Lifelong single-cell profiling of cranial neural crest diversification

Peter Fabian[#1], Kuo-Chang Tseng[#1], Mathi Thiruppathy[#1], Claire Arata[#1], Hung-Jhen Chen[1], Joanna Smeeton[1,2], Nellie Nelson[1], and J. Gage Crump*[1]

[1] Eli and Edythe Broad California Institute for Regenerative Medicine Center for Regenerative Medicine and Stem Cell Research, Department of Stem Cell Biology and Regenerative Medicine, University of Southern California Keck School of Medicine, Los Angeles, CA 90033, USA

[2] Department of Rehabilitation and Regenerative Medicine, Columbia University Irving Medical Center, Columbia University, New York, NY 10032, USA

[#]These authors contributed equally

*Correspondence: J. Gage Crump, gcrump@usc.edu, (323) 442-2693

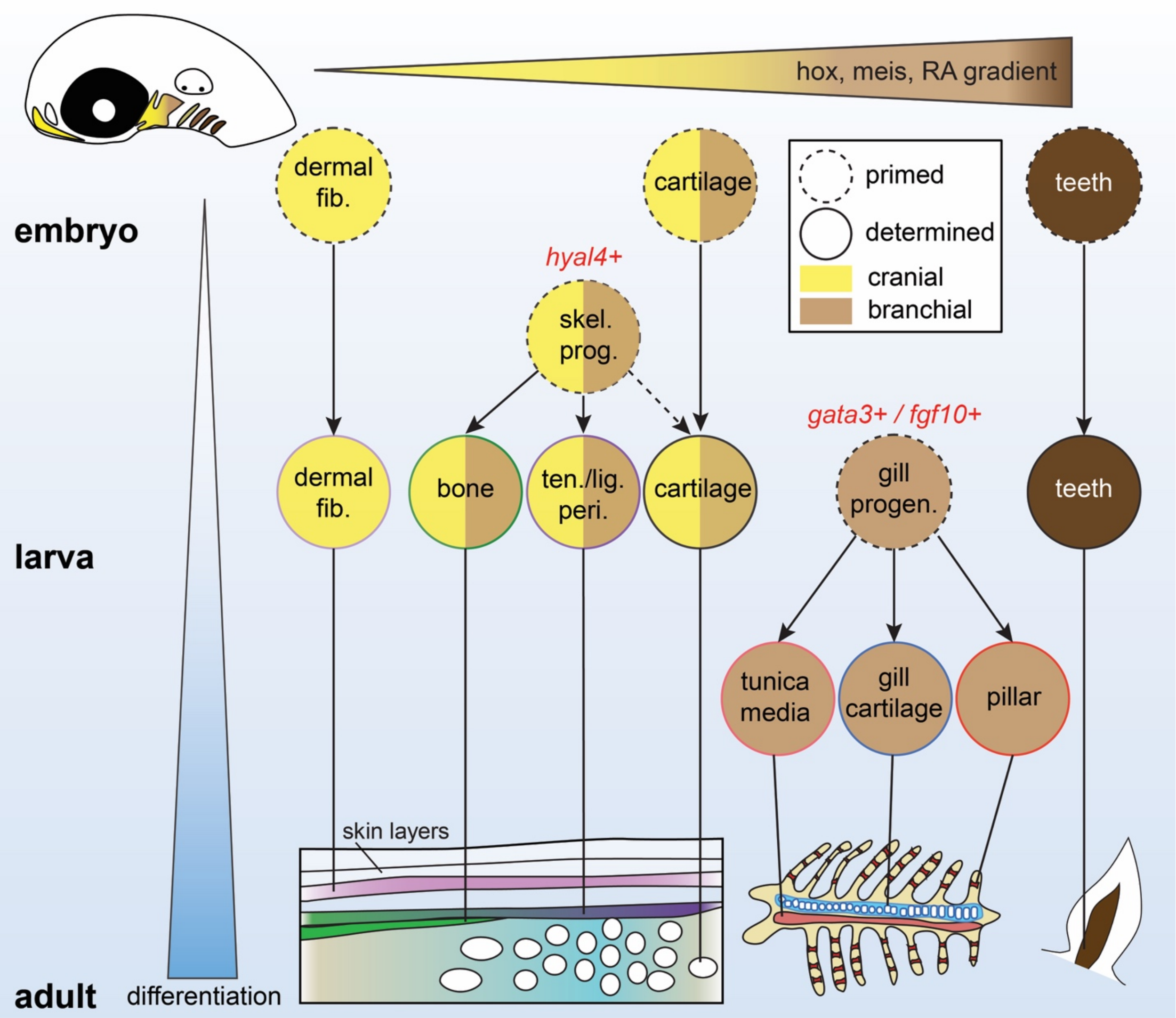
Running Title: Single-cell profiling of neural crest

Keywords: Cranial neural crest; single-cell genomics; craniofacial skeleton; chromatin accessibility; zebrafish

Highlights

- Single-cell transcriptome and chromatin atlas of cranial neural crest
- Progressive emergence of region-specific cell fate competency
- Chromatin accessibility mapping identifies candidate lineage regulators
- Gata3 function linked to gill-specific respiratory program

Graphical Abstract



Abstract

The cranial neural crest generates a huge diversity of derivatives, including the bulk of connective and skeletal tissues of the vertebrate head. How neural crest cells acquire such extraordinary lineage potential remains unresolved. By integrating single-cell transcriptome and chromatin accessibility profiles of cranial neural crest-derived cells across the zebrafish lifetime, we observe region-specific establishment of enhancer accessibility for distinct fates. Neural crest-derived cells rapidly diversify into specialized progenitors, including multipotent skeletal progenitors, stromal cells with a regenerative signature, fibroblasts with a unique metabolic signature linked to skeletal integrity, and gill-specific progenitors generating cell types for respiration. By retrogradely mapping the emergence of lineage-specific chromatin accessibility, we identify a wealth of candidate lineage-priming factors, including a Gata3 regulatory circuit for respiratory cell fates. Rather than multilineage potential being an intrinsic property of cranial neural crest, our findings support progressive and region-specific chromatin remodeling underlying acquisition of diverse neural crest lineage potential.

Main text

Cranial neural crest-derived cells (CNCCs) are a vertebrate-specific population, often referred to as the fourth germ layer, that have extraordinary potential to form diverse cell types. In addition to pigment cells and the peripheral nervous system, CNCCs form the ectomesenchyme that populates the pharyngeal arches and gives rise to much of the skeleton and connective tissue of the jaws and face[1]. Posterior arch CNCCs contribute to a distinct set of organs, including the thymus, parathyroid, and cardiac outflow tract, and in fishes cell types important for respiration, including specialized endothelial-like pillar cells that promote gas exchange[2]. In zebrafish, teeth develop from CNCCs of the most posterior seventh arch.

The extent to which diverse lineage potential is an intrinsic property of CNCCs, or acquired through later inductive signaling, has been investigated for over a century through labeling, grafting, and extirpation experiments, yet remains unresolved[3]. Individual avian CNCCs can generate multiple types of derivatives in vitro, including ectomesenchyme and neuroglial cells, suggesting multilineage potential is an intrinsic property[4]. However, upon cranial transplantation, trunk neural crest cells, which normally do not make mesenchymal derivatives, can contribute to the facial skeleton following extended culture[5] or misexpression of key transcription factors[6]. A recent study in skate also shows mesodermal contribution to the gill skeleton, a classically considered CNCC-derived structure[7]. These findings point to extrinsic inductive cues for CNCC fate determination. Here we take a genomics approach in zebrafish to understand when enhancers linked to diverse CNCC fates first gain accessibility, thus revealing that chromatin accessibility underlying multilineage potential is largely gained after CNCC migration.

Single-cell atlas of CNCC derivatives across the zebrafish lifetime

To understand the emergence and diversification of CNCC lineages across the lifetime of a vertebrate, we constructed a longitudinal single-cell atlas of gene expression and chromatin accessibility of zebrafish CNCC derivatives. We permanently labeled CNCCs using *Sox10:Cre; actab2:loxP-BFP-STOP-loxP-dsRed* (*Sox10>dsRed*) fish (Fig. 1a), in which genetic recombination indelibly labels CNCCs shortly after their specification at 10 hours post-fertilization (hpf)[8]. Previous single-cell analyses of CNCCs in zebrafish[9], chick[10], and mouse[11-13], and in vitro human CNCC-like cells[14], had focused on CNCC establishment, migration, and early fate choices between the neuroglial, pigment, and ectomesenchyme lineages. Here we investigate cellular diversity and lineage progression of CNCC ectomesenchyme across embryonic (1.5 and 2 days post-fertilization (dpf)), larval (3 and 5 dpf), juvenile (14 and 60 dpf), and adult (150-210 dpf) stages. After fluorescence activated cell sorting (FACS) of *Sox10>dsRed*+ cells from dissected heads, we performed single-cell RNA sequencing (scRNAseq) and single-nuclei assay for transposase accessible chromatin sequencing (snATACseq) at each stage using the 10X Genomics Chromium platform and paired-end Illumina next-generating sequencing (Fig. 1b). After filtering for quality, we obtained 58,075 cells with a median of 866 genes per cell for scRNAseq, and 88,177 cells with a median of 10,449 fragments per cell for snATACseq. To better resolve snATACseq data, we used the SnapATAC package[15], which integrates snATACseq with scRNAseq data to create "pseudo-multiome" datasets.

Analysis of CNCC cell clusters across all stages using UMAP dimensionality reduction recovered most known CNCC derivatives, including Schwann cells (glia), several neuronal subtypes, pigment cells, and diverse mesenchymal cell types (Fig. S1-8, Table S1). We also recovered otic placode and epithelial cells, likely reflecting additional non-CNCC expression of *Sox10:Cre*[8], and blood lineage cells, likely due to autofluorescence. Similar clusters were recovered using both scRNAseq and SnapATAC data. We then re-clustered the CNCC ectomesenchyme sub-population across stages, as this makes the most substantial and diverse cell contributions in the head. To confirm ectomesenchyme identity at 1.5 dpf, we also performed scRNAseq analysis of cells double-positive for the CNCC transgene *sox10:dsRed* and the ectomesenchyme transgene *fli1a*:e*GFP*. Co-clustering showed high concordance between *sox10:dsRed*+; *fli1a*:e*GFP*+ ectomesenchyme and the *Sox10>dsRed*+ ectomesenchyme subset, and between *Sox10>dsRed*+ ectomesenchyme scRNAseq subsets at all 7 stages (Fig. S8).

