A human embryonic limb cell atlas resolved in space and time

Authors:
Bao Zhang[1#], Peng He[2,3#], John E Lawrence[3,4#], Shuaiyu Wang[1,5#], Elizabeth Tuck[3], Brian Williams[6], Kenny Roberts[3], Vitalii Kleshchevnikov[3], Lira Mamanova[3], Liam Bolt[3], Krzysztof Polanski[3], Rasa Elmentaite[3], Eirini S Fasouli[3], Martin Prete[3], Xiaoling He[7], Nadav Yayon[2,3], Yixi Fu[1], Hao Yang[1], Chen Liang[1], Hui Zhang[8], David R. FitzPatrick[9], Helen Firth[3], Andrew Dean[10], Roger A Barker[7], Mekayla A Storer[11], Barbara J Wold[6], Hongbo Zhang[1,12,13], Sarah A Teichmann[3]

Affiliations:
[1]The Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

[2]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, UK

[3]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton CB10 1SA, UK

[4]Department of Trauma and Orthopaedics, Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Box 37, Hills Road, Cambridge, CB2 0QQ, UK

[5]Department of Obstetrics, Guangzhou Institute of Pediatrics, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China

[6]Division of Biology, California Institute of Technology, Pasadena, CA, 91125, USA

[7]John van Geest Centre for Brain Repair, Department of Clinical Neurosciences and Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK

[8]Institute of Human Virology, Key Laboratory of Tropical Disease Control of Ministry of Education, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

[9]MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, WGH, Edinburgh,EH4 2SP, UK

[10]Department of Clinical Neurosciences, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

[11]Wellcome-MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, UK.

[12]Advanced Medical Technology Center, the First Affiliated Hospital, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

[13]The Department of Histology and Embryology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

[#] These authors contributed equally: B. Zhang, P. He, J. E. Lawrence, S. Wang.

* email: zhanghongbo@mail.sysu.edu.cn; st9@sanger.ac.uk

Abstract

Human limbs emerge during the fourth post-conception week as mesenchymal buds which develop into fully-formed limbs over the subsequent months. Limb development is orchestrated by numerous temporally and spatially restricted gene expression programmes, making congenital alterations in phenotype common. Decades of work with model organisms has outlined the fundamental processes underlying vertebrate limb development, but an in-depth characterisation of this process in humans has yet to be performed. Here we detail the development of the human embryonic limb across space and time, using both single-cell and spatial transcriptomics. We demonstrate extensive diversification of cells, progressing from a restricted number of multipotent progenitors to myriad mature cell states, and identify several novel cell populations, including perineural fibroblasts and multiple distinct mesenchymal states. We uncover two waves of human muscle development, each characterised by different cell states regulated by separate gene expression programmes. We identify musculin (MSC) as a key transcriptional repressor maintaining muscle stem cell identity and validate this by performing MSC knock down in human embryonic myoblasts, which results in significant upregulation of late myogenic genes. Spatially mapping the cell types of the limb across a range of gestational ages demonstrates a clear anatomical segregation between genes linked to brachydactyly and polysyndactyly, and uncovers two transcriptionally and spatially distinct populations of the progress zone, which we term “outer” and “transitional” layers. The latter exhibits a transcriptomic profile similar to that of the chondrocyte lineage, but lacking the key chondrogenic transcription factors SOX5,6 & 9. Finally, we perform scRNA-seq on murine embryonic limbs to facilitate cross-species developmental comparison at single-cell resolution, finding substantial homology between the two species.

Introduction

Human limb buds emerge by the end of the 4th post conceptional week (PCW4) and develop to form arms and legs during the first trimester. By studying model organisms such as the mouse and chick, it is known that development of the limb bud begins in the form of two major components. The multipotent parietal lateral plate mesodermal (LPM) cells condense into the skeletal system and skeletal muscle progenitor (SkMP) cells migrate from the paraxial mesoderm to the limb field, forming muscle[1,2]. These multipotent progenitors are encapsulated within a layer of ectoderm, a subset of which (termed the apical ectodermal ridge/AER) governs mesenchymal cell proliferation and aids in the establishment of the limb axes through fibroblast growth factor (FGF) signalling[3]. As the limb bud elongates, these AER signals no longer diffuse to the most proximal mesenchymal cells and differentiation begins[4]. Throughout the remainder of the first trimester in humans, the limbs continue to mature in a proximal-distal manner, such that by PCW8 the anatomies of the stylopod, zeugopod and autopod are firmly established. This maturation is tightly controlled by a complex system of temporally and spatially restricted gene expression programmes[5–7]. As with any complex system, small perturbations in even a single programme can result in profound changes to the structure and function of the limb[8]. Indeed, approximately 1 in 500 humans are born with congenital limb malformations[9,10].

Although model organisms have provided key insights into cell fates and morphogenesis that are translatable to human development and disease, at present it remains unclear how precisely these models recapitulate human development. Furthermore, the lack of complementary spatial information in such studies precludes the assembly of a comprehensive tissue catalogue that provides a global view of human limb development in space and time. Encouragingly, the Human Developmental Cell Atlas community has recently applied cell-atlasing technologies such as single cell and spatial transcriptomics to several tissues to give novel insights into development and disease[11–15]. The application of these techniques to human embryonic and fetal tissue therefore holds much promise in furthering our understanding of the developing human limb[16,17].

In this study, we performed droplet-based single-cell transcriptomic sequencing (scRNA-seq) and spatial transcriptomic sequencing to reconstruct an integrated landscape of the human hindlimb during first trimester development. We then performed scRNA-seq on murine embryonic limbs in order to compare the process of limb development across species at this level of resolution. Our results detail the development of the human limb in space and time at high resolution and genomic breadth, identifying fifty-five cell types from 114,000 captured single cells, and spatially mapping these across four timepoints to shed new light on the dynamic process of limb maturation. In addition, our spatial transcriptomics data gives

insights into the key patterning and morphogenic pathways in the nascent and maturing limb, with a focus on genes associated with limb malformation.

Finally, our integrated analysis of human and murine limb development across corresponding time periods reveals extensive homology between a classical model organism and the human, underlining the importance and utility of such models in understanding human disease and development. Our study provides a unique resource for the developmental biology community, and can be freely accessed at https://limb-dev.cellgeni.sanger.ac.uk/.

Results

Cellular heterogeneity of the developing limb in space and time

To track the contribution of the different lineages in the developing limb, we collected single-cell embryonic limb profiles from PCW5 to PCW9 (Fig. 1a). This time window covers the early limb bud-forming stages as well as later stages of limb maturation (Fig. 1a). In total, we analysed 114,047 single-cells that passed quality control filters (Extended Data Fig. 1a). After cell cycle expression module removal by regression, and batch correction (see Methods), we identified 55 cell types and states (Fig. 1b; see Methods; Extended Data Fig. 1b and Extended Data Table 1 for marker genes).

32 of these cell states represent cells derived from the LPM. They contain mesenchymal, chondrocyte, osteoblast, fibroblast and smooth muscle cell states involved in the maturation of cartilage, bone and other connective tissues, consistent with previous investigations of the cellular makeup of the limb[18]. In addition to these LPM-derived cells, a further eight states form a complete lineage of muscle cells that migrate as *PAX*3+ progenitors from the somite. These go on to differentiate in the limb to form myoprogenitors and myotubes.

Other non-LPM cell states include four of the primitive and definitive erythrocytes, two types of myeloid cells, three types of vascular endothelial cells and three types of neural crest-derived cells. Finally, we identified three epithelial cell states, among which were the AER cells at the distal rim of the limb bud that express *SP8* and *WNT6* (Extended Data Fig. 2a,b). Examining the relative abundance of each of these cell states across different gestational ages revealed how the cellular landscape of the developing limb changes over time. Within each of the aforementioned lineages, a clear pattern emerged whereby progenitor states were chiefly isolated from PCW5 & 6, with more differentiated cell states emerging thereafter (Extended Data Fig. 2c,d).

To further dissect the cellular heterogeneity with spatial context, and to build on limb patterning principles established in model organisms, we performed spatial transcriptomic experiments for limb samples from PCW5 and PCW8. Using the 10X Genomics Visium chips, we were able to generate transcriptome profiles capturing on average between 1,000 and 5,000 genes per voxel (Extended Data Fig. 1c). We then applied the cell2location package[19] to transfer cell state labels from our single-cell atlas to deconvolute Visium voxels (See Methods). The resulting cell composition map of Visium slides at each time point demarcated the tissue section into distinct histological regions (Fig. 1c, d). In the PCW5.6

samples, interestingly, a clear zonal segregation of progenitor cell types was observed, dividing the progress zone into two layers that we name the “outer” and “transitional” progress zones. The outer progress zone cells (PZC) are located at the distal periphery of the limb bud. Encased in it are the transitional PZC together with *SOX*9-expressing chondroblasts of the developing autopod (Fig. 1c). This novel spatial distinction was accompanied by subtle transcriptomic differences, with the outer progress zone specifically expressing a number of genes implicated in digit patterning, including *LHX2* and *TFAP2B*. Mutations in the latter cause Char syndrome, a feature of which is postaxial polydactyly[20]. The transitional progress zone specifically expresses *IRX1,* a key gene in digit formation that establishes the boundary between chondrogenic and non-chondrogenic tissue[21,22]. By performing differential expression testing between outer and transitional PZC, we were able to characterise gene modules that define each population (Extended Data Fig. 3a). We then calculated the expression score for each of these modules among the other major lineages, revealing that the transitional PZC module was upregulated in the chondrocyte lineage, with no other clear upregulation of either module in any other lineage (Extended Data Fig. 3a, b). This suggests that these cells may be a transitional state between undifferentiated progress zone cells and committed chondroblasts. Indeed, the genes that define the transitional progress zone largely relate to skeletal system development (Extended Data Fig. 3c).

In addition to the PZC, prehypertrophic chondrocytes (PHC) expressing India Hedgehog (*IHH*) localised to the mid-diaphysis of the forming tibia and the metatarsals. At the proximal limit of the sample, both *MEIS2*-expressing proximal mesenchymal cells (PrMes) and *CITED-1*$^{+}$ mesenchymal cells (Mes2) were observed, in keeping with the early stages of limb development (Fig. 1c).

For analysis of the PCW 8.1 sample, we developed an inline pipeline to align and merge multiple visium sections (see methods). This allowed us to analyse the entire lower limb as one structure (Fig. 1d; Extended Data Fig. 4). In the osteochondral lineage, articular chondrocytes were located at the articular surfaces of the developing knee, ankle, metatarso-phalangeal and interphalangeal joints, while osteoblasts closely matched to the mid-diaphyseal bone collar of the tibia and femur. The perichondrial cells from which they differentiate matched to comparable region, though they extended along the full length tibia and femur (Fig. 1d); a finding confirmed by immunofluorescence staining for *RUNX2* and *THBS2* (Extended Data Fig. 5a). Prehypertrophic chondrocytes matched again to the mid-diaphysis of the tibia and analysis of gene expression revealed some collagen-X expression in this region, in keeping with chondrocyte progression to hypertrophy (Fig.1d). Additionally, we were able to capture glial cells expressing myelin genes (Extended Data Fig. 1c), and an accompanying *FOXS*1-expressing fibroblast subtype (named “perineural fibroblast’ by us here) enriched in the periphery of the sciatic nerve in the posterior compartment of the thigh and its tibial division in the deep posterior compartment of the leg (Fig. 1d, Extended Data Fig. 5b-d). (We were not able to capture neurons in our single-cell data, most likely due to the distant location of their cell bodies within the spinal ganglia and anterior horn of the spinal cord.)

Interestingly, cell states with related (but not identical) transcriptomic profiles did not necessarily occupy the same location, which we are able to quantify based on our cell2location deconvolution analysis of Visium and scRNAseq data. This reveals that within the fibroblast lineage, three clusters were co-located with *KRT*15-expressing basal cells and

SFN-expressing cells of the periderm[23], suggesting a role in the dermal lineage and prompting their annotation as dermal fibroblasts (DermFiB) and their precursors (F10$^+$DermFiBP & HOXC5$^+$DermFiBP)(Extended Data Fig. 5e). A further fibroblast cluster expressing *ADH1B* (ADH$^+$FiB) co-localised with muscle cells, with no equivalent population found in the dermal region (Extended Data Fig. 5f-h).

Similarly, we were able to spatially resolve two clusters with subtle transcriptomic differences within the tenocyte lineage. Both clusters expressed the classical tendon markers scleraxis (*SCX*) and tenomodulin (*TNMD*), with one population of cells expressing increased biglycan (BGN) and Keratocan (*KERA*); molecules which play a role in the organisation of the extracellular matrix, while the other population expressed higher levels of pro-glucagon (*GCG*) that is important in metabolism. Analysis with cell2location matched the former cluster to the long flexor tendons of the foot, as well as the hamstrings, quadriceps & patellar tendons around the knee joint. The latter cluster, however, matched to the perimysium, the sheath of connective tissue that surrounds a bundle of muscle fibres (Fig. 1d; Extended Data Fig. 5i,j). We therefore annotated these clusters as tenocyte (Teno) and perimysium, respectively.

Overall, these findings provide new insights into the subtle transcriptomic differences within cell compartments including the muscle, tendon, bone and stromal lineages. This integrated analysis serves as an example of how spatial transcriptomic methodologies can improve our understanding of tissue architecture and locate cell states themselves within the context of developmental dynamics of an anatomical structure such as the whole limb.

Patterning, morphogenesis and developmental disorders in the limb

During organogenesis of the limb, individual cell identities are in part, determined by their relative position within the limb bud. This developmental patterning is controlled by a complex system of temporally and spatially restricted gene expression programmes. For example, key aspects of proximal-distal patterning are controlled by the AER[24,25]. In contrast anterior-posterior axis specification is chiefly controlled by the zone of polarising activity (ZPA) through *SHH* signalling[26–28]. Within the autopod, precise regulation of digit formation is mediated through interdigital tissue apoptosis[29–31]. We utilised Visium spatial transcriptomic data to explore the locations of transcripts of all these classic pattern-forming genes on the same tissue section, finding notable consistency with previous *in situ* hybridisation experiments in the mouse (Extended Data Fig. 6a-e). This included several key genes known to govern the proximal identity, including *MEIS*1 & 2, *PBX*1 and *IRX*3[32–35], as well as genes regulating limb outgrowth and distal morphogenesis such as *WNT5A*, *GREM*1, *ETV*4 and *SALL*1[36–39]. Similarly, classical mammalian anterior-posterior (AP) genes were captured, including *HAND*1, *PAX*9, *ALX*4 and *ZIC*3 (anterior) and *HAND*2, *SHH* and *PTCH*1 and *GLI*1 (posterior)[26,40–46].

The homeobox (*HOX*) genes are a group of 39 genes split into four groups termed “clusters”, each of which is located on a separate chromosome. During limb development, genes in the A & D clusters act in concert with the aforementioned axis-determining genes to dictate limb patterning in mammals[47]. In mice, these genes are expressed in two waves. The first wave occurs in the nascent limb bud, with the expression of 5’ groups in both clusters exhibiting a

posterior prevalence. During the second wave of expression, this posterior prevalence is lost in the A cluster, persisting only in the D cluster[48,49]. Our spatial transcriptomic data captured the expression patterns of the A and D clusters early in the 6th post-conception week (Extended Data Fig. 6f). As expected, their expression matches the second wave of expression in mice, with a loss of posterior prevalence in the *HOXA* cluster and its maintenance in the *HOXD* cluster. For both clusters, an increase in group number corresponded to more distally restricted expression, with group 13 genes limited to the most distal part of the limb bud. An exception to this was *HOXA*11 expression, which showed no overlap with *HOXA*13, in keeping with the expression pattern of these two genes during the second *HOX* wave in mice. This mutual exclusivity in expression domain is thought to be due to *HOXA*13/*D*13-dependent activation of an enhancer that drives antisense transcription of *HOXA*11 in pentadactyl limbs[50]. Indeed, our data revealed a clear switch to the antisense transcript in the distal limb bud, although this switch did occur more proximally than the limit of *HOXA*13/*D*13 expression (Extended Data Fig. 6f). This suggests that in humans, additional mechanisms may be involved in the switch to antisense transcription.

In order to investigate gene expression patterns during digit formation, we obtained coronal sections through a PCW6.2 foot plate to reveal the five forming digits together with the intervening interdigital space (IDS; Fig. 2a). We then manually annotated digital and interdigital voxels based on H&E histology. Differential expression testing between digital and interdigital regions across two adjacent sections demonstrated an enrichment of classical survival-promoting genes in the digital regions, such as *IRX*1 & 2, while genes involved in interdigital cell death, such as *MSX* 1 & 2, *LHX*2 and *BMP*7 were upregulated in the IDS (Fig. 2b). Similarly, interdigital regions showed an enrichment of molecules involved in the retinoic acid (RA) pathway, such as retinol binding protein (*RBP*) 4 and Signalling Receptor And Transporter Of Retinol (*STRA*) 6. Conversely, the RA metabolising enzyme *CYP26B1* was profoundly upregulated in the digital regions. These findings underline the importance of RA in triggering interdigital cell death in the hand and foot plate[51].

In addition to these known players in digit formation, we also identified several genes which likely play novel roles in this process. For instance, Wnt inhibitory factor (*WIF*) 1 and its downstream target *DKK*1 were significantly upregulated in the IDS. These proteins act in a coordinated manner to trigger apoptosis through activation of p21 and p53 and inhibiting c-*MYC* and *BCL*2, suggesting a role in the regression of the interdigital tissue of the hand and foot plates[52]. In addition, there was profound upregulation of the monocyte chemoattractant *CCL*2 in the IDS (Fig. 2b), in keeping with programmed cell death in this region. Interestingly, we detected only a weak presence of macrophages in the IDS, with the majority of macrophage signatures mapping to vasculature-associated regions (Extended Data Fig. 6g,h). Similar to an aortic macrophage population uncovered in recent studies[53,54], the majority (56.6%) of these macrophages are *TREM2*$^{+}$, though we cannot exclude the possibility that rather than being a vascular population, these macrophages are in the process of migrating to become tissue-resident elsewhere[55]. Finally, we histologically annotated each digit in the PCW6.2 foot plate to search for genes that vary with digit identity (Extended Data Fig. 6i). We identified four genes that were upregulated in the great toe including *ID*2 and *ZNF*503, both of which are known to have anterior expression domains in the limb, as well as the regulator of cell proliferation *PLK*2 and the cancer-associated gene *LEMD*-1[56–59]. *HOXD*11 was downregulated in the great toe, in keeping with its posterior prevalence. We found no differentially expressed genes in the remaining digits.

We next cross-referenced the list of digit-IDS differentially expressed genes against a list of 2300 single gene health conditions. We found genes involved in several types of isolated (or non-syndromic) brachydactyly (BD) were significantly upregulated in the digital tissue (Fig. 2c). These included *IHH* (Type A1 BD), *BMPR1B* (Type A2) and *NOG* (Type B2)[60]. Several genes where variations produce complex syndromes which include brachydactyly as part of their phenotype were also upregulated in digital tissue. These included *COL11A2* (Oto-Spondylo-Mega-Epiphyseal Dysplasia / OSMED), *SOX9* (Cook's syndrome) and *FGFR3* (Achondroplasia)[61,62]. Conversely, genes which are varied in syndromes with syndactyly as part of their phenotype were significantly upregulated in the IDS. These include *DLX5* (split hand-foot malformation), *MYCN* (Feingold Syndrome type 1), *BMP4* (Microphthalmia VI) and *TWIST1* (Saethre-Chotzen Syndrome)[63–66]. We then searched the EMAP eMouse Atlas Project (http://www.emouseatlas.org) for spatial expression data of these genes in the mouse, and found markedly similar patterns for those with available data[67] (Extended data Fig. 7). Similarly, where murine models of the aforementioned heritable conditions exist, their phenotype is broadly comparable to the human (Extended Data Table 2)[60,62,65,68–83].

Our spatial atlas provides a valuable reference of gene expression under homeostatic conditions for comparison with genetic variations for which phenotypes may begin to penetrate during embryonic development.

Regulation of cell fate decisions of mesenchymal-derived lineages

Our single-cell and spatial atlases revealed a high diversity of mesenchymal-derived cell types and states. In order to better understand what transcriptional mechanisms may control their specification, we inferred cell-fate trajectories in the 32 mesenchyme-associated states by combining diffusion maps, partition-based graph abstraction (PAGA) and force-directed graph (FDG) (see methods). We combined this with transcription factor (TF) network inference using the SCENIC package to identify distinct modules of active TF networks associated with progression through each lineage[84].

