scientific reports

OPEN

A single-cell atlas of the murine limb skeleton integrating the developmental and adult stages

Tim Herpelinck[1,2,7], Liesbeth Ory[1,2,7], Tom Verbraeken[1,2], Gabriele Nasello[1,2], Mojtaba Barzegari[3], Johanna Bolander[4], Frank P. Luyten[1,2], Przemko Tylzanowski[1,2,5,8] & Liesbet Geris[1,2,3,6,8✉]

The recent growth of single-cell transcriptomics has made single-cell RNA sequencing (scRNA-seq) into a near-routine technique. Breakthroughs in 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. However, the skeleton remains an underrepresented organ system in organism-wide atlases. Given the skeleton's critical role as the central framework of the vertebrate body, its function in housing the hematopoietic niche, and its involvement in metabolic and homeostatic processes, its underrepresentation presents a significant gap in current reference atlas projects. To address this issue, we integrated ten separate murine, publicly available scRNA-seq datasets, which include limb skeletal cells and their developmental precursors, resulting in an atlas of 133,332 cells. This limb skeletal cell atlas describes cells within the mesenchymal lineage, focusing on the process from limb induction to adult bone formation, and encompasses 39 well-characterized cell types and states. By expanding the repertoire of time points and cell types within a single dataset, we enable more complete analyses of cell-cell communication or in silico perturbation studies. Together, these efforts present a valuable resource for researchers in skeletal biology, metabolism, and regenerative medicine, filling an important gap in current atlas mapping projects.

Keywords Atlas, Single-cell, Transcriptome, Limb, Morphogenesis, Bone, Bioinformatics

The skeleton is a highly advanced organ 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[1]. In addition, some skeletal tissues possess remarkable regenerative properties, with bone being able to spontaneously regenerate after fracture with minimal scar formation[2]. While complex in its functions, the skeletal system is composed of a relatively limited set of core cell types, particularly within the structural and regulatory compartments of bone tissue, such as osteoblasts, osteocytes, chondrocytes, bone lining cells, and osteoclasts[3]. Its functional diversity appears to arise from a high degree of intrinsic heterogeneity within these cell types, allowing for regionalized specialization across skeletal sites. In contrast, non-mesenchymal compartments such as the bone marrow, harbor a much broader variety of hematopoietic and immune cell types[4].

Advancements in ultra-high-throughput single-cell RNA sequencing (scRNA-seq) technologies, along with the development of computational algorithms to analyze the data, enable the creation of organism-wide transcriptome maps that are resolved in both time and space[5–8]. To provide a complete representation of a tissue, atlases must account for biological variability, including multiple models, strains, and diverse environmental conditions. This approach enhances the generalizability of atlas-based findings[9]. Several high-profile efforts, such as the Tabula Muris, Mouse Cell Atlas, and Human Cell Atlas, have successfully applied this principle, generating comprehensive references by integrating data across tissues, developmental stages, and experimental

[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, 7 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. [7]Tim Herpelinck and Liesbeth Ory contributed equally to this work and share first authorship. [8]Przemko Tylzanowski and Liesbet Geris share senior authorship. ✉email: liesbet.geris@kuleuven.be

conditions. These resources show how harmonized datasets can support automatic annotation and consistent classification across studies[10–13].

However, the skeletal system remains underrepresented in most existing atlases, often lacking detailed annotation of skeletal lineages or user-friendly methods for interpretation. This results in a significant gap in the transcriptional characterization of the skeletal system. While large atlases contain abundant data, much of it remains unexplored due to imprecise annotations, whereas specialized datasets remain confined within their specific research domains. To address this gap, we reannotated and integrated publicly available datasets focused on murine mesenchymal and skeletal cell types and states. These datasets span various sexes, strains, and anatomical locations, allowing us to incorporate the natural biological variability within the skeletal system. Integrating this diversity is essential to create an atlas that reflects the full spectrum of skeletal cell phenotypes and is generalizable across experimental contexts. By combining these datasets into a single reference framework of 133,332 cells, we provide a more robust and inclusive foundation for skeletal cell classification. We ensured consistency across all datasets by using the original labels and enhancing them with more detailed annotations through the use of marker genes, creating the Limb Skeletal Cell Atlas (LSCA).

Following the generation of the LSCA, we explored its applicability and predictive value, as schematically represented in Fig. 1. We demonstrated the potential of the LSCA in intercellular communication analysis and pseudotemporal trajectory inference. The predictive capacity of the LSCA was further tested by simulation of SRY-box transcription factor 9 (Sox9) inactivation and the analysis of its consequences, which were in line with the in vivo phenotype[14]. Collectively, these results support the notion that the LSCA can be used, amongst others, as a reliable reference for the automated annotation of new skeletal scRNA-seq datasets. The name for the Atlas was chosen to emphasize the focus on mesenchymal and skeletal tissues of the limb while also including some, but not all, other cell types such as muscle, endothelial and hematopoietic lineages. To facilitate accessibility and reproducibility, all notebooks required to analyze the datasets, build the atlas and perform the subsequent analyses have been made available online.

Results

Dataset annotation and integration to produce a limb skeletal cell atlas

To build a reference atlas of the limb skeleton, we selected ten publicly available mouse scRNA-seq datasets containing cells from the onset of limb development to mature bone[14–23]. These datasets encompass a broad range of developmental stages (E10.5 to 16 weeks), anatomical sites (forelimb, hindlimb, femur, tibia, humerus, cortical bone, periosteum, endosteum, bone marrow, and synovium), and skeletal compartments (growth plate, epiphysis, bone marrow stroma). The majority of the datasets were derived from C57BL/6 strains, with some datasets also including transgenic reporter lines such as Osx-Cre: GFP and CTSK-mGFP (Sup. Table 1). A comprehensive analysis was conducted on a total of 133 332 individual cells that met the quality control filters. The individual datasets were manually reannotated based on the original labels and canonical marker gene expression (Sup. Figure 1, Sup. Table 2). Next, we selected four top-performing methods from the benchmarking study by Luecken et al. to determine the most suitable integration approach for our data[24]. We tested both an unsupervised (single-cell Variational Inference; scVI)[25,26] and a semi-supervised (single-cell Annotation using Variational Inference; scANVI)[25,27–30] integration model, selected based on their consistent performance in complex integration tasks[24]. The scVI model allowed us to infer the underlying structure of the data without prior labels, while scANVI leveraged both labeled and unlabeled data to improve annotation accuracy. Additionally, we evaluated Scanorama, which uses a method similar to computer vision algorithms for panorama stitching, identifying and merging images with overlapping content into a cohesive panorama[31]. We also considered
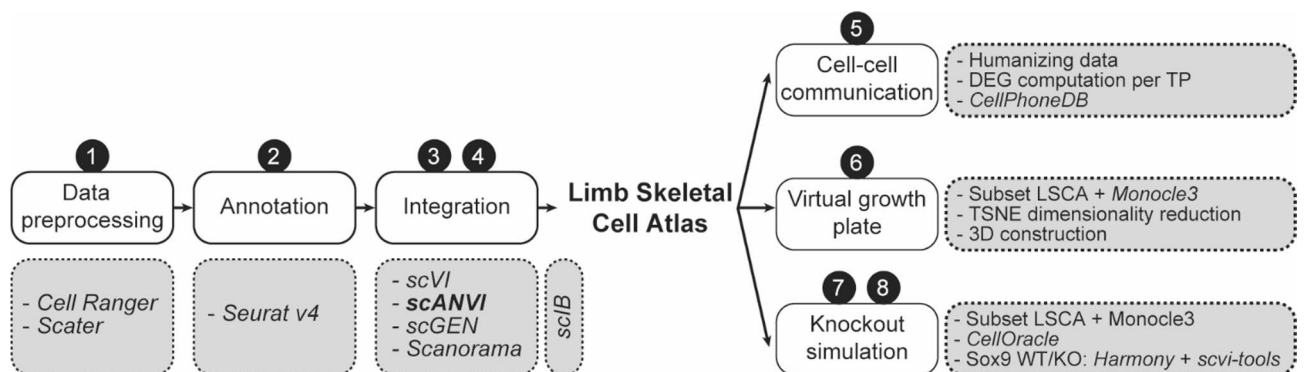

Fig. 1. Flow diagram of computation methods. (1) Publicly available datasets were preprocessed using Cell Ranger and Scater. (2) Individual datasets were then clustered and annotated using Seurat v4. (3) We tested scVI, scANVI, scGen and Scanorama and (4) evaluated them with the scIB benchmarking package. scANVI was used to create the final Limb Skeletal Cell Atlas. (5) Cell-cell interactions for early limb bud signaling pathways were predicted using CellPhoneDB and (6) the growth plate was reconstructed with the use of Monocle3 pseudotemporal ordering. (7) Finally, an in silico knockout simulation for Sox9 was performed with Monocle3 and CellOracle and (8) evaluated with in vivo wild-type and knockout data integrated/projected with Harmony and scvi-tools. *WT* wild-type, *KO* knockout

scGen, a generative model designed with a variational autoencoder architecture to achieve near-linear mappings for non-linear sources of variation, particularly useful for correcting batch effects using labeled data[32].

