



**HAL**  
open science

## TATTOO-seq delineates spatial and cell type-specific regulatory programs in the developing limb

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 in the developing limb. Science Advances , 2022, 8 (50), 10.1126/sciadv.add0695 . pasteur-04081196

**HAL Id: pasteur-04081196**

**<https://pasteur.hal.science/pasteur-04081196>**

Submitted on 25 Apr 2023

**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 4.0 International License

## GENETICS

## TATTOO-seq delineates spatial and cell type–specific regulatory programs in the developing limb

Sébastien Bastide<sup>1,2,3</sup>, Elad Chomsky<sup>4,5</sup>, Baptiste Saudemont<sup>1</sup>, Yann Loe-Mie<sup>1,6</sup>, Sandrine Schmutz<sup>7</sup>, Sophie Novault<sup>7</sup>, Heather Marlow<sup>1,8</sup>, Amos Tanay<sup>4</sup>, François Spitz<sup>1,3\*</sup>

The coordinated differentiation of progenitor cells into specialized cell types and their spatial organization into distinct domains is central to embryogenesis. Here, we developed and applied an unbiased spatially resolved single-cell transcriptomics method to identify the genetic programs underlying the emergence of specialized cell types during mouse limb development and their spatial integration. We identify multiple transcription factors whose expression patterns are predominantly associated with cell type specification or spatial position, suggesting two parallel yet highly interconnected regulatory systems. We demonstrate that the embryonic limb undergoes a complex multiscale reorganization upon perturbation of one of its spatial organizing centers, including the loss of specific cell populations, alterations of preexisting cell states' molecular identities, and changes in their relative spatial distribution. Our study shows how multidimensional single-cell, spatially resolved molecular atlases can allow the deconvolution of spatial identity and cell fate and reveal the interconnected genetic networks that regulate organogenesis and its reorganization upon genetic alterations.

## INTRODUCTION

During embryonic development, cells acquire different identities according to a specific patterning blueprint, which ensures that organs and other structures are correctly positioned along the body plan (1). Impaired establishment of spatial cues or the inability of cells to correctly acquire and interpret positional information can lead to developmental abnormalities (2, 3), and modulation of patterning centers' activity has been suggested to underlie evolutionary morphological diversification (4–7). The recent development of single-cell transcriptomics has allowed the characterization of cell state diversity within a tissue and to identify the associated gene signatures (8). However, spatial position is usually lost in the process, which is problematic in the context of embryonic patterning. Several methods have been developed to preserve or retrieve spatial information in single-cell data, each with specific strengths, limitations, and trade-offs. Retrospective mapping of single-cell clusters using spatial landmark genes (9, 10) relies on arbitrary binary thresholds and may become impractical when experimental conditions affect landmark genes. Highly multiplexed single-molecule RNA–fluorescence in situ hybridization (FISH) (multiplexed error-robust–FISH and sequential FISH) (11, 12) and in situ sequencing methods (fluorescent in situ sequencing and spatially-resolved transcript amplicon readout mapping) (13, 14) provide direct quantitative spatial measurements, but their detection power remains limited to a couple thousands of genes simultaneously. Spatially resolved mRNA captured via barcoded beads

(Slide-seq and Stereo-seq) (15, 16) offers high spatial resolution for tissue sections but yields sparse datasets that impede lowly expressed gene detection and does not currently provide single-cell resolution. Furthermore, both approaches remain technically challenging, requiring specialized equipment and reagents, and often inconvenient for large specimens. Here, we present a simple alternative approach, TATTOO-seq, allowing the characterization of thousands of cells by single-cell RNA sequencing (scRNA-seq), while independently recording their spatial position of origin and transcriptome.

As a proof of principle, we applied TATTOO-seq to characterize the genetic programs coordinating spatial patterning and cell identity in the developing mouse limb, a classic model of patterning and differentiation. In the limb, secreted fibroblast growth factors (FGFs) from the distal apical ectodermal ridge (AER) and sonic hedgehog (SHH) from the posterior zone of polarizing activity (ZPA) define a coordinate system along the proximal-distal (PD) and anterior-posterior (AP) axes, respectively (7, 17). Within this patterned field, mesenchymal progenitor cells, migrating myoblasts, and neurons will differentiate and organize into cartilage and bone, muscles, tendons, ligaments, connective tissues, nerves, and blood vessels (18–20). Undifferentiated mesenchymal precursors that populate the limb bud expand at different rates and adopt different fates depending on their location. Notably, chondrogenic condensations form in response to these positional cues and build a cartilage anlagen that will later on be ossified. Others make up the fibrous connective tissue that connects the musculoskeletal system: tendons and ligaments. Spatial information thus directly affects the number, position, shape, and length of the skeletal elements, as well as the attachment of muscle and connective tissues that make up a functional limb. Accordingly, the extensive diversity of limb morphologies observed in tetrapods often seems to result from changes to early limb patterning mechanisms or their interpretation by chondrogenic progenitors (4, 5, 7, 21). Similarly, alterations in the genes regulating the formation and activity of the limb patterning centers underlie many of the limb malformations found

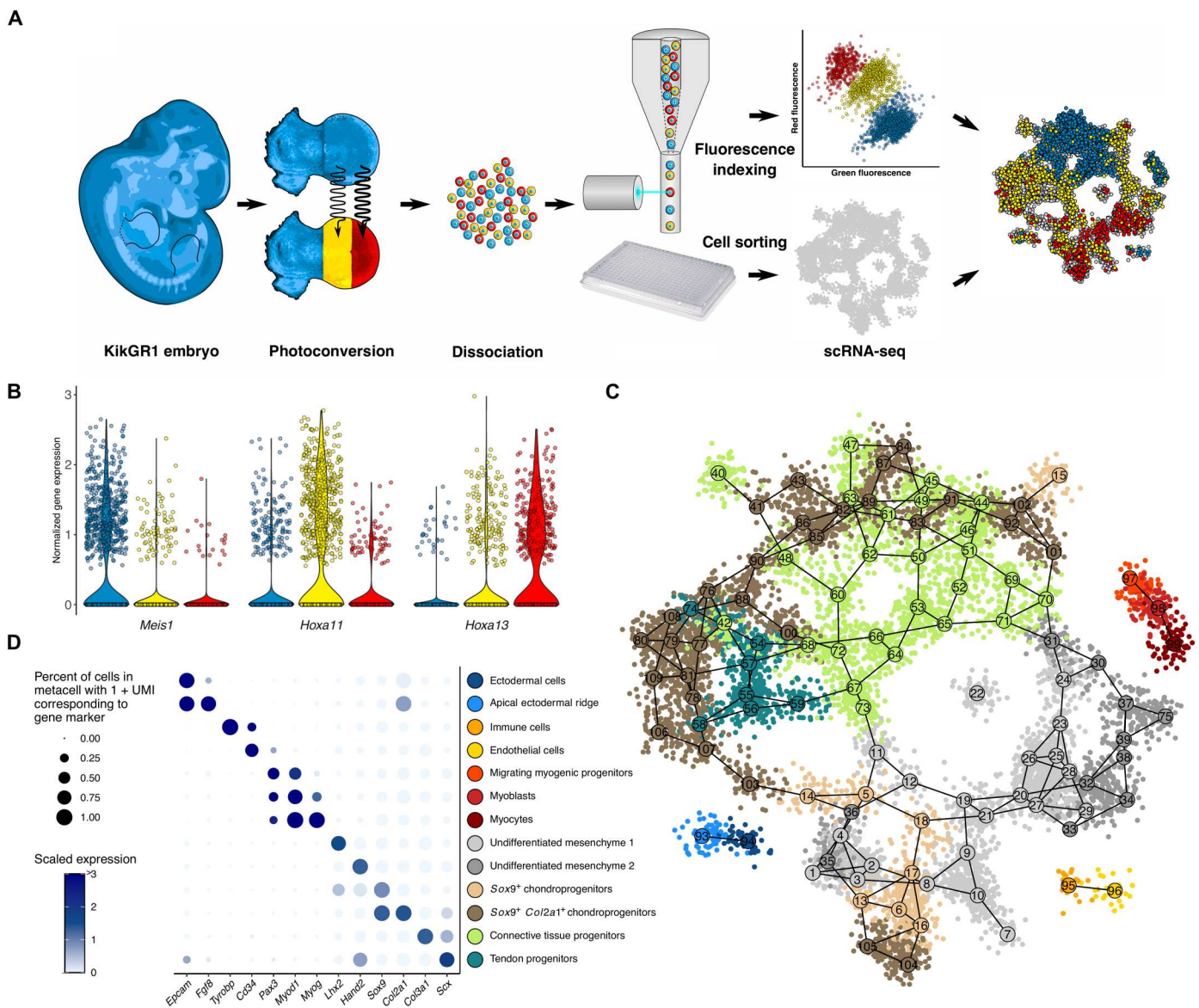
<sup>1</sup>(Epi)genomics of Animal Development, Department of Developmental and Stem Cell Biology, Institut Pasteur, Paris, France. <sup>2</sup>École Doctorale “Complexité du Vivant”, Sorbonne Université, 75005 Paris, France. <sup>3</sup>Department of Human Genetics, The University of Chicago, Chicago, IL, USA. <sup>4</sup>Department of Computer Science and Applied Mathematics, Weizmann Institute, Rehovot, Israel. <sup>5</sup>Department of Biological Regulation, Weizmann Institute, Rehovot, Israel. <sup>6</sup>Hub de Bioinformatique et Biostatistique, Département Biologie Computationnelle, Institut Pasteur, Paris, France. <sup>7</sup>Cytometry and Biomarkers, Center for Technological Resources and Research, Institut Pasteur, Paris, France. <sup>8</sup>Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, IL, USA.  
\*Corresponding author. Email: fspitz@uchicago.edu

in human patients [e.g., (22–25)]. Although decades of genetic studies have identified the signaling pathways (26, 27) and transcription factors (TFs) (28–31) controlling limb patterning and the differentiation of its cell types, we still lack a comprehensive understanding on how individual cells define and adapt their gene regulatory programs according to their position. Furthermore, while limb patterning has been extensively studied in the context of the formation of the skeletal elements, the influence of patterning signals on the organization and specialization of the other types of connective tissues are far less understood.

## RESULTS

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

TATTOO-seq is based on an optical registration system (Fig. 1A) using the constitutively expressed photoconvertible protein Kikume Green-Red (KikGR1) (32). The modulation of the degree of photoconversion of cellular KikGR1 proteins produces distinct levels of red-to-green fluorescence ratios, which are used as a positional index in a flexible, user-defined, spatial coordinate system. This index can be recorded as cells are fluorescence-activated cell sorting (FACS)-sorted into barcoded 384-well plates and processed by massively parallel scRNA-seq (MARS-seq) (33). This strategy enables the independent determination of single-cell positions and transcriptomes, with improved flexibility over simple photoactivation [as used for NICHE-seq (34)].



