# On the relation between genetic variation and osteoarthritis

Hollander, W. den

# 2

# EXPLORATION OF THE ARTICULAR CARTILAGE TRANSCRIPTOME REVEALS *CRLF1* AS NOVEL OSTEOARTHRITIS SUSCEPTIBILITY GENE

Wouter den Hollander[1*], Cindy Boer[2*], Irina Pulyakhina[3,4], Nils Bomer[1], Ruud van der Breggen[1], Wibowo Arindrarto[5], Nico Lakenberg[1], Joris Deelen[1], Marian Beekman[1], Jeroen F.J. Laros[5], Peter A.C. 't Hoen[3], Margreet Kloppenburg[6], Eline P.E. Slagboom[1], Rob G.H.H. Nelissen[7], Yolande F.M. Ramos[1] Joyce van Meurs[2‡], Ingrid Meulenbelt[1‡].

*[1]Dept. of Molecular Epidemiology, LUMC, Leiden, The Netherlands. [2]Dept. of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. [3]Dept. of Human Genetics, LUMC, Leiden, The Netherlands. [4]Wellcome Trust Centre for Human Genetics Oxford, Oxfordshire, United Kingdom. [5]Medical Statistics and Bioinformatics, LUMC, Leiden, The Netherlands. [6]Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands. [7]Dept. of Orthopaedics, LUMC, Leiden, The Netherlands.*

* *These authors contributed equally to this work*
‡ *These authors jointly supervised this work*

## ABSTRACT

Multiple osteoarthritis (OA) susceptibility single nucleotide polymorphisms (SNPs) mark imbalanced expression of positional genes in articular cartilage, reflected by unequally expressed alleles among heterozygotes (allelic imbalance, AI). Here, we explored the articular cartilage transcriptome from OA patients for AI events to identify putative disease driving genetic variation by RNA sequencing. We observed 2070 SNPs that consistently marked AI of 1031 unique genes in articular cartilage, of which 32 were additionally significantly differentially expressed (0.5>FC>2, FDR<0.05) between preserved and paired lesioned cartilage. Among those was *CRLF1* (FC=3.17, FDR=$7.86*10^{-5}$), which appeared subject to AI, marked by lower expression of the rs7256319 alternative allele T compared to the reference allele C. Additionally, the T allele harbored a protective signal in a combined genetic association meta-analysis (OR=0.881, $CI_{95}$=0.781-0.994, P=0.0393). Finally, we show that increased *CRLF1* signaling markedly affects expression of anabolic genes *in-vitro*, while catabolic markers remain unaffected.

## INTRODUCTION

Due to the increased proportion of elderly in the human population, osteoarthritis (OA) has become one of the major musculoskeletal diseases (1). While all joint tissues have been implicated in OA pathology, the disease is characterized primarily by progressive degradation and calcification of articular cartilage (2). Both gene targeted (3-5), as well as genome-wide research (6-9) showed that a multitude of genes are involved in the currently irreversible destruction of articular cartilage that precedes total joint replacement surgery; presently the only effective treatment for end-stage OA. In this regard, numerous studies have reported on altered regulation of gene expression that reflects, attenuates and/or stimulates OA-mediated cartilage degradation (10-13). Moreover, multiple OA risk alleles of single nucleotide polymorphisms (SNPs) were shown to consistently modulate OA pathology due to altered transcription of the respective genes in articular cartilage (14-18). Hence, it is clear that *in-cis* genetic regulation of transcription plays a substantial role in cartilage homeostasis and, therefore, in OA pathophysiology.

SNPs that confer risk for OA frequently act via allele-specific gene regulation as reflected by unequally expressed alleles among heterozygous carriers, commonly referred to as an allelic imbalance (AI) (15, 19-22). Notable examples are *DIO2* (19) and *GDF5* (15), genes of which the OA-associated risk allele affects transcription in articular cartilage. Additionally, AI among OA risk SNPs that are not situated within gene bodies have been addressed by measuring AI among SNPs in strong linkage disequilibrium (LD), as shown for *ALDH1A2* (22). Despite the evidence for *in-cis* genetic regulation of transcription in OA susceptibility, genome-wide association studies (GWAS) have thus far failed to explain the larger part of the hereditary component of OA (23). In this regard, a large number of the tested SNPs in GWAS likely bear no biological function in relation to the addressed phenotype or disease relevant tissues (24), resulting in massive inflation of, possibly biologically irrelevant, statistical tests and thus the multiple testing correction penalty. Consequently, large numbers of SNPs that do bear biological functionality in the context of OA are missed. Furthermore, SNPs that reside within LD blocks are hard to interpret, as association analysis is inherently unable to distinguish disease-relevant alleles from mere statistically associated alleles.

In previous studies, we and others have used targeted approaches to address AI events of putative, as well as established OA susceptibility genes (19-22, 25, 26). Given the successful identification of the transcriptional consequences of multiple OA-associated SNPs, we have here aimed to characterize, on a transcriptome-wide scale, novel SNPs that tag AI of genes expressed in articular cartilage. By means of RNA sequencing we addressed AI events and combined them with differential expression between preserved and lesioned articular cartilage, from the RAAK study (12), of the respective genes. Finally, we applied association analyses and functional experiments to reveal a novel OA susceptibility gene.

