

Lineage-Specific Differences and Regulatory Networks Governing Human Chondrocyte Development

Authors:

Daniel Richard^{1*}, Steven Pregizer^{2,3*}, Divya Venkatasubramanian^{2,3}, Pushpanathan Muthuirulan¹, Zun Liu¹, Terence D. Capellini^{1,4}, April M. Craft^{2,3,5,#}

Affiliations:

1 Human Evolutionary Biology, Harvard University, Cambridge, MA USA

2 Department of Orthopedic Research, Boston Children's Hospital, Boston, MA USA

3 Department of Orthopedic Surgery, Harvard Medical School, Boston, MA USA

4 Broad Institute of MIT and Harvard, Cambridge, MA USA

5 Harvard Stem Cell Institute, Cambridge, MA USA

*These authors contributed equally to this work

#Correspondence: april.craft@childrens.harvard.edu

Abstract

To address large gaps in our understanding of the molecular regulation of articular and growth plate cartilage development in humans, we used our directed differentiation approach to generate these distinct cartilage tissues from human embryonic stem cells. The resulting transcriptomic profiles of hESC-derived articular and growth plate chondrocytes were similar to fetal epiphyseal and growth plate chondrocytes, with respect to genes both known and previously unknown to cartilage biology. With the goal to characterize the regulatory landscapes accompanying these respective transcriptomes, we mapped chromatin accessibility in hESC-derived chondrocyte lineages, and mouse embryonic chondrocytes, using ATAC-sequencing. Integration of the expression dataset with the differentially accessible genomic regions revealed lineage-specific gene regulatory networks. We validated functional interactions of two transcription factors (RUNX2 in growth plate chondrocytes and RELA in articular chondrocytes) with their predicted genomic targets. The maps we provide thus represent a framework for probing regulatory interactions governing chondrocyte differentiation. This work constitutes a substantial step towards comprehensive and comparative molecular characterizations of distinct chondrogenic lineages, and sheds new light on human cartilage development and biology.

Introduction

Cartilage is a crucial component of the musculoskeletal system, providing structure and functioning in various capacities to support pain-free movement. Chondrocytes are the cells that produce and maintain the collagen- and proteoglycan-rich extracellular matrix (ECM) of cartilage tissues throughout the body. In the appendicular skeleton, chondrocytes give rise to growth plate cartilage, a transient tissue that provides a template for elongation of endochondral bones, and articular cartilage, a permanent tissue that covers joint surfaces to allow for frictionless joint movement. Articular cartilage arises from the epiphyseal cartilage at the ends of the developing bones. Abnormal development of growth plate or articular chondrocytes can result in chondro- or skeletal dysplasias, while injury to and aging of chondrocytes can contribute to the development of joint degeneration (i.e., osteoarthritis). Pharmaceutical or gene-based treatments for the majority these skeletal ailments are inadequate or simply do not exist due to the vast gaps in our knowledge regarding molecular mechanisms that govern the differentiation of chondrocytes into these two distinct lineages, especially in humans.

Much pioneering work focused on understanding the molecular regulation of chondrogenesis was performed *in vivo* using the mouse as a model system. A typical framework for these experiments involved genetic manipulation of a gene, either in the germline or conditionally, followed by careful phenotyping and gene expression readouts (e.g., *in situ* hybridization, immunohistochemistry, or quantitative polymerase chain reaction) to assess the effect on chondrocyte development. This paradigm led to the identification of signaling pathways such as TGF β /BMP (reviewed here¹), IHH^{2,3}, and PTHrP^{4,5}, as well as transcription factors such as Sox5/6/9^{6,7}, Runx2/3⁸⁻¹⁰, MEF2C¹¹, HIF2 α ¹², and FOXA2/3¹³ that are important for development growth plate and/or articular chondrocytes. Mouse models have indeed substantially contributed to our evolving understanding of gene functions and related diseases, as human mutations often recapitulate murine phenotypes and vice versa.

A growing number of studies within the past several decades have attempted to build on these seminal findings by exploring gene expression and gene regulatory mechanisms in chondrocytes on a broader scale. A significant number of these studies have focused on the Sox family, especially Sox9¹⁴⁻¹⁸. Early studies employed then-emerging technologies such as ChIP-chip¹⁸ and expression microarrays¹⁹⁻²⁵, while later studies incorporated ChIP-seq and/or RNA-seq^{14-17,26-30} as well as ATAC-seq³¹. Chondrocytes used for these experiments were derived from a variety of sources including rib or epiphyseal cartilage from embryonic, neonatal or juvenile rodents^{19-26,31-34}, as well as a rat chondrosarcoma cell line^{16,18}. These cells all serve as generally good models of growth plate development, but not necessarily articular chondrocyte development. Modeling articular chondrocyte development using rodent-derived cells is inherently challenging due to limited availability of the source tissue and failure of the cells to retain their phenotype during expansion in culture. Chondrocytes from neonatal bovine articular cartilage are more plentiful and have been used in at least one gene-regulatory study³⁰. A handful of studies have used chondrocytes derived from human sources, including fetal epiphyseal cartilage^{27,28} and mesenchymal stromal cell-derived cartilage^{28,29}. While helpful for illuminating nuances of human growth plate chondrocyte development, and albeit understandably less so for articular

chondrogenesis, it is clear that we still lack a complete understanding of how distinct chondrogenic lineages are specified and maintained in humans.

Using a directed differentiation approach inspired by embryonic chondrogenesis, we differentiate human pluripotent stem cells (i.e., hESCs/iPSCs) into growth plate and articular chondrocytes³⁵. Following induction of appropriate mesoderm and mesenchymal-like progenitors, chondrogenesis is induced in a high-density micro-mass format, eventually producing disks of cartilage tissues approximately 1 cm in diameter and 1-3 mm thick. Long-term culture of the micromasses with TGF β 3 results in the generation of articular-like cartilage tissue, while a transition to long-term treatment with BMP4 results in growth plate-like cartilage tissue. We previously defined 12 weeks as the end-stage of this micromass protocol, where the cells and tissues exhibit key characteristics of their *in vivo* counterparts, including the morphology and size of the cells, tissue/zonal organization, proteoglycan content, and expression of candidate cell-type specific markers such as lubricin (encoded by *PRG4*) and type X collagen (encoded by *COL10A1*). Lubricin, produced by superficial zone chondrocytes in articular cartilage, provides boundary lubrication and reducing friction between articulating cartilage surfaces³⁶. Type X collagen is specifically expressed by hypertrophic chondrocytes in growth plate cartilage. This directed differentiation platform thus provides an opportunity to investigate human chondrogenesis and cartilage development *in vitro*.

We used our established *in vitro* hPSC-based model of developing human cartilage to address the critical gaps in our knowledge of human cartilage development, with an important goal of providing a comprehensive guide of gene regulatory mechanisms governing articular versus growth plate chondrocyte cell fate. In this present study, we performed bulk RNA-sequencing (RNA-seq) in hESC-derived articular and growth plate cartilage to identify lineage-specific gene expression, and compared them to developing fetal cartilage. We also performed ATAC-sequencing (ATAC-seq) to define the accompanying regulatory landscapes in hESC-derived chondrogenic lineages and in mouse chondrocytes isolated by cell sorting^{31,37}. Integrating the transcriptomic and epigenetic datasets suggests gene regulatory networks specific to the growth plate or articular chondrocyte lineages. For two such networks, RUNX2 and RELA, and we provide evidence of transcription factor interaction with predicted target gene regulatory elements, validating the predictive gene regulatory networks uncovered by our analyses of these two distinct human chondrogenic lineages.

Results

Transcriptomic profiles of hESC-derived chondrocytes recapitulate those of human fetal chondrocytes

Having differentiated hESCs to produce articular chondrocytes and growth cartilage chondrocytes (Fig. 1A), we performed bulk RNA-seq on these cartilage tissues after 12 weeks in culture. To serve as *in vivo* references for the *in vitro* cartilage tissues, we performed bulk RNA-seq on human distal femur articular and growth plate chondrocytes isolated from embryonic day (E)67 fetal donor tissue. Principal component analysis indicated the 4 different chondrocyte sources (hESC-derived growth plate cartilage, hESC-derived articular cartilage, fetal growth plate cartilage, and fetal articular articular) clustered separately (Fig. 1B). Reassuringly, even though hESC differentiation and RNA-seq was performed on more than one occasion (4 independent differentiations

and 3 sequencing batches), transcriptome clustering was primarily dependent on cell type. That is, hESC-derived articular (orange icons) or growth plate tissues (light blue icons) from different experiments and batches cluster together, indicating reproducibility within the hESC model system. Principle component 1 (PC1) generally represents differences observed between the *in vitro* and *in vivo* samples (Fig. 1B, circles/triangles vs. squares), including minor contributions from sex-linked genes (hESC-derived samples are female and the fetal donor tissue was male). The top genes contributing to PC1 are enriched in GO biological processes that, with the exceptions of extracellular matrix/structure organization and immune responses, are detecting cellular responses to ions (Fig. S1A), suggesting differences associated with culturing cells in media not identical to milieu *in vivo*. PC2 represents differences between articular/epiphyseal and growth plate cartilages, which were more pronounced for hESC-derived chondrocytes than for their *in vivo* counterparts, as indicated by the greater distance in separation along the PC2 axis. The top genes contributing to PC2 are enriched in GO biological processes that are consistent with cartilage and skeletal system development and morphogenesis (Fig. S1B).

The top 40 differentially expressed genes (DEGs) between respective cartilage lineages, *in vitro* or *in vivo*, are shown in Fig. 1C-D, respectively (all DEGs are provided in Table S1). We performed gene-set enrichment analyses on the set of genes upregulated in hESC-derived articular or hESC-derived growth plate cartilage. The former was enriched for terms relating to ECM organization, response to TGF stimulus, and collagen processes, while the latter was enriched for terms relating to ossification, ECM organization and cartilage development (Table S1). We obtained similar enrichment terms when we performed the same analysis on genes upregulated in fetal epiphyseal or fetal growth plate cartilage (Table S1). Of the top 200 genes with the highest degree of differential expression between hESC-derived articular and growth plate chondrocytes, >70% exhibited similar differential expression in fetal epiphyseal and fetal growth plate chondrocytes, a trend that was statistically significant (Fig. 1E, compare top and bottom graphs; $p = 2.8e-16$). Notable genes whose expression in fetal cartilage is opposite that of hESC-derived cartilage are *MMP13*, a collagenase often observed in pathogenic osteoarthritic cartilage, and *SCARA5*, a dexamethasone-responsive gene implicated in adipogenesis³⁸. We obtained similar results in expression trends when starting with the top 200 genes with the highest degree of differential expression between fetal epiphyseal and fetal growth plate chondrocytes, and when considering those DEGs whose difference is most significant from either dataset (Fig. S1). Taken together, these data lend strong support to the notion that the two chondrogenic cell types derived from hESCs represent *bona fide* articular and growth plate chondrocytes.

Differentially expressed transcripts localize to specific regions within hESC-derived and fetal articular and growth plate cartilages

RNA-seq differences between fetal and hESC-derived cartilages may represent differences in the relative abundance of specific chondrocyte subtypes between the two samples. For example, mature articular cartilage has superficial, intermediate, deep zone and calcified chondrocytes, but fetal epiphyseal cartilage is less mature, and contains surface chondrocytes that will give rise to the future articular cartilage as well as

chondrocytes that will contribute to the secondary ossification center (future growth plate chondrocyte). Growth plate cartilage has resting, proliferating, pre-hypertrophic and hypertrophic chondrocytes. But as the developing cartilage is a continuous unit *in vivo*, the fetal epiphyseal chondrocytes and fetal growth plate both contain a portion of resting and proliferating chondrocytes, indicated by the approximate dissection point and dashed circle in Fig. 1A. Therefore, we used *in situ* hybridization to localize sites of differential gene expression in hESC-derived and fetal cartilages, and to estimate the fraction of cells expressing those transcripts (Figure 2). Type II collagen, encoded by the gene *COL2A1*, is a major structural component of both articular and growth plate cartilage, and as such, expression is observed in the cartilaginous structures both *in vitro* and *in vivo* (Fig. 2A-D). *PRG4* is expressed in the superficial layer of the hESC-derived TGFB-treated articular cartilage (Fig. 2A) and absent in the BMP4-treated growth plate cartilage (Fig. 2B). Similarly, *in vivo*, *PRG4* is expressed in the superficial zone of fetal articular cartilage, as well as the intra-articular ligaments and meniscus (Fig. 2C), and is absent in the growth plate (Fig. 2D). Tenomodulin (*TNMD*), a well-known marker of tendon fate³⁹, was a top DEG in hESC and fetal articular chondrocytes, where it is found in cells at the most superficial layers of the hESC-derived articular cartilage and the fetal knee cartilage (Fig. 2A, C, white arrows). *TNMD* is also expressed in the intra-articular ligaments *in vivo*, as expected (Fig. 2C double arrow). *COL10A1* mRNA is detected in the hESC-derived growth plate cartilage (Fig. 2F), but not hESC-derived articular cartilage (Fig. 2E), consistent with expression patterns found in the fetal knee, where *COL10A1* is expressed in the hypertrophic chondrocytes of the growth plate (Fig. 2H) but not in the epiphyseal chondrocytes (Fig. 2G). Thus, lower fold-change differences in genes (e.g., *PRG4*) in the fetal cartilages compared to the hESC-derived cartilages may reflect the incomplete terminal differentiation in the fetal cartilage, while others (e.g., *COL10A1*) may reflect higher proportions of hypertrophic chondrocytes among the cells recovered from hESC-derived growth plate cartilage compared to the fetal growth plate.

Additional DEGs were validated by qPCR in hESC-derived chondrocytes from 5 additional independent differentiations (Fig. 2I), and fetal chondrocytes from the distal femur and proximal tibia of three developmental timepoints (E59 (Carnegie Stage 23), E67 and E72; Fig 2J). As predicted by previous studies^{4,40-46}, expression levels of *FGF18* and *PTHLH* are significantly higher in hESC-derived articular chondrocytes, and levels of *FGFR3*, *PTH1R*, *PANX3*, and *ALPL* are significantly higher in the hESC-derived growth plate chondrocytes. Similar patterns are observed in fetal chondrocytes.

The transcriptomic data from hESC-derived cartilages identified genes that were not previously implicated in cartilage development in addition to confirming those that have been. We found *chitinase-3 like protein 1* (*CHI3L1*, also known as *YKL-40*) and mesenchyme homeobox 1 (*MEOX1*) to be top DEGs in the articular cartilage lineage, and confirmed their lineage-restricted expression *in vitro* and *in vivo* (Fig. 1I-J). *CHI3L1* expression has been described in cultured chondrocytes and osteoarthritic cartilage^{47,48}, however *MEOX1* is most well-known for its role in somitogenesis and axial skeleton formation⁴⁹. *Type XV collagen* (*COL15A1*), differentially expressed in hESC-derived articular cartilage tissues but not fetal cartilages, is a non-fibrillar basement membrane-associated collagen previously detected in the perichondrium and in mesenchymal stem cells undergoing osteogenic differentiation^{50,51}. The highest level of type XV collagen protein was localized in

the matrix of the superficial zone of hESC-derived articular cartilage, and we detected intracellular staining in the deeper zone of the articular cartilage and in some cells within the growth plate cartilage (Fig. 2M). We also localized Type XV collagen in developing human phalangeal (E70) and knee joints (E59), where we found it to be in the matrix of the epiphysis of the metacarpophalangeal joint, and at the surface of the knee joint cartilages, but absent in the matrix surrounding hypertrophic chondrocytes of the growth plates. These data suggest *COL15A1* expression may be specific to the superficial zone of articular cartilage. *EF-hand domain-containing protein 1 (EFHD1)* expression was significantly higher in both hESC-derived and fetal growth plate chondrocytes. EFHD1 is a calcium-binding protein localized to the inner mitochondrial membrane, previously undescribed in cartilage⁵². EFHD1 protein was localized to the cytoplasm of BMP4-treated hypertrophic chondrocytes, and hypertrophic chondrocytes in the fetal growth plates, but not in articular or epiphyseal cartilage, as predicted (Fig. 2N). These data indicate *EFHD1* is specifically expressed in hypertrophic chondrocytes of the growth plate.

The identification of known lineage-specific expression in the distinct cartilage tissues we generated from hESCs validated the utility of our established *in vitro* differentiation methods, and further illustrated the strength of our transcriptomic datasets in identifying novel markers and potential regulators of articular and growth plate cartilage.

Chromatin accessibility differences between hESC-derived articular chondrocytes and hESC-derived growth plate chondrocytes

As cell fate decisions are guided by transcriptional regulation, we sought to more deeply investigate the expression of transcription factors (TFs) and potential gene regulatory elements within the hESC-derived cartilage tissues. From our initial differential gene expression analyses, we identified 277 transcription factors (TFs) that were differentially expressed in at least one of the four cell types profiled and for which a binding motif has been described (Table S1, Fig. S2). To refine this list of potential chondrogenic lineage regulators, we performed ATAC-seq⁵³, a method used to characterize chromatin accessibility on a genome-wide basis, on a subset of hESC-derived chondrocytes that were used for transcriptomic analysis (i.e., the three biological replicate tissues per treatment sequenced in batch 2). To establish a set of evolutionarily conserved and therefore likely functional regulatory elements in chondrogenesis, we also generated ATAC-seq data from mouse embryonic chondrocytes expressing *Col2a1* (a general chondrocyte marker) or hypertrophic growth plate chondrocytes expressing *Col10a1*. *Col2a1*+ or *Col10a1*+ chondrocytes were isolated from E15.5 transgenic mice harboring fluorescent reporters driven by *Col2a1* or *Col10a1* regulatory elements using cell sorting (see Methods and previous work³⁷). The genome-wide overlap of peaks found in the two types of human and mouse chondrocytes is summarized in Table 1. *Col2a1*+ chondrocytes may contain chondrocytes that also express *Col10a1*, and it is expected that *Col10a1*+ chondrocytes also express *Col2a1*, but this population is more restricted in its hypertrophic lineage.

Profiling the hESC-derived chondrocytes by ATAC-seq and calling significant reproducible open-chromatin regions (i.e., peaks) revealed a total of 37,780 unique peaks, corresponding to putative regulatory elements. We categorized these regions on the basis of differential accessibility in either growth plate or articular chondrocytes, identifying 12,154 regions more accessible in growth plate chondrocytes and 11,571 more accessible in articular chondrocytes (Table S2). These differentially accessible regions (DARs) suggest cell-type specific regulatory activity and are the focus of subsequent analyses. To illustrate these data, the top 20 DEGs for each lineage in this subset of hESC-derived tissues, the accessibility of their corresponding promoters, and their respective *cis*-regulatory score (CRS, see below and *Methods*) are shown in Figure 3A. DARs identified near the *IHH* and *CHI3L1* loci in hESC-derived growth plate chondrocytes and hESC-derived articular chondrocytes are indicated by black rectangles in the 'BMP Diff Peaks' track, orange dashed line and the 'TGF Diff Peaks' track, blue dashed line, respectively (Fig. 3B-C). Tracks showing accessible regions detected in mouse embryonic chondrocytes are also shown (outlined with purple dashed lines) to visualize those regulatory elements that are conserved between species (examples indicated with arrows). Using the GREAT region-based association tool⁵⁴, we identified terms significantly associated with DARs from growth plate chondrocytes, including *anomaly of the limb diaphyses* and *ECM organization*. Likewise, we identified terms associated with DARs from articular chondrocytes, including *ECM organization*, *collagen metabolic process*, and *osteoarthritis* (Table S2).

A simplistic model of gene expression is not sufficient to explain the gene regulatory network information captured by this merged transcriptomic and epigenetic data approach

To begin to understand differential transcriptional regulation mechanisms in these two lineages, we used *de novo* motif analysis to identify over-represented TF motifs in DARs specific to either articular or growth plate chondrocytes (Table S2). We detected RNA expression for 15 TFs whose motifs are enriched in BMP DARs, but only 5 TFs were differentially expressed (Table S1, considering all batches). *FOSL2* and *PITX2* were expressed at significantly higher levels in the corresponding BMP-treated growth plate lineage, while *SOX11*, *FOXA1* and *RUNX1* were expressed at significantly higher levels in the opposite lineage (Tables S1 and S2). We detected RNA expression for 19 TFs whose motifs are enriched in TGFβ DARs, but again only 5 of these were differentially expressed (Tables S1, considering all batches). *ETV4*, *AP4* and *NFY-B* were expressed at significantly higher levels in the corresponding TGFβ-treated articular cartilage lineage, while *NFAT5* and *NHLH1* were expressed at significantly higher levels in the opposite lineage (Tables S1 and S2). Thus, the majority of motifs identified in these DARs were not for TFs that were also differentially expressed in the corresponding cell type.

We then examined the same two sets of DARs specifically for enrichment of motifs belonging to TFs differentially expressed in the corresponding cell types. This yielded a reduced set of TFs, several of which were also observed in our *de novo* analysis (Fig. S4, Table S2). For this latter approach we confirmed that motif enrichment is not substantially correlated with sequence complexity (Fig. S5). When we considered sets of lineage-specific DARs nearby genes exhibiting lineage-specific expression, we observed similar

enrichments for motif occurrences of several of these TFs in both region (lineage) sets, despite conditioning on lineage-specific expression of these factors (Fig. S4, right). For example, motifs for the top DE TFs, *POU2F2* (a TGFB-specific DE TF) and *MEF2C* (a BMP-specific DE TF) were significantly enriched in both TGFB-DARs and BMP-DARS when compared to randomly-sampled sequence sets (red lines, Fig. S4, right). Motifs for only two of these DE TFs, *RUNX2* and *RUNX3* were significantly enriched in the corresponding BMP lineage, and significantly depleted or not significant in the TGFB-DARs. This suggests that a simplistic model of gene expression, wherein upregulation of a given TF is associated with increased accessibility of elements to which it may bind, and subsequently increased expression of its putative targets, is not sufficient to explain the gene regulatory network information captured by our ATAC-seq/RNA-seq strategy.

Defining hypothetical gene regulatory networks via cataloging gene expression and chromatin accessibility differences between hESC-derived articular chondrocytes and hESC-derived growth plate chondrocytes

We sought to integrate our ATAC- and RNA-seq datasets in a way that better captured the regulatory behavior described in our sequencing datasets. Our approach defined three metrics of expression and accessibility at a given locus: 1) gene expression, 2) proximal (promoter) accessibility, and 3) distal (enhancer) activity, defined as a *cis*-regulatory score (see Methods). Based on the simplistic model of gene expression described above, we would have expected absolute correspondence between these three metrics for all DEGs, however there were clear deviations from this result (Figs. 3A and 4A). We reasoned that multiple regulatory principles may be at play and, inspired by recent work describing the *cis*-regulatory behavior of immunological genes in mice⁵⁵, we classified genes into four different regulatory behaviors based on the proportion of variance in expression explained by chromatin accessibility within their respective loci. Briefly, these consist of genes whose expression variance is best explained by: variance does not clearly associate with chromatin accessibility ('unexplained', cluster 1), a combination of promoter accessibility and distal *cis*-regulatory accessibility ('combo-centric', cluster 2), promoter accessibility alone ('promoter-centric', cluster 3), or distal *cis*-regulatory accessibility alone ('enhancer-centric', cluster 4; Fig. 4B). In general, genes falling into clusters 2-4 exhibited larger fold-changes in expression between articular and growth plate chondrocytes compared to genes falling into the 'unexplained variance' category (cluster 1; Fig. 4C). Likewise, a greater proportion of genes from clusters 2-4 (genes whose variance in expression can be attributed to promoter or enhancer accessibility or both) were differentially expressed, compared to those from cluster 1 (whose variance cannot be attributed to differentially accessibility in any putative regulatory elements, Fig. 4C). Further, we confirmed that sets of genes segregated with this method show increased sharing of direction (i.e., lineage bias) for the expected parameters (e.g., 'combo-centric' gene expression had a greater correspondence with our *cis*-regulatory bias metric than did 'promoter-centric' gene expression) (Fig. 4D).