As expected, the global embedding resembled a 'spoke-hub' system, whereby multipotent mesenchymal cells are embedded centrally, with committed cell types radiating outward as they begin to express classical cell-type specific marker genes (Fig. 3a, b). The central hub of mesenchymal cells consisted of five clusters with subtle differences in their transcriptome. A first population, here named Mesenchyme 1 (Mes1), expressed *PITX2*, a key initiator of hindlimb bud formation that is known to play a role in left-right identity and global limb patterning[85]. Proximal mesenchymal cells (PrMes) expressing the regulator of proximal identity, *MEIS2*, were also captured. Two further clusters (Mes2, Mes3) of mesenchyme expressed *CITED1*, a molecule which localises to the proximal domain of the murine limb bud and plays an unclear role in limb development[86]. In addition to *CITED1*, the Mes3 cluster also expressed the homeodomain protein *HOPX*. This regulatory molecule is thought to suppress adipogenesis in bone marrow stromal cells, and its role in the nascent limb bud has not to our knowledge been characterised to date[87]. The Mes4 cluster exhibited similar overall expression patterns to the other mesenchymal cells, with the addition of low levels of *PRAC1*, a molecule identified as maintaining a prostate gland stem cell niche but with no

known role in limb development[88]. In addition to these populations, we also identified a subpopulation of mesenchymal cells expressing *ISL*1 (ISL1^{+}Mes). These cells represent a mesenchymal niche within the nascent hindlimb bud which contributes to the posterior elements of the limb[89]. Finally, a cluster bridging mesenchymal cells and those in the chondrocyte lineage, but lacking expression of classical marker genes of either, was identified. This transitional state possibly represents cells forming mesenchymal condensations (MesCond) at the core of the limb, prior to commitment to the chondral lineage.

Examining the abundance of cell types by gestational age revealed how cellular heterogeneity within the mesenchymal compartment evolves during limb development (Fig. 3c). During PCW5, the majority of the cells captured were mesenchymal progenitors. This was particularly notable at PCW5.1 and 5.4, where mesenchyme accounted for 85% and 65% of all cells respectively. The relative abundance of mesenchymal cells in the limb declined thereafter, with almost none present at PCW8 & 9. A multitude of TF networks were predicted to be active in these progenitor populations (Fig. 3d; Extended Data Table 3). *ALX*4 and *ISL*1, genes known to dictate limb bud establishment, were both expressed across mesenchymal progenitor populations[42,89,90]. Similarly, *MEOX*2 and *ALX*1 were present in several populations, reflecting their role in limb patterning and morphogenesis[91,92]. *ZMAT*4 and *ZFHX*3 showed similarly diffuse activity, and the role of these TFs has not been clearly characterised in limb development. *ZFHX*3 has been implicated in demarcating the developing perichondrium and periosteum in the chicken, though interestingly showed no activity in these cell types in our data[93]. *ZMAT*4 has previously been shown to be upregulated in bone marrow mesenchymal cells when compared to limbal epithelial cells and mesenchymal cells, though no specific role in limb development has been identified[94]. The *SHH* regulator *GATA*6, another key patterning TF, was restricted to Mes1 and PrMes[95]. *SHOX*2 was active in the proximal mesenchyme, in keeping with its function of tissue specification in the proximal limb[96]. Finally, cells of the progress zone showed activation of a distinct module of TFs, including *LHX*2 and *LHX*9, as previously described in the mouse[97].

The chondrocyte lineage increased in number steadily over time, accounting for 25% of the cells captured at PCW5.6, increasing to 50% at PCW7.2. Within this lineage, a shift from progenitor to more mature cell types was observed during the period studied, with uncommitted osteochondral progenitors (OCP) and immature chondroblasts giving way to maturing, resting and prehypertrophic chondrocytes (Fig. 3c). This progression was accompanied by changes in regulon activity (Fig. 3d). For example, *SHOX* activity was highly specific for OCPs. This TF has been widely implicated in secondary ossification through its interactions with the master regulators of chondrogenesis - *SOX*5, 6 and 9[98]. Altered *SHOX* expression has been found in syndromes of altered skeletal growth, including Turner syndrome and idiopathic short and tall stature[99]. Its role in primary ossification is not well characterised, and its potential role in driving mesenchymal progenitors to form OCPs could shed further light on the mechanisms underlying such syndromes.

The switch from OCP to chondroblasts (and subsequent chondrocytes) was associated with the activation of *SOX*5,6 and 9, with the latter localising to chondrocyte condensations at PCW 5.6 and the developing tibia, fibula and digits at PCW6.2 (Fig. 3d,e). This trend was observed for other known regulators of chondrogenesis, including *THRB* and *NKX*3-2[100,101]. Interestingly, *FOXJ*1 was predicted to have similar activity in the chondrocyte lineage. In

addition to its established role in ciliation, this TF has been shown to regulate dental enamel development [102]. Furthermore, like *SOX*5 and 9, it is regulated by *IRX*1; a TF which specifies the digits and establishes the boundary between chondrogenic and non-chondrogenic tissue in the developing chick limb[21]. Several known regulators of chondrocyte hypertrophy were specific to PHCs, including Osterix (*SP*7), *RORA*, *DLX*3 and *RUNX*3, with the latter localising to the tibial diaphysis at PCW 6.2[103–106]. Finally, *RUNX*2 was, as expected, predicted to be active in osteoblasts and the perichondrial cells from which they are derived, with expression again localising to the tibial diaphysis.

Our experiments also captured the cells of the interzone; mesenchymal cells that reside at the sites of future synovial joints and give rise to their constituent parts. This cluster expressed the classical interzone marker *GDF*5 and emerged at the end of PCW5, giving rise to articular chondrocytes expressing lubricin (PRG4) at PCW8 (Fig. 3b,c). Intriguingly, the articular chondrocytes were not predicted to exhibit *SOX*5/6/9 activity, but instead possessed a separate TF programme focused on inhibiting classical osteochondral transcripts (Fig 3d). For example, *ELF*3 has been shown to inhibit SOX9 in cultured human chondrocytes, whilst *TFAP2C* has demonstrated the same effect in human colorectal cancer cell lines[107,108]. In murine models, the glucocorticoid receptor NR3C1 inhibits the osteoblastic transcripts *COL1A1* and Osteocalcin (*BGLAP*)[109,110].The role of *SOX9* in articular cartilage development and homeostasis is uncertain, with inducible loss in mice resulting in no degenerative change at the joint postnatally[111,112].

Tendon progenitors expressing high levels of *SCX* but low *TNMD* emerged during PCW5 before declining in number and being replaced by tenocytes and perimysium expressing high levels of *TNMD* from PCW7 onward (Fig. 3b, c). Several TFs that control tenogenesis were predicted to be active in these cell types, including *SIX*1 and *MKX*[113,114]. Additionally, *HEYL* activity was elevated in tenocytes (and to a lesser degree in perimysium) when compared to tendon progenitors (Fig. 3d). The role of this TF in tenogenesis has not been characterised, however it has been shown to suppress the expression of *MYOD*1 in muscle stem cells under increased loading conditions[115]. Our data suggest this mechanism may extend to tenocyte specification in the embryonic limb. *TEF* also showed upregulation in these cell types. This gene is known to play a role in extracellular matrix regulation in cardiac tissue, and may therefore play a similar role in the developing tendon [116] Interestingly, the homeobox transcription factor *HLX* was specific to perimysium. This TF plays a role in the development of many organs including the liver, diaphragm, bowel, spleen and myeloid lineage, but no function in tendon development has been described[117].

Finally, different fibroblast and smooth muscle populations within the limb exhibited clearly distinct TF activities (Extended Data Fig. 8a,b). Dermal fibroblasts showed activity in known regulators of this lineage, including *FOSL2*, *PRDM1, SP5* and *HOXC5*[118–120]. Perineural fibroblasts showed activity in *FOXS1,* a molecule previously associated with sensory nerves [121]. Smooth muscle cells (SmMC) and their precursors (SmMP) both showed activity in *GATA*6, which is thought to regulate their synthetic function [122]. In addition, SmMC showed activity in several additional TFs with known roles in smooth muscle maturation, such as *HEY*2 and *HES6*[123,124].

Regulation of embryonic and fetal myogenesis

Limb muscle originates from the dermomyotome in the somite[1,2]. Classically, its formation begins with delamination and migration from the somite regulated by *PAX3* and co-regulators such as *LBX1* and *MEOX2*, followed by two subsequent waves of myogenesis: embryonic and fetal[125]. During embryonic myogenesis, a portion of *PAX3*$^+$ embryonic skeletal muscle progenitors are destined to differentiate and fuse into multinucleated myotubes. These primary fibers act as the scaffold for the formation of secondary fibers derived from *PAX7*$^+$ fetal skeletal muscle progenitors, which are themselves derived from *PAX3*$^+$ muscle progenitors[126–128].

To dissect these limb muscle developmental trajectories in detail from our human data, we took cells from the eight muscle states, re-embedded them using diffusion mapping combined with PAGA and FDG. Three distinct trajectories with an origin in *PAX3*$^+$ skeletal muscle progenitors (*PAX3*$^+$ SkMP), emerged (Fig. 4a). The first trajectory (labelled 1st Myogenesis) starts from *PAX*3$^+$ SkMP and progresses through an embryonic myoblast state (MyoB1) followed by an early embryonic myocyte state (MyoC1), and finally arrives at mature embryonic myocytes. This trajectory is in keeping with embryonic myogenesis. Along the second trajectory, the *PAX3*$^+$ SkMP lead to *PAX*3$^+$*PAX*7$^+$ cells, followed by a heterogeneous pool of *PAX*7$^+$ SkMP cells that are mostly *MyoD* negative (Fig. 4a,b). This represents a developmental path that generates progenitors for subsequent muscle formation and regeneration. The final trajectory (labelled 2nd Myogenesis) connects cell states that express *PAX*7 first to fetal myoblasts (MyoB2), then early fetal myocytes (MyoC2) and finally mature fetal myocytes.

Comparing these myogenic pathways, we noticed that *PAX*3 expression gradually phases out along the trajectory of the embryonic myogenic pathway, while it is almost absent in the fetal myogenic pathway (Fig. 4b).This is consistent with a previous study that captured Pax3$^+$ Myog$^+$ cells in the mouse limb [129]. Interestingly, ID2 and ID3 that are known to attenuate myogenic regulatory factors [130,131] are also more highly expressed in embryonic myogenesis than fetal, which may imply different upstream regulatory networks. Additional genes such as *FST*, *RGS*4, *NEFM* and SAMD11 were also identified to be marking the first myogenic pathway while KIF19, TNFSF13B, KRT31 and RGR mark the second (Fig. 4b). In fact, Keratin genes have been found to facilitate sarcomere organisation[132].

Next, we performed SCENIC analysis to search for transcription factors driving each myogenic stage. A large number of stage-specific transcription factors were identified (Fig. 4d; Extended Data Table 4), including known muscle regulators such as *MSX*1, *PAX*3, *PAX*7, *PITX*2, *SIX*2, MYOD1, *MYOG* and *LBX1* (Extended Data Fig. 8.c,d). Indeed, *PAX3* had a higher activity score in cell states during embryonic myogenesis while *PAX7* has higher scores during fetal myogenesis. Although *PITX2* was reported to be transcribed at similar levels during embryonic limb myogenesis[133], we observe a higher activity score and abundance during embryonic myogenesis than fetal myogenesis (Extended Data Fig. 8.c), possibly related to its different regulatory roles[134]. Its related family member, *PITX1*, shows overlapping activity hotspots based on analyses using SCENIC. Interestingly, while known as a hindlimb-specific transcription factor, we find *PITX1* expressed in both forelimb and hindlimb muscle cells (Fig. 4e), including a fraction of *PAX3*$^+$ cells as early as PCW5 (Fig. 4f), suggesting a potential regulatory role in embryonic myogenesis.

Complementary to SCENIC analyses focusing on activators, we also investigated in several transcriptional repressors such as *MSC* (also known as Musculin, ABF-1 or MyoR), TCF21(Capsulin), and families of ID, HES and HEY proteins. We observed specific expressions of *MSC*, *HES*1 and *HEY*1 in PAX7$^+$ skeletal progenitors. The most prominent repressor, *MSC* is a bHLH transcription factor that has been shown to inhibit *MyoD*'s ability to activate myogenesis in 10T1/2 fibroblasts[135] and rhabdomyosarcoma cells[136]. In addition, in C2C12 murine myoblasts, MSC facilitates Notch's inhibition of myogenesis (although it appears to exhibit functional redundancy in this role)[137]. To test whether human MSC also plays a role in repressing *PAX7*$^+$ skeletal muscle progenitor maturation, in addition to its widely accepted role in B-cell development[138,139], we knocked down MSC in primary human embryonic limb myoblasts. Our RT-qPCR results showed profound upregulation of late myocyte genes (Fig. 4g). This suggests that *MSC* is key to maintaining limb muscle progenitor identity.

Spatially resolved microenvironments exhibit distinct patterns of cell-cell communication

In order to investigate communication between cell types, we utilised the CellphoneDB python package to identify stage-specific ligand-receptor interactions by cell type in the developing limb[140,141]. This output was then filtered to reveal signalling pathways between co-located populations of cells at the histological level, determined in an unbiased way using cell2location factor analysis (Fig. 5a).

In the early (PCW5.6) limb bud, NOTCH signalling was predicted to occur in its distal posterior aspect through the canonical ligand Jagged (*JAG*)-1 (Fig. 5b). This interaction occurs between adjacent cells, with *JAG*1 bound to the cell rather than being secreted, triggering proteolytic cleavage of the intracellular domain of NOTCH receptors with varying activity depending on the NOTCH receptor involved[142–144]. *JAG*1 is induced by *SHH* in the posterior distal limb bud, with its anterior expression inhibited by *GLI3R*[41,145]. Our spatial transcriptomic dataset confirms this expression pattern in the early limb bud, with several voxels containing both *JAG*1 and NOTCH family transcripts (Fig. 5b; white asterisks). In addition, *HES*1, a downstream target of NOTCH, was expressed in these voxels, supporting the predicted activity of this signalling pathway. This novel finding sheds further light on the mechanisms controlling limb morphogenesis and has implications for conditions where this signalling axis is disrupted, such as the posterior digit absence characteristic of Adams-Oliver syndrome and the 5th finger clinodactyly of Alagille syndrome[146,147].

In the limb bud samples at PCW6.2, we captured weak but reproducible signals of *FGF*8 in the AER epithelial cells while *FGF*10 was detected in the adjacent mesenchyme (Fig. 5c, d). It is known that *FGF*8 and *FGF*10 are expressed in the adjacent ectoderm and mesoderm respectively and form a feedback loop through *FGFR*2 that is essential for limb induction[148]. Indeed, *FGFR*2 expression overlapped with that of *FGF*8 and *FGF*10 (Fig. 5d) and showed higher abundance in the AER than the progress zone, consistent with previous findings[148]. Interestingly, our single-cell and spatial atlases also suggest that *FGFR*2 has the lowest expression in the outer progress zone, interposed by the AER and transitional progress zone (Fig. 5c,d), which indicates a potential repression mechanism in the outer progress zone. As expected, FGFR2 was also expressed in the osteochondral lineage (Fig. 5e). The

importance of this receptor in skeletal development is highlighted by the limb phenotypes observed in the FGFR2-related craniosynostoses, such as radiohumeral synostosis, arachnodactyly and bowed long bones[149]. *FGF*8, together with *FGF*5 and 9, were also expressed in regions with high expression of myogenic proteins, including their receptor FGFR4 (Fig. 5d-f; Extended Data Fig. 9a,b). This is in agreement with its previously reported role in activating MyoD expression in the chicken limb bud[150].

The distribution of other FGF family members in the nascent limb were in keeping with observations in model organisms (Fig. 5e). FGFR1 showed a broad expression pattern across many cell populations, including skeletal muscle, bone and tendon. This receptor plays multiple roles in limb development, and its inactivation in the mouse results in truncated limbs with three digits; a deformity most marked in the hindlimb[151]. FGFR3 was expressed in intermediate and prehypertrophic chondrocytes, reflecting its role as a negative regulator of endochondral ossification[152]. Our spatial transcriptomic data further confirm the cell-specific spatial distribution of FGFs and their corresponding receptors (Fig 5f).

Homology and divergence between human and murine limb development

Limb development has long been studied in model organisms, while assays directly performed on human samples are less common. To explore differences between mice and humans, and to understand evolutionary principles, we collected 13 mouse limb samples for scRNA-seq, and combined our newly generated data with 18 high-quality (Extended Data Fig. 10a) limb datasets from three previously published studies [133,153,154] to build a comprehensive mouse embryonic limb atlas (Extended Data Fig. 10b, c). To compare the mouse and human transcriptome, we used the diagonal alignment algorithm MultiMAP[155] to align single-cells based on matched orthologs, while also considering information from non-orthologous genes. The resulting integrative atlas (Fig. 6a-c) with aligned cell-type clusters show a highly conserved cell composition between Human and Mouse (Fig. 6b, Extended Data Fig. 10d).

But differential compositions are also observed in a small number of cell types. As expected, the mouse limb dataset has a greater abundance of *PAX*3$^+$ SkMP and Early LPMC cells (Fig. 6d, brown circle; Extended Data Fig. 10d) that are enriched in early embryonic limb development, primarily because more early (before E12) mouse samples were collected given the limited access to early embryonic human samples (before PCW5). In addition, mouse limbs contained a higher percentage of epithelial cells and immune cells (Fig 6d, orange circle), possibly due to faster maturation of the epidermal and early immune system in the mouse. Consistent with this, a recent comparison of mouse and chicken limbs at the single cell level found that gene modules of epithelial and immune cells exhibit higher evolutionary turnover[156]. Interestingly, whilst more cells of the cluster PAX3$^+$ SkMP in mouse are PAX3$^+$ MSC- (3448) than PAX3$^+$ MSC$^+$ (592), the reverse is true for human (116 PAX3$^+$ MSC- cells vs 1532 PAX3$^+$ MSC$^+$ cells), indicating differences in the transcription factor repertoire between mouse and human PAX3$^+$ SkMP.

To systematically compare pattern formation between mouse and human limbs, we dissected forelimbs and hindlimbs from a human embryo and a mouse counterpart, each

separated into proximal, middle and distal segments to compare with our first trimester human samples. This allowed us to address the differences between forelimb and hindlimb along the proximo-distal axis at matched time points in human and mouse development.

Overall, mice and humans demonstrate highly similar cell-type compositions along the P-D axis. In both human and mouse forelimbs, proximal mesenchymal cells are enriched towards the proximal end, while progress zone cells (TransPZC and OuterPZC) are highly enriched in the distal part as expected (Fig. 6e). Additionally, interzone cells are enriched in the middle segment, where we intentionally included the joints. The same is true for the hindlimb. Comparison of forelimbs and hindlimbs demonstrated that both humans and mice show minimal differences in terms of cell type composition (Fig. 6e). This suggests that the composition of cell types of the developing limb is highly conserved between humans and mice even when pinpointing the broad anatomical regions. To perform a more stringent comparison, we took cells from the thirty-two LPM-derived states to compare ortholog expression signatures between proximal and distal segments in mouse and human. Both species recapitulate known P-D biased genes such as *MEIS*1 (proximal) and *HOXD*13 (distal) and known forelimb/hindlimb biased genes such as *TBX*5 (forelimb) and *TBX*4 (hindlimb) (Extended Data Fig. 10e). Overall, we show that the gene neighbourhoods controlling forelimb/hindlimb identity and P-D axis formation are highly conserved in evolution.

Discussion

Our developmental limb atlas combines single-cell RNA and spatial transcriptomic analyses of embryonic limb cells from multiple time points in the first trimester in order to form the first detailed characterisation of human limb development across space and time. We identify fifty-five cell states within eight tissue lineages in the developing limb and place them into anatomical context, building on existing knowledge of cellular heterogeneity gained from model organisms[133]. Our spatial data also reveals the expression of key regulators of limb axis identity, including the homeobox genes, in the nascent human limb.

In addition to recapitulating model organism biology, our atlas enables the identification of novel cell states. We identify a population of perineural fibroblasts surrounding the sciatic nerve and its tibial division, and we confirm their location using immunostaining. We also characterise several populations of mesenchymal cells, each defined by the expression of marker genes that in many cases play unclear roles in limb formation and should spur further investigation. The scale and resolution of our atlas also enables the construction of a refined model of cell states and regulators in partially overlapped and paralleled primary and secondary myogenesis in the limb marked by different panels of regulators, with the identification and validation of MSC as a key player in muscle stem cell maintenance.