## RESULTS

**TRANSCRIPTOME-WIDE DISCOVERY OF ARTICULAR CARTILAGE ALLELIC IMBALANCE EVENTS.**

To understand how genetic variation contributes *in-cis* to transcriptional regulation in articular cartilage on a transcriptome-wide scale, we first called heterozygous SNPs (dbSNP144) using RNA sequencing data from articular cartilage derived from patients who underwent total joint replacement surgery of either the hip (N=22) or knee (N=25) joint due to primary OA (**Supplementary Table S1**). After filtering by the number of read counts per position ($\boldsymbol{R}_{reference} \geq 10$, $\boldsymbol{R}_{alternative} \geq 10$ and $\boldsymbol{R}_{total} \geq 25$), selecting for heterozygous SNPs present in at least 2 individuals, removing SNPs present in multiple distinct transcripts and discarding the HLA locus, we defined $\boldsymbol{\varphi}$ for 13853 SNPs as the measure of imbalance (**Figure 1A**), which denotes the fraction of $\boldsymbol{R}_{alternative}$ among $\boldsymbol{R}_{total}$. Possibly due to reference bias, a considerable number of SNPs marked AI by $\boldsymbol{\varphi} < 0.1$ or $\boldsymbol{\varphi} > 0.9$ (N=418) and were subsequently removed prior to further analyses. A meta-analysis per SNP (null hypothesis: median $\boldsymbol{\varphi}=0.49$) and subsequent multiple testing correction (FDR) revealed 2070 SNPs that significantly marked AI among respectively 1031 genes (**Figure 1B**, **Supplementary Table S2**). We assessed five SNPs that have been reported to mark AI in articular cartilage (**Supplementary Figure S1**). As such, we were able to replicate AI of *ALDH1A2*, marked by the alleles of rs3204689 ($\boldsymbol{\varphi}=0.42$, FDR=0.003) (22), and, additionally, report on multiple other SNPs at the *ALDH1A2* locus that do so. These SNPs were, expectedly, in high LD and thus suggest that the observed SNP marking AI might not be the actual regulatory SNP. In that regard, we were unable to detect significant AI reports for rs225014 ($\boldsymbol{\varphi}=0.49$, FDR=0.98, *DIO2*) (19), rs1676486 ($\boldsymbol{\varphi}=0.51$, FDR=0.49, *COL11A1*) (21), rs11177 ($\boldsymbol{\varphi}=0.47$, FDR=0.59, *GNL3*) (20) or rs6617 ($\boldsymbol{\varphi}=0.51$, FDR=0.63, *SPCS1*) (20). This observation may be due to lack of power in our dataset (insufficient heterozygous individuals), insufficient positional read depth, or may imply that the actual regulatory SNPs for these genes are in incomplete LD with the SNPs that mark AI reported by us and others. Finally, it could be that targeted assays are simply more sensitive.

**DIFFERENTIAL GENE EXPRESSION BETWEEN PRESERVED AND PAIRED OA LESIONED CARTILAGE.**

While SNPs marking AI in articular cartilage could contribute to OA pathophysiology in various ways, e.g. in cartilage development or homeostasis, it can be expected that those located in genes which additionally mark the articular cartilage's disease state are more likely to contribute to or attenuate disease progression. Therefore, we went back to the original expression data and determined differential expression, using the edgeR R package, in patients for which paired RNA sequencing data of both preserved and OA-lesioned articular cartilage was generated (6 hip joints, 15 knee joints). Of the 10468 Ensembl gene identifiers with at least 5 counts per million, 118 and 48 were observed to be respectively significantly (FDR<0.05) down- (FC<0.5) and upregulated (FC>2) in lesioned compared to

preserved cartilage (**Figure 2A**, **Supplementary Table S3**). As has been reported by microarray studies that have utilized a similar design, Gene Ontology (GO) term enrichment analysis (**Supplementary Table S4**) revealed significant enrichment for inflammatory (e.g. *SCUBE1*, *CFH* and *CXCL14,* **Figure 2B-D**), response to wound healing (e.g. *NOTCH3*, *BMP5* and *SERPINE1*, **Figure 2E-G**) and joint development associated pathways (e.g. *SPP1*, *MMP3* and *COL9A1*, **Figure 2H-J**). Of the 166 differentially expressed genes, 32 were additionally subject to AI, marked by 74 SNPs (**Supplementary Table S5**).

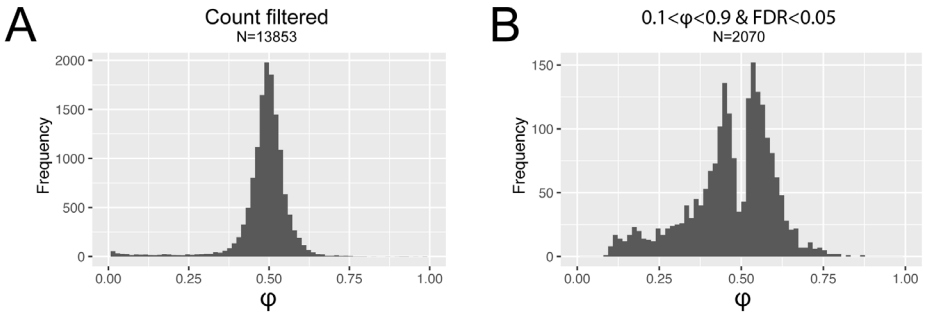



**Figure 1.** Distribution of AI events in articular cartilage. (**A**) All 13435 called variants after selecting for at least 2 heterozygotes, selecting SNPs present in only single genes and removing low counts. (**B**) After filtering by allelic fraction (0.1<ϕ<0.9) and FDR<0.05 for all heterozygotes in the same direction per SNP, 352 SNPs remained that marked consistent AI of 219 unique genes.

