

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

## A single cell transcriptional atlas of early synovial joint development

Qin Bian<sup>1,2</sup>, Yu-Hao Cheng<sup>1,3</sup>, Jordan P Wilson<sup>1</sup>, Dong Won Kim<sup>4</sup>, Hong Wang<sup>4</sup>,  
Seth Blackshaw<sup>1,4</sup>, Patrick Cahan<sup>1,2,3</sup>

<sup>1</sup>Institute for Cell Engineering, Johns Hopkins School of Medicine, Baltimore MD 21205 USA

<sup>2</sup>Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore MD 21205 USA

<sup>3</sup>Department of Molecular Biology and Genetics, Johns Hopkins School of Medicine, Baltimore MD 21205 USA

<sup>4</sup>Solomon H. Snyder Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore MD 21205 USA

Correspondence to: [patrick.cahan@jhmi.edu](mailto:patrick.cahan@jhmi.edu)

**Keywords:** synovial joint development; interzone; single cell RNA-Seq; articular cartilage; ligament; synovium; chondrocyte; meniscus; Gdf5

44 **SUMMARY**

45

46 Synovial joint development begins with the formation of the interzone, a region of condensed  
47 mesenchymal cells at the site of the prospective joint. Recently, lineage tracing strategies have  
48 revealed that Gdf5-lineage cells native to and from outside the interzone contribute to most, if  
49 not all, of the major joint components. However, there is limited knowledge of the specific  
50 transcriptional and signaling programs that regulate interzone formation and fate diversification  
51 of synovial joint constituents. To address this, we have performed single cell RNA-Seq analysis  
52 of 6,202 synovial joint progenitor cells from the developing murine knee joint from E12.5 to  
53 E15.5. By using a combination of computational analytics, *in situ* hybridization, and functional  
54 analysis of prospectively isolated populations, we have inferred the underlying transcriptional  
55 networks of the major developmental paths for joint progenitors. Our freely available single cell  
56 transcriptional atlas will serve as a resource for the community to uncover transcriptional  
57 programs and cell interactions that regulate synovial joint development.

58

59

60

61

62

63

64

65

66

67

68

69

70

71

## 72 Introduction

73  
74 Synovial joints are complex anatomical structures comprised of diverse tissues, including  
75 articular cartilage, synovium, fibrous capsule, and ligaments (Decker et al. 2014). Each of these  
76 tissues are susceptible to a range of diseases—both congenital and degenerative—and by  
77 common injuries that collectively have a profound global morbidity (den Hollander et al. 2019;  
78 Asahara et al. 2017). A better understanding of the inter- and intra-cellular networks that govern  
79 how these structures emerge during development will inform efforts to generate pluripotent stem  
80 cell derivatives for cell replacement therapy and disease modeling (Wang et al. 2019) and  
81 efforts to elicit regeneration *in situ* (Johnson et al. 2012). Moreover, an improved understanding  
82 of joint development will aid in identifying putative disease causing genes (Kelly et al. 2019).

83  
84 Over the past two decades, lineage tracing has revealed much regarding the cell populations  
85 contributing to murine synovial joint development. It begins with the formation of the interzone  
86 (IZ), a region of condensed mesenchymal cells at the site of the prospective joint. In the mouse  
87 hindlimb, the IZ is initiated from a Col2a1<sup>+</sup> Sox9<sup>+</sup> pool of cells recruited from the mesenchymal  
88 condensation of the emerging limb bud starting at E11.5 (Hyde et al. 2008; Soeda et al. 2010). It  
89 is generally believed that chondrocytes at the presumptive joint de-differentiate (i.e. undergo a  
90 chondrocyte-to-mesenchymal transition) and begin to exhibit the flattened and layered  
91 morphology that is indicative of the IZ. A history of expressing Gdf5, a TGFβ ligand and critical  
92 contributor to joint formation (Storm and Kingsley 1999), marks cells that initially form the IZ or  
93 that later immigrate into it, and that subsequently go on to contribute to all of the major joint  
94 constituents including articular chondrocytes, ligament, meniscus, and synovium (Shwartz et al.  
95 2016; Chen et al. 2016).

96  
97 To gain a more comprehensive understanding of these developmental programs, bulk  
98 microarray expression profiling and RNA-Seq have been applied to the developing limb (Taher  
99 et al. 2011), to whole joints including the elbow and knee (Pazin et al. 2012), to the meniscus  
100 (Pazin et al. 2014), to the tendon (Liu et al. 2015), to connective tissue (Orgeur et al. 2018), and  
101 to laser-capture micro-dissected regions of the interzone (Jenner et al. 2014). While these  
102 investigations have yielded new insights into the genetic programs underpinning limb and joint  
103 morphogenesis, they provide limited resolution of the expression states for individual cell types  
104 due to the heterogenous nature of the samples profiled. With the advent of single cell profiling, it  
105 is now possible to detect transient populations of cells, to reconstruct developmental

106 transcriptional programs, and to identify new cell populations (Guo et al. 2010; Kumar et al.  
107 2017). For example, Feng et al revealed molecular signatures and lineage trajectories of an  
108 interzone related  $Lgr5^+$  population in the murine E14.5 knee joint that contributes to the  
109 formation of cruciate ligaments, synovial membrane, and articular chondrocytes (Feng et al.  
110 2019).

111  
112 Here, we applied single-cell RNA-sequencing on  $Gdf5$ -lineage cells of the murine hindlimb to  
113 determine the transcriptional programs of early synovial joint development. In contrast to the  
114 recent study of Feng et al 2019, which focused on lineage divergence of a specific  $Lrg5^+$   
115 interzone population, we sought to characterize formation of the entire IZ and to discover the  
116 extent to which heterogeneity in the nascent interzone is resolved into the distinct lineages that  
117 are apparent later at cavitation. Therefore, we sequenced  $Gdf5$ -lineage cells from the  
118 presumptive joint of the hindlimb from E12.5 (prior to frank IZ formation) through E15.5  
119 (coinciding with cavitation). We combined computational analytics and *in situ* hybridization to  
120 infer the lineage relationships of joint progenitors and to identify the combinatorial transcriptional  
121 programs that mediate the elaboration of the interzone into the major synovial joint lineages. We  
122 found that early  $Gdf5$ -lineage enriched cells consist of sub-populations with chondrogenic or  
123 fibrous-lineage bias. Furthermore, we discovered within the chondrogenic-biased population  
124 were two distinct sub-populations that followed similar trajectories to de-differentiate into IZ  
125 cells, supporting a model of regionally and temporally complex IZ specification (Shwartz et al.  
126 2016). To aid the community in discovering additional transcriptional programs and in inferring  
127 cell interactions that contribute to synovial joint development, we have made this data freely and  
128 easily accessible with a web application at [http://www.cahanlab.org/resources/joint\\_ontology](http://www.cahanlab.org/resources/joint_ontology).

129

## 130 **Results**

131

132  **$Gdf5^{Cre^+}$  cells in the hind limb from E12.5 to E15.5  $Gdf5^{Cre}R26^{EYFP}$  mice are primarily**  
133 **located in the interzone, articular cartilage, ligament, menisci, and synovium, as well as**  
134 **in other non-joint tissues**

135

136  $Gdf5$ -lineage cells contribute to several components of synovial joint, including articular  
137 cartilage, meniscus, ligaments, and synovium. To isolate joint progenitors, we crossed  $Gdf5$   
138 promoter driven Cre mice with the R26 reporter mice in which loxP-flanked STOP sequence  
139 followed by the EYFP inserted into the  $Gt(ROSA)26Sor$  locus, allowing us to identify  $Gdf5$ -

140 lineage cells by YFP expression. We used fluorescent immunohistochemistry to determine the  
141 spatial and temporal pattern of YFP. At E12.5, YFP is mainly expressed in the presumptive joint  
142 area including part of the bone anlagen and the surrounding mesenchyme (**Fig 1**). At E13.5,  
143 YFP<sup>+</sup> cells are more centered in the interzone (IZ) and in the surrounding connective tissue;  
144 they are sparse in the anlagen of the femur and tibia. By E14.5, YFP<sup>+</sup> staining is mainly present  
145 at the area of future articular cartilage (AC), synovium and surrounding soft tissue. YFP  
146 expression becomes obvious in menisci one day later. YFP<sup>+</sup> cells are also seen in AC,  
147 epiphyseal cartilage, and synovium at E15.5.

148  
149 We observed “ectopic” YFP expression in non-joint tissues such as the dermis and muscle,  
150 consistent with prior reports (Roelofs et al. 2017). However, because our scRNA-Seq analysis  
151 pipeline includes a “cell typing” step (see below), we were able to identify these non-joint cells *in*  
152 *silico* and exclude them from our in-depth analyses that focus on the Gdf5-lineages of the joint.  
153 We refer to cells that passed our *in silico* filtering as Gdf5-lineage enriched (GLE) cells rather  
154 than YFP<sup>+</sup> Gdf5-lineage cells because we cannot absolutely prove that YFP expression tracks  
155 with Gdf5-lineage in this system. Nonetheless, our staining combined with prior reports  
156 examining Gdf5cre cells in the limb, indicate that GLE cells are major cellular contributors to the  
157 knee joint. Therefore, determining their transcriptomes will yield insights into the genetic circuitry  
158 that accompanies IZ formation and the emergence of articular components such as ligament  
159 and tendon.

160

### 161 **GLE cells form three distinct super clusters across two major developmental stages**

162

163 To define the transcriptional states of joint cells and their progenitors during landmark  
164 developmental events, we isolated YFP<sup>+</sup> cells from the hind limbs of male embryos from E12.5  
165 (the time just prior to frank IZ formation) to E15.5 (before cavitation). To minimize contamination  
166 with Gdf5-lineage cells from the ankle and digits, we manually dissected the region of the limb  
167 containing the presumptive joint and excluded the paw (**Supp Fig 1A**). Then, we collected Gdf5  
168 lineage cells by fluorescence activated cell sorting (FACs) of YFP<sup>+</sup> cells after enzymatically  
169 disassociating the presumptive knee joint region (**Supp Fig 1B**). We loaded approximately  
170 6,000 cells for single cell RNA-Seq library preparation using the 10x Genomics platform, and  
171 sequenced the transcriptome of ~1,000 to 5,000 cells at a target depth of 100,000 reads per cell  
172 (**Table 1**).

173

174 After performing quality control to remove potential doublets and low-quality libraries, we sought  
175 to identify the major transcriptional states in our data by clustering using the Leiden graph-  
176 based community detection algorithm (Traag et al. 2019). We found 14 clusters, many of which  
177 contained cells from multiple timepoints (**Supp Fig 2A**). To determine the cell type of each  
178 cluster, we used SingleCellNet to classify individual cells based on a well-annotated reference  
179 data set (Tan and Cahan 2019), and we used differential gene expression to identify marker  
180 genes of cell types that are not included in current single cell reference data sets (e.g. neural  
181 crest cells and melanocytes). This approach identified eight clusters made up of non-joint cell  
182 types including myoblasts, immune and red blood cells, neural crest cells and melanocytes, and  
183 endothelial cells (**Supp Fig 2B**). After removing these non-joint cells, we re-clustered the data  
184 and we performed differential gene expression analysis (**Supp Fig 2C**). All clusters had  
185 detectable levels of the osteochondral transcription factor (TF) Sox9 except one, which had high  
186 levels of genes associated with dermis, including Twist2 and Irx1 (**Supp Fig 2D**). To localize the  
187 cells in this cluster, we performed *in situ* hybridization (ISH), confirming that they are dermal  
188 cells (**Supp Fig 2E**), and we excluded these cells from further analysis. Finally, we performed  
189 cell cycle analysis by scoring each cell according to its likely phase (G1, G2M, or S) based on  
190 expression of canonical cell cycle-related genes (**Supp Fig 2C**). We removed two clusters,  
191 which were comprised predominately of cells in G2M or S phase, as we found that including  
192 these cells confounded downstream analysis. This cell trimming process resulted in a data set  
193 of 6,202 synovial joint GLE cells.

194  
195 Next, we asked whether there were discernible transcriptional profiles that spanned timepoints.  
196 To address this question, we clustered all of the GLE cells and uncovered three 'super-clusters'  
197 (SCs), two of which contain a plurality of cells from more than a single timepoint (**Fig 2A-B**).  
198 One of the clusters corresponds roughly to developmental time: SC1 is 98.2% E12.5 cells. The  
199 other two SCs are mixtures, with SC2 and SC3 predominately made up of cells from E13.5-  
200 E15.5. To gain a better understanding of these SCs, we examined the expression of genes with  
201 well-established roles in limb and joint development. Prrx1 and Pitx1 are preferentially  
202 expressed in the early SC1 (**Fig 2C**), consistent with their roles in specifying limb mesenchymal  
203 cells from lateral plate mesoderm (Bobick and Cobb 2012; Marcil et al. 2003; Wang et al. 2018).  
204 Shox2, regulating onset of early chondrogenesis (Bobick and Cobb 2012) has a similar  
205 expression pattern. Since many cells in SC1 express Sox9 but few express Col2a1, it is likely  
206 that this supercluster is comprised of a mixture of progenitor cells of mesenchymal character  
207 and chondroprogenitors. 25% of SC1 cells express the IZ marker Gdf5, and thus may represent

208 de-differentiated chondrocytes. SC2 is similar to SC1 in expression profile, but it also  
209 preferentially expresses Sox9, Gdf5, Col11a1 and Col2a1, suggesting that this SC is likely to  
210 contain a mixture of IZ cells and transient chondrocytes (Zhao et al. 1997). SC3 cells express  
211 fibrous related genes Col3a1, Col1a1, Lgals1 (Dasuri et al. 2004), Dcn (Havis et al. 2014),  
212 indicating that SC3 largely consists of fibroblast-related cells.

213

214 Gene set enrichment analysis largely corroborated our supervised annotation of the  
215 superclusters (**Fig 2D**). SC1 is enriched in limb and joint development-associated pathways  
216 including embryonic limb morphogenesis, Notch signaling (Jiang et al. 1998), and epithelial to  
217 mesenchymal transition. SC2 is enriched in extracellular matrix (ECM) organization, skeletal  
218 system development, and cartilage development. SC3 involvement in fibrous differentiation is  
219 supported by the enrichment of collagen fibril organization and elastic fiber formation.

220

221 Taken together, this analysis has revealed three major transcriptional states of GLE cells in  
222 synovial joint development. It has also hinted at substantial heterogeneity within SCs. To more  
223 clearly define the cell types and states of GLE cells, we next analyzed each SC separately, as  
224 described in the following sections.

225

## 226 **Two categories of early GLE cells: chondrogenic and mesenchymal**

227

228 By applying Leiden clustering to only SC1, we identified two sub-clusters: SC1\_A and SC1\_B  
229 (**Fig 3A**). SC1\_A has high expression levels of genes associated with chondrogenesis (e.g.  
230 Sox9 and Col2a1) and the IZ (e.g. Nog and Gdf5) (Ray et al. 2015; Hartmann and Tabin 2001;  
231 Storm and Kingsley 1996). SC1\_B exhibited high expression levels of genes associated with  
232 fibrous and mesenchymal cells such as Col3a1 and Col1a2 (Niederreither et al. 1992), as well  
233 Osr1, which is mainly expressed in the outer mesenchyme (**Fig 3B**) where it promotes fibroblast  
234 differentiation and inhibits chondrogenesis (Stricker et al. 2012). These results suggest that SC1  
235 is comprised of chondroprogenitors and early chondrocytes of the limb anlagen, nascent IZ  
236 cells, as well as the non-chondrogenic mesenchymal cells situated outside the anlagen. We  
237 tested and confirmed this conjecture using ISH for genes indicative of each cluster (**Fig 3C-D**).

238

239 To determine the lineage relationship between these clusters we performed RNA Velocity  
240 analysis (La Manno et al. 2018). Our results predicted that there is little-to-no transition between  
241 SC1\_A and SC1\_B (**Fig 3E**). To test this prediction, we prospectively isolated E12.5 YFP<sup>+</sup> cells

242 using antibodies specific for SC1\_A (CD9) or SC1\_B (PDGFRA), and measured lineage specific  
243 marker expression after culturing the cells *in vitro* for seven days. Cells from the PDGFRA<sup>+</sup>  
244 population exhibited a mesenchymal morphology, whereas cells from the CD9<sup>+</sup>/PDGFRA<sup>-</sup>  
245 population exhibited a chondrocyte-like morphology (**Fig 3F, left**). Consistent with their  
246 respective shapes and appearances, the PDGFRA<sup>+</sup> population yielded a substantially higher  
247 proportion cells positive for the tendon and ligament marker TNMD compared to the  
248 CD9<sup>+</sup>/PDGFRA<sup>-</sup> population, and a lower proportion of cells positive for the chondrogenesis  
249 regulator SOX9 as measured by immunofluorescence (**Fig 3F, right and Fig 3G**). While the  
250 CD9<sup>+</sup> population yielded more THY1-positive cells, neither group had a substantial fraction of  
251 positive cells. The fact that both populations were not mutually exclusive for TNMD and SOX9  
252 expression can be explained by incomplete lineage commitment, by the imperfect ability of  
253 PDGFRA to mark SC1\_A and of CD9 to mark SC1\_B, and by impurity in the FACS gating. With  
254 these caveats in mind, the data do support a model where the *in vitro* differentiation propensity  
255 of SC1\_A is towards a tenocyte/ligamentocyte fate, whereas the *in vitro* propensity of SC1\_B is  
256 towards a chondrocyte fate.

257

#### 258 **Diverse origins of nascent interzone**

259

260 While most SC1\_B cells expressed Sox9, we noticed that they were heterogeneous in terms of  
261 IZ- and chondrocyte-related genes, suggesting that this cluster consisted of sub-populations or  
262 sub-states. To examine this further, we clustered SC1\_B alone and identified four clusters:  
263 SC1\_B1 to SC1\_B4 (**Fig 4A**). SC1\_B4 was marked by high levels of Col2a1 and Matn1,  
264 indicating that it contained cells destined to become transient chondrocytes (Hyde et al. 2007)  
265 (**Fig 4B**). The three other clusters expressed both chondroprogenitor transcription factors (e.g.  
266 Sox5, Sox6, and Sox9), as well as the IZ marker Gdf5. These clusters varied in the extent to  
267 which they expressed other IZ-related genes: SC1\_B3 had high levels of Sfrp2, Vcan, and  
268 Trps1, whereas SC1\_B2 had the highest level of Ebf1, Jun respectively (**Fig 4B**) (Choocheep et  
269 al. 2010) (Norris et al. 2007; Choocheep et al. 2010; Kunath et al. 2002; Salva and Merrill 2017).

