



**HAL**  
open science

## TATTOO-seq delineates spatial and cell type-specific regulatory programs during limb patterning

Sébastien Bastide, Elad Chomsky, Baptiste Saudemont, Yann Loe-Mie, Sandrine Schmutz, Sophie Novault, Heather Marlow, Amos Tanay, François Spitz

► **To cite this version:**

Sébastien Bastide, Elad Chomsky, Baptiste Saudemont, Yann Loe-Mie, Sandrine Schmutz, et al.. TATTOO-seq delineates spatial and cell type-specific regulatory programs during limb patterning. 2022. pasteur-03663882

**HAL Id: pasteur-03663882**

**<https://pasteur.hal.science/pasteur-03663882>**

Preprint submitted on 10 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



27 **Abstract:**

28 The coordinated differentiation of progenitor cells into specialized cell types and their spatial organization  
29 into distinct domains is central to embryogenesis. Here, we applied a new unbiased spatially resolved  
30 single-cell transcriptomics method to identify the genetic programs that underlie the emergence of  
31 specialized cell types during limb development and their integration in space. We uncovered combinations  
32 of transcription factors whose expression patterns are predominantly associated with cell type specification  
33 or spatial position, enabling the deconvolution of cell fate and position identity. We demonstrate that the  
34 embryonic limb undergoes a complex multi-scale re-organization upon perturbation of one of its spatial  
35 organizing centers, including the loss of specific cell populations, specific alterations in the molecular  
36 identities of other pre-existing cell states and changes in their relative spatial distribution. Altogether, our  
37 study shows how multi-dimensional single-cell and spatially resolved molecular atlases could reveal the  
38 interconnected genetic networks that regulate the intricacies of organogenesis and its reorganization upon  
39 genetic alterations.

## 40 **Introduction**

41 The development of multicellular organisms involves the progressive specialization of cells into distinct  
42 functions according to a precise spatial and temporal blueprint. Genetically encoded programs ensure the  
43 formation of organs and other anatomical structures at the proper location through the coupling spatial  
44 identity and cell fate [1]. The ability of individual cells to acquire and interpret positional information is  
45 therefore central to the proper patterning of the embryo, and the differentiation of its multiple cell types in  
46 correctly positioned organs. Developmental abnormalities can be caused by genetic defects impacting a  
47 differentiation program *per se* [2–4], but they may also result from improper processing of spatial  
48 information [5, 6]. Furthermore, modulation of the position and activity of the different signaling centers  
49 that impose positional information during embryonic development or impairment of the ability of cells to  
50 respond to these cues have been suggested to contribute substantially to evolutionary diversity of  
51 morphologies (e.g. [7–10]). Typically, spatial position depicted as possessing an instructive role on cell  
52 fate. Yet, cell differentiation programs usually proceed in similar ways in cells exposed to different  
53 positional signals, as exemplified by the formation of muscles or bones throughout the body. This implies  
54 that spatial patterning and cell differentiation are not strictly hierarchically related but correspond to  
55 intertwined yet distinct processes.

56 In vertebrates, the development of the limb is a classic example of a developmental process coupling  
57 patterning and differentiation. In the mouse, forelimb buds emerge from the flank of the embryo around  
58 embryonic day E9.5 and a two-dimensional coordinate system is rapidly established by organizing centers  
59 secreting gradients of signaling molecules [11]. The proximal-distal (PD) axis is specified early by a  
60 gradient of distally secreted Apical Ectodermal Ridge (AER)-derived FGFs [12, 13] while the secretion of  
61 SHH from the posteriorly located Zone of Polarizing Activity (ZPA) defines the anterior-posterior (AP)  
62 axis [10, 14]. Within this patterned limb field, migrating somite-derived myoblasts and pools of lateral  
63 plate-derived mesenchymal cells progressively adopt different fates: muscles, cartilage, bone, tendons, and  
64 other types of connective tissue [15–17]. Although decades of genetic studies have identified the signaling  
65 pathways [12, 14, 18] and the transcription factors [19–21] engaged in limb patterning and in the  
66 differentiation of its main cell types, we still lack a comprehensive understanding on how individual cells  
67 define and adapt their cell type specific gene regulatory programs according to their position.

68 The recent development of single-cell transcriptomics has enabled the characterization of the diversity of  
69 cell states present in a tissue and the identification of their associated specific gene arrays [22]. Yet, the loss  
70 of spatial organization resulting from tissue dissociation still represents a major limitation for these  
71 approaches to the study of embryonic patterning. Different methods have been developed to resolve this  
72 conundrum, each with specific strengths and limitations. Unsupervised computational methods can be used  
73 to directly learn the spatial organization of the tissue from the transcriptional organization [23]. While this

74 approach may be successful when the spatial arrangement of cell states is simple, it is not suitable for more  
75 complex systems where multiple cell states reside at a given position. Highly multiplexed single-molecule  
76 RNA-FISH methods (MER-FISH, seqFISH) [24, 25] and *in situ* sequencing methods (FISSEQ, StarMAP)  
77 [26, 27] provide direct spatial and quantitative measurements, but their detection power remains limited to  
78 a couple of thousands of genes simultaneously. Spatially resolved capture of mRNA via barcoded beads  
79 (Slide-seq, Stereo-seq [28, 29]) offers high spatial resolution for tissue sections, but in their current state,  
80 these strategies do not provide single-cell resolution and yield sparse datasets that can impede the detection  
81 of lowly expressed genes. The combination of spatial transcriptomic methods, both high-resolution imaging  
82 and laser capture microdissection, and scRNA-seq has been successfully applied in numerous contexts (e.g.  
83 [30, 31]). However, these approaches remain technically challenging, require specialized equipment and  
84 reagents and may be inconvenient to explore large samples. In the context of developmental biology, the  
85 use of known landmark genes defined by *in situ* hybridization (ISH) experiments can allow to infer *a*  
86 *posteriori* the location of individually sequenced cells as shown by studies in zebrafish [32] and fly [33].  
87 However, such retrospective mapping depends on arbitrary, and sometimes hard-to-defined, detection  
88 thresholds to transform ISH signals obtained from non-linear signal amplification into binary domains of  
89 expression. Furthermore, this type of approach can only be applied when the pattern of expression of  
90 landmark genes is known and stable, which can hamper their use to analyze the consequences of mutations  
91 or other perturbations that are likely to impact the expression of the landmark genes as well. As an  
92 alternative to these methods, we developed TATTOO-seq, a simple approach which allows the  
93 characterization of thousands of cells by single-cell RNA-seq (scRNA-seq), while recording their spatial  
94 position of origin independently of their transcriptome.

95

### 96 **An unbiased strategy for spatially resolved scRNA-seq**

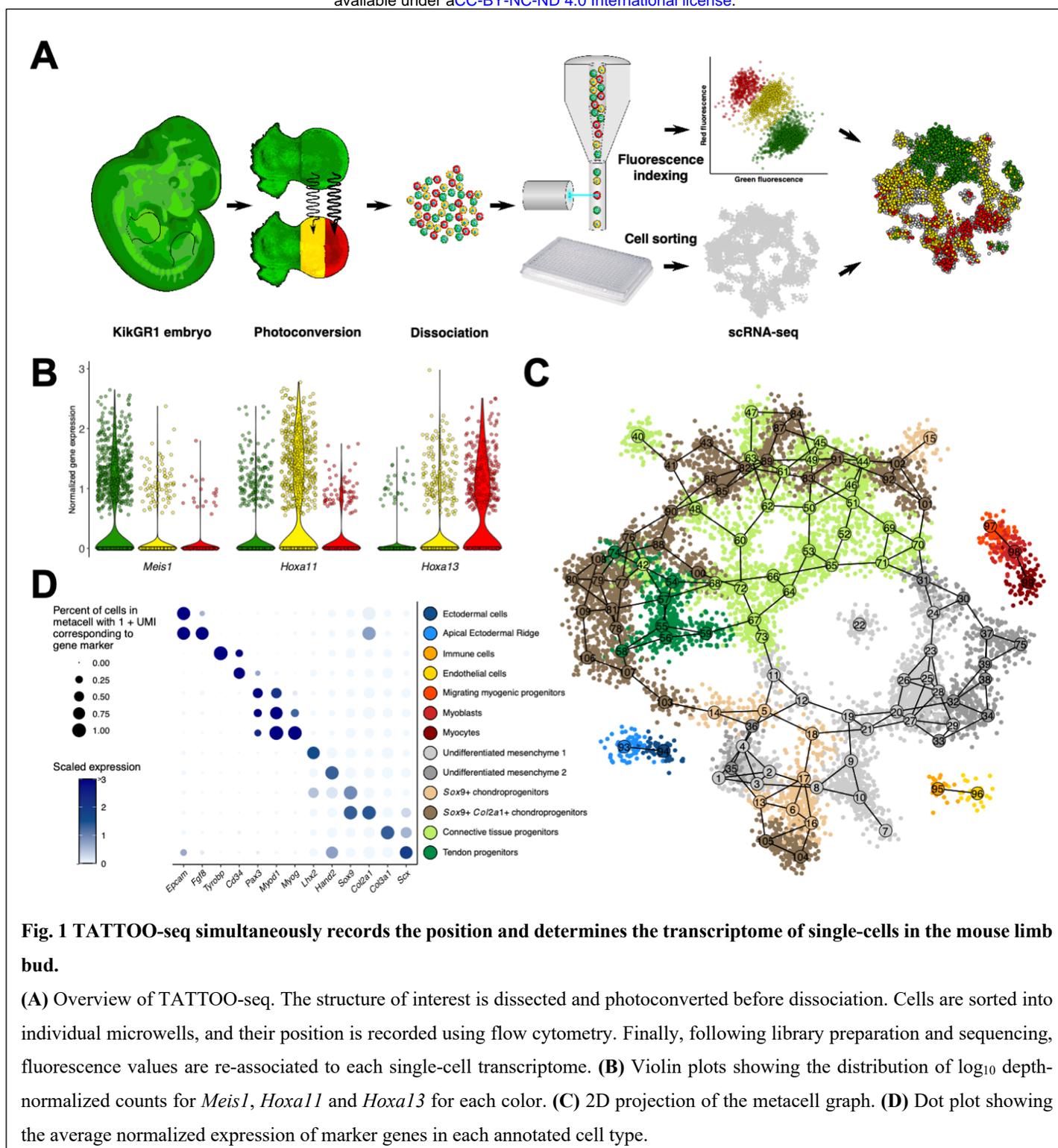
97 TATTOO-seq is based on an optical cell-tattooing system, provided by the constitutively expressed green-  
98 to-red photoconvertible protein KikGR1 [34] (Fig. 1A). Controlling the degree of exposure to 405 nm light  
99 enables to modulate the degree of photoconversion of the pool of KikGR1 proteins, and therefore obtain  
100 distinct levels of red-to-green fluorescence ratios. This property yields an improved flexibility and  
101 resolution over simple photo-activation or photo-bleaching. Instead of simply allowing selective binary  
102 sorting (as utilized for NICHE-seq [35]), we can overlay any defined spatial pattern on the structure of  
103 interest and use the relative fluorescence intensities as a positional index to generate a more complex spatial  
104 coordinate system. Thus, as photoconverted samples are dissociated into single cell suspensions, individual  
105 cells retain their colors. They are then FACS-sorted into 384-well plates. By registering the green and red  
106 fluorescence values of each single cell as they are allocated to specific wells, we can therefore retain direct  
107 information relative to their position of origin. We combined graded photoconversion with Massively

108 Parallel Single-cell RNA-sequencing (MARS-seq) [36], which associates a specific barcode to each  
109 individual cell to parallelly and independently determine single cell transcriptomes and positions of origin.  
110 We then used this strategy on forelimb buds from E11.5 mouse embryos (45-51 somites) and applied  
111 different photoconversion patterns with variable levels of complexity. While we were technically able to  
112 separate four degrees of photoconversion, for simplicity, we used three colors for most patterns. Under the  
113 conditions used, photoconversion was homogeneous throughout the thickness of the limb sample along the  
114 dorsal-ventral (DV). We did not detect significant color-based differences in cell viability, based on flow  
115 cytometry prior to sorting using live and dead cell markers (calcein and Sytox, respectively) (Fig. S1A). To  
116 characterize more in detail the developing limb transcriptomic landscape, we sorted ~10,000 live cells from  
117 differently photo-converted forelimbs (embryos n = 4, limbs n = 7) and performed MARS-seq as previously  
118 described [36]. Cells with fewer than 2000 UMIs were excluded from the analysis. The remaining ~8750  
119 cells had a median UMI count per cell of 5980 and a median number of detected genes of 3115 (Fig. S1C).  
120 We found a high degree of correlation ( $R > 0.97$ ) between samples that were exposed to different degrees  
121 of photo-conversion (Fig. S1B), indicating that photoconversion did not measurably impact gene  
122 expression. To assess the reliability of our approach in registering cell position, we compared the expression  
123 of *Meis1*, *Hoxa11* and *Hoxa13*, which are known markers of the proximal, medial and distal regions of the  
124 limb bud respectively [13], with color-patterns along the PD axis (Fig. 1B). As expected, we found a very  
125 specific overlap between cell position (as measured by color) and marker gene expression, indicating that  
126 TATTOO-seq accurately retains positional information.

127

### 128 **A spatial molecular atlas of the developing limb bud**

129 We used MetaCell [37] to identify transcriptionally homogeneous groups of cells and reconstructed a high-  
130 resolution transcriptional manifold model of the developing limb bud. We identified 109 metacells (Fig. 1C)  
131 and assigned them to the different cell populations present in the limb bud using known cell type specific  
132 marker genes (Fig. 1D). Non-mesenchymal cell states (white blood cells, endothelial cells, and ectodermal  
133 cells) formed well-separated clusters (Fig. S2A). Ectodermal cells cluster into two distinct metacells, one  
134 corresponding to the *Fgf8*-expressing AER and the other to the dorsal and ventral ectoderms (Fig. S2B).  
135 Additionally, we detected three myogenic progenitor cell states corresponding to different stages of  
136 differentiation, from *Pax3*-expressing progenitors migrating from the somites into the limb bud, to *Myod1*-  
137 expressing early differentiating myoblasts and *Myog*-expressing differentiated myocytes [16] (Fig. S2C).  
138 As expected, mesenchymal cell states constituted the bulk of the data and formed a continuum of related  
139 cell states. The chondrogenic lineage was represented by multiple metacells, comprising condensing  
140 chondroprogenitors expressing high levels of *Sox9*, cells engaged in differentiation expressing the  
141 chondrocyte-specific collagen *Col2a1* and a few *Acan*-expressing more mature chondrocytes (Fig. S2D).



142 Interestingly, each of these differentiation stages of chondrogenesis are represented by multiple cell states  
 143 spread throughout the mesenchymal metacell projection (Fig.1C, brown metacells). This dispersion  
 144 indicates that additional components further diversify cell identity on top of the core chondrogenic program.  
 145 Dense regular connective tissue progenitors (connective tissue progenitors hereafter) differentiating into  
 146 tenocytes and ligament fibroblasts show expression for *Col3a1* while more differentiated tendon

147 progenitors express *Scx* (Fig. S2E). The remaining mesenchymal cells correspond to uncommitted cells that  
148 do not express any of the canonical functional markers for chondrogenic or non-chondrogenic connective  
149 tissue. This undifferentiated mesenchyme exhibited a diversity of cell states that could be classified into  
150 two main classes based on the expression of *Lhx2* and *Hand2* (Fig. S2F). Of note, the inspection of dorsal  
151 and ventral markers (Fig. S3) suggested that, at our level of resolution, we do not capture major differences  
152 between cells positioned differentially along the DV axis. This molecular atlas revealed the diversity and  
153 complexity of the cell types and cell-states present in the limb bud. In particular, the data indicated  
154 remarkable variance in TF distribution throughout the model (Fig. S4A), showing a large number of  
155 differentially expressed TFs, even among metacells with similar cell type annotations (Fig. S4B).

156

### 157 **TATTOO-seq integrates multiple spatial axes into one tissue model**

158 To delineate the spatial localization of the different cells and their associated transcriptomic states in the  
159 limb bud, we applied patterns of photoconversion (proximal-distal, anterior-posterior, distance to AER)  
160 which allowed us to capture the dimensionality of the limb as specified by its different organizing centers  
161 (Fig. 2A). The annotation of the metacell graph with this positional information immediately revealed  
162 association of the different transcriptional states to a specific location (Fig.2A). With the predictable  
163 exception of the endothelial and ectodermal metacells, the color compositions of metacells were  
164 significantly more homogeneous than expected by chance (Fig. S5). This indicated that metacells reside at  
165 precise spatial locations which supports their biological relevance as cell states. Moreover, this showed that  
166 cell position is a strong component of mesenchymal cell identity as reflected by the expression of position-  
167 specific transcriptional programs.

168 The positional homogeneity observed within metacells across the different coordinate axes allowed us to  
169 associate each transcriptomic state coherently and confidently to its physical location in the limb bud. Even  
170 though the information provided by a single cell is limited to one axis of our coordinate system, metacells  
171 comprise cells originating from each of the three photoconversion axes. Since the abundance and location  
172 of the different cell type across the limb bud were highly similar between embryos of corresponding  
173 developmental stage, we could use this collective information to assign each metacell to a specific position  
174 in our two-dimensional coordinate system. For this, we calculated the spatial probability distribution that  
175 maximized the likelihood of observing the color composition of the metacells and position metacells into a  
176 more complex 14-bin spatial grid (Fig. 2B, see **methods**). Because our spatial model is two-dimensional,  
177 the area-normalized fraction of cells in a given bin represented an estimate of the thickness of the limb-bud  
178 at each spatial bin, assuming that the mean cell size is similar across all bins. As expected, the inferred limb  
179 bud thickness decreases as we move away from the embryo's trunk and towards the edge of the limb bud.  
180 Transcriptionally close metacells showed markedly different spatial distribution on this 2D limb model,

181 thereby providing an orthogonal validation of the pertinence of our metacells and demonstrates the ability  
182 of our method to infer positions of small cell populations.