Our atlas also leverages spatial data by placing subtly distinct single cell clusters into their anatomical context, shedding light on their true identity. In particular, two clusters of cells with subtly different transcriptomes mapped to the progress zone in two distinct bands, which we term “outer” and “transitional” layers. The gene module of the more proximal

transitional layer showed some similarity to that of chondroblasts, suggesting it may be beginning to differentiate towards these committed cell types. Similarly, two clusters in the tendon lineage map to the tendon and perimysium, giving insight into the subtle differences between these related tissues. Furthermore, through histological annotation of the developing autopod, we connected physiological gene expression patterns to single gene health conditions that involve altered hand phenotype, demonstrating the clinical relevance of developmental cell atlas projects. We further maximised the utility of this study by presenting an integrated cross-species atlas with unified annotations as a resource for the developmental biology community that we expect will strengthen future studies of limb development and disease that utilise murine models.

Whilst the combination of single cell and spatial transcriptomics is an established method for tissue atlasing, we recognise the challenges of combining different technologies. For example, our single-cell data captured large numbers of chondrocytes, including prehypertrophic chondrocytes which mapped to the mid-diaphysis of the forming bones. Interestingly, analysis of spatial gene expression revealed some collagen-X expression in these regions; a marker gene for mature, hypertrophic chondrocytes. Our scRNAseq experiments did not capture any cells expressing collagen-X, suggesting that permeabilisation and RNA capture with visium in this case was a superior method for profiling matrix-rich tissues such as mature cartilage. We expect these technical considerations to feed forward into future atlasing endeavours involving cartilage, bone and other dense tissues.

Methods

Human tissue sample collection

First trimester human embryonic tissue was collected from elective termination of pregnancy procedures at Addenbrookes Hospital, Cambridge, UK under full ethical approval (REC-96/085; for scRNA-seq and Visium), or at Guangzhou Women and Children's Medical Center, China under the license of ZSSOM-2019-075 approved by the human research ethics committee of Sun Yat-sen University (for experimental validation). Written, informed consent was given for tissue collection by the patient. Embyonic age (post conception weeks, PCW) was estimated using the independent measurement of the crown rump length (CRL), using the formula PCW (days) = 0.9022 × CRL (mm) + 27.372.

Human tissue processing and scRNA-seq data generation

Embryonic hindlimbs were dissected from the trunk under a microscope using sterile microsurgical instruments. Four samples (a hindlimb and a forelimb from both PCW5.6 and 6.1) were then further dissected into proximal, middle and distal thirds prior to dissociation. For the PCW5.1 sample, no further dissection was performed and the limb bud was dissociated as a whole. For all other samples, the limb was dissected into proximal and distal halves prior to dissociation.

Dissected tissues were mechanically chopped into a mash, and then were digested in Liberase TH solution (Roche, 05401135001, 50 µg/ml) at 37°C for 30-40 min till no tissue pierce visible. Digested tissues were filtered through 40 µm cell strainers followed by centrifugation at 750g for 5 min at 4°C. Cell pellets were resuspended with 2% FBS in PBS if the embryos were younger than PCW8, otherwise red blood cell lysis (eBioscience, 00-4300) was performed. The single-cell suspensions derived from each sample were then loaded onto separate channels of a Chromium 10x Genomics single cell 3'version 2 library chip as per the manufacturer's protocol (10x Genomics; PN-120233). cDNA sequencing libraries were prepared as per the manufacturer's protocol and sequenced using an Illumina Hi-seq 4000 with 2x150bp paired-end reads.

Mouse tissue sample collection and scRNA-seq data generation

Timed pregnant C57BL/6J wild type mice were ordered from Jackson Laboratories. Embryos were collected at E12.5, E13.5 and E16.5. Only right side forelimbs and hindlimbs were used in this study: n=5 at the E12.5 timepoint, n=5 at E13.5 and n=2 at E16.5. Hindlimbs and forelimbs were pooled separately in ice cold HBSS (Gibco, 14175-095), and dissected into proximal, mid and distal limb regions, which were again separately pooled in 200 µl of HBSS placed in a drop in the centre of a 6 cm culture plate. Tissues were then minced with a razor blade, and incubated with an addition of 120 µl of diluted DNAse solution (Roche, 04716728001) at 37C for 15 minutes. DNAse solution: 1 ml UltraPure water (Invitrogen, 10977-015) 110 µl 10X DNAse buffer, and 70 µl DNAse stock solution. 2 ml of diluted Liberase TH (Roche, 05401151001) was then added to the plate, and the minced tissue suspension was pipetted into a 15 ml conical centrifuge tube. The culture plate was rinsed with 2 ml, and again with 1 ml⊾ of fresh Liberase TH which was serially collected and added to the cell suspension. The suspension was incubated at 37°C for 15 minutes, triturated with a P1000 tip, and incubated for an additional 15 minutes at 37°C. Liberase TH solution: 50X stock was prepared by adding 2 ml PBS to 5 mg of Liberase TH. Working solution is made by adding 100 µl 50X stock to 4.9 ml PBS. After a final gentle trituration of the tissue with a P1000 tip, the suspension was spun at 380g in a swinging bucket rotor at 4°C for 5 minutes. After removing the supernatant, cells were resuspended in 5 ml of 2% fetal bovine serum in PBS, and filtered through a pre-wetted 40 µm filter (Falcon, 352340). After spinning again at 380g at 4°C for 5 minutes, the supernatant was removed and cells were resuspended in 200 ul 2% FBS in PBS. A small aliquot was diluted 1:10 in 2% FBS/PBS and mixed with an equal volume of Trypan Blue for counting on a hemocytometer. The full suspension was diluted to 1.2 million cells/ml for processing on the 10x Genomics Chromium Controller, with a target of 8000 cells/library. Libraries were processed according to the manufacturer's protocol, using the v3 Chromium reagents.

Visium spatial transcriptomic experiments of human tissue

Whole embryonic limb samples at PCW6-8 were embedded in OCT within cryo wells and flash-frozen using an isopentane & dry ice slurry. Ten-micron thick cryosections were then cut in the desired plane and transferred onto Visium slides prior to haematoxylin and eosin staining and imaged at 20X magnification on a Hamamatsu Nanozoomer 2.0 HT Brightfield.

These slides were then further processed according to the 10X Genomics Visium protocol, using a permeabilisation time of 18min for the PCW6 samples and 24 minutes for older samples. Images were exported as tiled tiffs for analysis. Dual-indexed libraries were prepared as in the 10X Genomics protocol, pooled at 2.25 nM and sequenced 4 samples per Illumina Novaseq SP flow cell with read lengths 28bp R1, 10bp i7 index, 10bp i5 index, 90bp R2.

Digit region analysis of Visium data

For differential gene expression testing, regions of interest were annotated based on H&E histology, and significant feature analysis performed using the 10X Genomics Loupe Browser 4.1.0, selecting for locally distinguishing features between the two regions. Results were visualised using the violin plot function in Loupe.

Alignment, quantification and quality control of human scRNA-seq data.

Droplet-based (10X) sequencing data were aligned and quantified using the Cell Ranger Single-Cell Software Suite (v.2.1.1, 10X Genomics) against the Cell Ranger hg38 reference genome refdata-cellranger-GRCh38-3.0.0, available at: http://cf.10xgenomics.com/supp/cell-exp/refdata-cellranger-GRCh38-3.0.0.tar.gz. The following quality control steps were performed: (i) cells that expressed fewer than 200 genes (low quality), and over 10,000 genes (potential doublets) were excluded; (ii) genes expressed by less than 5 cells were removed; (iii) cells in which over 10% of unique molecular identifier (UMIs) were derived from the mitochondrial genome were removed.

Alignment and quantification of human Visium data.

Raw FASTQ files and histology images were processed, aligned and quantified by sample using the Space Ranger software v.1.0.0. which uses STAR v.2.5.1b52 for genome alignment, against the Cell Ranger hg38 reference genome refdata-cellranger-GRCh38-3.0.0, available at: http://cf.10xgenomics.com/supp/cell-exp/refdata-cellranger-GRCh38-3.0.0.tar.gz.

Alignment, quantification and quality control of mouse scRNA-seq data.

Droplet-based (10x) sequencing data were aligned and quantified using the Cell Ranger Single-Cell Software Suite (v.3.0.2, 10x Genomics) against the Cell Ranger mm10 reference genome refdata-gex-mm10-2020-A, available at: https://cf.10xgenomics.com/supp/cell-exp/refdata-gex-mm10-2020-A.tar.gz. The following quality control steps were performed: (i) cells that expressed fewer than 200 genes (low quality) were excluded; (ii) genes expressed

by less than 5 cells were removed; (iii) cells in which over 10% of unique molecular identifier (UMIs) were derived from the mitochondrial genome were removed.

Doublet detection of human scRNA-seq data.

Doublets were detected with an approach adapted from a previous study[157]. In the first step of the process, each 10X lane was processed independently using the Scrublet to obtain per-cell doublet scores. In the second step of the process, the standard Scanpy processing pipeline was performed up to the clustering stage, using default parameters. Each cluster was subsequently separately clustered again, yielding an over-clustered manifold, and each of the resulting clusters had its Scrublet scores replaced by the median of the observed values. The resulting scores were assessed for statistical significance, with P values computed using a right tailed test from a normal distribution centred on the score median and a median absolute deviation (MAD)-derived standard deviation estimate. The MAD was computed from above-median values to circumvent zero truncation. The P values were corrected for false discovery rate with the Benjamini–Hochberg procedure, and a significance threshold of 0.1 was imposed. Cells with a significant corrected P values were detected as doublets and removed.

Data preprocessing and integration of human scRNA-seq data.

Preprocessing included data normalisation (pp.normalize_per_cell with 10,000 counts per cell after normalisation), logarithmise (pp.log1p), highly variable genes (HVGs) detection (pp.highly_variable_genes with batch key (sample identities)), data feature scaling (pp.scale), cell cycle regressing (tl.score_gene_cell_cycle and pp.regress_out), and principal component analysis (PCA) (tl.pca with 100 components) performed using the Python package Scanpy (v.1.6.0). bbknn (v.1.3.11) was used to correct for batch effect between sample identities with the following parameters (n_pcs = 70, neighbors_within_batch = 3, trim = 200, metric= “euclidean”).

Data preprocessing and integration of mouse scRNA-seq data.

Preprocessing included data normalisation (pp.normalize_per_cell with 10,000 counts per cell after normalization), logarithmise (pp.log1p), highly variable genes (HVGs) detection (pp.highly_variable_genes and select for highly correlated ones as described in [133]) per batch and merging, data feature scaling (pp.scale), cell cycle regressing (tl.score_gene_cell_cycle and pp.regress_out), and principal component analysis (PCA) (tl.pca with 100 components) performed using the Python package Scanpy (v.1.8.2). bbknn (v.1.5.1) was used to correct for batch effect between sample identities with the following parameters (n_pcs = 100, metric= “euclidean”).

Clustering and annotation of human scRNA-seq data.

We first performed dimension reduction using Uniform Manifold Approximation and Projection (UMAP) (scanpy tl.umap with default parameters) on 1903 highly variably genes. Next, we applied Leiden graph-based clustering (scanpy tl.leiden with default parameters) to perform unsupervised cell classification. To make sure all expected Leiden clusters could clearly be mapped onto their UMAP embedding coordinates, we performed the partition-based graph abstraction (PAGA) (tl.paga with the Leiden clusters) and rerun UMAP with the initial position from PAGA. Cluster cell identity was assigned by manual annotation using known marker genes and computed DEGs using self-designed framework. The detailed description of the method for DEGs detection is available on Github (https://github.com/ZhangHongbo-Lab/DEAPLOG).

Deconvolution of human Visium data- cell2location

To map cell types identified by scRNA-seq in the profiled spatial transcriptomics slides, we used the cell2location method [19]. In brief, this involved first training a negative binomial regression model to estimate reference transcriptomic profiles for all the cell types identified with scRNA-seq in the developing limb. Next, lowly expressed genes are excluded as per recommendations for use of cell2location. Next, we estimated the abundance of cell types in the spatial transcriptomics slides using the reference transcriptomic profiles of different cell types. To identify microenvironments of co-localising cell types, we used non-negative matrix factorisation (NMF) implementation in scikit-learn, utilising the wrapper in the cell2location package [158]. A cell type was considered part of a microenvironment if the fraction of that cell type in said environment was over 0.2.

Alignment and merging of multiple visium sections

In order to analyse the whole PCW 8.1 human hindlimb, we took three consecutive ten micron sections from different regions and placed them on different capture areas of the same Visium LP slide. The first section spanned the distal femur, knee joint and proximal tibia (sample C42A1), the second the proximal thigh (C42B1) and the third the distal tibia, ankle and foot (C42C1).

The images from these 3 visium capture areas were then aligned using the TrackEM plugin (Fiji)[159]. Following affine transformations of C42B1 and C42C1 to C42A1, the transformation matrices were exported to an in house pipeline (/22-03-29-visium-stitch/limb_reconst.ipynb) for complementary alignment of the spot positions from the SpaceRanger output to the reconstructed space. In addition we arbitrarily decided that for regions of overlapping spots we kept the spots from the centre portion (see supp) and the same decision was made for the reconstructed image in order to maintain the 1-1 relationship between the image and genetic profile. Next we merged the 3 library files by an in-house pipeline (22-03-29-visium-stitch/N1-join-limb.ipynb) and matched the reconstructed image to the uniform AnnData object.

Trajectory analysis of human scRNA-seq data.

Development trajectories were inferred by combining diffusion maps (DM), PAGA and force-directed graph (FDG). The first step of this process was to perform the first nonlinear dimensionality reduction using DM (scanpy tl.diffmap with 15 components) and recompute the neighborhood graph (scanpy pp.neighbors) based on the 15 components of DM. In the second step of this process, PAGA (scanpy tl.paga) was performed to generate an abstracted graph of partitions. Finally, FDG was performed with the initial position from PAGA (scanpt tl.draw_graph) to visualise the development trajectories.

Cell-cell communication analysis of human scRNA-seq data.

Cell–cell communication analysis was performed using the CellPhoneDB (v.2.1.4) for each dataset at the same stage of development. The stage matched Visium data was used to validate the spatial distance and expression pattern of significant ($P < 0.05$) ligand-receptor interactions.

Enrichment analysis of transcription factors (TFs).

To carry out transcription factor network inference, Analysis was performed as previously described[160] using the pySCENIC python package (v.0.10.3). For the input data, we filtered out the genes that were expressed in less than 10 percent of the cells in each cell cluster. Then, we performed the standard procedure including deriving co-expression modules (pyscenic grn), finding enriched motifs (pyscenic ctx) and quantifying activity (pyscenic aucell).

Integration of human and mouse scRNA-seq data.

Processed human and mouse data were merged together using outer join of the ortholos. The matched dataset was then integrated by MultiMAP using the MultiMAP_Integration() function, using separately pre-calculated PCs and the union set of mouse and human highly-variable genes. Downstream clustering and embedding were performed as usual and cell-type annotation was based on marker genes. Cell-type composition of proximal, middle and distal segments of the same limb was visualised using plotly.express.scatter_ternary() function. To capture the differential expression of sparsely captured genes, the odds ratio of the percentages of non-zero cells between groups of cells was used to select for proximal/distal or fore/hind biassed genes with a cutoff at 30 fold and 3 fold respectively.

Immunohistochemistry

The lower limbs were post-fixed in 4% PFA for 24 h at 4°C followed by paraffin embedding. A thickness of 4 μm sections were boiled in 0.01 M citrate buffer (pH 6.0) after dewaxing. Immunofluorescent staining was then carried out as described previously[161]. Primary antibodies for RUNX2 (1:50, Santa Cruz, sc-390715), THBS2 (1:100, Thermo Fisher, PA5-76418), COL2A1 (1:200, Santa Cruz, sc-52658), PITX1 (1:30, Abcam, Ab244308), PAX3 (1:1, DSHB, AB_528426 supernatant) and ALDH1A3 (1:50, Proteintech, 25167-1-AP) and MYH3 (1:3, DSHB, AB_528358 supernatant) were incubated overnight at 4°C. After washing, sections were incubated with appropriate secondary antibodies Alexa Flour 488 goat anti-mouse IgG1 (Invitrogen, A-21121), Alexa Flour 647 goat anti-mouse IgG2b (Invitrogen, A-21242), Alexa Flour 488 goat anti-mouse IgG (H+L) (Invitrogen, A-11029) and Alexa Flour 546 goat anti-rabbit IgG (H+L) (Invitrogen, A-11035) at room temperature for 1h, and were mounted using FluorSave Reagent (Calbiochem, 345789). For 3, 3-diaminobenzidine (DAB) staining, we used Streptavidin-Peroxidase broad spectrum kit (Bioss, SP-0022) and DAB solution (ZSGB-BIO, ZLI-9017) following the manufacturers' manuals. Primary antibodies PI16 (1:500, Sigma-Aldrich, HPA043763), FGF19 (1:500, Affinity, DF2651) and NEFH (1:1000, Cell Signaling, 2836) were applied. Single-plane images were acquired using an inverted microscope (Leica, DMi8).

MSC knockdown in human primary myoblasts

Isolation of human primary myoblast cells

The thighs from human embryos were processed as described[162], except that the dissociated cells were not treated with erythrocyte lysis solution, and were incubated with anti-human CD31 (eBioscience, 12-0319-41), CD45 (eBioscience, 12-0459-41) and CD184 (eBioscience, 17-9999-41) antibodies for cell sorting. Fluorescent activated cell sorting (FACS, BD, influx) sorted CD31-CD45-CD184+ cells were cultured in complete growth medium DMEM supplemented with 20% FCS and 1% penicillin/streptomycin (Gibco, 15140122).

siRNA transfection

Human primary myoblasts were seeded into a 6-well plate one night before transfection. When the cell density reached approximately 50% confluence, oligos of small interfering RNA (siRNA) against MSC (si-MSC) and negative control (NC) were transfected using Lipofectamine 3000 reagent (Invitrogen, L3000015) at a final concentration of 37.5 nM. After incubation for 16 h, the growth medium was replaced with differentiation medium containing 2% horse serum and 1% penicillin/streptomycin in DMEM. After culturing for an additional 6-8 h, the cells were collected for RNA extraction. Initially three siRNA oligos (Bioneer, 9242-1, 2, 3) were

tested, and the third one with sense sequences 5′-GAAGUUUCCGCAGCCAACA-3′ were used in this study.

RNA extraction and quantitative PCR (qPCR)

Total cell RNA was extracted with the EZ-press RNA purification kit (EZBioscience, B0004D), and the cDNA was synthetised using the PrimeScript RT Master Mix Kit (TaKaRa, RR036A). The qPCR was performed using PerfectStartTM Green qPCR Super Mix (TransGen Biotech, AQ601) on a Real-time PCR Detection System (Roche, LightCycle480 II). RPLP0 served as an internal control, and the fold enrichment was calculated using the formula $2-\Delta\Delta Ct$. The following primers (5'-3') were used:

RPLP0 forward: ATGCAGCAGATCCGCATGT, reverse: TTGCGCATCATGGTGTTCTT;
MSC forward: CAGGAGGACCGCTATGAGAA, reverse: GCGGTGGTTCCACATAGTCT;
MYOG forward: AGTGCCATCCAGTACATCGAGC, reverse: AGGCGCTGTGAGAGCTGCATTC;
MYH2 forward: GGAGGACAAAGTCAACACCCTG, reverse: GCCCTTTCTAGGTCCATGCGAA;
MYH3 forward: CTGGAGGATGAATGCTCAGAGC, reverse: CCCAGAGAGTTCCTCAGTAAGG;
MYH4 forward: CGGGAGGTTCACACAAAAGTCATA, reverse: CCTTGATATACAGGACAGTGACAA;
TNNT1 forward: AACGCGAACGTCAGGCTAAGCT, reverse: CTTGACCAGGTAGCCGCCAAAA.

