



Transcriptional profiling of mESC-derived tendon and fibrocartilage cell fate switch

Deepak A. Kaji¹, Angela M. Montero¹, Roosheel Patel² & Alice H. Huang¹  

The transcriptional regulators underlying induction and differentiation of dense connective tissues such as tendon and related fibrocartilaginous tissues (meniscus and annulus fibrosus) remain largely unknown. Using an iterative approach informed by developmental cues and single cell RNA sequencing (scRNA-seq), we establish directed differentiation models to generate tendon and fibrocartilage cells from mouse embryonic stem cells (mESCs) by activation of TGF β and hedgehog pathways, achieving 90% induction efficiency. Transcriptional signatures of the mESC-derived cells recapitulate embryonic tendon and fibrocartilage signatures from the mouse tail. scRNA-seq further identify retinoic acid signaling as a critical regulator of cell fate switch between TGF β -induced tendon and fibrocartilage lineages. Trajectory analysis by RNA sequencing define transcriptional modules underlying tendon and fibrocartilage fate induction and identify molecules associated with lineage-specific differentiation. Finally, we successfully generate 3-dimensional engineered tissues using these differentiation protocols and show activation of mechanotransduction markers with dynamic tensile loading. These findings provide a serum-free approach to generate tendon and fibrocartilage cells and tissues at high efficiency for modeling development and disease.

¹Department of Orthopaedics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ²Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ✉email: ah364@cumc.columbia.edu

Tendons are dense, connective tissues whose load-bearing function is enabled by a fibrous extracellular matrix (ECM) composed of aligned type I collagen and other minor components^{1–3}. With injury, tendon ECM is replaced by disorganized scar, leading to impaired function. Regenerative strategies for tendon are limited as the basic cell and molecular regulators underlying tendon differentiation remain largely unknown.

The defining transcription factors for tendon cell fate, identified from developmental studies, include *Scx*, *Mkx*, *Egr1*, and *Egr2*. Although *Scx* is the earliest marker for tendon progenitors, it is not required for tendon induction or maintenance^{4,5}. Similarly, null mutations in *Mkx*, *Egr1*, and *Egr2* do not result in overt embryonic tendon phenotypes^{6–8}. In addition to tendons, *Scx* is also detected in related connective tissues, such as the meniscus and the annulus fibrosus of the intervertebral disc^{9,10}. Compared to tendon, these tissues have mixed fibrous and cartilage (fibrocartilage) elements, characterized by the presence of type II collagen and proteoglycans, in addition to type I collagen¹¹.

To date, the TGF β pathway remains the primary signaling pathway identified for mammalian tendons, as it is required for tendon induction and maintenance^{12,13}. Interestingly, TGF β also induces the chondrogenic cell fate as it is required for induction of skeletal cartilage in vivo and is frequently used to induce cartilage in vitro^{14–17}. Although tendon, fibrocartilage, and cartilage cell fates exist along a continuum and arise from common mesenchymal progenitors^{16,18}, the transcriptional and molecular signals that regulate the switch between these tissues have not been defined. Large-scale transcriptomic profiling efforts such as ENCODE and single-cell RNA sequencing (scRNA-seq) atlases consistently omit dense connective tissues such as tendons and fibrocartilage^{19–21}. Only two transcriptomic studies for embryonic mouse tendon have been carried out using microarray and RNA sequencing (RNA-seq) of sorted ScxGFP cells; however, these analyses bypassed initiating events underlying *Scx* induction^{22,23}. Thus, the transcriptional and molecular regulators that govern tendon induction have still not been identified.

Stem cell differentiations are ideal models to investigate the regulators of cell fate and lineage specification. Currently, there are very few protocols for directed differentiation of tendon cells from pluripotent sources^{24,25}. These prior work with embryonic stem cells (ESCs) and induced pluripotent stem cells used limited markers to confirm tendon cell fate²⁵, and resulting induction efficiencies were limited (~6–18% efficiency) or not reported^{24,26}.

In this study, we established models of tendon and fibrocartilage induction by leveraging developmental signals to differentiate mouse embryonic stem cells (mESCs). Using single-cell RNA sequencing (scRNA-seq), we defined our differentiated *Scx* populations against their relevant in vivo embryonic counterparts. Informed by scRNA-seq, we refined the signaling environment to improve final induction efficiency to ~90% and uncovered retinoic acid as a molecular driver of TGF β -induced tendon versus fibrocartilage fates. We further profiled temporal trajectories of tendon and fibrocartilage induction using RNA sequencing (RNA-seq) to identify factors regulating induction. Finally, we successfully generated three-dimensional (3D) engineered tissues using these defined media and showed enhancement or maintenance of tendon and fibrocartilage differentiation, respectively, in concert with activation of mechanotransduction pathways in response to dynamic tensile loading. These results represent a comprehensive investigation of tendon and fibrocartilage induction from pluripotent progenitors and establish a model system for

studying tendon development and tendon mechanobiology in vitro.

Results

Derivation of ScxGFP tendon cells from mESCs by activation of TGF β and hedgehog signaling. mESCs derived from mouse blastocysts containing the ScxGFP reporter⁹ were aggregated to form embryoid bodies (D0–D2) in the presence of BMP inhibitor LDN-193189, followed by 4 days of paraxial mesoderm (PM) specification using Wnt agonist CHIR and LDN-193189 (D2–D6, Fig. 1a)²⁷. For tendon differentiation, we then cultured PM cells with tenogenic induction media containing TGF β 1 for an additional 4 days (D6–D10). FGF2 was used in all conditions throughout tenogenic differentiation (D6–D10) to maintain cell survival and served as baseline control (Base) for tendon induction^{27–29}. With FGF2-only, we observed a modest level of ScxGFP expression (15–35%) at D10 (Fig. 1b–e). The addition of TGF β 1 enhanced ScxGFP+ induction efficiency between 40–80% (Fig. 1b–e). Interestingly, we identified two ScxGFP+ populations in the TGF β 1 condition, which we designated ScxGFP high (ScxGFP^{hi}) and ScxGFP low (ScxGFP^{lo}). The addition of TGF β 1 increased the ScxGFP^{hi} population relative to Base (Fig. 1b–d). Since ScxGFP induction was observed in the Base condition, we considered the possibility this may be the result of autocrine TGF β signaling. We therefore included the TGF β inhibitor SB-43154 with FGF2, and found that inhibition of TGF β signaling inhibited ScxGFP^{hi} induction, but total ScxGFP induction was not significantly affected (Fig. 1b, e).

Despite robust ScxGFP induction, TGF β 1 supplementation resulted in ~50% overall ScxGFP induction efficiency at D10, suggesting ~50% of the culture remained undifferentiated or adopted alternative fates. To improve tenogenesis, we considered other relevant signals involved in tendon induction. Since axial tendons are derived from a subcompartment of the sclerotome termed the syndetome³⁰, we considered sclerotomal induction signals, such as sonic hedgehog, that may improve induction^{31–35}. Using a small molecule smoothened agonist to activate the hedgehog pathway (SAG, D6–D10), we found similar levels of ScxGFP induction between SAG and TGF β 1 conditions at D10 (Fig. 1c–e). However, more ScxGFP^{hi} cells were observed with TGF β 1 (almost one log fold change higher) compared to SAG. Inhibition of TGF β signaling with SB-431542 in the SAG condition (D6–D10) resulted in near-complete suppression of ScxGFP at D10 (Fig. 1c). The combination of TGF β 1 + SAG was synergistic, with the highest numbers of total ScxGFP+ cells as well as ScxGFP^{hi} cells (Fig. 1c–e). Since ScxGFP+ induction by SAG was TGF β -dependent, it was possible that increased ScxGFP induction was due entirely to extra autocrine release of TGF β ligands. We therefore tested increasing concentrations of TGF β 1 (10–80 ng/mL) in the absence of SAG, relative to 10 ng/mL TGF β 1 + SAG (D6–D10). While no difference was observed in ScxGFP induction with increasing TGF β 1 concentrations (10–80 ng/mL) at D10, the addition of SAG in combination with the lowest TGF β 1 concentration (10 ng/mL) increased both overall ScxGFP+ and ScxGFP^{hi} cell populations (Fig. 1f). This data suggests that SAG modifies cell responsiveness to TGF β 1 to enhance ScxGFP induction.

Having established the utility of SAG in enhancing ScxGFP efficiency, we next confirmed differentiation by gene expression analysis of mESC pluripotency genes (*Nanog*, *Oct4*, *Sox2*)³⁶. As expected, pluripotency genes were highly expressed in naive mESCs at D0, but significantly downregulated with paraxial mesoderm and tendon differentiation at D6 and D10, respectively (Supplementary Fig. 1). Expression of *Scx* was dramatically induced only in the D10 tenogenic condition, relative to both

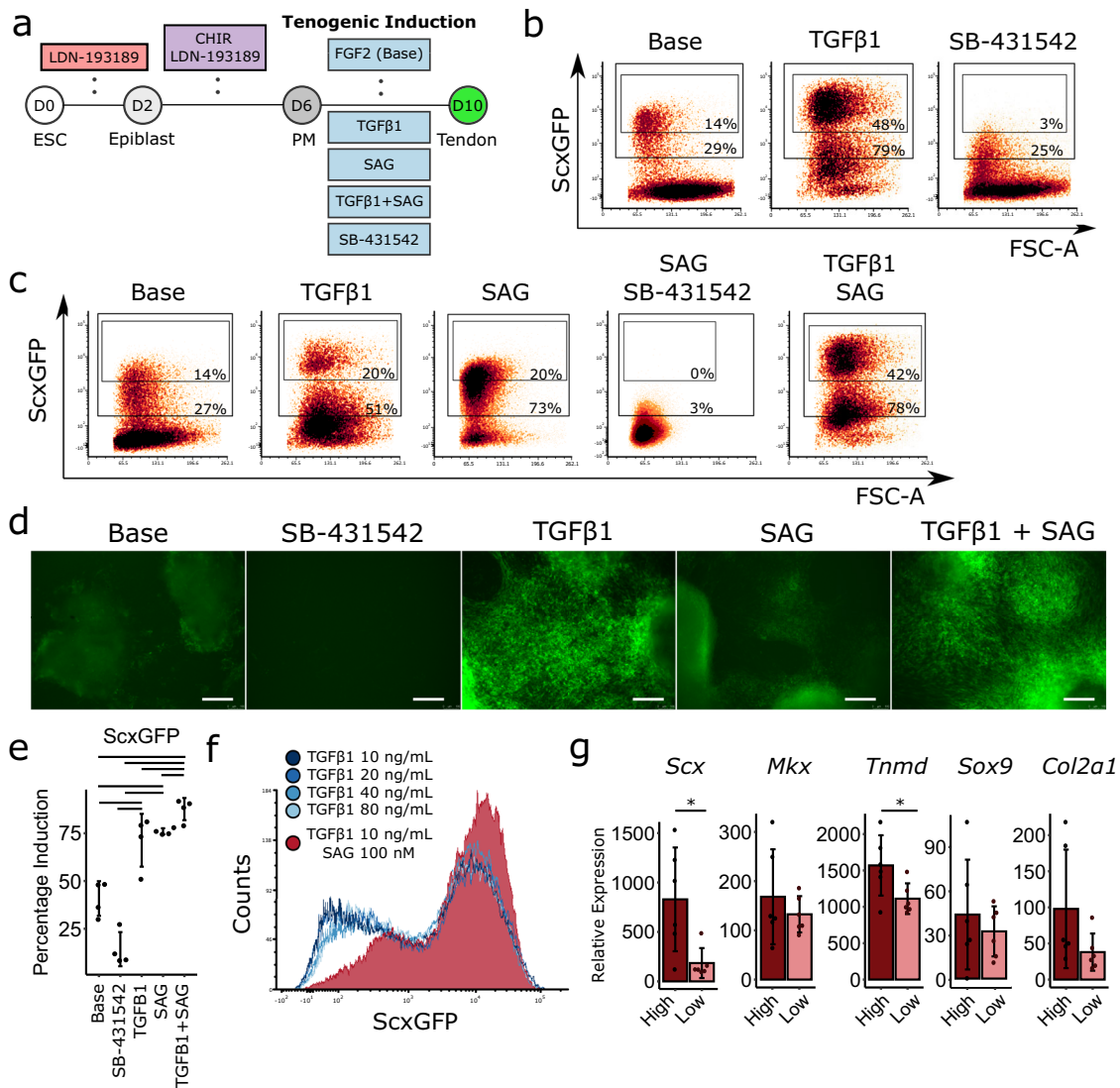


Fig. 1 Hedgehog activation modifies mESC responsiveness to TGFβ signaling and improves ScxGFP induction. **a** Schematic of tendon differentiation protocol. Top row indicates common media components. Bottom row indicates media component variables. **b, c** Flow cytometry plots for ScxGFP induction under tenogenic media conditions. All conditions include FGF2. **d** Representative images of ScxGFP expression for tenogenic media conditions. All conditions include FGF2. All experiments were repeated a minimum of three times. Scale bar: 100 μm. **e** Flow cytometry quantification of ScxGFP induction for tenogenic media conditions ($n = 4$ independent samples, one-way ANOVA with Tukey's post-hoc tests). All experiments were repeated a minimum of three times. Data represented as mean \pm SD. **f** Counts histogram showing ScxGFP induction with increasing concentrations of TGFβ1 compared to TGFβ1 + SAG condition. **g** Real-time qPCR quantification of tendon and cartilage genes in sorted ScxGFP^{hi} and ScxGFP^{lo} cells ($n = 6$ independent samples, unpaired two-sided Student's t -test). All experiments were repeated a minimum of three times. Data represented as mean \pm SD. Bars indicate $p < 0.05$ between groups. * $p < 0.05$. Source data are provided as a source data file.

naive D0 mESCs and D6 PM, confirming tendon lineage specification. To determine whether ScxGFP^{hi} and ScxGFP^{lo} populations represented distinct tendon cell subpopulations, we next performed cell sorting at D10 (TGFβ1 + SAG) and analyzed tendon and cartilage gene expression by qPCR (Fig. 1g). Interestingly, ScxGFP^{lo} cells expressed lower levels of *Scx* and *Tnmd* compared to ScxGFP^{hi} cells, but no significant difference in cartilage markers *Sox9* and *Col2a1* was observed ($p > 0.1$). This data suggests that ScxGFP^{lo} cells may represent a transitional population in the process of turning on ScxGFP and tendon markers.

