An integrated single-cell atlas of the skeleton from development through adulthood

Tim Herpelinck[1,2], Liesbeth Ory[1], Gabriele Nasello[1,2], Mojtaba Barzegari[3], Johanna Bolander[4], Frank P. Luyten[1,2], Przemko Tylzanowski[1,5] and Liesbet Geris[1,2,3,6]

[1] Skeletal Biology and Engineering Research Center, KU Leuven, Leuven, Belgium
[2] Prometheus, Division of Skeletal Tissue Engineering, KU Leuven, Leuven, Belgium
[3] Biomechanics Section, Department of Mechanical Engineering, KU Leuven, Leuven, Belgium
[4] Wake Forest Institute for Regenerative Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA
[5] Department of Biomedical Sciences, Laboratory of Molecular Genetics, Medical University of Lublin, Lublin, Poland
[6] Biomechanics Research Unit, GIGA In Silico Medicine, University of Liège, Liège, Belgium

ABSTRACT

The recent growth of single-cell transcriptomics has turned single-cell RNA sequencing (scRNA-seq) into a near-routine experiment. Breakthroughs in improving scalability have led to the creation of organism-wide transcriptomic datasets, aiming to comprehensively profile the cell types and states within an organism throughout its lifecycle. To date however, the skeleton remains a majorly underrepresented organ system in organism-wide atlases. Considering how the skeleton not only serves as the central framework of the vertebrate body but is also the home of the hematopoietic niche and a central player in major metabolic and homeostatic processes, this presents a major deficit in current reference atlas projects. To address this issue, we integrated seven separate scRNA-seq datasets containing skeletal cells and their developmental precursors, generating an atlas of over 800,000 cells. This skeletal cell atlas describes cells across the mesenchymal lineage from the induction of the limb field to adult bone, encompassing 50 different cell states. In addition, the original datasets were reannotated, enabling the discovery of novel, highly specific marker genes, some of which we have validated *in vivo* by whole-mount *in situ* hybridization. Furthermore, expanding the repertoire of available time points and cell types within a single dataset allowed for more complete analyses of cell-cell communication or *in silico* perturbation studies. Taken together, we present a missing piece in the current atlas mapping efforts, which will be of value to researchers in the fields of skeletal biology, hematopoiesis, metabolism and regenerative medicine.

INTRODUCTION

The skeleton is a highly advanced organ system with a wide variety of functions, ranging from protection of the internal organs and supporting locomotion to calcium homeostasis, housing the hematopoietic system and serving an endocrine function. In addition, some skeletal tissues possess remarkable regenerative properties, with bone being able to spontaneously regenerate after fracture without scar formation. While complex in its functions, the skeletal system is based on a surprisingly low number of different cell types, suggesting that its functional variety is a trait obtained through extensive intrinsic heterogeneity of cell states allowing advanced regionalized specialization.

Advancement towards ultra-high throughput single-cell RNA sequencing (scRNA-seq) platforms and the concomitant development of computational algorithms required to analyze the data, permits the generation of organism-wide transcriptomic maps, resolved both in time and space [1,2]. If a reference atlas is available, new datasets can be annotated automatically thus introducing fast, data-driven and consistent labeling of the cells [3–5]. Unfortunately, the skeleton is minimally represented in most of these atlases, often with insufficiently detailed annotation of the skeletal lineage. Contrarily, the skeletal biology community has generated scRNA-seq datasets but focused on very specific cell subpopulations. As a result, we find a significant discrepancy in the transcriptional characterization of the skeletal system: large atlases contain a wealth of data left largely unexplored due to the insufficiently precise annotation while the specialized datasets remain confined to their domain of study. To bridge this gap, we reannotated and merged publicly available atlases with specialized skeletal datasets into a single framework comprising over 800,000 cells [6–12] forming the Skeletal Cell Atlas (SCA).

Following the generation of the SCA, we explored its applicability and predictive value. We demonstrated the potential of the SCA in novel marker gene identification, pseudo-temporal trajectory inference and intercellular communication analysis. The predictive capacity of the SCA was further tested by simulation of Sox9 inactivation and the analysis of its consequences, which were in line with the *in vivo* phenotype. These results support the notion that our SCA can be used as a reliable reference for the automated annotation of new skeletal scRNA-seq datasets. To facilitate accessibility, all notebooks required to build the atlas are available online. Lastly, we developed a web app where users can query and download the data to increase the accessibility of scRNA-seq data to a wider audience.

RESULTS

Dataset annotation and integration to produce a skeletal cell atlas

To build a comprehensive reference atlas of the skeleton we selected seven publicly available mouse scRNA-seq datasets containing skeletal cells from the onset of limb development to mature bone. Prior to integration, we removed non-mesenchymal cells with the exception of endothelium and the apical ectodermal ridge (AER) from the Mouse Organogenesis Cell Atlas (MOCA) [12] based on the authors' classification. Cells belonging to the hematopoietic lineage were only retained if they were derived from the bone marrow, i.e. the sample was obtained through flushing of the marrow. The individual datasets were then manually reannotated based on canonical marker gene expression and integrated (n = 808,778; Fig. 1a, Extended Data Fig. 1a-g; Supplementary Table 1).

No cross-laboratory or technology-induced batch effects were observed (Extended Data Fig. 2a,b). From this we concluded that the integration of the atlas was cell-based, spanning 50 unique cell states.

Novel marker gene identification

During the reannotation of the mesenchymal cells in the MOCA [12], we identified a small cluster of mesenchymal cells expressing characteristic early limb mesenchyme markers including the forelimb marker *Tbx5* [13] and the hindlimb-specific transcription factors *Tbx4* and *Pitx1* [14,15](Fig. 1b). Based on the differential abundance of cells expressing hindlimb markers (Extended Data Fig. 3a), we

hypothesized this cluster could be hindlimb-specific. To investigate this, we selected the differentially expressed genes *Tbx20* and *Postn* as possible new markers, based on their abundance across multiple days of development, high specificity for this particular cluster (Extended Data Fig. 3b,c) and the absence of hindlimb-specific expression of these genes in the literature. To validate the digital prediction, we carried out whole-mount *in situ* hybridization (WISH) on chick limbs at the comparable developmental stages, exploiting the strong conservation of skeletal development between these organisms [16–18]. Indeed, we could confirm that these genes were specific to the early hindlimb mesenchyme at Hamilton-Hamburger stage 26 (HH26, Fig. 1c,d). We postulate that differential expression analysis within the atlas can be leveraged for the discovery of novel marker genes, specific in both time and anatomical location.

Virtual reconstruction of the growth plate

During development, long bones are lengthened through a rapid proliferation of chondrocytes at the ends of the bone along the longitudinal axis, thereby creating columnar structures. As the distance from the dividing front increases, these chondrocytes gradually become hypertrophic and are ultimately replaced by bone (reviewed in [19]). This growth center responsible for bone elongation is referred to as the growth plate (GP) and can, in a simplified view, be regarded as a cylindrical, chondrocytic structure.

To virtually restore this structural landmark, we first subset the atlas to select GP chondrocytes, which we use as the input for pseudotime analysis to reconstruct the transcriptional trajectory from proliferation to hypertrophy *in silico* (Fig. 2a). Cells were then binned by pseudotime. Upon each bin we performed dimensionality reduction by *t*-distributed stochastic neighborhood embedding (*t*-SNE). Importantly, we imposed a circle as boundary condition on the gradient descent function of each *t*-SNE. Alignment of these circular projections then recreated the cylindrical shape of the GP, while also visualizing transcriptional heterogeneity within the GP across bins of pseudotime (Fig. 2b). As such, plotting gene expression along the pseudotime axis combined with expression in pseudospace artificially restores the tissue architecture lost during tissue digestion (Fig. 2c). This model can be considered an extension of the virtual GP developed by Li and colleagues [8].