270

271 With the exception of SC1\_B4, we hypothesized that these clusters represented distinct stages  
272 of IZ formation. To explore this hypothesis, we performed RNA velocity analysis. Consistent with  
273 the notion that SC1\_B4 consists of transient chondrocytes, the trajectories of the other,  
274 presumptive IZ, clusters did not lead to it (**Fig 4C**). Rather, the only trajectories were an  
275 apparent convergence of IZ clusters SC1\_B1 and SC1\_B2 to at a singular IZ expression state



276 defined by high expression of IZ markers such as *Sfrp2* (Pazin et al. 2012) and *Vcan*  
277 (Choocheep et al. 2010) in SC1\_B3. To better understand the regulatory networks that  
278 contributed to this convergence, we subjected these clusters to Epoch analysis (manuscript in  
279 preparation). In brief, this tool takes as input pseudotime-ordered scRNA-Seq data. Then it  
280 identifies temporally regulated genes and periods of gene activity, reconstructs gene regulatory  
281 networks in a temporally sensitivity manner, and it proposes candidate regulators of transitions  
282 between expression states. To use this tool, we first ordered the cells along a pseudotemporal  
283 axis as defined by diffusion-based pseudotime (Haghverdi et al. 2016), with two roots, or  
284 starting points, selected based on the RNA velocity analysis. Then, we used Epoch to identify  
285 genes temporally regulated along each of these converging trajectories, or paths. Each path  
286 held three classes of genes: those with expression that peaked at early, in the middle, or later in  
287 the trajectory (**Fig 4D-E**). Path 1, which is defined by cells from SC1\_B1, starts with an  
288 expression of limb mesenchyme (high *Prrx2* (Leussink et al. 1995), *Zcchc12* (Li et al. 2009)),  
289 then expression of *Col1a2* and *Bmp2* peaks in the middle stage, and it ends in the high *Gdf5*  
290 and high *Sfrp2* state. Epoch predicted that the major regulators of the first stage are *Maf*, a  
291 known regulator of chondrocyte differentiation (MacLean et al. 2003), *Isl1* (Yang et al. 2006),  
292 *Ebf1* and *Sox4*, detected in IZ with unknown mechanism (Jenner et al. 2014; Bhattaram et al.  
293 2014) and *Lin28b*, an indicator of embryonic to adult transitioning (Zhang et al. 2016) (**Fig 4D**).

294  
295 The middle stage of Path 1 was predicted to be regulated by epithelial to mesenchymal  
296 transition regulator *Twist1* (Liu et al. 2017), and chondrogenic regulators *Klf2* (Cameron et al.  
297 2009) and *Ets2* (Karsenty and Wagner 2002). Other regulators included *Meis2*, which was  
298 previously reported as expressed in the knee IZ (Pazin et al. 2012) and *Nfib*, a homolog of *Nfia*  
299 which maintains the IZ domain (Singh et al. 2018).

300  
301 The later stage of Path1 is predicted to be regulated by *Sox9*, which is considered to decrease  
302 in expression during IZ formation (Soeda et al. 2010); IZ morphogenesis regulator: *Sox6* (Dy et  
303 al. 2010); *Tcf7l2*, which mediates crosstalk the between Hedgehog and Wnt signaling that  
304 promotes IZ differentiation (Rockel et al. 2016). Epoch also identified *Osr2*, *Barx2*, *Hoxd9*,  
305 *Wnt5a*, and *Trps1* as important contributors to the late stage of Path 1. Many of these factors  
306 have previously been reported to be associated with IZ: *Osr2* contributes to IZ expression of  
307 *Gdf5* (Gao et al. 2011); *Barx2* is upregulated in the presumptive IZ (Meech et al. 2005); *Hoxd9*  
308 regulates sesamoids formation from IZ (Khoa et al. 1999; Fromental-Ramain et al. 1996); *Wnt5a*  
309 was detected in digital IZ and is downregulated at cavitation (Church et al. 2002); *Trps1* acts

310 downstream of Gdf5 to promotes chondrogenesis (Itoh et al. 2008). Path 2, which is defined by  
311 cells from SC1\_B2, also starts with an expression state of mixed limb mesenchyme (high  
312 Prrx1,2 and Zcchc12), but a distinct set of IZ related TFs including Dlx5 (Ferrari and Koshier  
313 2006) and Hand2 (Askary et al. 2015) are involved in regulating early stage of transition. Many  
314 of the regulators and target genes of the middle and tertiary stages Path 2 are shared with Path  
315 1 (**Fig 4D-G**). For example, the middle stage of Path 2 is marked by peak expression of Col1a1,  
316 Nfia, and Basp1. Similarly, the final stages of both paths are marked by peak expression of  
317 Sox9, Mef2c, Tbx15. However, a notable difference in the paths is in the early stages where  
318 Path 1 is regulated by Irx3, Irx5 and Meis2, which are preferentially expressed in the proximal  
319 anterior portion of the developing limb (Li et al. 2014) (Capdevila et al. 1999). This suggests that  
320 the SC1\_B1 and SC1\_B2 start at distinct states reflecting remnants of spatial patterning of the  
321 condensing mesenchymal cells of the limb. However, as they differentiate, they leverage the  
322 same, or highly similar GRNs, to converge on the IZ state. Overall, the trajectories presented  
323 here indicate a process in which similar but separately patterned early chondroprogenitors  
324 follow parallel paths, with shared landmarks (e.g. an intermediate stage in which IZ regulators  
325 Sox9 peaks), before reaching an IZ-like state. All of the regulators predicted by Epoch are listed  
326 in **Supp Table 1, 2**.

327

## 328 **IZ formation**

329

330 Compared to SC1, many SC2 cells had high levels of more IZ-related genes such as Cd44 and  
331 Sfrp2; other SC2 cells exhibited more established chondrocyte profiles. To resolve this  
332 population heterogeneity, we performed clustering on SC2 and identified two groups (**Fig 5A**).  
333 SC2\_A was enriched in chondrocyte-related genes Col2a1, Col9a1, Sox9 (**Fig 5B**). We  
334 confirmed the expression co-localization of these genes in the anlagen by ISH (**Fig 5C**). SC2\_B,  
335 on the other hand, exhibited IZ features based on higher expression of Gdf5, Sfrp2, and Col3a1  
336 (**Fig 5B**). We confirmed the IZ localization of these genes by ISH (**Fig 5D**). To understand the  
337 lineage relationship between these clusters, we performed RNA Velocity analysis, finding that  
338 approximately half of the SC2\_A cells were transitioning to SC2\_B (**Fig 5E**), suggesting that  
339 GLE anlagen prechondrocytes continue to de-differentiate and contribute to the IZ.

340

341 The fact that a substantial fraction of the SC2\_A cells were not transitioning to IZ suggested  
342 population substructure. To explore this, we performed clustering on each of SC2\_A and  
343 SC2\_B, finding three and four subsets respectively. One of the SC2\_A sub-clusters, SC2\_A2,

344 consisting of about 20 cells, expressed *Ihh* and *Cd200*, suggestive of a pre-hypertrophic state  
345 (**Supp Fig 3A**). As these cells were not predicted to be related to any of the other clusters by  
346 RNA velocity, we excluded them from further analyses (**Fig 5F**). We found that *SC2\_A1* and  
347 *SC2\_B2* have similar expression patterns in chondrogenic genes *Col2a1*, *Col11a1*, *Sox9*,  
348 *Wwp2* (Zhao et al. 1997; Hyde et al. 2007; Akiyama and Lefebvre 2011) (**Fig 5G**). The other  
349 four clusters have lost expression of *Col2a1* and have upregulated expression of IZ-related  
350 genes including *Gdf5*, *Cd44* (Hartmann and Tabin 2001), *Sfrp2*, *Htra1* (Oka et al. 2004), and  
351 *Dkk3* (Witte et al. 2009) (**Fig 5G**). Repeating RNA velocity on these clusters recapitulated the  
352 results of a transition from chondrocyte to IZ state when applied to all SC2 cells (**Fig 5H**).

353  
354 To identify the GRN contributing to this transition, we applied Epoch to the group of cells that  
355 exhibited a concerted velocity from chondrogenic to IZ-like, as defined and ordered by diffusion  
356 based pseudotime (**Fig 5H,I,J**). Epoch analysis revealed different early regulators as compared  
357 to the programs identified in IZ initiation at E12.5 (**Fig 5K, Supp Table 3**). Here, in the early part  
358 of the path, some of the IZ progenitors appear to be outer IZ based on the higher level of  
359 expression of outer IZ related genes: *3110079O15Rik*, *Matn1*, *Susd5*, *Matn3*, *Fgfr3* and co-  
360 expression of *Gdf5* (**Supp Fig 3B**). Epoch predicted that the major regulators of the first stage  
361 included the known IZ regulator *Erg* (Iwamoto et al. 2007); *Lef1*, an effector of canonical Wnt  
362 signaling with multiple roles in early IZ specification (Guo et al. 2004); *Klf8* (Wang et al. 2011)  
363 and *Snail1*, two epithelial to mesenchymal transition (EMT) inducers, indicating a shared  
364 regulatory program between chondrocyte-to-mesenchymal transition and EMT (Vincent et al.  
365 2009; Lin et al. 2014) (**Fig 5L**). The middle stage is marked by upregulation of both fibrogenic  
366 genes such as *Fgfr2*, *Ddit3* (Caterson and Melrose 2018) and chondrogenic genes such as  
367 *Isl1* (Yang et al. 2006), *Tnc* (Grogan et al. 2013). This suggests that the middle stage is a  
368 transition state in which cells exhibit properties of both chondrocyte and IZ cell. *Isl1* and *Ddit3*  
369 also act as the predicted regulators for middle stage, as does *Klf4*, which promotes the  
370 expression of *Col1a1* that is required for IZ morphogenesis (Orgeur et al. 2018). Interestingly,  
371 the mesenchyme markers *Prrx2*, *Prrx1*, *Col3a1*, *Col1a1* expression turn back to the peak level  
372 at late stage, indicating there is a dedifferentiation of chondrocytes to mesenchymal cells in IZ  
373 development. The major regulators of the late stage include: *Scx*, an inducer of ligament/tendon  
374 differentiation (Anderson et al. 2006) and *Meox2*, which is contributes to tendon and soft  
375 connective tissue development (Acharya and Amit n.d.). Other regulators that have  
376 underexplored roles in IZ formation identified include *Pitx1*, *Meox1*, *Deaf1*, *Tbx5*, *Jund*, and  
377 *Zeb1*. *Def1* has been reported to bind to *Gdf5* and have a repressive effect on *Gdf5* expression

378 (Syddall et al. 2013). Tbx5 interacts with Fgf and Wnt in the limb bud to modulate limb and joint  
379 morphogenesis (Agarwal et al. 2003) (Rallis et al. 2003). Jund, along with Fos, forms a complex  
380 that directly regulates Wnt activity in the IZ (Kan and Tabin 2013). Zeb1 modulates TGF $\beta$   
381 signaling, and when mutated leads to multiple joint fusions (Takagi et al. 1998). In summary, we  
382 have found that IZ formation is characterized by continuous chondrocyte-to-mesenchymal  
383 process that includes cells of the anlagen. Our analysis has revealed many previously  
384 implicated regulators of this process, as well as many novel candidate genes.

385

### 386 **Development of articular fibrous components**

387

388 SC3, characterized by fibroblast-related genes and pathways, is distinct from the IZ-related  
389 SC2. As we did for SC1 and SC2, we performed a deeper analysis of SC3 by clustering it more  
390 finely into two sub-clusters. SC3\_A was mainly comprised by E13.5 cells and had high levels of  
391 cell growth related genes whereas SC3\_B was made up of E14.5 and E15.5 cells and had  
392 higher levels of fibroblast ECM related genes such as Postn and Col3a1 (**Fig 6A-B**). We  
393 confirmed the preferential expression of Col3a1 and Postn in the ligament, tendon and menisci  
394 by ISH (**Fig 5D** and **Fig 6C**). However, it was not clear whether intra-articular fibrous  
395 components (especially at E15.5) belonged to SC2 or SC3 based on the expression pattern of  
396 differential genes Postn, Col3a1, and Col2a1 (**Suppl Fig4A-E**). Nevertheless, the presence of  
397 Gdf5 at E15.5 articular surface and predominant expression in SC2 indicate intra-articular  
398 ligament cells were included in SC3 (**Suppl Fig4F**). Our data and analysis suggest that the Scx  
399 expressing cells of SC2 give rise to fibrochondrocytes, which contribute to the transitioning zone  
400 of articular cartilage, intra-articular ligament, and meniscus (**Suppl Fig4G**). In addition to Col3a1  
401 and Col1a1, expression of Dcn and Tnmd were better able to differentiate SC3 from SC2  
402 (**Suppl Fig4H,I**). Thus, Dcn<sup>+</sup>Tnmd<sup>+</sup>Scx<sup>+</sup> SC3 refers to fibroblasts that contribute to fibrous tissue  
403 of joint.

404

405 As SC3\_A and SC3\_B differed mainly by developmental stage, we sub-clustered each to  
406 search for more subtle differences in state or lineage, resulting in seven sub-groups (**Fig 6D**).  
407 We then annotated the likely cell type of each sub-cluster based on differential gene expression  
408 and enrichment analysis, as described below. SC3\_B2 is likely to represent myotendinous  
409 junction site cells, as it has high levels of Tbx3, which is required for muscle attachment  
410 (Colasanto et al. 2016), in addition to other muscle-related genes including Lsp1, Cygb, and  
411 Moxd1 (Singh et al. 2014) (**Fig 6E**). Cells of SC3\_B3 are likely to be tendon/ligament cells

412 based on preferential expression of *Tnmd* and *Scx* (Sugimoto et al. 2013; Soeda et al. 2010)  
413 (Subramanian and Schilling 2015), as well as other tendon associated genes including *Thbs4*  
414 (Havis et al. 2014) (Subramanian and Schilling 2014), *Htra1* (Oka et al. 2004), *Cilp* (Caterson  
415 and Melrose 2018), *Meox2* (Havis et al. 2014), *Abi3bp* (Zhang et al. 2014), and *Fmod* (Bi et al.  
416 2007). *SC3\_B4* is likely to include synovial fibroblasts and fibrocartilage cells of the enthesis  
417 (Zelzer et al. 2014) based on the preferential expression of *Cthrc1* and *Tsp2*, both of which are  
418 produced by synovial fibroblasts (Shekhani et al. 2016; Park et al. 2004) and chondrocyte-  
419 related genes *Aspn*, *1500015O10Rik*, *Mia*, and *Dlx5* (Ferrari and Kosher 2006). We hypothesize  
420 that *SC3\_B1* is comprised of synovial lining cells based on the enrichment of MAPK, IL-17, TNF  
421 signaling pathways, in contrast to the other *SC3\_B* clusters, which were enriched in ECM-  
422 receptor interaction, Focal adhesion, and the PI3K-Akt signaling pathway (**Fig 6F**). Taken  
423 together, our data suggest that *SC3\_B* subclusters represent cells of the myotendinous junction  
424 site (B2), tendon/ligament (B3), fibrocartilage cells of the synovium and enthesis (B4), and the  
425 synovial membrane (B1).

426

427 Next, we applied RNA velocity to infer lineage relationships among *SC3* cells. This analysis  
428 detected velocity primarily between *SC3\_A3* and the tendon/ligament cluster *SC3\_B3* (**Fig 6G-**  
429 **I**). This result was consistent with the lineage annotation of *SC\_B* subgroups we proposed  
430 above because there is little-to-no trajectory between the *SC3\_B* sub-clusters. To perform  
431 Epoch analysis and reconstruct the GRN that contributes to this progenitor-to-  
432 tenocyte/ligamentocyte transition, we first performed diffusion based pseudotime analysis on  
433 *SC3\_A3* and *SC3\_B3* (**Fig 6H-J**). Our data and analysis are consistent with prior studies which  
434 reported that tendon/ligament progenitors lose *Sox9* concomitant with *Scx* upregulation  
435 (Sugimoto et al. 2013; Soeda et al. 2010) and followed by *Tnmd* upregulation (Subramanian  
436 and Schilling 2015) (**Suppl Fig 4J**). Next, we used Epoch to identify genes temporally regulated  
437 along this pseudotime axis (**Fig 6I, Supp Table 4**). At the early stage, many chondrogenesis  
438 and IZ related genes were high including *Wwp2*, *Pitx1* (Wang et al. 2018), *Chadl*, *Col2a1*,  
439 *Col9a3*, *Dlk1* (Chen et al. 2011), *Wnt4*, *Vcan*, *Sox9*, and *Sfrp2*. This suggested that the  
440 development of articular fibrous components, especially tendon/ligament starts from a  
441 progenitor population with some chondrocyte features. By the later stage, the mature tendon  
442 markers *Aqp1*, *Tnmd*, and the fibroblast ECM genes *Col1a1*, *Col3a1*, *Aspn* were upregulated  
443 (**Fig 6J**). Epoch analysis predicted that the early stage was regulated by *Barx1*, which has an  
444 inhibitory effect on chondrogenic initiation during joint development (Church et al. 2005). Other  
445 predicted regulators included *Etv4*, which is detected in muscle-tendon interface with high

446 expression level and regulated by FGF signaling (Havis et al. 2016) and Hdac1, which was  
447 recently found to inhibit Scx expression in tendon progenitor cells (Zhang et al. 2018).  
448 Regulators of the middle stage included those previously associated with tendon development  
449 (e.g. Ebf1 --expressed in presumptive tendons surrounding chondrogenic condensation (Mella  
450 et al. 2004)) and other factors that have not previously been implicated in this process such as  
451 Dlx5, Dlx6, which are expressed in presumptive elbow joint and involved in osteogenesis  
452 (Ferrari and Kosher 2006; Lee et al. 2003). The predicted regulators of the final stage included  
453 TFs associated with inflammatory response: Cebpd, Egr1; osteogenesis: Sp7, Cbfb (Lien et al.  
454 2007), and tendon development: Klf10 (McConnell and Yang 2010), Klf2, Klf4, Aebp1  
455 (Blackburn et al. 2018), Ddit3 (Caterson and Melrose 2018), and Bhlhe40 (Peffer et al. 2015)  
456 (**Fig 6K**). In summary, the development of fibrous components of the synovial joint, particularly  
457 tendon/ligament, is characterized by the ordered loss of chondrogenic gene expression  
458 programs followed by the upregulation of tendon/ligament expression programs. Moreover, we  
459 found that the cells of different fibrous components can be distinguished by their transcriptional  
460 signatures.

