3-DT* (meta-$\varphi = 0.6$, FDR $= 1.7 \times 10^{-5}$) (Fig. 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-$\varphi = 0.76$, FDR $= 0.0033$; Fig. 2C) and MGP (meta-$\varphi = 0.33$, FDR $= 1.1 \times 10^{-19}$) (Supplementary Table S2, available at *Rheumatology* online).

AEI genes differentially expressed between preserved and lesioned 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 [12, 13]. We found nine AEI genes differentially expressed in cartilage, of which two genes (*MGP* and *SLC44A2*) were also differentially expressed in subchondral bone (Table 3). Of note the *TNC* gene (meta-$\varphi = 0.47$, FDR $= 2.08 \times 10^{-7}$, FC $= 1.4$) 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 (Fig. 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-$\varphi = 0.54$, FDR $= 0.00132$, FC $= 1.4$; Fig. 3B),

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	3.5×10^{-3}	Cartilage
CDK2AP1	PC	rs1060105	SBNO1	rs6633	0.94	chr12	123261262	T	0.12	0.44	2.2×10^{-10}	Cartilage
MALAT1	lincRNA	rs10896015	LTBP3	rs3200401	0.65	chr11	65504361	T	0.14	0.46	3.3×10^{-4}	Cartilage
SCAMP2	PC	rs35206230	CSK—LMAN1L	rs12487	0.72	chr15	74844353	C	0.18	0.45	0.03	Cartilage
FAM219B	PC	rs35206230	CSK-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	4.3×10^{-4}	Cartilage
RBM5	PC	rs62262139	RBM6	rs1138536	0.66	chr3	50115923	C	0.25	0.48	7.5×10^{-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	1.2×10^{-4}	Cartilage
HMGN1	PC	rs9981884	BRWD1	rs3167757	0.92	chr21	39342552	T	0.47	0.41	1.1×10^{-5}	Cartilage

AEI: allelic expression imbalance; ALT: alternative allele; FDR: false discovery rate; MAF: minor allele frequency; PC: protein-coding; chr: chromosome.