Intercellular communication inference across the skeleton

Intercellular communication driven by ligand-receptor interactions is pivotal for the cellular differentiation. Therefore, a wide variety of tools to infer intercellular signaling from scRNA-seq data has been developed (reviewed in [20]). Here, we used CellPhoneDB [21] as it considers multimeric receptor complexes when inferring ligand-receptor interactions. We analyzed Bone Morphogenetic Protein (BMP) signaling taking place between the limb bud and the Apical Ectodermal Ridge (AER). BMPs are members of the Transforming Growth Factor ligand superfamily and therefore signal through a tetrameric Type I-Type II receptor complex. Type I and Type II BMP receptors are expressed across the limb mesenchyme [22,23] and AER [22], which also acts as a source of BMPs [24]. However, due to the multimeric character of the receptor complexes resulting in many different combinations, it is impossible to experimentally determine which ligand-receptor pair is predominantly used. Screening with CellPhoneDB drastically narrowed down the number of possible combinations. The analysis revealed that the AER mainly signals to the limb bud through BMP4 and BMP7 (Fig. 3) and that BMP4 appears to favor binding a receptor complex containing BMPR2. Conversely, BMP signaling towards the AER from the limb mesenchyme is predicted to be limited. Also noteworthy is that the favored ligand-receptor combinations are largely invariable over time. In addition, the AER is predicted to be capable of autocrine signaling.

Simulation of transcription factor perturbation

Developmental biology was the first field to tackle the monumental task of reconstructing gene regulatory networks (GRNs). Historically, this was done through series of experiments in which transcription factor (TF) activity was perturbed by gain- or loss-of-function. CellOracle [25] is a novel algorithm designed for GRN inference from scRNA-seq data. In addition, it allows for perturbation of

TF expression to predict resultant gene expression changes. To test if our dataset was amenable to this type of analysis, we simulated an inactivation of Sox9, the master regulator of chondrocyte identity [26]. In our simulation, most mesoderm-derived cells were unable to contribute to the limb mesenchyme and cartilage, while myogenesis and endothelial development were unaffected (Fig. 4a,b). Conversely, unperturbed mesodermal cells progressed to both fates, as shown by the gradient of pseudotime or developmental flow (Fig. 4c). Attractor basins created as a result of the Sox9 absence were then visualized by the inner product of the perturbation and developmental flow vectors. The results are in line with the findings in mice where *in vivo* inactivation of Sox9 renders mesodermal cells unable to reach the chondrocyte state and causes accumulation of cartilage precursors, while progression into the myogenic lineage is unimpeded [26].

DISCUSSION

The Skeletal Cell Atlas represents a manually curated compendium of over 800,000 cells across seven datasets, of which we have shown its applicability in both data- and hypothesis-driven analyses. First, the repository was demonstrated to enable marker gene discovery with remarkable specificity, as demonstrated by the previously undescribed hindlimb-specific enrichment of *Tbx20* and *Postn* in the early limb bud. Additionally, we presented the unsupervised construction of a spatiotemporal map of the GP, which correctly recapitulates the molecular cascades of GP development, by imposing boundary conditions on the gradient descent of *t*-SNE dimensionality reduction. Similar results have been obtained previously by warping the principal component space [8]. Next, analysis of intercellular communication by BMP signaling was able to shed light on a longstanding question in the field of limb development. It is known that BMP signaling is required for AER regression, but not whether the AER, the mesenchyme or both act as the source of those BMPs [24]. Based on CellPhoneDB's cell-cell communication inference, taking into account the subunit architecture of both ligands and heteromeric receptors [20,21], our case study suggests the AER to be the dominant source. This finding does require further *in vivo* validation. Lastly, we demonstrated that artificial knockout of Sox9 corroborates previous work demonstrating the inability of Sox9-negative cells to differentiate into chondrocytes [26]. Given how the expression of effectively any TF can be modulated to assess developmental outcome, the SCA represents a valuable resource for both limb developmental biologists and skeletal biology researchers.

This meta-analysis does have limitations. The list of datasets used in its assembly is not exhaustive and except for the MOCA, all cells are derived from the appendicular skeleton. Considering the different developmental origin of the axial skeleton, it is very likely that some cell states and rare populations are still missing. In addition, the current release is restricted to the healthy murine skeleton. We therefore plan to continuously update and extend the SCA, as data availability permits, towards the entire skeleton in health and disease.

The goal of this study was to provide an initial reference of the musculoskeletal system. In future work, this first version will serve as the basis of a collaborative effort to complete the atlas. All data, analyses and code to replicate the figures are freely available. Our web portal is designed to be intuitive and to allow browsing of the data at a glance. In short, the Skeletal Cell Atlas provides a framework with characterization of most known cell populations in the skeleton and represents a foundation for future studies in a wide variety of disciplines.

MATERIALS AND METHODS

Data preprocessing

Quality control of the raw count matrices was performed for each individual dataset using scater [27]. Initial filtering was performed based on the library size and number of expressed genes per cell. Cells more than three standard deviations away from the median of either metric were filtered out. Subsequently, cells where the mitochondrial fraction of reads exceeded a proportion higher than three standard deviations from the median mitochondrial fraction were removed. Finally, low-abundance genes (expression lower than 1e-3) were filtered out and duplicate rows were removed if present.

Clustering and dimensionality reduction

All datasets were first analyzed individually prior to integration. The filtered count matrices were imported into Seurat v4 [5]. Normalization was performed using the *LogNormalize* parameter and a scale factor of 1e4. Subsequently, the data was centered and scaled using all genes followed by a calculation of principal components (PCs) using the top 2000 highly variable genes selected by the "vst" method. The optimal number of PCs to construct the Shared Nearest Neighbor (SNN) graph was visually determined based on the scree plot and varied for each dataset. Clustering was performed using the Louvain algorithm. The resolution was adapted to the individual dataset, where we defined the optimal number of clusters as the maximum number of cell states that could confidently be labeled based on marker gene expression. These marker genes were obtained from the *FindMarkers* function using the default settings. The Uniform Manifold Approximation Plots (UMAPs) were calculated using the same number of PCs as those used for the SNN graph.

Integration by reference mapping

The integrated atlas was constructed using Seurat v4 [5]. Anchors between all individual datasets were identified using reciprocal PCA (rPCA). After passing these anchors to *IntegrateData*, the integrated matrix was scaled, the first 30 PCs were calculated and used to construct the UMAP.

Pseudotime analysis

For pseudotime analysis of the growth plate, a subset of cells was taken from the atlas based on their cell type labels (proliferative, prehypertrophic and hypertrophic chondrocytes). The subset was then rescaled and UMAP projections were calculated using the first 20 PCs. Minimal distance was set to 0.5 and local connectivity to 1.5. This projection was then passed to monocle3 [12] using SeuratWrappers. Clustering was performed on the resulting CellDataSet (CDS) object using the default parameters. The trajectory graph was learned on the monocle-derived clusters by calling *learn_graph*.

Cell-cell interaction prediction

Prediction of cell-cell communication by ligand-receptor interactions between cell types was performed using CellPhoneDB [21]. Results were separated into individual tables for each developmental time point. Cell types not present at a specific developmental stage were removed to retain only biologically meaningful interactions.

Knockout simulation

For *in silico* knockout (KO) experiments in CellOracle [25], the atlas was subset to only include the developmental time points. This Seurat object was then converted to AnnData format using sceasy (https://github.com/cellgeni/sceasy). To reduce the amount of computational time and resources required by a large dataset, 30,000 cells were randomly selected and only highly variable genes (n=1552) were included. For gene regulatory network (GRN) inference, we used the built-in base GRN made from the mouse sci-ATAC-seq atlas [28]. Following k nearest neighbors (KNN) imputation based on the first 71 PCs, GRNs were imputed for each cluster. To simulate KO of a transcription factor, its expression was set to 0. After this KO, GRN inference was performed again. Signal perturbation propagation and transition probabilities were calculated using the standard settings. Visualization of the pseudotime gradient, simulation vector field and their inner product was performed as described in the CellOracle online documentation.