We next looked for TF motif enrichment in the DARs of genes belonging in clusters 2-4 (i.e., combo-centric, promoter-centric, and enhancer-centric genes) with the following initial restrictions: 1) motifs were only considered for TFs that were differentially expressed between hESC-derived articular and growth plate chondrocytes (TGFB=124, BMP=83; Table S1; and S2) enrichment of each motif was only considered for

DARs, or promoters, in which the direction of accessibility (growth plate vs. articular) matched the direction of expression (growth plate vs. articular) (Table S3). For each motif demonstrating enrichment according to these criteria, we then validated whether enrichment could not be significantly detected in the set of DARs/promoters for which the direction of accessibility was opposite to the direction of expression. This approach yielded a small number of motifs enriched in either promoter or enhancer sequences from cluster 2-4 genes, and that were biased towards articular chondrocytes or growth plate chondrocytes (Table S3).

We focused on those motifs most highly enriched in DARs from enhancer-centric (cluster 4) and combo-centric (cluster 2) genes (Fig. 5), as these groups exhibit the strongest trends in differential expression across lineages (Fig. 4C). We identified 7 TFs in the hESC-derived articular cartilage lineage whose expression and motif accessibility within enhancer-centric DEGS (Fig. 5A) or combo-centric DEGs (Fig. 5B) are significant and lineage-specific. *ETV1*, *FL1*, *RELA*, and *NFKB1* motifs are specifically enriched in both enhancer-centric and combo-centric gene DARS, while *RFX1* and *RFX2* motifs are enriched in enhancer-centric gene DARS, and *ATF7* motifs are enriched in combo-centric gene DARS. We used *RELA* as an example to illustrate these data. *RELA* motifs are significantly enriched in TGFB-specific combo-centric gene DARS ($p = 9.41e-6$), and significantly depleted in BMP-specific combo-centric gene DARS ($p = 0.0098$, Fig. 5C). In the hESC-derived growth plate lineage, we identified 7 TFs whose expression and motif accessibility within enhancer-centric DEGS (Fig. 5D) or combo-centric DEGs (Fig. 5E) are significant and lineage-specific. *CEBPB* and *RUNX2* motifs are specifically enriched in both enhancer-centric and combo-centric gene DARS, while *FOXF2*, *FOXO4* and *FOXA2* motifs are enriched in enhancer-centric gene DARS, and *DLX5* and *EMX2* motifs are enriched in combo-centric gene DARS. Visualizing these results, *RUNX2* motifs are significantly enriched in BMP-specific combo-centric gene DARS ($p = 4.03e-53$), but not significantly enriched in TGFB-specific combo-centric gene DARS (Fig. 5F).

Functional validation of TF and target interactions in human chondrocytes

To functionally validate putative gene regulatory interactions we identified in these studies, we performed ChIP-qPCR for several enhancer and promoter elements assigned to DEGs that have putative binding sites for *RELA* or *RUNX2* in hESC-derived chondrocytes (Tables 2 and 3). *RELA* and *NFKB1* are members of the same transcriptional complex. Having identified both of them in these conservative analyses, we postulated they have a co-operative functional role in articular chondrocyte biology, and chose *RELA* as the differential p-value between the specificity of motif enrichments between hESC-derived articular cartilage DARs and growth plate DARs was higher than *NFKB1*. We also chose to investigate downstream *RUNX2* targets in the growth plate cartilage lineage, as we wished to complement studies performed in the osteoblast lineages and in mice^{56,57} with human-specific data. We cross-referenced putative binding sites with ATAC-seq data collected from E15.5 mouse *Col2a1+* and *Col10a1+* chondrocytes, and published ChIP-seq data for several cell types (Table S4). We chose 7-10 targets that satisfied some or all of these criteria, choosing some targets that have been previously described in chondrocyte biology, and others with binding sites that have overlapping ChIP-seq peaks in other cell types.

Seven putative RELA target loci (Table 2) were chosen to confirm by ChIP-qPCR, including several genes known to be involved in articular cartilage identity and maintenance. These include *PRG4* (lubricin), a functional marker for the superficial zone of articular cartilage; *LOXL2* (lysyl oxidase-like 2), which induces anabolic gene expression and plays a potential protective role against OA⁵⁸; *DKK3* (Dickkopf-3), a noncanonical member of the Dkk family of Wnt antagonists that plays a role in articular cartilage maintenance⁵⁹ (PMID: 26687825); and *TLR2* (Toll-like receptor 2), which mediates articular cartilage homeostasis⁶⁰. We also chose to validate targets that are less well-studied, or newly identified in articular cartilage, including *LTBP2*, *COL15A1* (validated in Fig. 2M), and *GLIPR2*. Representative binding regions with RELA motifs are the *GLIPR2* promoter (Fig. 6A, overlapping with RELA ChIP-seq data and overlap with Col2a1+ mouse chondrocytes, indicated dotted box and red arrows) and an upstream enhancer of *LOXL2* (Fig. 6B, overlapping with RELA ChIP-seq data and histone acetylation peaks). *RELA* and these putative target genes are expressed at significantly higher levels in hESC-derived articular cartilage (Fig. 6C-D). The majority of RELA target genes were also expressed at significantly higher levels in the fetal epiphyseal chondrocytes compared to fetal growth plate chondrocytes (Table 1 and data not shown), however *RELA*, *COL15A1*, and *LOXL2* were not DEGs, likely due to under-represented terminally differentiated chondrocytes, and the overlap of unspecialized developing chondrocytes in both primary samples (Fig. 1A, Fig. 2). Notably, while *COL15A1* is not a DEG in the fetal chondrocytes, its protein expression appears higher in the matrix of the fetal epiphysis and superficial layers, compared to the matrix of the fetal growth plate (Fig. 2K).

Following ChIP-mediated pulldown of genomic regions bound by RELA in TGFB-treated articular cartilage, four of the six loci were enriched at least 5-fold (*PRG4*, *GLIPR2*, *DKK3*, *LOXL2*, *LTBP2*), and a fifth locus (*TLR2*) was enriched at least 2-fold compared to the negative control in the TGFB-treated articular cartilage (Table 2 and Fig. S6). The binding region for the *COL15A1* locus was only 1.3-fold enriched compared to the negative control in hESC-derived TGFB-treated articular cartilage, which suggests that RELA was not sufficiently bound to this locus in this sample.

Ten putative RUNX2 targets (Table 3) were chosen to confirm by ChIP-qPCR, including genes known to be important for chondrocyte and growth plate biology, including *ACAN* (Aggrecan), an essential proteoglycan in the extracellular matrix of both articular and growth plate cartilage^{61,62}; *COL10A1* (Type X collagen), a marker of hypertrophic chondrocytes important for endochondral bone formation⁶³; *WNT10B*, a Wnt family ligand thought to play a role in terminal chondrocyte differentiation and osteoblastogenesis^{64,65}; *ATOH8* (Atonal homolog 8), a transcription factor important for chondrocyte proliferation and differentiation in the cartilaginous elements of endochondral bone⁶⁶; and *RXRA* (Retinoid X receptor alpha), a retinoic acid receptor that plays a role in endochondral ossification⁶⁷. We also included targets previously undescribed in cartilage biology, including *C16ORF72*, *RCL1*, *GPR153*, *MAP4K3*, and *SCUBE1* based on previously described ChIP-seq interactions or homology with ATAC-seq peaks from mouse chondrocytes (Table 3). Representative gene regulatory elements with RUNX2 motifs are an upstream *ATOH* enhancer (Fig. 6E, overlapping with RUNX2 ChIP-seq data, Col2a1+ mouse chondrocytes, and histone acetylation marks, indicated dotted box and red arrows) and an upstream enhancer of *ACAN* (Fig. 6F), which overlaps with peaks found in mouse Col2a1+

chondrocytes and is homologous to an enhancer identified in mouse chondrocytes⁶⁸. *RUNX2* and the putative target DEGs are more highly expressed in hESC-derived growth plate cartilage (Fig. 6G-H), with the exception of *ACAN* which is expressed in both cartilage lineages. Similarly, *RUNX2* and all but one putative target genes were more highly expressed in fetal growth plate chondrocytes compared to fetal epiphyseal chondrocytes, the exception being *C16ORF72* which was expressed at a similar level. Following ChIP-mediated pulldown of genomic regions bound by *RUNX2* in BMP-treated growth plate cartilage, all 10 target loci chosen for validation were enriched at least 2-fold compared to the negative control (Table 3 and Fig. S6), confirming *RUNX2* binding events at these gene regulatory elements. Six of the ten loci (*C16ORF72*, *RCL1*, *WNT10B*, *MAP4K3*, *RXRA*, *SCUBE1*) were enriched at least 5-fold compared to the negative control.

As the great majority of putative DARs we predicted as harboring motifs recognized by *RELA* and *RUNX2* in hESC-derived articular and hESC-derived growth plate cartilage, respectively, were indeed enriched compared to the negative control loci, we consider the utility of these datasets to be extremely valuable for further exploration of the molecular mechanisms underlying human chondrocyte biology and cell fate decisions.

Discussion

We provide here unbiased molecular characterizations of both the transcriptomic signatures and gene regulatory landscapes of hESC-derived articular and growth plate chondrocytes. We also provide evidence that these hESC-derived lineages are molecularly similar to their *in vivo* counterparts, through transcriptomic profiling of human fetal epiphyseal and growth plate chondrocytes, and epigenetic profiling of mouse embryonic chondrocytes that were isolated from either *Col2a1*-reporter (representing the majority of all mouse chondrocytes) or *Col10a1*-reporter mice (representing mouse growth plate chondrocytes). Specifically, we found strong correlations between hESC-derived articular cartilage and fetal epiphyseal samples, and likewise between hESC-derived growth plate cartilage and fetal growth plate samples.

We performed extensive experimental validation of differentially expressed genes, confirming lineage-specific patterns across multiple independent hESC differentiation experiments and primary cell datasets. Receptor-ligand pairs, Fibroblast growth factor 18 (*FGF18*) and its receptor *FGFR3*⁴⁰⁻⁴³, and Parathyroid hormone-like hormone (*PTH1R*) and its receptor *PTH1R*^{4,44}, are known to be differentially expressed between articular and growth plate cartilage, respectively. We found these expression patterns, and those of other known markers, to hold true in both the hESC-derived and fetal chondrocytes (Figs. 1-2).

In addition to known markers, we identified novel genes that mark distinct cartilage lineages, such as *MEOX1* and *CHI3L1* in the articular cartilage lineage (Fig. 2I-J), and *EFHD1* in the growth plate cartilage lineage (Fig. 2N). *MEOX1* and *CHI3L1*, whose expression has been reported in the axial skeleton and in osteoarthritic cartilage respectively, had not yet been identified in developing articular cartilage. *EFHD1* was strongly localized to hypertrophic cells in hESC-derived and fetal growth plate chondrocytes. Previously studied in its role as a calcium sensor⁶⁹, *EFHD1* could play a role in mediating cellular response to calcium in hypertrophic chondrocytes⁷⁰. We also surprisingly found *tenomodulin* (*TNMD*) to be expressed in the superficial zone of

articular cartilage (Fig. 2A,C), co-expressed with *PRG4*. *TNMD*, closely related to *chondromodulin 1 (CNMD)*, is known as a functional marker for tenocytes³⁹. While there is conflicting evidence of *TNMD* expression in resting and proliferating chondrocytes of the growth plate cartilage^{71,72}, *TNMD* expression in the superficial zone of articular cartilage has not been previously described. Although this result was seemingly unexpected, both cartilage and tendons/ligaments rely on TGF β signaling, and they can arise from a common developmental progenitor^{64,73,74}. This interesting finding warrants further exploration of the developmental relationship between cartilage and the adjacent connective tissues in the joint. Further, these data uncover several other novel genes as yet unstudied in chondrocyte biology, underscoring the potential utility of tissue- or zone-specific markers, and the opportunity to investigate their function(s) in cartilage development or maintenance in human cells and other models.

Despite this strong overall correspondence between hESC-derived and fetal cartilages, there were some notable exceptions, including genes whose expression patterns were opposite those seen *in vitro*. This was an expected result, as the cartilage dissected from fetal samples are more heterogeneous than the hESC-derived tissues, contain fewer terminally differentiated specialized chondrocytes, and also likely have significant overlap in composition of resting and proliferative chondrocytes. For example, the dissected epiphyseal cartilage includes perichondrium, resting zone chondrocytes, proliferative chondrocytes, in addition to chondrocytes that will participate in events related to the secondary ossification center and, perhaps in less abundance, those that will eventually give rise to the neonatal and adult articular cartilage. Likewise, the growth plate cartilage includes proliferative, pre-hypertrophic, and hypertrophic chondrocytes, in addition to perichondrium cells (our micro-dissection approach aimed to omit osteoblasts and hematopoietic cells). Furthermore, it is yet unclear where hESC-derived articular and growth plate cartilage reflect developmental time relative to fetal cartilage. Some obvious differences in developmental stage are the detection of late stage growth plate marker gene expression such as Integrin Binding Sialoprotein (IBSP) in hESC-derived growth plate chondrocytes that was lacking in the fetal growth plate chondrocytes, and the presence of a distinct superficial zone of cartilage in the hESC-derived articular cartilage that is less developed and also less abundant in the fetal tissue used in this study (i.e., we indicated that the surface of the epiphysis corresponded to the site of the future superficial zone of articular cartilage). The potentially more developed/mature superficial zone in the hESC-derived articular cartilage may explain why superficial-zone-specific genes, such as *COL15A1*, a non-fibrillar basement membrane-associated collagen⁷⁵, were identified as differentially expressed in the hESC-derived articular cartilage but not in the fetal epiphyseal cartilage, despite protein localization being lineage-specific (Fig. 2K). Future studies focused on transcriptomic profiling chondrocytes at the single-cell level and/or over differentiation and time will address some of these standing questions.

Given our confidence in the divergent properties of the hESC-derived chondrogenic lineages, and that they reflect *in vivo* lineage properties, we also sought to use this system to define putative gene-regulatory networks (GRNs) which may govern lineage specification and gene expression patterns in developing chondrocytes. As TFs typically have key roles in governing lineage-specific expression patterns^{76,77}, we identified a number of factors which demonstrate biases in expression across lineages, and for which motif occurrence is enriched in

putative lineage-biased regulatory elements. Finding that a simple model of a GRN was insufficient to explain the behaviors observed in our epigenetic and expression datasets, we applied a per-locus approach to integrating our ATAC- and RNA-seq genes, defining sets of genes with different putative regulatory behaviors. We found that these groups exhibited different patterns of differential gene expression, associations with chromatin accessibility data, and, importantly, enriched occurrence of lineage-biased TFs (Figure 4). Notably, our finding that grouped genes differed in their degree of differential expression is consistent with a previous study which stratified immune genes on the basis of regulatory behaviors⁵⁵. We leveraged these findings to identify TFs exhibiting enrichments for DARs around DEGs we defined as either ‘enhancer-centric’ or ‘combo-centric’ enrichments exclusive to a particular lineage (Figure 5). This approach pinpointed a subset of TFs whose binding motifs are significantly enriched in the corresponding chondrogenic lineage (discussed below), of which we functionally tested *RELA* and *RUNX2*. *RELA*, also known as p65, belongs to the NF- κ B family of transcription factors that share a REL homology domain and can form transcriptionally active dimers with other family members. It is a transcriptional activator of *SOX9*, a master regulator of chondrocyte differentiation, as well as early differentiation and anabolic factors such as *SOX6* and *COL2A1*, late stage factor *HIF-2 α* , and the catabolic gene *ADAMTS5*. It also plays a role in cartilage homeostasis⁷⁸ and degradation in osteoarthritis^{79–84}. *RUNX2*, also known as *CBFA1*, *PEBP2*, or *AML3*, belongs to a class of transcription factors containing a Runt-homology domain⁸⁵. *RUNX2* has long been recognized as a ‘master’ skeletogenic factor, sitting atop a regulatory cascade governing osteoblast differentiation^{86–88}. Since its initial discovery, the role of *RUNX2* in skeletogenesis has expanded to include the regulation of chondrocyte hypertrophy in growth plate cartilage^{8–10}. It also has a similar, though pathogenic, role in articular chondrocytes, which acquire hallmarks of hypertrophy in joint diseases such as osteoarthritis^{89–91}. Remarkably, when we performed ChIP-qPCR against candidate regulatory regions with *RELA* or *RUNX2* binding sites, we found that a majority (16 of the 17 tested) did in fact bind the predicted TF in the expected lineage. These findings emphasize the co-use of the epigenetic and expression datasets generated in this study in defining putative gene regulatory networks which may be active in developing chondrocyte populations, and in identifying key regulatory factors controlling these networks.

Our transcriptomic profiling approach uncovered additional lineage-specific TFs in both hESC-derived cartilages, including those that were identified as differentially expressed in the larger group of samples, but not in the smaller subset of samples (batch 2) in which we also performed ATAC-seq. Many of these TFs and their associated family members exhibited similar expression patterns in the fetal cartilage specimens, and many have been previously identified in the context of cartilage and joint biology, once again validating the hESC-model system we’ve established. Such TF families identified in the TGF β -induced hESC-derived articular cartilage include the ETS factors, containing a conserved ETS DNA-binding domain, include the polyomavirus enhancer activator 3 (*PEA3*) family members (*ETV1*, *ETV4*, *ETV5*), and the ETS-related gene (*ERG*) family members (*ERG*, *FLI1*, *FEV*)⁹². Here, we specifically pinpointed *ETV1* and *FLI1* as regulators of enhancer- and combo-centric genes in the hESC-derived articular cartilage lineage (Fig. 5). *PEA3* family members are significantly differentially expressed in both hESC-derived articular cartilage and in fetal epiphyseal chondrocytes. They are FGF-responsive genes and there is some evidence that loss of these proteins results in reduced and disor-

ganized brachial cartilage⁹³. *ERG* and *FLI1* are differentially expressed in hESC-derived articular cartilage (but not significant in fetal data), while *FEV* is differentially expressed in growth plate cartilage (not in fetal). *ERG* has a role in long-term maintenance of articular cartilage, and, along with *FLI1*, regulates articular cartilage genes such as *PTHLH* and *PRG4*^{94,95}. The CREB family of TFs include *CREB5* and *CREB3L1*, both of which are differentially expressed in hESC-derived articular cartilage (*CREB5* is also differentially expressed in fetal epiphyseal cartilage). *CREB5* is a known regulator of *PRG4* expression in articular cartilage³⁰, and shares sequence homology with the ATF family of TFs, such as *ATF7* which we highlight as a regulator of combo-centric genes in the hESC-derived articular cartilage lineage and which has near-identical DNA binding motif compared to *CREB5* (Fig. 5). Nuclear factor of activated T-cells (NFAT) family members *NFATC2* and *NFATC4* are both differentially expressed in hESC-derived articular cartilage (*NFATC4* is also differentially expressed in the fetal epiphysis). *NFATC2* is also more highly expressed in superficial zone chondrocytes compared to deep zone chondrocytes in bovine cartilage³⁰, and NFAT family members play a role in chondrocyte gene expression and articular cartilage maintenance⁹⁶⁻⁹⁸. The homeobox proteins *MEOX1* and *MEOX2* and the LIM-homeobox protein *LHX9* were also DEGs in both hESC-derived articular cartilage and fetal epiphyseal chondrocytes. *MEOX1* and *MEOX2* are essential for the development of all somite compartments and for the normal development of the cranio-cervical joint⁴⁹. *LHX9* is induced by FGF-signaling and has been previously studied for its role in the progression of osteosarcomas⁹⁹. These TFs and TF families, among others we identified in these studies, warrant further exploration for their individual and joint roles in articular cartilage development and stability.

In the growth plate lineage, members of the DLX family of TFs, *DLX2*, *DLX5*, *DLX6*, were highly expressed in growth plate cartilage (*DLX5* and *DLX6* were also differentially expressed in the fetal growth plate), and are known to be critical regulators of cartilage differentiation during endochondral ossification (PMID: 17051482). In particular, *DLX5* has been shown to regulate the differentiation of immature proliferating chondrocytes into hypertrophic chondrocytes, and in osteoblast differentiation¹⁰⁰. Similarly, two *RUNX* family members, *RUNX2* and *RUNX3* are differentially expressed in both hESC-derived and fetal growth plate cartilage. *RUNX2*, as discussed above, is a critical TF for chondrogenic maturation and osteoblast differentiation, and can cooperate with *DLX5* and *SP7* for the proper skeletal development¹⁰¹. *RUNX3* works redundantly with *RUNX2* in chondrocyte maturation¹⁰. The forkhead box (FOX) proteins are a superfamily of TFs, of which several members are differentially expressed in either articular cartilage or growth plate lineages. Of this large family, *FOXA2*, expressed in the hESC-derived growth plate cartilage, is a critical regulator of hypertrophic differentiation in chondrocytes¹⁰² and has been implicated in cartilage degradation and OA progression¹⁰³. *Myocyte enhancer factor 2c (MEF2C)* is differentially expressed in both hESC-derived and fetal growth plate cartilage, and activates the genetic program for hypertrophy during endochondral ossification¹¹. These TFs, and others identified in the studies herein, can now be investigated for their biological role in growth plate biology and chondrocyte function.

The molecular data provided herein have, for the first time, unlocked key findings regarding human articular and growth plate cartilage development. We established and validated our *in vitro* human pluripotent stem cell

cartilage differentiation system as a predictive tool in investigating articular and growth plate cartilage lineages. This is particularly important for understanding how to specify and maintain articular cartilage, since diseased and sometimes even regenerating tissue following cartilage damage display hypertrophy-like changes¹⁰⁴. Novel genes that are expressed differentially between the two different tissues were also identified, some of which exhibit zone-specific expression patterns within developing cartilage. Continued efforts to identify genes and networks that regulate cartilage development will undoubtedly be propelled by these comprehensive comparative analyses of transcriptomic and epigenetic signatures of human articular and growth plate cartilage.

Acknowledgements

The authors would like to acknowledge funding sources for this work (NIAMS - R01-AR073821 (A.M.C.) and R01-AR070139 (T.D.C.)), and the Birth Defects Research Laboratory at the University of Washington and Advanced Bioscience Resources Inc. as sources for fetal donor tissue. The authors also thank Chaochang Li (Boston Children's Hospital) and Garrett Ruff (Cornell University) for imaging assistance, Evelyn Flynn (Boston Children's Hospital) for histological support, Alexander Okamoto (Harvard University) for primary tissue dissection, the Bauer Core Facility (Harvard University) for NGS support, and Jessica Lehoczy (Brigham and Women's Hospital, Harvard Medical School) and Matthew Warman (Boston Children's Hospital) for critical reading of the manuscript.