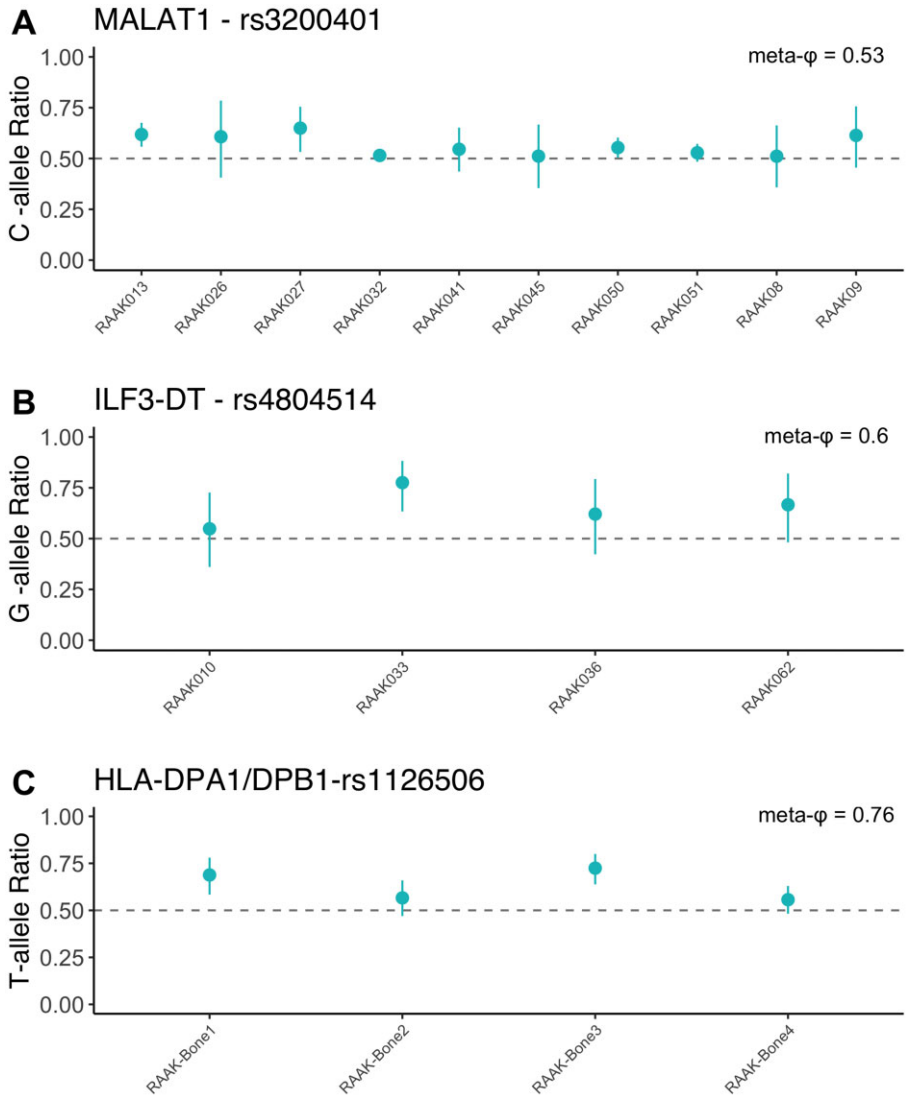

Figure 2. Allelic expression imbalance of two lincRNAs in articular cartilage and HLA genes in subchondral bone. (**A**) SNP rs3200401 showed increased expression of lincRNA MALAT1 of the risk allele C in OA cartilage (meta-φ=0.54). (**B**) LincRNA ILF3-DT with risk allele G displayed increased expression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

suggesting that a lower expression in lesioned cartilage could be potentially beneficial. Additionally, the *SLC44A2* gene was differentially expressed in subchondral bone and cartilage (Table 3), however with AEI was exclusively in cartilage (meta-$\varphi = 0.65$, FDR $= 1.13 \times 10^{-11}$, Fig. 3C), indicating that lower expression of *SLC44A2* in cartilage is conferring risk to OA likely by an aberrant cartilage-specific transcription factor.

Table 3. AEI genes differentially expressed in cartilage and bone

Cartilage AEI & DEG	Average expression (normalized read counts)	Fold change	FDR DEG	OA risk allele	AEI allele	r^2	Meta -φ	FDR AEI
COLGALT2	1919 (Upper quantile-75%)	1.42	1.92×10^{-5}	rs11583641_C	rs12023991_T	1	0.54	1.3×10^{-3}
ILF3-DT	142 (Upper quantile-75%)	0.69	3.41×10^{-4}	rs1560707_T	rs4804514_T	0.95	0.4	4.3×10^{-4}
MGP	38663 (Upper quantile-75%)	1.43	0.02	rs4764133_T	rs1800801_T	0.95	0.37	1.1×10^{-6}
RFLNA	91 (Down quantile-50%)	0.48	6.48×10^{-5}	rs4765540_C	rs12551_C	0.74	0.48	0.006
SLC44A2	1014 (Upper 75%)	0.80	0.028	rs1560707_T	rs1053007_G	0.95	0.34	1.6×10^{-9}
SMAD3	493 (Upper quantile-75%)	0.83	0.028	rs12901071_A	rs1061427_A	0.60	0.45	0.01
TNC	21350 (Upper quantile-75%)	1.41	0.010	rs2480930_A	rs2274836_T	1	0.47	2.1×10^{-7}
WWP2	5382 (Upper quantile 75%)	0.79	0.034	rs34195470_A	rs2270841_C	0.65	0.54	4.6×10^{-3}
CHST3	2509 (Upper quantile 75%)	1.23	2.9×10^{-2}	rs3740129_A	rs4148950_A	0.8	0.41	2.8×10^{-7}
Bone AEI & DEG	**Average expression (normalized read counts)**	**Fold change**	**FDR DEG**	**OA risk allele**	**AEI allele**	**r^2**	**Meta -φ**	**FDR AEI**
MGP	3826 (Upper quantile-75%)	1.53	NS	rs4764133_T	rs1800801_T	0.95	0.33	1.1×10^{-19}
SLC44A2	1383 (Upper quantile-75%)	0.84	0.019	rs1560707_T	rs1053007_G	0.95	NS	NS