At the adult stage, we recovered 17 distinct clusters using scRNAseq that corresponded to 16 clusters using SnapATAC; these were largely associated with the jaw skeleton or gills (Fig. 1c-e). Skeletal derivatives include bone, cartilage, teeth, and a population with properties of periosteum, tendon, and ligament. Gills are composed of primary filaments containing cartilage rods and primary veins surrounded by a tunica media, and numerous secondary filaments housing endothelial-like "pillar" cells that promote gas exchange. Unexpectedly, we recovered a specialized type of gill cartilage distinct from that in the rest of the head, as well as pillar and tunica media cells and putative gill progenitors. We also recovered smooth muscle, perivascular, and stromal cells (see Table S1 for cluster marker genes and Fig. S9-10 for in situ validation).

In addition to skeletal and gill populations, we recovered a distinct type of fibroblast enriched for the cell adhesion molecule *chl1a* and *wnt5a*. Strikingly, these fibroblasts are also enriched for genes encoding enzymes for all steps of phenylalanine and tyrosine breakdown (Fig. 1f, Fig. S11). In situ hybridization for two of these genes (*hpdb* and *pah*) reveals that these fibroblasts are in the dermis between the skin epidermis and *runx2b*+/*sp7*+ osteoblast lineage cells (Fig. 1g,h). Humans with mutations in *HGD*, which encodes an intermediate enzyme in the Phe/Tyr catabolic pathway, develop Alkaptonuria, or black bone disease, due to accumulation and pathological aggregation of homogentisic acid[16]. As the abundant melanocytes in the zebrafish skin use high levels of Tyr to synthesize melanin, one possibility is that these specialized dermal fibroblasts function to protect the skeleton by removing damaging Phe/Tyr metabolites.

Progressive emergence of CNCC derivatives and region-specific progenitors

To understand lineage decisions of CNCC mesenchyme across time, we first used the STITCH algorithm[17] to connect individual stages into developmental trajectories for scRNAseq and snATACseq datasets (Fig. 2a,b). As early as 3 dpf (particularly apparent for snATACseq), we observe divergence of CNCCs into skeletogenic versus gill lineages. A *hyal4*+ perichondrium population precedes branches for tendon/ligament, periosteum, and osteoblasts (Fig. S9), and an *fgf10b*+ gill progenitor population appears at 5 dpf and precedes branches for gill cartilage, pillar, and tunica media cells (Fig. S10). We also observe a distinct trajectory to dermal fibroblasts by 3 dpf (Fig. S11), as well as to *cxcl12a*+ stromal cells (Fig. S9) and teeth. We do not observe CNCC contributions to cardiomyocytes (Fig. S8), in contrast to reports for amniotes[18]. By creating an index for ectomesenchyme-enriched gene expression at 1.5 dpf, a stage preceding the onset of differentiation, we found no evidence for retention of ectomesenchyme identity at later stages, as shown by aggregated ectomesenchyme gene expression and the early ectomesenchyme marker *nr2f5*[19] (Fig. S12). Although formation of CNCC ectomesenchyme involves a reacquisition of the pluripotency network[14], we also did not observe expression of pluripotency genes *pou5f3* (*oct4*), *sox2*, *nanog*, and *klf4* at any stage of post-migratory ectomesenchyme, with the exception of *lin28aa* that displays broad expression at 1.5 dpf and is rapidly extinguished by 2 dpf (Fig. S12). Rather than maintenance of a multipotent ectomesenchyme population, our data point to progressive emergence of specialized *hyal4*+ perichondrium, *cxcl12a*+ stromal, and *fgf10a/b*+ gill populations at 3 dpf and beyond (Fig. S12).

To further dissect region-specific lineages, we used Monocle3[20] on scRNAseq datasets to construct pseudotime trajectories of anterior arch (i.e. skeletogenic) versus posterior arch (i.e. gill, *hoxb3a+/gata3*+) CNCC mesenchyme at 5-14 dpf (Fig. S13, dermal fibroblasts and teeth were removed). For skeletogenic clusters, cell distribution from 5 to 14 dpf suggested two distinct lineages: one involving chemokine-expressing stromal cells (*cxcl12a*+/*ccl25b*+) and a second emanating from *hyal4*+ cells (Fig. 2c-e, Fig. S13). In situ hybridization at 14 dpf revealed broad mesenchymal expression of *cxcl12a*, and expression of *hyal4* in perichondrium in a largely complimentary pattern to *postnb* and *col10a1a* expression in periosteum (Fig. S9). Branches from *hyal4*+ perichondrium led to periosteum, tendon and ligament cells, chondrocytes, and osteoblasts, consistent with studies showing perichondrium to be the precursor of the periosteum in endochondral bones[21,22].

For gill clusters, cell distribution from 5 to 14 dpf revealed two primary trajectories (Fig. 2f-h, Fig. S13). In the first, *cxcl12a*+/*ccl25b*+ stromal cells give rise to mesenchyme associated with retinoic acid metabolism (*aldh1a2*+/*rdh10a*+), with in situ hybridization revealing these cell types restricted to the base of secondary filaments (Fig. S10). In the second, *fgf10a*+ cells are connected to *fgf10b*+ cells, which then diverge into gill cartilage, pillar, tunica media, and perivascular populations. To test whether *fgf10b*+ cells are progenitors for specialized gill subtypes, we used CRISPR/Cas9 to insert a photoconvertible nuclear EOS protein into the endogenous *fgf10b* locus. We found *fgf10b*:nEOS to be robustly expressed in the forming gills, with expression becoming progressively restricted to the tips of gill filaments over time, similar to endogenous *fgf10b* expression (Fig. S10, S14). We then used UV light to convert *fgf10b*:nEOS fluorescence from green to red in a small number of filaments at 7 dpf and observed contribution to gill chondrocytes and pillar cells 3 days later, with new *fgf10b*:nEOS cells (i.e. green only) being generated at the tips of growing filaments (Fig. 2i). Similar results were seen in adult gill filaments (Fig. S14). These data support *fgf10b*+ cells being progenitors for gill-specific cell types from larval through adult stages.

To understand how CNCC mesenchyme changes from embryogenesis to adulthood, we next interrogated patterns of gene usage and chromatin accessibility (Fig. 2j, Fig. S15-16, Table S2). Gene ontogeny (GO) analysis of ectomesenchyme at 1.5 and 2 dpf revealed terms linked to cell division and metabolism, consistent with early expansion of this population. We also find enrichment of transcription factors for early ectomesenchyme (*dlx2a*, *twist1a*, *nr2f6b*) and arch patterning (*pou3f3b*, *hand2*), as well as transcription factor binding motifs for several types of nuclear receptors, in accordance with known roles of Nr2f members in ectomesenchyme development[19]. The *hyal4*+ population contains skeletal-associated terms (collagen fibril organization, skeletal system development, regulation of ossification, cartilage development), consistent with being a common progenitor for cartilage, tendon, ligament, and bone in pseudotime analysis. The *hyal4*+ population is enriched for transcription factors implicated in perichondrium biology (*mafa*, *foxp2*, *foxp4*)[23,24] and cartilage formation (*barx1*, *sox6*, *emx2*)[25-27], and motifs for Bmp signaling (SMAD) and transcription factors (NFAT, RUNX) known to regulate cartilage and bone[28]. For gill *fgf10a/b*+ progenitors, we recover terms for general growth (e.g.

translation, cellular biosynthetic process), response to Fgf signaling, and respiratory system development, consistent with lineage tracing showing *fgf10b:nEOS*-labeled cells giving rise to gill respiratory cell types through adult stages. We also observe enrichment of *gata2a*, *gata3*, and GATA motif accessibility, suggesting important roles of Gata factors in gill-specific lineages.

In contrast to *hyal4*+ and *fgf10a/b*+ populations that display hallmarks of progenitors, the *cxcl12a*+ stromal population is associated with terms for regeneration, response to injury and wounding, negative regulation of the Wnt signaling pathway, and, particularly at adult stages, response to stress and modulation of the immune response. This population is enriched for *osr1*, early response genes of the Fos/Jun family, C/EBP family members implicated in response to inflammation[29], and *egr1* that has recently been linked to injury-induced regenerative responses across the animal kingdom[30]. Recovery of motifs for STAT and C/EBP also point to immune system interactions. As *Cxcl12*+ stromal cells in murine bone marrow have been shown to only contribute to osteoblasts during bone regeneration[31], it will be interesting to test whether the *cxcl12a*+ stromal population in animals such as zebrafish that lack bone marrow also plays a role in skeletal regeneration[32].

Highly resolved embryonic spatial expression domains from integrated datasets

We next sought to understand the developmental origins of distinct cell types and lineage programs in CNCC ectomesenchyme. To do so, we first examined the ability of integrated transcriptomic and chromatin accessibility datasets to predict the expression patterns of potential ectomesenchyme patterning genes at 1.5 dpf, a stage before overt cell type differentiation. Compared to scRNAseq (Fig. 3a) or snATACseq alone (Fig. S17), SnapATAC pseudo-multiome analysis (Fig. 3b) was better able to separate CNCCs along the major positional axes, including the dorsal-ventral axis and the anterior-posterior axis (frontonasal, mandibular (arch 1), hyoid (arch 2), branchial (arch 3-6), and tooth-bearing (arch 7)).

Comparison of the predicted SnapATAC expression of known region-specific genes - *pou3f3b* (dorsal arches 1 and 2), *dlx5a* (intermediate arches), *hand2* (ventral arches), *meis2b* (arch 7), and *pitx1* (oral mandibular)[25,33,34] - revealed tight correlation to reported expression, including zebrafish-specific overlap of *dlx5a* and *hand2* in the mandibular arch (Fig. 3d). We also identified a previously unappreciated oral-aboral axis of the mandibular arch in zebrafish, marked by *pitx1* and *nr5a2* respectively, which we validated by in situ hybridization for *nr5a2* (Fig. 3e). Re-examination of genes identified from previous bulk RNA sequencing of zebrafish arches further revealed strong correlation of SnapATAC domains with reported expression for 23 of 27 genes (Fig. S18), with SnapATAC suggesting frontonasal and tooth-domain expression for two genes previously annotated as false positives[25]. We also observed correlation of the transcription factor binding motifs enriched in cluster-specific accessible chromatin with the activities of transcription factors of the same family, including POU3F3, MEIS2, HAND, DLX5, PITX1, and NR5A2 (Fig. 3c,d). This approach shows the power of integrated scRNAseq and snATACseq data to predict the spatial expression domains of the vast majority of CNCC ectomesenchyme genes at pharyngeal arch stages.