Data Availability Statement

The raw ATAC-seq and RNA-seq datasets reported in this paper will be available on GEO at the time of publication. Token is provided to reviewers.

Author contributions

Conceptualization, A.M.C.; Methodology, S.P., D.R.; Formal Analysis, D.R., S.P.; Investigation, D.R., S.P.; Validation, D.V.; Data Curation, S.P., D.R.; Writing – Original Draft, D.R., S.P., D.V.; Writing – Review & Editing, A.M.C., T.D.C.; Visualization, A.M.C., D.R., S.P.; Funding Acquisition, A.M.C., T.D.C.; Resources, P.M., Z.L.; Supervision, A.M.C., T.D.C.

Declarations

The authors declare no competing interests.

Figure Legends

Figure 1. hESC-derived articular and growth plate chondrocytes have distinct transcriptional profiles that are similar to their respective fetal cartilage counterparts.

(A) Brief methods to generate hESC-derived cartilage accompanied by toluidine-blue staining sections of hESC-derived articular (TGFB) and growth plate-like (BMP) cartilage tissues and the epiphyseal and growth plate cartilage of a developing fetal cartilage (E59 proximal tibia) show distinct chondrocyte morphology and proteoglycan-rich matrix. Fetal dissection location noted is approximate, and dotted circle highlights overlap of chondrocytes with similar phenotype likely present in both samples following dissection.

(B) PCA plot of RNA-seq expression data from hESC-derived and fetal cartilages. Legend indicates cell type and sequencing batch.

(C) Expression heatmap of the top 20 differentially-expressed genes up- and down-regulated when comparing hESC-derived articular and growth plate cartilage. Red/blue color scale indicates Z-score expression values across samples in each plot. Columns indicate biological replicates.

(D) Expression heatmap of the top 20 differentially-expressed genes up- and down-regulated when comparing fetal epiphyseal and growth plate cartilage tissues. Red/blue color scale indicates Z-score expression values across samples in each plot. Columns indicate biological replicates.

(E) The top 100 differentially-expressed genes up- and down-regulated in the hESC-derived cartilages (top) were compared with equivalent log(2)FC values from the fetal cartilage (bottom).

Figure 2. Validation of differential gene and protein expression in hESC-derived articular and growth plate cartilage and fetal epiphyseal and growth plate cartilage.

(A-H) Confocal microscopy of hESC-derived and epiphyseal and growth plate fetal cartilage following in situ hybridization for indicated probes (RNAscope¹⁰⁵) and counterstained with DAPI (adjacent images). White arrows indicate *TMND* staining at cartilage surface, white double arrow indicates *TNMD* staining in intra-articular ligament. Scale bar, 500 μ m. ial, intra-articular ligament; m, meniscus

(I-J) Quantitative RT-PCR of DEGs in hESC-derived cartilage (I, n=5 independent experiments with 3-6 replicates per experiment) and fetal cartilage (J). Chondrocytes were isolated from the epiphysis and growth plate (GP) of the distal femur and proximal tibia (n=3 per site at E59, E67 and E72). *p<0.05, **p<0.01, ***p<0.001

(K-L) Immunohistochemistry (IHC, brown staining) was used to validate the expression of indicated proteins within cartilage and joint tissues as indicated. Scale bar, 100 μ m. ial, intra-articular ligament; m, meniscus. Sections counterstained with Mayer's hematoxylin (blue).

Figure 3. Epigenetic profiling of hESC-derived articular and growth plate chondrocytes.

(A) The expression (left), gene-promoter accessibility (middle), and metric of cis-regulatory activity (cis-regulatory score, CRS, right) values of the top 20 differentially-expressed genes in each lineage. Red/blue color scale indicates Z-score expression/accessibility values across samples in each plot. Blue/orange scale indicates lineage-bias in cis-regulatory metric.

(B-C) Representative differentially-expressed gene locus in each lineage show differentially accessible regions (DARs) at the promoter corresponding to the same lineage. IHH (B) is differentially expressed in the growth plate lineage while CHI3L1 (C) is differentially expressed in the articular cartilage lineage. Arrows highlight DAR of interest in respective tracks.

Figure 4. Variance in expression of DEGs can be attributed to different classes of regulatory elements (gene regulatory behavior).

(A) Expression (left), gene-promoter accessibility (middle) and cis-regulatory metric (right) of all genes with expression logFC > 1 across lineages. Red/blue, blue-orange color scale indicates Z-score of indicated metric across samples in each indicated plot.

(B) For each gene in (A), the proportion of variance in expression which can be explained by regressing on individual accessibility metrics is shown in red/blue color scale (red = more variance). Hierarchical clustering dendrogram and cluster identity are shown on left and right (respectively), indicating the four clusters of regulatory behavior identified.

(C) LogFC values of genes clustered by regulatory behavior. Significance bars indicate Tukey post-hoc corrected p-values. Proportion of significant DE genes in each cluster are indicated (see Table S3). n.s., not significant; *p < 0.05; **p < 0.01; ***p < 0.001.

(D) Similar plots to (A), for genes clustered by regulatory behavior. Within each heatmap genes are hierarchically clustered by expression logFC. Color scales as in (A).

Figure 5. Identifying putative lineage-delineating transcription factors. Enrichment test results comparing the occurrence of the indicated motif in TGFB or BMP-biased DARs relative to randomized backgrounds.

(A) TFs differentially expressed (DE) in TGFB-treated articular chondrocytes, testing motif occurrence in TGFB or BMP-biased DARs around enhancer-centric DEGs. *p < 0.05; NS, not significant.

(B) TFs DE in TGFB-treated articular chondrocytes, testing motif occurrence in TGFB or BMP-biased DARs around combo-centric DEGs. *p < 0.05; NS, not significant.

(C) Enrichment histogram of RELA motif occurrence in BMP (left) and TGFB (right)-biased DARs around combo-centric genes DE in their respective lineages. Red line indicates target set value, black bars indicate occurrences in randomized sets.

(D) TFs DE in BMP-treated growth plate chondrocytes, testing motif occurrence in TGFB or BMP-biased DARs around enhancer-centric DEGs. *p < 0.05; NS, not significant.

(E) TFs DE in BMP-treated growth plate chondrocytes, testing motif occurrence in TGFB or BMP-biased DARs around combo-centric DEGs. *p < 0.05; NS, not significant.

(F) Enrichment histogram of RUNX2 motif occurrence in BMP (left) and TGFB (right)-biased DARs around combo-centric genes DE in their respective lineages. Red line indicates target set value, black bars indicate occurrences in randomized sets.

Figure 6. Putative targets of TF regulation in hESC-derived articular and growth plate chondrocytes.

(A-B) Two representative differentially accessible regions (blue dashed boxes) in hESC-derived articular chondrocytes (TGFB-treated) which contain *RELA* binding motifs: near the promoter of *GLIPR2* (A) and an intronic enhancer of *LOXL2* (B). Region of interest indicated with red arrow in relevant tracks.

(C) *RELA* is differentially expressed in TGFB-treated articular chondrocytes. ** $p < 0.01$.

(D) Expression of selected genes with putative *RELA* binding motifs was quantified by qRT-PCR. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

(E-F) Two representative differentially accessible regions (orange dashed boxes) in hESC-derived growth plate chondrocytes (BMP-treated) which contain *RUNX2* binding motifs: near the promoter of *ATOH8* (E) and an enhancer of *ACAN* (F). Region of interest indicated with red arrow in relevant tracks.

(G) *RUNX2* is differentially expressed in BMP-treatment. *** $p < 0.001$.

(H) Expression of selected genes with putative *RUNX2* binding motifs was quantified by qRT-PCR. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Tables

Table 1. Summary of ATAC-seq peaks from mouse and human chondrocytes.

			Mouse embryonic chondrocytes		
			All	Col2+	Col10+
			30,950	28,972	12,906
hESC-derived chondrocytes	All Peaks (T+B)	37,780	13,687		
	TGFB (All peaks)	31,137		13,381	9,223
	BMP (All peaks)	29,821		12,471	9,070
	TGFB (Unique)	11,571		3,971	2,385
	BMP (Unique)	12,154		2,584	1,754

Table 2. Summary of candidate RELA targets.

Putative Target	Chromosome location (hg19)	Distance from TSS of gene	Overlap with Mouse Col2a1+ peaks	Overlap with mouse Col10a1+ peaks	Validated ChIP-seq hits from other studies	Fold enrichment in ChIP-qPCR
PRG4	chr1:186201240-186201490	- 64.4 kb	Y		1	8.26
LOXL2	chr8:23268990-23269240	- 7520 bp			44	6.03
LTBP2	chr14:75083374-75083624	- 4380 bp			33	5.71
GLIPR2	chr9:36135932-36137932	+10 bp	Y		17	5.57
DKK3	chr11:12101707-12101957	- 71.1 kb			134	11.56
TLR2	chr4:154577179-154577429	+ 27.2 kb	Y		36	4.85
COL15A1	chr9:101733568-101733818	+ 26.6 kb	Y		1	1.28

Table 3. Summary of candidate RUNX2 targets.

Putative Target	Chromosome location (hg19)	Distance from TSS of gene	Overlap with mouse col2+ peaks	Overlap with mouse col10+ peaks	Validated ChIP-seq hits from other studies	Fold enrichment in ChIP-qPCR
ACAN	chr15:89312870-89313120	-33.0 kb	Y			3.98
ATOH8	chr2:85969150-85969400	-11.9 kb	Y		2	4.79
C16orf72	chr16:9166745-9166995	-18.5 kb			4	12.88
COL10A1	chr6:116439814-116440064	+7110 bp	Y	Y	4	2.44
RCL1	chr9:4837930-4838180	+44.8 kb			3	14.73
WNT10B	chr12:49366141-49368141	- 899 bp			4	5.96
GPR153	chr1:6319685-6321685	+ 699 bp	Y	Y	3	4.06
MAP4K3	chr2:39719320-39719570	-54.6 kb	Y	Y	1	11.36
RXRA	chr9:137178491-137178741	-39.3 kb			4	6.42
SCUBE1	chr22:43701886-43702136	-36.8 kb	Y	Y	1	9.27

Methods

Maintenance of hESCs

H9 hESCs (Wicell, RRID:CVCL_9773) were maintained on irradiated mouse embryonic fibroblasts in hESC media containing DMEM/F12 (Corning) supplemented with 20% knockout serum replacement (Gibco), nonessential amino acids (Gibco), L-glutamine (Gibco), Pen/Strep (Gibco), b-mercaptoethanol (Gibco), and human bFGF (10 ng/mL) in 6-well tissue culture plates. Cells were passaged when they reached ~80% confluency onto new feeders as cell clusters of about 3-10 cells, following dissociation with TrypLE (Gibco).

Generation of chondrocytes from hESCs

A detailed description of the protocol for generating chondrogenic cells and tissues from human pluripotent stem cells has been published³⁵. Briefly, embryoid bodies (EBs) were generated from H9 hESCs and cultured in suspension in the presence of BMP4 (1 ng/mL) and ROCK inhibitor (5 μ M) for 24 hours in StemPro-34 medium (Gibco) supplemented with L-glutamine (Gibco), L-ascorbic acid (Sigma-Aldrich), transferrin (Roche), and α -monothioglycerol (Sigma-Aldrich). On day 1, EBs were harvested and resuspended in StemPro-34 media with bFGF (5 ng/mL), BMP4 (3 ng/mL), Activin A (2 ng/mL), and ROCK inhibitor (5 μ M) to induce primitive streak-like mesoderm. After 44 hours, on day 3, the EBs were harvested from the induction media, cells were dissociated with TrypLE and cultured as monolayers (100,000 cells per well) in 96-well tissue culture plates (Corning) in StemPro-34 media containing bFGF (20 ng/mL), inhibitor of type I activin receptor-like kinase (ALK) receptors SB431542 (5.4 μ M), type I BMPR inhibitor dorsomorphin (4 μ M), and a Wnt inhibitor IWP2 (2 μ M). After 48 hours, on day 5, monolayer cultures were maintained in StemPro-34 media containing bFGF (20 ng/mL) until day 14 to generate chondrogenic mesoderm. Cultures were maintained in hypoxic 5% O₂, 5% CO₂, 90% N₂ environment for 11 days, and normoxic 5% CO₂/air condition for the remainder of the culture period.

Cartilage tissues were generated from the hESC-derived chondrogenic mesoderm culture on day 14 by plating cells in micromass culture. Briefly, 250,000 cells were seeded onto 24-well tissue culture plates (Corning) coated with Matrigel (Corning) in base chondrogenic media consisting of high glucose DMEM (Gibco) supplemented with 1% ITS-A, L-proline (40 μ g/ml) (Sigma-Aldrich), dexamethasone (0.1 μ M) (Sigma-Aldrich), L-ascorbic acid (100 μ g/mL) and TGF β 3 (10 ng/mL) for 2 weeks to generate chondroprogenitors. They were then maintained for an additional 10 weeks in TGF β 3 (10 ng/mL) to generate articular cartilage tissue or transitioned to base chondrogenic media containing L-ascorbic acid (100 μ g/mL) and BMP4 (50 ng/mL) to generate growth plate-like cartilage. Cells and/or tissues were collected after 12 total weeks in micromass.

Chondrocytes were generated for the transcriptomic and epigenetic studies in 4 independent experiments. For each cell type (articular and growth plate), a single micromass was collected per replicate, as described in Table 1. Additional experiments were performed to produce cartilage tissues for validations.

Fetal tissue dissection

Human fetal donor samples (E59, E67, E72) were collected from the first trimester termination via the University of Washington (UW) Birth Defects Research Laboratory (BRDL) in full compliance with the ethical guidelines of the NIH and with the approval of UW Review Boards for the collection and distribution of human tissue for research, and Harvard University and Boston Children's Hospital for the receipt and use of such materials. The samples were briefly washed in Hank's Balanced Salt Solution (HBSS) and transported in the same buffer at 4°C during shipment.

Cartilaginous tissues as described below were dissected under a light dissection microscope in 1x phosphate buffered saline (PBS) on ice and soft tissues were removed. Where appropriate, each epiphysis or growth plate chondrocyte population was microdissected and cells were isolated by collagenase treatment independently.

For fetal bulk-RNA seq, epiphyseal and growth plate cartilages from the left and right distal femur were microdissected from an E67 donor sample. For qPCR, epiphyseal and growth plate cartilage from the distal femur and proximal tibia of one knee joint were microdissected from E59, E67, and E72 human samples. Donor samples for histology/immunohistochemistry were obtained from E59 knee joints (UW BDRL) and E70 metacarpophalangeal and metatarsophalangeal joints (Advanced Bioscience Resources Inc.), formalin fixed and paraffin embedded as described in the Histology/Immunohistochemistry section.

RNA-sequencing (RNA-seq)

Cartilage derived from hESCs (Batch 1) was enzymatically digested with 0.2% type I collagenase (Sigma, St. Louis) for up to two hours at 37°C to solubilize the majority of ECM. Microdissected fetal cartilage was minced and subsequently incubated with 0.1% bacterial collagenase (Sigma, St. Louis) for 2-3 hours at 37°C to solubilize the majority of ECM. Liberated cells were pelleted by centrifugation and the supernatant completely removed. Fetal chondrocytes and hESC-derived cells from Batch 1 were lysed in guanidine thiocyanate buffer. Total RNA was purified using silica column-based kits (ThermoFisher). RNA quality and quantity was assessed via Bioanalyzer (Agilent, Santa Clara), with RIN values > 7. 100ng of total RNA was used as input for the TruSeq RNA Library Prep Kit v2 (Illumina, San Diego). Libraries from Batch 1 were sequenced on an Illumina NextSeq instrument using 2 x 150 bp paired end reads and the fetal samples (Batch 3) plus one technical replicate of an hESC-derived articular and growth plate library from Batch 1 were sequenced on an Illumina Hi-seq using 2 x 150 bp paired end reads. hESC-derived cartilage tissues from Batch 2 were lysed directly in Trizol (Thermo Fisher, Waltham) without prior ECM dissociation (see Paired ATAC-seq and RNA-seq section below). See Table S1 for additional details.

Paired ATAC-seq and RNA-seq

In order to perform simultaneous RNA-seq and ATAC-seq assays on the same micromass chondrocyte culture, we physically bisected cultures via forceps and a scalpel blade. One section of bisected micromass culture was immediately minced, then subsequently transferred to a 1.5 ml microfuge tube containing 200uL of

TRIzol™ reagent (Thermo Fisher Scientific, Waltham) and 5 mm stainless steel beads (Qiagen). Sections were homogenized using a tissue lyser (Qiagen) at 50 Hz for two 2 min periods with intermittent incubation on ice for 1 min. RNA was extracted from the tissue homogenate using phenol-chloroform extraction method, followed by purification using the Direct-zol RNA miniprep kit (Zymo Research, Irvine). RNA quality and quantity was assessed using Bioanalyzer (Agilent, Santa Clara) and Qubit (ThermoFisher, Waltham). Samples with RNA integrity numbers >7 were used for subsequent RNA-seq experiment. 100ng of total RNA were then sent to the Harvard University Bauer Core Facility for library preparation with the Kapa mRNA Library Prep kit (Roche, Basel). Generated RNA-seq libraries were then sequenced on a single sequencing lane of the Illumina NextSeq 500 using 2x38bp reads at the Harvard University Bauer Core Facility – see Table S1 for per-sample sequencing information. Sequencing yielded ~500 million reads per lane and an average of 50 million per sample. Quality control statistics and primer information are presented in Table S1.

The second section of micromass culture was again bisected, then transferred to a 1.5 ml DNA LoBind tube (Eppendorf) containing 200 μ l of DMEM media supplemented with 5% FBS. To generate a single-cell suspension, each sample was then subjected to 1% collagenase, type 2 (Worthington Biochemical, New Jersey) digestion for 2 hours at 37°C rocking, mixing every 30 min. Following digestion, suspension was vortexed, then subsequently centrifuged at 500xg at 4°C for 5 min. Pellets were resuspended in 200 μ l 5% FBS/DMEM for cell counting. All cell counting methods were performed using trypan blue and a hemocytometer and subsequent ATAC-seq steps were performed on those samples that had cell death rates well below 10%. Next, cells were re-suspended in concentrations of 50,000 cells in 1x PBS. Cell samples then subjected to the ATAC-seq protocol as described previously^{53,106}, modifying the protocol by using 2 μ l of transposase per reaction.

The transposase reaction product was then purified using the DNA Clean & Concentrator™ Kits (Zymo Research) following manufacturers protocols, eluted in 10 μ l of warmed ddH₂O, and stored at -20°C. All samples were next subjected to PCR amplification and barcoding following^{53,106}. Ten microliters of transposed DNA were then placed in a reaction containing NEBNext High-Fidelity PCR Master Mix, ddH₂O, and barcoding primers. Following PCR amplification, samples were subjected to double-sided size selection using the Magbind RXN Pure Plus beads (OMEGA) following manufacturer's instructions. The samples were eluted in 20 μ l of TE buffer, nano-dropped, and the fragment size distribution was analyzed by 2100 Bioanalyzer instrument (Agilent Technologies). Prior to sequencing, library concentrations were determined using the KAPA Library Quantification Complete Kit (KK4824). Samples were then sent out to the Harvard University Bauer Core Facility for sequencing on a single lane of an Illumina NextSeq 500. Quality control statistics and primer information are presented in Table S2.

Quantitative RT-PCR

Total RNA was extracted from *in vitro* tissue using the MagMAX mirVana Total RNA kit (Applied Biosystems) and from fetal tissue using the RNAqueous-Micro Total RNA kit (Invitrogen). RNA (0.1-1 μ g) was reverse tran-

scribed with Superscript IV VILO reverse transcriptase (Invitrogen) and treated with ezDNase enzyme (Invitrogen). Real-time quantitative PCR was performed on a ViiA 7 Real-Time PCR System with OptiFlex Optics System (Applied Biosystems) using PowerUp SYBR Green PCR kit (Applied Biosystems). Genomic DNA standards were used to evaluate the efficiency of the PCR and calculate the copy number of each gene relative to the expression of the gene encoding TATA-box binding protein (*TBP*). All data represent three biological replicates (independent experiments or fetal donor specimens) or more as indicated. Student's t-test was used to evaluate statistical significance, as indicated. Oligonucleotides are provided in Table S5.

Histology and immunohistochemistry

In vitro-derived cartilage tissues and primary human donor samples (E59 knee joint; E70 metacarpophalangeal and metatarsophalangeal joints) were fixed in 10% formalin and embedded in paraffin. 5- μ m sections were stained with toluidine blue to visualize sulfated glycosaminoglycans. Immunohistochemistry was performed using antibodies recognizing EF-Hand domain family D1 (HPA056959; RRID:AB_2683288; Sigma Aldrich), and type XV collagen (HPA017915; RRID:AB_1847100; Sigma Aldrich). Antigen retrieval was performed on the tissue using citrate buffer (pH 6.0, overnight at 50°C) for type XV collagen, and using pepsin (30 min. at 37°C) for EFHD1. Positive staining was visualized with DAB. Sections were counterstained with Mayer's hematoxylin (blue).

RNA Scope

Five μ m sections of tissues were deparaffinized by submersion in xylene and washed with 100% ethanol, and treated with the RNAscope hydrogen peroxide solution (Advanced Cell Diagnostics (ACD), Cat. No. 322381). Target antigen retrieval was performed by incubating sections in TEG buffer (25 mM Tris-HCl pH 8, 10 mM EDTA, and 50 mM Glucose) at 60°C for 4 hours, changing the buffer every 40 minutes. Sections were then rinsed in water, dipped in 100% ethanol for 3 min and air dried. RNAscope Protease 3 was applied to the sections at 40°C for 1 hour (ACD, Cat. No. 322381). *In situ* detection of *PRG4*, *TNMD*, *COL2A1*, and *COL10A1* mRNA was performed using the RNAscope™ Multiplex Fluorescent V2 Assay system (ACD, Cat. No. 323110) with the probes (Probe-Hs-COL2A1, Cat. No. 427878; Probe-Hs-PRG4-C3, Cat. No. 427861-C3; Probe-Hs-TNMD, Cat. No. 564409; Probe-Hs-COL10A1, Cat. No. 427851). Fluorescent signal was detected using a confocal microscope (Zeiss LSM 800).

RNA-seq processing

Paired-end sequencing reads were mapped to the human reference transcriptome (GRCh37.67) obtained from the ENSEMBL database using the 'quant' function of Salmon version 0.14.0¹⁰⁷ with the following parameters: "-l A --numBootstraps 100 --gcBias --validateMappings", all others parameters were left to defaults. Salmon quantification files were imported into R version 3.6.1¹⁰⁸ using the tximport library (version 1.14.0)¹⁰⁹ with the 'type' option set to 'salmon', all others set to default. Transcript counts were summarized at the gene-level using the corresponding transcriptome GTF file mappings obtained from ENSEMBL.

Count data was subsequently loaded into DESeq2¹¹⁰ version 1.26.0 using the ‘DESeqDataSetFromTximport’ function. For differential-expression analysis, a low-count filter was applied prior to normalization, wherein a gene must have had a quantified transcript count greater than 5 in at least three samples in order to be retained. PCA of samples across genes was done using the ‘vst’ function in DESeq2 with default settings and was subsequently plotted with ggplot2.