**GENETIC ASSOCIATION ANALYSIS OF AI SNPS IN DISEASE MODULATED GENES.**

As a proof of principle, we performed genetic association analysis among participants of the Genetics osteoARthritis and Progression study (GARP, N=380) and healthy controls from the Leiden Longevity Study (LLS, N=2315) with the top 10 AI SNPs (single most significant SNP per gene) among differentially expressed genes (**Supplementary Table S6**). After adjustment for age, sex and BMI, we observed a significant protective association for the alternative allele T of rs7256319 (C>T, MAF=0.39, OR=0.803, $CI_{95}$=0.681-0.946, P=0.009), located in the cytokine receptor-like factor 1 gene (*CRLF1*) encoding a member of the cytokine type I receptor family (CLF1). The CLF1 protein is reported to signal via the ciliary neurotrophic factor receptor (CNTFR) gene (*CNTFR*) after heterodimerization with cardiotrophin-like cytokine factor 1 (CLCF1) gene (*CLCF1*) (27). We were able to replicate the protective signal by meta-analysis, having additionally included three independent cohorts from the Rotterdam Study (RS1: 188 cases, 1903 controls, RS2: 57 cases, 742 controls RS3: 36 cases, 1185 controls), among which cases were defined as having underwent a total joint replacement surgery. This analyses confirmed the direction of effect and provided a significant association in the combined meta-analysis (OR=0.881, $CI_{95}$=0.781-0.994, P=0.0393, I (2)= 0%, $P_{heterogeneity}$=0.417, **Supplemental Figure S2**).

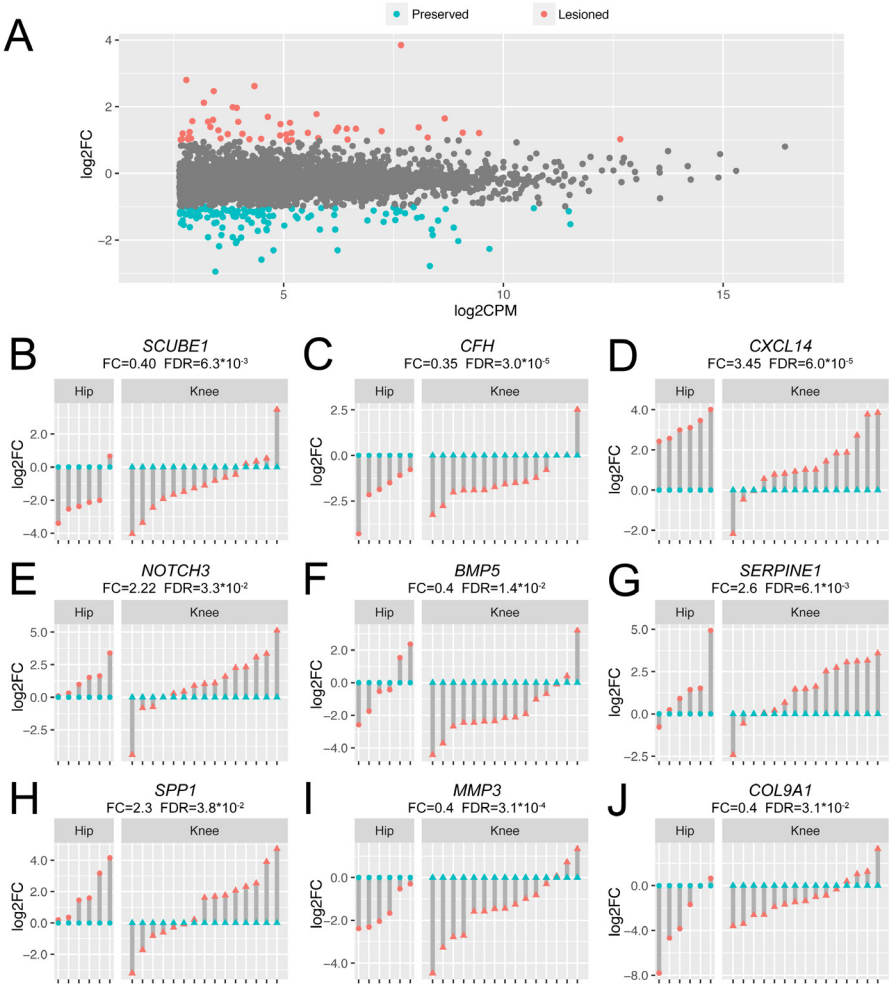



**Figure 2.** Paired differential gene expression between preserved and lesioned articular cartilage. (**A**) Volcano plot, showing FDR cutoff (0.05) and foldchange cutoff (0.5>FC>2.0) indicated in red. (**B-D**) Differential expression of selected genes from inflammatory pathways. (**E-G**) Differential expression of selected genes from developmental pathways. (**H-J**) Differential expression of selected genes from developmental related pathways. Hip and knee joints are depicted by triangles and circles, respectively. Preserved and lesioned samples are depicted by blue and red, respectively.