In addition to hedgehog activation, we also considered BMP signaling, since BMP signaling was previously shown to inhibit *Scx* expression during embryonic chick and mouse

development.^{16,37,38} We therefore hypothesized that some level of BMP activation may be modulating tenogenic induction in our cultures and that inhibition of BMP signaling may therefore improve induction. As expected, supplementation with BMP4 resulted in nearly total inhibition of ScxGFP+ cell induction (4%) at D10, consistent with the developmental literature (Supplementary Fig. 2). However, the inclusion of the BMP inhibitor, LDN-193189 in combination with TGFβ1 resulted in reduced ScxGFP induction compared to TGFβ1 treatment alone (27% decrease). Interestingly, while BMP4 treatment resulted in the loss of ScxGFP+ cell induction, the combination of TGFβ1 + BMP4 resulted in only modest reduction of ScxGFP+ cells compared to TGFβ1 alone (13% decrease). Analysis of gene expression by qPCR at D10 showed no differences in *Scx*, *Mkx*, or

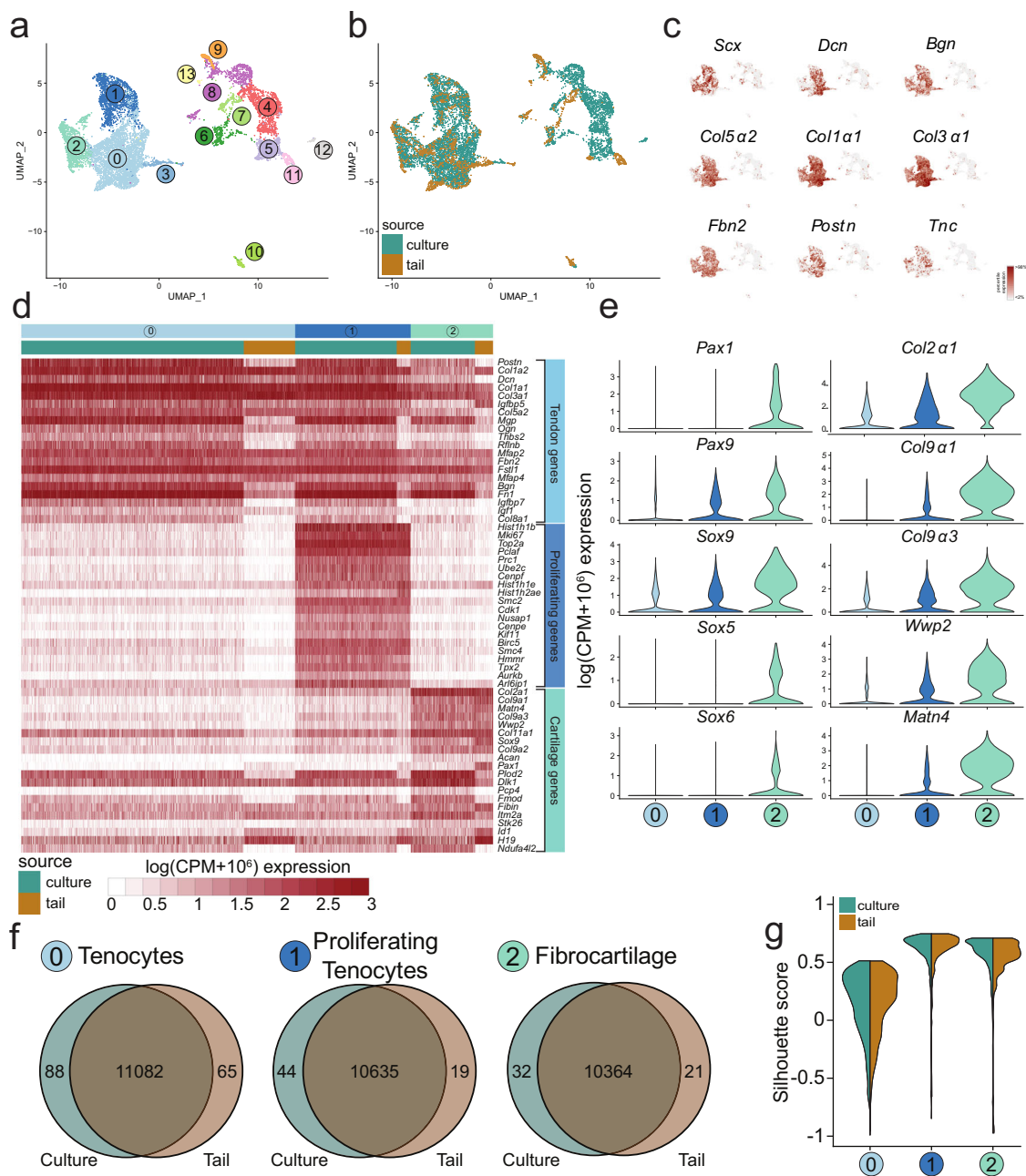


Fig. 2 Single-cell RNA sequencing of E14.5 tail cells and mESC-derived culture cells reveals tendon and fibrocartilage populations. **a, b** UMAP of all 14 clusters identified by modularity analysis from the E14.5 tail or the culture. **c** Feature plots of known tendon genes for clusters. **d** Top 20 differentially expressed genes for clusters 0, 1, and 2 visualized by heatmap. **e** Violin plots for transcription factors (left) and proteins (right) involved in cartilage matrix production and homeostasis. **f** Venn diagrams comparing differentially expressed genes between E14.5 tail cells and culture cells in Clusters 0, 1, and 2. **g** Silhouette scores between E14.5 tail cells (blue) and culture cells (brown) in Clusters 0, 1, and 2.

Tnmd expression between the TGF β 1, TGF β 1 + BMP4, or TGF β 1 + LDN groups (Supplementary Fig. 2). We further considered that BMP4 treatment may be inducing a chondrogenic phenotype, however, *Sox9* and *Col2a1* expression were also not significantly different between BMP4 alone compared to the TGF β 1-treated groups (Supplementary Fig. 2).

Single-cell RNA sequencing revealed high transcriptional similarity between D10 mESC-derived *Scx* cells and embryonic E14.5 *Scx* cells. To fully define the mESC-derived cell populations, we performed scRNA-seq. Using the TGF β 1 + SAG condition which generated the highest induction of *Scx*GFP+ cells,

we compared the entire culture population at D10 to embryonic mouse tail populations at E14.5. The E14.5 stage was chosen as tendons are fully differentiated⁴.

We successfully sequenced 2372 cells isolated from the E14.5 tail (4548 median genes/cell and 83,835 mean reads/cell) and 8382 cells isolated from the D10 mESC differentiation protocol (culture) (2658 median genes/cell and 18,879 mean reads/cell). Using the in vivo embryonic tail cells as our gold standard, we performed unsupervised clustering of E14.5 tail populations, which partitioned cells into 13 clusters (Fig. 2a). Clusters were then annotated based on the expression of select marker genes. Using the annotated tail clusters as the in vivo reference dataset, the cells from the D10 culture were assigned to the clusters using

Table 1 Percent contribution of E14.5 tail cells and D10 mESC-derived culture cells to each scRNA-seq cluster.

Cluster	E14.5 tail cells (% total)	D10 culture cells (% total)
0	685 (33.4%)	2964 (37.9%)
1	186 (9.1%)	1353 (17.3%)
2	49 (2.4%)	1088 (13.9%)
3	241 (11.7%)	855 (10.9%)
4	121 (5.9%)	556 (7.1%)
5	71 (3.5%)	435 (5.6%)
6	86 (4.2%)	257 (3.3%)
7	82 (4.0%)	155 (2.0%)
8	157 (7.6%)	53 (0.7%)
9	137 (6.7%)	2 (0.0%)
10	71 (3.5%)	58 (0.7%)
11	128 (6.2%)	1 (0.0%)
12	0 (0.0%)	43 (0.5%)
13	38 (1.8%)	1 (0.0%)

a nearest neighbors approach³⁹. The combined culture-tail dataset revealed many highly integrated clusters (Fig. 2a, b).

Next, we conducted differential gene expression (DGE) analysis to identify the genes defining each cluster (Supplementary Dataset 1). To identify tendon clusters, we assessed expression of select marker genes including *Scx*. Clusters 0, 1, and 2 all expressed *Scx* as well as other known tendon ECM markers such as *Dcn*, *Bgn*, *Col5a2*, *Col1a1*, *Col3a1*, *Fbn2*, *Postn*, and *Tnc* (Fig. 2c, d)^{22,40,41}. Interestingly, while cluster 3 (0.6% of culture cells) did not express *Scx*, it did express all of the other tendon markers consistent with clusters 0, 1, and 2. In total, *Scx*-expressing clusters comprised ~67% of the culture, within the range of *Scx*GFP induction with TGFβ1 + SAG (Fig. 1e).

While cluster 0 (38% of culture cells) was distinguished primarily by tendon marker expression, cluster 1 (17% of culture cells) was defined by expression of both tendon and proliferation genes (Fig. 2d). Gene set enrichment analysis showed statistically significant cell cycle terms (Supplementary Fig. 3), suggesting this cluster represents proliferating tendon cells. Surprisingly, when we attempted to regress out cell cycle effects, this cluster remained intact. While top-ranked genes (by log fold change) for cluster 1 were genes related to cell proliferation, we also detected significant low-level differences in sclerotomal genes such as *Pax9* and its downstream targets *Sox9* and *Col2a1* (Fig. 2e), suggesting that low-level activation of the sclerotomal program maintains a proliferative, progenitor state in cluster 1 tendon cells. While expression of tendon genes were also detected, cluster 2 (14% of culture) was defined primarily by cartilage genes, such as *Col2a1*, *Col9a1*, *Matn4*, *Col11a1*, *Sox9*, and *Acan* (Fig. 2d, e). The combined expression of tendon and cartilage genes suggested that cluster 2 represents fibrocartilage cells¹¹. DGE analysis of culture versus tail cells within clusters 0, 1, and 2 revealed low numbers of differentially expressed genes (Fig. 2f, Supplementary Dataset 2). Further, silhouette analysis of these clusters showed excellent integration between tail and culture (Fig. 2g), indicating successful recapitulation of tendon and fibrocartilage phenotypes comparable to in vivo cells.

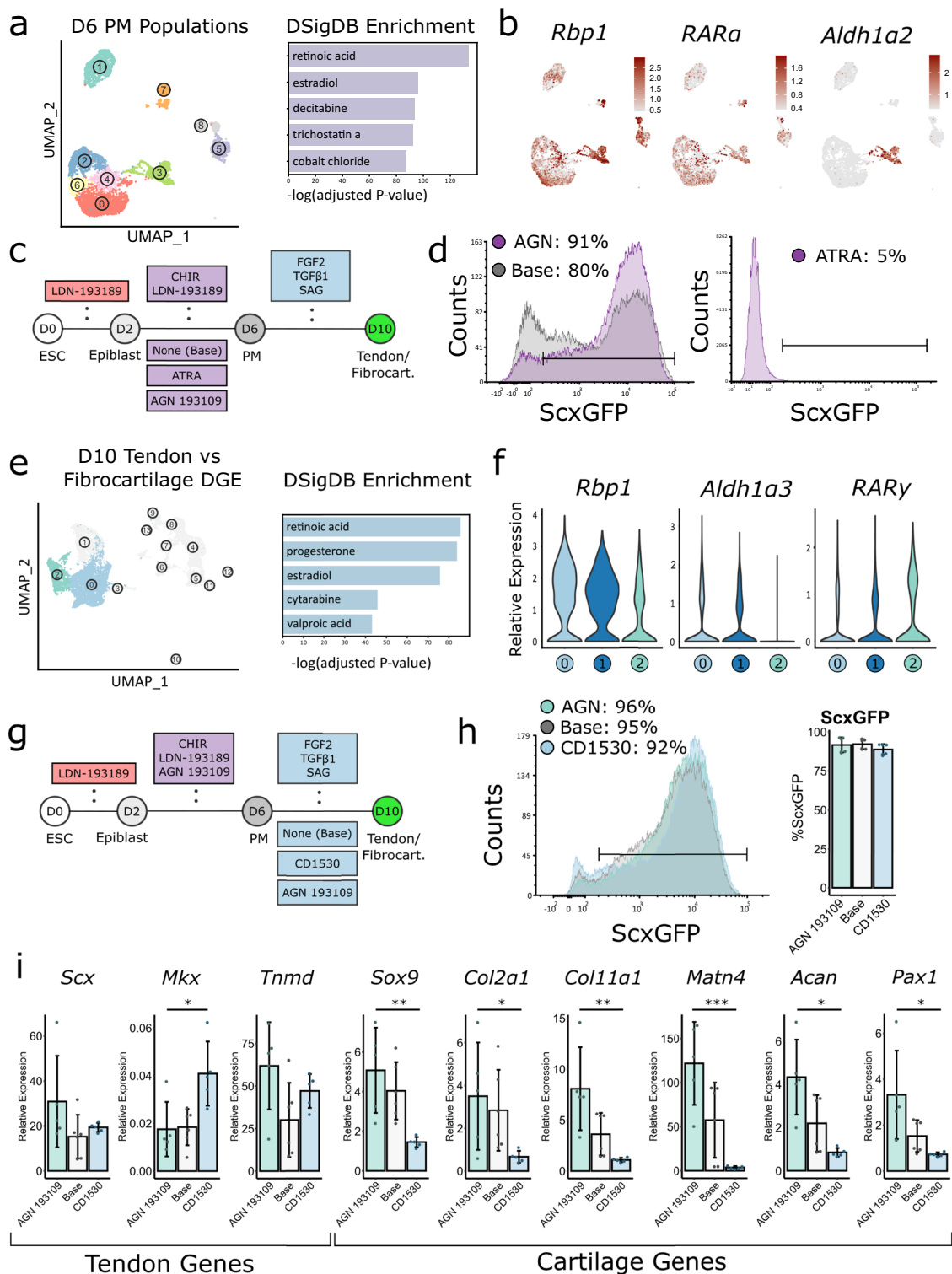
In addition to fibrous clusters, gene set enrichment and dot plot analysis also showed the presence of other expected cell types within the embryonic tail, including skin (clusters 4 and 7), blood (clusters 5 and 11), nerves (cluster 9), and muscle (cluster 8) (Supplementary Dataset 1, Supplementary Fig. 3, Supplementary Fig. 4). While larger clusters contained a mixture of tail and culture cells, the smaller clusters tended to be source-specific; for example, clusters 11 and 12 were composed largely of tail-only and culture-only cells, respectively (Table 1). Comparison of

