

SCIENTIFIC REPORTS

OPEN

Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis

Ye Liu^{1,2,3}, Jen-Chien Chang^{1,2}, Chung-Chau Hon^{4,5}, Naoshi Fukui^{6,7}, Nobuho Tanaka⁶, Zhenya Zhang³, Ming Ta Michael Lee^{8,9,10} & Aki Minoda^{1,2}

Osteoarthritis (OA) is a common joint disorder with increasing impact in an aging society. While genetic and transcriptomic analyses have revealed some genes and non-coding loci associated to OA, the pathogenesis remains incompletely understood. Chromatin profiling, which provides insight into gene regulation, has not been reported in OA mainly due to technical difficulties. Here, we employed Assay for Transposase-Accessible Chromatin with high throughput sequencing (ATAC-seq) to map the accessible chromatin landscape in articular knee cartilage of OA patients. We identified 109,215 accessible chromatin regions for cartilages, of which 71% were annotated as enhancers. By overlaying them with genetic and DNA methylation data, we have determined potential OA-relevant enhancers and their putative target genes. Furthermore, through integration with RNA-seq data, we characterized genes that are altered both at epigenomic and transcriptomic levels in OA. These genes are enriched in pathways regulating ossification and mesenchymal stem cell (MSC) differentiation. Consistently, the differentially accessible regions in OA are enriched for MSC-specific enhancers and motifs of transcription factor families involved in osteoblast differentiation. In conclusion, we demonstrate how direct chromatin profiling of clinical tissues can provide comprehensive epigenetic information for a disease and suggest candidate genes and enhancers of translational potential.

Osteoarthritis (OA) is a degenerative joint disease^{1,2} that is one of the most common causes of chronic disability in the world^{3,4}, of which the knee OA is the most common. Main features of OA include cartilage degradation, subchondral bone thickening, joint space narrowing and osteophytes formation⁵, resulting in stiffness, swelling, and pain in the joint. Currently available treatments are either pain relief or joint function improvement by strengthening the supporting muscles. However, OA progression ultimately leads to costly total joint replacement surgery, making it a growing global health burden.

Although the causes of OA are not well understood, risk factors such as age, weight, gender, and genetic factors have been identified⁴. Several models for OA initiation, such as mechanical injury, inflammatory mediators from synovium, defects in metabolism and endochondral ossification, have been proposed to explain pathogenesis of this disease^{6–12}. To date, Genome-wide association studies (GWAS) have identified more than 20 loci to be associated with the risk of developing OA¹³. While next generation sequencing data is being generated to

¹Epigenome Technology Exploration Unit, Division of Genomic Technologies, RIKEN Center for Life Science Technologies (CLST), Yokohama, Japan. ²Epigenome Technology Exploration Unit, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. ³Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan. ⁴Genome Information Analysis Team, Division of Genomic Technologies, RIKEN Center for Life Science Technologies (CLST), Yokohama, Japan. ⁵Laboratory for Genome Information Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. ⁶Clinical Research Center, National Hospital Organization Sagami Hospital, Kanagawa, Japan. ⁷Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan. ⁸Genomic Medicine Institute, Geisinger, Danville, PA, USA. ⁹Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. ¹⁰Laboratory for International Alliance on Genomic Research, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. Ye Liu and Jen-Chien Chang contributed equally. Correspondence and requests for materials should be addressed to M.T.M.L. (email: mlee2@geisinger.edu) or A.M. (email: akiko.minoda@riken.jp)

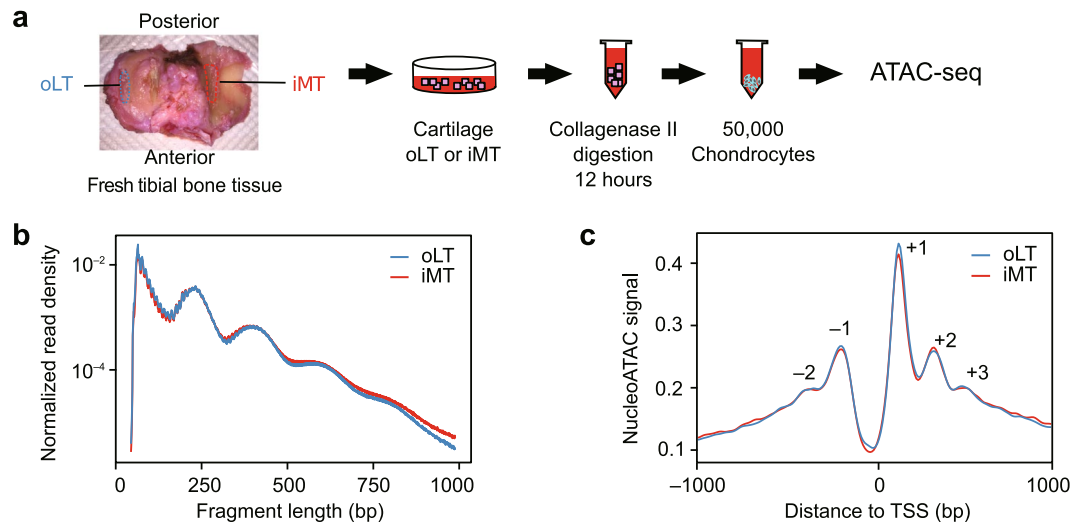


Figure 1. ATAC-seq of OA cartilage tissues. **(a)** Schematic diagram of the experimental flow. Chondrocyte from two regions of the articular cartilages were isolated for ATAC-seq. oLT (outer Libial Tibial): intact tissue. iMT (inner Medial Tibial): damaged tissue. **(b)** ATAC-seq sequencing fragment length distribution for pooled oLT ($n = 8$) and iMT ($n = 8$) libraries. The banding pattern, representing fragments from nucleosome free regions, mono-, di-, and tri-nucleosome, indicates a successful ATAC-seq experiment. **(c)** Normalized nucleoATAC signal aggregated over all genes shows distinct nucleosome positioning around transcription start sites (TSS). Positive and negative numbers indicate the TSS upstream and downstream nucleosomes.

discover rare variants with larger effect size, the identified variants are often located in the non-coding regions of the genome¹⁴, complicating the identification of the causal genes. Transcriptomic analyses of cartilage in diseased joints of OA patients (taken from replacement surgeries) provided an opportunity to pinpoint transcriptionally dysregulated genes and pathways relevant to OA^{15,16}. However, such studies have yet to fully reveal the underlying molecular mechanism of how the transcription of these genes are dysregulated.

Recently, epigenetic tools have been applied to gain further insight into the pathogenesis of OA. There have been reports of DNA methylation status in cartilage of diseased joints, revealing epigenetic marks as potential mediators of OA genetic risk^{17–23}. However, the change of gene expression is rarely associated with DNA methylation alterations at promoters^{16,24,25}. Many of the identified differentially methylated sites fall in enhancer regions^{19,21–23}, which are non-coding regulatory elements, disruption of which may lead to dysregulated transcription, and many are cell type-specific^{26,27}. Recent large-scale studies, such as FANTOM5²⁸, Roadmap Epigenomics Project²⁹ and GTEx³⁰ have enabled the prediction of regulatory networks between enhancers and their potential target genes (e.g. JEME³¹), which could be applied in a clinical context to explore the roles of enhancers in disease pathogenesis.

Here, we set out to investigate alterations of enhancers associated with OA by applying ATAC-seq³² on the knee joint cartilages from OA patients, using an optimized protocol for cartilage sample preparation. ATAC-seq maps the accessible chromatin regions, which are often regulatory regions such as promoters and enhancers that play roles in regulation of gene expression. By integrating our ATAC-seq data with the publicly available genetic, transcriptomic and epigenomic data, we identified dysregulated enhancers and their potential target genes. Our data highlights a number of OA risk loci and differentially methylated loci (DML) that potentially play roles in cartilage degradation during OA development.

Results

Mapping chromatin accessibility of chondrocytes in OA knee cartilages. To investigate chromatin signatures in articular cartilage associated with OA, we performed ATAC-seq on the chondrocytes isolated from the knee joints of patients. We have previously shown that the oLT region (outer region of the lateral tibial plateau, representing the intact cartilage) is a good control for comparing the iMT region (inner region of medial tibial plateau, representing the damaged cartilage) as a model for OA disease progression¹⁶, and the transcriptome and methylome of this model have been characterized by us and the others^{24,25,33,34}. In this study, we performed ATAC-seq on the chondrocytes isolated from oLT and iMT regions of 8 patients (Supplementary Figure S1a), generating 16 ATAC-seq libraries (pairs from 8 patients) on clinically relevant OA cartilage tissues (Fig. 1a).