Figures and figure legends

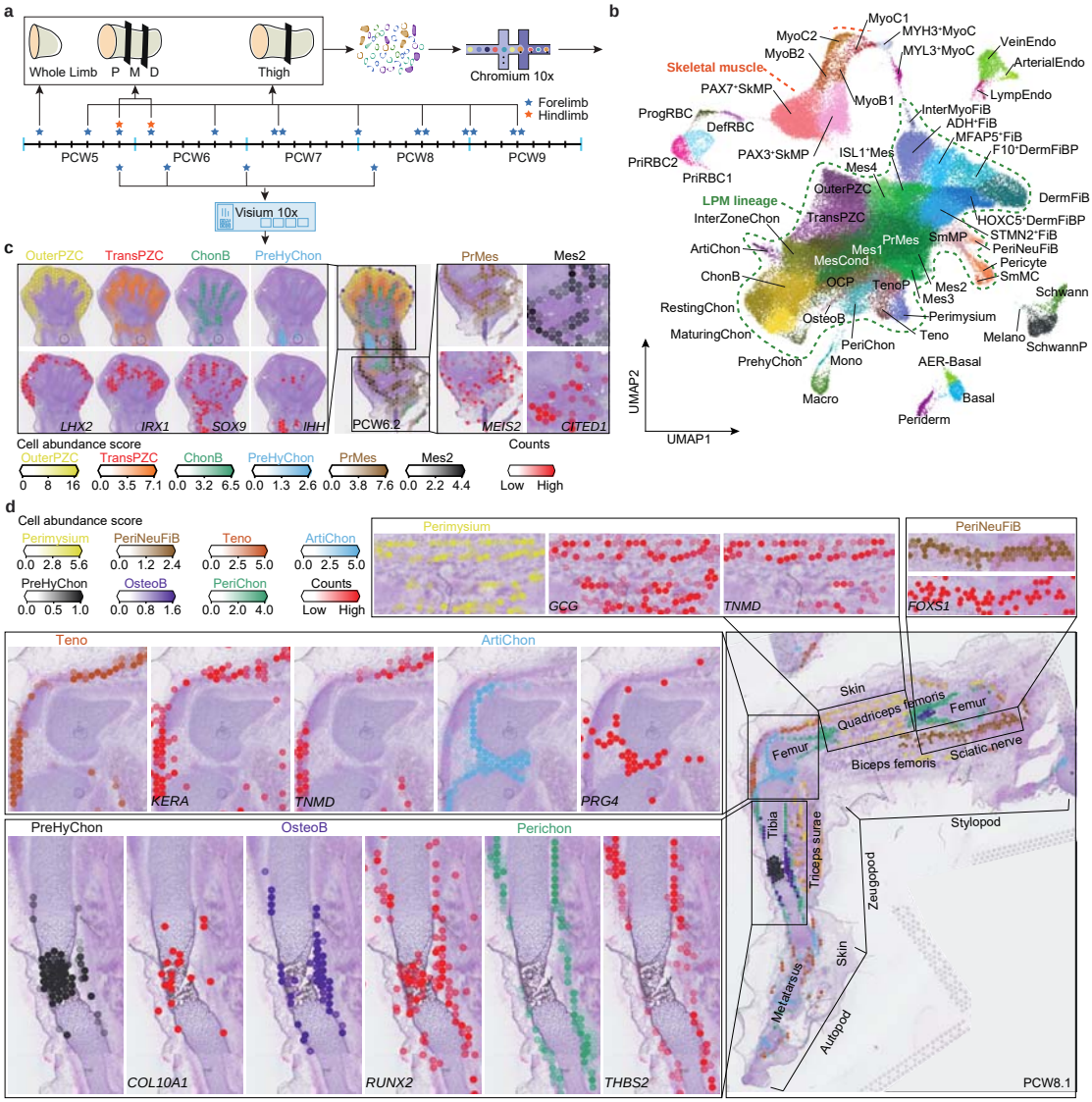
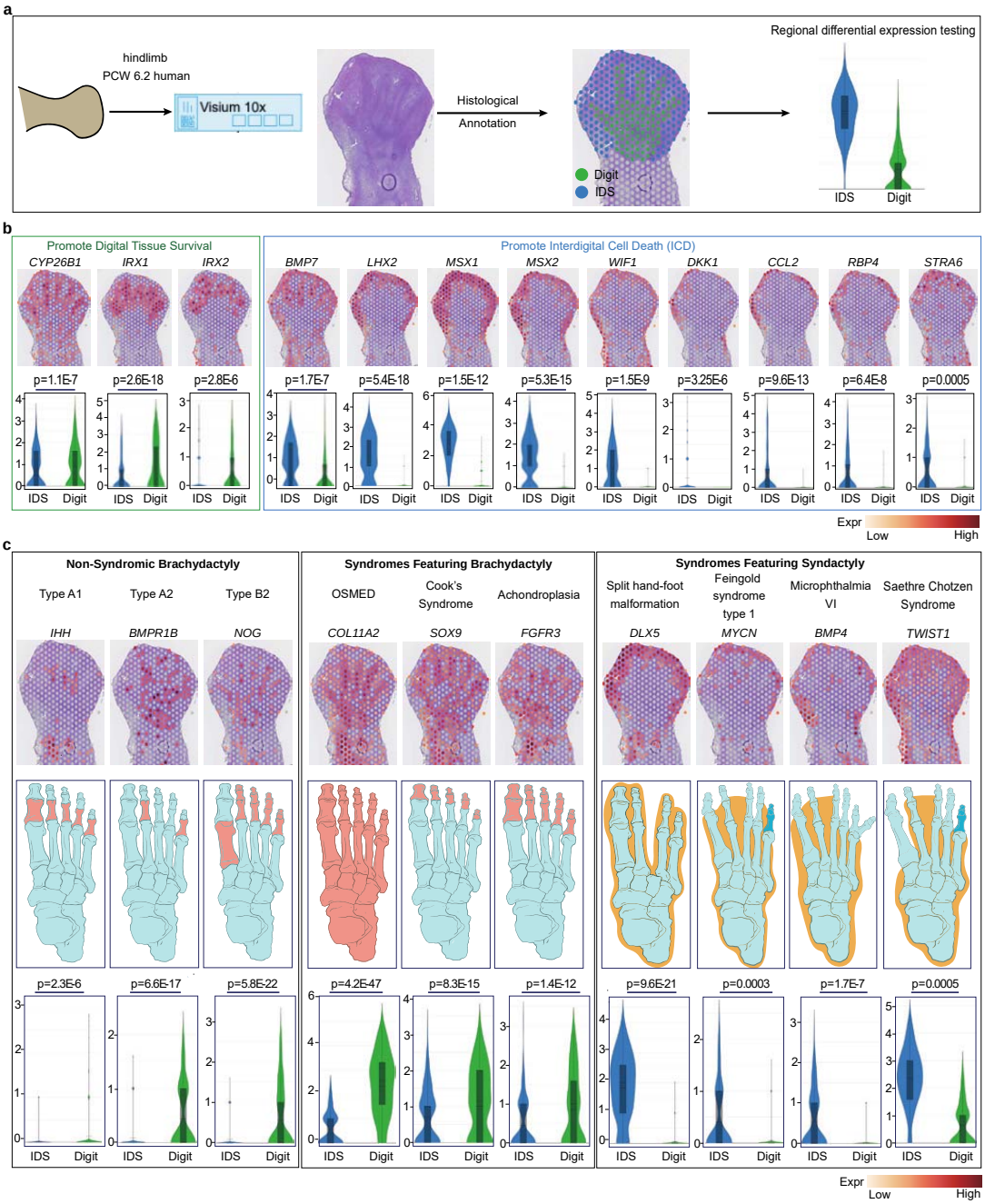


Figure 1. A single-cell temporal-spatial atlas of human embryonic limb.

a, Overview of human embryonic developmental time points sampled and experimental scheme. The asterisk marks the sampling time point. P, proximal; M, middle; D, distal; PCW, post conceptional week. **b**, Uniform manifold approximation and projection (UMAP) visualisation of 114,047 human embryonic limb cells, coloured by their cell-type annotation. The orange dotted line shows the skeletal muscle lineage, and the green dotted line shows the lateral plate mesoderm (LPM) lineage. AER, apical ectoderm region; Chon, chondrocyte; ArtiChon, articular Chon; ChonB, chondroblast; PrehyChon, prehypertrophic Chon; PeriChon, perichondrium cell; Endo, Endothelial; FiB, fibroblast; DermFiBP, dermal fibroblast progenitor; PeriNeuFiB, perineural FiB; LympEndo, lymphatic Endo; Macro, macrophage; Melano,

melanocyte; Mes, mesenchyme; PrMes, Proximal Mes; MesCond, mesenchymal condensate cell; Mono, monocyte; MyoB, myoblast; MyoC, myocyte; OCP, osteochondral progenitor; OsteoB, osteoblast; PZC, progress zone cell; TransPZC, transitional PZC; RBC, red blood cell; DefRBC, definitive RBC; PriRBC, primitive RBC; ProgRBC, progenitor RBC; SchwannP, schwann progenitor; SkMP, skeletal muscle progenitor; SmMP, smooth muscle progenitor; SmMC, smooth muscle cell; TenoP, tenocyte progenitor. **c-d**, Spatially resolved heatmaps across tissue sections from PCW 6.2 (**c**) and PCW 8.1 (**d**) human hindlimb showing spatial distribution of cells and marker genes of OuterPZC, TransPZC, ChonB, PreHyChon, PrMes, Mes2 (**c**), PeriChon, OsteoB, PreHyChon, ArtiChon, PeriNeuFiB, Perimysium and Teno (**d**) based on non-negative matrix factorisation. colour bars denote the cell abundance score and gene expression levels.

963



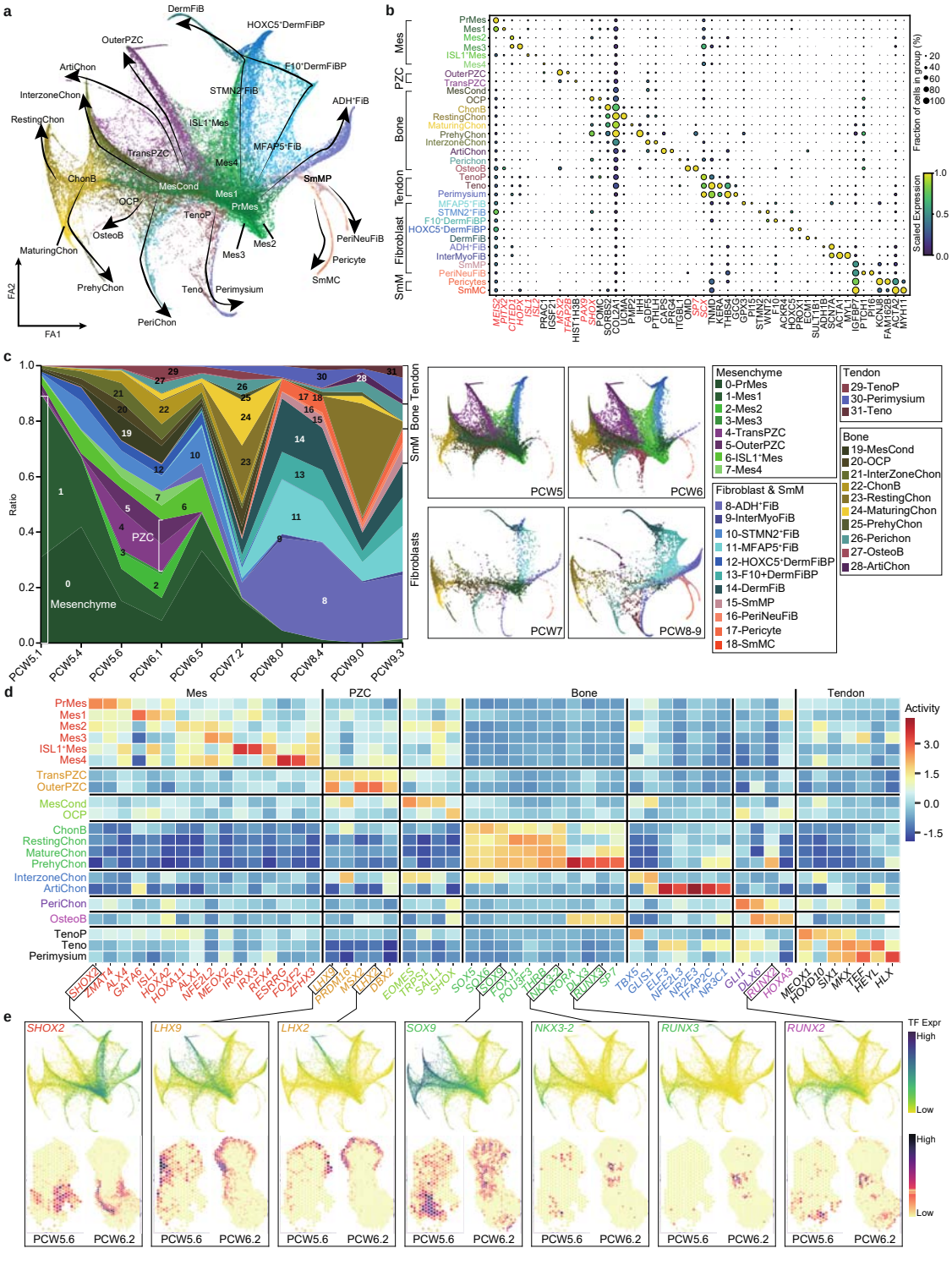
964

965 Figure 2. Spatial expression pattern of genes involved in digit formation and altered
966 hand phenotype.

967 **a**, Overview of experimental scheme to identify genes involved in digit formation and
968 interdigital cell death (ICD). IDS, interdigital space. **b-c**, Tissue sections from the PCW
969 6.2 human hindlimb showing spatial expression pattern of genes promoting digital
970 tissue survival (left panel) and ICD (right panel) (**b**) as well as genes linked to altered
971 digital phenotype(**c**). Red colouring of bones denotes those affected by brachydactyly;

light blue denotes clinodactyly. Violin plot showing gene significance of difference between digital region and IDS. Colour bars encode the gene expression levels.

976



977

Figure 3. Cell lineage diversification of the lateral plate mesoderm (LPM) and associated transcription factor (TF) activity during human embryonic limb development.

a, Force-directed graph representation of cells associated with the LPM lineage, coloured by global cell-type annotation. Black arrows indicate the direction of cell differentiation. Chon, chondrocyte; ArtiChon, articular Chon; ChonB, chondroblast; PrehyChon, prehypertrophic Chon; PeriChon, perichondrium cell; FiB, fibroblast; DermFiBP, dermal fibroblast progenitor; PeriNeuFiB, perineural FiB; Mes, mesenchyme; PrMes, Proximal Mes; MesCond, mesenchymal condensate cell; OCP, osteochondral progenitor; OsteoB, osteoblast; PZC, progress zone cell; TransPZC, transitional PZC; SmMP, smooth muscle progenitor; SmMC, smooth muscle cell; TenoP, tenocyte progenitor. **b**, Dot plot showing scaled expression of selected marker genes for each cell type. The size of the dot encodes the percentage of cells within a cell type in which that marker was detected, and its colour encodes the average expression levels. Red fonts denote known TFs. **c**, Fraction of cell type per time point. Left panel, Fraction of cell type per time point; right panel, Force-directed graph layout of cells from each post conceptional week, coloured by global cell type. **d**, Heat map illustrating the column-normalised mean activity of top5 TFs for each cell type from mesenchyme, progress zone cell, cartilage, bone and tendon lineage. **e**, Spatiotemporal expression patterns of selected TFs. Top panels; force-directed graph layout of cells coloured by expression level of selected transcription factors; bottom panels, tissue sections from the PCW 5.6 and PCW 6.2 human hindlimb showing spatial expression pattern of selected transcription factors.

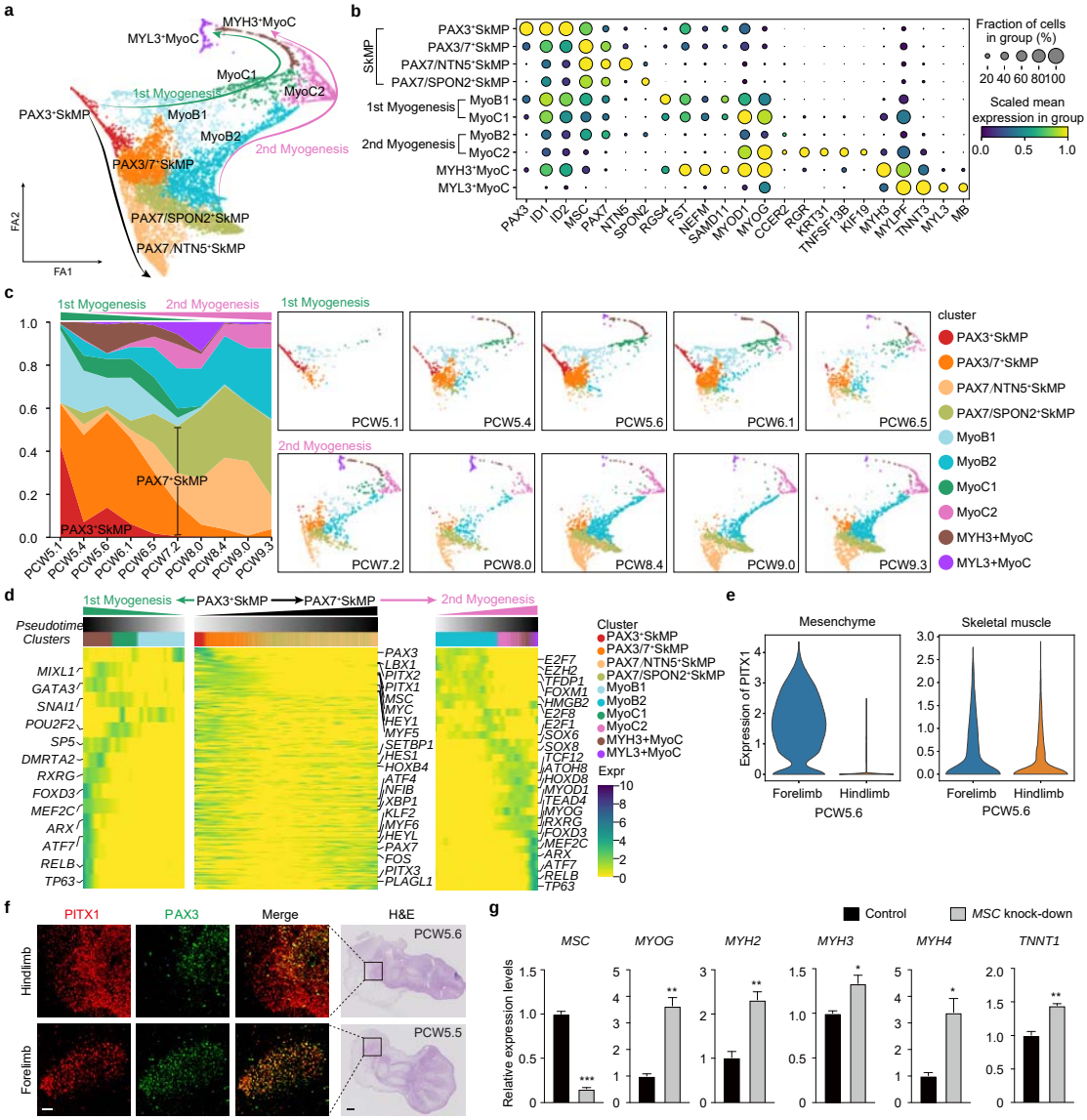


Figure 4. Cell trajectory and transcription factors (TFs) conversion of embryonic and fetal myogenesis during human embryonic limb development. a, Force-directed graph layout of skeletal muscle cells, coloured by sub-clusters. Green and pink arrows indicate the direction of first and second myogenesis, separately. SkMP, skeletal muscle progenitor; MyoB, myoblast; MyoC, myocyte. **b**, Dot plot showing scaled expression of selected marker genes for each sub-cluster. The size of the dot encodes the percentage of cells within a cell type in which that marker was detected, and its colour encodes the average expression level. **c**, Fraction of sub-cluster per time point. Left panel, Fraction of cell type per time point; right panel, Force-directed graph layout of cells from each time point, coloured by sub-clusters. **d**, Heatmap of enriched TFs expression (smoothed over 100 adjacent cells), with cells ordered by diffusion pseudotime and genes ordered according to pseudotime inferred from diffusion pseudotime of cells. Left panel, first myogenesis; middle panel, PAX3 positive

SkMP to PAX7 positive SkMP; right panel, second myogenesis. **e**, Violin plot showing the expression of PITX1 in human forelimb- and hindlimb. Left panel, expression in mesenchyme; right panel, expression in skeletal muscle. **f**, Immunofluorescent staining of PITX1 with PAX3 co-stained on lower (top panel) and upper (bottom panel) limb sections at indicated embryonic stages. Scale bar: Immunofluorescent staining, 50 μm; H&E staining, 200 μm. **g**, qPCR analysis of the fold enrichment of indicated myocyte differentiation genes upon knock-down of MSC in human primary embryonic myoblasts. Data are presented as mean ± SEM. *P < 0.05, **P < 0.01, ***P < 0.001 (two-sided Student's t test).

