nature

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41586-019-0969-x

In the format provided by the authors and unedited.

The single-cell transcriptional landscape of mammalian organogenesis

Junyue Cao[1,2,10], Malte Spielmann[1,10], Xiaojie Qiu[1,2], Xingfan Huang[1,3], Daniel M. Ibrahim[4,5], Andrew J. Hill[1], Fan Zhang[6], Stefan Mundlos[4,5], Lena Christiansen[6], Frank J. Steemers[6], Cole Trapnell[1,7,8]* & Jay Shendure[1,7,8,9]*

[1]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [2]Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA. [3]Department of Computer Science, University of Washington, Seattle, WA, USA. [4]Max Planck Institute for Molecular Genetics, RG Development & Disease, Berlin, Germany. [5]Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany. [6]Illumina, San Diego, CA, USA. [7]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. [8]Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA. [9]Howard Hughes Medical Institute, Seattle, WA, USA. [10]These authors contributed equally: Junyue Cao, Malte Spielmann. *e-mail: coletrap@uw.edu; shendure@uw.edu

Supplementary Note 1

We first sought to apply Monocle 3 to a single major cell type, cluster 25, whose 26,559 cells we annotate as limb bud mesenchyme on the basis of *Hoxd13, Fgf10* and *Lmx1b* expression (**Supplementary Table 4**). Visualizing the trajectory of cells of this cluster illustrates the dramatic expansion of limb mesenchymal cells over developmental time, with the main outgrowth between E10.5 and E12.5 (**Extended Data Fig. 7a**). Gene expression is highly dynamic during this expansion, with the levels of 4,763 protein-coding genes changing (FDR of 1%; **Supplementary Table 7**). The early stages of limb mesenchyme development are characterized by expression of some expected genes such as *Tbx15*[1], and *Gpc3*[2] and the later stages by *Msx1*[3], *Epha4*[4] and *Dach1*[5] (**Extended Data Fig. 7b**), but the vast majority of dynamically expressed genes are novel. Transcription factors significantly upregulated during limb mesenchyme development included those with roles in chondrocyte differentiation (*e.g. Sox9*[6] and *Yap1*[7]), muscle differentiation (*e.g. Tead4*[8]), and wound healing and limb regeneration (*e.g. Smarcd1*[9]) (**Extended Data Fig. 7c**).

Interestingly, forelimb and hindlimb cells were not obviously separated by unsupervised clustering (**Extended Data Fig. 7d**) or trajectory analysis (**Extended Data Fig. 7e**), but could be distinguished by the mutually exclusive expression of *Tbx5* in forelimb (2,085 cells, 7.9% of all limb mesenchyme cells) and *Pitx1* in hindlimb (1,885 cells, 7.1% of all limb mesenchyme cells) with only 22 cells expressing both markers (0.08% of all limb mesenchyme cells vs. ~0.6% expected if they were independent; **Extended Data Fig. 7f**)[10]. 285 genes were differentially expressed between cells assigned to the forelimb and hindlimb in this way (**Extended Data Fig. 7g**, **Supplementary Table 8**). Known marker genes such as *Tbx4* and the genes of the Hoxc cluster (*Hoxc4-10*)[11] were upregulated in hindlimb cells as expected, but we also identified genes not previously shown to be differentially expressed. For example, we observed *Epha3* and *Hs3st3b1* to be 5-fold enriched in forelimb, and *Pcdh17* and *Igf1* to be 3-fold enriched in hindlimb.

Although developmental time is a major axis of variation in the limb mesenchyme trajectory (**Extended Data Fig. 7a**), there is clearly additional structure. At least some of this appears to correspond to the two main spatial axes of limb development: the proximal-distal axis (the primary direction of outgrowth) and the anterior-posterior axis (corresponding to the five digits)[10]. With Monocle 3, we applied Moran's I test[12] to detect genes exhibiting autocorrelation across the limb mesenchyme trajectory (*i.e.* genes expressed in similar regions of the principal graph). We found, for example, that cells expressing *Sox6* and *Sox9* (proximal markers)[13,14], *Hoxd13* and *Tfap2b* (distal markers)[15], *Pax9* and *Alx4* (anterior markers), and *Shh* and *Hand2* (posterior markers), were differentially distributed across the trajectory (**Extended Data Fig. 7h**, **Extended Data Fig. 7i**). Whole-mount *in situ* hybridization of *Hoxd13* (a known distal marker) and *Cpa2* (a novel marker whose distribution in the Monocle 3 trajectory was similar to that of known distal markers), confirmed that both genes are expressed in the distal limb mesenchyme between E10.5 and E13.5 (**Extended Data Fig. 7j-l**). Altogether, we identified 1,783 genes exhibiting variable expression across the limb mesenchymal trajectory (FDR of 1%; Moran's I > 0.01). These genes clustered into eight patterns of expression, several of which matched the distributions of known markers for the proximal-distal and anterior-posterior axes (**Extended Data Fig. 7m**, **Supplementary Table**

9). These analyses illustrate how this single cell atlas of mouse organogenesis can be used to characterize the spatiotemporal dynamics of gene expression in specific systems.

