From neural border to migratory stage: A comprehensive single cell roadmap of the timing and regulatory logic driving cranial and vagal neural crest emergence

Aleksandr Kotov[1,2], Mansour Alkobtawi[1,2]†, Subham Seal[1,2]†, Vincent Kappès[1,2], Sofia Medina Ruiz[3], Hugo Arbès[1,2], Richard Harland[3], Leonid Peshkin[4], and Anne H. Monsoro-Burq[1,2,5*]

Affiliations

[1]Université Paris Saclay, Faculté des Sciences d'Orsay, CNRS UMR 3347, INSERM U1021, rue Henri Becquerel, F-91405 Orsay, France.

[2]Institut Curie Research Division, PSL Research University, CNRS UMR 3347, INSERM U1021, Rue Henri Becquerel, F-91405 Orsay, France.

[3] Molecular and Cell Biology Department, Genetics, Genomics and Development Division, University of California Berkeley, CA 94720, USA.

[4]Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

[5]Institut Universitaire de France, F-75005 Paris, France.

* corresponding author: anne-helene.monsoro-burq@curie.fr

Abstract

Neural crest cells exemplify cellular diversification from a multipotent progenitor population. However, the full sequence of molecular choices orchestrating the emergence of neural crest heterogeneity from the embryonic ectoderm remains elusive. Gene-regulatory-networks (GRN) govern early development and cell specification towards definitive neural crest. Here, we combine ultra-dense single cell transcriptomes with machine-learning and large-scale experimental validation to provide a comprehensive GRN underlying neural crest fate diversification from induction to early migration stages. During gastrulation, a transient neural border zone state precedes choice between neural crest and placodes following a "dual convergence model". Transcription factor connectome and bifurcation analyses demonstrate the early emergence of neural crest fates at neural plate stage, alongside an unbiased multipotent lineage persisting until after epithelial-mesenchymal transition. We decipher the circuits driving cranial and vagal neural crest formation and provide a broadly applicable strategy for investigating SC transcriptomes in vertebrate GRNs in development, evolution and disease.

Introduction

Neural crest cells form a population of multipotent and migratory progenitors found in vertebrate embryos, essential for the peripheral and enteric nervous system, craniofacial structures, endocrine and pigment cells among others. Together with ectodermal placodes, neural crest (NC) cells are evolutionary inventions that support many cell and tissue innovations promoting the vertebrate predatory lifestyle. Shortly after gastrulation, in the anterior-most part of the embryo, NC cells are induced from the dorsal-lateral "neural border zone" (NB), an ectoderm domain located between the non-neural ectoderm and the neural plate ectoderm (*1*, *2*). In addition to neural crest, NB territory also gives rise to posterior placodes, non-neural ectoderm and the dorsal part of the neural tube (*3*, *4*). Whether these four cell types arise from a common and multipotent early progenitor state, and how fate decisions are orchestrated at the NB during gastrulation remain poorly understood. During neurulation, NC specification and induction progresses as an anterior-to-posterior wave along the edges of the neural plate, with gene programs that define early and immature neural crest cells (e.g.

expression of *snail2, foxd3* and *sox8* genes) followed by later pre-migratory programs presaging emigration of NC cells from the NB epithelium as the neural folds elevate and close (e.g. expression of *sox10*, *twist1* and *cdh2 (N-cadherin)* genes; activation of specific metabolic and signaling pathways) (*5*, *6*). In addition to this pan-NC program, several regional molecular modules are activated along the anterior-posterior body axis and define sub-populations with specific potential (*7*, *8*). How these programs are interconnected with the pan-NC module, and how and when they are activated in pre-migratory NC cells is poorly described. Later, at the end of neurulation, NC cells leave the dorsal ectoderm by a stereotypical epithelium-to-mesenchymal transition (EMT) followed by extensive migration towards a variety of target tissues, where the NC cells differentiate into more than thirty different cell types, including peripheral and enteric neurons and glia, craniofacial osteocytes, chondrocytes, adipocytes and mesenchyme, chromaffin secretory cells and pigment cells.

Neural crest biology has been scrutinized during development and evolution, leading to the elucidation of elaborate gene regulatory networks (GRNs) during the last decade (*9*, *10*). These networks, however, remain incomplete and do not account for most of the defects observed in human neurocristopathies (*11*). This problem is ripe for single cell (SC) transcriptomics which would enable a full description of NC development, over sequential developmental stages, and in comparison to adjacent tissues (e.g. at the neural border) would define the developmental genetic trajectories of the complete NC lineage tree. Recent SC studies on NC cells have mainly explored NC after emigration, using chick, fish and mouse embryos (*12*) (Supplemental Table 1). In contrast, pre-migratory NC single cells have received limited exploration, mostly around the EMT stage and on small cell numbers at a specific level of the body axis (*7*, *13*). While the formation of the NB territory has been defined by expression of a few genes during gastrulation (e.g. *pax3* and *pax7)* (*2*, *9*, *14*), the timing of NB specification from the rest of the dorsal ectoderm and the circuits driving fate decisions between the four NB-derived cell fates (NC, placodes, non-neural ectoderm and dorsal neural tube) remain to be established (*3*, *15*, *16*). Furthermore, the timing of lineage decisions in the pre-migratory NC along the anterior-posterior axis, the maintenance of a multipotent NC sub-population, and the molecular mechanisms driving each state of the pre-migratory NC lineage tree remain unexplored. Here, we used single cell transcriptomes from eight consecutive developmental stages of *Xenopus tropicalis*, featuring 6135 NC cells and 17138 early ectoderm cells, to provide a comprehensive developmental profiling of the NB and the pre-migratory NC. We have uncovered several new NC sub-populations and highlighted their precise trajectories, resulting in eight NC sub-populations emigrating from anterior to vagal levels of the body axis. Interestingly, we find that some fates emerge much earlier than previously anticipated, that NC diversity is maintained upon EMT and that further diversification occurs at the onset of migration. We propose a temporal sequence of molecular events underlying the successive transcriptomic states and the fate decisions supporting the emergence of the NC cells from the neural border during gastrulation up until early migratory states at early organogenesis stage. Moreover, we identify key transcription factors involved in main lineage branching and validate several regulatory predictions *in vivo*. We therefore provide an extensive gene regulatory network describing the emergence of the neural crest lineage in the ectoderm of vertebrate embryos.

Results

General strategy for identification of neural crest and neural border zone cells.

Using deeper re-sequencing of single cell (SC) series from whole *X. tropicalis* embryos (*17*), taken at 8 consecutive developmental stages, followed by updated genome annotation and alignment, we have scrutinized 17138 early dorsal ectoderm cells and 6135 neural crest (NC) cells from early gastrulation stage to post-EMT early migration stage. The large cell number of our new dataset permits assessment of the cellular diversity in the NC population during early induction (at late gastrulation stages 12-13 and neural plate stages 13-

14), during neural fold elevation (stages 16-18), during EMT (neural tube stages 18-20) and at the earliest stages of NC cells emigration (tailbud stage 22).

During gastrulation (stages 11-13), the neural plate and the neural border zone are induced from the dorsal ectoderm due to BMP antagonists in the midline and Wnt/FGF signaling from the underlying mesoderm (*18*). As early as stage 10.5-11, the neural border initiates a few distinct gene expressions, such as *pax3/7* (*19*). However, a fuller definition of the NB transcriptome is still lacking. Unlike the neural border, the neural crest has long been defined by the expression of a set of typical genes, like *snail2*, *sox8/9/10*, and *foxd3*. Previous RT-qPCR or *in situ* hybridization showed *sox2* (neural), *snai2* and *sox10* (NC) expression at stages 11, 12 and 14 and expression of other NC markers used for cell selection (*sox8*, *sox9, zic1*; Fig. S1). Importantly, we detected the regional expression of such ectoderm genes in SC data at similarly early stages. This sensitivity was essential to capture the emergence of rare gene programs.

We first performed unsupervised clustering of the whole embryo dataset at each stage (a total of 177250 cells) and identified the NC clusters by expression of canonical NC genes, from induction (stage 12) to migration (stage 22) as well-established in the literature (Figs. 1A, S1; Table S2 and Materials and Methods). We next defined the genes most specific for NC and expressed broadly across the NC population, by selecting genes with high NC-specificity scores compared to the rest of the embryo, and low variation within the NC population. This highlighted *tfap2b, c9, c3* and *sox8* as the best pan-NC markers during neurulation, closely followed by *snai2* and *sox10*. Small temporal variations in these genes' expression suggested that early NC is best labeled by *snai2* at stages 13-16, followed by *tfap2b* from stage 14 to 20, or *c3, c9, sox8* between stages 14 to 18 (Figs. 1C, D, S1). However, in addition to being broadly expressed in the NC cells, *c3* and *c9* are also expressed in placodes and some ectoderm cells (Fig. 1D). This points to *tfap2b* as the most specific pan-NC gene shortly prior to EMT, an important notion for future selection of NC reporters for transgenesis.