**Fig. 1. TATTOO-seq simultaneously records the position and measures the transcriptome of single cells in the mouse limb bud. (A)** Overview of TATTOO-seq. **(B)** Expression of PD markers *Meis1*, *Hoxa11*, and *Hoxa13* (78) for each color along the PD axis. **(C)** Two-dimensional projection of the metacell graph. **(D)** Marker gene expression in each annotated cell type.

We used TATTOO-seq to study patterning and differentiation in embryonic day 10.5 (E10.5) to E11.5 mouse forelimb buds. Under our experimental conditions, photoconversion was homogeneous throughout the limb thickness and did not affect cell viability or gene expression (fig. S1). To characterize the developing limb bud transcriptional landscape, we sorted ~10,000 live cells from E11.5 forelimbs (45 to 51 somites; embryos,  $n = 4$ ; limbs,  $n = 7$ ) and performed MARS-seq (33). A total of 8750 cells were kept after quality control filtering [unique molecular identifier (UMI) > 2000] (fig. S2A). To limit batch effects, we used different photoconversion patterns on the two forelimbs of each embryo. We first examined the efficiency of the KikGR1-encoded optical registration method to correctly identify the position of origin of cells. We considered the expression of the positional markers *Meis1* (proximal), *Hoxa11* (medial), and *Hoxa13* (distal) (17) in samples photoconverted along the PD axis (Fig. 1B). The expression of these markers showed strong concurrence with the expected indexed color, validating that TATTOO-seq accurately retains positional information.

### A molecular atlas of the developing limb bud

We used MetaCell (35) to identify transcriptionally homogeneous groups of cells and reconstruct the developing limb bud's transcriptional manifold. We identified 109 metacells (Fig. 1C). As a control for batch- or stage-specific biases, we analyzed the contribution of each embryo to the different metacells (fig. S2, B and C). The vast majority of the metacells comprised cells from all four embryos, indicating that the corresponding cell states are present in the limb throughout the period analyzed (45 to 51 somites). We, however, noticed that metacells 5, 17 to 19, 21, and 27 showed a biased contribution from embryo #2 (45 somites), but not from embryo #5 (45 somites as well), a bias that could be explained by the specific sampling of medial and distal cell states in sample #2 or a batch effect. For example, the transcriptome of metacell 27 (mostly from embryo #2) is very similar to that of metacell 26 (which is depleted from cells from embryo #2). This suggests that metacells 26 and 27 represent the same cell state, up to a batch effect. We then assigned each metacell to the different cell populations present in the limb bud using known cell type-specific markers (Fig. 1D and fig. S3). The average expression in each metacell (as a ratio to the median expression in all metacells) is provided in table S1. Mesenchymal cell states constituted the bulk of the data and formed a continuum of related cell states. Multiple metacells represented the chondrogenic lineage, comprising *Sox9*<sup>+</sup>-condensing chondroprogenitors, *Col2a1*<sup>+</sup> early-stage chondrocytes, and some *Acan*<sup>+</sup> chondrocytes (fig. S3D). Each differentiation stage spread throughout the mesenchymal projection (Fig. 1C, brown metacells). Dense regular connective tissue progenitors differentiating into tenocytes and ligament fibroblasts showed expression for *Col3a1*, while more differentiated tendon progenitors expressed *Scx* (fig. S3E). Noteworthy, *Col3a1* transcripts were also detected in the most differentiated chondrocyte metacells. The remaining uncommitted mesenchymal cells expressed none of the canonical chondrogenic or connective tissue markers. This undifferentiated mesenchyme exhibited a diversity of cell states that could be classified into two main classes based on *Lhx2* and *Hand2* expression (fig. S3F). At this level of resolution, we did not capture any transcriptional difference along the dorsal-ventral (DV) axis in the mesenchyme (fig. S4, A and B), and although our clustering allowed separating the AER from the dorsal

and ventral ectoderm, we were not able to separate the dorsal *Wnt7a*-expressing ectoderm from the *En1*-expressing ventral ectoderm (36) (*En1* is also expressed in the AER; fig. S4C). Two metacells representing cell states derived from the same cell type often show multiple differentially expressed TFs (log fold change > 2). This highlights the existence of distinct regulatory programs within a given cell type, corresponding to differentiation stage, spatial position, or other factors (fig. S5).

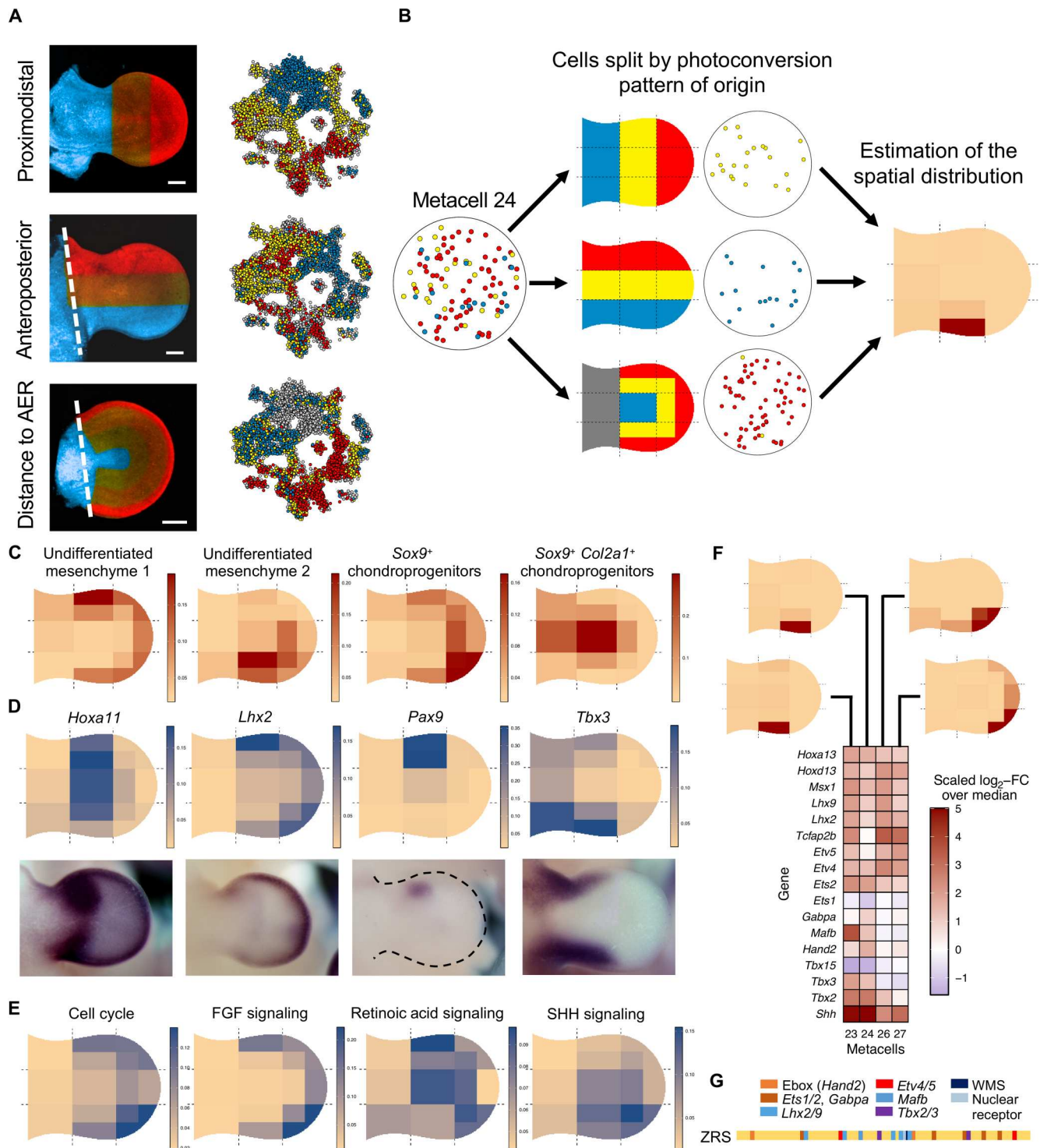
### The fine-scale organization of limb bud patterning

As mentioned above, the flexibility of optical registration allows using patterns that closely fit the spatial organization of the structure of interest and capture its multidimensional organization by different user-defined patterns of photoconversion. For the limb, we applied patterns of photoconversion to register cell positions along the PD and AP axes of the limb bud, with respect to the distance to the AER (Fig. 2A). The gating strategy is shown in fig. S6. Overlaying this color information about the metacell graph revealed a strong spatial component to its organization (Fig. 2A). With the exception of endothelial and ectodermal metacells, metacells' color composition was more homogeneous than expected by chance (fig. S7), indicating that metacells reside at precise spatial locations. This fact supports that metacells represent different biological states and are not clustering artifacts. It directly demonstrates that cell position is a strong component of mesenchymal cell identity as reflected by the expression of position-specific transcriptional programs.

Although the information provided by a single cell is limited to one axis of our coordinate system, metacells comprise cells from all three photoconversion axes. Each metacell can thus be positioned into a more complex 14-bin spatial grid (Fig. 2B and fig. S8; see Materials and Methods) by calculating the spatial probability distribution that maximized the likelihood of observing a metacell's color composition. Individual metacells' probability distributions were aggregated by cell type annotations, resulting in cell type spatial distributions (Fig. 2C).

This spatially resolved atlas showed that transcriptionally related metacells can have markedly different spatial distributions and allowed for the exploration of gene expression with increased precision and resolution. For each expressed gene, we derived a virtual in situ hybridization (vISH) pattern by computing the probability distribution over space for a random UMI of that gene to be found in each spatial bin. vISH results closely matched actual in situ hybridization experiments (Fig. 2D and fig. S9). We provide access to a precomputed atlas of vISH for ~17,000 mouse genes at [http://nobelmarlowlab.uchicago.edu:8888/TATTOOseq\\_vISH/](http://nobelmarlowlab.uchicago.edu:8888/TATTOOseq_vISH/). vISH further enabling to visualize genes and pathways across cell types at new depth and precision. We not only can readily compare the expression of two or more genes (fig. S10A) but can also break down gene expression patterns by cell type (fig. S10B). Process-specific gene expression profiles can also be combined to examine the spatial distribution of integrated pathways or biological processes, such as cell proliferation or signaling pathway-responsive cells (Fig. 2E).

As an example, we used vISH to map the ZPA, which organizes limb patterning along the AP axis (1, 26). Its defining gene, *Shh*, was robustly expressed in four metacells *Shh* (23, 24, 26, and 27) (Fig. 2F) which collectively mapped to ZPA at the limb's posterior margin. Metacells 23 and 24 were located medially, whereas metacells 26 and 27 were located more distally, suggesting that they may



**Fig. 2. Integrating TATTOO-seq data to build a spatial map of cell state and gene expression.** (A) Maximum intensity projection of z-stack images of each of the three photoconversion patterns. (B) Strategy to assemble information from different patterns of photoconversion to infer the position of cell states. (C) Spatial distribution for some aggregated mesenchymal cell types. (D) Examples of vISH obtained ab initio from TATTOO-seq data. Images were obtained from the EMBRYS database (79). (E) Spatial projection of summarized gene expression for relevant gene sets. (F) Spatial and transcriptomic dissection of the *Shh*-positive ZPA. FC, fold change. (G) Schema of the ZRS enhancer–regulating *Shh* and its cohort of TF-binding sites.