To compare the effectiveness of these integration methods, we employed metrics focused on batch correction while preserving the biological variability inherently present in the data (scIB)[24,25] (Sup. Figure 2). scANVI demonstrated a more favorable performance, achieving the highest label isolation score and effectively integrating the datasets without introducing cross-laboratory batch effects (Sup. Figure 3). The robust performance of scANVI ensured that the integration was cell-based, accurately representing 39 distinct cell states (Fig. 2a, b). Given the complexity of our data, this integration approach provides a reliable foundation for downstream analyses and applications. Twenty-six clusters originated from a mesenchymal lineage (Clusters #1–26) and encompassed mesenchyme, chondrocyte, osteoblast/osteocyte or fibroblast cell states. A further four states were associated with the muscle lineage (Clusters #27–30) and other clusters included endothelial (Clusters #31–32), epithelial (Clusters #33–34), hematopoietic (Clusters #35–37), neuronal (Cluster #38) and ectodermal (Cluster #39) cell states (Fig. 2a, b).

One of the drawbacks of single-cell RNA sequencing is the loss of spatial data. However, a pseudospatial context within the cells of the early limb bud can be reconstructed by identifying distinct gene expression patterns. For this purpose, we defined the proximal-distal axis by mapping the expression of the Hoxa/d9-13
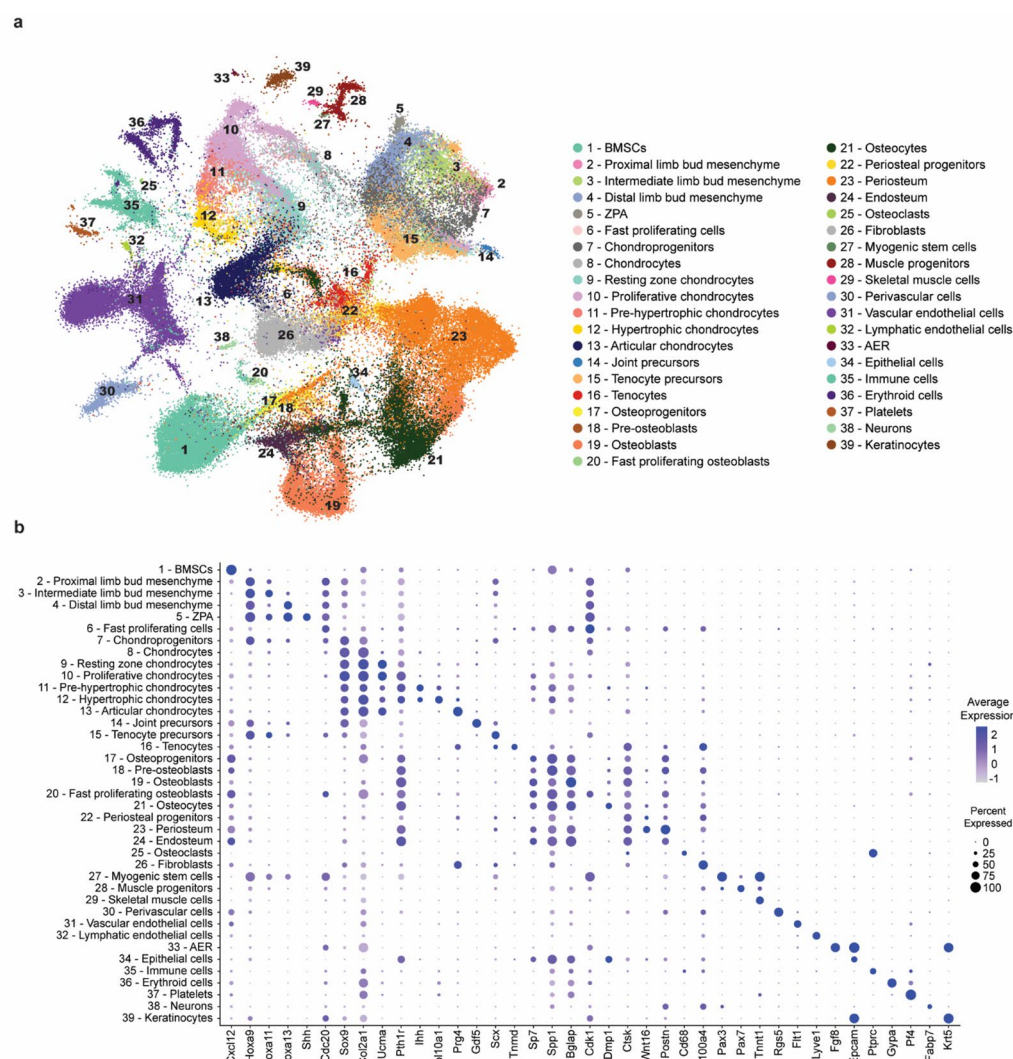

Fig. 2. An integrated compendium of skeletal cell types with detailed annotation. **a** UMAP visualization of the scANVI latent space of 133 332 murine limb mesenchyme- and skeleton-derived cells, colored by annotation. **b** Dot plot showing the expression of one selected marker gene per cluster. The color of the dot represents the mean expression level and its size represents the percentage of cells within the cluster in which that gene was detected. For visual clarity, a single representative gene was chosen per cell type, prioritizing specificity over abundance. This visualization is intended to illustrate the distinctiveness of each cluster. Full marker combinations used for annotation are provided in Supplementary Table 2. *BMSCs* bone marrow-derived stromal cells, *ZPA* zone of polarizing activity, *AER* apical ectodermal ridge

mRNAs, from the most proximal (Hoxd9) to the most distal (Hoxd13) (Fig. 2b)[33]. The expression of Sonic Hedgehog (Shh) mRNA was used to define the Zone of Polarizing Activity (ZPA) located at the posterior part of the limb bud[34]. A subset of ectodermal cells situated at the distal tip of the limb bud, called the Apical Ectodermal Ridge (AER), is a signaling center associated with proximal-distal limb growth. It could be identified by the expression of Fibroblast growth factor 8 (Fgf8) mRNA[35].

To validate the cluster annotations in the LSCA, we compared the differentially expressed genes of specific cell populations—AER, distal mesenchyme and hypertrophic chondrocytes—with published experimental data. We focused on genes that are not widely recognized as traditional markers for limb development but nevertheless exhibited distinct expression patterns within the clusters. For each cluster, we identified a key gene: Sp8[36] for the AER, Hottip[37] for the distal mesenchyme and Loxl4[38] for the hypertrophic chondrocytes in the growth plate (Sup. Figure 4). We then confirmed their expression in the corresponding tissues by consulting independent (whole mount) in situ hybridization data from the literature. While these genes are not established markers in the field, their specific expression patterns in the literature provided additional support for our cluster annotations and highlighted novel molecular features within these cell populations, further validating the LSCA.