We validated the NC selection method by designing a LightGBM (*20*) NC classifier. We tested it on another vertebrate organism, zebrafish, to evaluate its accuracy for NC annotation. Our binary classifier determined whether a cell was NC or not, using our annotated frog whole embryo dataset at seven developmental stages (12-22). The algorithm was tested on a zebrafish whole embryo SC dataset (*21*) at stage 14hpf (equivalent to frog stage 22) with a result of 0.95 for AUC score and 0.66 for F1 score (F1 score defines the weighted harmonic mean of the precision and recall, and it is equal to 0.05 for a random model applied to the zebrafish dataset, Fig. S2). This result was striking because it was robust despite significant species-specific variations in the expression of classical NC gene markers between frog and fish, and despite a strong batch effect between the datasets (595 genes from the list of the most important genes for NC classification were missing from the zebrafish dataset). For example, the expression of *c9* and *snai2,* which are pan-NC genes in *Xenopus,* was drastically different in the zebrafish NC dataset. Yet, the classifier efficiently recognized NC cells in the fish neurula. This result confirmed the accuracy and robustness of our NC cells selection criteria.

While the identity of early NC cells has been well described, very few NB-specific markers have been identified (e.g. *pax3/pax7* in frog and chick respectively) (*2*). The NB zone has instead been defined by the *overlapping gradients* of ectodermal expression of ventral-lateral genes (e.g. *msx1, tfap2a, hes4, tfap2a)* and dorsal-lateral ones (e.g. *zic1, zic3)*, and by the early *exclusion* of ventral-most non-neural ectoderm marker expression (NNE, e.g. *dlx3, foxi1e, ker81)* and dorsal-most neural plate ones (*sox2)* (*22*, *23*). We have thus identified NB cells as the group of cells co-expressing *zic1* (dorsal and NB marker) and *tfap2a* (ventral and NB marker) simultaneously at high levels (*19,* see Materials and Methods). Differential

expression analysis with both neural and non-neural progenitors revealed a clear NB gene signature enriched for *sox9, pax3, snai1/2, hes1, gmnn, lhx5* and *axin2* expression. NB zone differential analysis with neural crest and placodes revealed NB-specific expression of *has1, lhx5* and *tfap2c* genes. This indicates that the NB territory, which was so far mainly defined by the overlap between several gene patterns, also presents a specific identity from gastrula stage 11 (Fig. 1B, E). Thus, we have determined a clear signature of developing NC progenitors from stage 11 to stage 22.

Single cell heterogeneity of premigratory cranial and vagal neural crest during neurulation reveals early lineage diversification from an unbiased stem cell population which persists throughout EMT.

Next, we used Leiden clustering (*24*) on the principal component derived neighbor graph to separate the NC cells into distinct states. We defined the optimal number of clusters by manually increasing their number and checking for biological meaning, as revealed by specific gene expression, for example of *hox* genes, such as *hoxd3* for cardiac NC (cluster #12), or *hoxb6* for Enteric Nervous System progenitors (ENSp, cluster #13). We defined 16 clearly distinct states with specific gene signatures, compared to 8 states previously described in (*17*) (Fig. 2A, C, E). Using partition-based graph abstraction (PAGA) analysis (*25*), we described the principle graph of similarities between clusters' transcriptomes (Fig. 2B). This revealed the complexity of interconnections between each cluster from early neurulation to emigration stage.

In this cluster map, we first searched for NC cells emigrating from the neural tube, which are the best characterized so far. We identified 10 cell clusters with an explicit mesenchymal NC signature (#12, #9, #15, #11, #7, #16, #8, #14, #13, #6, all expressing various levels of *itga4/*integrin-a4, *vim/*vimentin and *fn1*/fibronectin, Fig. S4)**.** These clusters mainly included stage 18 to 22 “late” cells (Fig. S4), stages when the cranialmost NC undergoes EMT and early migration. Using their *hox* gene expression signature, we positioned these clusters along the body axis (Fig. 2D). We also identified five cranial NC cell subpopulations devoid of *hox* gene expression. Three were previously found in (*17*): #10 (*rpe65*$^+$*, zfhx4*$^+$*, alx1*$^+$ NC); #11 (*vim*$^+$*, itga4*$^+$ enriched NC) and #15 (*dlx2*$^+$*, efnb2*$^+$*, cyp26c1*$^+$*, epha2*$^+$ NC) (Fig. 3A). Additionally, we uncovered two previously undescribed clusters: #16, a muscle-like NC subpopulation, expressing actin and myosin-like genes (*actnc1*$^+$*, myl1*$^+$ cells) and #14, a cluster of migrating NC cells expressing the canonical early neural crest signature (*tfap2b, c9*) without additional specific gene expression, suggesting that these cells continue to remain unbiased and are potentially multipotent (Fig. 3B-D, F). Finally, we identified three subpopulations of *hox*-positive rhombencephalic neural crest previously grouped together (*17*). These cells form clusters #12, #13 and #9 and express *hoxa3, hoxb3, hoxd3* and are thus located at the vagal level. Cluster #12 includes cardiac NC (*egr2*$^+$*, mafb*$^+$) (*26*) while cluster #13 contains (*tnc*$^+$*, ltbp1*$^+$) ENS progenitors (*27*).

Characterization of a new migratory muscle-like NC cluster.

One key expectation from single cell transcriptomics is to identify and characterize novel rare cell states. Here, cluster #16 describes the emergence of a muscle-like NC subpopulation, previously undescribed at the single cell level and expressing simultaneously late pan-NC markers (*tfap2b, tfap2e, sox10, c9*) and muscle markers *actc1, myl, mylpf, cstb* and *des* (Figs. 2, 3F, S3*). Myl1* is expressed in striated muscle (*28*) and future lineage tracing analysis will determine if these cells generate NC-derived myofibroblasts or striated head muscles. To validate this finding biologically, we have used a well-established assay in *Xenopus laevis* embryos, using pluripotent ectoderm that can be directly reprogrammed into into neural crest progenitors, without any mesodermal contribution (induced neural crest assay, Fig. 3E) (*29*). The explanted ectoderm form spheroids that were cultured until the

equivalent of stage 20-22. We found that *myl1* and *myoD* started being expressed around stage 20 from the induced NC progenitors (Fig. 3F). This finding complements the palette of differentiated phenotypes elicited by this assay, and opens future avenues to study these NC-derived muscle cells.

This muscle-like cluster #16 was the only NC cluster to initiate differentiation. While it has been proposed that the melanocyte lineage was biased prior to NC emigration from the neural tube, we did not see a specific set of cell states that would support pre-migratory melanoblast formation, even rare ones, in our dataset (e.g. no significant detection of *dct, tyr, tyrp1* expression) (*30*). This led us to propose that our global analysis found no other indication of cell fate determination until stage 22 at cranial and vagal levels, in agreement with previous single cell lineage-tracing studies (*31*).

Developmental series analysis reveals that some NC states are generated over long time periods.

The second asset of this dataset is that it follows the cell states over a developmental time series. We could thus assess not only "pseudotime" trajectories (reflecting the heterogeneity of closely related cell transcriptomes at one time point) but also "real time" population dynamics for each cluster in the embryo. Indeed, each cluster is composed of a variable number of cells from different developmental stages, indicating that a given transcriptome state is reached by different cells over time (Fig. S4A). Because each state is characterized by a *hox/non-hox* signature, as an estimate for the cells' position along the body axis, surprisingly these temporal variations do not describe the anterior-posterior wave of NC cell emergence (Fig. 2A). Rather, the range of stages found in each cluster describes the time period over which each regional state of the NC-GRN is generated. For example, cluster #10, which gathers cranialmost NC cells, is generated over the entire duration of neurulation starting at neural plate stage 13, with most cells ranging from mid- to late- neurula stages 14-16-18, and even persisting after EMT stage during stages 20 and 22. This reflects how NC cells of a similar state can be continuously generated over the course of development at a given portion of the neuraxis. In contrast, clusters #11 and #15, two migratory clusters highly expressing *vimentin (vim$^+$),* are generated abruptly at one single post-EMT stage (stage 22) and eventually populate cranial areas as well. These could reflect an abrupt step of mesenchymalisation in the epithelial-to-mesenchymal transition process of cranial NC.

An unbiased NC state coexists with diversifying states upon EMT.