### *CRLF1* SIGNALING PLAYS AN ACTIVE ROLE IN END-STAGE OA ASSOCIATED CARTILAGE DEGRADATION.

The protective T allele marked consistent lower *CRLF1* expression (***φ***=0.29, FDR=4.02*$10^{-21}$, **Figure 3A, Supplementary Table 2**) compared to the reference allele C, which was confirmed by custom TaqMan assay performed in respectively 5 preserved and 5 lesioned articular cartilage samples, originating from 6 independent patients that underwent total knee replacement surgery (**Figure 3B**). In parallel, and similar to distinct publically available microarray datasets, expression of the three genes (*CRLF1, CNTFR* and

*CLCF1*) in the current dataset was significantly different between preserved and OA-lesioned cartilage (**Supplementary Table S3**) with a significant upregulation in OA affected cartilage for *CRLF1* (FC=3.17, FDR=7.86*$10^{-5}$, **Figure 3C**) as well as for *CLCF1* (FC=1.56, FDR=0.01, **Figure 3D**), while the receptor for the complex (CNTFR) revealed significant downregulation (FC=0.31, FDR=3.43*$10^{-6}$, **Figure 3E**). Together, these data demonstrated that the protective effect of the alternative allele T of rs7256319 likely acts via mitigation of *CRLF1/CLCF1* signaling with ongoing OA pathophysiology.

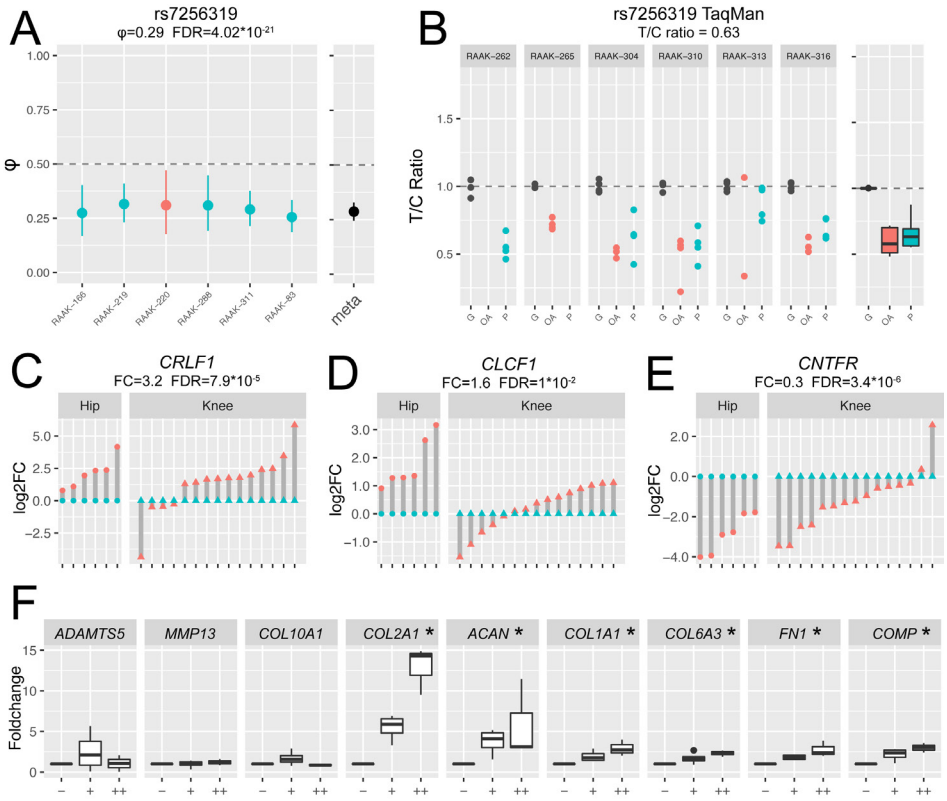



**Figure 3.** *CRLF1* signaling is modulated in articular cartilage during OA progression. (**A**) *CRLF1* expression is subject to AI, marked by lower expression of the alternative allele T of rs7256319 (ϕ=0.29, FDR=1.26*$10^{-14}$). (**B**) Replication of the observed AI tagged by rs7256319 by TaqMan genotyping in 6 additional knee samples ratified the observed lower expression of the alternative allele T. Genomic DNA (grey) was used as reference ratio. (**C**) *CRLF1* expression was significantly upregulated in OA lesioned articular cartilage compared with paired preserved articular cartilage (FC=3.17, $P_{FDR}$=4.06*$10^{-4}$). (**D**) *CLCF1* expression was significantly upregulated in OA lesioned articular cartilage compared with paired preserved articular cartilage (FC=1.57, $P_{FDR}$=0.02). (**E**) The CLF1/CLC complex receptor gene *CNTFR* was significantly downregulated in OA lesioned articular cartilage compared with paired preserved articular cartilage (FC=0.32, $P_{FDR}$=5.26*$10^{-5}$). (**F**) Expression of catabolic markers *ADAMTS5*, *MMP13* and *COL10A1* were not affected by addition of either 10 (+) or 50 (++) ng $mL^{-1}$ CLF1/CLC protein complex to the culture medium of primary chondrocytes isolated from 4 individual knee cartilage samples. Expression of anabolic markers *COL2A1*, *AGC*, *COL1A1*, *COL6A3*, *FN1* and *COMP* were significantly affected as reflected by a dose response curve upon addition of the protein complex. Preserved samples are indicated in blue, lesioned samples in red, knees as circles and hips as triangles.

**CLF1/CLCF1 MODULATES ARTICULAR CARTILAGE ANABOLISM *IN-VITRO*.**

Finally, we studied downstream effects of increased *CRLF1* signaling in cartilage micro-masses derived from primary human chondrocytes isolated from 4 patients who underwent a total joint replacement of the knee due to primary OA. After proliferation and formation of the extracellular matrix (ECM) on day 3, micro-masses were treated with 10 or 50 ng $mL^{-1}$ of human recombinant CLF1/CLCF1 protein complex for 4 consecutive days and RNA was isolated to assess transcriptional differences. We did not observe a difference in expression of the catabolic genes *ADAMTS5*, *MMP13* and *COL10A1* with either concentration of added CLF1/CLCF1 protein complex compared to control cultures. However, a marked dose response upregulation was observed for multiple anabolic genes, such as *COL2A1*, *AGC*, *COL1A1*, *COL6A3*, *FN1* and *COMP* (**Figure 3F**).