461

## 462 **Nascent joint development**

463

464 To better understand the potential lineage relationships of the superclusters, we applied RNA  
465 velocity to all GLE cells. We found that some chondroprogenitor SC1 cells were predicted to  
466 give rise to SC2\_A1, and that the  $Osr1^+Col3a1^+$  SC1\_A cells were predicted to differentiate to  
467 SC3, consistent with their chondrogenic or mesenchymal features, respectively (**Fig 7A**).  
468 Synthesizing these results with prior analyses yielded the following summary of our data. Early  
469 GLE cells contained a  $CD9^+$  chondrogenic population and a  $PDGFRA^+$  mesenchymal population  
470 (**Fig 7B**). The chondrogenic progenitors gave rise to the IZ, which is comprised of  
471  $Col2a1^+Sox9^+Col9a1^+Gdf5^{Low}$  cells (SC2\_A1, SC2\_B2) and  $Sfrp2^+Col3a1^+Gdf5^{high}$  cells  
472 (SC2\_A3, SC2\_B1), which are likely to correspond to the outer and intermediate IZ,  
473 respectively. Our data supports the notion that some outer IZ serves as precursor for  
474 intermediate IZ. In addition, newly recruited Gdf5-expressing IZ cells with enrichment in  $Sfrp2$ ,  
475  $Htra1$ ,  $Dkk3$  (SC2\_B3, B4) appear to develop to either more mature IZ cells or to fibrous cells of  
476 SC3. On the other hand, the mesenchymal progenitors of SC1\_A differentiate to  
477  $Col3a1^+Postn^+Dcn^+Tnmd^+$  fibrous component cells, including ligament, tendon and synovium  
478 (SC3). Intriguingly, a group of  $Scx^+Meox2^+Meox1^+Tbx5^-$  SC2 cells (SC2\_A3, SC2\_B1) was  
479 predicted to transit to SC3 (**Fig 7A**), suggesting that some of the fibrous components are

480 specified from multiple origins, in this case from both the early SC1\_A and the later, IZ SC2\_B  
481 sub-cluster.

482

## 483 **Discussion**

484

485 The synovial joint initiates from a thin layer of mesenchymal cells marked by Gdf5 expression.  
486 Through lineage tracing of Gdf5, it has become apparent that Gdf5-expressing IZ cells give rise  
487 to multiple joint lineages. However, the transcriptional programs that drive IZ formation and  
488 elaboration has remained underexplored. In this study, we applied scRNA-Seq to Gdf5 lineage  
489 cells during embryonic stages of synovial joint development to define the continuum of  
490 expression states that govern the process from interzone formation to joint cavitation. Here we  
491 have revealed the dynamic transcriptome changes and heterogeneity in GLE cells, we have  
492 inferred the lineage trajectories of subpopulations, and we have predicted the regulators of key  
493 developmental decisions. Several insights have emerged from our dataset and analyses that  
494 have implications for the field.

495

496 First, our results have revealed that GLE cells at E12.5 already exhibited a transcriptional  
497 heterogeneity, with one cluster tending towards a more mesenchymal state (SC1\_A) and one  
498 cluster tending towards a more chondroprogenitor state (SC1\_B). By prospectively isolating  
499 cells using markers that distinguished these clusters, we confirmed the *in vitro* lineage  
500 propensity of these cell populations. The degree of commitment of these cells *in vivo* remains to  
501 be determined. Second, we discovered further sub-structure within the SC1\_B  
502 chondroprogenitors: one cluster (SC1\_B1) preferentially expressed more proximal Meis2 and  
503 proximal/anterior Irx3, and one cluster (SC1\_B2) preferentially expressed more distal (or  
504 zeugopod-associated) Hoxa11os. By RNA Velocity analysis, we predicted that both of these  
505 clusters were transitioning to a pre-IZ state marked by expression of Sfrp2, Vcan, Trps1, and  
506 Snail1, and our Epoch analysis revealed that the gene regulatory networks associated with  
507 these transitions were highly similar. This raises notion that the later complex architecture of the  
508 IZ and its derivatives are presaged by limb spatial patterning. Third, we found that between  
509 E13.5-E15.5 there is a continual transition of chondroprogenitors to an IZ state that is  
510 reminiscent of the pre-anlagen limb bud mesenchyme as exemplified by up-regulation of Prrx1.  
511 Fourth, our data support the idea that the fibrous joint components have dual origins. The  
512 Osr1<sup>+</sup>Col3a1<sup>+</sup> mesenchymal progenitors detected in E12.5 (SC1\_A) were predicted to transition  
513 to fibrous components of the tendon and fibrochondrocytes of the synovium, whereas cells of

514 the putative intermediate IZ (SC2\_ B3 and SC2\_ B4) were predicted to transition to Tnmd<sup>+</sup>Scx<sup>+</sup>  
515 cells of the ligament/tendon cells (SC3\_ B3) (**Fig 7**).

516

517 We have made this data freely and easily accessible with a web application at  
518 [http://www.cahanlab.org/resources/joint\\_ontogeny](http://www.cahanlab.org/resources/joint_ontogeny). We believe that this resource will aid the  
519 community in discovering additional transcriptional programs and in inferring cell interactions  
520 that underpin synovial joint development. Further, we anticipate that this data can be used to  
521 yield improved protocols for the derivation of synovial lineages from pluripotent stem  
522 cells (Oldershaw et al. 2010; Craft et al. 2015; Kawata et al. 2019; Yamashita et al. 2015) by, for  
523 example, using it to identify candidate signaling pathways or by using the expression data as a  
524 reference against which to compare engineered cells.

525

526

## 527 **STAR Methods**

528

### 529 **Mice**

530 *Gdf5-cre* (Sperm Cryorecovery via IVF, FVB/NJ background) mouse strain was obtained from  
531 the Jackson laboratory. B6.129X1-Gt(ROSA)26Sortm1(EYFP)Cos/J (RosaEYFP) was gifted by  
532 the lab of Prof. Xu Cao (Johns Hopkins University). *Gdf5-cre::Rosa-EYFP* mice were generated  
533 by crossing heterozygote *Gdf5-cre* strain with homozygote *RosaEYFP* strain. The genotype of  
534 the mice was determined by PCR analyses of genomic DNA isolated from mouse tails using the  
535 following primers: *Gdf5*-directed *cre* forward, 5'GCCTGCATTACCGGTCGATGCAACGA3', and  
536 reverse, 5'GTGGCAGATGGCGCGGCAACACCATT3' (protocol provided by Prof. David  
537 Kingsley, HHMI and Stanford University). Day 5 wild type refers to C57/BL10 mouse. All the  
538 protocols were approved by the institutional review board of Johns Hopkins University.

539

### 540 **Mice gender identification**

541 We identified mouse gender by genotyping *Sry* Y gene using the primers: forward,  
542 5'CTGGAAATCTACTGTGGTCTG3', and reverse, 5'ACCAAGACCAGAGTTTCCAG3'.

543

### 544 **Cell isolation**

545 Mice were kept in light-reversed room (light turns on at 10 pm and turns off at 10 am). Timing  
546 was determined by putting one male mouse and two female mice in the same cage at 9 am and  
547 separating them at 4 pm on the same day. We count that midnight as E0.5 stage. On E12.5,



548 E13.5, E14.5 and E15.5, the pregnant mice were sacrificed by CO<sub>2</sub> at 3 pm. The cells were  
549 isolated using the protocol (Primary culture and phenotyping of murine Chondrocytes) with  
550 modification: The embryos (usually n=6-8) were rinsed three times in PBS on ice. Two  
551 presumptive joint part from hind limb between presumptive ankle and hip of each individual  
552 embryo were disassociated in a single 3 cm dish (Figure 1A) and incubated in digestion solution  
553 I (3 mg / mL collagenase D, DMEM high glucose culture medium, serum free) for 45 min at 37  
554 °C, and then in digestion solution II (1 mg / mL collagenase D, DMEM high glucose culture  
555 medium, serum free) for 3 hrs (one embryo per dish) at 37 °C. During the period of incubation,  
556 the mice gender was identified by genotyping and only male samples were chosen for further  
557 processing. The tissues with medium were gently pipetted to disperse cell aggregates and  
558 filtered through 40 μm cell strainer, then centrifuged for 10 min at 400 g. The pellet was  
559 suspended with 0.4% BSA in PBS.

560

### 561 **Cell fractionation**

562 All cells were fractionated by fluorescence-activated cell sorting (FACS). A MoFlo XDP sorter  
563 (Beckman Coulter, Miami, FL, USA) was used to collect YFP<sup>+</sup> cells, and Propidium iodide was  
564 used to exclude dead cells.

565

### 566 **Single cell RNA sequencing**

567 GemCode™ Single Cell platform (10X Genomics) was used to determine the transcriptomes of  
568 single cells (Zheng et al. 2017). Cells at 1000 / μl were obtained after sorting and placed on ice.  
569 Each time point, one sample was selected and profiled based on the viability and amount. A  
570 total of 6000 cells were loaded each time, followed by GEM-RT reaction, and cDNA  
571 amplification. Single cell libraries were constructed by attaching P7 and P5 primer sites and  
572 sample index to the cDNA. Single cell RNA sequencing was performed on Illumina NextSeq 500  
573 and HiSeq 2500 to a depth ranging from 347 to 489 million reads per sample.

574

### 575 **Analysis and visualization of scRNA seq data**

576 CellRanger (version 2.0.0) was used to perform the original processing of single cell sequencing  
577 reads, aligning them to the mm10 reference genome. We used the command line interface of  
578 Velocity, version 1.7.3, to count reads and attribute them as spliced, un-spliced, or ambiguous  
579 (La Manno et al. 2018). The resulting loom files for each sample were then concatenated and  
580 subjected to quality control processing, normalization, estimation of cell cycle phase, clustering,  
581 and differential gene expression analysis using Scanpy 1.4.3 (Wolf et al. 2018). Specifically, we

582 excluded cells in which mitochondrial gene content exceeded 5% of the total reads or cells in  
583 with fewer than 501 unique genes detected. Then, we excluded genes that were detected in  
584 fewer than 10 cells, resulting in a data set of 10,124 cells and 16,352 genes. Then, we  
585 performed an initial normalization on a per cell basis followed by log transformation, and scaling.  
586 We scored the phases of cell cycle using cell cycle-associated genes as previously described  
587 (Satija et al. 2015). Then we identified the genes that were most variably expressed across the  
588 whole data set, and within each timepoint, resulting in 3,593 genes. We performed PCA and  
589 inspected the variance ratio plots to determine the ‘elbow’, or number of PCs that account for  
590 most of the total variation in the data. We generated a graph of cell neighbors using diffusion  
591 maps (Coifman et al. 2005), and then we performed Leiden clustering (Traag et al. 2019), which  
592 we visualized with a UMAP embedding (McInnes and Healy 2018) . We also analyzed this with  
593 SingleCellNet (Tan and Cahan 2019), which had been trained using the Tabula Muris data set  
594 (Tabula Muris Consortium et al. 2018). We removed cells in clusters that were classified by  
595 SingleCellNet as ‘blood’, ‘erythroblast’, ‘endothelial’. We also removed cells in clusters that we  
596 identified as likely to be myoblast based on high levels of Myod1 and other muscle-specific  
597 genes, melanocyte (based on Pmel expression), and neural crest (based on Sox10 expression).  
598 Then, we repeated the pre-processing and analysis pipeline on the remaining 8,378 genes. We  
599 noted that two clusters, primarily from E12.5 and E13.5, were predicted to be in G2M phase; we  
600 removed these cells from further analysis. Finally, we removed cells in a cluster that we  
601 determined by ISH to consist mainly of dermis cells, resulting in final data set of 6,202 cells and  
602 16,352 genes. Super-clusters and all sub-clusters were identified by following the same pipeline  
603 as described above, except that the analysis was limited to the corresponding set of cells. For  
604 example, the superclusters were identified by first finding the genes that vary across both all  
605 cells, and within each time point. Then, a neighborhood graph was determined using the  
606 principal components (the number of which was decided by examining the variation ratio plot),  
607 followed by Leiden clustering, and visualized by UMAP embedding, and, for some subsets of  
608 data, diffusion map embedding. Differentially expressed genes were identified using the Scanpy  
609 rank\_genes\_groups function. Gene set enrichment analysis was performed using GSEAPY  
610 (<https://github.com/zqfang/GSEAPy>), a Python interface to enrichR (Chen et al. 2013; Kuleshov  
611 et al. 2016). The analysis pipeline of Velocity was applied to data subsets as mentioned in the  
612 main text. We used the Velocity results to manually assign roots for diffusion map pseudotime  
613 analysis. The results of pseudotime were imported into Epoch for gene regulatory network  
614 reconstruction (manuscript in preparation).

615

616 **Histochemistry, immunohistochemistry, and histomorphometry**

617 The specimens were fixed in 10% buffered formalin for 6-24 hrs at RT. D5 joints were  
618 decalcified in 10% ethylenediaminetetraacetic acid (EDTA) in PBS (pH 7.4) for 3 days at 4°C,  
619 washed with distilled water and equilibrated in 30% sucrose in PBS at 4°C overnight, then  
620 mounted in O.C.T and frozen at -80°C. Ten-micrometer-thick coronal-oriented or sagittal-  
621 oriented sections were performed by cryostat.

622 We performed Trichrome staining according to Trichrome Stain (Connective Tissue Stain) Kit  
623 protocol.

624 Immunostaining was performed using a standard protocol. Sections were incubated with  
625 primary antibodies to mouse GFP (1:200), TNMD (1:100), SOX9 (1:500), THY1 (1:100) in  
626 Antibody Diluent, at 4°C overnight followed with three 5 min washes in TBST. The slides were  
627 then incubated with secondary antibodies conjugated with fluorescence at room temperature for  
628 1 h while avoiding light followed with three 5 min washes in TBST and nuclear stained with  
629 mounting medium containing DAPI. Images were captured by Nikon ECLipse Ti-S, DS-U3 and  
630 DS-Qi2. See **Suppl Table 5**.

631

632 ***In situ* hybridization**

633 See **Suppl Table 6** for the information of oligonucleotides used for templates for antisense RNA  
634 probes. The T7 and SP6 primer sequence were added to 5- and 3- prime end, respectively. SP6  
635 RNA polymerase was used for probe transcription. Probes were synthesized with digoxigenin-  
636 labeled UTP and hybridized at 68°C overnight. Results were visualized by Alkaline  
637 phosphatase-conjugated anti-digoxigenin antibody and BCIP/NBT substrates.

638

639 **FACS for prospective isolation**

640 E12.5 embryonic hind limb cells or Day 5 knee joint cells were isolated as described in Cell  
641 isolation. After filtered through 40 μm cell strainer, cells were suspended in autoMACS rinsing  
642 solution at 1 million per mL. After spin down, E12.5 cells were then stained with PDGFRA (1 μg  
643 per 10 million cells) and CD9 (1 μg per 5 million cells) in 100 uL autoMACS rinsing solution in  
644 dark for 30 min followed by two times washes with autoMACS rinsing solution. Cells were re-  
645 suspended in autoMACS rinsing solution. A negative control without staining was used to setup  
646 gate. The following two E12.5 populations were collected at the same time: YFP<sup>+</sup>PDGFRA<sup>+</sup>  
647 population, YFP<sup>+</sup>PDGFRA<sup>-</sup>CD9<sup>+</sup> population. Day 5 four populations were collected based on  
648 four evenly distributed cell samples.

649

650 **Supplemental Table 1**

651 TF score for SC1\_B Path 1. TF: Transcription factor; epoch: 1=Early stage; 2=Middle stage;  
652 3=Later stage; weightMean: mean association strength of TF and targets genes; ntargets:  
653 Predicted number of target genes; peakTime: the pseudotime at which gene expression is  
654 highest.

655 **Supplemental Table 2**

656 TF score for SC1\_B Path 2.

657 **Supplemental Table 3**

658 TF score for SC2 Path.

659 **Supplemental Table 4**

660 TF score for SC3\_A3 to B3 Path.

661

662

663

664

665

666

667

668

**Figure legends**

669 **Figure 1:** Top: localization of Gdf5-lineage cells in murine hindlimb. Bottom: Cell density and  
670 morphology during joint formation as shown by Trichrome staining. Scale bar = 100.

671

672 **Figure 2: scRNA-Seq of Gdf5-lineage enriched cells during knee development.** Leiden  
673 clustering and UMAP embedding of the five distinct superclusters of GLE cells (A). The  
674 proportion of cells from each timepoint varies across superclusters (B). Expression of genes  
675 well-characterized in limb and joint development (C). Size of each dot reflects the percent of  
676 cells in which the gene is detected within the supercluster. The color indicates mean expression,  
677 including cells in which there is no detectable expression. (D) Supercluster gene set enrichment  
678 analysis, showing selected categories. Complete results are in Supplemental Table 1.

679

680 **Figure 3: SC1 is composed of chondrogenic and mesenchymal fated cells.** (A) Leiden  
681 clustering and diffusion map embedding SC1. (B) Dot plot expression of representative genes  
682 differentially expressed between SC1\_A and SC1\_B. (C) ISH detection for SC1\_A and SC1\_B  
683 representative genes. (E) RNA Velocity analysis. Arrows indicate the predicted future state of  
684 SC1 cells, showing a minimal transition between SC1\_A and SC1\_B. (F) In vitro culture of  
685 YFP<sup>+</sup>/Pdgfra<sup>+</sup> and YFP<sup>+</sup>/Cd9<sup>+</sup> hindlimb cells from e12.5 embryos shows distinct morphology of  
686 the cells (left). Immunofluorescence staining of tendon and ligament marker TNMD, fibroblast  
687 marker THY1, and chondrocyte regulator SOX9 (right). (G) Quantification of the proportion of  
688 cells positive for each marker.