The third major advantage of this dataset is its high cell number, likely to capture most, if not all, cell states during NC formation. This allowed us to find a group of NC cells with immature/pan-NC signature that persists throughout neurulation and EMT. At EMT/migration stage, cluster #14 expresses the canonical early neural crest signature (*tfap2b, c9*), as well as *mmp14* but low *vim, twi1,* and *fn1*. This cluster includes a variety of developmental stages (from stage 13 to 22) and represents a potentially "multipotent" stem-like NC cluster (Figs. 2A, 3B-D). We thus addressed whether all NC cells adopted a similar unique "stem-like" state upon EMT as it was proposed recently (*13*). Instead, we observed that the list of clusters with elevated expression of *vim, fn1*, and *mmp14* possessed distinct features: for example, clusters #10, #13, #15, are all undergoing transition to a mesenchymal *vim*$^+$ state but #10 and #15 are ecto-mesodermal (*twi1*$^+$), while *#13* instead highly expressed *tnc* (Fig. S3B, C). None of these three clusters are similar to stem-like cluster #14. Clearly, the diversity of states is maintained as cells transition from pre-EMT stages to EMT and early migration. In conclusion, by exploring a high-density dataset, we did not see that NC cells adopt a single homogeneous state upon EMT, but rather continue to remain highly diverse.

In summary, our data describes the NC cell populations along a significant segment of the body axis, from the anteriormost cranial NC to the vagal level. We find three important novel points characterizing NC cells around the time of EMT and initiation of migration. Firstly, there is a large diversity of states along the cranio-caudal axis. This diversity comes from cells adopting specific states, such as different cranial populations, multipotent-like state, enteric nervous system progenitors, at different time points during development, with some early-stage cells being intermingled with later ones into the same cluster. Secondly, state diversity is maintained as cells undergo transition from epithelial to mesenchymal phenotypes, while a stem-like subpopulation persists as a distinct cluster. Lastly, we identify a myoblast/myofibroblast-like cell population with strong NC stem-like features as the first one to differentiate upon EMT, but we do not detect indications that other NC derivatives would be set aside from the main premigratory NC population prior to or upon EMT.

A comprehensive lineage tree of cranial and vagal NC cells from induction to migration reveals early developmental heterogeneity dynamics and the maintenance of a stem-like NC population throughout neurulation.

With reference to the post-EMT NC clusters, we were able to define eight progenitor states for NC cells. These delineate developmental dynamics of cranial, vagal and enteric nervous system progenitors. These clusters mainly include cells from the second part of gastrulation (stage 12) to pre-EMT stage (stage 16). Unexpectedly, as early as at gastrula stages 12-13 we detected three NC cell states clearly distinct from the main early NC state: an *hnf1b*$^+$ cluster #3, a *cyp26c1*$^+$ cluster #4 and and *rpe65*$^+$ cluster #10 (Fig. 3A). However, these clusters only comprised small cell populations while most cells in stages 12-13 bear an "unbiased" canonical early NC signature: 85% for stage 12 and 68% for stage 13 (Fig. 3B). At mid-neurula stage 14, two additional trajectories emerge from the unbiased stem-cell like lineage, *nrp2*$^+$ cluster #7 and *mafb*$^+$ cluster #9. Using these specific signatures, we described the main characteristics of progenitors for each cranial to vagal NC subpopulation.

Three early premigratory states emerge at the end of gastrulation.

From the earliest stage of NC induction, three distinct trajectories emerged from the main canonical NC progenitor population (Fig. 2). First, the previously undescribed *hnf1b*$^+$ cluster #3 is a pre-migratory cluster composed mainly of stage 12, 13 and 14 cells expressing pan-NC markers such as *pax3, tfap2b* and *sox9,* as well as the dorsal neural gene *olig4, hnf1b,* the *ngfr*-related gene *nradd*, complement factor *cfb,* the posterior master regulator *cdx4* and a strong posterior hox gene signature (including posterior hox gene *hoxb8* and sacral hox genes *hoxc10/hoxd10*) (Figs. 2D, S3). Interestingly, this cluster seemed to respond to a combination of high WNT/low RA signals, in line with its posterior trunk fate specification (*32*). Indeed, early on, cluster #3 expressed canonical WNT signaling targets *axin2* and *cdx4* (*33*), as well as *dhrs3*, which attenuates RA signaling, is required for posterior axis formation and is usually mutually exclusive with *cyp26c1* expression (Fig. S3B, S5C) (*34*). In cluster #3's older cells, expression of genes responsive to canonical WNT signals, *cdx4* and *sp5,* was decreased, while expression of the non-canonical ligand *wnt11* was increased. Cluster #3's trajectory converged and merged with the enteric NC cells of cluster #13. While most of the enteric peripheral nervous system (ENS) arises from vagal NC, sacral NC cells contribute to cells in the hindgut. Cluster #3 posterior gene signature suggested that as early as mid-gastrulation stage 12, a minor NC progenitor population is already biased towards the future sacral NC fate, although those cells would emerge later on from the neural tube at tailbud-stage and give rise to the posterior ENS (*35–37*). Moreover, *wnt11* negatively regulates enteric NC differentiation (*38*), further suggesting that cluster #3 trajectory may consist of immature ENS progenitors. This cluster is the most posterior cell group found in this dataset (posterior trunk/sacral *hox* signature). Identification of cluster #3 specific markers will enable future lineage tracing

experiments to define the position and developmental contribution of these early progenitors of the enteric nervous system.

The second early-biased cluster, cluster #4, was composed of neural plate stage 13-14 progenitors of cranial NC populations with a unique retinoic acid (RA) signaling-related signature. It merged with cluster #6, (composed of stage 16-20 cells) and cluster #15 (composed of stage 22 cells). Cluster #4 specifically expressed genes encoding the RA-degrading enzyme *cyp26c1,* transcription factors *meis2* and *olig3,* neuropilin receptor *nrp1,* ephrin signaling receptor *epha2,* and ligand *efnb2* (Fig. 5S). Compared to the early unbiased cluster #1, cluster #4 exhibited increased expression of mature pan-NC markers *tfap2b, c9, snai2*, and decreased expression of multipotency marker *pou5f1* (Fig. 3D). Cluster #4 expressed *hoxb1* and *hoxb2* indicating an anterior rhombencephalon position. More precisely *epha2* expression pinned those cells to rhombomere 4 (Fig. 2D) (*39*). High *cyp26c1* expression indicated low RA levels as essential to define rhombomere 4 identity (*40*)**.** Last, *nrp1* specifically marks mouse NC from rhombomere 4 suggesting high conservation of the cluster signatures across vertebrates (*41*). Cluster #4's *cyp26c1/epha2* signature remained specifically expressed in the low-RA-dependent stream (#6, #15) from stage 12 to 22. Expression levels of genes encoding *epha2* and *efnb2* remained high while *cyp26c1* expression gradually decreased in later subpopulation of posterior CNC along with increased expression of genes encoding RA receptors (*rarg, rxrb, ncor1*; Figs. S5A, B) and cellular retinoic acid-binding protein 2 (*crabp2*). This revealed that the identity of rhombomere 4 NC was established as early as neural plate stages 13-14 under low RA signaling (cluster #4) while starting at neural fold stage 16 (cluster #6), later development of this lineage involved increased RA signaling. These observations highlighted the transcriptional outcomes of dynamic RA signaling in the progenitors of rhombencephalic neural crest forming the first (mandibular) and second (hyoid) arch stream respectively, anterior to the otic vesicle.

Appearing at stage 13, the third early-biased cluster is *hox*$^{-}$ cluster #10, expressing *dmbx1, otx1, nrp2, lmx1b, rpe65* and *pax3,* thus containing premigratory early cranial cells (Fig. S3A). Bias towards cluster #10 was slightly noticeable at stage 12, as around 5% of the *unbiased* cluster #1 co-expressed *dmbx1* and *otx1*, which might indicate early fate predetermination. Moreover, 10% of the cluster #*10* were cells of 13-14 stages (Fig. S4A), the rest being later stages. Lxh5 and Dmbx1 are early cranial NC-specific factors and drive expression of *tfap2b* and *sox8* (*10*). Here we found *lhx5* and *dmbx1* expressed separately in early unbiased cluster #1 and in cranial cluster #10 (Fig. S5D), and confirm *otx1* as a marker for anterior neural border (*2*). This means that progenitors for cranial NC start being specified at neural plate stage, much earlier than anticipated.

Two bipotent clusters appear at mid-neurulation stage forming cardiac, enteric and cranial progenitors

Two clusters were composed of progenitors for more than one trajectory ("bipotent clusters"): a partially biased vagal cluster #9 (*mafb*$^{+}$) resolving into both the cardiac NC cluster #12, and the ENSp NC cluster #13, and a cranial cluster #7 (*nrp2*$^{+}$), which is equally linked to the *rpe65*$^{+}$*, zfhx4*$^{+}$ cluster #10 and to the *vim*$^{+}$*, itga4*$^{+}$ enriched cluster #11 (Fig. 2C). From late neural plate stage 14, the unbiased cluster #5, with higher expression of *c9, sox8, tfap2b* in comparison to the earliest unbiased cluster #1, started to split into the vagal-biased #9 (*mafb*$^{+}$, *hoxd3*$^{+}$) and the cranial-biased #7 (*rpe65*$^{+}$*, dlx2*$^{+}$*, hox*$^{-}$) trajectories. Despite the global homogeneity within cluster #5, we detect a slight stratification for expression of neuropilins *nrp1* and *nrp2*, which would later mark NC cells emigrating from rhombomere 4 and from rhombomere 2 respectively (*41*). This suggested that within the unbiased cluster #5, a few cells initiated a trajectory towards cluster #9 (vagal NC, *nrp1, mafb*, *cfi*, *meis2*, *mdk*) while others are shifted towards an anterior hindbrain trajectory (*nrp2*, *id3*, *ifitm3*, *alk;* Fig.

