Title

A multi-omic atlas of human embryonic skeletal development.

Permalink

Journal

Authors

To, Ken

Fei, Lijiang

Pett, J

et al.

Publication Date

DOI

Peer reviewed

Article

A multi-omic atlas of human embryonic skeletal development

Ken To[1,2,16], Lijiang Fei[1,16], J. Patrick Pett[1,16], Kenny Roberts[1], Raphael Blain[3], Krzysztof Polański[1], Tong Li[1], Nadav Yayon[1,4], Peng He[1,4,5], Chuan Xu[1], James Cranley[1,6], Madelyn Moy[1], Ruoyan Li[1], Kazumasa Kanemaru[1], Ni Huang[1], Stathis Megas[1,7], Laura Richardson[1], Rakesh Kapuge[1], Shani Perera[1], Elizabeth Tuck[1], Anna Wilbrey-Clark[1], Ilaria Mulas[1], Fani Memi[1], Batuhan Cakir[1], Alexander V. Predeus[1], David Horsfall[1], Simon Murray[1], Martin Prete[1], Pavel Mazin[1], Xiaoling He[8,9], Kerstin B. Meyer[1], Muzlifah Haniffa[1,10,11], Roger A. Barker[8,9], Omer Bayraktar[1], Alain Chédotal[3,12,13], Christopher D. Buckley[14] & Sarah A. Teichmann[1,6,7,9,15 ✉]
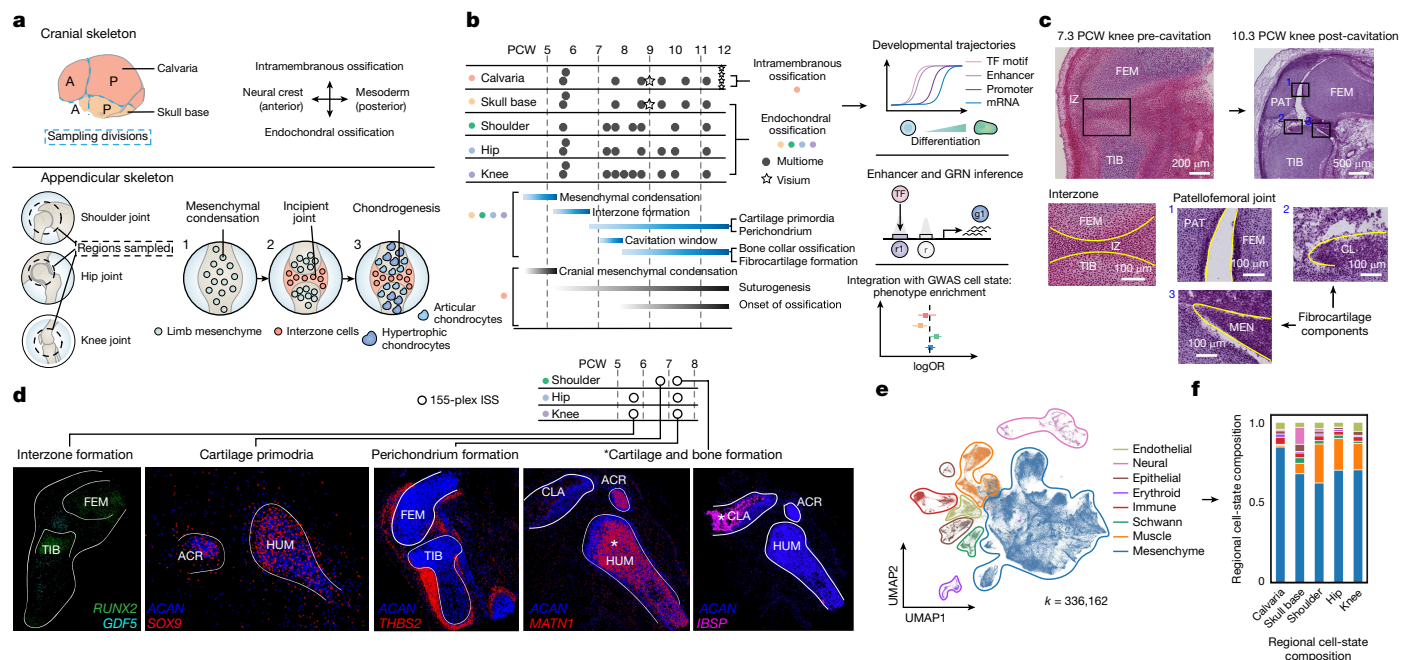
Human embryonic bone and joint formation is determined by coordinated differentiation of progenitors in the nascent skeleton. The cell states, epigenetic processes and key regulatory factors that underlie lineage commitment of these cells remain elusive. Here we applied paired transcriptional and epigenetic profiling of approximately 336,000 nucleus droplets and spatial transcriptomics to establish a multi-omic atlas of human embryonic joint and cranium development between 5 and 11 weeks after conception. Using combined modelling of transcriptional and epigenetic data, we characterized regionally distinct limb and cranial osteoprogenitor trajectories across the embryonic skeleton and further described regulatory networks that govern intramembranous and endochondral ossification. Spatial localization of cell clusters in our in situ sequencing data using a new tool, ISS-Patcher, revealed mechanisms of progenitor zonation during bone and joint formation. Through trajectory analysis, we predicted potential non-canonical cellular origins for human chondrocytes from Schwann cells. We also introduce SNP2Cell, a tool to link cell-type-specific regulatory networks to polygenic traits such as osteoarthritis. Using osteolineage trajectories characterized here, we simulated in silico perturbations of genes that cause monogenic craniosynostosis and implicate potential cell states and disease mechanisms. This work forms a detailed and dynamic regulatory atlas of bone and cartilage maturation and advances our fundamental understanding of cell-fate determination in human skeletal development.

Human bone development begins between 6 and 8 weeks after conception (post-conception weeks, PCW) during the transition from embryonic to fetal stages. In the cranium, calvarial progenitors differentiate into osteoblasts through intramembranous ossification and continue to house osteoprogenitors postnatally[1,2]. In the nascent synovial joint, an interzone condensation appears in the limb bud at 5–6 PCW[3] and forms a joint cavity between 7 and 8 PCW, varying in timing across joints, within which fibrous and ligamentous structures develop[4,5] (Fig. 1b,c). Cartilage scaffolds form on either side of synovial joints to facilitate development of the body plane until they are replaced by bone tissue as endochondral ossification ensues from 8 PCW[6,7]. These regionally distinct modes of ossification govern osteogenesis throughout the

human skeleton. To our knowledge, the cellular basis by which they form and mature remain incompletely described in human development at single-cell resolution. To address this, we applied single-nucleus paired RNA (snRNA) and assay for transposase-accessible chromatin (snATAC) sequencing (seq), and spatial methods, to decipher the regulatory landscape that mediates maturation of the distinct bone-forming and joint-forming niches in the embryonic cranium and limbs from 5 to 11 PCW. Through this, we uncovered previously undescribed cellular diversity in the osteogenic and chondrogenic lineages. We developed ISS-Patcher, a tool to impute cell labels from the droplet data on our high-resolution 155-plex in situ sequencing (ISS) datasets, which facilitated insights into spatially defined niches within the embryonic

Article

Fig. 1 | Multi-omics of human embryonic skeletal development. a, Anatomical sampling approach for the five main anatomical regions. Top, origin cells within the cranium are determined by the anterior (A)–posterior (P) axis. We sampled according to these canonical divisions. Incisions were made as shown by the blue dotted lines (top). Appendicular regions sampled are shown within the black dotted lines (bottom). Bottom, the diagram portrays (1) initial condensation of the mesenchyme, (2) joint-site determination by interzone cells in the incipient joint, and (3) chondrogenesis occurring within the region of the joint. **b**, Sample donor (n = 12) overview across age and anatomical regions sampled; atlasing modalities are represented in the legend. The timeline of morphological events and changes across the timeline sampled (bottom left) are shown; key analyses undertaken with Multiome data generated from this work are also displayed (right). g1, gene; r, region; TF, transcription factor. **c**, Haematoxylin and eosin staining of tissue sections of the knee joint to illustrate features before and after cavitation. The interzone (IZ) is present before cavitation and this is separated by the joint space following cavitation with emergence of soft-tissue structures including ligaments. CL, cruciate ligament; FEM, femur; MEN, meniscus; PAT, patella; TIB, tibia. Images are representative of sections from two donors. **d**, Marker genes detected by ISS probes applied to the tissue samples indicated in the schematic. The white lines outline regions of bone. The asterisks show cartilage in the humerus (HUM) and bone in the clavicle (CLA). ACR, acromion. Images are representative of three donors. **e**, Uniform manifold approximation and projection (UMAP) embedding of the dataset using MultiVI latent variables calculated from snRNA and snATAC data. The colours indicate cell cluster compartments. **f**, Relative cell-type abundance across anatomical locations. The bar plot of the proportion of the cell cluster compartment in each anatomical region sampled shows predominance of the mesenchyme compartment across anatomical regions.

synovial joint. Applying OrganAxis[8], a new spatial transcriptomics annotation tool, we also define the spatial trajectory of the developing cranial bone. We characterized novel cell states of the craniofacial region and additionally delineated processes of human Schwann cell and fibroblast development. Our resource and new computational toolset, including SNP2Cell, enabled predictions of the mechanisms of developmental conditions, such as craniosynostosis[9–11], and allowed association of gene-regulatory networks (GRNs) in region-specific mesenchymal clusters to ageing diseases of the human skeleton, such as osteoarthritis[12,13].

Cellular taxonomy of joint development

We performed paired droplet-based snRNA-seq and snATAC-seq (10X Genomics Multiome) on the hip, knee and shoulder joints across samples from 12 donors between 5 and 11 PCW (Fig. 1a,b). For the developing cranium, which had not been previously profiled across different ages, to our knowledge, we sampled the anterior and posterior regions of the calvaria and skull base separately, to divide the presumed intramembranous and endochondral bone-forming niches (Fig. 1a and Supplementary Table 1). We captured 336,162 high-quality droplets across 8 shared cellular compartments (Fig. 1c–e, Extended Data Fig. 1a–e and Supplementary Table 2). High concordance was observed between the transcriptome and ATAC peak profile across compartments (Extended Data Fig. 1f). Mesenchymal cells were predominant across all regions, whereas myogenic cells were absent in the calvaria (Fig. 1f). From these,

we defined over 100 fine-grained clusters (Supplementary Table 3) and captured a diversity of chondrogenic and osteogenic populations in our data compared with previous published single-cell data (see Methods; Extended Data Fig. 2 and Discussion in Supplementary Information). To resolve bone-lineage cell states and their spatial organization in the nascent synovial joint, we performed high-resolution 155-plex ISS of the whole intact early embryonic forelimbs (6.7 PCW) and hindlimbs (5.7 PCW), and late embryonic (7.3 PCW) knee and shoulder regions (Fig. 1d, Extended Data Fig. 3a–c and Supplementary Table 4). In addition, we conducted sequencing-based spatial transcriptomics (10X Genomics Visium CytAssist) of the developing coronal suture (9 PCW) and frontal bone (Extended Data Fig. 3d), allowing capture of osteolineage maturation across space. We then leveraged these data to systemically curate cell lineages within the mesenchymal compartment in a spatial context. This enabled the discovery of osteogenic cell states in the appendicular regions and skull base (endochondral ossification) and calvarium (intramembranous ossification), reflecting different mechanisms for osteoblastogenesis. Although postnatal mice leverage endochondral ossification during suture closure[14], it remains unclear whether this is typical in prenatal mice or human sutures. We found that chondrogenic clusters were relatively depleted in the calvarium (Fig. 3a and Extended Data Fig. 4a), consistent with previously reported mechanisms of intramembranous bone formation[15]. We also observed greater cell-abundance discrepancies across droplet and spatial data, particularly at later developmental stages (Supplementary Fig. 1 and Discussion in Supplementary Information).

Zonation of the embryonic synovial joint

Synovial joint-site determination occurs between 5 and 6 PCW in the limbs, in an initial mesenchymal condensation comprising *GDF5*-expressing interzone cells[13,16]. We applied differential abundance testing on developmental stages and identified InterzoneChon as the earliest joint cluster (Extended Data Fig. 4a–c and Discussion in Supplementary Information). Subclustering of the InterzoneChon (*GDF5*+*PITX2*+) population and RNA velocity dynamics analysis (see Methods) allowed inference of their pseudotime trajectory (Fig. 2a and Discussion in Supplementary Information). We applied SCENIC+ to predict gene programs and transcription factor accessibility changes across the seven interzone subclusters (Fig. 2b and Supplementary Table 5). The early interzone population (*PRRX1*) enriched for transcription factors associated with limb mesenchymal development (*TBX18*, *SHOX* and *LHX9*)[5] and had low *RUNX2* expression and target gene accessibility, but moderate *SOX5*, *SOX6* and *SOX9* expression and target accessibility, suggesting a poised trajectory favouring chondrogenesis over osteogenesis. The articular, fibro and *GDF5*hi interzone clusters highly expressed *GDF5*, and each had distinct gene signatures (see Discussion in Supplementary Information). We hypothesize that the *GDF5*hi interzone cluster, which showed low activity for chondrogenic and osteogenic transcription factors, is undifferentiated and has the potential to sustain influx into the forming joint[16]. We leveraged our newly developed ISS-Patcher function (see Methods) to infer cell labels in the 155-plex-clustered ISS data manifold (Fig. 2c). In the embryonic hindlimb, the early interzone cluster was diffusely distributed across regions of the interzone and cartilage scaffold, and was surrounded by the dermal interzone (Fig. 2d). The articular interzone cluster was predominantly enriched in sites of incipient knee articular cartilage, which colocalized with *SOX9* staining (Extended Data Fig. 4c). By contrast, the fibro interzone cluster[17], which expressed meniscus-related (*PTN*) and ligament-related (*POSTN* and *SCX*) genes (Extended Data Fig. 4b), was enriched in the shoulder interzone region adjacent to the articular surface of the humerus[18] (Fig. 2d). The comparative paucity of the fibro interzone cluster enriched in the hindlimb may be due to the earlier formation of the shoulder fibrocartilage than of the joints in the hindlimb[19,20]. From these spatial enrichment patterns, we demonstrated zonation of the presumptive joint, showing early chondrogenic and anti-osteogenic transcription factor enrichment in the interzone centre, and *RUNX2* enrichment in the developing cartilage scaffold (Extended Data Figs. 3d and 4d). Using gene-enrichment scoring, our data suggest that joint cavitation occurs subsequent to zonation and takes place between 7 and 8 PCW (Supplementary Fig. 2 and Discussion in Supplementary Information).

Emergence of fibroblast lineages

Fibroblast lineage cell states in the embryonic mouse limb arise from a master *HIC1*+ precursor population[21] that contributes minimally to osteochondral components and gives rise to a postnatal 'universal' *PI16*+ population, also identified in the adult human across tissues[21,22]. We sought to uncover the taxonomy of the fibroblast lineage in first trimester human joints. We identified a fibroblast progenitor (FibroPRO1), *HIC1*+ mesenchyme (*HIC1*+Mes) and dermal fibroblasts (DermFIB1 and DermFIB2; see Discussion in Supplementary Information) in the appendicular joints during the embryonic period (less than 8 PCW), surrounding the nascent joint, and with diffuse distribution in the limbs, respectively (Extended Data Fig. 5a–d). At approximately 8 PCW, FibroPRO2 expresses *PI16* and *DPT*, which are markers of pan-tissue adventitia-associated fibroblasts in postnatal health (Extended Data Fig. 5e,f). *BNC2*, a myofibroblast-associated transcription factor[23], had high activity in FibroPRO1 and FibroPRO2, consistent with its postnatal expression in the vascular adventitia[22]. In addition, *YBX1*, a transcription factor that has been shown to drive proliferation of mouse

embryonic fibroblasts, was also enriched. On the basis of developmental time and RNA velocity (see Methods), we predicted *HIC1*+Mes as an early fibroblast progenitor during the embryonic phase at less than 8 PCW (Extended Data Fig. 5a,b). Here it was inferred to give rise to tenocytes, synovial fibroblasts, dermal fibroblasts and myofibroblasts (Extended Data Fig. 5a,b). *HIC1*+Mes showed high activity of numerous proliferation-associated transcription factors including *WT1*, *SOX5* and *FOXC1*, which are associated with an invasive and activated synovial fibroblast phenotype[24–27] (Extended Data Fig. 5g), and mapped to the embryonic limb on our RNA-ISS data (Extended Data Fig. 5h). *HIC1*+Mes also demonstrated moderate accessibility in tenogenesis transcription factors such as *SCX* and *MKX* (Extended Data Fig. 5i), which is suggestive of tenogenic potential, consistent with fate-mapping in mice[28] and our trajectory analysis (Extended Data Fig. 5a–d), and is a finding that may warrant functional exploration in future work.

Formation of the cranial sutures

In the cranium, the incipient suture mesenchyme matures from 7 PCW, forming suture joints[29]. We identified calvaria-abundant early cranial progenitors (CranialMes, FacialMes and PArchMes; Extended Data Figs. 4a and 6a–c and Discussion in Supplementary Information) and *RUNX2*-expressing SutureMes1 and SutureMes2, which also expressed *CTSK*, *SIX2* and *AXIN2*, consistent with mouse cranium progenitors[30,31], suggesting that they form part of the intramembranous osteogenic lineage (see Discussion in Supplementary Information). Classical markers of fetal mouse cranial sutures (*TWIST1*, *ZIC1* and *ZIC4*) and *THBS2* were enriched in both SutureMes populations[31]. Using Cell2location, we localized these to the developing coronal suture joint (Fig. 2e). Osteoprogenitors (*HHIP*+PreOB) emerged at the opposing frontal and parietal bone boundaries of the SutureMes populations (Fig. 2e and Extended Data Fig. 6d). *ALX1*, a transcription factor required for cranium formation in the mouse[32] and neural crest cell (NCC) migration in human-derived cells, and numerous NCC transcription factors (*PAX3*, *BMP3* and *TSHZ2*) were found to be differentially enriched in cells captured from anterior regions of the cranium (Fig. 1a and Extended Data Fig. 6b,c,e). Although these data suggest that *ALX1*+ cells may potentially have a neural crest origin, owing to the transient presence of NCCs[33], we did not capture bona fide early embryonic *SOX10*+ NCCs. Analogous to osteogenic repressors expressed by the articular interzone, SutureMes showed high activity for anti-osteogenesis transcription factors also enriched in mouse cranial sutures (*TWIST1*, *LMX1B* and *NFATC2*)[34]. Simultaneously, osteogenic (*SP7* and *FOXO1*) transcription factor activity was also high, suggesting a GRN primed for bone formation. Comparable with molecular gradients of the embryonic knee (Fig. 2f,g), we observed *LMX1B* and *TWIST1* expression within the suture region, dissipating towards the flanking bone edges concurrent to enrichment of *RUNX2* (Fig. 2f and Extended Data Fig. 4e), suggesting similar mechanisms in sustaining the non-ossifying joint space in both the limb and the cranium. To reveal the enhancer-driven GRN of the loci surrounding the key osteogenic transcription factors *RUNX2* and *HHIP*, we visualized ATAC coverage in combination with SCENIC+-predicted transcription factor–peak and peak–gene links across clusters (Fig. 2h and Supplementary Fig. 3). Both *RUNX2* and *HHIP* were predicted to be inhibited by a shared set of anti-osteogenic transcription factors, including *LMX1B*, *TWIST1* and *ALX4*, via intermediate repressors targeting enhancers around their loci, illustrating the relationships maintaining the balance of osteogenic initiation. *HHIP* was highly accessible in *HHIP*+PreOB and was indirectly repressed by *LMX1B* via *TWIST1*. *RUNX2* was most accessible in *HHIP*+PreOB (intramembranous ossification) and preosteoblast (endochondral ossification) and was indirectly targeted by the same repressors via *TCF12* and *PRRX2*. Overall, this network illustrates the coherent regulation of bone-adjacent non-ossifying niches by key osteogenic regulators via multiple redundant paths.
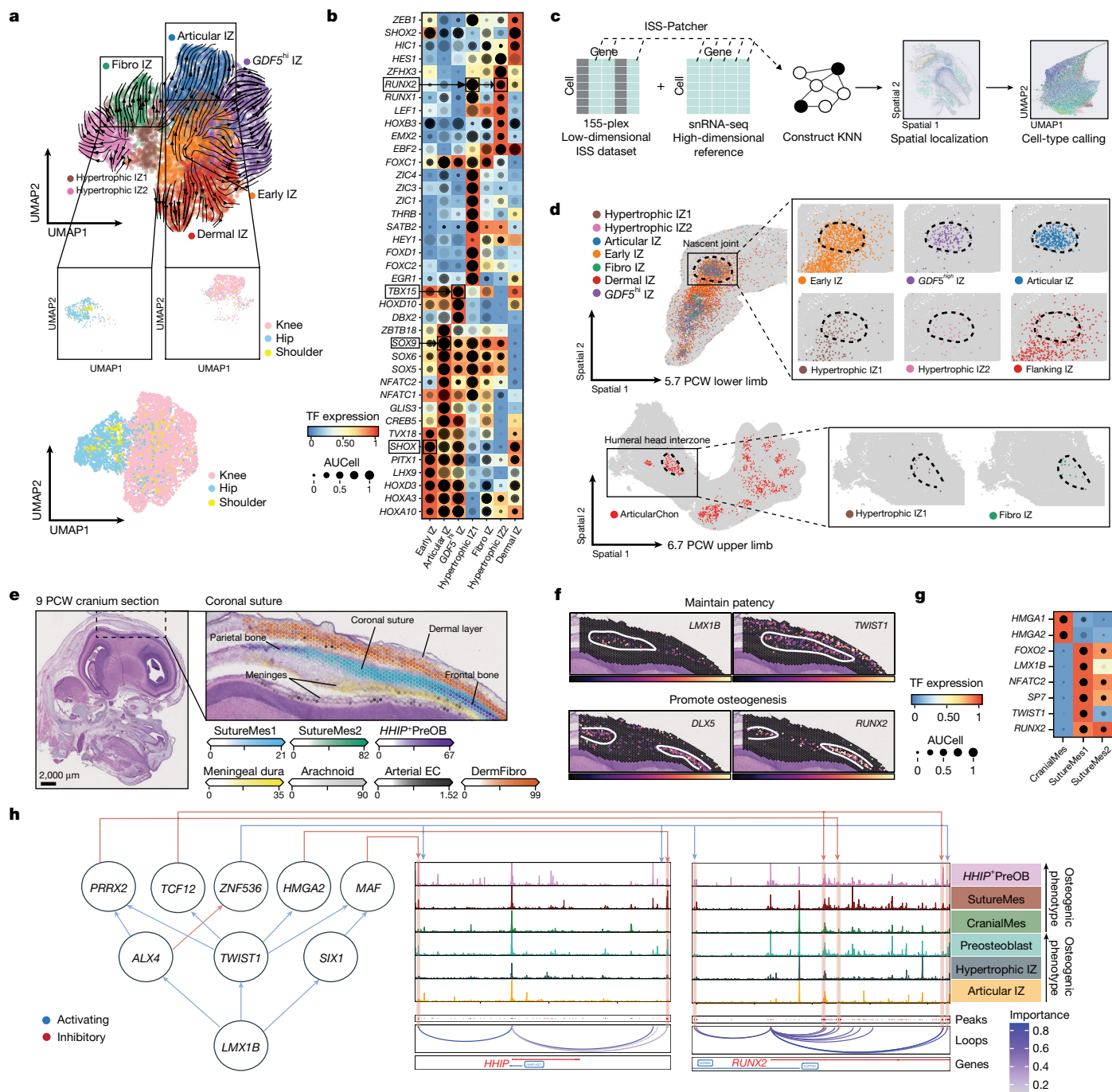
Fig. 2 | Formation of embryonic joints across space and time. a, RNA velocity on UMAP of RNA-subclustered cell states of the broad InterzoneChon cluster, with colours indicating each interzone cluster (top). UMAP coloured according to anatomical region (bottom). **b**, SCENIC+-predicted transcription factor expression (box colour). The dot size shows target gene accessibility (AUCell), and the dot shade (greyscale) shows target gene expression (GEX AUCell). **c**, ISS-Patcher workflow schematic (see Methods). KNN, *k*-nearest neighbour. **d**, Spatial plots of ISS-Patcher-imputed cell clusters. An ISS image of the hindlimb with the imputed interzone clusters overlaid is shown (top left), and individual imputed interzone clusters within the knee interzone are also shown (top right). An ISS image of the forelimb with the imputed articular chondrocyte cluster (bottom left), and individual imputed interzone clusters within the shoulder interzone (bottom right) are also displayed. **e**, A haematoxylin and eosin staining section of a cranium in sagittal view at 9 PCW (left), and a region on the adjacent section profiled using 10X CytAssist Visium data (right). Images are representative of one donor. Cell2location