Statistical analysis was performed using the ‘DESeq’ function of DESeq2 using all samples, with results subsequently summarized using the ‘results’ function for the BMP-TGF contrast with the ‘alpha’ parameter set to 0.05; p-values were adjusted using the Benjamini-Hochberg FDR method¹¹¹, with differentially-expressed genes (DEGs) defined at an adjusted p-value cutoff of 0.05. To visualize the expression patterns of these genes across tissues, count data was normalized using the ‘estimateSizeFactors’ function using the default ‘median ratio method’, with the normalized matrix output using the ‘counts’ function with the ‘normalized’ option. The top 200 DEGs (sorted by absolute logFC) were subset from this normalized matrix, z-score transformed using the ‘scale’ function in base R, and visualized with the ComplexHeatmap package version 2.4.3¹¹².

Note on digestion protocol for extracting RNA from cartilage

For transcriptomic analysis of hESC-derived chondrocytes, we isolated RNA from half the samples after first separating the cells from their extracellular matrix via enzymatic digestion (collagenase). From the remaining samples, we isolated RNA directly without prior disruption of the matrix. Differential expression analysis revealed a subset of genes up-regulated in collagenase-digested cells, which were substantially enriched for GO biological processes such as ‘response to lipopolysaccharides’, ‘response to oxidative stress’, and ‘cellular response to toxic substance’. Several of these processes were previously identified via a similar analysis of collagenase-digested osteoblasts¹¹³. Various “cellular responses” were among the enriched processes, indicating chondrocytes are sensitive to enzymatic disruption of their matrix. For applications in which matrix disruption is unavoidable, such as single cell RNA-seq, care should be taken to ensure the transcriptomic signatures of interest are not confused with artifacts associated with enzymatic digestion.

Shared direction analysis

In order to compare the direction of differential-expression between *in-vitro* and *in-vivo* samples, we took the *in-vitro* dataset and subset the top 100 most strongly up-regulated and down-regulated genes, for a final set of 200 (see Table S1). The equivalent data for these genes were subset from the *in-vivo* dataset. The resulting sets of log₂ fold-change values were plotted using ggplot2. A chi-square test for shared direction was performed in base R using a 2x2 contingency table of values (see Table S1). An equivalent analysis was performed using 200 genes defined in the *in-vivo* dataset, see Figure S1, Table S1.

Definition of differentially-expressed transcription factors

To identify transcription factors which are DEGs in our transcriptional data, we extracted all defined motif position-weight-matrices (PWMs) from the JASPAR 2020 database¹¹⁴ along with the motif database provided by HOMER version 4.11¹¹⁵. We then intersected these sets of genes with those differentially-expressed in our RNA-seq analysis to define our sets of differentially-expressed transcription factors (see Table S1). For our analyses integrating ATAC and RNA-seq datasets, we defined differentially-expressed transcription factors using the expression data from the RNA-seq for which paired ATAC was generated (i.e. 'Batch 2' samples).

Gene-set enrichment analyses

Genes associated with different expression/accessibility sets (as described in Results and Methods text) were tested for enrichment in GO Biological Process terms using the 'enrichGO' function from the clusterProfiler¹¹⁶ package version 3.13.1. The background gene sets used for individual enrichment tests are specific - pertaining to a particular analysis. For all differential-expression datasets, all genes in the human reference transcriptome (GRCh37.67), following quality filtering (see RNA-seq section of Methods), was used. For GO enrichments on defined differentially-accessible promoter windows, the background gene set was defined as the set of all promoters. Semantically-similar enriched GO terms were subsequently collapsed using the 'simplify' function from clusterProfiler, using default settings. The top enriched GO terms (sorted by adjusted p-value) for each region-associated gene set are reported in Table S2, limiting to the top twenty significant (adjusted p-value < 0.05) terms.

ATAC-seq Read Processing

Sequence read quality was checked with FastQC and subsequently aligned to the human reference hg19 genome assembly with Bowtie2 v2.3.2¹¹⁷ using default parameters for paired-end alignment. Reads were filtered for duplicates using picard (<https://github.com/broadinstitute/picard>) and subsequently used for peak calling using MACS2¹¹⁸ software (version 2.1.1.2) with the following flags: 'bampe call -f BAMPE -nolambda'. Reproducible called peaks were defined using an IDR threshold of < 0.05, as defined by the IDR statistical test¹¹⁹ (version 2.0.3), as well as a more stringent cutoff of 0.01 for peak sets containing more than 100,000 peaks as recommended by previous ENCODE processing pipelines. Briefly, the IDR method looks for overlaps in peak calls across pairs of replicate samples through comparing ranked peak lists (using MACS2 q-value) to define a reproducibility score curve. All paired ranks are assigned a pointwise score based on this curve, subsequently sorted, and all peaks falling below an 'irreproducible discovery rate' (IDR) threshold of 0.05 are taken as our final reproducible peak set.

For calculating differential accessibility of regions between treatments, all IDR-filtered peaks from both sets of treated tissue fragments (BMP and TGF) were padded to a fixed size of 1000bp (from called peak centres) and pooled using bedtools¹²⁰ version 2.29.1. For each pooled peak, a 1000bp window was defined (500bp up/downstream of peak centre); along this window a 250bp sliding window (chosen based on the averaged called peak size across individual IDR-filtered peaks from each set) was slid in 50bp increments to generate a set of overlapping regions (using the R language version 3.6.1). ATAC-seq read coverage within these regions

was calculated using the 'bedcov' function of samtools¹²¹ version 1.5 for each mapped .bam file corresponding to individual ATAC-seq samples. This sliding-window approach is used to identify the area of strongest cross-sample signal (i.e. ATAC-seq read coverage) around a pooled ATAC-seq peak, improving our confidence in defining differential-accessibility (i.e. avoiding regions around an ATAC-seq peak with lower read coverage, which may make our analysis more sensitive to noise). For sets of overlapping windows corresponding to a single ATAC-seq peak, we applied a smoothing algorithm to eliminate extreme values occurring in overlapping windows for individual samples. Windows whose read coverage fell outside one standard-deviation of the set (of sliding windows for a single peak) were assigned the average read coverage of the two windows adjacent. This adjustment does not impact the later differential-accessibility analysis, for which we take the original read coverage values calculated for a given window using the mapped bam file. Rather, this will impact the choice of sliding window assigned to represent this given peak, and is done to avoid consistently using the most-extreme windows to define differential accessibility (which may otherwise bias our results).

Following this smoothing, for each window we calculated the 75th percentile read coverage value across all pelvic elements and samples (this was found to be more robust than mean read coverages, even after smoothing adjustment) using the dplyr¹²² package version 1.0.7. The window with the greatest 75th-percentile coverage was then selected as the representative region for that pooled ATAC-seq peak. Subsequently, raw read coverages for all optimized windows across all ATAC-seq samples were imported as a matrix into DESeq2 version 1.26.0 using the 'DESeqDataSetFromMatrix' function, with differential-accessibility calculated using the 'DESeq' function with treatment-type as the main variable. Differentially-accessible (DA) peaks were generated using the 'results' function from DESeq2 using the BMP-TGF contrast, with significance assessed as a Benjamini-Hochberg adjusted p-value of < 0.05. To visualize accessibility in these DA peaks, read coverages were normalized using the 'estimateSizeFactors' function in DESeq2 using the default TMM normalization. Subsequently, the normalized read-counts matrix was z-score transformed using the 'scale' function in base R and plotted using the 'ComplexHeatmap' package version 2.4.3.

De-novo motif enrichments

Sequence sets for the sets of differentially-accessible regions (as defined above) were generated using reference sequences from hg19. HOMER (version 4.11) *de novo* motif analysis was performed on each sequence set using a 10x random shuffling as a background set. *De novo* motifs were compared to a vertebrate motif library included with HOMER, which incorporates the JASPAR database; matches are scored using Pearson's correlation coefficient of vectorized motif matrices (PWMs), with neutral frequencies (0.25) substituted for non-overlapping (e.g., gapped) positions. Best-matching motif PWMs for TGF- and BMP-biased region sets are shown in Figure 2C and Table S2, Sheets 4-5.

Targeted motif enrichments

We took our set of differentially-expressed transcription factors up-regulated in articular and growth-plate chondrocytes, defined using expression from Batch 2 samples, and obtained the position-weight-matrices (PWMs) for all factors. In instances where a factor had more than one defined PWM (e.g. due to overlap between JASPAR and HOMER databases), the information content (IC) of each matrix was calculated, with the matrix having the highest IC retained. To define the background nucleotide frequencies of our ATAC-seq region sets, we took the pooled set of fixed-size peaks used in the DA analysis ($n = 37780$) and generated a Markov background model using the 'fasta-get-markov' function from the MEME suite¹²³. Next, we took the sets of nucleotide sequences corresponding to DA peaks and scanned them for instances of particular TF motifs using the FIMO program from MEME, defining a p-value threshold for motif hits as $2e-4$ (calculated as $0.1/(2 \times 250\text{bp of sequence})$). To define a background expectation of motif hits for sequence sets equivalent to our target sets, we used the 'shuffle' command from bedtools, randomly shuffling our sequence set across the hg19 genome (exclusive of the true set of regions). For each randomly-shuffled set, we then scanned for motif hits with FIMO, performing this random-shuffling and scanning for $n = 100$ randomized sets. This was used to establish a random background distribution of expected motif hits for a given TF motif. Motif hit values were standardized and statistical significance was assessed using a CDF of the standard normal distribution as implemented in the 'pnorm' function in base R. For each DA sequence set, P-values for significant deviations from the background distribution were corrected for the number of transcription factors tested ($n = 194$) (Table S2). As further confirmation, for the top 10 most strongly-DE transcription factors in each lineage, we increased the number of randomly-shuffled background sets to $n = 1000$, finding that our enrichments were robust to the size of the background distribution.

We next checked to confirm whether these motif enrichments are reflective of complexity of PWMs, i.e. that TFs with lower sequence binding specificity result in more promiscuous motif hits. For a given TF, we compared the fraction of all sequences in a given DA set which had at least one motif hit to the information content (IC) of the factor's PWM. These values were plotted across all TFs tested using the 'ggscatter' function from ggpubr version 0.4.0, with Pearson correlation calculated with this function (see Figure S3).

GREAT analysis

GREAT⁵⁴ takes an input set of genomic regions along with a defined ontology of gene annotations; firstly, it defines regulatory domains for all genes genome-wide, then measures the fraction of the genome covered by the regulatory domains of genes associated with a particular annotation (e.g. 'cartilage development'). These fractions are used as the expectation in a binomial test counting the number of input genomic regions falling within a given set of regulatory domains, which results in the reported significance of association between an input region set and a particular gene ontology term. GREAT also performs a more traditional gene-based hypergeometric test to test for significance of region set-ontology association. The program returns a set of enriched ontologies sorted by the joint rankings of FDR-corrected binomial and hypergeometric tests, as reported here in Supplemental Table S2.

Defining promoter-accessibility

All hg19 Refseq gene TSS were obtained from the UCSC genome browser¹²⁴ and padded 5kb up/downstream to define regions around each promoter. For each promoter, 2kb windows were slid along this region in 50bp increments, with per-window ATAC-seq read coverage for all samples calculated using the 'bedcov' function of samtools (version 1.5) for each mapped .bam file. A similar smoothing method as that described above for per-peak windowed accessibility was also performed here using sliding promoter window accessibility. As above, this smoothing method does not impact the later differential-accessibility analysis, for which we take the original read coverage values calculated for a given window using the mapped bam file. Following smoothing, the window with the greatest 75th-percentile coverage was selected as the representative read coverage metric for a given promoter. Following per-promoter window selection, the final matrix of read coverages for all promoter windows across all samples was loaded into DESeq2 version 1.26.0 using the resulting in a final matrix of read coverages for all promoters across all samples. This matrix was subsequently loaded into DESeq2 version 1.26.0 using the 'DESeqDataSetFromMatrix' function, with differential-accessibility calculated using the 'DESeq' function with treatment-type as the main variable. Differentially-accessible (DA) promoters were generated using the 'results' function from DESeq2 using the BMP-TGF contrast, with significance assessed as a Benjamini-Hochberg adjusted p-value of < 0.05.

Defining cis-regulatory scores

We defined a 'cis-regulatory score' which seeks to capture information on accessibility patterns within putative regulatory elements around a given gene locus, a concept inspired by methods of integrating RNA-seq and ChIP-seq datasets¹²⁵. For a given gene, we collected all accessibility regions (i.e. the optimized windows described above) within 100kb up/downstream of the TSS. Given that we also, separately, considered promoter accessibility in describing the regulatory behavior of genes (see above, and below), we explicitly excluded any regions which fell within the optimized promoter regions defined above. For each captured accessibility region, we took the calculated fold-accessibility change (BMP-TGF contrast) and scaled it by the distance to the gene TSS using the following formula:

$$\text{Per-region score} = (\text{differential-accessibility logFC}) * (e^{-(0.5 + 4 * (\text{distance} / 100\text{kb}))})$$

Where the second distance-scaling term is taken from Tang et al. 2011¹²⁶. These per-region scores were summed across all captured regions across the locus to give a single, signed lineage-specific (i.e. BMP or TGF-treated) *cis*-regulatory score. These scores were visualized per-gene as an additional feature in the 'ComplexHeatmap' visualizations of expression (as seen in Figure 2B); blue/red scale was defined based on max/min values (respectively) for this score (positive being TGF-biased, negative being BMP-biased); points outside two standard deviations of the mean score were capped at 2*SD +/- the mean.

Defining gene regulatory behaviors

For a given gene, all accessible elements within a 100kb (up/downstream of the gene TSS) were collected, exclusive of the optimized promoter window (as described above). For each element, the normalized accessibility counts data (via DESeq2 using the default TMM normalization, as described above) was retrieved

for all hESC samples (BMP and TGF treated). These normalized counts were subsequently scaled by distance to the TSS using the above distance-scaling term. Similarly, the normalized accessibility counts data for promoter regions (as described above) was also retrieved across all samples. Finally, RNA-seq expression data for the given gene was retrieved and normalized using the 'vst' function in varistran¹²⁷ version 1.0.4 to obtain normalized expression counts across all samples.

For this gene, three different linear regression models for RNA-seq expression were generated using the 'lm' function in base R. (1) RNA ~ Promoter accessibility + Enhancer Accessibility + Promoter x Enhancer Interaction. (2) RNA ~ Promoter accessibility. (3) RNA ~ Enhancer accessibility. The proportion of total variance in RNA expression explained by each model was calculated by feeding individual model objects into the 'anova' function in base R and summing the individual sum-of-squares values for all model coefficients, then dividing by the total sum of squares. We took the differences between the total variance explained by the full model (1) and that explained by individual-component models (2) and (3) as an approximate measurement of the contributions of either enhancer or promoter accessibility. Additionally, we defined a metric for the additive effects of combining promoter and enhancer accessibility as the difference between the variance explained by the full model and that explained by the two models individually, while accounting for instances in which this difference is negative (i.e. when the two single-component models share overlapping information). Finally, the proportion of variance that could not be explained by the full model was also included as an additional metric.

These four per-gene metrics (contribution of enhancer information, promoter information, combination of enhancer and promoter, and unexplained variance) were defined across all genes captured in our RNA-seq analysis. The resulting matrix of genes was subsequently visualized and clustered using the 'Heatmap' function from ComplexHeatmap version 2.8.0 using a 'k' of 4 (as seen in Figure 4B). Clustered genes were defined as being either 'unexplained', 'combo-centric', 'promoter-centric', or 'enhancer-centric' based on the distribution of variance explained in the heatmap shown in Figure 4B. The percentage of clustered genes exhibiting significant DE between BMP and TGF-treated samples (Figure 4C, bottom) was calculated using the results of the differential-expression analysis above. For all genes in each cluster (regardless of significance) the absolute log₂FC values were compared across clusters using the 'aov' function in base R, followed by Tukey Post-hoc correction using the TukeyHSD function in base R. Differences in absolute log₂FC values were plotted using the 'plotmeans' function from gplots version 3.1.1 with default settings.

To test whether the directionality of accessibility and expression was shared in different clusters of genes we used the following approach. For each gene, the distance-scaled enhancer accessibility, promoter accessibility, and gene expression values were binarized into either 'biased towards TGF samples' or 'biased towards BMP samples'. Across all genes in a given cluster, the number of genes sharing direction in either (a) enhancer accessibility and RNA expression, or (b) promoter accessibility and RNA expression was compared to those switching direction using a chi-square test in base R. Finally, for each clustered set of genes, the

RNA-seq expression, promoter accessibility, and cis-regulatory score values were visualized using ComplexHeatmap, sorting first on RNA-seq expression.

Motif analyses with regulatory behaviors

For motif analyses in the context of our defined regulatory behaviors, we considered motif enrichments at putative regulatory regions (enhancers) around genes and gene promoter regions separately. We first took the groups of genes falling into ‘enhancer-centric’ and ‘combo-centric’ clusters, separated them into those up-regulated in articular/growth-plate chondrocytes, then defined 200kb windows centered on the TSS of each gene. Subsequently, all DA peaks sharing directionality (i.e. articular chondrocyte DA peaks around articular chondrocyte DEGs) within these windows were collected and aggregated into a final set of sequences. For ‘promoter-centric’ genes, we took the set of optimized promoter regions for DEGs and aggregated them. These sequence sets were then used with the AME webserver (part of the MEME suite)¹²³, along with the sets of differentially-expressed TF motif matrices, matching the sequence sets (i.e. articular chondrocyte-biased sequence sets were scanned with articular chondrocyte-biased TFs), using default settings. The resulting enrichments were filtered with a p-value threshold of < 0.05 and summarized (see Table S3, Sheet 6).

To test the specificity of these enrichments for sequences biased towards a particular lineage, we performed an additional level of motif analysis. We collected the sets of motifs enriched for genes falling in particular regulatory groups, and differentially-expressed in particular lineages, and utilized the same sequence sets pertaining to enhancers or promoters of genes. To define the background nucleotide frequencies of our ATAC-seq region sets, we took the pooled set of fixed-size peaks used in the DA analysis (n = 37780) and generated a Markov background model using the ‘fasta-get-markov’ function from the MEME suite¹²³. We similarly took the set of optimized promoter regions (based on the analysis described above) (n = 27739) and defined a Markov background nucleotide frequency model. We then took the regulatory elements of a particular set (e.g. DA peaks biased towards articular chondrocytes, around articular chondrocyte DEGs) and scanned for instances of a given motif using the FIMO program from MEME, defining a p-value threshold for motif hits as $2e-4$ (calculated as $0.1/(2 \times 250\text{bp of sequence})$). Similar to above, we defined a background expectation of motif hits for sequence sets equivalent to our target sets using the ‘shuffle’ command from bedtools version 2.29.1, randomly shuffling our sequence set across the hg19 genome (exclusive of the true set of regions). For scanning promoter sequences, we randomly shuffled the true set of promoters across all promoters used in the differential-promoter-accessibility analysis (n = 27739), to account for background biases in the occurrence of motifs within promoter regions generally.

For each randomly-shuffled set, we then scanned for motif hits with FIMO, performing this random-shuffling and scanning for n = 2000 randomized sets. This was used to establish a random background distribution of expected motif hits for a given TF motif. Motif hit values were standardized and statistical significance was assessed using a CDF of the standard normal distribution as implemented in the ‘pnorm’ function in base R. As we sought to test the specificity of motif enrichments, all motifs which were enriched for a given regulatory set

(e.g. enriched in enhancer sequences) were tested, regardless of any existing lineage bias detected in our AME analyses. This resulted in our testing $n = 96$ TF motifs against all enhancer sequences, and $n = 109$ TF motifs against all optimized promoter regions. For each lineage-biased sequence set (i.e. sequences associated with DEGs in articular chondrocytes or growth-plate chondrocytes), P-values for significant deviations from the background distribution were BH corrected for the number of transcription factors tested (either 96 or 109, based on the regulatory group of genes considered) (Table S3).

As a more stringent analysis, we also re-ran these enrichment tests using a constrained set of randomized regions. For scanning enhancer sequences, we randomly shuffled the true target set of regions across the entire set of peaks used in the DA analysis ($n = 37780$), to account for background biases in the occurrence of motifs across ATAC-seq peaks generally. We refer to this analysis as the 'ATAC-BACK' set. As above, 1000 randomly-generated sets were used to assess statistical significance along with BH corrections.

To visualize the differences in motif enrichments across sets of lineage-biased genes and regions, we took the sets of motifs enriched in either growth-plate (BMP) or articular (TGF) chondrocyte lineages and sorted them by the absolute difference in \log_{10} adjusted p-values (i.e. selecting those motifs with the largest change in motif enrichments across lineages). The top five motifs from each analysis were then plotted using ggplot2 (as seen in Figure 5 A, B, D, E). The 'ATAC-BACK' background sets were used to visualize the distribution of motif hits used in assessing significance of the indicated factors in Figure 5C and F).

Overlap with ChIP-seq data

For our two chosen factors of interest (RELA and RUNX2) we obtained ChIP-seq datasets from ChIP-Atlas¹²⁸, which aggregates ChIP-seq datasets from publicly-available datasets (see Table S4, Sheet 1 for full lists of accessions aggregated in each track). We broadly selected all ChIP-seq tracks available for our factors, retaining all peaks at a significance threshold of 0.05. Tracks were sorted and merged with bedtools. To test for overlap with our region sets we performed hypergeometric tests using the 'phyper' function in base R. Given our interest in lineage-specificity of transcription factor motifs, we used as a background the pooled set of DA peaks biased in either lineage ($n = 4720$). BH correction was applied for the number of sequence sets tested ($n = 4$) (Table S4, Sheet 1). We subsequently summarized the results of these overlaps, which are presented in Table S4 Sheets 2-5. Additionally, we took the result of FIMO predicted motif hits (as described above) for RELA and RUNX2 motif matrices on BMP- and TGF-biased peak sets, and included them along with the data on ChIP-seq overlaps (see Table S4). Indicated genes in Table S4 represents the closest nearby gene to a given peak.