689

690 **Figure 4: Two SC1\_B sub-populations converge to a common interzone-like state.** (A)  
691 Leiden clustering and diffusion map embedding of SC1B. (B) Dot plot of expression of  
692 representative genes differentially expressed between SC1B sub-clusters. (C) RNA velocity  
693 indicates converging trajectories of SC1\_ B1 and SC1\_ B2. Epoch analysis identifies three  
694 Epochs of gene expression in Path1(D) and Path 2 (E). Selected genes from each Epoch are  
695 listed on right, and Epoch identified regulators are in bold. (F-G) Minimal spanning tree  
696 representation of Epoch-reconstructed gene regulatory network.

697

698 **Figure 5: SC2 is composed of interzone fated cells.** (A) Leiden clustering and diffusion map  
699 embedding SC2. (B) Dot plot expression of representative genes differentially expressed  
700 between SC2\_A and SC2\_B. (C,D) ISH detection for SC2\_A and SC2\_B representative genes.  
701 (E) RNA Velocity analysis. Arrows indicate the predicted future state of SC2 cells, showing a  
702 minimal transition between SC2\_A and SC2\_B.

703

704 **Figure 5 (continued): IZ formation is a chondrocyte to mesenchymal cell transition**  
705 **process.** (F) Leiden clustering and diffusion map embedding of SC2, colored by subgroups or  
706 timepoints. (G) Dot plot of expression of representative genes differentially expressed between  
707 SC2 sub-clusters. (H) RNA velocity indicates a developmental path connecting sub-clusters. (I,  
708 J) Leiden clustering of SC2, colored by pseudotime (I) and groups (J). (L) Epoch analysis  
709 identifies three Epochs of gene expression based on group1 (orange) shown in (J). Selected  
710 genes from each Epoch are listed on right, and Epoch identified regulators are in bold. (M)  
711 Minimal spanning tree representation of Epoch-reconstructed gene regulatory network.

712

713 **Figure 6: SC3 is composed of articular fibrous component cells.** (A) Leiden clustering and  
714 diffusion map embedding SC3. (B) Dot plot expression of representative genes differentially  
715 expressed between SC3\_A and SC3\_B. (C) ISH detection for SC3\_A representative genes. (D)  
716 Sub-clustering of SC3 by Leiden. (E) Dot plot expression of representative genes differentially  
717 expressed among 7 sub-clusters. (F) Enrichment analysis of SC3\_B sub-clusters. (G) RNA  
718 Velocity analysis. Arrows indicate the predicted future state of SC3 cells, showing a minimal  
719 transition between SC3\_A3 and SC3\_B3. (H-I) Leiden clustering and diffusion map embedding  
720 SC3\_A3 and SC3\_B3, colored by groups (H) and pseudotime (I). (I) Epoch analysis identifies  
721 three Epochs of gene expression based on SC3\_A3 and SC3\_B3 populations. Selected genes  
722 from each Epoch are listed on right, and Epoch identified regulators are in bold. (J) Minimal  
723 spanning tree representation of Epoch-reconstructed gene regulatory network.

724

725 **Figure 7: Nascent joint development.** (A) RNA Velocity of 3 of SCs. (B) Cartoon of nascent  
726 joint development.

727

728 **Tables 1: Statistics on cells collected for scRNA-Seq.** 'Cells captured' was determined by  
729 10X Cell Ranger. GLE cells indicate the number of cells remaining after excluding cells unlikely  
730 to be GDF5-lineage, including immune cells, neural crest cells, and endothelial cells.

731

### 732 **Acknowledgements**

733 This work was supported by the National Institutes of Health under grant R35GM124725 to PC  
734 and by the Maryland Stem Cell Research Fund 2017-MSCRFF-3910 (Award ID: 90074850) to  
735 Qin Bian. Jordan Wilson was supported by NIH R25GM109441. This work was made possible  
736 by support from the Johns Hopkins Medicine Discovery Fund.

737

738

### 739 **References**

740

741 Acharya and Amit, N. Meox2 is necessary for axial and appendicular tendon development.  
742 Agarwal, P., Wylie, J.N., Galceran, J., et al. 2003. Tbx5 is essential for forelimb bud initiation  
743 following patterning of the limb field in the mouse embryo. *Development* 130(3), pp. 623–633.  
744 Akiyama, H. and Lefebvre, V. 2011. Unraveling the transcriptional regulatory machinery in  
745 chondrogenesis. *Journal of Bone and Mineral Metabolism* 29(4), pp. 390–395.  
746 Anderson, D.M., Arredondo, J., Hahn, K., et al. 2006. Mohawk is a novel homeobox gene  
747 expressed in the developing mouse embryo. *Developmental Dynamics* 235(3), pp. 792–801.  
748 Asahara, H., Inui, M. and Lotz, M.K. 2017. Tendons and ligaments: connecting developmental  
749 biology to musculoskeletal disease pathogenesis. *Journal of Bone and Mineral Research* 32(9),  
750 pp. 1773–1782.  
751 Askary, A., Mork, L., Paul, S., et al. 2015. Iroquois proteins promote skeletal joint formation by  
752 maintaining chondrocytes in an immature state. *Developmental Cell* 35(3), pp. 358–365.  
753 Bhattaram, P., Penzo-Méndez, A., Kato, K., et al. 2014. SOXC proteins amplify canonical WNT  
754 signaling to secure nonchondrocytic fates in skeletogenesis. *The Journal of Cell Biology* 207(5),  
755 pp. 657–671.  
756 Bi, Y., Ehrchiou, D., Kilts, T.M., et al. 2007. Identification of tendon stem/progenitor cells and  
757 the role of the extracellular matrix in their niche. *Nature Medicine* 13(10), pp. 1219–1227.  
758 Blackburn, P.R., Xu, Z., Tumelty, K.E., et al. 2018. Bi-allelic Alterations in AEBP1 Lead to  
759 Defective Collagen Assembly and Connective Tissue Structure Resulting in a Variant of Ehlers-  
760 Danlos Syndrome. *American Journal of Human Genetics* 102(4), pp. 696–705.  
761 Bobick, B.E. and Cobb, J. 2012. Shox2 regulates progression through chondrogenesis in the  
762 mouse proximal limb. *Journal of Cell Science* 125(Pt 24), pp. 6071–6083.  
763 Cameron, T.L., Belluoccio, D., Farlie, P.G., Brachvogel, B. and Bateman, J.F. 2009. Global  
764 comparative transcriptome analysis of cartilage formation in vivo. *BMC Developmental Biology*  
765 9, p. 20.

- 766 Capdevila, J., Tsukui, T., Rodríguez Esteban, C., Zappavigna, V. and Izpisua Belmonte, J.C.  
767 1999. Control of vertebrate limb outgrowth by the proximal factor Meis2 and distal antagonism  
768 of BMPs by Gremlin. *Molecular Cell* 4(5), pp. 839–849.
- 769 Catterson, B. and Melrose, J. 2018. Keratan sulfate, a complex glycosaminoglycan with unique  
770 functional capability. *Glycobiology* 28(4), pp. 182–206.
- 771 Chen, E.Y., Tan, C.M., Kou, Y., et al. 2013. Enrichr: interactive and collaborative HTML5 gene  
772 list enrichment analysis tool. *BMC Bioinformatics* 14, p. 128.
- 773 Chen, H., Capellini, T.D., Schoor, M., Mortlock, D.P., Reddi, A.H. and Kingsley, D.M. 2016.  
774 Heads, shoulders, elbows, knees, and toes: modular gdf5 enhancers control different joints in the  
775 vertebrate skeleton. *PLoS Genetics* 12(11), p. e1006454.
- 776 Chen, L., Qanie, D., Jafari, A., et al. 2011. Delta-like 1/fetal antigen-1 (Dlk1/FA1) is a novel  
777 regulator of chondrogenic cell differentiation via inhibition of the Akt kinase-dependent  
778 pathway. *The Journal of Biological Chemistry* 286(37), pp. 32140–32149.
- 779 Choocheep, K., Hatano, S., Takagi, H., et al. 2010. Versican facilitates chondrocyte  
780 differentiation and regulates joint morphogenesis. *The Journal of Biological Chemistry* 285(27),  
781 pp. 21114–21125.
- 782 Church, V., Nohno, T., Linker, C., Marcelle, C. and Francis-West, P. 2002. Wnt regulation of  
783 chondrocyte differentiation. *Journal of Cell Science* 115(Pt 24), pp. 4809–4818.
- 784 Church, V., Yamaguchi, K., Tsang, P., Akita, K., Logan, C. and Francis-West, P. 2005.  
785 Expression and function of Bapx1 during chick limb development. *Anatomy and embryology*  
786 209(6), pp. 461–469.
- 787 Coifman, R.R., Lafon, S., Lee, A.B., et al. 2005. Geometric diffusions as a tool for harmonic  
788 analysis and structure definition of data: diffusion maps. *Proceedings of the National Academy of*  
789 *Sciences of the United States of America* 102(21), pp. 7426–7431.
- 790 Colasanto, M.P., Eyal, S., Mohassel, P., et al. 2016. Development of a subset of forelimb  
791 muscles and their attachment sites requires the ulnar-mammary syndrome gene *Tbx3*. *Disease*  
792 *Models & Mechanisms* 9(11), pp. 1257–1269.
- 793 Craft, A.M., Rockel, J.S., Nartiss, Y., Kandel, R.A., Alman, B.A. and Keller, G.M. 2015.  
794 Generation of articular chondrocytes from human pluripotent stem cells. *Nature Biotechnology*  
795 33(6), pp. 638–645.
- 796 Dasuri, K., Antonovici, M., Chen, K., et al. 2004. The synovial proteome: analysis of fibroblast-  
797 like synoviocytes. *Arthritis Research & Therapy* 6(2), pp. R161-8.
- 798 Decker, R.S. 2017. Articular cartilage and joint development from embryogenesis to adulthood.  
799 *Seminars in Cell & Developmental Biology* 62, pp. 50–56.
- 800 Decker, R.S., Koyama, E. and Pacifici, M. 2014. Genesis and morphogenesis of limb synovial  
801 joints and articular cartilage. *Matrix Biology* 39, pp. 5–10.
- 802 Dy, P., Smits, P., Silvester, A., et al. 2010. Synovial joint morphogenesis requires the  
803 chondrogenic action of Sox5 and Sox6 in growth plate and articular cartilage. *Developmental*  
804 *Biology* 341(2), pp. 346–359.
- 805 Dymont, N.A., Breidenbach, A.P., Schwartz, A.G., et al. 2015. Gdf5 progenitors give rise to  
806 fibrocartilage cells that mineralize via hedgehog signaling to form the zonal enthesis.  
807 *Developmental Biology* 405(1), pp. 96–107.
- 808 Feng, C., Chan, W.C.W., Lam, Y., et al. 2019. Lgr5 and Col22a1 Mark Progenitor Cells in the  
809 Lineage toward Juvenile Articular Chondrocytes. *Stem cell reports*.
- 810 Ferrari, D. and Kosher, R.A. 2006. Expression of Dlx5 and Dlx6 during specification of the  
811 elbow joint. *The International Journal of Developmental Biology* 50(8), pp. 709–713.

- 812 Fromental-Ramain, C., Warot, X., Lakkaraju, S., et al. 1996. Specific and redundant functions of  
813 the paralogous Hoxa-9 and Hoxd-9 genes in forelimb and axial skeleton patterning. *Development*  
814 122(2), pp. 461–472.
- 815 Gao, Y., Lan, Y., Liu, H. and Jiang, R. 2011. The zinc finger transcription factors Osr1 and Osr2  
816 control synovial joint formation. *Developmental Biology* 352(1), pp. 83–91.
- 817 Grogan, S.P., Duffy, S.F., Pauli, C., et al. 2013. Zone-specific gene expression patterns in  
818 articular cartilage. *Arthritis and Rheumatism* 65(2), pp. 418–428.
- 819 Guo, G., Huss, M., Tong, G.Q., et al. 2010. Resolution of cell fate decisions revealed by single-  
820 cell gene expression analysis from zygote to blastocyst. *Developmental Cell* 18(4), pp. 675–685.
- 821 Guo, X., Day, T.F., Jiang, X., Garrett-Beal, L., Topol, L. and Yang, Y. 2004. Wnt/beta-catenin  
822 signaling is sufficient and necessary for synovial joint formation. *Genes & Development* 18(19),  
823 pp. 2404–2417.
- 824 Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F. and Theis, F.J. 2016. Diffusion pseudotime  
825 robustly reconstructs lineage branching. *Nature Methods* 13(10), pp. 845–848.
- 826 Hartmann, C. and Tabin, C.J. 2001. Wnt-14 plays a pivotal role in inducing synovial joint  
827 formation in the developing appendicular skeleton. *Cell* 104(3), pp. 341–351.
- 828 Havis, E., Bonnin, M.-A., Esteves de Lima, J., Charvet, B., Milet, C. and Duprez, D. 2016.  
829 TGFβ and FGF promote tendon progenitor fate and act downstream of muscle contraction to  
830 regulate tendon differentiation during chick limb development. *Development* 143(20), pp. 3839–  
831 3851.
- 832 Havis, E., Bonnin, M.-A., Olivera-Martinez, I., et al. 2014. Transcriptomic analysis of mouse  
833 limb tendon cells during development. *Development* 141(19), pp. 3683–3696.
- 834 den Hollander, W., Pulyakhina, I., Boer, C., et al. 2019. Annotating Transcriptional Effects of  
835 Genetic Variants in Disease-Relevant Tissue: Transcriptome-Wide Allelic Imbalance in  
836 Osteoarthritic Cartilage. *Arthritis & rheumatology (Hoboken, N.J.)* 71(4), pp. 561–570.
- 837 Hyde, G., Dover, S., Aszodi, A., Wallis, G.A. and Boot-Handford, R.P. 2007. Lineage tracing  
838 using matrilin-1 gene expression reveals that articular chondrocytes exist as the joint interzone  
839 forms. *Developmental Biology* 304(2), pp. 825–833.
- 840 Itoh, S., Kanno, S., Gai, Z., et al. 2008. Trps1 plays a pivotal role downstream of Gdf5 signaling  
841 in promoting chondrogenesis and apoptosis of ATDC5 cells. *Genes To Cells* 13(4), pp. 355–363.
- 842 Iwamoto, M., Tamamura, Y., Koyama, E., et al. 2007. Transcription factor ERG and joint and  
843 articular cartilage formation during mouse limb and spine skeletogenesis. *Developmental*  
844 *Biology* 305(1), pp. 40–51.
- 845 Jenner, F., Ijpma, A., Cleary, M., et al. 2014. Differential gene expression of the intermediate  
846 and outer interzone layers of developing articular cartilage in murine embryos. *Stem Cells and*  
847 *Development* 23(16), pp. 1883–1898.
- 848 Jiang, R., Lan, Y., Chapman, H.D., et al. 1998. Defects in limb, craniofacial, and thymic  
849 development in Jagged2 mutant mice. *Genes & Development* 12(7), pp. 1046–1057.
- 850 Johnson, K., Zhu, S., Tremblay, M.S., et al. 2012. A stem cell-based approach to cartilage repair.  
851 *Science* 336(6082), pp. 717–721.
- 852 Kan, A. and Tabin, C.J. 2013. c-Jun is required for the specification of joint cell fates. *Genes &*  
853 *Development* 27(5), pp. 514–524.
- 854 Karsenty, G. and Wagner, E.F. 2002. Reaching a genetic and molecular understanding of skeletal  
855 development. *Developmental Cell* 2(4), pp. 389–406.
- 856 Kawata, M., Mori, D., Kanke, K., et al. 2019. Simple and Robust Differentiation of Human  
857 Pluripotent Stem Cells toward Chondrocytes by Two Small-Molecule Compounds. *Stem cell*



- 858 *reports* 13(3), pp. 530–544.
- 859 Kelly, N.H., Huynh, N.P. and Guilak, F. 2019. Single cell RNA-sequencing reveals cellular  
860 heterogeneity and trajectories of lineage specification during murine embryonic limb  
861 development. *bioRxiv*.
- 862 Khoa, N.D., Hasunuma, T., Kobata, T., Kato, T. and Nishioka, K. 1999. Expression of murine  
863 HOXD9 during embryonic joint patterning and in human T lymphotropic virus type I tax  
864 transgenic mice with arthropathy resembling rheumatoid arthritis. *Arthritis and Rheumatism*  
865 42(4), pp. 686–696.
- 866 Kuleshov, M.V., Jones, M.R., Rouillard, A.D., et al. 2016. Enrichr: a comprehensive gene set  
867 enrichment analysis web server 2016 update. *Nucleic Acids Research* 44(W1), pp. W90-7.
- 868 Kumar, P., Tan, Y. and Cahan, P. 2017. Understanding development and stem cells using single  
869 cell-based analyses of gene expression. *Development* 144(1), pp. 17–32.
- 870 Kunath, M., Lüdecke, H.-J. and Vortkamp, A. 2002. Expression of *Trps1* during mouse  
871 embryonic development. *Mechanisms of Development* 119 Suppl 1, pp. S117-20.
- 872 La Manno, G., Soldatov, R., Zeisel, A., et al. 2018. RNA velocity of single cells. *Nature*  
873 560(7719), pp. 494–498.
- 874 Lee, M.-H., Kwon, T.-G., Park, H.-S., Wozney, J.M. and Ryoo, H.-M. 2003. BMP-2-induced  
875 Osterix expression is mediated by *Dlx5* but is independent of *Runx2*. *Biochemical and*  
876 *Biophysical Research Communications* 309(3), pp. 689–694.
- 877 Leussink, B., Brouwer, A., el Khattabi, M., Poelmann, R.E., Gittenberger-de Groot, A.C. and  
878 Meijlink, F. 1995. Expression patterns of the paired-related homeobox genes *MHox/Prx1* and  
879 *S8/Prx2* suggest roles in development of the heart and the forebrain. *Mechanisms of*  
880 *Development* 52(1), pp. 51–64.
- 881 Li, D., Sakuma, R., Vakili, N.A., et al. 2014. Formation of proximal and anterior limb skeleton  
882 requires early function of *Irx3* and *Irx5* and is negatively regulated by *Shh* signaling.  
883 *Developmental Cell* 29(2), pp. 233–240.
- 884 Li, H., Liu, Q., Hu, Xiang, et al. 2009. Human *ZCCHC12* activates AP-1 and CREB signaling as  
885 a transcriptional co-activator. *Acta biochimica et biophysica Sinica* 41(7), pp. 535–544.
- 886 Lien, C.-Y., Lee, O.K. and Su, Y. 2007. *Cbfb* enhances the osteogenic differentiation of both  
887 human and mouse mesenchymal stem cells induced by *Cbfa-1* via reducing its ubiquitination-  
888 mediated degradation. *Stem Cells* 25(6), pp. 1462–1468.
- 889 Lin, Y., Li, X.-Y., Willis, A.L., Liu, C., Chen, G. and Weiss, S.J. 2014. *Snail1*-dependent control  
890 of embryonic stem cell pluripotency and lineage commitment. *Nature Communications* 5, p.  
891 3070.
- 892 Liu, C.-F., Samsa, W.E., Zhou, G. and Lefebvre, V. 2017. Transcriptional control of chondrocyte  
893 specification and differentiation. *Seminars in Cell & Developmental Biology* 62, pp. 34–49.
- 894 Liu, H., Xu, J., Liu, C.-F., Lan, Y., Wylie, C. and Jiang, R. 2015. Whole transcriptome  
895 expression profiling of mouse limb tendon development by using RNA-seq. *Journal of*  
896 *Orthopaedic Research* 33(6), pp. 840–848.
- 897 Longobardi, L., Li, T., Tagliaferro, L., et al. 2015. Synovial joints: from development to  
898 homeostasis. *Current osteoporosis reports* 13(1), pp. 41–51.
- 899 MacLean, H.E., Kim, J.I., Glimcher, M.J., Wang, J., Kronenberg, H.M. and Glimcher, L.H.  
900 2003. Absence of transcription factor *c-maf* causes abnormal terminal differentiation of  
901 hypertrophic chondrocytes during endochondral bone development. *Developmental Biology*  
902 262(1), pp. 51–63.
- 903 Marcil, A., Dumontier, E., Chamberland, M., Camper, S.A. and Drouin, J. 2003. *Pitx1* and *Pitx2*