346 S9A, B*)*. This indicated that the SC transcriptomes identified an early bias in "bipotent" cell
347 populations, enabling analysis of gene regulation during specification of each fate.

348 ***The multipotent NC gene signature identifies a separate cell lineage from neural***
349 ***plate stage to the onset of migration.***

350 During the whole process of neurulation, from NC induction to NC migration stage, a
351 significant proportion of the cranial and vagal NC progenitors expressed a canonical NC stem
352 cell-like signature (*sox9, tfap2c, snai2)*, together with high expression of pluripotency markers
353 (*e.g. pou5f1*, *cmyc*, and *sox2).* No indication of bias towards a specific fate/alternative state
354 was detected in these clusters (#1, #5 and #14) (Fig. 3B-D). The signature of these potentially
355 multipotent cells, consisted of transcripts common to all NC cells, but was slightly variable
356 across stages. At induction stages, this cell lineage expressed enriched levels of *zic1, zic3,*
357 *tfap2c, cldn6* and *c3*, while at stages 18-22, it was enriched for *tfap2b, mycn, eef1b2, apoc1,*
358 *vim, rack1, tfap2e* and *c9* expression. About 72% of NC progenitors were found unbiased at
359 induction stages (gastrula and neural plate stages 12-14), 15% at pre-migratory and EMT
360 stages (stages 16-18) and 9% among tailbud stage 22 NC cells. Interestingly, according to
361 PAGA, the latest unbiased population (#14) is transcriptionally equally close to the vagal (#9)
362 and cranial (#7) transitional clusters described below, indicating that some stage 18-20-22
363 cells share the unbiased pan-NC signature irrespective of their cranial or vagal levels of origin.
364 Moreover, the whole unbiased lineage (clusters #5, #14) is mostly composed of premigratory
365 cells with low *vim/fn1* expression and remains *hox* negative, generating NC cells which
366 acquire a *hox* signature upon emigration state (*e.g.* hindbrain cluster #15 expressing *hoxa2*).

367 In sum, this detailed cluster analysis provides the temporal dynamics of gene
368 expression underlying the main transcriptional states defining the various premigratory neural
369 crest trajectories (Figs. 2, S3A). It highlights the early specification of several lineages, along
370 with the maintenance of a NC stem-like population at all axial levels between cranial and
371 vagal body axis regions. This analysis further provides a basis for the in-depth exploration of
372 gene program variations which may drive branching at each bifurcation.

373 **Connectome analysis**

374 ***Connectome predictions using GRNboost2***

375 Using the temporal dynamics of gene expression, several algorithms have been
376 designed to infer genetic co-regulation (*42*). One of them, GRNBoost2, is based on a gradient
377 boosting machine and retrieves the gene regulatory network (GRN) from the expression data-
378 matrix (*43*). Starting from a given list of TFs, this algorithm powerfully links genes with
379 similar expression patterns in SC transcriptomes, providing “syn-expression groups” (*44*) at
380 the single cell level. Thus, GRNBoost2 provides a network of potential bidirectional gene
381 relationships (including genes encoding TFs and other types of proteins). To expand
382 significantly the network of genes involved in the NC/NC progenitor GRNs, we used an input
383 list of 1417 TFs (*45*) and applied GRNboost2 on the NC dataset. After filtration by expression
384 levels and edge weights (predicted importance of the connection), we retrieved 16978 TF-
385 target connections defining a global NC network with a median 22 connections per TF (Fig.
386 4A). To define the main GRN nodes (= genes), we calculated the number of connections per
387 gene (*degree centrality*) and the degree to which a gene lies on paths between other genes
388 (*betweenness centrality*). Further, we determined the main nodes for each developmental stage
389 using nodes maximal expression (Table 1). Among the genes predicted to be important at NC
390 induction stages, we retrieved the previously known early neural border specifiers *pax3* and
391 *zic1*, as well as *zic3, olig4, sox9*, and the anterior NC marker *dmbx1*. For premigratory NC
392 stages, the model predicts prominent functions for the known NC specifiers *tfap2b, sox10 and*
393 *snai2*, the anterior NC markers *rpe65* and *alx1*, and for the hindbrain hox gene *hoxb3*. At the

394 stage of NC cell migration, *tfap2e*, *mycn, dlx2* and *egr2* (*krox20)* displayed high expression
395 and high degree centrality. This analysis thus retrieved several known key regulators in the
396 NC-GRN (*pax3, zic1, sox10, snai2 etc.)*, validating the pertinence of the model, while also
397 predicting strong new candidates (Table 1).

398 ***Large scale experimental validation of GRNboost2 model predictions***

399 GRNboost2 is currently one of the best GRN predictors (*42*). However, benchmarking
400 with experimental datasets is still lacking. In order to validate experimentally the network
401 predicted with GRNboost2 in the TF-towards-gene-target direction on a large scale, we
402 sequenced NB/NC microdissected explants after genetic perturbations (Fig. 4A, B). We
403 depleted selected gene products *in vivo* using previously validated antisense morpholino
404 oligonucleotides (MO) designed against *pax3* or *zic1* (Pax3 MO*,* Zic1 MO) (*9*). At mid-
405 neurula stage, individual wild-type, Pax3-, or Zic1-morphant explants were microdissected
406 from the anterior NB ectoderm (aNB) using sibling embryos and processed for small-RNAseq
407 [as in (*2*)] (Fig. 4B). Differential analysis found decreased expression of 1333 genes in Pax3
408 morphant NB and 1103 genes in Zic1 morphant NB respectively (Figs. 4B, S6). This
409 confirmed that these two NB specifiers are essential to activate a large NB/NC gene signature,
410 and provided the most complete list to date of Pax3 and Zic1 targets in the premigratory
411 neural crest (Tables S3, S5). We found that GRNBoost2 predicted 1076 genes to be linked to
412 Pax3 activation, among which 170 were decreased in Pax3 morphant neural crest (e.g. *zic1*,
413 *knop1*, *bms1* and *prps1*, Figs. 4B, S7). Similar modelling predicted 1136 genes linked to Zic1,
414 of which 188 were significantly changed in Zic1 MO cells (Figs. S6, S7). Interestingly,
415 *cyp26a1* was connected to *zic1* in the predicted network and was significantly decreased in
416 Zic1 MO. This supports previous characterization of *cyp26c1* as an immediate-early target of
417 Pax3 and Zic1 (*46*), and may relate to potential roles of Zic1 and Pax3 in antero-posterior
418 patterning through regulation of retinoid signaling (*47*).

419 We further defined Pax3 direct binding on target genes *in vivo* using ChIP-seq (Fig.
420 S8) on mid-neurula stage embryos (Fig. 4D). We retrieved 475 targets of Pax3 expressed in
421 the SC NC dataset, among a total of 657 targets found in the whole embryo (Table S6).
422 Known targets such as *cxcr4* and *prtg* were validated (*46*, *48*). Among the 475 targets, 80
423 were also predicted by GRN-Boost2 modelling including *psmd4, psen2, sp7, notch1, hnf1b.* In
424 sum, we provide here a vast Pax3-centered NC-GRN with three complementary approaches:
425 GRNboost2 predictions, morpholino-based depletion and Pax3 chromatin-binding (Fig. 4B-
426 D). However, the overlap between the gene lists retrieved as targets by three approaches was
427 only partial. A first possible reason is that GRNBoost2 predicted a larger group of connections
428 than did experimental MO depletion-based small-RNAseq or ChIPseq: based on gene
429 expression patterns, GRNBoost2 retrieved both TF-towards-target and target-towards TF
430 potential relationships. Another possibility is that the *pax3* expression pattern, used as the
431 basis for GRNboost2 predictions, is broader than most NC specifiers' early pattern [see in
432 (*2*)], as *pax3* is also expressed in the prospective dorsal neural tube, for example in stage 12-
433 13 neuroepithelial cells (see ectoderm cluster #7, #9 below). Nonetheless, these data provide a
434 NC-specific genome-wide connectome for two key NB specifiers, Pax3 and Zic1, much of
435 which we have validated experimentally.