results of the coronal suture, showing enrichment of cluster labels in each voxel, are also shown (right). **f**, Normalized transcription factor expression plotted in Visium voxels of the Visium data in panel **e**. **g**, SCENIC+-predicted transcription factor expression (box colour). The dot size shows target gene accessibility (AUCell), and the dot shade (greyscale) shows target gene expression (GEX AUCell) in suture progenitors. **h**, Coverage plots showing aggregated single-cell ATAC signals around the *HHIP* and *RUNX2* loci for osteoprogenitor cell states with increasing osteogenic phenotype in intramembranous (CranialMes to *HHIP*+PreOB) and endochondral (Articular IZ to preosteoblast) ossification. Below each coverage plot, loops predicted by SCENIC+ between the transcription start site and enhancers are shown (coloured by importance score). Selected upstream transcription factors predicted by SCENIC+ to bind and regulate via some of the enhancers are also shown (left). The network links inhibitors of osteogenesis (such as *TWIST1* and *LMX1B*) to pro-osteogenic genes (*RUNX2* and *HHIP*) via overall inhibitory connections.
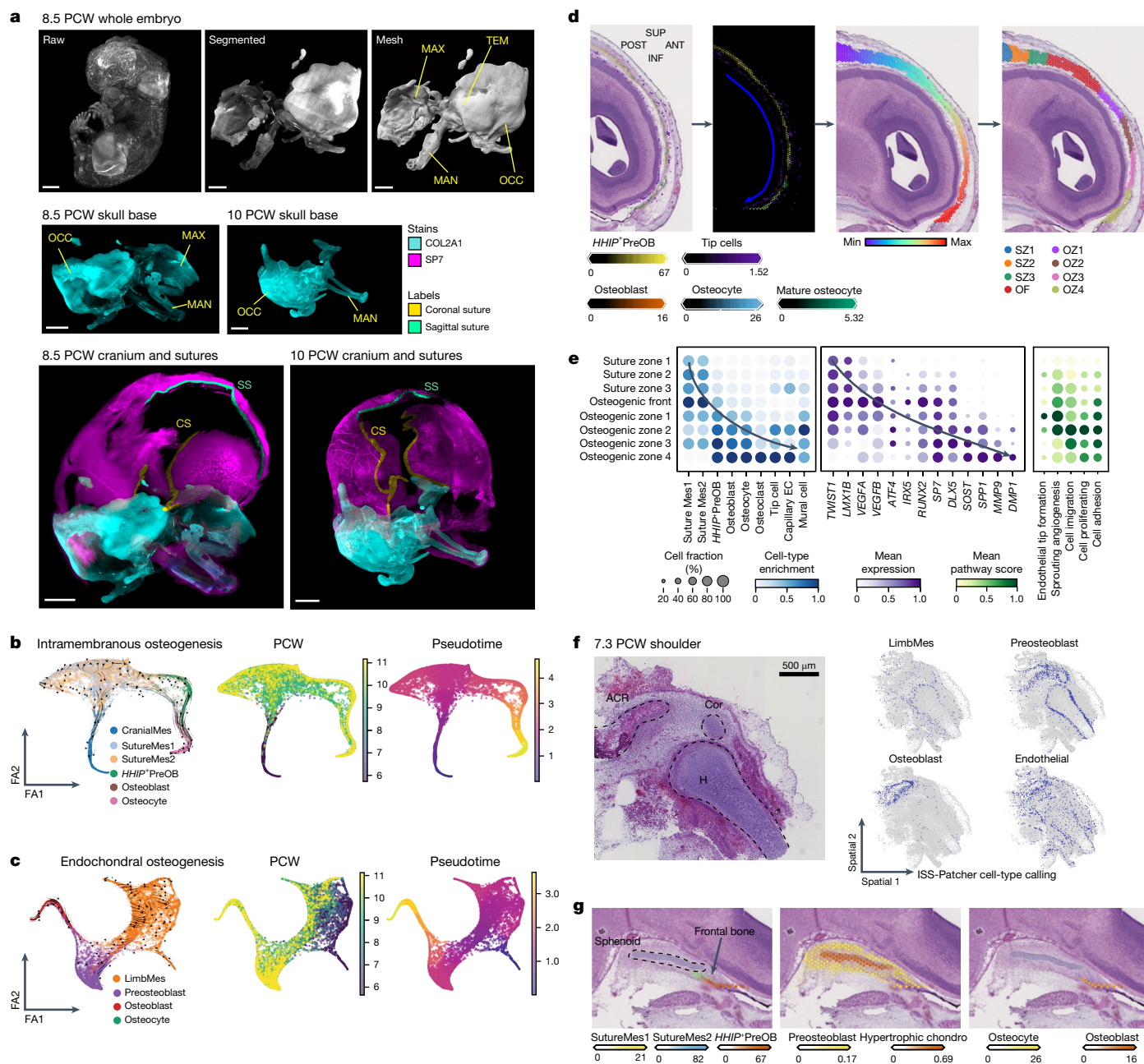
Fig. 3 | Intramembranous and endochondral osteogenic niches. a, Light-sheet fluorescence microscopy cleared images of the embryo and cranium at 8.5 and 10 PCW. A 3D view of rendered images of the 8.5 PCW embryo, the segmented chondrocranium and a mesh image for 3D rendering (top row); COL2A1 staining in the chondrocranium for both samples (middle row); and COL2A1 and osterix (SP7) staining overlaid across the whole cranium, with evident overlap in COL2A1 and SP7 staining in the skull base, but limited COL2A1 expression in the calvarium (bottom row) are shown. Coronal sutures (CSs) and sagittal sutures (SSs) are coloured according to the label legends. Images are representative of two donors. Scale bars, 2,000 μm. MAN, mandible; MAX, maxilla; OCC, occipital; TEM, temporal. **b,** Force-directed (FA) embedding showing the trajectory of intramembranous ossification in the calvaria suture mesenchyme. Embeddings show RNA velocity-predicted pseudotime trajectory arrows, PCW metadata and scFates-computed pseudotime, respectively (left to right). **c,** As in panel **b**, of a force-directed embedding showing the trajectory of endochondral ossification

in the limb mesenchyme and preosteoblast. **d,** Spatial binning of a 10X Visium CytAssist image of the frontal bone (sagittal section) using OrganAxis. The axis spans the coronal suture to the frontal bone (posterior–anterior). From left to right: haematoxylin and eosin staining image of the frontal bone, cell-cluster enrichment from Cell2location, axis values (rainbow scale) and manual bins based on histological features. Images are representative of three sections from one donor. ANT, anterior; INF, inferior; OF, osteogenic front; OZ, osteogenic zone; POST, posterior; SUP, superior; SZ, suture zone. **e,** Enrichment within spatial bins defined in panel **d,** for intramembranous cell states (left), selected marker gene expression (middle) and pathway enrichment (right). EC, endothelial cell. **f,** A 7.3 PCW humerus (H; coronal section) and imputed cell clusters showing endochondral ossification as a sequence of mapped cell states. Images are representative of sections from three donors. Cor, coracoid process. **g,** Sagittal section of the sphenoid bone and cell-state enrichment from Cell2location. Images are representative of one donor.

Trajectories of skeletal osteogenesis

Osteoblastogenesis commenced from approximately 7 to 8 PCW and was apparent in the cranium by 8 PCW (Fig. 3a, Extended Data Fig. 4a and Supplementary Videos 1 and 2). To study this, we inferred two major osteoblastogenic trajectories from distinct osteoprogenitors, which enriched osteogenic transcription factors and downregulated anti-osteogenic transcription factors along pseudotime (Fig. 3b,c,

Article

Extended Data Figs. 6f and 7a, Supplementary Tables 6 and 7 and Discussion in Supplementary Information). Endochondral ossification of the limb was predicted to stem from limb mesenchyme (LimbMes), a cluster sampled from, and mapping to, both forelimbs and hindlimbs and is transcriptionally similar to the lateral plate mesoderm (*WT1*) in the fetal human limb bud[5,22] (Fig. 3c, Extended Data Figs. 6a–c and 7a and Discussion in Supplementary Information). CranialMes and FacialMes were differentially abundant in the anterior portion of the calvarium and skull base, respectively (Extended Data Fig. 4a), and expressed the NCC-derived mesenchymal regulators *PAX3* and *ALX1* (ref. 35) (Extended Data Fig. 6a,e). We hypothesize that these clusters constitute previously undescribed human NCC-derived osteogenic populations (see Discussion in Supplementary Information). Suturogenesis was predicted to occur from CranialMes differentiating into SutureMes1 and/or SutureMes2 (Fig. 3b and Extended Data Fig. 7a), forming a predicted trajectory towards *HHIP*⁺PreOB. *HHIP* marks the osteogenic coronal suture mesenchyme in mice[36], and we demonstrate here that they enrich in a distinct population progeny to *TWIST1*-enriched SutureMes1 and/or SutureMes2 and is distributed in the ossifying cranial bone in human fetal development (Fig. 2f). Following suture formation, progressive waves of oriented differentiation emanate from the cranial sutures towards the developing bone front[31]. We applied OrganAxis (see Methods) to define a continuous maturation axis spanning the coronal suture to regions of the maturing frontal bone. Using zonal bins based on histological features, we evaluated cell-state mapping along the anterior–posterior axis (Fig. 3d). Enrichment of *TWIST1*⁺ SutureMes1 and/or SutureMes2 was observed in the suture zones (1–3; Fig. 3e). Within the osteogenic front, histological features of osteoprogenitors emerged along with *HHIP*⁺PreOB enrichment. Establishment of the osteogenic zones coincided with downregulation of anti-osteogenic (*LMX1B* and *TWIST1*) and upregulation of pro-osteogenic (*RUNX2*, *DLX5* and *SP7*) transcription factors, signifying a spatial molecular switch that zonates territories of the suture (Figs. 2f,g and 3d,e and Extended Data Fig. 6f). Osteoblast markers including *IRX5*, *SOST*, *SPP1*, *MMP9* and *DMP1* peaked towards the distant osteogenic zones[37], and enrichment of osteogenic transcription factors aligned with axis values away from the coronal suture (Fig. 3e and Extended Data Fig. 8a). In the limb ISS data, we applied ISS-Patcher and identified spatial localization of LimbMes, which is predicted to differentiate to preosteoblasts in the epiphysis and osteoblasts in the incipient bone from 7.3 PCW (Fig. 3f and Extended Data Fig. 6g). In the more mature skull base at 9 PCW, a comparable pattern was detected in the sphenoid where hypertrophic chondrocytes of the cartilage scaffold were surrounded by preosteoblasts (Fig. 3g).

Angiogenesis in the osteogenic niches

The basis for endothelia sprouting, which drives osteogenesis within the human intramembranous ossification niche, remains undefined[38]. Within the intramembranous niche, tip cells, mural cells and capillary endothelial cells progressively co-enriched along the osteogenic zones with osteoblast and osteocytes (Fig. 3e, Extended Data Fig. 8a–c and Discussion in Supplementary Information). Accordingly, SutureMes1 and/or SutureMes2 and *HHIP*⁺PreOB highly expressed *VEGFA* and *VEGFB*, which are modulators of vessel sprouting, and localized to the osteogenic front where endothelial cells highly expressed *VEGF* receptor genes (*FLT1*, *KDR* and *NPR1*; Extended Data Fig. 9a). These suggest that SutureMes1 and/or SutureMes2 and *HHIP*⁺PreOB may promote vascular invasion in intramembranous niches, akin to chondrocytes in endochondral ossification niches[39,40]. Through RNA in situ hybridization (RNA-ISH), we observed *VEGFA* and *KDR*–*FLT1* coexpression in the appendicular cartilage (Extended Data Fig. 10a–d). Our Visium data also demonstrated enrichment of *VEGFA* expression in the hypertrophic cartilage of the skull base, colocalizing with capillary endothelial cells mapped by Cell2location (Extended Data Fig. 6h). We

observed a spatial gradient of angiogenesis along the cranial frontal bone maturation axis by scoring for enrichment in sprouting angiogenic pathways (Fig. 3e), suggesting an association between angiogenesis and osteogenesis. Using NicheNet, we predicted osteolineage–tip cell interactions specific to the intramembranous niche and identified colocalization of these pairs using RNA-ISH (Extended Data Figs. 9b and 10e–j and Discussion in Supplementary Information). We used CellphoneDB to predict signalling interactions from ligands in the endothelial cell-to-osteogenic cell states[41]. Tip cells were predicted to signal via *NOTCH*, including JAG1/JAG2–NOTCH2 (Extended Data Fig. 9c and Discussion in Supplementary Information), which have been reported to promote differentiation of postnatal perivascular osteoprogenitors[42]. Mural and capillary endothelial cells expressed the ligand genes *FGF2* and *RSPO3*, which have been previously described to facilitate osteoblast differentiation via FGF2–FGFR2 (ref. 43) and RSPO3–LGR5 (Extended Data Fig. 9d and Discussion in Supplementary Information), and THBS1–CD36 (ref. 44). Endothelial cells also expressed *CCL14* and *CXCL12* − encoding the ligands for CCR1 (ref. 45) and DPP4 (ref. 46), respectively − which support in vitro osteoclast recruitment and differentiation. These spatially defined interactions suggest tip cell recruitment by VEGFA and EPHB2 from osteolineage cells in the bone front. We theorize that the vascularizing endothelial cells then promote osteoblastic differentiation, osteocyte mineralization and osteoclast recruitment in the maturing bone (Extended Data Figs. 6i and 9e). Other lineages, such as neurons, that may modulate osteogenesis were not captured in our droplet data, and future studies of the innervating neurons may shed light on the potential neuron–osteolineage interactions at work (Extended Data Fig. 9f and Discussion in Supplementary Information).

Developmental chondrocyte heterogeneity

Various types of cartilage, including hyaline, fibrous and elastic cartilage, are formed during development. Our data identified diverse chondrocyte clusters that exhibited strong region-specific abundance and gene modules[47] (Fig. 4a–c, Extended Data Fig. 11a–g, Supplementary Tables 3 and 8 and Discussion in Supplementary Information). Along with previously described populations, we identified new clusters: two region-specific chondrocyte progenitors (ChondroPro1 and ChondroPro2; Discussion in Supplementary Information) and *DLK1*⁺ chondrocytes (*DLK1*⁺Chon: *DLK1* and *CD63*). ChondroPro1 and ChondroPro2 were enriched in appendicular joints and the skull base, respectively, and expressed fibroblast differentiation markers (*POSTN*, *COL1A1*, *PRRX1* and *TWIST1*), consistent with findings in early chondrocyte progenitors in mice[17]. *DLK1*⁺Chon was enriched in ribosomal genes and *CD63*, which has been identified in the pre-hypertrophic layer in the limb, whereas *DLK1* itself is a marker for embryonic lineage progression from proliferative to pre-hypertrophic phenotypes[48]. Spatially, CyclingChon, *DLK1*⁺Chon and HyperChon were organized sequentially within the nascent bone, spanning from the epiphysis towards the diaphysis, the incipient primary ossification centre (Extended Data Fig. 8b), suggesting a transitional state within *DLK1*⁺Chon, between proliferative and pre-hypertrophic chondrocyte phenotypes. We also characterized previously undescribed craniofacial populations including facial (FacialChon) and mandibular chondrocytes (MandibularChon), which highly expressed *PAX3* and *SEMA3D*, respectively, implying potential origins from the neural crest[49–51]. Lineage tracing in zebrafish and mice[52] has shown that Schwann cells can differentiate into chondrocytes during embryogenesis. Our data captured a diverse Schwann compartment, and revealed *SOX9*⁺ endoneurial Schwann cells (*SOX9*⁺ enSC) characterized by the expression of chondrocyte (*SOX9*, *COL9A1*, *ACAN* and *COL2A1*) and classical Schwann cell markers (*MPZ* and *SOX10*; Fig. 4d,e and Extended Data Fig. 12a–c). *SOX9*⁺ enSC represented an end point in the trajectory analysis predictions, stemming from Schwann cell precursors, and expressed mesenchymal (*PRRX1*,
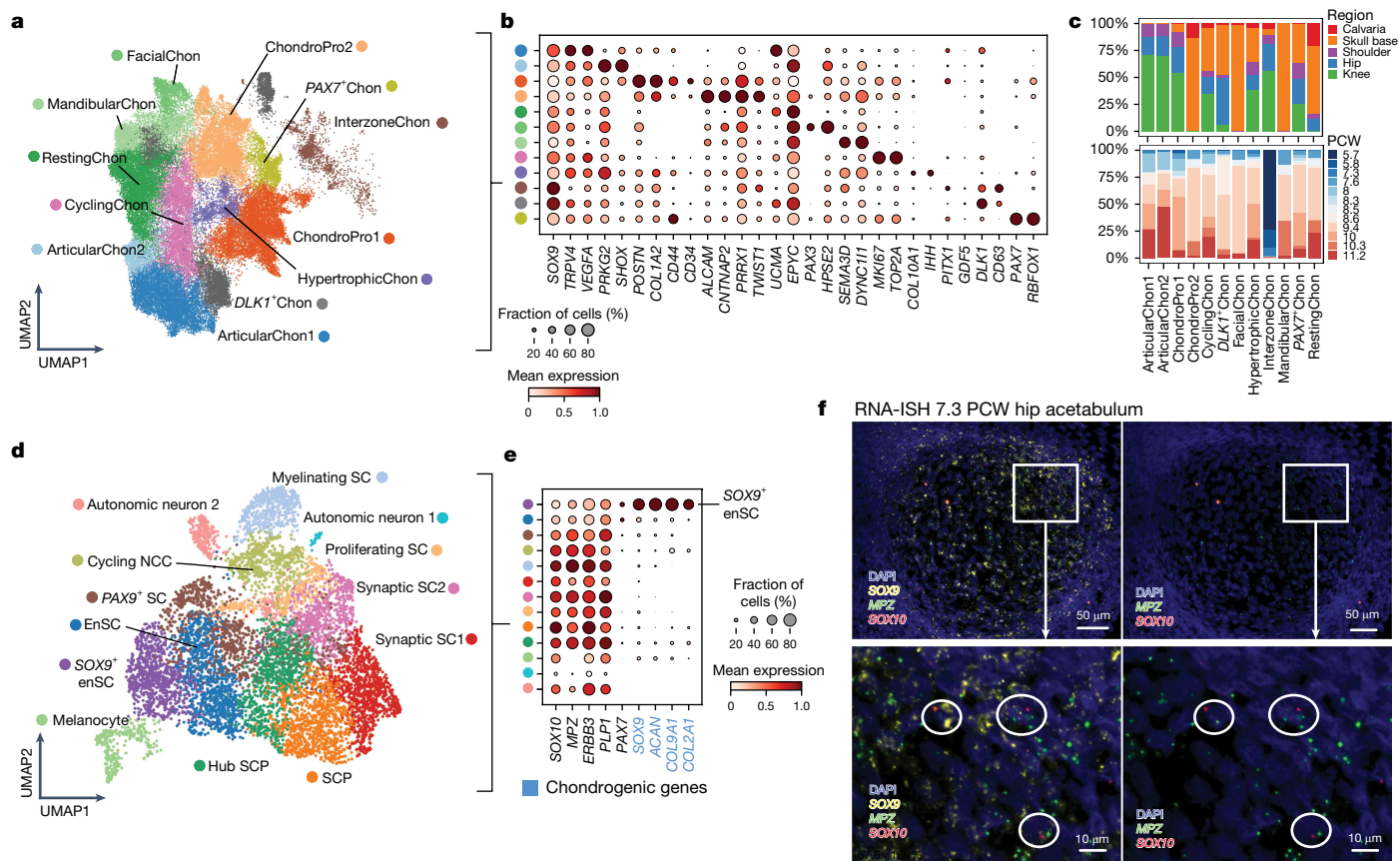
Fig. 4 | Chondrogenesis across anatomical regions. a, UMAP embedding of chondrocytes coloured by subtypes. **b**, Dot plot showing the mean expression (dot colour) and fraction of expressing cells (dot size) of selected marker genes. **c**, Proportion of chondrogenic cell clusters across anatomical regions and developmental age (PCW). **d**, UMAP embedding of Schwann lineage cell states. SC, Schwann cell; SCP, Schwann cell precursor. **e**, Dot plot showing mean expression (dot colour) and fraction of expressing cells (dot size) of chondrogenesis-associated genes across Schwann cell states. The gene names in blue indicate markers of $SOX9^+$ enSC. **f**, Three-plex RNA-ISH (RNAscope) of marker genes in the 7.3 PCW hip acetabulum demonstrating enrichment and colocalization of markers of $SOX9^+$ enSC (MPZ^+SOX9^+) and Schwann cells (MPZ^+SOX10^+). Images are representative of sections from one donor.