References:

1. Singh, M. K. *et al.* The T-box transcription factor Tbx15 is required for skeletal development. *Mech. Dev.* **122,** 131–144 (2005).
2. Paine-Saunders, S., Viviano, B. L., Zupicich, J., Skarnes, W. C. & Saunders, S. glypican-3 controls cellular responses to Bmp4 in limb patterning and skeletal development. *Dev. Biol.* **225,** 179–187 (2000).
3. Hara, K. & Ide, H. Msx1 expressing mesoderm is important for the apical ectodermal ridge (AER)-signal transfer in chick limb development. *Dev. Growth Differ.* **39,** 705–714 (1997).
4. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161,** 1012–1025 (2015).
5. Davis, R. J. *et al.* Dach1 mutant mice bear no gross abnormalities in eye, limb, and brain development and exhibit postnatal lethality. *Mol. Cell. Biol.* **21,** 1484–1490 (2001).
6. Akiyama, H., Chaboissier, M.-C., Martin, J. F., Schedl, A. & de Crombrugghe, B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* **16,** 2813–2828 (2002).
7. Deng, Y. *et al.* Yap1 Regulates Multiple Steps of Chondrocyte Differentiation during Skeletal Development and Bone Repair. *Cell Rep.* **14,** 2224–2237 (2016).
8. Joshi, S. *et al.* TEAD transcription factors are required for normal primary myoblast differentiation in vitro and muscle regeneration in vivo. *PLoS Genet.* **13,** e1006600 (2017).
9. Knapp, D. *et al.* Comparative transcriptional profiling of the axolotl limb identifies a tripartite regeneration-specific gene program. *PLoS One* **8,** e61352 (2013).
10. Zeller, R., López-Ríos, J. & Zuniga, A. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nat. Rev. Genet.* **10,** 845–858 (2009).

11. Nishimoto, S., Minguillon, C., Wood, S. & Logan, M. P. O. A combination of activation and repression by a colinear Hox code controls forelimb-restricted expression of Tbx5 and reveals Hox protein specificity. *PLoS Genet.* **10,** e1004245 (2014).
12. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37,** 17–23 (1950).
13. Vargesson, N., Luria, V., Messina, I., Erskine, L. & Laufer, E. Expression patterns of Slit and Robo family members during vertebrate limb development. *Mech. Dev.* **106,** 175–180 (2001).
14. Chimal-Monroy, J. *et al.* Analysis of the molecular cascade responsible for mesodermal limb chondrogenesis: Sox genes and BMP signaling. *Dev. Biol.* **257,** 292–301 (2003).
15. Petit, F., Sears, K. E. & Ahituv, N. Limb development: a paradigm of gene regulation. *Nat. Rev. Genet.* **18,** 245–258 (2017).

Guide to Supplementary Tables

Supplementary Table 1: Differential gene expression test results for aggregated "bulk" RNA-seq across different development stages. The "qval" is the false detection rate for the gene differential expression test (one-sided likelihood ratio test) across different development stages.

Supplementary Table 2: Differential gene expression test results from cell clusters or subclusters. For each gene, the "max.cluster" is the cluster id with the highest expression ("max.expr"). The "second.cluster" is the cell cluster with the second highest expression ("second.expr"). The "fold.change" is the fold change between the max expression and second max expression. The "qval" is the false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons) for the gene differential expression test across different cell clusters. Test_group is the group information for differential gene test: Main_clusters mean the test was performed on the 40 major clusters obtained by Louvain clustering; Subcluster_6 means the test was performed on all subclusters of major cluster 6 (epithelial cells); Subcluster_20 means the test was performed on all sub-clusters of major cluster 20 (endothelial cells).

Supplementary Table 3: Curated cluster id and annotated cell types together with reference gene markers for annotation.

Supplementary Table 4: Differential gene expression test results for the 38 major cell types. For each gene, the "max.tissue" is the cell type with the highest expression ("max.expr"). The "second.tissue" is the cell type with the second highest expression ("second.expr"). The "fold.change" is the fold change between the max expression and second max expression. The "qval" is the false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons) for the gene differential expression test across different cell types.

Supplementary Table 5: Summary of features of all 655 subclusters. For each subcluster, the table includes a subcluster id (*e.g.* "cluster.5-19" is subcluster 19 of major cluster 5), a primary subcluster name (*e.g.* "Excitatory neuron trajectory.7-of-55": this subcluster is the 7th of 55 subclusters that map to to the excitatory neuron trajectory, when the subclusters are ordered according to their average developmental pseudotime), the major cluster number, subcluster number, the name of the subtrajectory to which it mapped, subtrajectory stage (*e.g.* "7-of-55"), proportion of cells of its assigned subtrajectory that it contributes, total cell number, number of cells derived from E9.5/E10.5/E11.5/E12.5/E13.5 embryos, average developmental pseudotime time, list of minimum defining gene markers, doublet cell proportion and doublet cluster information, curated annotation, top matched cell type in MCA[10] and BCA[32].

Supplementary Table 6: Differential gene expression test results for AER development. The "qval" is the false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons) for the gene differential expression test.

Supplementary Table 7: Differential gene expression test results for limb development. The "qval" is the false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons) for the gene differential expression test.

Supplementary Table 8: Differential gene expression test results between forelimb and hindlimb cells. The "qval" is the false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons) for the gene differential expression test.

Supplementary Table 9: Differential gene expression test results from Moran's I test in limb mesenchyme cell development. The "qval" is the false detection rate (Moran's I test with adjustment for multiple comparisons) for the gene differential expression test.

Supplementary Table 10: Differential gene expression test results from Moran's I test during skeletal myogenesis. The "qval" is the false detection rate (Moran's I test with adjustment for multiple comparisons) for the gene differential expression test.

Supplementary Table 11: RT, ligation and PCR primer sequences for sci-RNA-seq3.