Single-cell multi-omics, spatial transcriptomics and systematic perturbation decode circuitry of neural crest fate decisions

Zhiyuan Hu[1,2]*, Sarah Mayes[1], Weixu Wang[3], José M. Santos-Pereira[1,6], Fabian Theis[3], and Tatjana Sauka-Spengler[1,4,5,7]*

Affiliations:
[1] MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford; Oxford, United Kingdom.
[2] Medical Research Institute, Frontier Science Center for Immunology and Metabolism, Wuhan University; Wuhan, China.
[3] Institute of Computational Biology, Helmholtz Center Munich; Munich, Germany.
[4] Stowers Institute for Medical Research; Kansas City, MO, USA.
[5] MRC WIMM Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom.
[6] Current address: Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain; Departamento de Biología Celular, Facultad de Biología, Universidad de Sevilla, Seville, Spain.
[7] Lead Contact
* Correspondence: TSauka-Spengler@stowers.org (T.S.-S.); zhiyuan.hu@whu.edu.cn (Z.H.)

Abstract

Cranial neural crest (NC) cells, which can migrate, adopt multiple fates, and form most of the craniofacial skeleton, are an excellent model for studying cell fate decisions. Using time-resolved single-cell multi-omics, spatial transcriptomics, and systematic Perturb-seq, we fully deciphered zebrafish cranial NC programs, including 23 cell states and three spatial trajectories, reconstructed and tested the complete gene regulatory network (GRN). Our GRN model, combined with a novel velocity-embedded simulation method, accurately predicted functions of all major regulons, with over a 3-fold increase in correlation between in vivo and in silico perturbations. Using our new approach based on regulatory synchronization, we discovered a post-epithelial-mesenchymal-transition endothelial-like program crucial for migration, identified motif coordinators for dual-fate priming, and quantified lineage-specific cooperative transcription factor functions. This study provides a comprehensive and validated

NC regulatory landscape with unprecedented resolution, offering general regulatory models for cell fate decisions in vertebrates.

Introduction

As the embryo develops, multipotent cells make crucial fate decisions to transition from precursor cells into derivative states[1]. An overarching goal is to delineate all gene regulatory events driving these cell fate decisions in a comprehensive and quantitative manner. Single-cell profiling, coupled with perturbation screens, is foundational for reconstructing the regulatory landscape of embryonic development and advancing in silico perturbation technologies. Although recent advancements in single-cell multi-omic techniques have empowered GRN inference[2,3] and perturbation simulation, previous GRN models are limited to being trained on unpaired datasets of profiling and perturbation or in vitro scenarios, which introduces bias or fails to capture the complexity of living organisms. Additionally, the absence of full-length transcript single-cell transcriptomic data and the long intervals between sample collections hinder the accuracy of velocity models and the accurate construction of developmental trajectories. Therefore, to achieve unbiased insights into cell fate decisions and improve perturbation simulation methods, it is essential to collect in vivo, temporally contiguous single-cell multi-omic data for GRN construction, along with full-length transcript data for velocity inference and sample-matched high-throughput Perturb-seq data.

The neural crest (NC) is a transient structure in vertebrate embryos that gives rise to diverse tissues. NC cells are categorized as cranial, trunk, vagal, sacral, or cardiac, depending on their location along the body axis and their specific derivatives[4,5]. Among them, cranial NC cells arise from the neural tube in the head region and produce critical tissues such as the craniofacial mesenchyme, which forms cartilage, bone, neurons, glia, and connective tissues in the face[6-8].

Defects in cranial NC cells are linked to congenital disorders such as Treacher Collins syndrome[9], Hirschsprung's disease[10], and craniofacial abnormalities[11].

Research on NC development has yielded several key findings. First, neuroepithelial precursors are primed for cranial NC fate by reactivating canonical pluripotency factors[12] and suppressing a neuronal program driven by ZIC, PAX2, and OTX factors[13]. Second, a three-phase model of NC cell fate decisions was proposed based on transcriptional analysis in mice. This model suggests that competing fate programs are initially coactivated, followed by increased synchronization of relevant programs, and finally, repression of competing programs to establish a specific fate[14]. Lastly, studies on in vitro cranial NC cells identified an active *cis*-regulatory element called the TWIST1-homeodomain "Coordinator", which mediates cooperative and selective binding between the mesenchymal regulator TWIST1 and homeodomain transcription factors (TFs), providing insights into DNA-guided regulatory mechanisms[15]. However, the in-depth and full regulatory dissection of cranial NC development is imperative to fully understand these cell fate decision models, identify new regulatory elements, and explore their nuanced roles using paired RNA and ATAC data at the single-cell level. This matched data is crucial because the computational integration is often flawed by misalignment between ATAC-inferred gene activities and transcriptional states of individual cells in a quickly developing system.

Despite the need, this comprehensive dissection of the cranial NC remains absent due to challenges such as scarce cell representation in embryogenesis data[16-19], enormous spatiotemporal and cell fate heterogeneity, and the rapid transitions of cell states within remarkably short developmental windows. Notably, no GRN has been constructed using

temporally continuous, paired single-cell multi-omic data and validated by global Perturb-seq for cranial NC cell fate decisions or, strikingly, for any in vivo embryonic system.

Here, to fill the gap, we present the paradigm that defines and functionally validates a complete GRN orchestrating cranial NC ontogeny in zebrafish, by leveraging recent advances in single-cell multi-omics, GRN inference[2,20,21], precise reverse-genetics perturbations[22-24], spatial transcriptomics[25,26], and deep learning algorithms[27,28]. This work reconstructs robust single-cell trajectories and a comprehensive single-cell NC GRN of cranial NC development with unparalleled resolution. Our GRN analysis identifies fate-specific time-sensitive roles of 186 regulons, including less studied yet essential factors (ETS, EBF, BACH, TEAD), from ten major programs. We further validate the robustness and predictive power of our NC-GRN, in contrast to previous qualitative GRN models[29], by combining in vivo Perturb-seq data and GRN-informed perturbation simulation by RegVelo (W. Wang et al., F. Theis, manuscript in preparation). We systematically probe all major TFs across progressive stages, identify new Coordinator motifs that drive DNA-based TF cooperativity, and reveal an endothelial-like program essential for NC migration and differentiation. Overall, this study presents the most comprehensive vertebrate GRN to date, builds on foundational theories such as Eric Davidson's GRN concept[30,31] and the Waddington regulatory landscape model, and serves as an unique reference for studying cell fate decisions and developing GRN-based methodologies.

Results

Single-cell multiomic and spatial landscape of normal and perturbed NC

There are four distinct phases of NC cell development: 1) specification at the NPB, 2) epithelial-to-mesenchymal transition (EMT) from the neural tube, 3) migration throughout the

periphery, 4) differentiation into distinct cell types (Fig. 1a)[32]. Zebrafish embryos develop rapidly and externally, enabling the analysis of contiguous time points in NC development over a matter of hours and in thousands of isogenic animals.

To decipher normal and perturbed NC development, we examined embryos with and without the critical TF Foxd3[33-35]. Foxd3 is required for differentiation of the craniofacial skeleton, and its loss promotes neuronal differentiation of NCs at the expense of mesenchymal derivatives[36,37]. We have also previously shown that FoxD3 acts bimodally—initially as a pioneer factor to prime NC specification and subsequently as a transcriptional repressor to direct NC cells into specific cell lineages[36]. We isolated wild-type cells (heterozygous *foxd3*$^{+/-}$, referred to as wild type throughout, i.e. Citrine-positive/mCherry-negative or Citrine-negative/mCherry-positive cells) and *foxd3*-mutant cells (homozygous *foxd3*$^{-/-}$, referred to as *foxd3* mutant throughout, i.e. Citrine-positive/mCherry-positive cells) by flow cytometry from the embryos crossing two gene-trap lines, gt(*foxd3*-Citrine) and gt(*foxd3*-mCherry)[37] (Fig. 1a).

To profile cranial NC cells from their inception to differentiation into distinct cell types, we analyzed the cells isolated across four ranges of combined time points using 10x Multiome, which profiled transcriptomes by single-nucleus RNA sequencing (snRNA-seq) and chromatin accessibility by single-cell ATAC-seq from the same cell (Fig. 1b, Extended Data Fig. 1a-c, Supplementary Table 1). Additionally, we analyzed seven distinct time points with ultra-deep Smart-seq3 single-cell RNA-seq (scRNA-seq). This dual approach allowed us to capture continuous cell state changes and link them to specific developmental times. Our Smart-seq3 data achieved a median coverage of 7,791 genes per cell to complement the sparse multiome-RNA data (Fig. 1c, d, Extended Data Fig. 1d). After quality filtering, our single-cell multiomic dataset encompassed 24 major cell populations, including NC cells and *foxd3*-expressing

neuromesodermal progenitors and derivatives that are not the focus of this study (Fig. 1b, c; Extended Data Fig. 1c). To collect enough cells for the early stage between 75% epiboly and 4 somite stage (ss), we combined wild-type and *foxd3*-mutant cells for multiome profiling and then used amplicon-seq combined with machine learning approaches to demultiplex cellular genotypes (Extended Data Fig. 1e).

We next dissected the NC ontogeny by single-cell multi-omic and spatial transcriptomic analysis in depth. Subsequent clustering of NC populations from 8 - 22 hpf and two genotypes yielded 23 distinct cell states, reproducible in both 10x Multiome and Smart-seq3 data (Fig. 1e-h). The NC data showed trajectories starting from NPB, delaminating NC (dNC), migratory NC (mNC), to differentiated states, such as non-hox head mesenchyme (head mes.), pharyngeal arches (arches), and pigment lineages, according to collection stages, marker gene expression and motifs enrichment (Fig. 1i, j, Extended Data Fig. 1f, Supplementary Table 2). These identified cell states were reproducible in Smart-seq3 data (Fig. 1e, Extended Data Fig. 1g-h). In this work, we defined the cell state as the current condition or status of a cell, which can be transitional, and the cell fate as the developmental path or future outcome of a cell. For example, dNC and mNC are cell states but not cell fates, while differentiated cell states (e.g., gch2-high pigment, sox6-high pigments) are cell states of the pigment fates.

To validate NC cell states and profile their spatial distribution systematically, we designed a MERSCOPE 300-gene panel and aligned multiome-RNA (i.e. 10x Multiome gene expression assay data) clusters with spatial transcriptomic data (Extended Data Fig. 1i-j). Our single-cell-level spatial transcriptomic data captured NC cell states (NPB, dNC, mNC, pigment, non-hox mesenchyme, and the first two pharyngeal arches etc.), and demonstrated that the NC cells migrated along the forebrain-optic and optic-midbrain boundaries and delaminated near the

dorsal neural cells (Fig. 1k). Our continuous dataset captured states of NC development that had previously been overlooked due to the intermittent nature of the organogenesis data[17], and also captured mutant-specific states (Fig. 1l).

Reconstruction of normal and perturbed cranial NC trajectories

To chart the NC developmental lineages and their underpinning regulatory machinery (Fig. 2a), we first inferred RNA-velocity-based trajectories from the multiome-RNA data. The RNA-velocity analysis predicted lineage dynamics using ratios of spliced to unspliced RNA[38]. To account for the loss of cytoplasmic mRNA and incomplete-length transcript information in the multiome-RNA data, we reproduced the multiome-RNA trajectories using the Smart-seq3 dataset.

The trajectories of wild-type NC cells recapitulated the NC ontogeny: from the NPB to NC delamination, followed by migration and differentiation into various structures, including head mesenchyme, pharyngeal arches, and pigment progenitors (Fig. 1i, 2b, 2c). The trajectories of cranial NC cells also reconstructed three distinct migratory routes along the cranial-caudal axis, named non-hox, hox2, and hox3/4. HCR validated major states in cranial NC, e.g., the non-hox head-mesenchyme (*postnb* and *foxc1a*, which are skeletogenic and cartilage markers)[39], pharyngeal arches 1-2 (*dlx2a*, *grem2b*), mNC_hox34 (*itga3a*), pigment lineages (*mitfa*), and migratory NC cells (*rxraa*) (Fig. 2d-f, Extended Data Fig. 2a-e). In addition to the consistency between latent time and real-world time, the high velocity confidence values confirmed the reliability of our trajectory analyses (Fig. 2g, Extended Data Fig. 2f).

The trajectories of *foxd3*-mutant NC cells were distinct from wild-type NC cells, and showed early disruptions that began during the NC delamination phase (Fig. 2h, i). The second migratory route was completely absent, and mutant cells in the first (midbrain, the first and

second rhombomeres) and third (the fourth rhombomere) migratory routes undertook non-canonical migratory patterns, consistent with earlier phenotypic findings[37]. Transcriptomic analysis revealed that mutant migratory cells established a new trajectory/derivative branch containing cells that expressed *elavl3* (Cluster Mutant_nohox_12_22ss) and preserved neuronal potential. Another *foxd3*-mutant cluster (Mutant_nohox_cycling) exhibited a partial mesenchymal identity (*alx4a* and *adamts6*-expressing cells) but did not express standard mesenchymal markers (Fig. 1j). Overall, *foxd3*-mutant clusters bore a closer resemblance to the neural floorplate and hindbrain roofplate fates than to canonical NC derivatives (Extended Data Fig. 2g).

NC GRN identifies major regulatory programs and key elements

SCENIC+ combines concurrent snRNA-seq and scATAC-seq data to identify regulons comprising TFs, *cis*-regulatory elements, and their target genes[2]. We applied SCENIC+ to our multiome data and identified 186 high-quality enhancer-driven regulons, corresponding to 155 TFs (Fig. 2j, Supplementary Table 3). We then quantified regulon activities in individual cells using the gene-rank-based AUCell method to obtain gene-based and region-based regulon activity profiles. Based on regulon activity and target gene functions, regulons were categorized into ten major programs, including NC specification (cluster 1: foxd3+, tfap2a+, sox2-), NPB/EMT (cluster 3: zeb1a/1b+, zic1/2b/4/6+, otx1/2a/2b+), migration (cluster 7: foxd3+/-, fli1a+, rxraa+), connective tissue/skeletal system (cluster 4: esrrga+, ets1+, egr2a/b+), hindbrain (cluster 9: sox9a/b+, hoxb1a+, sox3+), and chromatin remodeling/cell cycling (cluster 8: ctcf+) (Extended Data Fig. 2h-j, Supplementary Table 4).

Among the predicted regulons, we recognized twist1b(+) and zeb1a(+), which are associated with NC delamination[40], and sox10(+), which is associated with melanocyte fate determination[41] (Fig. 2j), where '(+)' denotes co-expression of TFs with their target genes. The

NC specifiers *tfap2a* and *foxd3* emerged as core TFs, defined by their high out-degree centrality within the overarching network (Extended Data Fig. 3a), consistent with loss of NC formation in *tfap2a;foxd3* double mutant zebrafish[42]. In addition to well-established drivers like *zeb1a*, *mitfa*, *sox9b*, less-studied factors such as *bach2b, ebf3a, tead1a, smarcc1a, bhlhe40, si:ch73-37h15.2*, also exhibited high out-degree centrality, suggesting their importance as key nodes or drivers in the NC GRN. For instance, bach2b(+) targets were enriched in processes related to Hippo signaling, transcription regulation, and cell migration (Supplementary Table 5). The TFs *rxraa* and *elk3*, which are downstream of *foxd3*, were downregulated in *foxd3* mutants, confirming that our predicted GRN reflects genuine regulatory dynamics (Fig. 2k, l).

We validated SCENIC+ *cis*-regulatory element candidates by examining the profiles of four histone modifications during the migratory phase (12-16 ss). Specifically, we detected H3K27ac and H3K4me1, which mark active and poised enhancers respectively, H3K27me3 (heterochromatic regions), and H3K4me3 (activation of transcription)[43-46]. Migratory *cis*-regulatory element candidates were enriched for H3K27ac and depleted for H3K27me3 in the wild-type migratory NC cells and displayed elevated H3K27me3 levels (repressed state) in mutant NC cells of Cluster Mutant_nohox_12_22ss (Extended Data Fig. 3b-c). These findings underscore the dynamics of histone modifications, particularly those associated with *foxd3,* during NC ontogeny.

GRN-informed perturbation probes new TF functions in silico and in vivo

So far, our work has charted the cranial NC molecular lineage map with its underlying gene regulatory circuits controlling cell states and fates across a continuous temporal scale. Using this map, we comprehensively probed regulon functions through a combination of global in silico predictions and targeted in vivo validations (Fig. 3a, b).

For in silico predictions, we used RegVelo to identify TFs impacting cell fate decisions (see Methods). RegVelo integrates the SCENIC+ GRN framework with Smart-seq3 data to predict the consequences of TF knockouts. This method combines a latent time autoencoder predictor with the velocity ordinary differential equation model, incorporating RNA velocity to simulate lineage-specific perturbation impacts. RegVelo achieved significantly increased accuracy in single-gene simulation (4.2- to 6.3-fold) and multiple-gene simulation (3.0- to 6.1-fold) compared to other methods using our NC-GRN. It also enabled more effective measurement of delayed effects compared to methods that only assess instant perturbation outcomes[2,20].

RegVelo simulations identified key drivers of cell migration (*nr2f5*, *ets1*, *elfl3*, *zic2b*, *sox6*, *ebf1b*), affecting both mesenchymal and pigment fates, drivers affecting non-hox head mesenchymal and arch 2 fate (*ets1*, *elk3*, *fli1a*, *rarga*, *erf*), crucial influencers in head mesenchymal fate decisions (*sox9b*, *nr2f2*, *tead1b*), and key drivers for pigment lineage differentiation (*tfec*, *mitfa*, *bhlhe40*, *sox10*, *elf1*) (Fig. 3c). Moreover, RegVelo results highlighted the potential previously unknown roles of transcriptional enhanced associate domain (TEAD) and early B cell factor (EBF) families in NC migratory and head mesenchymal fate decisions. Targets of EBF regulons (*ebf1b*, *ebf3a*, *ebf3a.1*) were enriched in processes such as cell-cell adhesion, cell migration, and skeleton morphogenesis, while ebf1b(+) regulon was enriched in histone methylation activity (Fig. 3d, Extended Data Fig. 4a). Tead1b(-) downstream targets showed enrichment in bone development, and TEAD3 (*tead3a/b*) regulons were enriched in processes including regulation of cell migration, cartilage development, and neural tube development, indicating a general role for TEAD3 in neural tube and NC development (Fig. 3e-f, Extended Data Fig. 4b, Supplementary Table 5).

For in vivo validation, we employed zebrafish F_0 crispant knockout[22] alongside direct-capture Perturb-seq, using chemically modified single guide RNAs (sgRNAs) to extend their lifespan to at least 3 days after injection and enhance detection rates[47] (Extended Data Fig. 4c-d). In addition, such modified sgRNAs were as detectable as naive sgRNAs via scRNA-seq and, meanwhile, ensured high knockout efficacy (Extended Data Fig. 4e-g).

Our knockout panels for the in vivo perturbation studies comprised five pigment regulons for proof-of-concept experiments and twelve newly identified migratory/mesenchymal regulons for functional validation. We selected TF drivers that had severe effects predicted by RegVelo, and that were defined as central nodes (Extended Data Fig. 2i-j). To address the limited number of NC cells per embryo and to mitigate inter-embryo variability, we pooled a minimum of 12 embryos per condition, ensuring a more accurate measurement of average effects. We achieved uniform sequencing depth per condition to minimize biases related to gene coverage (Extended Data Fig. 4h). Because inherent TF redundancy and mosaicism in crispant knockouts may obscure long-term effects, our perturb-seq analysis workflow was tailored to quantify short-term lineage-specific knockout effects (i.e., a few hours post effective knockout), inspired by CellOracle[20,48].

Our final Perturb-seq dataset (n = 39,453) comprised 22 single and combinatorial knockout conditions targeting 17 TFs (Fig. 3g-h, Extended Data Fig. 4h-l, Supplementary Table 1). Although exceptions were noted, likely due to redundant effects, most targeted TFs (15 out of 17) showed downregulation in the knockout conditions, confirming knockout efficiency (Extended Data Fig. 4l).

Overall, in vivo perturbation effects were highly correlated with RegVelo simulation results (Spearman correlation coefficients = 0.38 for single-gene knockout and 0.58 for multiple-gene knockout), validating the predictive power of our NC GRN (Fig. 3i). For example, single-gene knockouts of top pigment drivers (*tfec*, *mitfa*, *elf1*) predicted by RegVelo resulted in pigment lineage depletion, and the single-gene knockouts of migration drivers *nr2f5* led to both mesenchymal and pigment fate depletion. The single-gene in vivo perturbation effects of few ETS factors were not as significant as in silico coefficients, and we hypothesized that the in vivo effects were masked by functional compensation. Moreover, knockouts of Ebf3a.1, the newly identified driver from the EBF family, led to posterior fate depletion (arch 2, mNC_hox34, mNC_trunk), consistent with the regulon functional enrichment in cell migration (Fig. 3j). Next, we investigated the GRN architecture in depth, explored the global cell fate decision models, and addressed the functional compensation of lineage driver TFs by integrative analysis.

GRN-based regulatory dynamic quantification of EMT and fate decisions

Recognizing that cell-state-specific regulon activities presented only broad-brush GRN changes, we next aimed to provide a quantitative view of the regulatory dynamics over a continuous timeline. To reduce the impact of differential kinetics and enhance latent time prediction, we narrowed our focus to the non-hox trajectories.

We used transcriptomic and chromatin accessibility data to infer the latent time by MultiVelo[21] (Fig. 4a). When benchmarked against RNA velocity, multiomics-based latent time achieved better separations between the delamination and migration phases and better aligned with the real-world time (Extended Data Fig. 5a-c). We next predicted cell state transitions using CellRank[49,50], which recapitulated the non-hox trajectory from the NPB through delamination and migration, culminating in differentiated states (Fig. 4b, Extended Data Fig. 5d). Terminal

state probabilities, representing the likelihood of differentiated cell states, were minimal during the NPB and delamination phases but increased substantially during NC migration (Extended Data Fig. 5e), suggesting the emergence of fate bias during migration. At the gene level, velocity dynamics predominantly adhered to the induction model (52%, characterized by partial kinetics with only the induction phase) and Model 1 (38%, characterized by complete kinetics with both induction and repression phases and chromatin closing before transcriptional repression), indicating that gene transcriptional repression in this process is modulated primarily by chromatin remodeling (Extended Data Fig. 5f).

To overcome the high dropout rates in gene-level analysis, we moved to regulon-level GRN dynamics quantification, integrating latent time and regulon activities. AUCell scores were derived from either RNA or ATAC data to reflect regulon activities based on gene expression ("gene") and chromatin accessibility ("region"), respectively.

EMT is critical for NC delamination and migration. Our dynamic analysis of GRNs revealed that the region-based activities of zeb1a(+), zeb1b(+), and twist1b(+) regulons were initiated during the NPB phase, followed by expression activation in both NPB and delamination phases (Fig. 4c). Notably, EMT *cis*-regulatory element accessibility diminished during migration, indicating that EMT is a transient process strictly confined to epithelial cells. Furthermore, de novo motif discovery analysis identified a novel Coordinator motif harboring CTCF, SNAI, TWIST, ZIC, and SMARCC motifs, confirmed by TF footprinting (Extended Data Fig. 5g, h). Enrichment in ZIC motifs within chromatin-accessible regions in NPB cells (Figure 1J), suggests that early ZIC TF binding within the coordinator may be a prerequisite at the DNA-sequence level for EMT program activation later, indicated by SNAI and TWIST TF binding. Our findings propose a robust model wherein EMT initiates in epithelial cells, is operational

during delamination, and is quickly deactivated in fate-determined cells, modulated by CTCF and SMARCC factors alongside canonical EMT factors.

To dissect lineage-specific dynamics leading to NC differentiation, we incorporated cell fate probabilities into these temporal processes and defined cluster-specific regulons based on their regulon specificity scores (RSSs). We analyzed activity cascades of these regulons over latent time, which showed that regulon-wise chromatin accessibility preceded gene expression for pigment and mesenchymal lineage markers (Fig. 4d, e). We infer that enhancer-driven gene activation is a primary regulatory mechanism defining NC-derived mesenchymal lineages. Both transcriptomic and chromatin accessibility data showed that the pigment lineage activated mesenchymal regulons during migration (Fig. 4f, g). In mesenchymal lineages, early pigment regulons, i.e., tfap2b(+) and sox10(+), were upregulated during delamination and repressed during migration, while late pigment regulons, e.g., mitfa(+) and tfec(+), displayed mild activation, likely due to stochasticity during migration. This finding aligns with the previously proposed expression-based fate decision "co-activation" model, where the transcriptional modules of both competing fates get activated prior to the fate bifurcation event [14]. Importantly, our results further suggest that chromatin accessibility modulates the co-activation of transcriptional programs, and stochastic activation of dual programs plays a permissive role before cellular commitment to a specific fate (Fig. 4h).

Coordinator motifs enable dual fate potentials

To elucidate the regulatory mechanisms underlying cell fate decisions at the DNA sequence level, we applied de novo motif discovery on Tn5-bias-corrected profiles using base-resolution deep learning models, aiming to reveal NC *cis*-regulatory sequence syntax in vivo (motifs and their Coordinators). This analysis identified key motif dominance patterns implicated in NPB, delamination, migration, and differentiation processes (Fig. 4i, j), with ETS motifs found

prevalent in the mesenchymal fate definitions, RXR/RAR motifs in pharyngeal arch derivative formation, and bHLH motifs in defining pigment lineages.

Among the de novo discovered motifs, a Coordinator motif (Combi_593, bHLH+AP-2+ETS) suggested the priming role for Tfap2a in defining both mesenchymal and pigment fate potentials (Extended Data Fig. 5i). This Coordinator was enriched in the accessible chromatin regions of early cell states (NPB and dNC), indicating the acquisition of dual-fate potential during NC specification, even before the expression of migratory ETS factors (*elk3, fli1a* etc.). This finding provides the DNA sequence basis for the aforementioned "co-activation" theory [14]. Further investigation of TFs expressed in early NC populations identified ETS candidate factors (*gabpa, ets1, etv4, and etv5b*) that likely co-bind with Tfap2a, with Gabpa emerging as the most probable binding partner based on AlphaFold structural predictions (Extended Data Fig. 5j-m).

Additionally, we noted the enrichment of NR2/RXR/RAR motifs in migratory NC cells and in a *foxd3*-mutant cluster (Mutant_nohox_cycling), implying that NR2 factors may facilitate an alternative pathway for *foxd3*-mutant cells to activate RXR/RAR regulons and undergo migration (Fig. 4j).

foxd3 **deficiency compromises NC migration and promotes the neuronal fate**

Our data revealed that *foxd3* deficiency leads to substantial changes in transcriptomic profiles and chromatin states. As a proof-of-concept, we explored how *foxd3* affects *cis*-regulatory elements and downstream genes to compromise NC migration and promote the neuronal fate.

GRN predictions with wild-type NC uncovered migration programs that include factors from the RAR, RXR, ETS, and nuclear receptor (NR) families (Fig. 5a). Notably, these regulons

were activated in the migratory wild-type NC cells, but not in the *foxd3* mutant counterparts in Cluster Mutant_nohox_12_22ss (Fig. 5b). A comparison between wild-type and *foxd3* mutant cells showed differential regulon activity that began as early as the delamination phase and intensified during migration (Fig. 5c-f). A specific set of chromatin-related regulons (ctcf, creb1b, e4f1, creb1a) was active in wild-type but not mutant NC cells at delamination (Fig. 5c). These chromatin-related regulons may induce early-stage chromatin remodeling and repress neural fates in wild-type NC.