183

### 184 **TATTOO-seq characterizes the spatial patterning of the limb bud**

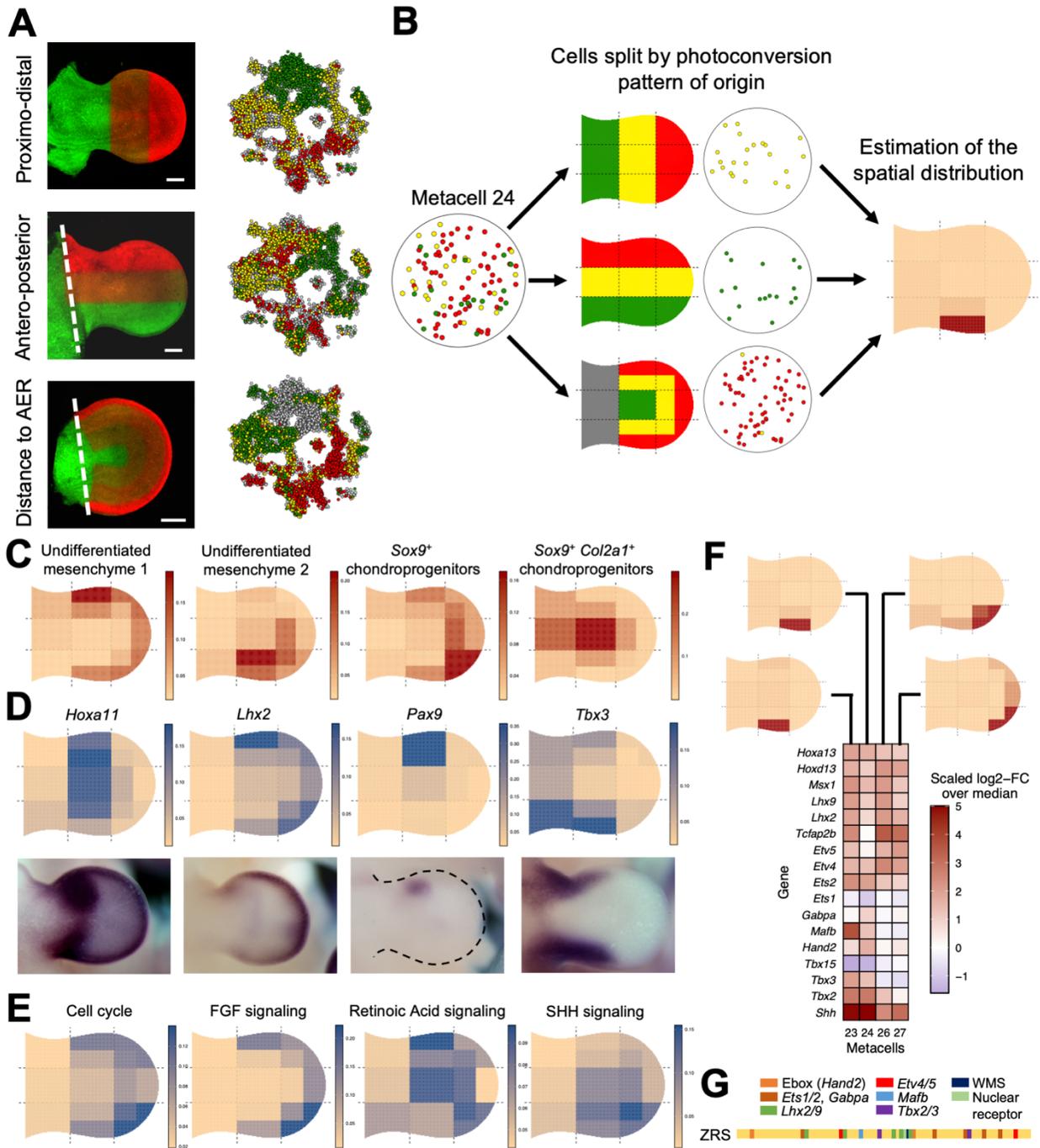
185 The possibility to assign each metacells to a specific spatial position allowed for novel ways to describe  
186 and investigate limb patterning. For instance, we determined the spatial distributions of cell types, by  
187 aggregating the spatial distributions of individual metacells assigned to the same cell type (Fig. 2C). With  
188 this approach, we showed that the two types of undifferentiated mesenchyme were located at different  
189 positions: mesenchyme 1 corresponded to cells present in the progress zone, while mesenchyme 2 cells  
190 were located in the zeugopodial posterior region consistent with their strong *Hand2* expression.

191 To allow the exploration of gene expression with greater precision and resolution, we developed several  
192 tools allowing to infer gene expression spatial patterns for the TATTOO-Seq spatial model. For a given  
193 gene, we can obtain a virtual *in situ* hybridization pattern (vISH) by computing the probability distribution  
194 over space for a random UMI of that gene to be found in each spatial bin. Remarkably, patterns obtained  
195 using vISH closely resembled that of the actual *in situ* hybridization experiments (Fig. 2D, Fig. S6). To  
196 facilitate exploration of limb gene expression pattern, we provide a pre-computed atlas of vISH for more  
197 than 17,000 mouse genes is available at [[http://nobelmarlowlab.uchicago.edu:8888/TATTOOseq\\_vISH/](http://nobelmarlowlab.uchicago.edu:8888/TATTOOseq_vISH/)].  
198 We further expanded vISH to examine the spatial distribution of integrated pathways or biological process,  
199 using the geometric mean of the counts of process-specific signature genes (plus a small regularization  
200 constant) as a representation of these processes. With this approach, one can display on the limb blueprint  
201 the position of zones of cell proliferation or signaling pathways responsive-cells (Fig. 2E).

202 The combined spatial components of this atlas can provide further ways to visualize quantitative differences  
203 between genes across cell types and space. We can use vISHs to directly compare the expression of two or  
204 more genes (Fig. S7A) or even break down gene expression patterns by cell type or cell states. (Fig. S7B).  
205 This cell type-resolved vISH showed, for example, that the proximal anterior and distal expression domain  
206 of *Msx2* expression correspond to two vastly different cell types (undifferentiated mesenchyme and drCT  
207 progenitors, respectively) and highlight two underlying modes of regulation for *Msx2* [38]. Similarly, the  
208 expression of *Asb4* could be decomposed in anterior, central, and proximal cell type-specific contiguous  
209 domains.

210 Interestingly, this approach revealed diverse transcriptomic states in cell populations usually considered as  
211 homogeneous. The Zone of Polarizing Activity (ZPA), a critical signaling center which organizes limb  
212 patterning along the AP axis and influences its outgrowth by maintaining AER-derived signaling [1, 14] is  
213 located at the distal posterior margin of the limb bud and is molecularly defined by the expression of *Shh*.

214



**Fig. 2 Integrating TATTOO-seq data to build a spatial map of cell states and genes expression.**

(A) Maximum intensity projection of z-stack images of each of the three photoconversion patterns: along the PD axis, along the AP axis and as a function of the distance to the AER. (B) Assembling the information from different patterns of photoconversion to infer the position of cell states. Three photoconversion patterns are combined into a 14-bin grid. (C) Spatial distribution for some aggregated mesenchymal cell types. (D) Virtual *in situ* hybridization (vISH) obtained *ab initio* from TATTOO-seq data. vISH accurately reproduces known ISH patterns. Images were obtained from the EMBRYOS database (38). (E) Spatial projection of summarized gene expression for relevant gene sets: cell cycle genes, FGF signaling downstream targets (*Etv4/5*, *Dusp6*, *Cyp26b1*, *Spry4*, *Il17rd/Sef*), SHH downstream targets (39) and direct RA downstream targets (category 3 from (40)). (F) Spatial and transcriptomic dissection of the *Shh*-positive ZPA. Heatmap showing the normalized expression of TFs that have been suggested to bind to the ZRS in the four ZPA metacells. Their positions are display at the top. (G) Schema of the ZRS enhancer regulating *Shh* and its cohort of TF binding sites

216 We found that four different metacells expressed robustly *Shh* (23, 24, 26 and 27) and therefore represent  
217 the ZPA (Fig. 2F). While collectively the cells associated with these metacells mapped to ZPA location at  
218 the posterior distal quadrant of the limb, the different metacells showed distinct distribution, further  
219 supporting the idea that they represent actual functional domains as opposed to clustering artefacts:  
220 metacells 23 and 24 were located medially whereas metacells 26 and 27 are located more distally. Several  
221 TFs showed differential expression between these metacells, including *Hand2*, *Hox* and ETS transcription  
222 factors, which have been shown to regulate *Shh* expression [39–41], as well novel ones (Fig. 2F). We noted  
223 the presence of evolutionary conserved potential binding sites for these TFs within the *Shh* limb enhancer  
224 ZRS [7, 42, 43] (Fig. 2H), suggesting that this partition of the ZPA in distinct sub-regions may correspond  
225 to alternative ZRS-associated TFs. This hitherto hidden level of spatial and regulatory diversity in the ZPA  
226 therefore may further contribute to the diversity of limb morphologies resulting from ZRS genetic variants  
227 [44–46] as they may change *Shh* expression modulate its impact on target genes in either the proximal or  
228 distal part of the ZPA [41].

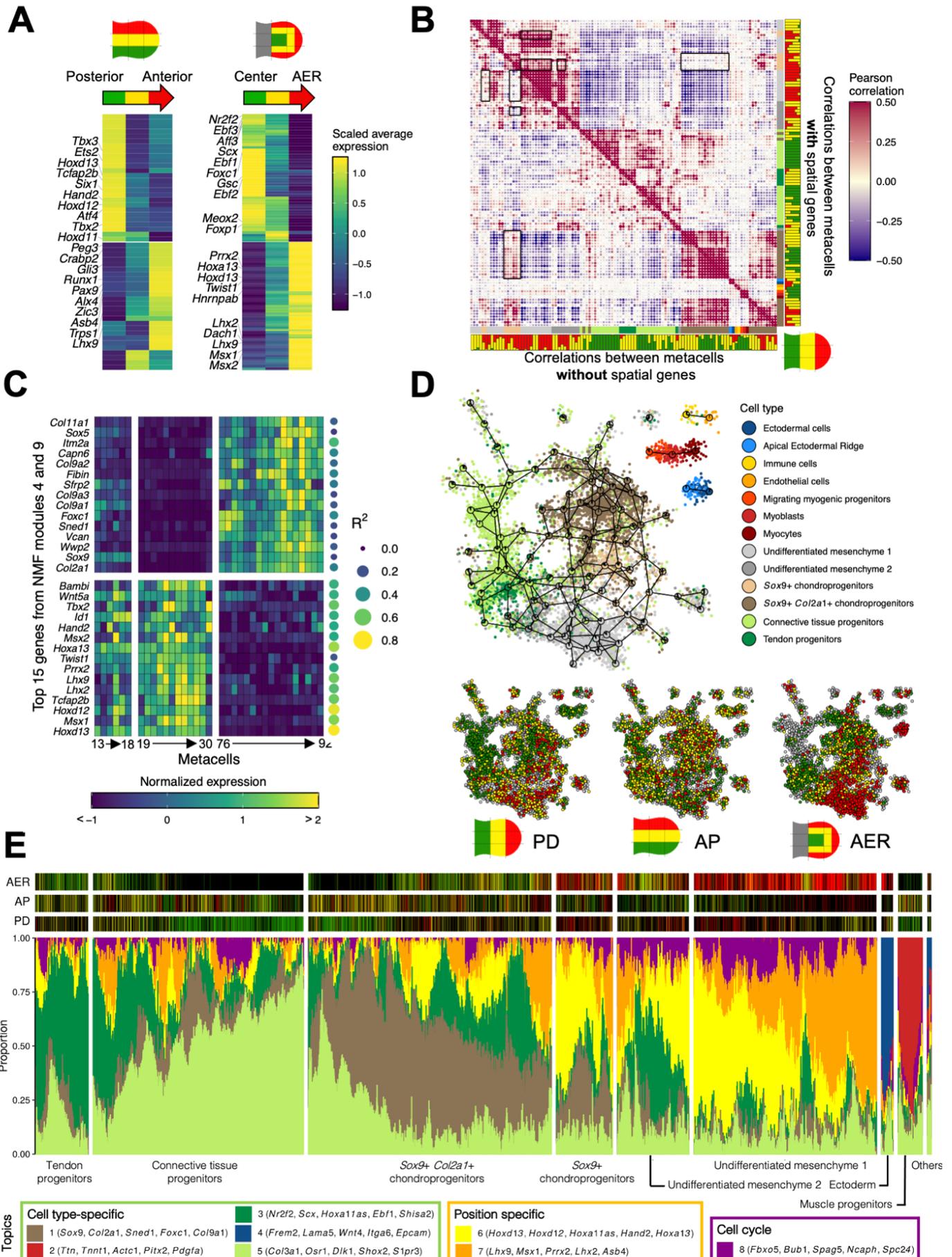
229

### 230 **A transcription factor code of spatial position**

231 We next sought to use the TATTOO-seq atlas to comprehensively identify transcription factors whose  
232 expression exhibits strong spatial trends in the medial and distal mesenchymal cells. Out of 1390 TFs  
233 detected in the autopod and zeugopod mesenchymal metacells (threshold for detection: combined counts  
234 >100 UMIs), 202 showed some variable expression, ie. their expression in at least one metacell is at least  
235 1.5-fold above its median expression across all mesenchymal metacells. We then used Spearman's rank  
236 correlation to assess monotonic trends in gene expression as a function of the distance to the AER and along  
237 the AP axis (Fig. 3A). We uncovered a multitude of TFs whose expressions are correlated with spatial  
238 organization, highlighting the spatial heterogeneity of regulatory states in the limb bud. TFs forming  
239 anterior-to-posterior gradients included previously known genes such as *Alx4*, *Asb4*, and *Zic3* [21], but also  
240 revealed TFs with unknown functions in limb development (*Thra*, *Lmo4*, *Cited2*). 26 TFs formed posterior-  
241 to-anterior gradients of expression, comprising known patterning genes such as the 5' - *Hoxd* genes (*Hoxd10*  
242 to *Hoxd13*, plus *Evx2*) and *Hand2*. 49 TFs were expressed at higher level in mesenchymal cells close to the  
243 AER, including many of the known targets of AER-secreted FGFs and BMPs (*Lhx2/9*, *Msx1/2*, *Etv4*) and  
244 distal *Hox* genes (*Hoxd13*, *Hoxa13*). 104 genes with including genes involved in connective tissue  
245 differentiation (*Sox5/6/9*, *Scx*, *Runx1/3* [18, 47–49]) showed the opposite pattern, with weaker expression  
246 underneath the AER.

247

248



**Fig. 3 Deconvoluting positional information and cell type specific regulatory programs.**

(A) Heatmaps showing the scaled average expression per spatial compartment for TFs that exhibit spatial trends in the mesenchyme. The top 10 most significant TFs for positive and negative Spearman correlation values are annotated. (B) Pairwise correlation heatmap for metacells using all genes defined as HVGs (top) and excluding genes that are defined as spatial (bottom). Cell types and color composition for the PD photoconversion pattern are indicated. (C) Heatmap showing the normalized expression of the top 15 genes (by weight) in NMF modules 4 and 9 across metacells 13 to 30 and 76 to 92. For each gene, the  $R^2$  of the spatial regression is indicated. (D) 2D projection after clustering the TATTOO-seq dataset with no spatial genes. Cells are colored based on their previous annotation and metacells are depicted as pie charts showing the distribution of annotations. The bottom part shows cell color on the 2D projection for the three photoconversion patterns. (E) fastTopic structure plot showing the gradients of topic usage for each cell. The cells are grouped by cell types (endothelial and immune cells are grouped under “Others”). The legend includes the top 5 genes by estimate for each topic. Cell colors for each pattern are represented on top (black if the cell originates from a different pattern of photoconversion).

250 **Deconvoluting spatial patterning and cell differentiation programs**

251 We noticed that several TFs showing spatial trends exhibited very variable expression levels within the  
252 same cell type, leading to the partition of these cell types in distinct cellular states, corresponding to distinct  
253 spatial positions. These observations could be explained by two types of underlying logics regarding how  
254 cells respond to positional cues. Firstly, the differentiation of a generic cell type (e.g chondrocytes,  
255 tenocytes) may be specified at different positions by distinct programs using different sets of TFs.  
256 Alternatively, spatial patterning genes may simply modulate the activity of an otherwise conserved core  
257 cell type-specific program. As an attempt to deconvolute the spatial and cell -type specific regulatory logics  
258 operating in the mesenchyme, we sought to classify genes as primarily carrying “spatial” or “cell type”-  
259 related information. We reasoned that genes encoding spatial information would have a similar spatial  
260 distribution of expression across cell states, while genes regulated in a strictly cell type-specific manner  
261 would show poor predictability from metacell position. We used linear regression to predict gene expression  
262 from the spatial probability distribution of each metacell for all genes and computed overall significance  
263 using F-test. After Bonferroni correction for multiple testing, we retained models for 4713 genes (out of  
264 17537, adjusted p-value < 0.01). Strikingly, these genes include major actors downstream of signaling  
265 pathways during limb development (Table S1). We then computed and compared two pairwise correlation  
266 matrices between metacells: using all highly variable genes (HVGs), previously defined for the clustering,  
267 and using only non-“spatially regulated” HVGs, respectively (Fig. 3B). We noted two types of changes: the  
268 correlation between two groups of distal metacells decreased without spatial HVGs, while the correlation  
269 between one of these groups and proximal chondrogenic metacells increases. To better characterize this  
270 behavior, we performed Non-negative Matrix Factorization (NMF) of the gene-by-metacell expression  
271 matrix in order to identify gene modules that could explain these changes. NMF module 4 exhibits high  
272 weights for known patterning genes with strong spatial regulation (14/15 top genes with  $R^2 > 0.45$ ) while  
273 module 9 exhibits high weights for chondrogenesis genes with overall low spatial regulation (Fig. 3C). Both  
274 sets of genes include several transcription factors. The exclusion of spatially regulated genes to compute

275 correlations decreased the influence of module 4 and therefore the similarity between metacells that simply  
276 reside at the same position and unmasked a high similarity between chondrocytes in the distal and proximal  
277 compartments. The same effect is visible by re-clustering single cells using only non-spatially regulated  
278 genes (Fig. 3D, Fig. S8). We observed that the new graph projection grouped cells together by annotated  
279 cell types, reducing the striking position-driven dispersal of the initial projection (Fig 1C). Correspondingly,  
280 cells from different positions (as shown by their color indexes) were considerably more intermixed (the  
281 original spatial signal is almost abolished for the AP axis).

282 To further investigate the interplay between cell position and cell fate, we used fastTopic [50, 51] to model  
283 each cell as a mixture of different ( $k = 8$ ) topics which allowed to describe cell states as combinations of  
284 more or less independent modules (Fig. 3E). This uncovered a spectrum of combinatorial topic utilization  
285 consistent with the continuum of cell states highlighted by MetaCell. Notably, two anticorrelated topics are  
286 present in cells forming the undifferentiated mesenchyme 1 (topics 6 and 7). Examination of the individual  
287 cell colors shows that the relative importance of these topics followed a medial-anterior to distal-posterior  
288 axis, and the strong correlation of these topics with space. Chondrogenic progenitors are highlighted by the  
289 high prevalence of topic1 (which signature genes comprise many core chondrogenic markers), which  
290 importance did not appear to depend on the position. However, these two spatial topics 6 and 7 are still  
291 present at different levels within chondrogenic progenitors and follow the same general spatial distribution  
292 than in the undifferentiated mesenchyme. Because mesenchymal condensation and chondrogenesis are  
293 initiated as the limb develops along the PD axis, there is a strong but confounding correlation between  
294 differentiation stage and PD position. To compare chondrocyte differentiation dynamics at different  
295 positions, we therefore employed a method similar to that used to reconstruct cell trajectories across time  
296 series during zebrafish development [52], by constructing a single-cell graph of chondrogenesis in a  
297 stepwise manner, using the subspaces defined for groups of cells originating from consecutive spatial  
298 positions (instead of consecutive time points). The resulting graph was then visualized using a force-  
299 directed layout, allowing us to compare the expression of chondrogenesis genes in spite of the confounding  
300 effect of spatial compartmentalization. Interestingly, this projection reflected the heterochrony of  
301 chondrogenesis between the different spatial compartments (Fig. S9A). However, chondrogenic cells at  
302 equivalent differentiation stages did not segregate on the graph by position of origin, indicating that  
303 spatially segregated chondrocytes follow a core and largely identical differentiation trajectory, as further  
304 suggested by the dynamic of marker gene expression analyzed with EIPiGraph [53] (Fig. S9B).

305 These series of analyses show how a comprehensive, spatially resolved atlas can help identifying the logic  
306 and components of the core positional and differentiation programs active in a developing system. It also  
307 suggests that spatial information is not exclusively instructive, but at least in the context of chondrogenesis  
308 in the developing limb bud appear to modulate an otherwise stereotypic differentiation program.