Downloaded from https://www.science.org at Institut Pasteur on April 26, 2023

represent previously unknown ZPA subdomains. As mentioned above, metacells 26 and 27 are likely to represent the same general state, and the differences are the possibility of consequences of a batch effect or a stage difference. Levels of *Tbx2* expression constitutes the main difference between distally located metacells 26 and 27, which is consistent with the distal extension of the *Tbx2* expression domain throughout the hand plate formation and the possible earlier stage of development of the cells assigned to metacell 27. Several TFs showed differential expression between metacells 23 and 24 and 26 and 27 (Fig. 2F and fig. S11), including *Hand2*, *Hox*, and *Ets* TFs, which have been shown to regulate *Shh* expression (37–39). Conserved binding motifs for these TFs are present within the zone of polarizing activity regulatory sequence (ZRS), the *Shh* limb enhancer (Fig. 2H) (4, 23, 40). The existence of different transcriptional regimes between the proximal (represented by metacells 23 and 24) and distal (represented by metacells 26 and 27) parts of the ZPA, including ZRS-associated TFs, raises the possibility that *Shh* may use different modes of regulation in different parts of the ZPA. On the proximal side, its activity may be primarily driven by *Tbx2/3* and *Hand2*, whose expression is higher proximally. Whereas distally, other factors (*Hoxd13*, *Tfap2b*, and ETS TFs) may play a relatively increased role. Supporting the multiplicity of ZRS-associated complements of TFs, mutations in the ZRS have been shown to affect its activity to different degrees along the PD axis (39). It should be emphasized that these different regimes of regulation of *Shh* through its ZRS enhancer, as proposed here, do not necessarily imply the partition of the ZPA into distinct domains. These regimes may coexist within the ZPA, with their relative contribution to *Shh* expression varying along the PD axis (23/24 versus 26/27) and possibly the AP axis (23 versus 24). This hitherto hidden diversity of regulatory regimes within the ZPA may contribute to the diversity of limb morphologies resulting from genetic variants in the ZRS or ZRS-associated TFs as they may modulate the shape and signaling strength of the ZPA (41–43).

### A TF code of spatial position

We next sought to use the TATTOO-seq atlas to comprehensively identify TFs whose expression exhibits strong spatial trends in the medial and distal mesenchymal cells. Of 1390 TFs detected in the autopod and zeugopod mesenchymal metacells (>100 UMIs in total), 202 showed variable expression ( $\geq 1$  metacell with expression of  $>1.5 \times$  median expression across all mesenchymal metacells). We used Spearman's rank correlation to assess monotonic trends in gene expression as a function of the distance to the AER and along the AP axis (Fig. 3A). We uncovered a multitude of TFs whose expressions are correlated with position, highlighting the spatial heterogeneity of regulatory states in the limb bud. TFs forming anterior-to-posterior gradients not only included previously known genes, such as *Alx4*, *Asb4*, and *Zic3* (31), but also revealed new genes with unknown functions in limb development (*Thra*, *Lmo4*, and *Cited2*). Twenty-six TFs formed posterior-to-anterior gradients of expression, comprising the 5'-*Hoxd* genes (*Hoxd10* to *Hoxd13* and *Evx2*) and *Hand2*. Forty-nine TFs were expressed close to the AER, including many of the known targets of AER-secreted FGFs and bone morphogenetic proteins (BMPs) (*Lhx2/9*, *Msx1/2*, and *Etv4*) and distal *Hox* genes (*Hoxd13* and *Hoxa13*). A total of 104 genes, including genes involved in connective tissue differentiation (*Sox5/6/9*, *Scx*, and *Runx1/3*) (29, 44–46)

showed the opposite trend, with weaker expression underneath the AER.

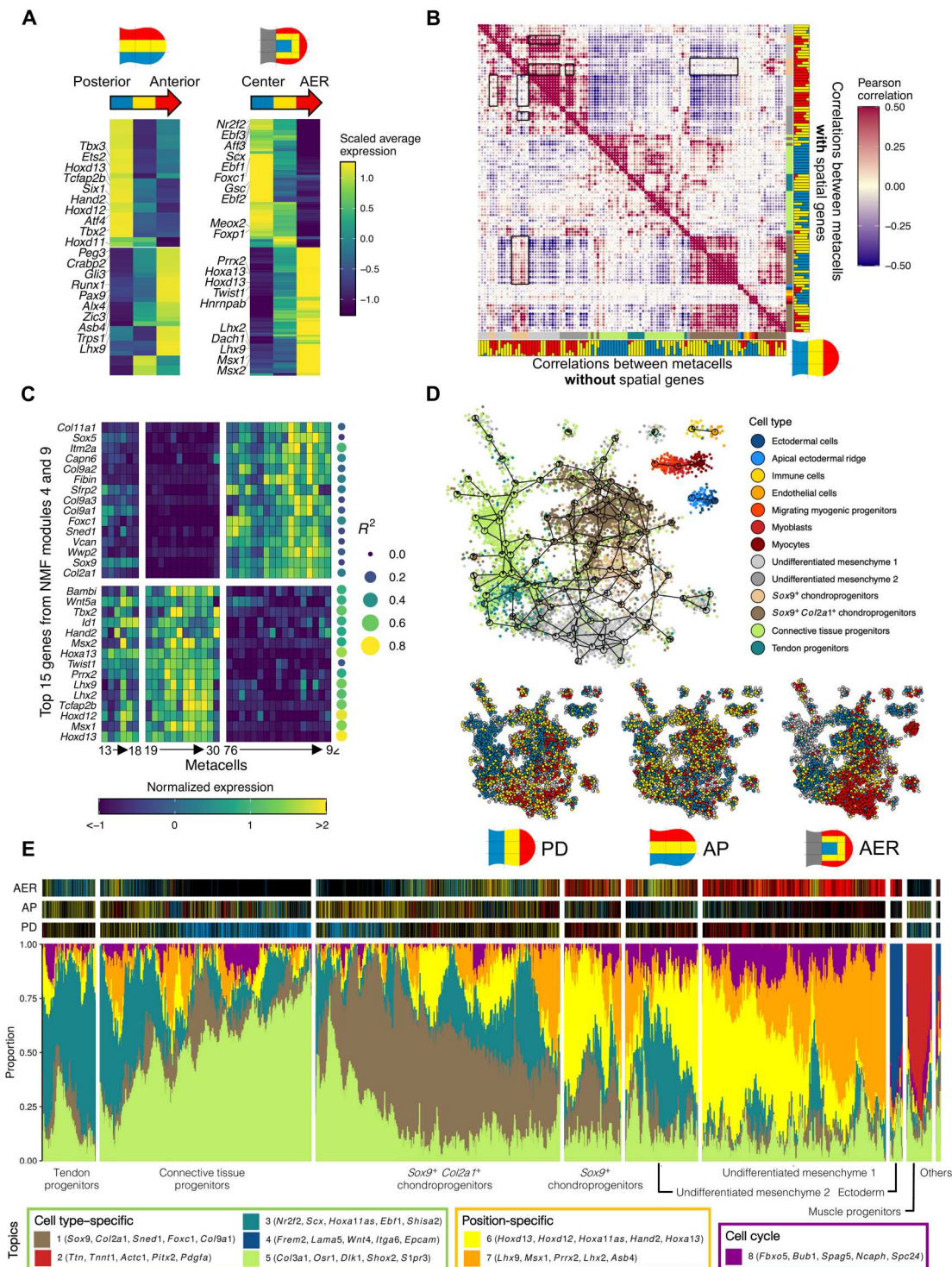
### Deconvoluting spatial patterning and cell differentiation programs

As an attempt to deconvolute the spatial and cell type-specific regulatory logics operating in the mesenchyme, we sought to classify genes as primarily carrying "spatial" or "cell type"-related information. We reasoned that genes encoding spatial information would have a similar spatial distribution of expression across cell states, while genes regulated in a strictly cell type-specific manner would show poor predictability from metacell position. We used linear regression to predict gene expression from the spatial probability distribution of each metacell for all genes and computed overall significance using *F* test. After Bonferroni correction for multiple testing, we retained models for 4713 genes (of 17,537, adjusted  $P < 0.01$ ). These genes include major components of limb signaling pathways (table S2). We then compared two pairwise correlation matrices between metacells computed using either all highly variable genes (HVGs) or only non-"spatially regulated" HVGs (Fig. 3B). We noted two types of changes: The correlation between two groups of distal metacells decreased without spatial HVGs, while the correlation between one of these groups and proximal chondrogenic metacells increases. Nonnegative matrix factorization (NMF) of the gene-by-metacell expression matrix revealed gene modules that could explain these changes. NMF module 4 exhibited high weights for known patterning genes with strong spatial regulation (14 of 15 top genes with  $R^2 > 0.45$ ), while module 9 exhibited high weights for chondrogenesis genes with overall low spatial regulation (Fig. 3C). The exclusion of spatially regulated genes to compute correlations decreased the influence of module 4 and, therefore, the similarity between metacells that simply reside at similar positions. It also unmasked a high similarity between chondrocytes in the distal and proximal compartments, which also appeared if single-cell data are clustered using only non-spatially regulated genes (Fig. 3D and fig. S12). This new graph projection grouped cells together by cell types, reducing the notable position-driven dispersal of the initial projection (Fig. 1C).

To further investigate the interplay between cell position and cell fate, we used fastTopics (47, 48) to model each cell as a mixture of different ( $k = 8$ ) topics, describing cell states as combinations of more or less independent modules (Fig. 3E). This uncovered a spectrum of combinatorial topic utilization consistent with the continuum of cell states highlighted by MetaCell. Notably, two anticorrelated topics reflected positional information in the undifferentiated mesenchyme 1 (topics 6 and 7). The relative importance of these topics followed a medial-anterior to distal-posterior axis, showing the strong correlation of these two topics with spatial position. These spatial topics 6 and 7 were also widely expressed at different levels in chondrogenic progenitors, providing position-associated diversity to cells otherwise defined by the core chondrogenic topic 1.