PRRX2, *PDGFRA* and *TWIST2*) and HOX (*HOXA9*, *HOXA10*, *HOXA11* and *HOXD10*) signatures (Extended Data Figs. 12d,e). Through RNA-ISH and RNA-ISS, we observed widespread colocalization of Schwann marker *MPZ* and *SOX9* transcripts within the hip cartilage of the developing hindlimb and a smaller number of cells that coexpress *SOX10* (Fig. 4e,f, Extended Data Fig. 12f and Supplementary Table 9). Accordingly, our chondrocyte clusters do not express *SOX10* or *MPZ* (Extended Data Fig. 12g), indicating that $SOX9^+$ Schwann populations are present within the cartilage. A recent study has also identified Schwann cells within the cartilage of the developing hindlimb digits[53]. Although lineage-tracing experiments are needed to investigate further, we theorize that akin to mice, Schwann lineage cells may be a non-canonical source of chondrocytes in human development. Our droplet data also captured a *PAX7*+ chondrocyte cluster, which co-expressed markers and gene modules of chondrocytes and muscle cells, and persisted following computational quality control against doublets and ambient RNA (Supplementary Figs. 4–8 and Methods and Discussion in Supplementary Information). Although RNA-ISH demonstrated potential localization of *PAX7* transcripts in chondrocytes of the limb, the signal intensity was relatively modest. Future work, involving post-FACS transcriptional analysis, is needed to investigate the validity of this population.
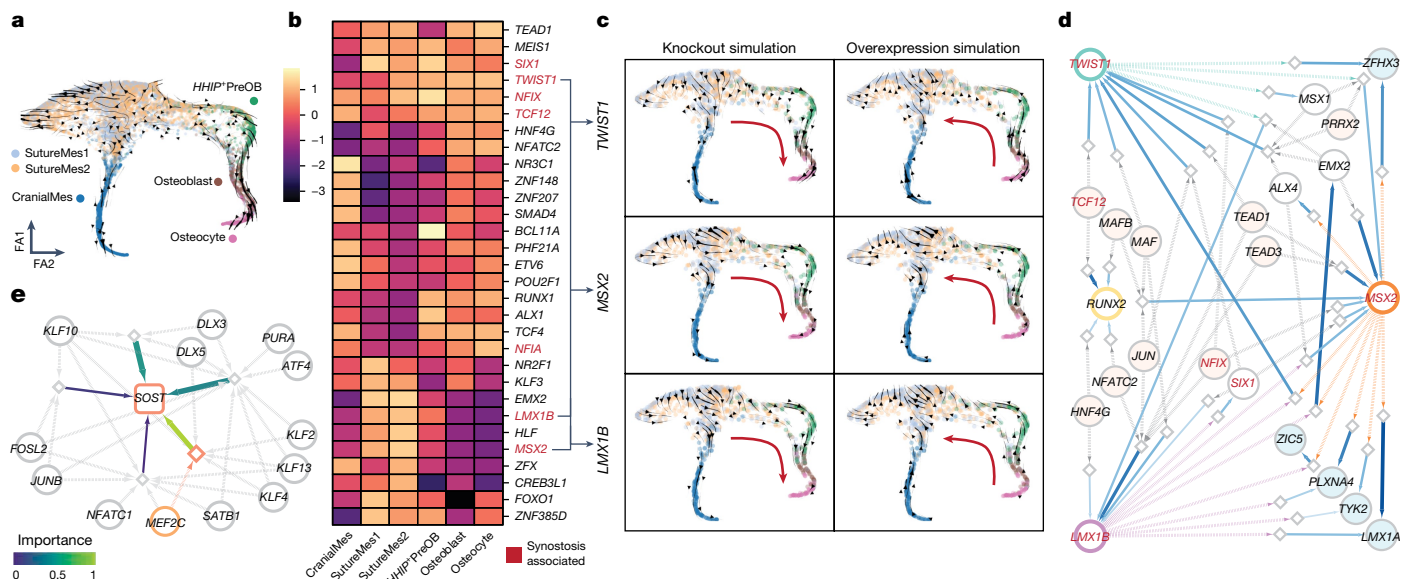
Developmental links to complex traits

Numerous conditions of the ageing skeleton have been linked to disrupted joint and bone changes during the embryonic stages of life.

Of note, enhancer-associated variants associated with adult osteoarthritis appear to act on anatomical region-specific regulatory networks to influence synovial joint morphology during development[54]. Using functional genome-wide association study (fGWAS; see Methods)[55], we found knee osteoarthritis signals distinctly enriched in chondrogenic states, except for InterzoneChon (Fig. 5a and Supplementary Table 10). By contrast, hip osteoarthritis enrichment was observed in only two chondrocyte populations (ChondroPro1 and Hypertrophic-Chon), but was enriched in preosteoblasts, osteoblasts and osteocytes. Previous genetic studies have theorized that early bone development affects subsequent risk of hip osteoarthritis, potentially through modulating hip shape and consequent mechanical forces[56], whereas the knee-specific findings point towards alterations to chondrogenesis. Deriving insights from clusters implicated in fGWAS, we used SNP2Cell to identify cluster-specific sub-GRNs enriching for osteoarthritis signals (see Methods; Fig. 5b and Discussion in Supplementary Information). Preosteoblasts expectedly displayed the largest average enrichment across the osteogenic pseudotime trajectory for hip osteoarthritis (Fig. 5c), consistent with fGWAS results. We also identified subnetworks for hip osteoarthritis and knee osteoarthritis, prioritizing preosteoblasts (Fig. 5d) and articular chondrocytes (Extended Data Fig. 13a), respectively, and revealed similar regulatory pathways that balance chondrogenic and osteogenic functions. In the articular chondrocyte–knee osteoarthritis network, several non-transcription factor genes with roles in cartilage makeup and chondrocyte differentiation (*COL27A1*, *PRKCA*, *SNORC* and *CRISPLD2*) were predicted to be regulated by *NFATC1* and *FOXA3* (Discussion in Supplementary

Fig. 5 | Links to complex diseases of the skeleton. a, Enrichment of GWAS signals (fGWAS) for osteoarthritis (OA) and its surrogate phenotypes (replacement (rep) of knee and hip joints), showing how patterns of cell-type enrichment differ between hip-associated and knee-associated single-nucleotide polymorphisms. Enrichment was tested for each cluster-specific logistic regression (Wald test β1 ≠ 0, d.f. = 1, Benjamini–Hochberg (BH) correction, false discovery rate (FDR) < 0.1). BOER, Boer et al.[62]; logOR, log odds ratio; THR, total hip replacement; TKR, total knee replacement. **b**, Schematic representation of the SNP2Cell method; scores were derived from GWAS summary statistics, and cell cluster marker scores are mapped and integrated across a GRN, highlighting enriched modules that are predicted to have a cluster-specific role in disease. **c**, Enrichment scores for hip osteoarthritis

of the top ten genes per cluster across the osteogenic cluster (heatmap; top), and a boxplot showing the median enrichment scores, with lower and upper quartiles and whiskers at a maximum of 0.1 the interquartile range across *n* = 1,967 genes and peaks with a *z*-score of more than 2 (bottom). Notches (95% confidence interval around the median) are approximate guides of significant differences between clusters based on their overlap. **d**, Preosteoblast-specific and hip osteoarthritis-specific subnetwork of the most enriched genes and peaks. The brighter colours correspond to greater enrichment scores, relative to scores obtained from random permutations. **e**, Gene set enrichment analysis for Gene Ontology biological process terms across ArticularChon knee osteoarthritis and preosteoblast hip osteoarthritis enrichment scores. NES, normalized enrichment score.

Information). For hip preosteoblast in hip osteoarthritis, the osteogenic regulator *RUNX2* showed significant enrichment, along with multiple NFAT genes (*NFATC1*, *NFATC2*, *NFATC4* and *NFAT5*), which in conjunction with additional transcription factors (*ZEB1*, *MAF* and *TEAD1*), implicated calcineurin and WNT signalling pathways, which are known to have a role in hip shape formation and osteoarthritis (Discussion in Supplementary Information). Pathway analysis also showed that cellular responses to lipids were enriched in hip preosteoblasts (Fig. 5e and Discussion in Supplementary Information). Overall, through application of fGWAS and our new tool SNP2Cell, we identified differential enrichment of knee and hip osteoarthritis GWAS signals in developmental chondrogenic and osteogenic single-cell profiles, respectively.

Deciphering monogenic craniosynostosis

Craniosynostosis is a congenital condition that involves disturbances in cranial ossification and suture formation during fetal and postnatal development, leading to premature cranial suture fusion and depletion of osteoprogenitor pools (Discussion in Supplementary Information), resulting in global developmental consequences. To predict normal developmental cell states that enrich genes affected in craniosynostosis, we cross-referenced pseudotime-associated differentially expressed genes enriched in the intramembranous and endochondral ossification pathways against a candidate database of over 2,700 genes known to cause congenital conditions in humans[5] (Extended Data Fig. 13b,

Supplementary Tables 6 and 11 and Discussion in Supplementary Information). The majority of enriched craniosynostosis (*n* = 13 of 22) genes were observed in progenitor populations of the intramembranous pseudotime trajectory (SutureMes1, SutureMes2 and *HHIP*+PreOB). Most of these enriched genes were also highly accessible, apart from *IHH*, suggesting that the embryonic period may be affected in craniosynostosis. To simulate effects of candidate craniosynostosis-associated transcription factor perturbation during normal osteogenesis, we applied CellOracle to predict velocity shifts within the intramembranous trajectory in silico for 485 detected transcription factors. Knockout simulations for *TWIST1*, *MSX2* and *LMX1B* were predicted to lead to high-velocity shifts in SutureMes2 (Fig. 6a–c). *TWIST1* and *LMX1B* also showed spatial enrichment in the coronal suture (Fig. 2f). The direction of inferred velocity changes was also consistent with knockout leading to osteogenesis (Discussion in Supplementary Information). Overall, these predictions may help to inform potential transcriptional effects associated with pathogenic features of transcription factor-mediated craniosynostosis. Future functional studies will be required to reveal mechanistic effects for each gene. We next reconstructed interaction networks across the prioritized transcription factors (*TWIST1*, *MSX2* and *LMX1B*) to resolve their co-regulatory relationship (Fig. 6d). The predictions revealed inter-regulation of shared coding and non-coding targets across these transcription factors. Of the connected nodes, numerous transcription factors (*SIX1*, *TCF12*, *NFIX* and *ALX4*) are known to be associated with craniosynostosis through loss-of-function

Fig. 6 | Monogenic conditions affecting bone development. a, Force-directed embedding of the intramembranous ossification trajectory. The arrows show directional prediction of differentiation with cytoTRACE, assuming that cells at an earlier time of the trajectory express more genes on average. Consistent with RNA velocity, differentiation from CranialMes to osteocytes is predicted. The arrows on the left side of the SutureMes cluster pointing away from the main trajectory may reflect suture maintenance. **b**, Heatmap of in silico transcription factor-knockout perturbation scores per cell cluster showing the top five transcription factors with the highest scores per cluster. A higher score indicates that transcriptomic changes induced by the perturbation are promoting osteogenesis. The top transcription factors include several genes with a known role in craniosynostosis (marked in red). **c**, Transcription factor perturbation vectors showing the direction of induced transcriptomic changes on the force-directed embedding for three selected genes encoding transcription factors: *TWIST1*, *MSX2* and *LMX1B*. Knockout simulation of all three genes promotes osteogenesis, whereas overexpression inhibits it,

suggesting that the transcription factors regulate suture maintenance. **d**, Enhancer-driven GRN showing inter-regulation between *TWIST1*, *MSX2* and *LMX1B*, inferred using SCENIC+. The circles and diamonds represent genes and regions with transcription factor-binding sites, respectively. Region–gene links are coloured and scaled according to peak2gene importance, whereas transcription factor–region links are coloured by transcription factor. The blue circles represent regulated targets, the cream circles indicate regulators and the white circles represent intermediate genes. Transcription factors with a known role in craniosynostosis are marked in red text. The densely connected network contains several of such genes and suggests a potentially similar function for others. **e**, Enhancer-driven GRN showing predicted regulation of *SOST*. A region (orange diamond) containing the enhancer known to be mutated in Van Buchem disease and affecting the regulation of *SOST* is highlighted in orange. Region–gene links are coloured and scaled according to peak2gene importance.

mutations, suggesting a tightly regulated network conferring suture patency in this region. Of note, *TCF12* has been reported to govern coronal suture development through heterodimer formation with *TWIST1* (Discussion in Supplementary Information), and severe phenotypes are observed in mice with doubly deleterious mutations[57]. We also explored enhancer-mediated regulation of normal development centred around *SOST* to evaluate the role of the ECR5 enhancer, which when mutated leads to Van Buchem disease, a cause of sclerosing dysplasia of bone (Fig. 6e). These predictions of normal development may inform cellular models of transcription factor and enhancer-driven monogenic conditions of the bone lineage. To explore potential cell-extrinsic influences on fetal bone development, we applied Drug2Cell[58] to score enrichment of genes for teratogenic drug targets within our osteogenic clusters (see Methods; Supplementary Table 12). This revealed overall greater target enrichment of known teratogenic drug targets within intramembranous progenitors and downstream osteoblast or osteocyte cell states, for example, SutureMes1 and *HHIP*+PreOB, than endochondral progenitors (Extended Data Fig. 14a and Discussion in Supplementary Information). Our analyses allow identification of the bone-lineage clusters that express genes of targets for teratogenic drugs during normal development, and may help to inform the design of future functional studies.

Discussion

We present a multi-omic cell atlas that captures the spatially resolved cellular taxonomy of human synovial and suture joint formation across

the first trimester. In mice, *GDF5*+ cells are proposed to give rise to all cellular components within the joint, including ligaments and tendons. Here we infer that fibroblasts and tenocytes of the synovial joints arise from embryonic *HIC1*+ fibroblasts, a population previously reported in the mouse embryonic limb[21], and a later (*PI16*+) fibroblast progenitor, which is transcriptionally comparable with human postnatal universal fibroblasts[22]. Together, these shed new light on the lineage origins of the developing human embryonic synovial joint in the first trimester. We identified numerous previously undescribed human cranial embryonic osteolineage cell states including CranialMes (*HAND2*)[59,60], FacialMes (*PAX3*)[35] and PArchMes (*LHX8*)[61], which express comparable markers reported in mice. We defined two spatially resolved *TWIST1*+ SutureMes populations, and *HHIP*-expressing preosteoblast, which mirror populations of the fetal mouse suture[22,36] and infer their developmental trajectory in spatial niches. Using OrganAxis, we illustrated the shift in the osteogenic populations across the cranium, uncovering the association between endothelial recruitment and intramembranous ossification. We have shown that the endochondral niche in the limb cartilage scaffold derives from an undifferentiated progenitor expressing *ISL1* or *TBX5*, which marks an equivalent population in the human limb bud[5], and forms preosteoblasts and osteoblasts. These inferred trajectories form a valuable reference for studies of human bone development in the first trimester (Discussion in Supplementary Information). Through our droplet and spatial data, which capture Schwann cell development in human embryonic bone and joint, we inferred that Schwann cells may confer chondrogenic potential, mirroring previously described mouse development[52]. In future, focused

Article

functional investigations that involve isolation of these human cell lineages will further inform their developmental biology. Leveraging our dataset, we uncovered developmental links to both monogenic and complex musculoskeletal diseases. We observed marked differences in the enrichment of hip and knee osteoarthritis GWAS signals across developmental clusters, implicating osteogenesis in hip osteoarthritis and chondrogenesis in knee osteoarthritis. Lastly, by systematically simulating in silico knockout of transcription factors known to be associated with craniosynostosis, we identified a network of regulators inhibiting osteogenesis in the sutures, which provides mechanistic insights of how previously reported monogenic loss-of-function mutations may act. This approach, applied to the osteogenic trajectories, has potential value in aiding exploration of gene effects across other diseases involving embryonic osteogenesis. Our cross-region multimodal human developmental skeletal atlas is a fundamental resource for the understanding of human cartilage and bone development in the first trimester. It also has the potential to inform in vitro endeavours to differentiate osteoblast and other mesenchymal cell states.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08189-z.