Intercellular communication inference across the developing limb

Intercellular communication driven by ligand-receptor interactions is one of the mechanisms regulating cellular differentiation. Therefore, a wide variety of tools to infer intercellular signaling from scRNA-seq data has been developed[39]. Here, we used CellPhoneDB[40] as it considers multimeric receptor complexes when inferring ligand-receptor interactions. We analyzed Bone Morphogenetic Protein (Bmp) and Shh signaling taking place within the limb bud, the AER and the ZPA (Fig. 3a).

Bmps are members of the Transforming Growth Factor (Tgf) ligand superfamily and signal through a tetrameric receptor complex. Type I and Type II Bmp receptors are expressed across the limb mesenchyme[41,42] and AER[41] which also acts as a source of Bmps[43]. However, due to the multimeric character of the receptor complexes resulting in many different combinations, it is challenging to experimentally determine which ligand-receptor pair is predominantly used. Screening with CellPhoneDB narrowed down the number of possible combinations. The analysis revealed that the AER mainly signals to the distal limb bud mesenchyme through Bmp4 and Bmp7, particularly at E11.5 (Fig. 3b, Sup. Figure 5–7) and that they appear to favor binding a receptor complex containing Bmpr1a. The regulation of the AER by Bmp signaling, whether from the AER itself, the mesenchyme, or both, remains a topic of ongoing discussion in the field[43]. Our analysis indicates that autocrine signaling is the primary mediator, particularly at E10.5. Additionally, we observed limited Bmp signaling from the limb mesenchyme towards the AER. Noteworthy is that the favored ligand-receptor combination is invariable over time (Fig. 3b, Sup. Figure 5–7).

Antero-posterior outgrowth and proximodistal polarization are interconnected to each other by a Shh epithelial-mesenchymal feedback loop[44]. Shh mRNA expression decreases as the limb bud grows, which explains the decrease in cell-cell communication from E10.5 to E11.5 and no Shh expression in E12.5. Shh signaling is regulated during limb development at many levels and one of them is a negative feedback loop involving binding of Shh to Hedgehog interacting protein (Hhip) which acts as a decoy receptor to inhibit Hedgehog signaling[45]. To demonstrate the usefulness of the LSCA, we investigated this regulatory loop and found that cell-cell communication is mostly limited to intermediate limb bud mesenchyme and chondroprogenitors (Fig. 3a, b).

Virtual reconstruction of the growth plate

Long bones are formed through the process of endochondral bone formation. Initially, mesenchymal cells condense and then undergo a series of differentiation steps going from resting chondrocytes over proliferating to pre-hypertrophic and hypertrophic chondrocytes. These cells form columnar structures that together make up the growth plate. The hypertrophic chondrocytes then either transdifferentiate into osteoblasts or undergo apoptosis and the space is replaced by the osteoblasts producing bone tissue[46].

As a part of the validation of our atlas, we virtually restored the growth plate. To accomplish that, we used pseudotime analysis to reconstruct the transcriptional trajectories from resting to hypertrophy zone in silico (Fig. 4a, Sup. Figure 8). Cells were then binned by pseudotime and for each bin, we performed dimensionality reduction by t-distributed stochastic neighborhood embedding (t-SNE). Importantly, we imposed a circle as a boundary condition for each t-SNE. Alignment of these circular projections then recreated the cylindrical shape of the growth plate, while also visualizing transcriptional heterogeneity within the growth plate across bins of pseudotime (Fig. 4b). This approach provided an intuitive way to visualize gene expression patterns, allowing researchers, particularly those without bioinformatics expertise, to interpret the data in a 2D or 3D-like reconstruction of the growth plate. Plotting gene expression along the pseudotime axis, combined with expression in pseudospace, artificially restored to a certain extent, the tissue architecture (Fig. 4c). While this reconstruction does not provide actual spatial data, it offers an easily interpretable representation of growth plate signaling. Future iterations of the LSCA could integrate spatial transcriptomics data to provide a more precise and spatially resolved view of the growth plate.