Overall, the libraries are of high quality, showing two- to four-fold enrichment in the Roadmap DHS (DHS enrichment score, Methods), with no substantial difference between oLT and iMT libraries (Supplementary Table S2). In addition, the expected nucleosome banding patterns were observed in the fragment size distribution for both oLT and iMT libraries (Fig. 1b). When we applied nucleoATAC to infer genome-wide nucleosome occupancy and positioning from ATAC-seq data³⁵, we found similar aggregated signals around transcription start sites (TSS) for both oLT and iMT libraries, corresponding to -2 , -1 , $+1$, $+2$, $+3$ nucleosomes as well as nucleosome depletion region at upstream of TSS (Fig. 1c). Thus, we concluded that our ATAC-seq libraries had good and indistinguishable quality between oLT and iMT regions.

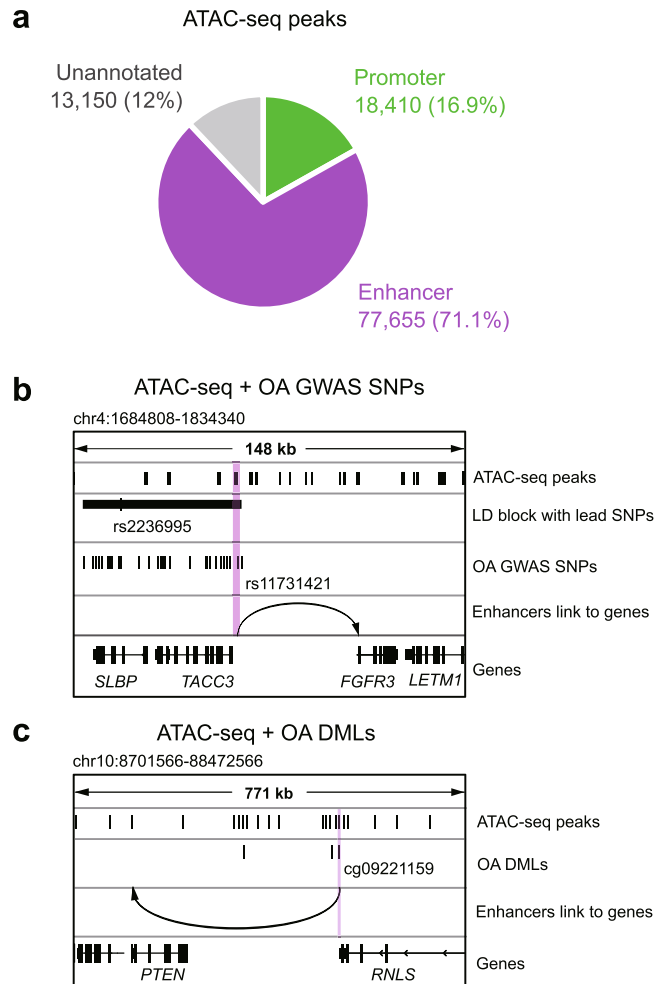


Figure 2. Identification of accessible chromatin regions with OA susceptible GWAS SNPs and differentially methylated loci. **(a)** Annotation of identified accessible chromatin regions. **(b, c)** An example of an OA GWAS SNP rs11731421 **(b)** and differentially methylated locus cg09221159 **(c)** overlapping with an open enhancer in cartilages from OA patients. The arrow points to the predicted target genes.

Accessible chromatin landscape highlights potential enhancers and their target genes relevant to OA.

Based on the 16 ATAC-seq libraries, we identified a set of unified accessible chromatin regions across all samples ($n = 109,215$ robust peaks, Methods, Supplementary Table S3); 77,655 (71.1%) of which were annotated as enhancers and 18,410 (16.9%) as promoters (Fig. 2a) based on Roadmap DHS annotations (Methods). To assess the relevance of these peaks to OA, we intersected these peaks against OA GWAS (Single Nucleotide Polymorphisms) SNPs and OA DML datasets. We found 149 peaks (31 promoters and 101 enhancers) overlapping with OA-associated GWAS SNPs and 2,508 peaks (595 promoters and 1,715 enhancers) overlapping with OA DML. The fact that majority of these overlapping peaks are enhancers, along with the notion that many disease-associated noncoding variants reside in enhancers^{14,36}, suggest the dysregulation of enhancer might play a role in OA pathogenesis. To further characterize these potentially OA-relevant enhancers, we utilized public resources to predict their target genes (Methods). We listed enhancers that overlap OA GWAS SNPs (Supplementary Table S4) and OA DML (Supplementary Table S5), as well as their predicted target genes, representing candidates of dysregulated transcriptional network during OA pathogenesis. The list includes previously identified OA associated genes, such as *FGFR3*^{37,38} (its enhancer overlaps an OA GWAS SNP, Fig. 2b) and *PTEN*³⁹ (its predicted enhancer overlaps with an OA DML, Fig. 2c), thus validating our approach. rs4646626 has previously been identified as a suggestive OA risk locus ($p = 9 \times 10^{-6,40}$), for which we found two proxy SNPs (rs10851632 and rs12905608), predicted to target the *ALDH1A2* gene, hitting the accessible enhancer region (chr15:57949887–57950974 and chr15:58021340–58022482, respectively) in our samples. Therefore, our study provides an accessible chromatin landscape of cartilage tissue for better interpretation of other genetic and epigenomic data relevant to OA and other skeletomuscular disease, although caution should be taken since public data from many populations are involved in the integrative analysis while our ATAC-seq dataset is acquired from the Japanese population.

Identification of differentially accessible enhancers in OA. To compare the accessible chromatin landscape between the intact (oLT) and the damaged (iMT) tissues, we performed principal component analysis