## DISCUSSION

Our approach in the current paper comprises a concept framework for complex traits to identify disease-relevant genetic variation, as reflected by allele associated transcription levels in the disease's pivotal tissue. We aim to present the reported observations as a legacy dataset for researchers from the field to probe for their gene or SNP of interest and have described by proof of concept how the *CRLF1* locus seems causally involved in the progression of OA-associated cartilage degradation.

*CRLF1*, the gene encoding for the CLF1 protein, harbors the rs7256319 C>T SNP, which marked imbalanced expression of its respective alleles in articular cartilage, reflected by consistent lower expression of the alternative allele T in comparison with the reference allele C among heterozygotes. As has been reported previously (12, 27), *CRLF1* was significantly upregulated in OA-affected compared to preserved articular cartilage, as was its signaling partner *CLCF1*, while the protein complex signaling receptor gene *CNTFR* was downregulated. The decreased expression of the alternative allele of rs7256319 in heterozygotes, significant upregulation of *CRLF1* in lesioned cartilage and the respective protective genetic association seem to indicate that mitigating *CRLF1*/*CLCF1* signaling will be beneficial towards ongoing cartilage degradation due to primary OA.

Less straightforward, however, may be the subsequent observation that increasing concentrations of CLF1/CLC protein complex, applied to primary chondrocytes derived cartilage constructs, resulted in a significant dose responsive upregulation of multiple anabolic cartilage genes, while catabolic genes remained unaffected. This, as well as upregulation of catabolic genes and downregulation of anabolic genes are generally considered relevant determinants in OA. We propose that increased expression of anabolic markers reflect the incapability of chondrocytes to return to a steady-state upon disease

associated CLF1/CLC-CNTFR signalling, likely required for adapting to mechanical stress and subsequent articular cartilage micro trauma. Alternatively, increased expression of these anabolic genes might influence the stiffness of the ECM and thereby affect mechanical resilience of articular cartilage. In any way, our data suggest that consistent innate increased expression of *CRLF1*, either during development of the cartilage anlagen and/or with ongoing pathology, affects the propensity of articular cartilage to engage an OA phenotype. Paradoxically, treatment of 2D cultured mouse ATDC5 cells with human CLF1/CLC complex was shown to induce downregulation of *Acan* and *Col2a1*, while catabolic genes remained unaffected (28). Aside from this apparent species-specific and/or difference in culture protocol (3D vs 2D) response, the role of the CLF1/CLC complex in OA seems depicted by the inability to accordingly regulate articular cartilage anabolism, as opposed to affect degradation.

Up for speculation remains how rs7256319 marks allelic *CRLF1* expression mechanistically, especially due to its location within an intron, suggesting intron retention and/or alternatively spliced transcripts, which we were, however, unable to detect consistently (data not shown). A number of transcription factors is predicted to bind at the rs7256319 locus (e.g. *JUN*, *REST* and *SP1*, (based on HaploReg 4.1)) and regulation might thus be affected by altered binding motifs. Possibly, regulation by SNPs in high LD with rs7256319 that are situated in other regulatory elements, such as rs3170474, which disrupts another set of transcription factor binding sets (e.g. *HEY1*, *INSM1* and *PAX5*), is located in the 3' UTR of *C19orf60* and was shown to be an expression quantitative trait locus (eQTL) in brain tissues, are responsible for the marked allelic imbalance.

Although outside the scope of our current efforts, we believe that alongside rs7256319, the 2070 SNPs that mark AI in articular cartilage are likely to contain multiple compelling association signals. While in canonical GWAS a strict genome-wide significance level of $5 \times 10^{-8}$ is imposed due to the vast amount of SNPs that is tested for, we here postulate that selecting for SNPs that are more likely to contribute to the disease *a priori*, by means of marking AI in articular cartilage, could aid the search for OA susceptibility loci substantially. Further downstream selection criteria, such as, but not limited to, significant differential expression between preserved and OA lesioned cartilage, will help tailor genetic association analyses even more and might attribute SNPs to specific disease facets, such as progression of cartilage degradation, as we have shown here. Of note, in this regard, it deserves mention that inherent to our study design, we have potentially missed genes that affect joint morphology, cartilage integrity during development and/or change expression during early stage OA.

A number of SNPs is known to mark AI in articular cartilage, as has been shown by gene-targeted approaches. We were able to replicate the earlier observed AI of *ALDH1A2* and to lesser extent of *DIO2*, marked by rs3204689 (22) and rs225014 (19) respectively, while rs1676486 (*COL11A1*) (21),

rs11177 (*GNL3*) (20) and rs6617 (*SPCS1*) (20) did not mark consistent AI in our dataset. Additionally, we did not observe heterozygotes for rs143383 (*GDF5*) (25) or rs3815148 (*HBP1*) (26). These findings indicate first and foremost that additional replication, preferably using a different technique, is required to increase confidence in the observed AI, as we have done for rs7256319. Secondly, it stresses the fact that observed AI reflects regulatory properties of the respective LD block and does not *per se* identify genetic variation that affects respective gene expression levels mechanistically. Furthermore, despite the applied filtering steps and statistics, the list of significant AI SNPs potentially contains a number of false positives, of which some could have originated from alignment bias. While future novel alignment and other bioinformatic approaches (29) might address these issues from a more fundamental perspective, we have here aimed to reduce them by inclusion of multiple filtering steps ($0.1<\boldsymbol{\varphi}<0.9$, present in at least two heterozygotes and null hypothesis adjustment).