Fold change direction related to the lesioned tissue. AEI: allelic expression imbalance; DEG: differentially expression gene; FDR: false discovery rate; NS: non-significant.

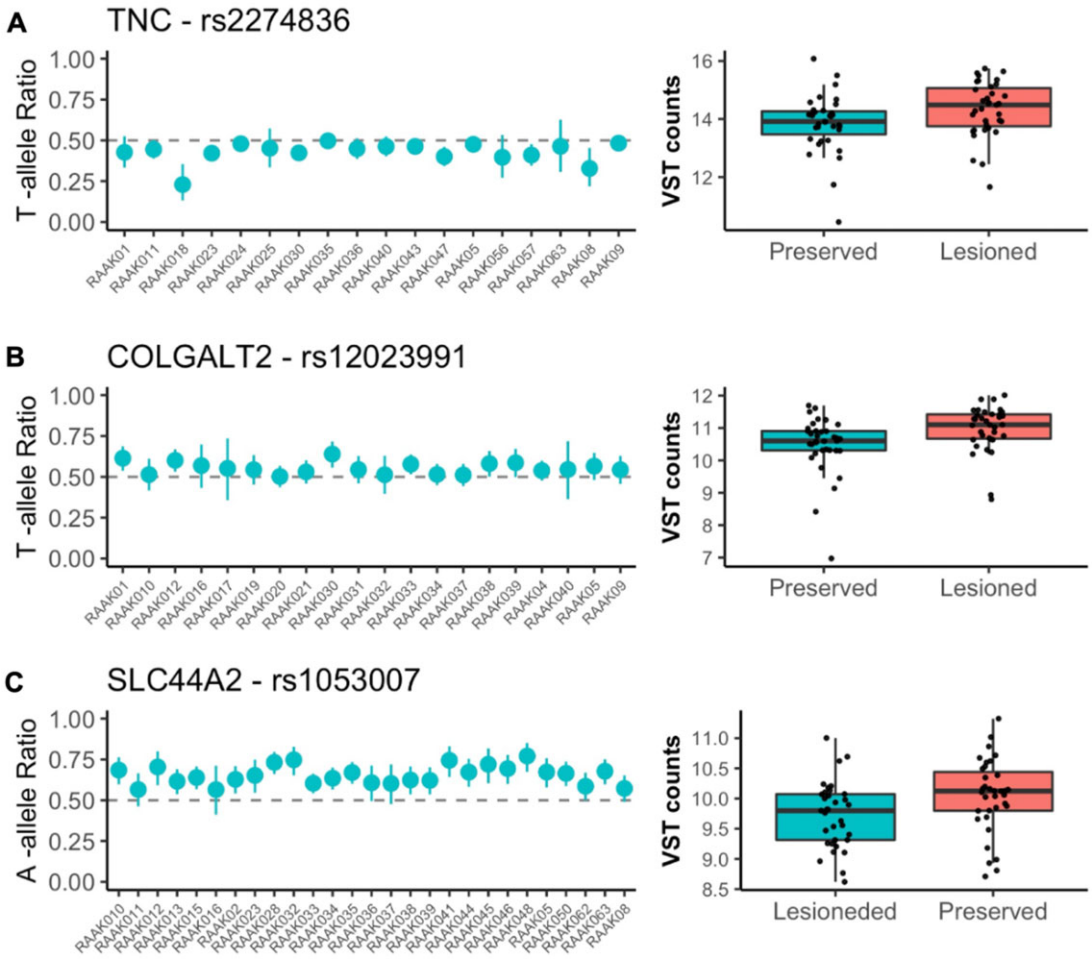

Figure 3. Genes with allelic expression imbalance (AEI) also showing differentially expressed between preserved and lesioned cartilage. (**A**) TNC upregulated in lesioned cartilage and with AEI downregulated effect (meta-φ=0.47); (**B**) *COLGALT2* (meta-φ=0.54); (**C**) *SLC44A2* (meta-φ=0.65)