Chromatin accessibility predicts cell type competency in early arches

We next sought to understand how the establishment of cell fate competency is linked to the earlier activity of arch patterning genes. To do so, we first computed unique patterns of chromatin accessibility (“peaks”) for each cell cluster at 14 dpf (Fig. 4a, Table S3). Modules of the top enriched peaks for each cell type were then mapped onto UMAP projections of SnapATAC data at 1.5, 2, 3, and 5 dpf (Fig. S19). To understand when cluster-specific peaks become established, as well as cluster relatedness, we developed the bioinformatics pipeline “Constellations". First, we calculated whether projections of cluster-specific peak modules are skewed toward particular regions of UMAP space at each earlier time-point, suggesting establishment of cluster-specific chromatin accessibility (a proxy for cell type competency). We then computed the relatedness of peak module projections in two dimensions for each mapped cluster at each stage (Fig. 4b). Analysis of cell competency trajectories shows that cell types can be grouped into five main classes: skeletogenic cells (including *hyal4*+ perichondral and *postnb*+ periosteal cells), stromal cells, dermal fibroblasts, gill cell types, and cartilage. Constellations analysis also reveals a temporal order of cell type competency establishment, with unique chromatin accessibility for cartilage and dermal fibroblast lineages emerging at 1.5 dpf; bone and perichondrium at 2 dpf; and periosteum, tendon and ligament, and gill progenitors and pillar cells at 3 dpf (Fig. 4c). This analysis suggests that chromatin accessibility prefiguring diverse CNCC cell types is progressively established rather than being inherited from earlier multipotent CNCCs.

Constellations analysis reveals candidate transcription factors for lineage priming

To discover potential transcription factors for establishing cell type competency, we analyzed the Constellations dataset for transcription factors whose expression and predicted binding motifs were co-enriched in particular clusters. We identified 287 transcription factor expression/motif pairs showing enrichment (Fig. S20, Table S4). The FOXC1 motif and *foxc1b* gene body activity were highly enriched in the cartilage trajectory, and LEF1/*lef1* in the dermal fibroblast trajectory (Fig. 5a). Projection of FOX motifs and merged Fox gene activity (*foxc1a*, *foxc1b*, *foxf1*, *foxf2a*, *foxf2b*) and LEF1/*lef1* onto SnapATAC UMAPs at 1.5 dpf reveals close correlation to mapping of the 14 dpf peak modules for cartilage and dermal fibroblasts at this stage (Fig. 5b,c), as well as the known fate map of cartilage precursors in the arches[35] (Fig. 5d,e). This confirms genetic evidence for roles of Foxc1 and Foxf1/2 in cartilage formation in zebrafish and mouse[36,37], and more specifically Foxc1 in establishing accessibility of cartilage enhancers in the developing face[28]. It also raises the possibility that Wnt signaling, mediated in part by Lef1, may play a role in early dermal fibroblast specification, consistent with enrichment of *wnt5a* in this population (Fig. S11).

We also find GATA3/*gata3* to be highly enriched in gill populations, with SnapATAC UMAP projections of GATA3 motif and *gata3* gene body activity at 5 dpf correlating with 14 dpf gill progenitor peaks (Fig. 5f). The enrichment of ETS2/*ets2*, which plays a role in endothelial development[38], in the gill pillar trajectory is consistent with ETS factors driving a mesenchyme-to-endothelia transition during formation of these vascular cells. Skeletogenic trajectories are uniquely marked by IRF8/*irf8*. Whereas loss of bone in mouse *Irf8*$^{-/-}$ mutants has been attributed to increased osteoclastogenesis[39], our analysis suggests that Irf8 may also have an early function in priming the skeletal lineage. Enrichment of CEBPA/*cebpa* in stromal trajectories may reflect the immunomodulatory role of this mesenchymal population[29]. These findings show the power of Constellations analysis to reveal potential factors for establishing regional chromatin accessibility important for later cell type differentiation.

Gill-specific lineages distinguished by early Gata3 activity

Given the selective enrichment of GATA3 motifs and *gata3* activity in gill lineages, we further investigated the presence of a Gata3 regulatory circuit directing CNCCs to gill fates. Whereas previous work had shown that *gata3* is expressed in and required for initial gill bud formation in zebrafish, larval lethality had precluded analysis of gill subtype differentiation[40]. We find *gata3* expression to be maintained in gill populations through adult stages in scRNAseq data, which we validated by in situ hybridization at 14 dpf and 2 years of age (Fig. S21). We then identified a non-coding region ~143kb downstream of the *gata3* gene, itself containing a predicted GATA3 binding site, that was selectively accessible in posterior arch CNCCs by 3 dpf, gill progenitors and pillar cells by 5 dpf, and gill cartilage cells by 14 dpf (Fig. 6a, Fig. S22). This *gata3-P1* element was sufficient to drive highly restricted GFP expression in posterior arch CNCCs starting at 1.5 dpf, which continued in gill progenitors, pillar cells, and chondrocytes through 60 dpf (Fig. 6c-e, Fig. S21).

Gill cartilage has a markedly distinct expression and chromatin accessibility profile from hyaline cartilage of the jaw, as shown by selective expression of *ucmaa* in gill cartilage versus *ucmab* in hyaline cartilage (Fig. S23). We identified a non-coding region ~5kb upstream of the *ucmaa* gene that was selectively accessible in gill cartilage starting at 14 dpf and contained a predicted GATA3-binding site (Fig. 6b, Fig. S22). This *ucmaa-P1* element drives highly restricted GFP expression in gill chondrocytes at 11 and 23 dpf, in contrast to a previously described *ucmab* enhancer[28] driving GFP expression in hyaline but not gill cartilage (Fig. 6f, Fig. S23). Although functional assays are needed to confirm Gata3 dependence, our findings are consistent with GATA factors establishing a positive autoregulatory circuit in posterior arch CNCCs that maintains *gata3* expression and promotes the later differentiation of gill-specific cell types (Fig. 6g).

Conclusion

Integration of transcriptome and chromatin accessibility data of the CNCC lineage has allowed us to connect patterning along major development axes to the emergence of the wide diversity of CNCC-derived cell types. Rather than lineage-specific chromatin accessibility being an intrinsic property of CNCCs, our Constellations analysis points to the progressive remodeling of chromatin underlying diverse cell type differentiation. Roles for inductive signaling in establishing enhancer accessibility in post-migratory arch CNCCs would help explain reports of mesodermal cells contributing to classically considered CNCC structures such as the skate gill skeleton[7]. Further, retrograde mapping of cell type-specific chromatin accessibility, combined with our highly resolved atlas of pharyngeal arch gene expression, reveals candidate transcription factors priming distinct CNCC lineages. Consistent with recent reports of organ-specific fibroblast heterogeneity[41], we also uncover a CNCC-derived dermal fibroblast population characterized by Phe/Tyr metabolism genes, which may be induced by early Wnt/Lef1 activity. Expression of some of the same Phe/Tyr catabolic genes is observed in a subset of axolotl limb fibroblasts[42], with the blackening of bone and cartilage in Alkaptonuria patients defective in Phe/Tyr breakdown suggesting general roles for these specialized dermal fibroblasts in protecting the skeleton. In the gill region, we identify a *fgf10*-expressing progenitor population characterized by sustained Gata3 activity, with later emergence of Ets2 activity in pillar cells providing a potential mechanism for the mesenchyme-to-endothelia transition of these specialized vascular cells. The presence of a similar Fgf10-expressing mesenchyme population in the mammalian lung[43] raises the possibility that an ancestral CNCC-derived gill respiratory program may have been co-opted by the mesoderm during later lung evolution. Single-cell profiling of transcriptome and chromatin accessibility across time thus provides a blueprint for understanding the diversification and post-embryonic production of the huge variety of CNCC-derived cell types throughout the head.

Materials and methods

Zebrafish lines

The Institutional Animal Care and Use Committee of the University of Southern California approved all animal experiments (Protocol 20771). Published lines include *Tg(Mmu.Sox10-Mmu.Fos:Cre)*^{zf384} [8]; *Tg(actab2:loxP-BFP-STOP-loxP-dsRed)*^{sd27} [44]; and *Tg(ucmab_p1:GFP)*^{el806}, *Tg(fli1a:eGFP)*^{y1}, and *Tg(sox10:DsRedExpress)*^{el10} [28]. Five transgenic lines were generated as part of this study: *Tg(fgf10b:nEOS)*^{el865}, *Tg(gata3_p1:GFP)*^{el857}, *Tg(gata3_p1:GFP)*^{e,858}, *Tg(ucmaa_p1:GFP)*^{el851} and *Tg(ucmaa_p1:GFP)*^{el854}. The *fgf10b:nEOS* knock-in line was made using CRISPR/Cas9-based integration [45]. Three gRNAs targeting sequences upstream of the *fgf10b* translational start site (5'-CATGATAACCCTTCCTAGAT-3', 5'-GAGCTCTTTGATAGCGGGCT-3', and 5'-GTTGAGCAGCATGTCCCATG-'3) were co-injected at 100 ng/uL into wild-type embryos with Cas9 RNA (100 ng/uL), an mbait-NLS-EOS plasmid (20 ng/uL) [46], and the published gRNA targeting the mbait sequence [45] to linearize the plasmid. A germline founder was identified based on nEOS fluorescence in the progeny of injected animals. For enhancer transgenic lines, we synthesized peaks for *gata3* (chr4:24918100-24918770) and *ucmaa* (chr4:7836670-783720) using iDT gBlocks and cloned these into a modified pDest2AB2 construct containing E1b minimal promoter, GFP, and an eye-CFP selectable marker [28] using In-Fusion cloning (Takara Bio). We injected plasmids and Tol2 transposase RNA (30 ng/uL each) into one-cell stage zebrafish embryos, raised these animals, and screened for founders based on eye CFP expression in the progeny. Two independent germline founders were identified for each that showed similarly specific activity in the gills.