### 309 **Integration of spatial information by position-specific regulatory landscapes**

310 The integration of spatial information into differentiation regulatory programs can be wired at different  
311 levels. *Sox9* is a key regulator of chondrogenesis [54] with a dynamic and complex regulation during limb  
312 development [55], involving a large regulatory landscape extending over 2Mb and comprising dozens of  
313 potential limb enhancers [56, 57]. To understand how *Sox9* expression is regulated in the different limb  
314 compartments, we sought to compare chromatin accessibility at this locus along the limb PD axis. We used  
315 scATAC-seq data from wild-type E11.5 limb buds [58] to identify chromatin-accessible elements in the  
316 large regulatory domain/TAD associated with *Sox9*. We assigned a spatial position and *Sox9* activity level  
317 to scATAC-Seq cells by transferring the annotations from our TATTOO-seq clusters with Seurat/Signac  
318 [59] We then grouped cells based on position (proximal, medial and distal) and *Sox9* expression (*Sox9*<sup>+</sup> and  
319 *Sox9*<sup>-</sup>) and conducted a differential accessibility analysis on these six groups. Despite the limited number  
320 of cells in each category (< 300 in the distal *Sox9*<sup>+</sup> category), we were able to detect 6 ATAC peaks (p-  
321 value < 0.05) in the *Sox9* regulatory domain/TAD which are differentially accessible along the PD axis  
322 (Fig. S10), 5 of which correspond to previously described *Sox9* enhancers [60]. While preliminary, this data  
323 suggests that spatial information is integrated on cell type specific regulatory networks via the usage of  
324 different, position-specific enhancers.

325

### 326 **High-resolution characterization of cell fate and patterning alteration in mutant limbs**

327 Patterning defects can lead to complex phenotypic consequences by perturbing cell differentiation programs  
328 as well as their spatial modulation and coordination. We sought to apply TATTOO-seq to better understand  
329 the impact of a mutation impacting the function of the AER, one of the limb's critical signaling centers. For  
330 this purpose, we used a mouse line (DEL(Poll-SHFM)) that harbors a deletion of the enhancers that control  
331 *Fgf8* expression in the AER [61]. FGF8 is the main FGF molecule controlling the growth and patterning  
332 function of the AER, and accordingly, deletion of *Fgf8* (or the elements that control its expression in the  
333 AER) leads to severe limb malformations [12, 61]. At E11.5, DEL(Poll-SHFM) homozygous mutant  
334 forelimbs are much shorter than those of wild-type mice, display shortened bones and lack anterior skeletal  
335 elements in the forelimb zeugopod and autopod (radius and digits 1-2) as well as the stylopod (humerus)  
336 (Fig. 4A). We used TATTOO-seq as a molecular phenotyping tool to detect how the reduction of FGF  
337 signaling from the AER changes intrinsic cell states and/or their spatial distribution in the growing limb  
338 bud to lead to the observed malformations.

339 We produced a TATTOO-seq atlas for DEL(Poll-SHFM) homozygous limb buds obtained from E11.5  
340 embryos carrying the KikGR1 transgene. Embryonic limb buds were dissected, photoconverted and  
341 dissociated as previously described for the wild-type (embryos n = 3, limbs n = 4). As mutant limbs are  
342 shorter, the AER photoconversion pattern was applied to the whole limb. After QC-based filtering, we

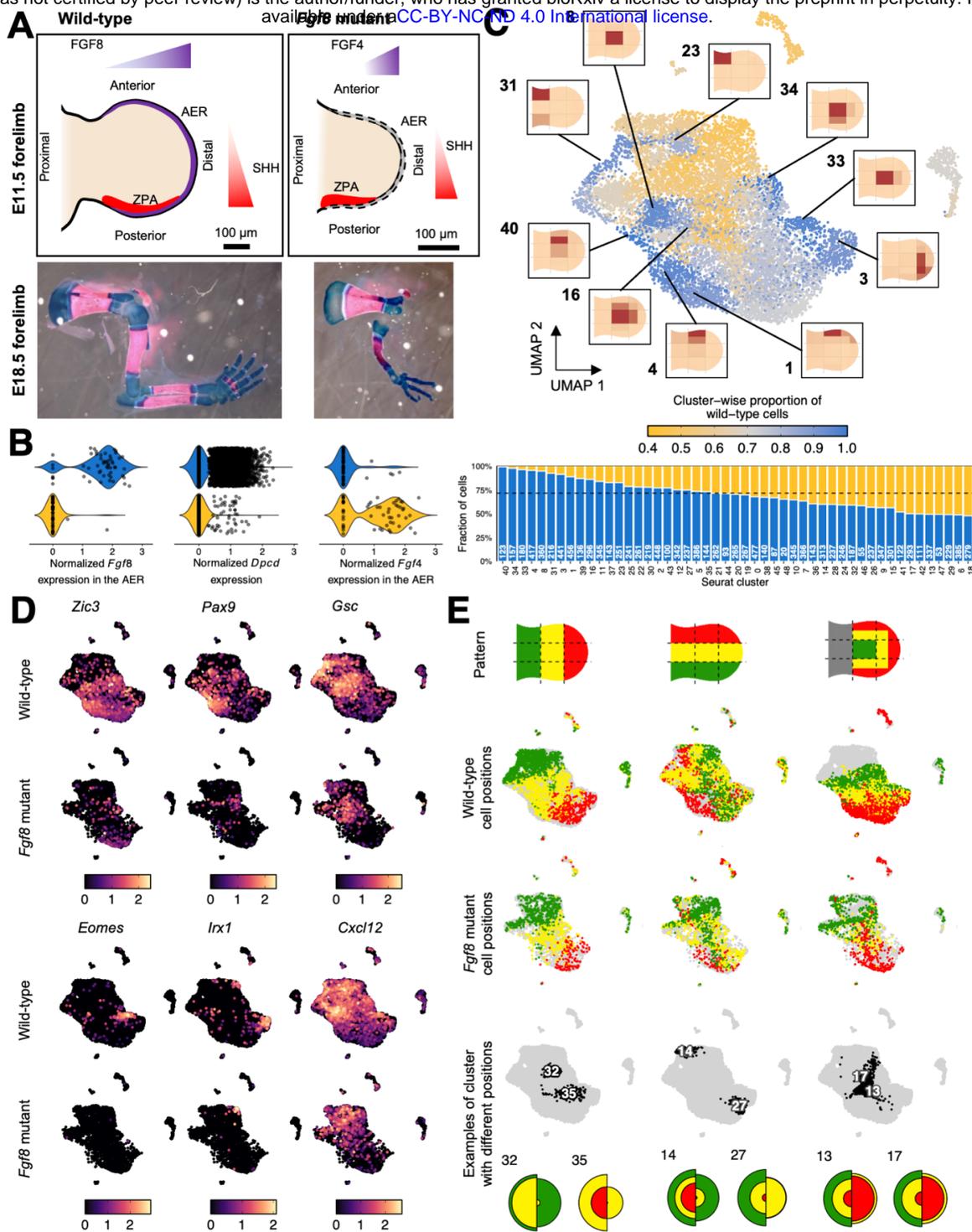
343 obtained high-quality transcriptomes for 3349 mutant cells. Although this number is smaller than that of  
344 the wild-type limb, because of the size difference, it is likely that the cell type “coverage” – represented by  
345 our ability to detect transcriptionally distinct populations – is comparable. Clustering and annotation of the  
346 mutant dataset was done as for the wild-type reference. We first examined the expression of genes known  
347 to be associated directly or indirectly with the mutation and *Fgf8* in the AER. As expected, *Fgf8* expression  
348 was abolished in the ectoderm and *Dpcd*, a bystander gene included in the DEL(PoLL-SHFM) deletion, was  
349 almost undetectable in the mutant cells (Fig. 4B). We could additionally observe changes reported by  
350 previous analyses of the mutant by candidate gene approaches using *in situ* hybridization, including a  
351 compensatory upregulation of *Fgf4* in the AER [12].

352 We then sought to characterize how the reduction of FGF signaling from the AER impacted the different  
353 limb bud cell populations. Given the spatial- and cell type-level resolution offered by TATOO-seq, we  
354 could investigate changes at multiple levels including the correct specification of the different cell types as  
355 well as changes in their transcriptomes and spatial distributions. Considering the smaller size of the mutant  
356 limbs, we first examined whether the differences present in the mutant limbs reflected a global  
357 developmental delay. For this comparison, we generated scRNAs-seq from E10.5 wild-type forelimbs  
358 (embryos n = 4, limbs n = 4; 2912 cells total after QC) and examined several molecular markers indicative  
359 of limb developmental progression in the different datasets (Fig. S11). These comparisons did not yield  
360 evidence suggesting a general heterochrony of limb development in the mutant limbs. To compare the cell  
361 populations present in the wild-type and mutant limbs, we clustered together two datasets, using a coarser  
362 clustering strategy with the Seurat package[59]. For each cluster, we computed the relative fraction of wild-  
363 type and mutant cells (Fig. 4C). Because we slightly biased the sorting strategy of the mutant cells to include  
364 more cells from the presumptive PZ (up to a 2-fold enrichment for red cells close to the AER), we could  
365 not calculate unambiguously the expected cluster-wide fractions under the null hypothesis. Yet, we found  
366 several robust differences that could not be accounted for by the sorting bias.

367 At this level of resolution, we did not detect any mutant-specific cell state. However, several cell types  
368 present in a normal limb bud were almost entirely absent from the mutant samples (Fig. 4C). We used the  
369 metacell clustering of the wild-type TATTOO-seq data to determine the spatial distribution of the individual  
370 cell states corresponding to Seurat clusters depleted in the mutant. Many of these depleted clusters were  
371 located in the anterior compartment of the limb, indicating notably that *Fgf8* AER deletion largely  
372 prevented the specification or growth of a spatially restricted set of cell states (Fig. 4D). The substantial  
373 depletion of these anterior cell states (characterized by the expression of *Zic3* distally, *Pax9* medially and  
374 *Gsc* proximally) provides a putative molecular basis for the absence of anterior skeletal elements at later  
375 stages. Interestingly, we also noted that posteriorly-located cell states (e.g. cluster 3), were strongly reduced  
376 in abundance in the mutant. Furthermore, the mutant exhibits a depletion of some *Cxcl12* and *Fgf18*-

377 expressing cell states who play important roles as secondary signaling centers for connective tissue  
378 differentiation [62, 63]. While most *Sox9*-expressing cell states were shared between the wild-type and the  
379 mutant, distal chondrogenic cell states expressing *Eomes* (base of digit 4 ) and *Irx1* (tip of digits 2-4 [65])  
380 were greatly reduced in the mutant (Fig. 4C and 4D). This shows that, although chondrogenesis can be  
381 initiated in an environment where FGF signaling is considerably diminished (possibly through alternative  
382 sets of *cis*-regulatory elements), spatial refinement of cell states requires genetically-encoded patterning  
383 systems. Interestingly, the depleted distal chondrogenic cells states have been proposed to act as regulation  
384 centers for TGF $\beta$  superfamily-related signaling, expressing the joint-specific ligand *Gdf5* [66], the BMP-  
385 induced activin ligand (*Inhba*) [67], and the BMP antagonists *Nog* and *Chrdll1* [68, 69]. Although the  
386 depletion of these cell states is only partial, no expression of these signaling-related genes was detected in  
387 the mutant data (Fig. S12A). Absence of *Gdf5* and *Nog* signalling has been shown to lead to scapulohumeral  
388 or humero-ulnar synostosis [69], and therefore this molecular defect reported by our data could account for  
389 a missing elbow joint in DEL(Poll-Fgf8) animals.

390 To compare the cell types present in both the mutant and wild-type, and identify system-level and cell type-  
391 specific adjustments to altered signaling environment, we first perform genome-wide differential  
392 expression analysis using the broader cell types identified in the wild-type limb atlas, so as to maintain a  
393 sufficient power to detect gene expression changes (Fig. S12B). We found pervasive adaptation of the  
394 transcriptional states of these cell types to the reduction of *FGF* signaling from the AER with changes in  
395 expression levels of multiple genes within “conserved” cell-states (Fig. S12C and Table S2). We then  
396 compared the spatial distribution of these cell-states in the mutant and in the wild-type (Fig. 4E and Fig.  
397 S13). Only a few positional changes were detected along the PD and AER axes and most of the spatial  
398 rearrangements were observed along the AP axis. This is consistent with the lack of anterior mesenchyme  
399 cell state in the mutant leading to mutant cell states exhibiting relatively more anterior colors than their  
400 wild-type counterpart. However, we also identified opposite positional shifts along the AP axis with some  
401 cell states located more posteriorly in the mutant than in the wild-type.



**Fig. 4 Using TATTOO-seq for high content phenotyping of a patterning mutant.**

(A) Top: schematic representation of the E11.5 wild-type and mutant forelimb buds. Bottom: skeletal preparation of wild-type and mutant forelimbs (alcian blue/alizarin red). (B) Violin plots showing the distribution of  $\log_{10}$  depth-normalized counts for *Fgf8*, *Dpcd* and *Fgf4* for each genotype. (C) UMAP of the combined wild-type and mutant datasets. Cells are colored based on the fraction of wild-type cells in their cluster. The bottom part shows the fraction of wild-type cells and the total number of cells for each Seurat cluster. For clusters showing a depletion in mutant cells, a spatial projection of the wild-type cells is shown. (D) UMAP showing  $\log_{10}$  depth-normalized counts for marker genes of the clusters showing a depletion in mutant cells. The data is split by genotype: wild-type (top row) and *Fgf8* mutant (bottom row). (E) UMAP showing cell colors for each photoconversion pattern split by genotype. The right part shows the location of some Seurat clusters and the distribution of colors for each genotype (left: wild-type, right: *Fgf8* mutant).

## 403 Discussion

404 The data and analyses provided here illustrate how the spatially resolved, single-cell level molecular  
405 signatures obtained with TATTOO-seq can offer a comprehensive and high-resolution description of the  
406 dynamic processes involved in embryonic patterning. The use of orthogonal analyses such as NMF, topic  
407 modelling and graph manifold assembly on TATTOO-seq data allowed us to appreciate the respective  
408 contributions of patterning and differentiation-specific gene modules to cell states, identify specific TFs  
409 which mediate spatial identity and constitute a first step in disentangling spatial and cell type specific gene  
410 regulatory programs. We show that positional information significantly modulates the transcriptome of  
411 limb chondrocytes, even though their differentiation appears to be controlled primarily at all positions by  
412 the same core program. Interestingly, this core differentiation program appears to operate at distinct  
413 positions through different, context-specific, modalities. These modalities associate position-specific TFs  
414 or TFs which activity is regulated by local signaling cues to distinct cis-regulatory elements which act as  
415 specific spatial integrators for common differentiation pathways. These spatial integrators are likely to  
416 subtend the molecular etiology of regionalized natural and pathological variations in morphology [70] and  
417 could be primary genetic substrate in the context of evolution for changing body shapes. The possibility to  
418 simultaneously capture cell position and identity with TATTOO-seq does not only enable to identify these  
419 elements, but provide also an integrative and multi-level description of developmental defects, as illustrated  
420 by the complex cellular, molecular and spatial changes we characterized in the mutant *Fgf8* limb buds. We  
421 expect that using such comprehensive, spatially resolved molecular fingerprint of a complex system may  
422 provide a much finer-grained, accurate representation of “mutant phenotypes”, which were so far limited  
423 to the analysis of a handful of markers and morphological features. Such approaches may better reveal how  
424 genetic perturbation altered the states of the different gene regulatory networks that control cell states and  
425 behaviors.

426 By independently integrating spatial position and single-cell transcriptome, TATTOO-seq constitutes a  
427 simple, robust, and flexible approach for characterizing the gene regulatory networks and cell types present  
428 during embryonic patterning and their crosstalk. Since spatial coordinates are not defined by reference-  
429 based retrospective mapping [32, 33] or pseudo-spatial ordering [23], TATTOO-seq neither requires  
430 landmark reference genes nor does it assume the homogeneity of gene expression at a given position, which  
431 are strong limitations of the aforementioned strategies. Instead of inferring individual cell positions, we  
432 labelled cells prior to dissociation based on their position, measured their position, and reconstructed cell  
433 states and their spatial distribution over the limb bud as well as gene expression patterns. TATTOO-seq  
434 shares conceptual similarities to TOMO-seq [71], though with single-cell resolution and with the simplicity  
435 and rapidity of optic labelling over physical sectioning. Because our approach relies on *in silico* cell  
436 aggregates, instead of individual cells, it is suitable for relatively sparse datasets and works without

437 imputation. Although the spatial resolution of this study is relatively coarse (at the order of 50 to 100  $\mu\text{m}$ ,  
438 depending on the quadrant considered), our approach could resolve finer spatial grids as it is only limited  
439 by the precision of the photoconversion and the capacity of flow cytometry to detect graded fluorescence  
440 levels. While a genetically encoded photoconvertible protein is at the basis of our method, the recent  
441 development of photoconvertible and clickable dyes or the use of photoconvertible membrane labelling  
442 dyes [72, 73] may allow further implementation of of TATTOO-seq based approaches in non-transgenic  
443 and non-model organisms or in patient-derived samples.  
444

## 445 **Material and methods**

### 446 **Animals**

447 Transgenic CAG-KikGR-1 mice [34] (KikGR1 hereafter, kind gift of Dr. Alexander Aulehla) were  
448 maintained inbred and genotyped either by PCR using internal primers (p1:  
449 GAAATGAAGATCGAGCTGCGTATGG and p2: CACCCTTCAGCACTCCATCACGCAC) and a  
450 standard program with 65°C annealing temperature or by assessing green fluorescence in distal phalanx  
451 biopsy. DEL(PoIL-SHFM) mice [61] were maintained on an inbred C57BL/6J background.

452 Homozygous KikGR1 males and wild type females were crossed to generate control embryos. For mutant  
453 analysis, KikGR1/+; DEL(PoIL-SHFM)/+ males and DEL(PoIL-SHFM)/+ females were crossed to  
454 generate experimental and control embryos. E0.5 was defined as noon of the day the vaginal plug was  
455 detected. Embryos were collected at E11.5 and dissected in ice cold PBS supplemented with magnesium  
456 and calcium chloride. KikGR1 heterozygous embryos displayed strong and widespread green fluorescence  
457 in all tissues at all observed stages.

### 458 **Photoconversion and imaging**

459 Samples were mounted in PBS supplemented with magnesium and calcium chloride between a glass slide  
460 and a coverslip separated by several layers of adhesive tape. Photoconversion was performed and all  
461 fluorescence images were acquired using a Zeiss inverted confocal microscope (LSM 800) using a 10X  
462 objective. Green and red fluorescence images were obtained by excitation with the 488 nm (1 mW, gain =  
463 580, digital gain = 2) and 561 nm (1 mW, gain = 580, digital gain = 2) laser diodes, respectively. The  
464 spectra of EGFP and mCherry were used for non-photoconverted and photoconverted version of KikGR1,  
465 respectively, and adjusted such that they did not overlap. For green-to-red photoconversion of KikGR1  
466 proteins, the 405 nm laser diode was used with variable power (100% laser power was equal to 5 mW). Full  
467 conversion (red color) was obtained after 4 iterations at 40% power while partial conversion was obtained  
468 after 4 iterations at 10% power (for the proximo-distal pattern), 9% power (for the antero-posterior pattern)  
469 or 10%/6% power (for the AER pattern). For the photoconversion along the PD and AP axes, the size of  
470 the photoconverted regions were determined to be 1/3 of the length of the limb bud (from the tip to the  
471 flank or from the anterior-most part to the posterior-most part, respectively) as measured with the ZEN  
472 software. For the AER axis, the photoconverted regions were determined to be approximately 1/4 of the  
473 radius of the dissected autopods and zeugopods.

### 474 **Tissue dissociation and fluorescence-activated cell sorting**

475 Limb buds were dissected and incubated for 5 min in 270  $\mu$ L of 0.22 Wunsch unit/mL Liberase TM (Roche,  
476 5401119001) in EDTA-free, calcium- and magnesium-free PBS with agitation (900 rpm) at 37°C in a low-  
477 binding microcentrifuge tube (Biozym, 710176). Mechanical disruption was then performed by pipetting

478 though a p200 gelatin-coated tip. The enzyme was inhibited by adding 30  $\mu$ L of 0.5 M EDTA and volume  
479 was adjusted to 1 or 1.5 mL with calcium- and magnesium-free PBS.