AEI genes interacting with 12 drugs

By assessing the effects of identified OA risk alleles on the expression of positional genes, a reliable inference can be made on potential druggable targets that could counteract this aberrant gene expression. Therefore, we next combine OA risk alleles subject to AEI with known drugs and their targets on all 29 AEI genes. Upon setting the DGBI IG score >0.5, we identified 12 drugs with a singular interaction with an AEI gene that had a mode of action counteracting the direction of the risk allele expression (Fig. 4; Supplementary Table S3,

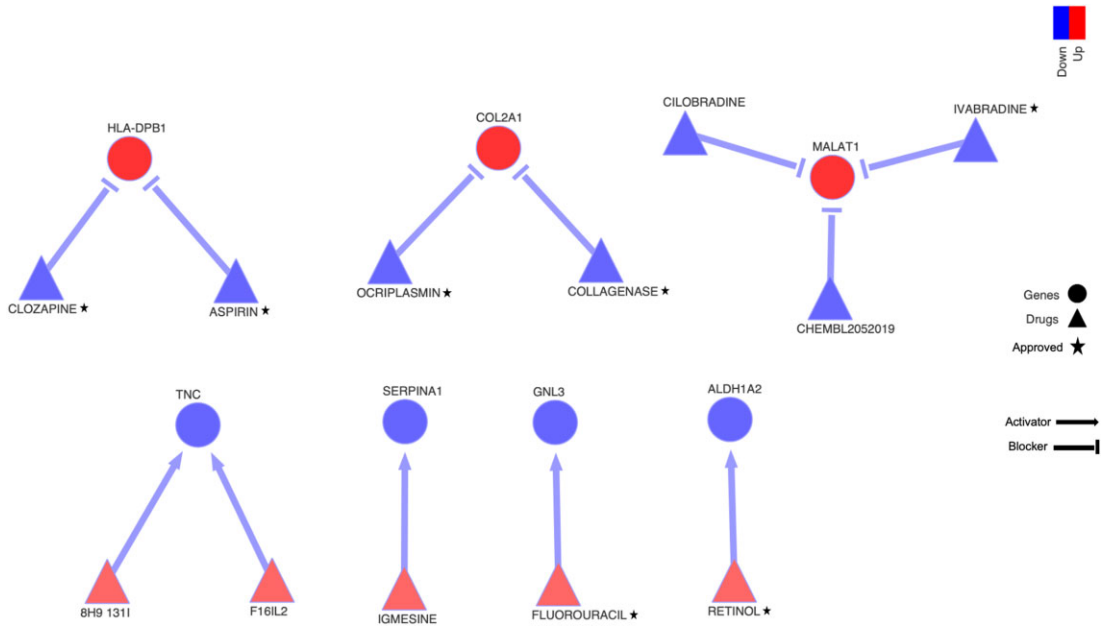

Figure 4. Seven AEI genes interacting with 12 drugs. Genes are represented by circles and drugs are triangles. The stars shows drug-approved federal agencies according to the database. Blue are genes with AEI downregulated effect and for drugs represents the blockers interaction. Red are genes with AEI upregulated effects and drugs with an activator interaction

available at *Rheumatology* online). Moreover, out of 15 000 drug–gene interactions present in the DGBI database, these 12 AEI drug–gene interactions showed a 2.1-fold significant enrichment (Fisher's exact test P-value = 0.044) suggesting that our AEI OA risk genes are part of more general pathophysiological processes for which drugs have been developed. In addition, as shown in Fig. 4, seven drugs have been already approved by federal agencies.