In situ hybridization and immunohistochemistry

All samples were prepared by fixation in 4% paraformaldehyde and embedded in paraffin, with decalcification for one week in 20% EDTA if over 14 dpf. All in situ patterns were confirmed in at least 3 independent animals. RNAscope probes were synthesized by Advanced Cell Diagnostics in channels 1 through 4. Channel 1 probes: *ifitm5, ucmaa, col10a1a*. Channel 2 probes: *postnb, myh11a, cxcl12a, sp7, gata3*. Channel 3 probes: *pah, lum, fgf10b, sox9a.* Channel 4 probe: *hyal4, acta2, ncam3*. Paraformaldehyde-fixed paraffin-embedded sections were deparaffinized, and the RNAscope Fluorescent Multiplex V2 Assay was performed according to manufacturer's protocols with the ACD HybEZ Hybridization oven. Colorimetric in situ hybridization was performed as described [32]. The *hpdb* riboprobe was generated by cloning a purchased gBlock fragment designed from transcript hpdb-201 using nucleotides 679-1395 (tggatga...gactccc) into pCR-BluntII-TOPO (Life Technologies). The *nr5a2* riboprobe was generated by PCR amplification of cDNA with primers 5'-ATGGGGAACAGGGGCATATG-3' and 5'-AGGGGTCGGGATACTCTGAT-3', the *ucmaa* riboprobe with primers 5'-TGGTACCAGCTCAAGACACT-3' and 5'-ATAGTACTGGCGGTGGTGAG-3', the *ucmab* riboprobe with primers 5'-ATGTCCTGGACTCAACCTGC-3' and 5'-GTTATCTCCCAGCGTGTCCA-3', and the *thbs4a* riboprobe with primers 5'-CCCATGTTTCTTCGGTGTGA-3' and 5'-GGTTTGGTACCAGCCTACAG-3'. Amplified products were cloned into pCR-BluntII-TOPO.

pCR-BluntII-TOPO plasmids were linearized by restriction digest (enzyme dependent on direction of blunt insertion), and RNA probe was synthesized using either T7 or Sp6 polymerase (Roche) depending on direction of blunt insertion. Immunohistochemistry for dsRed was performed with a 7 minute -20°C 100% acetone target retrieval and blocking in 2% normal goat serum (Jackson ImmunoResearch, cat. no. 005-000-121). Primary antibodies include rabbit anti-mCherry (1:100, Rockland Immunochemicals, cat. no. RL600-401-P16) and rabbit anti-mCherry (1:100, Novus Biologicals, cat. no. NBP2-25157) used at the same time. The secondary antibody was goat anti-rabbit Alexa Fluor 546.

Imaging

Confocal images of whole-mount or section fluorescent in situ hybridizations and live images of transgenic fish were captured on a Zeiss LSM800 microscope using ZEN software. Colorimetric in situs were imaged on a Zeiss AxioScan Z.1 For *fgf10:nEOS* experiments, we used the ROI function on the confocal microscope to specifically convert nEOS-expressing cells in the gill filaments of live animals using targeted UV irradiation, prior to the emergence of gill filament cartilage. At the specified days post-conversion, we euthanized the animal, fixed it in 4% PFA for 1 hour, and dissected the gill arches. We stained the gills with DRAQ5 nuclear dye (Abcam) for 30 min and imaged at 40X to locate converted cells. For all transgenic imaging experiments, expression patterns were confirmed in at least 5 independent animals.

Single-cell analysis and statistics

scRNAseq and snATACseq library preparation and alignment

Dissected heads from converted *Sox10:Cre; actab2:loxP-BFP-STOP-loxP-dsRed* fish were incubated in fresh Ringer's solution 5-10 min, followed by mechanical and enzymatic dissociation by pipetting every 5 minutes in protease solution (0.25% trypsin (Life Technologies, 15090-046), 1 mM EDTA, and 400 mg/ml Collagenase D (Sigma, 11088882001) in PBS) and incubated at 28.5°C for 20-30 minutes or until full dissociation. Reaction was stopped by 6X stop solution (6 mM CaCl2and 30% fetal bovine serum (FBS) in PBS). Cells were pelleted (2000 rpm, 5 min, 4 °C) and resuspended in suspension media (1% FBS, 0.8 mM CaCl2, 50 U/ml penicillin, and 0.05 mg/ml streptomycin (Sigma-Aldrich, St. Louis, MO) in phenol red-free Leibovitz's L15 medium (Life Technologies)) twice. Final volumes of 500 µl resuspended cells were placed on ice and sorted by fluorescence activated cell sorting (FACS) to isolate live cells that excluded the cytoplasmic stain Zombie green (BioLegend, 423111). For scRNAseq library construction, barcoded single-cell cDNA libraries were synthesized using 10X Genomics Chromium Single Cell 3' Library and Gel Bead Kit v.2 per manufacturer's instructions. Libraries were sequenced on Illumina NextSeq or HiSeq machine at a depth of at least 1,000,000 reads per cell for each library. Read2 was extended to 126 cycles for higher coverage. Cellranger v3.0.0 (10X Genomics) was used for alignment against GRCz11 (built with GRCz11.fa and GRCz11.98.gtf) and gene-by-cell count matrix generation with default parameters.

For snATACseq library construction, we used the same cell dissociation and sorting protocol as for scRNAseq, with isolation of live cells that excluded the cytoplasmic stain Zombie green (BioLegend, 423111) and collected live cells in 0.04% BSA/PBS. Nuclei isolation was performed per manufacturer's instructions (10X Genomic, protocol CG000169). Cells were incubated with lysis buffer on ice for 90 s, followed by integrity check of nuclei under fluorescence microscope with DAPI before library synthesis. Barcoded single-nuclei ATAC libraries were synthesized using 10X Genomics Chromium Single Cell ATAC Reagent Kit v1.1 per manufacturer's instructions. Libraries were sequenced on Illumina NextSeq or HiSeq machine at a depth of at least 75,000 reads per nucleus for each library. Both read1 and read2 were extended to 65 cycles. Cellranger ATAC v1.2.0 (10X Genomics) was used for alignment against genome (built with GRCz11.fa, JASPAR2020, and GRCz11.98.gtf), peak calling, and peak-by-cell count matrix generation with default parameters.

We included biological replicates at several stages to test the reproducibility of library preparation and increase depth of data. For scRNAseq, we performed two replicates at 5 and 14 dpf, and three replicates at 3 and 150 dpf. For snATACseq, we performed two replicates at 2, 3, and 14 dpf.

SnapATAC for peak refinement and gene activity matrix imputation

To refine the peak profile for better representation of diverse cell types across libraries, we performed a second round of peak calling using package Snaptools (v1.2.7) and SnapATAC (v1.0.0) [15]. We first removed low-quality cell and cell doublets by setting cutoffs based on percentage of reads in peaks (> 30 for 60 dpf, > 45 for 210 dpf, and > 50 for the rest) and fragment number within peaks (5,000 – 30,000 for 5 dpf, 1,000 – 11,000 for 14 dpf, and 1,000 – 20,000 for the rest). Potential cell debris or low-quality cells were removed by setting hard fragments-in-peak number cutoffs. Using the SnapATAC package, we then generated "pseudo-multiome" data at each stage. To recover every aligned fragment, we binned the genome into 5 kb sections and constructed the bin-by-cell matrices (bmats) for each library by Snaptools from the positional-sorted bam files generated by Cellranger ATAC v1.2.0. The cells were filtered, dimensionally reduced by diffusion map, and clustered with inputs of the first 34 dimensions followed the SnapATAC vignette. The specific peaks were called for each cluster by the wrapped MACS2 function in SnapATAC with parameter gsize = 1.5e9, shift = 100, ext = 200, and qval = 5e-2. The finalized and refined peak profile was derived by collapsing and merging all 175 individual peak files to 445,307 peaks. To impute the gene activity with the corresponding scRNAseq data, the bmats of each time point were used to calculate gene-activity-by-cell matrices (gmats) by SnapATAC. The gmats were used to find anchors within the scRNAseq data at the same time point by Seurat. We then transferred the expression data from scRNAseq through the anchors to derive the imputed gene-activity-by-cell matrices for each time point.

Data processing of scRNAseq and snATACseq

The count matrices of both scRNAseq and snATACseq data were analyzed by R package Seurat (v3.2.3) and Signac (v1.0.0). The count matrices of each sample were aggregated where replicates were available. For scRNAseq data, the matrices are normalized (NormalizeData) and scaled for the top 2,000 variable genes (FindVariableFeatures and ScaleData). The scaled matrices were dimensionally reduced to 50 principal components (60 components for 150 dpf), and then subjected to neighbor finding (FindNeighbors, k = 20) and clustering (FindClusters, resolution = 0.8). The data were visualized through UMAP with 50 principal components as input. For snATACseq data, the matrices are dimensionally reduced to 30 latent semantic indices (LSIs) through RunTFIDF and RunSVD functions. The neighbor finding, clustering, and visualization are performed as for scRNAseq data (algorithm = 3 for FindClusters) with input of the second to thirtieth LSIs. To calculate the motif accessibility, the enrichment of motifs in JASPAR2020 [47] was calculated by chromVAR [48] through function RunChromVar. To test the enriched genes and gene activities in both scRNAseq and snATACseq data as shown in Table S1, two-sided likelihood-ratio test is performed through FindAllMarkers function (min.pct = 0.25) with cutoff of adjusted p value smaller than 0.001.

STITCH network construction and force directed layout

To identify the overall cell trajectories in our scRNAseq and snATACseq data, we used the STITCH algorithm [17] to construct cell neighbor networks. As the dimensional reduction space of snATACseq data are LSIs, we modified the *stitch_get_knn* and *stitch_get_link* function of the STITCH package to make it compatible to LSI. For *stitch_get_knn* function of snATACseq data, we used the LSI matrix to find the k nearest neighbor of each cell for each time point. For *stitch_get_link* function of snATACseq data, we projected the LSI space of time point *t* to time point *t-1* by solving the right orthogonal matrix of the singular vector decomposition (SVD) for *t-1*. SVD ($M = U\Sigma V^T$) is the initial step of latent semantic analysis where M is the peak-by-cell matrix and U will be later transformed to LSI. For t and $t-1$, we have $M_t = U_t\Sigma_t V^T_t$ and $M_{t-1} = U_{t-1}\Sigma_{t-1}V^T_{t-1}$. To project the space from t to $t-1$, we derived a projected U_{t-1} as U^p_{t-1} through solving the equation $M_{t-1} = U^p_{t-1}\Sigma_t V^T_t$. Both U_t and U^p_{t-1} were further combined, normalized, and subjected to the default neighbor finding as *stitch_get_link*. To visualize the STITCH networks of both scRNAseq and snATACseq data, we used the force directed layout by ForceAtlas2 in Gephi (v0.9.2) to derive the 2-dimensional layouts.

Pseudotime analysis

We used the R package monocle3 (v0.2.3.0) to predict the pseudotemporal relationships within skeletogenic or gill populations. We first merged 5 and 14 dpf scRNAseq data, including an additional scRNAseq library of *sox10>dsRed*+ cells sorted from the dissected ceratohyal endochondral bone at 14 dpf (to further enrich skeletogenic populations), and performed clustering and dimensionality reduction. After removing dermal fibroblast (*pah*+) and teeth (*spock3*+) populations, we placed *hoxb3a+/gata3*+ cells into a "gill" cluster and all other cells into a "jaws" cluster. Cell paths were predicted by the *learn_graph* function of monocle3. We set the origin of the cell paths based on the enriched distribution of 5 dpf cells.