percentage composition of culture and tail cells in each cluster revealed that while neural cells comprised ~6% of the tail population, only 0.03% of culture cells were assigned to this cluster. Similarly low percentages of skin (2%) and muscle (0.7%) were observed in the culture population (Table 1), indicating minimal differentiation toward these non-tenogenic lineages using our protocol. For larger non-fibrous clusters (such as cluster 4, representing 7% of the culture population), silhouette analyses also showed that the culture signatures were not well integrated with tail signatures despite assignment to the same cluster (Supplementary Fig. 3). In the case of cluster 4, gene enrichment showed that culture cells expressed genes such as *Nanog*, that were not expressed by tail cells in these clusters (Supplementary Fig. 3). This suggested that the cluster 4 culture cells may represent a small population of pluripotent cells that escaped differentiation.

Retinoic acid signaling regulates the switch between tendon and fibrocartilage cell fates.

The scRNA-seq results confirmed the presence of differentiated tendon and fibrocartilage cells, however, a significant population of cells adopted or maintained other fates. We hypothesized that this reflected inefficient mESC differentiation to PM based on the composition of cluster 4 culture cells. To identify the signals that may antagonize PM induction, we performed scRNA-seq on PM culture cells at D6. 6695 D6 PM cells were successfully sequenced with 26,589 mean reads/cell and 3170 median genes/cell. Unsupervised clustering of D6 PM cells revealed that the majority of the cells were in a closely grouped supercluster (cluster 0, 2, 3, 4, and 6), while other cells had diverged (clusters 1, 5, 7, and 8, Fig. 3a).

To better define the signaling pathways driving these transcriptional differences, we performed differential gene expression and gene set enrichment analyses and identified the retinoic acid pathway as the most activated pathway at the D6 PM stage. *Rbp4*, a retinol-binding protein, was the most differentially expressed gene in the largest cluster by log fold change (Fig. 3b, Supplementary Dataset 3). We examined other retinoic acid pathway elements and found *Rbp1*, *Rara*, and cluster-specific expression of *Aldh1a2*. To test the effect of retinoic acid signaling at the PM induction stage, we included a retinoic acid pathway activator (all-trans retinoic acid, ATRA) or inverse agonist (AGN 193109) during the PM differentiation phase (D2-D6). All PM conditions were then cultured with TGFβ1 + SAG from D6-D10 (Fig. 3c). Remarkably, retinoic acid pathway activation during PM induction by ATRA eliminated almost all *Scx*GFP induction at D10 (Fig. 3d). Treatment with RARγ-specific agonist CD1530 during D2-D6 also showed suppression of *Scx*GFP, albeit to a lesser extent compared to ATRA (14% vs 5%), which targets all RAR isoforms (Supplementary Fig. 5). In contrast, treatment with AGN 193109 resulted in an increase in *Scx*GFP+ cells, compared to the Base condition at D10 (Fig. 3d). The failure of *Scx*GFP induction at D10 despite TGFβ1 + SAG treatment suggested a failure of PM specification at D6 with ATRA activation of retinoic acid signaling. Since retinoic acid signaling is known to maintain pluripotency as well as induce neural differentiation⁴²⁻⁴⁴, we tested these possibilities with ATRA treatment (D2-D6) compared to AGN 193109 treatment. Gene expression analysis at D6 showed downregulation of pluripotency marker *Oct4* in both AGN 193109 and ATRA conditions while epiblast markers *Eomes*⁴⁵ and *Fgf5*⁴⁵ were suppressed in ATRA only (Supplementary Fig. 5). Analysis of fate markers showed upregulation of paraxial mesoderm marker *Tbx6*⁴⁶ with AGN 193109 treatment compared to upregulation of neural marker *Sox1*⁴⁷ with ATRA treatment (Supplementary Fig. 5). These results suggested that activation of retinoic acid signaling resulted in differentiation toward a neural fate and therefore loss of cells that



were competent at D6 to undergo subsequent tendon/fibrocartilage differentiation, despite exposure to TGFβ1 + SAG at D6-D10.

Having improved ScxGFP induction efficiency, we next addressed fate specification between tendon and fibrocartilage. To uncover potential molecular drivers regulating the tendon/fibrocartilage fate switch, we performed DGE testing between tendon (cluster 0) and fibrocartilage (cluster 2) cells at D10, using the scRNA-seq dataset collected from the D10 TGFβ1 + SAG condition. Gene set enrichment analysis again identified the retinoic acid pathway, which was the top-ranked term by p-value

(Fig. 3e). We also found that *Rbp1* and *Aldh1a3* were more highly expressed in the tendon clusters relative to the fibrocartilage cluster (Fig. 3f). Conversely, the retinoic acid receptor, *Rary*, was upregulated in the fibrocartilage cluster relative to tendon clusters (Fig. 3f). Since the retinoic acid pathways are known to modulate chondrogenesis^{48–50}, we tested supplementation with the RARγ-specific agonist CD1530 and the inverse agonist AGN 193109, in combination with TGFβ1 + SAG during the tendon induction phase (D6–D10) (Fig. 3g). At D10, ScxGFP induction efficiency was equivalent between all three conditions (~90%, Fig. 3h).

Fig. 3 Retinoic acid pathway manipulation improves ScxGFP induction efficiency and direct tendon and fibrocartilage fate switch. **a** UMAP of clusters from D6 PM. DSigDB gene set enrichment analysis on most differentially expressed genes from each cluster in D6 PM. **b** Feature plots of *Rbp1*, *RAR α* , *Aldh1a2*. **c** Schematic for differentiation protocol with ATRA, AGN 193109, or no supplementation (Base) during PM induction. Top row indicates common media components. Bottom row indicates media component variables. **d** Counts histograms of ScxGFP expression comparing AGN 193109, ATRA, and Base conditions at D10. **e** UMAP of E14.5 tail cells and D10 culture cells with tendon and fibrocartilage clusters highlighted. DBSigDB gene set enrichment analysis on the most differentially expressed genes in the tendon cluster 0 relative to fibrocartilage cluster 2. **f** Violin plots of *Rbp1*, *Aldh1a3*, *RAR γ* for Clusters 0, 1, and 2. **g** Schematic for tendon and fibrocartilage differentiation with retinoic acid pathway manipulation using CD1530 and AGN 193109. Top row indicates common media components. Bottom row indicates media component variables **h** Counts histograms of ScxGFP expression and quantification of ScxGFP+ cells at D10 ($n = 4$ independent samples, one-way ANOVA). All experiments were repeated a minimum of three times. Data represented as mean \pm SD. **i** Real-time qPCR analysis of tendon and cartilage genes at D10 identified from the single-cell transcriptional signatures. D10 expression is normalized to D6 PM gene expression ($n = 6$ independent samples, unpaired two-sided Student's *t*-test between AGN 193109 and CD1530 conditions). All experiments were repeated a minimum of three times. Data represented as mean \pm SD. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source data are provided as a source data file.

Among tendon genes, only *Mkx* expression increased with pathway activation (CD1530), while *Scx* and *Tnmd* were not affected (Fig. 3i). However, treatment with CD1530 drastically reduced cartilage gene expression (Fig. 3i). Using the Mouse Proteome Profiler Array, we further screened and identified enhanced phosphorylation of receptor tyrosine kinases associated with chondrogenesis in the AGN 193109-treated group compared to CD1530, such as PDGFR α , PDGFR β , and ErbB2 (Supplementary Fig. 6)^{51–54}. These data indicate that retinoic acid signaling in the presence of TGF β 1 and SAG regulates the chondrogenic switch between mESC-derived tendon and fibrocartilage fates without adversely affecting the fibrogenic phenotype. These results further suggest that the biological response to retinoic acid signaling is highly context-specific and depends on the competency of the cells experiencing signaling as well as the presence of other signaling pathways (Supplementary Fig. 7).

Molecular profiling of tendon and fibrocartilage differentiation trajectories by RNA sequencing. Using these highly efficient differentiation protocols to specify tendon and fibrocartilage cells, we next profiled the molecular dynamics of fate induction for both lineages. RNA-seq was performed on D6 PM as the starting common control and subsequent timepoints (D6.25, 6.75, 7.5, 8.5, and 10) sampled for CD1530 and AGN 193109 conditions (105 M reads/sample, Fig. 4a). To determine the kinetics for tendon and fibrocartilage fate acquisition in the CD1530 and AGN 193109 conditions, respectively, we first examined the tendon, proliferating tendon, and fibrocartilage gene signatures (identified from scRNA-seq clusters 0, 1, and 2, respectively) (Supplementary Fig. 8). The proliferating tendon signature was downregulated in both trajectories after D8.5, consistent with terminal differentiation to post-mitotic lineages (Supplementary Fig. 8). This indicates that modulation of retinoic acid signaling pushed differentiation toward cluster 0 “tendon” or cluster 2 “fibrocartilage” phenotypes, without inducing a cluster 1 “proliferating tenocyte” population. Consistent with the expected phenotypes, tendon genes (*Scx*, *Mkx*, *Col1a1*, *Col3a1*) were highly upregulated in both conditions at D10 while cartilage genes (*Col2a1*, *Acan*, *Matn4*) were much higher in the AGN 193109 condition and minimally expressed in the CD1530 condition at D10 (Supplementary Fig. 8).

Principal component analysis (PCA) revealed a time dimension captured by PC0, while tendon versus fibrocartilage differentiation was captured by PC1 (Fig. 4b). Interestingly, transcriptional signatures associated with cell fate (PC1) were minimally affected until D8.5, after which both the tendon and fibrocartilage trajectories rapidly upregulated their respective signatures. By D10, tendon and fibrocartilage signatures were strongly activated. Divergence of fate began after D7.5 and increased dramatically between D8.5 and D10 (Fig. 4b). To

identify the genes contributing to this divergence, we identified the genes with the largest contributions to PC0 and PC1 (Supplementary Dataset 4). Among the largest contributors to PC0 were *Postn*, *Bgn*, *Col1a2*, *Col3a1*, *Col5a2*, and *Tnc*, consistent with the fibrous signature shared by both lineages. Genes that contributed to upward movement along PC1 included *Lox* (tendon marker associated with late embryonic differentiation⁵⁵), and *Igf1bp5* (highly expressed in scRNA-seq tendon clusters 0 and 1 and previously identified in embryonic tendon attachment cells^{22,56}) (Fig. 2, Supplementary Fig. 8). In contrast, genes that contributed to downward movement along PC1 included *Tgfb1*, *Col2a1*, and *Vcan* which are established cartilage markers (Supplementary Fig. 8). Interestingly, many genes contributing to PC0 also contributed to PC1 in the upward or downward direction. This suggests that many genes are highly upregulated in both lineages compared to PM, but differentially expressed between tendon and fibrocartilage.

We then performed DGE testing to identify statistically significant genes contributing to each fate trajectory (Figs. 4c, 5). Fate-independent (PC0) and -dependent (PC1) modules of highly correlated genes were determined by performing clustering of differentially expressed genes in separate correlation spaces. Since PC0 represented 79% of the variance of the dataset and tendon and fibrocartilage trajectories did not diverge until D8.5, the fate-independent modules revealed early induction gene patterns common to both cell types. Analysis of fate-independent modules identified multiple genes with intriguing kinetics, including one large module (Module 1) that showed rapid upregulation in gene expression between D6 and D7.5 before stabilizing (Fig. 4c). Transcription factors within this module included known markers of the syndetome (*Scx*) and sclerotome (*Pax1* and *Pax9*), as well as other transcription factors previously unassociated with tendon or cartilage differentiation (Fig. 4c, Supplementary Dataset 4). Interestingly, a separate module with a similar pattern (Module 4) contained the transcription factor *Sox9*. Genes within this module were upregulated more quickly than Module 1 genes, before subsequently downregulating. In this dataset, *Sox9* and *Pax1* were higher in AGN 193109 vs CD1530, supporting fate divergence (Fig. 4c). *Tbx6* and *Cdx1* were both expressed in PM and contributed to Module 3. Rapid downregulation of these genes is consistent with efficient differentiation. Lastly, Module 2 was only transiently expressed, as shown by representative markers *Myog* and *Tal1* (Fig. 4c, Supplementary Dataset 4). Although *Myog* is a key factor of muscle differentiation^{57,58}, its expression was also previously identified in embryonic tendon cells of the limb²³.