480 1  $\mu$ M Calcein Violet AM (Life Technologies, C34858) and 5 nM Sytox Red (Thermo Fisher Scientific,  
481 S34859) from Invitrogen were used as live and dead cell markers for the sorting, respectively. Excitation  
482 and emission spectra were well separated from those of the non-photoconverted and photoconverted  
483 KikGR1. Single-cell suspensions were filtered through a 35  $\mu$ m BD Falcon Cell-Strainer Cap (352235) into  
484 a FACS tube which was maintained at 4°C during the experiment. Sorting was done on a BD FACSAria  
485 III with flow rate the lowest value (“1.0”). A mild spectral compensation was applied to PE-A (red) and  
486 Violet-F-450/40-A (violet) channels from the FITC-A (green) channel.

487 Debris and doublets were excluded from the analysis by using a sequential gating strategy relying on first  
488 SSC-A vs FSC-A followed by FSC-H vs FSC-W. Live cells were selected based on live/dead staining  
489 gating on Violet-F-450/40-A (Calcein Violet AM) versus APC-A (Sytox Red). Only Calcein-positive  
490 Sytox-negative cells were included in the analysis. Cells with a low fluorescence value in both green (FITC-  
491 A) and red (PE-A) channels were also excluded.

#### 492 **Massively Parallel Single-Cell RNA-Seq**

493 Cells were processed by MARS-seq as previously described [36]. Briefly, live single cells were sorted into  
494 384-well capture plates containing lysis solution and barcoded poly(T) reverse-transcription (RT) primers.  
495 Unique Molecular Identifier (UMI) barcodes contain a cell-specific/well-specific label and an 8-bp random  
496 molecular tag (RMT). After evaporation of the lysis buffer, the RT reaction was performed in presence of  
497 ERCC RNA Spike-in (Ambion). Unused RT primers were digested using Exonuclease I (NEB) and wells  
498 were pooled by half plates. Second-strand cDNA synthesis was then performed, and products were *in vitro*  
499 transcribed overnight using T7 polymerase (NEB) for linear amplification. After RNA fragmentation, a  
500 partial Illumina Read1 sequencing adapter containing a plate-specific barcode was single-strand ligated  
501 using T4 RNA ligase I (NEB) and the product was reverse-transcribed. Finally, the product was purified  
502 and PCR-amplified by PCR (14 cycles) with primers containing the Illumina P5-Read1 and P7-Read2  
503 sequences. Concentration of the purified library was assessed with a Qubit fluorometer (Life Technologies)  
504 and mean molecule size was determined with a 2200 TapeStation instrument (Agilent Technologies).  
505 Libraries were pooled and paired-end sequenced using an Illumina NextSeq 500 sequencer at a median  
506 depth of ~45,000 reads per cell. Read1 (R1) was 70 nt long and covered the plate-specific barcode and the  
507 cDNA. Read2 (R2) was 16 nt long and covered the UMI and well-specific label.

#### 508 **Quality check and read mapping**

509 Reads were demultiplexed and count tables were built as described in [36] (scripts are available at  
510 <http://compgenomics.weizmann.ac.il/tanay/>). Cell-specific/well-specific labels and RMTs were extracted.  
511 Reads with low quality (Phred < 27) barcodes were discarded to prevent ambiguous or spurious assignments

512 of reads to cells or unique molecules. R2 reads whose cell-specific or well-specific labels were unknown  
513 or mutated were discarded. R1 reads were mapped to the mouse genome (mm9) using Ensembl gene  
514 annotations (downloaded in July 2017) using Bowtie2 [74].

### 515 **Metacell analysis**

516 We clustered the wild-type E11.5 dataset using MetaCell [37]. Briefly, an ideal metacell is a group of cells  
517 whose expression profiles are statistically equivalent to independent sampling from a single underlying  
518 transcriptional state. This is achieved by first creating a k-NN graph of the individual cells based on  
519 expression similarity and partitioning this graph into a disjoint sets of cells, metacells, which are both small  
520 (~100 cells each) and as close to the above ideal metacell as possible.

521 Because plates were processed in two batches, we filtered out cells with fewer than 2000 or 2500 UMIs  
522 (lower quality transcriptomes) or more than 15000 UMIs (potential doublets). We excluded mitochondrial  
523 genes (annotated with the prefix “mt-”) from the analysis. Variable genes were selected using the  
524 parameters  $T_{vm} = 0.1$ ,  $T_{tot} = 50$  and  $T_{top3} = 3$ . We then excluded cell cycles and histone genes, ribosomal  
525 genes, small nuclear riboprotein genes, some lncRNA (annotated with the suffix “Rik”) and poorly  
526 supported gene models (annotated with the prefixes “RP-”, “Gm” and “AC”) from the variable genes. In  
527 addition, we identified genes that showed a batchy expression. To do so, we performed a first clustering  
528 and, for each metacell with a sufficient number of cells from both batches ( $> 15$ ), we repeatedly (10 times)  
529 randomly sampled and aggregated UMIs from 10 cells (UMIs from individual cells were downsampled to  
530 eliminate the effect of sequencing depth). If the median fold-change for a gene was  $> 1.6$  in at least one  
531 metacell, the gene was considered batchy and blacklisted for feature selection (but not omitted from the  
532 analysis). When excluding spatial genes, we additionally blacklisted genes with adjusted p-value  $< 0.01$  for  
533 the linear regression. A k-NN graph was built using pearson correlation and  $k = 100$  and 500 bootstrap  
534 iterations were performed (0.75 resampling probability). Metacells were built with a minimum size of 20,  
535  $K = 30$  and  $\alpha = 3$ . Inhomogeneous metacells were split and outlier removed ( $Tlfc = 3$ ). Otherwise,  
536 default parameters were used.

### 537 **Identification of cell colors**

538 To infer a single cell color from the intensities of the PE (red, photoconverted) and FITC (green, non-  
539 photoconverted) channels, we examined the ratio between the log of these intensities:

$$540 \quad s = \frac{\log_2 PE}{\log_2 FITC}$$

541 The distribution of this ratio exhibited the expected number of peaks (3 or 4) across all samples, though the  
542 exact location of the peaks and their separation varied. We therefore modeled the distribution of each  
543 individual limb separately as a mixture of skewed normal distributions whose parameters were estimated  
544 using an expectation-maximization algorithm implementer in the R package mixsmsn [75] (“smsn.mix”,

545 family = “Skew.normal”,  $\nu = 3$ ). Three components were used for the PD and AP patterns and four were  
546 used for the AER pattern. The color of each individual cell was determined using the posterior probabilities.  
547 For the AER pattern, we merged the two red-most compartments (high ratio-of-logs values) to increase the  
548 number of cells in that category.

#### 549 **Modeling the TATTOO-seq photoconversion process**

550 We treated the limb-bud as a two-dimensional shape, constructed through concatenation of Bézier curves.  
551 We then assumed that the AP and PD photoconversion patterns split this shape into three regions of equal  
552 length along the relevant axis, whose fluorescence is either mostly green, mostly yellow, or mostly red.  
553 Based on pictures of the photoconverted limb-buds, we estimated that the part of the limb-bud that was  
554 dissected for AER pattern photoconversion roughly matched the yellow and red regions of the PD pattern.  
555 We then approximated the boundaries of the AER photoconversion regions using further concatenations of  
556 Bézier curves, choosing the parameters that best fit the pictures. Note that even though the AER  
557 photoconversion split the limb-bud into four regions, we combined the two red-most regions in our model,  
558 as the number of cells collected that had red-most fluorescence was extremely low.

559 We then created a generative model connecting a cell’s original position within the limb-bud with its  
560 measured fluorescence. We split the modeled limb-bud’s shape into distinct spatial bins (referred to  
561 henceforth as sbins). Given the sbin containing the cell and the pattern that was used for photoconversion,  
562 it is possible to infer a distribution over the cell’s fluorescence being measured as either green, yellow or  
563 red. It is also possible to infer the probability that the cell will be completely missed by the experiment (as  
564 is the case with proximal cells when the AER photoconversion pattern is used).

565 Theoretically, based on the above photoconversion model, it is possible to split the limb-bud into 17 sbins  
566 which set a single, deterministic fluorescence color under all photoconversion patterns. However, using  
567 such sbins to infer cell positions led to extremely poor results, probably due to the failure of such a strictly  
568 deterministic model from handling slight deviations between the model and the experimental reality (for  
569 example, the limb-bud’s shape and the exact dissection line differed slightly between embryos). Instead,  
570 we split the limb-bud into 36 sbins based on a square 6x6 equidistant grid along the AP-PD axes. Within  
571 each such sbin the fluorescence color under the AP and PD colors is deterministically defined, while the  
572 distribution of fluorescence under the AER pattern is assumed to be proportional to the area of the sbin that  
573 is modeled to be photoconverted to green, yellow, and red respectively. We then merged together sbins  
574 whose distributions were highly similar, ending up with 14 sbins.

#### 575 **Inferring cell positions**

576 The fluorescence information provided for each cell by TATTOO-seq only limits its position to one of three  
577 or four broad spatial regions. To gain a finer spatial resolution, we utilized the reproducibility of patterning

578 of the mouse embryo system. This allowed us to pool together spatial information collected across multiple  
579 limb buds and multiple photoconversion patterns and infer a distribution over the limb-bud's sbins.

580 We will mark by  $\lambda$  one of the photoconversion pattern, i.e.:

$$581 \quad \lambda \in \{\text{AER, AP, PD}\}$$

582 Given a subset of cells, we can calculate the regularized, empirical distribution of the measured fluorescence  
583 under each of the pattern  $\lambda$ :

$$584 \quad P_{emp,\lambda}(F)$$

585 We assumed that the subset of cells defines some transcriptional criterion  $M$  (e.g. a cell type) which implies  
586 a distribution over the model's sbins. Then, the sbins containing cells belonging to  $M$  will be distributed  
587 according to:

$$588 \quad P_{\theta,\lambda}(S|M)$$

589 where  $\theta$  is a parameter. Finally, the photoconversion model described in the previous section implies a  
590 distribution over the measured fluorescence given the sbin and the photoconversion patterns:

$$591 \quad P_{\lambda}(F | S = s)$$

592 Given the above, we can calculate the expected fluorescence of cells belonging to  $M$ :

$$593 \quad P_{\theta,\lambda}(F|M) = \sum_s P_{\lambda}(F | S = s) \cdot P_{\theta,\lambda}(S = s | M)$$

594 To infer the correct value of  $\theta$ , we minimize the Kullback-Leibler divergences between the inferred  
595 distributions and the empirical ones:

$$596 \quad \theta = \underset{\theta}{\operatorname{argmin}} \sum_{\lambda} D_{\text{KL}}[P_{\theta,\lambda}(F|M) || P_{emp,\lambda}(F)]$$

597 As the distributions  $P_{\theta,\lambda}(S|M)$  are discrete, the parameter  $\theta$  can be taken to be the complete definition of  
598 the distribution. To solve the minimization problem, we used a constrained interior-point algorithm,  
599 implemented in the `scipy.optimize.minimize()` function (from the `scipy` python package) with a method  
600 parameter of "trust-constr". To ensure convergence to a proper distribution function, the sum of the  
601 probabilities was constrained to 1:

$$602 \quad \sum_s P_{\theta,\lambda}(S = s | M) = 1$$

603 And the value of the probability of each sbin was constrained to a segment which strictly contain the [0,1]  
604 segment:

$$605 \quad -0.00005 \leq P_{\theta,\lambda}(S = s | M) \leq 1$$

606 When calculating the Kullback-Leibler divergences, the individual sbin probabilities were clamped to the  
607 [0.001, 1] segment. Following the convergence of the algorithm, any negative sbin probabilities were set  
608 to 0, and the probabilities renormalized to ensure they sum to 1.

### 609 **Inferring positions of metacells**

610 The cells of a single metacell form a natural candidate for applying the cell positioning algorithm described  
611 in the previous section. We expect that the expression profiles of most of the limb-bud's cell types depend  
612 on spatial signaling fields. We can therefore assume that cells with a highly similar expression pattern (as  
613 the cells of a single metacell are guaranteed to be) are also tightly localized within the limb-bud. The use  
614 of a metacell also provides a natural definition to the transcriptional criterion  $M$ , i.e. the inferred sbin  
615 distribution describes the expected spatial localization of new cells whose expression profile is close to the  
616 metacell's centroid.

### 617 **Metacell specific transcription factors**

618 The metacell-specificity of transcription factors was estimated using a method similar to the "roc" method  
619 in Seurat. For each gene, the empirical CDF was computed based on the rescaled (from 0 to 1) gene  
620 expression in metacells. The area under the curve (AUC) was then computed. Genes that have very specific  
621 patterns of expression (i.e., expression in very few metacells) display large values of AUC because most  
622 metacells exhibit low values of gene expression.

### 623 **Chondrogenic trajectory and pseudotime**

624 Each chondrogenic progenitor was assigned to a compartment along the PD axis based on its color. Cells  
625 originating from other photoconversion patterns, were assigned to the compartment for which their  
626 metacells had the highest probability. We performed PCA on the top 2000 variable genes within each  
627 compartment (thus excluding variation resulting from spatial compartment-specific genes) and built graphs  
628 connecting each cell to its 40 nearest neighbors within its compartment. Projecting medial cells into the  
629 distal PC space, we built a graph between distal and medial cells. Similarly, we built a graph connecting  
630 medial and proximal cells in the medial PC space. Edges were then filtered to keep at most 10 mutual edges.  
631 The resulting graph was projected using the DrL graph layout. ElPigraph [53] was then used to compute  
632 the pseudotime for each cell. Gene expression as a function of the pseudotime was computed and smoothed  
633 using loess regression.

634 **Acknowledgements**

635 We thank Dr. Fujimoto and the RIKEN Laboratory for Animal Resources and Genetic Engineering for  
636 given us access to the KikGR-1 transgenic line (ref. CDB0201T-1), and Dr. Alexander Aulehla (EMBL)  
637 for sharing with us these mice. We acknowledge greatly the contribution of the members of the animal  
638 facilities at the Institut Pasteur. We thank members of the Spitz, Marlow, and Tanay labs for sharing  
639 reagents, ideas, and comments. We also thank Dr. Chloé Moreau for helping with the pilot photoconversion  
640 experiments. We particularly acknowledge the contribution of Dr. Elodie Brient-Litzler and Dr. Jean-  
641 Christophe Olivo-Marin (Institut Pasteur, CiTech) in the development of a single-cell initiative and  
642 technological platform at the Institut Pasteur, which was essential to this project. This project was also  
643 directly supported by the Institut Pasteur, the Region Ile-de-France (program Sesame) and la Fondation  
644 pour la Recherche Medicale (programme Equipe 2018 to F.S.).

## 645 **References**

- 646 1. Wolpert L (1969) Positional information and the spatial pattern of cellular differentiation. *Journal of*  
647 *Theoretical Biology*, 25(1):1–47. [https://doi.org/10.1016/S0022-5193\(69\)80016-0](https://doi.org/10.1016/S0022-5193(69)80016-0)
- 648 2. Lefebvre V, Angelozzi M, Haseeb A (2019) SOX9 in cartilage development and disease. *Current*  
649 *Opinion in Cell Biology*, 61:39–47. <https://doi.org/10.1016/J.CEB.2019.07.008>
- 650 3. Lauderdale JD, Wilensky JS, Oliver ER, Walton DS, Glaser T (2000) 3' deletions cause aniridia by  
651 preventing PAX6 gene expression. *Proceedings of the National Academy of Sciences*, 97(25):13755–  
652 13759. <https://doi.org/10.1073/PNAS.240398797>
- 653 4. Lou Y, Javed A, Hussain S, Colby J, Frederick D, Pratap J, Xie R, Gaur T, Wijnen AJ van, Jones SN,  
654 Stein GS, Lian JB, Stein JL (2009) A Runx2 threshold for the cleidocranial dysplasia phenotype.  
655 *Human Molecular Genetics*, 18(3):556–568. <https://doi.org/10.1093/HMG/DDN383>
- 656 5. Goodman FR (2003) Congenital abnormalities of body patterning: embryology revisited. *The Lancet*,  
657 362(9384):651–662. [https://doi.org/10.1016/S0140-6736\(03\)14187-6](https://doi.org/10.1016/S0140-6736(03)14187-6)
- 658 6. Lettice LA, Hill AE, Devenney PS, Hill RE (2008) Point mutations in a distant sonic hedgehog cis-  
659 regulator generate a variable regulatory output responsible for preaxial polydactyly. *Human Molecular*  
660 *Genetics*, 17(7):978–985. <https://doi.org/10.1093/hmg/ddm370>
- 661 7. Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, Tissières V, Pickle CS,  
662 Plajzer-Frick I, Lee EA, Kato M, Garvin TH, Akiyama JA, Afzal V, Lopez-Rios J, Rubin EM, Dickel  
663 DE, Pennacchio LA, Visel A (2016) Progressive Loss of Function in a Limb Enhancer during Snake  
664 Evolution. *Cell*, 167(3):633–642.e11. <https://doi.org/10.1016/j.cell.2016.09.028>
- 665 8. Towers M (2018) Evolution of antero-posterior patterning of the limb: Insights from the chick. *genesis*,  
666 56(1):e23047. <https://doi.org/10.1002/dvg.23047>
- 667 9. Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ (2004) *Bmp4* and Morphological Variation of  
668 Beaks in Darwin's Finches. *Science*, 305(5689):1462–1465. <https://doi.org/10.1126/science.1098095>
- 669 10. Zuniga A (2015) Next generation limb development and evolution: old questions, new perspectives.  
670 *Development*, 142(22):3810–3820. <https://doi.org/10.1242/dev.125757>
- 671 11. Bénazet JD, Zeller R (2009) Vertebrate limb development: moving from classical morphogen gradients  
672 to an integrated 4-dimensional patterning system. *Cold Spring Harbor perspectives in biology*,  
673 1(4)<https://doi.org/10.1101/cshperspect.a001339>
- 674 12. Lewandoski M, Sun X, Martin GR (2000) Fgf8 signalling from the AER is essential for normal limb  
675 development. *Nature Genetics*, 26(4):460–463. <https://doi.org/10.1038/82609>
- 676 13. Tabin C, Wolpert L (2007) Rethinking the proximodistal axis of the vertebrate limb in the molecular  
677 era. *Genes and Development*, 21(12):1433–1442. <https://doi.org/10.1101/gad.1547407>