1. O'Rahilly, R. & Gardner, E. The initial appearance of ossification in staged human embryos. *Am. J. Anat.* **134**, 291–307 (1972).
2. Faro, C., Benoit, B., Wegrzyn, P., Chaoui, R. & Nicolaides, K. H. Three-dimensional sonographic description of the fetal frontal bones and metopic suture. *Ultrasound Obstet. Gynecol.* **26**, 618–621 (2005).
3. Edwards, J. C. et al. The formation of human synovial joint cavities: a possible role for hyaluronan and CD44 in altered interzone cohesion. *J. Anat.* **185**, 355–367 (1994).
4. Ogata, S. & Uhthoff, H. K. The development of synovial plicae in human knee joints: an embryologic study. *Arthroscopy* **6**, 315–321 (1990).
5. Zhang, B. et al. A human embryonic limb cell atlas resolved in space and time. *Nature* https://doi.org/10.1038/s41586-023-06806-x (2023).
6. Pazzaglia, U. E. et al. Long bone human anlage longitudinal and circumferential growth in the fetal period and comparison with the growth plate cartilage of the postnatal age. *Microsc. Res. Tech.* **82**, 190–198 (2019).
7. Pazzaglia, U. E. et al. Study of endochondral ossification in human fetal cartilage anlagen of metacarpals: comparative morphology of mineral deposition in cartilage and in the periosteal bone matrix. *Anat. Rec.* **301**, 571–580 (2018).
8. Yayon, N. et al. A spatial human thymus cell atlas mapped to a continuous tissue axis. Preprint at *bioRxiv* https://doi.org/10.1101/2023.10.25.562925 (2023).
9. Wilkie, A. O. M., Johnson, D. & Wall, S. A. Clinical genetics of craniosynostosis. *Curr. Opin. Pediatr.* **29**, 622–628 (2017).
10. Wilkie, A. O. M. Craniosynostosis: genes and mechanisms. *Hum. Mol. Genet.* **6**, 1647–1656 (1997).
11. Twigg, S. R. & Wilkie, A. O. A genetic–pathophysiological framework for craniosynostosis. *Am. J. Hum. Genet.* **97**, 359–377 (2015).
12. Rice, S. J. et al. Genetic risk of osteoarthritis operates during human skeletogenesis. *Hum. Mol. Genet.* **32**, 2124–2138 (2023).
13. Capellini, T. D. et al. Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. *Nat. Genet.* **49**, 1202–1210 (2017).
14. Sahar, D. E., Longaker, M. T. & Quarto, N. Sox9 neural crest determinant gene controls patterning and closure of the posterior frontal cranial suture. *Dev. Biol.* **280**, 344–361 (2005).
15. Bartoletti, G., Dong, C., Umar, M. & He, F. Pdgfra regulates multipotent cell differentiation towards chondrocytes via inhibiting Wnt9a/β-catenin pathway during chondrocranial cartilage development. *Dev. Biol.* **466**, 36–46 (2020).
16. Shwartz, Y., Viukov, S., Krief, S. & Zelzer, E. Joint development involves a continuous influx of Gdf5-positive cells. *Cell Rep.* **15**, 2577–2587 (2016).
17. Bian, Q. et al. A single cell transcriptional atlas of early synovial joint development. *Development* **147**, dev185777 (2020).
18. Sugimoto, Y. et al. Scx⁺/Sox9⁺ progenitors contribute to the establishment of the junction between cartilage and tendon/ligament. *Development* **140**, 2280–2288 (2013).
19. Hita-Contreras, F. et al. Development of the human shoulder joint during the embryonic and early fetal stages: anatomical considerations for clinical practice. *J. Anat.* **232**, 422–430 (2018).
20. Mérida-Velasco, J. A. et al. Development of the human knee joint. *Anat. Rec.* https://doi.org/10.1002/(sici)1097-0185(199706)248:2%3C269::aid-ar14%3E3.0.co;2-n (1997).
21. Arostegui, M., Scott, R. W., Böse, K. & Underhill, T. M. Cellular taxonomy of Hic1⁺ mesenchymal progenitor derivatives in the limb: from embryo to adult. *Nat. Commun.* **13**, 4989 (2022).
22. Buechler, M. B. et al. Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021).
23. Bobowski-Gerard, M. et al. Functional genomics uncovers the transcription factor BNC2 as required for myofibroblastic activation in fibrosis. *Nat. Commun.* **13**, 5324 (2022).
24. Sontake, V. et al. Wilms' tumor 1 drives fibroproliferation and myofibroblast transformation in severe fibrotic lung disease. *JCI Insight* **3**, e121252 (2018).
25. Shi, Y. et al. Transcription factor SOX5 promotes the migration and invasion of fibroblast-like synoviocytes in part by regulating MMP-9 expression in collagen-induced arthritis. *Front. Immunol.* **9**, 749 (2018).
26. Yu, Z., Xu, H., Wang, H. & Wang, Y. Foxc1 promotes the proliferation of fibroblast-like synoviocytes in rheumatoid arthritis via PI3K/AKT signalling pathway. *Tissue Cell* **53**, 15–22 (2018).
27. Wang, J. et al. Forkhead box C1 promotes the pathology of osteoarthritis by upregulating β-catenin in synovial fibroblasts. *FEBS J.* **287**, 3065–3087 (2020).
28. Arostegui, M., Wilder Scott, R. & Michael Underhill, T. Hic1 identifies a specialized mesenchymal progenitor population in the embryonic limb responsible for bone superstructure formation. *Cell Rep.* **42**, 112325 (2023).
29. Morriss-Kay, G. M. & Wilkie, A. O. M. Growth of the normal skull vault and its alteration in craniosynostosis: insights from human genetics and experimental studies. *J. Anat.* **207**, 637–653 (2005).
30. Zhao, H. et al. The suture provides a niche for mesenchymal stem cells of craniofacial bones. *Nat. Cell Biol.* **17**, 386–396 (2015).
31. Farmer, D. T. et al. The developing mouse coronal suture at single-cell resolution. *Nat. Commun.* **12**, 4797 (2021).
32. Zhao, Q., Behringer, R. R. & de Crombrugghe, B. Prenatal folic acid treatment suppresses acrania and meroanencephaly in mice mutant for the Cart1 homeobox gene. *Nat. Genet.* **13**, 275–283 (1996).
33. Rocha, M. et al. From head to tail: regionalization of the neural crest. *Development* **147**, dev193888 (2020).
34. Cesario, J. M. et al. Anti-osteogenic function of a LIM-homeodomain transcription factor LMX1B is essential to early patterning of the calvaria. *Dev. Biol.* **443**, 103–116 (2018).
35. Zalc, A., Rattenbach, R., Auradé, F., Cadot, B. & Relaix, F. Pax3 and Pax7 play essential safeguard functions against environmental stress-induced birth defects. *Dev. Cell* **33**, 56–66 (2015).
36. Holmes, G. et al. Single-cell analysis identifies a key role for Hhip in murine coronal suture development. *Nat. Commun.* **12**, 7132 (2021).
37. Cain, C. J. et al. Loss of Iroquois homeobox transcription factors 3 and 5 in osteoblasts disrupts cranial mineralization. *Bone Rep.* **5**, 86–95 (2016).
38. Percival, C. J. & Richtsmeier, J. T. Angiogenesis and intramembranous osteogenesis. *Dev. Dyn.* **242**, 909–922 (2013).
39. Kronenberg, H. M. Developmental regulation of the growth plate. *Nature* **423**, 332–336 (2003).
40. Kusumbe, A. P., Ramasamy, S. K. & Adams, R. H. Coupling of angiogenesis and osteogenesis by a specific vessel subtype in bone. *Nature* **507**, 323–328 (2014).
41. 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).
42. Ramasamy, S. K., Kusumbe, A. P., Wang, L. & Adams, R. H. Endothelial Notch activity promotes angiogenesis and osteogenesis in bone. *Nature* **507**, 376–380 (2014).
43. Miraoui, H. et al. Fibroblast growth factor receptor 2 promotes osteogenic differentiation in mesenchymal cells via ERK1/2 and protein kinase C signaling. *J. Biol. Chem.* **284**, 4897–4904 (2009).
44. Kevorkova, O. et al. Low-bone-mass phenotype of deficient mice for the cluster of differentiation 36 (CD36). *PLoS ONE* **8**, e77701 (2013).
45. Yu, X., Huang, Y., Collin-Osdoby, P. & Osdoby, P. CCR1 chemokines promote the chemotactic recruitment, RANKL development, and motility of osteoclasts and are induced by inflammatory cytokines in osteoblasts. *J. Bone Miner. Res.* **19**, 2065–2077 (2004).
46. Ishida, M. et al. DPP-4 inhibitor impedes lipopolysaccharide-induced osteoclast formation and bone resorption in vivo. *Biomed. Pharmacother.* **109**, 242–253 (2019).
47. Shum, L. & Nuckolls, G. The life cycle of chondrocytes in the developing skeleton. *Arthritis Res.* **4**, 94–106 (2002).
48. Taipaleenmäki, H. et al. The crosstalk between transforming growth factor-β1 and delta like-1 mediates early chondrogenesis during embryonic endochondral ossification. *Stem Cells* **30**, 304–313 (2012).
49. Taïhi, I., Nassif, A., Isaac, J., Fournier, B. P. & Ferré, F. Head to knee: cranial neural crest-derived cells as promising candidates for human cartilage repair. *Stem Cells Int.* **2019**, 9310318 (2019).
50. Chilton, J. K. & Guthrie, S. Cranial expression of class 3 secreted semaphorins and their neuropilin receptors. *Dev. Dyn.* **228**, 726–733 (2003).
51. Berndt, J. D. & Halloran, M. C. Semaphorin 3d promotes cell proliferation and neural crest cell development downstream of TCF in the zebrafish hindbrain. *Development* **133**, 3983–3992 (2006).
52. Xie, M. et al. Schwann cell precursors contribute to skeletal formation during embryonic development in mice and zebrafish. *Proc. Natl Acad. Sci. USA* **116**, 15068–15073 (2019).
53. Lawrence, J. E. G. et al. Single cell transcriptomics reveals chondrocyte differentiation dynamics in vivo and in vitro. Preprint at *bioRxiv* https://doi.org/10.1101/2023.12.20.572425 (2023).
54. Richard, D. et al. Evolutionary selection and constraint on human knee chondrocyte regulation impacts osteoarthritis risk. *Cell* **181**, 362–381.e28 (2020).
55. Elmentaite, R. et al. Cells of the human intestinal tract mapped across space and time. *Nature* **597**, 250–255 (2021).

56. Wilkinson, J. M. & Zeggini, E. The genetic epidemiology of joint shape and the development of osteoarthritis. *Calcif. Tissue Int.* **109**, 257–276 (2021).
57. Sharma, V. P. et al. Mutations in TCF12, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. *Nat. Genet.* **45**, 304–307 (2013).
58. Kanemaru, K. et al. Spatially resolved multiomics of human cardiac niches. *Nature* **619**, 801–810 (2023).
59. Liu, N. et al. DNA binding-dependent and -independent functions of the Hand2 transcription factor during mouse embryogenesis. *Development* **136**, 933–942 (2009).
60. Xiong, W. et al. Hand2 is required in the epithelium for palatogenesis in mice. *Dev. Biol.* **330**, 131–141 (2009).
61. Zhou, C. et al. Lhx8 mediated Wnt and TGFβ pathways in tooth development and regeneration. *Biomaterials* **63**, 35–46 (2015).
62. Boer, C. G. et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell* **184**, 6003–6005 (2021).

Article

Methods

Sample acquisition and ethics

Developing human limb and cranium tissue samples were obtained from elective terminations under REC 96/085 with written and informed consent obtained from all sample donors (East of England, with full approval from the Cambridge Central Research Ethics Committee). In brief, samples were kept suspended in PBS and at −4 °C on ice during dissection. Shoulder, hip and knee joints were dissected en-bloc from the limbs. For the shoulder joint, a proximal incision was made at the distal third of the clavicle, and a distal incision was created at the neck of the humerus. For embryonic shoulder samples where distinctive bone features had not formed, approximations were made to capture the entirety of the glenohumeral and acromioclavicular joints. For the cranium samples (less than 8 PCW), two regions were dissected for each of the calvaria and skull base, separated at the posterior border of the frontal bone in both cases. For older cranial samples (more than 8 PCW), tissues were dissected to separate the frontal, parietal, sphenoid, ethmoid, occipital and temporal bones where feasible. Samples were initially embedded in optimal cutting temperature medium and frozen at −80 °C on an isopentane-dry ice slurry. Cryosections were then cut at a thickness of 10 μm using a Leica CM1950 cryostat and placed onto SuperFrost Plus slides (VWR) for ISS or Visium CytAssist, or used directly for single-nucleus processing. For samples used in whole-mount immunostaining, samples were obtained from terminations of pregnancy with written and informed consent obtained from all sample donors. Samples were provided by INSERM's HuDeCA Biobank and utilized in compliance with French regulations. Full authorization to use these tissues was granted by the French agency for biomedical research (Agence de la Biomédecine, Saint-Denis La Plaine, France; PFS19-012) and the INSERM Ethics Committee (IRB00003888).

ISS and high-resolution imaging

ISS was performed using the 10X Genomics CARTANA HS Library Preparation Kit (1110-02, following protocol D025) and the In Situ Sequencing Kit (3110-02, following protocol D100), which comprise a commercialized version of HybRISS[63]. Probe panel design was based on fold-change thresholds in cell states of the limbs (Supplementary Table 3). In brief, cryosections of developing limbs were fixed in 3.7% formaldehyde (252549, Merck) in PBS for 30 min and washed twice in PBS for 1 min each before permeabilization. Sections were briefly digested with 0.5 mg ml^{-1} pepsin (P7012, Merck) in 0.1 M HCl (10325710, Fisher) at 37 °C for 15 s (5 PCW) or 30 s (6 PCW and older), then washed twice again in PBS, all at room temperature. Following dehydration in 70% and 100% ethanol for 2 min each, a 9-mm diameter (50 μl volume) SecureSeal hybridization chamber (GBL621505-20EA, Grace Bio-Labs) was adhered to each slide and used to hold subsequent reaction mixtures. Following rehydration in buffer WB3, probe hybridization in buffer RM1 was conducted for 16 h at 37 °C. The 158-plex probe panel included 5 padlock probes per gene, the sequences of which are proprietary (10X Genomics CARTANA). The section was washed with PBS-T (PBS with 0.05% Tween-20) twice, then with buffer WB4 for 30 min at 37 °C, and three times again with PBS-T. Probe ligation in RM2 was conducted for 2 h at 37 °C, and the section was washed three times with PBS-T, then rolling circle amplification in RM3 was conducted for 18 h at 30 °C. Following PBS-T washes, all rolling circle products (RCPs) were hybridized with LM (Cy5-labelling mix with DAPI) for 30 min at room temperature, the section was washed with PBS-T and dehydrated with 70% and 100% ethanol. The hybridization chamber was removed and the slide mounted with SlowFade Gold Antifade Mountant (S36937, Thermo). Imaging of Cy5-labelled RCPs at this stage acted as a quality control step to confirm RCP ('anchor') generation and served to identify spots during decoding. Imaging was conducted using a Perkin Elmer Opera Phenix Plus High-Content Screening System in confocal mode with 1-μm z-step size, using a 63× (NA 1.15, 0.097 μm pixel^{-1})

water-immersion objective. For channels: DAPI (excitation of 375 nm and emission of 435–480 nm), Atto 425 (excitation 425 nm and emission 463–501 nm), Alexa Fluor 488 (excitation 488 nm and emission 500–550 nm), Cy3 (excitation 561 nm and emission 570–630 nm) and Cy5 (excitation 640 nm and emission 650–760 nm). Following imaging, each slide was de-coverslipped vertically in PBS (gently, with minimal agitation such that the coverslip 'fell' off to prevent damage to the tissue). The section was dehydrated with 70% and 100% ethanol, and a new hybridization chamber was secured to the slide. The previous cycle was stripped using 100% formamide (AM9342, Thermo), which was applied fresh each minute for 5 min, then washed with PBS-T. Barcode labelling was conducted using two rounds of hybridization, first an adapter probe pool (AP mixes AP1-AP6, in subsequent cycles), then a sequencing pool (SP mix, customized with Atto 425), each for 1 h at 37 °C with PBS-T washes in between and after. The section was dehydrated, the chamber removed, and the slide mounted and imaged as previously described. This was repeated another five times to generate the full dataset of seven cycles (anchor and six barcode bits).

Whole-mount immunostaining, tissue clearing and image analysis

Specimens were decalcified by incubation during 1 week in EDTA 0.5 M under agitation at room temperature. The solution was renewed halfway through the incubation period. The samples were washed twice in PBS 1X during 1 day. Samples were dehydrated for 1 h at room temperature in ascending concentrations of methanol in H$_2$O (20%, 40%, 60% and 80%). Then, samples were placed overnight under white light (11 W and 3,000 K°) and rolling agitation (004011000, IKA) with a 6% hydrogen peroxide solution in 100% methanol. Samples were rehydrated for 1 h at room temperature in descending concentrations of methanol (80%, 60%, 40% and 20%), washed twice and blocked in 0.2% PBS-gelatin with 0.5% Triton X-100 (PBSGT) solution during 1 week. Samples were transferred to a solution containing the primary antibodies (osterix, 1/500; ab209484, Abcam and collagen2, 1/500; ab185430, Abcam) diluted in PBSGT and were incubated at 37 °C with agitation at 20 rpm for 14 days. This was followed by six washes of 1 h in PBSGT at room temperature. Next, secondary antibodies were diluted in PBSGT and passed through a 0.22-μm filter. Samples were incubated at 37 °C in the secondary antibody solution for 7 days and washed six times during 1 h in PBSGT at room temperature.

The iDISCO+ protocol was used to clear the samples[64]. Samples were placed in TPP (Techno Plastic Products) tubes, dehydrated for 1 h in methanol (20%, 40%, 60%, 80% and 100% (2x)) under rotating agitation (SB3, Stuart). Methanol volumes were equal to about five times the sample volume. The samples were next incubated in a solution of 67% DCM and 33% MeOH overnight followed by 100% DCM for 30 min at room temperature on a rotator, then put in 100% DBE. Cleared samples were imaged with a Blaze light-sheet microscope (Miltenyi Biotec) equipped with sCMOS camera 5.5MP (2,560 × 2,160 pixels) controlled by Imspector Pro 7.5.3 acquisition software (Miltenyi Biotec). The light sheet, of 4 μm thickness, was generated by lasers at four different wavelengths (488 nm, 561 nm, 639 nm and 785 nm). 1× or 4× objectives with different magnification lenses of ×0.6, ×1 and ×1.66 were used. Samples were supported by a sample holder from Miltenyi and placed in a tank filled with DBE and illuminated by the laser light sheet from one or both sides depending on the size of the samples. LightSpeed Mode was used during acquisition to acquire these images in a reasonable time and at a suitable resolution. Mosaics of 3D image tiles were assembled with an overlap of 10% between the tiles. The images were acquired in a 16 bits TIFF format. Images were initially processed using MACS iQ View Software, which performed automatic alignment of the tiles. Stack images were converted to imaris file (.ims) using ImarisFileConverter. To isolate a specific structure in Imaris, we used the surface tool with manual selection, and then used the surface to mask the image. Images and videos were taken by using either the function Snapshot

and Animation in Imaris. Adobe Photoshop (v25.2) was used to colour the suture areas.

Multiplexed smFISH

Cryosections were processed using a Leica BOND RX to automate staining with the RNAscope Multiplex Fluorescent Reagent Kit v2 assay (Advanced Cell Diagnostics and Bio-Techne), according to the manufacturers' instructions. Probes can be found in Supplementary Table 7. Before staining, fresh frozen sections were post-fixed in 4% paraformaldehyde in PBS for 45 min at 4 °C, then dehydrated through a series of 50%, 70%, 100% and 100% ethanol for 5 min each. Following manual pre-treatment, automated processing included digestion with Protease III for 15 min before probe hybridization. Tyramide signal amplification with Opal 520, Opal 570 and Opal 650 (Akoya Biosciences), TSA-biotin (TSA Plus Biotin Kit, Perkin Elmer) and streptavidin-conjugated Atto 425 (Sigma-Aldrich) was used to develop RNAscope probe channels. Stained sections were imaged as for ISS above.

Flow cytometry cell sorting

We applied whole-cell dissociation of fresh donor tissue as previously described[5]. Before cell extraction, the sample tissues (approximately 9 PCW shoulder joints) were dissected to obtain bone samples, and soft tissues were microdissected away. The resultant cell suspension was stained with DAPI (Invitrogen) for live-viability, FGFR3 antibody (1:50; MA5-38521, Thermo Fisher Scientific) and TACR3 antibody (1:50; BS-0166R, Thermo Fisher Scientific), and secondary antibodies. DAPI-positive singlet cells were gated for DAPI staining by FACS using a BigFoot Spectral Cell Sorter (Thermo Fisher Scientific) and its proprietary software. Sequential gating for FGFR3 and TACR3 was then conducted to identify double-positive cells. Positive controls for FGFR3 and TACR3 were conducted using human peripheral blood mononuclear cells, and unstained cells were used as negative controls.

Image-based ISS decoding

We used the ISS decoding pipeline outlined in Li et al.[65]. This pipeline consists of five distinct steps. First, we performed image stitching using Acapella scripts provided by Perkin Elmer, which generated two-dimensional maximum intensity projections of all channels for each cycle. Next, we used Microaligner[66] (v1.0.0) to register all cycles based on DAPI signals using the default parameters. For cell segmentation, we utilized a scalable algorithm that leverages CellPose[67] (v3.0) as the segmentation method. The expected cell size is set to 70 pixels in diameter and further expanded 10 pixels to mimic the cytoplasm. To decode the RNA molecules, we used the PoSTcode algorithm[68] (v1.0) with the following parameters: rna_spot_size = 5, prob_threshold = 0.6, trackpy_percentile = 90 and trackpy_separation = 2. Furthermore, we assigned the decoded RNA molecules to segmented cells using STRtree (v2.0.6) and subsequently generated AnnData objects following the approach described by Virshup et al.[69]. Finally, only cells with more than four RNA molecules were kept for downstream analysis.

Visium processing and library preparation

Visium CytAssist Spatial Gene Expression for Fresh Frozen (10x Genomics) was performed following the manufacturer's protocol. Regions of interest were selected based on the presence of microenvironments of bone formation relevant to the droplet data (for example, coronal suture) and aligned to the CytAssist machine gasket accordingly. Images were captured using a Hamamatsu S60 slide scanner at ×40 magnification before conducting the Visium CytAssist protocol for subsequent alignment. Libraries were mapped with SpaceRanger (10X Genomics).

Single-nucleus isolation and library preparation

Single nuclei were isolated from fresh frozen samples through cryosectioning followed by mechanical dissociation as described in previous work[70]. In brief, 10-μm sections were homogenized in homogenization buffer (250 mM sucrose, 25 mM KCl, 5 mM MgCl$_2$, 10 mM Tris-HCl, 1 mM dithiothreitol, 1× protease inhibitor, 0.4 U μl^{-1} RNaseIn, 0.2 U μl^{-1} SUPERaseIn and 0.1% Triton X-100 in nuclease-free water) using a glass Dounce tissue grinder set (Merck). Samples were dissociated with 10–20 strokes of a loose pestle 'A' followed by 10 strokes of a tight pestle 'B' when tissue fragments remained. The resulting mixture was passed through a 50-μm cell strainer, followed by centrifugation (500g for 5 mins), the pellet was then resuspended in 300 μl of storage buffer (1× PBS, 4% BSA and 0.2 U μl^{-1} Protector RNaseIn) and passed through a 20-μm cell strainer. Nuclei were visualized and assessed for viability under microscopy following staining with trypan blue solution and were further processed for 10X Genomics Chromium Single Cell Multiome ATAC + Gene Expression according to the manufacturer's protocol. Nucleus suspensions were loaded with a targeted nuclei recovery of 16,000 droplets per reaction. For some of the nucleus samples, mixtures of samples from different sample donors were pooled within one reaction and later demultiplexed by genotype. Quality control of cDNA and final libraries was done using Bioanalyzer High Sensitivity DNA Analysis (Agilent). Libraries were sequenced using a NovaSeq 6000 (Illumina) with a minimum sequencing depth of 20,000 read pairs per droplet.