We hypothesized that some migratory TFs, such as *rxraa*, were directly primed by *foxd3*. We focused on *rxraa,* because it possessed high out-degree centrality and was downstream of *foxd3* in the SCENIC+-inferred GRN. To locate the *cis*-regulatory element candidates activating these migratory TFs, we used the deep learning approach ChromBPNet, which learns to factorize Tn5 bias from ATAC profiles[51]. Bias-free ChromBPNet models yielded contribution scores that facilitated the identification of putative TF binding sites. Analysis of putative *rxraa cis*-regulatory elements and their base-resolved contribution scores identified an element that increased in accessibility post delamination and displayed high contribution scores at putative Tfap2a and Foxd3 binding sites, suggesting that this *rxraa* enhancer may be primed by Foxd3 and Tfap2a during delamination (Fig. 5g). Notably, both *rxraa* expression and accessibility of this *cis*-regulatory element were reduced in the *foxd3* mutant (Fig. 2k, Fig. 5g).

Comparing mutant to the wild-type migratory NC cells, we identified a group of mutant cells expressing a neuronal transcriptional signature that retained activities of neural and NPB regulons, such as zic1/4/6(+) and otx2a(+), typically repressed in wild-type migratory NC cells (Fig. 5b, h). The expression of an NPB TF, *zic2a*, absent from wild-type canonical NC at the same stage, persisted in *foxd3* mutants (Fig. 5i). Moreover, a neuronal marker, *elavl3*,

continued to be expressed in mutant cells, and the *cis*-regulatory element controlling its expression remained accessible (Fig. 5j-l). These findings suggest that *foxd3* functions to repress neural and neuronal regulons while activating migratory regulons driving canonical NC development. Our single-cell-level spatial transcriptomic analysis of mutants validated the *foxd3*-mutant-specific NC states, for both pigment- and neural-biased mutant states (Fig. 5m-o). Specifically, a *tfap2a/elavl3/cxcl14*-expressing and *foxd3* (mutant transcript)-expressing neural-fate-biased cell was located next to the midbrain optic domain, close to the neural niche populated by *foxd3*-negative cells (Fig. 5o). Note that the *foxd3* mutant NC cells still expressed partial unfunctional *foxd3* mRNA, which can be detected by sequencing and spatial transcriptomics, but did not produce functional Foxd3 proteins[36]. These data suggest that *foxd3*-mutant NC populations, which cannot obtain mesenchymal fates, are inclined to the neural fate due to environmental cues.

Redundant bHLH factors and elf1 drive pigment fate decision

Previous models have established the differentiation programs of pigment progenitor cells into iridophores, melanophores, and xanthophores at later developmental stages[52,53]. However, the initial specification of the pigment lineage within the migratory NC remains unresolved. Our analysis revealed that *tfec* was the earliest expressed bHLH factor, preceding the precursor factor *mitfa* and others (*tfeb* and *bhlhe40*), as evidenced in both multiome-RNA and Smart-seq3 datasets (Fig. 6a-b, Extended Data Fig. 6a). Perturb-seq analyses indicated that *tfec* and *tfec*;*mitfa* knockouts resulted in more profound depletion effects compared to *mitfa* (Fig. 3g, 6c-d). The upstream *cis*-regulatory element candidates of *tfec* harbored Sox10 TF binding sites, suggesting that *tfec* is downstream of *sox10* and is the earliest fate definer of pigment cell lineages (Extended Data Fig. 6b). Moreover, our data revealed a self-auto-regulatory loop and redundancy in pigment lineage specification, suggesting how cells become "trapped" in the differentiated fate (Extended Data Fig. 6c-d).

In addition to the traditionally recognized bHLH factors, we identified an ETS factor, *elf1*, as one of the top regulators whose knockout disturbed pigment fate decisions both in silico and in vivo (Fig. 3c, 3g, 6e). Moreover, *elf1* crispants were enriched for mesenchymal fates (head mesenchyme, arch 1, and arch 2), accompanied by the upregulation of *fli1a*, *erf*, and *etv2* and the downregulation of *mitfa*, suggesting that *elf1* represses the migratory states and mesenchymal fates (Fig. 6f, Extended Data Fig. 4l).

Predictive GRNs revealed TF effectors with individual and convergent functions

Having validated our GRN-informed in-silico and in-vivo methods using the proof-of-concept analysis of pigment lineage commitment, we investigated individual and convergent functions of TFs in non-pigment derivatives. We examined ETS (*erf, erfl3, ets1, fli1a, etv2, elk3, elf1*), NR (*nr2f2, nr2f5*), RXR/RAR (*rxraa, rarga*), SWI/SNF (*smarcc1a*), EBF (*ebf3a.1*), and TEAD (*tead3a/b, tead1b*) factors based on results of functional analysis, de novo motif discovery, and perturbation simulation.

To study the novel individual functions of these regulons, we analyzed crispant knockouts of these migration and mesenchymal factors, along with their ohnologues (Extended Data Fig. 6e). In the case of *nr2f5*, *nr2f2*, and *smarcc1a*, we observed short-term post-knockout effects, primarily manifesting as a reduction in the non-hox head mesenchymal fate (Fig. 6g). Conversely, *tfec* knockout, employed as a negative control due to its lack of impact on mesenchymal fate, consistently enriched the head mesenchymal fate (Fig. 6h). Moreover, knockouts of both *rxraa* and *rarga* not only reduced cranial NC cells but also significantly hindered migration in arches 1 and 2 (Fig. 6i). Apart from *rarga* and *nr2f2*, four single-TF (*rxraa, nr2f5, smarcc1a, erf*) knockouts negatively influenced both pigment and non-pigment lineages, implying that they possess global roles (Fig. 3g).

To describe TF functional convergence, we initially defined the effects with two metrics: redundant effects characterized by target sharing (indicated by high Jaccard indexes) and high motif similarity between regulon activators; and the non-redundant cooperative effects, potentially synergy, characterized by target sharing coupled with low motif similarity. For example, our GRN analysis indicated that tfec(+), mitfa(+), tfeb(+), and bhlhe40(+) performed redundant functions in pigment lineage specification.

To take temporal changes into account for TF cooperativity, we introduced the third metric termed "regulatory synchronization (SyncReg)". This metric characterizes the dynamic similarity between regulon pairs (Fig. 7a). We developed the temporal SyncReg analysis that maps regulon temporal activity patterns onto a two-dimensional plane, extracts features by neural networks and computes similarity among these temporal patterns. Using this approach, we identified diverse regulon temporal groups, including a group of regulons that shared similar functions in migration and mesenchymal differentiation: erf(+), rarga(+), nr2f2(+), rxraa(+), elk3(+), fli1a(+), nr2f5(-), and smarcc1a(-) (Fig. 7b). Notably, our findings suggested that the regulatory synchronization was independent of motif similarity or shared gene expression profiles, highlighting the combined role of TF binding and gene expression in defining specific NC states (Extended Data Fig. 6f). Moreover, GRN-informed RegVelo simulations of revealed stronger effects of two-gene knockout compared to single-cell knockout. For instance, co-perturbation of *tfec;mitfa*, *tfec;bhlhe40*, and *mitfa;bhlhe40* led to a 3- to 4-fold increase in the perturbation effect on downstream derivative states (Supplementary Table 6). Similarly, *sox9b*, a major driver of mesenchymal cell fate, exhibited strong depletion in the non-hox head mesenchymal lineage when co-perturbed with *sox6*, *nr2f2*, *nr2f5*, as well as migratory regulons (*ets1, elk3, fli1a*) (Fig. 7c).

An ETS-mediated endothelial-like program in NC migration and mesenchymal differentiation

We noticed that a subset of ETS regulons was activated in migratory NC cell states (Fig. 5c, 7b). Although *ets1* is known to be associated with NC migration in vertebrates[54,55], the collective function and potential synergistic interactions among ETS factors remain largely unexplored. Given that ETS factors are involved in vascular development and tumor migration[54,56,57], we considered that they may similarly influence migration and fate determination in NC cells. ETS factor zebrafish mutants have defects in vascular development, yet lack NC migration defects[58]. ETS motifs (5'-aGGAAg-3') were enriched in our scATAC-seq of wild-type migratory NC cells, but not in *foxd3* mutants, suggesting that *foxd3* may activate ETS factors and, in turn, enhancers containing ETS motifs (Extended Data Fig. 7a). Alternatively, the diminished accessibility of ETS motifs is a direct consequence of the *foxd3* knockout. Moreover, these ETS motifs were strongly enriched in endothelial cells (Fig. 7d), suggesting that an endothelial-like ETS-enriched regulatory program was activated in migratory NC cells.

Our Perturb-seq analysis further elucidated the roles of these ETS factors (*erf*, *erfl3*, *fli1a*, *elk3*, *etv2*) by quantifying the short-term effects of their knockout. Notably, the two-gene knockouts of *erf* and *erfl3* (paralogues) resulted in the disruption of both mesenchymal and pigment lineages at a single-cell resolution (Fig. 7e), whilst single *erf* knockout showed no obvious effects. The resultant phenotypes in the double *erf*;*erfl3* knockouts resembled the craniofacial characteristics associated with Chitayat syndrome (Extended Data Fig. 7b).

The individual gene knockouts (*elk3, fli1a*, *etv2*) did not result in any reduction of head mesenchymal fate (Fig. 7f, g), likely due to redundancy among ETS factors. Such redundancy

or convergent functions were also supported by GRN evidence, including high Jaccard indexes, motif resemblance, and the topological network centering on *fli1a* with strong self-auto-regulatory loops and mutual interactions (Extended Data Fig. 7c-f). Therefore, we performed combinatory perturbation and found that double *fli1a*;*elk3* knockout resulted in significant depletion of cranial mesenchymal identities (both non-hox and arch 2), and *fli1a*;*etv2* knockout resulted in non-hox head mesenchymal lineage, demonstrating functional compensation, particularly between fli1a(+) and etv2(+) in *fli1a*;*etv2* knockouts and between fli1a(+) and elk3(+) in *fli1a*;*elk3* knockouts (Fig. 3g, 7f, 7g).

Next, we aimed to determine whether ETS factors synergize with other concurrently active factors. We documented the coordinator of bHLH, AP-2, and ETS motifs, which was enriched in the NPB and delamination populations. Beyond the interactions with bHLH and AP-2 factors, de novo motif discovery also identified a putative coordinator of ETS and NR2F/RXR/RAR motifs (Fig. 7h). This was consistent with our GRN analysis results that ETS regulons shared high proportions of target genes with rxraa(+) and rarga(+) (Fig. 7i, Extended Data Fig. 7c-e) and the topological network surrounding *fli1a* indicated strong mutual interactions with RXR/RAR factors (Extended Data Fig. 7f). The overlapped targets between ETS factors (*elk3*, *fli1a*, *ets1*) and RXR/RAR factors (*rxraa*, *rarga*) were enriched in cell migratory processes, cartilage development, and bone morphogenesis (adjusted p-value $<$ 0.037). We hypothesized that ETS and RXR/RAR cooperate in modulating mesenchymal differentiation of cartilage and bone derivatives, as well as migration. Indeed, our GRN-informed two-gene perturbation simulation indicated that the combinations of ETS and RXR/RAR/NR factors resulted in high depleting coefficients in non-hox head mesenchyme and arch 2 fates (Fig. 7c). Moreover, *elk3* exhibited co-binding potentials with *rxraa*, and *ets1* exhibited co-binding potentials with *rarga* (Extended Data Fig. 7g-h). These findings suggest

that synergy between ETS and RXR/RAR factors, facilitated by DNA co-binding, plays a crucial role in NC migration and differentiation of head mesenchyme, cartilage, and bone.

Discussion

Our study reconstructed a comprehensive GRN underlying cranial NC development. We developed novel quantitative and cellular tools to define and test the lineage, state, and fate-specific roles of all essential factors in the NC landscape, fully reconstituting Waddington epigenetic landscapes in vivo[59,60] (Fig. 7j). This spatiotemporal-resolved NC developmental landscape was absent in previous organogenesis studies[17-19]. We identified stable states, known as attractors, within this landscape and characterized the NPB-to-dNC potential surface as monostable, resembling a single valley or 'creode', while the mNC-to-derivatives exhibited bistable or multistable surface, where the valley splits into multiple states. These transitions are modulated by chromatin remodeling, as evidenced by the activities of CTCF and other remodeling factors. Furthermore, our study elucidated how GRNs contribute to shaping this landscape. For instance, we observed that bHLH and ETS factors establish positive feedback loops, stabilizing attractor states such as pigment and mesenchyme. On the other hand, loss of Foxd3, which inhibits the neural fate, stabilized a new NPB-“early mutant NC”-“neural fate” creode and the destabilization of the original mesenchymal creode. Using our understanding of GRNs, we demonstrated how modulating single or multiple regulons can alter the landscape, inducing or repressing specific fates.

We found that ETS-like factors and the entire endothelial pathway are essential for NC migration, a role that was previously masked by functional redundancies. We confirmed the directive role of RXR and RAR factors in driving formation of the pharyngeal arch derivative, and NR2F and SWI/SNF factors in driving the non-hox head mesenchymal fate. Our GRN analysis revealed new insights into the roles of less studied families: BACH, TEAD, and EBF.

TEAD factors, known downstream effectors of Hippo signaling, that have been previously linked to NC metabolic activities[61] and endothelial nutrient acquisition for angiogenic growth[62]. BACH factors (*bach2a/b*) are likely essential co-factors in early NC development due to their numerous targets, despite lacking strong RegVelo predictions. EBF factors emerge as another crucial family, influencing various early processes such as NPB, and later governing migration and cranial mesenchymal derivative fates.

We proposed negative regulatory switches governing distinct NC trajectories. Our previous work in enteric nervous system regulation suggested an intricate interplay between different NC programs, as evidenced by Foxd3 expression hindering the development of neuronal NC derivatives[63]. In our study conducted at the bulk level in the premigratory NC, we proposed the existence of distinct NC programs, hinting at the role of the collective inactivation of canonical NC TFs in driving the de-repression of a putative neuronal NC program[13]. We propose viewing this newly discovered trajectory within the framework of negative regulatory switches in the Waddington landscape, governed by major switch factors such as Foxd3.

In contrast to one previously reported mesenchymal bias in the early NC[14], our work distinguishes an endothelial-like migratory program from EMT, elucidating their temporal and regulatory distinctions. EMT activation precedes migration, with post-delamination cranial NC cells adopting an endothelial-like phenotype, characterized by the activity of specific ETS factors (*fli1a, elk3, ets1, etv2*). Whilst most ETS factors have pro-mesenchymal bias, we reported *elf1* as an exception that terminates migration and promotes pigment fates. This is consistent with the DNA-binding anti-cooperativity of ELF1 caused by an electrostatic repulsion mechanism, in contrast to FLI1 and ETS1[64]. Thus, our methodology efficiently discerns functional regulons even when they share similar motifs. The roles of ETS factors in

NC development are intriguing, as these factors have been implicated in the vasculature and tumor development[57,65,66]. Our findings confirm the role of numerous ETS factors in NC migration and mesenchymal differentiation, and, more importantly, unveil their interplay within the GRN.

Finally, we introduced a comprehensive framework of multi-metric methods to explore the functional relatedness between TFs. Unlike conventional definitions focused solely on driving the same lineage commitment, our approach incorporates metrics such as target sharing, motif similarity, regulatory synchronization, DNA sequence-based co-regulation, and putative TF protein co-binding. We demonstrated the effectiveness of this approach in identifying TF convergent roles among pigment lineage drivers (*tfec, mitfa, bhlhe40*) and mesenchymal drivers (ETS, RXR/RAR, NR2F). Our study systematically uncovered new coordinator motifs for NC EMT and dual-fate priming, in addition to the previously described bHLH-homeodomain coordinator motif[15]. ETS factors were found to be involved in early NC development, where a coordinated regulatory scaffold is primed by *tfap2a*, potentially in conjunction with early ETS factors like *gabpa*. Concurrently, the expression of migratory ETS factors is activated by Foxd3. The coordinated action of *tfap2a* and *foxd3*, central to our NC-GRN model, underscores the intricate regulatory choreography essential for cranial NC specification and migration.

In summary, our study provides a comprehensive analysis of cranial NC developmental lineages, linking cellular states with molecular machinery and providing insights into the regulatory architecture governing cell fate decisions.

Methods

Zebrafish husbandry

The handling of animals followed protocols approved by the UK Home Office, in compliance with UK law (Animals [Scientific Procedures] Act 1986) and adhering to the guidelines outlined in the Guide for the Care and Use of Laboratory Animals. All research involving vertebrate animals took place at the Oxford University Biomedical Services facilities. Adult fish were cared for following established procedures[67]. In summary, the adult fish were maintained under the following conditions: a 14-hour light and 10-hour dark cycle (with lights on from 7 am to 9 pm and off from 9 pm to 7 am), housed in a closed recirculating water system at a temperature range of 27-28.5°C, fed 3-4 times daily, and maintained at a density of 5 fish per 1 L of water. Embryos were staged according to previously published methods[68]. Embryo staging was conducted under dissecting stereomicroscopes (Olympus).

Embryo dissociation and cell sorting

Embryos were obtained by crossing two zebrafish lines Gt(foxd3-mCherry)^{ct110R} and Gt(foxd3-Citrine)^{ct110} and incubated until desired stages. Embryos were dissociated with Liberase TM (Roche) for 25 min at 32°C. The cell suspension was filtered by 100-µm mini strainer (PluriSelect) and stained with 1-3 µg/ml DAPI. Live single cells that were either Citrine-positive (Citrine+Cherry-) or double-positive (Citrine+Cherry+) were isolated using fluorescence-activated cell sorting (FACS) on an MA900 instrument (Sony).

10x Multiome sequencing

<u>Single-cell multiomics sequencing</u>

To isolate nuclei, the Low Cell Input Nuclei Isolated protocol (10x Genomics, CG000365) was followed. The sorted cells were washed once with 0.04% BSA and then cells were lysed by adding 45 µl chilled lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1%

Tween-20, 0.1% IGEPAL CA-630, 0.01% Digitonin, 1% BSA, 1 mM DTT and 1 U/µl RNase inhibitor) into 50 µl cell suspension and keeping cells on ice for 20 s. Cell lysis was stopped by adding chilled wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 1% BSA, 1 mM DTT and 1 U/µl RNase inhibitor), centrifuging 500 rcf for 5 min at 4°C and resuspended in chilled diluted nuclei buffer (1x Nuclei Buffer, 1 mM DTT, 1 U/µl RNase inhibitor). The nuclei were counted, and quality checked by using Trypan Blue and C-Chip (NanoEnTek). The single-cell multiome ATAC libraries and gene expression libraries were prepared and sequenced by following the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide (10x Genomics, CG000338).

Single-cell multiome data pre-processing

Fastq data from 10x Multiome were processed using cellranger-arc count v2.0.0 and mapped to a customized GRCz11 Ensembl 105 reference genome, which incorporated sequences of Citrine and mCherry for the specific fish genotypes. The ambient RNA was removed from each 10x sample via SoupX's autoEstCont and adjustCounts functions[69]. In GRCz11 Ensembl 105, ENSDARG00000099849/*ebf3a* on chromosome 14 corresponds to *ebf1a* on ZFIN, and ENSDARG00000100244/*ebf3a.1* on chromosome 12 corresponds to *ebf3a* on ZFIN. To avoid confusion, we adopted the naming convention used by GRCz11 Ensembl 105.

In quality-control steps, cells were filtered based on multiome-RNA and multiome-ATAC data quality. For RNA data, quality parameters were selected as nCount_RNA between1000 to 15000, nFeature_RNA between 500 and 6000, and the percentage of mitochondrial genes (percent.mt) below 20%[70]. Following these quality checks, RNA-based doublets were identified and removed using the R package DoubletFinder[71].

The multiome-ATAC data underwent pre-processing using the R package ArchR[72]. The bespoke ArchR reference genome was created by incorporating

BSgenome.Drerio.USCS.danRer11, the custom-modified GRCz11 Ensembl 105 reference genome, and the DANIO-CODE genome blacklist[73]. The blacklist was adapted from DanRer10 via the UCSC Liftover tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). ArchR arrow files were generated from the cellranger-arc output named atac_fragment.tsv.gz, applying a minimum transcription start site enrichment score (minTSS) of 4 and a threshold of 2500 for the minimum number of mapped ATAC-seq fragments per cell (minFrags). Multiome-ATAC-based doublets were detected using ArchR's addDoubletScores function, using parameters of k = 10 and the LSI "tf-logidf" method, and subsequently filtered out by the filterDoublets function. Cells meeting the criteria for both RNA and ATAC quality control, as well as passing doublet filtering, were selected for downstream analysis.

10x Multiome genotyping

Multiome genotyping amplicon-seq

To overcome the issue of limited cell numbers at the early stage, we multiplexed Citrine-positive and double-positive cells for one 10x run. Although only s8 (75% epiboly - 3 ss) contained mixed genotypes, other single-genotype samples (s1 - s7) underwent the same amplicon-seq procedure to be used as the training sets. From the mixed-genotype and other single-genotype samples, amplicon-seq libraries were prepared from the 10x cDNA samples using NEBNext Ultra II Q5 Master Mix (NEB) and 10μM primers (IDT). This panel contained 7 amplicons, covering Cherry (n = 2), *foxd3* mutated allele (n = 2) and *foxd3* wild-type allele (n = 3). The PCR program was 98°C for 30 s, 25 cycles of (98°C for 10 s, 68°C for 30 s, 72°C for 30 s), 72°C for 2 min and hold at 10°C. The PCR products were purified using AMPure XP beads (Beckman Coulter). The amplicon-seq libraries were sequenced by MiSeq 600-cycle platform (Illumina).

Genotype demultiplexing of 10x Multiome

Our amplicon-seq samples comprised a training set (s1 - s7) with known genotypes and a testing set (s8) for which genotypes were not known. To create the reference genome for genotyping, we assembled the wild-type and mutant *foxd3* allele reference sequences by applying Trinity[74] on bulk RNA-seq data GSE106676[36]. Then the amplicon-seq fastq data were aligned to these reference sequences and quantified using the Alevin function of Salmon v1.8.0[75]. Salmon output was organized into a count matrix for NC cells in training or testing sets respectively, where the rows were amplicons, and the columns were cells. The count matrices were transformed by $\log_2(x+1)$ transformation.

To construct the model to classify genotypes, we fitted a logistic regression model using the training set with the glm function and parameters of family = "binomial" or "poisson", as $g_i \sim \beta_0 + \beta_1 e_{1,i} + \cdots + \beta_n e_{n,i}$, where g_i indicated the genotype (wild type or mutant) of the i-th cell, and $e_{k,i}$ indicated the expression level $log_2(count + 1)$ of k-th amplicon in the i-th cell. The model performance was evaluated using the roc function from pROC package[76]. The binomial model (AUC: 0.781, 0.742-0.820) slightly outperformed the Poisson model (AUC: 0.764, 0.724-0.804). Next the best cutoff of the binomial model was decided by using the ROCR function[77].

Next this logistic classification model was applied to the testing set and classify the cells into the wild-type ("Citrine-positive") or the mutant ("double-positive") groups. For the cells that were not captured in the amplicon-seq data, we assumed that cells with the same genotype tend to share similar expression profiles. Thus, the *k*-nearest-neighbour method (function class::knn) was used to predict missing values based on the Seurat-calculated PCA space (Extended Data Fig. 1e).

Smart-seq3

<u>Smart-seq3 single-cell RNA-Seq</u>

During FACS, cells were sorted into individual wells containing 3-µl lysis buffer dispensed by the MANTIS liquid handler (FORMULATRIX). The lysis buffer consisted of 5% Polyethylene Glycol 8000 (Sigma), 0.1% Triton X-100 (Sigma), 0.5 U/µl RNAse inhibitor (Takara), 0.5 µM OligoT_{30}VN (IDT), 0.5 mM/each dNTPs (Thermo Fisher) and nuclease-free water (Thermo Fisher). Note that these concentrations reflect their concentrations in the 4-µl reverse transcription reaction. The plates were briefly centrifuged, quickly frozen on dry ice and stored at -80°C. The Smart-seq3 libraries were prepared according to the Smart-seq3 protocol [78]. In brief, the plates of sorted cells were transferred on dry ice from -80°C, incubated at 72°C for 10 min, and cooled down to 4°C. Quickly, 1 µl/well reverse transcription master mix containing 25 mM Tris-HCl pH8 (Invitrogen), 30 mM NaCl (Invitrogen), 2.5 mM $MgCl_2$ (Invitrogen), 1 mM GTP (Thermo Fisher), 8 mM DTT (Thermo Fisher), 0.5 U/µl RNAse inhibitor (Takara), 2 µM TSO (IDT), 2 U/µl Maxima H-minus (Life Technologies) was added. The reverse transcription program was 42°C 90 min, 10x (50°C 2 min, 42°C 2 min), 85°C 5 min and 4°C hold. MANTIS added 6 µl/well PCR mix containing 1x KAPA HiFi HotStart buffer, 0.3 mM/each dNTPs, 0.5 mM $MgCl_2$, 0.5 µM forward primer, 0.1 µM reverse primer and 1 U/µl polymerase (Roche). The PCR program was 98°C 3 min, 23x (98°C 20 s, 65°C 30 s, 72°C 4 min), 85°C 5 min and 4°C hold. These cDNA samples were cleaned up by 0.6:1 AMPure XP beads and relocated into 384-well plates using FxP (Beckman Coulter), quantified by Quant-iT PicoGreen (Invitrogen) on CLARIOstar (BMG Labtech), quality checked by TapeStation 4200 and High Sensitivity D5000 assays (Agilent) and normalized to 0.3 ng/µl using MANTIS.

Next, 150 pg/µl cDNA was tagmented by adding 0.5 µl Smart-seq3 in-house Tagmentation buffer (40 mM Tris-HCl pH7.5, 20 mM $MgCl_2$, 20% Dimethylformamide and UltraPure water in 4x buffer), 0.08 µl Amplicon Tagmentation Mix (Illumina) and UltraPure water for a total reaction volume of 2 µl and incubating at 55°C for 10 min. The tagmentation was stopped by adding 0.5 µl of 0.2 % SDS and incubated at room temperature for 5 min. Then 1.5 µl Nextera

index primer pairs and 3 µl PCR master mix (1x Phusion HF buffer, 0.2 mM/each dNTPs, 2 U/µl Phusion HF) were added to each well of the tagmentation product and the PCR program was 72°C 3 min, 98°C 3 min, 12x (98°C 10 s, 55°C 30 s, 72°C 30 s), 72°C 5 min and 4°C hold. The libraries were pooled, cleaned up by 0.6:1 AMPure XP beads and sequenced on NovaSeq for 150-bp paired end reads (Illumina).