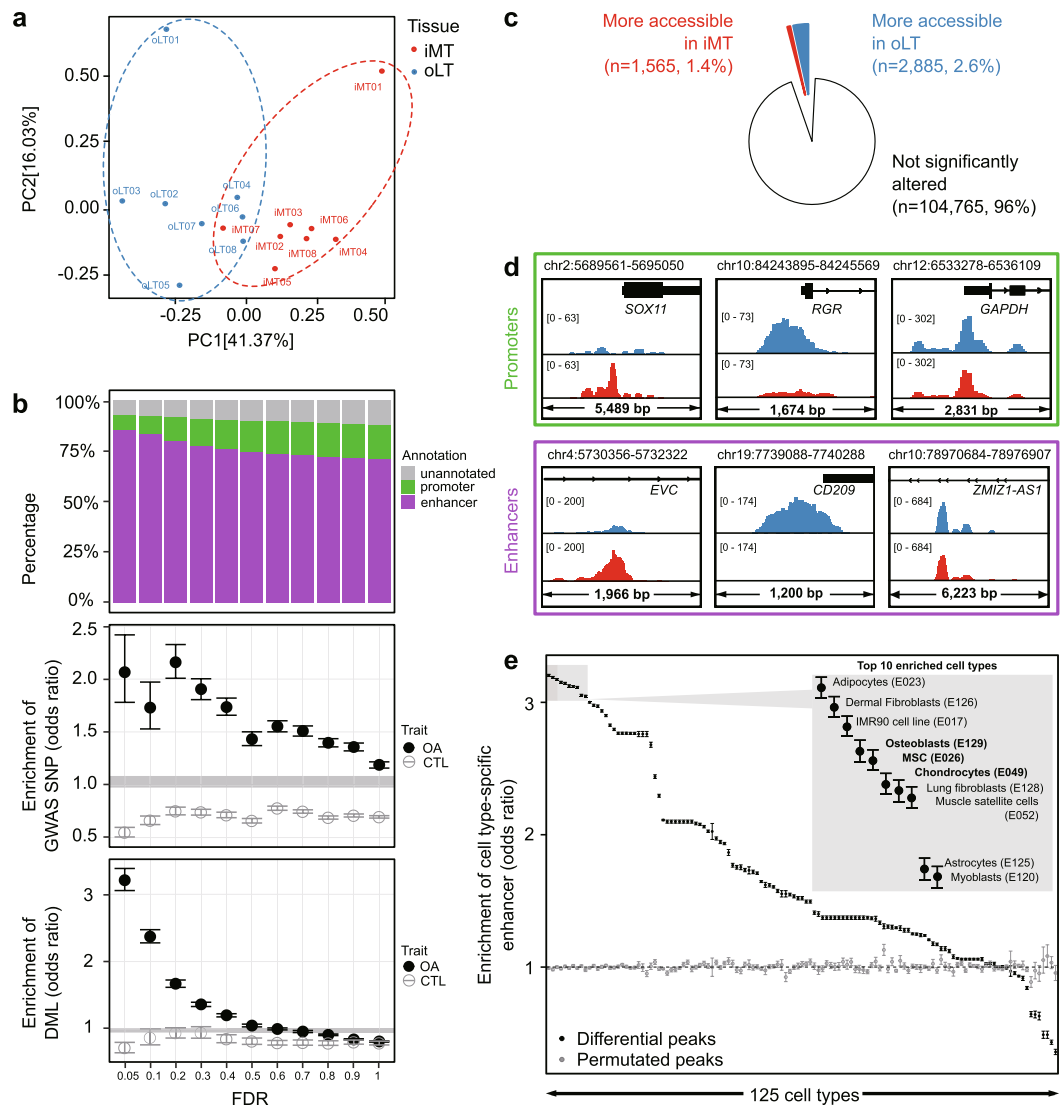


Figure 3. Bone- and chondrocyte-specific enhancers are dysregulated in OA cartilage. **(a)** Principal component analysis using ATAC-seq peaks shows general separation between oLT and iMT. **(b)** Peaks with smaller FDR (i.e. more confident to be differentially accessible) are enriched for enhancers (*top*), OA GWAS SNPs (*middle*), and OA differential methylated loci (DML, *bottom*). CTL: Control trait (Parkinson's disease). Circles and error bars represent means and 95% confidence intervals for 100 permutations. Shaded grey area represents 0.05–0.95 percentiles of baseline enrichment by measuring randomly shuffled peaks. **(c)** Pie chart shows proportions of the ATAC-seq peaks that are more accessible in iMT (red) and oLT (blue). **(d)** Genome browser views of example loci (all patients pooled) showing differential accessible regions at promoters (*top*) and enhancers (*bottom*) with more accessible in iMT (*left*), more accessible in oLT (*middle*), and not significantly altered between oLT and iMT (*right*). **(e)** Enrichment of cell type-specific enhancers in differentially accessible peaks. Top 10 are listed (inset), which includes bone- and chondrocyte-related cell types (bold). Error bars represent 95% confidence interval.

on the peak signals across the 16 samples. The first principal component (41.37% of variance) can be attributed to tissue damage variations (i.e. oLT versus iMT), while the second principal component (16.03% of variance) can be attributed to patient-to-patient variations (Fig. 3a). This observation suggests the accessible chromatin landscapes of damaged and intact tissues are readily distinguishable from each other, despite the variations among individual patients.

To identify the chromatin signatures relevant to OA, we performed differential accessibility analysis between oLT and iMT using edgeR⁴¹. First and foremost, we assessed the global relevance of the differential accessibility of peaks to OA. In Fig. 3b, we investigated the relationship between the extent of differential accessibility of peaks (measured by False Discovery Rate (FDR)) and their 1) compositions, 2) enrichment in OA GWAS SNPs, and 3) enrichment in OA DML. We found that the peaks that are more differentially accessible (i.e. smaller FDR) contain substantially higher fraction of enhancers (Fig. 3b, *top panel*), suggesting enhancers are more likely to be dysregulated than promoters during OA disease progression. Moreover, peaks that are more differentially accessible

Enhancer peak ID	Proxy SNP ID	Lead SNP ID	r ²	GWAS P value	Phenotype	Population	GWAS literature	log ₂ (fold change)	FDR	Predicted target gene(s)
chr7:32611048–32612186	rs10807862	rs7805536	0.8	4.4 × 10 ⁻⁸	OA	EUR	⁴⁰	-0.96	0.01	ZNRF2P1; DPY19L1P1; AVL9; LSM5
chr7:32616344–32617105	rs10951345	rs7805536	0.8	4.4 × 10 ⁻⁸	OA	EUR	⁴⁰	-0.77	0.03	DPY19L1P1; AVL9
chr10:119370132–119371726	rs17098787	rs11198893	0.9	9.0 × 10 ⁻⁶	OA (knee and hip)	EUR	⁸⁷	-0.59	0.02	RP11-215A21.2
chr7:44238166–44238952	rs67391165	rs3757837	1.0	8.0 × 10 ⁻¹⁰	OA (hip)	EUR	⁸⁸	0.62	0.03	MYL7; YKT6
chr8:125503356–125504061	rs7832357	rs4512391	1.0	1.1 × 10 ⁻⁶	OA (knee and hip)	EUR	⁸⁷	-0.54	0.04	NA

Table 1. Differentially accessible enhancers overlapping with OA GWAS SNPs.

Enhancer peak ID	DML	Delta beta*	Log ₂ (fold change)	FDR	Predicted target gene(s)
chr1:19398124–19398841	cg06360604	-0.177	1.194	0.003	AKR7A2; AKR7A3; CAPZB; PQLC2
chr16:66511371–66512170	cg02934719	-0.168	0.827	0.021	BEANI; TK2
chr11:74025533–74026295	cg01117339	-0.164	0.623	0.037	COA4; RAB6A; UCP2
chr1:85606965–85607684	cg00418071	-0.157	1.214	0.001	CYR61; DDAH1; ZNHIT6; SYDE2
chr4:15880913–158811140	cg11637968	-0.218	0.878	0.032	ETFDH; FAM198B; FNIP2; PPID; C4orf46
chr5:32767956–32768626	cg26647771	-0.198	0.924	0.010	GOLPH3; NPR3; TARS
chr1:226708432–226710048	cg04503570	-0.166	0.500	0.046	PSEN2
chr2:1720779–1722443	cg08216099	-0.193	0.732	0.016	PXDN
chr10:119523073–119524950	cg18591136	-0.152	0.508	0.040	RGS10
chr6:168412123–168412714	cg26379705	-0.174	0.736	0.009	SMOC2
chr6:146850220–146851275	cg18291422	-0.184	0.960	0.012	STXBP5
chr8:81688679–81689538	cg04491064	0.164	-0.959	0.046	CHMP4C
chr12:124700740–124701544	cg04658679	0.160	-0.514	0.030	FAM101A; UBC
chr4:26412447–26413348	cg22992279	0.159	-0.737	0.028	RBPJ
chr11:35344856–35345647	cg13971030	0.157	-0.656	0.038	SLC1A2
chr12:132846887–132848225	cg21232015	0.151	-0.615	0.038	ZNF10; ZNF268; ZNF605; ZNF84; GOLGA3; ANKLE2; CHFR

Table 2. Differentially accessible enhancers overlapping with OA Differentially Methylated Loci. *Positive or negative delta beta indicates hypermethylation or hypomethylation in iMT, respectively.

tend to be more enriched (i.e. higher odds ratio) in both OA GWAS SNPs and DML (Methods, Fig. 3b, middle and lower panels). The differentially accessible enhancers (FDR ≤ 0.05) overlapping either OA GWAS SNP (n = 5; from 4 independent loci) or OA DML (n = 16, concordant in direction) are summarized in Tables 1 and 2, respectively. We note that 92% of the differentially accessible peaks (FDR ≤ 0.05) overlapping DML are concordant in the direction of change (hypomethylation with increased accessibility or hypermethylation with decreased accessibility). As a negative control, we examined the GWAS SNPs and DMLs associated with Parkinson's disease⁴², which is a neurodegenerative disease pathologically unrelated to cartilage, and did not observe any substantial enrichment. Taken together, these observations suggest the differential chromatin accessibility measured in this study is relevant to OA disease progression in a specific manner.

We next determined significantly differentially accessible peaks with FDR ≤ 0.05. Out of the 4,450 differentially accessible peaks, 1,565 are more accessible and 2,885 are less accessible in the damaged tissues compared to the intact tissues (Fig. 3c,d, and Supplementary Figure S2a). The identified differentially accessible peaks cluster in two groups (Supplementary Figure S2b), and are generally consistent among patients (Supplementary Figure S2c). Majority (85.3%) of these differentially accessible peaks are enhancers and only 7.6% are promoters. It is noted that the promoter accessibility of *SOX11* and *RGR* are altered in the OA damaged tissues (more accessible for *SOX11* and less accessible for *RGR*, Fig. 3d), which are consistent with the previous findings that they are up- and down-regulated in OA damaged tissues, respectively²⁴.