Data preprocessing

Sequencing data were aligned to the human reference genome (GRCh38-2020-A-2.0.0) using CellRanger-ARC software (v2.0.0). The called barcodes from 10X Multiome lanes with pooled genotypes from multiple sample donors were demultiplexed per genotype using BAM outputs through Souporcell (v2.0)[71]. Subsequently, the Souporcell outputs were clustered by genotype for metadata assignment to each barcode. Visium data were mapped to SpaceRanger (v1.1.0) using default input settings, and low-resolution CytAssist images were aligned to hi-resolution microscopy images of the processed slides using 10X Genomics LoupeBrowser (v7.0) according to capture frame marker regions. For gene expression data, SoupX (v1.6.0)[72] was applied to remove background ambient RNA. For CellRanger-ARC called matrices that contained more than 16,000 droplets (exceeding the number expected from targeted droplet recovery), we increased the estimated global rho value by 0.1 to account for the potential of additional ambient RNA. Droplets were filtered for more than 200 genes and less than 5% mitochondrial and ribosomal reads. Doublet removal is described below. For single-cell ATAC-seq, we applied ArchR[73] (v1.0.2) to process the outputs from CellRanger-ARC. Initial per-droplet quality control was performed considering the number of unique nuclear fragments, signal-to-background ratio and the fragment size distribution. Moreover, droplets with transcription start site enrichment score < 7 and number of fragments < 1,000 were removed. Doublets were discarded using the default settings. Initial clustering was performed at a resolution of 0.2 with the top 40 dimensions from iterative latent semantic indexing. Then, pseudo-bulk replicates were made for each broad cluster per region from the initial clustering results. Peak calling (501-bp fixed-width peaks) was performed based on pseudo-bulk coverages by MACS2 (v2.2.7.1). Then, a cell-by-peak count matrix was obtained and exported. We applied muon[74] (v0.1.2) for normalization, latent semantic indexing dimension reduction and clustering analysis using BBKNN[75] (v1.5.1) to correct for batch effects from anatomical regions and sample donors to obtain an ATAC embedding. Gene scores based on chromatin accessibility around gene bodies were calculated. We then applied MultiVI[76] (via scVI v0.6.8) to construct a joint embedding for snRNA-seq and single-cell ATAC-seq. We also applied EmptyDropMultiome[77] (v1.0.0) to repeat droplet calling to identify nucleus-containing droplets in our Multiome data to reduce the ambient RNA noise ('soup'). By generalizing EmptyDrops to the multi-omic setting, we used the smallest droplets to create an RNA and an ATAC ambient RNA 'soup' profile, and then tested each droplet for statistical deviations from each of these two profiles, retaining only droplets that were statistically significantly different from the soup profile.

Article

Doublet detection

All potential doublets detected in both RNA and ATAC modalities were removed from our data. For RNA data, Scrublet (v0.2.3)[78] was applied to estimate doublet probability, and a score of more than 0.3 was used as a cut-off value. To apply a stringent doublet threshold, we conducted an adapted Scrublet workflow as previously described[79]. In brief, per-droplet Scrublet scores were first determined for CellRanger-ARC count matrices from each 10X Multiome (gene expression) lane independently. The droplets were then overclustered through the standard scanpy workflow using default parameters up to Leiden clustering. Each individual cluster was further clustered. A per-cluster median of Scrublet scores was computed. A normal distribution of doublet score, centred at the score median with a standard deviation estimated from the median absolute deviation, was used to compute P values for each of the clusters. After false discovery rate adjustment using Benjamini–Hochberg correction, a $P > 0.65$ was deemed as a cut-off value of good-quality cells, as doublets were significant outliers. For ATAC data, we first applied doublet detection methods from ArchR to remove putative ATAC doublets. In addition, homotypic and heterotypic doublets were characterized by running AMULET (v1.1.0) on individual snATAC-seq libraries, and droplets with $q < 0.01$ were removed.

Droplet cluster annotation

We adopted a hierarchical clustering approach by first conducting Leiden clustering on the global integrated scVI (v0.9.1; hidden layers = 256, latent variables = 52, dispersion = 'gene-batch') RNA embeddings to obtain broad clusters. To validate these, we used Celltypist to train a model on cell states in the embryonic limb bud[5,80,81], and transferred labels onto our embedding for inspection. We utilized this information in addition to canonical marker genes to annotate broad clusters and subset sublineages. For sublineages (chondrocytes, fibroblasts, osteogenesis-related clusters, Schwann cells, immune cells and endothelial cells), we further embedded each subset using scVI (hidden layers = 256, latent variables = 52 and dispersion = 'gene-batch') and conducted Leiden clustering (resolution = 0.6), followed by differentially expressed gene (DEG) analyses (method = 'wilcoxon') to obtain cluster markers. We additionally utilized the inferred spatial location of cell states (described below) to inform annotations.

Differential abundance testing

We applied the Python implementation of the MILO package (v0.1.1) for differential abundance testing (http://github.com/emdann/milopy)[82]. We used the scVI latent representation to create a k-nearest neighbour graph of droplets in the relevant compartment and subsequently applied milopy to allocate droplets to overlapping neighbourhoods, with these droplets originating from multiple samples (brc_code). Each neighbourhood was then annotated as a cluster based on majority voting. We binarized values for anterior–posterior positions and calvarium-appendicular covariates to allow testing across these variables. We then determined log fold-change values for differential abundance and false discovery rate values based on the Bejanmini–Hochberg correction.

Spatial mapping using Cell2location

We performed Cell2location (v0.1.4) for deconvolution of Visium CytAssist voxels using our annotated Multiome data as inputs. Sample donor was used as the batch variable, and each library was considered a covariate in the regression model. For spatial mapping, we estimated 30 cells per voxel based on histological data, and set a hyperparameter detection alpha value of 20 for per-voxel normalization.

ISS-Patcher

ISS-Patcher is a package for approximating features not experimentally captured in low-dimensional data based on related high-dimensional data. It was developed as an approach to approximate expression signatures for genes missing in ISS data using matched snRNA-seq data as a reference in this study. First, a shared feature space between both datasets was identified by subsetting the 155–158 genes present in the ISS pool, followed by separate normalization to median total cell counts, log-transformation and z-scoring for both modalities. Then, the 15 nearest neighbours in scRNA-seq space were identified for each ISS cell with the Annoy Python package, and the genes absent from ISS were imputed as the average raw counts of the scRNA-seq neighbours.

Visium axis annotation using OrganAxis

Our Visium cranium sample was annotated with TissueTag[8] using a semi-automatic mode to generate a one-dimensional maturation axis. Regions of the developing bone were first manually annotated based on haematoxylin and eosin features. Tissue regions that did not include bone-forming niches were excluded from annotation. The annotation categories that were stored included multiple regions of the coronal suture (level 0 to level 2 annotation), stemming from the central-most portion, an osteogenic front (level 3 annotation) with histological features of osteoprogenitors and osteogenic zones (level 4 to level 7 annotation) from the emergence of histological osteoblasts. All annotations were saved as TissueTag output format, which logs the annotation resolution, the pixels per micrometre and the pixel value interpretation of annotation names (for example, 0 = 'suture') and colours (for example, 'osteogenic front': 'red'). To robustly and efficiently migrate TissueTag annotations to the Visium objects, we first transferred TissueTag annotations from pixel space to a high-resolution hexagonal grid space (15-μm spot diameter and 15-μm point-to-point centre distance with no gap between spots) based on the median pixel value of the centre of the spot (radius/4) in the annotated image. Next, to generate continuous annotations for Visium data, for each spot in the hexagonal high-resolution grid, we measured the mean Euclidean distance to the ten nearest points from each annotated structure in the level 0 annotation and the distance from the closest point for structures in level 1 annotation. All annotations were mapped to the Visium spots by proximity of the spot annotation grid to the nearest corresponding spot in the Visium array.

GRN analysis

The SCENIC+[83] (v1.0.0) pipeline was used to predict transcription factors and putative target genes as well as regulatory genomic regions with binding sites. The fragment matrix of peaks called with MACS2 and processed within ArchR[73] together with the corresponding RNA count matrix were used as inputs. Meta-cells were created by clustering droplets into groups of around 10–15 droplets based on their RNA profiles and subsequent aggregation of counts and fragments. The pipeline was applied to subsets of the dataset corresponding to individual lineages: first, CisTopic (pycistopic v1.0.2) was applied to identify region topics and differentially accessible regions from the fragment counts as candidate regions for transcription factor binding. CisTarget (pycistarget v1.0.2) was then run to scan the regions for transcription factor-binding sites, and GRNBoost2 (arboreto v0.1.6)[84] was used to link transcription factors and regions to target genes based on co-expression or accessibility. Enriched transcription factor motifs in the regions linked to target genes were used to construct transcription factor–region and transcription factor–gene regulons. Finally, regulon activity scores were computed with AUCell based on target gene expression and target region accessibility, and regulon specificity scores derived from them. Networks of transcription factors, regions and target genes (enhancer-driven GRNs) were constructed by linking individual regulons. Transcription factor–enhancer–gene links for all subsets (osteogenesis, chondrogenesis, fibrogenesis, early joint progenitors, immune and Schwann) can be found in Supplementary Table 9.

Trajectory analysis

For pseudotime trajectory construction in the osteogenic subcompartment, non-cycling droplets were subsetted, and X_scVI was used as projections for palantir to obtain multiscale diffusion space. A neighbourhood graph was generated on the diffusion space using sc.pp.neighbors, and the first two principal components were used as initial positions to create ForceAtlas2 embeddings using sc.tl. draw_graph. scFates[85] (v1.0.3) was used to predict a principle graph that captures the differentiation path. The force-directed embeddings and principle graph were exported into R, and monocle3 (v1.0.0)[86,87] was used to compute differentially expressed genes along pseudotime using a graph-based test (morans' I)[87,88], which allows identification of genes upregulated at any point in pseudotime. The results were visualized with heatmaps using the complexHeatmap (v2.6.2)[89] and seriation (v1.3.0)[90] packages, after smoothing gene expression with smoothing splines in R (smooth.spline; d.f. = 12). Velocity analysis[91] was performed using scvelo[92] (v0.2.3). Spliced and unspliced read counts were computed with velocyto (v0.17.17) from the unprocessed data, before using scvelo.pp.moments, scvelo.tl.velocity and scvelo.tl.velocity_graph to compute velocities for the preprocessed droplets. cytoTRACE[93] was used (through the CellRank[94] (v2.0.2) implementation) to obtain another prediction of directionality, independent of RNA velocity (based on the assumption that the number of expressed genes decreases throughout differentiation).

Cavitation enrichment score

To approximate the timing of cavitation onset, we computed a cavitation enrichment score using sc.tl.score_genes() in scanpy on a specific gene set within the mesenchymal and muscle compartments of the hip, shoulder and knee joints comprising *CD44*, *HAS2*, *ABCC5*, *HMMR*, *MSN* and *UDPGD*, derived from literature and Gene Ontology terms, which encompass hyaluronan biosynthetic processes and hyaluronan synthase activity. We excluded genes with low expression levels in our data, such as *HAS3*. For pathway analysis, we retrieved gene sets corresponding to all 18,640 Gene Ontology terms, and computed the correlation between their enrichment scores and cavitation enrichment scores.

In silico transcription factor perturbations

CellOracle[95] (v0.12.0) was used with the osteogenesis trajectory created with scFates[85], and the regulons predicted with SCENIC+[83] for the same cells were imported into CellOracle as a base GRN. Cells were aggregated into meta-cells of 10–15 cells, and linear models explaining transcription factor from target gene expression were fitted with CellOracle per cell cluster. Regulon-based transcription factor perturbation vectors were inferred using the cell cluster-specific models to predict effects of transcription factor overexpression and knockout. Diffusion pseudotime[96] was then computed for intramembranous and endochondral ossification lineages separately by selecting corresponding starting points. The pseudotime gradients were used to derive pseudotime-based differentiation vectors, and the pseudotime-perturbation vector cross-product was computed to obtain perturbation scores. These perturbation scores indicate whether the in silico perturbation of a transcription factor is consistent with or opposes differentiation along a lineage (osteogenesis). The simulations were carried out systematically, overexpressing and knocking out all transcription factors in the GRN. For each transcription factor and condition, the perturbation scores were then averaged per cell cluster and summarized in a table to screen for transcription factors promoting or inhibiting osteogenesis.

fGWAS analysis

fGWAS analysis[97] was applied to identify disease-relevant cell clusters as described in detail[55] (https://github.com/natsuhiko/PHM). The model makes use of full summary statistics from GWAS, linking single-nucleotide polymorphisms (SNPs) to genes, and captures a general trend between gene expression and disease association of linked loci for each cell cluster. At the same time, the model also corrects for linkage disequilibrium and other relevant factors. We used full GWAS summary statistics obtained from the EBI GWAS Catalog, open targets, and knee and hip osteoarthritis as well as total knee and hip replacement from ref. 62 (https://msk.hugeamp.org/downloads.html; Supplementary Table 8).

SNP2Cell

We used a network propagation[98] approach to integrate GWAS summary statistics and cell cluster marker gene-based scores for prioritizing disease-relevant and cell cluster-specific subunits of our transcription factor network. First, scores per SNP were computed from downloaded summary statistics and weighted by linkage disequilibrium. Then, the scores were mapped to a GRN, here an enhancer-driven GRN computed with SCENIC+ for the corresponding lineage. As the used networks contain transcription factors and target genes, and also regions with transcription factor-binding sites as nodes, SNP scores were mapped to both genes and regions, representing distal regulatory elements. The scores were then propagated across the network using a random walk with restart (or personalized page rank) process. This integrates the contribution of individual SNPs, with signals converging around relevant network nodes. The procedure was repeated with 1,000 randomly permuted scores to compute permutation-test results and z-scores. Next, differential expression-based marker gene scores for each cell cluster were propagated in the same way, resulting in cell cluster specificity scores for each network node. The SNP and expression-based scores were then combined per cell cluster (as in ref. 99) by using the minimum for each node. The final scores were thresholded, and the resulting connected components were obtained as enriched subnetworks. The method has been compiled into a tool that we called SNP2Cell, which is available at https://github.com/Teichlab/snp2cell.

Cell–cell interactions

Ligand–receptor interactions were inferred using 'cpdb_analysis_method.call' in CellPhoneDB (v4.0.0). We included genes expressed in more than 10% of cells within each cluster. Inferred interactions with a $P > 0.001$ were removed. We used NicheNet (v1.1.1) to identify different interactions between endochondral and intramembranous niches. We first calculated DEGs of osteogenic clusters and tip cells across the two niches using the Wilcoxon test implemented in Seurat, and minimum log fold change per cluster was used to summarize the differentially expressed ligands and receptors. The top 1,000 DEGs were used to calculate ligand activities. We prioritized the ligand–receptor links using default settings. The top ten ligands and their top-scoring receptors were visualized using heatmaps.

Drug2Cell analysis

Drug and target gene information for humans (*Homo sapiens*) were gathered from the ChEMBL database (https://www.ebi.ac.uk/chembl/). For the teratogenic drugs targeting, we searched the clinically approved molecules that target genes encoding their reported targets and curated a list of 65 clinically approved drugs from the chEMBL database, which carried warnings of teratogenicity (Supplementary Table 6). Drug scores were calculated as previously described[58]. Subsequently, we introduced drug categories for each drug according to broad clinical utility. The Drug2Cell Python package is available at GitHub (https://github.com/Teichlab/drug2cell).

CellHint label harmonization

First, fastq files from the Zhang et al.[5] dataset were remapped using STARSOLO to a common genome reference (GRCh38-2020-A-2.0.0) as per the workflow performed for the Multiome data. Cellbender was applied to remove background counts represented as simulated

Article

ambient RNA. We intersected this matrix with barcodes from the post-quality control counts matrix from Zhang et al., and scVI was then used to integrate this with our snRNA-seq data, accounting for categorical covariates of sample donor and droplet technology (cell or nuclei), as well as continuous variables of total counts, the percentage of ribosomal and mitochondrial counts, and cell cycle scores 'S_score' and 'G2M_score' computed using the scanpy package. Latent variables obtained from this were then used to determine neighbourhoods followed by dimensionality reduction in UMAP. Cluster labels from Zhang et al. were then used as labels for CellHint harmonization in the cellhint. harmonize() alignment function. Cellhint.treeplot() was used to examine and semi-automatically align the labels across the two datasets. Gene expression profiles of marker genes were used to verify alignment of clusters across the two datasets.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

High-throughput raw sequencing data in this study are available from ArrayExpress (www.ebi.ac.uk/arrayexpress) with the accession number E-MTAB-14385. Processed snRNA–scATAC-seq, Visium and ISS data are available for visualization and can be downloaded from https://developmental.cellatlas.io/skeleton-development. Source data are provided with this paper.

Code availability

ISS-Patcher is available at https://github.com/Teichlab/iss_patcher. SNP2Cell is available at https://github.com/Teichlab/snp2cell. The custom code for the other analyses performed in this study is available at GitHub (https://github.com/Teichlab/skeletal_dev_atlas).

63. Lee, H., Marco Salas, S., Gyllborg, D. & Nilsson, M. Direct RNA targeted in situ sequencing for transcriptomic profiling in tissue. *Sci. Rep.* **12**, 7976 (2022).
64. Renier, N. et al. Mapping of brain activity by automated volume analysis of immediate early genes. *Cell* **165**, 1789–1802 (2016).
65. Li, T. et al. WebAtlas pipeline for integrated single cell and spatial transcriptomic data. *Nature* https://doi.org/10.1038/s41592-024-02371-x (2024).
66. VasylVaskivskyi/microaligner: image registration (alignment) software for large microscopy images. *GitHub* https://github.com/VasylVaskivskyi/microaligner (2022).
67. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
68. Gataric, M. et al. PoSTcode: probabilistic image-based spatial transcriptomics decoder. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.12.464086 (2021).
69. Virshup, I., Rybakov, S., Theis, F. J., Angerer, P. & Wolf, F. A. anndata: Access and store annotated data matrices. *J. Open Source Softw.* https://doi.org/10.21105/joss.04371 (2021).
70. Litviňuková, M. et al. Cells of the adult human heart. *Nature* https://doi.org/10.1038/s41586-020-2797-4 (2020).
71. Heaton, H. et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).
72. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, giaa151 (2020).
73. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
74. Bredikhin, D., Kats, I. & Stegle, O. MUON: multimodal omics analysis framework. *Genome Biol.* **23**, 42 (2022).
75. Polański, K. et al. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).
76. Ashuach, T. et al. MultiVI: deep generative model for the integration of multimodal data. *Nat. Methods* **20**, 1222–1231 (2023).
77. Megas, S., Lorenzi, V. & Marioni, J. C. EmptyDropsMultiome discriminates real cells from background in single-cell multiomics assays. *Genome Biol.* **25**, 121 (2024).
78. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019).
79. Popescu, D.-M. et al. Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).
80. Xu, C. et al. Automatic cell type harmonization and integration across Human Cell Atlas datasets. Preprint at *bioRxiv* https://doi.org/10.1101/2023.05.01.538994 (2023).
81. Domínguez Conde, C. et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022).
82. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2021).
83. Bravo González-Blas, C. et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* https://doi.org/10.1038/s41592-023-01938-4 (2023).
84. Moerman, T. et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159–2161 (2019).
85. Faure, L., Soldatov, R., Kharchenko, P. V. & Adameyko, I. scFates: a scalable Python package for advanced pseudotime and bifurcation analysis from single-cell data. *Bioinformatics* **39**, btac746 (2023).
86. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
87. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* https://doi.org/10.1038/s41586-019-0969-x (2019).
88. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).
89. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btw313 (2016).
90. Hahsler, M., Hornik, K. & Buchta, C. Getting things in order: an introduction to the R package seriation. *J. Stat. Softw.* https://doi.org/10.18637/jss.v025.i03 (2008).
91. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
92. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
93. Gulati, G. S. et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405–411 (2020).
94. Lange, M. et al. CellRank for directed single-cell fate mapping. *Nat. Methods* **19**, 159–170 (2022).
95. Kamimoto, K. et al. Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742–751 (2023).
96. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
97. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
98. Cowen, L., Ideker, T., Raphael, B. J. & Sharan, R. Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.* **18**, 551–562 (2017).
99. Paull, E. O. et al. Discovering causal pathways linking genomic events to transcriptional states using tied diffusion through interacting events (TieDIE). *Bioinformatics* **29**, 2757–2764 (2013).