Smart-Seq3 data pre-processing

Fastq files for each Smart-seq3 pool libraries were merged into read 1, read 2, index 1, and index 2 four fastq files. Next, we applied zUMIs[79] to the fastq files with the expected barcodes file containing the Nextera v2 index combinations, the GRCz11 Ensembl 105 reference genome built by STAR v2.7.3a[80] and the parameters including barcode binning of 1, no automatic barcode detection, and the minimum number of reads per cell of 0.

Spatial transcriptomic analysis

MERSCOPE sample preparation and experiments

Zebrafish embryos were collected, dechorionated, and deyolked before being embedded in optimal cutting temperature (OCT) compound (VWR). OCT blocks were snap frozen on dry ice/ethanol bath and sectioned at 10-µm thickness onto MERSCOPE slides using a microtome-cryostat (Leica). Sections were hot fixed with 4% PFA at 37°C for 30 min, following the MERSCOPE Fresh and Fixed Frozen Tissue Sample Preparation User Guide, and imaged with the MERSCOPE Instrument (Vizgen).

MERSCOPE gene panel design

To identify cell types and NC cell states using MERSCOPE, we designed a gene panel comprising both global markers for non-NC cells and specific marker genes for NC cell states. We used SPAPROS, a probe selection algorithm, on multiome-RNA data, Smart-seq3 data and the organogenesis dataset of 10-, 14-, and 18-hpf embryos[17,81]. The reads per kilobase million

(RPKM) and transcripts per million (TPM) count matrices were computed for stage-specific pseudo-bulk samples to filter genes based on expression levels. After submitting the initial 304-gene panel to Vizgen for a designability check, redundant markers for the same non-NC cell type were identified and eliminated, resulting in a refined set of 300 genes. These selected gene sets underwent validation using SPAPROS metrics and were assessed for their ability to delineate clusters in low-dimensional representation.

Assigning cell type identities

Spatial transcriptomic data were visualized using MERSCOPE Visualizer software (Vizgen) for quality assessment. Suboptimal segmentation results were addressed by re-segmenting based on polyT and DAPI using VPT and Cellpose2 human-in-the-loop methods[82]. Re-segmented data were exported to h5ad format using Visualizer and preprocessed with Squidpy[83], filtering cells with minimal counts of 10 and minimal volume of 50. Clustering was performed using the Leiden algorithm with 20 principal components and resolution of 1.5. Major cell types were identified by mapping organogenesis data onto spatial transcriptomic data with Tangram[26]. NC-dominating clusters were isolated and re-clustered with a resolution of 1.5 and 5-nn graph. The Tangram function map_cells_to_space was employed with uniform density_prior in clusters mode. To distinguish NC-derived from non-NC-derived head mesenchymal cells, we used NC markers, namely *grem2b, foxd3, sox10, tfap2a, zeb2a*, and *crestin*. Cells expressing at least one NC marker were categorized as NC derivatives, while those lacking NC marker expression were assigned to non-NC-derived head mesenchyme. Cell locations were visualized using Squidpy's spatial_scatter function.

ChIP-seq and data analysis

MicroChIP-seq

To validate the SCENIC+-predicted CREs, the related histone modifications were profiled in wild-type and *foxd3*-mutated cells from 12-16 ss embryos. Fish lines were crossed, and embryos were incubated in the same way as the 10x Multiome samples. The manufacturer's protocols of True MicroChIP-seq kit (Diagenode) and Chromatin EasyShear Kit - High SDS (Diagenode) and the Fish'n ChIPs (Chromatin Immunoprecipitation in the Zebrafish Embryo) protocol[84] were followed. In brief, dissociated cells were fixed using 1% paraformaldehyde for 8 min at room temperature and stopped by glycine prior to FACS. Citrine-positive (wild-type) and double-positive (*foxd3*-mutated) post-fixation cells from 12-16 ss embryos were isolated by FACS. The chromatin was sheared by 3 cycles of 30-s ON/30-s OFF on the Bioruptor Pico (Diagenode). The antibodies used were H3K27me3 (Millipore, cat no. 07-449), H3K27ac (Abcam, cat no. ab4729), H3K4me1 (Diagenode, cat no. pAb037050 / C15410037), and H3K4me3 (Diagenode, cat no. pAb003050 / C15410003). Controls of 'IgG-cit', 'IgG-dp', or 'input-cit' were performed.

ChIP-seq data analysis

Raw reads were quality-controlled using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Following quality control, the reads were trimmed using Trim Galore (https://github.com/FelixKrueger/TrimGalore), aligned with Bowtie2[85], and the resulting BAM files were indexed using Samtools[86]. For peak calling and bigwig generation, each experimental condition was compared to a corresponding control, which could be 'IgG-cit', 'IgG-dp', or 'input-cit'. Peaks were identified from these sorted BAM files using MACS2[87], and then annotated with HOMER[88]. Finally, Deeptools[89] was employed to ensure data quality, normalize the data, and provide visual representations of the results. The samples that were outliers in the Spearman correlation analysis were excluded from downstream analysis.

RNA in situ hybridization and immunocytochemistry

Hybridization chain reaction

Fluorescently labeled hairpins, DNA probe sets, amplification buffers, hybridization buffers, and wash buffers, whether procured commercially or produced in-house, were used for the multiplexed targeted visualization of mRNA in situ, employing the HCR kit (version 3, Molecular Instruments)[90]. Zebrafish embryos, harvested at the desired developmental stages, underwent dechorionation followed by fixation in 4% paraformaldehyde at 4°C for 24 hours. The HCR procedure adhered to the 'In situ HCR v3.0 protocol for whole-mount zebrafish larvae' (Molecular Instruments). Post HCR, the embryos were embedded in 1% low melting-point agarose and imaged via the 780 upright, 780 LSM inverted or 980 LSM inverted confocal microscope (Zeiss). Image analysis was performed using Fiji and Zeiss Zen.

Whole-mount immunofluorescent staining

Embryos at desired stages were fixed with 4% paraformaldehyde (Methanol-free, Thermo Fisher) for 1 hour at room temperature, followed by rinsing in PBT (2% DMSO, 0.5% Triton-X in PBS). Blocking was carried out in a solution of 10% goat or donkey serum in PBT for one hour at RT. The primary antibodies were diluted in blocking solution: goat anti-tfap2a antibody 1:50 (LS-C87212, LSBio), rabbit anti-Elavl3/4 antibody 1:200 (GTX128365, GeneTex), and chicken anti-GFP 1:200 (ab13970, Abcam), and applied to the embryos for overnight incubation at 4°C. After thorough washes with PBT, the secondary antibodies, donkey anti-rabbit 647 nm (A31573, Thermo Fisher), donkey anti-goat 568 nm (A11057, Thermo Fisher), donkey anti-chicken 488 nm (A78948, Thermo Fisher), and Hoechst 33258 (ab228550, Abcam, 1:1000 dilution) were prepared in PBT and added to the embryos for a 2-hour incubation at room temperature in the dark. Post extensive washing with PBT, embryos were embedded for confocal imaging.

Single-cell multiomic computational analysis

Clustering analyses and cell type annotations

Our first step was the pre-processing and clustering analysis of the 10x Multiome-RNA dataset using Seurat v4.2 package[70]. RNA data normalization was achieved with the "LogNormalize" method with a scale factor set at 10,000. We then selected the top 2000 highly variable features using the "vst" method. This was followed by data matrix centering, scaling, and principal component (PC) computation. For clarity, the first 25 PCs were chosen based on the elbow plot method for further analysis. Clusters were delineated using the functions FindNeighbors and FindClusters set at a resolution of 0.6. The marker genes were identified using the specific parameters (only.pos = TRUE, min.pct = 0.25, and logfc.threshold = 0.25).

For the multiome-ATAC dataset, the peak-cell count matrix was refined through dimensionality reduction using the ArchR Iterative Latent Semantic Indexing (LSI) method, followed by visualization with Uniform Manifold Approximation and Projection (UMAP). Marker peaks were identified using the ArchR getMarkerFeatures function with the Wilcoxon test method. The motif enrichment analysis of these marker peaks was conducted using the ArchR peakAnnoEnrichment function, comparing against the DANIO-CODE motifs or tailor-made motif collections.

To annotate the cell clusters, we transferred the labels from a zebrafish organogenesis scRNA-Seq dataset[17]. This involved normalizing, merging the top 2000 variable feature lists, scaling data, running PCA and computing anchors with the first 50 PCs. The labels of cluster names and tissue names were transferred by Seurat's TransferData from the organogenesis dataset to our multiome-RNA dataset. The non-NC clusters were annotated by the dominating labels. Three clusters with low numbers of features detected were annotated as "mutant low-feature", "neural crest low-feature" and unclassified. The neural crest clusters were identified based on

the label transferring results and manually annotated as "neural crest", "neural crest mutant", "neural crest early", "neural crest pigment", and "neural crest low-feature".

To identify potential intermediate cells between the NC and neural plate populations, we subset the multiome-RNA data for both NC and neural clusters and performed the clustering analysis. This analysis did not reveal any intermediate population, but *jhy*-expressing apoptotic clusters were identified and subsequently removed.

To retain only high-quality NC cells for further analysis, we excluded the cells expressing genes from the panel (*jhy*, *cyt1*, *krt5*, and cyt11) commonly associated with apoptosis and epidermal populations. The NC multiome-RNA dataset underwent the same pre-processing steps as the complete multiome-RNA dataset. We employed the top 15 PCs for clustering and visualization. Louvain clustering was implemented using a resolution of 1. Clusters were subsequently annotated, drawing on stages, genotypes, and cell cycle states as determined by the Seurat CellCycleScoring function. Some gene expression figures were made with the R package scCustomize (https://samuel-marsh.github.io/scCustomize/index.html).

Integration analysis of NC data

To compare the cell states across three datasets (zebrafish organogenesis[17], our multiome data, and Smart-seq3 data), Seurat canonical correlation analysis (CCA) was used to identify anchors from the first 50 PCs for integration. Because cell populations between multiome-RNA and Smart-seq3 datasets are largely overlapped, Seurat-CCA in this scenario was unlikely affected by the overcorrection error.

To quantify the inter-group similarity, we applied the getIDEr function from CIDER R package[91]. The organogenesis cell labels were simplified by merging clusters containing the differentiating neurons, epidermal, pharyngeal arch, diencephalon, pronephric duct, mesoderm, heart, muscle, and lateral line, respectively. The apoptotic clusters were from the organogenesis

dataset. The curated annotations were used as initial cluster and the inter-project batch effects were removed using the limma trend method[92]. The inter-group similarity matrix was visualized using R package pheatmap (https://cran.r-project.org/web/packages/pheatmap/index.html).

RNA velocity analysis

For 10x Multiome data, bam files generated by cellranger-arc were processed by the velocyto run10x command to generate loom files[93]. For Smart-seq3 data, the bam files for velocyto were created by zUMIs and processed by "veloctyo run --vv --umi-extension Gene -d 1 --without-umi" to use both internal reads and UMI reads.

To perform velocity analysis for wild-type NC development, the NC multiome data were divided into the wild-type and the mutant datasets. The wild-type NC data were pre-processed by the scVelo preprocessing functions filter_genes, normalize_per_cell, filter_genes_dispersion, and log1p[38]. Neighbors were computed by scVelo preprocessing functions neighbors and moments functions with n_pcs of 30 and n_neighbors of 10. RNA velocity was then estimated by the velocity and velocity_graph functions. We computed the PAGA velocity graph using scVelo with minimum_spanning_tree = False and threshold_root_end_prior = 0.5. The mutant NC data were pre-processed in the same way. To visualize the PAGA results, we used the threshold of 0.04.

Multiomic velocity analysis

To use both gene expression and chromatin accessibility to estimate velocity, we used MultiVelo[21]. The multiome dataset was subset to retain only the most anterior lineage. Cells with counts between 1000 and 20000 were kept for MultiVelo analysis. The RNA data were processed as recommended by scVelo functions. As for the ATAC data, the peak-count matrix called by SCENIC+ pycisTopic was used. The feature linkage bedpe file was generated by

Signac[94]. The peaks were annotated by Homer v20201202 annotatePeaks.pl and parsed by an in-house script. The ATAC data, the peak annotation file and the linkage file were used as input to the MultiVelo aggregate_peaks_10x function. Cells were filtered for ATAC counts between 1000 and 60000. The aggregated peaks were normalized by TF-IDF. RNA data were normalized again after filtering. To smooth the ATAC data, Seurat WNN of 30 *k*-nearest neighbors was used.

Both RNA and ATAC datasets were then used to fit the multiomic dynamical model by function recover_dynamics_chrom with max_iter=5 and init_mode="invert". The latent time was computed using MultiVelo velocity_graph and latent_time functions. The pie summary plot was visualized by MultiVelo pie_summary function. Thus, 30 nearest neighbors were used.

Lineage inference and probability calculation

To compute the fate probability, MultiVelo velocity vectors were used as input of Cellrank2[49,50] to compute the transition matrix with 'deterministic' model. To refine the clusters, we applied leiden clustering using resolution of 1 and manually curated and merged the clusters. Next the estimator was initialized and fit using the Schur decomposition and GPCCA algorithms to estimate macrostates by 8 states. The terminal states were manually set to 'Pigment_sox6_high', 'mNC_arch1', 'mNC_head_mesenchymal', 'Pigment_gch2_high', and the initial state was predicted to be "NPB nohox 1". Next the fate probability matrix was computed, and the probabilities were extacted from the "obsm['terminal_states_memberships']" entry. For visualization, we used cellrank.pl.heatmap and cellrank.pl.gene_trends.

De novo motif discovery for the *foxd3* motif

Because the *foxd3* motif in the Danio Code motif database failed cisTarget analysis, we reasoned that the ortholog-based motif was not applicable in the NC development scenario. Thus, we sought to *de novo* construct the *foxd3* motif from the biotin-ChIP data [95]. The *foxd3* biotin-ChIP raw data (SRR6268024/5/6/7 as ChIP sample and SRR6268028/29/30/31 as input control) were downloaded from the Sequence Read Archive (SRA). Raw data were preprocessed by Trim Galore v0.6.5 and aligned by Bowtie2 v2.4.2 to GRCz11. The duplicates were removed by Samblaster v0.1.24[96]. The narrow peaks were called from the biotin-ChIP sample and the input control using MACS2 and the regions in the blacklist were excluded. We next used both GLAM2 from MEME v5.4.1 and MEMEChIP to predict motifs from the MACS2-called peaks[97,98]. Based on the hypothesis that the *foxd3* motif should be enriched in the migratory NC populations, we identified the GLAM2 motif1, which satisfied the enrichment and resembled the ortholog motif "5`-TGTTTGTTT-3`". This GLAM2-predicted *foxd3* motif was used for the SCENIC+ analysis.

GRN analysis by SCENIC+

The comprehensive GRN underlying both normal and *foxd3*-mutant development was constructed using SCENIC+[2]. SCENIC+ could integrate both snRNA-seq data and scATAC-seq data to infer enhancer-driven regulons.

Our approach involved preparing scRNA-seq data, the scATAC-seq pycisTopic object, and the pycisTarget motif enrichment dictionary to create the SCENIC+ object. Our dataset contains 10,604 wild-type and 5,946 mutant NC cells. For the multiome-RNA dataset, preprocessing involved adjusting the *foxd3* read counts to 10% in the *foxd3*-mutated cells to emulate the loss of function. The multiome-ATAC dataset underwent preprocessing with pycisTopic, wherein CellRanger-ARC fragment output files were processed through a series of pycisTopic functions (export_pseudobulk, peak_calling, get_consensus_peaks, compute_qc_stats, and

create_cistopic_object_from_fragments) to form the pycisTopic object. To determine the optimal number of topics in the topic modeling, the parameter sweeping was done from 5 to 100 in increments of 5, with the 95-topic model ultimately being chosen. Post topic modeling, differentially accessible regions per cell type, serving as enhancer candidates, were identified from the pycisTopic object.

Subsequently, the cisTarget database was established for pycisTarget analysis. Because the DANIO-CODE study focused on the conservation of DNA binding regions over general ortholog information, the motif database incorporating DANIO-CODE motifs along with the de novo *foxd3* motif was selected for pycisTarget analysis. The feather database was constructed using scripts from the create_cisTarget_databases repository. Motif enrichment analysis on the enhancer candidates predicted by pycisTopic was conducted using the SCENIC+ wrapper function run_pycistarget.

The SCENIC+ object was created by merging three key components: RNA data, the pycisTopic object, and the motif enrichment results. To reconstruct the GRN, the step-by-step SCENIC+ protocol was followed, including the steps of cistrome generation, region-to-gene and TF-to-gene relationship calculations, and actual GRN construction. To expedite the region-to-gene and TF-to-gene relationship calculations, a stochastic gradient boosting model (sGBM) was employed, configured with a learning rate of 0.01, 5,000 estimators, max features set at 0.1, and subsample rates of 0.5 and 0.2 for region-to-gene and TF-to-gene calculations respectively. The observation subsampling fraction was reduced to 0.2, acknowledging the relatively large number of observations in our dataset.

Following GRN construction, the activity of gene-based and region-based eRegulons (AUCell scores) was quantified using the function score_eRegulons, and their regulon specificity scores (RSSs) were derived using the function regulon_specificity_scores. The network was exported

by SCENIC+ functions (create_nx_graph and export_to_cytoscape) and visualized by Cytospace (https://cytoscape.org). To check the overlap between regulons, the Jaccard indexes were calculated using SCENIC+ jaccard_heatmap function as $J(R_1, R_2) = |R_1 \cap R_2|/|R_1 \cup R_2|$, where R_1 and R_2 represent the sets of elements in regulon 1 and regulon 2, respectively.

Tn5-bias-corrected de novo motif discovery

To aid de novo motif discovery in our multiome-ATAC data, we used ChromBPNet to mitigate Tn5 bias. ChromBPNet used convolutional neural networks to predict base-resolution cluster-specific chromatin accessibility profiles, enabling precise in vivo motif identification [51]. Key cell states in cranial NC trajectories were represented by selecting clusters including merged_dNC_nohox, merged_NPB_nohox, mNC_nohox, mNC_head_mesenchymal, mNC_arch1, Mutant_nohox_early, Mutant_nohox_12_22ss, and Pigment_sox6_high. To secure sufficient depth for dependable Chrombpnet model training, these clusters comprised cell counts exceeding 950, except for Mutant_nohox_12_22ss (N = 663) and Mutant_nohox_early (N = 828).

Pseudo-bulk bam files were generated from cellranger-arc bam and cluster annotations using pyflow-scATACseq (https://github.com/crazyhottommy/pyflow-ATACseq). MACS2 was employed for peak calling with a significance threshold of q = 0.01 and an effective genome size of 1.4e9, and the identified peaks were systematically filtered using the blacklist. Chromosomes were divided using the "chrombpnet prep splits" command with the Fold 0 parameter, defining training chromosomes (chr1, chr3, and chr6) and validation chromosomes (chr8 and chr20). Non-peak background regions were created by "chrombpnet prep nonpeaks". Cluster-specific ChromBPNet Tn5 bias models were trained using 'chrombpnet bias pipeline -d "ATAC" -b 0.5', followed by training cluster-specific bias-factorized models through

'chrombpnet pipeline -d "ATAC"' with cluster-specific Tn5 bias models. Contribution scores were predicted by "chrombpnet contribs_bw" with the bias-factorized models.

Subsequent to contribution score prediction, de novo motif discovery was conducted using the TFmodisco-lite command 'modisco motifs' with the input of predicted contribution profile scores[99]. All newly discovered motifs were concatenated and clustered using the Gimme command 'gimme cluster -t 0.999 -N 4'[100]. For motif annotation, the average motifs were meticulously compared against the DANIO-CODE motifs via TOMTOM 'tomtom -no-ssc -oc . -verbosity 4 -text -min-overlap 5 -dist pearson -evalue -thresh 10.0'[101], followed by manual curation to eliminate artificial motifs and reduce redundancy. Compared to the publicly available motif databases, the de novo discovered motifs might better represent in vivo scenarios. However, the cluster-specific discreet pseudo-bulk samples used by motif prediction can undermine the variability along continuous trajectories. Thus, de novo motif analysis and SCENIC+ GRN reconstruction were applied to be complementary to each other.

To validate the occupancy of de novo discovered motifs by TFs in vivo, we used marginal footprinting analysis with the ChromBPNet function footprints. This method computed the bias-corrected footprint of each motif per cellular cluster using Fold 0.

TF-TF interaction prediction

To identify TFs that co-bind with Tfap2a at the motif Combi_593, we used AlphaPullDown v1.0.4[102], incorporating AlphaFold-Multimer[103], for predicting protein-protein interactions. Our focus was on TFs capable of binding to the ETS motif in Combi_593 and expressed in early NC states. Initially, TFs were selected based on their expression levels in early NC states (NPB_nohox, dNC_hoxa2b, and dNC_nohox) from Smart-seq3 data, using a cutoff of average scaled expression greater than 0.5 across all three states. TFs including Ets1, Etv4, Etv5b, Gabpa, Tfec, Usf1, and Usf1l were chosen for further analysis. Amino acid sequences for these

TFs were retrieved from UniProt, and individual features were generated using AlphaPullDown's function create_individual_features.py. AlphaPullDown was applied with the parameters "run_multimer_jobs.py --mode=pulldown --num_cycle=3 --num_predictions_per_model=1". The potential for interaction with Tfap2a was assessed by examining metrics such as inter-chain predicted aligned errors (interPAE, indicating how confidently the relative positions of residue pairs were predicted), ipTM (average quality for all interfaces), pDockQ/mpDockQ, and PI-score, in the AlphaPullDown output.

Functional and temporally dynamic analysis of regulome

Functional enrichment analysis

To systematically analyze regulon functionality, SCENIC+-predicted regulons were initially grouped into broad functional clusters through agglomerative clustering, utilizing the regulon-based AUCell matrix (Extended Data Fig. 2h). For an in-depth interpretation of these regulon clusters, FishEnrichR[104] was employed to conduct gene ontology enrichment analysis on the target genes within each regulon cluster. This approach helped identify regulon clusters that played potential roles in early NC, pigment lineage, NC migration, hindbrain NC, cardiac NC, posterior NC, and chromatin remodeling.

Centrality analysis

Within each cluster, the TFs of the regulons were prioritized based on out-degree centrality of the TF-TF sub-network, which was calculated by applying networkx.out_degree_centrality function on the TF-only skeleton of the SCENIC+-predicted GRN[105]. High out-degree centrality in the TF-TF network indicated higher positions of TFs in the GRN hierarchy (Extended Data Fig. 2i-j); however, this method did not incorporate temporal information.

Motif similarity analysis

To assess motif similarity, we first generated the Markov background model (order 3) using MEME (version 5.4.1) with the command 'fasta-get-markov -m 3'. Next, we converted the position weight matrix (PWM) format to MEME format using the chen2meme function. Motif similarity was then computed using TOMTOM with the parameters '-dist kullback -motif-pseudo 0.1 -text -min-overlap 1'[101].

Temporally dynamic analysis of GRNs by SyncReg

To capture the dynamic nature of NC development, we designed an integrative analysis combining temporal metrics and regulon activities, termed "regulatory synchronization (SyncReg)". This approach, akin to temporal gene trend analysis, offered the advantage of minimizing the influence of dropout in regulon-level assessments.

Initially, we evaluated three algorithms, scVelo[38], MultiVelo[21], and RegVelo, to determine the best performer in predicting latent time for non-hox trajectories. This comparison involved two steps. Firstly, we correlated the predicted latent time with the ordinal variable representing development stages ("real-world" time) using Spearman correlation coefficients. Next, we compared the predicted latent time between pairs of methods to discern differences in their predictions. Lastly, we selected MultiVelo over scVelo due to its higher Spearman correlation coefficient and better representation of dNC-mNC transition (Extended Data Fig. 4a-c). Although MultiVelo and RegVelo showed similar performance in latent time prediction, we opted for MultiVelo to capture chromatin changes over time.

We integrated MultiVelo-predicted latent time (x-axis) with regulon activities assessed by AUCell scores (y-axis) into a two-dimensional framework. Feature extraction from this dynamic framework was conducted using a VGG16 model[106], followed by cosine similarity calculation between each pair of regulons. Subsequently, hierarchical clustering with the complete-linkage method was applied to group the regulons based on the resulting similarity

matrix. Dendextend visualization facilitated the comparison of different hierarchical clustering outcomes, providing a refined understanding of regulon interactions and temporal behavior.

Temporal activity of individual regulons was depicted using the R package ggplot2, with trend estimation achieved through cubic spline basis in a generalized additive model (GAM). SCENIC+ AUCell methods quantified the regulon activities, and cluster-specific regulons were identified based on the top twenty regulons for each cluster, determined by gene-based RSSs.

In vivo Perturb-seq

Design of chemically modified sgRNA

The sgRNA designs were sourced from various origins, including CHOPCHOP in-house design[108], the IDT sgRNA design tool, and existing literature. Specifically, the sgRNAs for *fli1a*, *tfec*, *rxraa*-sgRNA6, *rarga*-sgRNA3, and *erfl3*-sgRNA1 were derived from the IDT sgRNA design tool. The *nr2f5*-sgRNA4 corresponded to ZDB-CRISPR-210310-12[109], while the *tyr* sgRNAs were obtained from the MIC-Drop study[22]. The selection criteria for sgRNAs, adapted from the same study[22], included: a) 2-4 sgRNAs were employed per gene to maximize overall efficiency; b) spacer sequences were fixed at 20 nucleotides (or 19 if 20-nt spacers were unavailable); c) spacer sequences were required to have no fewer than three mismatches in off-target alignments; d) in cases where spacer sequences had off-target alignments with three mismatches, these mismatches were mandated to occur in the seed region for designs executed using CHOPCHOP. Highly expressed onhologs[110] were also considered for multi-gene knockouts (Extended Data Fig. 6e). Details of the spacer sequences for sgRNAs are cataloged in Supplementary Table 7.

As the ability of being template switched is the key in sgRNA capture rates, three different designs of sgRNA were tested for their template switching efficiency (Extended Data Fig. 4c,

d). There were not obvious differences. The Alt-R sgRNA design was chosen due to the cost efficiency and the durability in cells.

Ribonucleoprotein microinjection and generation of F_0 Crispants

The ribonucleoprotein (RNP) mixture was prepared by incubating the combination of sgRNAs, EnGen Cas9 (NEB), and nuclease-free water at room temperature for 10 min, followed by cooling on ice. Phenol red was added into the RNP mixture prior to injection. The final injection mixture contained 250 ng/µl sgRNAs (approximately 7.8 µM), 3.36 µM EnGen Cas9, and 0.05% (w/v) Phenol red. Embryos obtained by crossing Gt(*foxd3*-mCherry)^{ct110R} or Gt(*foxd3*-Citrine)^{ct110} and wild-type fish were injected with 1-2 nl of the injection mixture and incubated until reaching 20-21 ss.

Validation of sgRNA knockout efficiency

To assess the effectiveness of RNP injection, embryos at the 1-cell stage were injected with RNP containing three mCherry sgRNAs. The presence of fluorescence in the embryos was then examined one day post-injection using the MVX10 stereomicroscope (Olympus) (Extended Data Fig. 4e). A complete elimination of mCherry fluorescence indicated successful knockout.

Furthermore, the efficiency of knockout was also evaluated using RNP injections targeting only *tyr* and the *tfec;mitfa;tfeb* gene panel (Extended Data Fig. 4f-g). Embryos at both day 2 and day 5 post-injection were embedded in low-melting point agarose and subjected to imaging with the MVX10 stereomicroscope (Olympus) to confirm the efficacy of the gene disruption.