- 678 14. Riddle RD, Johnson RL, Laufer E, Tabin C (1993) Sonic hedgehog mediates the polarizing activity of  
679 the ZPA. *Cell*, 75(7):1401–1416. [https://doi.org/10.1016/0092-8674\(93\)90626-2](https://doi.org/10.1016/0092-8674(93)90626-2)
- 680 15. Shimizu H, Yokoyama S, Asahara H (2007) Growth and differentiation of the developing limb bud  
681 from the perspective of chondrogenesis. *Development Growth and Differentiation*, 49(6):449–454.  
682 <https://doi.org/10.1111/j.1440-169X.2007.00945.x>
- 683 16. Buckingham M, Rigby PWJ (2014) Gene Regulatory Networks and Transcriptional Mechanisms that  
684 Control Myogenesis. *Developmental Cell*, 28(3):225–238.  
685 <https://doi.org/10.1016/j.devcel.2013.12.020>
- 686 17. Nassari S, Duprez D, Fournier-Thibault C (2017) Non-myogenic contribution to muscle development  
687 and homeostasis: The role of connective tissues. *Frontiers in Cell and Developmental Biology*,  
688 5(MAR):22. <https://doi.org/10.3389/fcell.2017.00022>
- 689 18. Wang YJ, Belflower RM, Dong YF, Schwarz EM, O’Keefe RJ, Drissi H (2005) Runx1/AML1/Cbfa2  
690 mediates onset of mesenchymal cell differentiation toward chondrogenesis. *Journal of Bone and*  
691 *Mineral Research*, 20(9):1624–1636. <https://doi.org/10.1359/JBMR.050516>
- 692 19. Akiyama H, Kim JE, Nakashima K, Balmes G, Iwai N, Deng JM, Zhang Z, Martin JF, Behringer RR,  
693 Nakamura T, Crombrugge B De (2005) Osteo-chondroprogenitor cells are derived from Sox9  
694 expressing precursors. *Proceedings of the National Academy of Sciences of the United States of*  
695 *America*, 102(41):14665–14670. <https://doi.org/10.1073/pnas.0504750102>
- 696 20. Tzchori I, Day TF, Carolan PJ, Zhao Y, Wassif CA, Li LQ, Lewandoski M, Gorivodsky M, Love PE,  
697 Porter FD, Westphal H, Yang Y (2009) LIM homeobox transcription factors integrate signaling events  
698 that control three-dimensional limb patterning and growth. *Development*, 136(8):1375–1385.  
699 <https://doi.org/10.1242/dev.026476>
- 700 21. Yokoyama S, Furukawa S, Kitada S, Mori M, Saito T, Kawakami K, Belmonte JCI, Kawakami Y, Ito  
701 Y, Sato T, Asahara H (2017) Analysis of transcription factors expressed at the anterior mouse limb bud.  
702 *PLoS ONE*, 12(5)<https://doi.org/10.1371/journal.pone.0175673>
- 703 22. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L,  
704 Steemers FJ, Trapnell C, Shendure J (2019) The single-cell transcriptional landscape of mammalian  
705 organogenesis. *Nature*, 566(7745):496–502. <https://doi.org/10.1038/s41586-019-0969-x>
- 706 23. Nitzan M, Karaïskos N, Friedman N, Rajewsky N (2019) Gene expression cartography. *Nature*,  
707 576(7785):132–137. <https://doi.org/10.1038/s41586-019-1773-3>
- 708 24. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L (2014) Single-cell in situ RNA profiling by  
709 sequential hybridization. *Nature Methods*, 11(4):360–361. <https://doi.org/10.1038/nmeth.2892>

- 710 25. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X (2015) Spatially resolved, highly multiplexed  
711 RNA profiling in single cells. *Science*, 348(6233):aaa6090–aaa6090.  
712 <https://doi.org/10.1126/science.aaa6090>
- 713 26. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, Turczyk BM, Yang JL, Lee HS,  
714 Aach J, Zhang K, Church GM (2015) Fluorescent in situ sequencing (FISSEQ) of RNA for gene  
715 expression profiling in intact cells and tissues. *Nature Protocols*, 10(3):442–458.  
716 <https://doi.org/10.1038/nprot.2014.191>
- 717 27. Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan  
718 C, Liu J, Nolan GP, Bava FA, Deisseroth K (2018) Three-dimensional intact-tissue sequencing of  
719 single-cell transcriptional states. *Science*, 361(6400)<https://doi.org/10.1126/science.aat5691>
- 720 28. Rodrigues SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen  
721 F, Macosko EZ (2019) Slide-seq: A scalable technology for measuring genome-wide expression at high  
722 spatial resolution. *Science*, 363(6434):1463–1467. <https://doi.org/10.1126/science.aaw1219>
- 723 29. Chen A, Liao S, cheng M, Ma K, Wu L, Lai Y, Yang J, Li W, Xu J, Hao S, Chen X, Liu X, Huang X,  
724 Lin F, Tang X, Li Z, Hong Y, Fu D, Jiang Y, Peng J, Liu S, Shen M, Liu C, Li Q, Wang Z, Wang Z,  
725 Huang X, Yuan Y, Volpe G, Ward C, Munoz-Canoves P, Thiery JP, Zhao F, Li M, Kuang H, Wang O,  
726 Lu H, Wang B, Ni M, Zhang W, Mu F, Yin Y, Yang H, Lisby M, Cornall RJ, Uhlen M, Esteban MA,  
727 Li Y, Liu L, Wang J, Xu X (2021) Large field of view-spatially resolved transcriptomics at nanoscale  
728 resolution. *bioRxiv*,
- 729 30. Baccin C, Al-Sabah J, Velten L, Helbling PM, Grünschläger F, Hernández-Malmierca P, Nombela-  
730 Arrieta C, Steinmetz LM, Trumpp A, Haas S (2020) Combined single-cell and spatial transcriptomics  
731 reveal the molecular, cellular and spatial bone marrow niche organization. *Nature Cell Biology*,  
732 22(1):38–48. <https://doi.org/10.1038/s41556-019-0439-6>
- 733 31. Lohoff T, Ghazanfar S, Missarova A, Koulena N, Pierson N, Griffiths J, Bardot E, Eng C, Tyser R,  
734 Argelaguet R, Guibentif C, Srinivas S, Briscoe J, Simons B, Hadjantonakis A, Göttgens B, Reik W,  
735 Nichols J, Cai L, Marioni J (2020) Highly multiplexed spatially resolved gene expression profiling of  
736 mouse organogenesis. *bioRxiv*, :2020.11.20.391896. <https://doi.org/10.1101/2020.11.20.391896>
- 737 32. Satija R, Farrell JA, Gennert D, Schier AF, Regev A (2015) Spatial reconstruction of single-cell gene  
738 expression data. *Nature Biotechnology*, 33(5):495–502. <https://doi.org/10.1038/nbt.3192>
- 739 33. Karaïskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, Kocks C, Rajewsky N, Zinzen RP  
740 (2017) The *Drosophila* embryo at single-cell transcriptome resolution. *Science*, 358(6360):194–199.  
741 <https://doi.org/10.1126/science.aan3235>

- 742 34. Nowotschin S, Hadjantonakis AK (2009) Use of KikGR a photoconvertible green-to-red fluorescent  
743 protein for cell labeling and lineage analysis in ES cells and mouse embryos. *BMC Developmental*  
744 *Biology*, 9(1):49. <https://doi.org/10.1186/1471-213X-9-49>
- 745 35. Medaglia C, Giladi A, Stoler-Barak L, Giovanni M De, Salame TM, Biram A, David E, Li H, Iannacone  
746 M, Shulman Z, Amit I (2017) Spatial reconstruction of immune niches by combining photoactivatable  
747 reporters and scRNA-seq. *Science*, 358(6370):1622–1626. <https://doi.org/10.1126/science.aao4277>
- 748 36. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S,  
749 Tanay A, Amit I (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of  
750 tissues into cell types. *Science*, 343(6172):776–779. <https://doi.org/10.1126/science.1247651>
- 751 37. Baran Y, Bercovich A, Sebe-Pedros A, Lubling Y, Giladi A, Chomsky E, Meir Z, Hoichman M, Lifshitz  
752 A, Tanay A (2019) MetaCell: Analysis of single-cell RNA-seq data using K-nn graph partitions.  
753 *Genome Biology*, 20(1):206. <https://doi.org/10.1186/s13059-019-1812-2>
- 754 38. Brugger SM, Merrill AE, Torres-Vazquez J, Wu N, Ting MC, Cho JYM, Dobias SL, Yi SE, Lyons K,  
755 Bell JR, Arora K, Warrior R, Maxson R (2004) A phylogenetically conserved cis-regulatory module in  
756 the *Msx2* promoter is sufficient for BMP-dependent transcription in murine and *Drosophila* embryos.  
757 *Development*, 131(20):5153–5165. <https://doi.org/10.1242/dev.01390>
- 758 39. Lettice LA, Williamson I, Wiltshire JH, Peluso S, Devenney PS, Hill AE, Essafi A, Hagman J, Mort R,  
759 Grimes G, DeAngelis CL, Hill RE (2012) Opposing Functions of the ETS Factor Family Define *Shh*  
760 Spatial Expression in Limb Buds and Underlie Polydactyly. *Developmental Cell*, 22(2):459–467.  
761 <https://doi.org/10.1016/j.devcel.2011.12.010>
- 762 40. Osterwalder M, Speziale D, Shoukry M, Mohan R, Ivanek R, Kohler M, Beisel C, Wen X, Scales SJ,  
763 Christoffels VM, Visel A, Lopez-Rios J, Zeller R (2014) *HAND2* targets define a network of  
764 transcriptional regulators that compartmentalize the early limb bud mesenchyme. *Developmental Cell*,  
765 31(3):345–357. <https://doi.org/10.1016/j.devcel.2014.09.018>
- 766 41. Lettice LA, Devenney P, Angelis C De, Hill RE (2017) The Conserved Sonic Hedgehog Limb Enhancer  
767 Consists of Discrete Functional Elements that Regulate Precise Spatial Expression. *Cell Reports*,  
768 20(6):1396–1408. <https://doi.org/10.1016/j.celrep.2017.07.037>
- 769 42. Lettice LA, Heaney SJH, Purdie LA, Li L, Beer P de, Oostra BA, Goode D, Elgar G, Hill RE, Graaff E  
770 de (2003) A long-range *Shh* enhancer regulates expression in the developing limb and fin and is  
771 associated with preaxial polydactyly. *Human Molecular Genetics*, 12(14):1725–1735.  
772 <https://doi.org/10.1093/hmg/ddg180>
- 773 43. Sagai T, Hosoya M, Mizushina Y, Tamura M, Shiroishi T (2005) Elimination of a long-range cis-  
774 regulatory module causes complete loss of limb-specific *Shh* expression and truncation of the mouse  
775 limb. *Development*, 132(4):797–803. <https://doi.org/10.1242/dev.01613>

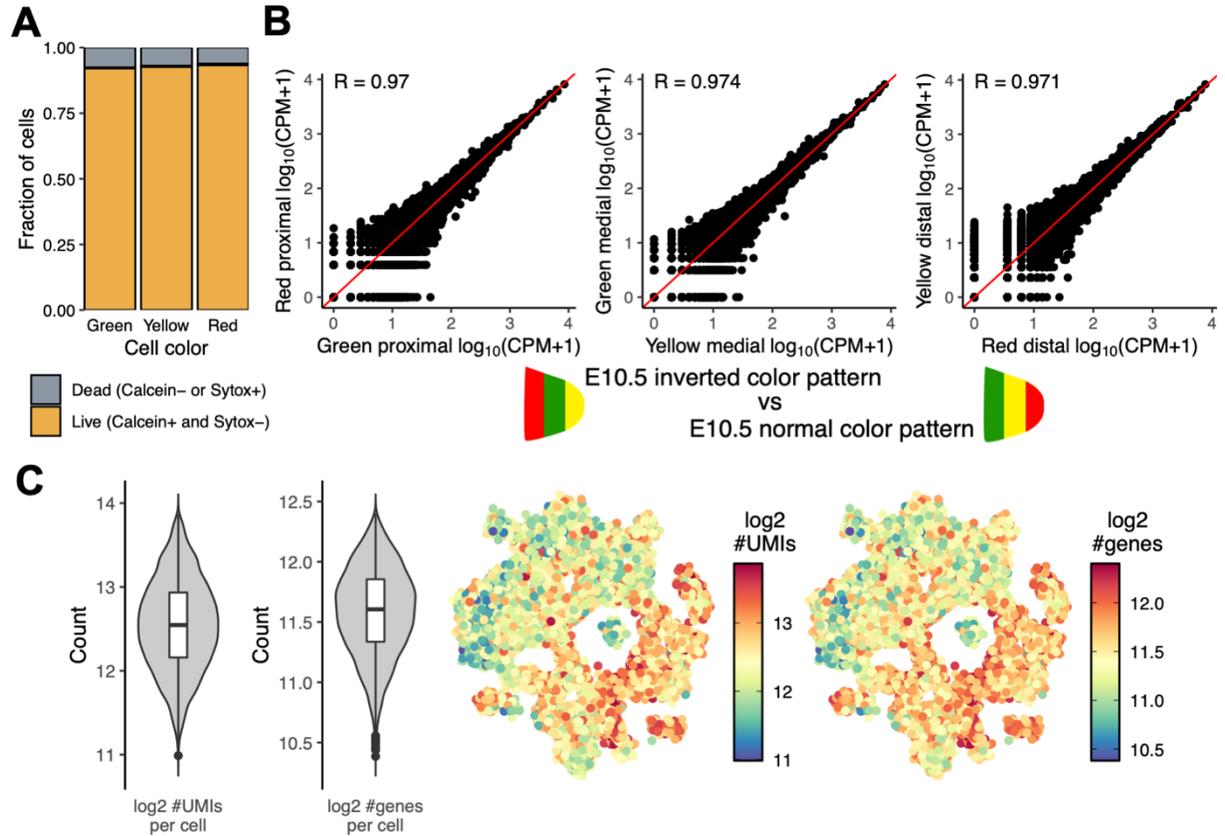
- 776 44. Gurnett CA, Bowcock AM, Dietz FR, Morcuende JA, Murray JC, Dobbs MB (2007) Two novel point  
777 mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial  
778 polydactyly. *American Journal of Medical Genetics, Part A*, 143(1):27–32.  
779 <https://doi.org/10.1002/ajmg.a.31563>
- 780 45. Wiczorek D, Pawlik B, Li Y, Akarsu NA, Caliebe A, May KJW, Schweiger B, Vargas FR, Balci S,  
781 Gillessen-Kaesbach G, Wollnik B (2010) A specific mutation in the distant sonic hedgehog (SHH) cis-  
782 regulator (ZRS) causes Werner Mesomelic Syndrome (WMS) while complete ZRS duplications  
783 underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal  
784 thumb. *Human Mutation*, 31(1):81–89. <https://doi.org/10.1002/humu.21142>
- 785 46. Xu C, Yang X, Zhou H, Li Y, Xing C, Zhou T, Zhong D, Lian C, Yan M, Chen T, Liao Z, Gao B, Su  
786 D, Wang T, Sharma S, Mohan C, Ahituv N, Malik S, Li QZ, Su P (2020) A novel ZRS variant causes  
787 preaxial polydactyly type I by increased sonic hedgehog expression in the developing limb bud.  
788 *Genetics in Medicine*, 22(1):189–198. <https://doi.org/10.1038/s41436-019-0626-7>
- 789 47. Ikeda T, Kawaguchi H, Kamekura S, Ogata N, Mori Y, Nakamura K, Ikegawa S, Chung U Il (2005)  
790 Distinct roles of Sox5, Sox6, and Sox9 in different stages of chondrogenic differentiation. *Journal of*  
791 *Bone and Mineral Metabolism*, 23(5):337–340. <https://doi.org/10.1007/s00774-005-0610-y>
- 792 48. Soung DY, Dong Y, Wang YJ, Zuscik MJ, Schwarz EM, O’Keefe RJ, Drissi H (2007)  
793 Runx3/AML2/Cbfa3 regulates early and late chondrocyte differentiation. *Journal of Bone and Mineral*  
794 *Research*, 22(8):1260–1270. <https://doi.org/10.1359/jbmr.070502>
- 795 49. Schweitzer R, Chyung JH, Murtaugh LC, Brent AE, Rosen V, Olson EN, Lassar A, Tabin CJ (2001)  
796 Analysis of the tendon cell fate using Scleraxis, a specific marker for tendons and ligaments.  
797 *Development*, 128(19):3855–3866.
- 798 50. Dey KK, Hsiao CJ, Stephens M (2017) Visualizing the structure of RNA-seq expression data using  
799 grade of membership models. *PLOS Genetics*, 13(3):e1006599.  
800 <https://doi.org/10.1371/journal.pgen.1006599>
- 801 51. Carbonetto P, Luo K, Dey K, Hsiao J, Stephens M (2021) fastTopics: fast algorithms for fitting topic  
802 models and non-negative matrix factorizations to count data. R package version 0.4-11.
- 803 52. Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM (2018) Single-cell mapping  
804 of gene expression landscapes and lineage in the zebrafish embryo. *Science*, 360(6392):981–987.  
805 <https://doi.org/10.1126/science.aar4362>
- 806 53. Albergante L, Mirkes E, Bac J, Chen H, Martin A, Faure L, Barillot E, Pinello L, Gorban A, Zinovyev  
807 A (2020) Robust and Scalable Learning of Complex Intrinsic Dataset Geometry via EIPiGraph.  
808 *Entropy*, 22(3):296. <https://doi.org/10.3390/e22030296>

- 809 54. Wright E, Hargrave MR, Christiansen J, Cooper L, Kun J, Evans T, Gangadharan U, Greenfield A,  
810 Koopman P (1995) The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos.  
811 *Nature Genetics*, 9(1):15–20. <https://doi.org/10.1038/ng0195-15>
- 812 55. Boehm B, Rautschka M, Quintana L, Raspopovic J, Jan Ž, Sharpe J (2011) A landmark-free  
813 morphometric staging system for the mouse limb bud. *Development*, 138(6):1227–1234.  
814 <https://doi.org/10.1242/dev.057547>
- 815 56. Andrey G, Schöpflin R, Jerković I, Heinrich V, Ibrahim DM, Paliou C, Hochradel M, Timmermann B,  
816 Haas S, Vingron M, Mundlos S (2017) Characterization of hundreds of regulatory landscapes in  
817 developing limbs reveals two regimes of chromatin folding. *Genome Research*, 27(2):223–233.  
818 <https://doi.org/10.1101/gr.213066.116>
- 819 57. Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, Adrian J, Kawli T, Davis CA, Dobin A, Kaul  
820 R, Halow J, Van Nostrand EL, Freese P, Gorkin DU, Shen Y, He Y, Mackiewicz M, Pauli-Behn F,  
821 Williams BA, Mortazavi A, Keller CA, Zhang X-O, Elhajjajy SI, Huey J, Dickel DE, Snetkova V, Wei  
822 X, Wang X, Rivera-Mulia JC, Rozowsky J, Zhang J, Chhetri SB, Zhang J, Victorsen A, White KP,  
823 Visel A, Yeo GW, Burge CB, Lécuyer E, Gilbert DM, Dekker J, Rinn J, Mendenhall EM, Ecker JR,  
824 Kellis M, Klein RJ, Noble WS, Kundaje A, Guigó R, Farnham PJ, Cherry JM, Myers RM, Ren B,  
825 Graveley BR, Gerstein MB, Pennacchio LA, Snyder MP, Bernstein BE, Wold B, Hardison RC,  
826 Gingeras TR, Stamatoyannopoulos JA, Weng Z (2020) Expanded encyclopaedias of DNA elements in  
827 the human and mouse genomes. *Nature*, 583(7818):699–710. <https://doi.org/10.1038/s41586-020-2493-4>
- 829 58. Desanlis I, Kherdjemil Y, Mayran A, Bouklouch Y, Gentile C, Sheth R, Zeller R, Drouin J, Kmita M  
830 (2020) HOX13-dependent chromatin accessibility underlies the transition towards the digit  
831 development program. *Nature Communications*, 11(1):1–10. <https://doi.org/10.1038/s41467-020-16317-2>
- 833 59. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoerckius M, Smibert  
834 P, Satija R (2019) Comprehensive Integration of Single-Cell Data. *Cell*, 177(7):1888-1902.e21.  
835 <https://doi.org/10.1016/j.cell.2019.05.031>
- 836 60. Despang A, Schöpflin R, Franke M, Ali S, Jerković I, Paliou C, Chan WL, Timmermann B, Wittler L,  
837 Vingron M, Mundlos S, Ibrahim DM (2019) Functional dissection of the Sox9–Kcnj2 locus identifies  
838 nonessential and instructive roles of TAD architecture. *Nature Genetics*, 51(8):1263–1271.  
839 <https://doi.org/10.1038/s41588-019-0466-z>
- 840 61. Marinić M, Aktas T, Ruf S, Spitz F (2013) An Integrated Holo-Enhancer Unit Defines Tissue and Gene  
841 Specificity of the Fgf8 Regulatory Landscape. *Developmental Cell*, 24(5):530–542.  
842 <https://doi.org/10.1016/j.devcel.2013.01.025>