436 We next focused experimental validation on a later NC specifier, TFAP2e, highly
437 expressed in cranial and trunk NC cells at a late neurula stage, predicted to be an important
438 node at EMT stage 19 (Fig. S9D). Using a similar approach, we processed for small-RNAseq
439 pre-migratory NC explants microdissected at the end of neurulation (stage 17-18) from wild-
440 type or *tfap2e* morphant embryos (Fig. S6) (*49*). Among 848 targets of TFAP2e predicted by
441 GRNBoost2 from scRNA-seq, 99 showed decreased expression after TFAP2e depletion in NC
442 *in vivo* (e.g. *tfap2b, sox10, dact1, sncaip;* Figs. 4E, S7). Moreover, using ChIP-seq for
443 TFAP2e, we identified 642 targets expressed in the NC dataset, among 805 targets for the

444 whole embryo, including top-scored *rmb20, pim1, arl5b* and *tfap2a* (Fig. 8S, Table S7). Out
445 of 848, 76 genes linked to *tfap2e* by synexpression analysis were directly bound by TFAP2e in
446 the late neurula *in vivo*. Thus, we also provide here a large-scale TFAP2e-centered NC-GRN
447 using GRNboost2-predicted, morpholino-validated and ChIPseq-validated targets.

448 ***Benchmarking SC-based machine-learning and in vivo experimental approaches to***
449 ***build the neural crest GRN.***

450 Our SC analyses provide a genome-wide connectome based on a total of 16978
451 interactions computed between 405 TFs (out of the initial list of 1417 TFs expressed in
452 developing embryos) and 4532 other genes expressed in neural crest during gastrulation and
453 neurulation. The biological significance of the network was validated using experimental
454 analysis for three different NB/NC specifiers, Pax3, Zic1 and TFAP2e, as transcriptomic
455 interactions predicted using GRNBoost2 retrieved many interrelations supported either by
456 direct TF binding (ChIP) or by TF depletion *in vivo*. However, when comparing the targets
457 predicted with SC transcriptome modelling and experimental validation *in vivo*, we find that
458 the fraction of gene correlations predicted with GRNboost2also validated by ChIPseq was 7-
459 9%, and also confirmed by TF depletion (which includes both direct and indirect targets) was
460 11-16%. Changing the network filtration criteria (weights, expression level) did not affect the
461 proportion of validated genes. This means that despite the significant discovery power
462 displayed by GRNBoost2 on a complex dataset, there is margin for increasing its specificity
463 and accuracy. For example, GRNBoost2 did not link *tfap2e* to *hnf1b* (although confirmed by
464 both ChIPseq and MO depletion), possibly because those genes have quite different
465 expression patterns during the course of neurulation (Fig. S9D). Likewise, GRNBoost2 did
466 not link *tfap2e* to *notch1* either (also confirmed with ChIPseq and MO) although they display
467 closely related patterns (Fig. S9D). Another hypothesis for the incomplete coincidence of the
468 predictions between MO/ChIP and GRNboost2 is that the dataset compiled for GRNboost2
469 includes cells of all available stages (due to the need for a large number of cells for modelling)
470 while the MO/ChIP experiments were carried out at specific stages. Additionally, another
471 limitation of GRNboost2 is its weak ability to predict the relationship between genes with
472 different expression time-frames, i.e. when a gene A activates a gene B and stops being
473 expressed in B-expressing cells. In this case, it is best to apply a branching analysis taking
474 pseudotime into account. Yet, these three TF-centered networks exemplify how a large single
475 cell dataset can be combined with *in vivo* experiments and bulk small-RNA-seq to efficiently
476 identify a substantial number of novel interactions in the NC-GRN: we provide here 264
477 targets of Pax3 in NC, covered by at least two methods, 206 for TFAP2e and 188 for Zic1
478 (Fig. S8). To further render this extensive resource available, we provide an interactive
479 visualization tool of all the interactions predicted between TFs and tables with their targets
480 validated with ChIP-seq and MO-based small-RNAseq (available upon publication).

481 **Branching towards biased premigratory neural crest is controlled by key**
482 **transcription factors.**

483 To explore the temporal dynamics of transcription factor expression that may specify
484 decision points in the development of pre-migratory neural crest, we used tree inference and
485 advanced pseudotime downstream analysis, with a focus on fate biasing using scFates (*50*). It
486 allowed us to explore the branch-specific transcriptional regulation from the calculated
487 pseudotime, in order to determine not only the gene-to-gene dependency but also the temporal
488 order in which their functions were accomplished. ElPiGraph approximates datasets with
489 complex topologies allowing us to build the graph structure. One limitation of ElPiGraph is
490 that it cannot be applied to large datasets with many potential branches. Thus, using the
491 principal graph constructed with PAGA (Fig. 2B), we sub-selected cells around three main
492 bifurcation points in the NC lineage tree and applied scFates, ElPiGraph and differential
493 expression analysis. We explored firstly the split of trajectories from Premigratory Unbiased

494 NC (cluster #5) to Migratory Bipotent Cranial NC (cluster #7) and Migratory Bipotent Vagal
495 NC (cluster #9); secondly the bifurcation of Migratory Bipotent Vagal NC #9 into EMT
496 Cardiac NC (#12) and Migratory ENSp (#13), and thirdly the emergence of two distinct
497 cranial states, Premigratory Early Cranial #10 and Migratory Cranial #11, from Migratory
498 bipotent cranial cluster #7 (Fig. 5A-F).

499 ***Cranial versus vagal bifurcation at the end of neural plate stage.***

500 Cranial NC cells emerge from the neural tube anterior to the otic vesicle while vagal
501 NC cells form from the hindbrain region adjacent to somites 1–7 (*51*). Our data indicates that
502 the first emergence of cells biased to vagal and cranial populations arise from the unbiased
503 population of cluster #5 around early neural fold stage 14. Although we did not see separation
504 of cluster #5 into two populations at the chosen level of clustering, we still observe an early
505 internal predisposition marked by the expression of early cranial (*nrp2)* and vagal (*mafb*)
506 markers in sub-regions of cluster #5 (Fig. S9A, B). In this analysis, cells that express *nrp2*
507 highly were closer to the cranial state than to the vagal stage (7 times more according to
508 PAGA; Fig S9A, B). Similarly, we explored cluster #5 cells which highly express *mafb* and
509 found them 1.5 times closer to the vagal state than the cranial state. Thus, although cluster #5
510 is rather homogeneous, a few genes might be responsible for early predisposition to the vagal
511 or cranial fates as early as stage 13. Moreover, we found that *mafb* and *nrp2* were the best
512 early predictors of branching for the vagal and cranial populations (Fig. 5A, B). Using
513 branching analysis, we retrieved early vagal gene programs as the sets of genes activated
514 before bifurcation (early; *mafb, klb, mdk*) and continuing to be expressed in each specific
515 branch (late; *prdm1, cfi, hoxd3*) (Fig 5A). Similarly, we uncovered early (*nrp2*, *alk, rnd1,*
516 *adam19*) and late (*shisa3, frmd6*) gene programs for cranial NC cells of cluster #7 (Fig. 5B).
517 This is in line with previous *in vivo* analysis of Alk function in cranial NC cells before and
518 during migration (*52*) and of Mafb role during cardiac NC cells specification (*26*).

519 ***Neural fold stage bifurcations: vagal to enteric split and cranial subdivision.***

520 We also defined gene programs for two later bifurcations: early vagal NC progenitors
521 of cluster #9 bifurcating into ENS (cluster #13) and cardiac NC (cluster #12) on one hand, and
522 bipotent cranial NC (cluster #7) bifurcating into *rpe65*$^{+}$ Cranial NC cells (cluster #10) and
523 Migratory Cranial *hox*$^{-}$; *dlx2*$^{+}$ NC cells of cluster #11 on the other hand. In contrast to the
524 unbiased cluster #5, clusters #7 and #9 are relatively homogeneous without obvious
525 propensity towards other later cell states. However, branching analysis revealed clear early
526 and late branch-specific actors. For the cardiac NC fate, we defined early genes, *mafb, mycn,*
527 *prdm1, nolc1* and *eef1d* and late ones, *egr2, hoxd3, epha4* and *abtb2*. For enteric nervous
528 system precursors (ENSp), early actors were *olig4* and *fbn2*. Due to its significantly increased
529 expression in later ENSp cells, the branching analysis identified *pax3* as a late actor, but the
530 expression-pseudotime heatmap showed that *pax3* was already expressed prior to branching in
531 ENSp progenitors and increased afterwards (Fig. 5C). Therefore, we also assigned *pax3* to the
532 group of early genes for ENSp. Later ENSp gene programs consist of *tnc, ltbp1, wnt11,* and
533 *hoxb6*. In order to validate experimentally if "early" factors affected "late" genes expression,
534 particularly if early TFs were bound to the late genes sequences, we examined which of the
535 late ENSp genes were targets of the early TF Pax3 gene, by ChIP-seq and MO analyses. We
536 identified several genes affected by Pax3 depletion, among them *krt8* (increased), *cldn1* and
537 posterior hox genes *hoxb6* and *hoxa7* (decreased) (Fig. 5G) while several were identified as
538 direct targets of Pax3 (e.g. *cfb* and *bmp5*, Figs. 5G, S9C).