Simulation of transcription factor perturbation

Developmental biology has been a key field for studying gene regulatory networks (GRNs)[47]. Traditionally, this involved a series of experiments where transcription factor activity is altered by gain- or loss-of-function and the resulting in vivo effects are analyzed. With the advances in bioinformatics, the CellOracle[48] algorithm was developed to study GRN inference from scRNA-seq data. In addition, it allows us to explore in silico the effects of the perturbation of transcription factor expression on target gene expression. To test if our dataset was amenable to this type of analysis, we decided to compare an in silico prediction to published in vivo knock-out data[14].
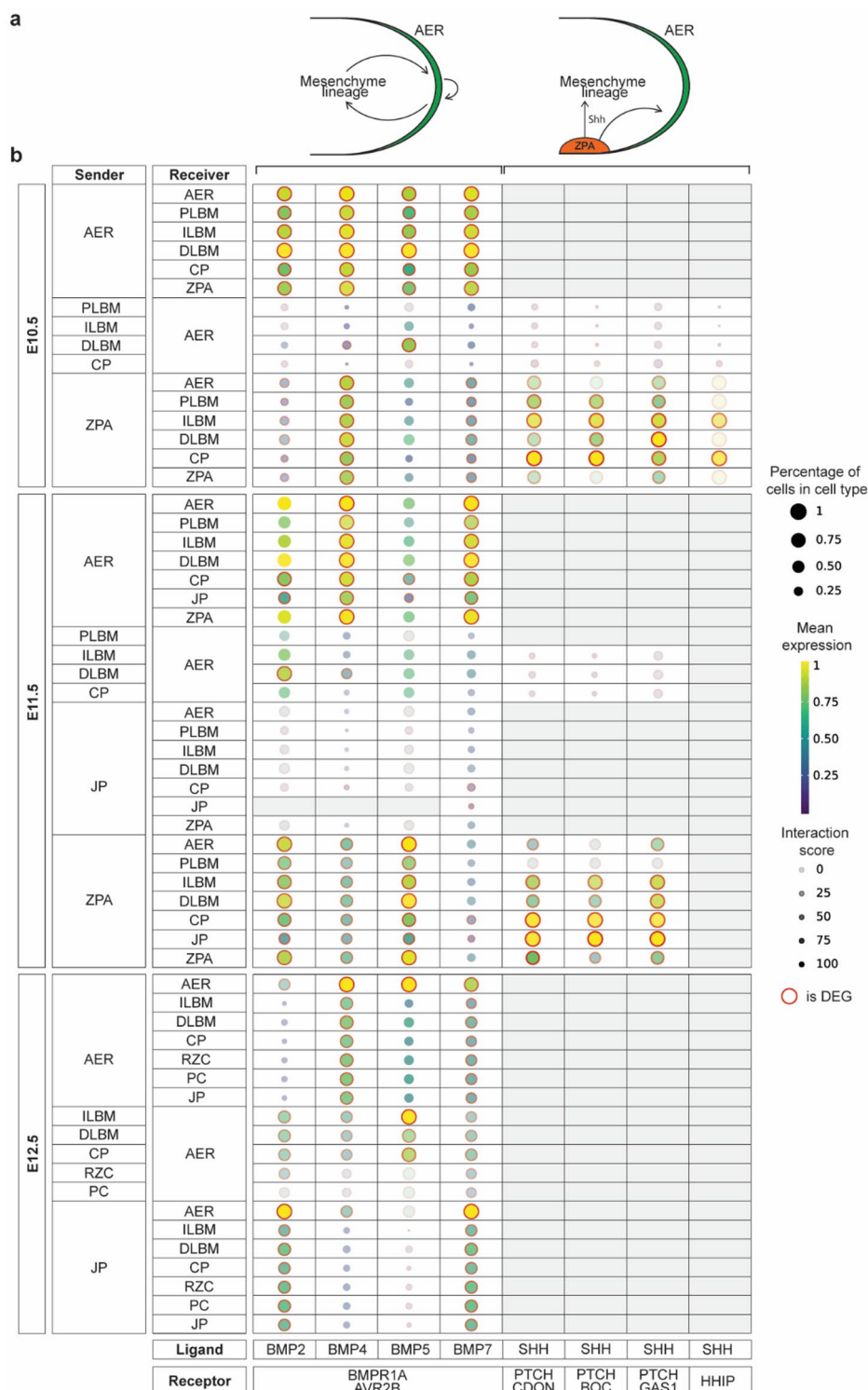
Fig. 3. BMP and SHH signaling case study between AER, ZPA and mesenchyme. **a** Schematic illustration for visualizing AER and ZPA signaling. **b** Predicted ligand-receptor interactions across developmental timepoints (E10.5, E11.5 and E12.5). The dot plot illustrates ligand-receptor interactions between sender and receiver cell types (y-axis) and ligand-receptor pairs (x-axis). Dot color indicates mean gene expression, while dot size represents the proportion of cells within each cell type expressing the gene. Translucency reflects interaction specificity and differentially expressed genes are marked with an outer red ring. *DEG* differentially expressed genes, *AER* apical ectodermal ridge, *PLBM* proximal limb bud mesenchyme, *ILBM* intermediate limb bud mesenchyme, *DLBM* distal limb bud mesenchyme, *CP* chondroprogenitors, *RZC* resting zone chondrocytes, *PC* proliferative chondrocytes, *PHC* pre-hypertrophic chondrocytes, *JP* joint precursors, *ZPA* zone of polarizing activity
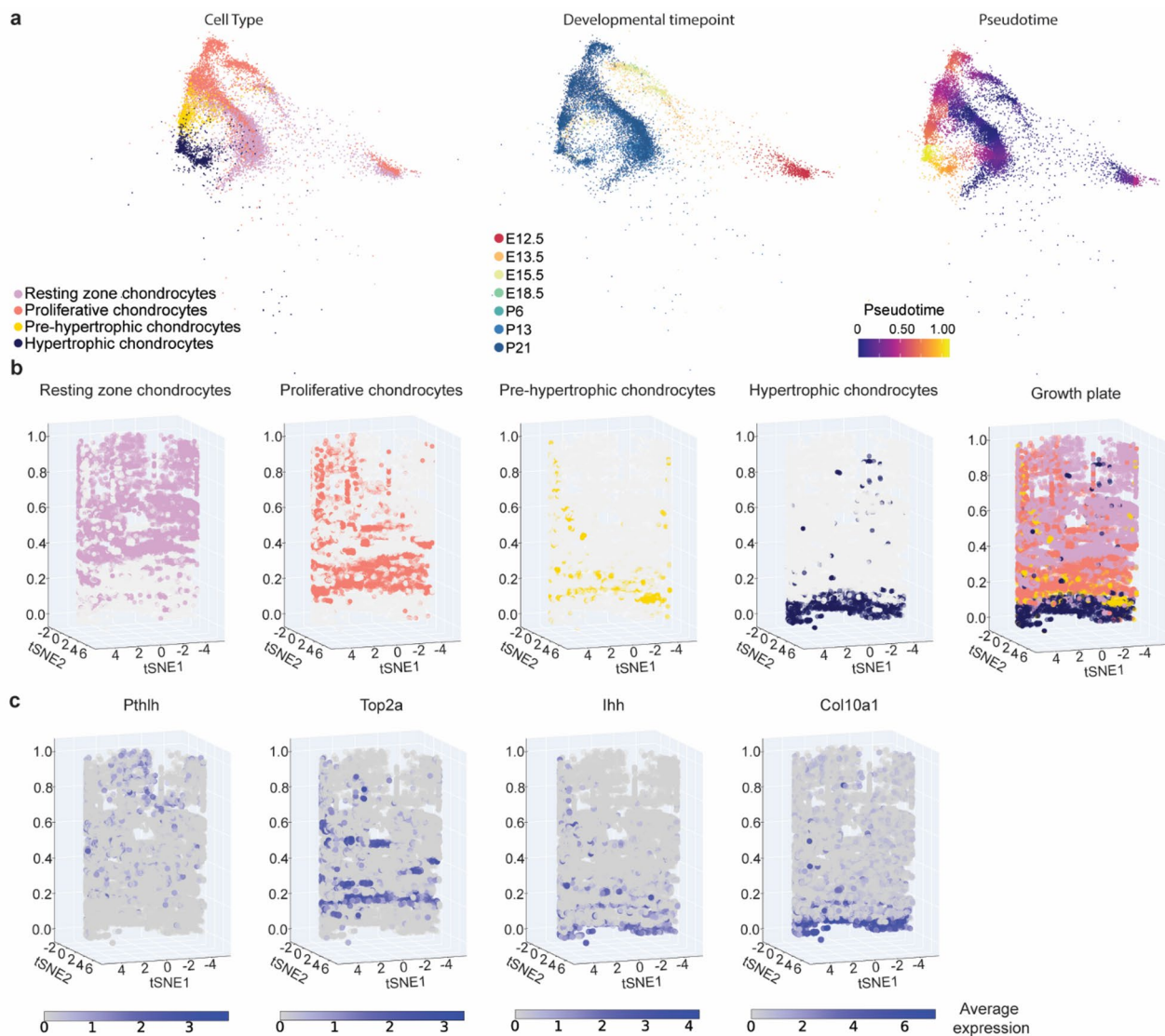
Fig. 4. Pseudospatiotemporal reconstruction of the transcriptional dynamics within the growth plate. **a** UMAP visualization of the scANVI latent space of the integrated growth plate data subset from the main atlas, colored by annotation (left), developmental timepoint (middle) and Monocle 3 pseudotime value (right). **b** Growth plate chondrocytes were grouped in 50 bins based on similar pseudotime values. t-SNE dimensional reduction was performed on each bin, using a circle with a radius of 20 as the boundary condition for gradient descent, thus reconstructing the cylindrical shape of the growth plate upon stacking the bins. **c** Expression of known marker genes across different growth plate zones in pseudotime and pseudospatial dimensions

Sox9 is involved in many stages of chondrogenesis, from regulating the initial mesenchymal condensation to maintaining the proliferation of columnar chondrocytes and preventing their premature transdifferentiation into osteoblasts. It also plays a crucial role in inducing chondrocyte hypertrophy, making it indispensable for the proper formation and maintenance of functional growth plates. Inactivation of Sox9 leads to a shortening of the columnar and hypertrophic zones in the growth plate and accelerates ossification due to premature pre-hypertrophy and matrix mineralization[14,49].