Gene ontology, motif family, and TF analysis of CNCC mesenchyme

Analysis was performed on ectomesenchyme, perichondrium, gill progenitor, and stromal populations at each stage based on markers from the scRNAseq data (Fig. S1-7). The enriched genes of each cluster are tested by running a two-sided Wilcoxon rank sum test against all other clusters using an adjusted p value ≤ 0.001. These enriched gene sets are subjected to gene ontology analysis for terms of biological processes (BP) by R package ViSEAGO (v1.2.0) [49]. The heatmap is generated by GOterms_heatmap function using values of log10(adjusted p.value). To generate the heatmap of motif families for each cluster, we first averaged and aggregated the motif accessibilities for each cell according to the motif family by TRANSFAC [50]. The means of each motif family are used for the heatmap. To generate the heatmap of TFs for each cluster, we subsetted the TFs from the enriched gene sets, and used the mean of each TF for every cluster for the heatmap.

Constellations analysis and calculation of cluster skewedness and correlation

The tissue module scores of the snATACseq data were calculated based on the enriched peak sets and their module scores for each cluster identified at 14 dpf by R packages Seurat and Signac. The enriched peak sets were calculated by the FindAllMarkers function using two-sided likelihood ratio test with fragment numbers in peak region as latent variables. We used the peaks with adjusted p values smaller than 0.001 as the enriched peaks for a cluster. As there are 23 clusters (tissues) identified at 14 dpf, we ended up with 23 peak sets, which we applied to calculate the tissue module scores to earlier time points (1.5, 2, 3, and 5 dpf) using AddChromatinModule function. To determine whether a tissue score at a time point distributes in a statistically significant, and hence biologically interesting, way, we calculated the skewedness of the distribution of a tissue score by the R package parameter (v0.12.0). We considered a tissue score to be distributed in a meaningful way if it was strongly right skewed by a hard cutoff of skewedness greater than 1. For 1.5 dpf, the cutoff of skewedness was lowered to 0.4 to accommodate overall lower skewedness at that time point, but with additional filter of max module score > 15 to avoid tissue module scores with extremely low values.

To profile the relationship of all tissue scores, we constructed a distance matrix of all 23 tissue module scores across all the time points (1.5, 2, 3, 5, and 14 dpf). For the distance *D* between score of tissue *A* at time point t_1 and score of tissue *B* at time point t_2, the distance *D* can be described as *D* = *D*(tissue) + (*a* x *D*(time point)). *D*(tissue) stands for the distance between the score of tissue *A* and *B* by averaging the Euclidian distance between score *A* and *B* at time point t_1 and the Euclidian distance between score *A* and *B* at time point t_2. *D*(time point) stands for the distance between time point t_1 and t_2 derived by the distance between the dendrogram of all the tissue scores at t_1 and the dendrogram at t_2. Since *D*(time point) is relatively smaller than *D*(tissue), we multiply *D*(time point) by 8 to make the distance between time points comparable to the distance between tissue scores. The distance matrix was dimensionally reduced and visualized by UMAP.

To detect the potential factors that contribute to the patterning of tissue-specific peaks, we performed linear regression of each tissue module score against all motif accessibilities and the gene activities of transcription factors. We used ZFIN and JASPAR2020, converted by homology data from MGI, to build up a list of transcription factors in zebrafish. We then curated and paired every motif in JASPAR2020 with its potential binding transcription factors. The coefficients of regression results were used as indications of whether a motif or transcription factor is positively correlated with a tissue module score with upper cutoff of adjusted p value 0.05. We transformed the coefficients of all the negative related motifs and transcription factors to 0 to filter out irrelevant motifs and transcription factors. To visualize the correlation of each pair of motif and transcription factor, we plotted the coefficient magnitudes of motifs by dot sizes and transcription factor gene body activities by a red color scale on Constellation maps.

Acknowledgments

We thank Megan Matsutani and Jennifer DeKoeyer Crump for fish care, Jeffrey Boyd and the BCC FACS core, the CHLA Sequencing core, the USC HPC computing core, and Andrew McMahon, Yang Chai, and Unmesh Jadhav for helpful comments on the manuscript.

Competing interests

No competing interest declared.

Funding

Funding was provided by the National Institute of Health (NIDCR R35 DE027550 to J.G.C.; NIDCR K99 DE029858 to P.F; NIDCR F31 DE029682-02 to C.A.; NICHD T32 HD060549 to M.T.).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. Processed and raw sequencing data have been deposited at GEO as GSE178969.

Author Contributions

P.F., K.-C.T., M.T., C.A., H.-J.C., J.S., N.N., and J.G.C. performed the experiments. J.G.C. oversaw the project and wrote the manuscript.

References

1 Platt, J. B. Ectodermic origin of the cartilages of the head. *Anat. Anz.* **8**, 506-509 (1893).

2 Mongera, A. *et al.* Genetic lineage labeling in zebrafish uncovers novel neural crest contributions to the head, including gill pillar cells. *Development* **140**, 916-925, doi:10.1242/dev.091066 (2013).

3 Dupin, E., Calloni, G. W., Coelho-Aguiar, J. M. & Le Douarin, N. M. The issue of the multipotency of the neural crest cells. *Dev Biol* **444 Suppl 1**, S47-S59, doi:10.1016/j.ydbio.2018.03.024 (2018).

4 Baroffio, A., Dupin, E. & Le Douarin, N. M. Common precursors for neural and mesectodermal derivatives in the cephalic neural crest. *Development* **112**, 301-305 (1991).

5 McGonnell, I. M. & Graham, A. Trunk neural crest has skeletogenic potential. *Curr Biol* **12**, 767-771, doi:10.1016/s0960-9822(02)00818-7 (2002).

6 Simoes-Costa, M. & Bronner, M. E. Reprogramming of avian neural crest axial identity and cell fate. *Science* **352**, 1570-1573, doi:10.1126/science.aaf2729 (2016).

7 Sleight, V. A. & Gillis, J. A. Embryonic origin and serial homology of gill arches and paired fins in the skate, Leucoraja erinacea. *Elife* **9**, doi:10.7554/eLife.60635 (2020).

8 Kague, E. *et al.* Skeletogenic fate of zebrafish cranial and trunk neural crest. *PLoS One* **7**, e47394, doi:10.1371/journal.pone.0047394 (2012).

9 Mitchell, J. M. *et al.* The alx3 gene shapes the zebrafish neurocranium by regulating frontonasal neural crest cell differentiation timing. *Development*, doi:10.1242/dev.197483 (2021).

10 Williams, R. M. *et al.* Reconstruction of the Global Neural Crest Gene Regulatory Network In Vivo. *Dev Cell* **51**, 255-276 e257, doi:10.1016/j.devcel.2019.10.003 (2019).

11 Soldatov, R. *et al.* Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* **364**, doi:10.1126/science.aas9536 (2019).

12 Xu, J. *et al.* Hedgehog signaling patterns the oral-aboral axis of the mandibular arch. *Elife* **8**, doi:10.7554/eLife.40315 (2019).

13 Yuan, Y. *et al.* Spatiotemporal cellular movement and fate decisions during first pharyngeal arch morphogenesis. *Sci Adv* **6**, doi:10.1126/sciadv.abb0119 (2020).

14 Zalc, A. *et al.* Reactivation of the pluripotency program precedes formation of the cranial neural crest. *Science* **371**, doi:10.1126/science.abb4776 (2021).

15 Fang, R. *et al.* Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat Commun* **12**, 1337, doi:10.1038/s41467-021-21583-9 (2021).

16 Fernandez-Canon, J. M. *et al.* The molecular basis of alkaptonuria. *Nat Genet* **14**, 19-24, doi:10.1038/ng0996-19 (1996).

17 Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-987, doi:10.1126/science.aar4362 (2018).

18 Tang, W., Martik, M. L., Li, Y. & Bronner, M. E. Cardiac neural crest contributes to cardiomyocytes in amniotes and heart regeneration in zebrafish. *Elife* **8**, doi:10.7554/eLife.47929 (2019).

19 Barske, L. *et al.* Essential Role of Nr2f Nuclear Receptors in Patterning the Vertebrate Upper Jaw. *Dev Cell* **44**, 337-347 e335, doi:10.1016/j.devcel.2017.12.022 (2018).

20 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502, doi:10.1038/s41586-019-0969-x (2019).

21 Bandyopadhyay, A., Kubilus, J. K., Crochiere, M. L., Linsenmayer, T. F. & Tabin, C. J. Identification of unique molecular subdomains in the perichondrium and periosteum and their role in regulating gene expression in the underlying chondrocytes. *Dev Biol* **321**, 162-174, doi:10.1016/j.ydbio.2008.06.012 (2008).

22 Colnot, C., Lu, C., Hu, D. & Helms, J. A. Distinguishing the contributions of the perichondrium, cartilage, and vascular endothelium to skeletal development. *Dev Biol* **269**, 55-69, doi:10.1016/j.ydbio.2004.01.011 (2004).

23 Huang, W., Lu, N., Eberspaecher, H. & De Crombrugghe, B. A new long form of c-Maf cooperates with Sox9 to activate the type II collagen gene. *J Biol Chem* **277**, 50668-50675, doi:10.1074/jbc.M206544200 (2002).

24 Zhao, H. *et al.* Foxp1/2/4 regulate endochondral ossification as a suppresser complex. *Dev Biol* **398**, 242-254, doi:10.1016/j.ydbio.2014.12.007 (2015).

25 Askary, A. *et al.* Genome-wide analysis of facial skeletal regionalization in zebrafish. *Development* **144**, 2994-3005, doi:10.1242/dev.151712 (2017).

26 Lefebvre, V., Li, P. & de Crombrugghe, B. A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J* **17**, 5718-5733, doi:10.1093/emboj/17.19.5718 (1998).

27 Nichols, J. T., Pan, L., Moens, C. B. & Kimmel, C. B. barx1 represses joints and promotes cartilage in the craniofacial skeleton. *Development* **140**, 2765-2775, doi:10.1242/dev.090639 (2013).

28 Xu, P. *et al.* Foxc1 establishes enhancer accessibility for craniofacial cartilage differentiation. *Elife* **10**, doi:10.7554/eLife.63595 (2021).

29 Poli, V. The role of C/EBP isoforms in the control of inflammatory and native immunity functions. *J Biol Chem* **273**, 29279-29282, doi:10.1074/jbc.273.45.29279 (1998).

30 Gehrke, A. R. *et al.* Acoel genome reveals the regulatory landscape of whole-body regeneration. *Science* **363**, doi:10.1126/science.aau6173 (2019).

31 Matsushita, Y. *et al.* A Wnt-mediated transformation of the bone marrow stromal cell identity orchestrates skeletal regeneration. *Nat Commun* **11**, 332, doi:10.1038/s41467-019-14029-w (2020).

32 Paul, S. *et al.* Ihha induces hybrid cartilage-bone cells during zebrafish jawbone regeneration. *Development* **143**, 2066-2076, doi:10.1242/dev.131292 (2016).

33 Barske, L. *et al.* Evolution of vertebrate gill covers via shifts in an ancient Pou3f3 enhancer. *Proc Natl Acad Sci U S A* **117**, 24876-24884, doi:10.1073/pnas.2011531117 (2020).