Analysis of fate-dependent modules revealed tendon modules whose expression peaked at D8.5 (Module 5) and D7.5 (Module 6) in the CD1530 condition, with little change or decrease in the AGN 193109 condition (Fig. 5, Supplementary Dataset 4).

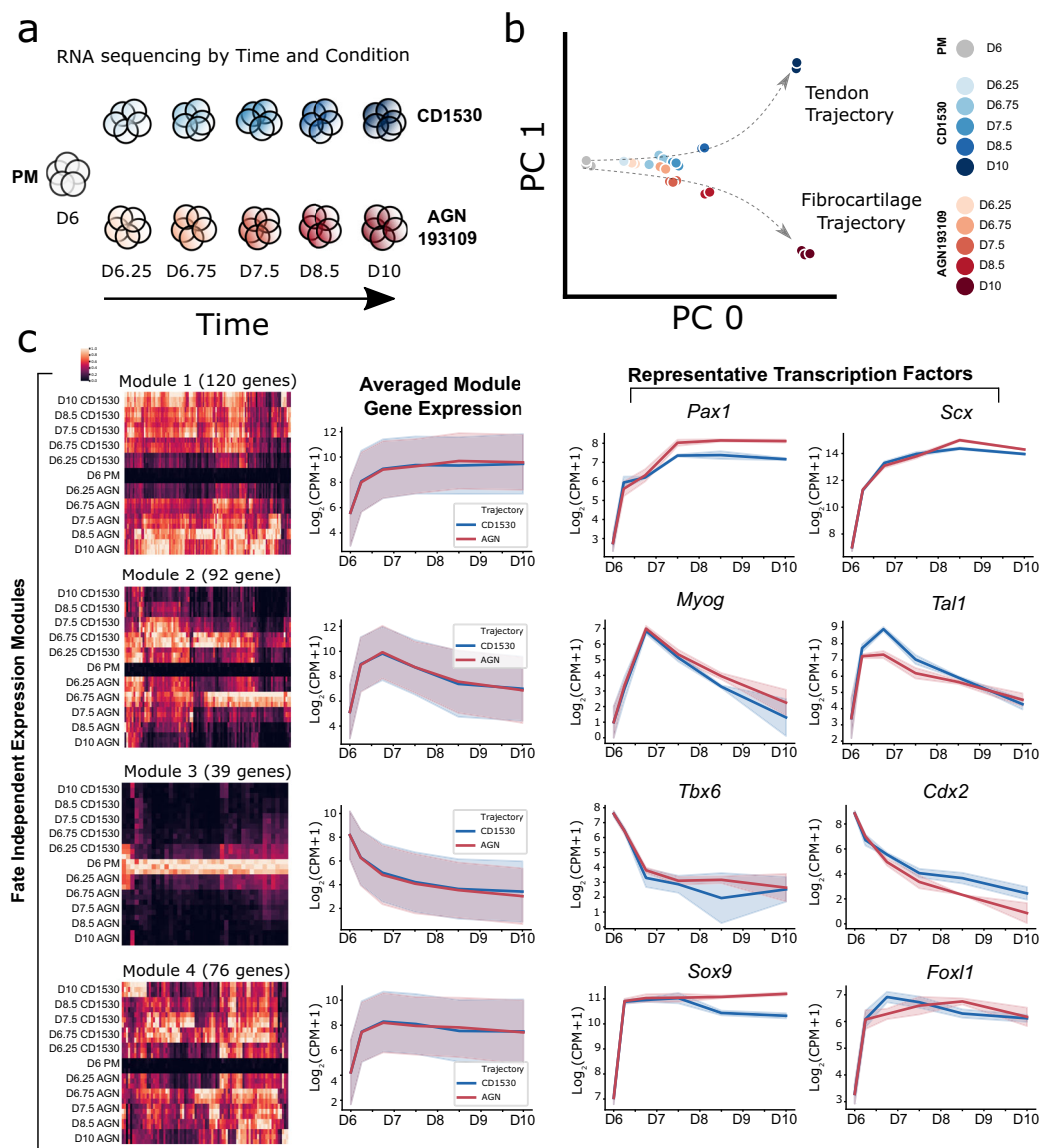


Fig. 4 Transcriptional profiling and trajectory analysis by RNA sequencing reveal transcription factor modules associated with differentiation. **a**

Schematic for RNA-seq of CD1530 (blue) and AGN 193109 (red) temporal trajectories starting from the common starting point of D6 PM. Samples were collected at indicated timepoints and conditions for RNA-seq. **b** PCA of tendon and fibrocartilage trajectories. **c** Fate-independent gene modules represented as column-normalized heatmaps, averaged module gene expression line plots, and representative transcription factors. ($n = 3$ independent samples). Standard deviation is represented by shaded regions.

Expression patterns for the fibrocartilage trajectory modules (Modules 7 and 8) showed transiently expressed and steadily upregulated genes, respectively. Notably, the homeobox gene *Dlx5*, a regulator of chondrocyte differentiation and maturation^{59–61}, was found in Module 8 alongside well-established cartilage markers *Acan*, *Matn4*, and *Sox6* (Fig. 5, Supplementary Fig. 8, Supplementary Dataset 3). Although these genes also increased in the CD1530 condition compared to D6 PM, the levels at D10 were markedly lower compared to AGN 193109.

From trajectory module analyses, we identified several genes that have not been previously implicated in tendon or fibrocartilage development. To determine whether there may be in vivo relevance, we collected tails from E11.5–E14.5 embryos, spanning the stages of tendon and annulus fibrosus progenitor induction (E11.5 and E12.5) through differentiation (E14.5, Supplementary Fig. 9). Real-time qPCR analysis of the top genes contributing to PC0 showed dramatic upregulation at E14.5

compared to E11.5 or E12.5 stages for all genes (Supplementary Fig. 9).

Extended culture of tendon and fibrocartilage cells leads to loss of fate and senescence.

To determine whether tendon and fibrocartilage phenotypes could be maintained under extended culture, we passaged cells at confluence (D10 and D14) and continued culturing in the presence of CD1530 or AGN 193109 (Supplementary Fig. 10). By D18, we found that while cells remained viable, they were no longer proliferative and ScxGFP expression was reduced in both media conditions. Analysis of phenotypic markers at D18 by qPCR showed a loss of tenogenic markers *Scx* and *Mkx* for the CD1530 condition. *Col2a1* was minimally expressed regardless of passage demonstrating no acquisition of a cartilage phenotype. In contrast, *Scx* expression modestly increased with extended 2D culture in the presence of AGN 193109, while *Tnmd* and *Col2a1* expression levels were unchanged (Supplementary Fig. 10). This suggests that while the

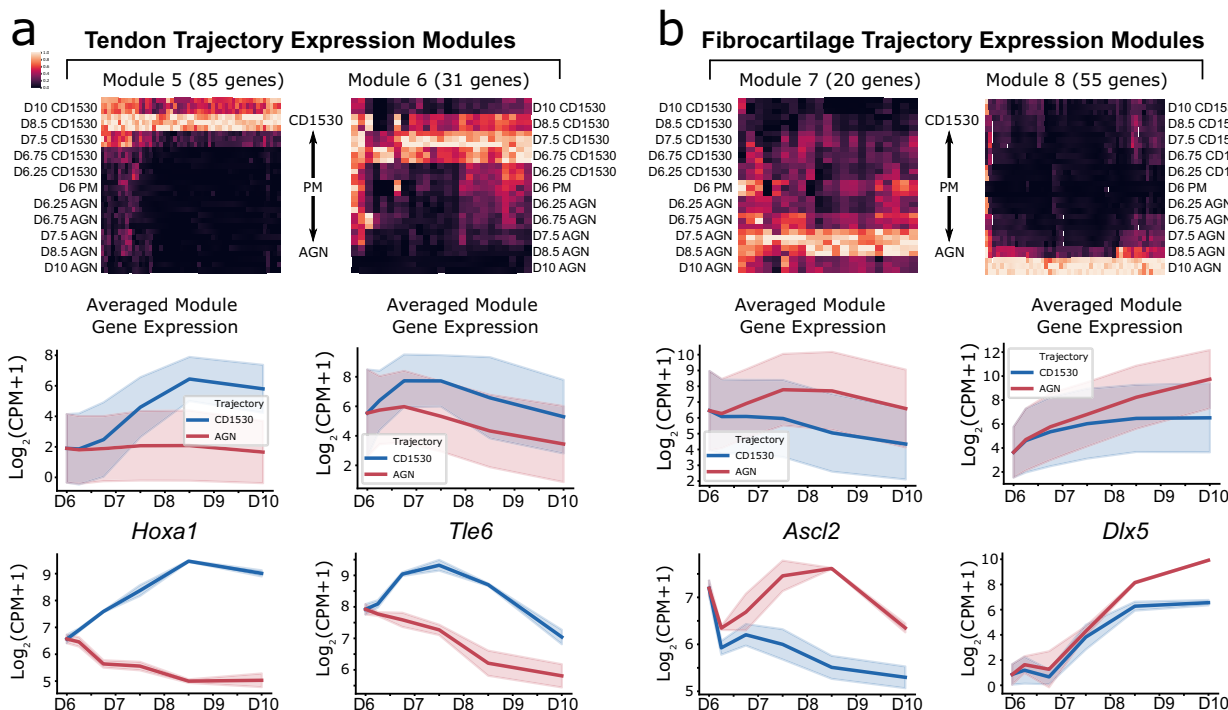


Fig. 5 Transcriptional profiling and trajectory analysis by RNA sequencing reveals transcription factor modules associated with fate-specific tendon and fibrocartilage induction. Fate-dependent gene modules for (a) tendon and (b) fibrocartilage trajectories represented as column-normalized heatmaps, averaged module gene expression line plots, and representative transcription factors. ($n = 3$ independent samples). Standard deviation is represented by shaded regions.

fibrocartilage AGN 193109-containing media was sufficient to maintain the phenotype of the cells, maintenance of the tenogenic phenotype may require additional cues.

Differentiation of tendon and fibrocartilage cells under 3D uniaxial static tension. Since tendon cells normally experience uniaxial mechanical tension during development, we next tested whether we could better sustain the tendon phenotype using a 3D tension culture system. mESCs were therefore differentiated toward D6 PM and then directly embedded into collagen gels held under static tension (Fig. 6a, b)⁶². We observed gradual cell-mediated contraction of the gels over a course of 11 days (D6–D17) with no difference in contraction between CD1530 and AGN 193109 conditions at any timepoint (Fig. 6b). To compare tendon and fibrocartilage induction in 3D culture relative to standard 2D culture, we harvested cells at D10 for qPCR analyses. In the CD1530 condition, *Scx* and *Acan* expression were similarly induced between 2D and 3D, while *Coll1a1* increased in 3D gels (Supplementary Fig. 11). Gene expression for osteogenic, myogenic, and adipogenic markers did not show any differences between 2D or 3D culture, indicating no aberrant differentiation along these other lineages. With AGN 193109 treatment, we observed similar levels of tendon and cartilage markers between 2D and 3D conditions, but a surprising increase in the osteogenic marker *Osx* with 3D culture (Supplementary Fig. 11). Analysis of tendon-specific transcription factors identified from RNA-seq-derived modules showed upregulation of both *Hoxa1* and *Wt1* with 3D culture in CD1530. In contrast, no difference was observed for the fibrocartilage-specific transcription factors, *Alx4* and *Dlx5* with AGN 193109-treatment (Supplementary Fig. 11).

Since collagen is the main ECM component in both tendon and fibrocartilage tissues, we next assessed organized collagen deposition using second-harmonic generation (SHG) on whole-mount gels at D17. Acellular collagen gels were also imaged to

distinguish the collagen gel biomaterial from cell-secreted collagen. Quantification of SHG images showed significant deposition of aligned collagen in the cell-seeded gels relative to acellular gels, with no detectable difference between CD1530 and AGN 193109 conditions (Fig. 6c). Whole-mount and multiphoton fluorescent imaging revealed the presence of ScxGFP+ longitudinally aligned cells in both conditions, but the appearance of rounded nodules only in the AGN 193109 condition (Fig. 6d, e). Staining of transverse sections showed that these rounded nodules stained intensely for type II collagen, indicating cartilage-like matrix (Fig. 7f). In contrast, cartilage nodules were not detected in any CD1530 constructs. Collectively, these data showed that 3D static tension supports tenogenic induction and differentiation, with the alignment of cells and deposition of collagen under CD1530 conditions. While ScxGFP cells also aligned in the AGN 193109 condition, the fibrocartilage phenotype was not appreciably altered under static tension based on gene expression and the presence of cartilaginous nodules, suggesting that biochemical cues are predominant for induction and maintenance of fibrocartilage.