In summary, we present a framework and resulting dataset for researchers in the OA research field to probe for disease relevant genetic variation that affects respective gene expression in the disease's pivotal affected tissue. Furthermore, conceptual downstream analyses and experiments revealed *CRLF1* to be actively involved in OA associated articular cartilage degradation, as reflected by lower expression of the rs7256319 protective allele T among heterozygous carriers and aberrant tissue homeostasis upon increased CLF1/CLCF1 signalling *in-vitro*.

## MATERIALS AND METHODS

### COHORTS

Ethical approval for the RAAK study was obtained from the medical ethics committee of the LUMC (P08.239) and informed consent was obtained from 68 participants. From 21 patients preserved as well as lesioned cartilage was sampled (6 hip patients, 15 knee patients), complemented by an additional 21 preserved (14 hip patients, 7 knee patients) and 5 lesioned (2 hip, 3 knee patients) samples (**Supplementary Table S1**). For cartilage sampling details see (12, 14). The GARP study consists of 380 siblings with symptomatic OA at multiple joint sites. BMI, age and sex was available in 378 subjects with mean BMI = 27.03, age range 50-75 years, 82.0% female) (9, 30). The LLS study consists of 2415 individuals of whom genetic data is available for 1583 offspring of long-lived individuals and 732 of their partners. BMI, age and sex was available in 1976 subjects with mean BMI = 25.41, age range 30-80 years, 54.9% female. (31) Genotyping and quality control of the data from GARP and LLS participants was done as described previously (11). In the current paper, to select for symptomatic OA patients, we performed in silico association of rs7256319 to those patients of the Rotterdam studies that underwent a joint replacement surgery as result of their OA.

**RNA-SEQUENCING DATA**.

Post RNA isolation (Qiagen RNeasy Mini Kit, RIN>7), paired-end 100 bp RNA library sequencing (Illumina TruSeq RNA Library Prep Kit, Illumina HiSeq 2000) resulted in an average of 10 million clusters. Reads were aligned using GSNAP against the hg19 reference genome, while known Dutch SNPs (GoNL) were masked to aid with potential reference alignment bias. AI events were assessed on SNPs called using SNVMix2 with default settings (32) with minimum coverage of 25 and at least 10 reads (***R***) per allele. AI is reported by the average fraction (***φ***) of the alternative allele reads ($\boldsymbol{R}_{alternative}$) among the total number of reads ($\boldsymbol{R}_{total}$=$\boldsymbol{R}_{alternative}$+$\boldsymbol{R}_{reference}$) at the position of the respective genetic variation per sample (***i***):

$$\boldsymbol{\varphi} = \frac{1}{n}\sum_{i=1}^{n}\frac{R_{i,alternative}}{R_{i,reference}}$$

To detect SNPs that robustly mark imbalance two binomial tests were performed per heterozygote and per SNP under the null hypothesis that the amount of imbalance is either greater or smaller than 0.49. Subsequently, P-values per SNP were corrected for multiple testing (FDR) by the number of heterozygotes of the respective SNP and considered significant if all FDR corrected P-values were <0.05 and in the same direction among all heterozygotes. Using the edgeR package, fragments per gene were used to assess the dispersion by quantile-adjusted conditional maximum likelihood (qCML) (33). Subsequently, differential gene expression analysis was performed pairwise between preserved and lesioned samples for which we had RNA of both (N=21, **Supplementary Table S1**) followed by FDR correction. GO term enrichment analysis was performed using the online available tool DAVID (34).

**TAQMAN ASSAY**

Conventional TaqMan genotyping was performed on both genomic DNA and articular cartilage cDNA (35) from 6 (2 female, 4 male) patients who underwent total joint replacement surgery of the knee due to primary OA. An allele-specific custom TaqMan assay for rs7256319 (Thermo Fisher Scientific) was used to quantify the allele ratio in cDNA samples and were normalized against the gDNA ratio, which was used as a 1:1 allele ratio reference.

**CELL CULTURE AND RT-QPCR**

Within the ongoing RAAK study human primary chondrocytes were isolated from macroscopically intact cartilage of OA patients (4 females) undergoing total knee arthroplasty, as described previously (19). In short, cartilage was rinsed with phosphate buffered saline (PBS), cut into small pieces, and incubated overnight with 3 mg $mL^{-1}$ collagenase I (Worthington Biochemical Corporation, USA) in

Dulbecco's Modified Eagle's Medium (DMEM; Gibco) containing antibiotics, 10% fetal bovine serum (Gibco) and 0.5 ng $mL^{-1}$ recombinant human basic FGF (PeproTech). Next day, cells were filtered through a 100 μm mesh to remove undigested cartilage fragments and debris. Chondrocytes were passaged twice before harvest and incubation in 3D-pellets (2.5 x $10^5$ cells/pellet) in serum-free chondrogenic medium in the absence or presence of 10 or 50 ng $mL^{-1}$ recombinant human CRLF1/CLC complex (R&D Systems). Chondrogenic medium with or without recombinant proteins was refreshed at day 2 and day 6, and pellets were harvested for RNA isolation at day 7 (Qiagen RNeasy Mini Kit). Gene expression was determined with RT-qPCR using FastStart SYBR Green Master reaction mix (Roche Applied Science) according to the manufacturer's protocol and corrected for ADP-ribosylation factor related protein 1 (*ARFRP1*; primer sequences are shown in **Supplementary Table S7**).