Extended Data Fig. 1 | Droplet dataset overview. a. UMAP-embedding of RNA droplets with broad cell cluster labels. **b**. Dotplot with marker genes for each cluster. **c**. Dotplot with marker genes for each cell compartment. **d**. Skeletal region metadata displayed on RNA (left), ATAC (middle) and co-embedding (Right). **e**. Developmental stage metadata (/PCW) displayed on RNA (left), ATAC (middle) and co-embedding (Right). **f**. Heatmap of concordance of cell compartment labels across RNA and ATAC modalities, scale bar shows the proportion of intersected cells across RNA and ATAC compartment (scaled by row).
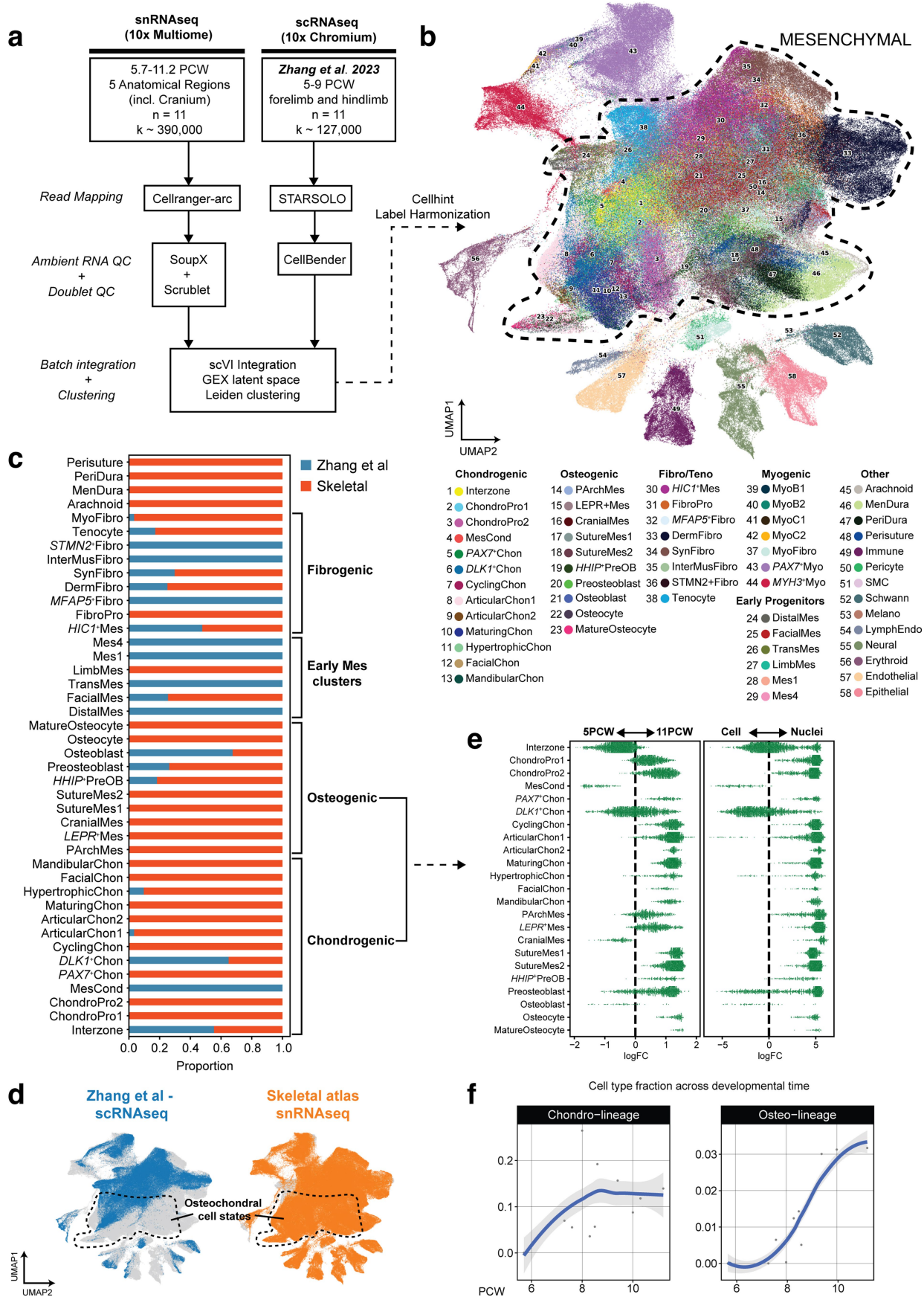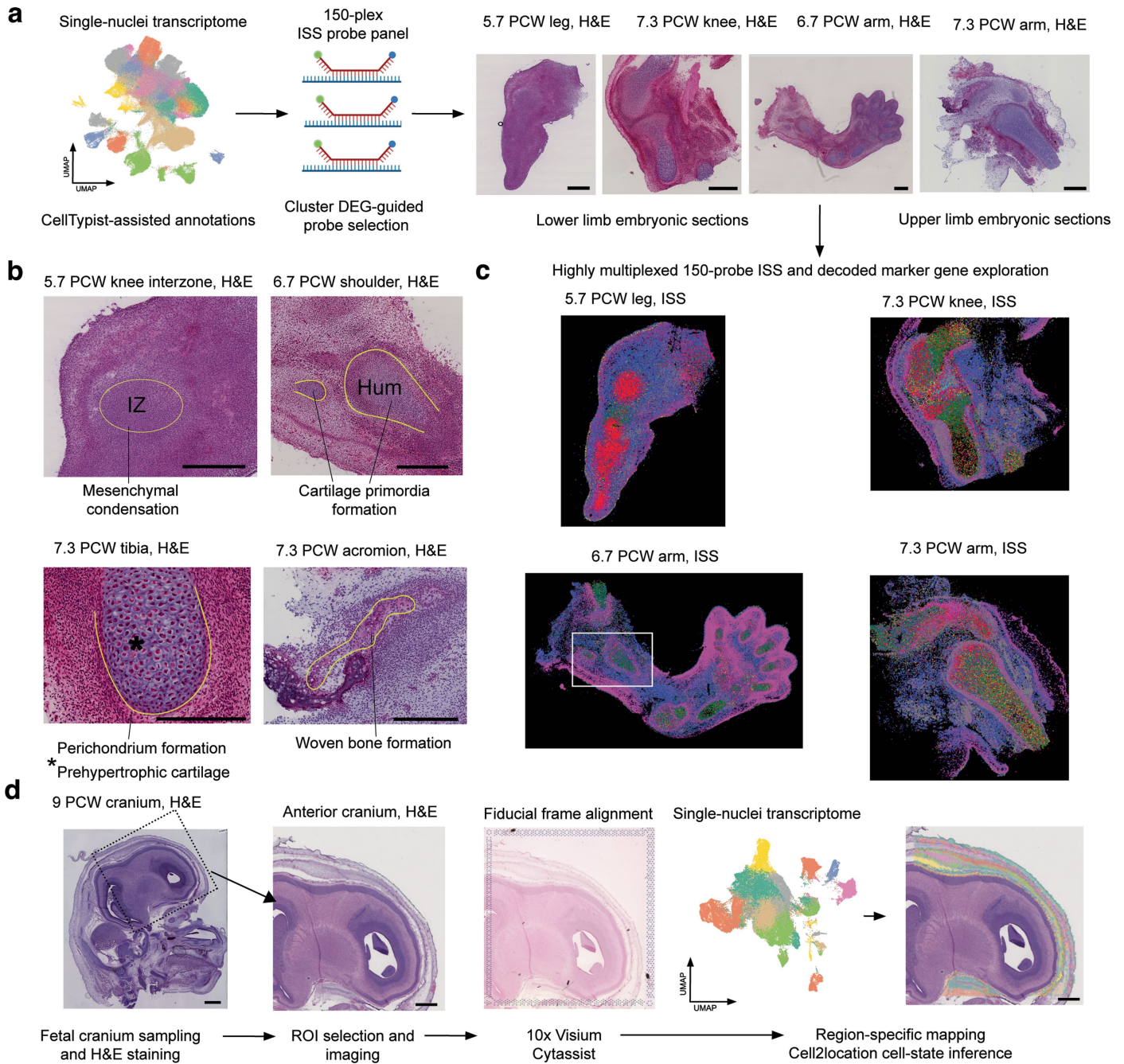
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Integrative analysis of skeletal atlas with published limb atlas data. a. Computational workflow for integrating Zhang et al. data by remapping to a common reference, followed by quality control and batch integration using scVI and CellHint (see Methods). **b**. UMAP of the integrated datasets with harmonised cluster labels. **c**. Barplot of proportion of cell type from each of the two studies. **d**. UMAP of the integrated datasets showing droplets originating from each dataset. Dotted line shows regions of osteochondral cell states. **e**. MILO analysis of differential abundance of each cell type across age (PCW) and modality. **f**. Line graph of cell type fraction across developmental time (/PCW) in the chondrogenesis and osteogenesis lineages, showing enrichment of both lineages toward the latter parts of the first trimester. Data: mean with 95% confidence interval, n = 12 sample of donors.

a

Single-nuclei transcriptome

UMAP / UMAP

CellTypist-assisted annotations

150-plex ISS probe panel

Cluster DEG-guided probe selection

5.7 PCW leg, H&E · 7.3 PCW knee, H&E · 6.7 PCW arm, H&E · 7.3 PCW arm, H&E

Lower limb embryonic sections · Upper limb embryonic sections

b

5.7 PCW knee interzone, H&E

IZ

Mesenchymal condensation

6.7 PCW shoulder, H&E

Hum

Cartilage primordia formation

7.3 PCW tibia, H&E

*

Perichondrium formation
*Prehypertrophic cartilage

7.3 PCW acromion, H&E

Woven bone formation

c

Highly multiplexed 150-probe ISS and decoded marker gene exploration

5.7 PCW leg, ISS · 7.3 PCW knee, ISS

6.7 PCW arm, ISS · 7.3 PCW arm, ISS

d

9 PCW cranium, H&E · Anterior cranium, H&E · Fiducial frame alignment · Single-nuclei transcriptome

UMAP / UMAP

Fetal cranium sampling and H&E staining → ROI selection and imaging → 10x Visium Cytassist → Region-specific mapping Cell2location cell-state inference

Extended Data Fig. 3 | Spatial transcriptomics analysis and workflow.
a. Workflow schematic showing UMAP-embedding of RNA droplets of the appendicular joints manually annotated informed by Celltypist utilising Zhang et al., 2023 labels. Per-cluster DEG from these were used to design a 150-plex probe panel, subsequently applied to sectioned samples of 5-7 PCW synovial joints (Right). Scale bars = 500 μm, Images are representative of three donors.

b. Regions of cellular niches within the samples utilised for ISS imaging. Scale bars = 500 μm, Images are representative of three donors. **c**. Visualisation of probe distribution on ISS imaging data applied to histological sections in a, white box shows nascent shoulder joint in 6.7 PCW arm. **d**. Workflow schematic for 10x Visium Cytassist applied to 9 PCW cranium, Scale bars = 2000 μm.
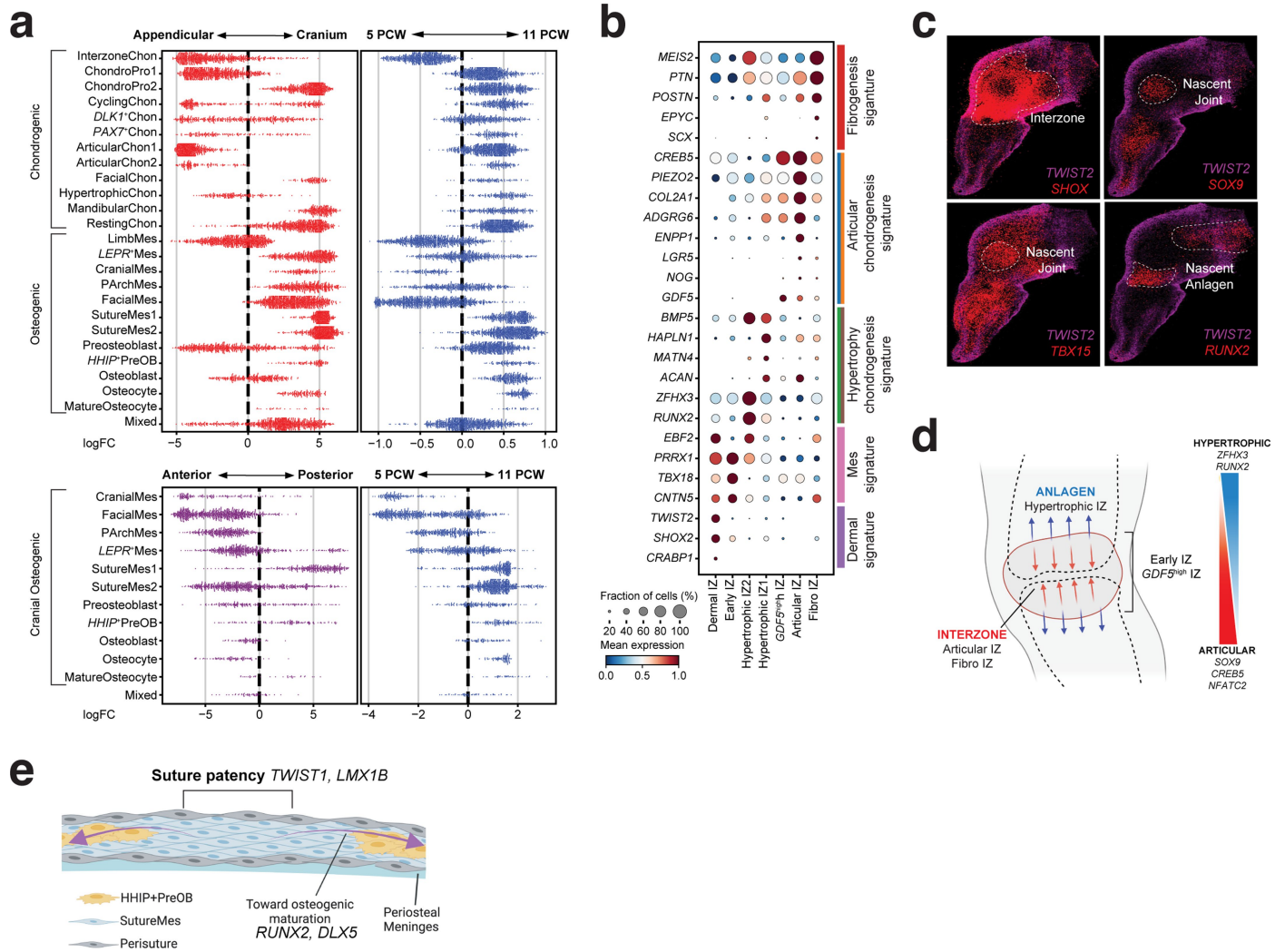
Extended Data Fig. 4 | Cell states in the limb and cranial joints. a. Differential abundance (MILO, see Methods) of cell states across lineages against multiple comparators. Top: Osteogenesis and chondrogenesis clusters across region and PCW. Bottom: Cranial osteogenesis subclusters across region and PCW. Mixed: other related cell types that are not included in each plot. **b**. Dotplot of normalised expression of genes associated with skeletal development within InterzoneChon subclusters (IZs). **c**. RNA-ISS of the 5.7 PCW hindlimb, showing expression patterns of select genes. **d**. Schematic of knee joint interzone formation, gradients showing genes associated with zonated hypertrophic and articular phenotypes. **e**. Schematic illustrating TF gradients across regions of the coronal suture. Schematics in panels **d,e** were created using BioRender (https://biorender.com).
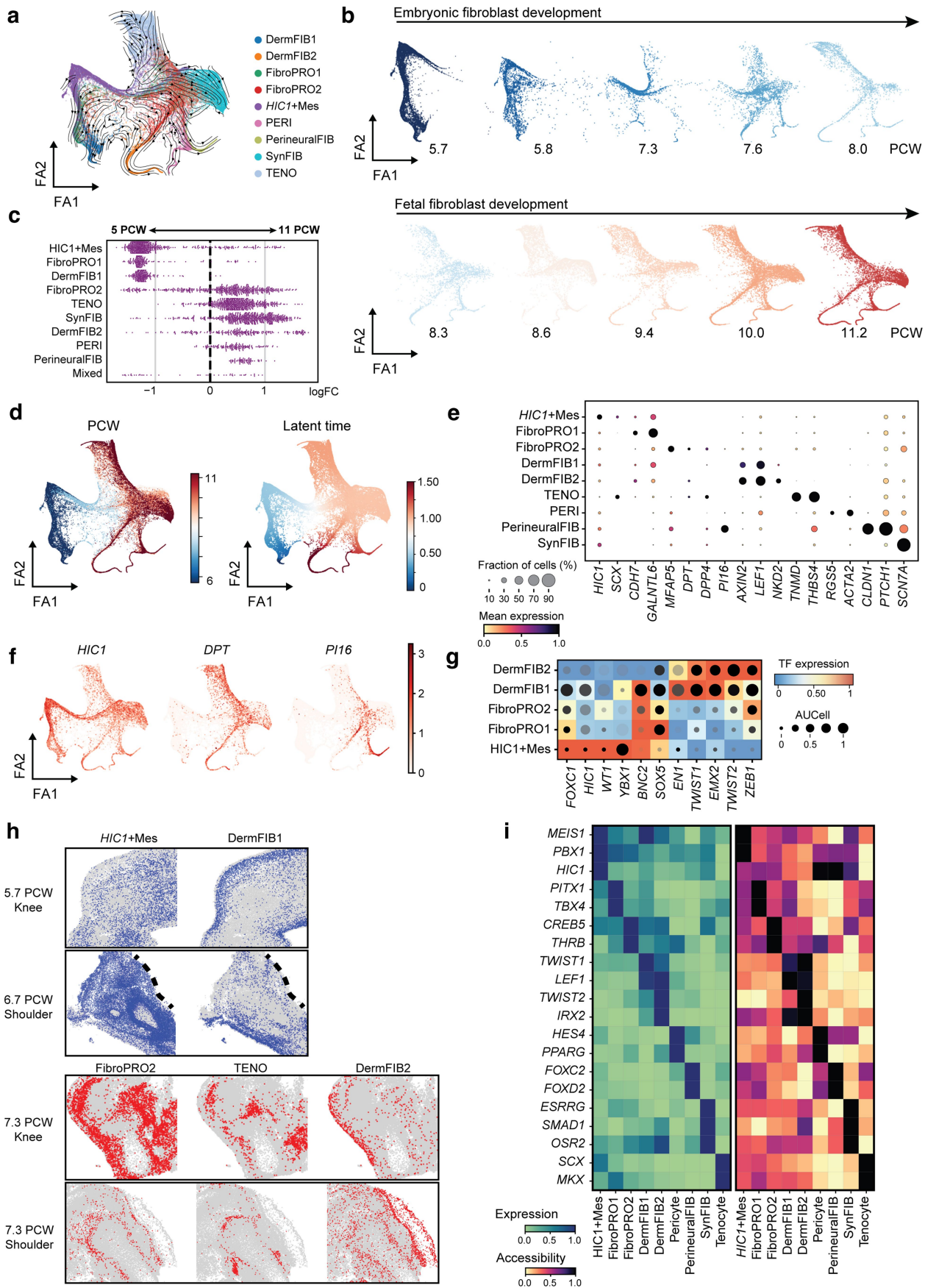
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Fibroblast cell states in the appendicular joints.
a. Force-directed embedding of fibroblast cell states. Arrows show RNA velocity. **b**. Developmental stage (/PCW) displayed over force directed embedding. **c**. MILO cell type abundance swarm plot for fibroblast cell states over developmental time. **d**. Developmental stage (/PCW) and pseudotime (latent-time) on force-directed embedding of fibroblast clusters. **e**. Dotplot showing marker genes expression per cell state. **f**. Visualisation of marker genes on force-directed embedding. **g**. TF activity across select fibroblast clusters. Colour shows normalised expression, dot size shows target gene accessibility (ATAC AUCell) and dot shade (grayscale) shows target gene expression (GEX AUCell). **h**. Spatial location of the imputed fibroblast clusters in the embryonic limb (<7 PCW, top) and foetal limb (>7 PCW, bottom). Dotted line shows sectioning artefact on one side of the 6.7 PCW shoulder sample. **i**. Heatmap showing the gene expression and accessibility (gene score from ArchR) of representative TFs in each cell type.
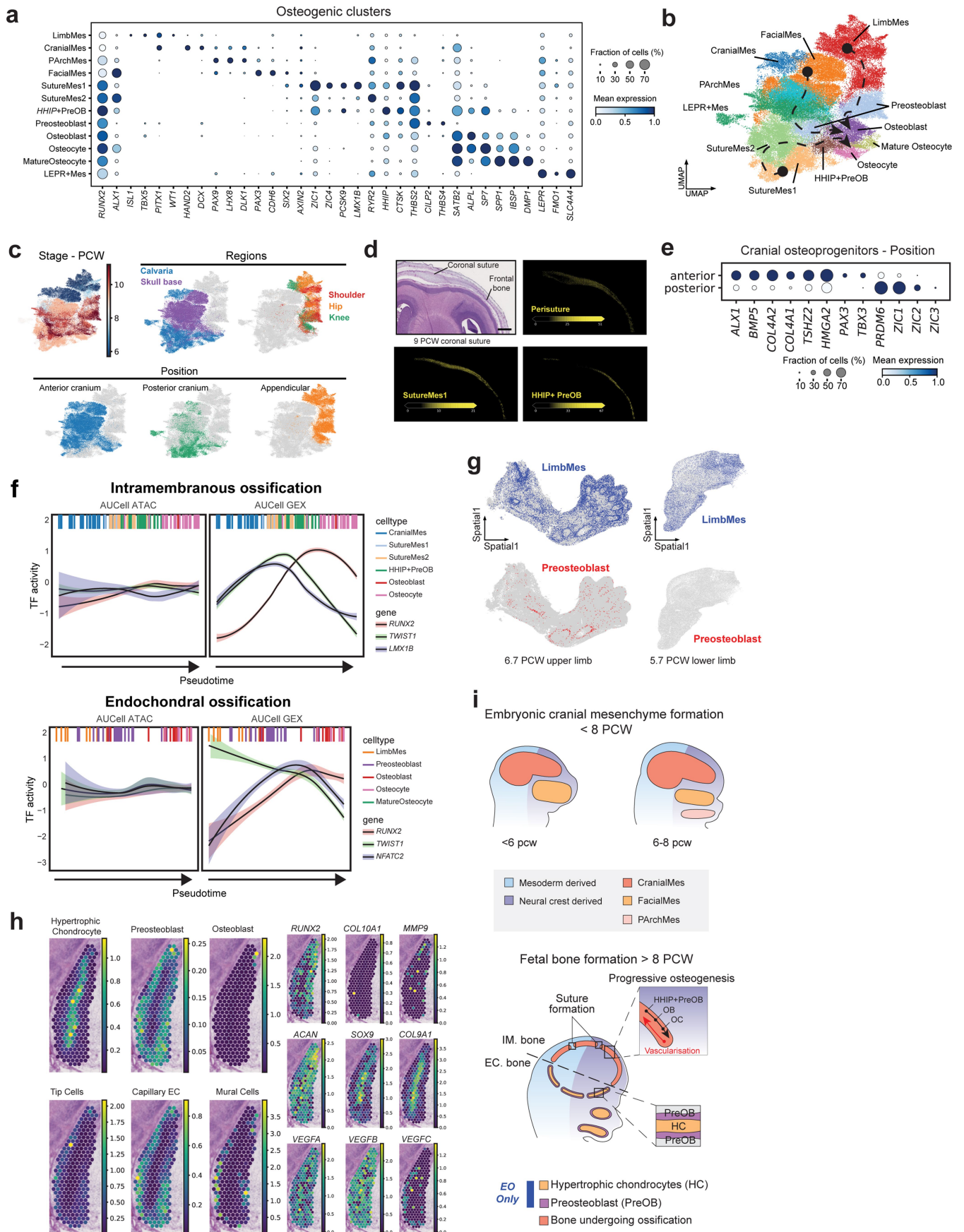
Extended Data Fig. 6 | See next page for caption.