Dynamic tensile loading of 3D engineered tissues activate mechanotransduction pathways. During embryogenesis, contractile muscle forces play a critical role in tendon and fibrocartilage development^{63–65}. To determine whether our engineered 3D tissues respond to mechanical loading, we adapted the STREX Cell Stretching System by designing 3D-printed molds to create custom PDMS dishes (Fig. 7a). D6 PM cell-seeded constructs were then generated and cultured within these dishes, which could be mounted onto the bioreactor system for dynamic tensile loading. Gels cultured within the custom dishes showed normal contraction and successful induction of ScxGFP (Fig. 7a–c). We validated the bioreactor for 3D gels by testing a range of bioreactor inputs (2–15% stretch, 0.5 Hz) comparing elongation

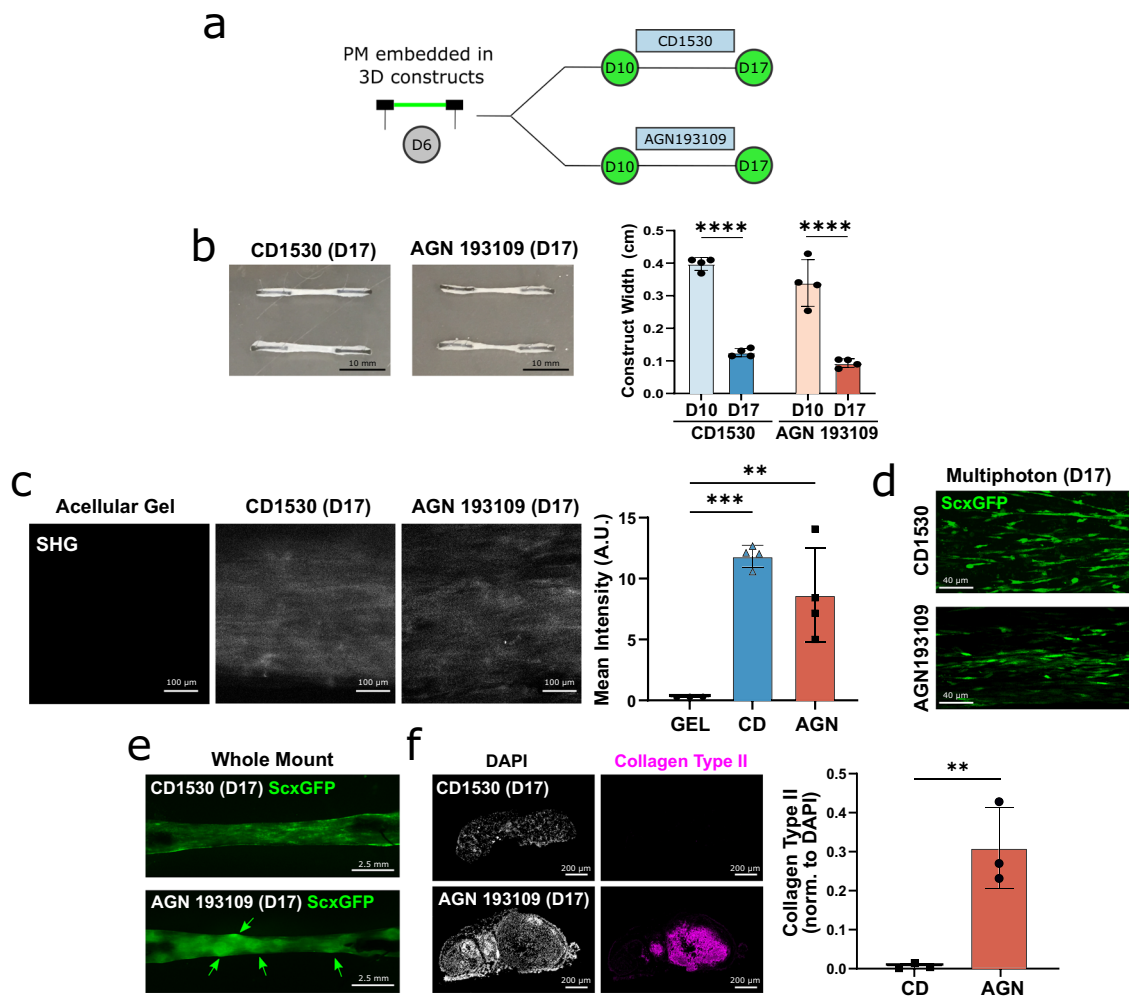


Fig. 6 3D static tension aligns cells and collagen deposition in engineered tendon and fibrocartilage tissues. **a** Schematic of experimental design. **b** Quantification of gel contraction over 17 days of culture in 3D ($n = 4$ independent gels, one-way ANOVA with Tukey's posthoc tests). Data represented as mean \pm SD. **c** SHG imaging and quantification for aligned collagen deposition ($n = 4$ independent gels, one-way ANOVA with Tukey's posthoc tests). Data represented as mean \pm SD. **d** Multiphoton imaging of aligned ScxGFP⁺ cells at D17 under 3D conditions. **e** Whole mount ScxGFP fluorescence imaging of 3D gels. **f** Immunostaining and quantification for collagen type II in transverse cryosections ($n = 3$ independent gels, unpaired two-sided Student's *t*-tests). Data represented as mean \pm SD. ** $p < 0.01$, **** $p < 0.0001$. Source data are provided as a source data file.

experienced by the gels against applied inputs. Image analysis showed a linear correlation of gel elongation with increasing applied stretch ($R^2 = 0.96$, Fig. 7d). Four hours of dynamic loading at 6% gel elongation (which is within the physiologic range for tendons *in vivo*⁶⁶) resulted in upregulation of known mechanotransduction markers for both CD1530 (*Wwtr1*, *Mrtfa*) and AGN 193109 (*Yap1*, *Wwtr1*, *Trpv4*, *Mrtfa*) conditions, while 3% loading had no effect. These results establish a platform for studying tendon developmental mechanobiology.

Discussion

In this study, we used an iterative approach informed by developmental cues and single-cell RNA sequencing to establish highly efficient methods for deriving tendon- and fibrocartilage-specific cells from mESCs through manipulation of the TGF β , hedgehog, and retinoic acid signaling pathways. Using these models of tendon and fibrocartilage differentiation, we further define the trajectory of transcriptional signatures during differentiation and identify potential regulators distinguishing these phenotypes. Finally, we applied these media protocols to generate 3D engineered tendon and fibrocartilage tissues and established a system

for modeling tendon mechanobiology under static and dynamic culture conditions (Fig. 8).

To date, healing of dense, connective tissues (including tendon, ligament, meniscus, and the annulus fibrosus) remains poor in adults and treatment is largely limited to surgical repair^{67–69}. Recently, a few studies showed that regenerative healing is possible during fetal and neonatal stages^{70–74}, however primary fetal and neonatal cells are not clinically viable options for transplantation. Although differentiated cells can be derived from pluripotent sources such as ESCs, from a translational perspective, there are ethical considerations for the use of these cells and practical challenges such as risk of teratoma formation^{75,76}. These risks may be overcome by improving differentiation efficiencies or including selection markers⁷⁶. In this study, we demonstrate that a directed differentiation approach efficiently generates tendon and fibrocartilage cells with transcriptional signatures comparable to *in vivo* embryonic cells, and can therefore be used to model developmental processes. While we refer to our non-chondrogenic Scx-expressing cell populations as tendon cells, it is likely that the 'tendon' phenotype here can be broadly applied to related tissues such as ligaments, which connect bone to bone. Although differences between tendons and ligaments emerge

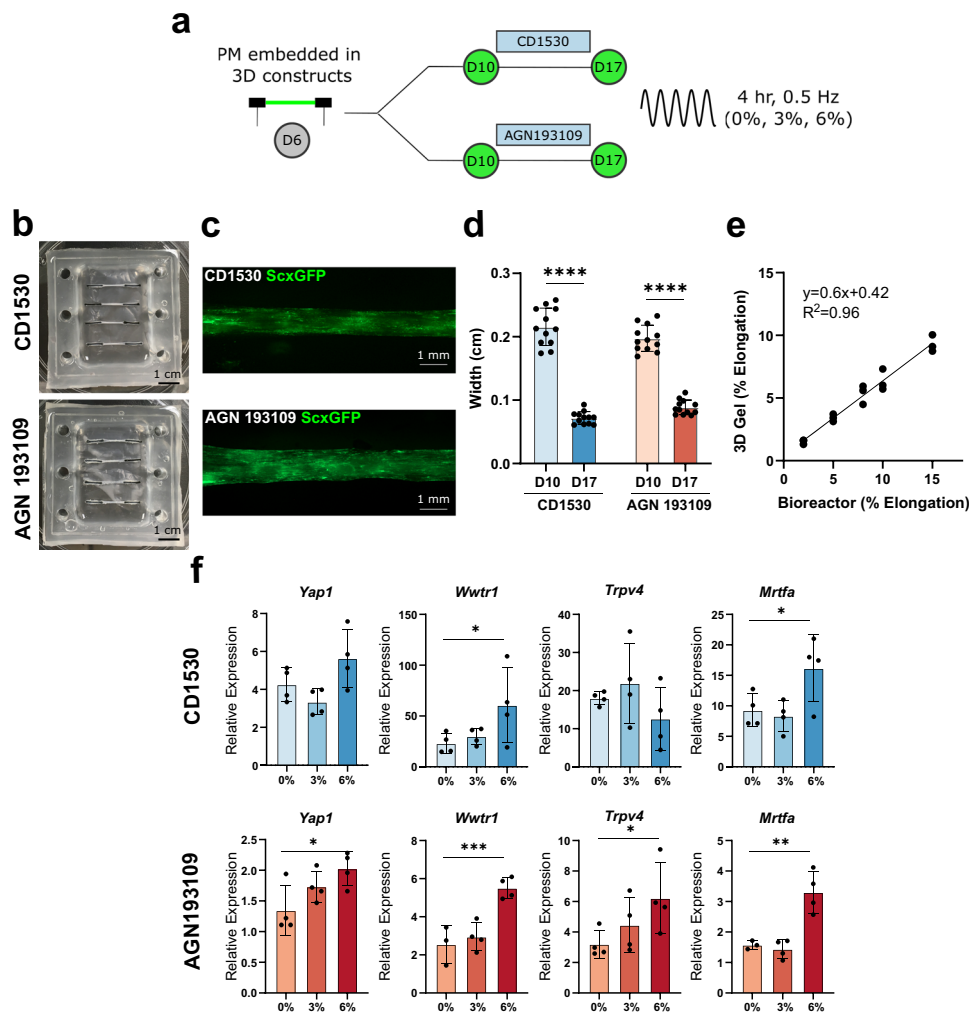


Fig. 7 Dynamic tensile loading of 3D engineered tendon and fibrocartilage tissues activates mechanotransduction markers. **a** Schematic of experimental design. Gels were dynamically loaded at D17. **b** Successful generation of contracted 3D cell-seeded gels in custom PDMS dishes at D17. **c** Whole-mount ScxGFP fluorescence imaging of 3D gels at D17. **d** Quantification of gel contraction ($n = 12$ independent gels, one-way ANOVA with Tukey's posthoc tests). Data represented as mean \pm SD. **e** Validation of % elongation experienced by 3D gels with bioreactor loading ($n = 3$ independent gels, Pearson's correlation). **f** Real-time qPCR analysis of known mechanotransduction markers with 4 h of dynamic loading ($n = 4$ independent gels, one-way ANOVA with Fisher's LSD tests). Data represented as mean \pm SD. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Source data are provided as a source data file.

postnatally with adaptation to local loading environments (and indeed, differences between anatomic tendons also emerge postnatally⁷⁷), at these relatively early embryonic stages, there is no evidence that tendons and ligaments are transcriptionally distinct²³. Ongoing research will focus on applying these protocols to human iPSCs, which may have more direct translational potential for tendon/ligament and fibrocartilage (meniscus/annulus fibrosus) repair.

Surprisingly, a pure cartilage population was not observed in our cultures despite the known role of TGF β signaling in chondrogenesis^{14,15}, as cells expressing a chondrogenic signature also strongly expressed fibrous markers. These opposing responses to TGF β may be due to the distinctive competence of paraxial mesoderm cells (vs. other progenitor cell types such as mesenchymal stem cells), since previous studies using mouse somitic cells also showed that TGF β exerts a strong fibrogenic effect at the expense of chondrogenesis⁷⁸. During limb development, TGF β also induces bipotent *Scx*+/*Sox9*+ progenitors that are subsequently allocated to either tendon (*Scx*+ only) or cartilage (*Sox9*+ only)¹⁶. This fate decision is mediated in part by

BMP signaling, which turns off *Scx* in a subset of progenitors while maintaining *Sox9*. Although BMP activation was shown to inhibit *Scx* expression in early limb development and BMP ligands are used to promote cartilage differentiation in vitro^{17,28,37}, manipulation of BMP signaling did not affect *Scx*GFP induction in our system when TGF β signaling was also active. Our results suggest that while BMP4 signaling in the absence of TGF β does indeed inhibit *Scx* expression, TGF β may also induce the production of other members of the BMP family that further promote tenogenic induction (since supplementation with LDN-193189 resulted in a marked reduction of *Scx*GFP induction). Potential members may be BMPs 12, 13, and 14 (formerly known as GDFs 7, 6, and 5, respectively), which were shown to be involved in embryonic tendon development^{79–82}. Recent studies also demonstrate a critical role for BMP signaling in tendon growth postnatally, indicating that BMP signaling is required for normal tendon development⁸³. In addition, the sustained presence of TGF β prevented the inhibitory effect of BMP4 signaling since only a modest decrease in *Scx*GFP cells was observed with no effect on tendon gene expression. This may be