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 a drug–gene interaction database to identify potential drugs that could counteract the effects of 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 the subchondral bone. We advocate that these genes should be prioritized for further translation towards underlying biological mechanisms such as 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 with 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 a 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 [19]. *MALAT1* regulates inflammation, cellular responses to oxidative stress and cellular senescence [20]. Noteworthy, *MALAT1* is decreased in senescent cells [21] and in bleomycin-induced murine fibrosis where myeloid deletion of MALAT1 increases susceptibility to fibrosis and the number of profibrotic M2 macrophages [22]. MALAT1 was identified as sponges for miR-150-5p and it was found up-regulated in OA chondrocytes compared with normal chondrocytes [23]. Nevertheless, MALAT1 was not significantly differentially expressed between lesioned and preserved cartilage [24] or in subchondral bone [25] in our dataset. We found an allelic imbalance of the allele C (in LD with the risk allele G) with an upregulated effect (Fig. 2B). It is tempting to speculate that the *MALAT1* risk allele could be partially regulating inflammation or even senescence pathways in OA. Being highly tissue-specific and condition-specific expression patterns lincRNAs have been proposed as a new tool for a particular and personalized therapeutic approach [26]. 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 used 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 a 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. Even though 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 an 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 Tenascin that is abundantly expressed (75th percentile) in the extracellular matrix [27]. TNC is a key molecule in cartilage remodelling and participates in chondrogenesis and cartilage development [28]. *TNC* may be involved in mechanotransduction in response to mechanical stress [27]. The here-reported AEI indicated that the OA risk allele T (rs2274836 SNP) is associated with downregulation of *TNC* in the cartilage of OA patients, but *TNC* was upregulated in lesioned as compared with preserved OA cartilage. The latter suggests 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* cell models [29] showing that *TNC* expression levels were still normal in the early OA phase, but its expression disappeared in 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 [30], suggesting that *TNC* deficiency enhances cartilage degeneration. Together these data indicate that aberrant *TNC* signalling may confer an important risk to OA, and as such should be prioritized for functional follow-up research towards underlying OA disease mechanisms.

COLGALT2 was among the nine AEI genes that additionally showed differential expression between preserved and lesioned cartilage (meta-$\varphi = 0.54$ and FC = 1.42, Fig. 3B; Table 2). *COLGALT2* encodes procollagen galactosyltransferase 2, an enzyme that post-translationally glycosylates collagen. Notable is that this gene has also previously been identified as a methylation quantitative trait loci (mQTL) [31], indicating that expression changes in *COLGALT2* are mediated by differential enhancer methylation [31]. Based on these findings, the authors further suggested that the upregulated expression of *COLGALT2* in OA cartilage is reflecting a reparative response. Recently it was shown, however, by application of molecular biology tools including CRISPR-Cas9 [5], that the OA risk allele of SNP rs11583641 is mediating lower levels of DNA methylation in cartilage at the *COLGALT2* enhancer, which ultimately leads to increased expression, hence demonstrating that upregulation of *COLGALT2* is associated to OA risk [5]. Our data support these latter findings, as we showed that the AEI rs12023991-T is associated with higher *COLGALT2* expression (meta-$\varphi = 0.54$, FDR $= 1.3 \times 10^{-3}$) and is in complete LD with the OA risk allele rs11583641-C. Moreover, given that *COLGALT2* is upregulated in lesioned relative to preserved OA cartilage (FC = 1.42, FDR $= 1.92 \times 10^{-5}$), we advocate that the upregulation of *COLGALT2* with OA pathophysiology reflects an unbeneficial response.