Whole-mount *in situ* hybridization (WISH)

Making of digoxigenin-labeled RNA probes

The digoxigenin-labeled riboprobes were constructed from PCR probes generated by performing PCR using primers with an attached T7 promotor sequence (Supplementary Table 2) and chick limb cDNA. RNA isolation and cDNA conversion were accomplished by following the protocol of the RNeasy Mini Kit (Qiagen) and the protocol of the PrimeScript™ RT reagent Kit (Perfect Real Time). Digoxigenin (DIG) labeling was performed with the DIG RNA Labeling Kit (SP6/T7) (Roche, cat. no. 11175025910): 2 µL 10X Transcription buffer, 2 µL 0.1M dithiothreitol, 2 µL Nucleotide mix for DIG labeling, 2 µL PCR product, 0.5 µL RnaseOut and 1.5 µL T7 RNA polymerase were mixed in a total volume of 20 µL. After 2 hours of incubation at 37°C, 2 µL DNaseI was added for 15 minutes at 37°C to degrade the DNA template. 2 µL Yeast tRNA and water were added to a total volume of 50 µL. The samples were mixed with 180 µL water, 50 µL 10M ammonium acetate and 400 µL 100% ethanol. After 30 minutes of incubation on wet ice, they were spun for 20 minutes at 14000 rpm at 4°C. The supernatant was removed, and the pellet was washed with 70% ethanol. The samples were spun a second time, the supernatant was removed, and the pellet air-dried. The pellet was re-dissolved in 50 µL water, aliquoted and stored at -80°C.

Whole-mount in situ hybridization of chick embryos

Isolation of chicken embryos was performed at stages HH18 to HH28, fixed in 4% paraformaldehyde (PFA) overnight and stored in methanol at -20°C. Embryos or limbs were bleached for an hour with 6% hydrogen peroxide, rehydrated through a methanol series (100%-75%-50%-25%) for 10 minutes each and washed in phosphate-buffered saline with 0.1% Tween20 (PBT). This was followed by a (100-150 µg/mL) Proteinase K treatment at room temperature (RT) for 10 minutes and the samples were re-fixed with 4% PFA. Then, the embryos/limbs were hybridized with DIG-labeled riboprobes in hybridization buffer (5X SSC pH 4.5; 50% formamide; 50 µg/mL yeast tRNA; 50 µg/mL heparin porcine) overnight at 65°C.

On the second day of the whole-mount *in situ* hybridization experiment, embryos/limbs were washed for two times 30 minutes with solution 1 (50% formamide, 5X SSC pH 4.5, 1% sodium dodecyl sulfate (SDS)) at 65°C, three times 5 minutes with solution 2 (0.5 M NaCl, 10 mM Tris-HCl pH 7.5, 0.1% Tween 20) on a rocking plate at RT, two times 15 minutes with 100 µg/mL RNase A in solution 2 on a rocking plate at RT, two times 30 minutes with solution 3 (50% formamide, 5X SSC pH 4.5) at 65°C and three times with maleic acid buffer with 0.1% Tween20 (MABT) for 5 minutes at RT. The samples were pre-blocked with 2% blocking reagent (Roche, cat. no. 11096176001) in 20% fetal bovine serum (FBS) at RT and then incubated overnight at 4°C with a 1:1000 dilution of an alkaline phosphatase-conjugated anti-DIG antibody (Roche, cat. no. 11093274910) in blocking solution. After several washes in MABT, an additional washing step at 4°C was performed for 2 days. To limit the background signal and internal phosphatase activity the embryos/limbs were washed with 2 mM Levamisole. The coloration was achieved using BM purple AP substrate (Roche, cat. no. 11442074001) at 30°C. After

staining, the samples were washed in 20 mM ethylenediaminetetraacetic acid (EDTA)/PBT, post-fixed with 4% PFA overnight at 4°C and stored in 100% ethanol. Embryos were photographed using a stereomicroscope (Zeiss). A minimum of three limbs per gene was assayed for reproducibility (*n*=3).

DATA AND CODE AVAILABILITY

All datasets used to generate the atlas are publicly available. Their accession numbers are listed in Supplementary Table 1. The Skeletal Cell Atlas can be downloaded or interactively explored at www.skeletalcellatlas.org. All code used to perform the analyses and notebooks to generate the figures is available at https://github.com/HerpelinckT/SkeletalCellAtlas. The code to create the web app can be found at: https://github.com/mbarzegary/skeletal-cell-atlas.

FIGURE LEGENDS

Figure 1. An integrated compendium of skeletal cell types with detailed annotation. a, UMAP visualization of 808,778 murine mesenchyme- and skeleton-derived cells, colored by annotation. **b**, UMAP projection of cell type annotation of data taken from the MOCA. Arrow: early limb mesenchyme enriched in hindlimb cells. **c,d**, WISH of marker genes for the cluster indicated in **b** in chick limbs at HH26, demonstrating clear hindlimb specificity in early limb mesenchyme. FL: forelimb, HL: hindlimb.

Figure 2. Spatiotemporal reconstruction of the transcriptional dynamics within the growth plate. a, UMAP visualization of the integrated GP data subset from the main atlas, colored by annotation (left) and monocle3 pseudotime value (right). **b**, GP chondrocytes were grouped in 50 bins based on similar pseudotime value. *t*-SNE dimensional reduction was performed on each bin, using a circle with radius 20 as the boundary condition for gradient descent, thus recreating the cylindrical shape of the GP upon stacking of the bins. **c**, pseudotime analysis of known marker genes for different zones of the growth plate in pseudotime-space (left) and average expression along pseudotime (right).

Figure 3. BMP signaling case study between AER and mesenchyme. Dotplot of predicted ligand-receptor complex interactions. Within each time point, the AER is first assessed as a receiver of BMP signal by evaluating expression of its receptor expression (blue). This is then inverted, testing for BMP ligand expression in the AER (red) and to which receptor complexes it is likely to signal in the mesenchyme (blue). Dot color indicates $\log_2$ of average expression of ligand and receptor mRNAs. Dot size reflects significance of the predicted interaction.

Figure 4. Sox9 KO simulation during development. a, reference annotation of developing limb data used for KO analysis. **b**, vector field graph indicating cell trajectories after Sox9 KO. **c**, gradient of pseudotime in the absence of perturbation. **d**, inner vectorial product of vectors in **b** and **c**. Red indicates attractor states, while blue represents unlikely states.

EXTENDED DATA FIGURES

Extended Figure 1. Individual datasets after reannotation. a-g, UMAP visualization of the individual datasets with their updated annotations.

Extended Figure 2. Assessment of batch effects influencing integration. a-c, UMAP visualization of the atlas after integration where neither research group (**a**) or sequencing technology (**b**) appear to affect integration.

Extended Figure 3. Hindlimb-enriched mesenchymal cluster marker genes. a, Violin plot comparing *Tbx5*, *Tbx4* and *Pitx1* showing enrichment of early hindlimb mesenchyme markers within the cluster. **b**, expression of *Tbx20* and *Postn*, both predicted as marker genes for this cluster over time showing the expression is relatively constant over multiple days of development and is specific to the cluster, as shown in **c**.

SUPPLEMENTARY TABLES

Supplementary table 1: scRNA-seq datasets used to construct the Skeletal Cell Atlas.

Supplementary table 2: primer sequences used for WISH (related to Fig. 1).

ACKNOWLEDGMENTS

We would like to thank Inge Van Hoven for experimental assistance and Ronny Moreas for technical support in deployment of the web app. Research was funded by the Research Foundation Flanders (FWO Vlaanderen) T.H.: 1S80021N, J.B.: 1518618N, MatheMorphosis: G0D3420N, and G088218N; by the European Union's Horizon 2020 Framework Program (H2020/2014-2021) ERC INSITE (772418) and SC1-DTH ISW (101016503), as well as by the special research fund of the KU Leuven (C24/17/077). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. This work is part of Prometheus, the KU Leuven R&D division for skeletal tissue engineering (http://www.kuleuven.be/prometheus).