Article

Extended Data Fig. 6 | Osteogenesis cell states. a. Dotplot of marker genes including those previously described in mouse suture cells across osteogenic clusters. **b**. UMAP embedding of osteogenesis lineage clusters. Arrows show manually annotated paths from progenitor to mature osteogenesis clusters. **c**. Stage (PCW), region and position metadata overlaid on UMAP embedding. **d**. Cell2location cell state enrichment in the coronal suture and frontal bone. Scale bar = 2000 μm. **e**. Dotplot with marker genes in relative anatomical (anterior-posterior) positions in the cranial osteogenic cell states. **f.** Line plot of GEX and ATAC AUC scores across pseudotime for selected TFs and clusters in endochondral and intramembranous osteogenesis lineages. Bands show the 95% CI. Top: higher GEX and ATAC of inhibitory TFs *TWIST1* and *LMX1B* is observed at the beginning of the trajectory (CranialMes), while *RUNX2* shows higher expression and accessibility towards the end of the trajectory. The crossing point at SutureMes1/2 suggests a critical regulatory balance at this cell state. **g**. Imputed osteogenic cell states in the appendicular ISS data **g**. Schematic showing cranial progenitor maturation across embryonic (top) and foetal (bottom) stages. **h**. Cell2location cell cluster enrichment in the sphenoid bone of the skull base (endochondral niche) (Left), gene expression in the same region (right). **i**. Schematic of cranial formation in the embryonic and fetal cranium, showing suture formation and progressive osteogenesis in the fetal cranium, with hypertrophic chondrocytes in the skull base. EO: endochondral osteogenesis, IO: intramembranous osteogenesis.
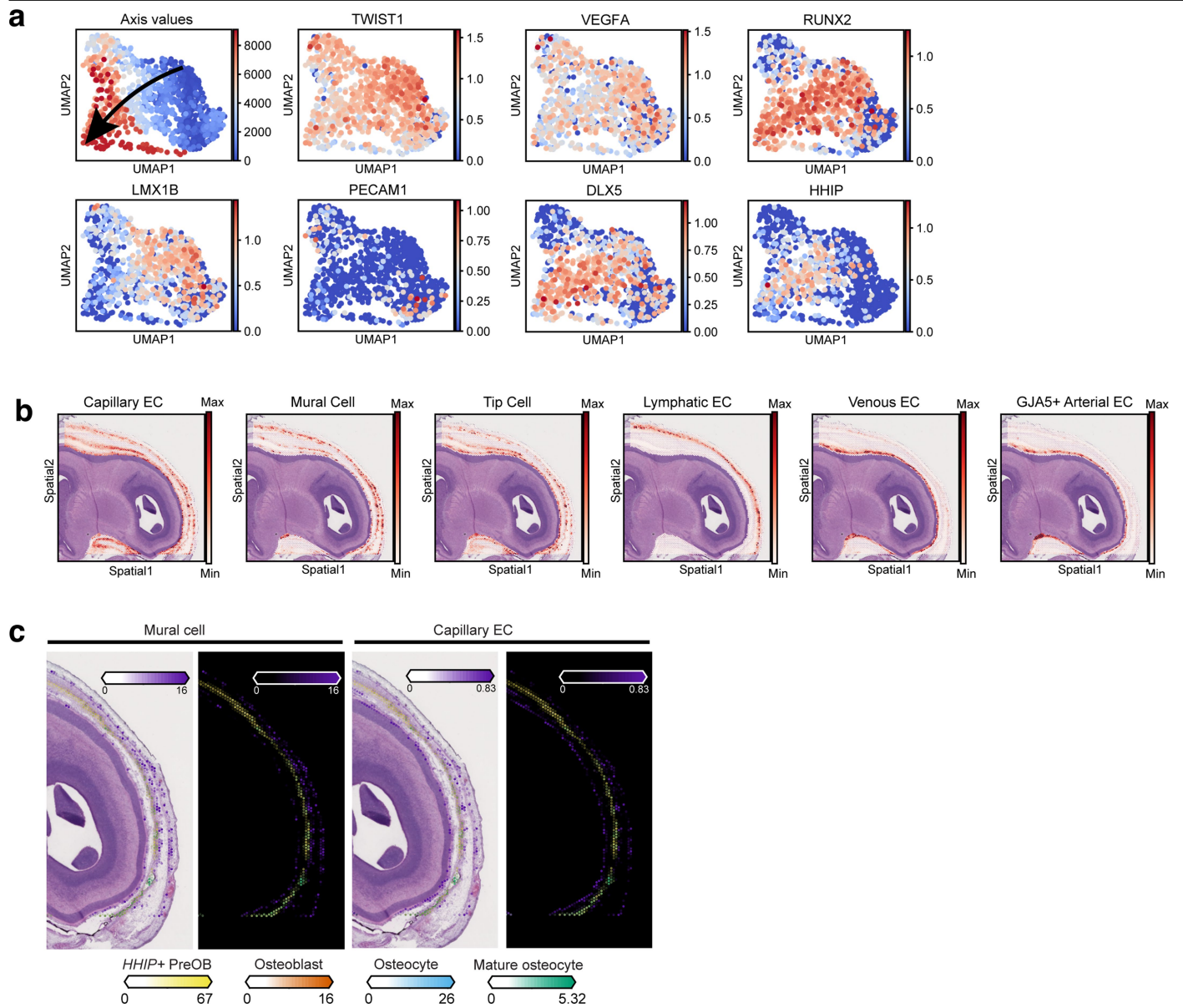
a

Intramembranous osteogenesis
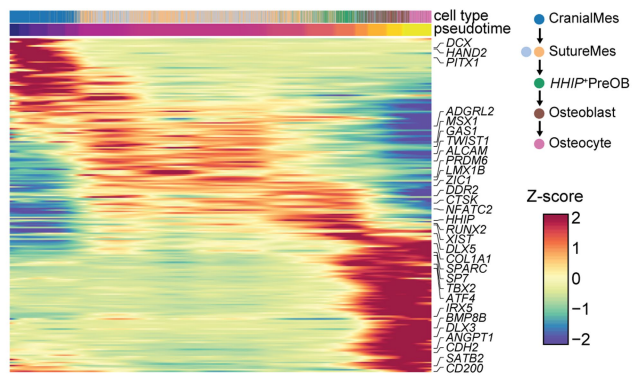
Endochondral osteogenesis
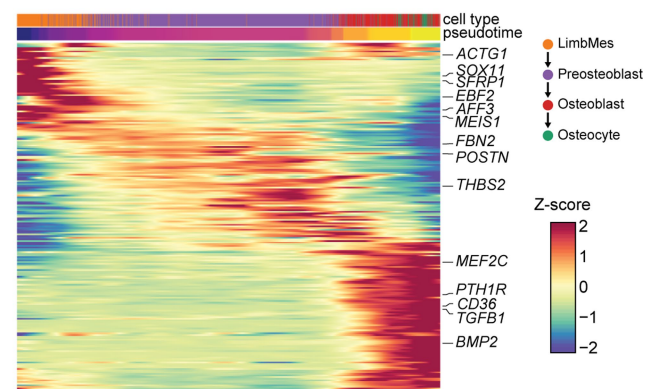
Extended Data Fig. 7 | Gene changes across osteogenesis. a. Heatmaps of top differentially expressed genes along the Intramembranous and Endochondral trajectory pseudotime (spatial autocorrelation test with monocle3, see Methods).

a

b

c

Extended Data Fig. 8 | Cellular interactions within the suture niche.
a. Clustering of Visium voxels with organ-axis values and gene expression.
b. Cell2location endothelial cell state enrichment in the anterior portion of the cranium **c**. Cell2location enrichment of osteogenic cell states alongside mural (left) and capillary EC (right) on the anterior portion of the frontal bone.
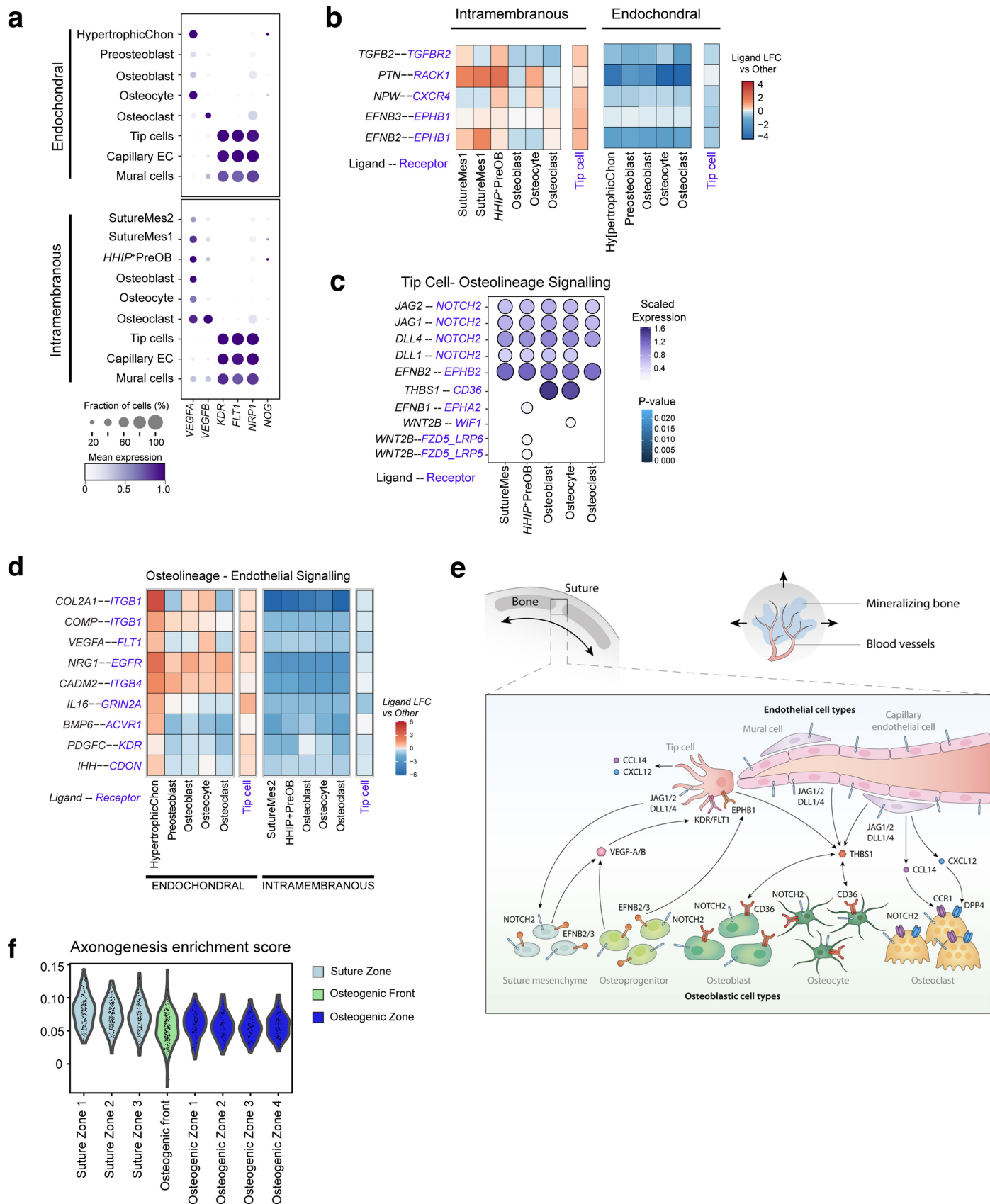
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Cellular interactions within the suture niche.
a. Dotplot of gene expression of receptor-ligands related to angiogenesis in IM and EC clusters. **b**. NicheNet predictions of differentially enriched cell-cell interaction pairs between endothelial tip cells and osteolineage cells in the IM pathway, compared to the EC pathway. **c**. CellphoneDB predictions of Ligand-receptor pairs between endothelial tip cells and cells of the osteogenic lineage.
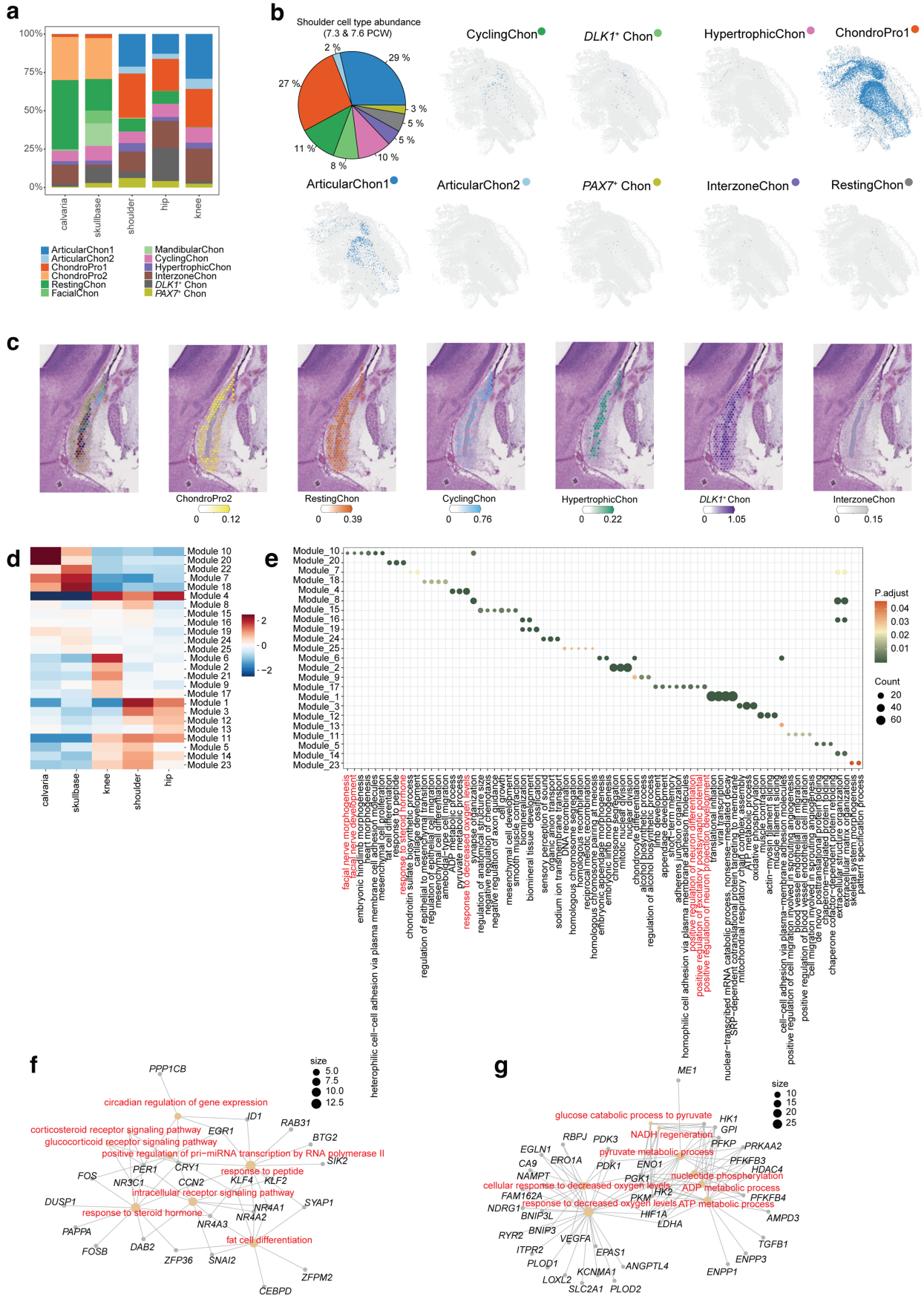
d. NicheNet inferred differential cell-cell interactions between osteogenic cell states and Tip cells in the EC and IM niches. **e**. Schematic of endothelial and osteogenic cell state interactions at the boundaries of the suture mesenchyme. **f**. Enrichment of axonogenesis gene set from the Gene-ontology 2023 database across the organ-axis spatial bins.
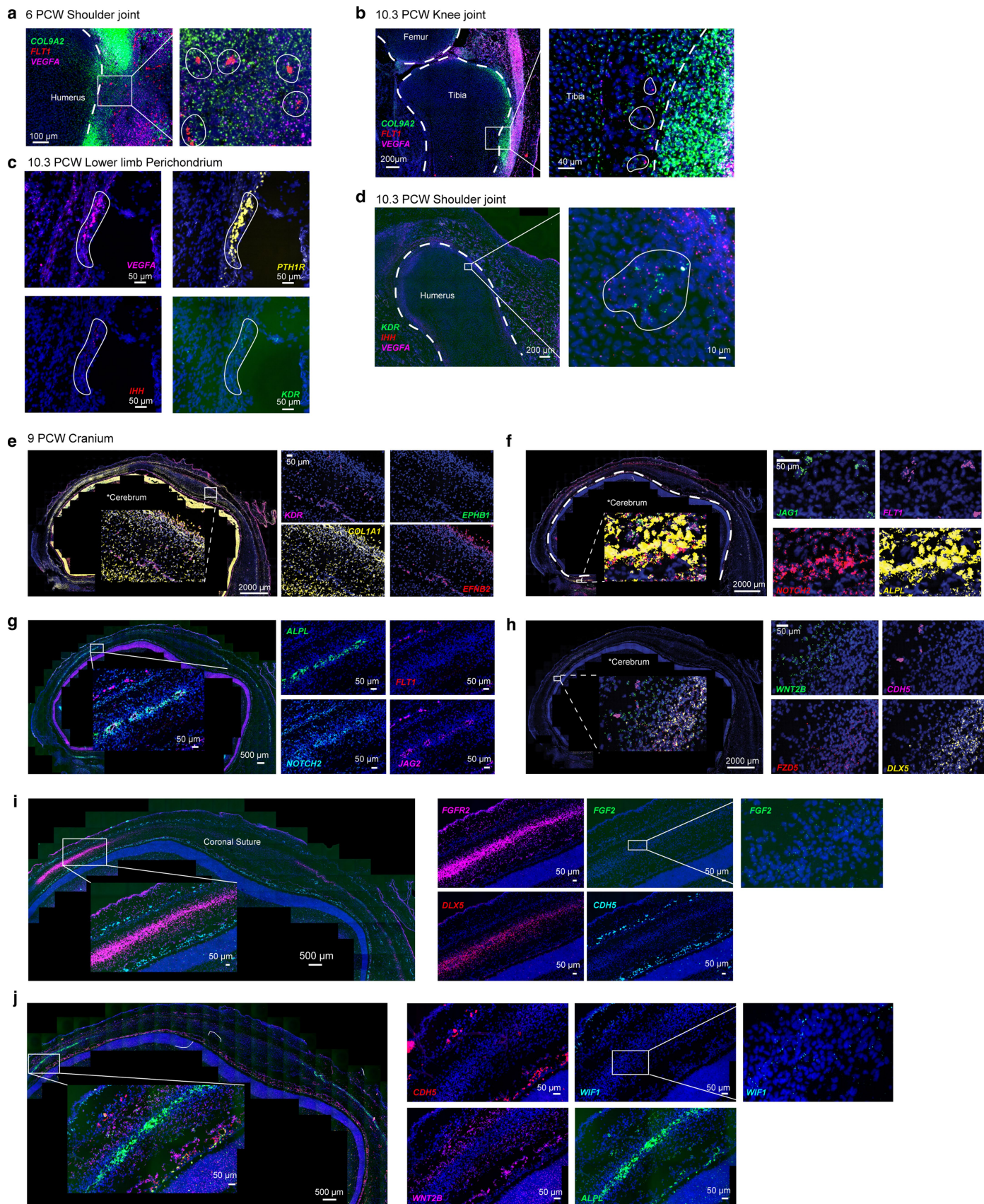
Extended Data Fig. 10 | Cell-Cell interactions in the skeletal joints. RNA-ISH of tissue sections from **a**. shoulder joint, **b**. knee joint, **c**. lower limb, **d**. shoulder joint, **e-j**. cranium, showing expression patterns of *COL9A2*, *FLT1*, *VEGFA*, *PTH1R*, *IHH*, *KDR*, *EPHB1*, *EFNB2*, *COL1A1*, *JAG1*, *NOTCH2*, *ALPL*, *JAG2*, *WNT2B*, *CDH5*, *FZD5*, *DLX5*, *FGFR2*, *FGF2*, and *WIF1*. Images are representative of sections from three donors (**a-j**).

a

b Shoulder cell type abundance (7.3 & 7.6 PCW)