Mouse chondrocyte isolation

ATAC-seq experiments were performed on transgenic Col2a1-ECFP/Col10a1-mCherry reporter mice (a gift from Dr. Cliff Tabin at Harvard Medical School)¹²⁹, which has an enhanced cyan fluorescent protein reporter under the control of the promoter of Col2a1, and an enhanced mCherry fluorescent protein reporter under the

control of the promoter of Col10a1. Col2a1-ECFP/Col10a1-mCherry male and female mice were used to establish timed matings, and at E15.5 pregnant females were euthanized to acquire embryos. At this time point, chondrocytes are easily extracted from surrounding extracellular matrix for ATAC-seq with negligible effects on the epigenome³¹. Embryos were dissected under a microscope in 1X PBS on ice and the proximal and distal portions of the right and left femur and tibia of the hind limb were stripped clear of soft tissues. Each proximal or distal cartilaginous end comprising of the articular chondrocytes, epiphyseal chondrocytes, and metaphyseal chondrocytes was then micro-dissected from the bony diaphysis and separately pooled from a single litter, consisting on average of eight animals. All samples were collected in micro-centrifuge Eppendorf tubes containing 200µL 5% FBS/DMEM. To generate a single-cell chondrocyte suspension, each pooled sample was then subjected to 1% Collagenase II (Worthington Biochemical, LS004176) digestion for 2 hours at 37°C rocking, mixing every 30 minutes. After placing on ice, samples were next filtered using a micro-centrifuge filter set-up by gently mashing the residual tissues through the filter followed by rinsing with 5% FBS/DMEM. Samples were then spun down at 500Xg at 4°C for 5 minutes. Col2a1-ECFP or Col10a1-mCherry positive chondrocytes were collected by fluorescence activated cell sorting (FACS) using BD FACS Aria Cell Sorters at Harvard University Bauer Core Facility (HUBCF). All chondrocyte counting methods were performed using trypan blue and a hemocytometer and subsequent ATAC-seq steps were performed on collected chondrocyte samples that had cell death rates well below 10%. On average we acquired 150,000–200,000 living cells for Col2a1-ECFP positive chondrocytes per harvest; and acquired 2500-5000 living cells for Col10a1-mCherry positive chondrocytes per harvest. Next, cells were re-suspended in concentrations of 50,000 cells in 1x PBS for CFP positive chondrocytes; and entire mCherry positive chondrocytes were resuspended in PBS for next step of ATAC-seq process. Cell samples then subjected to the ATAC-seq protocol as described previously^{53,106}, modifying the protocol by using 2 µl of transposase per reaction. The transposase reaction product was then purified using the Omega MicroElute DNA Clean Up Kit following manufacturer's protocols, eluted in 10 µl of warmed ddH₂O, and stored at -20°C. All samples were next subjected to PCR amplification and barcoding following^{53,106}. Ten microliters of transposed DNA were then placed in a reaction containing NEBNext High-Fidelity PCR Master Mix, ddH₂O, and primers. Because of low cell number for Col10a1-mCherry positive chondrocytes, additional PCR cycles of PCR amplification is needed to generate enough PCR fragments for the next steps of ATAC-seq. To determine the optimal number of cycles to amplify the library with minimal artifacts associated with saturation PCR of complex libraries, the appropriate number of PCR cycles is determined using qPCR to stop amplification prior to saturation^{53,106}. Following amplification, samples were transferred to new tubes and treated using the OMEGA Bead Purification Protocol following manufacturer's instructions. The samples were eluted in 30 µl of TE, nano-dropped, diluted to 5 ng/ml and run on a Bioanalyzer. Prior to sequencing, sample concentrations were determined using the KAPA Library Quantification Complete Kit (KK4824). Samples were then sent out to the Harvard University Bauer Core Facility for sequencing on one lane of the Illumina NextSeq 500. Sequencing yielded ~400 million reads per lane and an average of 50 million per sample. Quality control statistics and primer information are presented in Table S2.

Sequence read quality was checked with FastQC and subsequently aligned to the mouse reference mm10 genome assembly with Bowtie2 v2.3.2¹¹⁷ using default parameters for paired-end alignment. Reads were filtered for duplicates using picard (<https://github.com/broadinstitute/picard>) and subsequently used for peak calling using MACS2¹¹⁸ (version 2.1.1.2) with the following flags: 'bampe call -f BAMPE -nolambda'. Reproducible called peaks were defined using an IDR threshold of < 0.05, as defined by the IDR statistical test¹¹⁹ (version 2.0.3). However, the Col10a1 datasets, for which we noted elevated levels of cell death prior to ATAC-seq (possibly attributable to the hypertrophic nature of these cells) had substantially greater variability across samples compared to our Col2a1 datasets. Thus, for Col10a1 called peaks we used a less stringent approach to replicate consolidation – requiring that a called peak is overlapped in at least two different samples (using bedtools intersect) in order to be considered for subsequent analyses.

Chromatin Immunoprecipitation – quantitative PCR (ChIP-qPCR)

Chromatin preparation, ChIP, and qPCR was performed at Active Motif (Carlsbad, CA). In brief, hESC-derived articular and growth plate chondrocytes were isolated from their respective cartilage tissues, fixed with 1% formaldehyde for 15 min and quenched with 0.125 M glycine. Chromatin was isolated by the addition of lysis buffer, followed by disruption with a Dounce homogenizer. Lysates were sonicated and the DNA sheared to an average length of 300-500bp using the EpiShear™ Probe Sonicator (Active Motif, cat # 53051) with an EpiShear™ Cooled Sonication Platform (Active Motif, cat # 53080). Genomic DNA (Input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking (overnight at 65°C) followed by ethanol precipitation and SPRI bead clean up (Beckman Coulter). The resulting DNA was quantified by ClarioStar (BMG Labtech). Extrapolation to the original chromatin volume allowed quantitation of the total chromatin yield.

Aliquots of chromatin (25 ug) were precleared with protein A agarose beads (Invitrogen). Genomic DNA regions of interest were isolated using antibodies against RelA (Active Motif, catalog number 39369; RRID:AB_2793231) in articular chondrocytes, and against RUNX2 (CST, catalog number 8486; RRID:AB_10949892) in growth plate chondrocytes, where the two TFs are differentially expressed respectively. Complexes were washed, eluted from the beads with SDS buffer, and subjected to RNase and proteinase K treatment. Crosslinks were reversed by incubation overnight at 65°C, and ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation.

Quantitative PCR (qPCR) reactions were carried out in triplicate using SYBR Green Supermix (Bio-Rad, Cat # 170-8882) on a CFX Connect™ Real Time PCR system. One negative control primer pair was used (Unt12, Human negative control primer set 1, Active Motif, catalog number 71001) as well as one positive control (DPF1 for RUNX2, BIRC3 for RelA), plus the target sites of interest. The resulting signals were normalized for primer efficiency by carrying out qPCR for each primer pair using unprecipitated genomic DNA. Oligonucleotides are provided in Table S5.

Supplemental Table Legends

Table S1. Sequencing information for RNA-seq runs

Sheet 1: Quality summary statistics for RNA-seq samples.

Sheet 2: DEG data for hESC-derived (*in vitro*) and fetal (*primary*) chondrocyte samples. Columns C-H: DESeq2 statistics for hESC-derived samples. Log2FC are calculated as BMP-vs-TGF. Columns I-N: DESeq2 statistics for primary fetal samples. Log2FC are calculated as Growth plate vs. epiphyseal (GP-vs-EP). Columns O-AG: Normalized transcript counts for the indicated gene in the indicated sample.

Sheet 3: Testing shared direction of differential expression between primary and *in vitro* hESC-derived samples.

Sheet 4: DEG gene-set enrichments. The (up to) top 30 significantly-enriched gene sets obtained using sets of genes up-regulated in the given tissue.

Sheet 5: Differential expression information for transcription factors differentially expressed in at least one of four cell types. Column 'A' denotes classification of different TFs.

Sheet 6: Differential expression analysis run only on 'Batch 2' samples, for which paired ATAC-seq libraries were generated.

Sheet 7: Sets of transcription factors which were DEG in the Batch 2 analysis.

Table S2. Sequencing information for ATAC-seq runs

Sheet 1: Quality summary statistics for ATAC-seq samples.

Sheet 2: Sets of differentially-accessible peaks, separated by their lineage bias. hg19 coordinates shown.

Sheet 3: GREAT results for differentially-accessible peak sets. Description of column names and tests are available from the GREAT website.

Sheet 4: HOMER *de-novo* motif analysis results for DA peaks in growth plate chondrocytes

Sheet 5: HOMER *de-novo* motif analysis results for DA peaks in articular chondrocytes

Sheet 6: Results of motif testing of all differentially-expressed transcription factors with the indicated DA peaksets.

Table S3. Gene regulatory behaviors

Sheet 1: Summarized statistics comparing genes grouped by regulatory behavior

Sheet 2: Significant genes assigned to cluster 1 - "Poorly explained". DESeq2 DEG statistics for all genes are shown, along with the amount of variance in gene expression which was ascribed to either enhancer cis-regulatory scores, promoter accessibility, the interaction term, and variance which could not be explained, "Unexplained".

Sheet 3: Significant genes assigned to cluster 2 - "Combo-centric"

Sheet 4: Significant genes assigned to cluster 3 - "Promoter-centric"

Sheet 5: Significant genes assigned to cluster 4 - "Enhancer-centric"

Sheet 6: The set of differentially-expressed transcription factors for which motif enrichments were tested in regulatory sequences associated with genes falling in different regulatory groups. Columns I-K indicate whether a significant enrichment was observed for promoter-centric, enhancer-centric, or combo-centric genes. Column L indicates the direction of differential expression of a given TF.

Sheet 7: Results of motif enrichment testing for promoters of promoter-centric genes. Columns E-G refer to the enrichment statistics for BMP-biased DEGs. Columns I-K refer to enrichment statistics for TGF-biased DEGs. Sets of TFs are grouped based on their lineage-specificity behaviors.

Sheet 8: Results of motif enrichment testing for putative enhancers of enhancer-centric genes. Columns E-I refer to the enrichment statistics for BMP-biased DEGs. Columns J-O refer to enrichment statistics for TGF-biased DEGs. "GENOME_TAG" refers to enrichment results when using the entire genome as a background set. Sets of TFs are grouped based on their lineage-specificity behaviors.

Sheet 9: Results of motif enrichment testing for putative enhancers of combo-centric genes. Columns E-I refer to the enrichment statistics for BMP-biased DEGs. Columns J-O refer to enrichment statistics for TGF-biased DEGs. "GENOME_TAG" refers to enrichment results when using the entire genome as a background set. Sets of TFs are grouped based on their lineage-specificity behaviors.

Table S4. Overlapping ChIP-seq datasets with DA peaksets.

Sheet 1: Overlapping ChIP-seq datasets with DA peaksets. Accessions listed in this sheet for RELA and RUNX2 correspond to those accessions listed in Sheets 2-5.

Sheets 2 and 3: Overlaps of RELA ChIP-seq datasets with BMP- and TGF-biased peak sets. Column A: Whether or not the indicated region (Columns B-D) contained a FIMO predicted motif hit for RELA. 'Num_hits': the number of ChIP-seq peaks (pooled across individual datasets) which overlapped the given region. "Hits": the individual overlapping ChIP-seq peaks, separated by '@'. "sources" and "ID" refers to the individual datasets corresponding to each ChIP-seq peak separated by '@' in 'Hits' column.

Sheets 4 and 5: Overlaps of RUNX2 ChIP-seq datasets with BMP- and TGF-biased peak sets.

Table S5. Oligonucleotides used in this report

References

1. Wang W, Rigueur D, Lyons KM. TGF β signaling in cartilage development and maintenance. *Birth defects research Part C, Embryo today : reviews*. 2014;102(1):37-51. doi:10.1002/BDR.C.21058
2. Long F, Zhang XM, Karp S, Yang Y, McMahon AP. Genetic manipulation of hedgehog signaling in the endochondral skeleton reveals a direct role in the regulation of chondrocyte proliferation. *Development (Cambridge, England)*. 2001;128(24):5099-5108. doi:10.1242/DEV.128.24.5099
3. St-Jacques B, Hammerschmidt M, McMahon AP. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes & development*. 1999;13(16):2072-2086. doi:10.1101/GAD.13.16.2072
4. Karaplis AC, Luz A, Glowacki J, et al. Lethal skeletal dysplasia from targeted disruption of the parathyroid hormone-related peptide gene. *Genes & development*. 1994;8(3):277-289. doi:10.1101/GAD.8.3.277
5. Lanske B, Karaplis AC, Lee K, et al. PTH/PTHrP receptor in early development and Indian hedgehog-regulated bone growth. *Science (New York, NY)*. 1996;273(5275):663-666. doi:10.1126/SCIENCE.273.5275.663
6. Bi W, Deng JM, Zhang Z, Behringer RR, de Crombrugge B. Sox9 is required for cartilage formation. *Nature genetics*. 1999;22(1):85-89. doi:10.1038/8792
7. Akiyama H, Chaboissier MC, Martin JF, Schedl A, de Crombrugge B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes & development*. 2002;16(21):2813-2828. doi:10.1101/GAD.1017802
8. Inada M, Yasui T, Nomura S, et al. Maturation disturbance of chondrocytes in Cbfa1-deficient mice. *Developmental Dynamics*. 1999;214(4). doi:10.1002/(SICI)1097-0177(199904)214:4<279::AID-AJA1>3.0.CO;2-W
9. Kim IS, Otto F, Zabel B, Mundlos S. Regulation of chondrocyte differentiation by Cbfa1. *Mechanisms of development*. 1999;80(2):159-170. doi:10.1016/S0925-4773(98)00210-X
10. Yoshida CA, Yamamoto H, Fujita T, et al. Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. *Genes & development*. 2004;18(8):952-963. doi:10.1101/GAD.1174704
11. Arnold MA, Kim Y, Czubyrt MP, et al. MEF2C transcription factor controls chondrocyte hypertrophy and bone development. *Developmental cell*. 2007;12(3):377-389. doi:10.1016/J.DEVCEL.2007.02.004
12. Schipani E, Ryan HE, Didrickson S, Kobayashi T, Knight M, Johnson RS. Hypoxia in cartilage: HIF-1 α is essential for chondrocyte growth arrest and survival. *Genes & development*. 2001;15(21):2865-2876. doi:10.1101/GAD.934301
13. Ionescu A, Kozhemyakina E, Nicolae C, Kaestner KH, Olsen BR, Lassar AB. FoxA family members are crucial regulators of the hypertrophic chondrocyte differentiation program. *Developmental cell*. 2012;22(5):927-939. doi:10.1016/J.DEVCEL.2012.03.011
14. Ohba S, He X, Hojo H, McMahon AP. Distinct Transcriptional Programs Underlie Sox9 Regulation of the Mammalian Chondrocyte. *Cell reports*. 2015;12(2):229-243. doi:10.1016/J.CELREP.2015.06.013
15. He X, Ohba S, Hojo H, McMahon AP. AP-1 family members act with Sox9 to promote chondrocyte hypertrophy. *Development (Cambridge, England)*. 2016;143(16):3012-3023. doi:10.1242/DEV.134502
16. Liu CF, Lefebvre V. The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic acids research*. 2015;43(17):8183-8203. doi:10.1093/NAR/GKV688
17. Oh C do, Lu Y, Liang S, et al. SOX9 regulates multiple genes in chondrocytes, including genes encoding ECM proteins, ECM modification enzymes, receptors, and transporters. *PloS one*. 2014;9(9). doi:10.1371/JOURNAL.PONE.0107577
18. Oh C do, Maity SN, Lu JF, et al. Identification of SOX9 interaction sites in the genome of chondrocytes. *PloS one*. 2010;5(4). doi:10.1371/JOURNAL.PONE.0010113
19. Lui JC, Chau M, Chen W, et al. Spatial regulation of gene expression during growth of articular cartilage in juvenile mice. *Pediatric research*. 2015;77(3):406-415. doi:10.1038/PR.2014.208
20. Chau M, Lui JC, Landman EBM, et al. Gene expression profiling reveals similarities between the spatial architectures of postnatal articular and growth plate cartilage. *PloS one*. 2014;9(7). doi:10.1371/JOURNAL.PONE.0103061
21. Yamane S, Cheng E, You Z, Hari Reddi A. Gene expression profiling of mouse articular and growth plate cartilage. *Tissue engineering*. 2007;13(9):2163-2173. doi:10.1089/TEN.2006.0431

22. Tan Z, Niu B, Tsang KY, et al. Synergistic co-regulation and competition by a SOX9-GLI-FOXA phasic transcriptional network coordinate chondrocyte differentiation transitions. *PLoS genetics*. 2018;14(4). doi:10.1371/JOURNAL.PGEN.1007346
23. James CG, Appleton CTG, Ulici V, Michael Underhill T, Beier F. Microarray analyses of gene expression during chondrocyte differentiation identifies novel regulators of hypertrophy. *Molecular biology of the cell*. 2005;16(11):5316-5333. doi:10.1091/MBC.E05-01-0084
24. Cameron TL, Belluoccio D, Farlie PG, Brachvogel B, Bateman JF. Global comparative transcriptome analysis of cartilage formation in vivo. *BMC developmental biology*. 2009;9(1). doi:10.1186/1471-213X-9-20
25. James CG, Stanton LA, Agoston H, Ulici V, Underhill TM, Beier F. Genome-wide analyses of gene expression during mouse endochondral ossification. *PloS one*. 2010;5(1). doi:10.1371/JOURNAL.PONE.0008693
26. Duan X, Cai L, Schmidt EJ, et al. RNA-seq analysis of chondrocyte transcriptome reveals genetic heterogeneity in LG/J and SM/J murine strains. *Osteoarthritis and cartilage*. 2020;28(4):516-527. doi:10.1016/J.JOCA.2020.01.001
27. Li B, Balasubramanian K, Krakow D, Cohn DH. Genes uniquely expressed in human growth plate chondrocytes uncover a distinct regulatory network. *BMC genomics*. 2017;18(1). doi:10.1186/S12864-017-4378-Y
28. Vail DJ, Somoza RA, Caplan AI, Khalil AM. Transcriptome dynamics of long noncoding RNAs and transcription factors demarcate human neonatal, adult, and human mesenchymal stem cell-derived engineered cartilage. *Journal of tissue engineering and regenerative medicine*. 2020;14(1):29-44. doi:10.1002/TERM.2961
29. Cheung K, Barter MJ, Falk J, Proctor CJ, Reynard LN, Young DA. Histone ChIP-Seq identifies differential enhancer usage during chondrogenesis as critical for defining cell-type specificity. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2020;34(4):5317-5331. doi:10.1096/FJ.201902061RR
30. Zhang CH, Gao Y, Jadhav U, et al. Creb5 establishes the competence for Prg4 expression in articular cartilage. *Communications biology*. 2021;4(1). doi:10.1038/S42003-021-01857-0
31. Guo M, Liu Z, Willen J, et al. Epigenetic profiling of growth plate chondrocytes sheds insight into regulatory genetic variation influencing height. *eLife*. 2017;6:e29329.
32. Ohba S, He X, Hojo H, McMahon AP. Distinct Transcriptional Programs Underlie Sox9 Regulation of the Mammalian Chondrocyte. *Cell reports*. 2015;12(2):229-243. doi:10.1016/J.CELREP.2015.06.013
33. He X, Ohba S, Hojo H, McMahon AP. AP-1 family members act with Sox9 to promote chondrocyte hypertrophy. *Development (Cambridge, England)*. 2016;143(16):3012-3023. doi:10.1242/DEV.134502
34. Oh C do, Lu Y, Liang S, et al. SOX9 regulates multiple genes in chondrocytes, including genes encoding ECM proteins, ECM modification enzymes, receptors, and transporters. *PloS one*. 2014;9(9). doi:10.1371/JOURNAL.PONE.0107577
35. Craft AM, Rockel JS, Nartiss Y, Kandel RA, Alman BA, Keller GM. Generation of articular chondrocytes from human pluripotent stem cells. *Nature biotechnology*. 2015;33(6):638-645. doi:10.1038/NBT.3210
36. Lee Y, Choi J, Hwang NS. Regulation of lubricin for functional cartilage tissue regeneration: a review. *Biomaterials research*. 2018;22(1). doi:10.1186/S40824-018-0118-X
37. Richard D, Liu Z, Cao J, et al. Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk. *Cell*. 2020;0(0). doi:10.1016/j.cell.2020.02.057
38. Lee H, Lee YJ, Choi H, et al. SCARA5 plays a critical role in the commitment of mesenchymal stem cells to adipogenesis. *Scientific Reports 2017 7:1*. 2017;7(1):1-13. doi:10.1038/S41598-017-12512-2
39. Docheva D, Hunziker EB, Fässler R, Brandau O. Tenomodulin is necessary for tenocyte proliferation and tendon maturation. *Molecular and cellular biology*. 2005;25(2):699-705. doi:10.1128/MCB.25.2.699-705.2005
40. Hagan AS, Boylan M, Smith C, et al. Generation and validation of novel conditional flox and inducible Cre alleles targeting fibroblast growth factor 18 (Fgf18). *Developmental dynamics : an official publication of the American Association of Anatomists*. 2019;248(9):882-893. doi:10.1002/DVDY.85
41. Nakajima A, Shimizu S, Moriya H, Yamazaki M. Expression of fibroblast growth factor receptor-3 (FGFR3), signal transducer and activator of transcription-1, and cyclin-dependent kinase inhibitor p21 during endochondral ossification: differential role of FGFR3 in skeletal development and fracture repair. *Endocrinology*. 2003;144(10):4659-4668. doi:10.1210/EN.2003-0158

42. Ellman MB, Yan D, Ahmadinia K, Chen D, An HS, Im HJ. Fibroblast growth factor control of cartilage homeostasis. *Journal of cellular biochemistry*. 2013;114(4):735-742. doi:10.1002/JCB.24418
43. Davidson D, Blanc A, Filion D, et al. Fibroblast growth factor (FGF) 18 signals through FGF receptor 3 to promote chondrogenesis. *The Journal of biological chemistry*. 2005;280(21):20509-20515. doi:10.1074/JBC.M410148200
44. John Martin T. Parathyroid Hormone-Related Protein, Its Regulation of Cartilage and Bone Development, and Role in Treating Bone Diseases. *Physiological reviews*. 2016;96(3):831-871. doi:10.1152/PHYSREV.00031.2015
45. Iwamoto T, Nakamura T, Doyle A, et al. Pannexin 3 regulates intracellular ATP/cAMP levels and promotes chondrocyte differentiation. *The Journal of biological chemistry*. 2010;285(24):18948-18958. doi:10.1074/JBC.M110.127027
46. Miao D, Scutt A. Histochemical localization of alkaline phosphatase activity in decalcified bone and cartilage. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*. 2002;50(3):333-340. doi:10.1177/002215540205000305
47. Knorr T, Obermayr F, Bartnik E, Zien A, Aigner T. YKL-39 (chitinase 3-like protein 2), but not YKL-40 (chitinase 3-like protein 1), is up regulated in osteoarthritic chondrocytes. *Annals of the Rheumatic Diseases*. 2003;62(10):995-998. doi:10.1136/ard.62.10.995
48. Ling H, Recklies AD. The chitinase 3-like protein human cartilage glycoprotein 39 inhibits cellular responses to the inflammatory cytokines interleukin-1 and tumour necrosis factor-alpha. *The Biochemical journal*. 2004;380(Pt 3):651-659. doi:10.1042/BJ20040099
49. Skuntz S, Mankoo B, Nguyen MTT, et al. Lack of the mesodermal homeodomain protein MEOX1 disrupts sclerotome polarity and leads to a remodeling of the cranio-cervical joints of the axial skeleton. *Developmental biology*. 2009;332(2):383-395. doi:10.1016/J.YDBIO.2009.06.006
50. Muona A, Eklund L, Väisänen T, Pihlajaniemi T. Developmentally regulated expression of type XV collagen correlates with abnormalities in Col15a1(-/-) mice. *Matrix biology : journal of the International Society for Matrix Biology*. 2002;21(1):89-102. doi:10.1016/S0945-053X(01)00187-1
51. Lisignoli G, Codeluppi K, Todoerti K, et al. Gene array profile identifies collagen type XV as a novel human osteoblast-secreted matrix protein. *Journal of cellular physiology*. 2009;220(2):401-409. doi:10.1002/JCP.21779
52. Mun SA, Park J, Park KR, et al. Structural and Biochemical Characterization of EFhd1/Swiprosin-2, an Actin-Binding Protein in Mitochondria. *Frontiers in cell and developmental biology*. 2021;8. doi:10.3389/FCELL.2020.628222
53. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Current protocols in molecular biology*. 2015;109(1):21-29.
54. McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology*. 2010;28(5):495. doi:10.1038/nbt.1630
55. Yoshida H, Lareau CA, Ramirez RN, et al. The cis-Regulatory Atlas of the Mouse Immune System. *Cell*. 2019;176(4):897-912.e20. doi:10.1016/J.CELL.2018.12.036
56. Wu H, Whitfield TW, Gordon JAR, et al. Genomic occupancy of Runx2 with global expression profiling identifies a novel dimension to control of osteoblastogenesis. *Genome Biology*. 2014;15(3):1-17. doi:10.1186/GB-2014-15-3-R52/FIGURES/6
57. Hojo H, Saito T, He X, et al. Runx2 Regulates Chromatin Accessibility to Direct Skeletal Cell Programs. *SSRN Electronic Journal*. Published online December 21, 2021. doi:10.2139/SSRN.3988393
58. Alshenibr W, Tashkandi MM, Alsaqer SF, et al. Anabolic role of lysyl oxidase like-2 in cartilage of knee and temporomandibular joints with osteoarthritis. *Arthritis research & therapy*. 2017;19(1). doi:10.1186/S13075-017-1388-8
59. Snelling SJB, Davidson RK, Swingler TE, et al. Dickkopf-3 is upregulated in osteoarthritis and has a chondroprotective role. *Osteoarthritis and cartilage*. 2016;24(5):883-891. doi:10.1016/J.JOCA.2015.11.021
60. Sillat T, Barreto G, Clarijs P, et al. Toll-like receptors in human chondrocytes and osteoarthritic cartilage. *Acta orthopaedica*. 2013;84(6):585-592. doi:10.3109/17453674.2013.854666
61. Dateki S. ACAN mutations as a cause of familial short stature. *Clinical pediatric endocrinology : case reports and clinical investigations : official journal of the Japanese Society for Pediatric Endocrinology*. 2017;26(3):119-125. doi:10.1297/CPE.26.119