Direct capture Perturb-seq

For the Perturb-seq process, embryo dissociation was performed similarly to the 10x Multiome. Citrine-positive live cells were sorted into 100 µl of 0.1% UltraPure BSA (Invitrogen) in PBS, and then centrifuged at 600× g for 5 min at 4°C. Care was taken to remove the supernatant without disturbing the cell pellet, leaving approximately 40 µl of supernatant. The cell pellet

was reconstituted in the remaining supernatant. Cell count and quality were assessed using Trypan Blue and C-Chips (NanoEnTek). The protocol CG000510 Chromium NextGEM Single Cell 5' v2 CRISPR User Guide (Rev B, 10x Genomics) was adhered to for the preparation of single-cell gene expression and CRISPR libraries. Sequencing of the libraries was carried out on NextSeq and NovaSeq platforms (Illumina).

Perturb-seq data pre-processing

The raw datasets underwent preprocessing with the 'cellranger count' function from CellRanger (version 7.1.0), using the SC5P-R2 chemistry. Sequencing reads were aligned to the custom reference genome, which integrated the GRCz11 Ensembl 105 reference with additional entries for the mCherry and Citrine genes. Cells were demultiplexed into distinct knockout panels based on sgRNA counts over one. Due to the specific arrangement of panels and 10x sequencing runs in our study, each 10x run was devoid of nested knockout panels. Consequently, cells harboring sgRNAs from two disparate knockout panels were identified as doublets. Ambient RNA contamination was regressed on a sample-by-sample basis using SoupX[69]. RNA-based doublet cells were identified using DoubletFinder[71], and both doublets and sgRNA-negative cells were systematically identified and excluded.

Perturb-seq data were clustered using the Seurat package, with cell state annotations assisted through integration with multiome-RNA data via reciprocal PCA (RPCA) (Extended Data Fig. 4i, j). The Seurat-RPCA was selected over Seurat-CCA here due to the largely missing early populations in the Perturb-seq data and the large numbers of cells. The NC (neural crest) subset of the Perturb-seq data was then isolated and subjected to further clustering. Cell type annotations were established based on the RPCA integration outcomes between the Perturb-seq NC and multiome-RNA NC datasets. Clusters with low feature counts were omitted from

subsequent analyses. To minimize technical discrepancies, all knockout panels were sequenced to achieve relatively uniform depths (Extended Data Fig. 4h).

Quantification of knockout effects

The filtered Perturb-seq NC data were visualized using the PHATE method[111]. To measure the extent of gene disruption, the MELD method[48] was used to quantify the differences in abundance between the control (Cherry no-targeting) and each knockout condition/panel. Given that each knockout condition encompassed multiple embryos ($n \geq 12$), the Perturb-seq data essentially reflected an average effect, thereby minimizing inter-animal variance. While the control and some conditions involved cells from several 10x sequencing runs, aiding in mitigating batch effects, technical duplicates were still produced through re-sampling to compute a mean MELD likelihood value.

To gauge the short-term impacts of the knockout, the latent time for the affected cells was determined. Two integration approaches, RPCA[112] and scVI[113], were employed to generate shared nearest neighbor graphs. The average latent time of neighboring cells was computed for each cell, serving as the imputed latent time. The latent times estimated via RPCA and scVI demonstrated a significant correlation (Extended Data Fig. 4k), and the RPCA-estimated latent time was selected for further downstream analysis and visual representation. For each lineage, the median of estimated latent time was used to dichotomize cells in this lineage-specific clusters into early and late stages. Finally, the depletion-enrichment likelihood (range: 0-1) was computed by MELD, and the average likelihood of the late-stage cells for each lineage represented the lineage-specific short-term effect for each knockout panel.

Data availability

The 10x Multiome, Smart-seq3, and Merscope data are available to view on the UCSC Cell browser (https://zhiyhu.github.io/neural_crest_cellbrowser/). The interactive zebrafish NC-

GRN viewer with options to select regulons and view their downstream targets is available at https://zhiyhu.github.io/NC-GRN_shiny/.

Code availability

All analysis code is available at the GitHub repository (https://github.com/zhiyhu/neural-crest-scmultiomics).

Acknowledgments

We thank the MRC WIMM Flow Cytometry Facility, MRC WIMM Advanced Single Cell OMICS Facility, MRC WIMM Centre for Computational Biology, MRC WIMM Sequencing Facility, and Wolfson Imaging Centre for their help and technical support in this study. We also thank the HPC core facility of the Medical Research Institute at Wuhan University for their technical support. We thank Dr Vanessa Chong-Morrison, Dr Oana Pelea, and Dr Martyna Lukoseviciute for their help in MERSCOPE experiments and sgRNA design and we thank Dr Kate Attfield for providing the platform for some 10x runs. We thank A. Andersen for invaluable advice on the manuscript. Scientific illustrations (Figures 1a, 2a, 3a-b, 4h, 7a, and 7j) were created by Uta Mackensen. This research was supported by the Wellcome Trust ref 215615/Z/19/Z (to T.S.-S) and Stowers Institute for Medical Research Institutional Support (to T.S.-S).

Author contributions

Conceptualization: T.S.-S, Z.H.; Methodology: Z.H., T.S.-S; Investigation: Z.H., S.M., W.W., J.M.S.-P., F.T., T.S.-S; Visualization: Z.H., T.S.-S; Funding acquisition: T.S.-S; Project administration: T.S.-S; Supervision: T.S.-S; Writing – original draft: Z.H., W.W., T.S.-S; Writing – review & editing: Z.H., T.S.-S.

Declaration of interests

The authors declare no competing interests.

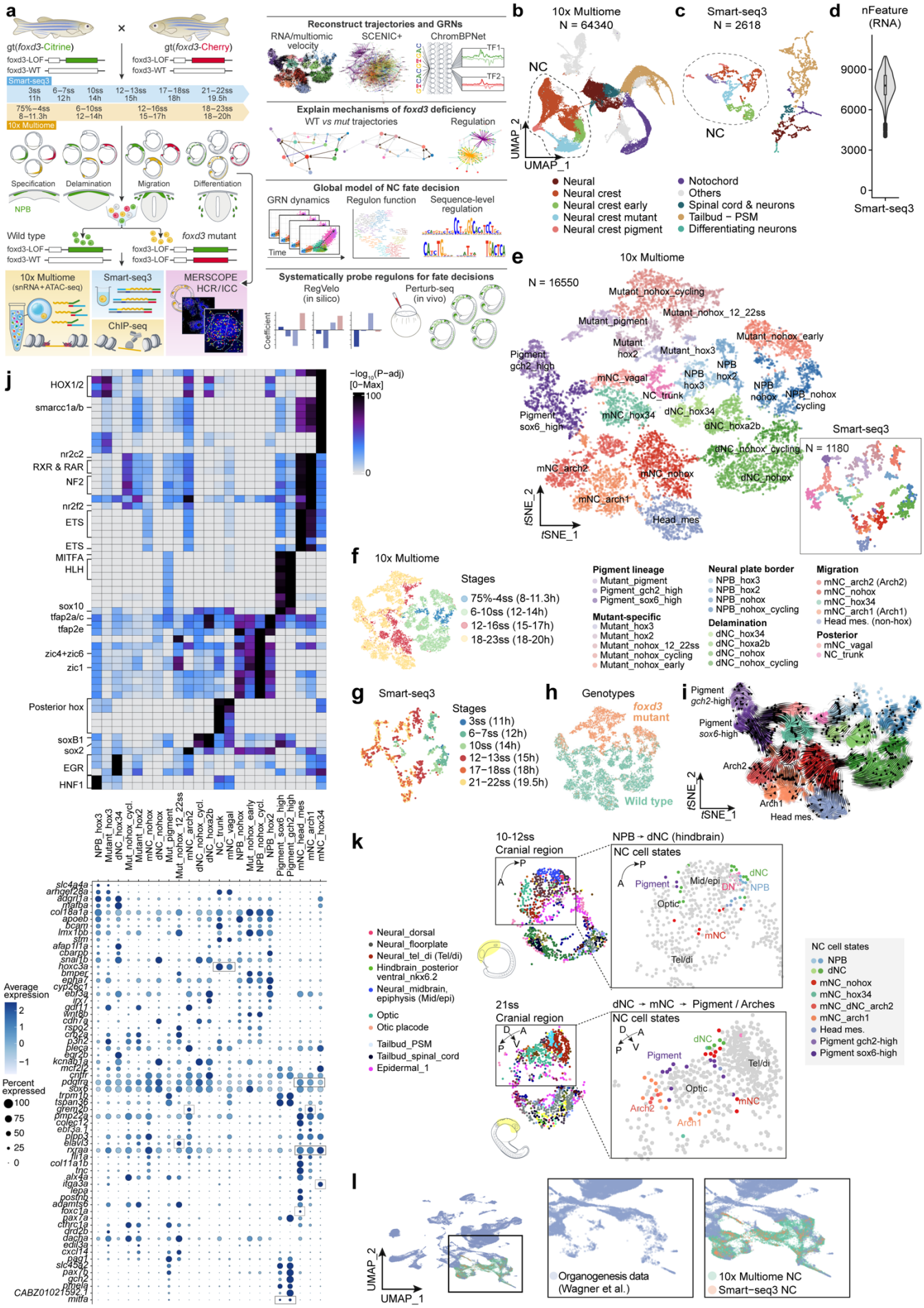

a
gt(foxd3-Citrine)
gt(foxd3-Cherry)
foxd3-LOF
foxd3-WT
Smart-seq3
3ss 11h
6–7ss 12h
10ss 14h
12–13ss 15h
17–18ss 18h
21–22ss 19.5h
75%–4ss 8–11.3h
6–10ss 12–14h
12–16ss 15–17h
18–23ss 18–20h
10x Multiome
Specification
Delamination
Migration
Differentiation
NPB
Wild type
foxd3 mutant
10x Multiome (snRNA+ATAC-seq)
Smart-seq3
MERSCOPE HCR/ICC
ChIP-seq
Reconstruct trajectories and GRNs
RNA/multiomic velocity
SCENIC+
ChromBPNet
Explain mechanisms of foxd3 deficiency
WT vs mut trajectories
Regulation
Global model of NC fate decision
GRN dynamics
Regulon function
Sequence-level regulation
Time
Systematically probe regulons for fate decisions
RegVelo (in silico)
Perturb-seq (in vivo)
Coefficient
b
10x Multiome
N = 64340
NC
UMAP_2
UMAP_1
Neural
Neural crest
Neural crest early
Neural crest mutant
Neural crest pigment
Notochord
Others
Spinal cord & neurons
Tailbud – PSM
Differentiating neurons
c
Smart-seq3
N = 2618
NC
d
nFeature (RNA)
9000
6000
3000
0
Smart-seq3
e
10x Multiome
N = 16550
Mutant_nohox_cycling
Mutant_nohox_12_22ss
Mutant_pigment
Mutant_nohox_early
Mutant hox2
Mutant_hox3
Pigment gch2_high
NPB hox3
NPB hox2
mNC_vagal
NPB nohox
NPB_nohox cycling
NC_trunk
Pigment sox6_high
mNC_hox34
dNC_hox34
dNC_hoxa2b
dNC_nohox_cycling
Smart-seq3
N = 1180
mNC_arch2
mNC_nohox
dNC_nohox
mNC_arch1
Head_mes
tSNE_2
tSNE_1
f
10x Multiome
Stages
75%-4ss (8-11.3h)
6-10ss (12-14h)
12-16ss (15-17h)
18-23ss (18-20h)
Pigment lineage
Mutant_pigment
Pigment_gch2_high
Pigment_sox6_high
Mutant-specific
Mutant_hox3
Mutant_hox2
Mutant_nohox_12_22ss
Mutant_nohox_cycling
Mutant_nohox_early
Neural plate border
NPB_hox3
NPB_hox2
NPB_nohox
NPB_nohox_cycling
Delamination
dNC_hox34
dNC_hoxa2b
dNC_nohox
dNC_nohox_cycling
Migration
mNC_arch2 (Arch2)
mNC_nohox
mNC_hox34
mNC_arch1 (Arch1)
Head mes. (non-hox)
Posterior
mNC_vagal
NC_trunk
g
Smart-seq3
Stages
3ss (11h)
6–7ss (12h)
10ss (14h)
12–13ss (15h)
17–18ss (18h)
21–22ss (19.5h)
h
Genotypes
foxd3 mutant
Wild type
i
Pigment gch2-high
Pigment sox6-high
Arch2
Arch1
Head mes.
tSNE_2
tSNE_1
j
HOX1/2
smarcc1a/b
nr2c2
RXR & RAR
NF2
nr2f2
ETS
ETS
MITFA
HLH
sox10
tfap2a/c
tfap2e
zic4+zic6
zic1
Posterior hox
soxB1
sox2
EGR
HNF1
−log10(P−adj) [0−Max]
100
0
NPB_hox3
Mutant_hox3
dNC_hox34
Mut_nohox_cycl.
Mutant_hox2
mNC_nohox
dNC_nohox
Mut_pigment
Mut_nohox_12_22ss
mNC_arch2
dNC_nohox_cycl.
dNC_hoxa2b
NC_trunk
mNC_vagal
NPB_nohox
Mut_nohox_early
NPB_nohox_cycl.
NPB_hox2
Pigment_sox6_high
Pigment_gch2_high
mNC_head_mes
mNC_arch1
mNC_hox34
slc4a4a
arhgef28a
adgrl1a
mafba
col18a1a
apoeb
bcam
lmx1bb
stm
afap1l1a
cbarp
snai1b
hoxc3a
bmper
epha7
cyp26c1
ebf3a
lfx
gdf11
wnt8b
cdh1a
rspo2
crb2a
p3h2
plecs
egr2b
kcna0.1a
mcf2l
cnfr
pdgfra
sox6
trpm1b
tspan36
gren2b
pmp22a
colec12
ebf3a.1
plpp3
elavl3
rxraa
fli1a
col11a1b
tnc
atx4a
itga3a
lepa
postn
adamts6
foxc1a
pax7a
cthrc1a
drd2b
dachs
edil3a
cxcl14
pag1
slc45a2
pax7b
gch2
ptuela
CABZ01021592.1
mitfa
Average expression
2
1
0
−1
Percent expressed
100
75
50
25
0
k
10-12ss
Cranial region
NPB → dNC (hindbrain)
NC cell states
Pigment
Mid/epi
dNC
DN
NPB
Optic
mNC
Tel/di
Neural_dorsal
Neural_floorplate
Neural_tel_di (Tel/di)
Hindbrain_posterior ventral_nkx6.2
Neural_midbrain, epiphysis (Mid/epi)
Optic
Otic placode
Tailbud_PSM
Tailbud_spinal_cord
Epidermal_1
21ss
Cranial region
dNC → mNC → Pigment / Arches
NC cell states
Pigment
dNC
Tel/di
Optic
mNC
Arch2
Arch1
NC cell states
NPB
dNC
mNC_nohox
mNC_hox34
mNC_dNC_arch2
mNC_arch1
Head mes.
Pigment gch2-high
Pigment sox6-high
l
UMAP_2
UMAP_1
Organogenesis data (Wagner et al.)
10x Multiome NC
Smart−seq3 NC

Fig. 1: Single-cell multiomic and spatial transcriptomic profiling reveals the dynamics of zebrafish NC development.

a, Schematic diagram of the study design. LOF, loss of function; WT, wild type.

b-c, Single-cell atlas of zebrafish foxd3-expressing NC and posterior populations, shown via uniform manifold approximation and projection (UMAP) visualization of the multiome-RNA (b) and Smart-seq3 (c) datasets, with individual cells (dots) color-coded by major cell types. PSM, pre-somite mesoderm; NC, neural crest.

d, Boxplot detailing gene coverage across Smart-seq3 data, highlighting the upper, middle, and lower quartiles.

e-h, *t*-distributed Stochastic Neighbor Embedding (*t*-SNE) plots of NC cell atlas in multiome-RNA and Smart-seq3 datasets, color-coded by cell states (e), developmental stages (f, g) and genotypes (h). ss, somites stage; h, hour post fertilization.

i, RNA velocity analysis results in wild-type NC cells.

j, Motif enrichment (top) and marker gene expression (bottom) in multiome-ATAC and -RNA datasets, against 23 NC cell states (x-axis).

k, MERSCOPE cell state mapping analysis shows the spatial location of cells (dots, colored by broad cell types or NC cell states). A, anterior; P, posterior; D, dorsal; V, ventral; Tel/di, telencephalon/diencephalon; DN, dorsal neural.

l, Integration of zebrafish organogenesis data [17] with our multiome-RNA and Smart-seq3 NC datasets. Cells (dots) are color-coded by datasets.

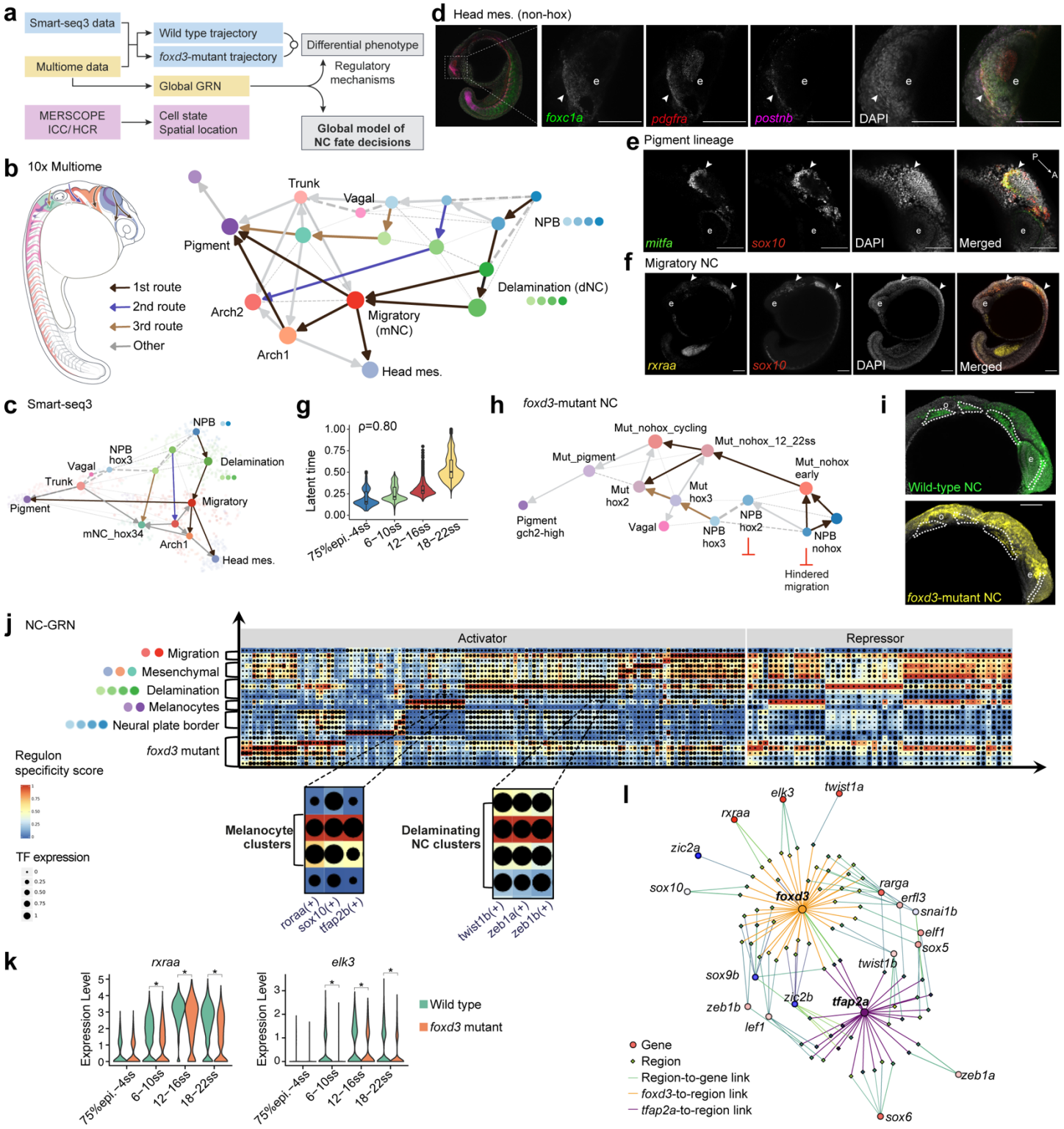


Fig. 2: Developmental trajectories and the gene regulatory network (GRN) of zebrafish cranial NC.

a, Schematic of the single-cell multiomic analysis workflow.

b-c, Developmental trajectories of wild-type NC reconstructed using multiome-RNA data (b) and full-length transcript Smart-seq3 data (c), with cell states annotated (NPB, neural plate border; head mes., head mesenchyme; mNC, migratory NC).

d, Triple HCR of *foxc1a*, *pdgfra*, and *postnb* pinpoints cell positions within the whole and cranial regions of a 21-ss wild-type embryo.

e, Cranial pigment-lineage derivatives, marked by arrows, are confirmed by *sox10* and *mitfa* expression patterns in the cranial region of a 21-ss wild-type embryo.

f, Migratory NC cells are confirmed in the anterior and hindbrain regions using HCR of *rxraa* and *sox10* in a 21-ss wild-type embryo.

g, Boxplot shows positive correlation of latent time with developmental stages.

h, Developmental trajectories of *foxd3*-mutant NC cells using multiome-RNA data, highlighting the impact of *foxd3* deficiency on NC development. Mut, mutant.

i, Comparative images illustrating hindered NC migration in *foxd3* mutants versus wild type. The green fluorescence (Citrine-positive/mCherry-negative) indicates wild-type NC cells and derivatives, while the yellow (Citrine-positive/mCherry-positive) indicates *foxd3*-mutant NC cells and derivatives. The *foxd3*-mutant NC cells did not populate the target regions, denoted by dashed lines.

j, Heatmap of NC-GRN reconstructed from multiome-RNA and -ATAC data, displaying regulon specificity scores (RSS) for regulons (x-axis) across NC cell states (y-axis). The zoom-in panels are the examples of six previously recognized regulons.

k, Downregulation of *rxraa* and *elk3* in *foxd3*-mutant NC cells. The star indicates Benjamini-Hochberg (BH)-corrected p-values < 4e-28 by limma voom. Epi., epiboly.

l, Network visualizing downstream genes and cis-regulatory regions of *foxd3* and *tfap2a*.

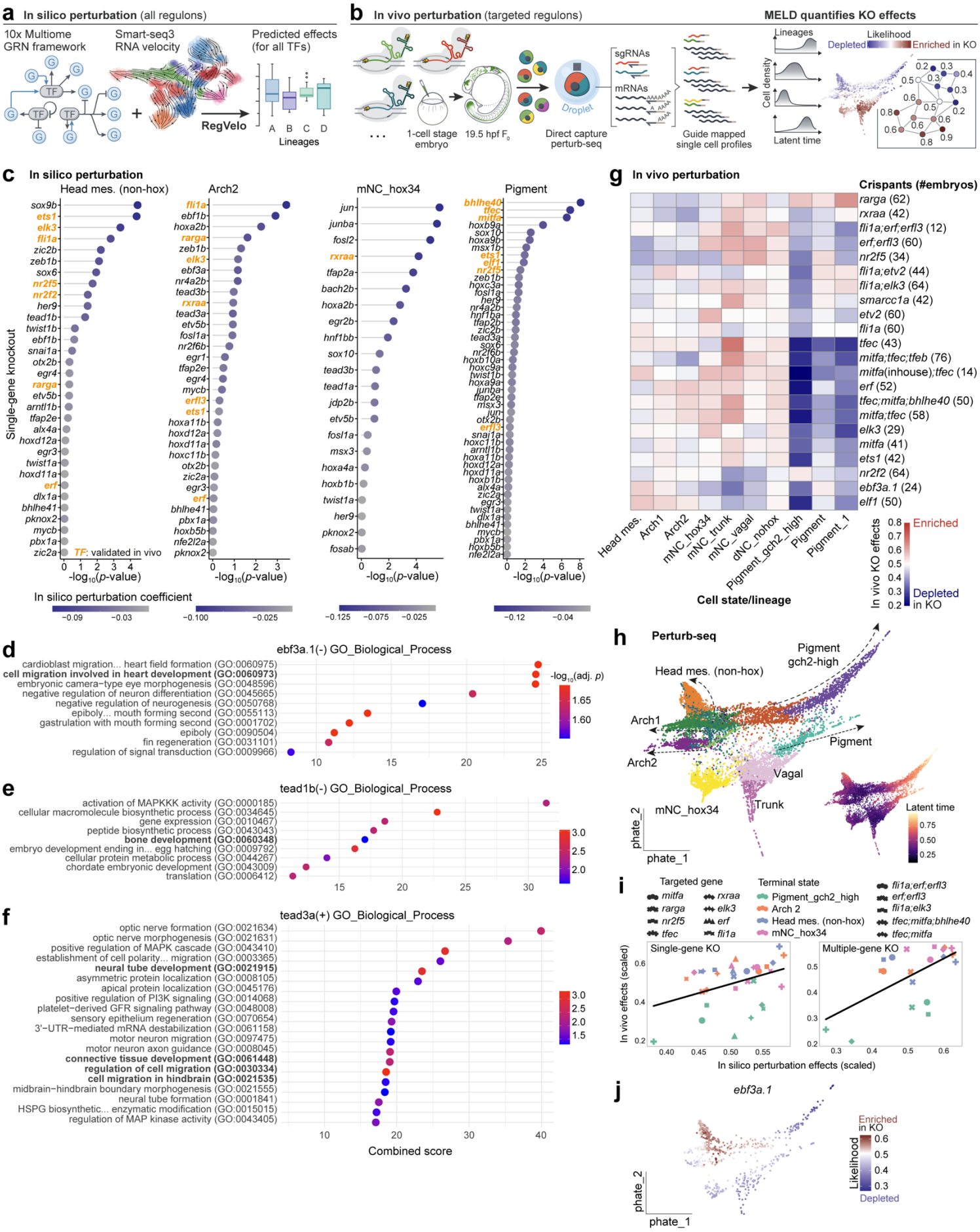

a In silico perturbation (all regulons)
10x Multiome GRN framework
Smart-seq3 RNA velocity
Predicted effects (for all TFs)
RegVelo
Lineages
b In vivo perturbation (targeted regulons)
MELD quantifies KO effects
1-cell stage embryo
19.5 hpf F0
Droplet
Direct capture perturb-seq
sgRNAs
mRNAs
Guide mapped single cell profiles
Lineages
Cell density
Latent time
Likelihood
Depleted
Enriched in KO
c In silico perturbation
Head mes. (non-hox)
Arch2
mNC_hox34
Pigment
Single-gene knockout
TF: validated in vivo
-log10(p-value)
In silico perturbation coefficient
g In vivo perturbation
Crispants (#embryos)
rarga (62)
rxraa (42)
fli1a;erf;erfl3 (12)
erf;erfl3 (60)
nr2f5 (34)
fli1a;etv2 (44)
fli1a;elk3 (64)
smarcc1a (42)
etv2 (60)
fli1a (60)
tfec (43)
mitfa;tfec;tfeb (76)
mitfa(inhouse);tfec (14)
erf (52)
tfec;mitfa;bhlhe40 (50)
mitfa;tfec (58)
elk3 (29)
mitfa (41)
ets1 (42)
nr2f2 (64)
ebf3a.1 (24)
elf1 (50)
Head mes.
Arch1
Arch2
mNC_hox34
mNC_trunk
mNC_vagal
dNC_nohox
Pigment_gch2_high
Pigment
Pigment_1
Cell state/lineage
In vivo KO effects
Enriched
Depleted in KO
d ebf3a.1(-) GO_Biological_Process
cardioblast migration... heart field formation (GO:0060975)
cell migration involved in heart development (GO:0060973)
embryonic camera−type eye morphogenesis (GO:0048596)
negative regulation of neuron differentiation (GO:0045665)
negative regulation of neurogenesis (GO:0050768)
epiboly... mouth forming second (GO:0055113)
gastrulation with mouth forming second (GO:0001702)
epiboly (GO:0090504)
fin regeneration (GO:0031101)
regulation of signal transduction (GO:0009966)
-log10(adj. p)
e tead1b(-) GO_Biological_Process
activation of MAPKKK activity (GO:0000185)
cellular macromolecule biosynthetic process (GO:0034645)
gene expression (GO:0010467)
peptide biosynthetic process (GO:0043043)
bone development (GO:0060348)
embryo development ending in... egg hatching (GO:0009792)
cellular protein metabolic process (GO:0044267)
chordate embryonic development (GO:0043009)
translation (GO:0006412)
f tead3a(+) GO_Biological_Process
optic nerve formation (GO:0021634)
optic nerve morphogenesis (GO:0021631)
positive regulation of MAPK cascade (GO:0043410)
establishment of cell polarity... migration (GO:0003365)
neural tube development (GO:0021915)
asymmetric protein localization (GO:0008105)
apical protein localization (GO:0045176)
positive regulation of PI3K signaling (GO:0014068)
platelet−derived GFR signaling pathway (GO:0048008)
sensory epithelium regeneration (GO:0070654)
3'−UTR−mediated mRNA destabilization (GO:0061158)
motor neuron migration (GO:0097475)
motor neuron axon guidance (GO:0008045)
connective tissue development (GO:0061448)
regulation of cell migration (GO:0030334)
cell migration in hindbrain (GO:0021535)
midbrain−hindbrain boundary morphogenesis (GO:0021555)
neural tube formation (GO:0001841)
HSPG biosynthetic... enzymatic modification (GO:0015015)
regulation of MAP kinase activity (GO:0043405)
Combined score
h Perturb-seq
Head mes. (non-hox)
Pigment gch2-high
Arch1
Arch2
Pigment
Vagal
Trunk
mNC_hox34
Latent time
phate_1
phate_2
i Targeted gene
mitfa
rarga
nr2f5
tfec
rxraa
elk3
erf
fli1a
Terminal state
Pigment_gch2_high
Arch 2
Head mes. (non-hox)
mNC_hox34
fli1a;erf;erfl3
erf;erfl3
fli1a;elk3
tfec;mitfa;bhlhe40
tfec;mitfa
Single-gene KO
Multiple-gene KO
In vivo effects (scaled)
In silico perturbation effects (scaled)
j ebf3a.1
Enriched in KO
Likelihood
Depleted
phate_1
phate_2