The prioritization of druggable targets based on genetic support has been proven to highly increase the 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 12 drugs interacting with at least one AEI gene (Fig. 4). Of note, the SERPINA2 gene interacts with the Igmesine drug, which could potentially reverse the allelic downregulation effect by enhancing SERPINA2 expression. Nevertheless, this drug has not yet been approved by federal agencies to be used; therefore, more studies to confirm this candidate should be performed. On the other hand, we found seven drugs that have been already approved to be used, including retinol, which interacts with the *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 [32], 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 [33], 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, suppresses chondrocyte apoptosis, and reduces extracellular matrix degradation [34, 35]. Recently, the use of revesratrol to suppress the activation of MALAT1 showed a decrease in pro-inflammatory genes (NF-κB1 and IL-6) by modulating a microRNA (miR-9) [36]. Considering that ivabradine is a blocker modulator, we can speculate that this drug could act similarly to reverstrol with MATLAT1 to reverse the effects of OA response. Nevertheless, more *in-vitro* and *in-vivo* studies will be essential to clarify this hypothesis.

Taken together, applying AEI in disease-relevant tissues showed that we can improve the selection of possible causal genes associated with 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 an approved drug. Thus, by pinpointing the risk allele that influences gene expression in disease-relevant tissues, we can pave the way for further post-GWAS *in-vitro* models in OA.

Supplementary data

Supplementary data are available at *Rheumatology* online.

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 underlying this article are available in the article and in its online supplementary material.

Funding

The research leading to these results has received funding from: TreatOA which is funded by the European Commission framework 7 program grant 200800; BBMRI Metabolomics Consortium is funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO, grant no. 184.021.007 and 184033111); Dutch Scientific Research council NWO/ZonMW VICI scheme (91816631/528).

Disclosure statement: All authors have declared no competing interests.

Acknowledgements

We thank all study participants of the RAAK study. The Leiden University Medical Centre has and is 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.