- 843 62. Liu Z, Xu J, Colvin JS, Ornitz DM (2002) Coordination of chondrogenesis and osteogenesis by  
844 fibroblast growth factor 18. *Genes and Development*, 16(7):859–869.  
845 <https://doi.org/10.1101/gad.965602>
- 846 63. Nassari S, Blavet C, Bonnin MA, Stricker S, Duprez D, Fournier-Thibault C (2017) The chemokines  
847 CXCL12 and CXCL14 differentially regulate connective tissue markers during limb development.  
848 *Scientific Reports*, 7(1):1–15. <https://doi.org/10.1038/s41598-017-17490-z>
- 849 64. Kwon GS, Hadjantonakis AK (2007) Eomes::GFP - A tool for live imaging cells of the trophoblast,  
850 primitive streak, and telencephalon in the mouse embryo. *Genesis*, 45(4):208–217.  
851 <https://doi.org/10.1002/dvg.20293>
- 852 65. Zülch A, Becker MB, Gruss P (2001) Expression pattern of *Irx1* and *Irx2* during mouse digit  
853 development. *Mechanisms of Development*, 106(1–2):159–162. [https://doi.org/10.1016/S0925-](https://doi.org/10.1016/S0925-4773(01)00411-7)  
854 [4773\(01\)00411-7](https://doi.org/10.1016/S0925-4773(01)00411-7)
- 855 66. Storm EE, Kingsley DM (1999) GDF5 coordinates bone and joint formation during digit development.  
856 *Developmental Biology*, 209(1):11–27. <https://doi.org/10.1006/dbio.1999.9241>
- 857 67. Merino R, Macias D, Gañan Y, Rodriguez-Leon J, Economides AN, Rodriguez-Esteban C, Izpisua-  
858 Belmonte JC, Hurlé JM (1999) Control of digit formation by activin signalling. *Development*,  
859 126(10):2161–2170.
- 860 68. Allen JM, McGlenn E, Hill A, Warman ML (2013) Autopodial development is selectively impaired by  
861 misexpression of chordin-like 1 in the chick limb. *Developmental Biology*, 381(1):159–169.  
862 <https://doi.org/10.1016/j.ydbio.2013.06.003>
- 863 69. Degenkolbe E, König J, Zimmer J, Walther M, Reißner C, Nickel J, Plöger F, Raspopovic J, Sharpe J,  
864 Dathe K, Hecht JT, Mundlos S, Doelken SC, Seemann P (2013) A GDF5 Point Mutation Strikes Twice  
865 - Causing BDA1 and SYNS2. *PLoS Genetics*, 9(10):1003846.  
866 <https://doi.org/10.1371/journal.pgen.1003846>
- 867 70. White JD, Indencleef K, Naqvi S, Eller RJ, Hoskens H, Roosenboom J, Lee MK, Li J, Mohammed J,  
868 Richmond S, Quillen EE, Norton HL, Feingold E, Swigut T, Marazita ML, Peeters H, Hens G, Shaffer  
869 JR, Wysocka J, Walsh S, Weinberg SM, Shriver MD, Claes P (2021) Insights into the genetic  
870 architecture of the human face. *Nature Genetics*, 53(1):45–53. [https://doi.org/10.1038/s41588-020-](https://doi.org/10.1038/s41588-020-00741-7)  
871 [00741-7](https://doi.org/10.1038/s41588-020-00741-7)
- 872 71. Kruse F, Junker JP, Oudenaarden A van, Bakkens J (2016) Tomo-seq: A method to obtain genome-wide  
873 expression data with spatial resolution. *Methods in Cell Biology*, 135:299–307.  
874 <https://doi.org/10.1016/bs.mcb.2016.01.006>
- 875 72. Jun J V., Haney CM, Karpowicz RJ, Giannakoulis S, Lee VMY, Petersson EJ, Chenoweth DM (2019)  
876 A “clickable” Photoconvertible Small Fluorescent Molecule as a Minimalist Probe for Tracking

- 877 Individual Biomolecule Complexes. *Journal of the American Chemical Society*, 141(5):1893–1897.  
878 <https://doi.org/10.1021/jacs.8b13094>
- 879 73. Carlson AL, Fujisaki J, Wu J, Runnels JM, Turcotte R, Celso C Lo, Scadden DT, Strom TB, Lin CP  
880 (2013) Tracking Single Cells in Live Animals Using a Photoconvertible Near-Infrared Cell Membrane  
881 Label. *PLoS ONE*, 8(8):69257. <https://doi.org/10.1371/JOURNAL.PONE.0069257>
- 882 74. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*,  
883 9(4):357–359. <https://doi.org/10.1038/nmeth.1923>
- 884 75. Prates MO, Cabral CRB, Lachos VH (2013) Mixsmsn: Fitting finite mixture of scale mixture of skew-  
885 normal distributions. *Journal of Statistical Software*, 54(12):1–20.  
886 <https://doi.org/10.18637/jss.v054.i12>
- 887 76. Riddle RD, Ensini M, Nelson C, Tsuchida T, Jessell TM, Tabin C (1995) Induction of the LIM  
888 homeobox gene *Lmx1* by WNT6a establishes dorsoventral pattern in the vertebrate limb. *Cell*,  
889 83(4):631–640. [https://doi.org/10.1016/0092-8674\(95\)90103-5](https://doi.org/10.1016/0092-8674(95)90103-5)
- 890 77. Lan Y, Kingsley PD, Cho ES, Jiang R (2001) *Osr2*, a new mouse gene related to *Drosophila odd-*  
891 *skipped*, exhibits dynamic expression patterns during craniofacial, limb, and kidney development.  
892 *Mechanisms of Development*, 107(1–2):175–179. [https://doi.org/10.1016/S0925-4773\(01\)00457-9](https://doi.org/10.1016/S0925-4773(01)00457-9)
- 893 78. Mercader N, Selleri L, Criado LM, Pallares P, Parras C, Cleary ML, Torres M (2009) Ectopic *Meis1*  
894 expression in the mouse limb bud alters P-D patterning in a *Pbx1*-independent manner. *International*  
895 *Journal of Developmental Biology*, 53(8–10):1483–1494. <https://doi.org/10.1387/ijdb.072430nm>
- 896 79. Yokoyama S, Ito Y, Ueno-Kudoh H, Shimizu H, Uchibe K, Albin S, Mitsuoka K, Miyaki S, Kiso M,  
897 Nagai A, Hikata T, Osada T, Fukuda N, Yamashita S, Harada D, Mezzano V, Kasai M, Puri PL,  
898 Hayashizaki Y, Okado H, Hashimoto M, Asahara H (2009) A Systems Approach Reveals that the  
899 Myogenesis Genome Network Is Regulated by the Transcriptional Repressor RP58. *Developmental*  
900 *Cell*, 17(6):836–848. <https://doi.org/10.1016/j.devcel.2009.10.011>
- 901 80. Lu P, Yu Y, Perdue Y, Werb Z (2008) The apical ectodermal ridge is a timer for generating distal limb  
902 progenitors. *Development*, 135(8):1395–1405. <https://doi.org/10.1242/dev.018945>
- 903 81. Mercader N, Selleri L, Criado LM, Pallares P, Parras C, Cleary ML, Torres M (2009) Ectopic *Meis1*  
904 expression in the mouse limb bud alters P-D patterning in a *Pbx1*-independent manner. *International*  
905 *Journal of Developmental Biology*, 53(8–10):1483–1494. <https://doi.org/10.1387/ijdb.072430nm>
- 906 82. Salbaum JM (1998) *Punc*, a novel mouse gene of the immunoglobulin superfamily, is expressed  
907 predominantly in the developing nervous system. *Mechanisms of Development*, 71(1–2):201–204.  
908 [https://doi.org/10.1016/S0925-4773\(98\)00005-7](https://doi.org/10.1016/S0925-4773(98)00005-7)  
909

910

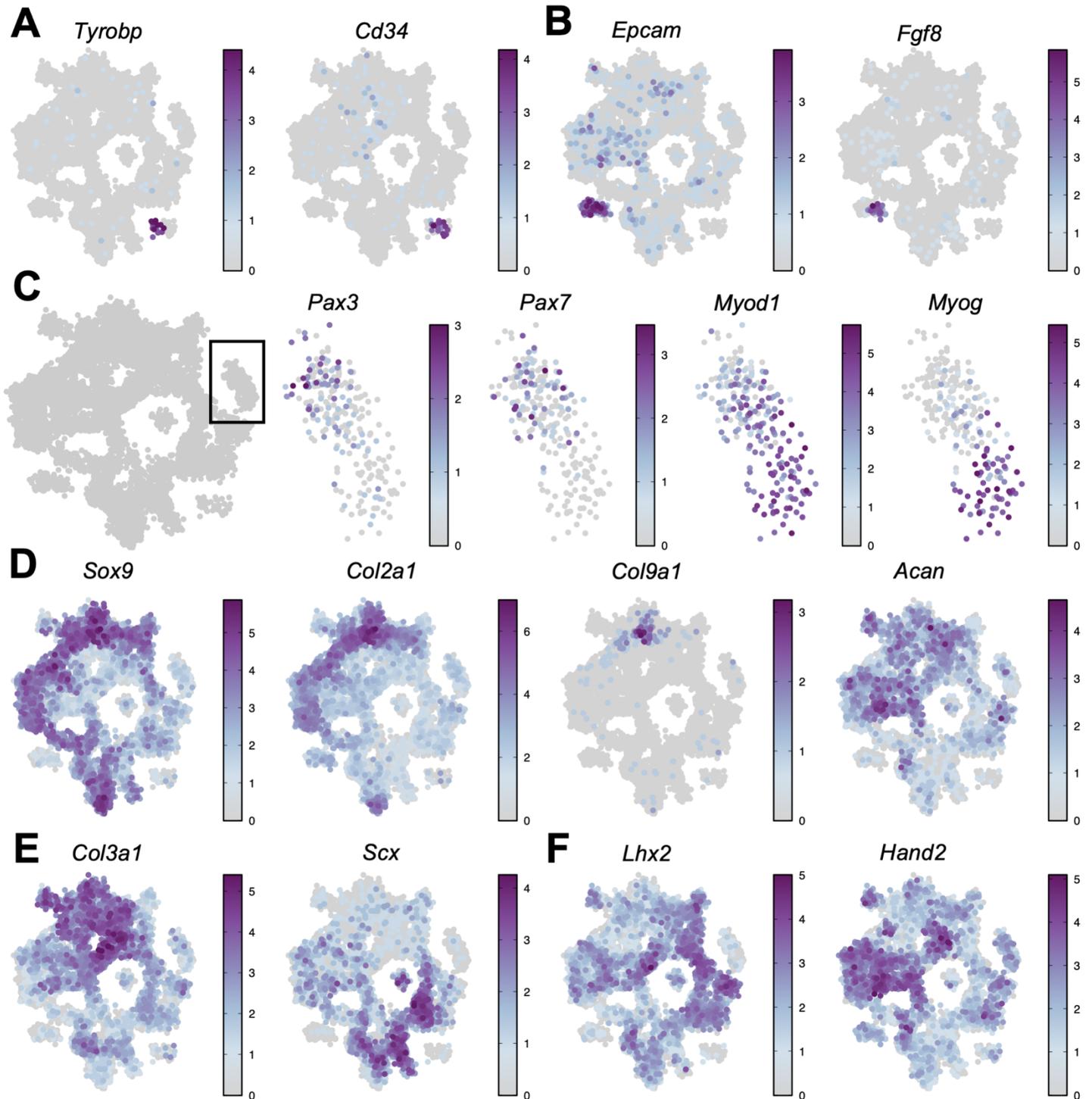


911

912 **Fig. S1 Assessing the impact of the TATTOO-seq method on cell viability and transcriptome and**  
 913 **checking the quality of the data.**

914 **(A)** Fraction of live and dead cells (using Calcein and Sytox) assessed by flow cytometry for each color.  
 915 The fraction of dead cells is independent from the color. **(B)** Comparison of aggregated gene expression for  
 916 different degrees of photoconversion on the same spatial compartment. **(C)** Violin and 2D projection  
 917 showing the distribution of the number of genes and number of UMIs detected per cell. The number of  
 918 detected genes and UMIs is slightly higher in distal cells which are more actively proliferating.

919

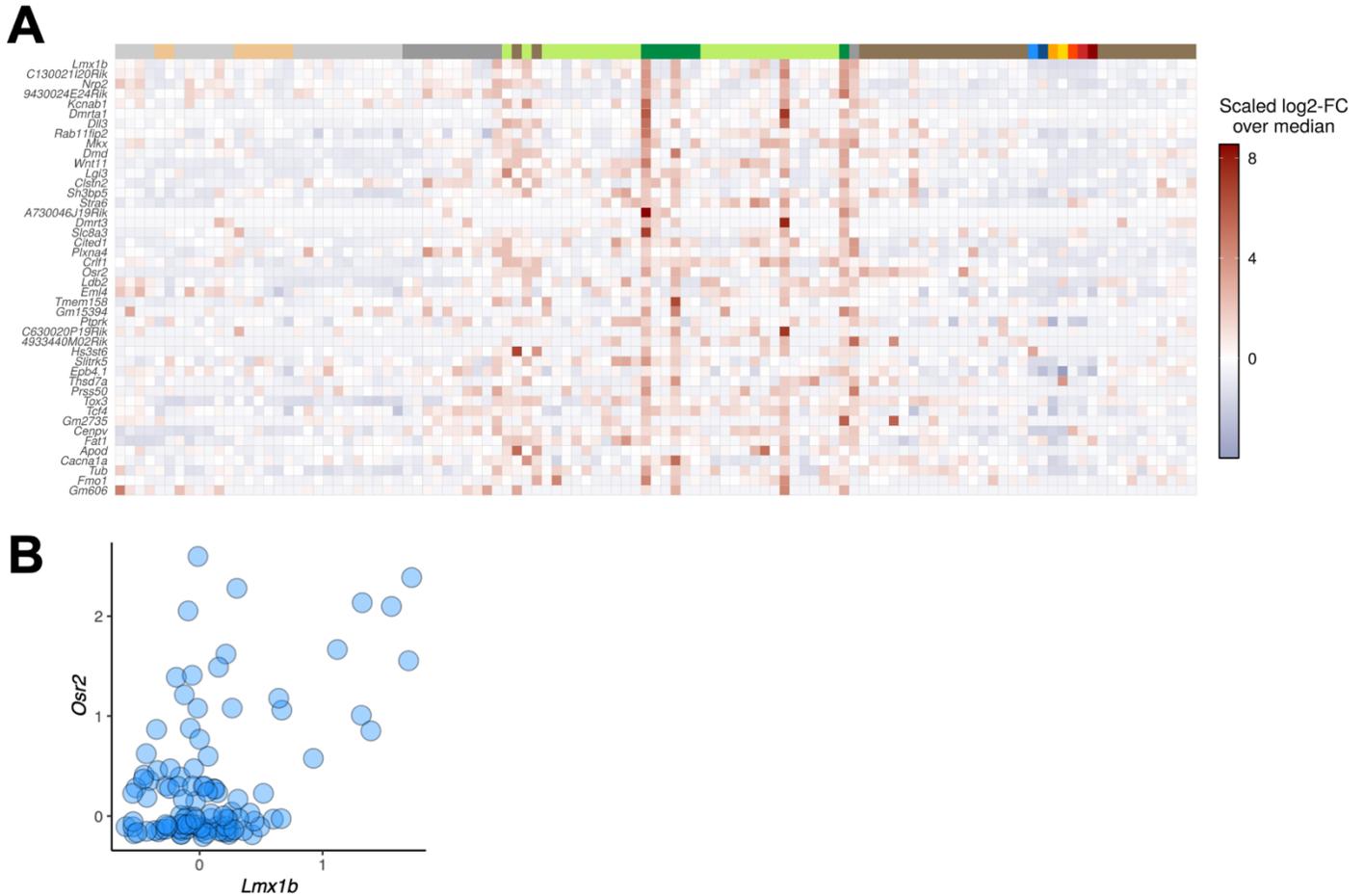


920 **Fig. S2 Expression of cell type specific marker genes.**

921 2D projection of the metacell graph. Cell color represents  $\log_{10}$  depth-normalized counts for marker genes

922 of (A) immune and endothelial cells, (B) epithelial cells and AER cells, (C) stages of muscle progenitor

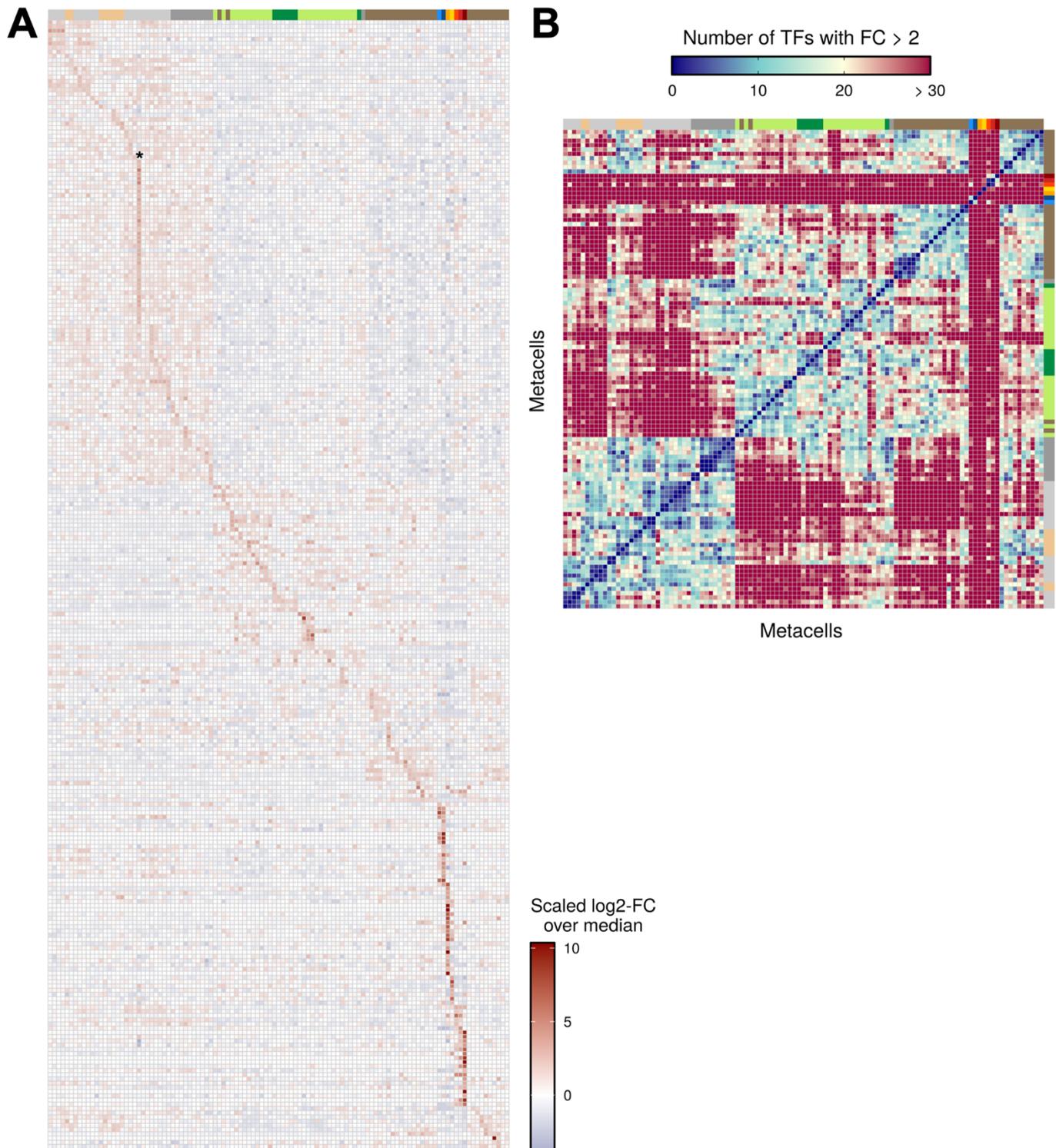
923 differentiation, **(D)** chondrogenic progenitors, **(E)** dense regular connective tissue progenitors and **(F)**  
924 different populations of undifferentiated mesenchyme.  
925



926 **Fig. S3 DV organization of the limb cell states.**

927 **(A)** *Lmx1b* was used as a marker of the dorsal mesenchyme [76]. Heatmap showing the normalized  
928 expression of genes whose expression is loosely correlated or anti-correlated to that of *Lmx1b* ( $|R| > 0.4$ ).