Majority of the differentially accessible peaks are enhancers, which are known to regulate transcription. Cell type-specific enhancers largely drive the transcriptional program required to carry out specific functions for each cell type⁴³ and their dysregulation may lead to diseases⁴⁴. To assess the cell type specificity of the differentially accessible enhancers identified in this study, we examined their enrichment of cell type-specific enhancers of 125 cell types defined by the Roadmap Epigenomics Project²⁹. We found these differential accessible enhancers are highly enriched for enhancers that are specific to bone-related cell types in damaged tissues (e.g. chondrocyte and osteoblast), as well as mesenchymal stem cells (MSC) and fibroblasts, which have been reported to have similar profiles to MSC⁴⁵ (Fig. 3e).

In summary, we demonstrate that the differential accessible chromatin regions identified in this study are enriched in OA-relevant enhancers supported by multiple genetic and epigenomic evidence. These differentially

accessible enhancers and their predicted target genes could be used for prioritization of candidate genes to be tested for studying OA disease progression.

Motif enrichment analysis reveals transcription factors relevant to OA. In order to gain more insights into which regulatory pathways may be dysregulated in OA, we next examined the enrichment of transcription factor binding motifs in the differentially accessible regions. Enrichment analyses were performed separately for the regions that are significantly more or less accessible in the damaged tissues, using all accessible regions as background. Transcription factors with significantly enriched motifs are summarized (Supplementary Figure S3a), and their binding prediction in robust peaks are listed (Supplementary Table S3). We note most of these transcription factors belong to ETS and bZIP family; many of which are known to regulate genes involved in bone or cartilage development, including AP-1⁴⁶, CEBP⁴⁷, MafK⁴⁸, STAT3⁴⁹, and ERG^{50,51}. Moreover, we have previously proposed that ETS-1 may be involved in OA based on our DML study²⁵. Taken together, the motif enrichment analysis of the differentially accessible regions in the damaged tissues is consistent with the hypothesis that the transcriptional program for chondrocyte differentiation may be disrupted during OA progression, and suggests that cell type-specific enhancers may be dysregulated through the ETS and bZIP family transcription factors.

Integrative transcriptomics and epigenomics analysis reveals pathways involved in OA. To evaluate the effects of the dysregulated regulatory regions on gene expression in OA, we reanalyzed the RNA-seq dataset from a different cohort that used the same disease model (i.e. oLT vs. iMT)¹⁵ and integrated it with the differentially accessible regions identified in this study (Fig. 4a). We first verified that the dysregulated regulatory regions have detectable effects on the gene expression; the genes with more accessible promoters or enhancers have significantly higher expression fold-changes between oLT and iMT ($p < 0.001$, Student's *t*-test) than the ones with less accessible promoters or enhancers (Fig. 4b). To further investigate the congruence between these transcriptomic and epigenomic changes, we overlapped the significant differentially expressed genes ($n = 3,293$, $FDR \leq 0.05$) onto the genes with differentially accessible promoters ($n = 255$) or enhancers ($n = 2,406$) (Fig. 4c). We find that their overlaps are statistically significant ($p < 0.001$ in both promoters and enhancers, Fisher's exact test), which further demonstrate the chromatin accessibility dataset from this study is generally consistent with the transcriptomic dataset. As a result, we identified 371 genes that are consistently dysregulated both at the epigenomic and transcriptomic levels, representing a shortlist of OA-related candidate genes supported by multiple lines of evidence (Fig. 4c, Table 3, Supplementary Table S6).

To elucidate the biological pathways dysregulated in OA, we examined the enrichment of gene ontology (GO) terms of the 371 OA-related candidate genes using enrichr⁵² (Fig. 4d, top 30 terms listed). Overall, we observe the enrichment of GO terms related to cell fate and differentiation, including MSC differentiation, ossification and bone development (Fig. 4d, dysregulated genes in each GO terms are listed in Supplementary Table S7). In the 'positive regulation of ossification pathway', two genes were identified to be more accessible and upregulated: *SOX11* (only the promoter is more accessible) and *WNT5A* (both the promoter and the enhancer are more accessible) (Supplementary Figure S3b). It has been shown that WNT5A protein can induce matrix metalloproteinase production and cartilage destruction⁵³, and its upregulation is consistent with ossification being an important process and signature of OA progression. Other susceptible genes and pathways that support the ossification during OA from this analysis include *LRP5*, *FGFR2* and *BMPRI1* in the endochondral bone morphogenesis pathway, which have been reported as OA associated genes previously^{18–20,54–56}. In conclusion, our integrative analysis of ATAC-seq and publicly available RNA-seq datasets indicates that dysregulated chondrocyte differentiation and endochondral ossification are associated with OA progression.

Discussion

Conventional epigenomic profiling at the chromatin level, such as chromatin immunoprecipitation sequencing (ChIP-seq) and DNase-seq, is informative in providing insights into the molecular mechanism underlying the regulation of gene expression. However, applying these methods to clinically relevant tissue is less feasible due to the requirement of large number of cells. Especially with the cartilage tissues, the limited tissue sampling size and the extracellular matrix make collection of sufficient cells difficult. One major advantage of ATAC-seq is that it can be achieved with only thousands of cells, making the direct chromatin profiling of clinical samples feasible. In this study, we applied ATAC-seq on OA samples to obtain a chromatin accessibility map in articular cartilage, and identified regulatory regions associated with OA. The lack of normal knee tissues due to difficulties associated with collecting them is a limitation for studying OA. However, our previous studies of the pathology and transcriptome showed that the oLT regions are very similar to normal^{16,24}. Thus, the oLT regions could serve as a suitable alternative to normal control, which could also reduce the inter-individual variations.

Our strategy of integrating the epigenomic data of clinically relevant tissues with the publicly available genetic and transcriptomic data allowed us to better understand how the identified loci may contribute to OA pathogenesis. Most of these accessible chromatin regions are annotated enhancers and we linked them to their putative target genes using public datasets. Previous studies have suggested the potential role of OA-associated epigenetic changes within enhancers in disease pathogenesis^{22,23}. With this enhancer-gene map in chondrocyte, we can now better interpret the previously identified OA GWAS SNPs or OA differential methylated loci located lie outside of the coding regions. For example, we have verified 7 OA associated SNPs (rs10851630, rs10851631, rs10851632, rs12905608, rs12910752, rs4238326, and rs35246600) reside in 3 chondrocyte accessible enhancers of the predicted target gene *ALDH1A2* (Supplementary Table S4). Since *ALDH1A2* is inactivated in prechondrogenic mesenchyme during the cartilage development⁵⁷, these SNPs may contribute to OA through disrupting the enhancer of *ALDH1A2* and inappropriately activating a cell differentiation pathway. Consistently, OA genetic risk variants in *ALDH1A2* locus has also been functionally characterized in a recent study⁵⁸. In addition, we identified