904 are required for development of hindlimb buds. *Development* 130(1), pp. 45–55.  
905 McConnell, B.B. and Yang, V.W. 2010. Mammalian Krüppel-like factors in health and diseases.  
906 *Physiological Reviews* 90(4), pp. 1337–1381.  
907 McInnes, L. and Healy, J. 2018. UMAP: Uniform Manifold Approximation and Projection for  
908 Dimension Reduction. *arXiv*.  
909 Meech, R., Edelman, D.B., Jones, F.S. and Makarenkova, H.P. 2005. The homeobox  
910 transcription factor *Barx2* regulates chondrogenesis during limb development. *Development*  
911 132(9), pp. 2135–2146.  
912 Mella, S., Soula, C., Morello, D., Crozatier, M. and Vincent, A. 2004. Expression patterns of the  
913 *coe/ebf* transcription factor genes during chicken and mouse limb development. *Gene Expression*  
914 *Patterns* 4(5), pp. 537–542.  
915 Niederreither, K., D’Souza, R.N. and de Crombrughe, B. 1992. Minimal DNA sequences that  
916 control the cell lineage-specific expression of the pro alpha 2(I) collagen promoter in transgenic  
917 mice. *The Journal of Cell Biology* 119(5), pp. 1361–1370.  
918 Norris, R.A., Damon, B., Mironov, V., et al. 2007. Periostin regulates collagen fibrillogenesis  
919 and the biomechanical properties of connective tissues. *Journal of Cellular Biochemistry* 101(3),  
920 pp. 695–711.  
921 Oka, C., Tsujimoto, R., Kajikawa, M., et al. 2004. HtrA1 serine protease inhibits signaling  
922 mediated by Tgfbeta family proteins. *Development* 131(5), pp. 1041–1053.  
923 Oldershaw, R.A., Baxter, M.A., Lowe, E.T., et al. 2010. Directed differentiation of human  
924 embryonic stem cells toward chondrocytes. *Nature Biotechnology* 28(11), pp. 1187–1194.  
925 Orgeur, M., Martens, M., Leonte, G., et al. 2018. Genome-wide strategies identify downstream  
926 target genes of chick connective tissue-associated transcription factors. *Development* 145(7).  
927 Park, Y.W., Kang, Y.M., Butterfield, J., Detmar, M., Goronzy, J.J. and Weyand, C.M. 2004.  
928 Thrombospondin 2 functions as an endogenous regulator of angiogenesis and inflammation in  
929 rheumatoid arthritis. *The American Journal of Pathology* 165(6), pp. 2087–2098.  
930 Pazin, D.E., Gamer, L.W., Capelo, L.P., Cox, K.A. and Rosen, V. 2014. Gene signature of the  
931 embryonic meniscus. *Journal of Orthopaedic Research* 32(1), pp. 46–53.  
932 Pazin, D.E., Gamer, L.W., Cox, K.A. and Rosen, V. 2012. Molecular profiling of synovial joints:  
933 use of microarray analysis to identify factors that direct the development of the knee and elbow.  
934 *Developmental Dynamics* 241(11), pp. 1816–1826.  
935 Peffers, M.J., Fang, Y., Cheung, K., Wei, T.K.J., Clegg, P.D. and Birch, H.L. 2015.  
936 Transcriptome analysis of ageing in uninjured human Achilles tendon. *Arthritis Research &*  
937 *Therapy* 17, p. 33.  
938 Rallis, C., Bruneau, B.G., Del Buono, J., et al. 2003. *Tbx5* is required for forelimb bud formation  
939 and continued outgrowth. *Development* 130(12), pp. 2741–2751.  
940 Ray, A., Singh, P.N.P., Sohaskey, M.L., Harland, R.M. and Bandyopadhyay, A. 2015. Precise  
941 spatial restriction of BMP signaling is essential for articular cartilage differentiation.  
942 *Development* 142(6), pp. 1169–1179.  
943 Rockel, J.S., Yu, C., Whetstone, H., et al. 2016. Hedgehog inhibits  $\beta$ -catenin activity in synovial  
944 joint development and osteoarthritis. *The Journal of Clinical Investigation* 126(5), pp. 1649–  
945 1663.  
946 Roelofs, A.J., Zupan, J., Riemen, A.H.K., et al. 2017. Joint morphogenetic cells in the adult  
947 mammalian synovium. *Nature Communications* 8, p. 15040.  
948 Salva, J.E. and Merrill, A.E. 2017. Signaling networks in joint development. *Developmental*  
949 *Dynamics* 246(4), pp. 262–274.

- 950 Satija, R., Farrell, J.A., Gennert, D., Schier, A.F. and Regev, A. 2015. Spatial reconstruction of  
951 single-cell gene expression data. *Nature Biotechnology* 33(5), pp. 495–502.
- 952 Shekhani, M.T., Forde, T.S., Adilbayeva, A., et al. 2016. Collagen triple helix repeat containing  
953 1 is a new promigratory marker of arthritic pannus. *Arthritis Research & Therapy* 18, p. 171.
- 954 Shwartz, Y., Viukov, S., Krief, S. and Zelzer, E. 2016. Joint Development Involves a Continuous  
955 Influx of Gdf5-Positive Cells. *Cell reports* 15(12), pp. 2577–2587.
- 956 Singh, P.N.P., Yadav, U.S., Azad, K., Goswami, P., Kinare, V. and Bandyopadhyay, A. 2018.  
957 NFIA and GATA3 are crucial regulators of embryonic articular cartilage differentiation.  
958 *Development* 145(2).
- 959 Singh, S., Canseco, D.C., Manda, S.M., et al. 2014. Cytoglobin modulates myogenic progenitor  
960 cell viability and muscle regeneration. *Proceedings of the National Academy of Sciences of the*  
961 *United States of America* 111(1), pp. E129-38.
- 962 Soeda, T., Deng, J.M., de Crombrugge, B., Behringer, R.R., Nakamura, T. and Akiyama, H.  
963 2010. Sox9-expressing precursors are the cellular origin of the cruciate ligament of the knee joint  
964 and the limb tendons. *Genesis* 48(11), pp. 635–644.
- 965 Storm, E.E. and Kingsley, D.M. 1999. GDF5 coordinates bone and joint formation during digit  
966 development. *Developmental Biology* 209(1), pp. 11–27.
- 967 Storm, E.E. and Kingsley, D.M. 1996. Joint patterning defects caused by single and double  
968 mutations in members of the bone morphogenetic protein (BMP) family. *Development* 122(12),  
969 pp. 3969–3979.
- 970 Stricker, S., Mathia, S., Haupt, J., Seemann, P., Meier, J. and Mundlos, S. 2012. Odd-skipped  
971 related genes regulate differentiation of embryonic limb mesenchyme and bone marrow  
972 mesenchymal stromal cells. *Stem Cells and Development* 21(4), pp. 623–633.
- 973 Subramanian, A. and Schilling, T.F. 2015. Tendon development and musculoskeletal assembly:  
974 emerging roles for the extracellular matrix. *Development* 142(24), pp. 4191–4204.
- 975 Subramanian, A. and Schilling, T.F. 2014. Thrombospondin-4 controls matrix assembly during  
976 development and repair of myotendinous junctions. *eLife* 3.
- 977 Sugimoto, Y., Takimoto, A., Akiyama, H., et al. 2013. Scx+/Sox9+ progenitors contribute to the  
978 establishment of the junction between cartilage and tendon/ligament. *Development* 140(11), pp.  
979 2280–2288.
- 980 Syddall, C.M., Reynard, L.N., Young, D.A. and Loughlin, J. 2013. The identification of trans-  
981 acting factors that regulate the expression of GDF5 via the osteoarthritis susceptibility SNP  
982 rs143383. *PLoS Genetics* 9(6), p. e1003557.
- 983 Tabula Muris Consortium, Overall coordination, Logistical coordination, et al. 2018. Single-cell  
984 transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* 562(7727), pp. 367–372.
- 985 Taher, L., Collette, N.M., Murugesu, D., Maxwell, E., Ovcharenko, I. and Loots, G.G. 2011.  
986 Global gene expression analysis of murine limb development. *Plos One* 6(12), p. e28358.
- 987 Takagi, T., Moribe, H., Kondoh, H. and Higashi, Y. 1998. DeltaEF1, a zinc finger and  
988 homeodomain transcription factor, is required for skeleton patterning in multiple lineages.  
989 *Development* 125(1), pp. 21–31.
- 990 Tan, Y. and Cahan, P. 2019. SingleCellNet: A Computational Tool to Classify Single Cell RNA-  
991 Seq Data Across Platforms and Across Species. *Cell systems*.
- 992 Traag, V.A., Waltman, L. and van Eck, N.J. 2019. From Louvain to Leiden: guaranteeing well-  
993 connected communities. *Scientific reports* 9(1), p. 5233.
- 994 Vincent, T., Neve, E.P.A., Johnson, J.R., et al. 2009. A SNAIL1-SMAD3/4 transcriptional  
995 repressor complex promotes TGF-beta mediated epithelial-mesenchymal transition. *Nature Cell*

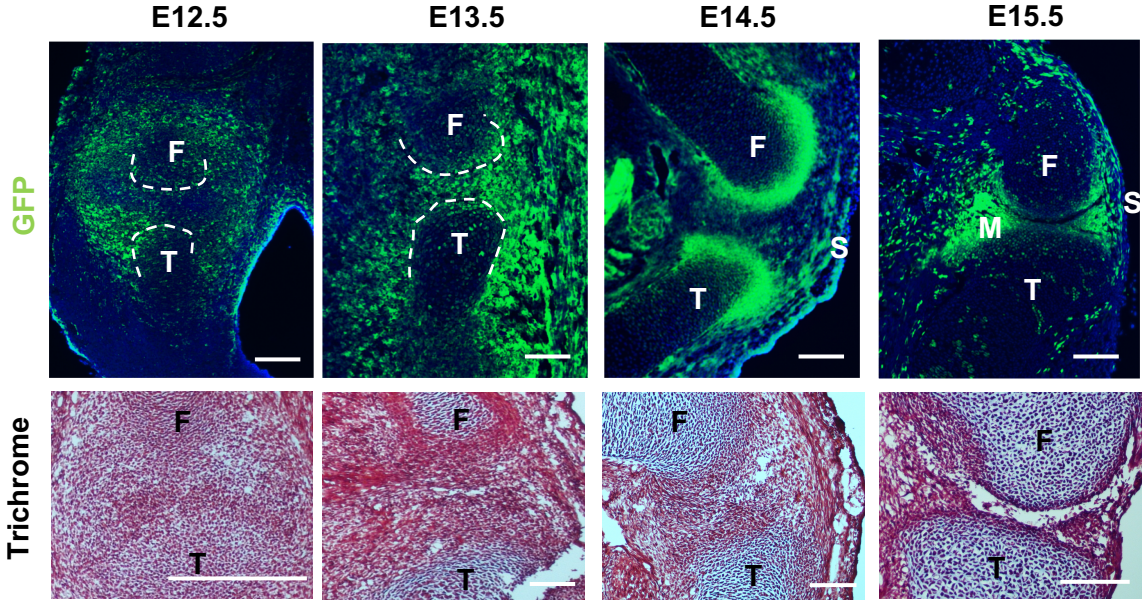
- 996 *Biology* 11(8), pp. 943–950.
- 997 Wang, J.S., Infante, C.R., Park, S. and Menke, D.B. 2018. PITX1 promotes chondrogenesis and  
998 myogenesis in mouse hindlimbs through conserved regulatory targets. *Developmental Biology*  
999 434(1), pp. 186–195.
- 1000 Wang, T., Nimkingratana, P., Smith, C.A., Cheng, A., Hardingham, T.E. and Kimber, S.J. 2019.  
1001 Enhanced chondrogenesis from human embryonic stem cells. *Stem Cell Research* 39, p. 101497.
- 1002 Wang, X., Lu, H., Urvalek, A.M., et al. 2011. KLF8 promotes human breast cancer cell invasion  
1003 and metastasis by transcriptional activation of MMP9. *Oncogene* 30(16), pp. 1901–1911.
- 1004 Witte, F., Dokas, J., Neuendorf, F., Mundlos, S. and Stricker, S. 2009. Comprehensive  
1005 expression analysis of all Wnt genes and their major secreted antagonists during mouse limb  
1006 development and cartilage differentiation. *Gene Expression Patterns* 9(4), pp. 215–223.
- 1007 Wolf, F.A., Angerer, P. and Theis, F.J. 2018. SCANPY: large-scale single-cell gene expression  
1008 data analysis. *Genome Biology* 19(1), p. 15.
- 1009 Yamashita, A., Morioka, M., Yahara, Y., et al. 2015. Generation of scaffoldless hyaline  
1010 cartilaginous tissue from human iPSCs. *Stem cell reports* 4(3), pp. 404–418.
- 1011 Yang, L., Cai, C.-L., Lin, L., et al. 2006. Isl1Cre reveals a common Bmp pathway in heart and  
1012 limb development. *Development* 133(8), pp. 1575–1585.
- 1013 Zelzer, E., Blitz, E., Killian, M.L. and Thomopoulos, S. 2014. Tendon-to-bone attachment: from  
1014 development to maturity. *Birth Defects Research. Part C, Embryo Today: Reviews* 102(1), pp.  
1015 101–112.
- 1016 Zhang, C., Zhang, E., Yang, L., et al. 2018. Histone deacetylase inhibitor treated cell sheet from  
1017 mouse tendon stem/progenitor cells promotes tendon repair. *Biomaterials* 172, pp. 66–82.
- 1018 Zhang, F., Guo, X., Zhang, Y., et al. 2014. Genome-wide copy number variation study and gene  
1019 expression analysis identify ABI3BP as a susceptibility gene for Kashin-Beck disease. *Human*  
1020 *Genetics* 133(6), pp. 793–799.
- 1021 Zhang, J., Ratanasirintraoort, S., Chandrasekaran, S., et al. 2016. LIN28 regulates stem cell  
1022 metabolism and conversion to primed pluripotency. *Cell Stem Cell* 19(1), pp. 66–80.
- 1023 Zhao, Q., Eberspaecher, H., Lefebvre, V. and De Crombrughe, B. 1997. Parallel expression of  
1024 Sox9 and Col2a1 in cells undergoing chondrogenesis. *Developmental Dynamics* 209(4), pp. 377–  
1025 386.
- 1026 Zheng, G.X.Y., Terry, J.M., Belgrader, P., et al. 2017. Massively parallel digital transcriptional  
1027 profiling of single cells. *Nature Communications* 8, p. 14049.
- 1028

**Table 1**

Stage	YFP <sup>+</sup> cells	YFP <sup>+</sup> cells %	cells captured	GLE cells	Reads	Reads/cell	Genes/cell
E12.5	74.8K	35.59%	3,107	1964	489,942,625	157,689	3,403
E13.5	113.9K	11.26%	4,786	2433	347,861,418	72,683	1,755
E14.5	54.1K	3.7%	1,888	993	359,671,420	190,503	2,835
E15.5	50.0K	7.57%	1,099	812	365,318,792	332,410	2,835

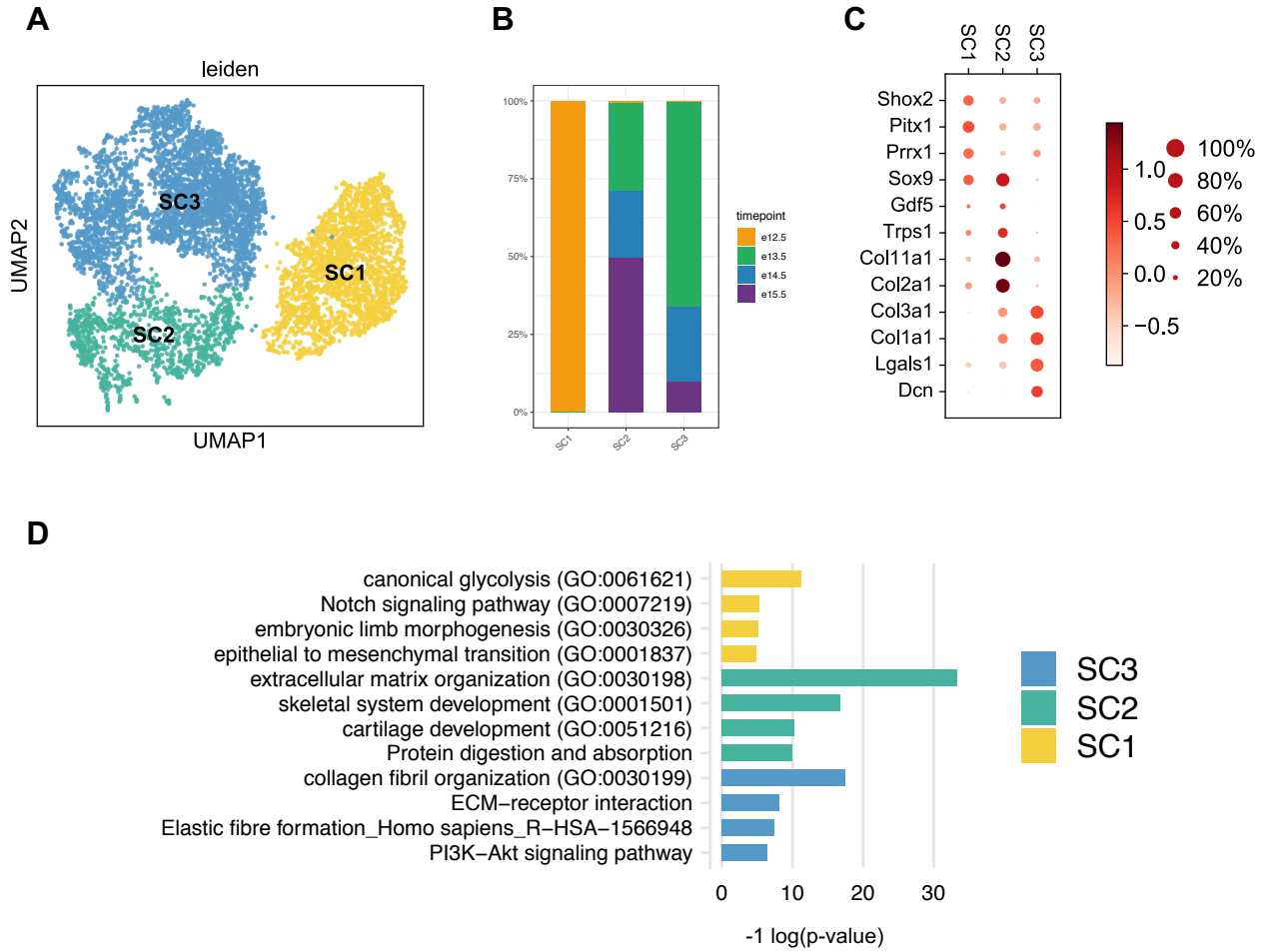
**Table 1:** Statistics on cells collected for scRNA-Seq. 'Cells captured' was determined by 10X CellRanger. GLE cells indicate the number of cells remaining after excluding cells unlikely to be GDF5-lineage, including immune cells, neural crest cells, and endothelial cells.