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

As chondrogenesis is initiated as the limb develops along the PD axis, chondroprogenitors at different stages of their differentiation lie in different spatially regulated transcriptomic subspaces, which is heavily reflected in the single-cell clustering and may mask lineage



**Fig. 3. Deconvoluting positional information and cell type-specific regulatory programs.** (A) Average expression per spatial compartment for TFs that exhibit spatial trends in the mesenchyme. Top 10 TFs with most significant positive and negative Spearman correlation values are annotated. (B) Pairwise correlation heatmap for metacells. Top: Using all genes defined as HVGs. Bottom: Excluding spatial genes. Cell types are indicated using the same color code as for Fig. 1 and color composition for the PD photoconversion pattern. (C) 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) Two-dimensional projection after clustering the TATTOO-seq dataset with no spatial genes. Top: Cells are colored by annotation. Bottom: Cell colors for each photoconversion pattern. (E) fastTopics structure plot. Top: Cell colors for each pattern (black if the cell originates from a different pattern of photoconversion).

relationships between cells located at different positions. To compare chondrocyte differentiation trajectories in different PD compartments, we constructed a single-cell graph of chondrogenesis in a stepwise manner, using the subspaces defined for groups of cells originating from consecutive spatial positions, adapting an approach used to reconstruct cell trajectories across time series (fig. S13A) (49). By allowing us to compare chondrogenesis gene expression despite the confounding effect of spatial compartmentalization, this method showed that spatially segregated chondrocytes display overlapping core chondrogenic transcriptomic states. Further analysis suggested that spatially segregated chondrocytes follow a largely stereotypic differentiation trajectory marked by the same core genes (fig. S13B).

Among those genes, *Sox9* is a key regulator of chondrogenesis (50). It shows a dynamic and complex regulation during limb development (51), involving a large regulatory landscape extending over 2 Mb and comprising dozens of potential limb enhancers (52, 53). To assess whether this landscape was identically activated along the limb PD axis, we leveraged published wild-type E11.5 limb bud single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) data (54) to identify chromatin-accessible elements. We used Seurat (55) to transfer the position along the PD axis and *Sox9* expression levels from our TATTOO-seq dataset to the scATAC-seq data and generated pseudo-bulk ATAC profiles based on position and *Sox9* expression (Fig. 4). Despite the limited number of cells in each category (<300 distal *Sox9*<sup>+</sup> cells), we were able to detect seven ATAC peaks ( $P < 0.05$ ; table S3) in the *Sox9* regulatory domain/topologically associating domain that are differentially represented along the PD axis (Fig. 4), five of which correspond to previously described *Sox9* enhancers (56). While preliminary, these data suggest that spatial information could be relayed to cell-fate controlling genes via position-specific enhancers, enabling specific modulation of other stereotypic differentiation regulatory networks.

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

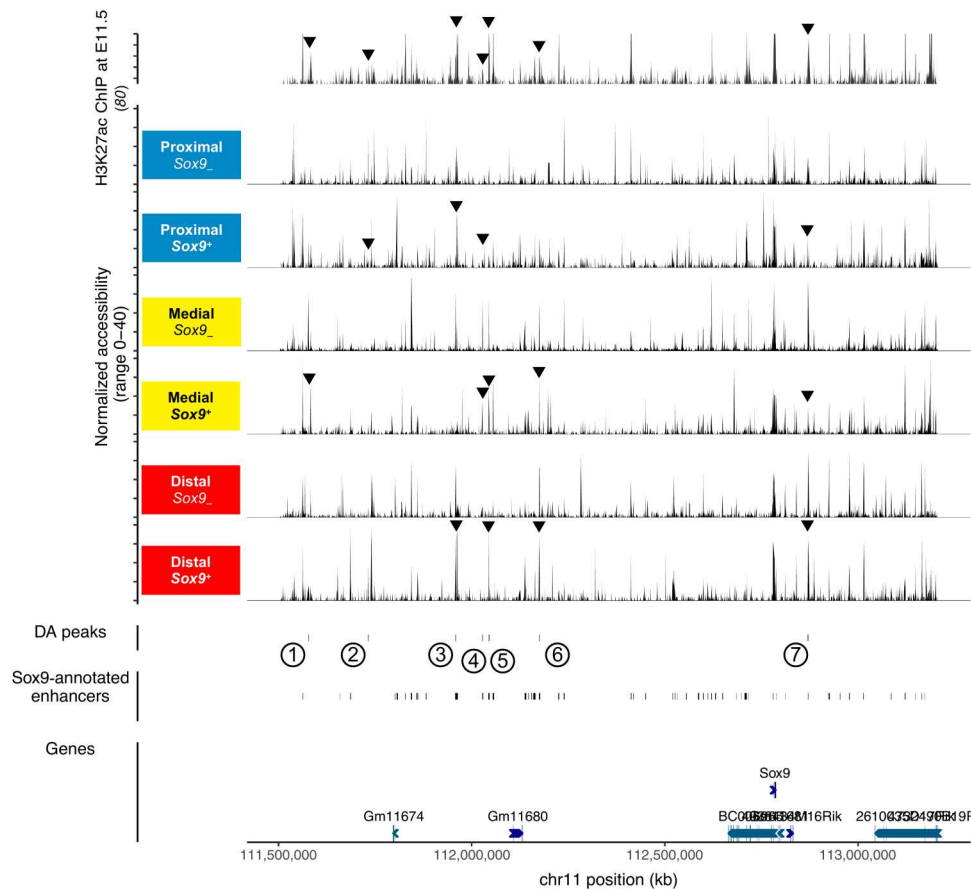
Patterning defects can lead to complex phenotypic consequences by perturbing cell differentiation programs, as well as their spatial modulation and coordination. We sought to apply TATTOO-seq to assess how limb patterning and differentiation are reorganized when positional information is perturbed. FGF8 is the main FGF molecule controlling the growth and patterning function of the AER (27). DEL(Poll-SHFM) (57) mice carry a deletion of the main *Fgf8* enhancers in the AER and fail to express *Fgf8* in the AER. They display shortened bones and lack proximal and anterior skeletal elements (humerus, radius, and digits 1 and 2) (Fig. 5A). To examine the molecular and positional changes that ultimately lead to the observed malformations, we produced a TATTOO-seq atlas for E11.5 DEL(Poll-SHFM) homozygous forelimb buds (embryos,  $n = 3$ ; limbs,  $n = 4$ ). The gating strategy (fig. S14) used for the mutant samples is the same as that of wild-type samples, but since mutant limbs are shorter, the AER pattern was applied to the whole limb, including the proximal cells. After filtering, we obtained high-quality transcriptomes for 3349 mutant cells. To compare the cell populations and their transcriptome between wild-type and mutant limbs, we produced a coarser joint clustering using Seurat (55), as the high-resolution clustering provided by MetaCell places mutant and wild-type cell in different metacells.

As a control, we first examined the expression of genes studied by single-gene in situ hybridization experiments in *Fgf8* mutants (27, 57). In our dataset, we found that *Dpca*, a bystander gene included in the DEL(Poll-SHFM) deletion, was almost undetectable in the mutant cells (Fig. 5B). Furthermore, as expected, *Fgf8* expression was abolished in the mutant AER cluster, while we detected a compensatory up-regulation of *Fgf4*, as reported previously (27).

We then investigated how reduced AER function affected the presence of the different cell types, their transcriptomes, and spatial distributions. The average gene expression for each Seurat cluster (both aggregated and split between by genotype) is indicated in table S4. The comparison of stage-specific marker expression with E10.5 data did not suggest a simple developmental delay (fig. S15). We first determined the relative fraction of wild-type and mutant cells for each cluster (Fig. 5C). Although we could not unambiguously calculate the expected cluster-wise fractions under the null hypothesis (AER pattern samples were enriched for distal cells, by up to twofold), several robust differences could not be explained by this limited sorting bias. At this level of resolution, we did not detect mutant-specific cell states. However, several wild-type cell states were almost absent from the mutant samples (Fig. 5C). Many depleted clusters (expressing *Zic3* distally, *Pax9* medially, and *Gsc* proximally) were located in the limb's anterior compartment (Fig. 5D), which likely explained the absence of anterior skeletal elements at later stages. However, we also noted that posteriorly located cell states (e.g., cluster 3) and some secondary signaling centers for connective tissue differentiation defined by *Cxcl12* and *Fgf18* (58, 59) were strongly diminished in the mutant. Examining chondrogenic cell states, we observed that although chondrogenesis appear to proceed overall normally in the mutant, the reduction in distal FGF signaling led to a reduced diversity of spatially distributed chondrogenic cell states: We found fewer distal *Sox9*-positive cell states expressing *Eomes* [base of digit 4 (60)] and *Irx1* [tip of digits 2 to 4 (61)] (Fig. 5, C and D). These distal chondrogenic cell states have been proposed to act as regulation centers expressing different BMP pathway components involved in joint formation, including *Gdf5*, *Inhb*, *Nog*, and *Chrdl1* (62–65). Although the depletion of these cell states is only partial, no expression of these signaling-related genes was detected in the mutant data (fig. S16A), which could account for the missing elbow joint in mutant limbs (27, 57). To facilitate the exploration of those cell states, fig. S17 shows the contribution of each wild-type metacell to each Seurat cluster.

Besides changes in specific cell state populations, we also identified a pervasive adaptation of the transcriptional states of otherwise conserved cell types, with the expression of multiple genes being affected by directly or indirectly by the reduction of FGF signaling from the AER (fig. S16C and table S5). Regarding the spatial distribution of the different limb cell states, only a few positional changes were detected along the PD and AER axes, and most of the spatial rearrangements were observed along the AP axis (Fig. 5E and fig. S18). Consistent with the absence of most anterior mesenchyme cell states, mutant cell states exhibited relatively more anterior colors than their wild-type counterparts. However, we also identified some cell states that were located more posteriorly in the mutant than in the wild-type, indicated complex alterations of patterning information, maybe resulting from the effect of reduced FGF signaling on the maintenance of ZPA activity and reduced *Shh* signaling (66, 67).





**Fig. 4. Putative spatial regulation of the accessibility landscape at the *Sox9* locus.** Cell identity and position were transferred from our TATTOO-seq data to limb scATAC-seq data (54) using Seurat's TransferData function. All peaks in the *Sox9* regulatory domain (chr11: 111,503,853 to 113,206,397, mm10) were tested for differential accessibility. The top track shows the H3K27ac chromatin immunoprecipitation (ChIP) sequencing signal in E11.5 whole limb buds (two replicates were aggregated) from (80). Differentially accessible (DA) peaks are indicated and previously described *Sox9* enhancers (56).