539 Last, similar analysis on clusters 7-10-11 showed that early genes specific for
540 Premigratory Early Cranial cluster #10 branch were *rpe65, dmbx1, rgr, lmx1a,* and *zfhx4*,
541 while late gene programs consisted of *alx1* and *bmper* (Fig. 5E, F). Interestingly, we found
542 *pax3* expressed early in cranial bifurcation from #7 towards #10 and #11. *Pax3* was expressed

543 in cluster #7 progenitors and specifically enriched in *rpe65*$^+$ cluster #10. To explore this
544 further, we also used ChIP-seq and MO to determine if there were cluster #10-specific genes
545 among Pax3 direct targets and found that *comt, slc23a2* and *rpe65* were affected after *pax3*
546 depletion while *bmper, f10, sema3a, dmbx1, slc16a12* and *kif26a* were directly bound by Pax3
547 (Fig. 5G). On the other hand, the branch leading to Migratory Cranial *hox*$^-$*;dlx2*$^+$ NC cells
548 (#11) expressed *mef2c* and *eef1d* before and after bifurcation, and *dlx2, shb, itga4* only after
549 bifurcation. We found that *tfap2e* started to be expressed before bifurcation and was enriched
550 in cluster #11 in comparison to cluster #10. Consequently, using MO-mediated depletion and
551 ChIP-seq analysis, we found that TFAP2e depletion affected the expression of nine of the *hox*$^-$
552 *;dlx2*$^+$ branch-specific genes, among them the early gene *mef2c*, and the late genes *dlx2,*
553 *mmp14* and *vim*, and that TFAP2e directly bound four other cluster #11 genes: *c9, vim,*
554 *mmp14 and mycn* (Fig. 5G, S9C). Direct regulation of *vim* and *mmp14* suggested that TFAP2e
555 may be important for cranial NC EMT and migration. TFAP2e was previously reported to be
556 important in NC induction (*49*). To test the later role of TFAP2e in NC migration, we devised
557 an *in vivo* assay using low-level depletion of TFAP2e still allowing initial NC induction.
558 Specifically, a pre-EMT NC explant microdissected from embryos co-injected with low-levels
559 of TFAP2e MO, mRNA encoding an dexamethasone-inducible form of TFAP2e and a lineage
560 tracer, was grafted into a wild-type control embryo at stage 17 (Fig. 5H). Morphant NC
561 remained at the grafted site while wild-type cells efficiently populated the craniofacial areas
562 (Fig. 5Hh,h'). Importantly, when TFAP2e was induced in morphant NC upon EMT stage, cell
563 migration was restored and lineage-traced cells were found along NC cranial migration routes
564 (Fig. 5Hh"). Together, these results validate branching analysis predictions and demonstrated
565 that TFAP2e regulates expression of EMT effectors and cranial NC migration *in vivo*.

566 In conclusion, using computational approaches we have defined and analyzed three
567 main bifurcation points in the premigratory NC dataset: from unbiased to vagal and cranial
568 NC, from vagal NC to cardiac and ENSp fates, and from early cranial to either *rpe65*$^+$ or
569 *dlx2*$^+$ cranial NC. For each branch we define specific gene programs, including early actors
570 predicted to trigger specific states. Using ChIP-seq and MO-based small-RNAseq, we validate
571 the link between numerous branch-specific genes and two early regulators, Pax3 and TFAP2e.
572 We demonstrate that Pax3 displays early functions both in ENSp and in anterior cranial
573 *rpe65*$^+$ NC branching. In parallel, TFAP2e activates branch-specific gene programs in *dlx2*$^+$
574 cranial NC as well as cell migration *in vivo*.

575 **During gastrulation, the neural border zone emerges from overlapping ventral**
576 **and dorsal programs, revealing multiple coexisting circuits for neural crest and placode**
577 **lineage specification.**

578 ***Neural border cells exhibit a gene signature regulated by main nodes of the***
579 ***ectoderm connectome.***

580 Aiming to uncover the molecular mechanisms of cell fate decisions during NC
581 induction at the neural-non neural ectoderm border zone, we analyzed a second dataset
582 consisting of ectoderm cells for stages 11-13 (Fig. 1E). In this dataset we identified 10 cell
583 clusters: *has1*$^+$ immature ectoderm at mid-gastrula stage 11 (cluster #0 - ec), two non-neural
584 ectoderm populations (early *gata2*$^+$ cluster #2 - nne1 and later *tll2*$^+$ cluster #4 - nne2), *sox2*$^+$*,*
585 *zic1*$^+$ neural ectoderm (cluster #1 - ne), *lhx2*$^+$ eye primordium (cluster #8 - eye), *sox9*$^+$*, c3*$^+$
586 NC progenitors (cluster #6 - nc), *pitx1*$^+$, *olfm4*$^+$ placode progenitors (cluster # 5 - pc),
587 *neurog1*$^+$*, elavl3*$^+$ immature neurons (cluster #7 - neurons) and *zeb2*$^+$ neural plate (cluster #9
588 - np; Figs. 1E, 6A, B). Earlier work showed that NC and PC can develop from a common
589 "neural border zone", defined as an ectodermal area located in-between *sox2*$^+$ NE and
590 *epidermal keratin*$^+$ NNE and co-expressing *tfap2a, zic1* and *pax3* (Fig. S1) (*53*). The early
591 developmental dynamics of this ectodermal area has not yet been described at the single-cell
592 level, and it remains unknown if the cells in this area resemble adjacent progenitors or if they

593 exhibit a specific gene signature. On force-directed graph plot, this domain did not appear as
594 an invididual cluster, rather as cells spread out between clusters 1, 2, 5 and 6 (Fig. 6A). PAGA
595 graph analysis revealed that the *tfap2a*$^{+}$*, zic1*$^{+}$ cell group (nb, red cells) was equally similar to
596 ne (#1) and nne (#2), and closer by 5.4 times to nc (#6) than to pc cluster (#5, Fig. 6A, C). We
597 also observed a low density of cells at the levels of NB zone specifically between stage 11 and
598 12 (Fig. 6A). Because low density was not observed in neural or NNE areas, a stage-related
599 sampling issue could be excluded. Rather this could be related to a faster transcriptional
600 transition of the NB cells compared to NNE cells between stages 11 and 12/13 stages. This
601 might suggest that early ectoderm cells go through NB state quickly before switching towards
602 defined NC or PC states, resulting into fewer cells being captured in this transitional state
603 upon sequencing. Alternatively, such clustering could be obtained if the NB zone ectoderm
604 was highly heterogeneous and contained a mixture of NNE-like, NE-like and NB state cells.

605 We then sought to identify a detailed signature for the NB zone. Several genes were
606 enriched as early as stage 11: *tfap2c, pax3, sox9, hes1, gmnn* and *myc* (Fig 1E, Fig. S1).
607 Because *pax3* was previously found as one of the earliest NB zone specific genes (*2*), we
608 examined the impact of Pax3 depletion on this NB zone signature *in vivo*. We found that
609 expression of the other genes forming the NB signature, *sox9, axin2, zic3* and *zic1,* was
610 decreased in Pax3 morphant NB while *lhx5.l,* highly expressed in earlier NE and NNE*,* was
611 increased (Fig. 6C). These results consolidated Pax3 as a major switch for the early NB
612 signature. In addition, Zic1 depletion also led to decreased expression of markers of the NB
613 signature *zic3* and *gmmn* (Fig. 6D). In this case, *lhx5.s* was strongly decreased, contrarily to
614 the effect of Pax3 depletion, indicating possible cooperative roles between Pax3 and Zic1 to
615 balance a precise *lhx5* level compatible with neural border zone formation. Last, we defined
616 other main nodes in the whole Ectoderm connectome (10085 gene connections) and found
617 TFAP2a among the most significant ones (Table 2). We retrieved 682 direct targets of
618 TFAP2a expressed in the NC dataset (Fig. S8), among the 1393 targets bound by TFAP2a in
619 the whole embryo, including NB signature genes *pax3, sox9, myc and tfap2a* itself.