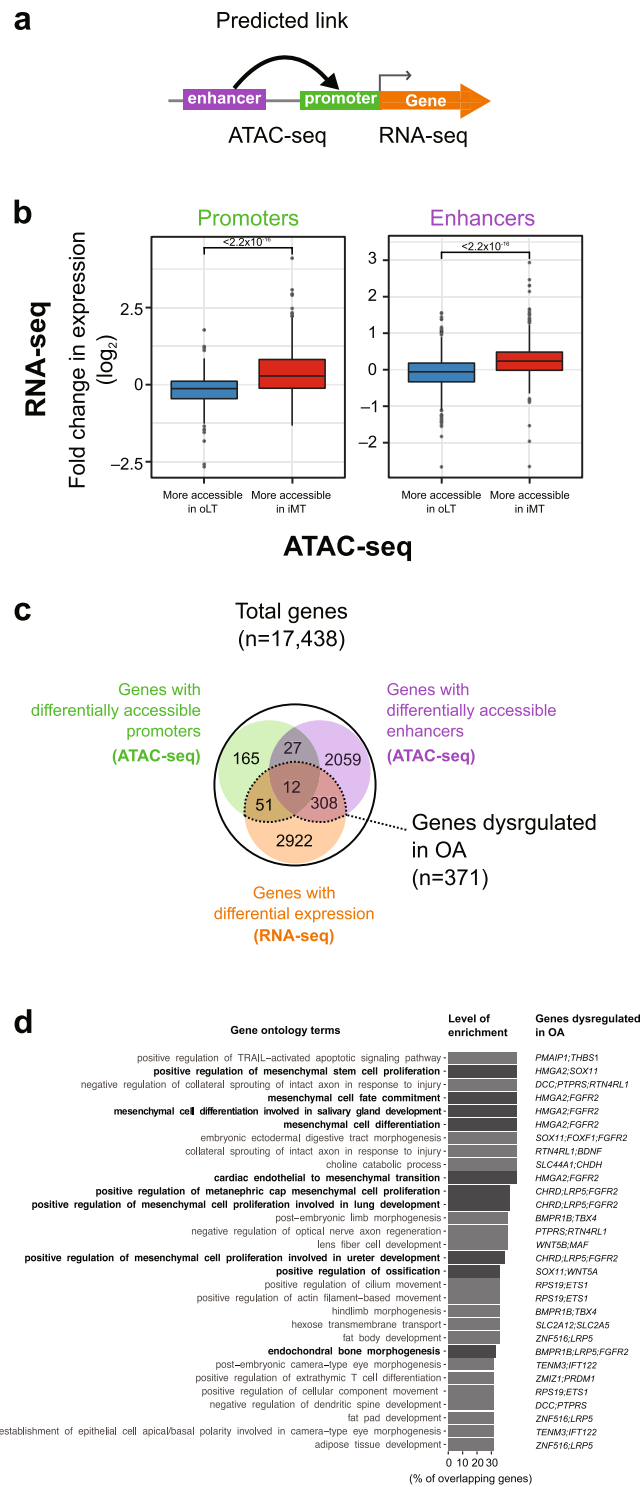


Figure 4. Integrative transcriptomic and epigenomic analysis reveals dysregulated genes and pathways in OA. **(a)** A scheme of ATAC-seq and RNA-seq integration analysis. **(b)** Differential chromatin accessibility (ATAC-seq) at both promoter (*left*) and enhancer (*right*) in OA is generally consistent with differential expression (RNA-seq). Fold change between iMT and oLT is plotted. Box plots show the median, quartiles, and Tukey whiskers. **(c)** Venn diagram summarizing protein coding genes that are dysregulated both at the transcriptomic (RNA-seq) and epigenomic (ATAC-seq) levels in OA with concordant direction of change. **(d)** GO enrichment analysis of 371 genes from (c). Top 30 terms in GO biological process ranked by the level of enrichment and overlapping dysregulated genes are listed. Terms related to MSC, bones, and chondrocytes are highlight in bold.

Genes	Differentially expressed	Differentially accessible promoter	Differentially accessible enhancer
<i>BMPR1B</i>	✓	✓	✓
<i>WNT5A</i>	✓	✓	✓
<i>ZMIZ1</i>	✓	✓	
<i>CHRD</i>	✓	✓	
<i>DCC</i>	✓	✓	
<i>HMGA2</i>	✓	✓	
<i>PMAIP1</i>	✓	✓	
<i>PRDM1</i>	✓	✓	
<i>SOX11</i>	✓	✓	
<i>TENM3</i>	✓	✓	
<i>WNT5B</i>	✓	✓	
<i>BDNF</i>	✓		✓
<i>CHDH</i>	✓		✓
<i>ETS1</i>	✓		✓
<i>FGFR2</i>	✓		✓
<i>FOXF1</i>	✓		✓
<i>IFT122</i>	✓		✓
<i>LRP5</i>	✓		✓
<i>MAF</i>	✓		✓
<i>PTPRS</i>	✓		✓
<i>RPS19</i>	✓		✓
<i>RTN4RL1</i>	✓		✓
<i>SLC2A12</i>	✓		✓
<i>SLC2A5</i>	✓		✓
<i>SLC44A1</i>	✓		✓
<i>TBX4</i>	✓		✓
<i>THBS1</i>	✓		✓
<i>ZNF516</i>	✓		✓

Table 3. OA associated genes enriched in top 30 GO terms.

several aberrantly methylated enhancers that may be associated with OA. One example is cg09221159 within the enhancer for *PTEN* gene which is hypomethylated in the damaged cartilage²⁵. *PTEN* is involved in the positive regulation of the apoptotic signaling pathway and its activation is consistent with the notion that chondrocyte apoptosis may contribute to the failure in appropriately maintaining the cartilage. Thus, our analyses show that the chromatin accessibility map can provide an additional layer of evidence for determining which loci, especially those in the non-coding regions, are associated with OA, which may have been ignored in previous studies.

In general, our differential enhancer analysis shows MSC, chondrocyte and osteoblast-specific enhancers are dysregulated in the damaged tissues. Furthermore, motif enrichment analysis of differentially accessible loci has identified many dysregulated transcription factors, the functions of which are known to be in chondrocyte development regulation. For example, the transcription factor Jun-AP1, which is enriched in the more accessible regions of the damaged tissues, is known to regulate chondrocyte hypertrophic morphology, which contributes to longitudinal bone growth, consistent with the notion of endochondral ossification⁵⁹. A recent study showed injection of adipose-derived stromal cells overexpressing an AP-1 family transcription factor Fra-1 can inhibit OA progression in mice⁶⁰. Consistently, in our ATAC-seq analysis, other members of AP-1 family (e.g. BTF, FOSL2) also showed enrichment in differentially accessible regions. In addition, if we combine the RNA-seq data¹⁵, 15 out of 18 transcription factor genes identified in the motif analysis are expressed (CPM \geq 2) in the cartilage tissue, of which *ETS1* is differentially expressed between the two cartilage types. These transcription factors could be candidate targets for further experiments.

In the integrative analysis of ATAC-seq and RNA-seq, many dysregulated genes related to lineage differentiation of MSC pathways were observed. For example, we found that *BMPR1B* (bone morphogenetic protein receptor type 1B) is upregulated and both its promoter and enhancer are more accessible in the damaged tissue. Its activation is consistent with the ossification pathway activation, since it encodes a transmembrane serine kinase that binds to BMP ligands that positively regulate endochondral ossification and abnormal chondrogenesis^{61,62}. Consistently, an osteoblast marker gene *MSX2* (Msh Homeobox 2) involved in promoting osteoblast differentiation^{63,64}, is upregulated and both its promoter and enhancer are more accessible in the damaged tissue, suggesting the osteoblast differentiation may be activated in OA. Furthermore, we found *ROR2* (receptor tyrosine kinase like orphan receptor 2) is down-regulated and its enhancers are less accessible in the damaged samples. Since it is required for cartilage development^{65,66}, it suggests that the normal chondrocyte development and cartilage formation may be compromised in OA. Consistently, we have also determined *FGFR2* and *STAT1*, which are known

to inhibit chondrocyte proliferation⁶⁷, are upregulated. In summary, our ATAC-seq data and integrative analyses support the OA model of abnormal MSC differentiation and endochondral ossification over other models, such as inflammatory mediums from the synovium.

The pathogenesis for OA is not yet fully understood, despite multiple genes and pathways that have been characterized to be dysregulated^{13,15,68–70}. In this study, by integrating clinically relevant epigenomic data with genetic and transcriptomic data, we provide multiple lines of evidence supporting a number of OA candidate genes and pathways that may be crucial to OA pathogenesis, which could potentially be used for clinical diagnostic or as therapeutic targets.

Conclusions

We present in this work an application of ATAC-seq in OA in a clinical relevant setting. The chromatin accessibility map in cartilage will be a resource for future GWAS and DNA methylation studies in OA and other musculoskeletal diseases. We identified altered promoters and enhancers of genes that might be involved in the pathogenesis of OA. Our analyses suggest aberrant enhancer usage associated with MSC differentiation and chondrogenesis in OA. Understanding these molecular basis of OA is necessary for future therapeutic intervention.

Methods

Knee joint tissues collection. Human knee joint tibial plates were collected from patients who undertook joint replacement surgery due to severe primary OA at the National Hospital Organization Sagamihara Hospital (Kanagawa, Japan). Demographic information for patients is listed in Supplementary Table S1. Diagnosis of OA was based on the criteria of the American College of Rheumatology⁷¹, and all the knees were medially involved in the disease. Tissues were stored in 4 °C Dulbecco's Modified Eagle Medium (DMEM) after removal for less than one hour prior to cartilage sectioning. Informed consent was obtained from each patient enrolled in this study. This study was approved by the institutional review board of all the participating institutions (Sagamihara National Hospital and RIKEN).