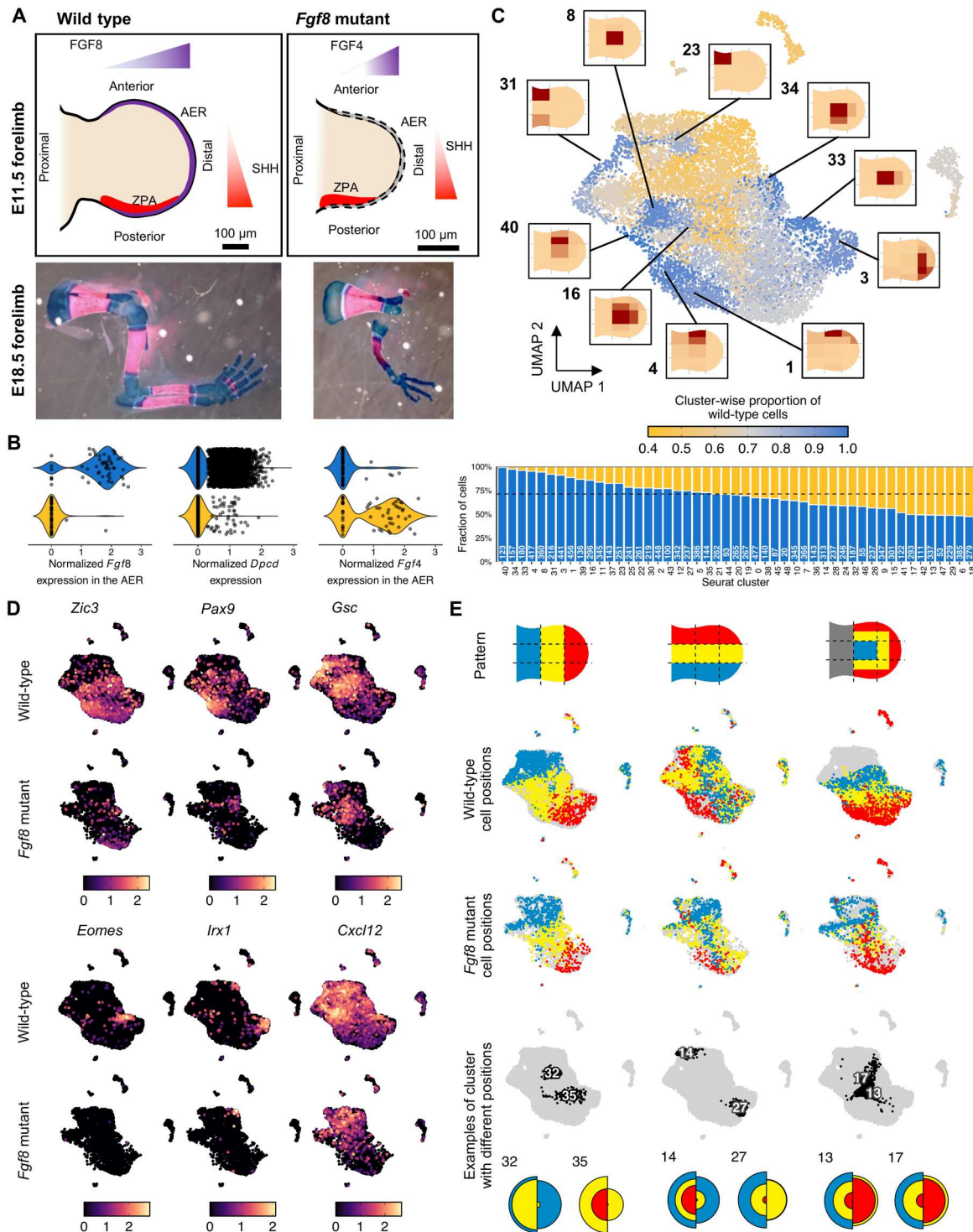
## DISCUSSION

siTATTOO-seq constitutes a simple, robust, and flexible approach to integrating spatial position and single-cell transcriptomes. The data and analyses provided here illustrate how TATTOO-seq can help making progress toward an integrated and high-resolution description of the dynamic processes involved in embryonic patterning. Orthogonal analyses such as NMF, topic modeling, and graph manifold assembly on TATTOO-seq data reveal the respective contributions of patterning and differentiation-specific gene modules to cell states, identify specific TFs that mediate spatial identity, and constitute a first step in disentangling spatial and cell type-specific gene regulatory programs. We show that positional information substantially modulates the transcriptome of limb chondrocytes and their core differentiation program and suggest that this core differentiation program operates at distinct positions through different, context-specific, modalities, which associate position-specific TFs to position-specific cis-regulatory elements. These differentially activated enhancers may act as specific positional integrators for common differentiation pathways and might be primary targets for regionalized natural and pathological variations in morphology (68). TATTOO-seq further offers a new integrative and multilevel description of developmental defects, as illustrated by the complex cellular, molecular, and spatial changes revealed

in the *Fgf8*-mutant limb buds. We expect that such a comprehensive, spatially resolved molecular data could provide a much finer-grained representation of "mutant phenotypes" and reveal how genetic or environmental perturbations may propagate through developmental gene regulatory networks and rewire them.

TATTOO-seq requires only standard equipment, namely, a confocal microscope and an FACS. While coupled here to a microplate-based single-cell technology, it can easily be adapted to other protocols, including droplet-based protocols, for example, using cell multiplexing/hashing strategies to label sorted cells of the same color. TATTOO-seq shares conceptual similarities to Tomo-seq (69), although it provides single-cell resolution and the simplicity and rapidity of optic labeling over physical sectioning. While the approach as shown here can be used on thick samples (up to 500  $\mu\text{m}$ ), the shape of the sample and, ideally, its patterning geometry must be compatible with the flat geometry of the microscope. For nonflat structures, the use of thick vibratome sections is recommended.

Because our approach relies on in silico cell aggregates, instead of individual cells, it is suitable for relatively sparse datasets and works without imputation. TATTOO-seq does not require spatial landmark reference genes nor does it assume the homogeneity of gene expression at a given position, which are strong limitations of



**Fig. 5. 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) Expression of *Fgf8*, *Dpccd*, and *Fgf4* for each genotype. (C) Top: Uniform Manifold Approximation and Projection (UMAP) of the combined wild-type and mutant datasets colored by cluster-wise fraction of wild-type cells. Bottom: The fraction of wild-type cells. (D) UMAP showing marker gene expression for mutant cell-depleted clusters. (E) Top: UMAP showing cell colors for each photoconversion pattern split by genotype. Bottom: Color distribution in some Seurat clusters for each genotype (left, wild-type mutant; right, *Fgf8* mutant).

retrospective mapping (9, 70) or pseudo-spatial ordering (71). Thus, it is immediately suitable for the analysis of mutant phenotypes without the need to build a mutant reference map of gene expression.

Although the spatial resolution used in this study is relatively coarse (50 to 100  $\mu\text{m}$ , depending on the photoconversion pattern), this resolution is in par with other spatial transcriptomics methods such as Slide-seq (15) or the commercial Visium platform (10x Genomics), which are based on barcoded beads or spots covering 50- to 100- $\mu\text{m}$  regions of a tissue section. While recent progress in bead/spot arraying can theoretically offer a resolution of 10  $\mu\text{m}$  (72), these approaches do not offer true single-cell resolution, are still better suited to identify the spatial distributions of single-cell clusters acquired separately, and may come at a much higher cost. In comparison, our approach could resolve finer spatial grids as specified by the user, since it is only limited by the precision of the photoconversion and the capacity of flow cytometry to detect graded fluorescence levels. However, as the number of relative fluorescence levels that can be reliably separated is not infinite (at least four with our current set up), a higher spatial resolution comes with a reduction of the area that can be studied. While a genetically encoded photoconvertible protein is at the basis of our method, the recent development of photoconvertible and clickable dyes or the use of photoconvertible membrane labeling dyes (73, 74) may allow further implementation of TATTOO-seq-based approaches in nontransgenic and nonmodel organisms or in patient-derived samples.

## MATERIALS AND METHODS

### Animals

Transgenic CAG-KikGR-1 mice (32) (KikGR1 hereafter, gift of A. Aulehla) were maintained inbred and genotyped either by polymerase chain reaction (PCR) using internal primers (p1, GAAATGAAGATCGAGCTGCGTATGG; p2, CACCCTTCAGCACTC CATCACGCAC) and a standard program with 65°C annealing temperature or by assessing green fluorescence in distal phalanx biopsy. DEL(PoL-SHFM) mice (57) were maintained on an inbred C57BL/6J background.

Homozygous KikGR1 males and wild-type females were crossed to generate control embryos. For mutant analysis, KikGR1/+; DEL(PoL-SHFM)/+ males and DEL(PoL-SHFM)/+ females were crossed to generate experimental and control embryos. E0.5 was defined as noon of the day when the vaginal plug was detected. Embryos were collected at E11.5 and dissected in ice-cold phosphate-buffered saline (PBS) supplemented with magnesium and calcium chloride. KikGR1 heterozygous embryos displayed strong and widespread green fluorescence in all tissues at all observed stages. All animal procedures were approved and performed accordingly to the ethical principles and regulatory guidelines and protocols set up by the Institutional Animal Care and Use Committees at the Institut Pasteur and the University of Chicago.

### Photoconversion and imaging

Samples were mounted in PBS supplemented with magnesium and calcium chloride between a glass slide and a coverslip separated by several layers of adhesive tape. Photoconversion was performed, and all fluorescence images were acquired using a Zeiss inverted confocal microscope (LSM 800) using a 10 $\times$  objective. Green and

red fluorescence images were obtained by excitation with the 488-nm (1 mW, gain = 580, digital gain = 2) and 561-nm (1 mW, gain = 580, digital gain = 2) laser diodes, respectively. The spectra of enhanced green fluorescent protein and mCherry were used for nonphotoconverted and photoconverted version of KikGR1, respectively, and adjusted such that they did not overlap. For green-to-red photoconversion of kikGR1 proteins, the 405-nm laser diode was used with variable power (100% laser power was equal to 5 mW). Full conversion (red color) was obtained after four iterations at 40% power, while partial conversion was obtained after four iterations at 10% power (for the PD pattern), 9% power (for the AP pattern), or 10 and 6% power (for the four-color AER pattern). For the photoconversion along the PD and AP axes, the size of the photoconverted regions was determined to be one-third of the length of the limb bud (from the tip to the flank or from the anterior-most part to the posterior-most part, respectively) as measured with the ZEN software. For the AER axis, the photoconverted regions were determined to be approximately one-fourth of the radius of the dissected autopods and zeugopods. While we could easily create 4° of photoconversion and separate them by flow cytometry, we grouped the two red-most colors into one color bin to increase the number of cells. We always applied the partial photoconversion (9 or 6% power) to both the “partially” and “fully” photoconverted target regions to ensure that cells at the border between the two regions are photoconverted to the same degree as the fully photoconverted region. To maximize the homogeneity of the photoconversion, we measured the thickness of the limb and chose the plane of photoconversion to coincide with the middle of the limb. The total duration of the photoconversion process was ~3 min in our particular setup. E10.5 samples were processed identically to E11.5 samples.

### Tissue dissociation and FACS

Limb buds were dissected and incubated for 5 min in 270  $\mu\text{l}$  of Liberase (0.22 Wunsch unit/ml; Roche, 5401119001) in EDTA-free, calcium-free, and magnesium-free PBS with agitation (900 rpm) at 37°C in a low-binding microcentrifuge tube (Biozym, 710176). Mechanical disruption was then performed by pipetting through a p200 gelatin-coated tip. The enzyme was inhibited by adding 30  $\mu\text{l}$  of 0.5 M EDTA, and volume was adjusted to 1 or 1.5 ml with calcium- and magnesium-free PBS. While this method allows the efficient recovery of mesenchymal cells, it is not optimized to recover ectodermal cells. Tight junctions could show limited sensitivity to short-duration collagenase treatment, thereby explaining why we recovered few ectodermal cells.