Fig. 3: In silico and in vivo perturbation methods probe novel differentiation circuits.

a, Schematic of RegVelo in silico perturbation methodology.

b, Diagram outlining in vivo direct-capture Perturb-seq workflow.

c, Identification of top TFs through in silico simulation for mesenchymal and pigment lineages, highlighting experimentally tested factors.

d-f, Gene ontology enrichment analysis for target genes of ebf3a.1(-), tead1b(-), and tead3a(+) regulons. Each regulon is specified in the subtitles. The x-axis represents the FishEnrichR combined scores, and the y-axis indicates the gene ontology groups.

g, In vivo lineage-specific perturbation likelihood across knockout panels and cell states calculated using Perturb-seq data.

h, PHATE representation depicting cell states and predicted latent time for all NC cells in the Perturb-seq dataset.

i, Scatterplots showing the strong correlation between in silico RegVelo prediction and in vivo Perturb-seq likelihood.

j, Perturbation likelihood results highlighting depletion of hindbrain and posterior NC in *ebf3a.1* knockouts.

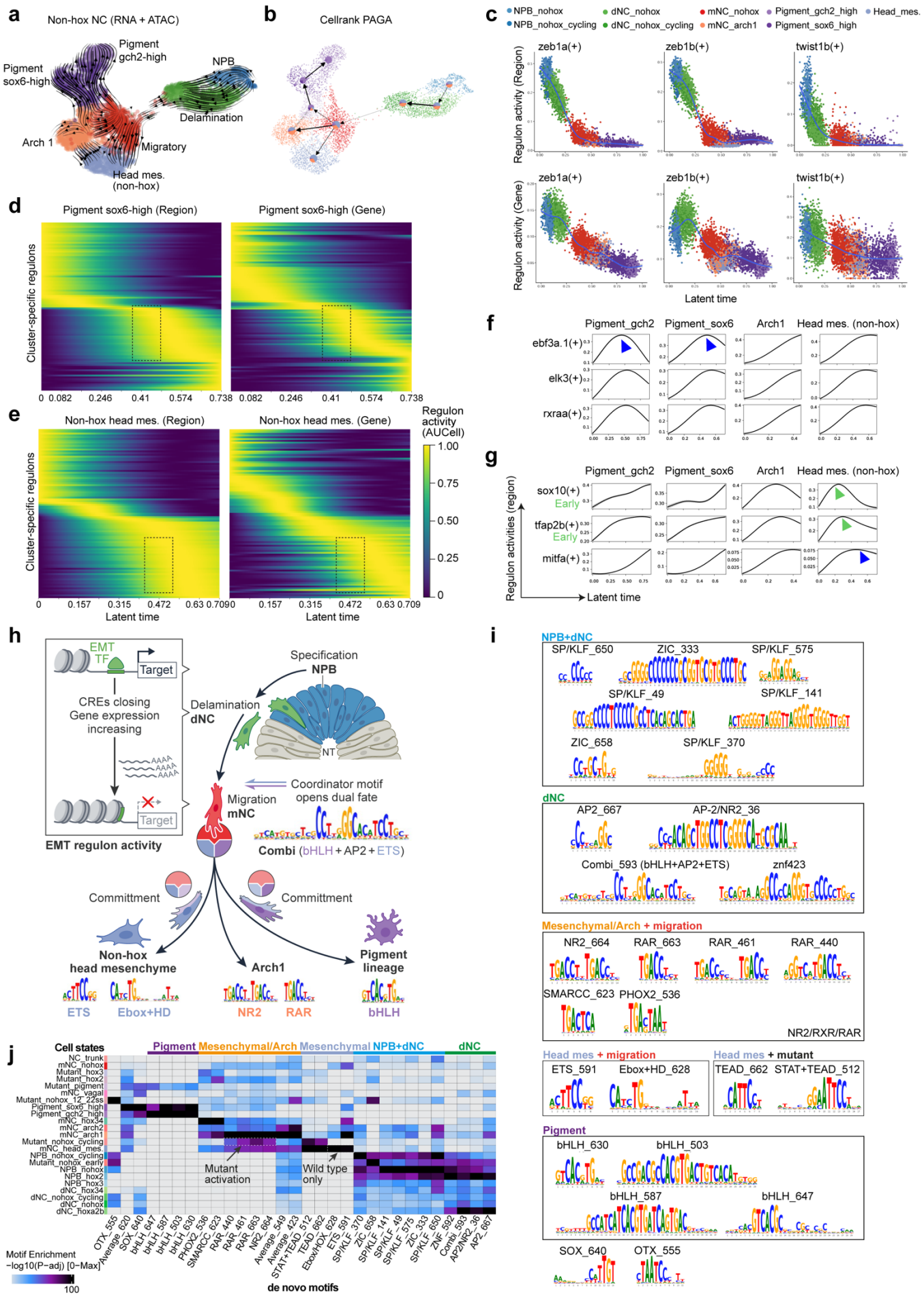

a
Non-hox NC (RNA + ATAC)
Pigment gch2-high
Pigment sox6-high
NPB
Delamination
Arch 1
Migratory
Head mes. (non-hox)
b
Cellrank PAGA
c
NPB_nohox
NPB_nohox_cycling
dNC_nohox
dNC_nohox_cycling
mNC_nohox
mNC_arch1
Pigment_gch2_high
Pigment_sox6_high
Head_mes.
zeb1a(+)
zeb1b(+)
twist1b(+)
Regulon activity (Region)
Regulon activity (Gene)
Latent time
d
Pigment sox6-high (Region)
Pigment sox6-high (Gene)
Cluster-specific regulons
0 0.082 0.246 0.41 0.574 0.738
e
Non-hox head mes. (Region)
Non-hox head mes. (Gene)
Regulon activity (AUCell)
1.00
0.75
0.50
0.25
0
0 0.157 0.315 0.472 0.63 0.709
f
Pigment_gch2
Pigment_sox6
Arch1
Head mes. (non-hox)
ebf3a.1(+)
elk3(+)
rxraa(+)
g
sox10(+)
Early
tfap2b(+)
Early
mitfa(+)
Regulon activities (region)
h
EMT TF
Target
CREs closing
Gene expression
increasing
EMT regulon activity
Specification
NPB
Delamination
dNC
NT
Migration
mNC
Coordinator motif
opens dual fate
Combi (bHLH + AP2 + ETS)
Committment
Committment
Non-hox
head mesenchyme
Arch1
Pigment
lineage
ETS
Ebox+HD
NR2
RAR
bHLH
i
NPB+dNC
SP/KLF_650
ZIC_333
SP/KLF_575
SP/KLF_49
SP/KLF_141
ZIC_658
SP/KLF_370
dNC
AP2_667
AP-2/NR2_36
Combi_593 (bHLH+AP2+ETS)
znf423
Mesenchymal/Arch + migration
NR2_664
RAR_663
RAR_461
RAR_440
SMARCC_623
PHOX2_536
NR2/RXR/RAR
Head mes + migration
ETS_591
Ebox+HD_628
Head mes + mutant
TEAD_662
STAT+TEAD_512
Pigment
bHLH_630
bHLH_503
bHLH_587
bHLH_647
SOX_640
OTX_555
j
Cell states
Pigment
Mesenchymal/Arch
Mesenchymal
NPB+dNC
dNC
NC_trunk
mNC_hohox
Mutant_hox3
Mutant_hox2
Mutant_pigment
mNC_vagal
Mutant_nohox_12_22ss
Pigment_sox6_high
Pigment_gch2_high
mNC_hox34
mNC_arch2
mNC_arch1
Mutant_nohox_cycling
mNC_head_mes
NPB_nohox_cycling
Mutant_nohox_early
NPB_nohox
NPB_hox2
NPB_hox3
dNC_hox34
dNC_nohox_cycling
dNC_nohox
dNC_hoxa2b
Mutant activation
Wild type only
OTX_555
Average_620
SOX_640
bHLH_647
bHLH_587
bHLH_503
bHLH_630
PHOX2_536
SMARCC_623
RAR_440
RAR_461
RAR_663
NR2_664
Average_549
Average_423
STAT+TEAD_512
TEAD_662
Ebox/HOX_628
ETS_591
SP/KLF_370
ZIC_658
SP/KLF_141
SP/KLF_49
SP/KLF_575
ZIC_333
SP/KLF_650
ZNF_592
Combi_593
AP2/NR2_36
AP2_667
Motif Enrichment
−log10(P−adj) [0−Max]
100
de novo motifs

Fig. 4: Unraveling EMT and differentiation dynamics through temporal GRN and deep learning analyses.

a, Multiomic velocity streams of non-hox NC trajectories.

b, CellRank PAGA results depicting the trajectories (directed arrows) and terminal fate probability (pie plots) for each cell state.

c, Trend plots exhibiting region-based and gene-based regulon activities over latent time, with cells (dots) colored by states, and blue lines indicating spline regression results.

d-e, Heatmaps showing region- and gene-based regulon activity cascades (n = 95) along latent time for *sox6*-high pigment (d) and head mesenchymal lineages (e), with priming events marked by boxes.

f-g, Trend plots for mesenchymal/migratory (f) and pigment regulons (g) in four lineages. The y-axes indicate regulon activity, and the x-axes latent time.

h, Diagram illustrating EMT regulatory dynamics and differentiation regulatory mechanisms in NC development.

i, ChromBPNet-derived motifs catalogued by enrichment in cell states. HD, homeodomain.

j, Motif enrichment analysis across NC cell states. NPB, neural plate border; dNC, delamination NC; mNC, migratory NC.

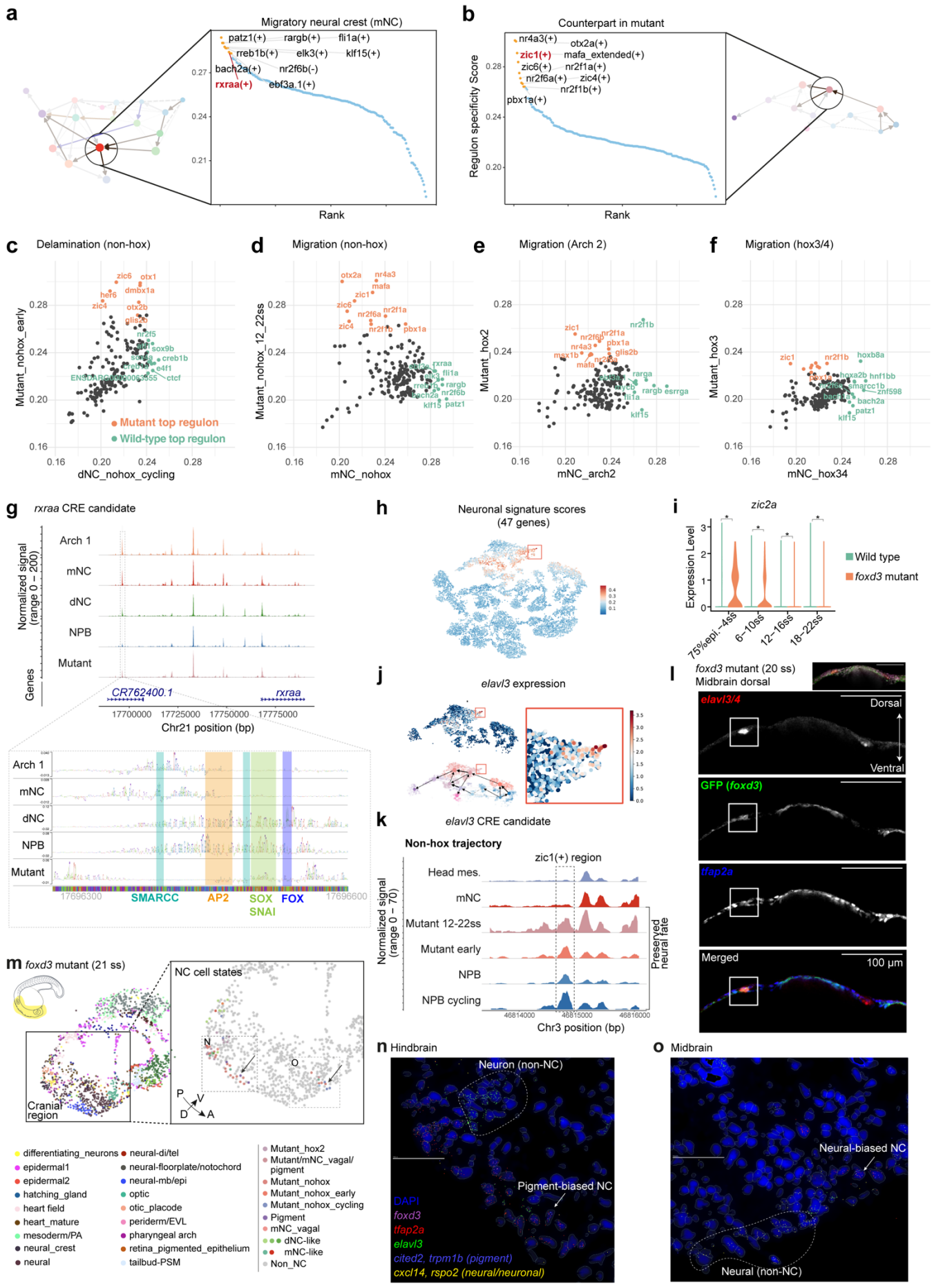

a
Migratory neural crest (mNC)
patz1(+) rargb(+) fli1a(+)
rreb1b(+) elk3(+) klf15(+)
bach2a(+) nr2f6b(-)
rxraa(+) ebf3a.1(+)
Rank
b
Counterpart in mutant
nr4a3(+) otx2a(+)
zic1(+) mafa_extended(+)
zic6(+) nr2f1a(+)
nr2f6a(+) zic4(+)
nr2f1b(+)
pbx1a(+)
Regulon specificity Score
Rank
c Delamination (non-hox)
Mutant_nohox_early
dNC_nohox_cycling
Mutant top regulon
Wild-type top regulon
d Migration (non-hox)
Mutant_nohox_12_22ss
mNC_nohox
e Migration (Arch 2)
Mutant_hox2
mNC_arch2
f Migration (hox3/4)
Mutant_hox3
mNC_hox34
g rxraa CRE candidate
Normalized signal (range 0 – 200)
Arch 1
mNC
dNC
NPB
Mutant
Genes
CR762400.1
rxraa
Chr21 position (bp)
SMARCC
AP2
SOX
SNAI
FOX
h Neuronal signature scores (47 genes)
i zic2a
Expression Level
Wild type
foxd3 mutant
75%epi.–4ss
6–10ss
12–16ss
18–22ss
j elavl3 expression
k elavl3 CRE candidate
Non-hox trajectory
zic1(+) region
Head mes.
mNC
Mutant 12-22ss
Mutant early
NPB
NPB cycling
Preserved neural fate
Chr3 position (bp)
l foxd3 mutant (20 ss)
Midbrain dorsal
elavl3/4
Dorsal
Ventral
GFP (foxd3)
tfap2a
Merged
100 µm
m foxd3 mutant (21 ss)
NC cell states
Cranial region
differentiating_neurons
epidermal1
epidermal2
hatching_gland
heart field
heart_mature
mesoderm/PA
neural_crest
neural
neural-di/tel
neural-floorplate/notochord
neural-mb/epi
optic
otic_placode
periderm/EVL
pharyngeal arch
retina_pigmented_epithelium
tailbud-PSM
Mutant_hox2
Mutant/mNC_vagal/pigment
Mutant_nohox
Mutant_nohox_early
Mutant_nohox_cycling
Pigment
mNC_vagal
dNC-like
mNC-like
Non_NC
n Hindbrain
Neuron (non-NC)
Pigment-biased NC
DAPI
foxd3
tfap2a
elavl3
cited2, trpm1b (pigment)
cxcl14, rspo2 (neural/neuronal)
o Midbrain
Neural-biased NC
Neural (non-NC)

Fig. 5: *foxd3* deficiency disturbs migration and releases a neural fate.

a-b, Differential regulon activity between migratory NC cells and counterpart cells in *foxd3* mutants, with directed graphs showcasing wild-type and mutant NC trajectories (a, b, respectively). The selected cell states are highlighted for regulon specificity score scatterplots.

c-f, Scatterplots illustrating differences in regulon activities between wild-type and mutant cells, across stages and lineages (by titles). The x- and y-axes indicate two cell states from wild type and mutant, respectively.

g, The rxraa cis-regulatory element candidate contains putative TF binding sites of several factors (SMARCC, AP2, SOX, SNAI, and FOX). The upper track set indicates chromatin accessibility profiles and the lower one displays dynamic tracks based on ChromBPNet contribution score profiles.

h, *t*-SNE plot depicting the strength of a neuronal transcriptional signature in NC cells (dots).

i, Upregulation of *zic2a* in early-stage *foxd3*-mutant NC cells compared to wild type. The star indicates BH-corrected p-values < 0.05 by limma voom.

j, Identification of a rare population expressing neuronal marker *elavl3* in *foxd3*-mutant cells.

k, Genomic tracks presenting the chromatin accessibility of an *elavl3* cis-regulatory element candidate harboring *zic1* TF binding site in NPB and mutant cell states, implying preservation of neural fate in *foxd3* mutants.

l, Antibody staining results of *elavl3*, *tfap2a*, and GFP (indicative of *foxd3*) in a 20-ss *foxd3*-mutant embryo, marking neuronal NC cells.

m, Spatial representation of MERSCOPE data of 21-ss *foxd3*-mutant embryos indicating the major mutant NC states.

n-o, Two zoomed-in regions in the cranial domain elucidate the pigment-biased (*trpm1b*, *foxd3* (mutant), *tfap2a*-coexpressing) and the neural-biased (*cxcl14*, *elavl3*, *foxd3* (mutant), *tfap2a*-

coexpressing) mutant NC cell states in the hindbrain (n) and midbrain regions (o). The arrow indicates the same cell as in (m). Scale bars, 100 μm.

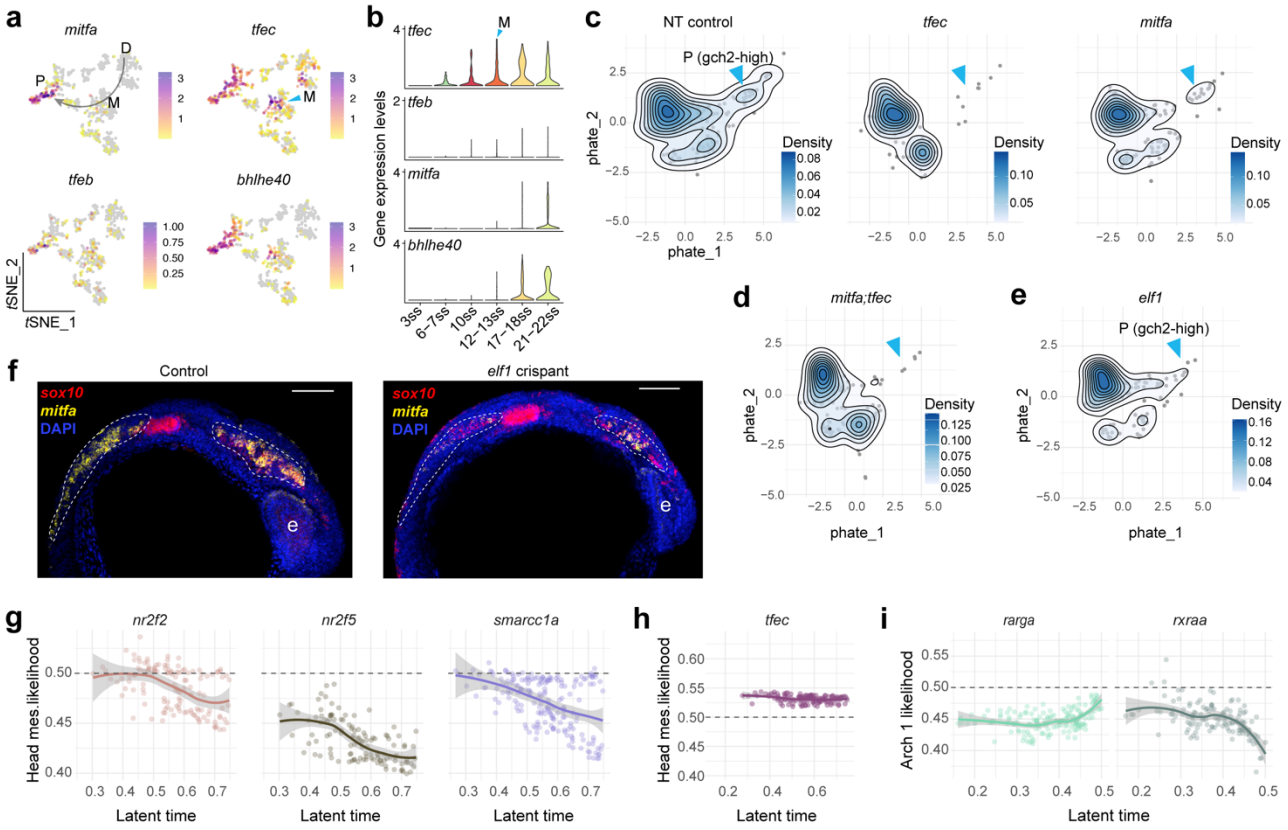


Fig. 6: TF effectors with individual and convergent functions for differentiation revealed by predictive GRN.

a-b, Expression patterns of representative bHLH factors in sox6-high pigment lineage within Smart-seq3 data presented in *t*-SNE space (a) and violin plots by stages (b). M, migration; D, delamination; P, pigment.

c-e, Density plots illustrating cell distribution differences among non-targeting (NT) control, *tfec*, *mitfa* (c), *tfec*;*mitfa* (d), and *elf1* knockout (e) panels in PHATE representations.

f, HCR demonstrates reduction in *mitfa*-expressing pigment lineage in *elf1* crispants.

g-i, Knockout (KO) effects visualized along predicted latent time for migratory/mesenchymal factors (g), *tfec* (h), and pharyngeal arch factors (i), with lines and surrounding grey areas indicating spline regression results and confidence intervals, respectively. The y-axis denotes knockout effects measured by lineage-specific likelihood, indicating enrichment (above 0.5) or depletion (below 0.5).