620 ***The NB zone contributes to NC and PC in parallel to convergent contributions from***
621 ***neural plate and non-neural ectoderm progenitors.***

622 Based on the force-directed graph we found three possible developmental routes
623 leading to NC and PC early states from stage 11 ectoderm cells: a) NE → NC; b) NB zone →
624 PC and NC; c) NNE → PC (Fig. 7A, C-H). This was in line with both models currently
625 proposed for NC and PC formation, namely the "neural border zone model" (route b) and the
626 "neural vs non-neural" model (routes a & c) (Fig. 7A) (*16*, *53*, *54*). Interestingly, no direct
627 route was found between nne1 and nc (Fig. 7D), or between ne and pc (Fig.7G), retrieving
628 known biological features, such as the partial neuralization needed for NC induction or the
629 close link between placodes and NNE (*17*, *18*). Because current tools cannot assess multiple
630 branching simultaneously, we applied branching analysis for each potential developmental
631 route in order to define and compare the gene programs underlying NC vs PC fate decisions
632 along each route. For route (a), we found that branching towards NP from the NE state
633 involved early genes *elav3 and sall2* and late genes *nkx6* and *tubb2b.* This was consistent with
634 current knowledge on neural plate induction (*sall2, nkx6*) and primary neurogenesis (*elav3,*
635 *tubb2b*), thus further validating the approach (Fig. 7C) (*55*). Route (a) branching towards NC
636 involved early genes *zic1* and *foxd3* and late genes *c3, myc, sox8, sox9, pcdh8* and *snai2* (Fig.
637 7D). In comparison, NC cells emerging from the NB zone by route (b) started to express *c3*
638 and *sox9* early, before bifurcation, while neural crest markers *foxd3*, *sox8, pcdh8, snail2* were
639 enriched after splitting (Fig. 7E). Similarly, we explored gene programs for both routes of
640 placodal development: from NB zone, route (b) showed early enrichment of *tcf7l1, hesx1* and
641 late for *stmn1, pax6, pitx1/2* (Fig. 7F). PC development from NNE, i.e. route (c), exhibited
642 early enrichment of placode specifiers *six1, otx2*, and late expression of placode markers
643 *egflam, pax6, pitx1/2*, Fig. 7G). Last, the route (c) branching for non-neural ectoderm

644 formation confirmed known developmental dynamics with early enrichment for *gata2* and late
645 for ectoderm stem cell marker *tp63* and epithelial cells *cldn1* (Fig. 7H) (*56*). This hierarchy of
646 gene expression along the different branches thus sopens avenues to further elaborate each of
647 the NC and PC GRNs.

648 ***Different gene programs can lead different progenitors towards a similar state.***

649 Interestingly, according to the route studied, we found that (i) distinct genes were
650 activated to obtain the same state, and that (ii) some genes were activated with different
651 expression dynamics relative to different bifurcations. For example, during the NB→NC
652 transition (route b, Fig. 7G), *sox9* and *c3* were activated early (before bifurcation) suggesting
653 that they could play a part in the fate decision network from NB progenitors. In contrast,
654 during the NE→NC gene program (Fig. 7D) *sox9* and *c3* were late genes while *foxd3* and *zic1*
655 were expressed early. This observation suggested a new model of fate decisions in the
656 developing ectoderm, where parallel and distinct genetic programs activated in distinct
657 ectoderm progenitors may lead to a similar state.

658 We validated Zic1-related genetic regulations along route a: *in vivo* depletion of early-
659 branch gene Zic1 increased expression of nc markers *sox8, foxd3* and *zic1* itself while it did
660 not affect other NC late branch genes *sox9, c3, myc*, or *snai2* (Fig. 7B). Zic1 depletion also
661 increased neuronal differentiation genes *tubb2b.l* and *tubb2b.s* in the np branch. This result
662 supported a role for Zic1 in NP and NC development but also revealed complex interplay with
663 Foxd3 (also early-branch) as Foxd3 enhanced expression may account for the activation of
664 late-branch NC marker *sox8.*

665 Last, NB zone specific gene *pax3* was expressed prior to bifurcation in the route (b),
666 NB→NC gene program, and controlled expression of late NC branch markers *sox9, sox8,*
667 *foxd3, pcdh8* and *c3* (Fig. 7B). Moreover, Ectoderm connectome described NC genes
668 connected to the rest of the network through Pax3 and Sox9 (Fig. 7I), suggesting that Sox9
669 might play a yet undescribed function downstream of Pax3 in NC induction and upstream of
670 the other late NC branch markers. This was also in agreement with *sox9* being an early gene in
671 the NB->NC branch. We tested the epistasis relationships between Pax3 and Sox9 in NC fate
672 induction by combining Sox9 depletion or gain-of-function in the induced-neural crest assay
673 (iNC is Pax3/Zic1-based NC induction from pluripotent ectoderm cells, Fig. 7J). *Snail2,* early
674 marker of NC induction, starts to be expressed in the prospective NC from gastrula stage 12.5
675 both *in vivo* and in iNC, then is enhanced at neurula stage 14 (Fig. 7D, E, I). At both stages,
676 co-activation of Sox9 strongly increased *snail2* expression, while Sox9 depletion reduced
677 *snail2* activation (Fig. 7I, J). This result indicated that Sox9 is required for efficient NC
678 induction by Pax3 and Zic1. Interestingly, when iNC explants were analyzed prior to the
679 normal onset of *snail2* expression, at mid-gastrula stage 11/11.5, Sox9 activation drastically
680 increased *snail2* expression, suggesting that Sox9 synergizes with Pax3 and Zic1 at the onset
681 of NC induction (Fig. 7J).

682 In conclusion, we have defined a new transcriptional signature for the incompletely
683 described Neural Border zone. We established a global Ectoderm connectome and validated
684 experimentally NC-related nodes, in particular highlighting a novel function of Sox9
685 enhancing NC induction by Pax3 and Zic1. We characterized three different developmental
686 routes branching towards NC and PC from the neural, neural border and non-neural ectoderm
687 progenitors. We thus propose a model in which multiple co-existing paths can lead to common
688 NC or PC states in gastrula-stage ectoderm.

689

690 **Discussion**

691 In this work, we exploit the unprecedented resolution of high-density single cell
692 transcriptomes collected from an 8-stage developmental series to unravel the emergence of the
693 neural crest lineage from the ectoderm during gastrulation followed by the diversification of
694 neural crest progenitors during neurulation and upon EMT. Modeling gene transcription
695 dynamics around each cell state allows inference of the underlying molecular networks. We
696 selected several important nodes among those regulations for large-scale and *in vivo*
697 experimental validation showing the good yet perfectible accuracy of machine learning-
698 predicted networks (Fig. 8A-B). This study highlights the previously unknown temporal
699 hierarchy of lineage decisions in neural crest development. Firstly, we characterize how the
700 neural crest is activated either from a transient neural border state, or from a neural plate state
701 at mid-gastrulation stage, and suggest a model that reconciles current debates upon the
702 possible routes leading from immature ectoderm to neural crest and placodes (Fig. 8C).
703 Secondly, we delineate the early and later neural crest trajectories that emerge during
704 neurulation and define key regulators of branching, leading to eight transitional states and
705 eight early migration states (Fig. 8D).

706 ***A reconciliatory "Model of Dual Convergence" describes the converging trajectories***
707 ***initiating neural crest and placode states.***

708 The molecular signature of the neural border ectoderm has been overtly simple, with
709 only Pax3 (in frog and fish) or Pax7 (in chick) as relatively specific markers for this domain
710 during gastrulation and early neurulation stages (*9*, *14*). Here, we have characterized the
711 neural border ectoderm state by two features: the lower level of expression of genes expressed
712 by adjacent neural (dorsal) and non-neural (ventral) cells (*sox2*, *lhx2*, *sall2*, *gata2, ker19*) and
713 the increased expression of a large gene list including *zic1, tfap2a/c, pax3*, *sox9, hes1, cmyc*.
714 We find that this state seems more transient than other ectoderm states, suggesting that these
715 fate decisions occur quickly. In frog embryos, the end of gastrulation is clearly defined by
716 blastopore closure, and this allows more precise exploration of timing compared to organisms
717 with simultaneous gastrulation and neurulation such as chick embryos. In frog, fate choices in
718 the dorsal ectoderm happen during the second half of gastrulation (between stage 11 and
719 12.5). As the neural plate forms and the blastopore closes (stage 13), fate decisions are clearly
720 established between neural, non-neural, placode and neural crest with robust molecular
721 signatures (Fig. 6). Modeling neural crest emergence from the neural border cell state
722 confirmed the central role for Pax3 and suggested novel epistatic relationships between Pax3
723 and Sox9 upstream of the definitive neural crest state, defined by persistent expression of
724 *snail2*. With a temporal series of induced neural crest assay, Pax3 depletion and Pax3 ChIP,
725 we validate the key role of Pax3 in the early NB/NC-GRN (Fig. 4) and the modeling-driven
726 hypothesis that Sox9 displays an early function at the root of the NC-GRN (Fig. 7).
727 Importantly, we propose a novel model for the pattern of fate decisions between the four main
728 ectoderm fates, neural, non-neural, placodal, neural crest. Instead of contrasting “neural
729 border” and “non-neural vs neural" hypotheses, we find that these routes are not exclusive and
730 find trajectories supporting the emergence of neural crest from either the neural border or the
731 nascent neural ectoderm on one hand, as well as two trajectories leading to placodes from
732 either the neural border or the non-neural ectoderm. In each case, the gene programs
733 underlying those alternative trajectories involve a subset of common genes and a few specific
734 factors (Fig. 7). For example, specific expression of *tcf7l1* and *stmn1* is found in placodes
735 arising from NB zone, compared to the NNE route (Fig. 7F, G). For neural crest, early *sox9*
736 expression in the NB route contrasts with post-bifurcation expression in the NP route (Fig.
737 GH7D, E). Thus, our SC transcriptome modeling reconciles and combines previously
738 alternatives in a "Dual convergence Model" of neural crest and placode patterning, together
739 with specific gene signatures for future functional exploration (Fig. 8C).