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

Debris and doublets were excluded from the analysis using a sequential gating strategy relying on first side scatter (SSC-A) versus forward scatter (FSC-A), followed by FSC-H versus FSC-W. Live cells were selected on the basis of live/dead staining gating on Violet-F-450/40-A (Calcein Violet AM) versus allophycocyanin-A (Sytox Red). Only Calcein-positive Sytox-negative cells were included in the analysis. Cells showed similar viability percentages regardless of the mounting and photoconversion status (~90 to 97%). Cells with a low fluorescence value in both green (FITC-A) and red (PE-A) channels were also excluded. No further binning on red and green fluorescence is performed. Cell colors are inferred a posteriori as described below (see the "Identification of cell colors" section). E10.5 samples were processed identically to E11.5 samples.

### Massively parallel scRNA-seq

Cells were processed by MARS-seq as previously described (33). Briefly, live single cells were sorted into 384-well capture plates containing lysis solution and barcoded poly(T) reverse transcription (RT) primers. UMI barcodes contain a cell-specific/well-specific label and an 8-base pair random molecular tag (RMT). After evaporation of the lysis buffer, the RT reaction was performed in the presence of External RNA Controls Consortium (ERCC) RNA Spike-in (Ambion). Unused RT primers were digested using Exonuclease I [New England Biolabs (NEB)], and wells were pooled by half plates. Second-strand cDNA synthesis was then performed, and products were in vitro transcribed overnight using T7 polymerase (NEB) for linear amplification. After RNA fragmentation, a partial Illumina Read1 sequencing adapter containing a plate-specific barcode was single strand-ligated using T4 RNA ligase I (NEB), and the product was reverse-transcribed. Last, the product was purified and PCR-amplified by PCR (14 cycles) with primers containing the Illumina P5-Read1 and P7-Read2 sequences. Concentration of the purified library was assessed with a Qubit fluorometer (Life Technologies), and mean molecule size was determined with a 2200 TapeStation instrument (Agilent Technologies). Libraries were pooled and paired-end sequenced using an Illumina NextSeq 500 sequencer at a median depth of ~45,000 reads per cell. Read1 was 70 nt long and covered the plate-specific barcode and the cDNA. Read2 was 16 nt long and covered the UMI and well-specific label. Raw and processed data sets are available from the Gene Expression Omnibus (GEO) repository (GSE202326).

### Quality check and read mapping

Reads were demultiplexed and count tables were built as described in (33) (scripts are available at <https://github.com/tanaylab/MARS-SEQ> and <https://doi.org/10.5281/zenodo.7316279>). Cell-specific/well-specific labels and RMTs were extracted. Reads with low-quality (Phred < 27) barcodes were discarded to prevent ambiguous or spurious assignments of reads to cells or unique molecules. Read2 reads whose cell-specific or well-specific labels were unknown or mutated were discarded. Read1 reads were mapped to the mouse genome (mm9) using Ensembl gene annotations (downloaded in July 2017) using Bowtie2 (75).

### Metacell analysis

We clustered the wild-type E11.5 dataset using MetaCell (35). Briefly, an ideal metacell is a group of cells whose expression profiles are statistically equivalent to independent sampling from a single underlying transcriptional state. This is achieved by first creating

a  $k$ -nearest neighbor ( $k$ -NN) graph of the individual cells based on expression similarity and by partitioning this graph into a disjoint sets of cells, metacells, which are both small (~100 cells each) and as close to the above ideal metacell as possible.

Because plates were processed in two batches, we filtered out cells with fewer than 2000 or 2500 UMIs (lower-quality transcriptomes) or more than 15,000 UMIs (potential doublets). We excluded mitochondrial genes (annotated with the prefix "mt-") from the analysis. Variable genes were selected using the parameters  $T_{vm} = 0.1$ ,  $T_{tot} = 50$ , and  $T_{top3} = 3$ . We then excluded cell cycles, histone genes, ribosomal genes, small nuclear riboprotein genes, some long noncoding RNA (annotated with the suffix "Rik"), and poorly supported gene models (annotated with the prefixes "RP-," "Gm," and "AC") from the variable genes. In addition, we identified genes that showed a batchy expression. To do so, we performed a first clustering, and for each metacell with a sufficient number of cells from both batches (>15), we repeatedly (10 times) randomly sampled and aggregated UMIs from 10 cells (UMIs from individual cells were downsampled to eliminate the effect of sequencing depth). If the median fold change for a gene was >1.6 in at least one metacell, then the gene was considered batchy and blacklisted for feature selection (but not omitted from the analysis). When excluding spatial genes, we additionally blacklisted genes with adjusted  $P < 0.01$  for the linear regression. A  $k$ -NN graph was built using Pearson correlation and  $k = 100$ , and 500 bootstrap iterations were performed (0.75 resampling probability). Metacells were built with a minimum size of 20,  $k = 30$ , and  $\alpha = 3$ . Inhomogeneous metacells were split, and outlier was removed ( $T_{lfc} = 3$ ). Otherwise, default parameters were used.

### Identification of cell colors

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

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

The distribution of this ratio exhibited the expected number of peaks (3 or 4) across all samples, although the exact location of the peaks and their separation varied. We therefore modeled the distribution of each individual limb separately as a mixture of skewed normal distributions whose parameters were estimated using an expectation-maximization algorithm implementer in the R package `mixsmsn` (76) ("`smsn.mix`", family = "Skew.normal", nu = 3). Three components were used for the PD and AP patterns, and four were used for the AER pattern. The color of each individual cell was determined using the posterior probabilities. For the AER pattern, we merged the two red-most compartments (high ratio-of-log values) to increase the number of cells in that category.

### Modeling the TATTOO-seq photoconversion process

We treated the limb-bud as a two-dimensional shape, constructed through concatenation of Bézier curves. We then assumed that the AP and PD photoconversion patterns split this shape into three regions of equal length along the relevant axis, whose fluorescence is mostly green, mostly yellow, or mostly red. On the basis of pictures of the photoconverted limb buds, we estimated that the part of the limb bud that was dissected for AER pattern photoconversion

roughly matched the yellow and red regions of the PD pattern. We then approximated the boundaries of the AER photoconversion regions using further concatenations of Bézier curves, choosing the parameters that best fit the pictures. Note that although the AER photoconversion split the limb bud into four regions, we combined the two red-most regions in our model, as the number of cells collected that had red-most fluorescence was extremely low.

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

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

### Inferring cell positions

The fluorescence information provided for each cell by TATTOO-seq only limits its position to one of three or four broad spatial regions. To gain a finer spatial resolution, we used the reproducibility of patterning of the mouse embryo system. This allowed us to pool together spatial information collected across multiple limb buds and multiple photoconversion patterns and infer a distribution over the limb bud's sbins.

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

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

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

$$P_{\text{emp},\lambda}(F)$$

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

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

where  $\theta$  is a parameter. Last, the photoconversion model described in the previous section implies a distribution over the measured

fluorescence given the sbin and the photoconversion patterns

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

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

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

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

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

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

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

In addition, the value of the probability of each sbin was constrained to a segment that strictly contain the  $[0,1]$  segment

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

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

As an initial validation for the cell positioning algorithm, we used all collected cells, effectively inferring the fraction of the cells residing in each of the spatial bins. Assuming that the mean cell size is similar across all bins, the area-normalized fraction provides an estimate of the thickness of the limb bud thickness at each spatial bin. As expected, the inferred limb bud thickness decreases as we move away from the embryo's trunk and toward the edge of the limb bud.

### Inferring positions of metacells

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

### Metacell-specific TFs

The metacell specificity of TFs was estimated using a method similar to the "roc" method in Seurat. For each gene, the empirical cumulative distribution function (CDF) was computed on the basis

of the rescaled (from 0 to 1) gene expression in metacells. The area under the curve (AUC) was then computed. Genes that have very specific patterns of expression (i.e., expression in very few metacells) display large values of AUC because most metacells exhibit low values of gene expression.

### Chondrogenic trajectory and pseudo-time

Each chondrogenic progenitor was assigned to a compartment along the PD axis based on its color. Cells originating from other photoconversion patterns were assigned to the compartment for which their metacells had the highest probability. We performed principal components analysis on the top 2000 variable genes within each compartment (thus excluding variation resulting from spatial compartment-specific genes) and built graphs connecting each cell to its 40 nearest neighbors within its compartment. Projecting medial cells into the distal principal component (PC) space, we built a graph between distal and medial cells. Similarly, we built a graph connecting medial and proximal cells in the medial PC space. Edges were then filtered to keep at most 10 mutual edges. The resulting graph was projected using the DrL graph layout. ElPigraph (77) was then used to compute the pseudo-time for each cell. Gene expression as a function of the pseudo-time was computed and smoothed using loess regression.

### Supplementary Materials

This PDF file includes:

Figs. S1 to S18

References

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S5

### REFERENCES AND NOTES

1. L. Wolpert, Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47 (1969).
2. F. R. Goodman, Congenital abnormalities of body patterning: Embryology revisited. *Lancet* **362**, 651–662 (2003).
3. L. A. Lettice, A. E. Hill, P. S. Devenney, R. E. Hill, Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum. Mol. Genet.* **17**, 978–985 (2008).
4. E. Z. Kvon, O. K. Kamneva, U. S. Melo, I. Barozzi, M. Osterwalder, B. J. Mannion, V. Tissières, C. S. Pickle, I. Plajzer-Frick, E. A. Lee, M. Kato, T. H. Garvin, J. A. Akiyama, V. Afzal, J. Lopez-Rios, E. M. Rubin, D. E. Dickel, L. A. Pennacchio, A. Visel, Progressive loss of function in a limb enhancer during snake evolution. *Cell* **167**, 633–642.e11 (2016).
5. M. Towers, Evolution of antero-posterior patterning of the limb: Insights from the chick. *Genesis* **56**, e23047 (2018).
6. A. Abzhanov, M. Protas, B. R. Grant, P. R. Grant, C. J. Tabin, Bmp4 and morphological variation of beaks in Darwin's finches. *Science* **305**, 1462–1465 (2004).
7. A. Zuniga, Next generation limb development and evolution: Old questions, new perspectives. *Development* **142**, 3810–3820 (2015).
8. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
9. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
10. K. Achim, J.-B. Pettit, L. R. Saraiva, D. Gaviouchkina, T. Larsson, D. Arendt, J. C. Marioni, High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
11. E. Lubeck, A. F. Coskun, T. Zhiyentayev, M. Ahmad, L. Cai, Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
12. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa609 (2015).
13. J. H. Lee, E. R. Daugherty, J. Scheiman, R. Kalhor, T. C. Ferrante, R. Terry, B. M. Turczyk, J. L. Yang, H. S. Lee, J. Aach, K. Zhang, G. M. Church, Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).
14. X. Wang, W. E. Allen, M. A. Wright, E. L. Sylwestrak, N. Samusik, S. Vesuna, K. Evans, C. Liu, C. Ramakrishnan, J. Liu, G. P. Nolan, F.-A. Bava, K. Deisseroth, Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
15. S. G. Rodrigues, R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, C. R. Vanderburg, J. Welch, L. M. Chen, F. Chen, E. Z. Macosko, Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
16. A. Chen, S. Liao, M. Cheng, K. Ma, L. Wu, Y. Lai, X. Qiu, J. Yang, J. Xu, S. Hao, X. Wang, H. Lu, X. Chen, X. Liu, X. Huang, Z. Li, Y. Hong, Y. Jiang, J. Peng, S. Liu, M. Shen, C. Liu, Q. Li, Y. Yuan, X. Wei, H. Zheng, W. Feng, Z. Wang, Y. Liu, Z. Wang, Y. Yang, H. Xiang, L. Han, B. Qin, P. Guo, G. Lai, P. Muñoz-Cánoves, P. H. Maxwell, J. P. Thiery, Q.-F. Wu, F. Zhao, B. Chen, M. Li, X. Dai, S. Wang, H. Kuang, J. Hui, L. Wang, J.-F. Fei, O. Wang, X. Wei, H. Lu, B. Wang, S. Liu, Y. Gu, M. Ni, W. Zhang, F. Mu, Y. Yin, H. Yang, M. Lisby, R. J. Cornall, J. Mulder, M. Uhlén, M. A. Esteban, Y. Li, L. Liu, X. Xu, J. Wang, Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).
17. C. Tabin, L. Wolpert, Rethinking the proximodistal axis of the vertebrate limb in the molecular era. *Genes Dev.* **21**, 1433–1442 (2007).
18. H. Shimizu, S. Yokoyama, H. Asahara, Growth and differentiation of the developing limb bud from the perspective of chondrogenesis. *Dev. Growth Differ.* **49**, 449–454 (2007).
19. M. Buckingham, P. W. J. Rigby, Gene regulatory networks and transcriptional mechanisms that control myogenesis. *Dev. Cell* **28**, 225–238 (2014).
20. S. Nassari, D. Duprez, C. Fournier-Thibault, Non-myogenic contribution to muscle development and homeostasis: The role of connective tissues. *Front. Cell Dev. Biol.* **5**, 22 (2017).
21. K. L. Cooper, K. E. Sears, A. Uygun, J. Maier, K.-S. Baczkowski, M. Brosnahan, D. Antczak, J. A. Skidmore, C. J. Tabin, Patterning and post-patterning modes of evolutionary digit loss in mammals. *Nature* **511**, 41–45 (2014).
22. D. Basel, A. DePaepe, M. Kilpatrick, P. Tsipouras, Split hand foot malformation is associated with a reduced level of Dactylin gene expression. *Clin. Genet.* **64**, 350–354 (2003).
23. L. A. Lettice, S. J. H. Heaney, L. A. Purdie, L. Li, P. de Beer, B. A. Oostra, D. Goode, G. Elgar, R. E. Hill, E. de Graaff, A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
24. J. Kohlhaase, A. Wischermann, H. Reichenbach, U. Froster, W. Engel, Mutations in the SALL1 putative transcription factor gene cause Townes-Brocks syndrome. *Nat. Genet.* **18**, 81–83 (1998).
25. P. Hill, B. Wang, U. Rüther, The molecular basis of Pallister Hall associated polydactyly. *Hum. Mol. Genet.* **16**, 2089–2096 (2007).
26. R. D. Riddle, R. L. Johnson, E. Laufer, C. Tabin, Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–1416 (1993).
27. M. Lewandoski, X. Sun, G. R. Martin, Fgf8 signalling from the AER is essential for normal limb development. *Nat. Genet.* **26**, 460–463 (2000).
28. H. Akiyama, J.-E. Kim, K. Nakashima, G. Balmes, N. Iwai, J. M. Deng, Z. Zhang, J. F. Martin, R. R. Behringer, T. Nakamura, B. De Crombrughe, Osteo-chondroprogenitor cells are derived from Sox9 expressing precursors. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14665–14670 (2005).
29. Y. J. Wang, R. M. Belflower, Y.-F. Dong, E. M. Schwarz, R. J. O'Keefe, H. Drissi, Runx1/AML1/Cbfa2 mediates onset of mesenchymal cell differentiation toward chondrogenesis. *J. Bone Miner. Res.* **20**, 1624–1636 (2005).
30. I. Tzchori, T. F. Day, P. J. Carolan, Y. Zhao, C. A. Wassif, L. Q. Li, M. Lewandoski, M. Gorivodsky, P. E. Love, F. D. Porter, H. Westphal, Y. Yang, LIM homeobox transcription factors integrate signaling events that control three-dimensional limb patterning and growth. *Development* **136**, 1375–1385 (2009).
31. S. Yokoyama, S. Furukawa, S. Kitada, M. Mori, T. Saito, K. Kawakami, J. C. I. Belmonte, Y. Kawakami, Y. Ito, T. Sato, H. Asahara, Analysis of transcription factors expressed at the anterior mouse limb bud. *PLOS ONE* **12**, e0175673 (2017).
32. S. Nowotschin, A.-K. Hadjantonakis, Use of KikGR a photoconvertible green-to-red fluorescent protein for cell labeling and lineage analysis in ES cells and mouse embryos. *BMC Dev. Biol.* **9**, 49 (2009).
33. D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay, I. Amit, Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
34. C. Medaglia, A. Giladi, L. Stoler-Barak, M. De Giovanni, T. M. Salame, A. Biram, E. David, H. Li, M. Iannacone, Z. Shulman, I. Amit, Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* **358**, 1622–1626 (2017).