Here, we demonstrate that the LSCA can be used to generate in silico gene knockouts, which can then be compared directly to single-cell in vivo knockout datasets and phenotypic observations, offering a robust framework to explore developmental and regulatory disruptions. To evaluate the effectiveness of the Atlas in exploring the consequences of targeted gene inactivation, we applied the CellOracle algorithm to create a virtual Sox9 knockout and compared our in silico predictions to data from an in vivo knockout study (Fig. 5a, b). For this analysis, we subsetted the LSCA to include time points between E10.5 and P21, focusing on relevant cell types such as periosteal cells, chondrocytes at various differentiation stages, osteogenic cells, endothelial cells (vascular and lymphatic), limb bud mesenchymal populations, and myogenic cells. We modeled the data of late inactivation of Sox9, driven by the Acan-Cre expressed in the growth plate. We defined the accessible cell states as those that cells can differentiate under the current regulatory conditions, while inaccessible states cannot be
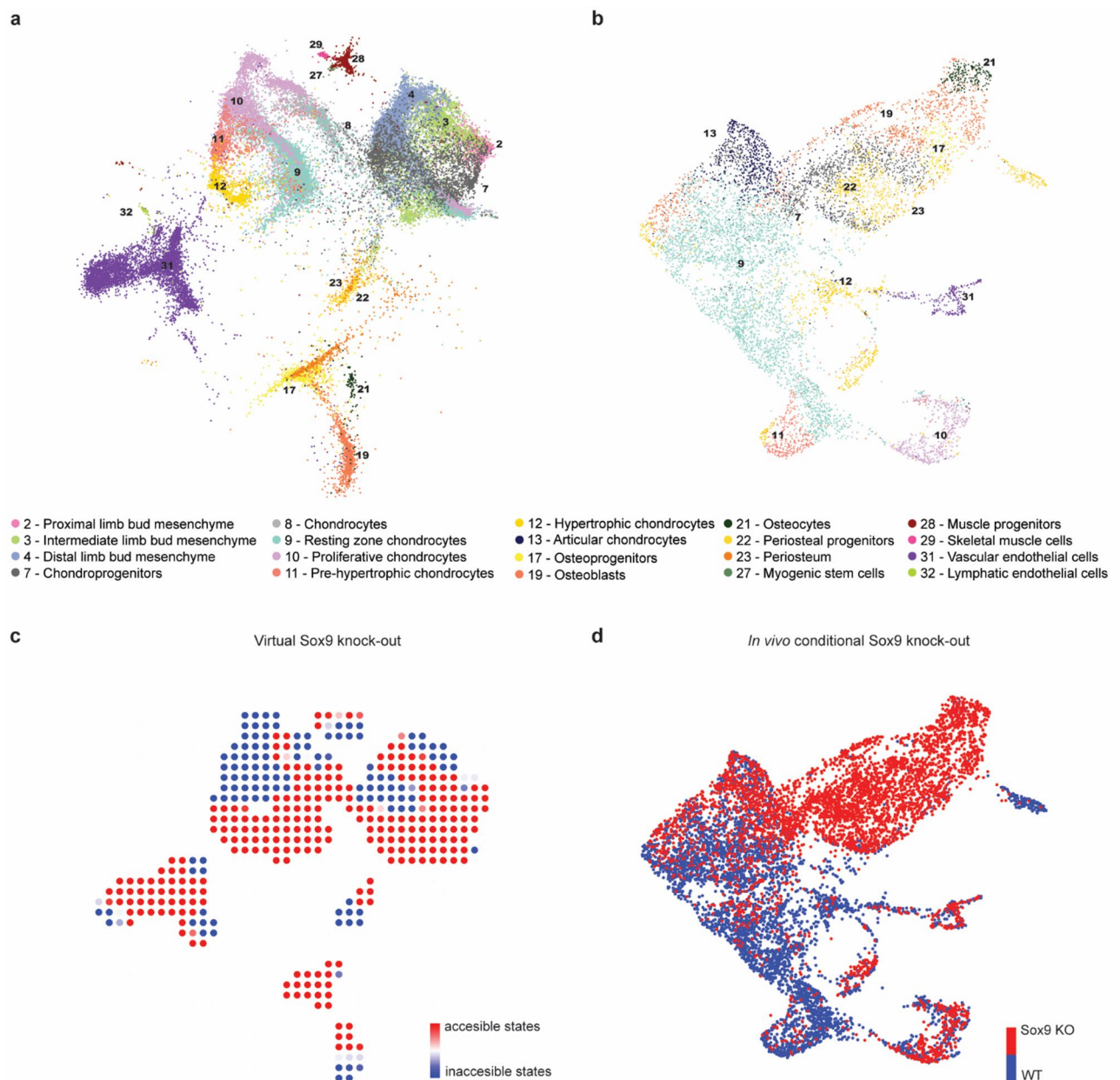
Fig. 5. In silico predictions of Sox9 knockout (KO) compared to in vivo data. **a** Reference annotation of the subsetted atlas, which includes only the relevant cell types from developmental stages E10.5 to P21, used for knockout analysis. **b** UMAP plot integrating WT and Sox9 KO in vivo data. **c** Inner vector product of two vector fields: the pseudotime gradient from unperturbed conditions and the cell state transition probability following in silico Sox9 perturbation. Red regions indicate accessible cell states, while blue regions represent inaccessible cell states. The loss of chondroprogenitor and prehypertophic chondrocyte cell identities as accessible states reflects the biological profile of the ground truth shown in **d**. Myogenic, osteogenic and endothelial differentiation are largely unaffected. D, UMAP of integrated Sox9 KO and WT in vivo data, representing ground truth affected cell identities colored by condition (red: Sox9 KO, blue: WT)

reached due to the absence or inactivation of essential transcription factors. Using the Atlas, we identified regions within the dataset that are either developmentally accessible or inaccessible to Sox9's regulatory influence.

The simulations showed that upon removal of Sox9, most mesoderm-derived cells failed to contribute to cartilage formation, while myogenesis, osteogenesis, and endothelial differentiation remained largely unaffected (Fig. 5c, Sup. Figure 9, 10). Specifically, in the virtual Sox9 inactivation model, cell type 11—pre-hypertrophic chondrocytes—is represented as a blue region, indicating that mesenchymal cells cannot access this cell state. These results are supported by the in vivo findings, where Sox9 inactivation prevents mesodermal cells from differentiating into chondrocytes, leading to an accumulation of cartilage precursors and osteogenic

cells (Fig. 5d). Phenotypically, the absence of Sox9 in early limb buds disrupts mesenchymal condensation, chondrocyte differentiation, and the establishment or differentiation of the osteoblast lineage[50]. This is reflected in our in silico Sox9 knockout, where the distal mesenchyme predominantly remains in an inaccessible state, indicating impaired mesenchymal condensation and chondrogenic cell fate commitment. Chondrocytes, including proliferating and pre-hypertrophic populations, as well as a substantial portion of osteoblasts are predicted to be inaccessible states, mirroring the absence of their differentiation (Fig. 5c, Sup. Figure 9, 10). Together, these findings underscore the value of the LSCA as a powerful tool for in silico modeling, enabling a deeper understanding of the mechanisms underlying gene function and their developmental consequences.

A limitation of the virtual knockout is that it accounts only for a neighborhood developmental flow but not lineage dependency. As a result, it will mark a region as an inaccessible cell state if its development depends on Sox9, however, it will mark regions dependent on this (inaccessible) cell type as accessible. For instance, cell type 12—hypertrophic chondrocytes—is shown as a red region in the virtual knockout, indicating that this cell state is predicted to be present in Sox9-perturbed situations. However, the in vivo data show that this cell type is absent in the knockout (red) because it originates from pre-hypertrophic cells.

Discussion

We have generated a Limb Skeletal Cell Atlas (LSCA) representing a manually curated compendium of 133,332 cells across ten datasets and explored its applicability in both data- and hypothesis-driven analyses. To assemble the LSCA, we curated data from ten publicly available datasets[14–23] representing murine cells spanning various developmental stages and tissues. The integration of these datasets was carefully performed using advanced computational methods, including normalization, dimensionality reduction, clustering, and multiple integration techniques (scVI, scANVI, Scanorama, and scGen). A flow diagram summarizing the methodology used to construct the Atlas outlines each step of the preprocessing, clustering, integration, and validation processes (Fig. 1).

First, analysis of intercellular communication by Bmp signaling sheds light on a longstanding question in the field of limb development. Specifically, 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[43]. Based on CellPhoneDB's cell-cell communication inference, considering the subunit architecture of both ligands and heteromeric receptors[39,40]. Our case study suggests the AER to be the dominant source of Bmp. This finding does, however, require further in vivo validation.

Next, we constructed the spatiotemporal map of the growth plate that recapitulated the in vivo situation. We achieved that by imposing boundary conditions on the gradient descent of t-SNE dimensionality reduction with pseudotime calculations, which revealed a progression through distinct chondrocyte subtypes. Similar results have been obtained previously by warping the principal component space[51].

Finally, we demonstrated that the in silico inactivation of Sox9 aligns with in vivo results, showing the absence of cells dependent on Sox9 for differentiation into chondrocytes[14,52]. While these findings align with established in vivo observations, the LSCA should be considered primarily as a tool to explore hypotheses that will require further experimental validation. Given that we can assess the developmental outcome of computationally simulated transcription factor knockouts, the LSCA represents a valuable resource for both limb developmental biologists and skeletal biology researchers.