Processing of cartilage samples. Fresh cartilage was separated from the subchondral bone by a scalpel after surgery and stored in 4 °C DMEM for three hours until they were processed. 100 mg cartilage from both outer region of lateral tibial plateau (oLT) and inner region of medial tibial plateau (iMT) (Fig. 1a and Supplementary Figure S1a) for each patient was digested with 0.2% type II collagenase (Sigma-Aldrich) in DMEM at 37 °C with rotation for 12 hours to fully remove the collagen matrix debris, and stopped by washing with chilled PBS. Immediately after digestion, chondrocytes were counted by a cell counting chamber Incyto C-Chip (VWR) and collected by centrifugation at 500 g for 5 minutes at 4 °C.

ATAC-seq of the cartilage. Approximately 50,000 cells were taken after the cartilage processing and used for ATAC-seq library preparation (Fig. 1a) according to the original protocol^{32,72}. Briefly, transposase reaction was carried out as previously described³² followed by 9 to 12 cycles of PCR amplification. Amplified DNA fragments were purified with MinElute PCR Purification Kit (QIAGEN) and size selected twice with Agencourt AMPure XP (1:1.5 and 1:0.5 sample to beads; Beckman Coulter). Libraries were quantified by KAPA Library Quantification Kit for Illumina Sequencing Platforms (KAPA Biosystems), and size distribution was inspected by Bioanalyzer (Agilent High Sensitivity DNA chip, Agilent Technologies). Library quality was assessed before sequencing by qPCR enrichment of a housekeeping gene promoter (*GAPDH*) over a gene desert region⁷². ATAC-seq libraries were sequenced on an Illumina HiSeq 2500 (50 bp, paired-end) by GeNAS (Genome Network Analysis Support Facility, RIKEN, Yokohama, Japan).

ATAC-seq data processing, peak calling and quality assessment. An ATAC-seq data processing pipeline for read mapping, peak calling, signal track generation, and quality control was implemented⁷³. Briefly, fastq files for all patients were grouped by tissue compartment (oLT or iMT) and input into the pipeline separately, with parameters *-true_rep -no_idr*. Reads were mapped to the hg38 reference genome. Peaks for individual samples, as well as for pooled oLT and pooled iMT, were called by MACS2 with default parameters in the pipeline. Basic sequencing information and library quality metrics are listed in Supplementary Table S2. NucleoATAC was applied to infer genome-wide nucleosome positions and occupancy from the ATAC-seq data³⁵. Briefly, NucleoATAC were run for oLT and iMT separately with default parameters, using bam files merging from all patients as input, and the outputted *nucleoatac_signal.bedgraph* were used for aggregated plot around TSS.

Previously annotated DNase I Hypersensitive Sites (DHS) were used to assess the quality of our libraries. Briefly, a set of DHS defined in Roadmap Epigenomics Project⁷⁴ were downloaded (http://egg2.wustl.edu/roadmap/web_portal/DNase_reg.html) and lifted over from hg19 to hg38 using UCSC liftOver tool⁷⁵ (refer as Roadmap DHS thereafter). For each library, we calculated the ratio of reads that fall within Roadmap DHS versus randomly chosen genomic regions of the same total size (DHS enrichment score). It is noted that the DHS enrichment score is in good correlation with enrichment qPCR for *GAPDH* (Supplementary Figure S1b).

Defining a set of unified accessible chromatin regions. Peaks called from pooled oLT and pooled iMT were merged as raw peak regions using *bedtools v2.27.0*⁷⁶ (*mergeBed -d 300*). Reads fall in raw peak regions were counted for individual samples using *bedtools*⁷⁶. Read counts were normalized as count-per-million (CPM) based on relative log expression normalization (RLE) method implemented in *edgeR v3.18.1*^{41,77}. In the end, we defined a set robust peaks ($n = 109,215$) with ≥ 2 CPM in at least 6 oLT or 6 iMT samples for all downstream analyses.

Peak annotation and differentially accessible peak identification. Peaks were annotated as promoters or enhancers based on their intersection with promoter or enhancer as defined in Roadmap DHS²⁹. Noted that

a peak was preferentially annotated as a promoter if it intersects both a promoter and a proximal enhancer DHS ($n = 1,693$, 1.6% of total robust peaks). Differentially accessible peaks between damaged (i.e. iMT) and intact (i.e. oLT) tissues were identified using edgeR^{41,77}. Briefly, read counts were normalized by the effective library size with the RLE method and DHS enrichment score of each sample were incorporated as continuous covariates in the design matrix of the generalized linear model implemented in edgeR^{41,77}. The Benjamini-Hochberg adjusted p -values (i.e. false discovery rate, FDR) was taken to measure the extent of differential accessibility as used in Fig. 3b. A cutoff of $FDR \leq 0.05$ was used to define differentially accessible peaks.

Processing of trait associated SNPs and DMLs. GWAS lead SNPs of all phenotypes of OA with genome-wide significance $p \leq 1 \times 10^{-5}$ were obtained from GWASdb v2 (as of 20 Jun 2018)⁷⁸, NHGRI-EBI GWAS Catalog (as of 20 Jun 2018)⁷⁹, and a recent study⁸⁰, totaling 197 lead SNPs. SNPs in linkage disequilibrium (LD) with these lead SNPs in matched populations (i.e. proxy SNPs, with $r^2 \geq 0.8$ within of 500 kb) were searched using PLINK v1.90b5.4 (<https://www.cog-genomics.org/plink2>), based on the 1000 Genomes Project Phase 3 data obtained from MAGMA⁸¹. It results in 4,483 GWAS SNPs associated with OA, consisting of 167 independent loci with non-overlapping LD blocks. As a negative control, GWAS SNPs associated with Parkinson's disease were obtained in the same manner. Coordinates for these SNPs in hg38 were obtained from dbSNP Build 150.

Differentially methylated loci associated with OA were obtained from previous studies^{25,70} and all identified DML were pooled and mapped to hg38 ($n = 9,265$). Parkinson's disease associated DMLs as a negative control trait were obtained from a previous study⁴².

Definition of cell type-specific enhancers. Bed files for enhancer clusters with coordinated activity in 127 epigenomes, as well as the density of clusters per cell type, were obtained from online database (http://egg2.wustl.edu/roadmap/web_portal/DNase_reg.html). DNase I regions selected with $p < 1 \times 10^{-2}$, lifted over from hg19 to hg38 using UCSC liftOver tool. For each cell type, clusters with two-fold density than average (across all cell types) were defined as specific and pooled.

Enrichment analysis for differentially accessible peaks. The enrichment of GWAS SNPs and DML in differentially accessible peaks were assessed as described in previous studies^{82,83}. Briefly, trait associated (i.e. OA and Parkinson's disease) SNPs (i.e. lead and proxy SNPs) and DMLs are referred to as foreground positions. For SNPs, ten millions of randomly selected common SNPs in dbSNP Build 150 were used to define the background positions. For DML, all probes in Illumina Infinium HumanMethylation450K were used as the background positions. Enrichment of foreground SNPs/DML in each set of regions (i.e. differentially accessible peaks at various FDR cutoffs) was evaluated by first counting the number of the foreground and background positions intersecting these regions (refer as $observed_{fore}$ and $observed_{back}$). The counting was repeated for 100 rounds of permutation, i.e. regions of the same number and sizes shuffled into the masked genomic regions using bedtools shuffle⁷⁶ (refer as $shuffled_{fore}$ and $shuffled_{back}$). Masked genomic regions was defined as the whole genome excluding the annotated gaps. Only the 22 autosomes were included in analyses. The odds ratio of foreground signal enrichment for each round of permutation was calculated as $(observed_{fore}/observed_{back})/(shuffled_{fore}/shuffled_{back})$. As a negative control, for the trait, the analyses were repeated by replacing the foreground SNPs and DML with those associated with Parkinson's disease.

In Fig. 3e, the enrichment of cell type-specific enhancers in differentially accessible peaks ($FDR \leq 0.05$) were assessed by Fisher's exact test. Same number of peaks were randomly sampled from all robust peaks as controls.