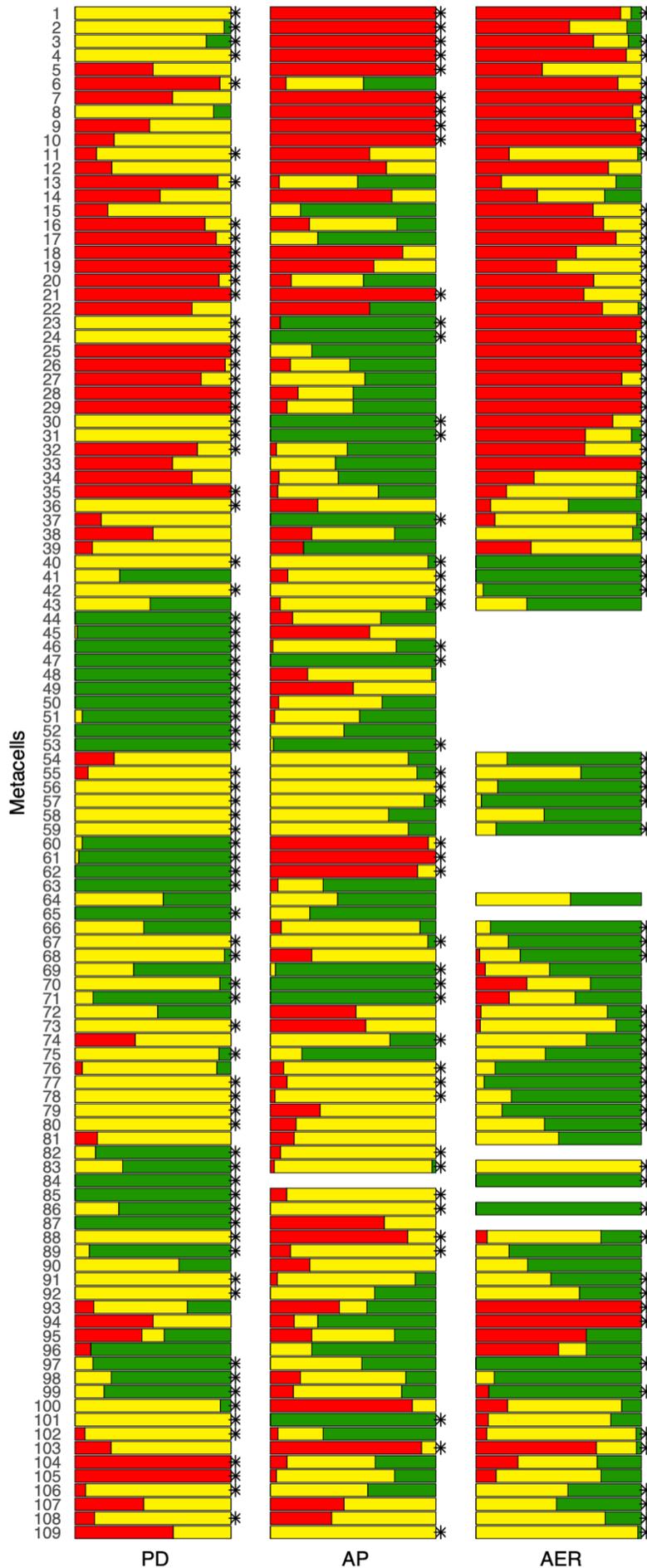
929 **(B)** Scatter plot showing the expression of *Lmx1b* and *Osr2* for all metacells. Metacells exhibiting high  
930 levels of *Lmx1b* showed high levels of *Osr2*, a key regulator in connective tissue differentiation, suggesting  
931 that connective-tissue differentiation at this stage is preferentially occurring dorsally, as suggested by *Osr2*  
932 *in situ* experiments [77].



933 **Fig. S4 Metacells exhibit a hierarchical and specific combinatorial code of TFs.**

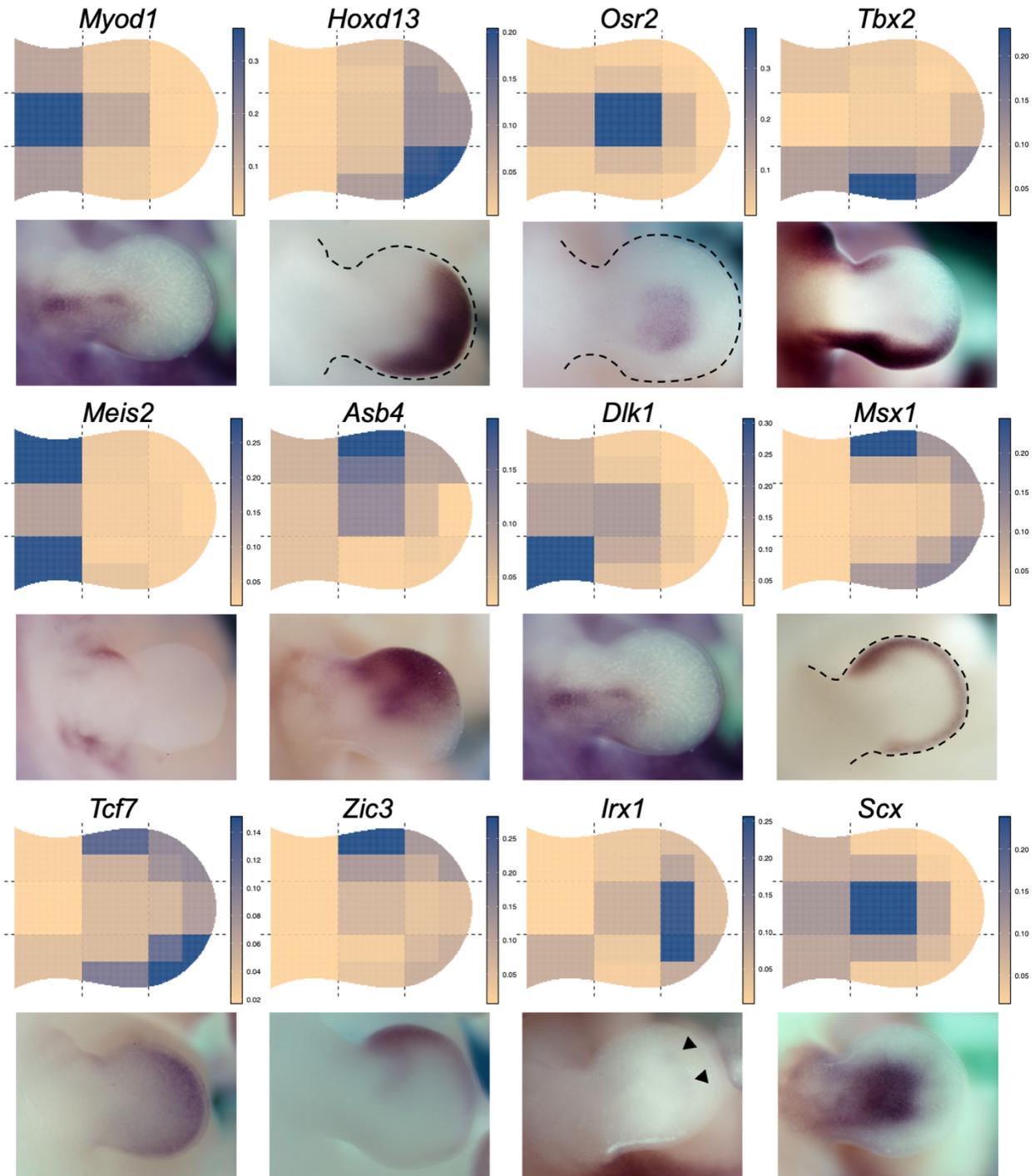
934 **(A)** Heatmap showing the normalized expression of genes with loosely specific gene expression (AUC >  
935 0.5) across all metacells. \* marks metacell 22 which groups cells of overall lower quality. **(B)** Heatmap

936 showing the pairwise number of TFs that show different expression levels ( $FC > 2$ ). Cell types are annotated  
937 using the same color code as in **Fig. 1**.



939 **Fig. S5 Metacells are spatially homogeneous.**

940 We employed Shannon entropy as a metric for measuring color heterogeneity. Monte Carlo simulations  
941 (n = 100000) were used to test for positional homogeneity. Stacked bar plots showing the distribution of  
942 color for each metacell in all three photoconversion patterns (PD, AP and AER). Clusters labeled with a  
943 start (\*) display significantly lower Shannon entropy than expected by chance (Bonferroni corrected p-  
944 value  $\leq 0.01$ ).

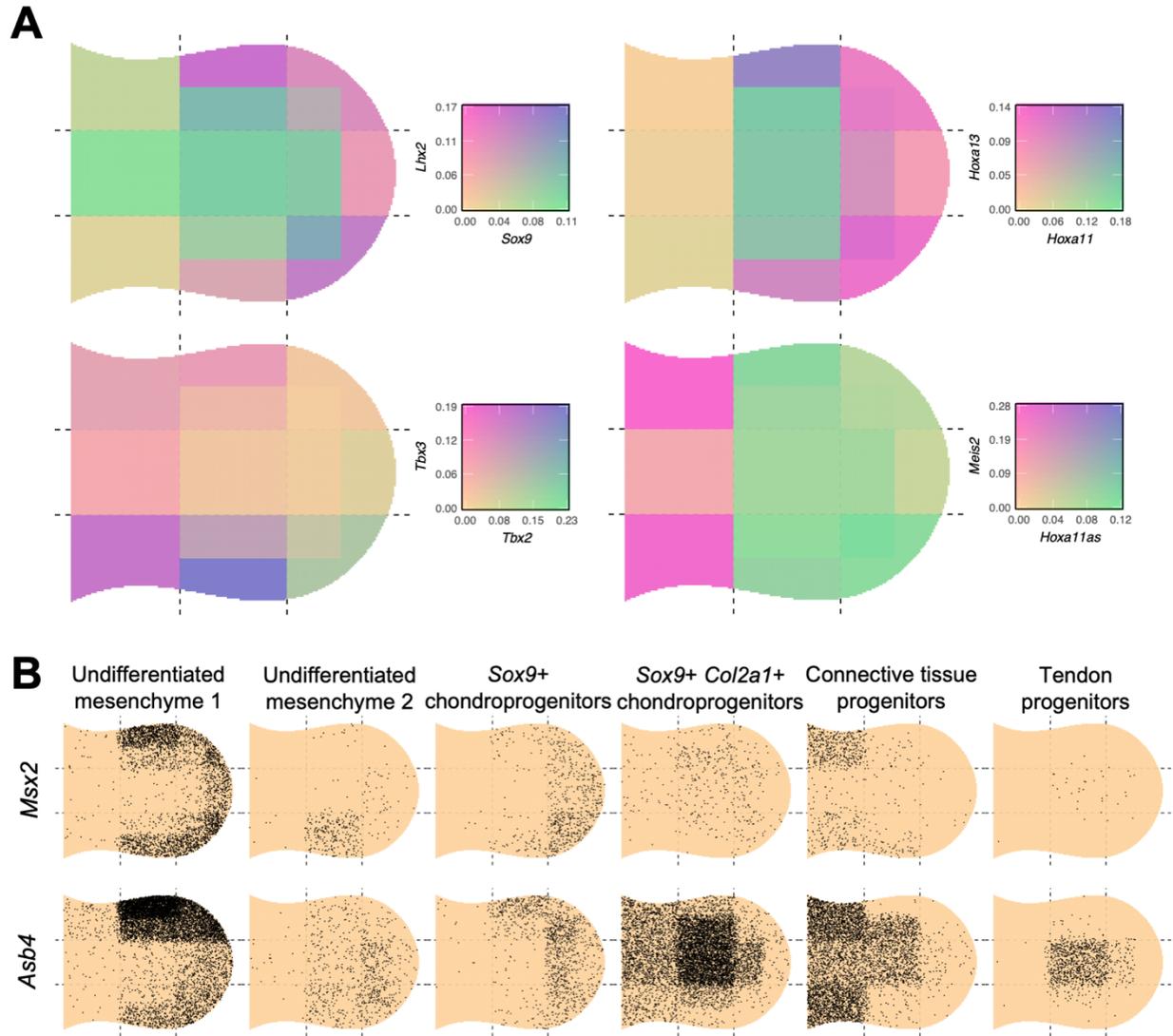


945

946 **Fig. S6 vISH match the gene expression patterns obtained by classical *in situ* hybridization.**

947 vISH and the corresponding classical ISH patterns for twelve TFs involved in limb development. ISH

948 images were obtained from the EMBRYOS database [78, 79].



949

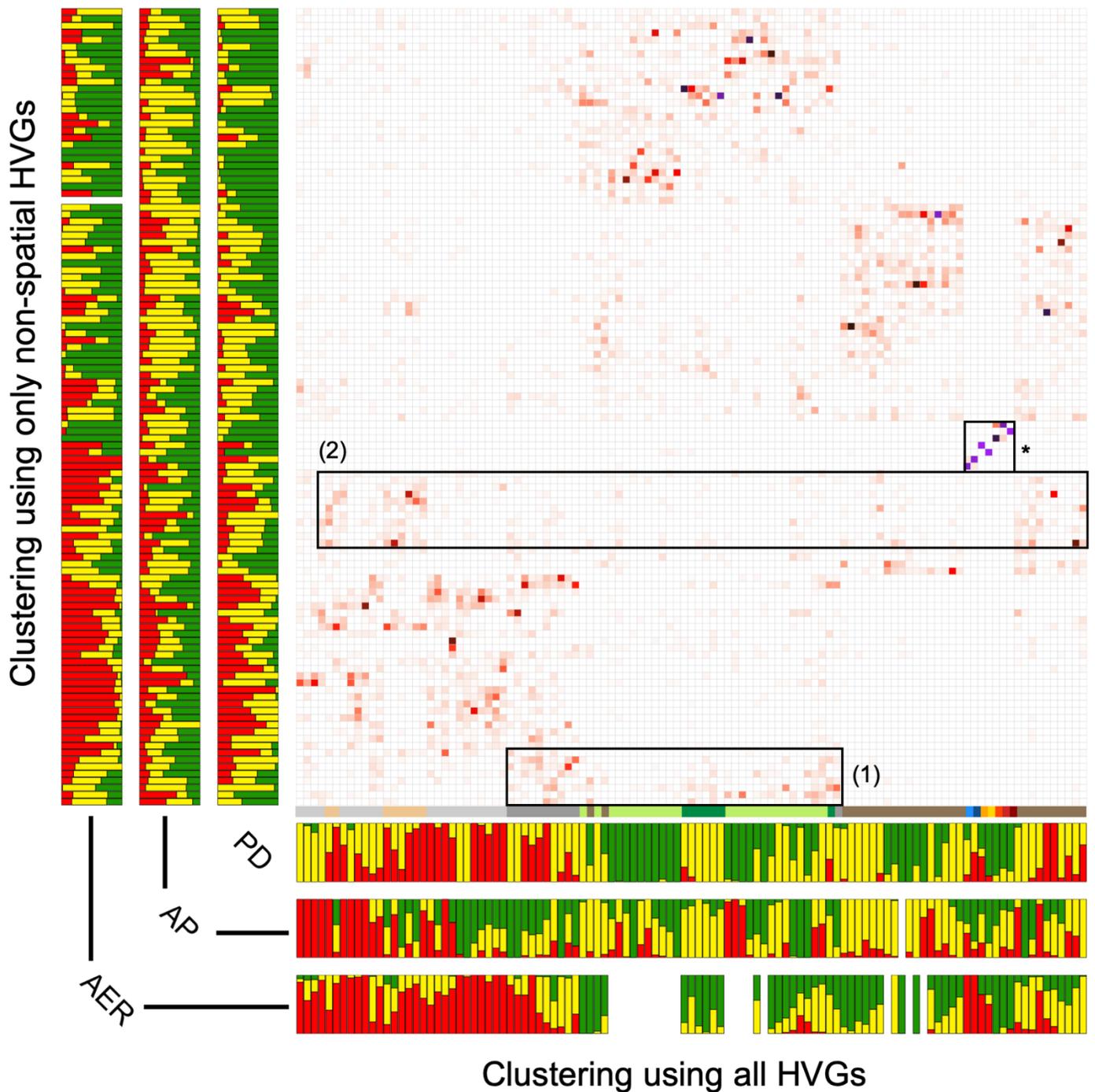
950 **Fig. S7 Double virtual *in situ* hybridization (vISH) and cell type-specific vISH.**

951 (A) Double vISH uses a bivariate color scale to represent the spatial overlap of expression patterns.