REFERENCES

[1] M. Asp, S. Giacomello, L. Larsson, C. Wu, D. Fürth, X. Qian, E. Wärdell, J. Custodio, J. Reimegård, F. Salmén, C. Österholm, P.L. Ståhl, E. Sundström, E. Åkesson, O. Bergmann, M. Bienko, A. Månsson-Broberg, M. Nilsson, C. Sylvén, J. Lundeberg, A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart, Cell. 179 (2019) 1647-1660.e19. https://doi.org/10.1016/j.cell.2019.11.025.

[2] R. Elmentaite, N. Kumasaka, K. Roberts, A. Fleming, E. Dann, H.W. King, V. Kleshchevnikov, M. Dabrowska, S. Pritchard, L. Bolt, S.F. Vieira, L. Mamanova, N. Huang, F. Perrone, I. Goh Kai'En, S.N. Lisgo, M. Katan, S. Leonard, T.R.W. Oliver, C.E. Hook, K. Nayak, L.S. Campos, C. Domínguez Conde, E. Stephenson, J. Engelbert, R.A. Botting, K. Polanski, S. van Dongen, M. Patel, M.D. Morgan, J.C. Marioni, O.A. Bayraktar, K.B. Meyer, X. He, R.A. Barker, H.H. Uhlig, K.T. Mahbubani, K. Saeb-Parsy, M. Zilbauer, M.R. Clatworthy, M. Haniffa, K.R. James, S.A. Teichmann, Cells of the human intestinal tract mapped across space and time, Nature. 597 (2021) 250–255. https://doi.org/10.1038/s41586-021-03852-1.

[3] G. Pasquini, J.E. Rojo Arias, P. Schäfer, V. Busskamp, Automated methods for cell type annotation on scRNA-seq data, Comput. Struct. Biotechnol. J. 19 (2021) 961–969. https://doi.org/10.1016/j.csbj.2021.01.015.

[4] D.S. Fischer, L. Dony, M. König, A. Moeed, L. Zappia, L. Heumos, S. Tritschler, O. Holmberg, H. Aliee, F.J. Theis, Sfaira accelerates data and model reuse in single cell genomics, Genome Biol. 22 (2021) 1–21. https://doi.org/10.1186/s13059-021-02452-6.

[5] Y. Hao, S. Hao, E. Andersen-Nissen, W.M. Mauck, S. Zheng, A. Butler, M.J. Lee, A.J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E.P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L.M. Fleming, B. Yeung, A.J. Rogers, J.M. McElrath, C.A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data, Cell. 184 (2021) 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

[6] K.K. Sivaraj, H.W. Jeong, B. Dharmalingam, D. Zeuschner, S. Adams, M. Potente, R.H. Adams, Regional specialization and fate specification of bone stromal cells in skeletal development, Cell Rep. 36 (2021). https://doi.org/10.1016/j.celrep.2021.109352.

[7] Q. Bian, Y.H. Cheng, J.P. Wilson, E.Y. Su, D.W. Kim, H. Wang, S. Yoo, S. Blackshaw, P. Cahan, A single cell transcriptional atlas of early synovial joint development, Development. 147 (2020). https://doi.org/10.1242/dev.185777.

[8] J. Li, H. Luo, R. Wang, J. Lang, S. Zhu, Z. Zhang, J. Fang, K. Qu, Y. Lin, H. Long, Y. Yao, G. Tian, Q. Wu, Systematic Reconstruction of Molecular Cascades Regulating GP Development Using Single-Cell RNA-Seq, Cell Rep. 15 (2016) 1467–1480. https://doi.org/10.1016/j.celrep.2016.04.043.

[9] S. Debnath, A.R. Yallowitz, J. McCormick, S. Lalani, T. Zhang, R. Xu, N. Li, Y. Liu, Y.S. Yang, M. Eiseman, J.H. Shim, M. Hameed, J.H. Healey, M.P. Bostrom, D.A. Landau, M.B. Greenblatt, Discovery of a periosteal stem cell mediating intramembranous bone formation, Nature. 562 (2018) 133–139. https://doi.org/10.1038/s41586-018-0554-8.

[10] A. Böhm, N. Dirckx, R.J. Tower, N. Peredo, S. Vanuytven, K. Theunis, E. Nefyodova, R. Cardoen, V. Lindner, T. Voet, M. Van Hul, C. Maes, Activation of Skeletal Stem and Progenitor Cells for Bone Regeneration Is Driven by PDGFR b Signaling, Dev. Cell. (2019) https://doi.org/10.1016/j.devcel.2019.08.013. https://doi.org/10.1016/j.devcel.2019.08.013.

[11] N. Baryawno, D. Przybylski, M.S. Kowalczyk, Y. Kfoury, N. Severe, K. Gustafsson, K.D. Kokkaliaris, F. Mercier, M. Tabaka, M. Hofree, D. Dionne, A. Papazian, D. Lee, O. Ashenberg, A. Subramanian, E.D. Vaishnav, O. Rozenblatt-Rosen, A. Regev, D.T. Scadden, A

Cellular Taxonomy of the Bone Marrow Stroma in Homeostasis and Leukemia, Cell. 177 (2019) 1915-1932.e16. https://doi.org/10.1016/j.cell.2019.04.040.

[12] 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 (2019) 496–502. https://doi.org/10.1038/s41586-019-0969-x.

[13] P. Agarwal, J.N. Wylie, J. Galceran, O. Arkhitko, C. Li, C. Deng, R. Grosschedl, B.G. Bruneau, Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo, Development. 130 (2003) 623–633. https://doi.org/10.1242/dev.00191.

[14] L.A. Naiche, V.E. Papaioannou, Loss of Tbx4 blocks hindlimb development and affects vascularization and fusion of the allantois, Development. 130 (2003) 2681–2693. https://doi.org/10.1242/dev.00504.

[15] D.P. Szeto, A.K. Rvan, S.M. O'Connell, M.G. Rosenfeld, P-OTX: A PIT-1-interacting homeodomain factor expressed during anterior pituitary gland development, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 7706–7710. https://doi.org/10.1073/pnas.93.15.7706.

[16] F. Petit, K.E. Sears, N. Ahituv, Limb development: A paradigm of gene regulation, Nat. Rev. Genet. 18 (2017) 245–258. https://doi.org/10.1038/nrg.2016.167.

[17] A. Zuniga, Next generation limb development and evolution: Old questions, new perspectives, Dev. 142 (2015) 3810–3820. https://doi.org/10.1242/dev.125757.

[18] A. Zuniga, R. Zeller, Dynamic and self-regulatory interactions among gene regulatory networks control vertebrate limb bud morphogenesis, in: Curr. Top. Dev. Biol., Academic Press, 2020: pp. 61–88. https://doi.org/10.1016/bs.ctdb.2020.02.005.

[19] H.M. Kronenberg, Developmental regulation of the growth plate, Nature. 423 (2003) 332–336. https://doi.org/10.1038/nature01657.

[20] E. Armingol, A. Officer, O. Harismendy, N.E. Lewis, Deciphering cell–cell interactions and communication from gene expression, Nat. Rev. Genet. 22 (2021) 71–88. https://doi.org/10.1038/s41576-020-00292-x.

[21] M. Efremova, M. Vento-Tormo, S.A. Teichmann, R. Vento-Tormo, CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes, Nat. Protoc. 15 (2020) 1484–1506. https://doi.org/10.1038/s41596-020-0292-x.

[22] Y. Kawakami, T. Ishikawa, M. Shimabara, N. Tanda, M. Enomoto-lwamoto, M. Iwamoto, T. Kuwana, A. Ueki, S. Noji, T. Nohno, BMP signaling during bone pattern determination in the developing limb, Development. 122 (1996) 3557–3566. https://doi.org/10.1242/dev.122.11.3557.