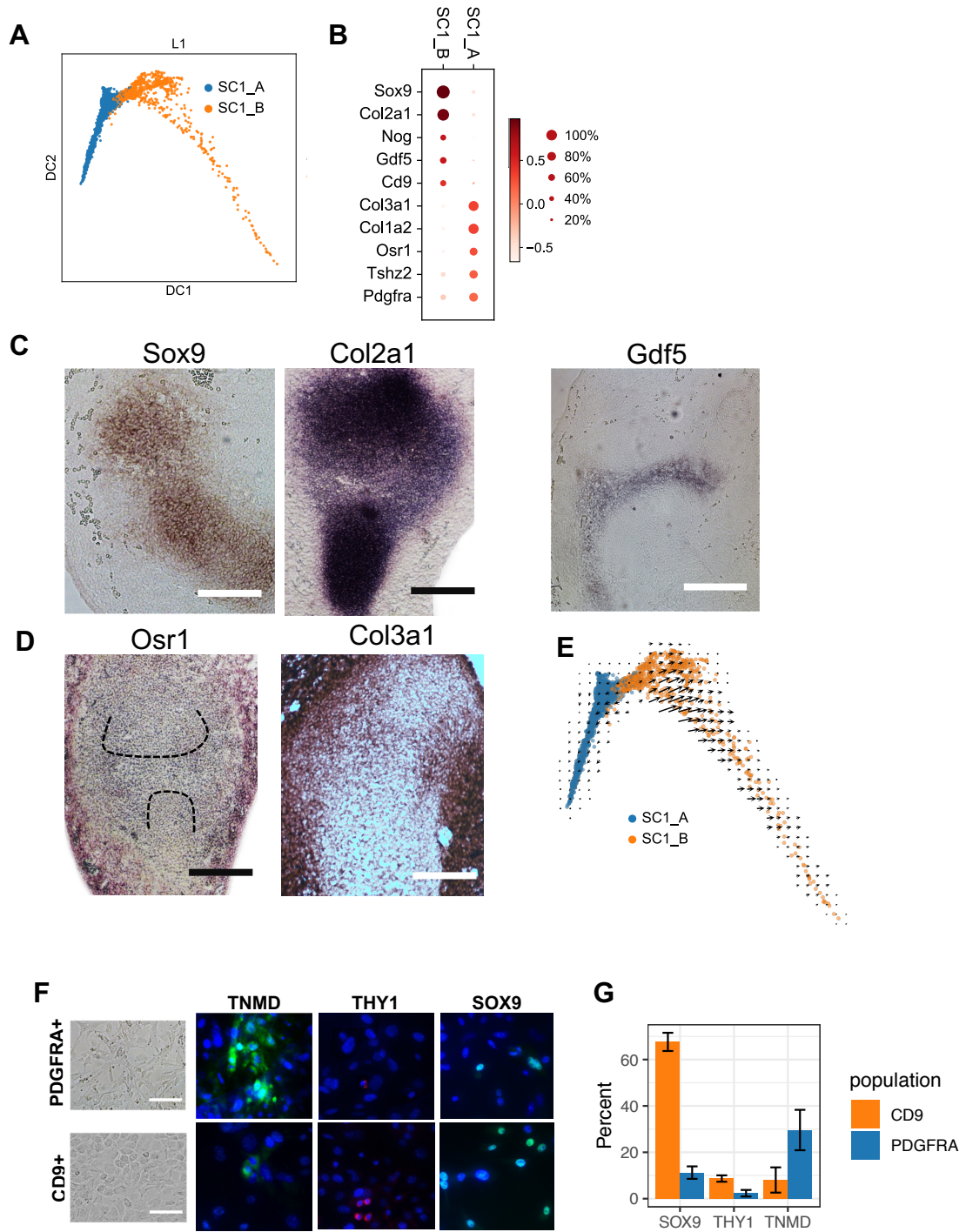
Figure 1



Scale bar=100μM

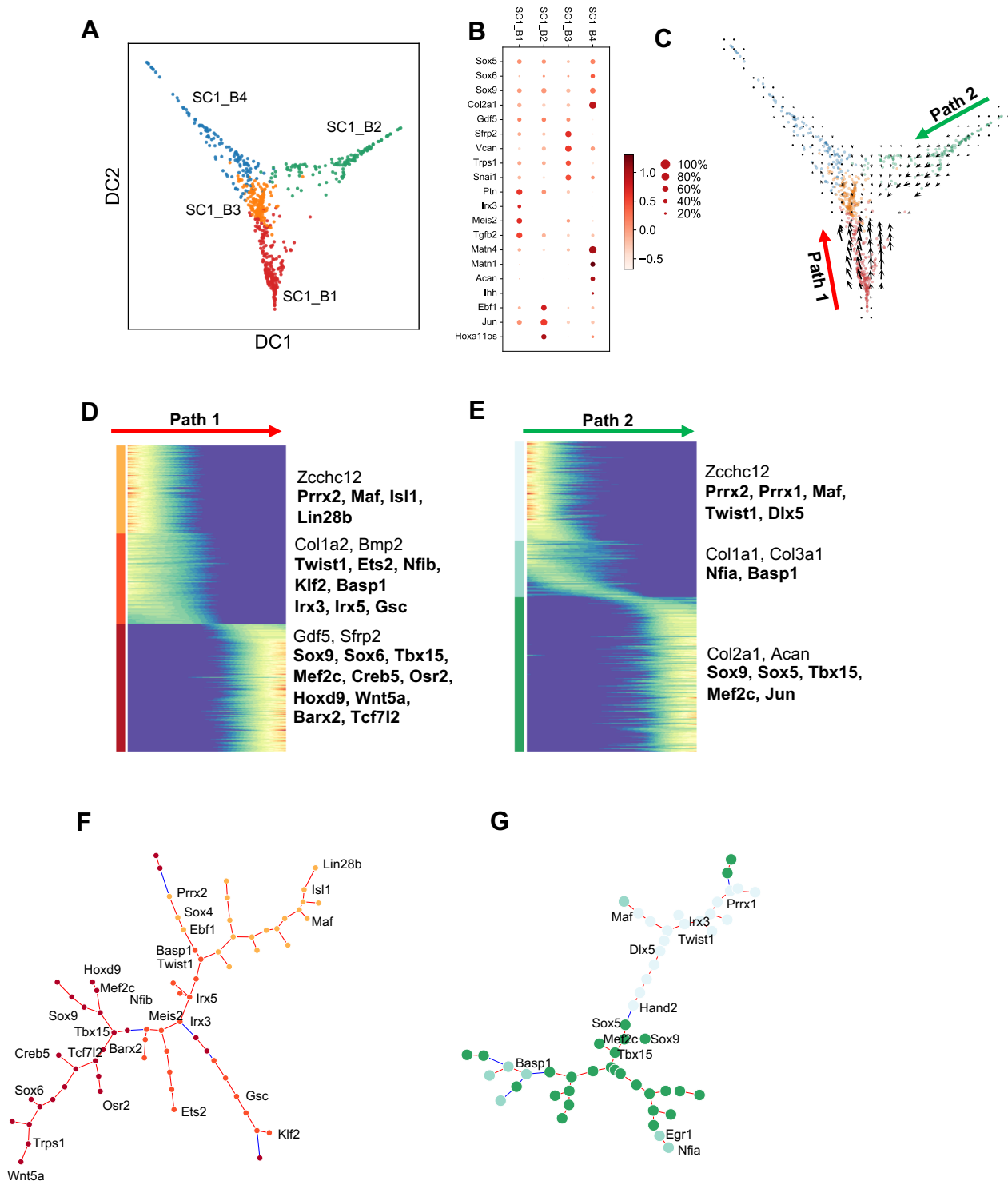
**Figure 2**



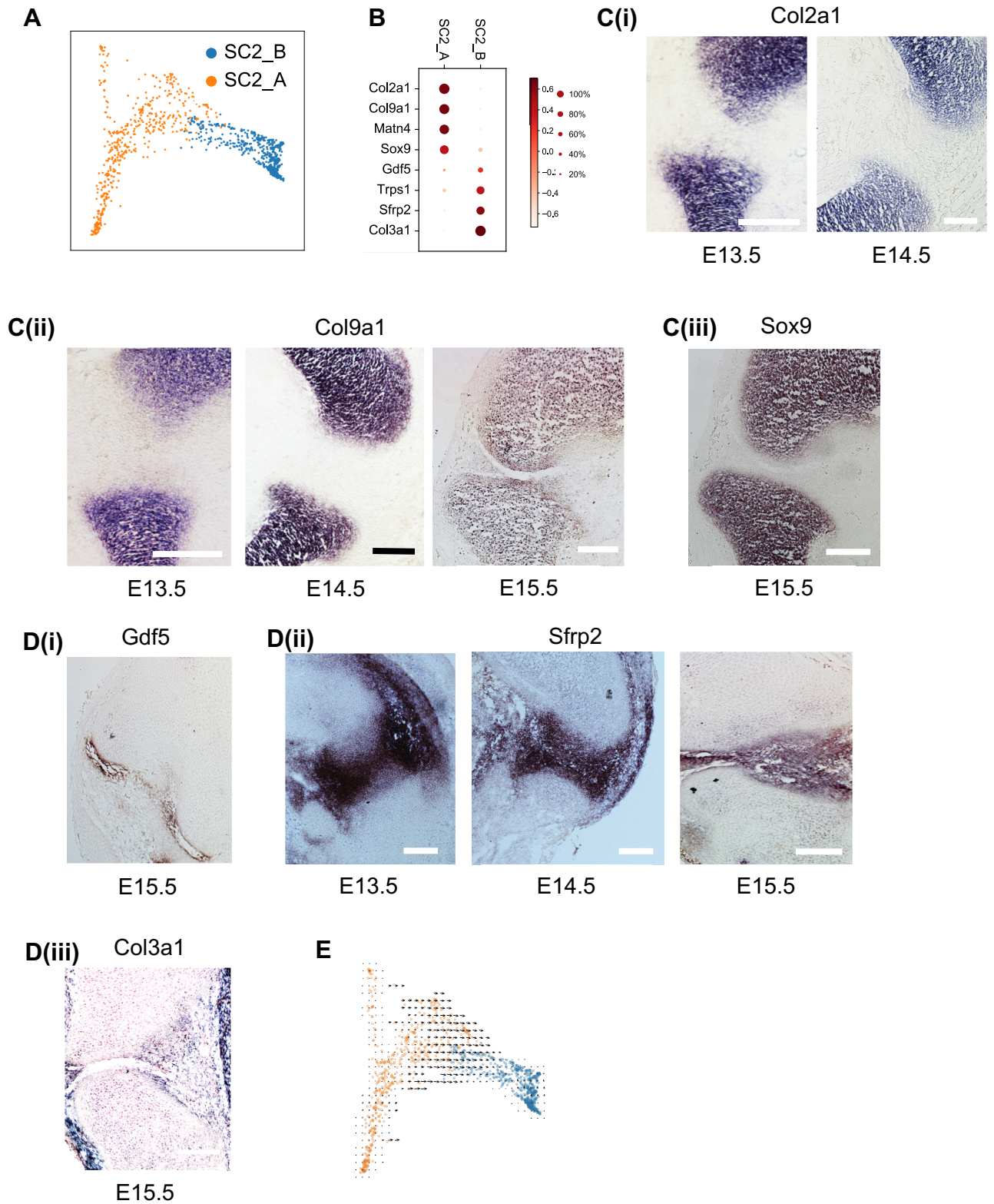
**Figure 3**



**Figure 4**



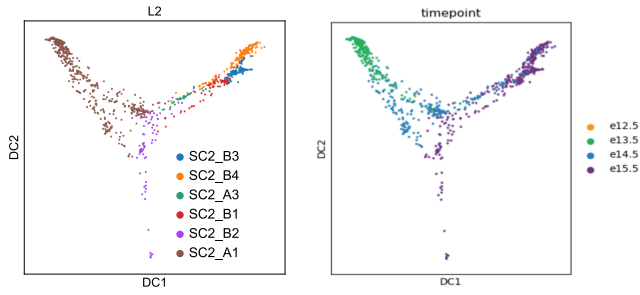
**Figure 5**



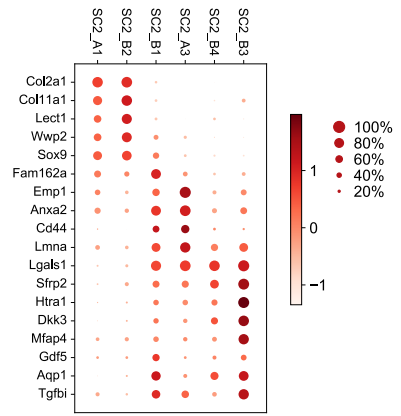
Scale bar=100 $\mu$ M

**Figure 5**

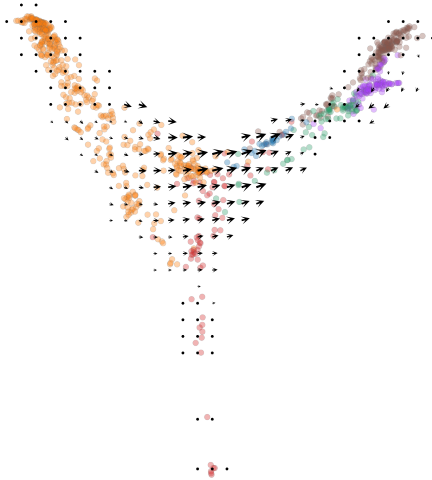
**F**



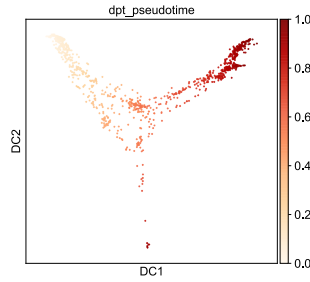