References

1. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage 2013;21:1145–53.
2. Mandl LA. Osteoarthritis year in review 2018: clinical. Osteoarthritis Cartilage 2019;27:359–64.
3. Woolf AD, Erwin J, March L. The need to address the burden of musculoskeletal conditions. Best Pract Res Clin Rheumatol 2012; 26:183–224.
4. MacGregor AJ, Spector TD. Twins and the genetic architecture of osteoarthritis. Rheumatology 1999;38:583–8.
5. Kehayova YS, Watson E, Wilkinson JM, Loughlin J, Rice SJ. Genetic and epigenetic interplay regulates *COLGALT2*, contributing to osteoarthritis genetic risk. Arthritis Rheumatol 2021;73:1856–65.
6. Loughlin J. Genetic contribution to osteoarthritis development. Current Opin Rheumatol 2015;27:284–8.
7. Boer CG, Hatzikotoulas K, Southam L *et al.* Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell 2021;184:4784–4818.e17.
8. Shepherd C, Reese AE, Reynard LN, Loughlin J. Expression analysis of the osteoarthritis genetic susceptibility mapping to the matrix Gla protein gene MGP. Arthritis Res Ther 2019;21:149.
9. Shepherd C, Zhu D, Skelton AJ *et al.* Functional characterization of the osteoarthritis genetic risk residing at *ALDH1A2* identifies rs12915901 as a key target variant. Arthritis Rheumatol 2018;70: 1577–87.
10. Hollander W, Pulyakhina I, Boer C *et al.* Annotating transcriptional effects of genetic variants in disease-relevant tissue: transcriptome-wide allelic imbalance in osteoarthritic cartilage. Arthritis Rheumatol 2019;71:561–70.
11. Ramos YFM, den Hollander W, Bovée JVMG *et al.* Genes involved in the osteoarthritis process identified through genome wide expression analysis in articular cartilage; the RAAK study. PLoS ONE 2014;9:e103056.
12. Coutinho de Almeida R, Ramos YFM, Mahfouz A *et al.* RNA sequencing data integration reveals an miRNA interactome of osteoarthritis cartilage. Ann Rheum Dis 2019;78:270–7.
13. Tuerlings M, Hoolwerff M, Houtman E *et al.* 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 Rheumatol 2021;73:789–99.
14. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet 2014;46:818–25.
15. Goya R, Sun MGF, Morin RD *et al.* SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. Bioinformatics 2010;26:730–6.
16. Buniello A, MacArthur JAL, Cerezo M *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019; 47:D1005–D1012.
17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
18. Freshour SL, Kiwala S, Cotto KC *et al.* Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res 2021;49:D1144–51.
19. Fei J, Jadaliha M, Harmon TS *et al.* Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. J Cell Sci 2017;130:4180–92.
20. Puvvula PK. LncRNAs regulatory networks in cellular senescence. Int J Mol Sci 2019;20:2615.
21. Grammatikakis I, Panda AC, Abdelmohsen K, Gorospe M. Long noncoding RNAs (lncRNAs) and the molecular hallmarks of aging. Aging 2014;6:992–1009.
22. Omote N, Sauler M. Non-coding RNAs as regulators of cellular senescence in idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease. Front Med 2020;7:603047.
23. Zhang Y, Wang F, Chen G, He R, Yang L. LncRNA MALAT1 promotes osteoarthritis by modulating miR-150-5p/AKT3 axis. Cell Bioscience 2019;9:54.
24. Hoolwerff M, Metselaar PI, Tuerlings M *et al.* Elucidating epigenetic regulation by identifying functional *cis* -acting long noncoding RNAs and their targets in osteoarthritic articular cartilage. Arthritis Rheumatol 2020;72:1845–54.
25. Tuerlings M, van Hoolwerff M, van Bokkum JM *et al.* Long noncoding RNA expression profiling of subchondral bone reveals *AC005165.1* modifying *FRZB* expression during osteoarthritis. Rheumatology 2022;61:3023–32.
26. Arun G, Diermeier SD, Spector DL. Therapeutic targeting of long non-coding RNAs in cancer. Trends Mol Med 2018;24:257–77.
27. Imanaka-Yoshida K, Aoki H. Tenascin-C and mechanotransduction in the development and diseases of cardiovascular system. Front Physiol 2014;5:283.
28. Hasegawa M, Yoshida T, Sudo A. Tenascin-C in osteoarthritis and rheumatoid arthritis. Front Immunol 2020;11:577015.
29. Nakoshi Y, Hasegawa M, Akeda K *et al.* Distribution and role of tenascin-C in human osteoarthritic cartilage. J Orthop Sci 2010;15: 666–73.
30. Ikemura S, Hasegawa M, Iino T *et al.* Effect of tenascin-C on the repair of full-thickness osteochondral defects of articular cartilage in rabbits. J Orthop Res 2015;33:563–71.
31. Rice SJ, Cheung K, Reynard LN, Loughlin J. Discovery and analysis of methylation quantitative trait loci (mQTLs) mapping to novel osteoarthritis genetic risk signals. Osteoarthritis Cartilage 2019;27: 1545–56.
32. Scotece M, Koskinen-Kolasa A, Pemmari A *et al.* Novel adipokine associated with OA: retinol binding protein 4 (RBP4) is produced by cartilage and is correlated with MMPs in osteoarthritis patients. Inflamm Res 2020;69:415–21.
33. Xiang X, Zhou Y, Sun H *et al.* Ivabradine abrogates TNF-α-induced degradation of articular cartilage matrix. Int Immunopharmacol 2019;66:347–53.
34. Pan C, Huang W, Chen Q *et al.* LncRNA Malat-1 from MSCs-derived extracellular vesicles suppresses inflammation and cartilage degradation in osteoarthritis. Front Bioeng Biotechnol 2021;9:772002.
35. Kong X, Wang J, Cao Y *et al.* The long noncoding RNA MALAT-1 functions as a competing endogenous RNA to regulate MSL2 expression by sponging miR-338-3p in myasthenia gravis. J Cell Biochem 2019;120:5542–50.
36. Zhang G, Zhang H, You W *et al.* Therapeutic effect of Resveratrol in the treatment of osteoarthritis via the MALAT1/miR-9/NF-κB signaling pathway. Exp Ther Med 2020;19:2343–52.