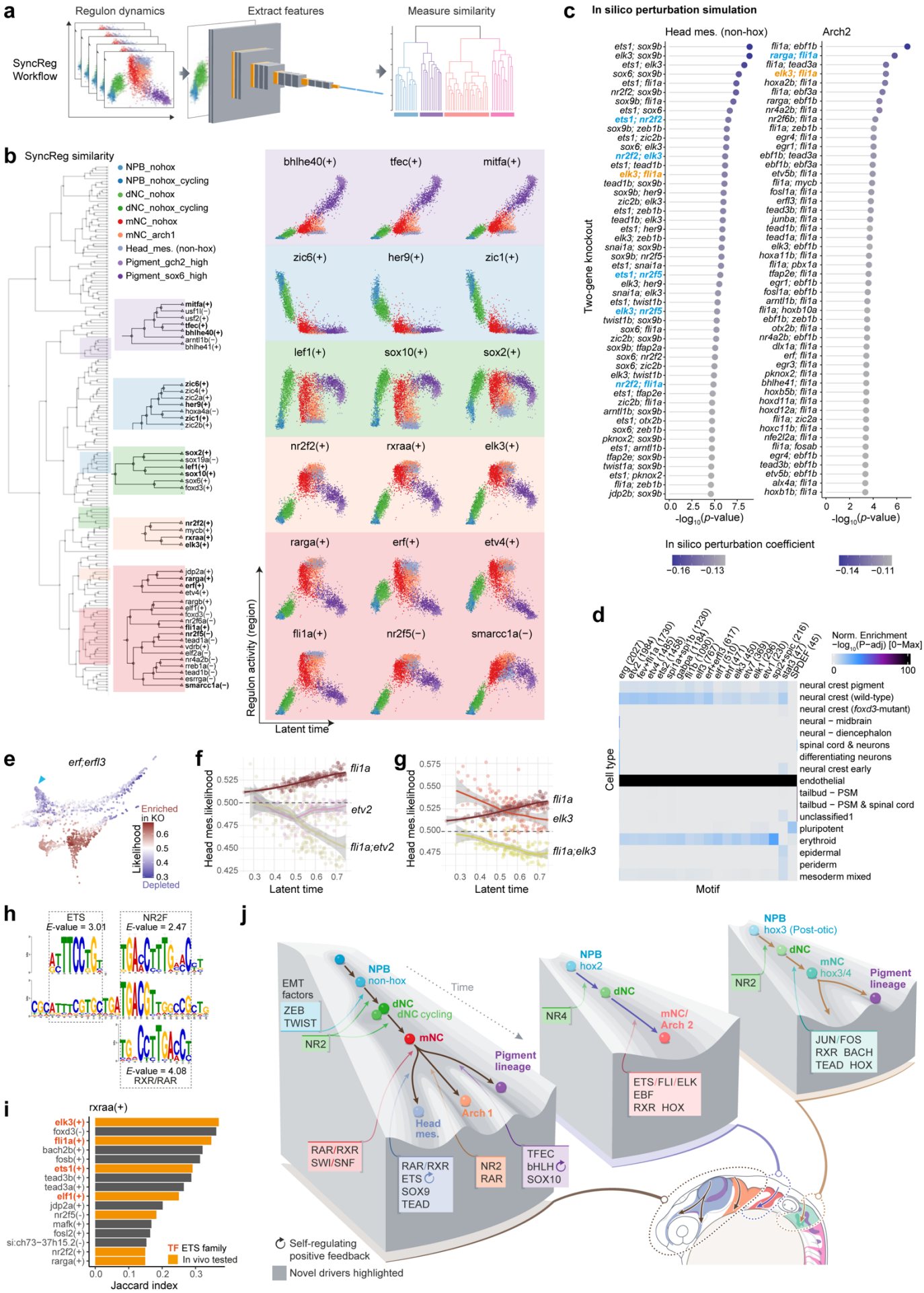

a
Regulon dynamics
Extract features
Measure similarity
SyncReg Workflow
b
SyncReg similarity
NPB_nohox
NPB_nohox_cycling
dNC_nohox
dNC_nohox_cycling
mNC_nohox
mNC_arch1
Head_mes. (non-hox)
Pigment_gch2_high
Pigment_sox6_high
bhlhe40(+)
tfec(+)
mitfa(+)
zic6(+)
her9(+)
zic1(+)
lef1(+)
sox10(+)
sox2(+)
nr2f2(+)
rxraa(+)
elk3(+)
rarga(+)
erf(+)
etv4(+)
fli1a(+)
nr2f5(−)
smarcc1a(−)
Regulon activity (region)
Latent time
c
In silico perturbation simulation
Head mes. (non-hox)
Arch2
Two-gene knockout
-log10(p-value)
In silico perturbation coefficient
d
Norm. Enrichment
Cell type
Motif
neural crest pigment
neural crest (wild-type)
neural crest (foxd3-mutant)
neural − midbrain
neural − diencephalon
spinal cord & neurons
differentiating neurons
neural crest early
endothelial
tailbud − PSM
tailbud − PSM & spinal cord
unclassified1
pluripotent
erythroid
epidermal
periderm
mesoderm mixed
e
erf;erfl3
Enriched in KO
Likelihood
Depleted
f
Head mes.likelihood
fli1a
etv2
fli1a;etv2
Latent time
g
Head mes.likelihood
fli1a
elk3
fli1a;elk3
Latent time
h
ETS
E-value = 3.01
NR2F
E-value = 4.25
E-value = 4.08
RXR/RAR
i
rxraa(+)
elk3(+)
foxd3(-)
fli1a(+)
bach2b(+)
fosb(+)
ets1(+)
tead3b(+)
tead3a(+)
elf1(+)
jdp2a(+)
nr2f5(-)
mafk(+)
fosl2(+)
si:ch73−37h15.2(-)
nr2f2(+)
rarga(+)
TF ETS family
In vivo tested
Jaccard index
j
EMT factors
ZEB
TWIST
NR2
NPB
non-hox
dNC
dNC cycling
mNC
Time
Pigment lineage
Head mes.
Arch 1
RAR/RXR
SWI/SNF
RAR/RXR
ETS
SOX9
TEAD
NR2
RAR
TFEC
bHLH
SOX10
NPB
hox2
dNC
NR4
mNC/
Arch 2
ETS/FLI/ELK
EBF
RXR HOX
NPB
hox3 (Post-otic)
dNC
mNC
hox3/4
Pigment lineage
NR2
JUN/FOS
RXR BACH
TEAD HOX
Self-regulating positive feedback
Novel drivers highlighted

Fig. 7: Transcriptional synergistic and redundant effects explored in regulon-mediated cell fate determination.

a, Schematic diagram of SyncReg temporal dynamic analysis.

b, SyncReg dynamics of all regulons identified temporally concurrent regulons, situated in the proximal branches of a dendrogram constructed based on dynamic similarity, with cells in dynamic trend plots colored by states. Examples of regulon temporal groups are presented, including the eight perturbed migratory/mesenchymal regulons in the orange and red panels.

c, TF pairs of top two-gene perturbation effects ranked by the $-\log_{10}$(p-value). KO, knockout. Blue highlights potential cooperative TF pairs (different motifs), while orange highlights putative redundant factors (sharing motifs).

d, Motif enrichment analysis of multiome-ATAC data indicates that the ETS motifs are significantly enriched in wild-type NC cells and absent in *foxd3* mutants.

e, Perturbation likelihood results highlighting depletion of non-hox head mesenchymal (blue arrow) and pigment lineages in *erf*;*erfl3* double knockouts.

f-g, Comparative analysis of knockout effects in head mesenchymal lineage for single-gene and two-gene knockout panels, with cells and regression lines color-coded by knockout conditions.

h, Analysis of a putative coordinator motif of ETS and NR2F/RXR/RAR with motifs compared against DANIO-CODE.

i, Identification of top regulons that are highly overlapping with rxraa(+), measured by Jaccard indexes, and highlighting ETS factors and in vivo perturbed regulons.

j, Conceptual diagram of the proposed regulatory model governing NC cell fate decisions.

References

1. Casey, M.J., Stumpf, P.S. & MacArthur, B.D. Theory of cell fate. *Wiley Interdiscip Rev Syst Biol Med* **12**, e1471 (2020).
2. Bravo Gonzalez-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* **20**, 1355-1367 (2023).
3. Wang, L. *et al.* Dictys: dynamic gene regulatory network dissects developmental continuum with single-cell multiomics. *Nat Methods* (2023).
4. Rothstein, M., Bhattacharya, D. & Simoes-Costa, M. The molecular basis of neural crest axial identity. *Developmental Biology* **444**, S170-S180 (2018).
5. Martik, M.L. & Bronner, M.E. Riding the crest to get a head: neural crest evolution in vertebrates. *Nature Reviews Neuroscience* **22**, 616-626 (2021).
6. D'Amico-Martel, A. & Noden, D.M. Contributions of placodal and neural crest cells to avian cranial peripheral ganglia. *Am J Anat* **166**, 445-68 (1983).
7. Le Douarin, N. & Kalcheim, C. *The neural crest*, (Cambridge university press, 1999).
8. Hörstadius, S. The neural crest : its properties and derivatives in the light of experimental research. (1950).
9. Trainor, P.A. Craniofacial birth defects: The role of neural crest cells in the etiology and pathogenesis of Treacher Collins syndrome and the potential for prevention. *Am J Med Genet A* **152A**, 2984-94 (2010).
10. Moore, S.W. The contribution of associated congenital anomalies in understanding Hirschsprung's disease. *Pediatric Surgery International* **22**, 305-315 (2006).
11. Long, H.K. *et al.* Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. *Cell Stem Cell* **27**, 765-783 e14 (2020).
12. Zalc, A. *et al.* Reactivation of the pluripotency program precedes formation of the cranial neural crest. *Science* **371**(2021).
13. Williams, R.M. *et al.* Reconstruction of the Global Neural Crest Gene Regulatory Network In Vivo. *Developmental Cell* **51**, 255-276.e7 (2019).
14. Soldatov, R. *et al.* Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* **364**(2019).
15. Kim, S. *et al.* DNA-guided transcription factor cooperativity shapes face and limb mesenchyme. *Cell* **187**, 692-711 e26 (2024).
16. Farrell, J.A. *et al.* Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).
17. Wagner, D.E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-987 (2018).
18. Farnsworth, D.R., Saunders, L.M. & Miller, A.C. A single-cell transcriptome atlas for zebrafish development. *Dev Biol* **459**, 100-108 (2020).
19. Raj, B. *et al.* Emergence of Neuronal Diversity during Vertebrate Brain Development. *Neuron* **108**, 1058-1074.e6 (2020).
20. Kamimoto, K. *et al.* Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742-751 (2023).
21. Li, C., Virgilio, M.C., Collins, K.L. & Welch, J.D. Multi-omic single-cell velocity models epigenome-transcriptome interactions and improves cell fate prediction. *Nat Biotechnol* **41**, 387-398 (2023).
22. Parvez, S. *et al.* MIC-Drop: A platform for large-scale in vivo CRISPR screens. *Science* **373**, 1146-1151 (2021).
23. Saunders, L.M. *et al.* Embryo-scale reverse genetics at single-cell resolution. *Nature* **623**, 782-791 (2023).
24. Replogle, J.M. *et al.* Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nature Biotechnology* **38**, 954-961 (2020).

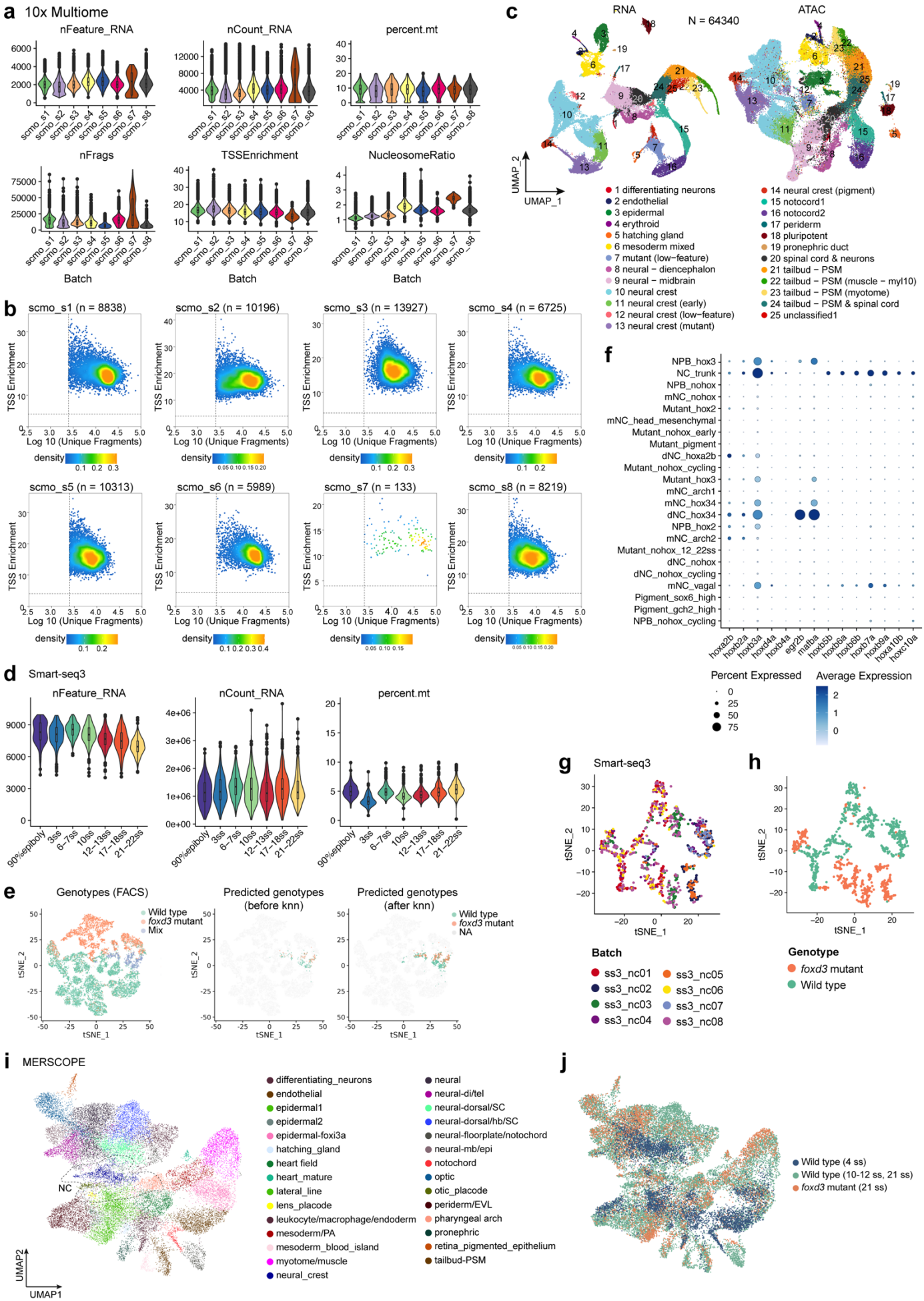
25. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proceedings of the National Academy of Sciences* **116**, 19490-19499 (2019).
26. Biancalani, T. *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods* **18**, 1352-1362 (2021).
27. Maizels, R.J., Snell, D.M. & Briscoe, J. Deep dynamical modelling of developmental trajectories with temporal transcriptomics. *bioRxiv*, 2023.07.06.547989 (2023).
28. Brennan, K.J. *et al.* Chromatin accessibility in the Drosophila embryo is determined by transcription factor pioneering and enhancer activation. *Dev Cell* **58**, 1898-1916 e9 (2023).
29. Martik, M.L. & Bronner, M.E. Regulatory Logic Underlying Diversification of the Neural Crest. *Trends Genet* **33**, 715-727 (2017).
30. Davidson, E.H. & Erwin, D.H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796-800 (2006).
31. Davidson, E.H. Emerging properties of animal gene regulatory networks. *Nature* **468**, 911-20 (2010).
32. Rocha, M., Singh, N., Ahsan, K., Beiriger, A. & Prince, V.E. Neural crest development: insights from the zebrafish. *Dev Dyn* **249**, 88-111 (2020).
33. Dottori, M., Gross, M.K., Labosky, P. & Goulding, M. The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. *Development* **128**, 4127-38 (2001).
34. Kos, R., Reedy, M.V., Johnson, R.L. & Erickson, C.A. The winged-helix transcription factor FoxD3 is important for establishing the neural crest lineage and repressing melanogenesis in avian embryos. *Development* **128**, 1467-1479 (2001).
35. Stewart, R.A. *et al.* Zebrafish foxd3 is selectively required for neural crest specification, migration and survival. *Dev Biol* **292**, 174-88 (2006).
36. Lukoseviciute, M. *et al.* From Pioneer to Repressor: Bimodal foxd3 Activity Dynamically Remodels Neural Crest Regulatory Landscape In Vivo. *Developmental Cell* **47**, 608-628 e6 (2018).
37. Hochgreb-Hagele, T. & Bronner, M.E. A novel FoxD3 gene trap line reveals neural crest precursor movement and a role for FoxD3 in their specification. *Dev Biol* **374**, 1-11 (2013).
38. Bergen, V., Lange, M., Peidli, S., Wolf, F.A. & Theis, F.J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology* **38**, 1408-1414 (2020).
39. Fabian, P. *et al.* Lifelong single-cell profiling of cranial neural crest diversification in zebrafish. *Nature Communications* **13**(2022).
40. Das, A. & Crump, J.G. Bmps and Id2a Act Upstream of Twist1 To Restrict Ectomesenchyme Potential of the Cranial Neural Crest. *PLOS Genetics* **8**, e1002710 (2012).
41. Johnson, S.L., Nguyen, A.N. & Lister, J.A. mitfa is required at multiple stages of melanocyte differentiation but not to establish the melanocyte stem cell. *Developmental Biology* **350**, 405-413 (2011).
42. Wang, W.D., Melville, D.B., Montero-Balaguer, M., Hatzopoulos, A.K. & Knapik, E.W. Tfap2a and Foxd3 regulate early steps in the development of the neural crest progenitor population. *Dev Biol* **360**, 173-85 (2011).
43. Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-411 (2002).

44. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-6 (2010).
45. Pekowska, A. *et al.* H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J* **30**, 4198-210 (2011).
46. Ferrari, K.J. *et al.* Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol Cell* **53**, 49-62 (2014).
47. Pelea, O., Mayes, S., Ferry, Q.R.V., Fulga, T.A. & Sauka-Spengler, T. Specific Modulation of CRISPR Transcriptional Activators through RNA-Sensing Guide RNAs in Mammalian Cells and Zebrafish Embryos. *biorxiv* (2023).
48. Burkhardt, D.B. *et al.* Quantifying the effect of experimental perturbations at single-cell resolution. *Nature Biotechnology* **39**, 619-629 (2021).
49. Lange, M. *et al.* CellRank for directed single-cell fate mapping. *Nat Methods* **19**, 159-170 (2022).
50. Reuter, B., Fackeldey, K. & Weber, M. Generalized Markov modeling of nonreversible molecular kinetics. *The Journal of Chemical Physics* **150**(2019).
51. Nair, S. *et al.* Transcription factor stoichiometry, motif affinity and syntax regulate single-cell chromatin dynamics during fibroblast reprogramming to pluripotency. *bioRxiv*, 2023.10.04.560808 (2023).
52. Howard, A.G.t. *et al.* An atlas of neural crest lineages along the posterior developing zebrafish at single-cell resolution. *Elife* **10**(2021).
53. Saunders, L.M. *et al.* Thyroid hormone regulates distinct paths to maturation in pigment cell lineages. *Elife* **8**(2019).
54. Tahtakran, S.A. & Selleck, M.A. Ets-1 expression is associated with cranial neural crest migration and vasculogenesis in the chick embryo. *Gene Expr Patterns* **3**, 455-8 (2003).
55. Martik, M.L. *et al.* Evolution of the new head by gradual acquisition of neural crest regulatory circuits. *Nature* **574**, 675-678 (2019).
56. De Val, S. *et al.* Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors. *Cell* **135**, 1053-64 (2008).
57. Sizemore, G.M., Pitarresi, J.R., Balakrishnan, S. & Ostrowski, M.C. The ETS family of oncogenic transcription factors in solid tumours. *Nat Rev Cancer* **17**, 337-351 (2017).
58. Kok, F.O. *et al.* Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell* **32**, 97-108 (2015).
59. Ferrell, J.E., Jr. Bistability, bifurcations, and Waddington's epigenetic landscape. *Curr Biol* **22**, R458-66 (2012).
60. Bhattacharya, S., Zhang, Q. & Andersen, M.E. A deterministic map of Waddington's epigenetic landscape for cell fate specification. *BMC Syst Biol* **5**, 85 (2011).
61. Bhattacharya, D., Azambuja, A.P. & Simoes-Costa, M. Metabolic Reprogramming Promotes Neural Crest Migration via Yap/Tead Signaling. *Developmental Cell* **53**, 199-211.e6 (2020).
62. Ong, Y.T. *et al.* A YAP/TAZ-TEAD signalling module links endothelial nutrient acquisition to angiogenic growth. *Nat Metab* **4**, 672-682 (2022).
63. Ling, I.T.C. & Sauka-Spengler, T. Early chromatin shaping predetermines multipotent vagal neural crest into neural, neuronal and mesenchymal lineages. *Nature Cell Biology* **21**, 1504-1517 (2019).
64. Madison, B.J. *et al.* Electrostatic repulsion causes anticooperative DNA binding between tumor suppressor ETS transcription factors and JUN-FOS at composite DNA sites. *J Biol Chem* **293**, 18624-18635 (2018).
65. Pham, V.N. *et al.* Combinatorial function of ETS transcription factors in the developing vasculature. *Dev Biol* **303**, 772-83 (2007).

66. Liu, F. & Patient, R. Genome-wide analysis of the zebrafish ETS family identifies three genes required for hemangioblast differentiation or angiogenesis. *Circ Res* **103**, 1147-54 (2008).
67. Westerfield, M. *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Danio Rerio)*, (University of Oregon Press, 2000).
68. Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B. & Schilling, T.F. Stages of embryonic development of the zebrafish. *Developmental Dynamics* **203**, 253-310 (2005).
69. Young, M.D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* **9**(2020).
70. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
71. McGinnis, C.S., Murrow, L.M. & Gartner, Z.J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems* **8**, 329-337.e4 (2019).
72. Granja, J.M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411 (2021).
73. Baranasic, D. *et al.* Multiomic atlas with functional stratification and developmental dynamics of zebrafish cis-regulatory elements. *Nature Genetics* **54**, 1037-1050 (2022).
74. Grabherr, M.G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652 (2011).
75. Srivastava, A., Malik, L., Smith, T., Sudbery, I. & Patro, R. Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. *Genome Biology* **20**, 65 (2019).
76. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
77. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: visualizing classifier performance in R. *Bioinformatics* **21**, 3940-3941 (2005).
78. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nature Biotechnology* **38**, 708-714 (2020).
79. Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. *GigaScience* **7**(2018).
80. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2012).
81. Kuemmerle, L.B. *et al.* Probe set selection for targeted spatial transcriptomics. *bioRxiv*, 2022.08.16.504115 (2022).
82. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat Methods* **19**, 1634-1641 (2022).
83. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat Methods* **19**, 171-178 (2022).
84. Lindeman, L.C., Vogt-Kielland, L.T., Aleström, P. & Collas, P. Fish'n ChIPs: Chromatin Immunoprecipitation in the Zebrafish Embryo. in *Chromatin Immunoprecipitation Assays: Methods and Protocols* (ed. Collas, P.) 75-86 (Humana Press, Totowa, NJ, 2009).
85. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359 (2012).
86. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
87. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).

88. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* **38**, 576-589 (2010).
89. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**, W160-W165 (2016).
90. Choi, H.M.T. *et al.* Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**(2018).
91. Hu, Z., Ahmed, A.A. & Yau, C. CIDER: an interpretable meta-clustering framework for single-cell RNA-seq data integration and evaluation. *Genome Biology* **22**, 337 (2021).
92. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47-e47 (2015).
93. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
94. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nature Methods* **18**, 1333-1341 (2021).
95. Lukoseviciute, M. *et al.* From Pioneer to Repressor: Bimodal foxd3 Activity Dynamically Remodels Neural Crest Regulatory Landscape In Vivo. *Developmental Cell* **47**, 608-628.e6 (2018).
96. Faust, G.G. & Hall, I.M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503-2505 (2014).
97. Machanick, P. & Bailey, T.L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696-7 (2011).
98. Frith, M.C., Saunders, N.F., Kobe, B. & Bailey, T.L. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol* **4**, e1000071 (2008).
99. Shrikumar, A. *et al.* Technical Note on Transcription Factor Motif Discovery from Importance Scores (TF-MoDISco) version 0.5.6.5. *arXiv*, 1811.00416v5 (2020).
100. Bruse, N. & Heeringen, S.J.v. GimmeMotifs: an analysis framework for transcription factor motif analysis. *bioRxiv*, 474403 (2018).
101. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. Quantifying similarity between motifs. *Genome Biology* **8**, R24 (2007).
102. Yu, D., Chojnowski, G., Rosenthal, M. & Kosinski, J. AlphaPulldown-a python package for protein-protein interaction screens using AlphaFold-Multimer. *Bioinformatics* **39**(2023).
103. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.10.04.463034 (2022).
104. Kuleshov, M.V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research* **44**, W90-W97 (2016).
105. Hagberg, A., Swart, P. & Chult, D. Exploring Network Structure, Dynamics, and Function Using NetworkX. (2008).
106. Liu, S. & Deng, W. Very deep convolutional neural network based image classification using small training sample size. in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)* 730-734 (2015).
107. Qiu, X. *et al.* Mapping transcriptomic vector fields of single cells. *Cell* **185**, 690-711 e45 (2022).
108. Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research* **47**, W171-W174 (2019).
109. Xu, B. *et al.* Unifying developmental programs for embryonic and postembryonic neurogenesis in the zebrafish retina. *Development* **147**(2020).

110. Singh, P.P. & Isambert, H. OHNOLOGS v2: a comprehensive resource for the genes retained from whole genome duplication in vertebrates. *Nucleic Acids Research* **48**, D724-D730 (2020).
111. Moon, K.R. *et al.* Visualizing structure and transitions in high-dimensional biological data. *Nat Biotechnol* **37**, 1482-1492 (2019).
112. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* **42**, 293-304 (2024).
113. Lopez, R., Regier, J., Cole, M.B., Jordan, M.I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nature Methods* **15**, 1053-1058 (2018).


a
10x Multiome
nFeature_RNA
nCount_RNA
percent.mt
nFrags
TSSEnrichment
NucleosomeRatio
Batch
b
scmo_s1 (n = 8838)
scmo_s2 (n = 10196)
scmo_s3 (n = 13927)
scmo_s4 (n = 6725)
scmo_s5 (n = 10313)
scmo_s6 (n = 5989)
scmo_s7 (n = 133)
scmo_s8 (n = 8219)
TSS Enrichment
Log 10 (Unique Fragments)
density
c
RNA
ATAC
N = 64340
UMAP_2
UMAP_1
1 differentiating neurons
2 endothelial
3 epidermal
4 erythroid
5 hatching gland
6 mesoderm mixed
7 mutant (low−feature)
8 neural − diencephalon
9 neural − midbrain
10 neural crest
11 neural crest (early)
12 neural crest (low−feature)
13 neural crest (mutant)
14 neural crest (pigment)
15 notocord1
16 notocord2
17 periderm
18 pluripotent
19 pronephric duct
20 spinal cord & neurons
21 tailbud − PSM
22 tailbud − PSM (muscle − myl10)
23 tailbud − PSM (myotome)
24 tailbud − PSM & spinal cord
25 unclassified1
f
NPB_hox3
NC_trunk
NPB_nohox
mNC_nohox
Mutant_hox2
mNC_head_mesenchymal
Mutant_nohox_early
Mutant_pigment
dNC_hoxa2b
Mutant_nohox_cycling
Mutant_hox3
mNC_arch1
mNC_hox34
dNC_hox34
NPB_hox2
mNC_arch2
Mutant_nohox_12_22ss
dNC_nohox
dNC_nohox_cycling
mNC_vagal
Pigment_sox6_high
Pigment_gch2_high
NPB_nohox_cycling
hoxa2b
hoxb2a
hoxb3a
hoxd4a
hoxb4a
egr2b
mafba
hoxb5b
hoxb6a
hoxb6b
hoxb7a
hoxb9a
hoxa10b
hoxc10a
Percent Expressed
Average Expression
d
Smart-seq3
nFeature_RNA
nCount_RNA
percent.mt
90%epiboly
3ss
6–7ss
10ss
12–13ss
17–18ss
21–22ss
e
Genotypes (FACS)
Predicted genotypes (before knn)
Predicted genotypes (after knn)
Wild type
foxd3 mutant
Mix
NA
tSNE_1
tSNE_2
g
Smart-seq3
Batch
ss3_nc01
ss3_nc02
ss3_nc03
ss3_nc04
ss3_nc05
ss3_nc06
ss3_nc07
ss3_nc08
h
Genotype
foxd3 mutant
Wild type
i
MERSCOPE
NC
UMAP2
UMAP1
differentiating_neurons
endothelial
epidermal1
epidermal2
epidermal-foxi3a
hatching_gland
heart field
heart_mature
lateral_line
lens_placode
leukocyte/macrophage/endoderm
mesoderm/PA
mesoderm_blood_island
myotome/muscle
neural_crest
neural
neural-di/tel
neural-dorsal/SC
neural-dorsal/hb/SC
neural-floorplate/notochord
neural-mb/epi
notochord
optic
otic_placode
periderm/EVL
pharyngeal arch
pronephric
retina_pigmented_epithelium
tailbud-PSM
j
Wild type (4 ss)
Wild type (10-12 ss, 21 ss)
foxd3 mutant (21 ss)

Supplementary Figure 1. Quality control and cell type overview across 10x Multiome, Smart-seq3, and spatial transcriptomics datasets.