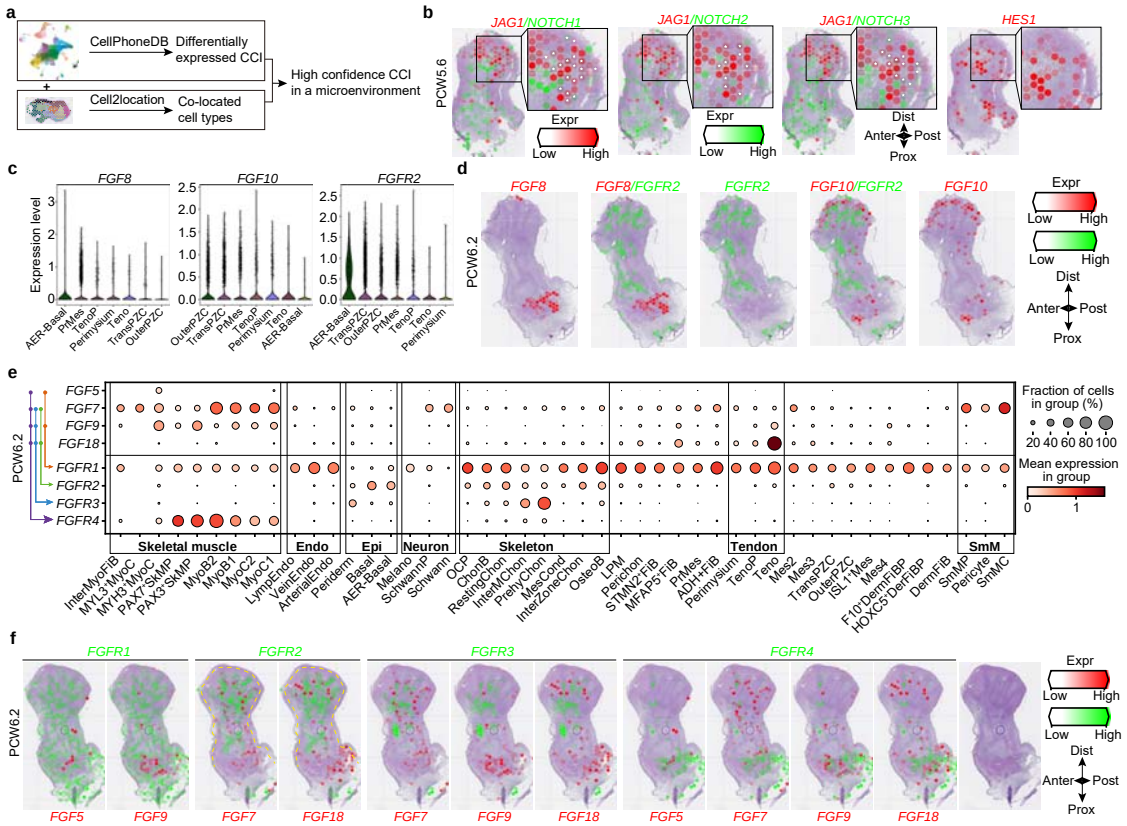


Figure 5. Spatially resolved cell-cell communication.

a, Overview of cell-cell communication pipeline combining scRNA-seq and spatial data. **b**, Spatially resolved heatmaps across tissue sections from the PCW 5.6 human hindlimb showing spatial expression of JAG1 and its receptors as well as HES1. Anter, anterior; Post, posterior; Prox, proximal; Dist, distal. **c**, Violin plots showing log-normalised expression of FGF8/10 and its receptor (FGFR10) in cell types. AER, apical ectoderm region; PrMes, proximal mesenchyme; TenoP, tenocyte progenitor; PZC, progress zone cell. **d**, Spatially resolved heatmaps across tissue sections from the PCW 6.2 human hindlimb showing spatial expression of FGF8/10 and its receptor (FGFR10). **e**, Dot plots showing scaled expression of FGFs (FGF5/7/9/18 and its corresponding receptors (FGFR1/2/3/4) in cell types. MyoB, myoblast; MyoC, myocyte; LympEndo, lymphatic Endo; Melano, Melanocyte; SchwannP, Schwann progenitor; OsteoB, osteoblast; Teno, tenocyte; TenoP, Teno progenitor; SmMP, smooth muscle progenitor; SmMC, smooth muscle cell. **f**, Spatially resolved heatmaps across tissue sections from the PCW 6.2 human hindlimb showing spatial expression of FGFs (FGF5/7/9/18 and its corresponding receptors (FGFR1/2/3/4).

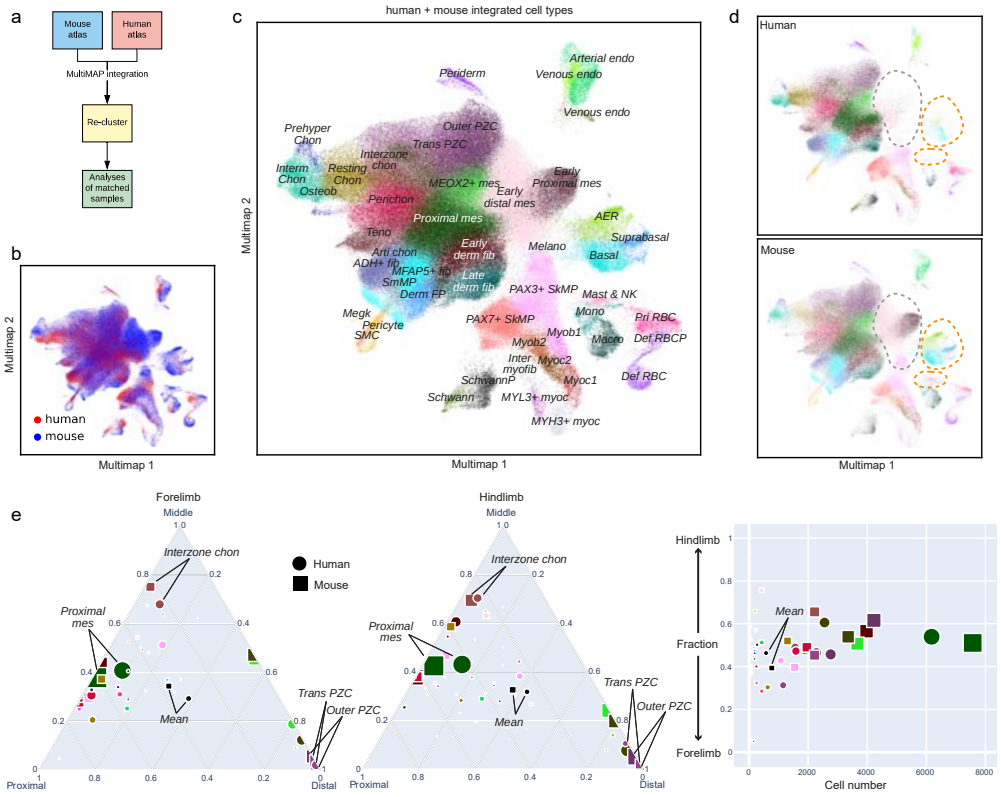
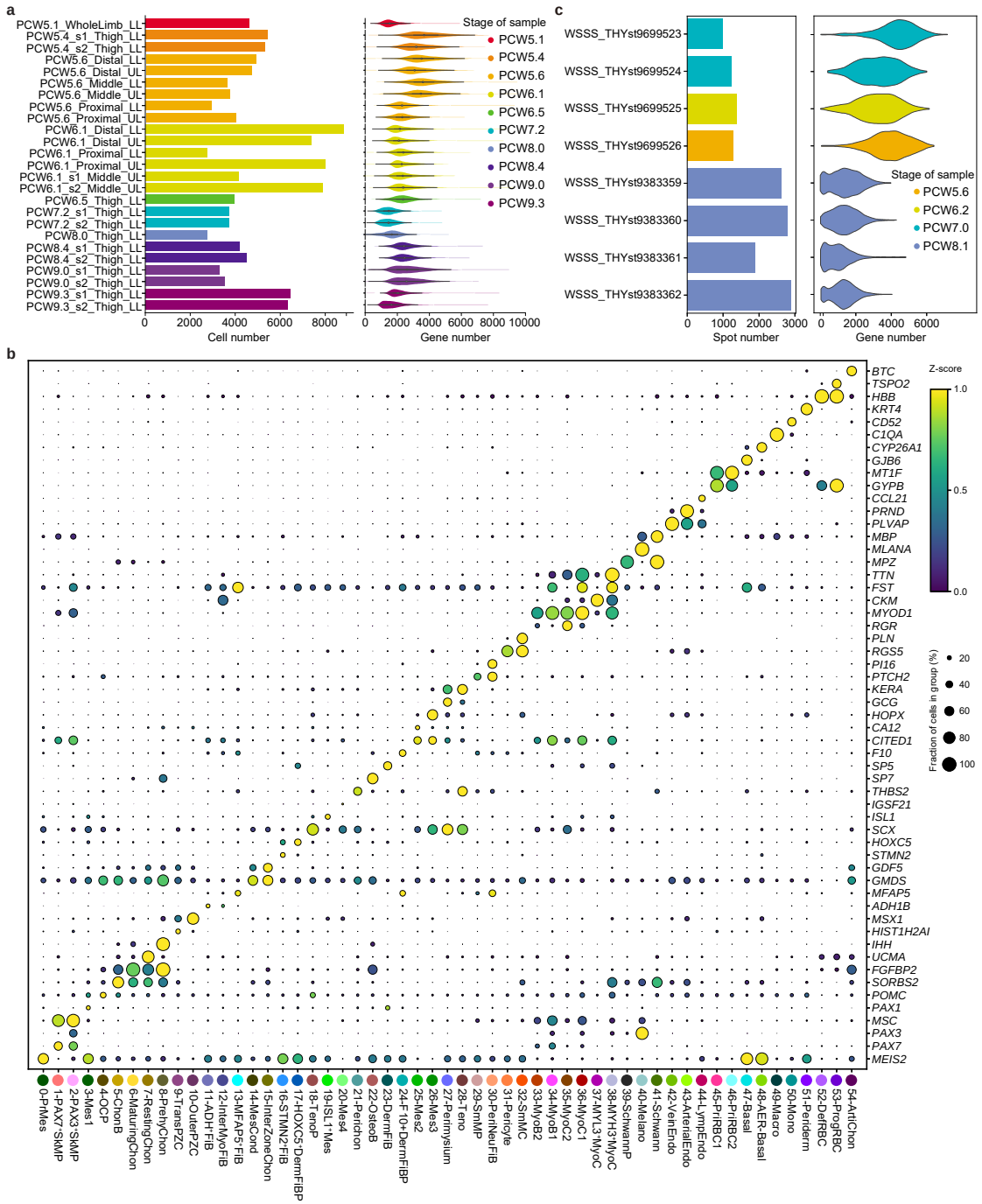


Figure 6. Comparison of single cell atlas between human and mouse limb.

a, Overview of analysis pipeline to integrate human and mouse scRNA-seq data. **b-d**, Multimap layout of integrated cells, coloured by species (**b**) and cell-type annotation (**c and d**). Dotted circles highlight populations that differ between species due to sampling (brown) or biological (orange) reasons. **e,** Triangular diagram showing the cell-type distribution in proximal, middle and distal region of human and mouse forelimb (left panel) and hindlimb (middle panel), and a scatter plot showing the fraction of each cell type's hindlimb representation (right panel). Each cell type is represented by a circle (human) and a square (mouse).

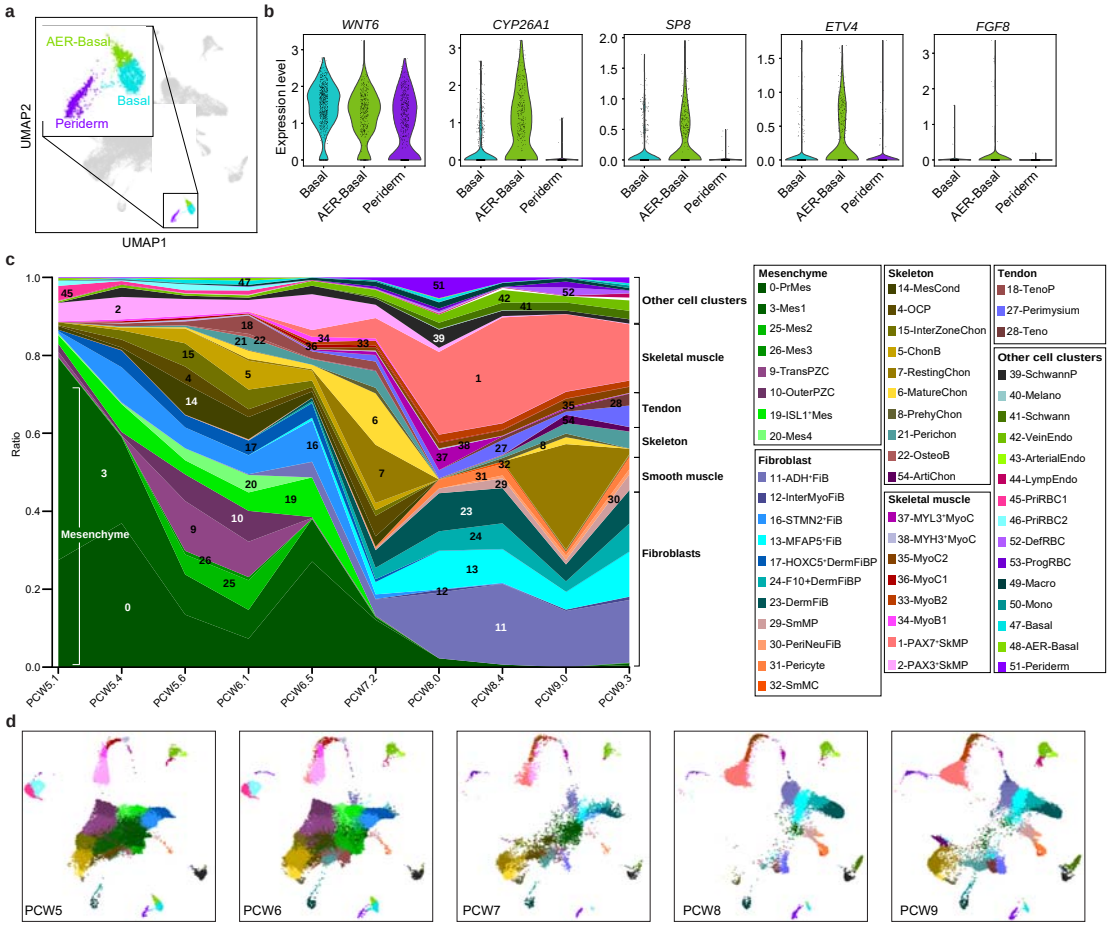
Extended data figures and legend



Extended Figure 1. Data quality and preprocess of human scRNA-seq data.

a, Histogram and violin plot showing the cell and gene number of cells per each sample, separately, coloured by stage of post conceptional week (PCW). **b**, Dot plot showing the scaled expression of selected marker genes for each cell type. The size

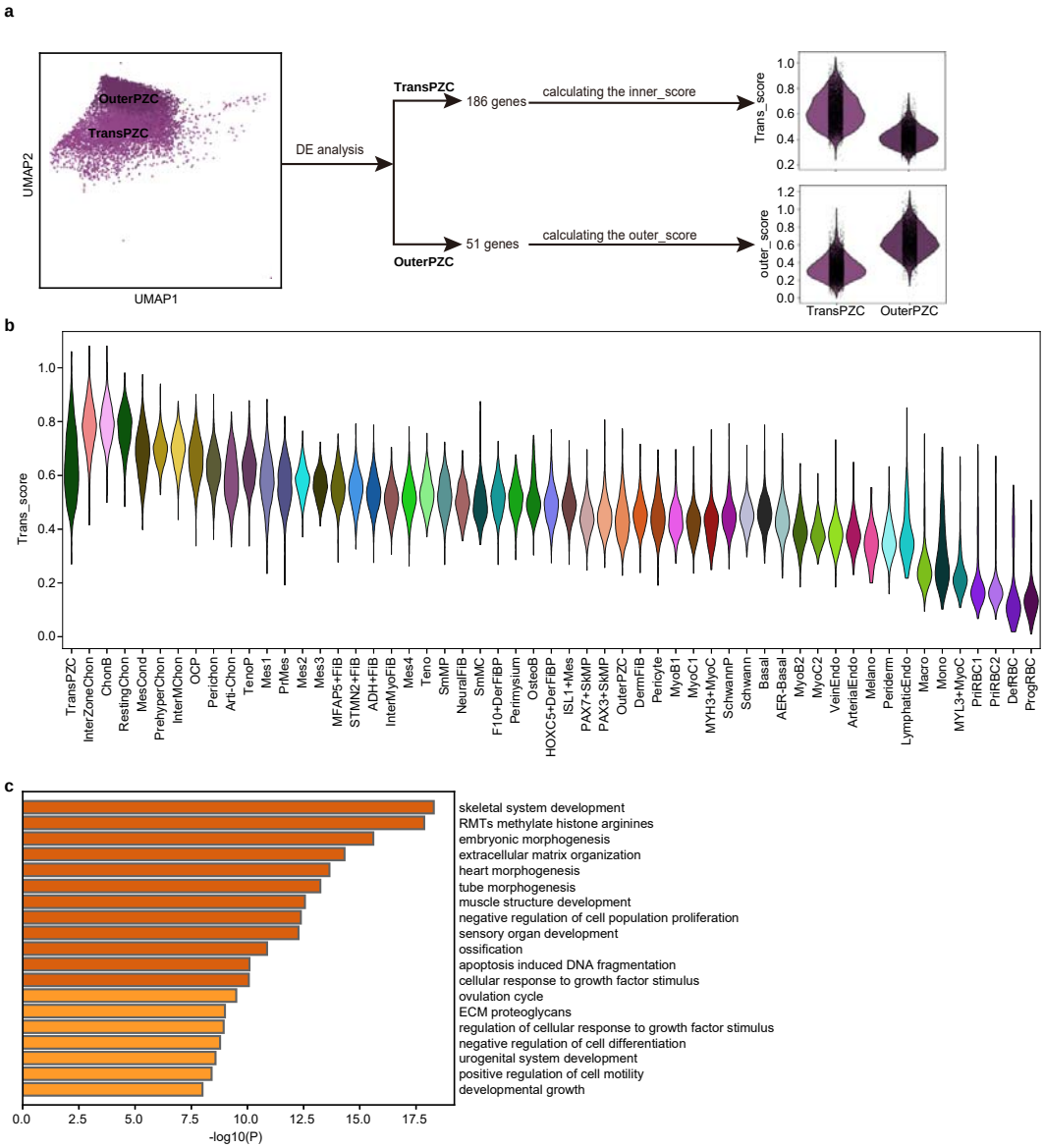
of the dot encodes the percentage of cells within a cell type in which that marker was detected, and its colour encodes the average expression level. AER, apical ectoderm region; Chon, chondrocyte; ArtiChon, articular Chon; ChonB, chondroblast; PrehyChon, prehypertrophic Chon; PeriChon, perichondrium cell; Endo, Endothelial; FiB, fibroblast; DermFiBP, dermal fibroblast progenitor; PeriNeuFiB, perineural FiB; LympEndo, lymphatic Endo; Macro, macrophage; Melano, melanocyte; Mes, mesenchyme; PrMes, Proximal Mes; MesCond, mesenchymal condensate cell; Mono, monocyte; MyoB, myoblast; MyoC, myocyte; OCP, osteochondral progenitor; OsteoB, osteoblast; PZC, progress zone cell; TransPZC, transitional PZC; RBC, red blood cell; DefRBC, definitive RBC; PriRBC, primitive RBC; ProgRBC, progenitor RBC; SchwannP, schwann cell progenitor; SkMP, skeletal muscle progenitor; SmMP, smooth muscle progenitor; SmMC, smooth muscle cell; TenoP, tenocyte progenitor. **c**, Histogram and violin plot showing the spot and gene number per spot for each 10X Visiumn spatial sample, separately, coloured by stage of post conceptional week (PCW).