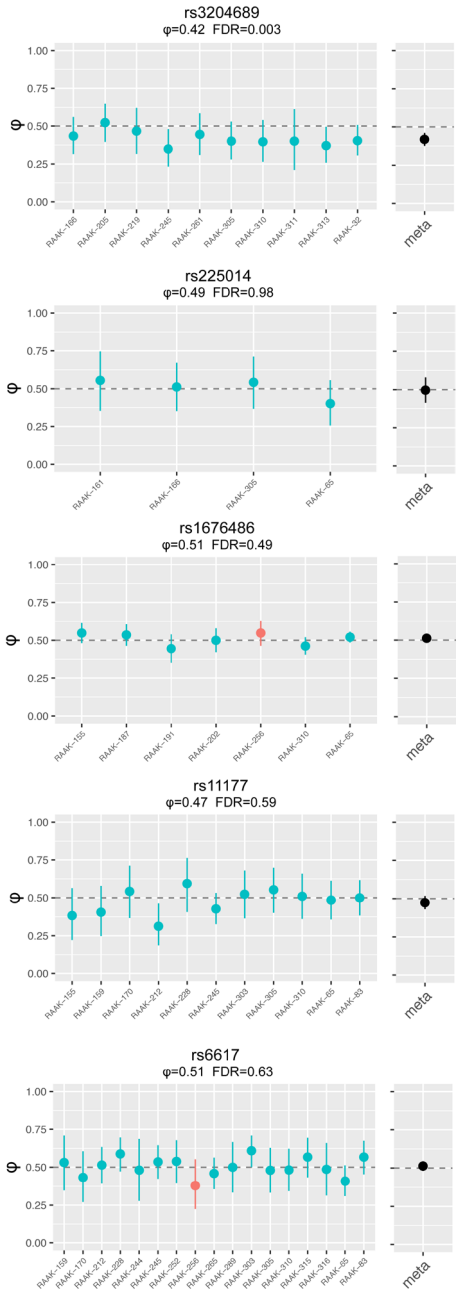
## ACKNOWLEDGEMENTS

The Leiden University Medical Centre, the Dutch Arthritis Association and Pfizer Inc., Groton, CT, USA support the GARP study, whilst the LLS was supported by the Netherlands Organization of Scientific Research (MW 904-61-095, 911-03-016, 917-66-344 and 911-03-012), Leiden University Medical Centre, and by the "Centre of Medical System Biology" and the "Netherlands Consortium of Healthy Aging" in the framework of the Netherlands Genomics Initiative (NGI). Furthermore, the research leading to the RAAK biobank and the current results has received funding from the Dutch Arthritis Association (DAA 2010_017) and the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement no. 259679. We thank Nico Lakenberg, Ruud van der Breggen and Eka Suchiman, for their help in preparing DNA and RNA samples.