(a) Violin plots show uniform quality across samples in the 10x multiome data, indicated by metrics such as the number of detected genes per cell (nFeature_RNA), RNA UMI counts per cell (nCount_RNA), mitochondrial gene percentage (percent.mt), ATAC fragment count (nFrags), transcription start site enrichment (TSSEnrichment), and nucleosome ratio (NucleosomeRatio). The upper and lower edges of the boxplots indicate the 25% and 75% quantiles, and the middle line shows the median.

(b) The quality of filtered 10x multiome-ATAC data is demonstrated by the log10 of unique fragments and TSS enrichment, with optimal samples densely distributed in the second quadrant. Dashed lines represent quality thresholds.

(c) UMAP plots show that cell type annotations are reliably reflected by both multiome-RNA and multiome-ATAC datasets.

(d) Violin plots show consistent quality across different developmental stages in the Smart-seq3 data.

(e) *t*-SNE plots visualizing the results of computational demultiplexing, with cells color-coded according to ground-truth genotypes, logistic regression model predictions, and KNN refinement.

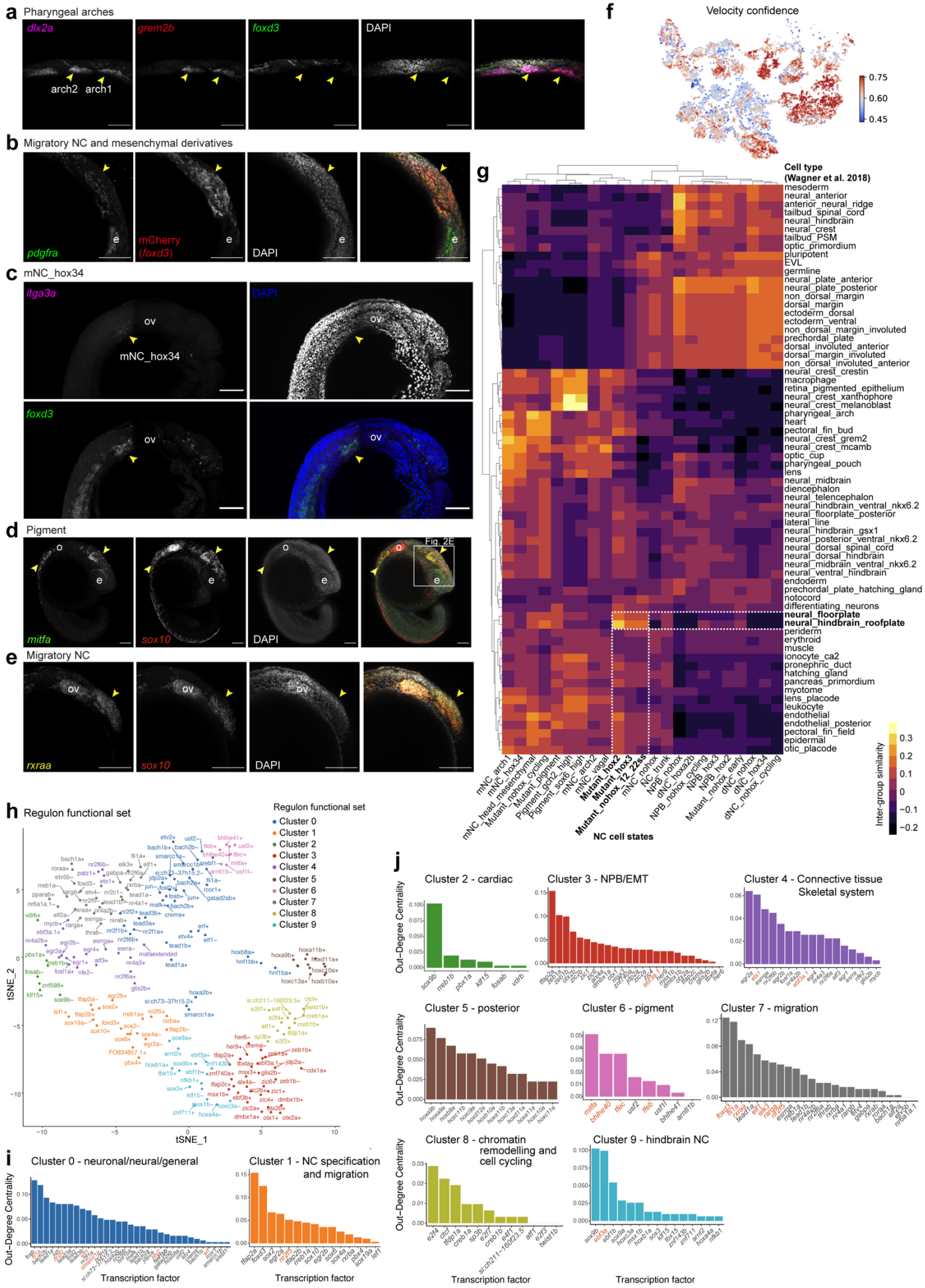
(f) HOX gene family expression patterns, visualized in the dotplot, facilitate the axial-level cell state annotation.

(g) *t*-SNE plot from Smart-seq3 data reproduced distinct cellular patterns between wild-type and *foxd3*-mutant NC cells.

(h) Overlap of cells from eight distinct batches in Smart-seq3 data signifies the absence of batch effects, attesting to the data quality.

(i) Cell type annotations of the MERSCOPE dataset. PA, pharyngeal arch; SC, spinal cord; hb, hindbrain; mb, midbrain; epi, Epiphysis; EVL, enveloping layer; PSM, pre-somite mesoderm.

(j) Sample annotations in the MERSCOPE dataset.


a
Pharyngeal arches
dlx2a
grem2b
foxd3
DAPI
arch2
arch1
b
Migratory NC and mesenchymal derivatives
pdgfra
mCherry (foxd3)
DAPI
e
c
mNC_hox34
itga3a
DAPI
ov
mNC_hox34
foxd3
d
Pigment
mitfa
sox10
DAPI
Fig. 2E
e
Migratory NC
rxraa
sox10
DAPI
f
Velocity confidence
0.75
0.60
0.45
g
Cell type (Wagner et al. 2018)
mesoderm
neural_anterior
anterior_neural_ridge
tailbud_spinal_cord
neural_hindbrain
neural_crest
tailbud_PSM
optic_primordium
pluripotent
EVL
germline
neural_plate_anterior
neural_plate_posterior
non_dorsal_margin
dorsal_margin
ectoderm_dorsal
ectoderm_ventral
non_dorsal_margin_involuted
prechordal_plate
dorsal_involuted_anterior
dorsal_margin_involuted
non_dorsal_involuted_anterior
neural_crest_crestin
macrophage
retina_pigmented_epithelium
neural_crest_xanthophore
neural_crest_melanoblast
pharyngeal_arch
heart
pectoral_fin_bud
neural_crest_grem2
neural_crest_mcamb
optic_cup
pharyngeal_pouch
lens
neural_midbrain
diencephalon
neural_telencephalon
neural_hindbrain_ventral_nkx6.2
neural_floorplate_posterior
lateral_line
neural_hindbrain_gsx1
neural_posterior_ventral_nkx6.2
neural_dorsal_spinal_cord
neural_dorsal_hindbrain
neural_midbrain_ventral_nkx6.2
neural_ventral_hindbrain
endoderm
prechordal_plate_hatching_gland
notocord
differentiating_neurons
neural_floorplate
neural_hindbrain_roofplate
periderm
erythroid
muscle
ionocyte_ca2
pronephric_duct
hatching_gland
pancreas_primordium
myotome
lens_placode
leukocyte
endothelial
endothelial_posterior
pectoral_fin_field
epidermal
otic_placode
mNC_arch1
mNC_hox34
mNC_head_mesenchymal
Mutant_nohox_cycling
Mutant_pigment
Pigment_gch2_high
Pigment_sox6_high
mNC_arch2
mNC_vagal
Mutant_hox2
Mutant_hox3
Mutant_nohox_12_22ss
mNC_nohox
NC_trunk
NPB_nohox
dNC_hox2b
NPB_nohox_cycling
NPB_hox3
NPB_hox2
Mutant_nohox_early
dNC_nohox
dNC_hox4
dNC_nohox_cycling
NC cell states
Inter-group similarity
0.3
0.2
0.1
0
−0.1
−0.2
h
Regulon functional set
Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
tSNE_1
tSNE_2
i
Cluster 0 - neuronal/neural/general
Cluster 1 - NC specification and migration
Out−Degree Centrality
Transcription factor
j
Cluster 2 - cardiac
Cluster 3 - NPB/EMT
Cluster 4 - Connective tissue Skeletal system
Cluster 5 - posterior
Cluster 6 - pigment
Cluster 7 - migration
Cluster 8 - chromatin remodelling and cell cycling
Cluster 9 - hindbrain NC
Out−Degree Centrality
Transcription factor

Supplementary Figure 2. Cell state validation, marker gene expression, and systematic functional assessment of regulons in NC development.

(a) Pharyngeal arch 1 and arch 2 regions are marked by *dlx2a* and *grem2b* expression in a 21-ss wild-type embryo. *dlx2a*, *grem2b* - arch 1 and arch 2.

(b) Overlapping expression of *pdgfra* and *foxd3*-mCherry indicates cranial mesenchymal cells. *pdgfra* - migratory NC and mesenchymal derivatives.

(c) Multiplexed HCR of *itga3a* and *foxd3* validates the mNC_hox34 cluster.

(d) Cranial and vagal pigment-lineage derivatives, marked by yellow arrows, are confirmed by *sox10* and *mitfa* expression patterns in a 21-ss wild-type embryo. The multiome-RNA NC dataset identifies mitfa as the definitive pigment lineage marker.

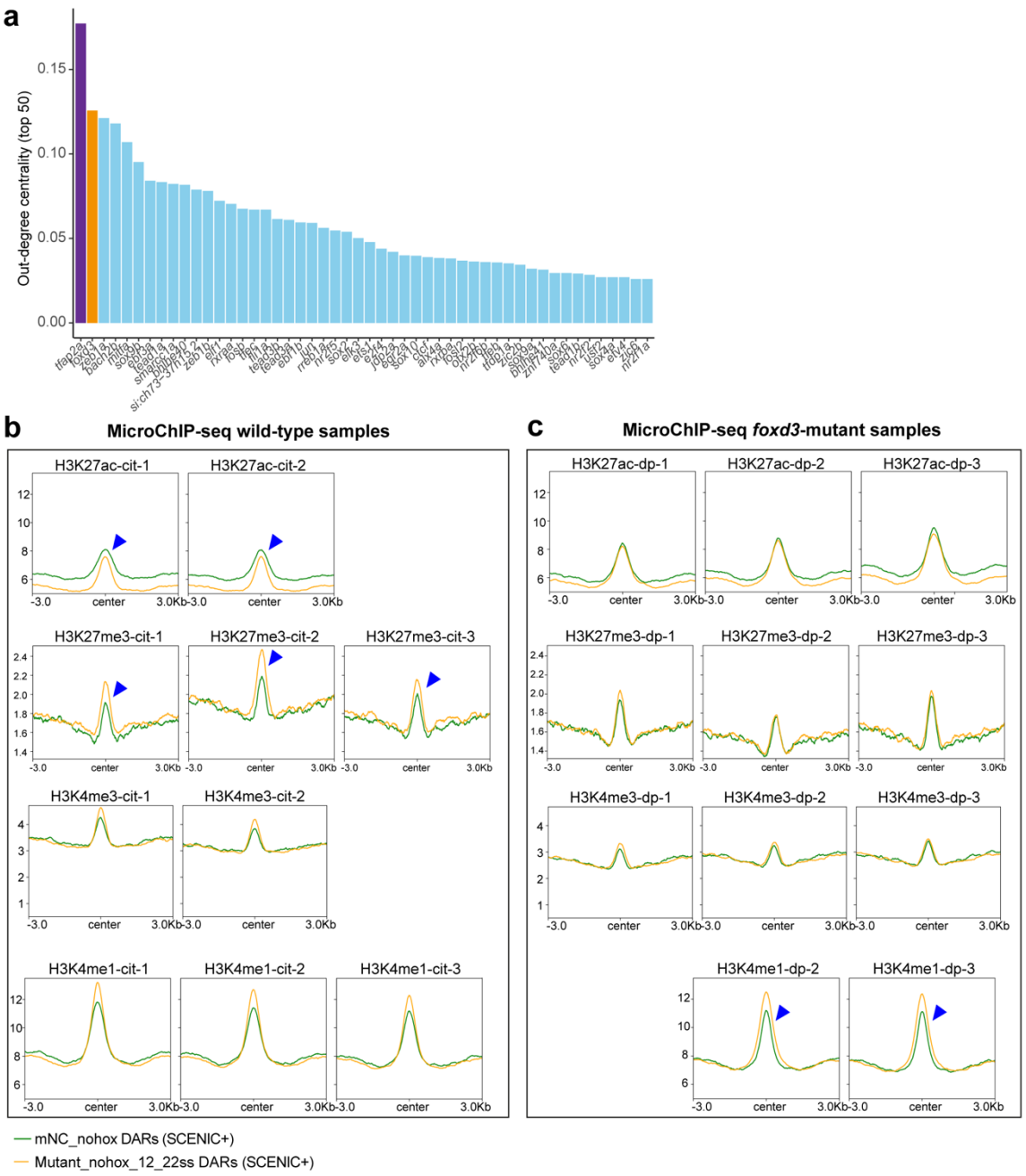
(e) Migratory NC cells are confirmed in hindbrain regions using HCR of *rxraa* and *sox10* in a 21-ss wild-type embryo. *rxraa* - migratory NC and derivatives. Labels: eye (e), otic vesicle (o), with a 100 µm scale bar. Arrows indicate regions of interest.

(f) *t*-SNE plot shows the velocity confidence of wild-type NC cells.

(g) The inter-group similarity heatmap computed by CIDER shows the resemblance of *foxd3*-mutant NC cells and floorplate or roofplate cells in terms of differential gene expression patterns.

(h) *t*-SNE plot exhibits the results from agglomerative clustering of regulon region-based activities, with regulons color-coded by functional sets.

(i-j) Barplots show the rankings of TFs based on out-degree centrality for each regulon cluster (Clusters 0-1, panel i; Clusters 2-9, panel j), highlighting those TFs selected for in vivo Perturb-seq validation.

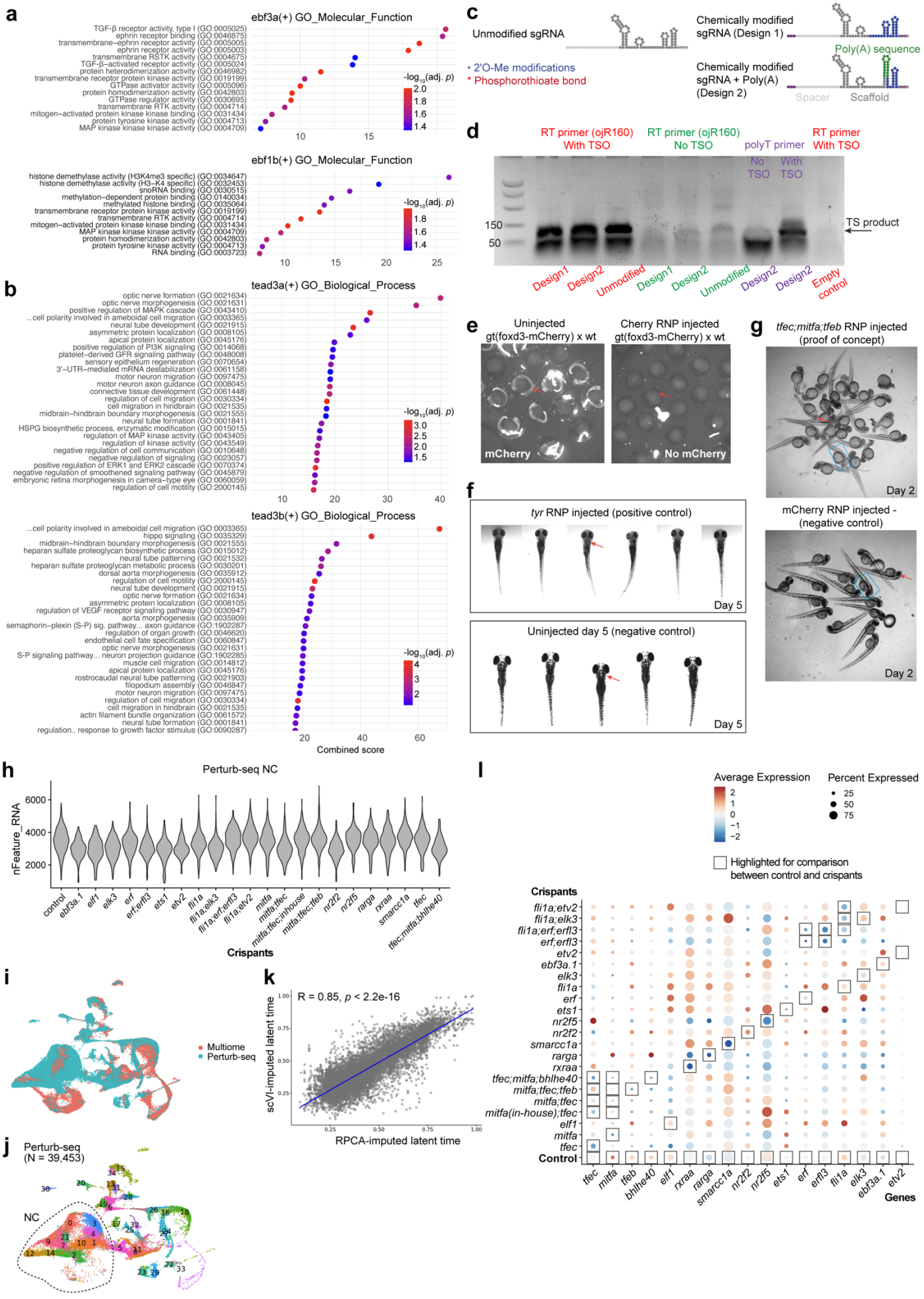


Supplementary Figure 3. Key TF effectors and histone modification in candidate enhancers of SCENIC+-predicted NC-GRNs

(a) Out-degree centrality analysis identifies *tfap2a* and *foxd3* as pivotal regulators in the NC-GRN.

(b-c) MicroChIP-seq profiles of enhancer candidates (differentially accessible regions, DARs) for mNC_nohox and Mutant_nohox_12_22ss reveal that mNC enhancer candidates are enriched with the H3K27ac modification (active enhancers) in *foxd3*-expressing wild-type migratory cells. On the other

hand, the mutant-specific enhancer candidates are H3K27me3-rich (repressive) in *foxd3*-expressing wild-type migratory cells.


a
ebf3a(+) GO_Molecular_Function
TGF-β receptor activity, type I (GO:0005025)
ephrin receptor binding (GO:0046875)
transmembrane−ephrin receptor activity (GO:0005005)
ephrin receptor activity (GO:0005003)
transmembrane RSTK activity (GO:0004675)
TGF−β−activated receptor activity (GO:0005024)
protein heterodimerization activity (GO:0046982)
transmembrane receptor protein kinase activity (GO:0019199)
GTPase activator activity (GO:0005096)
protein homodimerization activity (GO:0042803)
GTPase regulator activity (GO:0030695)
transmembrane RTK activity (GO:0004714)
mitogen−activated protein kinase kinase binding (GO:0031434)
protein tyrosine kinase activity (GO:0004713)
MAP kinase kinase kinase activity (GO:0004709)
-log10(adj. p)
2.0
1.8
1.6
1.4
10
15
ebf1b(+) GO_Molecular_Function
histone demethylase activity (H3K4me3 specific) (GO:0034647)
histone demethylase activity (H3−K4 specific) (GO:0032453)
snoRNA binding (GO:0030515)
methylation−dependent protein binding (GO:0140034)
methylated histone binding (GO:0035064)
transmembrane receptor protein kinase activity (GO:0019199)
transmembrane RTK activity (GO:0004714)
mitogen−activated protein kinase kinase binding (GO:0031434)
MAP kinase kinase kinase activity (GO:0004709)
protein homodimerization activity (GO:0042803)
protein tyrosine kinase activity (GO:0004713)
RNA binding (GO:0003723)
-log10(adj. p)
1.8
1.6
1.4
10
15
20
25
b
tead3a(+) GO_Biological_Process
optic nerve formation (GO:0021634)
optic nerve morphogenesis (GO:0021631)
positive regulation of MAPK cascade (GO:0043410)
...cell polarity involved in ameboidal cell migration (GO:0003365)
neural tube development (GO:0021915)
asymmetric protein localization (GO:0008105)
apical protein localization (GO:0045176)
positive regulation of PI3K signaling (GO:0014068)
platelet−derived GFR signaling pathway (GO:0048008)
sensory epithelium regeneration (GO:0070654)
3'−UTR−mediated mRNA destabilization (GO:0061158)
motor neuron migration (GO:0097475)
motor neuron axon guidance (GO:0008045)
connective tissue development (GO:0061448)
regulation of cell migration (GO:0030334)
cell migration in hindbrain (GO:0021535)
midbrain−hindbrain boundary morphogenesis (GO:0021555)
neural tube formation (GO:0001841)
HSPG biosynthetic process, enzymatic modification (GO:0015015)
regulation of MAP kinase activity (GO:0043405)
regulation of kinase activity (GO:0043549)
negative regulation of cell communication (GO:0010648)
negative regulation of signaling (GO:0023057)
positive regulation of ERK1 and ERK2 cascade (GO:0070374)
negative regulation of smoothened signaling pathway (GO:0045879)
embryonic retina morphogenesis in camera−type eye (GO:0060059)
regulation of cell motility (GO:2000145)
-log10(adj. p)
3.0
2.5
2.0
1.5
10
20
30
40
tead3b(+) GO_Biological_Process
...cell polarity involved in ameboidal cell migration (GO:0003365)
hippo signaling (GO:0035329)
midbrain−hindbrain boundary morphogenesis (GO:0021555)
heparan sulfate proteoglycan biosynthetic process (GO:0015012)
neural tube patterning (GO:0021532)
heparan sulfate proteoglycan metabolic process (GO:0030201)
dorsal aorta morphogenesis (GO:0035912)
regulation of cell motility (GO:2000145)
neural tube development (GO:0021915)
optic nerve formation (GO:0021634)
asymmetric protein localization (GO:0008105)
regulation of VEGF receptor signaling pathway (GO:0030947)
aorta morphogenesis (GO:0035909)
semaphorin−plexin (S-P) sig. pathway... axon guidance (GO:1902287)
regulation of organ growth (GO:0046620)
endothelial cell fate specification (GO:0060847)
optic nerve morphogenesis (GO:0021631)
S-P signaling pathway... neuron projection guidance (GO:1902285)
muscle cell migration (GO:0014812)
apical protein localization (GO:0045176)
rostrocaudal neural tube patterning (GO:0021903)
filopodium assembly (GO:0046847)
motor neuron migration (GO:0097475)
regulation of cell migration (GO:0030334)
cell migration in hindbrain (GO:0021535)
actin filament bundle organization (GO:0061572)
neural tube formation (GO:0001841)
regulation.. response to growth factor stimulus (GO:0090287)
-log10(adj. p)
4
3
2
20
40
60
Combined score
c
Unmodified sgRNA
Chemically modified sgRNA (Design 1)
Poly(A) sequence
• 2'O-Me modifications
* Phosphorothioate bond
Chemically modified sgRNA + Poly(A) (Design 2)
Spacer
Scaffold
d
RT primer (ojR160) With TSO
RT primer (ojR160) No TSO
polyT primer
No TSO
With TSO
RT primer With TSO
150
50
TS product
Design1
Design2
Unmodified
Design1
Design2
Unmodified
Design2
Design2
Empty control
e
Uninjected gt(foxd3-mCherry) x wt
Cherry RNP injected gt(foxd3-mCherry) x wt
mCherry
No mCherry
f
tyr RNP injected (positive control)
Day 5
Uninjected day 5 (negative control)
Day 5
g
tfec;mitfa;tfeb RNP injected (proof of concept)
Day 2
mCherry RNP injected - (negative control)
Day 2
h
Perturb-seq NC
nFeature_RNA
6000
4000
2000
control
ebf3a.1
elf1
elk3
erf
erf;erfl3
ets1
etv2
fli1a
fli1a;elk3
fli1a;erf;erfl3
fli1a;etv2
mitfa
mitfa;tfec
mitfa;tfec;inhouse
mitfa;tfec;tfeb
nr2f2
nr2f5
rarga
rxraa
smarcc1a
tfec
tfec;mitfa;bhlhe40
Crispants
i
Multiome
Perturb-seq
k
R = 0.85, p < 2.2e-16
scVI-imputed latent time
RPCA-imputed latent time
j
Perturb-seq (N = 39,453)
NC
l
Average Expression
2
1
0
−1
−2
Percent Expressed
25
50
75
Highlighted for comparison between control and crispants
Crispants
fli1a;etv2
fli1a;elk3
fli1a;erf;erfl3
erf;erfl3
etv2
ebf3a.1
elk3
fli1a
erf
ets1
nr2f5
nr2f2
smarcc1a
rarga
rxraa
tfec;mitfa;bhlhe40
mitfa;tfec;tfeb
mitfa;tfec
mitfa(in-house);tfec
elf1
mitfa
tfec
Control
tfec
mitfa
tfeb
bhlhe40
elf1
rxraa
rarga
smarcc1a
nr2f2
nr2f5
ets1
erf
erfl3
fli1a
elk3
ebf3a.1
etv2
Genes

Supplementary Figure 4. sgRNA design and knockout efficacy of chemically modified sgRNAs and Perturb-seq data synopsis.

(a-b) Gene ontology enrichment analysis results for target genes of EBF (a) and TEAD regulons (b). Each regulon is specified in the subtitle.

(c) Schematic representations of unmodified, chemically modified, and chemically modified plus poly(A) sgRNA structures.

(d) Gel electrophoresis confirms that chemical modifications do not impair the template switching (TS) reaction in the scRNA-seq protocol.

(e) Contrast images validate the elimination of mCherry fluorescence (indicated by red arrows) in gt(foxd3-mCherry) x wild-type embryos following RNP injection with three design-1 sgRNAs targeting the mCherry sequence.