35. Y. Baran, A. Bercovich, A. Sebe-Pedros, Y. Lubling, A. Giladi, E. Chomsky, Z. Meir, M. Hoichman, A. Lifshitz, A. Tanay, MetaCell: Analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol.* **20**, 206 (2019).
36. C. A. Loomis, R. A. Kimmel, C. X. Tong, J. Michaud, A. L. Joyner, Analysis of the genetic pathway leading to formation of ectopic apical ectodermal ridges in mouse *Engrailed-1* mutant limbs. *Development* **125**, 1137–1148 (1998).
37. L. A. Lettice, I. Williamson, J. H. Wiltshire, S. Peluso, P. S. Devenney, A. E. Hill, A. Essafi, J. Hagman, R. Mort, G. Grimes, C. L. DeAngelis, R. E. Hill, Opposing functions of the ETS factor family define *Shh* spatial expression in limb buds and underlie polydactyly. *Dev. Cell* **22**, 459–467 (2012).
38. M. Osterwalder, D. Speziale, M. Shoukry, R. Mohan, R. Ivanek, M. Kohler, C. Beisel, X. Wen, S. J. Scales, V. M. Christoffels, A. Visel, J. Lopez-Rios, R. Zeller, HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. *Dev. Cell* **31**, 345–357 (2014).
39. L. A. Lettice, P. Devenney, C. De Angelis, R. E. Hill, The conserved sonic hedgehog limb enhancer consists of discrete functional elements that regulate precise spatial expression. *Cell Rep.* **20**, 1396–1408 (2017).
40. T. Sagai, M. Hosoya, Y. Mizushima, M. Tamura, T. Shiroishi, Elimination of a long-range cis-regulatory module causes complete loss of limb-specific *Shh* expression and truncation of the mouse limb. *Development* **132**, 797–803 (2005).
41. C. A. Gurnett, A. M. Bowcock, F. R. Dietz, J. A. Morcuende, J. C. Murray, M. B. Dobbs, Two novel point mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial polydactyly. *Am. J. Med. Genet. Part A.* **143A**, 27–32 (2007).
42. D. Wieczorek, B. Pawlik, Y. Li, N. A. Akarsu, A. Caliebe, K. J. W. May, B. Schweiger, F. R. Vargas, S. Balci, G. Gilleßen-Kaesbach, B. Wollnik, A specific mutation in the distant sonic hedgehog (*SHH*) cis-regulator (ZRS) causes Werner mesomelic syndrome (WMS) while complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb. *Hum. Mutat.* **31**, 81–89 (2010).
43. C. Xu, X. Yang, H. Zhou, Y. Li, C. Xing, T. Zhou, D. Zhong, C. Lian, M. Yan, T. Chen, Z. Liao, B. Gao, D. Su, T. Wang, S. Sharma, C. Mohan, N. Ahituv, S. Malik, Q. Z. Li, P. Su, A novel ZRS variant causes preaxial polydactyly type I by increased sonic hedgehog expression in the developing limb bud. *Genet. Med.* **22**, 189–198 (2020).
44. T. Ikeda, H. Kawaguchi, S. Kamekura, N. Ogata, Y. Mori, K. Nakamura, S. Ikegawa, U.-I. Chung, Distinct roles of *Sox5*, *Sox6*, and *Sox9* in different stages of chondrogenic differentiation. *J. Bone Miner. Metab.* **23**, 337–340 (2005).
45. D. Y. Soung, Y. Dong, Y. J. Wang, M. J. Zusick, E. M. Schwarz, R. J. O’Keefe, H. Drissi, *Runx3/AML2/Cbfa3* regulates early and late chondrocyte differentiation. *J. Bone Miner. Res.* **22**, 1260–1270 (2007).
46. R. Schweitzer, J. H. Chung, L. C. Murtaugh, A. E. Brent, V. Rosen, E. N. Olson, A. Lassar, C. J. Tabin, Analysis of the tendon cell fate using *Scleraxis*, a specific marker for tendons and ligaments. *Development* **128**, 3855–3866 (2001).
47. K. K. Dey, C. J. Hsiao, M. Stephens, Visualizing the structure of RNA-seq expression data using grade of membership models. *PLoS Genet.* **13**, e1006599 (2017).
48. P. Carbonetto, K. Luo, K. Dey, J. Hsiao, M. Stephens, fastTopics: Fast algorithms for fitting topic models and non-negative matrix factorizations to count data. R package version 0.4–11. (2021).
49. D. E. Wagner, C. Weinreb, Z. M. Collins, J. A. Briggs, S. G. Megason, A. M. Klein, Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
50. E. Wright, M. R. Hargrave, J. Christiansen, D. Cooper, J. Kun, T. Evans, U. Gangadharan, A. Greenfield, P. Koopman, The *Sry*-related gene *Sox9* is expressed during chondrogenesis in mouse embryos. *Nat. Genet.* **9**, 15–20 (1995).
51. B. Boehm, M. Rautschka, L. Quintana, J. Raspopovic, Ž. Jan, J. Sharpe, A landmark-free morphometric staging system for the mouse limb bud. *Development* **138**, 1227–1234 (2011).
52. G. Andrey, R. Schöpflin, I. Jerković, V. Heinrich, D. M. Ibrahim, C. Paliou, M. Hochradel, B. Timmermann, S. Haas, M. Vingron, S. Mundlos, Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).
53. The ENCODE Project Consortium, J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shores, J. Adrian, T. Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He, M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X.-O. Zhang, S. I. Elhajjaj, J. Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri, J. Zhang, A. Victorson, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lécuycy, D. M. Gilbert, J. Dekker, J. Rinn, E. M. Mendenhall, J. R. Ecker, M. Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigó, P. J. Farnham, J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E. Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos, Z. Weng, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
54. I. Desanlis, Y. Kherdjemil, A. Mayran, Y. Bouklouch, C. Gentile, R. Sheth, R. Zeller, J. Drouin, M. Kmita, HOX13-dependent chromatin accessibility underlies the transition towards the digit development program. *Nat. Commun.* **11**, 2491 (2020).
55. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoekius, P. Smibert, R. Satija, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
56. A. Despang, R. Schöpflin, M. Franke, S. Ali, I. Jerković, C. Paliou, W.-L. Chan, B. Timmermann, L. Wittler, M. Vingron, S. Mundlos, D. M. Ibrahim, Functional dissection of the *Sox9–Kcnj2* locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* **51**, 1263–1271 (2019).
57. M. Marinić, T. Aktas, S. Ruf, F. Spitz, An integrated holo-enhancer unit defines tissue and gene specificity of the *Fgf8* regulatory landscape. *Dev. Cell* **24**, 530–542 (2013).
58. Z. Liu, J. Xu, J. S. Colvin, D. M. Ornitz, Coordination of chondrogenesis and osteogenesis by fibroblast growth factor 18. *Genes Dev.* **16**, 859–869 (2002).
59. S. Nassari, C. Blavet, M.-A. Bonnin, S. Stricker, D. Duprez, C. Fournier-Thibault, The chemokines CXCL12 and CXCL14 differentially regulate connective tissue markers during limb development. *Sci. Rep.* **7**, 17279 (2017).
60. G. S. Kwon, A.-K. Hadjantonakis, Eomes::GFP—A tool for live imaging cells of the trophoblast, primitive streak, and telencephalon in the mouse embryo. *Genesis* **45**, 208–217 (2007).
61. A. Zülch, M. B. Becker, P. Gruss, Expression pattern of *Irx1* and *Irx2* during mouse digit development. *Mech. Dev.* **106**, 159–162 (2001).
62. E. E. Storm, D. M. Kingsley, GDF5 coordinates bone and joint formation during digit development. *Dev. Biol.* **209**, 11–27 (1999).
63. R. Merino, D. Macias, Y. Gañan, J. Rodriguez-Leon, A. N. Economides, C. Rodriguez-Esteban, J. C. Izpisua-Belmonte, J. M. Hurle, Control of digit formation by activin signalling. *Development* **126**, 2161–2170 (1999).
64. J. M. Allen, E. McGlenn, A. Hill, M. L. Warman, Autopodial development is selectively impaired by misexpression of chordin-like 1 in the chick limb. *Dev. Biol.* **381**, 159–169 (2013).
65. E. Degenkolbe, J. König, J. Zimmer, M. Walther, C. Reißner, J. Nickel, F. Plöger, J. Raspopovic, J. Sharpe, K. Dathe, J. T. Hecht, S. Mundlos, S. C. Doelken, P. Seemann, A GDF5 point mutation strikes twice - Causing BDA1 and SYNS2. *PLoS Genet.* **9**, e1003846 (2013).
66. A. M. Moon, M. R. Capechchi, *Fgf8* is required for outgrowth and patterning of the limbs. *Nat. Genet.* **26**, 455–459 (2000).
67. P. H. Crossley, G. Minowada, C. A. MacArthur, G. R. Martin, Roles for FGF8 in the induction, initiation, and maintenance of chick limb development. *Cell* **84**, 127–136 (1996).
68. J. D. White, K. Indencleef, S. Naqvi, R. J. Eller, H. Hoskens, J. Roosenboom, M. K. Lee, J. Li, J. Mohammed, S. Richmond, E. E. Quillen, H. L. Norton, E. Feingold, T. Swigut, M. L. Marazita, H. Peeters, G. Hens, J. R. Shaffer, J. Wysocka, S. Walsh, S. M. Weinberg, M. D. Shriver, P. Claes, Insights into the genetic architecture of the human face. *Nat. Genet.* **53**, 45–53 (2021).
69. F. Kruse, J. P. Junker, A. van Oudenaarden, J. Bakkers, Tomo-seq: A method to obtain genome-wide expression data with spatial resolution. *Methods Cell Biol.* **135**, 299–307 (2016).
70. N. Karaiskos, P. Wahle, J. Alles, A. Boltengagen, S. Ayoub, C. Kipar, C. Kocks, N. Rajewsky, R. P. Zinnen, The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **358**, 194–199 (2017).
71. M. Nitzan, N. Karaiskos, N. Friedman, N. Rajewsky, Gene expression cartography. *Nature* **576**, 132–137 (2019).
72. R. R. Stickers, E. Murray, P. Kumar, J. Li, J. L. Marshall, D. J. Di Bella, P. Arlotta, E. Z. Macosko, F. Chen, Highly sensitive spatial transcriptomics at near-cellular resolution with slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
73. J. V. Jun, C. M. Haney, R. J. Karpowicz, S. Giannakoulis, V. M.-Y. Lee, E. J. Pettersson, D. M. Chenoweth, A “clickable” photoconvertible small fluorescent molecule as a minimalist probe for tracking individual biomolecule complexes. *J. Am. Chem. Soc.* **141**, 1893–1897 (2019).
74. A. L. Carlson, J. Fujisaki, J. Wu, J. M. Runnels, R. Turcotte, C. Lo Celso, D. T. Scadden, T. B. Strom, C. P. Lin, Tracking single cells in live animals using a photoconvertible near-infrared cell membrane label. *PLoS ONE* **8**, e69257 (2013).
75. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
76. M. O. Prates, C. R. B. Cabral, V. H. Lachos, Mixmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *J. Stat. Softw.* **54**, 1–20 (2013).
77. L. Albergante, E. Mirkes, J. Bac, H. Chen, A. Martin, L. Faure, E. Barillot, L. Pinello, A. Gorban, A. Zinovyev, Robust and scalable learning of complex intrinsic dataset geometry via EPI-Graph. *Entropy* **22**, 296 (2020).
78. N. Mercader, L. Selleri, L. M. Criado, P. Pallares, C. Parras, M. L. Cleary, M. Torres, Ectopic *Meis1* expression in the mouse limb bud alters P-D patterning in a *Pbx1*-independent manner. *Int. J. Dev. Biol.* **53**, 1483–1494 (2009).

79. S. Yokoyama, Y. Ito, H. Ueno-Kudoh, H. Shimizu, K. Uchibe, S. Albini, K. Mitsuoaka, S. Miyaki, M. Kiso, A. Nagai, T. Hikata, T. Osada, N. Fukuda, S. Yamashita, D. Harada, V. Mezzano, M. Kasai, P. L. Puri, Y. Hayashizaki, H. Okado, M. Hashimoto, H. Asahara, A systems approach reveals that the myogenesis genome network is regulated by the transcriptional repressor RP58. *Dev. Cell* **17**, 836–848 (2009).
80. J. Malkmus, L. Ramos Martins, S. Jhanwar, B. Kircher, V. Palacio, R. Sheth, F. Leal, A. Duchesne, J. Lopez-Rios, K. A. Peterson, R. Reinhardt, K. Onimaru, M. J. Cohn, A. Zuniga, R. Zeller, Spatial regulation by multiple *Gremlin1* enhancers provides digit development with *cis*-regulatory robustness and evolutionary plasticity. *Nat. Commun.* **12**, 5557 (2021).
81. R. D. Riddle, M. Ensini, C. Nelson, T. Tsuchida, T. M. Jessell, C. Tabin, Induction of the LIM homeobox gene *Lmx1* by WNT6a establishes dorsoventral pattern in the vertebrate limb. *Cell* **83**, 631–640 (1995).
82. Y. Lan, P. D. Kingsley, E. S. Cho, R. Jiang, *Osr2*, a new mouse gene related to *Drosophila* *odd-skipped*, exhibits dynamic expression patterns during craniofacial, limb, and kidney development. *Mech. Dev.* **107**, 175–179 (2001).
83. P. Lu, Y. Yu, Y. Perdue, Z. Werb, The apical ectodermal ridge is a timer for generating distal limb progenitors. *Development* **135**, 1395–1405 (2008).
84. J. M. Salbaum, *Punc*, a novel mouse gene of the immunoglobulin superfamily, is expressed predominantly in the developing nervous system. *Mech. Dev.* **71**, 201–204 (1998).

**Acknowledgments:** We thank T. Fujimori and the RIKEN Laboratory for Animal Resources and Genetic Engineering for giving us access to the KikGR-1 transgenic line (reference CDB0201T-1) and A. Aulehla (EMBL) for sharing with us these mice. We acknowledge greatly the contribution of the members of the animal facilities at the Institut Pasteur. We thank members of the Spitz, Marlow, and Tanay laboratories for sharing reagents, ideas, and comments. We also thank C. Moreau for helping with the pilot photoconversion experiments. We particularly

acknowledge the contribution of E. Brient-Litzler and J.-C. Olivo-Marin (Institut Pasteur, CITech) in the development of a single-cell initiative and single-cell technological platform at the Institut Pasteur, which was essential to this project. **Funding:** The development of this platform was supported by the Region Ile-de-France (program Sesame 16016329 “Paris Single Cell Centre” and DIM Elicit 2017 “The molecular microscope: Spatially-resolved single-cell transcriptomics in 3D samples”). This work was directly supported by the Institut Pasteur, the University of Chicago and la Fondation pour la Recherche Médicale (programme Equipe Labellisée 2018, DEQ20180339222, to F.S.). **Author contributions:** F.S. and H.M. conceived the project. F.S., A.T., and H.M. designed the experimental strategies and supervised the experiments and data analyses. S.B. and B.S. performed all experiments and optimized the protocols. E.C. and A.T. designed and developed the computational tools to map cells on a spatial model, with help of S.B. S.B. designed and performed all additional computational data analysis. Y.L.-M. processed the different scRNA-seq data and provided additional computational support. S.S. and S.N. provided expert advice for FACS setup, processing, and analysis. S.B. and F.S. wrote the manuscript with input from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Original raw and processed data described in the paper can be found on the GEO repository (accession number: GSE202326). Previously published publicly available datasets used in the described analysis are also available on GEO: limb scATAC-seq (GSE145657) and limb H3K27ac chromatin immunoprecipitation sequencing (GSE151488). All other relevant data supporting the key findings of this study are available within the article or in the Supplementary Materials.

Submitted 18 May 2022

Accepted 14 November 2022

Published 14 December 2022

10.1126/sciadv.add0695



## TATTOO-seq delineates spatial and cell type–specific regulatory programs in the developing limb

Sbastien Bastide, Elad Chomsky, Baptiste Saudemont, Yann Loe-Mie, Sandrine Schmutz, Sophie Novault, Heather Marlow, Amos Tanay, and Franois Spitz

*Sci. Adv.*, **8** (50), eadd0695.  
DOI: 10.1126/sciadv.add0695

### View the article online

<https://www.science.org/doi/10.1126/sciadv.add0695>

### Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

---

*Science Advances* (ISSN ) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.  
Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY-NC).