Linking enhancer peaks to their potential target genes. A promoter peak was assigned to a gene if it intersected with the transcription start sites (TSS) of its transcript as defined in the FANTOM CAGE associated transcriptome⁸². We note that a promoter peak might be associated with multiple genes. An enhancer peak was defined as linked to a target gene if it overlapped an expression quantitative trait loci (eQTL) of the corresponding gene (GTEx V7, in any tissues with $p < 1 \times 10^{-5}$)⁸⁴, or is supported by putative enhancer-promoter linkage predicted by JEME method (<http://yiplab.cse.cuhk.edu.hk/jeme/based>) on FANTOM5 and Roadmap Epigenomics Project data³¹. Out of the 77,655 enhancer peaks, 36,228 of them were linked to at least one gene, of which 15,391 (42.5%) were assigned to one gene, 8,073 (22.3%) to two genes, and 12,764 (35.2%) to three or more genes.

Integration with publicly available transcriptome data. We reanalyzed an RNA-seq dataset (ArrayExpress E-MTAB-4304) from an independent OA patient cohort, in which the RNA was extracted from cartilage tissues of both iMT and oLT for 8 patients¹⁵. Read counts on transcripts were estimated by Kallisto v0.43.1⁸⁵ using default parameters on FANTOM CAGE associated transcriptome⁸². Estimated read counts of a gene, defined as the sum of the estimated read counts of its associated transcripts, were used as the input for differential gene expression analysis using edgeR v3.18.1. In total, 3,293 genes were defined as significantly differentially expressed between iMT and oLT ($FDR \leq 0.05$). A gene is defined as "consistently dysregulated both at the epigenomic and transcriptomic levels" when it is upregulated (or downregulated) in RNA-seq with more (or less) accessible promoters or enhancers in ATAC-seq ($n = 371$).

Gene ontology analysis. Enrichr⁵² was used to identify gene sets enriched in genes implicated in OA. The selected gene lists were input to query enrichment for gene ontology (Biological Process 2017b) in the database. Terms with p -value ≤ 0.05 and $FDR \leq 0.25$ were considered significant ($n = 421$ of 4437). Top 30 terms ranked by percentage of overrepresented genes are listed in Fig. 4d. All terms are listed in Supplementary Table S7. It is of note that terms with fewer number of genes tend to rank higher in this way.

Transcription factor binding motif analysis. DNA motif analysis for differentially accessible regions was performed using HOMER v4.9 with default parameters⁸⁶. The enriched *de novo* and known motifs as well as its matching transcription factor are searched and scanned in more- and less-accessible peaks separately (parameters: *-mask -size given*), with all robust peaks as background region. P-value $< 1 \times 10^{-5}$ was considered significant.

Ethics, consent and permissions. This study was approved by the ethics review board of all the participating institutions (Sagamihara National Hospital and RIKEN) and all experiments were carried out in accordance with their guidelines and regulations.

Consent to publish. Informed consent was obtained from each patient enrolled in this study.

Availability of Data and Materials

Raw and processed data generated are available on the Gene Expression Omnibus under the accession code GSE108301.

References

- Findlay, D. M. & Kuliwaba, J. S. Bone-cartilage crosstalk: a conversation for understanding osteoarthritis. *Bone Res.* **4**, 16028 (2016).
- Goldring, M. B. & Goldring, S. R. Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis. in. *Annals of the New York Academy of Sciences* **1192**, 230–237 (2010).
- Kwoh, C. K. In *The Epidemiology of Aging*, 523–536 (2012).
- Martel-pelletier, J. *et al.* Osteoarthritis. *Nat. Rev. Dis. Prim.* **2**, 16072 (2016).
- Krane, S. M., Quigley, J. & Krane, S. Petulant cellular acts: destroying the ECM rather than creating it. *J. Clin. Invest.* **107**, 31–32 (2001).
- Kapoor, M. In *Osteoarthritis*, 1–28 (Springer International Publishing, 2015).
- Man, G. S. & Mologhianu, G. Osteoarthritis pathogenesis - a complex process that involves the entire joint. *J. Med. Life* **7**, 37–41 (2014).
- Kuyinu, E. L., Narayanan, G., Nair, L. S. & Laurencin, C. T. Animal models of osteoarthritis: Classification, update, and measurement of outcomes. *Journal of Orthopaedic Surgery and Research* **11** (2016).
- Kawaguchi, H. The canonical Wnt signal paradoxically regulates osteoarthritis development through the endochondral ossification process. *Integr. Mol. Med.* **3**, 672–674 (2016).
- Dreier, R. Hypertrophic differentiation of chondrocytes in osteoarthritis: the developmental aspect of degenerative joint disorders. *Arthritis Res. Ther.* **12**, 216 (2010).
- Kawaguchi, H. Endochondral ossification signals in cartilage degradation during osteoarthritis progression in experimental mouse models. *Mol. Cells* **25**, 1–6 (2008).
- Cox, L. G. E., Van Donkelaar, C. C., van Rietbergen, B., Emans, P. J. & Ito, K. Alterations to the subchondral bone architecture during osteoarthritis: Bone adaptation vs endochondral bone formation. *Osteoarthr. Cartil.* **21**, 331–338 (2013).
- Uhalte, E. C., Wilkinson, J. M., Southam, L. & Zeggini, E. Pathways to understanding the genomic aetiology of osteoarthritis. *Human Molecular Genetics* **26**, R193–R201 (2017).
- Corradin, O. & Scacheri, P. C. Enhancer variants: Evaluating functions in common disease. *Genome Med.* **6**, 85 (2014).
- Dunn, S. L. *et al.* Gene expression changes in damaged osteoarthritic cartilage identify a signature of non-chondrogenic and mechanical responses. *Osteoarthr. Cartil.* **24**, 1431–1440 (2016).
- Chou, C. H. *et al.* Direct assessment of articular cartilage and underlying subchondral bone reveals a progressive gene expression change in human osteoarthritic knees. *Osteoarthr. Cartil.* **21**, 450–461 (2013).
- Reynard, L. N. Analysis of genetics and DNA methylation in osteoarthritis: What have we learnt about the disease? *Seminars in Cell and Developmental Biology* (2017).
- Fernández-Tajes, J. *et al.* Genome-wide DNA methylation analysis of articular chondrocytes reveals a cluster of osteoarthritic patients. *Ann. Rheum. Dis.* (2014).
- Rushton, M. D. *et al.* Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. *Arthritis Rheumatol* (2014).
- Den Hollander, W. *et al.* Knee and hip articular cartilage have distinct epigenomic landscapes: Implications for future cartilage regeneration approaches. *Ann. Rheum. Dis.* (2014).
- Den Hollander, W. *et al.* Transcriptional associations of osteoarthritis-mediated loss of epigenetic control in articular cartilage. *Arthritis Rheumatol* (2015).
- den Hollander, W. & Meulenbelt, I. DNA Methylation in Osteoarthritis. *Curr. Genomics*. <https://doi.org/10.2174/1389202916666150817212711> (2015).
- van Meurs, J. B. J. Osteoarthritis year in review 2016: genetics, genomics and epigenetics. *Osteoarthritis and Cartilage*, <https://doi.org/10.1016/j.joca.2016.11.011> (2017).
- Chou, C. H. *et al.* Insights into osteoarthritis progression revealed by analyses of both knee tibiofemoral compartments. *Osteoarthr. Cartil.* **23**, 571–580 (2015).
- Zhang, Y. *et al.* Genome-wide DNA methylation profile implicates potential cartilage regeneration at the late stage of knee osteoarthritis. *Osteoarthr. Cartil.* **24**, 835–843 (2016).
- Bernstein, B. E. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Herz, H. M. Enhancer deregulation in cancer and other diseases. *BioEssays* **38**, 1003–1015 (2016).
- Forrest, A. R. R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
- Bernstein, B. E. *et al.* The NIH roadmap epigenomics mapping consortium. *Nature Biotechnology* **28**, 1045–1048 (2010).
- Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nature Genetics* **45**, 580–585 (2013).
- Cao, Q. *et al.* Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. *Nat. Genet.* **49**, 1428–1436 (2017).
- Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **2015**, 21.29.1–21.29.9 (2015).
- Chou, C.-H. *et al.* Genome-wide expression profiles of subchondral bone in osteoarthritis. *Arthritis Res. Ther.* **15**, R190 (2013).
- Zhang, Y. *et al.* Identification of DNA methylation changes associated with disease progression in subchondral bone with site-matched cartilage in knee osteoarthritis. *Sci. Rep.* **6**, 34460 (2016).
- Schep, A. N. *et al.* Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res.* **25**, 1757–1770 (2015).
- Farh, K. K. H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
- Zhou, S. *et al.* Conditional Deletion of Fgfr3 in Chondrocytes leads to Osteoarthritis-like Defects in Temporomandibular Joint of Adult Mice. *Sci. Rep.* **6** (2016).