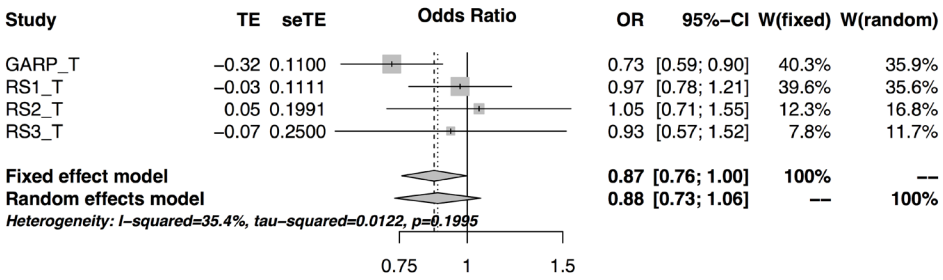
## REFERENCES


1. Bijlsma,J.W., Berenbaum,F., & Lafeber,F.P. Osteoarthritis: an update with relevance for clinical practice. *The Lancet* 377, 2115-2126 (2011).
2. Man,G.S. & Mologhianu,G. Osteoarthritis pathogenesis- a complex process that involves the entire joint. *J. Med. Life* 7, 37-41 (2014).
3. Saito,T. *et al.* Transcriptional regulation of endochondral ossification by HIF-2[alpha] during skeletal growth and osteoarthritis development. *Nat Med* 16, 678-686 (2010).
4. Sherwood,J. *et al.* A homeostatic function of CXCR2 signalling in articular cartilage. *Ann. Rheum. Dis.* (2014).
5. Blaney Davidson,E.N. *et al.* TGF-beta is a potent inducer of Nerve Growth Factor in articular cartilage via the ALK5-Smad2/3 pathway. Potential role in OA related pain? *Osteoarthritis Cartilage* 23, 478-486 (2015).
6. Jeffries,M.A. *et al.* Genome-wide DNA methylation study identifies significant epigenomic changes in osteoarthritic cartilage. *Arthritis Rheumatol.* (2014).
7. Rushton,M.D. *et al.* Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. *Arthritis Rheumatol.* (2014).
8. Zeggini,E. *et al.* Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* 380, 815-823 (2012).
9. Meulenbelt,I. *et al.* Identification of DIO2 as a new susceptibility locus for symptomatic osteoarthritis. *Human Molecular Genetics* 17, 1867-1875 (2008).
10. Xu,Y. *et al.* Identification of the pathogenic pathways in osteoarthritic hip cartilage: commonality and discord between hip and knee OA. *Osteoarthritis. Cartilage.* 20, 1029-1038 (2012).
11. den Hollander,W. *et al.* Transcriptional associations of osteoarthritis mediated loss of epigenetic control in articular cartilage. *Arthritis & Rheumatology*n/a (2015).
12. Ramos,Y.F. *et al.* Genes Involved in the Osteoarthritis Process Identified through Genome Wide Expression Analysis in Articular Cartilage; the RAAK Study. *PLoS. ONE.* 9, e103056 (2014).
13. Zhang,R. *et al.* Gene expression analyses of subchondral bone in early experimental osteoarthritis by microarray. *PLoS. ONE.* 7, e32356 (2012).
14. Bomer,N. *et al.* Underlying molecular mechanisms of DIO2 susceptibility in symptomatic osteoarthritis. *Ann. Rheum. Dis.* 74, 1571-1579 (2015).
15. Reynard,L.N., Bui,C., Canty-Laird,E.G., Young,D.A., & Loughlin,J. Expression of the osteoarthritis-associated gene GDF5 is modulated epigenetically by DNA methylation. *Hum. Mol. Genet.* 20, 3450-3460 (2011).
16. Styrkarsdottir,U. *et al.* Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat Genet* advance online publication, (2014).
17. Gee,F., Clubbs,C.F., Raine,E.V., Reynard,L.N., & Loughlin,J. Allelic expression analysis of the osteoarthritis susceptibility locus that maps to chromosome 3p21 reveals cis-acting eQTLs at GNL3 and SPCS1. *BMC. Med. Genet.* 15, 53 (2014).
18. Raine,E.V., Dodd,A.W., Reynard,L.N., & Loughlin,J. Allelic expression analysis of the osteoarthritis susceptibility gene COL11A1 in human joint tissues. *BMC. Musculoskelet. Disord.* 14, 85 (2013).
19. Bos,S.D. *et al.* Increased type II deiodinase protein in OA-affected cartilage and allelic imbalance of OA risk polymorphism rs225014 at DIO2 in human OA joint tissues. *Annals of the Rheumatic Diseases* 71, 1254-1258 (2012).
20. Gee,F., Clubbs,C.F., Raine,E.V., Reynard,L.N., & Loughlin,J. Allelic expression analysis of the osteoarthritis susceptibility locus that maps to chromosome 3p21 reveals cis-acting eQTLs at GNL3 and SPCS1. *BMC. Med. Genet.* 15, 53 (2014).
21. Raine,E.V., Dodd,A.W., Reynard,L.N., & Loughlin,J. Allelic expression analysis of the osteoarthritis susceptibility gene COL11A1 in human joint tissues. *BMC. Musculoskelet. Disord.* 14, 85 (2013).
22. Styrkarsdottir,U. *et al.* Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat Genet* 46, 498-502 (2014).
23. Loughlin,J. Osteoarthritis year 2010 in review: genetics. *Osteoarthritis Cartilage* 19, 342-345 (2011).
24. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648-660 (2015).
25. Southam,L. *et al.* An SNP in the 5'-UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage. *Hum. Mol. Genet.* 16, 2226-2232 (2007).

26. Raine,E.V.A., Wreglesworth,N., Dodd,A.W., Reynard,L.N., & Loughlin,J. Gene expression analysis reveals HBP1 as a key target for the osteoarthritis susceptibility locus that maps to chromosome 7q22. *Annals of the Rheumatic Diseases* 71, 2020-2027 (2012).
27. Elson,G.C. *et al.* CLF associates with CLC to form a functional heteromeric ligand for the CNTF receptor complex. *Nat. Neurosci.* 3, 867-872 (2000).
28. Tsuritani,K. *et al.* Cytokine receptor-like factor 1 is highly expressed in damaged human knee osteoarthritic cartilage and involved in osteoarthritis downstream of TGF-beta. *Calcif. Tissue Int.* 86, 47-57 (2010).
29. Castel,S.E., Levy-Moonshine,A., Mohammadi,P., Banks,E., & Lappalainen,T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* 16, 195 (2015).
30. Riyazi,N. *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann. Rheum. Dis.* 64, 438-443 (2005).
31. Westendorp,R.G. *et al.* Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. *J. Am. Geriatr. Soc.* 57, 1634-1637 (2009).
32. Goya,R. *et al.* SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics.* 26, 730-736 (2010).
33. Robinson,M.D., McCarthy,D.J., & Smyth,G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 26, 139-140 (2010).
34. Huang,d.W., Sherman,B.T., & Lempicki,R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44-57 (2009).
35. Wilkins,J.M. *et al.* Extreme context-specificity in differential allelic expression. *Hum. Mol. Genet.* (2007).




**Supplementary Figure S1.** The extent of AI plotted for SNPs that have been reported to mark AI in known OA susceptibility genes.




**Supplementary Figure S2.** Forest plot of meta-analysis performed in GARP and RS for rs7256319.

*Supplementary tables upon request.*

**Supplementary Table S1.**

Sample characteristics.

**Supplementary Table S2.**

All observed SNPs in at least two heterozygotes.

**Supplementary Table S3.**

Differential gene expression between preserved and paired lesioned samples.

**Supplementary Table S4.**

Significantly enriched GO terms among the differentially expressed genes (N=166, 0.5>FC>2, FDR<0.05).

**Supplementary Table S5.**

Primer sequences used for RT-qPCR analysis.

**Supplementary Table S6.**

Overview of differentially expressed genes (0.5>FC>2) between preserved and paired OA lesioned articular cartilage that are additionally influenced by *in-cis* genetic regulation as reflected by AI. FC=Foldchange. FDR=Benjamini-Hochberg corrected P-value.

**Supplementary Table S7.**

qPCR primer sequences.