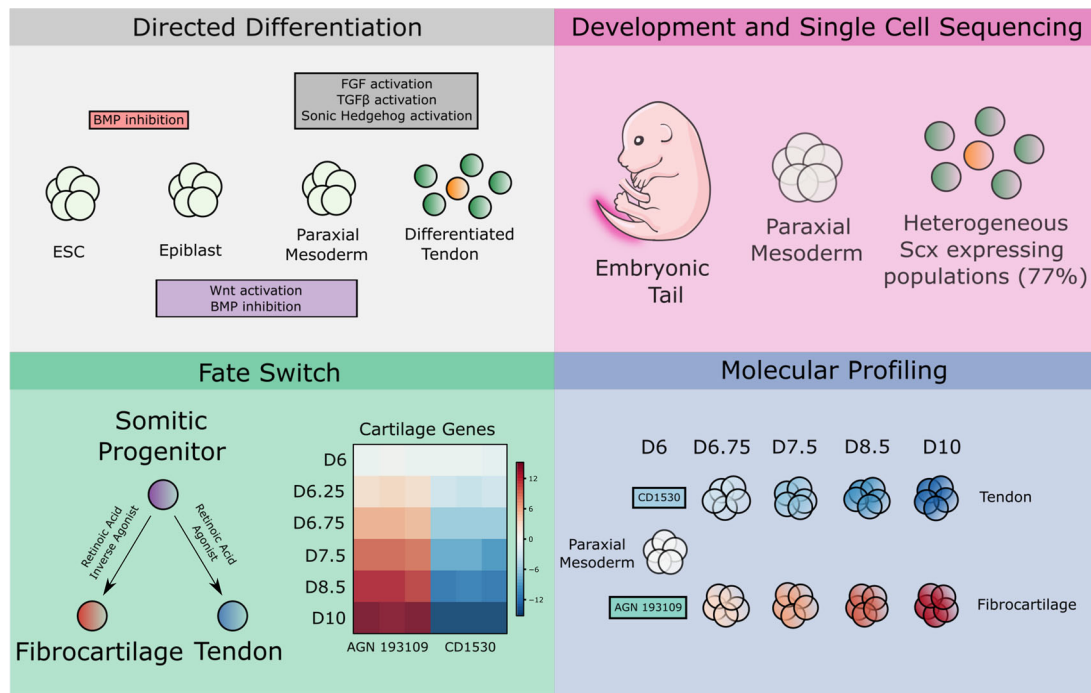


Fig. 8 Conceptual schematic highlighting key findings. Directed differentiation inspired by developmental events was used to induce ScxGFP tendon cells. Single-cell RNA-seq identified distinctive tendon and fibrocartilage cells with high transcriptional similarity to in vivo embryonic counterparts. Retinoic acid signaling was identified as fate-switch regulator between tendon and fibrocartilage cells. Molecular profiling identified regulators of tendon and fibrocartilage induction.

due to differences in timing/sequence of molecular factors or indicate that TGF β exerts a more dominant or antagonistic role.

In these studies, we also identified retinoic acid signaling as a key driver of tendon vs fibrocartilage fates in the presence of TGF β . Although retinoic acid signaling has been extensively studied in the context of skeletal and cartilage development, there are few studies in the context of tendon and fibrocartilage tissues. For tendon, retinoic acid molecules are implicated in tendon-muscle patterning in the eye and cranium^{84,85}. Activation of retinoic acid signaling (in the absence of TGF β) was also suggested to maintain stemness of tendon stem/progenitor cells while restricting differentiation along mesenchymal lineages, including tendon⁸⁶. Of the three retinoic acid receptors (RAR α , RAR β , and RAR γ), RAR γ is essential in the cartilage growth plate for skeletal development, but functions in combination with either RAR α or RAR β since single null RAR γ mutants are normal⁸⁷. The RARs can act as transcriptional activators or repressors, depending on ligand binding. In the absence of ligand, RAR γ functions as a repressor which results in chondrogenic activity. Repressor activity can be enhanced by using synthetic retinoids that act as inverse agonists, such as AGN 193109. Ligand binding with a standard agonist on the other hand (such as ATRA), results in repression of chondrogenesis⁴⁸. Here, we applied synthetic retinoic acid molecules (AGN 193109 and CD1530) in the presence of TGF β to differentiated paraxial mesoderm and showed that the anti-chondrogenic effects of RAR γ agonism were independent of fibrous differentiation. We therefore propose that TGF β drives the activation of fibrous markers while retinoic acid signaling toggles chondrogenesis. Unlike with BMP signaling, there does not seem to be an antagonistic relationship between TGF β and retinoic acid pathways, although it has been suggested that the anti-chondrogenic activities of retinoic acid activation is through suppression of Smad2/3 phosphorylation⁸⁸. Interestingly, recent research indicates that the tenogenic/fibrous functions of TGF β signaling is likely through non-canonical (ie. non-Smad)

pathways, suggesting a mechanism by which TGF β signaling can distinguish fibrous vs chondrogenic differentiation^{83,89}. Interestingly, the activity of retinoic acid signaling appears highly context-dependent. During paraxial mesoderm specification, we found that early activation of retinoic acid signaling resulted in differentiation toward a neural fate. When paraxial mesoderm was successfully induced however, retinoic acid signaling functioned in concert with TGF β to regulate tendon vs fibrocartilage. In our cultures, activation or inhibition of retinoic acid signaling at these latter stages did not induce a stem-like phenotype as was previously reported in adult-derived tendon stem/progenitor cells⁸⁶, since both CD1530 and AGN 193109 treatment groups showed robust upregulation of tenogenic differentiation markers. Further, sustained culture in 2D in the presence of CD1530 did not maintain proliferation potential. Other studies showed that within the context of BMP signaling, retinoic acid signaling does not inhibit chondrogenesis, but induces chondrocyte hypertrophy instead⁹⁰. In limb bud cultures, retinoic acid signaling resulted in the suppression of chondrogenesis without impacting myogenesis⁵⁰. Thus, the effect of retinoic acid activation or antagonism highly depends on other signaling pathways to provide contextualizing signals, as well as on the competence of the cell types experiencing the signaling. Future studies will further define interactions between TGF β , BMP, and retinoic acid signaling pathways in tendon/fibrocartilage differentiation in vivo and in vitro and test additional molecular regulators of these fates.

Successful induction efficiencies of >90% enabled the use of RNA-seq to profile transcriptional trajectories defining tendon and fibrocartilage differentiation. We were particularly interested in early kinetics since the transcription factors that regulate tendon and fibrocartilage induction have yet to be identified. Transcriptional profiling of tendon and fibrocartilage fate trajectories revealed several transcription factors associated with induction, both fate-independent and -dependent. These included

several transcription factors that correlated with *Scx* expression pattern. Analysis of *in vivo* embryonic mouse tails showed upregulation of several of these molecules in concert with tendon and fibrocartilage differentiation. A subset of these were also identified in previous microarray and sequencing datasets of embryonic tendon cells^{22,23}, further demonstrating the potential of our differentiation models to recapitulate *in vivo* developmental processes. Future work will test the function of these transcription factors in tendon and fibrocartilage specification through gain or loss of function experiments. These findings provide new insight in the molecular regulation of these clinically challenging, dense connective tissues. Further, since definitive phenotypic markers remain limited for these tissues, our data now provide comprehensive transcriptional signatures associated with the acquisition of tendon and fibrocartilage cell fates.

Although the development and maintenance of fibrous vs cartilaginous tissues is typically thought to be driven by mechanical loading demands (tensile and compressive forces, respectively)^{2,3,11}, our studies also suggest that molecular factors alone can control the switch between tendon and fibrocartilage fates under the same 2D environment. However, prolonged culture under 2D conditions resulted in progressive loss of tenogenic phenotype, consistent with prior reports^{91–93}. We overcame this limitation by inducing tenogenesis under 3D culture conditions and showed improved differentiation compared to 2D culture. Surprisingly, the fibrocartilage condition appeared largely insensitive to 3D environment since tendon and cartilage genes were relatively unchanged with 2D or 3D culture. Although compressive loading is thought to drive chondrogenesis, cartilaginous nodules were still observed under static tension, suggesting that biochemical cues can drive fibrocartilage formation independent of mechanical signals. This difference in mechanical dependency between tendon and fibrocartilage is intriguing and supported by some evidence in the developmental literature. For tendons, loss of muscles results in early and complete loss of forearm tendons in the mouse limb⁶³. In contrast, the meniscus is formed and maintained when muscles are absent, but dynamic muscle contraction is required for the maintenance of matrix markers^{94,95}. Using our system, we can dissect mechanotransduction mechanisms underlying tendon and fibrocartilage development and precisely control loading parameters, which is not feasible *in vivo* during mouse embryonic stages.

Here, we demonstrate the utility of these high-efficiency protocols for studying tendon and fibrocartilage development, mechanobiology, and signaling, however, they can also be applied for high-throughput small molecule screening to identify drug targets for clinical applications. Moreover, the wide availability of genetic tools in the mouse system also allows the use of mESCs derived from mutant lines for disease modeling or test loss of gene function *in vitro* in the context of tendon and fibrocartilage induction. Future studies will test these applications to advance scientific understanding and develop potential therapies for regenerative repair of these dense connective tissues.

Methods

Tenogenic differentiation of mESCs by activation of TGF β and hedgehog signaling. mESCs were derived from *ScxGFP* tendon reporter mice⁹. Briefly, mouse blastocysts (E3.5–3.75) from *ScxGFP* mice were individually plated onto irradiated primary mouse embryonic fibroblasts (MEFs, 1×10^5 cells/cm², R&D Systems Cat #PSC001) for 4–5 days in ES media (DMEM with 1 mM non-essential amino acids, 1 mM sodium pyruvate, 2 mM L-glutamine, 10 μ M beta-mercaptoethanol, 15% ES-qualified fetal calf serum, 50 μ g/mL penicillin/streptomycin, 1000 U/mL LIF). Outgrowths of the inner cell mass were then manually isolated and replated and passaged several times until characteristic ES colonies were observed⁹⁶. A total of 13 lines were obtained and genotyped for *ScxGFP*, which showed 8 *ScxGFP*⁺ and 5 *ScxGFP*⁻ mESC lines. Experiments were carried out using 3 *ScxGFP*⁺ lines and a *ScxGFP*⁻ line was cultured and differentiated in parallel for each experiment to set flow cytometry gates for GFP.

Undifferentiated mESCs were maintained with fetal bovine serum (FBS) and 20 ng/ μ L leukemia inhibiting factor (LIF) on MEF feeders⁹⁷. All experiments were conducted by passage 12 or below. To differentiate mESCs, single-cell suspensions were cultured in Serum-Free Differentiation (SFD) media for 2 days with 500 nM LDN-193189 for embryoid body formation⁹⁸. At day 2, SFD with 500 nM LDN-193189 and 5 μ M CHIR was used for 4 days. On day 6, tenogenesis was induced using SFD with 10 ng/mL TGF β 1, 100 nM SAG, and 10 ng/mL FGF2. *ScxGFP* induction was determined using flow cytometry and detection of *ScxGFP*^{hi} and *ScxGFP*^{lo} cells was performed based on TGF β 1+ conditions as described below.

Optimization of tendon differentiation by manipulation of retinoic acid signaling. To enhance paraxial mesoderm differentiation, embryoid body formation was carried out as described above, followed by 4 days of SFD with 500 nM LDN-193189, 5 μ M CHIR, and either 100 nM AGN 193109 or 30 nM ATRA. Tendon-specific differentiation was subsequently induced using 4 days of SFD with 10 ng/mL TGF β 1, 100 nM SAG, 10 ng/mL FGF2, and 1 μ M CD1530. Fibrocartilage-specific differentiation was induced for 4 days in SFD with 10 ng/mL TGF β 1, 100 nM SAG, 10 ng/mL FGF2, and 100 nM AGN 193109.

Flow cytometry. Single-cell suspensions of dissociated cells were obtained at day 10 of differentiation and stained 1:1000 with DAPI (ThermoFisher) in 2% FBS in PBS. Flow cytometry was carried out on an LSRIIA instrument (BD Sciences) using FACSDiva v8.0.3. Flow cytometry analysis was carried out using FCS Express v7. Gating for *ScxGFP* positivity was performed with a negative control mESC line (derived from littermates) that did not contain the *ScxGFP* transgene and represented as a percentage of the parent gate. Flow cytometry gating strategy for *ScxGFP* is shown in Supplementary Fig. 12. *ScxGFP* high and low gates were formed by identifying local minima in TGF β 1 containing conditions.

RNA isolation, reverse transcription, and real-time qPCR. RNA was isolated by Trizol/chloroform (Invitrogen) extraction, followed by cDNA synthesis (SuperScript VILO, Invitrogen), and gene expression assessed by qRT-PCR (SYBR, Applied Biosystems). Data were calculated using $2^{-\Delta\Delta C_t}$ with *Gapdh*. All primer sequences can be found in Supplementary Table 1.

Single-cell RNA sequencing. Single-cell suspensions from cultured cells and E14.5 embryonic mouse tail cells were prepared and dead cell depletion performed (EasySep Dead Cell Removal (Annexin V) Kit, STEMCELL). Live cells were input into a 10x Genomics Chromium device. Single-cell RNA libraries were prepared using the Chromium Single Cell 3' v3 Reagent Kit (10x Genomics) user guide. Sample libraries were pooled together at equimolar ratios and sequenced on an Illumina NextSeq 550 using 75-bp paired-end reads (high-output 75 cycle kit, cat no. 20024906), aiming for a sequencing depth of 20,000 reads/cell.

All subsequent data processing was conducted using the *cellranger* suite (version 3.1.0, 10X Genomics). Obtained FASTQ reads were aligned to the genome reference and mapped per cell using *cellranger* count. Multi-mapping reads were ignored during quantification. Gene-cell matrices were extracted and read into R and further analyzed using Seurat (version 3.0).

To exclude low-quality cells, the following quality control filters were applied and cells removed: cells with >7.5% of transcripts mapping to the mitochondrial genome, cells in the bottom 2% quantile (low quality) and top 2% quantiles (likely doublets/multiplets) of the transcripts detected per cell. Data of the resulting cells was normalized via the SCTransform method.