34 Bessa, J. *et al.* meis1 regulates cyclin D1 and c-myc expression, and controls the proliferation of the multipotent cells in the early developing zebrafish eye. *Development* **135**, 799-803, doi:10.1242/dev.011932 (2008).

35 Crump, J. G., Swartz, M. E., Eberhart, J. K. & Kimmel, C. B. Moz-dependent Hox expression controls segment-specific fate maps of skeletal precursors in the face. *Development* **133**, 2661-2669, doi:10.1242/dev.02435 (2006).

36 Hong, H. K., Lass, J. H. & Chakravarti, A. Pleiotropic skeletal and ocular phenotypes of the mouse mutation congenital hydrocephalus (ch/Mf1) arise from a winged helix/forkhead transcriptionfactor gene. *Human molecular genetics* **8**, 625-637 (1999).

37 Xu, P. *et al.* Fox proteins are modular competency factors for facial cartilage and tooth specification. *Development* **145**, doi:10.1242/dev.165498 (2018).

38 Wei, G. *et al.* Ets1 and Ets2 are required for endothelial cell survival during embryonic angiogenesis. *Blood* **114**, 1123-1130, doi:10.1182/blood-2009-03-211391 (2009).

39 Zhao, B. *et al.* Interferon regulatory factor-8 regulates bone metabolism by suppressing osteoclastogenesis. *Nat Med* **15**, 1066-1071, doi:10.1038/nm.2007 (2009).

40 Sheehan-Rooney, K., Swartz, M. E., Zhao, F., Liu, D. & Eberhart, J. K. Ahsa1 and Hsp90 activity confers more severe craniofacial phenotypes in a zebrafish model of

hypoparathyroidism, sensorineural deafness and renal dysplasia (HDR). *Dis Model Mech* **6**, 1285-1291, doi:10.1242/dmm.011965 (2013).

41 Muhl, L. *et al.* Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat Commun* **11**, 3953, doi:10.1038/s41467-020-17740-1 (2020).

42 Gerber, T. *et al.* Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration. *Science* **362**, doi:10.1126/science.aaq0681 (2018).

43 Malpel, S., Mendelsohn, C. & Cardoso, W. V. Regulation of retinoic acid signaling during lung morphogenesis. *Development* **127**, 3057-3067 (2000).

44 Kobayashi, I. *et al.* Jam1a-Jam2a interactions regulate haematopoietic stem cell fate through Notch signalling. *Nature* **512**, 319-323, doi:10.1038/nature13623 (2014).

45 Kimura, Y., Hisano, Y., Kawahara, A. & Higashijima, S. Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. *Sci Rep* **4**, 6545, doi:10.1038/srep06545 (2014).

46 Thomas, E. D. & Raible, D. W. Distinct progenitor populations mediate regeneration in the zebrafish lateral line. *Elife* **8**, doi:10.7554/eLife.43736 (2019).

47 Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92, doi:10.1093/nar/gkz1001 (2020).

48 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978, doi:10.1038/nmeth.4401 (2017).

49 Brionne, A., Juanchich, A. & Hennequet-Antier, C. ViSEAGO: a Bioconductor package for clustering biological functions using Gene Ontology and semantic similarity. *BioData Min* **12**, 16, doi:10.1186/s13040-019-0204-1 (2019).

50 Wingender, E., Dietze, P., Karas, H. & Knuppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* **24**, 238-241, doi:10.1093/nar/24.1.238 (1996).

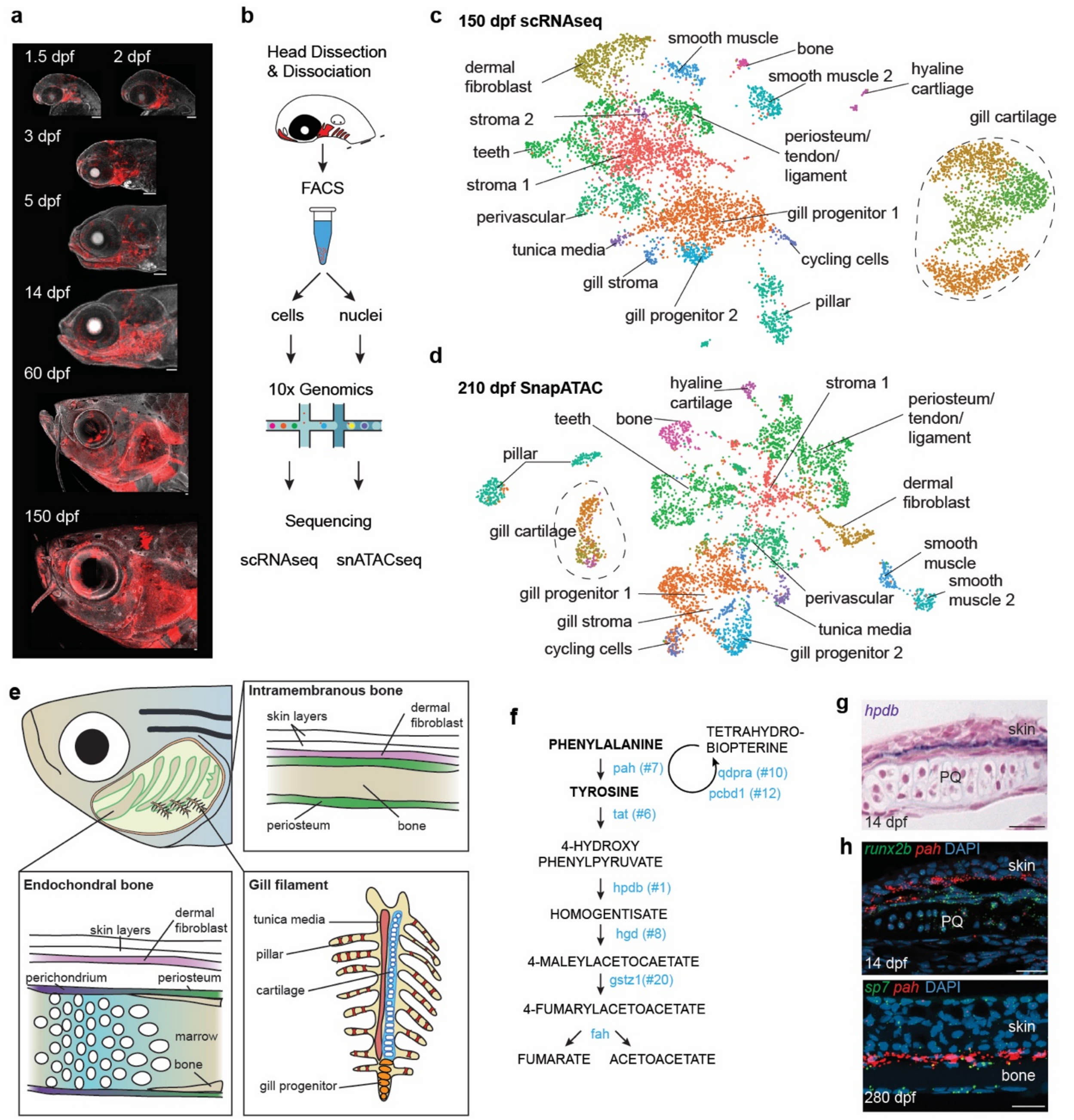


Figure 1. Single-cell transcriptomes and chromatin accessibility of CNCCs across the zebrafish lifetime. a, *Sox10:Cre; actab2:loxP-BFP-STOP-loxP-dsRed* labeling of CNCCs (red) in zebrafish heads across 7 stages. **b**, Scheme of cell or nuclei dissociation, fluorescence activated cell sorting (FACS), and processing on the 10X Genomics platform for sequencing. **c,d**, UMAPs of scRNAseq and SnapATAC datasets at adult stages. **e**, Diagram of an adult zebrafish head cut out to show major cell types of the interior skeletal elements and gill system. **f**, Pathway

for Phe and Tyr breakdown. 7 of 8 genes encoding catabolic enzymes are in the top 20 selectively enriched genes for the dermal fibroblast cluster (numbers show rank). **g**, Section colorimetric in situ hybridization for *hpdb* RNA shows expression in the dermis between the skin and palatoquadrate (PQ) endochondral bone. **h**, Section RNAscope in situ hybridizations show *pah* expression in dermal fibroblasts between the *runx2b*+ periosteum of PQ and skin at 14 dpf, and between *sp7*+ osteoblasts of intramembranous bone and skin at 280 dpf. DAPI labels nuclei in blue. Scale bars, 100 μm (a), 20 μm (g,h).