Extended Figure 2. The expression of marker genes for apical ectoderm region (AER) in epidermis and Fraction of cell type per time point.

a, Uniform manifold approximation and projection (UMAP) visualisation of AER-basal cells, basal cells and periderm cells. **b**, Violin plot showing the log-normalised expression of WNT6, CYP26A1, SP8, ETV4 and FGF8 in AER-basal cells, basal cells and periderm cells, coloured by cell type. **c**, Fraction of cell type per time point, coloured by cell type and grouped by tissue type. AER, apical ectoderm region; Chon, chondrocyte; ArtiChon, articular Chon; ChonB, chondroblast; PrehyChon, prehypertrophic Chon; PeriChon, perichondrium cell; Endo, Endothelial; FiB, fibroblast; DermFiBP, dermal fibroblast progenitor; PeriNeuFiB, perineural FiB; LympEndo, lymphatic Endo; Macro, macrophage; Melano, melanocyte; Mes, mesenchyme; PrMes, Proximal Mes; MesCond, mesenchymal condensate cell; Mono, monocyte; MyoB, myoblast; MyoC, myocyte; OCP, osteochondral progenitor; OsteoB, osteoblast; PZC, progress zone cell; TransPZC, transitional PZC; RBC, red blood cell; DefRBC, definitive RBC; PriRBC, primitive RBC; ProgRBC, progenitor RBC; SchwannP, schwann progenitor; SkMP, skeletal muscle progenitor; SmMP, smooth

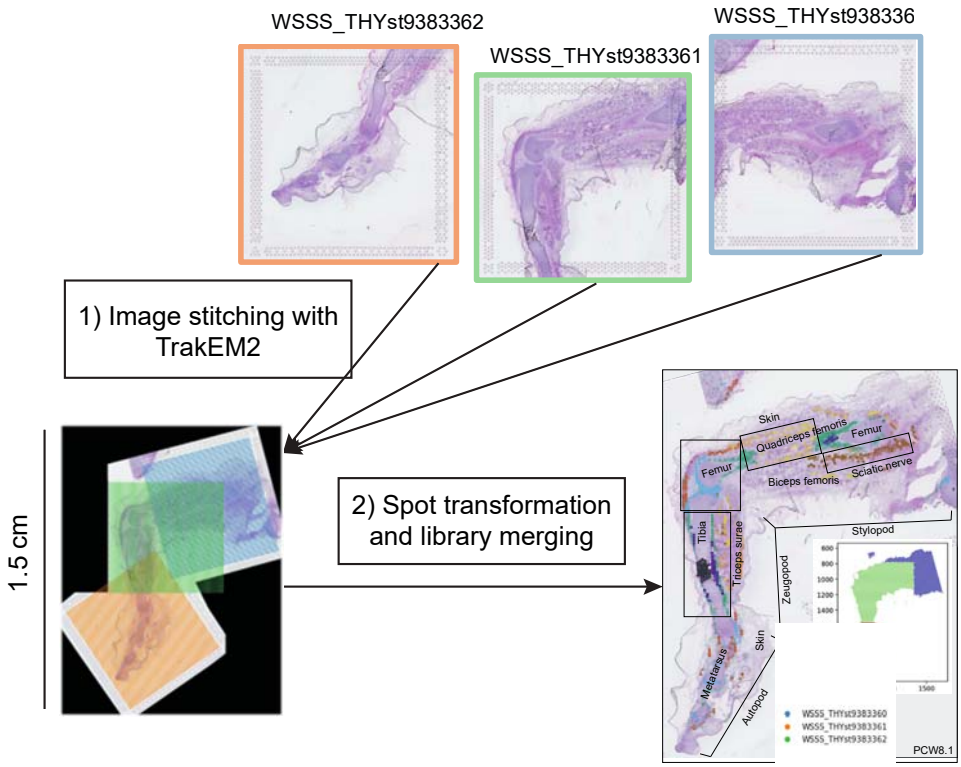
muscle progenitor; SmMC, smooth muscle cell; TenoP, tenocyte progenitor. **d**, Uniform manifold approximation and projection (UMAP) visualisation of fraction of cell type per post conceptional week (PCW), coloured by cell type in **c**.



Extended Figure 3. Cellular and functional heterogeneity of progress zone cells (PZC).

a, Overview of analysis pipeline to identify and score gene signature of outerPZC and transition PZC (transPZC). **b**, Violin plot showing the score of gene signature of tranPZC. Colored by cell type. AER, apical ectoderm region; Chon, chondrocyte; ArtiChon, articular Chon; ChonB, chondroblast; PrehyChon, prehypertrophic Chon; PeriChon, perichondrium cell; Endo, Endothelial; FiB, fibroblast; DermFiBP, dermal fibroblast progenitor; PeriNeuFiB, perineural FiB; LympEndo, lymphatic Endo; Macro, macrophage; Melano, melanocyte; Mes, mesenchyme; PrMes, Proximal Mes; MesCond, mesenchymal condensate cell; Mono, monocyte; MyoB, myoblast; MyoC, myocyte; OCP, osteochondral progenitor; OsteoB, osteoblast; PZC, progress zone cell; TransPZC, transitional PZC; RBC, red blood cell; DefRBC, definitive RBC;

PriRBC, primitive RBC; ProgRBC, progenitor RBC; SchwannP, schwann progenitor; SkMP, skeletal muscle progenitor; SmMP, smooth muscle progenitor; SmMC, smooth muscle cell; TenoP, tenocyte progenitor. **c**, Functional enrichment analysis of gene signature of transPZC.

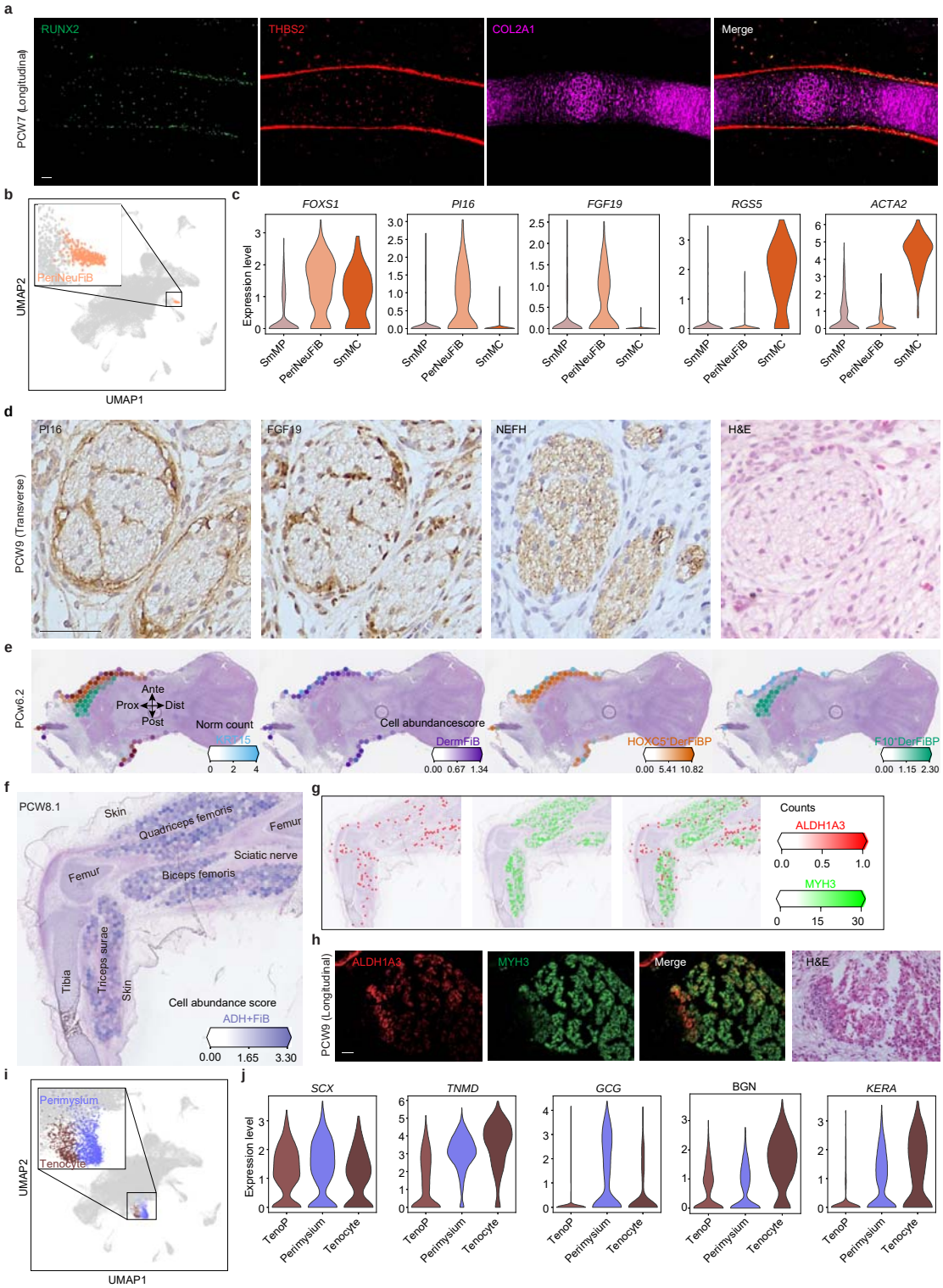


1148

1149

1150 Extended Figure 4. Workflow of concatenating Visium spatial transcriptomic data

1151 A two-step workflow to stitch Visium data is presented. Three independent capture areas
1152 each profiling an independent slice of the same sample are represented in three different
1153 colors. The final product is adapted from Fig. 1d with sample origin highlighted.

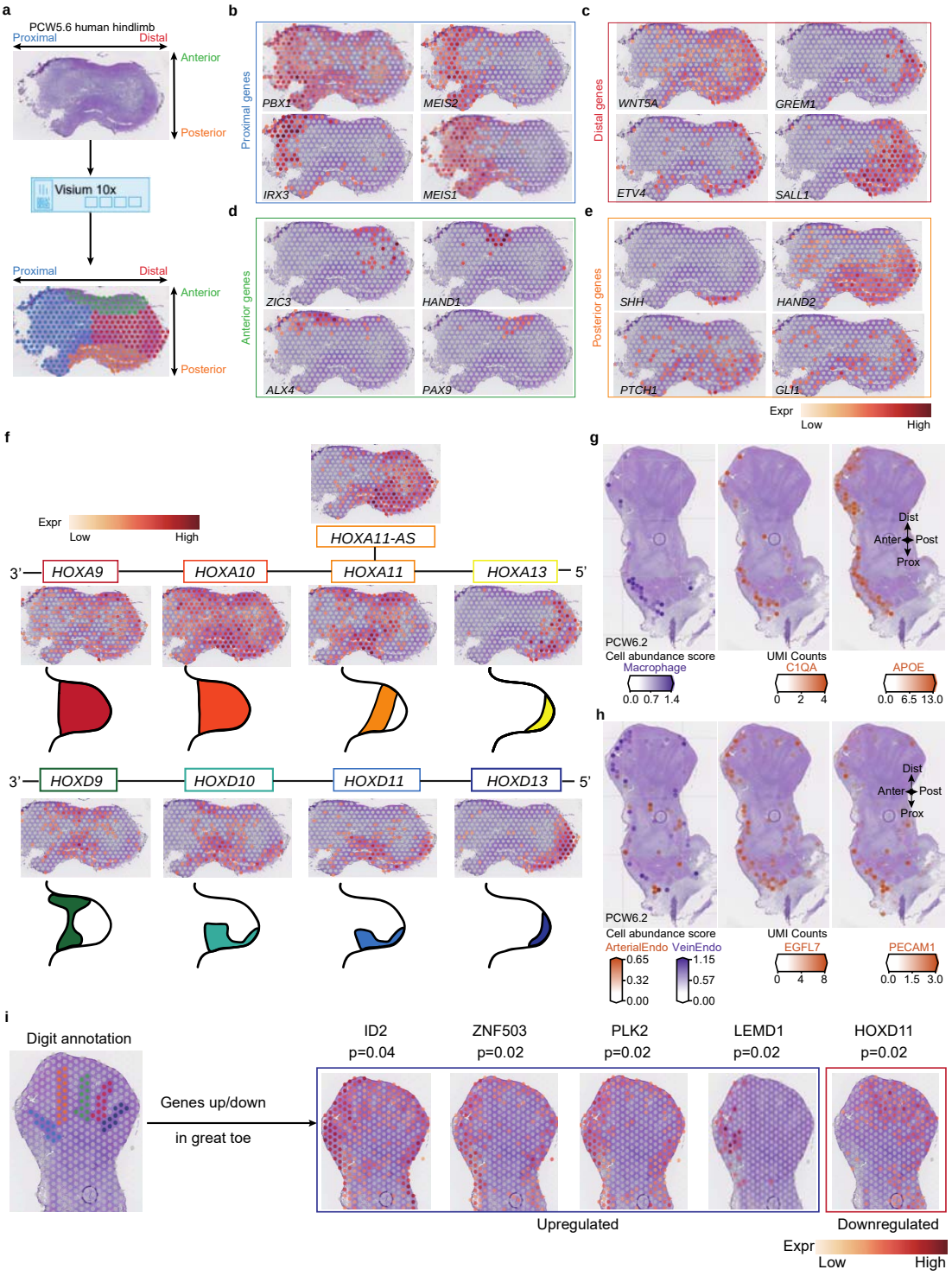


Extended Figure 5. Experimental validation of cell types and corresponding marker genes.

a, Immunofuorescence staining of RUNX2, THBS2 and COL2A1 on the longitudinal section of the tibia from a PCW7 embryo. Scale bar, 50 μm. **b**, Uniform manifold

approximation and projection (UMAP) visualisation of perineural fibroblasts (PeriNeuFiB); **c**, Violin plot showing the log-normalised expression of FOXS1, PI16, FGF19, RGS5 and ACTA2 in smooth muscle progenitors (SmMP), PeriNeuFiB and smooth muscle cells (SmMC), coloured by cell type. **d**, Immunohistochemical staining of PI16, FGF19 showing the perineural fibroblasts in the sciatic nerve at PCW9. The neurofilament was stained with NEFH antibody. A neighbouring section stained with H&E solution is also shown. Scale bar, 50 μm. **e**, Spatially resolved heatmaps across tissue sections from the PCW 6.2 human hindlimb showing spatial distribution of dermal fibroblast (DermFiB), HOXC5 positive fibroblast progenitor ($HOXC5^+$ FiBP) and F10 positive fibroblast progenitor ($F10^+$FiBP) as well as the spatial expression of KRT15. Prox, proximal; Dist, distal; Ante, anterior; Post, posterior. **f**, Spatially resolved heatmaps across tissue sections from the PCW 8.1 human hindlimb showing spatial distribution of ADH positive fibroblasts (ADH^+ FiB). **g**, Spatially resolved heatmap across tissue section from the PCW 8.1 human hindlimb showing spatial normalised expression of ALDH1A3 and MYH3. **h**, Immunofluorescence staining of MYH3 and ALDH1A3 on the skeletal muscle tissue (as also shown by H&E staining) from a PCW9 longitudinal section. Scale bar, 50 μm. **i**, UMAP visualisation of tenocytes and perimysium cells. **j**, Violin plot showing the log-normalised expression of SCX, TNMD, GCG, BGN and KERA in tenocyte progenitors (TenoP), Perimysium cells and tenocyte, coloured by cell type.

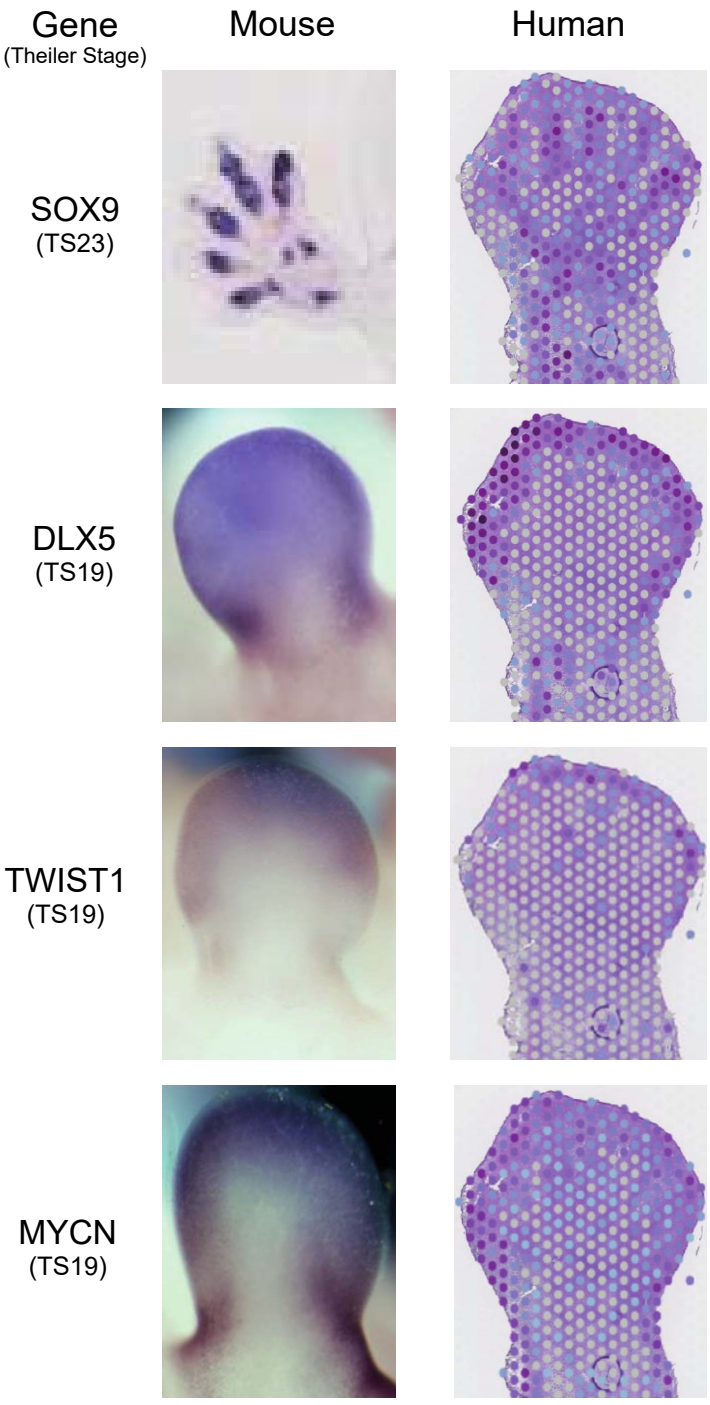
1179



1180

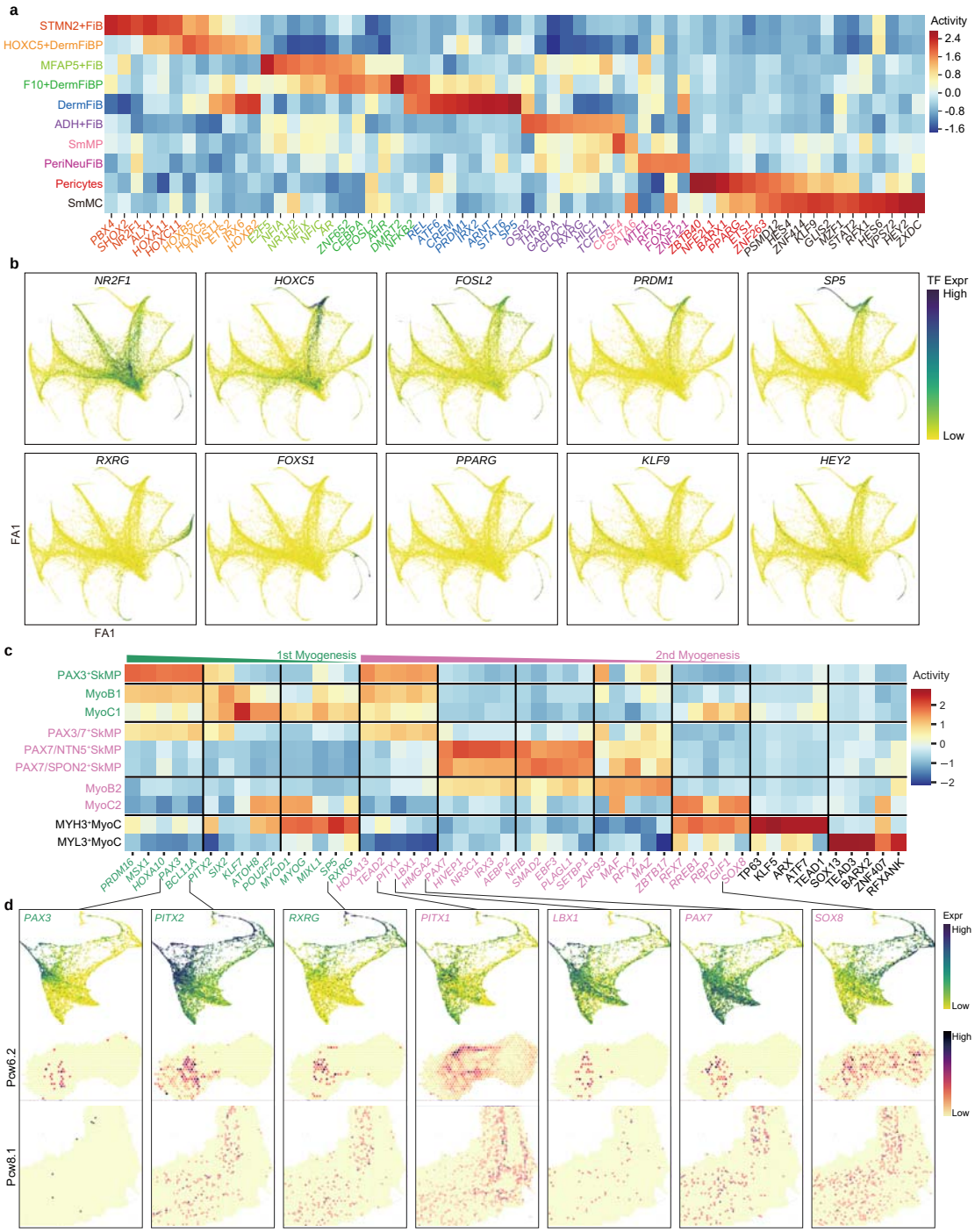
Extended Figure 6. Spatial expression patterns of genes that determine human limb axis formation and morphogenesis.

a, Overview of analysis pipeline to identify genes specific to spatial location. **b-e**, Spatially resolved heatmaps across tissue sections from the PCW 5.6 human hindlimb showing spatial normalised expression of genes that are specific to the proximal (**b**), distal (**c**), anterior (**d**) and posterior (**e**) axes. **f**, Spatially resolved heatmaps across tissue section from the PCW 5.6 human hindlimb showing spatial normalised expression of homeobox (HOX) A (top panel) and D (bottom panel) family genes. **g-h**, Spatially resolved heatmaps across tissue sections from the PCW 6.2 human hindlimb showing spatial distribution of macrophage (**g**) and endothelial cells (vein endothelial cells (VeinEndo) and arterial endothelial cells (ArterialEndo), (**h**) and expression of marker genes for macrophage and endothelial cells. Anter, anterior; Post, posterior; Prox, proximal; Dist, distal. **i**, Spatially resolved heatmaps across tissue sections from the PCW 5.6 human hindlimb showing spatial normalised expression of genes associated with digital formation.