62. Lauing KL, Cortes M, Domowicz MS, Henry JG, Baria AT, Schwartz NB. Aggrecan is required for growth plate cytoarchitecture and differentiation. *Developmental biology*. 2014;396(2):224-236. doi:10.1016/J.YDBIO.2014.10.005
63. Gu J, Lu Y, Li F, et al. Identification and characterization of the novel Col10a1 regulatory mechanism during chondrocyte hypertrophic differentiation. *Cell death & disease*. 2014;5(10). doi:10.1038/CDDIS.2014.444
64. Andrade AC, Nilsson O, Barnes KM, Baron J. Wnt gene expression in the post-natal growth plate: regulation with chondrocyte differentiation. *Bone*. 2007;40(5):1361-1369. doi:10.1016/J.BONE.2007.01.005
65. Bennett CN, Longo KA, Wright WS, et al. Regulation of osteoblastogenesis and bone mass by Wnt10b. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(9):3324-3329. doi:10.1073/PNAS.0408742102
66. Schroeder N, Wuelling M, Hoffmann D, Brand-Saberi B, Vortkamp A. Atoh8 acts as a regulator of chondrocyte proliferation and differentiation in endochondral bones. *PLoS one*. 2019;14(8). doi:10.1371/JOURNAL.PONE.0218230
67. Boyde A, Lacourt M, Richard H, et al. Investigating the role of Retinoid X Receptor in cartilage development and homeostasis. *Osteoarthritis and Cartilage*. 2019;27:S177-S178. doi:10.1016/J.JOCA.2019.02.267
68. Li IMH, Liu K, Neal A, Clegg PD, de Val S, Bou-Gharios G. Differential tissue specific, temporal and spatial expression patterns of the Aggrecan gene is modulated by independent enhancer elements. *Scientific reports*. 2018;8(1). doi:10.1038/S41598-018-19186-4
69. Hou T, Jian C, Xu J, et al. Identification of EFHD1 as a novel Ca(2+) sensor for mitoflash activation. *Cell calcium*. 2016;59(5):262-270. doi:10.1016/J.CECA.2016.03.002
70. Wang D, Canaff L, Davidson D, et al. Alterations in the sensing and transport of phosphate and calcium by differentiating chondrocytes. *The Journal of biological chemistry*. 2001;276(36):33995-34005. doi:10.1074/JBC.M007757200
71. Shukunami C, Takimoto A, Miura S, Nishizaki Y, Hiraki Y. Chondromodulin-I and tenomodulin are differentially expressed in the avascular mesenchyme during mouse and chick development. *Cell and tissue research*. 2008;332(1):111-122. doi:10.1007/S00441-007-0570-8
72. Brandau O, Meindl A, Fässler R, Aszódi A. A novel gene, tendin, is strongly expressed in tendons and ligaments and shows high homology with chondromodulin-I. *Developmental dynamics: an official publication of the American Association of Anatomists*. 2001;221(1):72-80. doi:10.1002/DVDY.1126
73. Pryce BA, Watson SS, Murchison ND, Staverosky JA, Dünker N, Schweitzer R. Recruitment and maintenance of tendon progenitors by TGFbeta signaling are essential for tendon formation. *Development (Cambridge, England)*. 2009;136(8):1351-1361. doi:10.1242/DEV.027342
74. Koyama E, Shibukawa Y, Nagayama M, et al. A distinct cohort of progenitor cells participates in synovial joint and articular cartilage formation during mouse limb skeletogenesis. *Developmental biology*. 2008;316(1):62-73. doi:10.1016/J.YDBIO.2008.01.012
75. Clementz AG, Harris A. Collagen XV: exploring its structure and role within the tumor microenvironment. *Molecular cancer research: MCR*. 2013;11(12):1481-1486. doi:10.1158/1541-7786.MCR-12-0662
76. Nutt SL, Kee BL. The Transcriptional Regulation of B Cell Lineage Commitment. *Immunity*. 2007;26(6):715-725. doi:10.1016/J.IMMUNI.2007.05.010
77. Ludwig LS, Lareau CA, Bao EL, et al. Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. *Cell Reports*. 2019;27(11):3228-3240.e7. doi:10.1016/J.CELREP.2019.05.046
78. Yu S, Li P, Li B, Miao D, Deng Q. RelA promotes proliferation but inhibits osteogenic and chondrogenic differentiation of mesenchymal stem cells. *FEBS letters*. 2020;594(9):1368-1378. doi:10.1002/1873-3468.13739
79. Zhao Q, Lin L, Guo Y, et al. Matrix metalloproteinase-13, NF-κB p65 and interleukin-1β are associated with the severity of knee osteoarthritis. *Experimental and therapeutic medicine*. 2020;19(6). doi:10.3892/ETM.2020.8618
80. Olivetto E, Otero M, Marcu KB, Goldring MB. Pathophysiology of osteoarthritis: canonical NF-κB/IKKβ-dependent and kinase-independent effects of IKKα in cartilage degradation and chondrocyte differentiation. *RMD open*. 2015;1(Suppl 1). doi:10.1136/RMDOPEN-2015-000061
81. Kobayashi H, Chang SH, Mori D, et al. Biphasic regulation of chondrocytes by RelA through induction of anti-apoptotic and catabolic target genes. *Nature Communications* 2016 7:1. 2016;7(1):1-12. doi:10.1038/NCOMMS13336

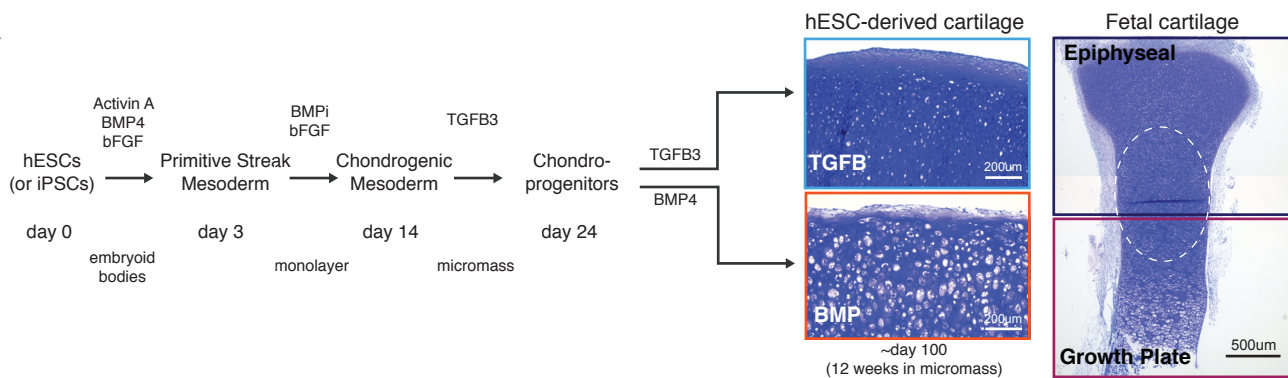
82. Saito T, Fukai A, Mabuchi A, et al. Transcriptional regulation of endochondral ossification by HIF-2 α during skeletal growth and osteoarthritis development. *Nature Medicine* 2010 16:6. 2010;16(6):678-686. doi:10.1038/NM.2146
83. Kobayashi H, Hirata M, Saito T, Itoh S, Chung U il, Kawaguchi H. Transcriptional Induction of ADAMTS5 Protein by Nuclear Factor- κ B (NF- κ B) Family Member RelA/p65 in Chondrocytes during Osteoarthritis Development. *The Journal of Biological Chemistry*. 2013;288(40):28620. doi:10.1074/JBC.M113.452169
84. Ushita M, Saito T, Ikeda T, et al. Transcriptional induction of SOX9 by NF- κ B family member RelA in chondrogenic cells. *Osteoarthritis and Cartilage*. 2009;17(8):1065-1075. doi:10.1016/J.JOCA.2009.02.003
85. Ogawa E, Maruyama M, Kagoshima H, et al. PEBP2/PEA2 represents a family of transcription factors homologous to the products of the Drosophila runt gene and the human AML1 gene. *Proceedings of the National Academy of Sciences of the United States of America*. 1993;90(14):6859-6863. doi:10.1073/PNAS.90.14.6859
86. Komori T, Yagi H, Nomura S, et al. Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell*. 1997;89(5):755-764. doi:10.1016/S0092-8674(00)80258-5
87. Ducy P, Zhang R, Geoffroy V, Ridall AL, Karsenty G. Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. *Cell*. 1997;89(5):747-754. doi:10.1016/S0092-8674(00)80257-3
88. Otto F, Thornell AP, Crompton T, et al. Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell*. 1997;89(5):765-771. doi:10.1016/S0092-8674(00)80259-7
89. Chen D, Kim DJ, Shen J, Zou Z, O'Keefe RJ. Runx2 plays a central role in Osteoarthritis development. *Journal of orthopaedic translation*. 2019;23:132-139. doi:10.1016/J.JOT.2019.11.008
90. Catheline SE, Hoak D, Chang M, et al. Chondrocyte-Specific RUNX2 Overexpression Accelerates Post-traumatic Osteoarthritis Progression in Adult Mice. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. 2019;34(9):1676-1689. doi:10.1002/JBMR.3737
91. Liao L, Zhang S, Gu J, et al. Deletion of Runx2 in Articular Chondrocytes Decelerates the Progression of DMM-Induced Osteoarthritis in Adult Mice. *Scientific reports*. 2017;7(1). doi:10.1038/S41598-017-02490-W
92. Findlay VJ, LaRue AC, Turner DP, Watson PM, Watson DK. Understanding the role of ETS-mediated gene regulation in complex biological processes. *Advances in cancer research*. 2013;119:1-61. doi:10.1016/B978-0-12-407190-2.00001-0
93. Herriges JC, Verheyden JM, Zhang Z, et al. FGF-Regulated ETV Transcription Factors Control FGF-SHH Feedback Loop in Lung Branching. *Developmental cell*. 2015;35(3):322-332. doi:10.1016/J.DEVCEL.2015.10.006
94. Iwamoto M, Tamamura Y, Koyama E, et al. Transcription factor ERG and joint and articular cartilage formation during mouse limb and spine skeletogenesis. *Developmental biology*. 2007;305(1):40-51. doi:10.1016/J.YDBIO.2007.01.037
95. Larmour C, Ohta Y, Okabe T, et al. Ets-transcription factors in formation and maintenance of permanent articular cartilage. *Osteoarthritis and Cartilage*. 2013;21:S20. doi:10.1016/J.JOCA.2013.02.063
96. Tomita M, Reinhold MI, Molkentin JD, Naski MC. Calcineurin and NFAT4 induce chondrogenesis. *The Journal of biological chemistry*. 2002;277(44):42214-42218. doi:10.1074/JBC.C200504200
97. Greenblatt MB, Ritter SY, Wright J, et al. NFATc1 and NFATc2 repress spontaneous osteoarthritis. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(49):19914-19919. doi:10.1073/PNAS.1320036110
98. Tardif G, Pelletier JP, Fahmi H, et al. NFAT3 and TGF- β /SMAD3 regulate the expression of miR-140 in osteoarthritis. *Arthritis research & therapy*. 2013;15(6). doi:10.1186/AR4387
99. Li SQ, Tu C, Wan L, et al. FGF-induced LHX9 regulates the progression and metastasis of osteosarcoma via FRS2/TGF- β / β -catenin pathway. *Cell division*. 2019;14(1). doi:10.1186/S13008-019-0056-6
100. Ferrari D, Kosher RA. Dlx5 is a positive regulator of chondrocyte differentiation during endochondral ossification. *Developmental biology*. 2002;252(2):257-270. doi:10.1006/DBIO.2002.0862
101. Komori T. The functions of Runx family transcription factors and Cbfb in skeletal development. *Oral Science International*. 2015;12(1):1-4. doi:10.1016/S1348-8643(14)00032-9

102. Ionescu A, Kozhemyakina E, Nicolae C, Kaestner KH, Olsen BR, Lassar AB. FoxA family members are crucial regulators of the hypertrophic chondrocyte differentiation program. *Developmental cell*. 2012;22(5):927-939. doi:10.1016/J.DEVCEL.2012.03.011
103. Charlier E, Neuville S, Plener Z, et al. The role of FoxA2 transcription factor as potential regulator of articular cartilage hypertrophy and OA progression. *Osteoarthritis and Cartilage*. 2019;27:S157-S158. doi:10.1016/J.JOCA.2019.02.231
104. van der Kraan PM, van den Berg WB. Chondrocyte hypertrophy and osteoarthritis: role in initiation and progression of cartilage degeneration? *Osteoarthritis and cartilage*. 2012;20(3):223-232. doi:10.1016/J.JOCA.2011.12.003
105. Wang F, Flanagan J, Su N, et al. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *The Journal of molecular diagnostics : JMD*. 2012;14(1):22-29. doi:10.1016/J.JMOLDX.2011.08.002
106. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*. 2013;10(12):1213.
107. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nature methods*. 2017;14(4):417. doi:10.1038/NMETH.4197
108. R Development Core Team. R: A Language and Environment for Statistical Computing. Published online 2008. <http://www.r-project.org>
109. Sonesson C, Love MI, Robinson MD, Floor SN. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research 2015 4:1521*. 2016;4:1521. doi:10.12688/f1000research.7563.1
110. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8
111. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society Series B (Methodological)*. Published online 1995:289-300.
112. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics (Oxford, England)*. 2016;32(18):2847-2849. doi:10.1093/bioinformatics/btw313
113. Ayturk UM, Jacobsen CM, Christodoulou DC, et al. An RNA-seq protocol to identify mRNA expression changes in mouse diaphyseal bone: Applications in mice with bone property altering Lrp5 mutations. *Journal of Bone and Mineral Research*. 2013;28(10):2081-2093. doi:10.1002/JBMR.1946
114. Mathelier A, Fornes O, Arenillas DJ, et al. JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*. 2016;44(D1):D110-D115. doi:10.1093/nar/gkv1176
115. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*. 2010;38(4):576-589.
116. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology*. 2012;16(5):284-287. doi:10.1089/omi.2011.0118
117. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9(4):357-359. doi:10.1038/nmeth.1923
118. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome biology*. 2008;9(9):R137. doi:10.1186/gb-2008-9-9-r137
119. Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics*. 2011;5(3):1752-1779. doi:10.1214/11-AOAS466
120. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033
121. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352
122. CRAN - Package dplyr. Accessed March 23, 2022. <https://cran.r-project.org/web/packages/dplyr/index.html>
123. Bailey TL, Boden M, Buske FA, et al. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*. Published online 2009. doi:10.1093/nar/gkp335

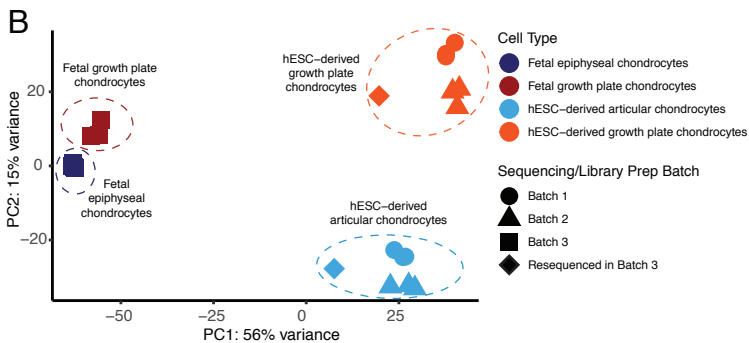
124. Karolchik D, Barber GP, Casper J, et al. The UCSC genome browser database: 2014 update. *Nucleic acids research*. 2013;42(D1):D764--D770.
125. Wang S, Sun H, Ma J, et al. Target analysis by integration of transcriptome and ChIP-seq data with BE-TA. *Nature Protocols* 2013 8:12. 2013;8(12):2502-2515. doi:10.1038/NPROT.2013.150
126. Tang Q, Chen Y, Meyer C, et al. A comprehensive view of nuclear receptor cancer cistromes. *Cancer research*. 2011;71(22):6940-6947. doi:10.1158/0008-5472.CAN-11-2091
127. Harrison PF. Varistran: Anscombe's variance stabilizing transformation for RNA-seq gene expression data. *Journal of Open Source Software*. 2017;2(16):257. doi:10.21105/JOSS.00257
128. Oki S, Ohta T, Shioi G, et al. CHIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO reports*. 2018;19(12):e46255. doi:10.15252/EMBR.201846255
129. Chokalingam K, Hunter S, Gooch C, et al. Three-dimensional in vitro effects of compression and time in culture on aggregate modulus and on gene expression and protein content of collagen type II in murine chondrocytes. *Tissue Engineering - Part A*. 2009;15(10):2807-2816. doi:10.1089/TEN.TEA.2008.0560

Figure 1

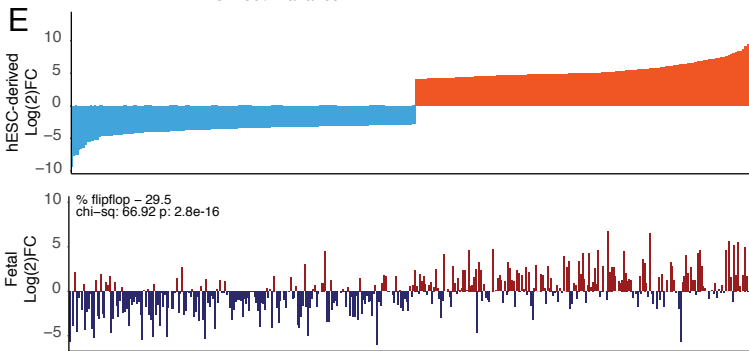
A



B



C



D

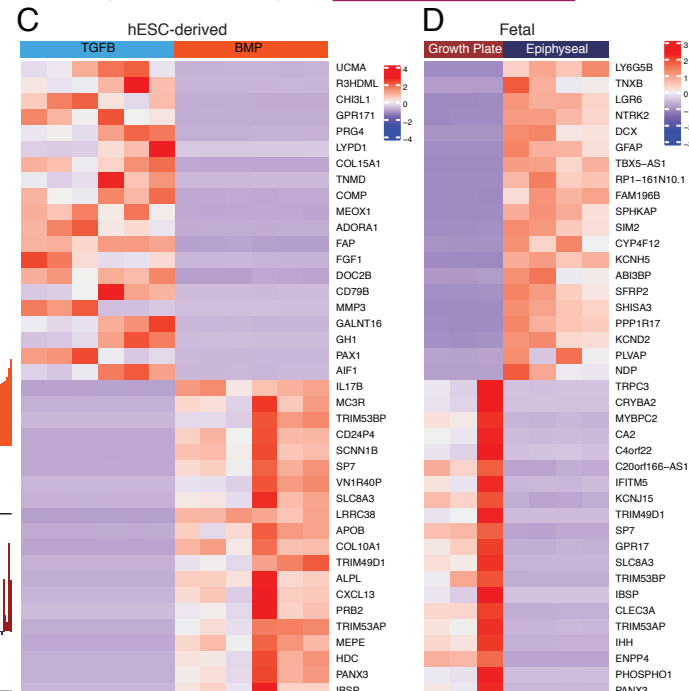


Figure 1. hESC-derived articular and growth plate chondrocytes have distinct transcriptional profiles that are similar to their respective fetal cartilage counterparts.

(A) Brief methods to generate hESC-derived cartilage accompanied by toluidine-blue staining sections of hESC-derived articular (TGFB) and growth plate-like (BMP) cartilage tissues and the epiphyseal and growth plate cartilage of a developing fetal cartilage (E59 proximal tibia) show distinct chondrocyte morphology and proteoglycan-rich matrix. Fetal dissection location noted is approximate, and dotted circle highlights overlap of chondrocytes with similar phenotype likely present in both samples following dissection.

(B) PCA plot of RNA-seq expression data from hESC-derived and fetal cartilages. Legend indicates cell type and sequencing batch.

(C) Expression heatmap of the top 20 differentially-expressed genes up- and down-regulated when comparing hESC-derived articular and growth plate cartilage. Red/blue color scale indicates Z-score expression values across samples in each plot. Columns indicate biological replicates.

(D) Expression heatmap of the top 20 differentially-expressed genes up- and down-regulated when comparing fetal epiphyseal and growth plate cartilage tissues. Red/blue color scale indicates Z-score expression values across samples in each plot. Columns indicate biological replicates.

(E) The top 100 differentially-expressed genes up- and down-regulated in the hESC-derived cartilages (top) were compared with equivalent log(2)FC values from the fetal cartilage (bottom).