CyclingChon   *DLK1*+ Chon   HypertrophicChon   ChondroPro1

ArticularChon1   ArticularChon2   *PAX7*+ Chon   InterzoneChon   RestingChon

c

ChondroPro2   RestingChon   CyclingChon   HypertrophicChon   *DLK1*+ Chon   InterzoneChon
0   0.12   0   0.39   0   0.76   0   0.22   0   1.05   0   0.15

d

e

f

g

Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Spatial locations and gene signatures of chondrocytes. a. Barplot showing the distribution of cell cluster abundance of chondrocytes per anatomical region and per PCW. **b**. Pie chart (left) showing the percentage of cell clusters in the shoulder at 7.3 and 7.6 PCW. Spatial plots (right) showing predicted cell clusters from ISS-Patcher cell cluster imputation. Colours represent the corresponding cell clusters. **c**. Histological view and annotations of chondrocytes in the sphenoid sections from 10x Cytassist Visium. **d**. Heatmap showing kernel values of module genes per anatomical region. **e**. Top 3 enriched biological process GO terms for each module. The GO terms were selected based on gene number, and terms of the same gene number remain. Colour represents the adjusted p-values, while the dot size represents the gene number of each GO term (see Supplementary Table 11). The statistical test was performed using a two-sided hypothesis. P values were adjusted with the Benjamini–Hochberg method. **f**. Network visualisation of enriched genes of calvaria and skull base-specific GO terms. **g**. Network visualisation of enriched genes of appendicular-specific GO terms.
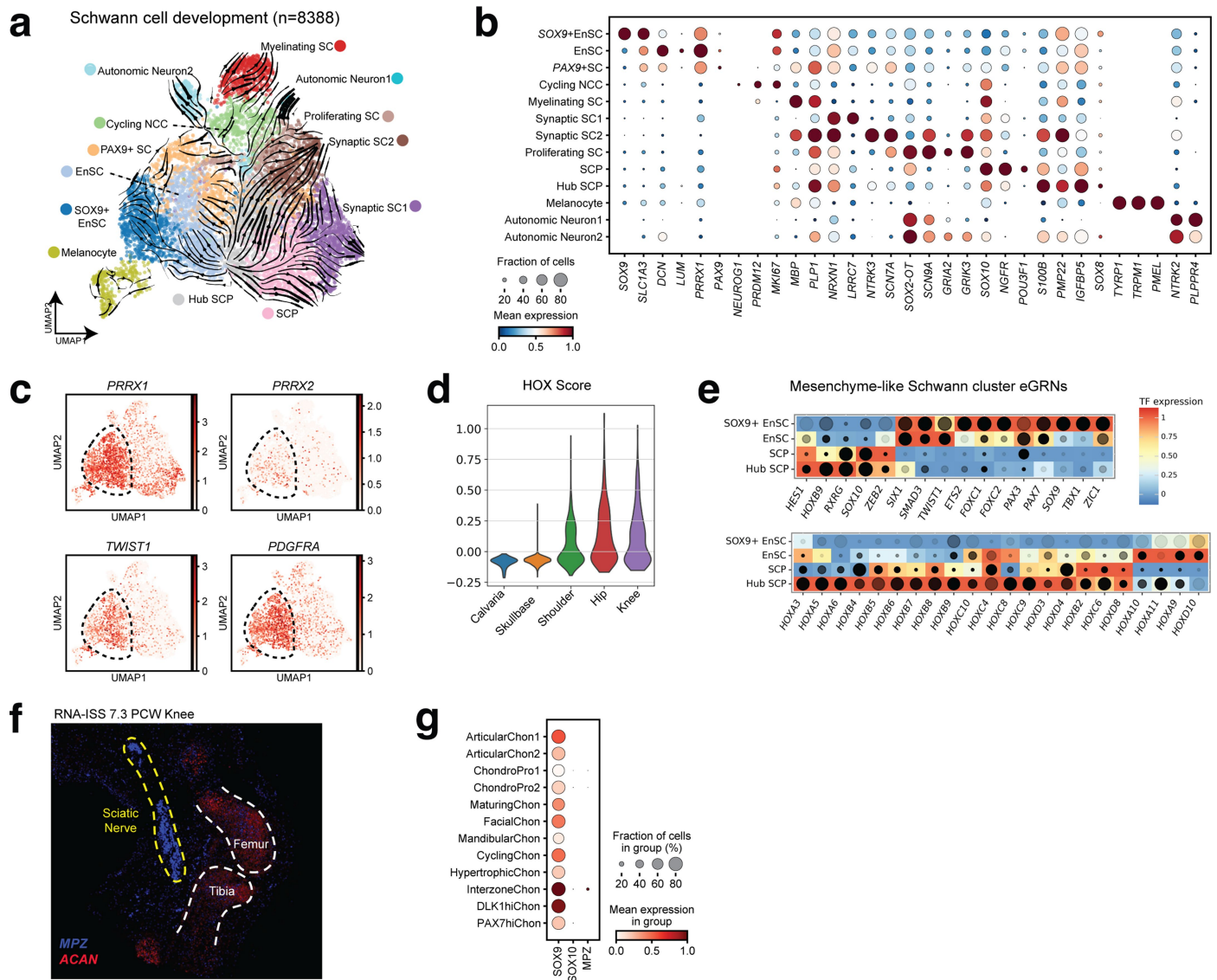
Extended Data Fig. 12 | Schwann cell lineage development. a. RNA-velocity streamlines displayed on embedding of Schwann lineage cell states, with numerous cell states emerging from Hub SCP (Schwann cell precursor). **b**. Dotplot showing marker gene expression across Schwann cell states. **c**. Expression of mesenchyme associated genes in the Schwann UMAP. **d**. Enrichment of *HOX* genes per anatomical region of the Schwann compartment. **e**. TF expression across the putative chondrogenic lineage of the Schwann clusters. Colour shows normalised expression, dot size shows target gene accessibility (AUCell) and dot colour (grayscale) shows target gene expression (GEX AUCell). **f**. Marker genes of ISS RNA probes in the ISS image of the knee joint demonstrating nerve-associate enrichment of Schwann markers, and the presence of *MPZ* staining in the developing bone. Images are representative of sections from three donors. **g**. Dotplot showing gene expression of Schwann cell markers across chondrocyte clusters.

a Knee OA signal in Articular Chondrocyte GRN

b Monogenic Musculoskeletal Disease Genes
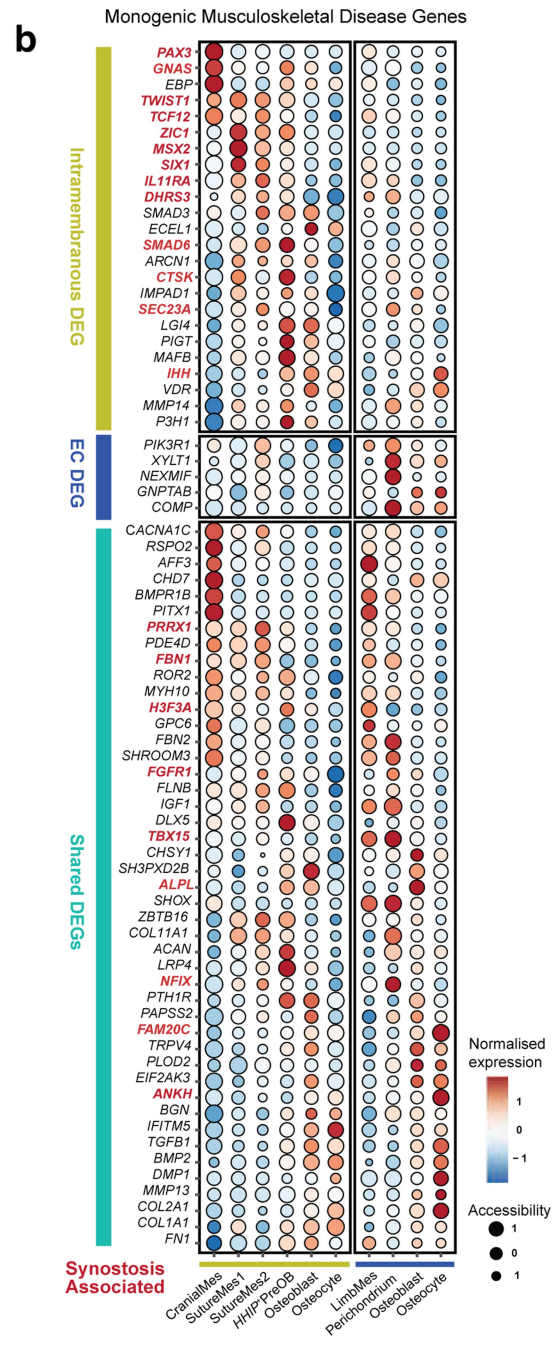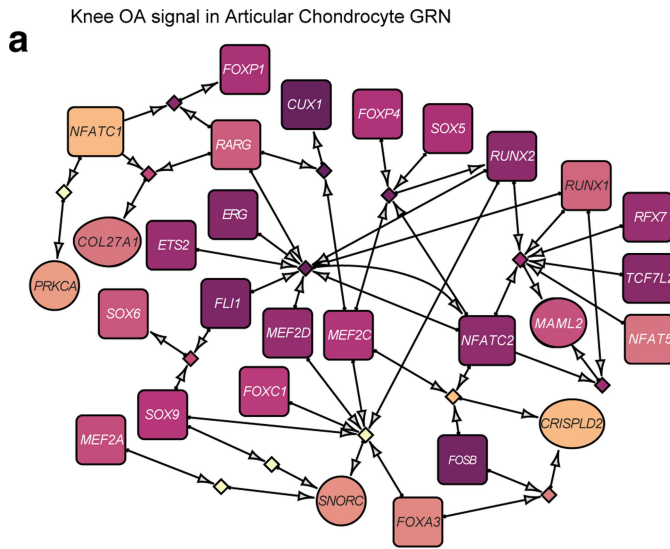
Extended Data Fig. 13 | Musculoskeletal disease gene regulation. a. eGRN extracted from SNP2Cell (see Methods) showing TFs, target genes and enhancers enriched for knee OA in articular chondrocytes. Brighter colours correspond to a stronger enrichment. **b**. Genes with known mutations in musculoskeletal monogenic conditions that are differentially expressed along the intramembranous, endochondral, or both ossification trajectories. The expression of the genes and accessibility of linked enhancers are shown. Marked in red: genes with a known role in craniosynostosis, the premature fusion of bone plates in the skull.

a
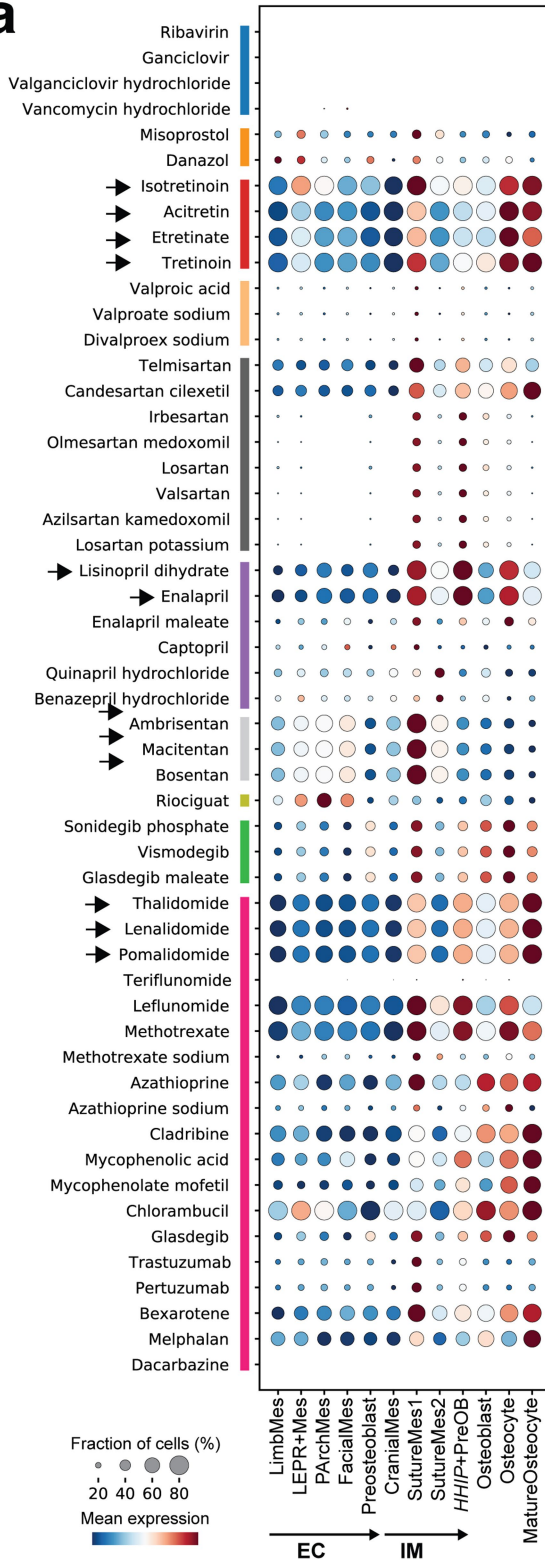
Extended Data Fig. 14 | Drug targets in osteogenesis cell states. a. Dotplot with drug2cell (chEMBL) enrichment of teratogenic drug target gene expression across cell states of the osteogenic trajectory. Selected highly enriched drugs target gene expression are marked with arrows (y-axis) and cell states unique to either endochondral (EC) or intramembranous (IM) ossification are marked on the x-axis. Patterns of different drugs predicted to affect target genes at different parts of the trajectory are visible.

nature portfolio

Corresponding author(s):   Sarah A. Teichmann

Last updated by author(s):   Oct 8, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The following software was used to process the raw data: CellRanger-ARC v2.0.0 Souporcell v2.0 SoupX v1.6.0 Scrublet v0.2.3 AMULET v1.1.0 SpaceRanger v1.1.0 LoupeBrowser v7.0 Imspector Pro 7.5.3 acquisition software ImarisFileConverter 10.1 EmptyDropMultiome 1.0.0 Velocyto 0.17.17
Data analysis	The following software was used for data analysis: Scanpy v1.9.6 ArchR v1.0.2 Muon v0.1.2 Scenic+ v1.0.0 CellOracle v0.12.0 scFates v1.0.3 CellRank v2.0.2

Monocle3 v1.0.0
Scvelo v0.2.3
CellHint v1.0.0
Milopy v0.1.1
Cell2location v0.1.4
CellPhoneDB v4.0.0
NicheNet v1.1.1
R 4.0.4
Python 1.7.1
FlowJo v9
Macs2 2.2.7.1
MultiVI (scvi: 0.6.8)
scVI 0.9.1
Celltypist 1.6.3
Palantir 1.3.3
Bbknn 1.5.1
PyCisTopic 1.0.2
PyCisTarget 1.0.2
Arboreto 0.1.6
ComplexHeatmap 2.6.2
Seriation 1.3.0
Starsolo 2.7.10a

And the following software was used for image processing:
MACS® iQ View Software v1.0
Adobe Photoshop 25.2
Microaligner 1.0.0
CellPose 3.0
PoSTcode v1.0
STRtree 2.0.6

Drug2cell code is available at https://github.com/Teichlab/drug2cell.
TissueTag code is available at https://github.com/Teichlab/TissueTag.
EmptyDropsMultiome code is available at https://github.com/MarioniLab/EmptyDropsMultiome.
fGWAS code is available at https://github.com/cellgeni/nf-fgwas.
ISS-Patcher is available at https://github.com/Teichlab/iss_patcher.
SNP2CELL is available at https://github.com/Teichlab/snp2cell.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The annotated droplet and spatial data objects for this study can be accessed here: http://developmental.cellatlas.io/skeleton-development.

High-throughput raw sequencing data in this study is available from ArrayExpress (www.ebi.ac.uk/arrayexpress) with the following accession number: E-MTAB-14385.
Human reference genome version 2020-A (GRCh38-2020-A-2.0.0) was downloaded using cellranger-arc v2.0.0.

Data analysed from Zhang et al 2023, are accessible from the following portal https://developmental.cellatlas.io/embryonic-limb .

Full GWAS summary statistics for hip-, knee- and overall osteoarthritis, as well as total knee- and hip replacement from the Genetics of Osteoarthritis (GO) Consortium were obtained from the musculoskeletal knowledge portal (https://msk.hugeamp.org/downloads.html).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Due to the early embryonic stages of the tissue samples; samples received within the first trimester from our tissue resource did NOT contain sex information. Findings apply to all sexes and sex and gender were not considered in research design. The studied processes during early embryonic and foetal development are relevant for all sexes.
Reporting on race, ethnicity, or	Donors were not distinguished based on race, ethnicity or other socially relevant groupings.

other socially relevant groupings	

Population characteristics	Population characteristics of the donors were not considered, however, donations were selected to span 5-11 post conception weeks and only healthy tissues were used. Embryonic age determined through estimated physical parameters of the embryo, e.g. Crown-rump Length were utilised to determine the approximate age of the samples and were recorded for analyses.
Recruitment	Embryonic samples were donated voluntarily by women who underwent termination of pregnancy from the Cambridge Centre for Brain Repair, Cambridge, UK
Ethics oversight	Wellcome Sanger Institute, UK AND Research and Development Department Cambridge University Hospitals NHS Foundation Trust Cambridge Biomedical Campus The human embryonic and fetal material were provided under the following references: Title: Invitro studies of postmortem human foetal neural tissue REC reference: 96/085 IRAS project ID: 95602 AND For whole-mount immunostaining, samples used in this study were obtained from terminations of pregnancy with written and informed consent from all sample donors. Samples were provided by INSERM's HuDeCA Biobank and utilised in compliance with French regulations. Authorization to use these tissues was granted by the French agency for biomedical research (Agence de la Biomédecine, Saint-Denis La Plaine, France; N° PFS19-012) and the INSERM Ethics Committee (IRB00003888).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. Sample acquisition depended on the availability of suitable tissue donations, resource availability to study samples and the study period. Available donations were selected to span the range between 5 to 11 post conception weeks in approximately equal time steps with multiple samples per time point. In addition, samples were taken at multiple selected time points for spatial transcriptomics analyses. Analyses in this manuscript are based on an integration of samples across time points, locations and modalities and sampling density across time points was chosen to be comparable to similar studies in the field.
Data exclusions	No samples were excluded from the study. Collected data was subjected to quality control after profiling. Cells with low quality according to established procedures and RNA/ATAC based metrics were excluded from the analysis. Droplets were filtered for >200 genes, and <5% mitochondrial and ribosomal reads. Moreover, cells with TSSenrichment score<7 and nFrags< 1000 were removed. Doublet removal was performed using Scrublet for RNA and ArchR as well as AMULET for ATAC.
Replication	Sample donors and number of replicates are stated in supplementary table 1. All attempts at replication were successful. Study results were checked for consistency with previous findings in the literature. For annotated cell states, we observed they were replicated across multiple donors. For computational analyses, parameters were explored extensively to ensure reproducibility of findings. For imaging data displayed, the number of replicates are stated in each figure legend.
Randomization	Randomization was not relevant to this study. Sample acquisition depended upon availability of tissue donations. Donor sample developmental stages were selected to span the available first-trimester stages (5-11PCW), ensuring approximate coverage of each post-conception week for the droplet data.
Blinding	Blinding was not required for this work as we did not allocate treatment and control groups, and did not administer treatment or test conditions.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

SP7 antibody (supplier name: abcam; Catalog: ab209484; Clone name: EPR21034; dilution: 1:500)
COL2A1 antibody (supplier name: abcam, Catalog: ab185430, Clone name: 2B1.5; dilution: 1:500)
FGFR3 antibody (supplier name: Thermo Fisher Scientific; Catalog: MA5-38521; Clone name: 2H10B4 ; Lot number: YL4148366; dilution: 1:50)
TACR3 antibody (supplier name: Thermo Fisher Scientific; Catalog: BS-0166R; Clone name: bs-0166R-A488; Lot number: BB03107336; dilution: 1:50)

Validation

literature on validation experiments are curated and accessible on the corresponding manufacturer website:
FGFR3: https://www.thermofisher.com/antibody/product/FGFR3-Antibody-clone-2H10B4-Monoclonal/MA5-38521
TACR3: https://www.thermofisher.com/antibody/product/TAC3R-Antibody-clone-2H10B4-Monoclonal/BS-0166R
SP7: https://www.abcam.com/en-gb/products/primary-antibodies/sp7-osterix-antibody-epr21034-ab209484
COL2A1: https://doc.abcam.com/datasheets/inactive/ab185430/en-us/collagen-ii-antibody-2b15-ab185430.pdf

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Prior to cell extraction, the sample tissues (~9 PCW shoulder joints) were dissected to obtain bone samples, and soft tissues were micro-dissected away. The resultant cell suspension was stained with DAPI (Invitrogen) for live-viability, FGFR3 antibody (MA5-38521, Thermo Fisher Scientific) and TAC3R antibody (BS-0166R, Thermo Fisher Scientific), and secondary

	antibodies.
Instrument	BigFoot Spectral Cell Sorter (Thermo Fisher Scientific)
Software	Thermo Fisher Scientific BigFoot proprietary software and flow Jo
Cell population abundance	The characterization of sorted population percentages were determined using flowjo, the sorted cells were not further subject to experimentation
Gating strategy	DAPI positive singlet cells were gated for DAPI staining by FACS. Sequential gating for FGFR3 and TAC3R was then conducted to identify double-positive cells. Positive controls for FGFR3 and TAC3R was conducted using human PBMCs and unstained cells were used as negative controls.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.