952 (B) vISH can be deconvolved into cell type-specific vISH. To take into account the different dynamic

953 ranges across cell types, up to 10,000 UMIs are sampled according to the expected number of counts for

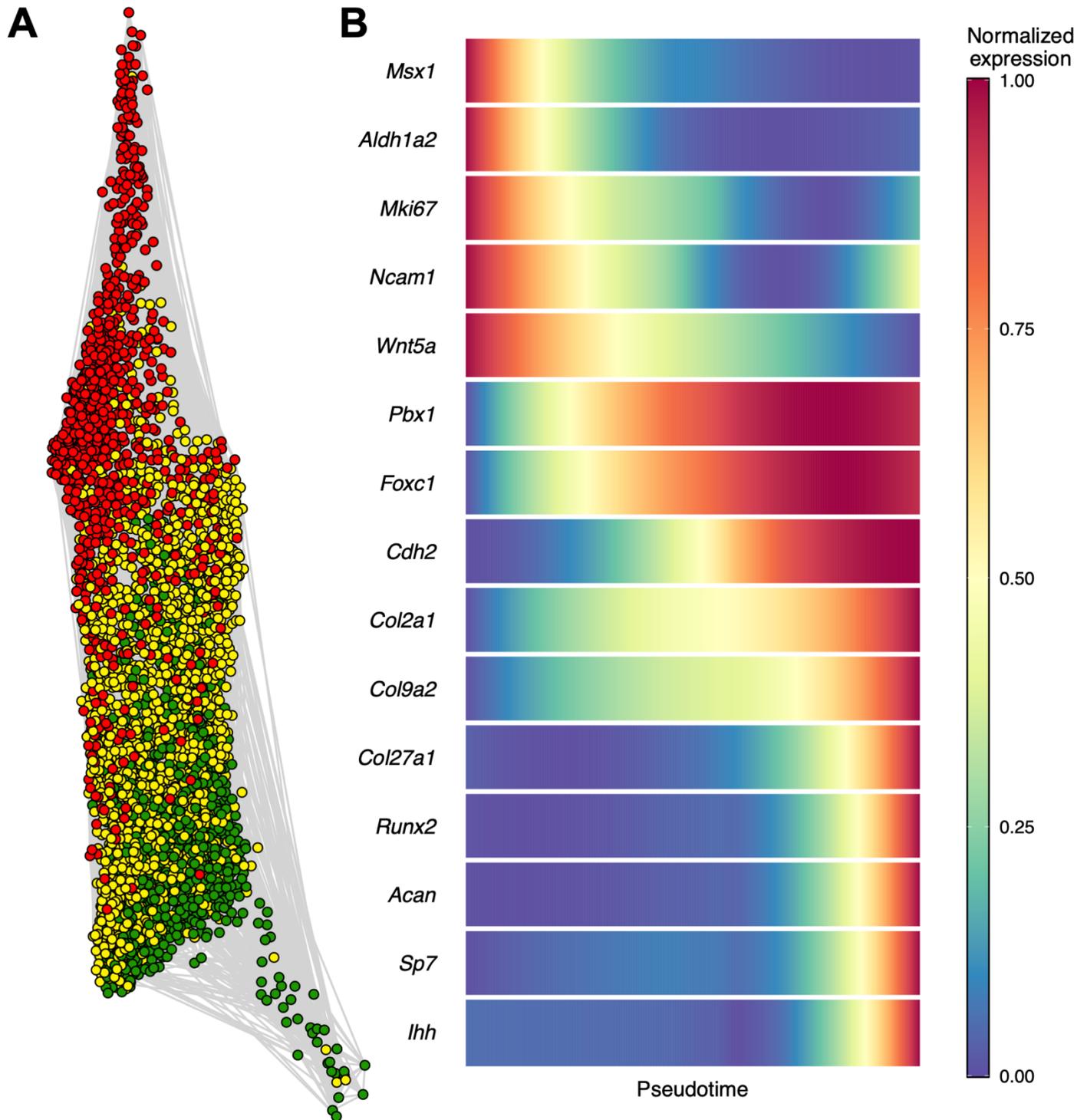
954 each spatial bin.



955 **Fig. S8 Confusion matrix between the two dataset clusterings.**

956 Confusion matrix representing the redistribution of cells between the clustering using all the highly variable  
957 genes and the clustering using only non-spatial highly variable genes. Darker values (black, purple) indicate  
958 high correspondence between metacells of the two clusterings. The color distribution is indicated on the  
959 side of the heatmap. Non-mesenchymal clusters are unaffected (\*) while some mesenchymal clusters are

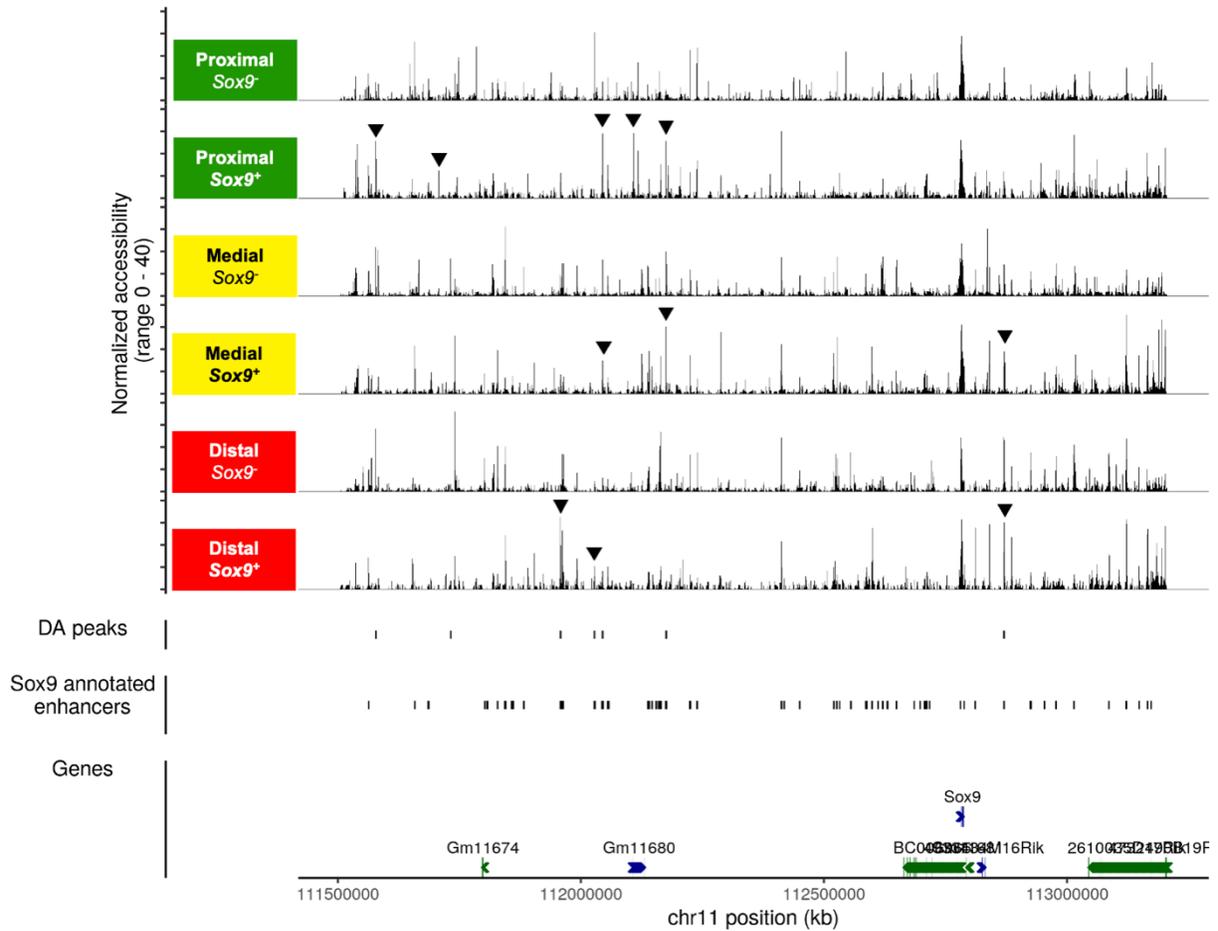
960 split and merged. Some no-HVGs metacells clusters receive contribution from type 2 undifferentiated  
961 mesenchyme and dense regular connective tissue progenitors metacells (1), and from distal chondrogenic  
962 progenitors and proximal *Sox9*<sup>+</sup> *Col2a1*<sup>+</sup> chondroprogenitors (2). Blacklisting of spatial HVGs additionally  
963 results in extensive cell type internal reorganization. Abrogating spatial information potentially leads to  
964 artefactual clustering if cell types are homogeneous.  
965



966 **Fig. S9 Eliminating confounder effects of cell position reveals overlapping chondrogenic**  
967 **transcriptomic states.**

968 **(A)** 2D projection of the k-NN ( $k = 10$ ) graph of chondrocytes using the Distributed Recursive (Graph)  
969 Layout algorithm. The procedure used to generate the graph is described in the **methods** section.

970 **(B)** Pseudotime was assessed using ELPiGraph and gene expression as a function of pseudotime was  
971 estimated and smoothed using a loess regression for various stage-specific chondrogenesis markers.

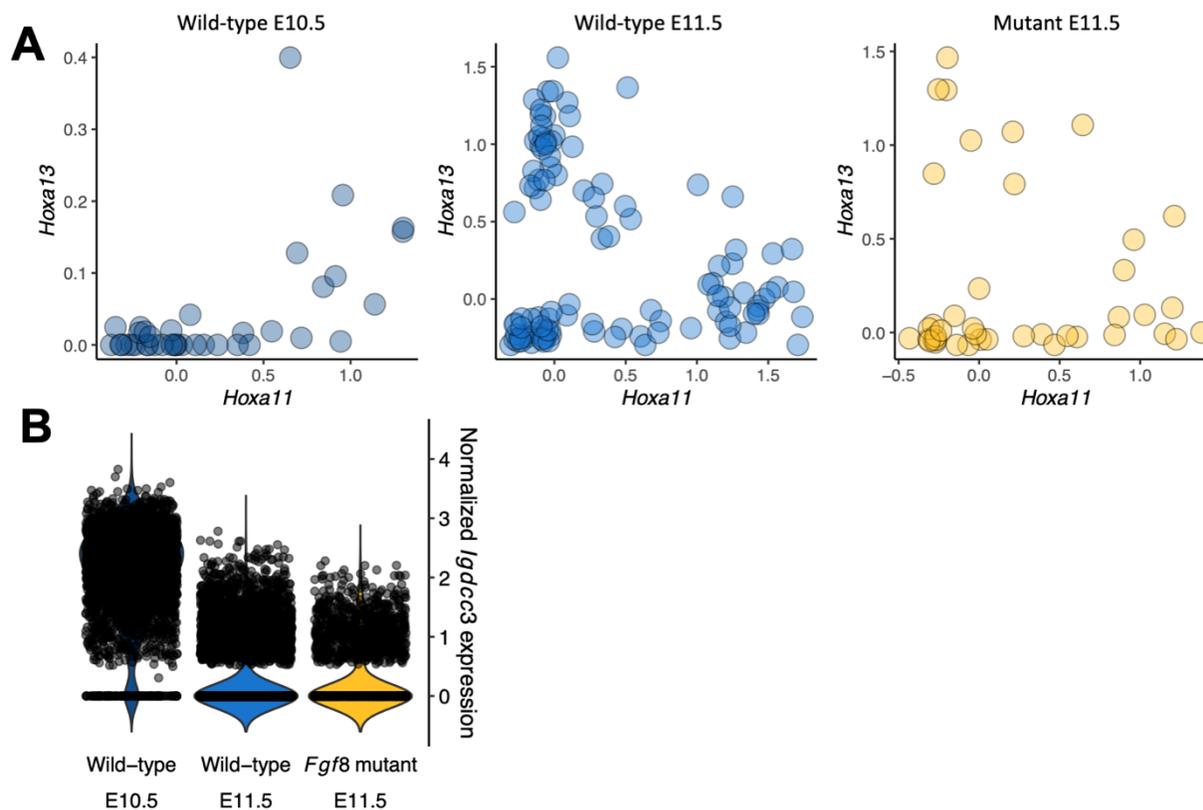


972

973 **Fig. S10 Transferred cell identity and position from TATTOO-seq to scATAC-seq points to spatially**  
974 **regulated accessibility landscape at the *Sox9* locus.**

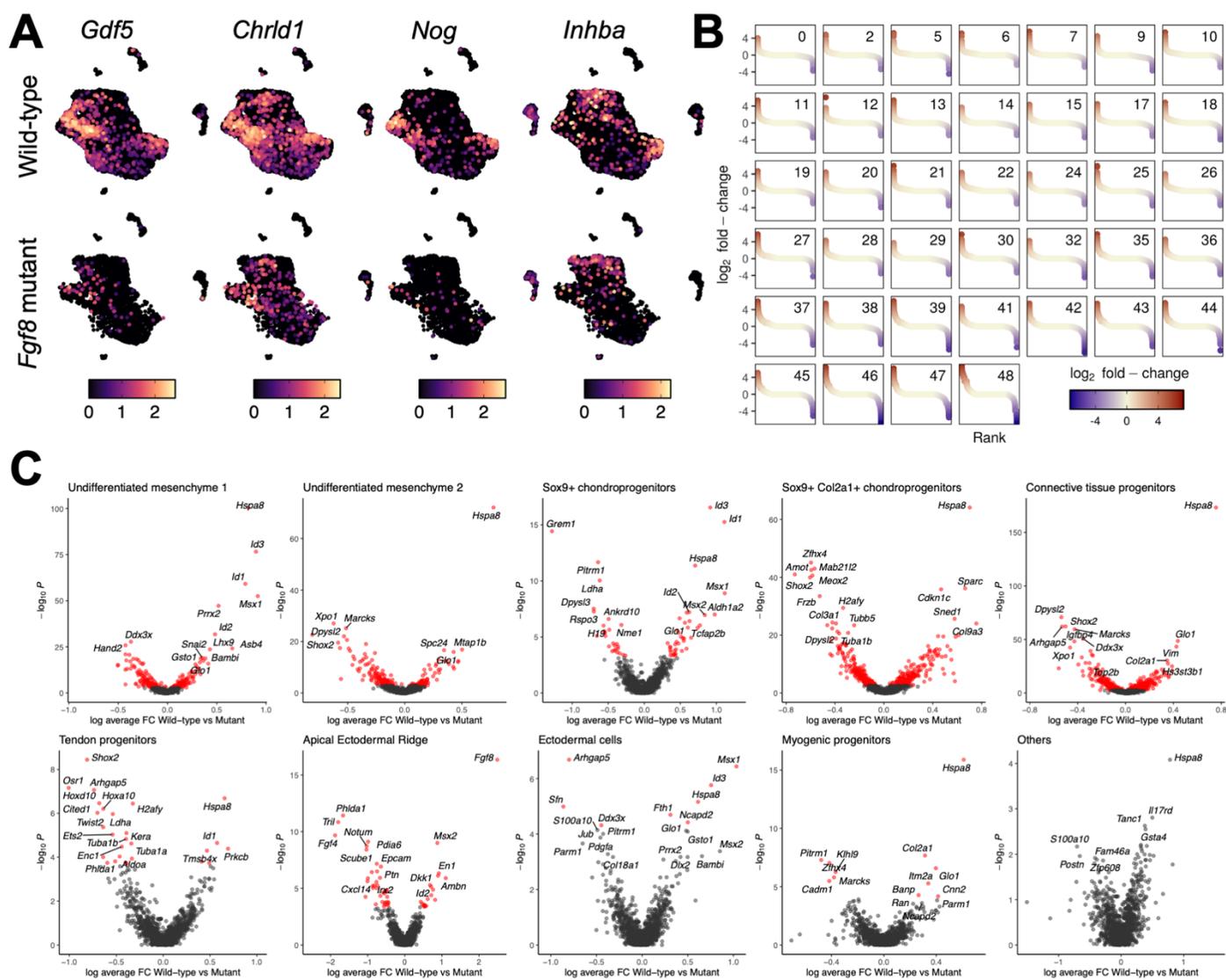
975 Cell identity and position was transferred from our TATTOO-seq data to limb scATAC-seq data using  
976 Seurat's TransferData function. Accessibility profiles were aggregated by transferred position and *Sox9*  
977 expression levels and displayed at the top. All peaks in the *Sox9* regulatory domain were tested for  
978 differential accessibility. Differentially accessible (DA) peaks are indicated as well as previously described  
979 *Sox9* enhancers [60].

980



981 **Fig. S11 Differences between wild-type and *Fgf8* mutant cells do not reflect a global developmental**  
982 **delay in mutant limbs.**

983 (A) Scatter plot showing the normalized expression of *Hoxa13* and *Hoxa11* in E10.5 and E11.5 wild-type  
984 metacells and E11.5 mutant metacells. In the mutant limbs, *Hoxa13* and *Hoxa11* were expressed in distinct  
985 cell-states, corresponding to the largely mutually exclusive territories observed in wild-type E11.5 limbs,  
986 and contrasting with the nested expression found earlier at E10.5 [80, 81]. (B) Violin plot showing the log<sub>10</sub>  
987 depth-normalized expression of *Igdcc3* in E10.5 and E11.5 wild-type cells and E11.5 mutant cells. *Igdcc3*  
988 has been reported to be a sensitive temporal marker of limb development, showing strong expression at  
989 E10.5 but minimal expression at E11.5. We found its expression greatly reduced in both wild-type and  
990 mutant E11.5 scRNA-seq data, while it is robustly detected in E10.5 samples.



991 **Fig. S12 Transcriptional differences between wild-type and mutant limb cells.**

992 (A) UMAP showing log<sub>10</sub> depth-normalized counts for some genes associated with the TGFβ superfamily

993 signaling showing a reduction of expression in mutant cells. The data is split by genotype: wild-type (top

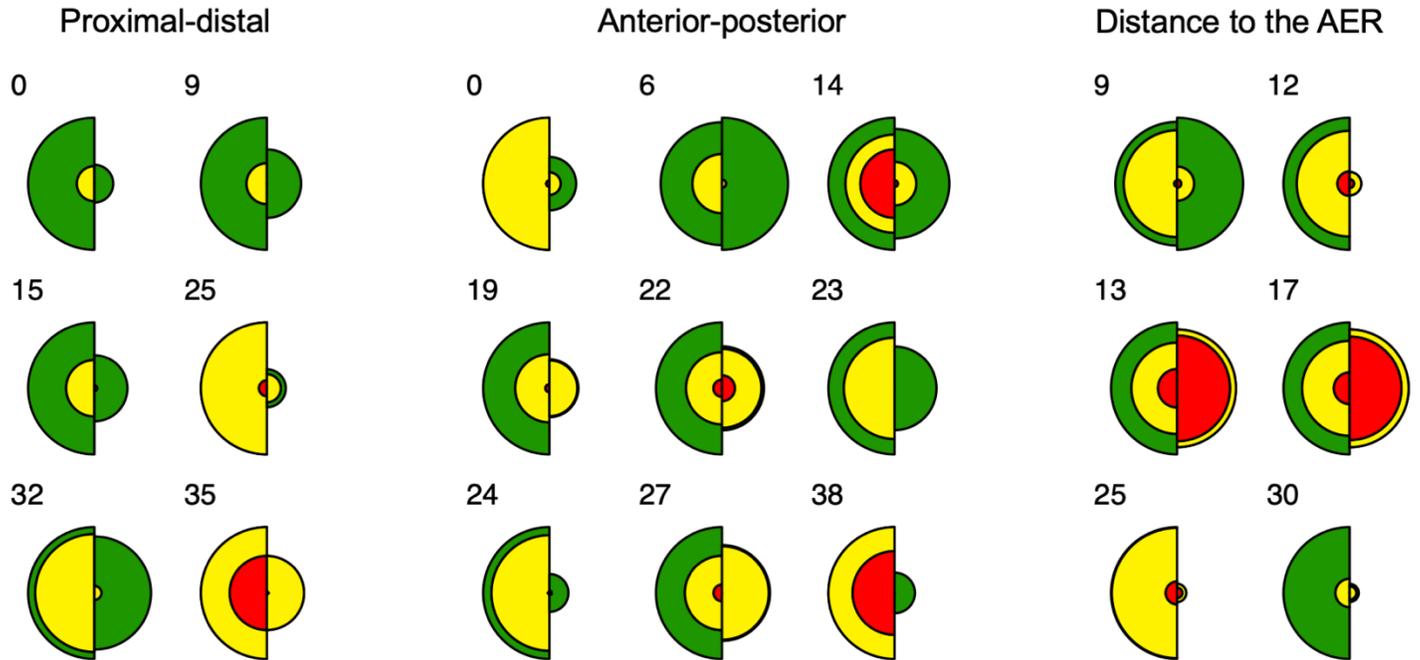
994 row) and *Fgf8* mutant (bottom row). (B) Plot showing the distribution of log<sub>2</sub> fold-changes between wild-

995 type and mutant cells within each Seurat cluster. (C) Volcano plot showing the p-values of statistical tests

996 for differential expression as a function of log<sub>2</sub> fold-change between wild-type and mutant cells in each cell

997 type. Genes for which BH-corrected p-value < 0.01 are represented in red and genes names are indicated

998 for the top 20 most significant genes.



999 **Fig. S13 Spatial distribution changes between wild-type and mutant limb cells.**

1000 Circular histograms representing the distribution of colors for wild-type and mutant cells in Seurat clusters

1001 that show altered spatial distributions.

1002