[23] H. Zou, R. Wieser, J. Massagué, L. Niswander, Distinct roles of type I bone morphogenetic protein receptors in the formation and differentiation of cartilage, Genes Dev. 11 (1997) 2191–2203. https://doi.org/10.1101/gad.11.17.2191.

[24] S. Pizette, L. Niswander, BMPs negatively regulate structure and function of the limb apical ectodermal ridge, Development. 126 (1999) 883–894. https://doi.org/10.1242/dev.126.5.883.

[25] K. Kamimoto, C.M. Hoffmann, S.A. Morris, CellOracle: Dissecting cell identity via network inference and in silico gene perturbation, BioRxiv. (2020) 2020.02.17.947416. https://doi.org/10.1101/2020.02.17.947416.

[26] W. Bi, J.M. Deng, Z. Zhang, R.R. Behringer, B. De Crombrugghe, Sox9 is required for cartilage formation, Nat. Genet. 22 (1999) 85–89. https://doi.org/10.1038/8792.

[27] D.J. McCarthy, K.R. Campbell, A.T.L. Lun, Q.F. Wills, Scater: Pre-processing, quality control,

normalization and visualization of single-cell RNA-seq data in R, Bioinformatics. 33 (2017) 1179–1186. https://doi.org/10.1093/bioinformatics/btw777.

[28] D.A. Cusanovich, A.J. Hill, D. Aghamirzaie, R.M. Daza, H.A. Pliner, J.B. Berletch, G.N. Filippova, X. Huang, L. Christiansen, W.S. DeWitt, C. Lee, S.G. Regalado, D.F. Read, F.J. Steemers, C.M. Disteche, C. Trapnell, J. Shendure, A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility, Cell. 174 (2018) 1309-1324.e18. https://doi.org/10.1016/j.cell.2018.06.052.

Figure 1

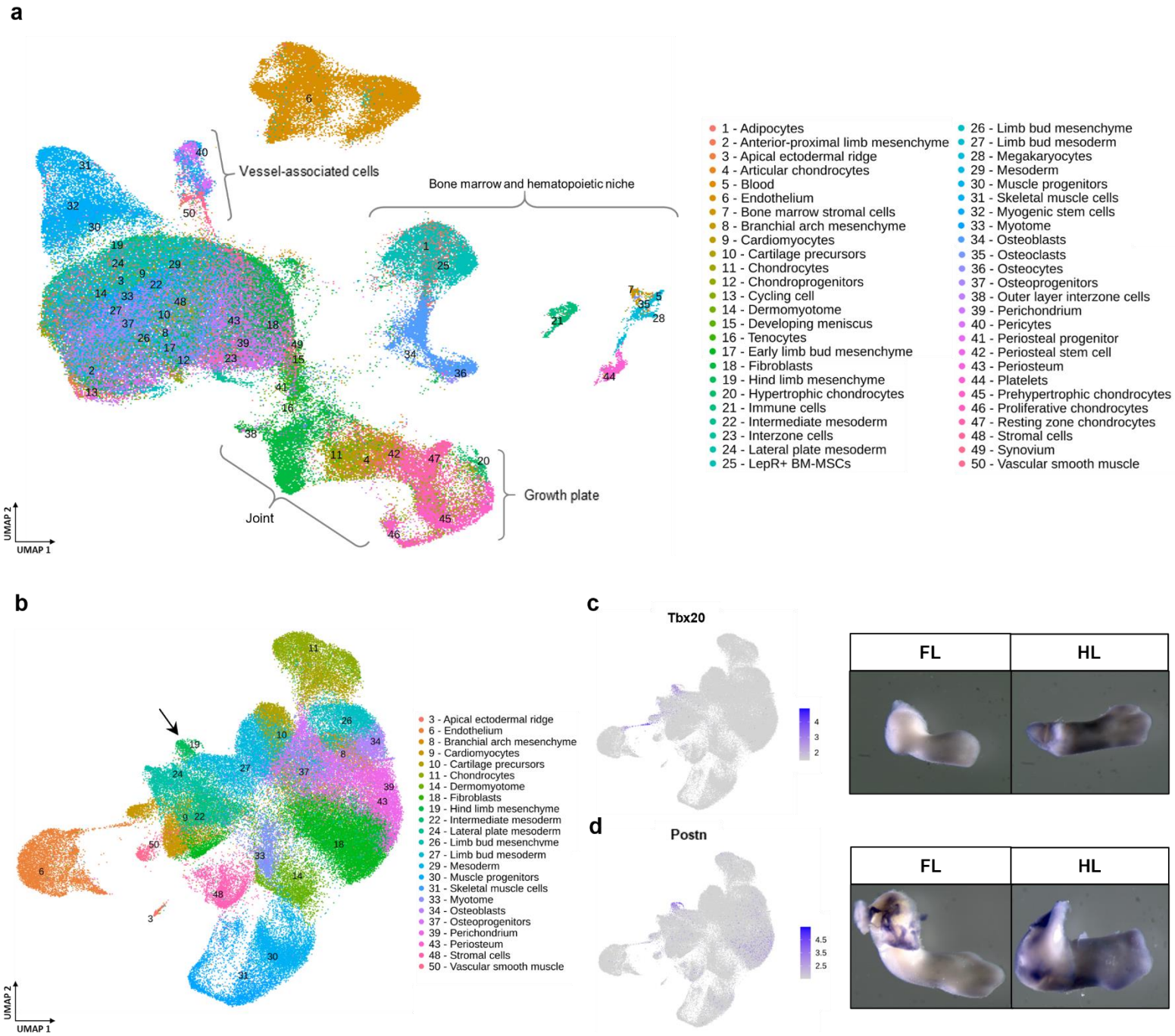

a
Vessel-associated cells
Bone marrow and hematopoietic niche
Joint
Growth plate
UMAP 2
UMAP 1
1 - Adipocytes
2 - Anterior-proximal limb mesenchyme
3 - Apical ectodermal ridge
4 - Articular chondrocytes
5 - Blood
6 - Endothelium
7 - Bone marrow stromal cells
8 - Branchial arch mesenchyme
9 - Cardiomyocytes
10 - Cartilage precursors
11 - Chondrocytes
12 - Chondroprogenitors
13 - Cycling cell
14 - Dermomyotome
15 - Developing meniscus
16 - Tenocytes
17 - Early limb bud mesenchyme
18 - Fibroblasts
19 - Hind limb mesenchyme
20 - Hypertrophic chondrocytes
21 - Immune cells
22 - Intermediate mesoderm
23 - Interzone cells
24 - Lateral plate mesoderm
25 - LepR+ BM-MSCs
26 - Limb bud mesenchyme
27 - Limb bud mesoderm
28 - Megakaryocytes
29 - Mesoderm
30 - Muscle progenitors
31 - Skeletal muscle cells
32 - Myogenic stem cells
33 - Myotome
34 - Osteoblasts
35 - Osteoclasts
36 - Osteocytes
37 - Osteoprogenitors
38 - Outer layer interzone cells
39 - Perichondrium
40 - Pericytes
41 - Periosteal progenitor
42 - Periosteal stem cell
43 - Periosteum
44 - Platelets
45 - Prehypertrophic chondrocytes
46 - Proliferative chondrocytes
47 - Resting zone chondrocytes
48 - Stromal cells
49 - Synovium
50 - Vascular smooth muscle
b
3 - Apical ectodermal ridge
6 - Endothelium
8 - Branchial arch mesenchyme
9 - Cardiomyocytes
10 - Cartilage precursors
11 - Chondrocytes
14 - Dermomyotome
18 - Fibroblasts
19 - Hind limb mesenchyme
22 - Intermediate mesoderm
24 - Lateral plate mesoderm
26 - Limb bud mesenchyme
27 - Limb bud mesoderm
29 - Mesoderm
30 - Muscle progenitors
31 - Skeletal muscle cells
33 - Myotome
34 - Osteoblasts
37 - Osteoprogenitors
39 - Perichondrium
43 - Periosteum
48 - Stromal cells
50 - Vascular smooth muscle
c
Tbx20
d
Postn
FL
HL
FL
HL