(f) Contrast images exhibit reproducible phenotypes in *tyr* F_0 crispants on day 5 post injection, with red arrows highlighting the phenotypes.

(g) Contrast images illustrate the albino phenotype in F_0 crispants on day 2 following the triple knockout of *tfec;mitfa;tfeb*, with blue boxes denoting absent pigment cells and red arrows indicating missing eye pigment.

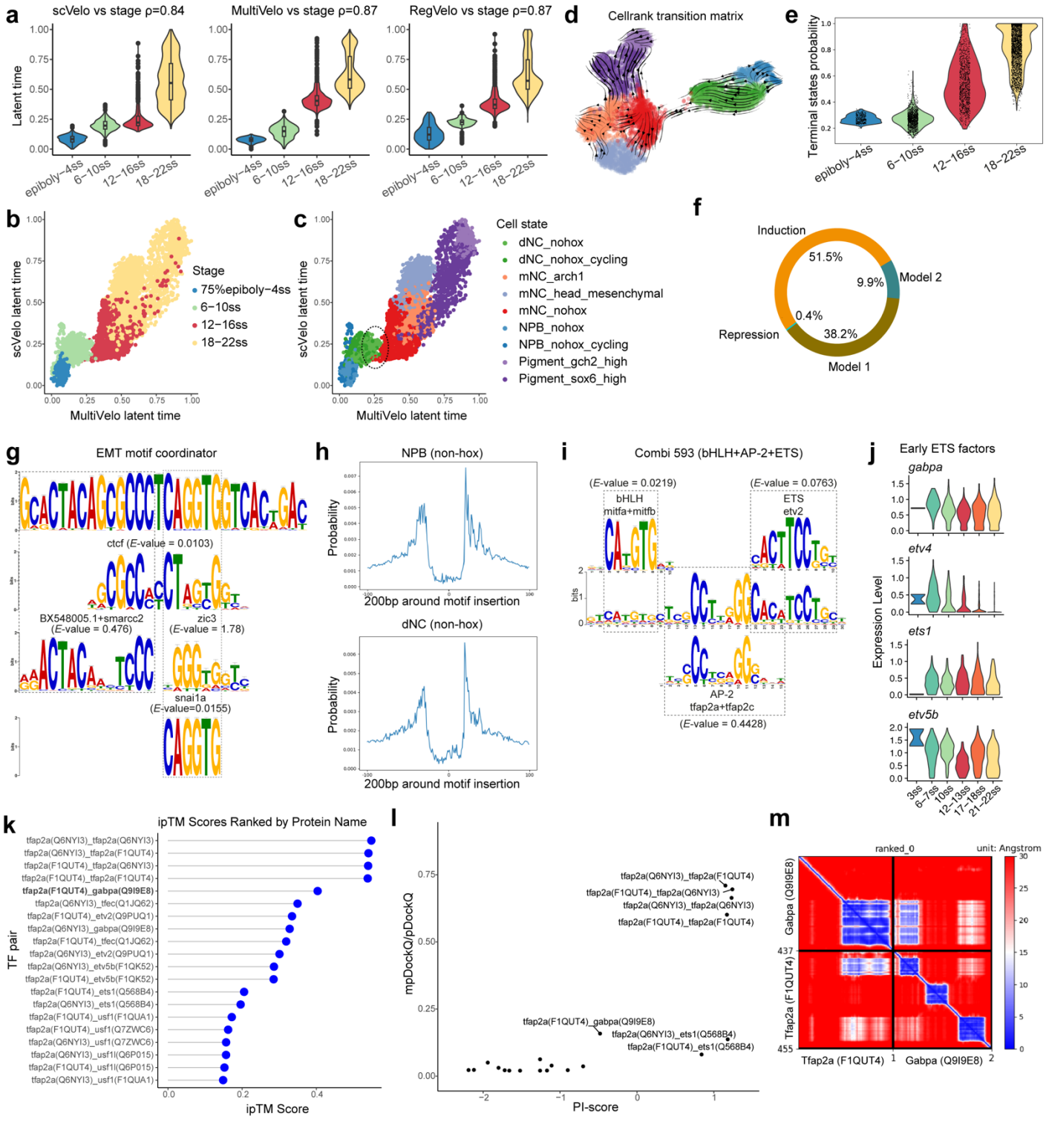
(h) Violin plot displays uniform gene coverage across knockout conditions.

(i) UMAP visualization presents the integrated landscape of multiome-RNA and Perturb-seq data.

(j) Perturb-seq data clusters are depicted within the same integrated space. The NC clusters were identified based on multiome cell annotations.

(k) Scatterplot confirms significant correlation between latent time predictions using RPCA and scVI.

(l) Dotplot shows gene expression levels for targeted genes in non-targeting control versus knockout conditions.


a
scVelo vs stage ρ=0.84
MultiVelo vs stage ρ=0.87
RegVelo vs stage ρ=0.87
Latent time
epiboly−4ss
6−10ss
12−16ss
18−22ss
d
Cellrank transition matrix
e
Terminal states probability
b
scVelo latent time
MultiVelo latent time
Stage
75%epiboly−4ss
6−10ss
12−16ss
18−22ss
c
Cell state
dNC_nohox
dNC_nohox_cycling
mNC_arch1
mNC_head_mesenchymal
mNC_nohox
NPB_nohox
NPB_nohox_cycling
Pigment_gch2_high
Pigment_sox6_high
f
Induction
51.5%
9.9%
Model 2
0.4%
Repression
38.2%
Model 1
g
EMT motif coordinator
ctcf (E-value = 0.0103)
BX548005.1+smarcc2
(E-value = 0.476)
zic3
(E-value = 1.78)
snai1a
(E-value=0.0155)
h
NPB (non-hox)
Probability
200bp around motif insertion
dNC (non-hox)
i
Combi 593 (bHLH+AP-2+ETS)
(E-value = 0.0219)
bHLH
mitfa+mitfb
(E-value = 0.0763)
ETS
etv2
AP-2
tfap2a+tfap2c
(E-value = 0.4428)
bits
j
Early ETS factors
gabpa
etv4
ets1
etv5b
Expression Level
3ss
6−7ss
10ss
12−13ss
17−18ss
21−22ss
k
ipTM Scores Ranked by Protein Name
TF pair
tfap2a(Q6NYI3)_tfap2a(Q6NYI3)
tfap2a(Q6NYI3)_tfap2a(F1QUT4)
tfap2a(F1QUT4)_tfap2a(Q6NYI3)
tfap2a(F1QUT4)_tfap2a(F1QUT4)
tfap2a(F1QUT4)_gabpa(Q9I9E8)
tfap2a(Q6NYI3)_tfec(Q1JQ62)
tfap2a(F1QUT4)_etv2(Q9PUQ1)
tfap2a(Q6NYI3)_gabpa(Q9I9E8)
tfap2a(F1QUT4)_tfec(Q1JQ62)
tfap2a(Q6NYI3)_etv2(Q9PUQ1)
tfap2a(Q6NYI3)_etv5b(F1QK52)
tfap2a(F1QUT4)_etv5b(F1QK52)
tfap2a(F1QUT4)_ets1(Q568B4)
tfap2a(Q6NYI3)_ets1(Q568B4)
tfap2a(F1QUT4)_usf1(F1QUA1)
tfap2a(F1QUT4)_usf1(Q7ZWC6)
tfap2a(Q6NYI3)_usf1(Q7ZWC6)
tfap2a(Q6NYI3)_usf1l(Q6P015)
tfap2a(F1QUT4)_usf1l(Q6P015)
tfap2a(Q6NYI3)_usf1(F1QUA1)
ipTM Score
l
mpDockQ/pDockQ
PI-score
tfap2a(Q6NYI3)_tfap2a(F1QUT4)
tfap2a(F1QUT4)_tfap2a(Q6NYI3)
tfap2a(Q6NYI3)_tfap2a(Q6NYI3)
tfap2a(F1QUT4)_tfap2a(F1QUT4)
tfap2a(F1QUT4)_gabpa(Q9I9E8)
tfap2a(Q6NYI3)_ets1(Q568B4)
tfap2a(F1QUT4)_ets1(Q568B4)
m
ranked_0
unit: Angstrom
Gabpa (Q9I9E8)
Tfap2a (F1QUT4)
437
455
1
2

Supplementary Figure 5. Multiomic velocity analysis and benchmarking against RNA velocity in wild-type non-hox NC trajectories and identification of potential early ETS factors co-binding with Tfap2a.

(a) Relationship between chronological development stages and latent time predicted by scVelo, MultiVelo, and RegVelo, revealing that scVelo does not differentiate between dNC (6-10 ss) and mNC (12-16 ss) cell states, although all latent time prediction results show a positive association. The Spearman correlation coefficients (ρ) are indicated.

(b-c) Comparative evaluation of scVelo and MultiVelo latent times, with cells distinguished by developmental stages (b) and cell states (c). The dashed circle indicates the overlap between NPB and dNC in RNA-velocity-based latent time.

(d) Vector field depicts the CellRank transition matrix for wild-type non-hox NC trajectories, with cells colored by cell states.

(e) Violin plot demonstrates the relationship between chronological stages and CellRank terminal states probability, which is positively associated with later developmental stages.

(f) Pie plot showing the proportions of genes following specific transcriptional regulatory models (indicated by colors). Models 1 and 2 represent that chromatin closing happens before, or after, transcriptional repression.

(g) EMT motif coordinator showing presumable combinations of CTCF, *smarcc2*, *zic3*, *snai1a*-like motifs.

(h) Footprinting analysis traces of the EMT motif coordinator in NPB (non-hox) and dNC (non-hox) clusters.

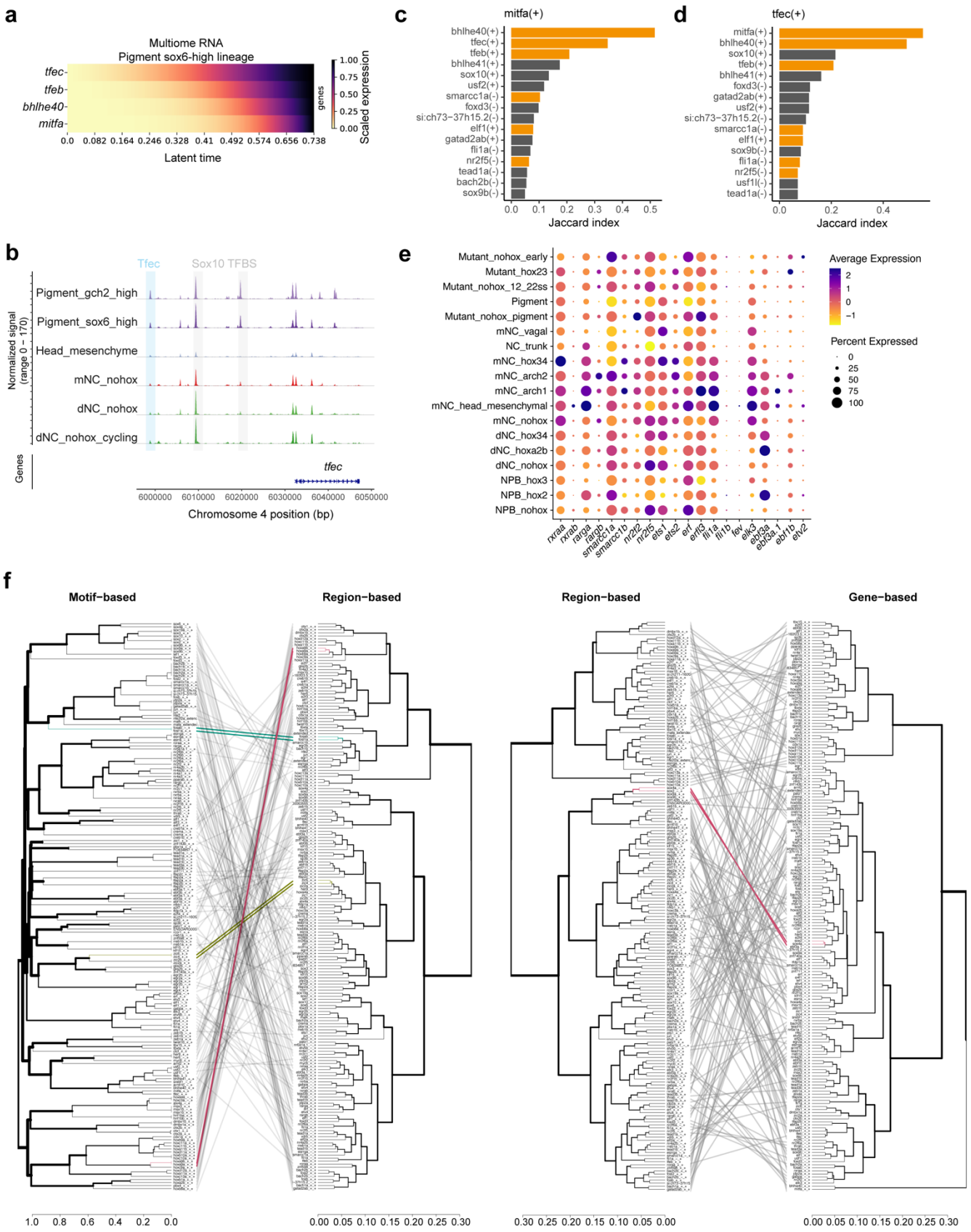
(i) Analysis of putative Coordinator motifs of bHLH, AP-2, ETS with motifs compared against DANIO-CODE.

(j) Violin plots illustrating early expression of four ETS factors.

(k) IpTM scores from AlphaPullDown prediction results involving four early ETS factors, early bHLH factors, and Tfap2a.

(l) Results of TF co-binding predictions displaying PI-score and mpDockQ/pDockQ. High-scoring TF pairs are labeled.

(m) PAE plots illustrating presumable co-binding between Tfap2a and Gabpa.


a
Multiome RNA
Pigment sox6-high lineage
tfec-
tfeb-
bhlhe40-
mitfa-
genes
Scaled expression
1.00
0.75
0.50
0.25
0.00
0.0 0.082 0.164 0.246 0.328 0.41 0.492 0.574 0.656 0.738
Latent time
c
mitfa(+)
bhlhe40(+)
tfec(+)
tfeb(+)
bhlhe41(+)
sox10(+)
usf2(+)
smarcc1a(-)
foxd3(-)
si:ch73−37h15.2(-)
elf1(+)
gatad2ab(+)
fli1a(-)
nr2f5(-)
tead1a(-)
bach2b(-)
sox9b(-)
0.0 0.1 0.2 0.3 0.4 0.5
Jaccard index
d
tfec(+)
mitfa(+)
bhlhe40(+)
sox10(+)
tfeb(+)
bhlhe41(+)
foxd3(-)
gatad2ab(+)
usf2(+)
si:ch73−37h15.2(-)
smarcc1a(-)
elf1(+)
sox9b(-)
fli1a(-)
nr2f5(-)
usf1l(-)
tead1a(-)
0.0 0.2 0.4
Jaccard index
b
Tfec
Sox10 TFBS
Pigment_gch2_high
Pigment_sox6_high
Head_mesenchyme
mNC_nohox
dNC_nohox
dNC_nohox_cycling
Normalized signal
(range 0 – 170)
Genes
tfec
6000000 6010000 6020000 6030000 6040000 6050000
Chromosome 4 position (bp)
e
Mutant_nohox_early
Mutant_hox23
Mutant_nohox_12_22ss
Pigment
Mutant_nohox_pigment
mNC_vagal
NC_trunk
mNC_hox34
mNC_arch2
mNC_arch1
mNC_head_mesenchymal
mNC_nohox
dNC_hox34
dNC_hoxa2b
dNC_nohox
NPB_hox3
NPB_hox2
NPB_nohox
rxraa rxrab rarga rargb smarcc1a smarcc1b nr2f2 nr2f5 ets1 ets2 erf erfl3 fli1a fli1b fev elk3 ebf3a ebf3a.1 ebf1b etv2
Average Expression
2
1
0
−1
Percent Expressed
0
25
50
75
100
f
Motif−based
Region−based
Region−based
Gene−based
1.0 0.8 0.6 0.4 0.2 0.0
0.00 0.05 0.10 0.15 0.20 0.25 0.30
0.30 0.25 0.20 0.15 0.10 0.05 0.00
0.00 0.05 0.10 0.15 0.20 0.25 0.30

Supplementary Figure 6. *tfec* is an initial marker in pigment lineage development.

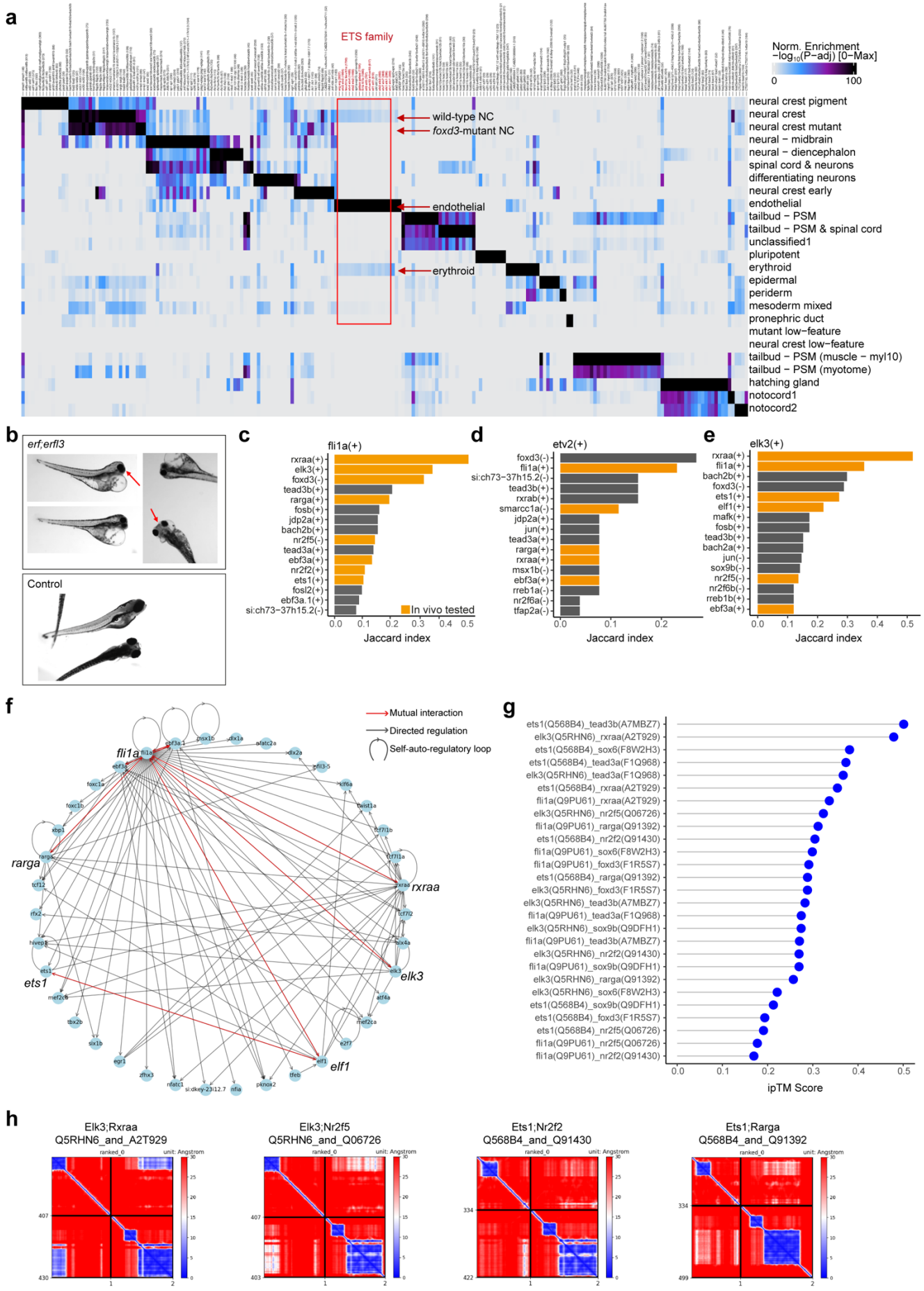
(a) Expression patterns of representative bHLH factors in sox6-high pigment lineage within multiome-RNA data.

(b) The enhancer candidates upstream of *tfec* contain transcription factor binding sites (TFBSs) for Sox10 and Tfec itself.

(c-d) Bar plots present the regulons with the highest Jaccard index, indicative of shared target genes with mitfa(+) (c) and tfec(+) (d), implying potential redundancy among *tfec, mitfa, bhlhe40*, and *tfeb*. Regulons targeted in Perturb-seq are distinguished by orange bars.

(e) Expression pattern of targeted genes and their ohnologs in Smart-seq3 data. Highly expressed ohnologs of the targeted genes are selected for consideration in double knockout strategies, given the potential functional redundancies. For example, the ohnologs *erf* and *erfl3* were designed for double knockout. The ohnologs of *fli1a*, i.e., *fli1b* and *fev*, were not expressed and excluded from knockout experiments.

(f) Comparative analysis of similarity metrics computed from motifs, region-based regulon dynamics, and gene-based regulon dynamics. The dendrograms of regulons were constructed based on one of the similarity metrics (referenced in subtitles), with gray lines illustrating connections between identical regulons across two dendrograms. Colored lines underscore the shared nearest neighbors between the two dendrograms.


a
ETS family
Norm. Enrichment
−log10(P−adj) [0−Max]
0
100
wild-type NC
foxd3-mutant NC
endothelial
erythroid
neural crest pigment
neural crest
neural crest mutant
neural − midbrain
neural − diencephalon
spinal cord & neurons
differentiating neurons
neural crest early
endothelial
tailbud − PSM
tailbud − PSM & spinal cord
unclassified1
pluripotent
erythroid
epidermal
periderm
mesoderm mixed
pronephric duct
mutant low−feature
neural crest low−feature
tailbud − PSM (muscle − myl10)
tailbud − PSM (myotome)
hatching gland
notocord1
notocord2
b
erf;erfl3
Control
c
fli1a(+)
rxraa(+)
elk3(+)
foxd3(-)
tead3b(+)
rarga(+)
fosb(+)
jdp2a(+)
bach2b(+)
nr2f5(-)
tead3a(+)
ebf3a(+)
nr2f2(+)
ets1(+)
fosl2(+)
ebf3a.1(+)
si:ch73−37h15.2(-)
In vivo tested
0.0 0.1 0.2 0.3 0.4 0.5
Jaccard index
d
etv2(+)
foxd3(-)
fli1a(+)
si:ch73−37h15.2(-)
tead3b(+)
rxrab(+)
smarcc1a(-)
jdp2a(+)
jun(+)
tead3a(+)
rarga(+)
rxraa(+)
msx1b(-)
ebf3a(+)
rreb1a(-)
nr2f6a(-)
tfap2a(-)
0.0 0.1 0.2
Jaccard index
e
elk3(+)
rxraa(+)
fli1a(+)
bach2b(+)
foxd3(-)
ets1(+)
elf1(+)
mafk(+)
fosb(+)
tead3b(+)
bach2a(+)
jun(-)
sox9b(-)
nr2f5(-)
nr2f6b(-)
rreb1b(+)
ebf3a(+)
0.0 0.1 0.2 0.3 0.4 0.5
Jaccard index
f
Mutual interaction
Directed regulation
Self-auto-regulatory loop
fli1a
rarga
rxraa
elk3
ets1
elf1
g
ets1(Q568B4)_tead3b(A7MBZ7)
elk3(Q5RHN6)_rxraa(A2T929)
ets1(Q568B4)_sox6(F8W2H3)
ets1(Q568B4)_tead3a(F1Q968)
elk3(Q5RHN6)_tead3a(F1Q968)
ets1(Q568B4)_rxraa(A2T929)
fli1a(Q9PU61)_rxraa(A2T929)
elk3(Q5RHN6)_nr2f5(Q06726)
fli1a(Q9PU61)_rarga(Q91392)
ets1(Q568B4)_nr2f2(Q91430)
fli1a(Q9PU61)_sox6(F8W2H3)
fli1a(Q9PU61)_foxd3(F1R5S7)
ets1(Q568B4)_rarga(Q91392)
elk3(Q5RHN6)_foxd3(F1R5S7)
elk3(Q5RHN6)_tead3b(A7MBZ7)
fli1a(Q9PU61)_tead3a(F1Q968)
elk3(Q5RHN6)_sox9b(Q9DFH1)
fli1a(Q9PU61)_tead3b(A7MBZ7)
elk3(Q5RHN6)_nr2f2(Q91430)
fli1a(Q9PU61)_sox9b(Q9DFH1)
elk3(Q5RHN6)_rarga(Q91392)
elk3(Q5RHN6)_sox6(F8W2H3)
ets1(Q568B4)_sox9b(Q9DFH1)
ets1(Q568B4)_foxd3(F1R5S7)
ets1(Q568B4)_nr2f5(Q06726)
fli1a(Q9PU61)_nr2f5(Q06726)
fli1a(Q9PU61)_nr2f2(Q91430)
0.0 0.1 0.2 0.3 0.4 0.5
ipTM Score
h
Elk3;Rxraa
Q5RHN6_and_A2T929
Elk3;Nr2f5
Q5RHN6_and_Q06726
Ets1;Nr2f2
Q568B4_and_Q91430
Ets1;Rarga
Q568B4_and_Q91392
ranked_0
unit: Angstrom

Supplementary Figure 7. Individual and convergent functions of ETS TFs in NC development.

(a) Motif enrichment analysis of multiome-ATAC data indicates that the ETS motifs are significantly enriched in wild-type NC cells and absent in *foxd3* mutants. Interestingly, the same motifs are also enriched in endothelial and erythroid populations in our global multiome-ATAC dataset.

(b) Phenotypic analysis of *erf;erfl3* double knockout embryos indicates significant craniofacial abnormalities (highlighted by red arrows).

(c, d, e) Bar plots display notable functional redundancy among ETS factors (*etv2, fli1a, elk3*), as evidenced by their regulons sharing the highest Jaccard Indices. A high degree of target gene overlap with RXR/RAR/NR2F regulons is observed for several ETS transcription factors (*elk3, fli1a, ets1*), suggesting potential synergistic effects.

(f) Egograph from the SCENIC+-predicted GRN for *fli1a* demonstrates self-auto-regulatory loops and mutual regulatory interactions between *fli1a* and *elk3*, indicating intricate regulatory networks among ETS factors.

(g) Results of TF co-binding predictions by AlphaPullDown involving three ETS factors and key mesenchyme/migration-related factors (*rarga, rxraa, foxd3, sox6, sox9b, tead3a, tead3b, nr2f5, nr2f2*).

(h) PAE plots depicting presumable co-binding TFs.

Supplementary Table 1. Sample information of single-cell sequencing and spatial transcriptomics.

Supplementary Table 2. Marker genes of NC cell states.

Supplementary Table 3. Cell-type-specific regulon specificity scores of NC GRN.

Supplementary Table 4. Biological processes identified by the gene ontology enrichment analysis conducted on the target genes of each functional cluster.

Supplementary Table 5. Gene ontology enrichment analysis of BACH, EBF and TEAD regulons.

Supplementary Table 6. RegVelo predictions results for single-cell and two-gene knockouts.

Supplementary Table 7. Primer sequences and sgRNA spacer sequences.