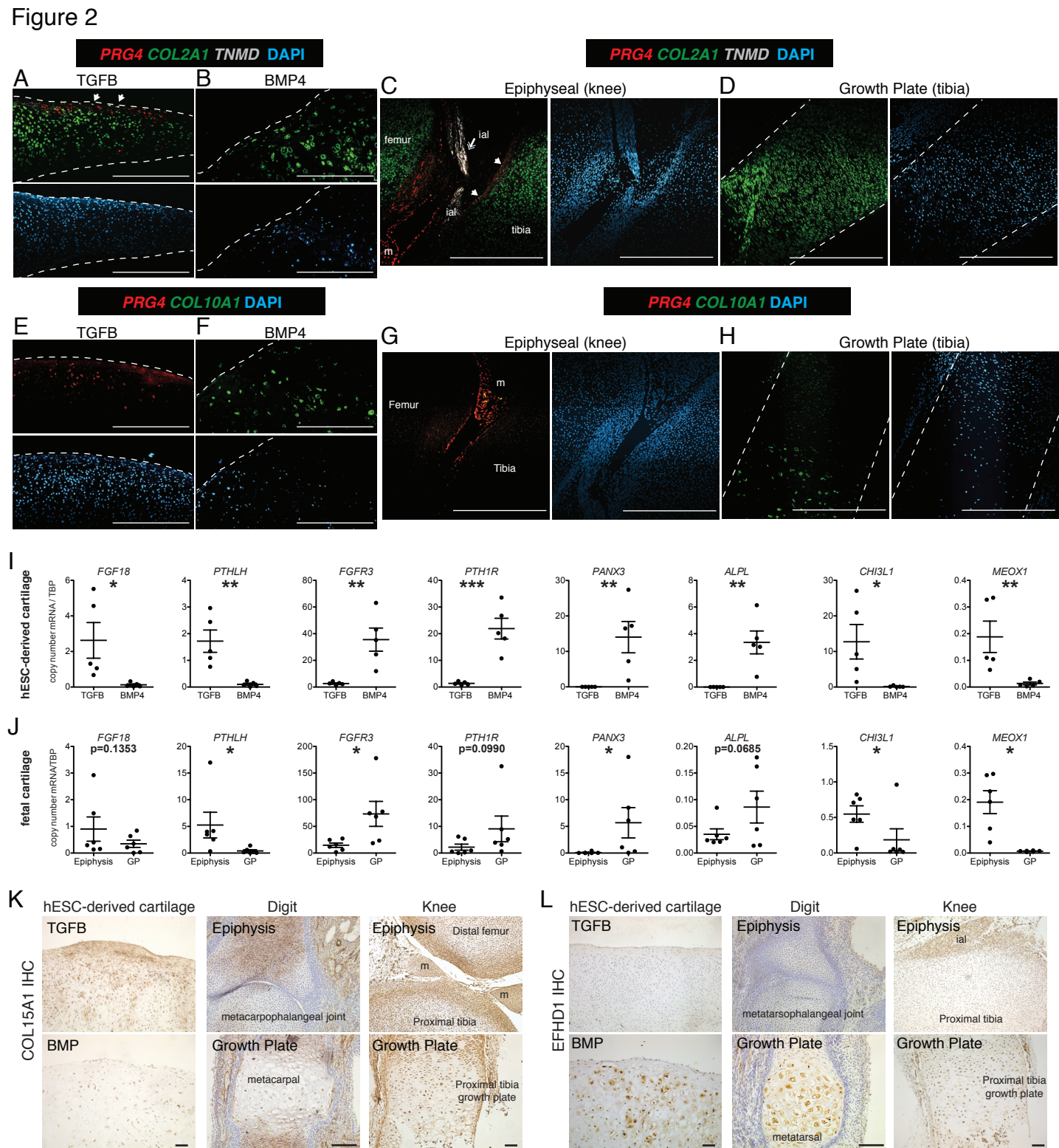


Figure 2. Validation of differential gene and protein expression in hESC-derived articular and growth plate cartilage and fetal epiphyseal and growth plate cartilage.

(A-H) Confocal microscopy of hESC-derived and epiphyseal and growth plate fetal cartilage following in situ hybridization for indicated probes (RNAscope) and counterstained with DAPI (adjacent images). White arrows indicate TNMD staining at cartilage surface, white double arrow indicates TNMD staining in intra-articular ligament. Scale bar, 500 μ m. ial, intra-articular ligament; m, meniscus

(I-J) Quantitative RT-PCR of DEGs in hESC-derived cartilage (I, n=5 independent experiments with 3-6 replicates per experiment) and fetal cartilage (J). Chondrocytes were isolated from the epiphysis and growth plate (GP) of the distal femur and proximal tibia (n=3 per site at E59, E67 and E72). * $p<0.05$, ** $p<0.01$, *** $p<0.001$

(K-L) Immunohistochemistry (IHC, brown staining) was used to validate the expression of indicated proteins within cartilage and joint tissues as indicated. Scale bar, 100 μ m. ial, intra-articular ligament; m, meniscus. Sections counterstained with Mayer's hematoxylin (blue).

Figure 3

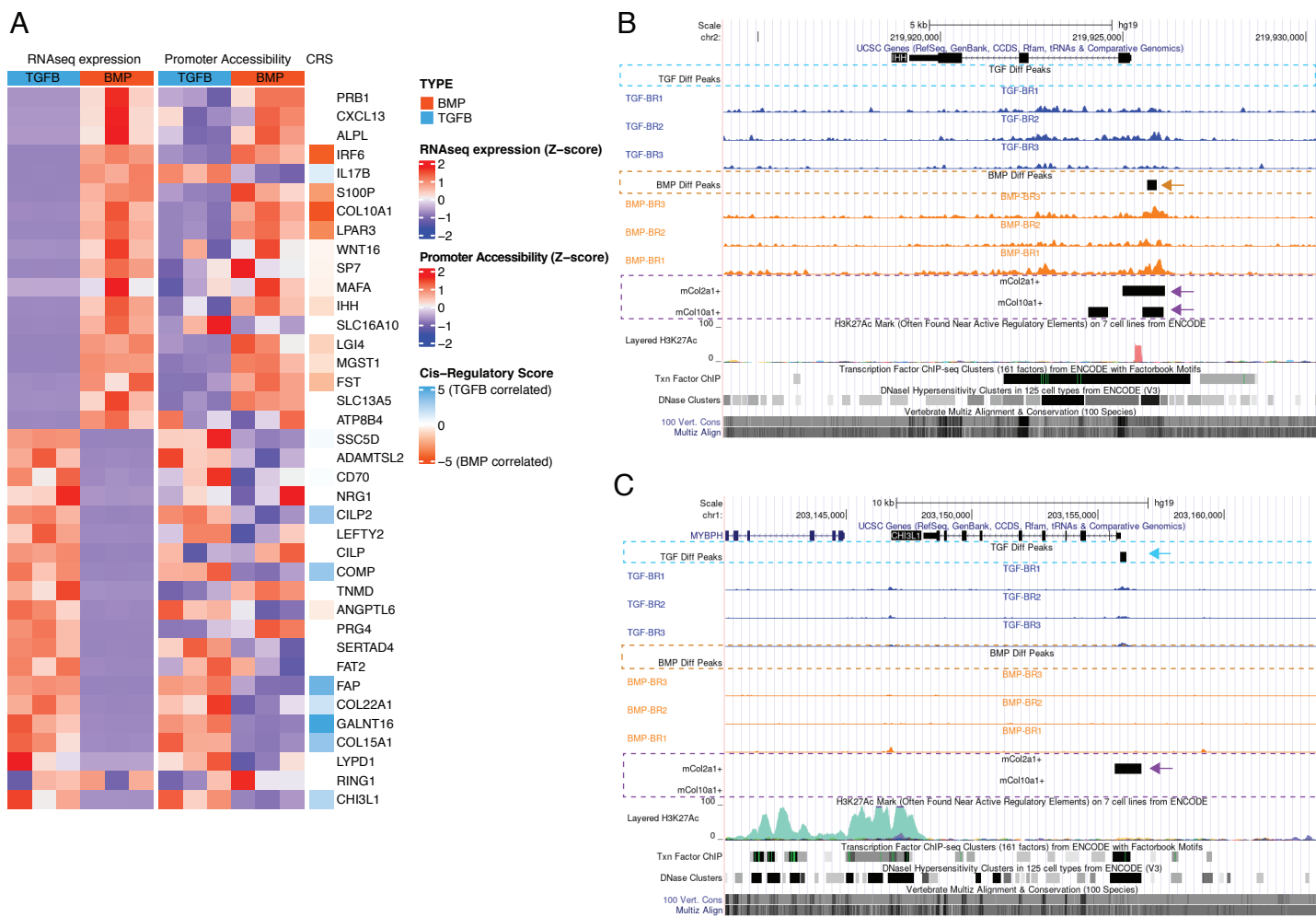


Figure 3. Epigenetic profiling of hESC-derived articular and growth plate chondrocytes. (A) The expression (left), gene-promoter accessibility (middle), and metric of cis-regulatory activity (cis-regulatory score, CRS, right) values of the top 20 differentially-expressed genes in each lineage. Red/blue color scale indicates Z-score expression/accessibility values across samples in each plot. Blue/orange scale indicates lineage-bias in cis-regulatory metric. (B-C) Representative differentially expressed gene locus in each lineage show differentially accessible regions (DARs) at the promoter corresponding to the same lineage. IHH (B) is differentially expressed in the growth plate lineage while CHI3L1 (C) is differentially expressed in the articular cartilage lineage. Arrows highlight DAR of interest in respective tracks.

Figure 4

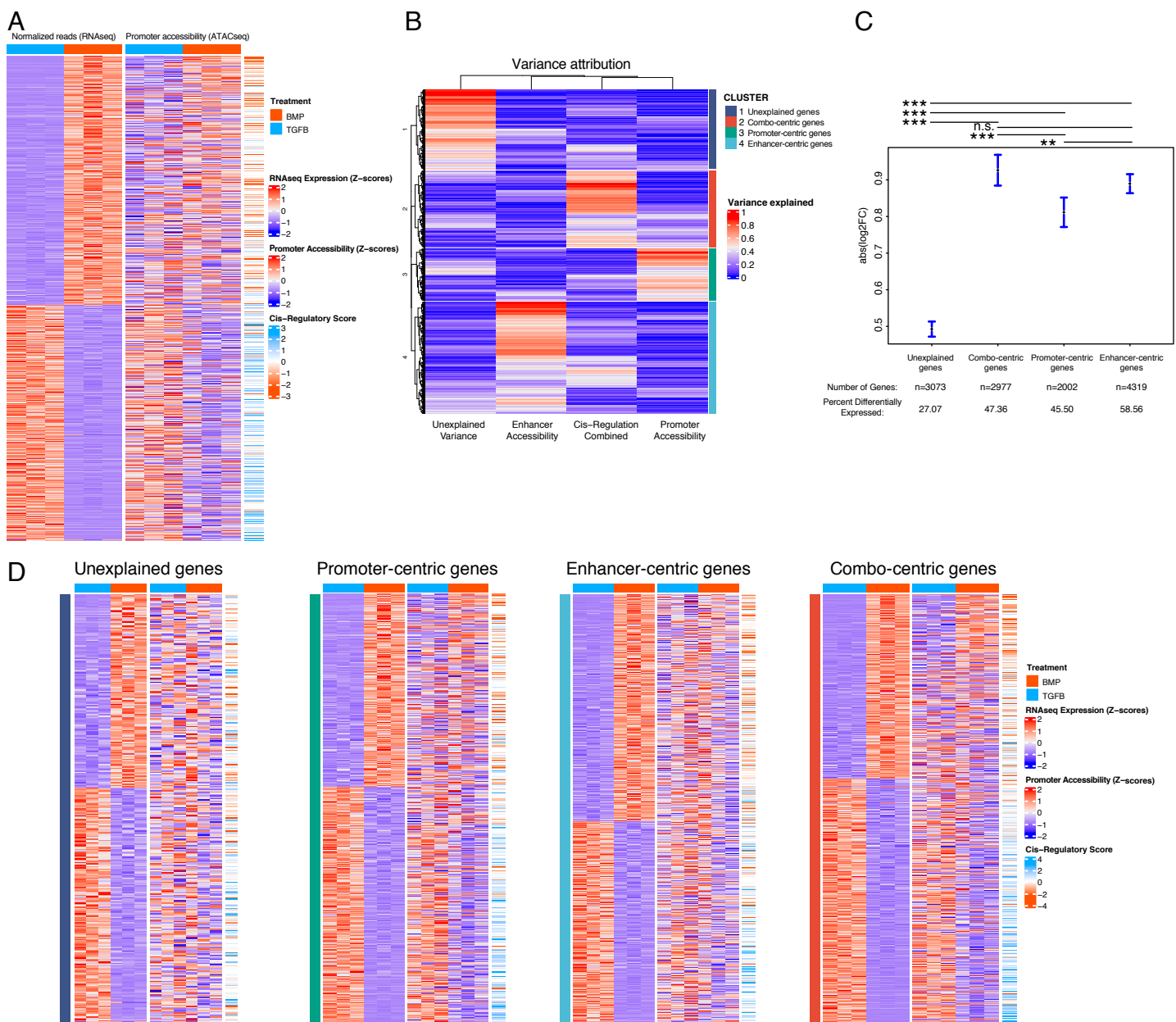


Figure 4. Variance in expression of DEGs can be attributed to different classes of regulatory elements (gene regulatory behavior). (A) Expression (left), gene-promoter accessibility (middle) and cis-regulatory metric (right) of all genes with expression logFC > 1 across lineages. Red/blue, blue-orange color scale indicates Z-score of indicated metric across samples in each indicated plot. (B) For each gene in (A), the proportion of variance in expression which can be explained by regressing on individual accessibility metrics is shown in red/blue colour scale (red = more variance). Hierarchical clustering dendrogram and cluster identity are shown on left and right (respectively), indicating the four clusters of regulatory behavior identified. (C) LogFC values of genes clustered by regulatory behavior. Significance bars indicate Tukey post-hoc corrected p-values. Proportion of significant DE genes in each cluster are indicated (see Table S3). n.s., not significant; *p < 0.05; **p < 0.01; ***p < 0.001. (D) Similar plots to (A), for genes clustered by regulatory behavior. Within each heatmap genes are hierarchically clustered by expression logFC. Color scales as in (A).

Figure 5

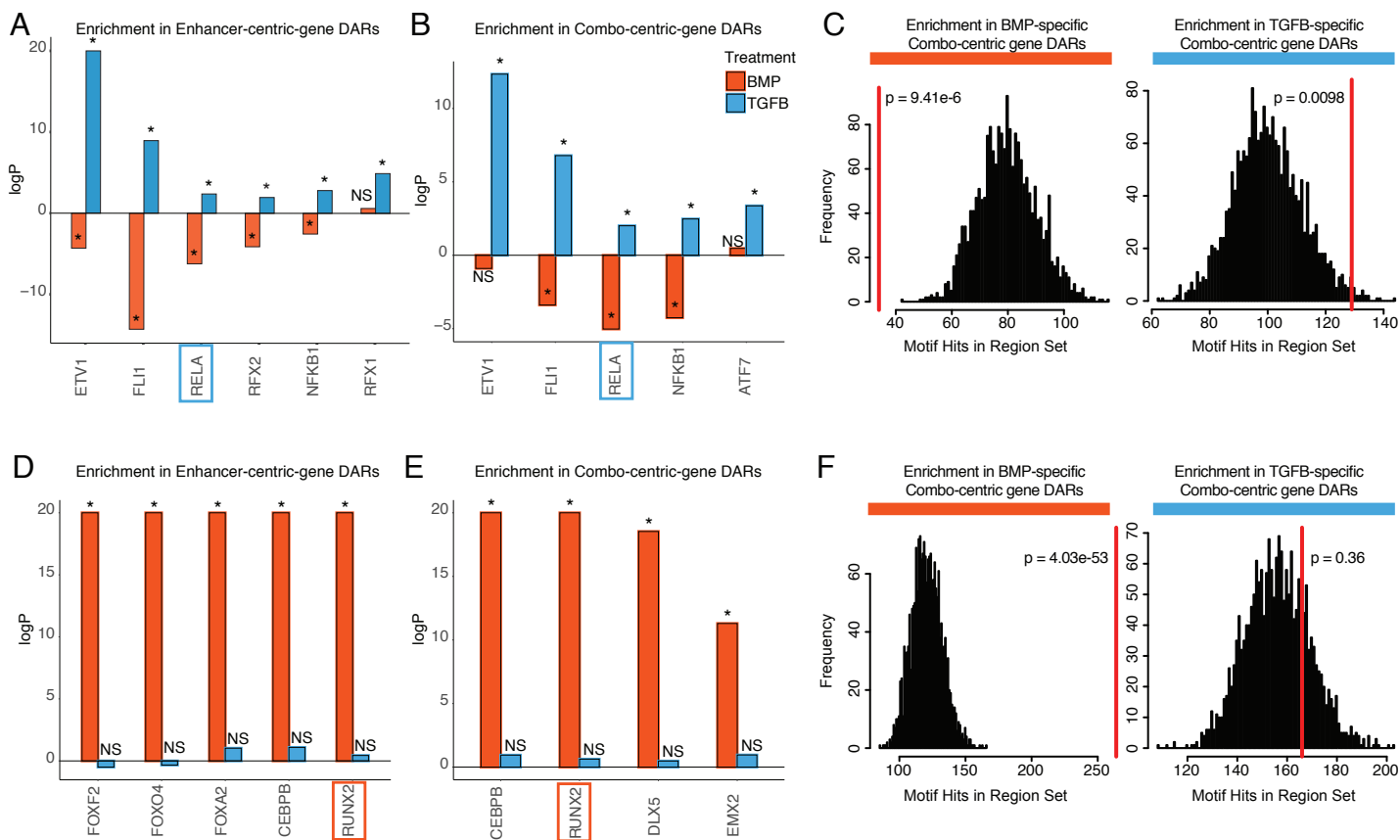


Figure 5. Identifying putative lineage-delineating transcription factors. Enrichment test results comparing the occurrence of the indicated motif in TGFB or BMP-biased DARs relative to randomized backgrounds. (A) TFs differentially expressed (DE) in TGFB-treated articular chondrocytes, testing motif occurrence in TGFB or BMP-biased DARs around enhancer-centric DEGs. * $p < 0.05$; NS, not significant. (B) TFs DE in TGFB-treated articular chondrocytes, testing motif occurrence in TGFB or BMP-biased DARs around combo-centric DEGs. * $p < 0.05$; NS, not significant. (C) Enrichment histogram of RELA motif occurrence in BMP (left) and TGFB (right)-biased DARs around combo-centric genes DE in their respective lineages. Red line indicates target set value, black bars indicate occurrences in randomized sets. (D) TFs DE in BMP-treated growth plate chondrocytes, testing motif occurrence in TGFB or BMP-biased DARs around enhancer-centric DEGs. * $p < 0.05$; NS, not significant. (E) TFs DE in BMP-treated growth plate chondrocytes, testing motif occurrence in TGFB or BMP-biased DARs around combo-centric DEGs. * $p < 0.05$; NS, not significant. (F) Enrichment histogram of RUNX2 motif occurrence in BMP (left) and TGFB (right)-biased DARs around combo-centric genes DE in their respective lineages. Red line indicates target set value, black bars indicate occurrences in randomized sets.

Figure 6

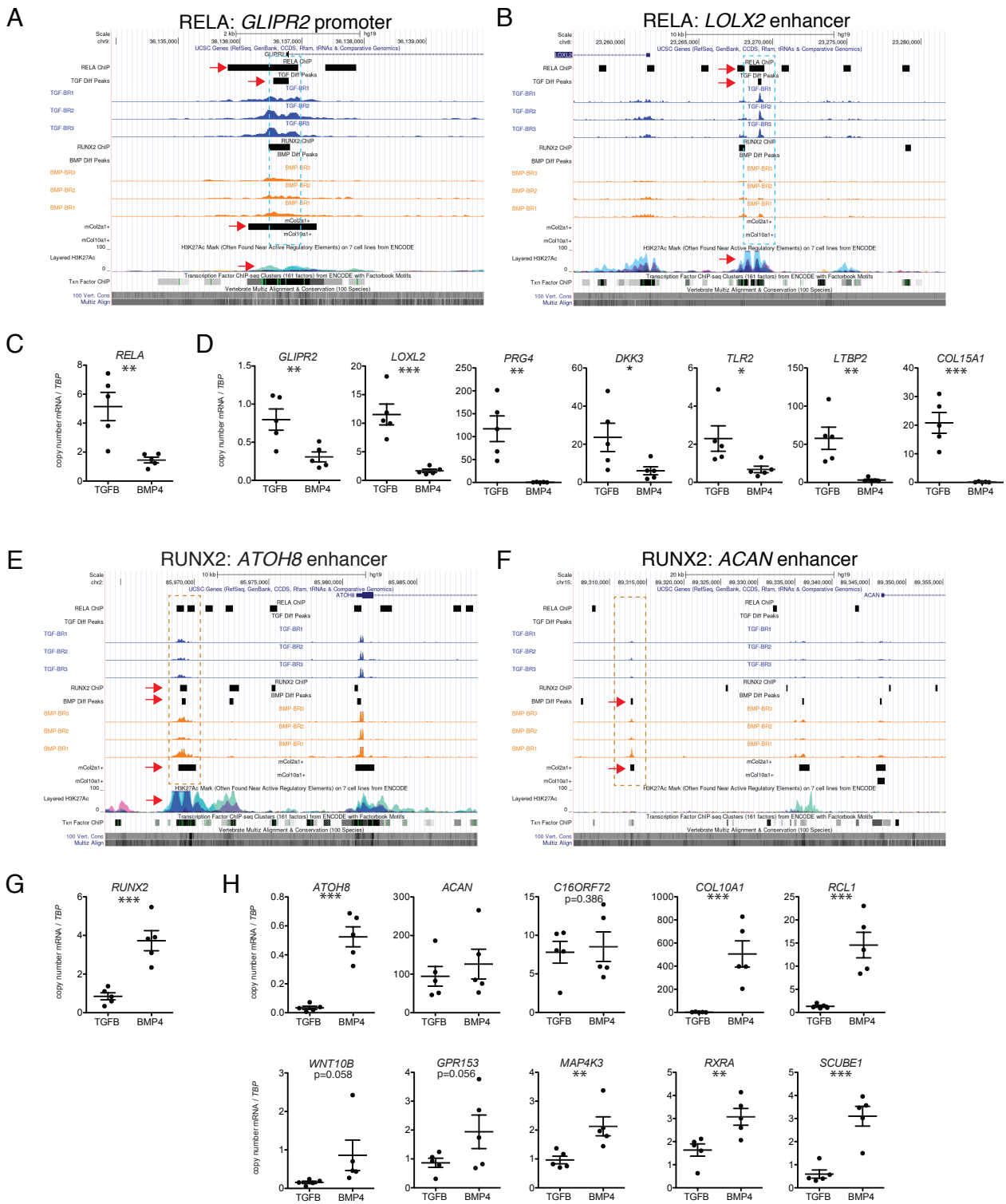


Figure 6. Putative targets of TF regulation in hESC-derived articular and growth plate chondrocytes. (A-B) Two representative differentially accessible regions (blue dashed boxes) in hESC-derived articular chondrocytes (TGF β -treated) which contain RELA binding motifs: near the promoter of *GLIPR2* (A) and an intronic enhancer of *LOXL2* (B). Region of interest indicated by red arrows in relevant tracks. (C) RELA is differentially expressed in TGF β -treated articular chondrocytes. ** $p < 0.01$. (D) Expression of selected genes with putative RELA binding motifs was quantified by qRT-PCR. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (E-F) Two representative differentially accessible regions (orange dashed boxes) in hESC-derived growth plate chondrocytes (BMP-treated) which contain RUNX2 binding motifs: near the promoter of *ATOH8* (E) and an enhancer of *ACAN* (F). Region of interest indicated by red arrows in relevant tracks. (G) RUNX2 is differentially expressed in BMP-treated growth plate chondrocytes. *** $p < 0.001$. (H) Expression of selected genes with putative RUNX2 binding motifs was quantified by qRT-PCR. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Supplemental Figures

Figure S1. Gene-set GO biological process enrichments for the top 200 gene loadings of PC1 and PC2.

Figure S2. Direction sharing of top DEGs (by $\log(2)$ FC or p-value) between in vitro-derived and primary tissue samples.

Figure S3. Transcription factor expression and overlap between hESC-derived and human fetal chondrocytes.