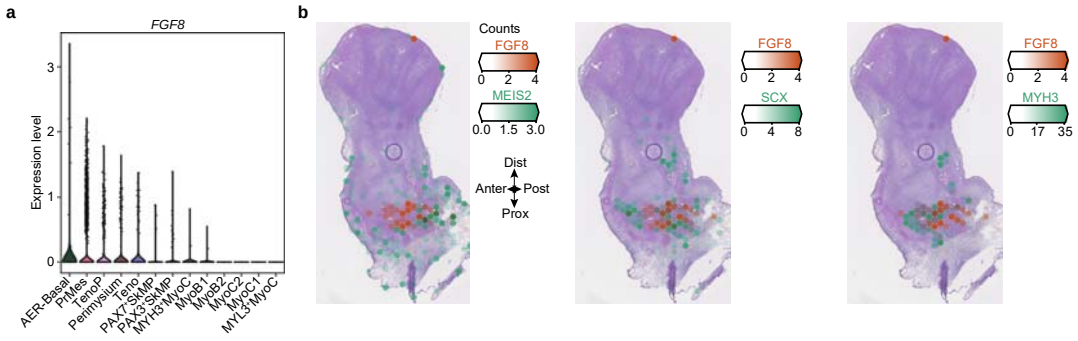
Extended Figure 7. Comparison of human and mouse expression of genes linked to altered digital phenotype.

Comparison of in-situ hybridisation (ISH) expression of disease-linked genes in the mouse embryo (where available) to visium expression data of human embryo at PCW6.2. Mouse Theiler stage given in parentheses.



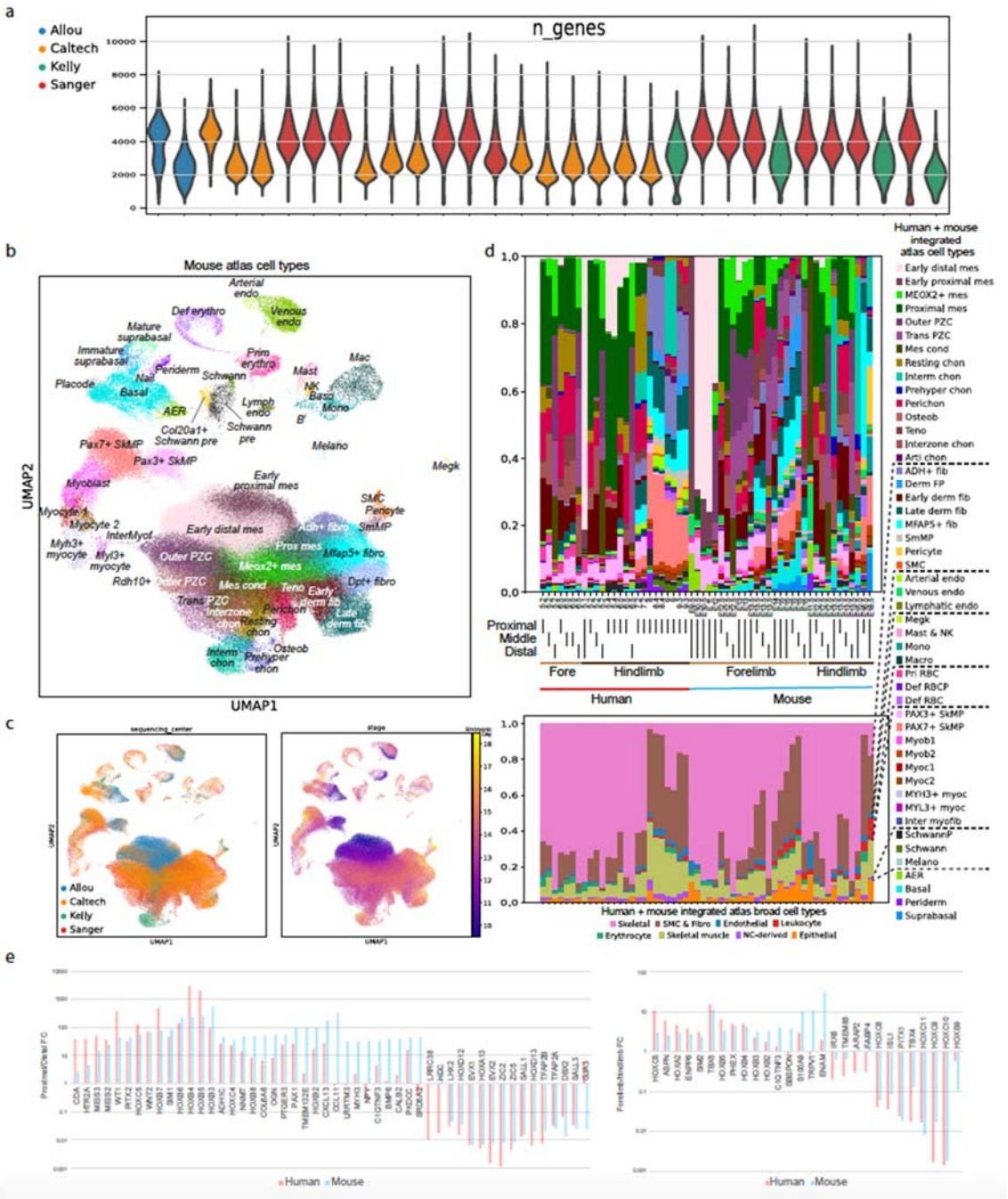
Extended Figure 8. transcription factors (TFs) conversion of soft connective lineage derived from lateral plate mesoderm (LPM) during human embryonic limb development.

a, Heat map illustrating the column-normalised mean activity of top5 TFs for each cell type from soft connective lineage. FiB, fibroblast; DermFiB, dermal FiB; DermFiBP, dermal FiB progenitor; SmMP, smooth muscle progenitor; PeriNeuFiB, perineural FiB; SmMC, smooth muscle cell. **b**, Force-directed graph layout showing the log-normalised expression of selected TFs at single-cell resolution. Heat map illustrating the column-normalised mean activity of top5 TFs for each cell type of myogenesis. SkMP, skeletal muscle progenitor; MyoB, myoblast; MyoC, myocyte. **d**, Spatiotemporal expression patterns of selected TFs. Top panels; force-directed graph layout of cells coloured by expression level of selected transcription factors; bottom panels, tissue sections from the PCW 6.2 and PCW 8.1 human hindlimb showing spatial expression pattern of selected transcription factors.



Extended Figure 9. The expression pattern of FGF8.

a, Violin plots showing log-normalised expression of FGF8 in cell types. AER, apical ectoderm region; PrMes, proximal mesenchyme; TenoP, tenocyte progenitor; PZC, progress zone cell. SkMP, skeletal muscle progenitor; MyoB, myoblast; MyoC, myocyte. **b**，Spatially resolved heatmaps across tissue sections from the PCW 6.2 human hindlimb showing spatial co-expression of FGF8 with markers of PrMes (MEIS2), TenoP (SCX) and MyoC (MYH3). Anter, anterior; Post, posterior; Prox, proximal; Dist, distal.



Extended Figure 10. Comparing human and mouse embryonic limbs.

a, Violin plots of sample quality for all the scRNA-seq data in the integrated atlas, colored by study or sequencing center. **b-c,** The integrated mouse scRNA-seq data projected on a shared UMAP plane, colored by cell types and cell states (b) or metadata (c). **d,** Cell-type (upper) and broad cell-type (lower) proportions of each scRNA-seq library. **e**, Genes enriched in proximal/distal/forelimb/hindlimb segments in human and mouse.

References

1. Chevallier, A., Kieny, M. & Mauger, A. Limb-somite relationship: origin of the limb musculature. *J. Embryol. Exp. Morphol.* **41**, 245–258 (1977).
2. Christ, B., Jacob, H. J. & Jacob, M. Experimental analysis of the origin of the wing musculature in avian embryos. *Anat. Embryol.* **150**, 171–186 (1977).
3. Tabin, C. & Wolpert, L. Rethinking the proximodistal axis of the vertebrate limb in the molecular era. *Genes Dev.* **21**, 1433–1442 (2007).
4. Sun, X., Mariani, F. V. & Martin, G. R. Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature* **418**, 501–508 (2002).
5. Zuniga, A. Next generation limb development and evolution: old questions, new perspectives. *Development* **142**, 3810–3820 (2015).
6. Hawkins, M. B., Henke, K. & Harris, M. P. Latent developmental potential to form limb-like skeletal structures in zebrafish. *Cell* **184**, 899–911.e13 (2021).
7. Lopez-Rios, J. The many lives of SHH in limb development and evolution. *Semin. Cell Dev. Biol.* **49**, 116–124 (2016).
8. Petit, F., Sears, K. E. & Ahituv, N. Limb development: a paradigm of gene regulation. *Nat. Rev. Genet.* **18**, 245–258 (2017).
9. Moore, K. L., Persaud, T. V. N. & Torchia, M. G. *The Developing Human - E-Book: Clinically Oriented Embryology*. (Elsevier Health Sciences, 2018).
10. Wilkie, A. O. M. Why study human limb malformations? *J. Anat.* **202**, 27–35 (2003).
11. Kildisiute, G. *et al.* Tumor to normal single-cell mRNA comparisons reveal a pan-neuroblastoma cancer cell. *Sci Adv* **7**, (2021).
12. Elmentaite, R. *et al.* Cells of the human intestinal tract mapped across space

1271 and time. *Nature* **597**, 250–255 (2021).

1272 13. Elmentaite, R. *et al.* Single-Cell Sequencing of Developing Human Gut Reveals
1273 Transcriptional Links to Childhood Crohn's Disease. *Dev. Cell* **55**, 771–783.e5
1274 (2020).

1275 14. Garcia-Alonso, L. *et al.* Mapping the temporal and spatial dynamics of the
1276 human endometrium in vivo and in vitro. *bioRxiv* 2021.01.02.425073 (2021)
1277 doi:10.1101/2021.01.02.425073.

1278 15. Jardine, L. *et al.* Blood and immune development in human fetal bone marrow
1279 and Down syndrome. *Nature* **598**, 327–331 (2021).

1280 16. Behjati, S., Lindsay, S., Teichmann, S. A. & Haniffa, M. Mapping human
1281 development at single-cell resolution. *Development* **145**, (2018).

1282 17. Haniffa, M. *et al.* A roadmap for the Human Developmental Cell Atlas. *Nature*
1283 **597**, 196–205 (2021).

1284 18. Xi, H. *et al.* A Human Skeletal Muscle Atlas Identifies the Trajectories of Stem
1285 and Progenitor Cells across Development and from Human Pluripotent Stem
1286 Cells. *Cell Stem Cell* **27**, 158–176.e10 (2020).

1287 19. Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial
1288 transcriptomics. *Nat. Biotechnol.* 1–11 (2022).

1289 20. Satoda, M., Pierpont, M. E., Diaz, G. A., Bornemeier, R. A. & Gelb, B. D. Char
1290 syndrome, an inherited disorder with patent ductus arteriosus, maps to
1291 chromosome 6p12-p21. *Circulation* **99**, (1999).

1292 21. Díaz-Hernández, M. E., Bustamante, M., Galván-Hernández, C. I. & Chimal-
1293 Monroy, J. Irx1 and Irx2 Are Coordinately Expressed and Regulated by Retinoic
1294 Acid, TGFβ and FGF Signaling during Chick Hindlimb Development. *PLoS One*
1295 **8**, e58549 (2013).

22. Zülch, A., Becker, M. B. & Gruss, P. Expression pattern of Irx1 and Irx2 during mouse digit development. *Mech. Dev.* **106**, (2001).

23. Richardson, R. J. *et al.* Periderm prevents pathological epithelial adhesions during embryogenesis. *J. Clin. Invest.* **124**, 3891–3900 (2014).

24. Saunders, J. W., Jr. The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm. *J. Exp. Zool.* **108**, 363–403 (1948).

25. Towers, M. & Tickle, C. Growing models of vertebrate limb development. *Development* **136**, (2009).

26. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–1416 (1993).

27. Zúñiga, A., Haramis, A. P., McMahon, A. P. & Zeller, R. Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. *Nature* **401**, 598–602 (1999).

28. Towers, M., Mahood, R., Yin, Y. & Tickle, C. Integration of growth and specification in chick wing digit-patterning. *Nature* **452**, 882–886 (2008).

29. Ballard, K. J. & Holt, S. J. Cytological and cytochemical studies on cell death and digestion in the foetal rat foot: the role of macrophages and hydrolytic enzymes. *J. Cell Sci.* **3**, 245–262 (1968).

30. Hernández-Martínez, R., Castro-Obregón, S. & Covarrubias, L. Progressive interdigital cell death: regulation by the antagonistic interaction between fibroblast growth factor 8 and retinoic acid. *Development* **136**, (2009).

31. Salas-Vidal, E., Valencia, C. & Covarrubias, L. Differential tissue growth and patterns of cell death in mouse limb autopod morphogenesis. *Dev. Dyn.* **220**, (2001).

32. Capdevila, J., Tsukui, T., Rodríquez, E. C., Zappavigna, V. & Jc, I. B. Control of

vertebrate limb outgrowth by the proximal factor Meis2 and distal antagonism of BMPs by Gremlin. *Mol. Cell* **4**, (1999).

33. Penkov, D. *et al.* Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with Hox genes/proteins. *Cell Rep.* **3**, (2013).
34. Delgado, I. *et al.* Control of mouse limb initiation and antero-posterior patterning by Meis transcription factors. *Nat. Commun.* **12**, 1–13 (2021).
35. Formation of Proximal and Anterior Limb Skeleton Requires Early Function of Irx3 and Irx5 and Is Negatively Regulated by Shh Signaling. *Dev. Cell* **29**, 233–240 (2014).
36. Yamaguchi, T. P., Bradley, A., McMahon, A. P. & Jones, S. A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* **126**, (1999).
37. Guo-hao Lin, L. Z. Apical ectodermal ridge regulates three principal axes of the developing limb. *J. Zhejiang Univ. Sci. B* **21**, 757 (2020).
38. Mao, J., McGlinn, E., Huang, P., Tabin, C. J. & McMahon, A. P. Fgf-dependent Etv4/5 activity is required for posterior restriction of Sonic Hedgehog and promoting outgrowth of the vertebrate limb. *Dev. Cell* **16**, (2009).
39. Kawakami, Y. *et al.* Sall genes regulate region-specific morphogenesis in the mouse limb by modulating Hox activities. *Development* **136**, 585 (2009).
40. Fernandez-Teran, M. *et al.* Role of dHAND in the anterior-posterior polarization of the limb bud: implications for the Sonic hedgehog pathway. *Development* **127**, 2133–2142 (2000).
41. McGlinn, E. *et al.* Pax9 and Jagged1 act downstream of Gli3 in vertebrate limb development. *Mech. Dev.* **122**, (2005).

1346 42. Kuijper, S. *et al.* Function and regulation of Alx4 in limb development: complex
1347 genetic interactions with Gli3 and Shh. *Dev. Biol.* **285**, (2005).
1348 43. Quinn, M. E., Haaning, A. & Ware, S. M. Preaxial polydactyly caused by Gli3
1349 haploinsufficiency is rescued by Zic3 loss of function in mice. *Hum. Mol. Genet.*
1350 **21**, (2012).
1351 44. Galli, A. *et al.* Distinct roles of Hand2 in initiating polarity and posterior Shh
1352 expression during the onset of mouse limb bud development. *PLoS Genet.* **6**,
1353 e1000901 (2010).
1354 45. Fuse, N. *et al.* Sonic hedgehog protein signals not as a hydrolytic enzyme but as
1355 an apparent ligand for Patched. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10992–10999
1356 (1999).
1357 46. Huangfu, D. & Anderson, K. V. Signaling from Smo to Ci/Gli: conservation and
1358 divergence of Hedgehog pathways from Drosophila to vertebrates. *Development*
1359 **133**, 3–14 (2006).
1360 47. The role of Hox genes during vertebrate limb development. *Curr. Opin. Genet.*
1361 *Dev.* **17**, 359–366 (2007).
1362 48. Tarchini, B. & Duboule, D. Control of Hoxd genes' collinearity during early limb
1363 development. *Dev. Cell* **10**, 93–103 (2006).
1364 49. Dollé, P., Izpisúa-Belmonte, J. C., Falkenstein, H., Renucci, A. & Duboule, D.
1365 Coordinate expression of the murine Hox-5 complex homoeobox-containing
1366 genes during limb pattern formation. *Nature* **342**, 767–772 (1989).
1367 50. Kherdjemil, Y. *et al.* Evolution of Hoxa11 regulation in vertebrates is linked to the
1368 pentadactyl state. *Nature* **539**, (2016).
1369 51. Díaz-Hernández, M. E., Rios-Flores, A. J., Abarca-Buis, R. F., Bustamante, M.
1370 & Chimal-Monroy, J. Molecular Control of Interdigital Cell Death and Cell

Differentiation by Retinoic Acid during Digit Development. *Journal of Developmental Biology* **2**, 138–157 (2014).

52. Ko, Y. B. *et al.* WIF1 can effectively co-regulate pro-apoptotic activity through the combination with DKK1. *Cell. Signal.* **26**, 2562–2572 (2014).
53. Cochain, C. *et al.* Single-Cell RNA-Seq Reveals the Transcriptional Landscape and Heterogeneity of Aortic Macrophages in Murine Atherosclerosis. *Circ. Res.* **122**, 1661–1674 (2018).
54. Williams, J. W. *et al.* Limited proliferation capacity of aortic intima resident macrophages requires monocyte recruitment for atherosclerotic plaque progression. *Nat. Immunol.* **21**, 1194–1204 (2020).
55. Turnbull, I. R. *et al.* Cutting edge: TREM-2 attenuates macrophage activation. *J. Immunol.* **177**, 3520–3524 (2006).
56. Lorda-Diez, C. I., Torre-Pérez, N., García-Porrero, J. A., Hurle, J. M. & Montero, J. A. Expression of Id2 in the developing limb is associated with zones of active BMP signaling and marks the regions of growth and differentiation of the developing digits. *Int. J. Dev. Biol.* **53**, (2009).
57. McGlinn, E. *et al.* Expression of the NET family member Zfp503 is regulated by hedgehog and BMP signaling in the limb. *Dev. Dyn.* **237**, 1172–1182 (2008).
58. Ma, S., Charron, J. & Erikson, R. L. Role of Plk2 (Snk) in Mouse Development and Cell Proliferation. *Mol. Cell. Biol.* **23**, 6936 (2003).
59. Sasahira, T., Kurihara, M., Nakashima, C., Kirita, T. & Kuniyasu, H. LEM domain containing 1 promotes oral squamous cell carcinoma invasion and endothelial transmigration. *Br. J. Cancer* **115**, 52–58 (2016).
60. Temtamy, S. A. & Aglan, M. S. Brachydactyly. *Orphanet J. Rare Dis.* **3**, 1–16 (2008).