Figure 2

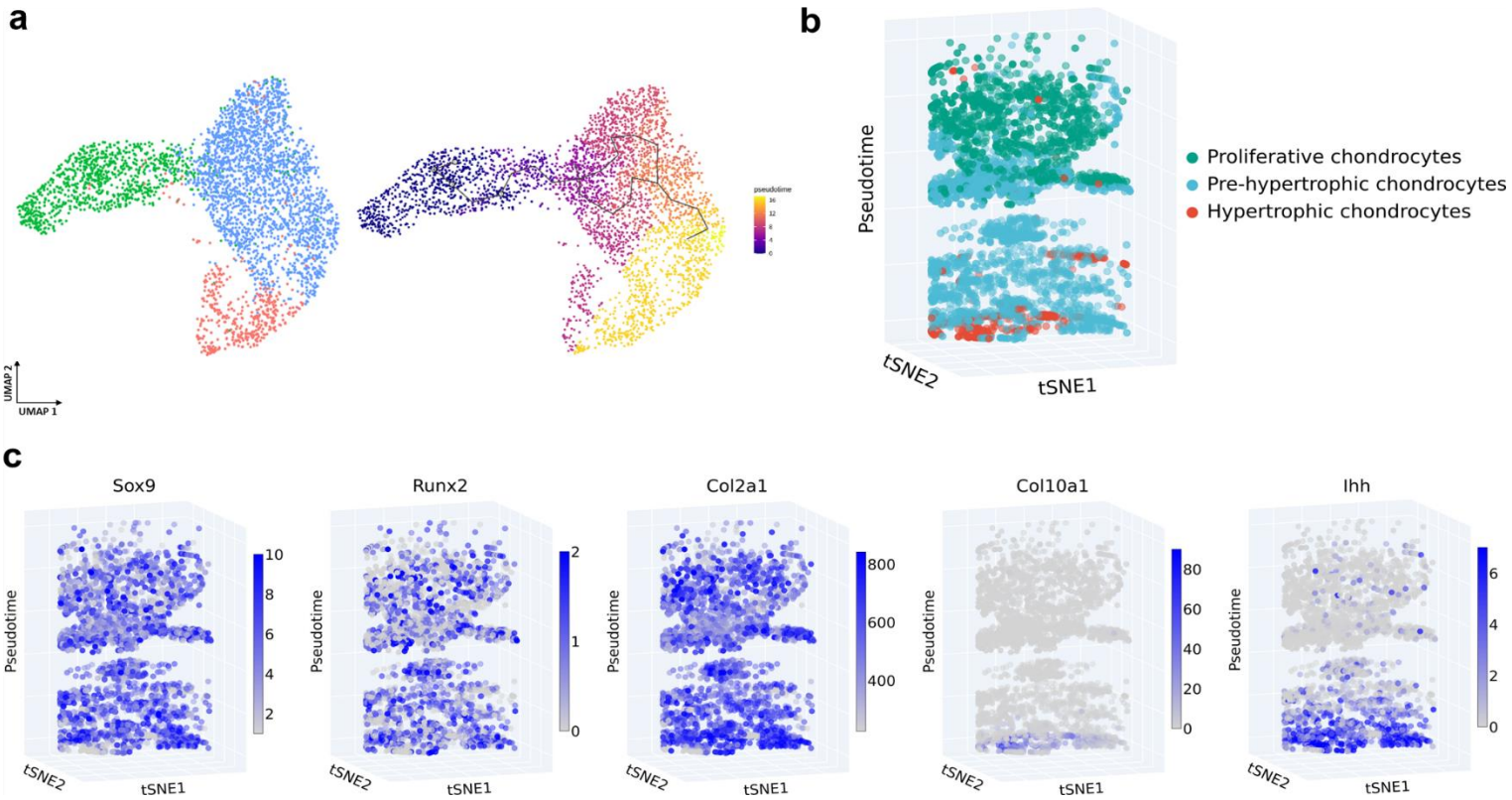

a
UMAP 2
UMAP 1
pseudotime
b
Pseudotime
tSNE2
tSNE1
Proliferative chondrocytes
Pre-hypertrophic chondrocytes
Hypertrophic chondrocytes
c
Sox9
Runx2
Col2a1
Col10a1
Ihh
Pseudotime
tSNE2
tSNE1