To confirm the identity of cultured cells, we used Seurat's integrated projection method, setting the scRNA-seq data from the embryonic tail as a reference and the cultured cells as the query. To account for batch effect differences, we used the Seurat alignment method to integrate the datasets. Developer defaults were used to conduct integration. The alignment leverages canonical correlation analysis to find linear relationships between features between datasets. Using these linear relationships, the method identifies a shared correlation space where accurate comparisons between datasets can be made. For each dataset, highly variable features were identified by taking the union of the top 3000 genes with the highest dispersion across both datasets. Canonical correlation analysis was applied upon these highly variable features to identify common sources of variation between the datasets. The first 30 correlational components were used to align the datasets and the resulting dataset was used for further analysis.

Principle component analysis was performed on the batch-corrected dataset. After visual inspection of the resulting components and contributions via scree-plot, 30 dimensions were used for unsupervised clustering and 2D projection of the data. Unsupervised clustering was done using the smart local moving algorithm () at a resolution of 0.2 to identify the major clusters. 2D projection and visualization of the clustering results were done using the uniform manifold approximation and projection (UMAP) method. UMAPs were generated using the same number of dimensions as with the clustering. Other gene expression plots were generated using Seurat base functions and ggplot2 v3.3.0. Gene ontology was performed using enrichR (2016 version)⁹⁹.

RNA sequencing. Time course RNA-seq was performed by preparing three biological replicates of samples from the undifferentiated PM and samples from the

tendon and fibrocartilage differentiation protocols at days 0.25 (6 h), 0.75 (18 h), 1.5, 2.5, and 4. RNA extraction was performed using Qiagen's RNeasy Plus Universal mini kit according to the manufacturer's instructions. Libraries were prepared using the NEB Next Ultra NRA Library Prep Kit for Illumina Sequencing in accordance with the manufacturer's instructions. The sequencing libraries were validated on the Agilent TapeStation (Agilent Technologies), and quantified by using Qubit 2.0 Fluorometer (Invitrogen) as well as by quantitative PCR (KAPA Biosystems).

The sequencing libraries were clustered on 1 lane of a flowcell. After clustering, the flowcell was loaded on the Illumina 4000 HiSeq instrument according to the manufacturer's instructions. The samples were sequenced using a 2 × 150 bp Paired End (PE) configuration. Image analysis and base calling were conducted by the HiSeq Control Software v3.6.3. Raw sequence data (.bcl files) generated from Illumina HiSeq was converted into fastq files and de-multiplexed using Illumina's bcl2fastq v2.20 software. One mismatch was allowed for index sequence identification.

Spliced transcripts were aligned and mapped to the *Mus musculus* reference genome (GRCm38) using STAR Aligner v2.7¹⁰⁰. Feature counting was subsequently performed with featureCounts v1.6.0¹⁰¹. Normalization, estimation of dispersion effects, and differential gene expression (DGE) were performed in R software v3.6.3 using edgeR v3.28.1¹⁰². Hierarchical clustering was implemented in SciPy v1.3.0¹⁰³. All images were visualized in Python v3.7.3 with Seaborn v0.9.0 and Matplotlib v3.1.0¹⁰⁴.

Proteomic analysis of receptor tyrosine kinases. The Mouse Phospho-Receptor Tyrosine Kinase (RTK) Array Kit (Catalog # ARY014, R&D Systems) was used to detect phosphorylation levels of 39 RTKs in day 10 differentiated cells in accordance with the manufacturer's protocol. Two hundred micrograms of total protein lysate from each sample was used per membrane. Protein concentration was determined using Pierce BCA Protein Assay Reagent (ThermoFisher Scientific). Following processing, each array membrane was exposed to X-ray film for 1 h and developed using a JPI X-ray film processor. Relative density of positive array hits was analyzed using ImageJ v1.53e software gel analysis and background subtraction applied using the included negative PBS control.

Collection of mouse embryos. Mouse embryos were collected from ScxGFP dams based on the time of conception according to Institutional Animal Care and Use Committee at the Icahn School of Medicine at Mount Sinai approved procedures (Protocol #IACUC-2014-0031) and consistent with animal care guidelines. Whole-mount imaging of embryos was carried out immediately at the time of harvest using Leica stereoscope fitted with GFP filter. Tails were then immediately placed in Trizol for RNA isolation.

Generation of 3D engineered tissues and dynamic loading. mESCs were differentiated toward D6 PM under 2D conditions, trypsinized, and embedded into 3D type I collagen gels at 250,000 cells per 2 mg/mL PureCol⁶². Cell-seeded gels were cast into sterilely prepared PDMS plates between two prolene sutures pinned in place by minuten pins. After casting, gels were placed in a humidified 37 °C incubator for 1 h to allow the gel to set before flooding with SDF media containing 10 ng/mL FGF2, 10 ng/mL TGFβ1, and 100 nM SAG. Tendon-specific conditions also included 1 μM CD1530 while fibrocartilage-specific conditions included 100 nM AGN 193109. For dynamic loading experiments, a custom 3D-printed mold was used to cast PDMS dishes that fit on the STREX Cell Stretching System. Cell-seeded gels were then created in these custom dishes as described above and cultured for 7 days in CD1530- or AGN 193109-containing media until full contraction. Validation of loading parameters was carried out by imaging gels during varying bioreactor inputs (2–15% elongation). The elongation experienced by gels was measured for each input using ImageJ. A calibration curve was then generated. Based on the literature indicating physiologic tendon loading of 6% elongation, we tested 3% and 6% elongation at 0.5 Hz for 4 h. Cell-seeded gels were collected immediately after loading for qPCR. Imaging of gels prior and after loading did not show changes in overall morphology or visible indications of damage.

Second-harmonic generation two-photon imaging. Imaging was performed on whole-mount constructs in Opti-MEM medium without phenol red using the Olympus FVMPE-RS multiphoton laser scanning microscope with a water immersive objective at ×25. Acellular gels were imaged in parallel and used as background controls. Multiphoton excitation was performed with Dual Line Insight X3 laser at an excitation of 880 nm. Second-harmonic generation (SHG) and ScxGFP were recorded by GaAsP PMT detectors (4 G and 3 G, respectively). All parameters (ie, laser intensity, gain, high voltage, and offset) were selected to minimize background noise without oversaturation and were held constant for all constructs. Images were converted using the Bio-Formats plugin and batch-processed in ImageJ. Mean pixel intensity were analyzed to quantify SHG signal and collagen density.

Immunofluorescence analyses. Engineered tendon constructs were fixed in 4% paraformaldehyde for 30 min, infiltrated in 15% and 30% sucrose (24 h), and embedded in OCT medium. Transverse cryosections (12 μm) were collected and

immunostaining for collagen type II carried out using anti-COL2α1 (II-II6B3, 1:100, Developmental Studies Hybridoma Bank) with secondary detection by streptavidin-Cy3 (016-160-084, 1:200, Jackson ImmunoResearch) with the M.O. M. kit (Vector Labs). Fluorescence imaging was carried out using the Zeiss Axioimager with Apotome optical sectioning and quantification performed in ImageJ.

Statistical analyses. Quantitative results are presented as mean ± standard deviation (SD). ANOVA was used for multiple comparisons; where significance was detected, posthoc testing was then carried out (Graphpad Prism). All other quantitative analyses were analyzed using unpaired Students *t*-tests. Statistical analyses are indicated in figure legends. Significant outliers were detected using Grubb's test (Graphpad Prism). Significance was determined at *p* < 0.05.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The scRNA-seq and RNA-seq data generated in this study have been deposited in the GEO repository database under accession codes "GSE154525" and "GSE154397". All other quantitative datasets including qPCR analyses, protein analyses, and immunostaining quantification are available in Supplementary Information/Source Data file provided with this paper. Source data are provided with this paper.

Received: 18 July 2020; Accepted: 24 June 2021;

Published online: 09 July 2021

References

- Zhang, G. et al. Development of tendon structure and function: regulation of collagen fibrillogenesis. *J. Musculoskelet. Neuronal Interact.* **5**, 5–21 (2005).
- Benjamin, M., Kaiser, E. & Milz, S. Structure-function relationships in tendons: a review. *J. Anat.* **212**, 211–228 (2008).
- Benjamin, M. & Ralphs, J. R. The cell and developmental biology of tendons and ligaments. *Int. Rev. Cytol.* **196**, 85–130 (2000).
- Huang, A. H. et al. Requirement for Scleraxis in the recruitment of mesenchymal progenitors during embryonic tendon elongation. *Development* <https://doi.org/10.1242/dev.182782> (2019).
- Murchison, N. et al. Regulation of tendon differentiation by scleraxis distinguishes force-transmitting tendons from muscle-anchoring tendons. *Development* **134**, 2697–2708 (2007).
- Liu, W. et al. The atypical homeodomain transcription factor Mohawk controls tendon morphogenesis. *Mol. Cell. Biol.* **30**, 4797–4807 (2010).
- Ito, Y. et al. The Mohawk homeobox gene is a critical regulator of tendon differentiation. *Proc. Natl Acad. Sci. USA* **107**, 10538–10542 (2010).
- Lejard, V. et al. EGR1 and EGR2 involvement in vertebrate tendon differentiation. *J. Biol. Chem.* **286**, 5855–5867 (2011).
- Pryce, B., Brent, A., Murchison, N., Tabin, C. & Schweitzer, R. Generation of transgenic tendon reporters, ScxGFP and ScxAP, using regulatory elements of the scleraxis gene. *Dev. Dyn.* **236**, 1677–1682 (2007).
- Yoshimoto, Y. et al. Scleraxis is required for maturation of tissue domains for proper integration of the musculoskeletal system. *Sci. Rep.* **7**, 45010 (2017).
- Benjamin, M. & Evans, E. J. Fibrocartilage. *J. Anat.* **171**, 1–15 (1990).
- Pryce, B. et al. Recruitment and maintenance of tendon progenitors by TGFβ signaling are essential for tendon formation. *Development* **136**, 1351–1361 (2009).
- Tan, G. K. et al. Tgfbeta signaling is critical for maintenance of the tendon cell fate. *eLife* <https://doi.org/10.7554/eLife.52695> (2020).
- Johnstone, B., Hering, T. M., Caplan, A. L., Goldberg, V. M. & Yoo, J. U. In vitro chondrogenesis of bone marrow-derived mesenchymal progenitor cells. *Exp. Cell Res.* **238**, 265–272 (1998).
- Pittenger, M. F. et al. Multilineage potential of adult human mesenchymal stem cells. *Science* **284**, 143–147 (1999).
- Blitz, E., Sharif, A., Akiyama, H. & Zelder, E. Tendon-bone attachment unit is formed modularly by a distinct pool of Scx- and Sox9-positive progenitors. *Development* **140**, 2680–2690 (2013).
- Craft, A. M. et al. Generation of articular chondrocytes from human pluripotent stem cells. *Nat. Biotechnol.* **33**, 638–645 (2015).
- Sugimoto, Y. et al. Scx+/Sox9+ progenitors contribute to the establishment of the junction between cartilage and tendon/ligament. *Development* **140**, 2280–2288 (2013).
- Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).
- Tabula Muris, C. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).