The current release is restricted to the healthy murine skeleton, with gaps in limb sites and developmental stages, as well as factors such as genetic background, sex, and the enrichment of sub-populations, which may influence the census and its downstream applications. These limitations highlight opportunities for future research to build upon our findings, addressing these gaps and expanding our understanding. We will continuously update and extend the LSCA (made available through the LSCA Github and web app, cfr infra), as data availability permits, toward the entire limb skeleton in health and disease. The goal of this study was to provide an initial reference of the skeletal system of the limb, which will serve as the basis for a collaborative effort to expand the Atlas.

While our murine LSCA offers a detailed view of limb and skeletal populations, direct comparisons to human data remain limited due to species-specific differences in development and annotation. Recent human limb single-cell atlases provide valuable opportunities for cross-species comparisons, and future work could integrate human datasets to enhance cross-species analyses and expand the LSCA's utility for developmental and disease research[53–55].

In our study, we employed three robust methods to validate our atlas and demonstrate its practical applications, each with its own limitations. CellPhoneDB effectively infers cell-cell interactions from single-cell RNA sequencing data, though it faces challenges with indirect protein abundance measurements and the use of different names for the same receptor or ligand. Monocle3 excels in resolving detailed pseudotemporal trajectories, although handling large, complex datasets could be better suited with alternative graph-based methods. CellOracle performs well calculating in silico gene perturbations, but a recent critique highlights its failure to account for distal regulatory interactions and a flaw that skews benchmarking scores[56]. Importantly, all data, analyses, and codes required to replicate the results are freely available via GitHub. Users can easily adjust analysis choices by modifying the provided code to incorporate alternative methods or algorithms that best fit their needs, ensuring our Atlas can be tailored to a wide range of research applications.

As a part of our effort, we created a web portal to allow browsing of the data. This interactive web-based resource makes the mRNA expression data easily accessible, allowing users to explore cell types expressing genes of interest or uncover transcriptomic subpopulations within a cell type. Cells can be filtered by cell type and by metadata from the original studies, such as sequencing technique, developmental timepoint or age, tissue origin, and study ID. Gene expression can be visualized, with options to download results and access the full atlas for further analysis.

Importantly, while the manuscript presents a fixed, representative subset of datasets to build and validate the Limb Skeletal Cell Atlas, the web portal is designed as a dynamic and continuously evolving resource. We will regularly update the portal to incorporate newly published datasets, thereby expanding the repertoire of tissues, developmental stages, and cell types represented. This ensures that the atlas remains current and increasingly comprehensive for the research community.

In short, the Limb Skeletal Cell Atlas provides a robust and flexible framework for the characterization of most known cell populations in the limb skeleton and serves as a foundation for future studies in a wide variety of disciplines.

Materials and methods

Data preprocessing

We selected ten publicly available scRNAseq datasets containing wild-type murine cells spanning limb and skeletal tissues during development and adulthood[14–23]. or datasets where aligned count matrices were not provided, raw sequencing files were processed using Cell Ranger to generate gene expression matrices[57]. Initial Quality control (QC) of the raw count matrices was performed independently for each dataset using the Scater package[58]. QC metrics were computed per cell, including total UMI counts (library size), the number of detected genes, and the proportion of reads mapping to mitochondrial genes. Cells were excluded if they exhibited a library size or number of detected genes deviating by more than three median absolute deviations (MADs) from the median of their respective distributions. Similarly, cells with a mitochondrial read fraction exceeding three MADs above the median were removed, as these likely represent damaged or dying cells. Genes with low abundance (expression below 1E-3) were excluded, and duplicate gene entries, if present, were removed.

Clustering and dimensionality reduction

All datasets were analyzed individually prior to integration. The filtered count matrices were imported into Seurat v4[59]. 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 elbow 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.

Integration with scvi-tools, scanorama and ScGen

The integrated atlas was constructed using scVI[26] and scANVI[27–30] selected for their strong performance across complex datasets and compatibility with downstream analyses[24]. For preprocessing the datasets, we filtered for the 5000 most variable genes. First, we used scVI as an unsupervised tool to find common axes of variation between the datasets, helping to capture underlying data structures without prior labels(19). To refine the integration further, we then used scANVI, a semi-supervised tool, leveraging both labeled and unlabeled data to improve the accuracy of integration. scANVI builds on the shared axes of variation found by scVI and integrates cell type information, allowing for a data manifold that better represents the latent biological structure. The parameters were used as described in the scvi-tools[25] tutorial of '*Atlas-level integration of lung data*'. For Scanorama, we used parameters from the step-by-step tutorial provided by the National Bioinformatics Infrastructure Sweden (NBIS), as described on the Scanorama GitHub page[31]. We used the scGen batch removal tutorial with standard parameters from the scGen documentation page (readthedocs)[32].

Integration metrics

To calculate integration metrics, the scIB-metrics package was used[24] which contains scalable implementations of the metrics used in the scIB benchmarking suite[60–63]. Bio conservation was captured with the use of classical label conservation metrics, which assess local neighborhoods (graph cLISI, extended from cLISI), global cluster matching (Adjusted Rand Index: ARI, normalized mutual information: NMI) and a metric evaluating rare cell identity annotations (isolated label scores). Batch correction scores were measured via the k-nearest-neighbor batch effect test (kBET), k-nearest-neighbor (kNN) graph connectivity and the average silhouette width (ASW) across batches. Independently of cell identity labels, batch removal is measured using the graph integration local inverse Simpson's Index (graph iLISI, extended from iLISI) and PCA regression.

Validation of differentially expressed genes

To validate the cell annotation and evaluate the potential of the LSCA for novel marker gene detection, differential expression analysis was performed for all cell identities, retaining the top 10 results returned by the *FindMarkers()* function. As these genes are statistically the most specific for their respective cell state, they represent prime candidate biomarkers. Reassuringly, known canonical marker genes were present for each annotation. We then focused on the DEG lists of the Apical Ectodermal Ridge (AER), distal limb mesenchyme, and hypertrophic chondrocyte clusters. An extensive literature search was performed for the expression patterns of DEGs not considered canonical markers for these cell types. Based on this search, we were able to confirm the specific expression of Sp8 in the AER, Hottip in the distal mesenchyme and Loxl4 in hypertrophic chondrocytes as shown by (whole mount) in situ hybridization experiments on mouse embryos (Supp. Figure 4). These genes showed clear and specific expression patterns, supporting the accuracy of the cluster.

Cell-cell interaction prediction

Prediction of cell-cell communication by ligand-receptor interactions between cell types was performed using CellPhoneDB v5[40]. CellphoneDB is an open-access resource that provides a curated collection of receptors, ligands, and their interactions. The atlas was subsetted for each developmental time point before downstream cell-cell communication analysis. For each subset, the translate function was used to humanize the data and normalized. Differentially expressed genes were computed for each cell type and used as input for ligand-receptor pair calculations.

Virtual growth plate reconstruction

We first subsetted the atlas for growth plate chondrocytes: Resting, Proliferative, Pre-hypertrophic and Hypertrophic chondrocytes. This subset within the latent space of the scANVI integration was then passed to Monocle 3[64]. 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. Roots were manually chosen with order_cells. The resulting pseudotime was added to the metadata of the growth plate subset within the latent space of scANVI and used as input for the growth plate reconstruction. Cells were binned by pseudotime using the cut function from- pandas[65,66]. Upon each bin, we performed dimensionality reduction by t-distributed stochastic neighborhood embedding (t-SNE). We imposed a circle as a boundary condition on the gradient descent function of each t-SNE. With the concatenation function, we were able to stack the circular projections in an ordered way to recreate the cylindrical shape of the growth plate.