Figure 3

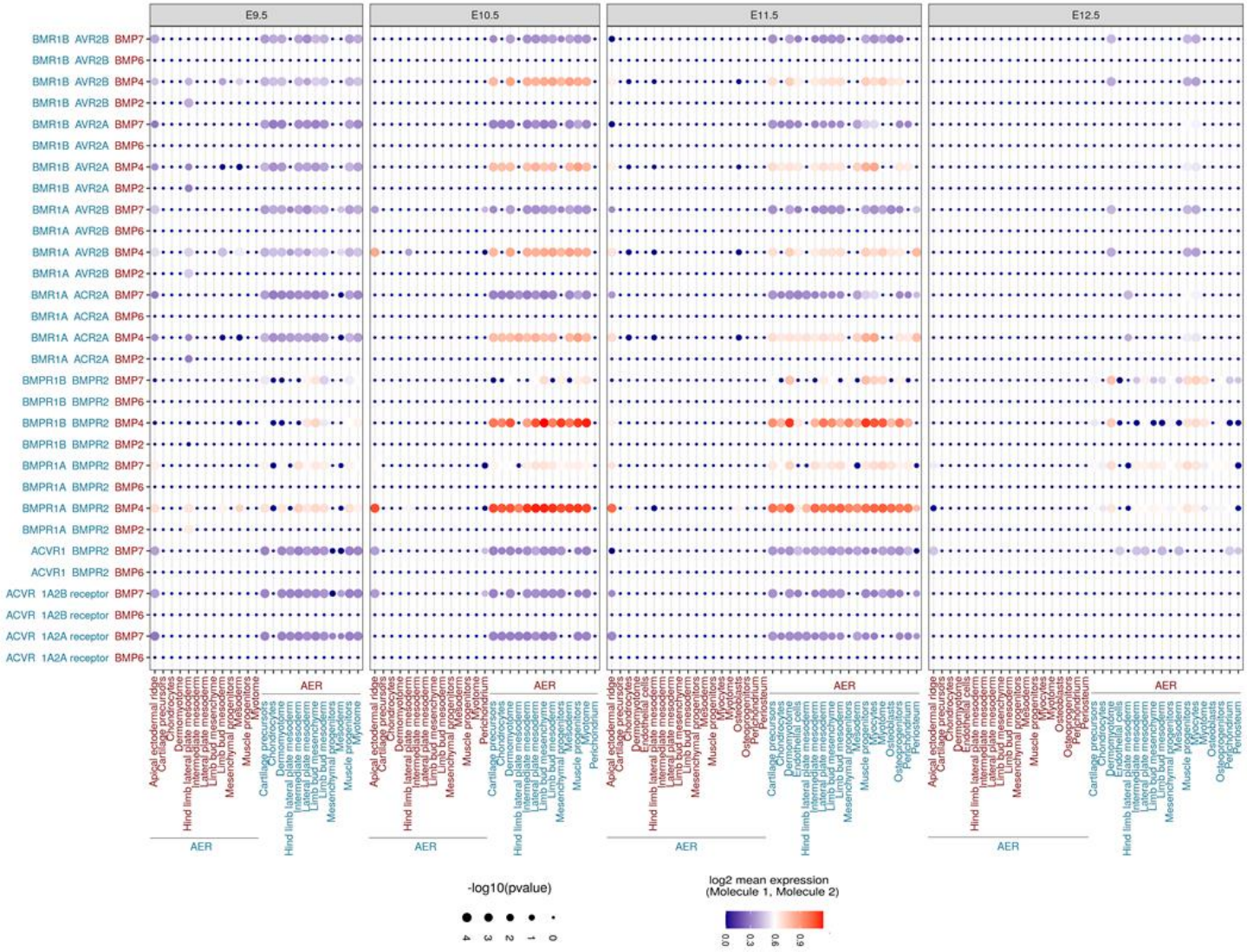
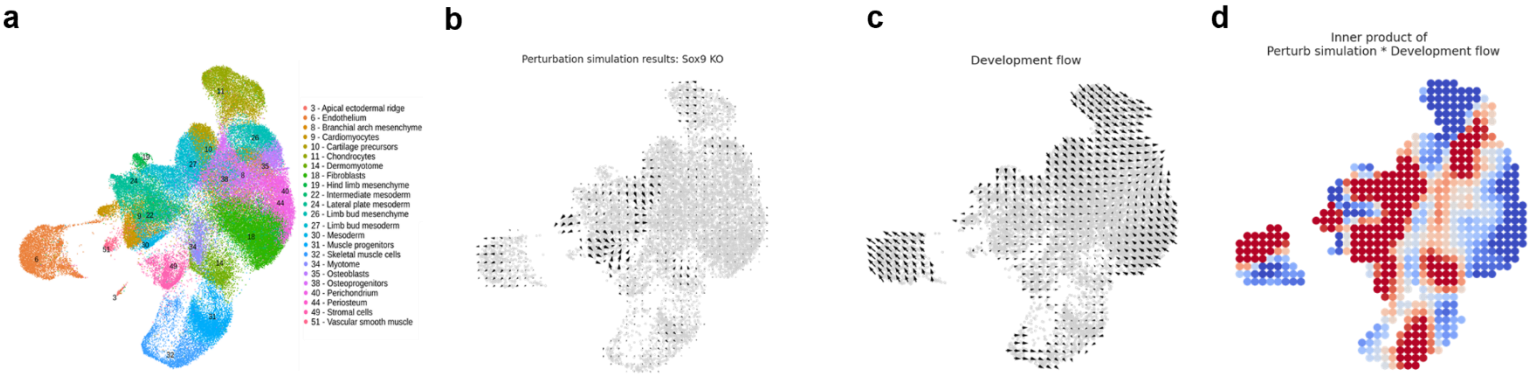
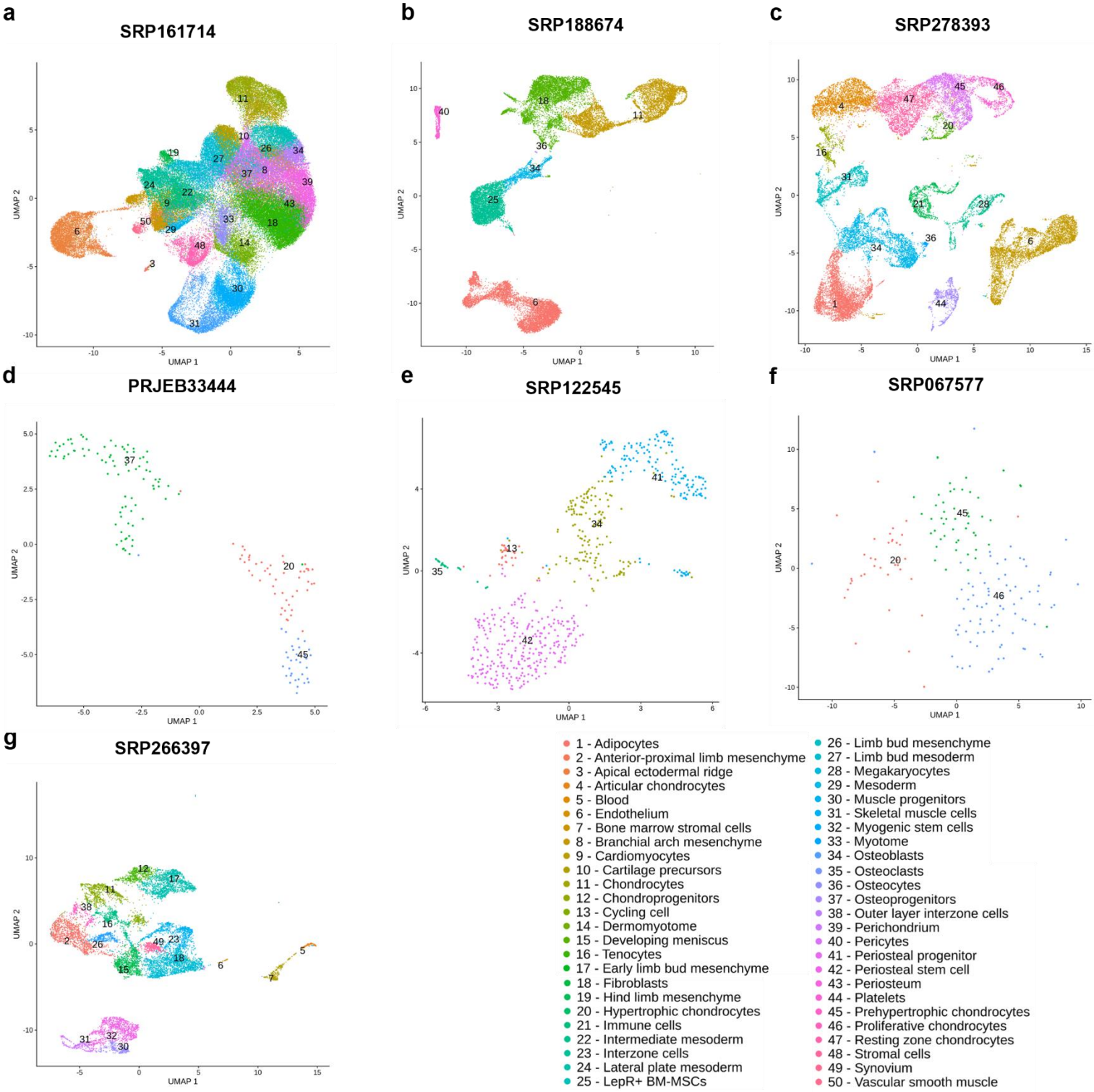

E9.5
E10.5
E11.5
E12.5
BMR1B AVR2B BMP7
BMR1B AVR2B BMP6
BMR1B AVR2B BMP4
BMR1B AVR2B BMP2
BMR1B AVR2A BMP7
BMR1B AVR2A BMP6
BMR1B AVR2A BMP4
BMR1B AVR2A BMP2
BMR1A AVR2B BMP7
BMR1A AVR2B BMP6
BMR1A AVR2B BMP4
BMR1A AVR2B BMP2
BMR1A ACR2A BMP7
BMR1A ACR2A BMP6
BMR1A ACR2A BMP4
BMR1A ACR2A BMP2
BMPR1B BMPR2 BMP7
BMPR1B BMPR2 BMP6
BMPR1B BMPR2 BMP4
BMPR1B BMPR2 BMP2
BMPR1A BMPR2 BMP7
BMPR1A BMPR2 BMP6
BMPR1A BMPR2 BMP4
BMPR1A BMPR2 BMP2
ACVR1 BMPR2 BMP7
ACVR1 BMPR2 BMP6
ACVR 1A2B receptor BMP7
ACVR 1A2B receptor BMP6
ACVR 1A2A receptor BMP7
ACVR 1A2A receptor BMP6
AER
-log10(pvalue)
4 3 2 1 0
log2 mean expression
(Molecule 1, Molecule 2)
0.0 0.3 0.6 0.9