**G**



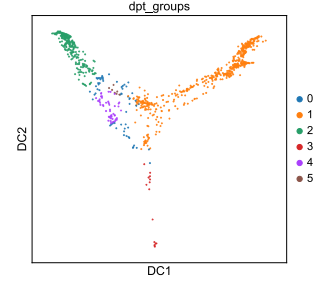
**H**



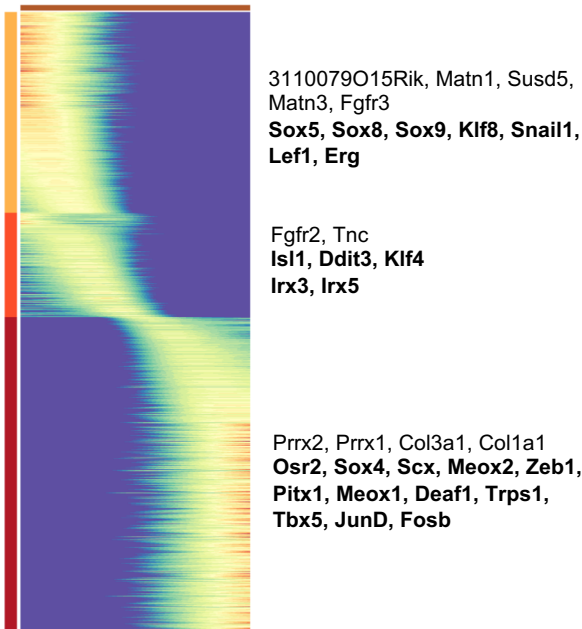
**I**



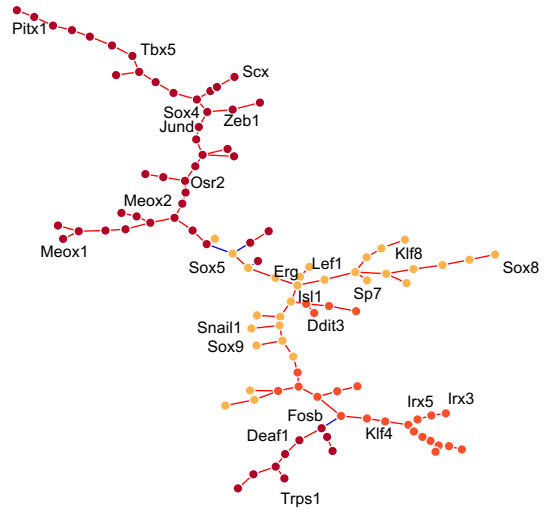
**J**



**K**

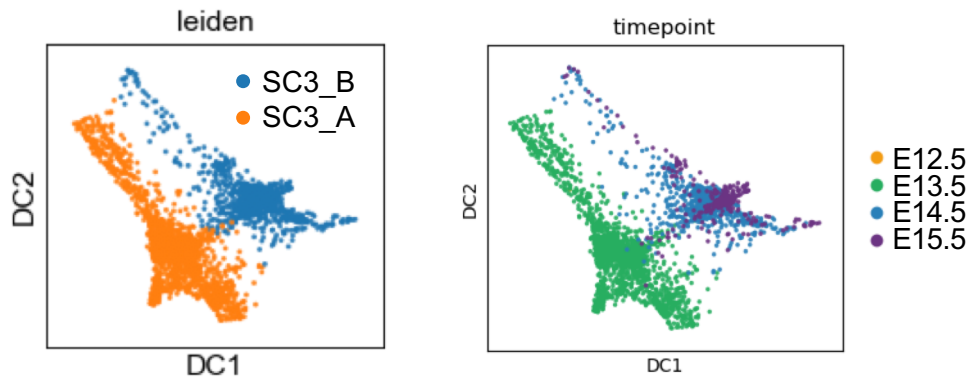


**L**

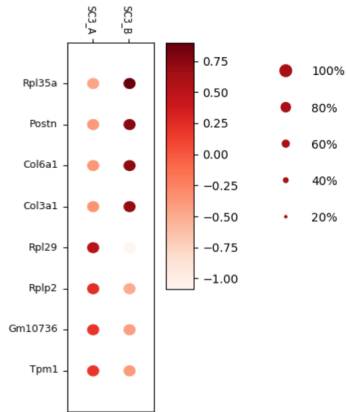


**Figure 6**

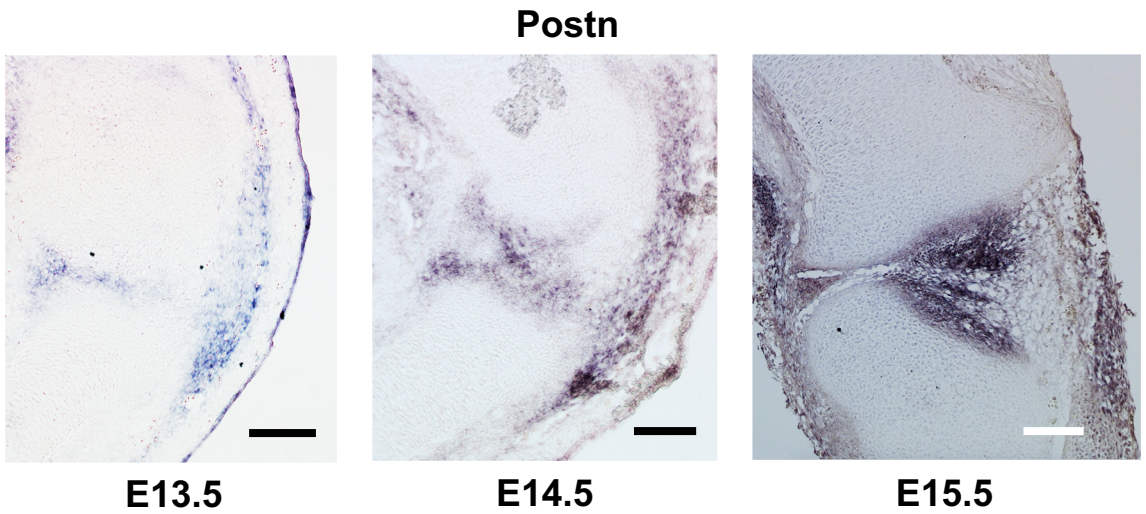
**A**



**B**

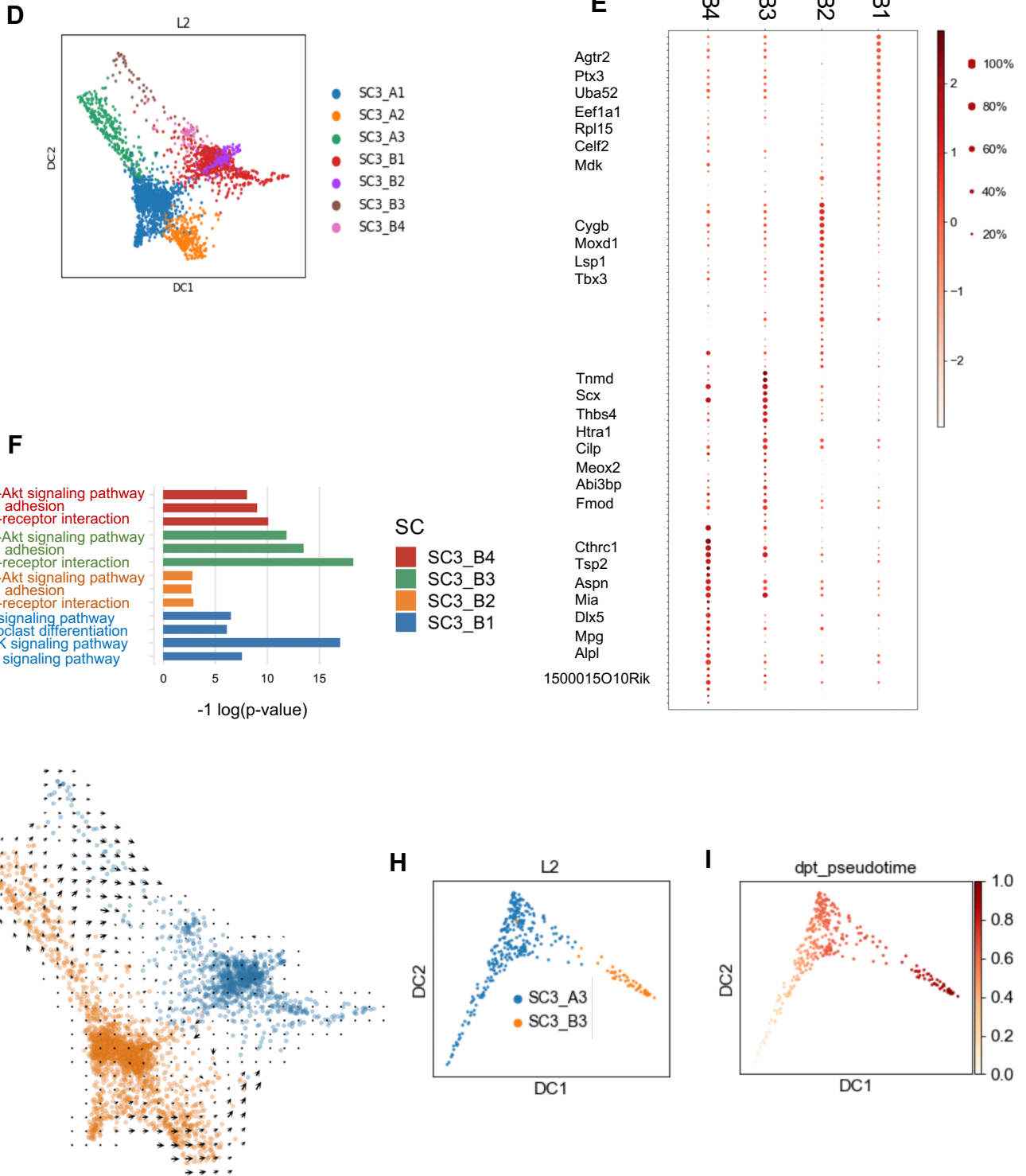


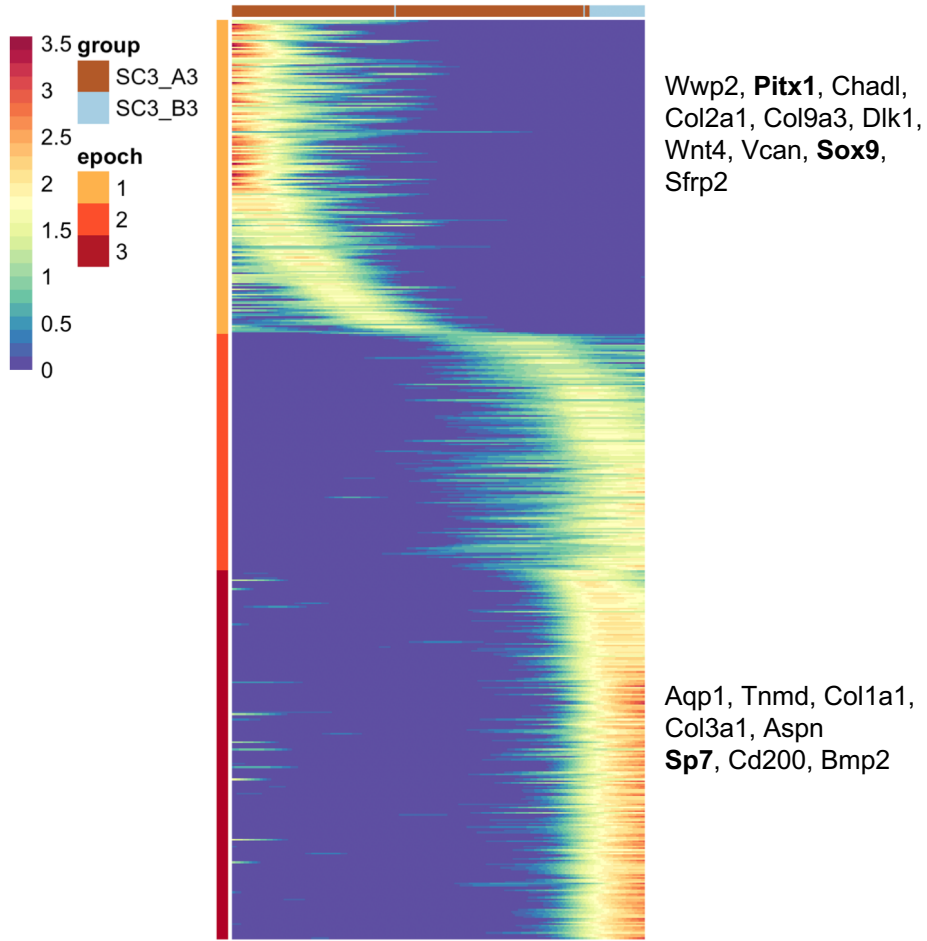
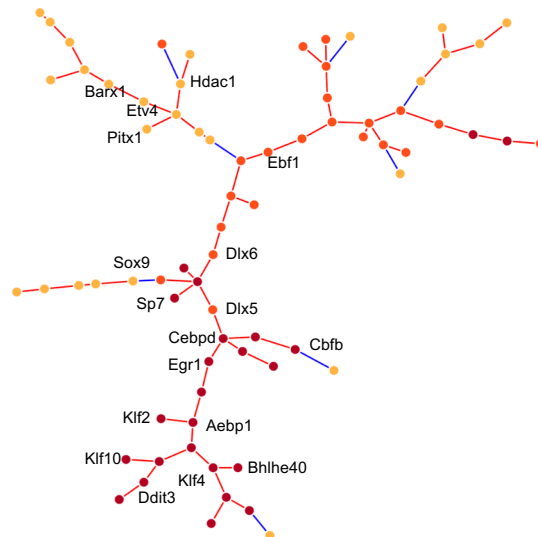
**C**