Knockout simulation

For in silico knockout experiments in CellOracle[48] the atlas was subsetted to only include time points E10.5-P21 and relevant cell types (Periosteal progenitors, Periosteum, Endosteum, Resting zone chondrocytes, Proliferative chondrocytes, Pre-hypertrophic chondrocytes, Hypertrophic chondrocytes, Chondrocytes, Chondroprogenitors, Osteoprogenitors, Pre-osteoblasts, Osteoblasts, Osteocytes, Vascular endothelial cells, Lymphatic endothelial cells, Proximal limb bud mesenchyme, Intermediate limb bud mesenchyme, Distal limb bud mesenchyme, Myogenic stem cells, Muscle progenitors and Skeletal muscle cell). This subset within the latent space of the scANVI integration of the atlas was then passed to Monocle 3[64]. 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. Roots were manually chosen with order cells. The resulting pseudotime was added to the metadata of the knockout subset within the latent space of scANVI and used as input for the virtual knockout. To reduce the amount of computational time and resources required by a large dataset, 20,000 cells were randomly selected and only highly variable genes ($n = 5000$) were included. For gene regulatory network (GRN) inference, we used the built-in base GRN made from the mouse sci-ATAC-seq atlas[67] CellOracle offers several pre-built base GRN options and provides pipelines to create a custom base GRN using your own scATAC-seq data. For mouse analyses, they recommend using the GRN built from the mouse sciATAC-seq Atlas dataset, which includes a variety of tissues and cell types. Following k nearest neighbors (KNN) imputation based on the first 27 PCs, GRNs were imputed for each cluster. To simulate the knockout of a transcription factor, its expression was set to 0. After this knockout, 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.

In vivo Sox9 knockout and WT data analysis and integration

Both wild-type (WT) and Sox9 knockout (KO) dataset of Haseeb et al.[14]were preprocessed separately to ensure high-quality data for subsequent analyses. The preprocessing steps included quality control measures to remove low-quality cells, as described in "Data preprocessing". Following preprocessing, cells from both WT and Sox9 KO datasets were clustered and labeled independently. Clustering was conducted to identify distinct cell populations within each condition, and labels were assigned based on known cell markers, as described in "Clustering and dimensionality reduction".

To integrate the WT and Sox9 KO datasets, the Harmony package[60] was employed. Harmony facilitates the integration of multiple datasets by correcting batch effects and aligning data across different conditions. This integration enabled a comprehensive comparison of cell populations and gene expression profiles between WT and Sox9 KO conditions. Harmony is particularly suited for integrating datasets from the same study with well-defined batch and biological structures, making it ideal for this analysis[24].

As an additional validation step, we employed reference mapping using scANVI to align the in vivo data with the reference latent space of the Atlas (Sup. Figure 9). We followed the *'Reference mapping with SCANVI'* tutorial, utilizing the scANVI model trained for integration to predict cell type labels[25,30].

Data availability

The datasets analyzed for this study can be found in Sup. Table 1. The Limb 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/TElabSBE/LimbSkeletalCellAtlas.

References

1. Šromová, V., Sobola, D. & Kaspar, P. A Brief Review of Bone Cell Function and Importance. *Cells* **12**(21), 2576. https://doi.org/10.3390/cells12212576 (2023).
2. Einhorn, T. A. The cell and molecular biology of fracture healing. *Clin. Orthop. Relat. Res.* **355**, S7–21. https://doi.org/10.1097/00003086-199810001-00003 (1998).
3. Florencio-Silva, R., Sasso, G. R. D. S., Sasso-Cerri, E., Simões, M. J. & Cerri, P. S. Biology of Bone Tissue: Structure, Function, and Factors That Influence Bone Cells. *BioMed Res. Int.* 421746. https://doi.org/10.1155/2015/421746 (2015).
4. Kwon, M., Kim, B. S., Yoon, S., Oh, S.-O. & Lee, D. Hematopoietic Stem Cells and Their Niche in Bone Marrow. *Int. J. Mol. Sci.* **25**(13), 6837. https://doi.org/10.3390/ijms25136837 (2024).
5. Asp, M. et al. A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **179**, 1647–1660. https://doi.org/10.1016/j.cell.2019.11.025 (2019).
6. Elmentaite, R., Kumasaka, N., Roberts, K. et al. Cells of the human intestinal tract mapped across space and time. *Nature* **597**, 250–255 https://doi.org/10.1038/s41586-021-03852-1 (2021).
7. Zhang, B., He, P., Lawrence, J.E.G. et al. A human embryonic limb cell atlas resolved in space and time. *Nature* **635**, 668–678. https://doi.org/10.1038/s41586-023-06806-x (2024).
8. Markman, S. et al. A single-cell census of mouse limb development identifies complex Spatiotemporal dynamics of skeleton formation. *Dev Cell* **58**, 565–581. https://doi.org/10.1016/j.devcel.2023.02.013 (2023).
9. Hrovatin, K. et al. Considerations for Building and using integrated single-cell atlases. *Nat. Methods*. **22**, 41–57. https://doi.org/10.1038/s41592-024-02532-y (2025).
10. Pasquini, G., Rojo Arias, J. E., Schäfer, P. & Busskamp, V. Automated methods for cell type annotation on scRNA-seq data. *Comput. Struct. Biotechnol. J.* **19**, 961–969 (2021). https://doi.org/10.1016/j.csbj.2021.01.015
11. Han, X. et al. Mapping the mouse cell atlas by Microwell-Seq. *Cell* **172**, 1091–1107. https://doi.org/10.1016/j.cell.2018.02.001 (2018).
12. Regev, A. et al. The human cell atlas. *Elife* **6**, 27041. https://doi.org/10.7554/eLife.27041 (2017).
13. Schaum, N. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula muris. *Nature* **562**, 367–372. https://doi.org/10.1038/s41586-018-0590-4 (2018).
14. Haseeb, A. et al. SOX9 keeps growth plates and articular cartilage healthy by inhibiting chondrocyte dedifferentiation/ osteoblastic redifferentiation. *Proc Natl. Acad. Sci. U S A* **118**(8). https://doi.org/10.1073/pnas.2019152118 (2021).
15. Sivaraj, K. K. et al. Regional specialization and fate specification of bone stromal cells in skeletal development. *Cell Rep.* **36**(2), 109352. https://doi.org/10.1016/j.celrep.2021.109352 (2021).
16. Agoro, R. et al. Single cell cortical bone transcriptomics define novel osteolineage gene sets altered in chronic kidney disease. *Front Endocrinol. (Lausanne)* **14**, 1063083. https://doi.org/10.3389/fendo.2023.1063083 (2023).
17. Nookaew, I. et al. Refining the identity of mesenchymal cell types associated with murine periosteal and endosteal bone. *J. Biol. Chem.* **300**(4), 107158. https://doi.org/10.1016/j.jbc.2024.107158 (2024).
18. Böhm, A. M. et al. Activation of skeletal stem and progenitor cells for bone regeneration is driven by PDGFRβ signaling. *Dev Cell* **51**(2), 236–254. https://doi.org/10.1016/j.devcel.2019.08.013 (2019).
19. Knights, A. J. et al. Synovial fibroblasts assume distinct functional identities and secrete R-spondin 2 in osteoarthritis. *Ann. Rheum. Dis.* **82**(2), 272–282. https://doi.org/10.1136/ard-2022-222773 (2022).
20. Desanlis, I., Paul, R. & Kmita, M. Transcriptional trajectories in mouse limb buds reveal the transition from anterior-posterior to proximal-distal patterning at early limb bud stage. *J. Dev. Biol.* **8**(4), 31. https://doi.org/10.3390/jdb8040031 (2020).
21. Kelly, N. H., Huynh, N. P. T. & Guilak, F. Single cell RNA-sequencing reveals cellular heterogeneity and trajectories of lineage specification during murine embryonic limb development. *Matrix Biol.* **89**, 1–10. https://doi.org/10.1016/j.matbio.2019.12.004 (2020).
22. Debnath, S. et al. Discovery of a periosteal stem cell mediating intramembranous bone formation. *Nature* **562**(7725), 133–139. https://doi.org/10.1038/s41586-018-0554-8 (2018).
23. Baryawno, N. et al. A cellular taxonomy of the bone marrow stroma in homeostasis and leukemia. *Cell* **177**(7), 1915–1932. https://doi.org/10.1016/j.cell.2019.04.040 (2019).
24. Luecken, M. D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**(1), 41–50. https://doi.org/10.1038/s41592-021-01336-8 (2022).
25. Gayoso, A. et al. A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* **40**, 163–166. https://doi.org/10.1038/s41587-021-01206-w (2022).
26. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**(12), 1053–1058. https://doi.org/10.1038/s41592-018-0229-2 (2018).
27. The Tabula Sapiens Consortium* ,The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**(6594), eabl4896. https://doi.org/10.1126/science.abl4896 (2022).
28. Kingma, D. P., Rezende, D. J., Mohamed, S. & Welling, M. Semi-supervised learning with deep generative models. *Adv. Neural Inform. Process. Syst.* **4**. https://doi.org/10.48550/arXiv.1406.5298 (2014).
29. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*. https://doi.org/10.48550/arXiv.1312.6114 (2014).
30. Xu, C. et al. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol* **17**, e9620. https://doi.org/10.15252/msb.20209620 (2021).
31. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using scanorama. *Nat. Biotechnol.* **37**(6), 685–691. https://doi.org/10.1038/s41587-019-0113-3 (2019).
32. Lotfollahi, M., Wolf, F. A. & Theis, F. J. scGen predicts single-cell perturbation responses. *Nat. Methods* **16**(8), 715–721. https://doi.org/10.1038/s41592-019-0494-8 (2019).
33. Zakany, J. & Duboule, D. The role of hox genes during vertebrate limb development. *Curr. Opin. Genet. Dev.* **17**(4), 359–366. https://doi.org/10.1016/j.gde.2007.05.011 (2007).
34. Panman, L. & Zeller, R. Patterning the limb before and after SHH signalling. *J. Anat.* **202**(1), 3–12. https://doi.org/10.1046/j.1469-7580.2003.00138.x (2003).
35. Lewandoski, M., Sun, X. & Martin, G. R. Fgf8 signalling from the AER is essential for normal limb development. *Nat Genet* **26**(4), 460–463. https://doi.org/10.1038/82609 (2000).
36. Bell, S. M. et al. R-spondin 2 is required for normal laryngeal-tracheal, lung and limb morphogenesis. *Development* **135**(6), 1049–58. https://doi.org/10.1242/dev.013359 (2008).
37. Lai, K. M. V. et al. Diverse phenotypes and specific transcription patterns in Twenty mouse lines with ablated LincRNAs. *PLoS One* **10**(4), e0125522. https://doi.org/10.1371/journal.pone.0125522 (2015).
38. Ito, H. et al. Molecular cloning and biological activity of a novel Lysyl Oxidase-related gene expressed in cartilage. *J. Biol. Chem.* **276**(26), 24023–24029. https://doi.org/10.1074/jbc.M100861200 (2001).
39. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88. https://doi.org/10.1038/s41576-020-00292-x (2021).
40. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**(4), 1484–1506. https://doi.org/10.1038/s41596-020-0292-x (2020).