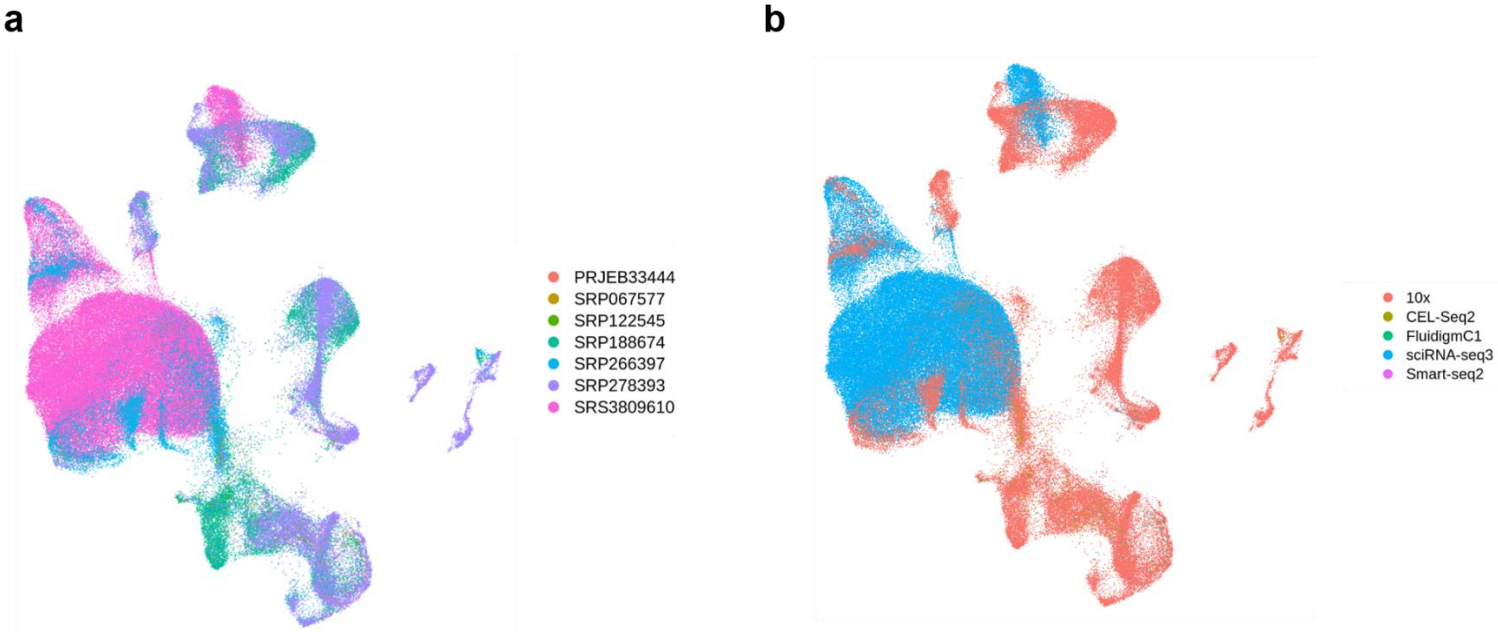
Figure 4


a
3 - Apical ectodermal ridge
6 - Endothelium
8 - Branchial arch mesenchyme
9 - Cardiomyocytes
10 - Cartilage precursors
11 - Chondrocytes
14 - Dermomyotome
18 - Fibroblasts
19 - Hind limb mesenchyme
22 - Intermediate mesoderm
24 - Lateral plate mesoderm
26 - Limb bud mesenchyme
27 - Limb bud mesoderm
30 - Mesoderm
31 - Muscle progenitors
32 - Skeletal muscle cells
34 - Myotome
35 - Osteoblasts
38 - Osteoprogenitors
40 - Perichondrium
44 - Periosteum
49 - Stromal cells
51 - Vascular smooth muscle
b
Perturbation simulation results: Sox9 KO
c
Development flow
d
Inner product of
Perturb simulation * Development flow

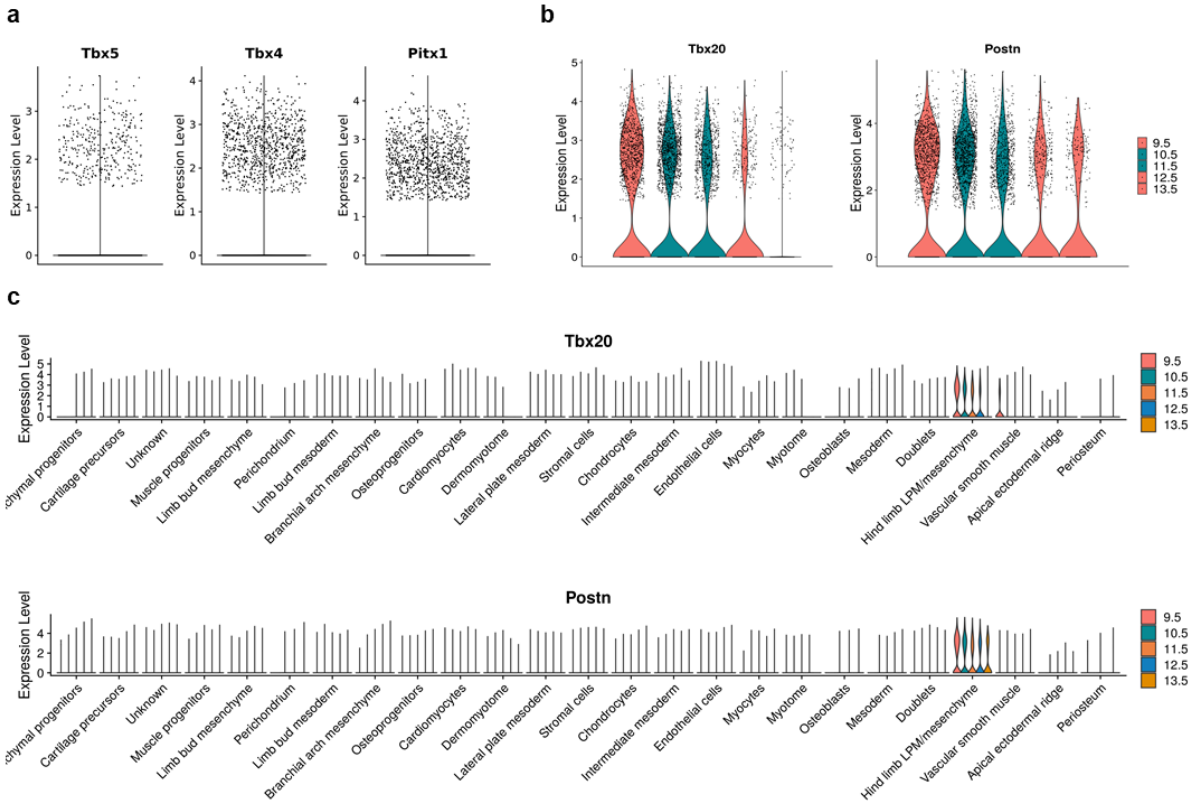
Extended Data Figure 1



Extended Data Figure 2



Extended Data Figure 3



Supplementary Table 1

Study (database ID)	Study (PMID)	Study (First, Last Author)	Technology	Age
SRP161714	30787437	Cao, Trapnell & Shendure	sciRNA-seq3	E9.5, E10.5, E11.5, E12.5, E13.5
SRP188674	31130381	Baryawno, Regev & Scadden	10x	8-10 weeks
SRP278393	34260921	Sivaraj, Adams	10x	P21
PRJEB33444	31543445	Bohm, Maes	Smart-seq2	E15.5
SRP122545	30250253	Debnath, Landau & Greenblatt	CEL-seq2	P7
SRP067577	27160914	Li, Wu	FluidigmC1	P7
SRP266397	32580935	Bian, Cahan	10x	E12.5, E13.5, E14.5, E15.5

Supplementary Table 2

Gene	Forward primer	Reverse primer
Tbx20	ATTAACCCTCACTAAAGGGAGAAGCTCACGAACAACGAGC	TAATACGACTCACTATAGGGAAAGTGGGGCCAGTACCTTG
Postn	ATTAACCCTCACTAAAGGGATGAGCCTGGTGTGACAAACA	TAATACGACTCACTATAGGGTGTACTCAGTTCCAGCTCCCT