1396 61. Falardeau, F., Camurri, M. V. & Campeau, P. M. Genomic approaches to
1397 diagnose rare bone disorders. *Bone* **102**, 5–14 (2017).
1398 62. Cooks, R. G., Hertz, M., Katznelson, M. B. & Goodman, R. M. A new nail
1399 dysplasia syndrome with onychonychia and absence and/or hypoplasia of distal
1400 phalanges. *Clin. Genet.* **27**, (1985).
1401 63. Bahubali D. Gane, P. N. Split-hand/feet malformation: A rare syndrome. *Journal*
1402 *of Family Medicine and Primary Care* **5**, 168 (2016).
1403 64. Marcelis, C. L. M. & de Brouwer, A. P. M. Feingold Syndrome 1. in
1404 *GeneReviews® [Internet]* (University of Washington, Seattle, 2019).
1405 65. Mutations in BMP4 Cause Eye, Brain, and Digit Developmental Anomalies:
1406 Overlap between the BMP4 and Hedgehog Signaling Pathways. *Am. J. Hum.*
1407 *Genet.* **82**, 304–319 (2008).
1408 66. El Ghouzzi, V. *et al.* Saethre-Chotzen mutations cause TWIST protein
1409 degradation or impaired nuclear location. *Hum. Mol. Genet.* **9**, (2000).
1410 67. Richardson, L. *et al.* EMAGE mouse embryo spatial gene expression database:
1411 2014 update. *Nucleic Acids Res.* **42**, D835–44 (2014).
1412 68. Gao, B. *et al.* A mutation in Ihh that causes digit abnormalities alters its
1413 signalling capacity and range. *Nature* vol. 458 1196–1200 (2009).
1414 69. Yi, S. E., Daluiski, A., Pederson, R., Rosen, V. & Lyons, K. M. The type I BMP
1415 receptor BMPRIB is required for chondrogenesis in the mouse limb.
1416 *Development* vol. 127 621–630 (2000).
1417 70. Stafford, D. A., Brunet, L. J., Khokha, M. K., Economides, A. N. & Harland, R. M.
1418 Cooperative activity of noggin and gremlin 1 in axial skeleton development.
1419 *Development* **138**, 1005–1014 (2011).
1420 71. Brunet, L. J., McMahon, J. A., McMahon, A. P. & Harland, R. M. Noggin,

Cartilage Morphogenesis, and Joint Formation in the Mammalian Skeleton. *Science* vol. 280 1455–1457 (1998).

72. Melkoniemi, M. *et al.* Autosomal recessive disorder otospondylomegaepiphyseal dysplasia is associated with loss-of-function mutations in the COL11A2 gene. *Am. J. Hum. Genet.* **66**, 368–377 (2000).
73. Li, S. W. *et al.* Targeted disruption of Col11a2 produces a mild cartilage phenotype in transgenic mice: comparison with the human disorder otospondylomegaepiphyseal dysplasia (OSMED). *Dev. Dyn.* **222**, 141–152 (2001).
74. Bi, W. *et al.* Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6698–6703 (2001).
75. Shazeeb, M. S. *et al.* Skeletal Characterization of the Fgfr3 Mouse Model of Achondroplasia Using Micro-CT and MRI Volumetric Imaging. *Scientific Reports* vol. 8 (2018).
76. Carey, J. C., Cassidy, S. B., Battaglia, A. & Viskochil, D. *Cassidy and Allanson's Management of Genetic Syndromes*. (John Wiley & Sons, 2021).
77. Sowińska-Seidler, A., Socha, M. & Jamsheer, A. Split-hand/foot malformation - molecular cause and implications in genetic counseling. *J. Appl. Genet.* **55**, 105–115 (2014).
78. Merlo, G. R. *et al.* Mouse model of split hand/foot malformation type I. *Genesis* **33**, 97–101 (2002).
79. Celli, J., van Bokhoven, H. & Brunner, H. G. Feingold syndrome: Clinical review and genetic mapping. *American Journal of Medical Genetics* vol. 122A 294–300 (2003).

80. Sawai, S. *et al.* Defects of embryonic organogenesis resulting from targeted disruption of the N-myc gene in the mouse. *Development* **117**, 1445–1455 (1993).

81. Selever, J., Liu, W., Lu, M.-F., Behringer, R. R. & Martin, J. F. Bmp4 in limb bud mesoderm regulates digit pattern by controlling AER development. *Developmental Biology* vol. 276 268–279 (2004).

82. Gripp, K. W., Zackai, E. H. & Stolle, C. A. Mutations in the human TWIST gene. *Hum. Mutat.* **15**, 479 (2000).

83. Bialek, P. *et al.* A Twist Code Determines the Onset of Osteoblast Differentiation. *Developmental Cell* vol. 6 423–435 (2004).

84. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

85. Marcil, A., Dumontier, E., Chamberland, M., Camper, S. A. & Drouin, J. Pitx1 and Pitx2 are required for development of hindlimb buds. *Development* **130**, 45–55 (2003).

86. Dunwoodie, S. L., Rodriguez, T. A. & Beddington, R. S. Msg1 and Mrg1, founding members of a gene family, show distinct patterns of gene expression during mouse embryogenesis. *Mech. Dev.* **72**, 27–40 (1998).

87. Hng, C. H. *et al.* HOPX regulates bone marrow-derived mesenchymal stromal cell fate determination via suppression of adipogenic gene pathways. *Sci. Rep.* **10**, 1–14 (2020).

88. Hu, W. Y. *et al.* Isolation and functional interrogation of adult human prostate epithelial stem cells at single cell resolution. *Stem Cell Res.* **23**, (2017).

89. Distinct populations within Isl1 lineages contribute to appendicular and facial skeletogenesis through the β-catenin pathway. *Dev. Biol.* **387**, 37–48 (2014).

90. Narkis, G. *et al.* Isl1 and Ldb co-regulators of transcription are essential early determinants of mouse limb development. *Dev. Dyn.* **241**, (2012).

91. Regionalized Twist1 activity in the forelimb bud drives the morphogenesis of the proximal and preaxial skeleton. *Dev. Biol.* **362**, 132–140 (2012).

92. Reijntjes, S., Stricker, S. & Mankoo, B. S. A comparative analysis of Meox1 and Meox2 in the developing somites and limbs of the chick embryo. *Int. J. Dev. Biol.* **51**, (2007).

93. Feregrino, C., Sacher, F., Parnas, O. & Tschopp, P. A single-cell transcriptomic atlas of the developing chicken limb. *BMC Genomics* **20**, 1–15 (2019).

94. Polisetti, N. *et al.* Gene expression profile of epithelial cells and mesenchymal cells derived from limbal explant culture. *Mol. Vis.* **16**, 1227 (2010).

95. Kozhemyakina, E., Ionescu, A. & Lassar, A. B. GATA6 is a crucial regulator of Shh in the limb bud. *PLoS Genet.* **10**, (2014).

96. Vickerman, L., Neufeld, S. & Cobb, J. Shox2 function couples neural, muscular and skeletal development in the proximal forelimb. *Dev. Biol.* **350**, (2011).

97. Tzchori, I. *et al.* LIM homeobox transcription factors integrate signaling events that control three-dimensional limb patterning and growth. *Development* **136**, 1375–1385 (2009).

98. Aza-Carmona, M. *et al.* SHOX interacts with the chondrogenic transcription factors SOX5 and SOX6 to activate the aggrecan enhancer. *Hum. Mol. Genet.* **20**, 1547–1559 (2011).

99. Rao, E. *et al.* Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat. Genet.* **16**, 54–63 (1997).

100. Desjardin, C. *et al.* Chondrocytes Play a Major Role in the Stimulation of Bone

Growth by Thyroid Hormone. *Endocrinology* **155**, 3123–3135 (2014).

101. Kawato, Y. *et al.* Nkx3.2 promotes primary chondrogenic differentiation by upregulating Col2a1 transcription. *PLoS One* **7**, (2012).

102. Venugopalan, S. R. *et al.* Hierarchical interactions of homeodomain and forkhead transcription factors in regulating odontogenic gene expression. *J. Biol. Chem.* **286**, 21372–21383 (2011).

103. Dreher, S. I., Fischer, J., Walker, T., Diederichs, S. & Richter, W. Significance of MEF2C and RUNX3 Regulation for Endochondral Differentiation of Human Mesenchymal Progenitor Cells. *Front Cell Dev Biol* **8**, 81 (2020).

104. Ghoul-Mazgar, S. *et al.* Expression pattern of Dlx3 during cell differentiation in mineralized tissues. *Bone* **37**, 799–809 (2005).

105. Woods, A., James, C. G., Wang, G., Dupuis, H. & Beier, F. Control of chondrocyte gene expression by actin dynamics: a novel role of cholesterol/Ror-alpha signalling in endochondral bone growth. *J. Cell. Mol. Med.* **13**, 3497–3516 (2009).

106. Yan, J. *et al.* Smad4 deficiency impairs chondrocyte hypertrophy via the Runx2 transcription factor in mouse skeletal development. *J. Biol. Chem.* **293**, 9162–9175 (2018).

107. Wang, X. *et al.* TFAP2C promotes stemness and chemotherapeutic resistance in colorectal cancer via inactivating hippo signaling pathway. *J. Exp. Clin. Cancer Res.* **37**, 1–16 (2018).

108. Otero, M. *et al.* ELF3 modulates type II collagen gene (COL2A1) transcription in chondrocytes by Inhibiting SOX9-CBP/p300-driven histone acetyltransferase activity. *Connect. Tissue Res.* **58**, 15 (2017).

109. Glucocorticoids Suppress Bone Formation by Attenuating Osteoblast

Differentiation via the Monomeric Glucocorticoid Receptor. *Cell Metab.* **11**, 517–531 (2010).

110. Glucocorticoid receptor isoforms generate transcription specificity. *Trends Cell Biol.* **16**, 301–307 (2006).

111. Pitsillides, A. A. & Beier, F. Keep your Sox on, chondrocytes! *Nat. Rev. Rheumatol.* **17**, 383–384 (2021).

112. Henry, S. P., Liang, S., Akdemir, K. C. & de Crombrugghe, B. The postnatal role of Sox9 in cartilage. *J. Bone Miner. Res.* **27**, 2511–2525 (2012).

113. Chen, X. *et al.* Force and scleraxis synergistically promote the commitment of human ES cells derived MSCs to tenocytes. *Sci. Rep.* **2**, 1–9 (2012).

114. Komura, S. *et al.* Induced pluripotent stem cell-derived tenocyte-like cells promote the regeneration of injured tendons in mice. *Sci. Rep.* **10**, 1–12 (2020).

115. Fukuda, S. *et al.* Sustained expression of HeyL is critical for the proliferation of muscle stem cells in overloaded muscle. *Elife* **8**, (2019).

116. Ambrosino, C. *et al.* TEF-1 and C/EBPβ are major p38α MAPK-regulated transcription factors in proliferating cardiomyocytes. *Biochem. J* **396**, 163 (2006).

117. Piragyte, I. *et al.* A metabolic interplay coordinated by HLX regulates myeloid differentiation and AML through partly overlapping pathways. *Nat. Commun.* **9**, 1–17 (2018).

118. Budnick, I. *et al.* Defining the identity of mouse embryonic dermal fibroblasts. *Genesis* **54**, 415 (2016).

119. Chiang, M. F. *et al.* Inducible deletion of the Blimp-1 gene in adult epidermis causes granulocyte-dominated chronic skin inflammation in mice. *Proc. Natl. Acad. Sci. U. S. A.* **110**, (2013).

120. Reich, N. *et al.* The transcription factor Fra-2 regulates the production of extracellular matrix in systemic sclerosis. *Arthritis Rheum.* **62**, (2010).

121. Montelius, A. *et al.* Emergence of the sensory nervous system as defined by Foxs1 expression. *Differentiation* **75**, (2007).

122. Lepore, J. J., Cappola, T. P., Mericko, P. A., Morrisey, E. E. & Parmacek, M. S. GATA-6 regulates genes promoting synthetic functions in vascular smooth muscle cells. *Arterioscler. Thromb. Vasc. Biol.* **25**, (2005).

123. Sakata, Y. *et al.* Transcription factor CHF1/Hey2 regulates neointimal formation in vivo and vascular smooth muscle proliferation and migration in vitro. *Arterioscler. Thromb. Vasc. Biol.* **24**, (2004).

124. Klose, R. *et al.* Loss of the serine protease HTRA1 impairs smooth muscle cells maturation. *Sci. Rep.* **9**, 1–10 (2019).

125. Buckingham, M. & Rigby, P. W. J. Gene regulatory networks and transcriptional mechanisms that control myogenesis. *Dev. Cell* **28**, 225–238 (2014).

126. Ontell, M. & Kozeka, K. The organogenesis of murine striated muscle: a cytoarchitectural study. *Am. J. Anat.* **171**, 133–148 (1984).

127. Buckingham, M. *et al.* The formation of skeletal muscle: from somite to limb. *J. Anat.* **202**, 59–68 (2003).

128. Hutcheson, D. A., Zhao, J., Merrell, A. & Haldar, M. Embryonic and fetal limb myogenic cells are derived from developmentally distinct progenitors and have different requirements for β-catenin. *Genes* (2009).

129. Singh, A. J. *et al.* FACS-Seq analysis of Pax3-derived cells identifies non-myogenic lineages in the embryonic forelimb. *Sci. Rep.* **8**, 7670 (2018).

130. Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L. & Weintraub, H. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell* **61**,

49–59 (1990).

131. Roschger, C. & Cabrele, C. The Id-protein family in developmental and cancer-associated pathways. *Cell Commun. Signal.* **15**, 7 (2017).

132. Muriel, J. M. *et al.* Keratin 18 is an integral part of the intermediate filament network in murine skeletal muscle. *Am. J. Physiol. Cell Physiol.* **318**, (2020).

133. He, P. *et al.* The changing mouse embryo transcriptome at whole tissue and single-cell resolution. *Nature* **583**, 760–767 (2020).

134. Hernandez-Torres, F., Rodríguez-Outeiriño, L., Franco, D. & Aranega, A. E. Pitx2 in Embryonic and Adult Myogenesis. *Front Cell Dev Biol* **5**, 46 (2017).

135. Lu, J., Webb, R., Richardson, J. A. & Olson, E. N. MyoR: A muscle-restricted basic helix–loop–helix transcription factor that antagonizes the actions of MyoD. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 552–557 (1999).

136. MacQuarrie, K. L., Yao, Z., Fong, A. P. & Tapscott, S. J. Genome-wide binding of the basic helix-loop-helix myogenic inhibitor musculin has substantial overlap with MyoD: implications for buffering activity. *Skelet. Muscle* **3**, 26 (2013).

137. Buas, M. F., Kabak, S. & Kadesch, T. Inhibition of myogenesis by Notch: evidence for multiple pathways. *J. Cell. Physiol.* **218**, 84–93 (2009).

138. Chiu, Y.-K. *et al.* Transcription factor ABF-1 suppresses plasma cell differentiation but facilitates memory B cell formation. *J. Immunol.* **193**, 2207–2217 (2014).

139. Massari, M. E. *et al.* Characterization of ABF-1, a novel basic helix-loop-helix transcription factor expressed in activated B lymphocytes. *Mol. Cell. Biol.* **18**, 3130–3139 (1998).

140. 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**, 1484–1506 (2020).

141. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 (2018).

142. Schroeter, E. H., Kisslinger, J. A. & Kopan, R. Notch-1 signalling requires ligand-induced proteolytic release of intracellular domain. *Nature* **393**, 382–386 (1998).

143. D'Souza, B., Miyamoto, A. & Weinmaster, G. The many facets of Notch ligands. *Oncogene* **27**, 5148–5167 (2008).

144. Canonical Notch ligands and Fringes have distinct effects on NOTCH1 and NOTCH2. *J. Biol. Chem.* **295**, 14710–14722 (2020).

145. Crosnier, C. *et al.* JAGGED1 gene expression during human embryogenesis elucidates the wide phenotypic spectrum of Alagille syndrome. *Hepatology* **32**, (2000).

146. Mašek, J. & Andersson, E. R. The developmental biology of genetic Notch disorders. *Development* **144**, 1743–1763 (2017).

147. Turnpenny, P. D. & Ellard, S. Alagille syndrome: pathogenesis, diagnosis and management. *Eur. J. Hum. Genet.* **20**, 251–257 (2011).

148. Xu, X. *et al.* Fibroblast growth factor receptor 2 (FGFR2)-mediated reciprocal regulation loop between FGF8 and FGF10 is essential for limb induction. *Development* (1998).

149. Azoury, S. C., Reddy, S., Shukla, V. & Deng, C.-X. Fibroblast Growth Factor Receptor 2 (FGFR2) Mutation Related Syndromic Craniosynostosis. *Int. J. Biol. Sci.* **13**, 1479 (2017).

150. Marics, I., Padilla, F., Guillemot, J.-F., Scaal, M. & Marcelle, C. FGFR4 signaling is a necessary step in limb muscle differentiation. *Development* **129**, 4559–4569

(2002).

151. Verheyden, J. M., Lewandoski, M., Deng, C., Harfe, B. D. & Sun, X. Conditional inactivation of Fgfr1 in mouse defines its role in limb bud establishment, outgrowth and digit patterning. *Development* **132**, 4235 (2005).

152. Su, N., Jin, M. & Chen, L. Role of FGF/FGFR signaling in skeletal development and homeostasis: learning from mouse models. *Bone research* **2**, (2014).

153. 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 Biology* vol. 89 1–10 (2020).

154. Allou, L. *et al.* Non-coding deletions identify Maenli lncRNA as a limb-specific En1 regulator. *Nature* **592**, 93–98 (2021).

155. Jain, M. S. *et al.* MultiMAP: dimensionality reduction and integration of multimodal data. *Genome Biol.* **22**, 1–26 (2021).

156. Feregrino, C. & Tschopp, P. Assessing evolutionary and developmental transcriptome dynamics in homologous cell types. *Dev. Dyn.* (2021) doi:10.1002/dvdy.384.

157. Popescu, D. M. *et al.* Decoding human fetal liver haematopoiesis. *Nature* **574**, (2019).

158. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

159. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* vol. 9 676–682 (2012).

160. Van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).

161. Wang, S. *et al.* Muscle Stem Cell Immunostaining. *Curr. Protoc. Mouse Biol.* **8**,

e47 (2018).

162. Lapan, A. D. & Gussoni, E. Isolation and characterization of human fetal myoblasts. *Methods Mol. Biol.* **798**, 3–19 (2012).

Data Availability

All of our newly generated raw data are publicly available on ArrayExpress (mouse scRNA-seq, E-MTAB-10514; human Visium, E-MTAB-10367; human scRNA-seq, E-MTAB-8813). Previously published raw data can be found from ENCODE portal (ENCSR713GIS) and GEO (GSE137335 and GSE142425). Processed data and be downloaded and visualized at our data portal (https://limb-dev.cellgeni.sanger.ac.uk).

Code availability

All in-house codes can be found on github (https://github.com/Teichlab/limbcellatlas/).

Acknowledgements

We thank Jana Lalakova for illustrating human limb syndromes. We thank Ken To for proofreading the manuscript. We thank Matt Thomson's lab for help with mouse 10X loading. We thank members of the Teichmann lab, Zhang lab, Marioni lab, Haniffa lab and Behjati lab for discussion and feedback. This work was supported by the National Key Research and Development Program (grant 2019YFA0801703), National Natural Science Foundation of China (grant 31871370), Science and Technology Program of Guangzhou (grant 202002030429), and Advanced Medical Technology Center, The First Affiliated Hospital, Zhongshan School of Medicine, Sun Yat-sen University (to H.Z.); China Postdoctoral Science Foundation (grant 2021M700936), and Natural Science Foundation of Guangdong (grant 2019A1515011342) (to S.W.). P.H holds a non-stipendiary research fellowship at St Edmund's College, University of Cambridge. J.E.L is funded by the wellcome trust under the clinical PhD programme.

Author contributions

S.A.T. and H.Z. supervised the project; S.A.T. initiated and designed the project; X.H. and Y.F. carried out human tissue collection; B.W. carried out mouse tissue collection; B.W.,L.M., L.B., R.E., and E.F. performed scRNA-seq; E.T. performed Visium spatial experiments; S.W.,K.R., and E.T did in situ staining and functional experiments; H.Y., C.L. and H.Z. provided experimental support. B.Z.,P.H and J.E.L. analysed sequencing data and generated figures. V.K., K.P., M.P. and N.Y. provided computational support. M.S. and B.J.W.contributed to interpretation of the results. B.Z.,P.H, J.E.L., S.W., H.Z. and S.A.T. wrote the manuscript. All authors contributed to the discussion and editing of the manuscript.

Competing interests

In the past three years, S.A.T. has consulted for or been a member of scientific advisory boards at Roche, Qiagen, Genentech, Biogen, GlaxoSmithKline and ForeSite Labs. The remaining authors declare no competing interests.