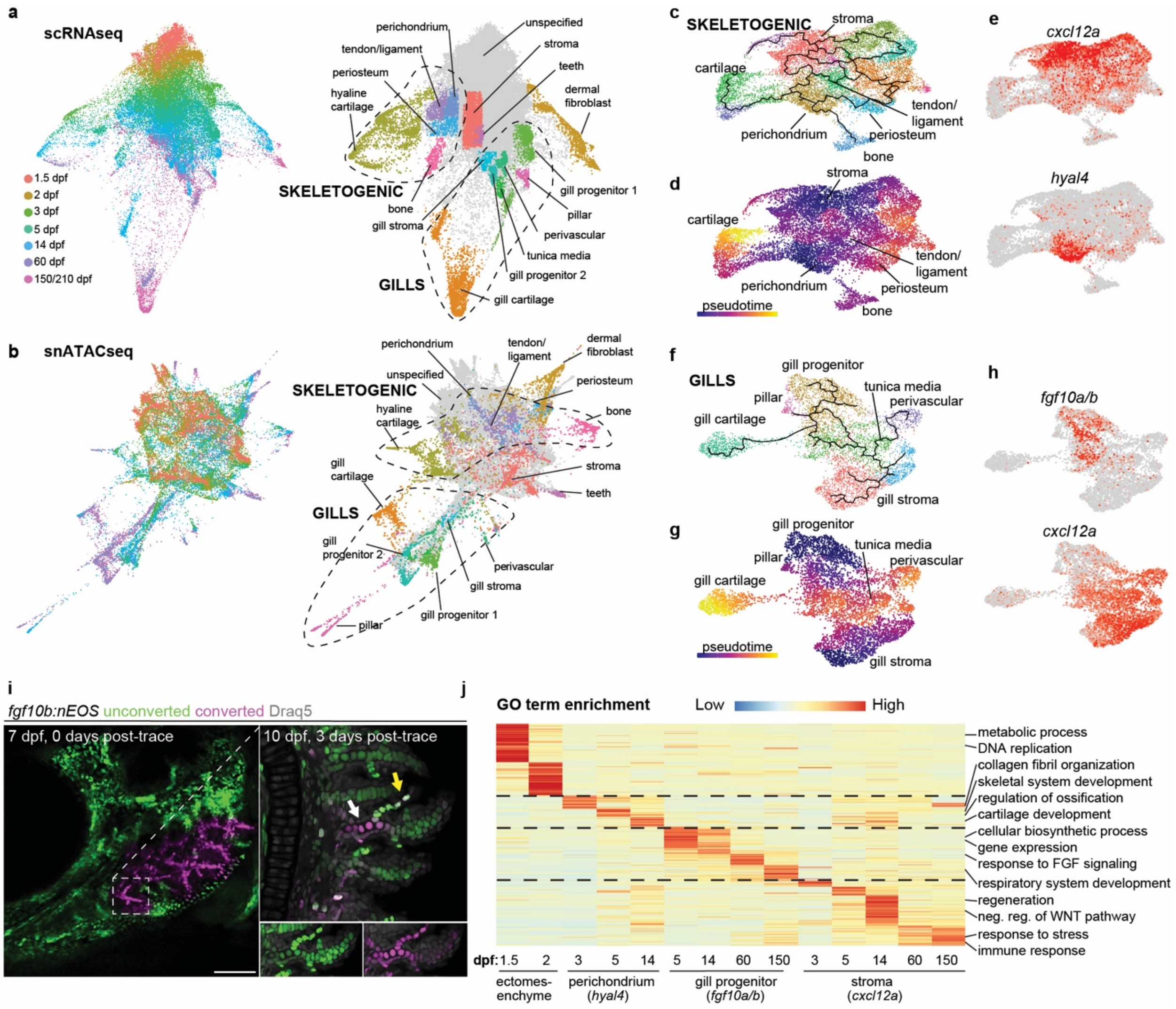


Figure 2. Progressive emergence of region-specific lineage programs. a,**b**, STITCH plots

connect individual scRNAseq and snATACseq datasets across the zebrafish lifetime. Cell type annotations show divergence of CNCC ectomesenchyme into skeletogenic, gill, dermal fibroblast, and stromal branches. **c-e**, Pseudotime analysis using Monocle3 of the skeletogenic subset (combined 5 and 14 dpf) shows a *cxcl12a*+ stromal branch, and a *hyal4*+ branch connected to cartilage, bone, periosteum, and tendon and ligament. **f-h**, Pseudotime analysis of gill subsets (combined 5 and 14 dpf) shows a *cxcl12a*+ stromal branch, and a *fgf10a/b*+ branch connected to gill pillar, tunica media, and perivascular cells, as well as a distinct type of gill cartilage. **i**, Following UV-mediated photoconversion of *fgf10b:nEOS* from green to magenta in a subset of filaments, re-imaging 3 days later revealed contribution of converted cells to pillar cells (yellow arrow, white reflects mixture of converted magenta and new unconverted green *fgf10b:nEOS*) and gill chondrocytes (white arrow) (*n* = 3). Draq5 labels nuclei in grey. **j**, Gene Ontology (GO) terms of biological processes for the respective clusters at the stages indicated. Heatmap reflects the negative log of the adjusted p value. Ectomesenchyme is defined as the average of all cells at 1.5 and 2 dpf stages. Scale bar, 100 μm.

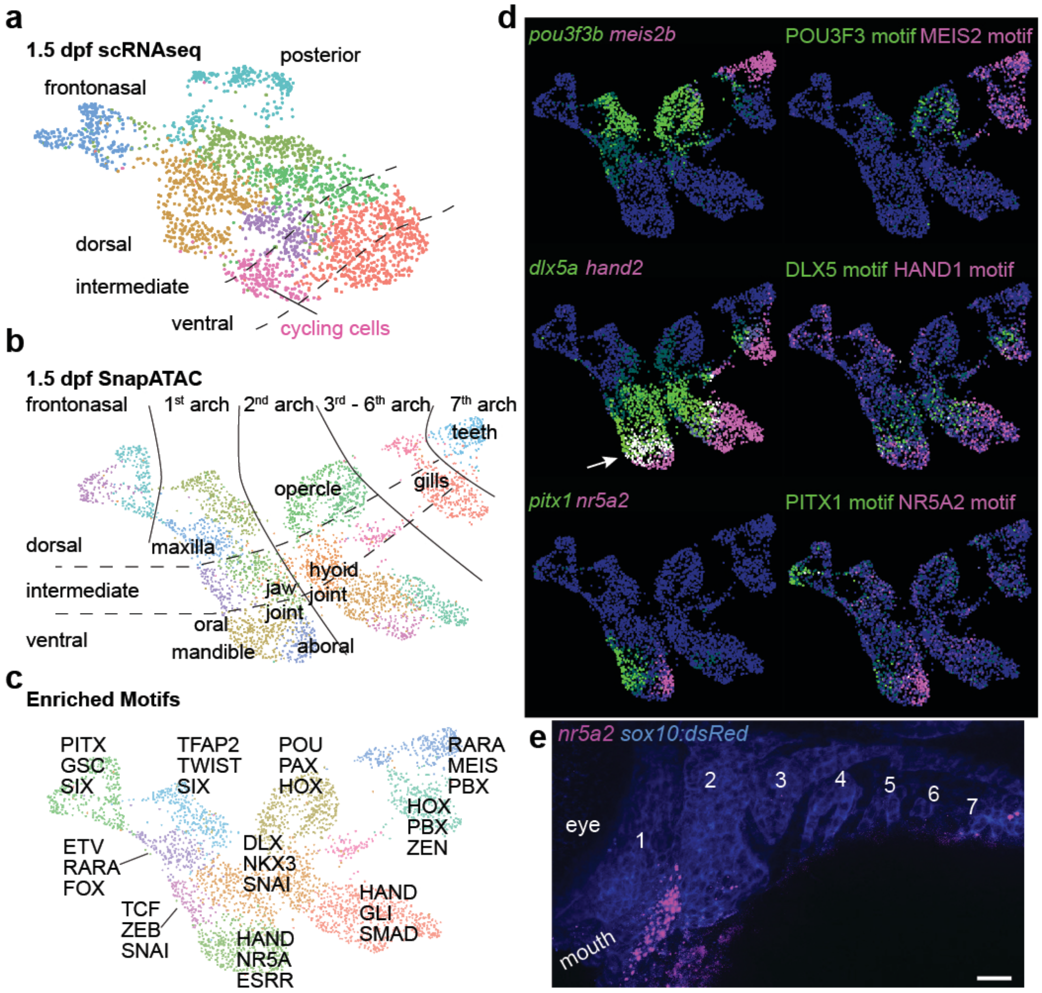


Figure 3. Highly resolved embryonic spatial expression domains from integrated datasets. **a**,**b**, UMAPs at 1.5 dpf generated by scRNAseq versus integration of scRNAseq and snATACseq datasets using SnapATAC. SnapATAC outperforms scRNAseq in resolving dorsoventral (vertical), anterior-posterior (horizontal), and major arch landmarks including a previously unappreciated oral-aboral axis in zebrafish. **c**, Select enriched transcription factor binding motifs for each cluster. **d**, Gene body activities for transcription factors and their corresponding DNA-binding motifs reveal tight correspondence with published expression patterns, including zebrafish-specific overlap of *dlx5a* and *hand2* in the mandibular arch (arrow). **e**, Fluorescent in situ hybridization shows restricted expression of *nr5a2* in the aboral domain of the mandibular arch as predicted by SnapATAC. *sox10:dsRed* labels CNCCs of the arches (numbered, anti-dsRed antibody stain). Scale bar, 20 μm.

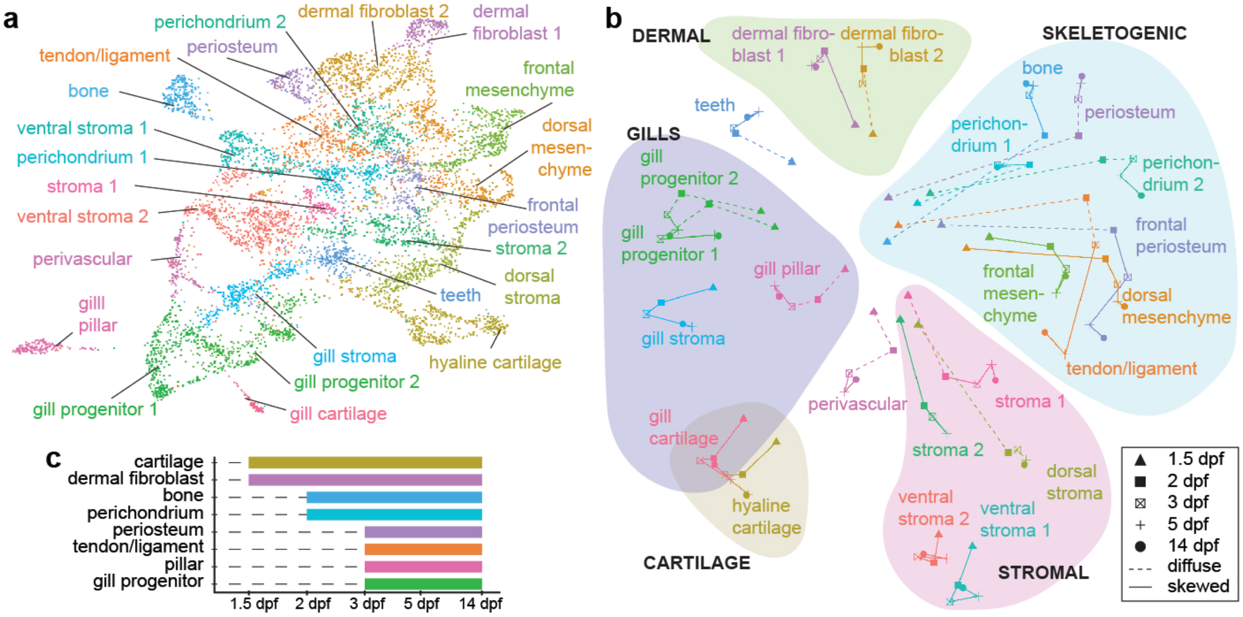


Figure 4. Cell type competency mapping through retrograde chromatin accessibility analysis. a, UMAP of 14 dpf SnapATAC data shows cell clusters for which top accessible peak modules were calculated for Constellations analysis. **b,** Constellations analysis involves mapping of cluster-specific chromatin accessibility from 14 dpf back to earlier stages and then plotting relatedness of mapped accessibility in two dimensions. Diffuse refers to a stage where cluster-specific chromatin accessibility does not map to a discrete portion of UMAP space, and skewed where it does (see Methods for details). Groups of related cell types are color-coded. **c**, Graphical representation from the Constellations analysis of when chromatin accessibility of major cell types first shows a skewed distribution in UMAP space, suggestive of establishment of competency.

a, UMAP of 14 dpf SnapATAC data shows cell clusters for which top accessible peak modules were calculated for Constellations analysis. **b**, Constellations analysis involves mapping of cluster-specific chromatin accessibility from 14 dpf back to earlier stages and then plotting relatedness of mapped accessibility in two dimensions. Diffuse refers to a stage where cluster-specific chromatin accessibility does not map to a discrete portion of the UMAP space, and skewed to where it does (see Methods for details). Groups of related cell types are color-coded. **c**, Graphical representation from the Constellations analysis of when chromatin accessibility of major cell types first shows a skewed distribution in UMAP space, suggestive of establishment of competency.