Scale bar=100 $\mu$ M

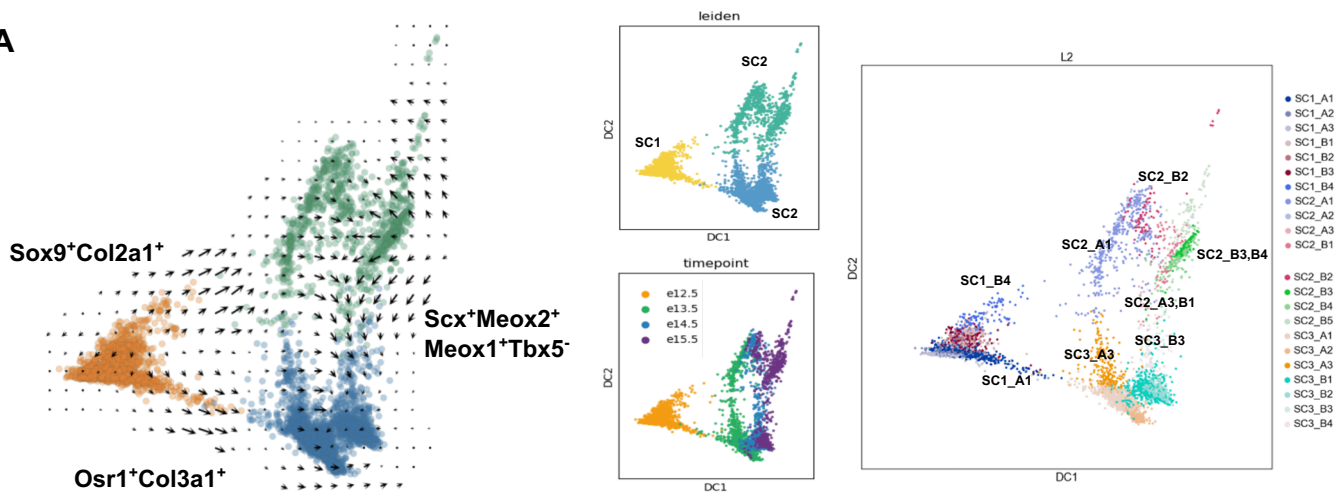
**Figure 6**



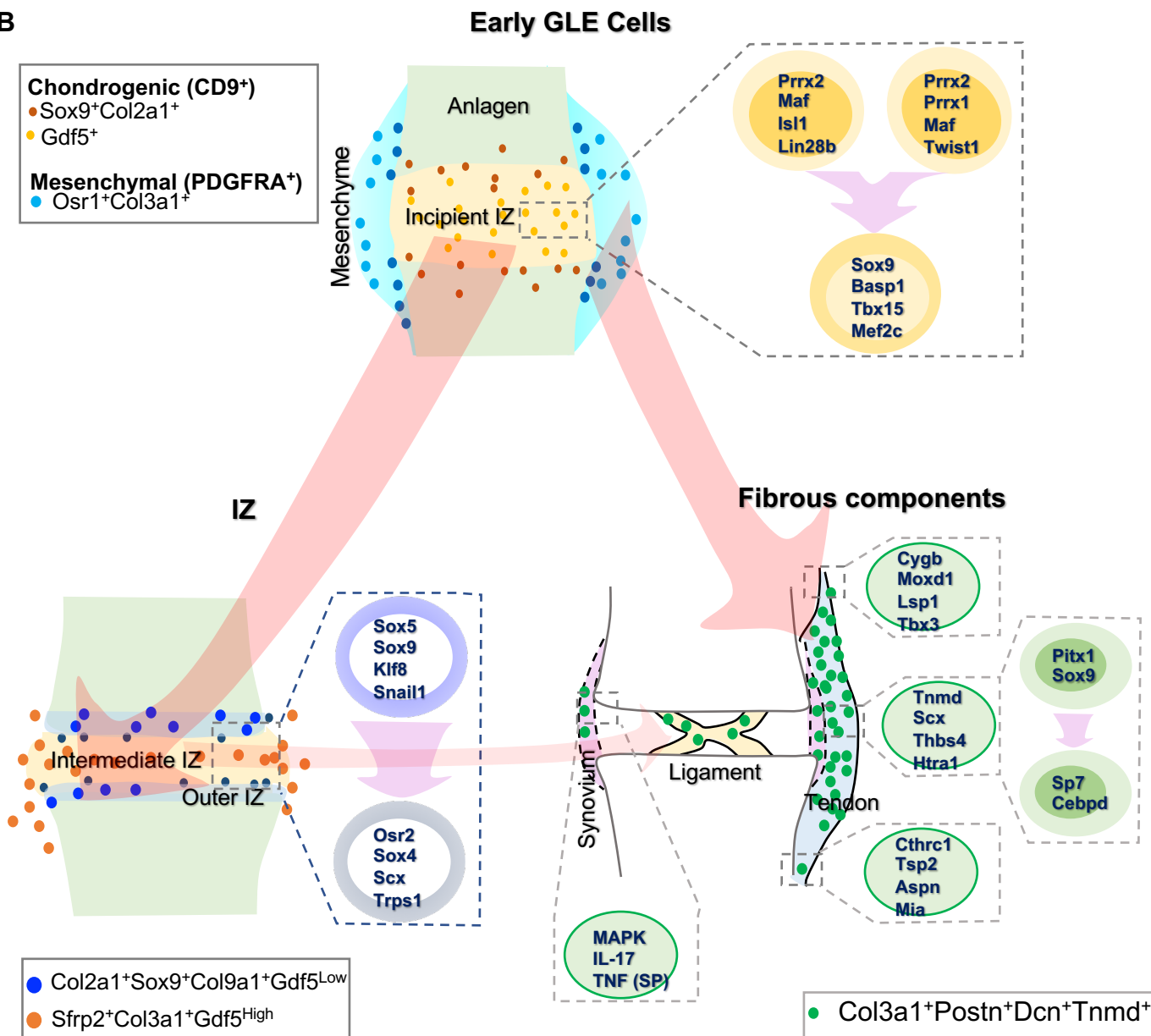
**J****K**

**Figure 7**

**A**

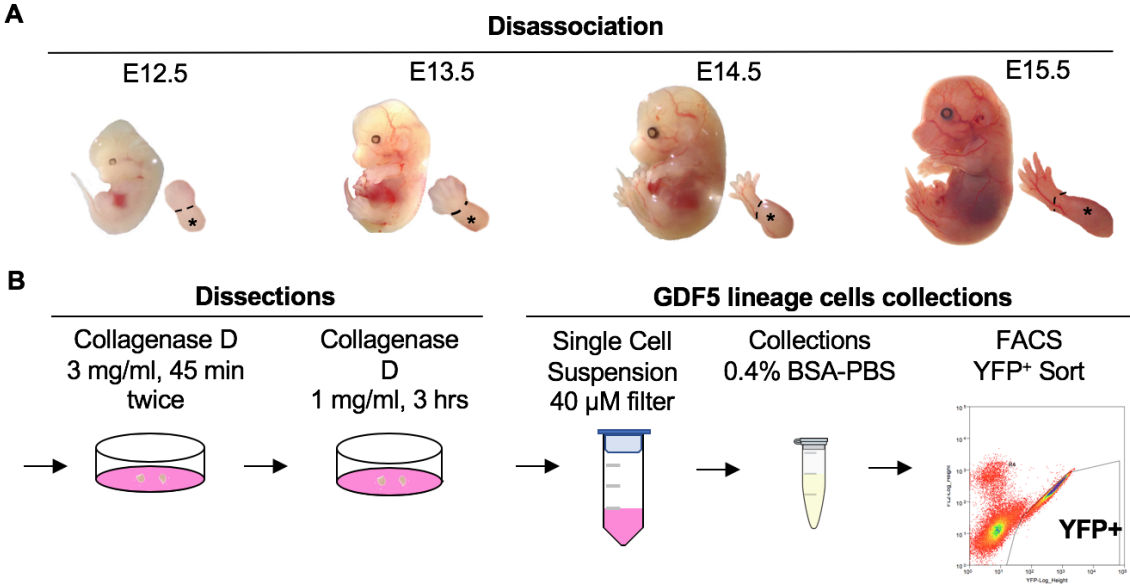


**B**



Supplemental Information

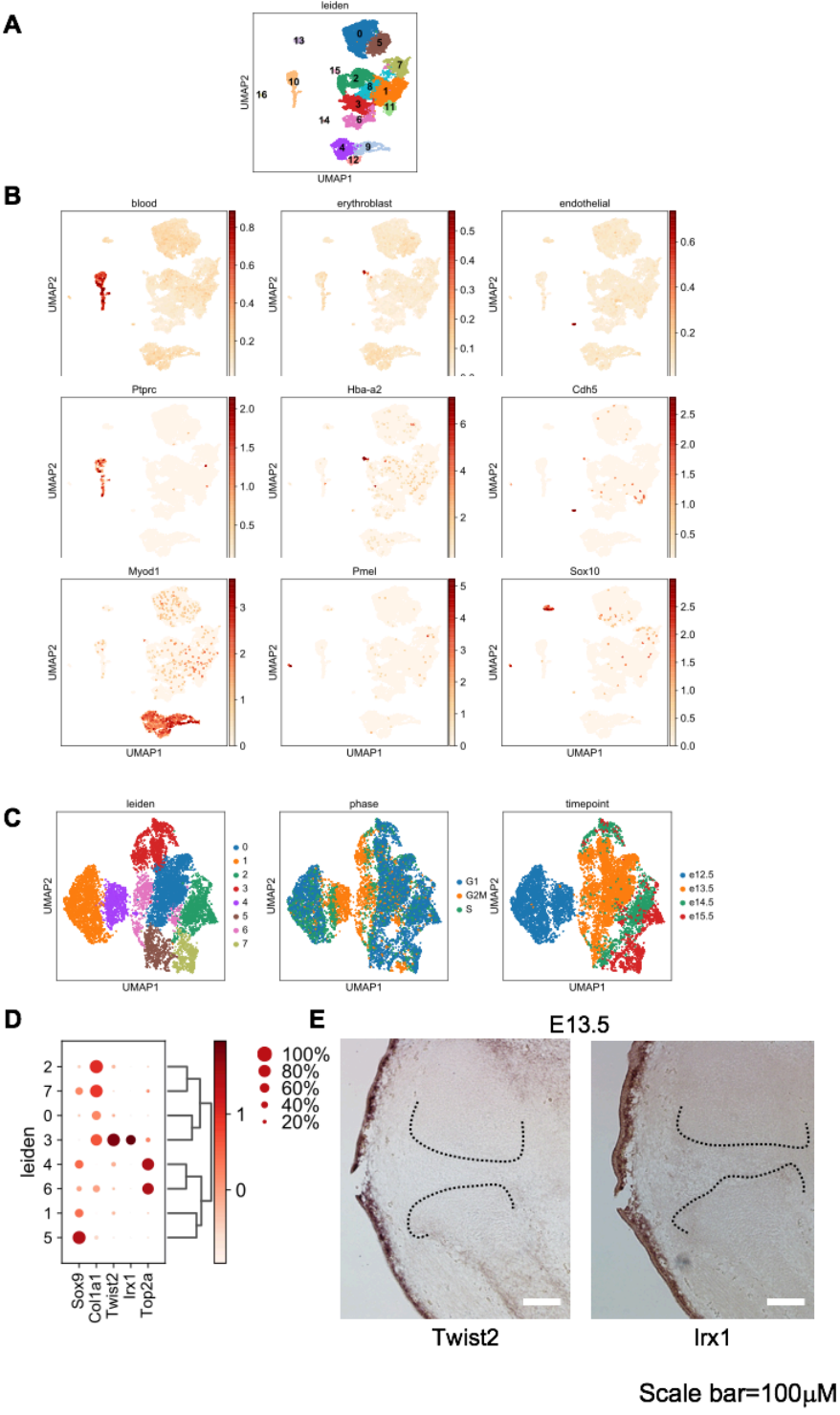
Supplemental Figure 1



**Supplemental Figure 1:** (A) Developmental stage and region of hind limb dissected for cell isolation. (B) Schematic of single cell dissociation and YFP<sup>+</sup> cell isolation.

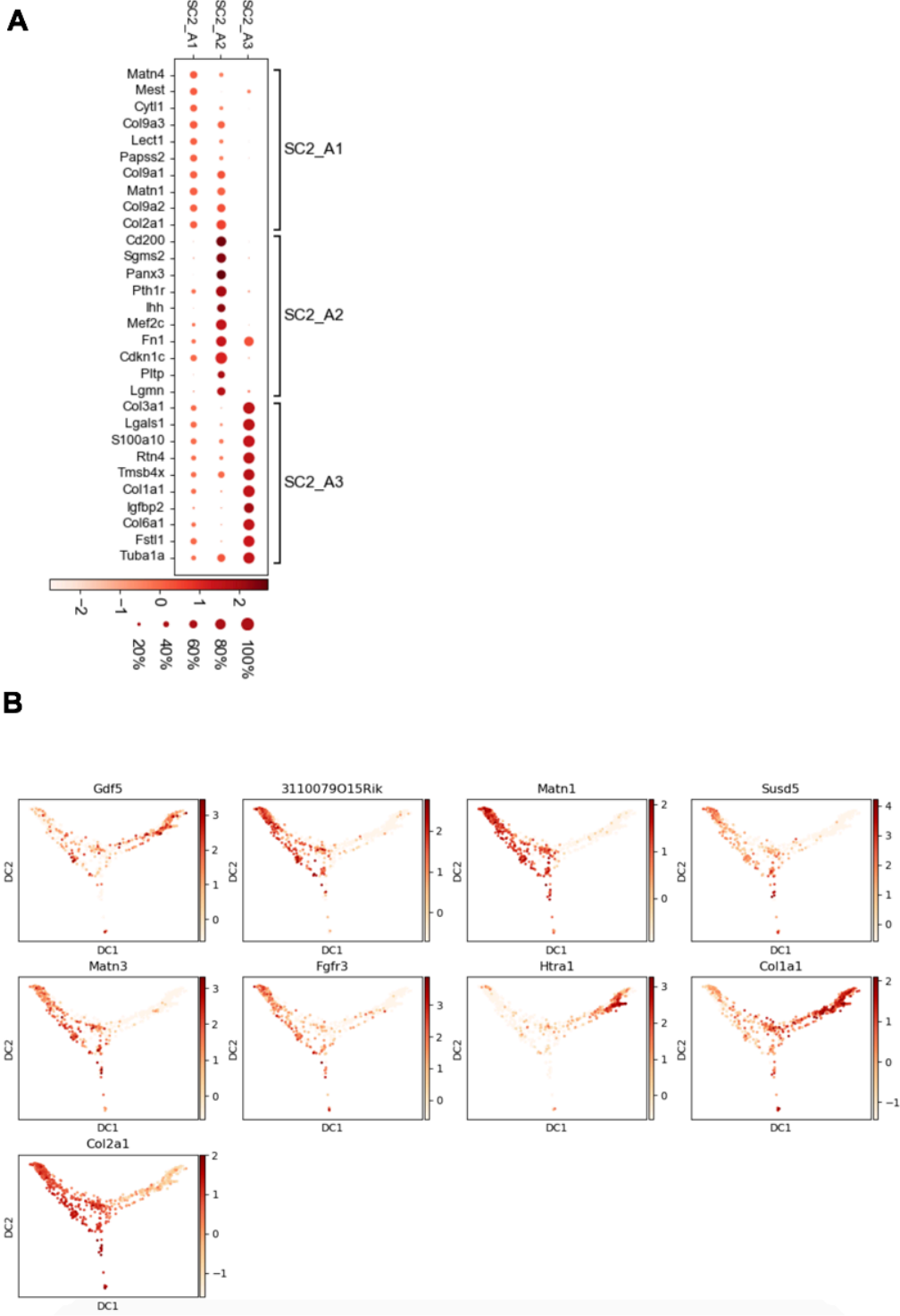


**Supplemental Figure 2**



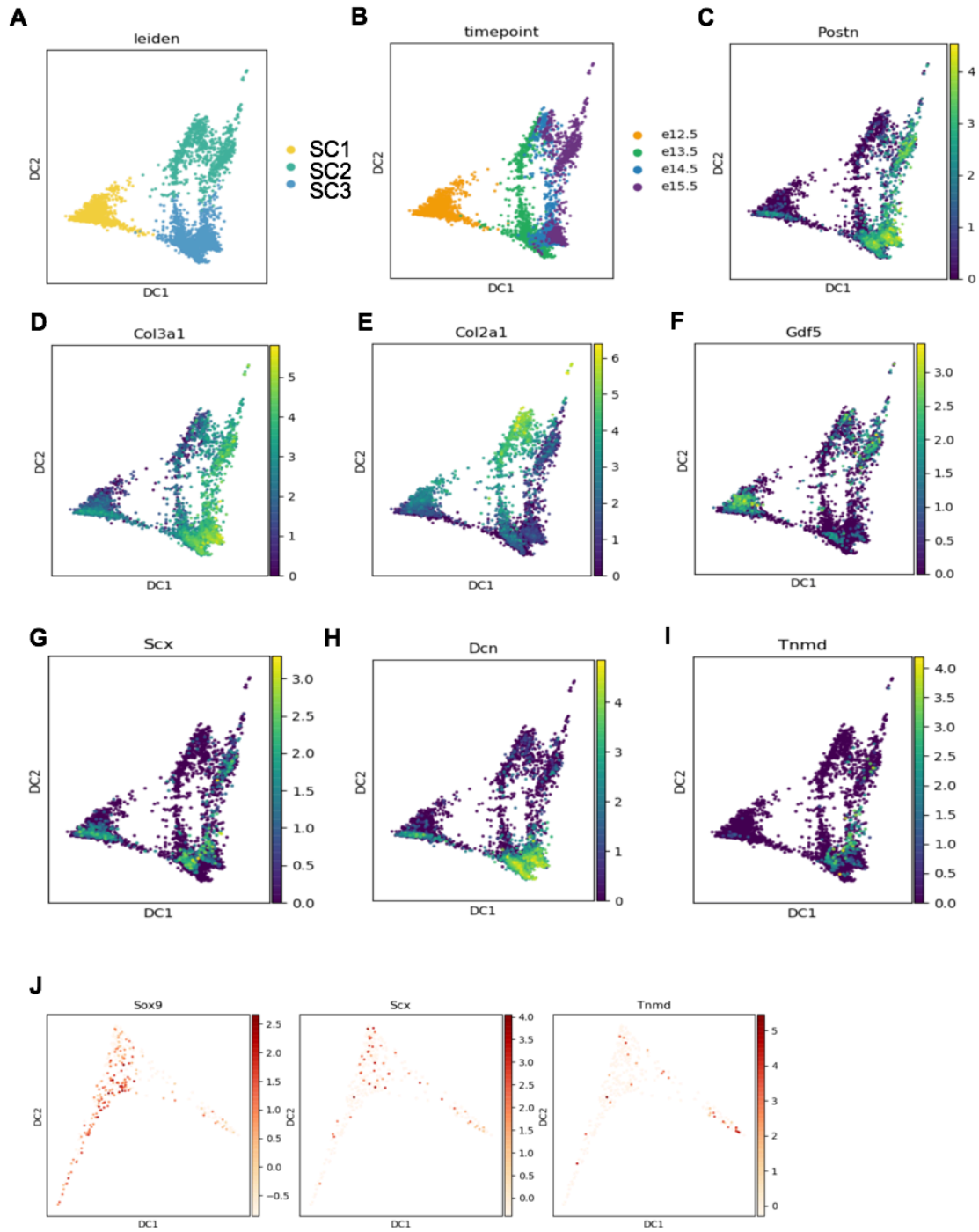
**Supplemental Figure 2:** Initial clustering and identification of non-joint cells and clusters of cells defined by stage of cell cycle.

**Supplementary Figure 3**



**Supplemental Figure 3:** Expression distribution of outer IZ and intermediate IZ representative genes of SC2.

### Supplementary Figure 4



**Supplemental Figure 4:** (A, B) Leiden clustering of SCs, colored by groups (A) and timepoints (B). (C-I) Expression distribution of representative genes on 3 of SCs. (J) Typical tendon/ligament developmental marker genes expression on SC3\_A3 and SC3\_B3 arranged by pseudotime.

Supplemental Table 5

## Reagent Source

REAGENT	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Anit-GFP antibody (Rabbit polyclonal to GFP)	Abcam	Ab6556
Anti-TNMD antibody (Rabbit polyclonal to TNMD)	Abcam	Ab203676
Anti-SOX9 antibody (Rabbit monoclal to SOX9)	Abcam	Ab185966
Anti-THY1 antibody (Rabbit monoclal to THY1)	Abcam	Ab3105
Goat anti-rabbit IgG H&L(Alexa Fluor 488)	Abcam	Ab150077
Goat anti-rabbit IgG secondary antibody, Cy3 conjugate	Boster	BA1032
CD140a (PDGFRA) monoclonal antibody, APC	eBioscience	17-1401-81
CD9 monoclonal antibody, eFluor 450	eBioscience	48-0091-82
<b>Chemicals, enzymes, recombinant proteins</b>		
Antibody Diluent	Agilent Dako	S080981-2
Vectashield Antifade Mounting Medium with DAPI	Vector Laboratories	H-1200
Trichrome Stain (Connective Tissue Stain) kit	Abcam	ab150686
Propidium Iodide	Sigma	P4864
GenElute™ Mammalian Total RNA Miniprep Kit	Sigma Aldrich	RTN70
autoMACS™ Rinsing Solution	Miltenyi Biotec	130-091-222
SuperScript™ III Reverse Transcriptase	Invitrogen	18080044
Collagenase D	Roche	11088858001
Extracta™DNA Prep for PCR	QuantaBio	95091-025
DreamTaq Green PCR Master Mix (2X)	Thermo Scientific™	K1081
iTaq Universal SYBR Green Supermix	BIO-RAD Laboratories	1725121
RNase A	Roche	10109142001

<b>Proteinase K</b>	Sigma-Aldrich	P2308-25MG
<b>DIG RNA Labeling Mix</b>	Roche	11277073910
<b>SP6 RNA Polymerase</b>	Millipore Sigma	10810274001
<b>BCIP 5-bromo-4-chloro-3-indolyl-phosphate, 4-toluidine salt</b>	Millipore Sigma	11585002001
<b>NBT Substrate powder(nitro-blue tetrazolium chloride)</b>	Thermo Scientific™	34035
<b>DNA, MB-grade from fish sperm</b>	Millipore Sigma	11467140001
<b>tRNA from brewer's yeast</b>	Millipore Sigma	10109525001
<b>MEM <math>\alpha</math>, no nucleosides</b>	Gibco™	12561056
<b>Smooth Muscle Cell Differentiation Medium</b>	Sigma-Aldrich	311D-250
<b>Synoviocyte Growth Medium</b>	Sigma-Aldrich	415F-500
<b>Horse Serum</b>	Gibco™	16050114
<b>Recombinant Mouse BMP-6 Protein</b>	R&D	6325-BM-020
<b>Chicken Embryo Extract Powder</b>	Gemini	100-163P
<b>TGF Beta 3</b>	Lonza	PT-4124

### Supplemental Table 6

#### Oligonucleotides for PCR amplification of templates for antisense RNA probes

Gene Symbol	Forward Primer	Reverse Primer	Probe size (bp)
<i>Osr1</i>	AAGCGTCAGAAGTCTAGTTCG	GCTTCTTTTCTGGGGATAGCTT	634
<i>Col2a1</i>	GTCCTACTGGAGTGACTGGTCC	CCAGATTCTCCTTTGTCACCTC	738
<i>Irx1</i>	ATGTCCTTCCCGCAG	TCAGGCAGACGGGAG	1444
<i>Sfrp2</i>	AGCAACTGCAAGCCCATC	ATGGAGAGAAGCCACCCC	803
<i>Twist2</i>	CGCCAGGTACATAGACTTCCTC	GTAAAGAACAGGAGTATGCGGG	675
<i>Col9a1</i>	AGAGGCCAGATTGATGCG	CATCAAATCCCCGAGCAC	843
<i>Postn</i>	TTTAGAGCAGCCGCCATC	CTGCAGCTTCAAGGAGGC	811
<i>Col3a1</i>	CTCAGGGTATCAAGGGTGAAAG	AGACTTTTCACCTCCAACCTCCA	739
<i>Sox9</i>	ATGAATCTCCTGGACCCC	TCAGGGTCTGGTGAGCTGTG	1532
<i>Gdf5</i>	GCCTTGTTCTAGTGTTTGGTC	CAGCCCCTGTAATGAACATCTC	899