21. Han, X. et al. Mapping the mouse cell atlas by microwell-seq. *Cell* **173**, 1307 (2018).
22. Havis, E. et al. Transcriptomic analysis of mouse limb tendon cells during development. *Development* **141**, 3683–3696 (2014).
23. Liu, H. et al. Whole transcriptome expression profiling of mouse limb tendon development by using RNA-seq. *J. Orthop. Res.* **33**, 840–848 (2015).
24. Komura, S. et al. Induced pluripotent stem cell-derived tenocyte-like cells promote the regeneration of injured tendons in mice. *Sci. Rep.* **10**, 3992 (2020).
25. Nakajima, T. et al. Modeling human somite development and fibrodysplasia ossificans progressiva with induced pluripotent stem cells. *Development* <https://doi.org/10.1242/dev.165431> (2018).
26. Dale, T. P. et al. Tenogenic differentiation of human embryonic stem cells. *Tissue Eng. Part A* **24**, 361–368 (2018).
27. Zhao, J. et al. Small molecule-directed specification of sclerotome-like chondroprogenitors and induction of a somitic chondrogenesis program from embryonic stem cells. *Development* **141**, 3848–3858 (2014).
28. Craft, A. M. et al. Specification of chondrocytes and cartilage tissues from embryonic stem cells. *Development* **140**, 2597–2610 (2013).
29. Chal, J. et al. Recapitulating early development of mouse musculoskeletal precursors of the paraxial mesoderm in vitro. *Development* <https://doi.org/10.1242/dev.157339> (2018).
30. Brent, A., Schweitzer, R. & Tabin, C. A somitic compartment of tendon progenitors. *Cell* **113**, 235–248 (2003).
31. Marcelle, C., Stark, M. R. & Bronner-Fraser, M. Coordinate actions of BMPs, Wnts, Shh and noggin mediate patterning of the dorsal somite. *Development* **124**, 3955–3963 (1997).
32. Fan, C. M. et al. Long-range sclerotome induction by sonic hedgehog: direct role of the amino-terminal cleavage product and modulation by the cyclic AMP signaling pathway. *Cell* **81**, 457–465 (1995).
33. Borycki, A. G., Mendham, L. & Emerson, C. P. Jr. Control of somite patterning by Sonic hedgehog and its downstream signal response genes. *Development* **125**, 777–790 (1998).
34. Murtaugh, L. C., Chyung, J. H. & Lassar, A. B. Sonic hedgehog promotes somitic chondrogenesis by altering the cellular response to BMP signaling. *Genes Dev.* **13**, 225–237 (1999).
35. Teillet, M. et al. Sonic hedgehog is required for survival of both myogenic and chondrogenic somitic lineages. *Development* **125**, 2019–2030 (1998).
36. Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
37. Schweitzer, R. et al. Analysis of the tendon cell fate using Scleraxis, a specific marker for tendons and ligaments. *Development* **128**, 3855–3866 (2001).
38. Blitz, E. et al. Bone ridge patterning during musculoskeletal assembly is mediated through SCX regulation of Bmp4 at the tendon-skeleton junction. *Dev. Cell* **17**, 861–873 (2009).
39. Waltman, L. & van Eck, N. J. A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* **86**, 471 (2013).
40. Taye, N., Karoulias, S. Z. & Hubmacher, D. The “other” 15–40%: the role of non-collagenous extracellular matrix proteins and minor collagens in tendon. *J. Orthop. Res.* **38**, 23–35 (2020).
41. Zhang, G. et al. Decorin regulates assembly of collagen fibrils and acquisition of biomechanical properties during tendon development. *J. Cell Biochem.* **98**, 1436–1449 (2006).
42. Tagliaferri, D. et al. Retinoic acid induces embryonic stem cells (ESCs) transition to 2 cell-like state through a coordinated expression of Dux and Duxbl1. *Front Cell Dev. Biol.* **7**, 385 (2019).
43. De Angelis, M. T., Parrotta, E. I., Santamaria, G. & Cuda, G. Short-term retinoic acid treatment sustains pluripotency and suppresses differentiation of human induced pluripotent stem cells. *Cell Death Dis.* **9**, 6 (2018).
44. Janesick, A., Wu, S. C. & Blumberg, B. Retinoic acid signaling and neuronal differentiation. *Cell Mol. Life Sci.* **72**, 1559–1576 (2015).
45. Iwafuchi-Doi, M. et al. Transcriptional regulatory networks in epiblast cells and during anterior neural plate development as modeled in epiblast stem cells. *Development* **139**, 3926–3937 (2012).
46. Chapman, D. L., Agulnik, I., Hancock, S., Silver, L. M. & Papaioannou, V. E. Tbx6, a mouse T-Box gene implicated in paraxial mesoderm formation at gastrulation. *Dev. Biol.* **180**, 534–542 (1996).
47. Pevny, L. H., Sockanathan, S., Placzek, M. & Lovell-Badge, R. A role for SOX1 in neural determination. *Development* **125**, 1967–1978 (1998).
48. Shimono, K. et al. Potent inhibition of heterotopic ossification by nuclear retinoic acid receptor- γ agonists. *Nat. Med.* **17**, 454–460 (2011).
49. Horton, W. E., Yamada, Y. & Hassell, J. R. Retinoic acid rapidly reduces cartilage matrix synthesis by altering gene transcription in chondrocytes. *Dev. Biol.* **123**, 508–516 (1987).
50. Pacifici, M., Cossu, G., Molinaro, M. & Tato, F. Vitamin A inhibits chondrogenesis but not myogenesis. *Exp. Cell Res.* **129**, 469–474 (1980).
51. Ataliotis, P. Platelet-derived growth factor A modulates limb chondrogenesis both in vivo and in vitro. *Mech. Dev.* **94**, 13–24 (2000).
52. Fisher, M. C., Clinton, G. M., Maihle, N. J. & Dealy, C. N. Requirement for ErbB2/ErbB signaling in developing cartilage and bone. *Dev. Growth Differ.* **49**, 503–513 (2007).
53. Jia, H. et al. EGFR signaling is critical for maintaining the superficial layer of articular cartilage and preventing osteoarthritis initiation. *Proc. Natl Acad. Sci. USA* **113**, 14360–14365 (2016).
54. Kwon, H. R., Kim, J. H., Woods, J. P. & Olson, L. E. An activating mutation in Pdgfrb causes skeletal stem cell defects with osteopenia and overgrowth in mice. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.01.21.427619> (2021).
55. Marturano, J. E., Arena, J. D., Schiller, Z. A., Georgakoudi, I. & Kuo, C. K. Characterization of mechanical and biochemical properties of developing embryonic tendon. *Proc. Natl Acad. Sci. USA* **110**, 6370–6375 (2013).
56. Kult, S. et al. Bi-fated tendon-to-bone attachment cells are regulated by shared enhancers and KLF transcription factors. *eLife* <https://doi.org/10.7554/eLife.55361> (2021).
57. Hasty, P. et al. Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature* **364**, 501–506 (1993).
58. Nabeshima, Y. et al. Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature* **364**, 532–535 (1993).
59. Hsu, S. H. et al. Dlx5- and Dlx6-mediated chondrogenesis: differential domain requirements for a conserved function. *Mech. Dev.* **123**, 819–830 (2006).
60. Zhu, H. & Bendall, A. J. Dlx5 is a cell autonomous regulator of chondrocyte hypertrophy in mice and functionally substitutes for Dlx6 during endochondral ossification. *PLoS ONE* **4**, e8097 (2009).
61. Ferrari, D. & Kosher, R. A. Dlx5 is a positive regulator of chondrocyte differentiation during endochondral ossification. *Dev. Biol.* **252**, 257–270 (2002).
62. Gehwolf, R. et al. 3D-embedded cell cultures to study tendon biology. *Methods Mol. Biol.* **2045**, 155–165 (2019).
63. Huang, A. H. et al. Musculoskeletal integration at the wrist underlies the modular development of limb tendons. *Development* **142**, 2431–2441 (2015).
64. Pan, X. S., Li, J., Brown, E. B. & Kuo, C. K. Embryo movements regulate tendon mechanical property development. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* <https://doi.org/10.1098/rstb.2017.0325> (2018).
65. Mikic, B., Isenstein, A. L. & Chhabra, A. Mechanical modulation of cartilage structure and function during embryogenesis in the chick. *Ann. Biomed. Eng.* **32**, 18–25 (2004).
66. Magnusson, S. P. et al. Differential strain patterns of the human gastrocnemius aponeurosis and free tendon, in vivo. *Acta Physiol. Scand.* **177**, 185–195 (2003).
67. Bron, J. L., Helder, M. N., Meisel, H. J., Van Royen, B. J. & Smit, T. H. Repair, regenerative and supportive therapies of the annulus fibrosus: achievements and challenges. *Eur. Spine J.* **18**, 301–313 (2009).
68. Abram, S. G. F. et al. Arthroscopic partial meniscectomy for meniscal tears of the knee: a systematic review and meta-analysis. *Br. J. Sports Med.* **54**, 652–663 (2020).
69. Thomopoulos, S., Parks, W. C., Rifkin, D. B. & Derwin, K. A. Mechanisms of tendon injury and repair. *J. Orthop. Res.* <https://doi.org/10.1002/jor.22806> (2015).
70. Ansoorge, H. et al. Recapitulation of the Achilles tendon mechanical properties during neonatal development: a study of differential healing during two stages of development in a mouse model. *J. Orthop. Res.* **30**, 448–456 (2012).
71. Beredjickian, P. et al. Regenerative versus reparative healing in tendon: a study of biomechanical and histological properties in fetal sheep. *Ann. Biomed. Eng.* **31**, 1143–1152 (2003).
72. Howell, K. et al. Novel model of tendon regeneration reveals distinct cell mechanisms underlying regenerative and fibrotic tendon healing. *Sci. Rep.* **7**, 45238 (2017).
73. Kaji, D. A., Howell, K. L., Balic, Z., Hubmacher, D. & Huang, A. H. TGF β signaling is required for tenocyte recruitment and functional neonatal tendon regeneration. *eLife* <https://doi.org/10.7554/eLife.51779> (2020).
74. Torre, O. M., Das, R., Berenblum, R. E., Huang, A. H. & Iatridis, J. C. Neonatal mouse intervertebral discs heal with restored function following herniation injury. *FASEB J.* **32**, 4753–4762 (2018).
75. Lui, P. P. Stem cell technology for tendon regeneration: current status, challenges, and future research directions. *Stem Cells Cloning* **8**, 163–174 (2015).
76. Yamanaka, S. Pluripotent stem cell-based cell therapy-promise and challenges. *Cell Stem Cell* **27**, 523–531 (2020).
77. Disser, N. P. et al. Widespread diversity in the transcriptomes of functionally divergent limb tendons. *J. Physiol.* **598**, 1537–1550 (2020).
78. Sohn, P., Cox, M., Chen, D. & Serra, R. Molecular profiling of the developing mouse axial skeleton: a role for Tgfb2 in the development of the intervertebral disc. *BMC Dev. Biol.* **10**, 29 (2010).
79. Mikic, B., Entwistle, R., Rossmeier, K. & Bierwert, L. Effect of GDF-7 deficiency on tail tendon phenotype in mice. *J. Orthop. Res.* **26**, 834–839 (2008).

80. Mikic, B., Rossmeier, K. & Bierwert, L. Sexual dimorphism in the effect of GDF-6 deficiency on murine tendon. *J. Orthop. Res.* **27**, 1603–1611 (2009).
81. Mikic, B., Rossmeier, K. & Bierwert, L. Identification of a tendon phenotype in GDF6 deficient mice. *Anat. Rec.* **292**, 396–400 (2009).
82. Mikic, B., Schalet, B. J., Clark, R. T., Gaschen, V. & Hunziker, E. B. GDF-5 deficiency in mice alters the ultrastructure, mechanical properties and composition of the Achilles tendon. *J. Orthop. Res.* **19**, 365–371 (2001).
83. Schlesinger, S. Y. et al. Loss of Smad4 in the scleraxis cell lineage results in postnatal joint contracture. *Dev. Biol.* **470**, 108–120 (2021).
84. Comai, G. E. et al. Local retinoic acid signaling directs emergence of the extraocular muscle functional unit. *PLoS Biol.* **18**, e3000902 (2020).
85. McGurk, P. D., Swartz, M. E., Chen, J. W., Galloway, J. L. & Eberhart, J. K. In vivo zebrafish morphogenesis shows Cyp26b1 promotes tendon condensation and musculoskeletal patterning in the embryonic jaw. *PLoS Genet.* **13**, e1007112 (2017).
86. Webb, S., Gabrelow, C., Pierce, J., Gibb, E. & Elliott, J. Retinoic acid receptor signaling preserves tendon stem cell characteristics and prevents spontaneous differentiation in vitro. *Stem Cell Res. Ther.* **7**, 45 (2016).
87. Williams, J. A. et al. Retinoic acid receptors are required for skeletal growth, matrix homeostasis and growth plate function in postnatal mouse. *Dev. Biol.* **328**, 315–327 (2009).
88. Xu, Q. & Kopp, J. B. Retinoid and TGF-beta families: crosstalk in development, neoplasia, immunity, and tissue repair. *Semin Nephrol.* **32**, 287–294 (2012).
89. Clayton, S. W., Ban, G. I., Liu, C. & Serra, R. Canonical and noncanonical TGF-beta signaling regulate fibrous tissue differentiation in the axial skeleton. *Sci. Rep.* **10**, 21364 (2020).
90. Li, X. et al. Retinoic acid stimulates chondrocyte differentiation and enhances bone morphogenetic protein effects through induction of Smad1 and Smad5. *Endocrinology* **144**, 2514–2523 (2003).
91. Chien, C., Pryce, B., Tufa, S. F., Keene, D. R. & Huang, A. H. Optimizing a 3D model system for molecular manipulation of tenogenesis. *Connect Tissue Res.* <https://doi.org/10.1080/03008207.2017.1383403> (2017).
92. Gaut, L. et al. Mechanical and molecular parameters that influence the tendon differentiation potential of C3H10T1/2 cells in 2D- and 3D-culture systems. *Biol. Open* <https://doi.org/10.1242/bio.047928> (2020).
93. Kuo, C. K. & Tuan, R. S. Mechanoactive tenogenic differentiation of human mesenchymal stem cells. *Tissue Eng. Part A* **14**, 1615–1627 (2008).
94. Pazin, D. E., Gamer, L. W., Cox, K. A. & Rosen, V. Molecular profiling of synovial joints: use of microarray analysis to identify factors that direct the development of the knee and elbow. *Dev. Dyn.* **241**, 1816–1826 (2012).
95. Mikic, B., Wong, M., Chiquet, M. & Hunziker, E. B. Mechanical modulation of tenascin-C and collagen-XII expression during avian synovial joint formation. *J. Orthop. Res.* **18**, 406–415 (2000).
96. Hogan, B. L., Beddington, R. S., Constantini, F. & Lacy, E. in *Manipulating the Mouse Embryo: A Laboratory Manual* 266–272 (Cold Spring Harbor Laboratory Press, 1994).
97. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
98. Gouon-Evans, V. et al. BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. *Nat. Biotechnol.* **24**, 1402–1411 (2006).
99. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
100. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
101. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
102. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
103. Bar-Joseph, Z., Gifford, D. K. & Jaakkola, T. S. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics* **17**, S22–S29 (2001). Suppl 1.
104. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

Acknowledgements

This work was supported by funding from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) to A.H. (R01 AR069537) and D.K. (F31 AR073626), and from NYSTEM to A.H. (C32570GG). We also acknowledge Dr. Ronen Schweitzer for providing ScxGFP mice, the Mouse Genetics and Gene Targeting Core for the derivation of mESCs from ScxGFP mice, the Flow Cytometry Core, qPCR Core, and the Human Immune Monitoring Center for single-cell RNA sequencing. The monoclonal antibody II-II6B3 developed by T.F. Linsenmayer was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at the University of Iowa, Department of Biology, Iowa City, IA 52242.

Author contributions

D.K. and A.H. conceived and designed the experiments. D.K. and A.M. performed tissue culture experiments and analyses. D.K. and R.P. performed bioinformatic analyses of the single-cell and bulk RNA sequencing datasets. The manuscript was written by D.K. and A.H. and edited and approved by all authors.

Competing interests

The authors have no competing interests to declare.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-24535-5>.

Correspondence and requests for materials should be addressed to A.H.H.

Peer review information *Nature Communications* thanks Guy Genin, Qiong Wu and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



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

© The Author(s) 2021