38. Tang, J. *et al.* Fibroblast Growth Factor Receptor 3 Inhibits Osteoarthritis Progression in the Knee Joints of Adult Mice. *Arthritis Rheumatol.* **68**, 2432–2443 (2016).
39. Iwasa, K. *et al.* PTEN regulates matrix synthesis in adult human chondrocytes under oxidative stress. *J. Orthop. Res.* **32**, 231–237 (2014).
40. Zeggini, E. *et al.* Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* **380**, 815–23 (2012).
41. 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).
42. Fernandez-Santiago, R. *et al.* Aberrant epigenome in iPSC-derived dopaminergic neurons from Parkinson's disease patients. *EMBO Mol. Med.* **7**, 1529–1546 (2015).
43. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
44. Kron, K. J., Bailey, S. D. & Lupien, M. Enhancer alterations in cancer: a source for a cell identity crisis. *Genome Med* **6**, 77 (2014).
45. Hematti, P. Mesenchymal stromal cells and fibroblasts: a case of mistaken identity? *Cytotherapy* (2012).
46. Vincenti, M. P. & Brinckerhoff, C. E. Transcriptional regulation of collagenase (MMP-1, MMP-13) genes in arthritis: integration of complex signaling pathways for the recruitment of gene-specific transcription factors. *Arthritis Res.* **4**, 157 (2002).
47. Okuma, T. *et al.* Regulation of mouse chondrocyte differentiation by CCAAT/enhancer-binding proteins. *Biomed Res* **36**, 21–29 (2015).
48. Hong, E., Di Cesare, P. E. & Haudenschild, D. R. Role of c-Maf in Chondrocyte Differentiation. *Cartilage* **2**, 27–35 (2011).
49. Hayashi, S. *et al.* p21 deficiency is susceptible to osteoarthritis through STAT3 phosphorylation. *Arthritis Res. Ther.* **17**, 314 (2015).
50. Iwamoto, M. *et al.* Transcription factor ERG and joint and articular cartilage formation during mouse limb and spine skeletogenesis. *Dev. Biol.* **305**, 40–51 (2007).
51. Iwamoto, M. *et al.* The role of ERG (ets related gene) in cartilage development. *Osteoarthr. Cartil.* **9**, S41–S47 (2001).
52. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
53. Huang, G., Chubinskaya, S., Liao, W. & Loeser, R. F. Wnt5a induces catabolic signaling and matrix metalloproteinase production in human articular chondrocytes. *Osteoarthr. Cartil.* **25**, 1505–1515 (2017).
54. Papatheanasiou, I., Malizos, K. N. & Tsezou, A. Low-density lipoprotein receptor-related protein 5 (LRP5) expression in human osteoarthritic chondrocytes. *J. Orthop Res* **28**, 348–353 (2010).
55. Ellman, M. B. *et al.* Fibroblast growth factor control of cartilage homeostasis. *Journal of Cellular Biochemistry* **114**, 735–742 (2013).
56. Wu, M., Chen, G. & Li, Y.-P. TGF- β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res.* **4**, 16009 (2016).
57. Hoffman, L. M. *et al.* BMP action in skeletogenesis involves attenuation of retinoid signaling. *J. Cell Biol.* **174**, 101–113 (2006).
58. Shepherd, C. *et al.* Functional characterisation of the osteoarthritis genetic risk residing at ALDH1A2 identifies rs12915901 as a key target variant. *Arthritis Rheumatol.* (2018).
59. He, X., Ohba, S., Hojo, H. & McMahon, A. P. AP-1 family members act with Sox9 to promote chondrocyte hypertrophy. *Development* **134**, 502 (2016).
60. Schwabe, K. *et al.* Inhibition of Osteoarthritis by Adipose-Derived Stromal Cells Overexpressing Fra-1 in Mice. *Arthritis Rheumatol.* **68**, 138–151 (2016).
61. Yoon, B. S. *et al.* Bmpr1a and Bmpr1b have overlapping functions and are essential for chondrogenesis *in vivo*. *Proc. Natl. Acad. Sci.* **102**, 5062–5067 (2005).
62. Li, J. & Dong, S. The signaling pathways involved in chondrocyte differentiation and hypertrophic differentiation. *Stem Cells Int.* **2016** (2016).
63. Matsubara, T. *et al.* BMP2 regulates osterix through Msx2 and Runx2 during osteoblast differentiation. *J. Biol. Chem.* **283**, 29119–29125 (2008).
64. Cheng, S. L., Shao, J. S., Charlton-Kachigian, N., Loewy, A. P. & Towler, D. A. Msx2 Promotes Osteogenesis and Suppresses Adipogenic Differentiation of Multipotent Mesenchymal Progenitors. *J. Biol. Chem.* **278**, 45969–45977 (2003).
65. Dickinson, S. & Hollander, A. The Wnt5a Receptor, Receptor Tyrosine Kinase-Like Orphan Receptor 2, Is a Predictive Cell Surface Marker of Human Mesenchymal Stem Cells with an Enhanced Capacity for Chondrogenic Differentiation. *Stem Cells* **26**, 2399–2407 (2017).
66. Schille, C., Bayerlová, M., Bleckmann, A. & Schambony, A. Ror2 signaling is required for local upregulation of GDF6 and activation of BMP signaling at the neural plate border. *Development* **143**, 3182–3194 (2016).
67. Karuppaiah, K. *et al.* FGF signaling in the osteoprogenitor lineage non-autonomously regulates postnatal chondrocyte proliferation and skeletal growth. *Development* **143**, 1811–1822 (2016).
68. Shen, J., Abu-Amer, Y., O'Keefe, R. J. & McAlinden, A. Inflammation and epigenetic regulation in osteoarthritis. *Connective Tissue Research* **58**, 49–63 (2017).
69. Usami, Y., Gunawardena, A. T., Iwamoto, M. & Enomoto-Iwamoto, M. Wnt signaling in cartilage development and diseases: Lessons from animal studies. *Lab. Investig.* **96**, 186–196 (2016).
70. Steinberg, J. *et al.* Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. *Sci. Rep.* **7** (2017).
71. Altman, R. *et al.* Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. *Arthritis Rheum.* **29**, 1039–1049 (1986).
72. Milani, P. *et al.* Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells. *Sci. Rep.* **6**, 25474 (2016).
73. Lee, J. *et al.* Kundajelab/Atac_Dnase_Pipelines: 0.3.0 (2016).
74. Boyle, A. P. *et al.* High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* **132**, 311–322 (2008).
75. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
76. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
77. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–97 (2012).
78. Li, M. J. *et al.* GWASdbv2: an update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.* **44**, 869–876 (2016).
79. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
80. Zeggini, E. *et al.* Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. *Nat. Genet.* (2018).
81. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* (2015).
82. Hon, C. *et al.* An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* **543**, 199–204 (2017).
83. Iyer, M. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* (2015).

84. Consortium, G. Te. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* (80-.). **348**, 648–660 (2015).
85. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
86. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–89 (2010).
87. Panoutsopoulou, K. *et al.* Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. *Ann. Rheum. Dis.* (2011).
88. Evangelou, E. *et al.* A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. *Ann. Rheum. Dis.*, <https://doi.org/10.1136/annrheumdis-2012-203114> (2014).

Acknowledgements

We thank Dr. Shiro Ikegawa (RIKEN & Tokyo University) for his support on the project. We thank Yanfei Zhang (Geisinger), Tommy Terooatea (RIKEN), Mitsunori Yahata (RIKEN), Haruka Yabukami (RIKEN) and Prashanti Jeyamohan (RIKEN) for active discussions, technical assistance and support. We thank RIKEN GenAS for the sequencing service. This work was supported by a Research Grant from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) to the RIKEN Center for Life Science Technologies, RIKEN Epigenetics Program, and Ye Liu was supported by the RIKEN Junior Research Associate Program.

Author Contributions

The study was conceived and designed by A.M. and M.T.M.L. Sample collection were done by N.F. and N.T. Experiments were performed by Y.L. Analysis and interpretation of data were carried out by Y.L., J.C. and C.H. The study was supervised by A.M., M.T.M.L. and Z.Z. Y.L. and J.C. drafted and A.M., C.H. and M.T.M.L. revised the manuscript. All authors read and approved the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-33779-z>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018