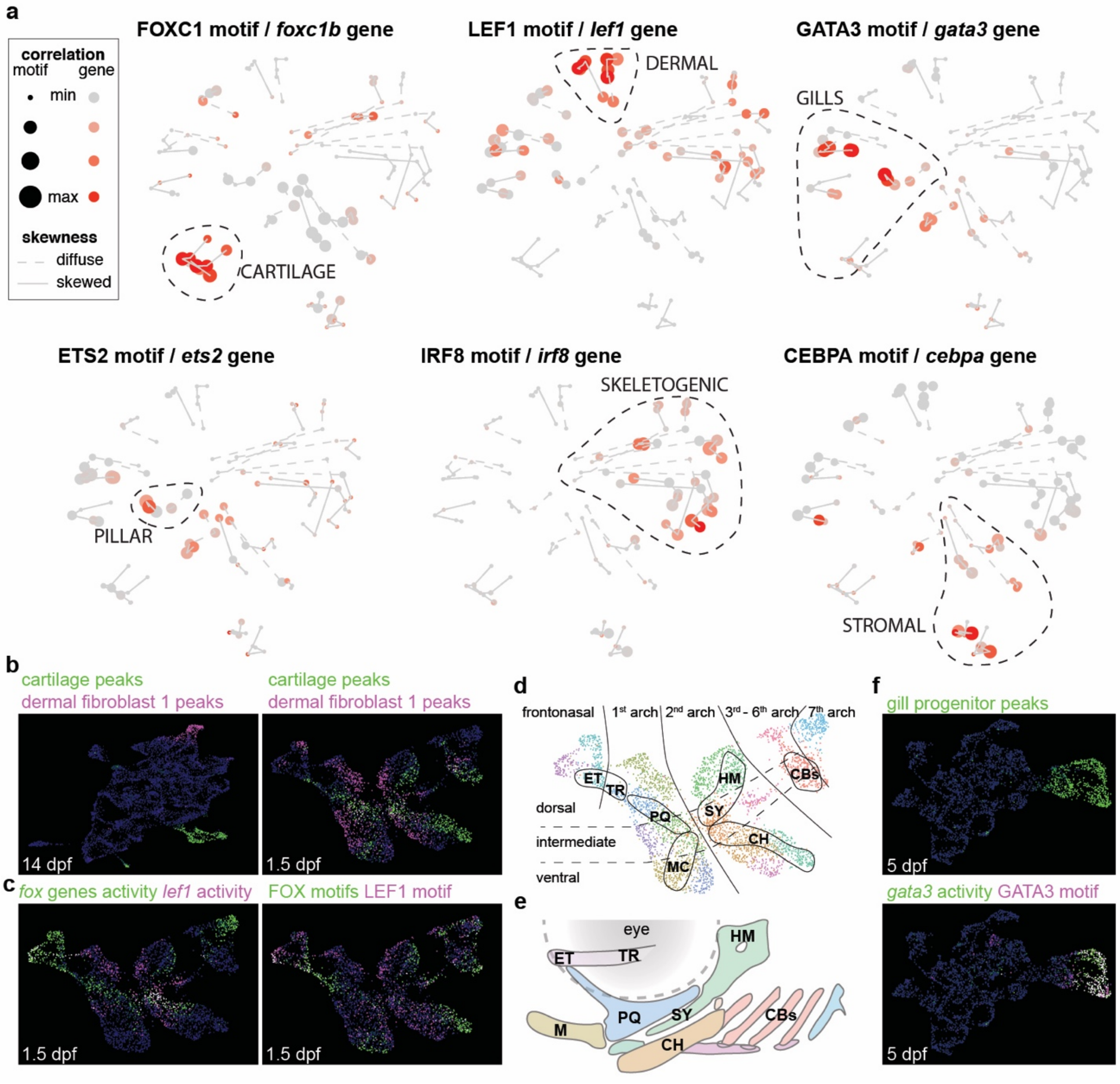


Figure 5. Constellations analysis reveals candidate transcription factors for lineage priming. a, Mapping onto the Constellations plot of transcription factors with correlated gene body activity and DNA-binding motif enrichment in specific clusters. Sizes of circles denote correlation of peak module mapping to motifs, and red color to gene body activities. **b**, Top peak modules for hyaline cartilage and dermal fibroblast 1 clusters at 14 dpf mapped onto 14 dpf and 1.5 dpf SnapATAC UMAPs. **c**, Summed gene body activities of *foxc1a*, *foxc1b*, *foxf1*, *foxf2a*, *foxf2b* and summed FOXC1, FOXF1, and FOXF2 motifs at 1.5 dpf correlate with cartilage peak mapping,

and *lef1* gene body activity and LEF1 motif with dermal fibroblast peak mapping. **d**,**e**, Retrograde cartilage accessibility mapping at 1.5 dpf allows predictions of the arch origins of the individual cartilaginous elements of the week-old skeleton: ceratobranchials (CBs), ceratohyal (CH), ethmoid (ET), hyomandibula (HM), Meckel's (M), palatoquadrate (PQ), symplectic (SY), and trabecula (TR). **f,** Mapping of the top peak module for 14 dpf gill progenitor clusters onto 5 dpf SnapATAC UMAP shows correlation with *gata3* gene body activity and GATA3 motif.

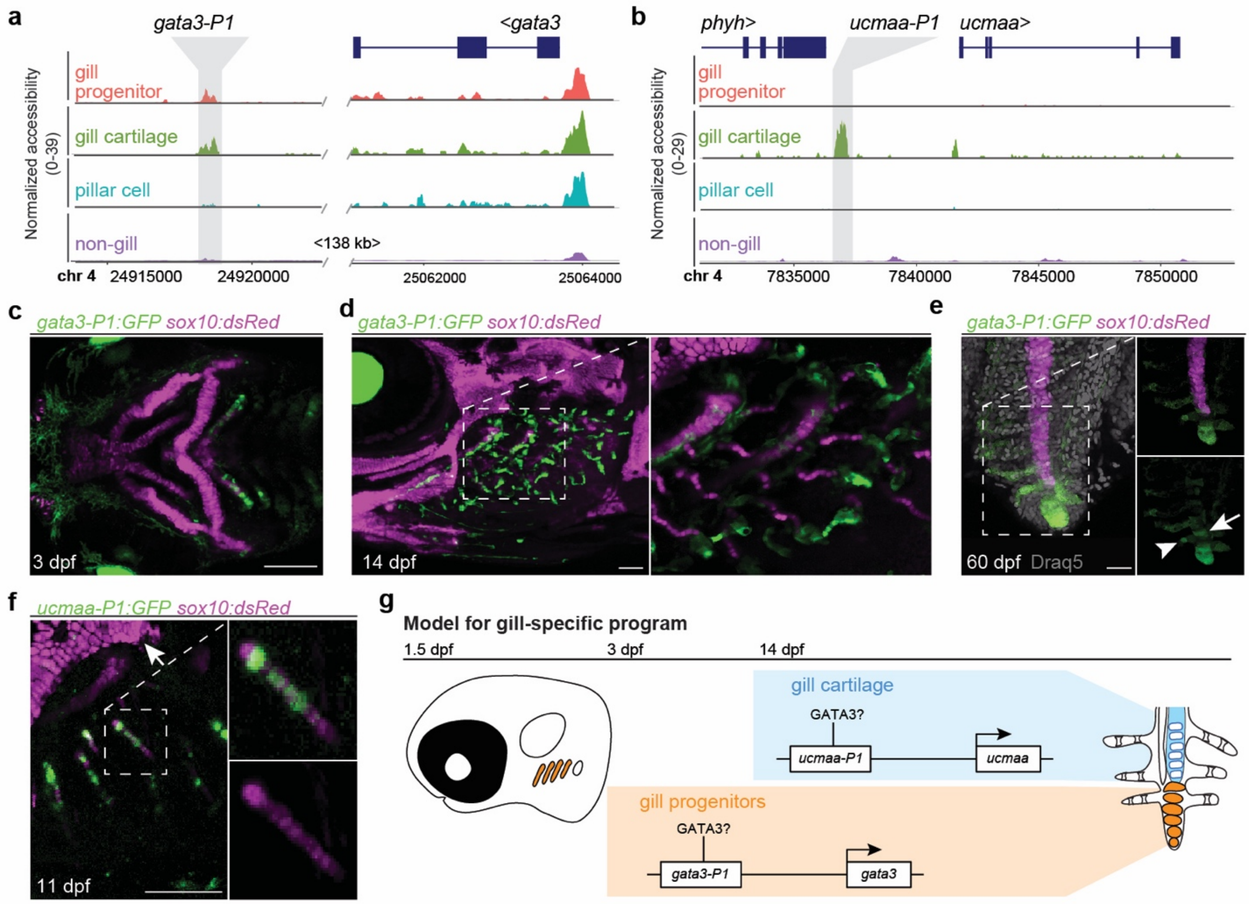


Figure 6. Gata3 activity distinguishes the gill-specific lineage. a,b, Genome tracks show chromatin accessibility for cell clusters from snATACseq data at 14 dpf. Grey shading shows gill-specific accessible regions near the *gata3* and *ucmaa* genes. Chromosome positions refer to the GRCz11 genome assembly. **c-f**, *gata3-P1:GFP* drives expression in the posterior gill-forming

arches at 3 dpf, the gill filament system at 14 dpf, and gill progenitors at the tips of primary filaments at 60 dpf, as well as some pillar cells (arrowhead) and gill chondrocytes (arrow) near the growing tips. *sox10:dsRed* labels cartilage and Draq5 nuclei. **f**, *ucmaa-P1:GFP* drives highly restricted expression in *sox10:dsRed*+ gill chondrocytes (boxed region shown in merged and single channels to the right) but not hyaline cartilage (top left). **g**, Model shows initiation of *gata3* expression in the posterior gill-forming arches through the *gata3-P1* enhancer, maintenance of *gata3* in gill progenitors at the tips of growing filaments, and expression of *ucmaa* in gill cartilage through the *ucmaa-P1* enhancer. Both *gata3-P1* and *ucmaa-P1* contain predicted Gata3 binding sites. Scale bars, 100 µm (c,d,f), 20 µm (e).