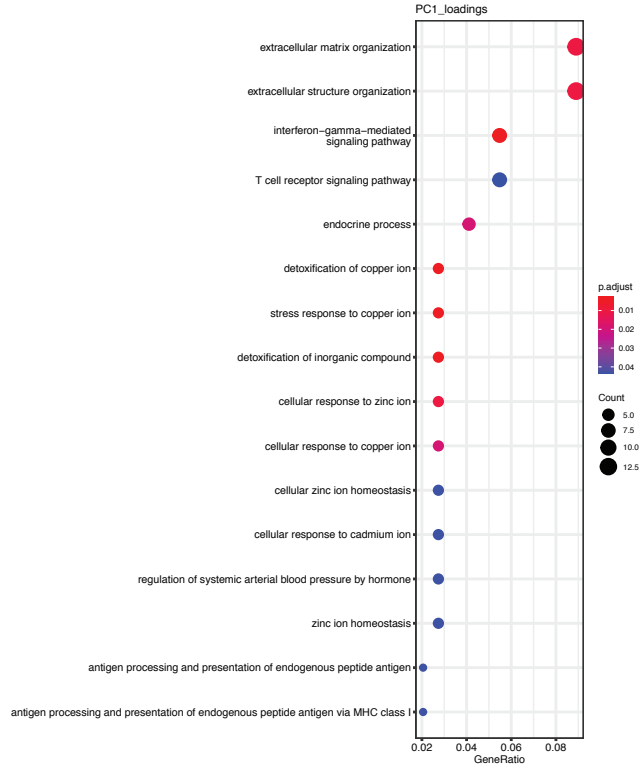
Figure S4. Top differentially expressed TFs in hESC-derived ACs and GPCs.

Figure S5. Motif enrichment is not correlated with sequence complexity.

Figure S6. Validation of TF interaction with putative regulatory elements by ChIP-qPCR.

Supplemental Figure 1

A



B

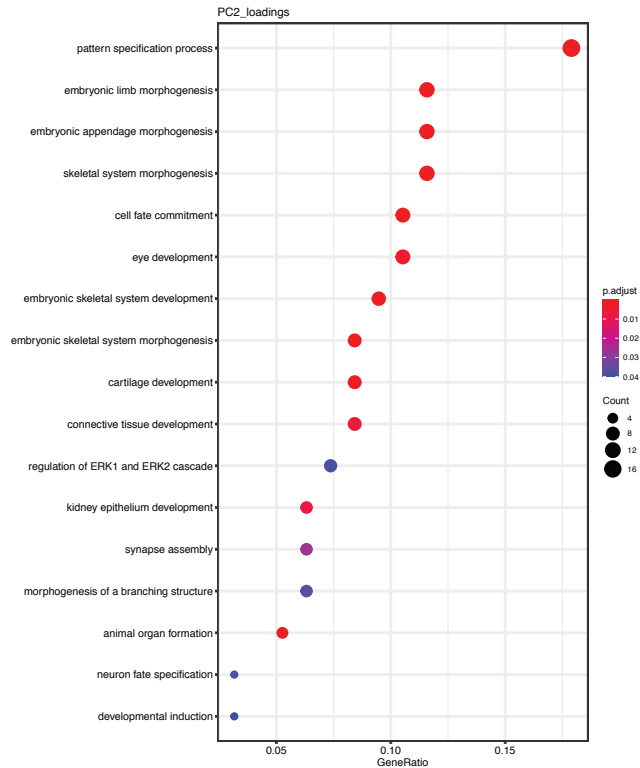


Figure S1. Gene-set GO biological process enrichments for the top 200 gene loadings of PC1 and PC2. (A) Gene-set GO biological process enrichments for the top 200 gene loadings of PC1. (B) Gene-set GO biological process enrichments for the top 200 gene loadings of PC2.

Supplemental Figure 2

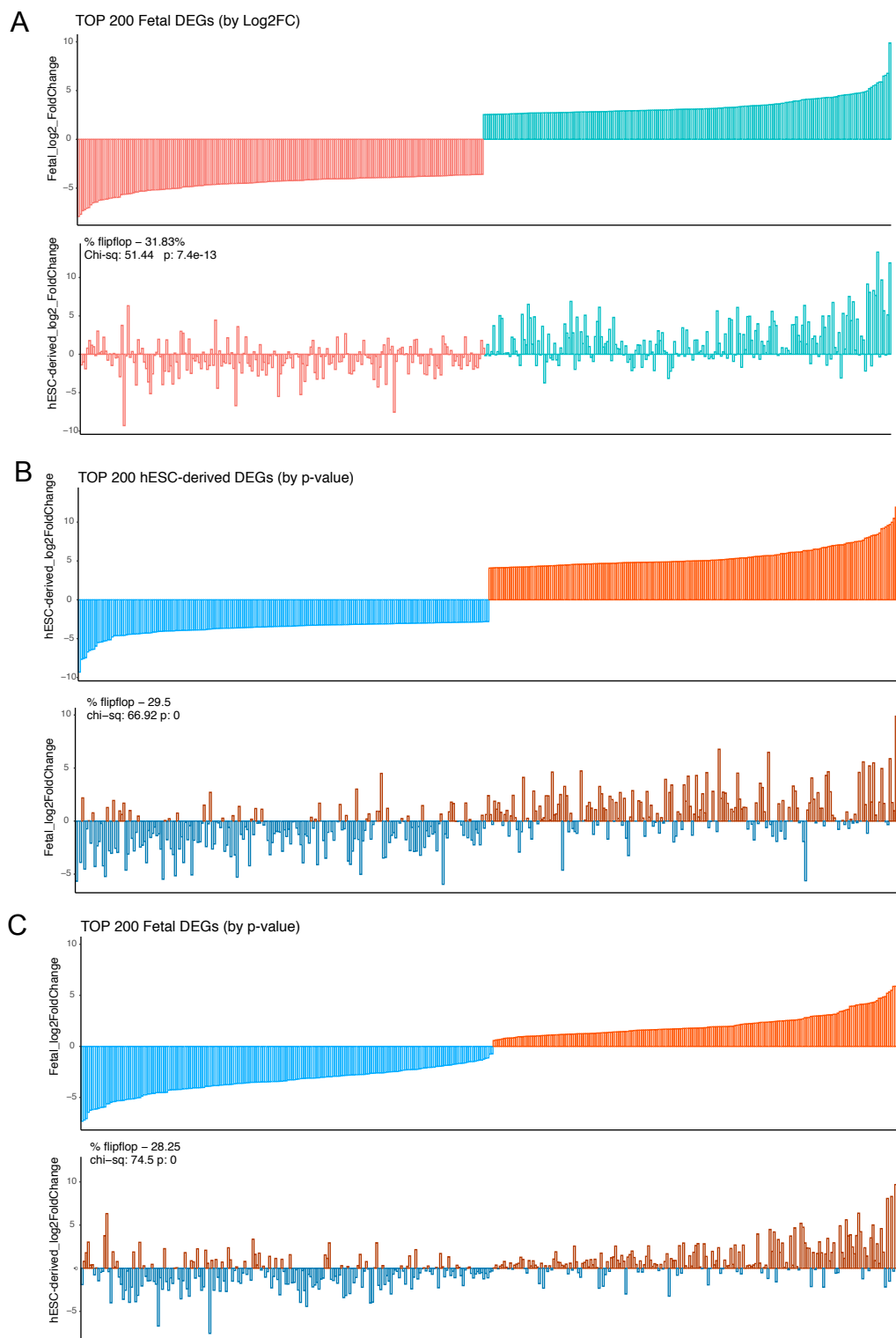


Figure S2. Direction sharing of top DEGs (by $\log_2(\text{FC})$ or p-value) between in vitro-derived and primary tissue samples. (A) The top 200 differentially-expressed genes ordered by $\log_2(\text{FC})$ (100 each up- and down-regulated) in the fetal articular and fetal growth plate cartilage (top) were compared with equivalent $\log_2(\text{FC})$ values from the hESC-derived cartilages (bottom). (B) The $\log_2(\text{FC})$ of the top 200 differentially-expressed genes with lowest p-value (100 each up- and down-regulated) in the hESC-derived cartilages (top) were compared to the equivalent $\log_2(\text{FC})$ values from the fetal articular and fetal growth plate cartilage (bottom). (C) The $\log_2(\text{FC})$ of the top 200 differentially-expressed genes with lowest p-value (100 each up- and down-regulated) in the fetal cartilages (top) were compared to the equivalent $\log_2(\text{FC})$ values from the hESC-derived cartilages (bottom).

Supplemental Figure 3

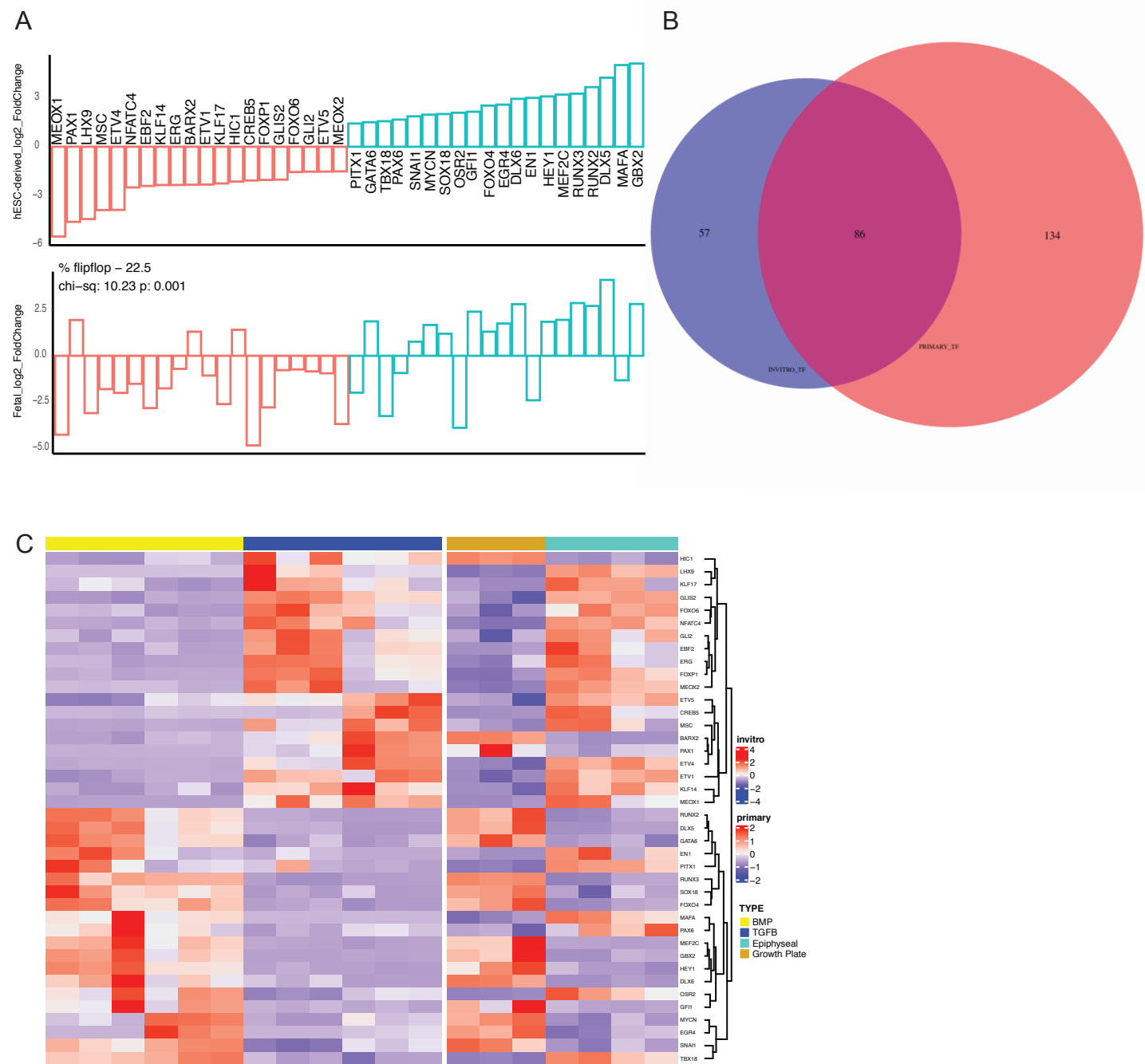


Figure S3. Transcription factor expression and overlap between hESC-derived and human fetal chondrocytes. (A) Direction sharing of the top 20 differentially-expressed transcription factors up- and down-regulated when comparing HESC-derived articular and growth-plate chondrocytes (top), along with the equivalent \log_2 FC values from the fetal tissue samples (bottom).

(B) Venn diagram of the overlap between transcription factors defined as differentially expressed in either hESC-derived (purple) or fetal (salmon) chondrocytes.

(C) Expression heatmap of the top 20 differentially-expressed transcription factors up- and down-regulated when comparing hESC-derived articular and growth-plate chondrocytes (left plots), with the equivalent expression of fetal articular and fetal growth plate cartilage (right). Red/blue color scale indicates Z-score expression values across samples in each plot. Columns indicate biological replicates.

Supplemental Figure 4

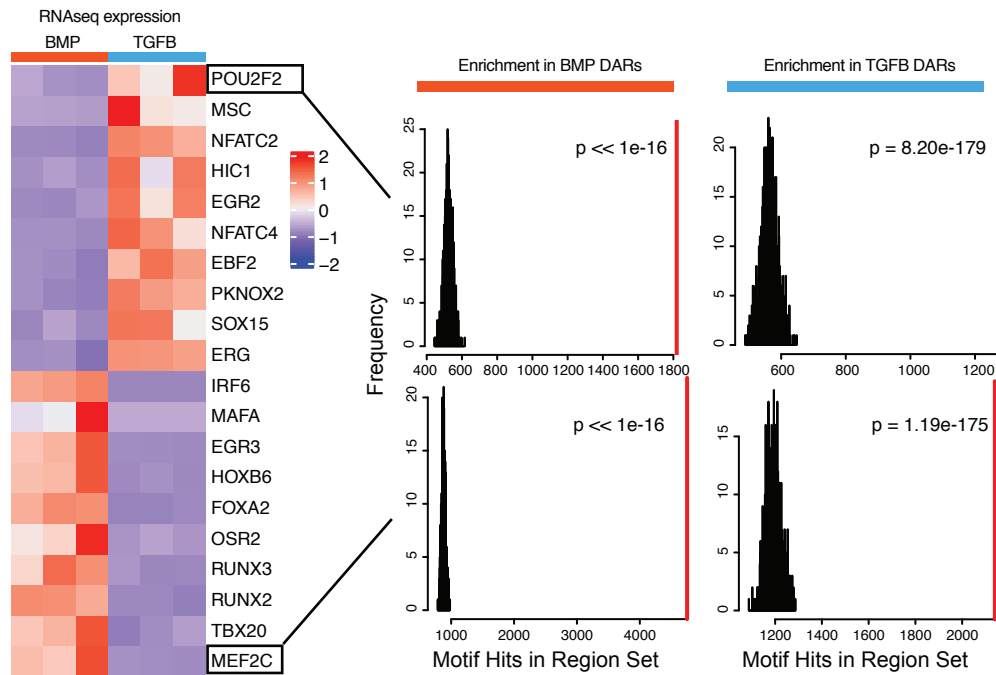


Figure S4. Top differentially expressed TFs in hESC-derived ACs and GPCs. (Left) Expression heatmap of the top 10 differentially-expressed transcription factors in each lineage. (Right) For indicated TF, the number of hits of that motif in putative enhancer sequences biased towards either lineage (red line), compared to randomly-sampled sequence sets (black lines/histogram).

Supplemental Figure 5

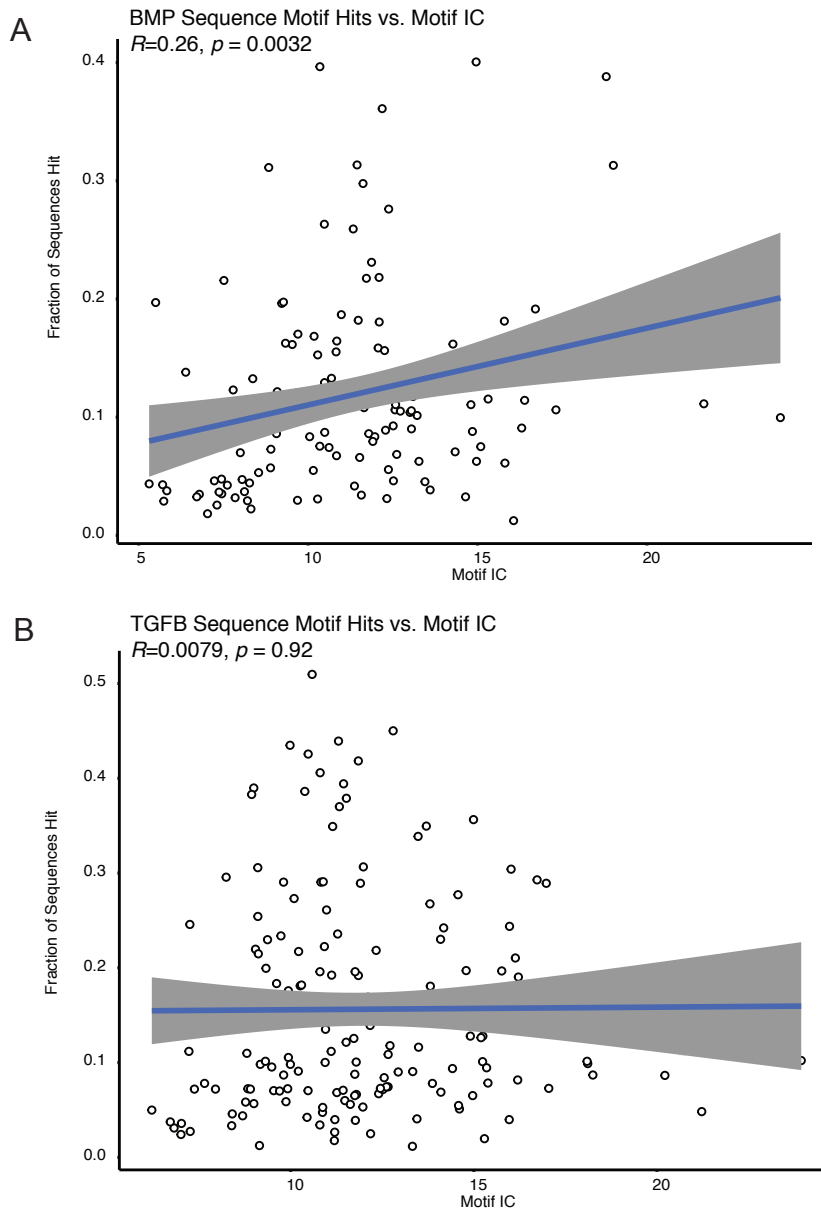


Figure S5. Motif enrichment is not correlated with sequence complexity.

(A) The number of motif hits within the BMP-biased peaks for a particular transcription factor, compared against the information content (IC) of that factor's binding motif. The Pearson correlation (R) and significance values of the relationship between IC and motif hits are shown in the top left.

(B) The number of motif hits within the TGFB-biased peaks for a particular transcription factor, compared against the information content (IC) of that factor's binding motif.

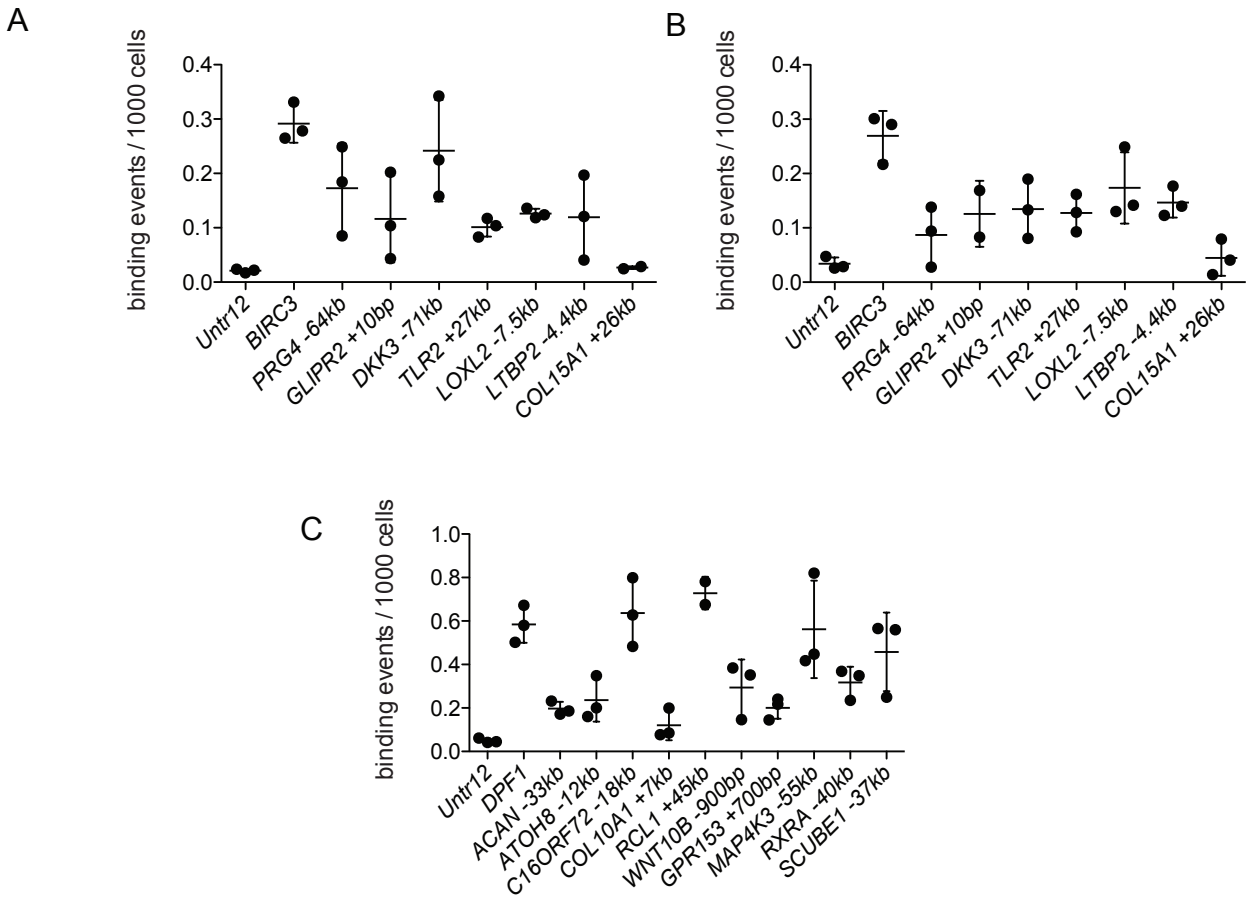


Figure S6. Validation of TF interaction with putative regulatory elements by ChIP-qPCR. (A-B) ChIP-qPCR of RELA results for differentially accessible peaks near target genes show enrichment of these sequences compared to a negative (gene desert) control (Untr12). BIRC3 is a positive RELA control. A and B represent two independent pools of TGFB-treated articular chondrocytes. (C) ChIP-qPCR of RUNX2 results for differentially accessible peaks near target genes show enrichment of these sequences compared to a negative (gene desert) control (Untr12). DPF1 is a positive RUNX2 control. Primers are provided in Supplemental Table 5.