41. Kawakami, Y. et al. BMP signaling during bone pattern determination in the developing limb. *Development* **122**(11), 3557–3566. https://doi.org/10.1242/dev.122.11.3557 (1996).
42. Zou, H., Wieser, R., Massagué, J. & Niswander, L. Distinct roles of type I bone morphogenetic protein receptors in the formation and differentiation of cartilage. *Genes Dev.* **11**(17), 2191–203. https://doi.org/10.1101/gad.11.17.2191 (1997).
43. Pizette, S. & Niswander, L. BMPs negatively regulate structure and function of the limb apical ectodermal ridge. *Development* **126**(5), 883–894. https://doi.org/10.1242/dev.126.5.883 (1999).
44. Carballo, G. B., Honorato, J. R., De Lopes, G. P. F. & Spohr, T. C. L. D. S. E. A highlight on Sonic Hedgehog pathway. *Cell. Commun. Signal.* **16**(1), 11. https://doi.org/10.1186/s12964-018-0220-7 (2018).
45. Tickle, C. & Towers, M. Sonic Hedgehog signaling in limb development. *Front. Cell. Dev. Biol.* **5**, 14. https://doi.org/10.3389/fcell.2017.00014 (2017).
46. Kronenberg, H. M. Developmental regulation of the growth plate. *Nature* **423**(6937), 332–336. https://doi.org/10.1038/nature01657 (2003).
47. Levine, M. & Davidson, E. H. Gene regulatory networks for development. *Proc. Natl. Acad. Sci. U.S.A.* **102**(14), 4936–4942. https://doi.org/10.1073/pnas.0408031102 (2005)
48. Kamimoto, K. et al. Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742–751. https://doi.org/10.1038/s41586-022-05688-9 (2023).
49. Dy, P. et al. Sox9 directs hypertrophic maturation and blocks osteoblast differentiation of growth plate chondrocytes. *Dev. Cell* **22**(3), 597–609. https://doi.org/10.1016/j.devcel.2011.12.024 (2012).
50. 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**(21), 2813–2828. https://doi.org/10.1101/gad.1017802 (2002).
51. Li, J. et al. Systematic reconstruction of molecular cascades regulating GP development using Single-Cell RNA-Seq. *Cell Rep.* **15**(7), 1467–1480. https://doi.org/10.1016/j.celrep.2016.04.043 (2016).
52. Bi, W., Deng, J. M., Zhang, Z. & Behringer, R. R. & De crombrugghe, B. Sox9 is required for cartilage formation. *Nat. Genet.* **22**(1), 85–89. https://doi.org/10.1038/8792 (1999).
53. Zhang, B. et al. A human embryonic limb cell atlas resolved in space and time. *Nature* **635**(8039), 668–678. https://doi.org/10.1038/s41586-023-06806-x (2024).
54. To, K. et al. A multi-omic atlas of human embryonic skeletal development. *Nature* **635**(8039), 657–667. https://doi.org/10.1038/s41586-024-08189-z (2024).
55. Bandyopadhyay, S. et al. Mapping the cellular biogeography of human bone marrow niches using Single-Cell transcriptomics and proteomic imaging. *Cell* **187**(12), 3120–3140. https://doi.org/10.1016/j.cell.2024.04.013 (2024).
56. Nourisa, J., Passemiers, A. & Tomforde, S. Critical issues found in dissecting cell identity via network inference and in silico gene perturbation. *BioRxiv.* https://doi.org/10.1101/2024.10.16.618746 (2024).
57. 10x. Genomics Cell Ranger v8.
58. McCarthy, D. J., Campbell, K. R., Lun, A. T. L., Wills, Q. F. & Scater Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**(8), 1179–1186. https://doi.org/10.1093/bioinformatics/btw777 (2017).
59. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**(13), 3573–3587. https://doi.org/10.1016/j.cell.2021.04.048 (2021).
60. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat. Methods* **16**(12), 1289–1296. https://doi.org/10.1038/s41592-019-0619-0 (2019).
61. Hubert, L. & Arabie, P. Comparing partitions. *J. Classif* **2**, 193–218. https://doi.org/10.1007/BF01908075 (1985).
62. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**(85), 2825–2830. https://doi.org/10.48550/arXiv.1201.0490 (2011).
63. Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A. & Theis, F. J. A test metric for assessing single-cell RNA-seq batch correction. *Nat. Methods* **16**(1), 43–49. https://doi.org/10.1038/s41592-018-0254-1 (2019).
64. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**(7745), 496–502. https://doi.org/10.1038/s41586-019-0969-x (2019).
65. McKinney, W. Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference*, 56–61. https://doi.org/10.25080/majora-92bf1922-00a (2010).
66. The pandas development team. pandas-dev/pandas: Pandas. Preprint at. (2020).
67. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**(5), 1309–1324. https://doi.org/10.1016/j.cell.2018.06.052 (2018).

Acknowledgements

Author contributions

TH, JB, FL, PT and LG conceptualized the study; TH, PT and LG designed the study; TH, LO, TV and GN provided (adapted) code for the analysis of the datasets and downstream analysis of the LSCA; LO and TV contributed to the analysis of the datasets and the LSCA; GN provided a docker work environment; MB adapted and optimized the code to create an interactive web application, TH and LO wrote the first draft of the manuscript and all the other authors contributed to the article. All authors approved the submitted version.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-05277-6.

Correspondence and requests for materials should be addressed to L.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.