740 ***A combination of Omics and in vivo strategies validates large sets of gene***
741 ***regulations driving the dynamics of neural crest diversification.***

742 The second outcome of our study is to define the temporal dynamics of trajectories that
743 result into eight neural crest states present upon early migration stage along the cranial and
744 vagal axial positions. The first key observation is the presence of a main population of NC
745 unbiased towards any particular state, expressing markers of the immature neural crest cells,
746 from which all the other trajectories emerge (Fig. 3). This unbiased cell trajectory is
747 maintained during and after EMT suggesting that a very plastic, stem-like NC cell population
748 emigrates and is subjected to the signals from the microenvironment prior to fate choices. The
749 second critical observation is that, for the anterior part of the body axis considered here,
750 trajectories do not emerge in a spatially linear sequence from anterior to posterior as
751 previously anticipated in a model where NC would follow an anterior-posterior wave of
752 maturation. Two early trajectories arise from both the anterior (progenitors of posterior cranial
753 NC - cluster #15) and the posterior-most positions (minor vagal trajectory progenitors, cluster
754 #3) at neural plate stage (stages 13-14) prior to neural fold elevation. This is followed at mid-
755 neurula stage (stage 15-16) by the emergence of the three other main cranial and vagal
756 trajectories leading to cranial clusters #10 and #11, and to vagal NC clusters #12 and #13.
757 Together with the maintenance of an immature stem-like cell population from induction to
758 emigration, this sequence of trajectory determination suggests that the main cue controlling
759 the temporal dynamics of states hierarchy in the cranial and vagal NC-GRN is not a function
760 of the time elapsed since NC cell induction, or correlated to Hox gene positional information,
761 but rather would involve response to external signals. This hypothesis is supported by the
762 strong retinoic-acid response signature which defines the posterior cranial lineage from neural
763 plate to emigration stages, and maintains this trajectory distinct from the unbiased NC
764 population.

765 Our temporal analysis highlights three important points deepening our understanding
766 of NC biology. Firstly, there have been long standing debates about the timing of NC fate
767 decisions, prior or after EMT from the neural tube, in a variety of animal models (*57*).
768 Importantly, we did not detect distinctive expression of predictive fate markers before EMT
769 (e.g. for neuronal, glial skeletogenic or melanocyte fates). This suggested that, if some NC
770 progenitors were biased towards a given fate prior to EMT, they did not exhibit a detectable
771 signature in our dataset. However, our observations are in agreement with several lineage
772 tracing studies showing the high multipotency of most NC cells when marked prior to EMT
773 (*58*). The first differentiation markers are found after emigration, as we detected myosin-like
774 expression in a small subset of cells suggesting the emergence of previously poorly described
775 NC-derived myofibroblasts shortly after EMT (Fig. 3E). It also supports that our dataset is
776 sensitive enough to detect other fate-specific markers if they were expressed at the end of
777 neurulation. Secondly, our data supports the early diversification into several distinct cell
778 states prior, during and after EMT, contrasting with recent suggestion that upon EMT the NC
779 progenitors would regroup into a single common multipotent state (*13*). The high cell content
780 of our dataset proves otherwise, suggesting that this previous observation made on a smaller
781 subset of cranial NC did not fully capture the diversity of pre-migratory NC states. Lastly,
782 temporal trajectory analysis unravels the branch-specific dynamics of gene expression
783 underlying bifurcations and state diversification. For each bifurcation, we provide a list of key
784 genes likely to control branching choices (Figs. 5,7). We further validate these predictions in
785 several instances, by experimental modulation of pivotal transcription factors function in the
786 premigratory neural crest (Pax3, Zic1, TFAP2a, TFAP2e), followed by *in vivo* or deep
787 sequencing analysis. This strategy proves versatile and powerful to validate the predictions of
788 SC transcriptomics on a large scale.

789 In sum, our study provides a comprehensive view of the hierarchy of molecular
790 decisions driving the cranial and vagal neural crest gene regulatory network from induction at
791 the neural border to early migration, with unprecedented resolution and deep learning-aided
792 experimental validation. We propose a new "Dual Convergence Model" for neural crest and
793 placode lineage emergence, and provide a detailed roadmap of the main molecular events in

794 the premigratory and early migrating NC-GRN. Using a dedicated interactive network
795 visualization interface, any gene of interest can be queried. Moreover, the detailed view of
796 lineage formation provided here will prove an essential reference for monitoring induction of
797 neural crest derivatives, for example from patient-derived induced pluripotent stem cells,
798 when reliable specification protocols preferably recapitulate the steps of embryonic
799 development.

800 **Materials and Methods**

801 ***Experimental Design***

802 Single cell transcriptomes from developing frog embryos were scrutinized for neural
803 crest progenitor development using machine-learning tools to infer the gene regulatory
804 network connectome and the gene programs underlying branching of fates. These predictions
805 were largely validated *in vivo* using micro-manipulations in *X. laevis* embryos followed by
806 RNA-seq or ChIPseq.

807 ***Single cell sequencing***

808 No new materials were collected for this study. Instead, we re-sequenced the RNA
809 libraries for developmental stages NF11 to NF22 (*59*) used in (*17*). Please refer to the methods
810 therein for staging, embryo collection, dissociation, cell collection, barcoding (inDrop v2 and
811 v3) and library preparation methods. For re-sequencing, we used two flow cells (two lanes
812 each) of NovaSeq S2 at 100 cycles setting generating a total of 12 billion reads. All datasets
813 are deposited under NCBI Gene Expression Omnibus number GSE198494.

814 ***Reference genome and gene symbol assignments***

815 For bioinformatics analysis of SC dataset, we used the *X.tropicalis* v.10 genome
816 assembly, gene models v. 10.7. Specifically, we used the file Xentr10-Gene-Sym-HUMAN-
817 BLOSUM45.txt downloaded from Xenbase on Nov 5th, 2021. Thousands of genes in that
818 transcriptome version lack interpretable gene symbols, though in many cases an unambiguous
819 protein identity can be identified by sequence homology. To fill the holes in gene symbol
820 assignments we assigned protein gene symbols to each transcript using a modified reciprocal
821 best HMMER hit approach (*60*) based on a target reference set of curated human proteins.

822 ***scRNA-seq reads processing***

823 Compared to (*17*), we selected a different aligner and scRNA pipeline since Bowtie
824 (*61*), which was used in the InDrops pipeline (*62*), was not designed to align transcripts to the
825 genome, and therefore is not splice-aware. A splice-aware algorithm does not align RNA-seq
826 reads to introns, and identifies possible downstream exons trying to align to those instead.
827 Moreover, using Bowtie for aligning reads to a transcriptome raises questions about choosing
828 the correct transcript. Therefore, we chose STAR, since it is one of the best RNA-seq aligners
829 and outperforms other aligners in terms of correctly and incorrectly aligned reads (*63*). Also, it
830 is fast and there is no need to do preprocessing by removing bad quality bases or adapters as
831 STAR does that internally (*64*). We used the DropEst pipeline (*65*). Recent benchmarking
832 shows that DropEst exceeds other options in terms of sensitivity and efficiency (running time
833 and memory use) (*66*). As a result, for v3 we used Indrops demultiplexing and soft filtering,
834 then we generated tagged files similar to dropTag without considering reads base quality for
835 barcodes, followed by alignment and quantification with STAR
836 (outFilterMatchNminOverLread 0.44 --outFilterScoreMinOverLread 0.44) and dropEst using
837 X.trop v10. After filtration by counts and genes numbers (>200 genes; >300 counts), we
838 gathered a dataset of 177250 cells. In the cells of interest (Ectoderm and NC cells), mean
839 counts number was 1778, and mean gene number was 1035.

840 ***scRNA-seq postprocessing***

841 To process scRNA data we used Scanpy (*25*), a comprehensive scRNA pipeline that is
842 functionally similar to Seurat (*67*). Firstly, we calculated the percentage of mitochondrial

843 (MT) and ribosomal (RB) genes per cell,by manually calculating the fraction of mitochondrial
844 reads and ribosomal reads. As high MT proportions indicate poor quality cells (*68*),
845 presumably due to loss of cytoplasmic RNA from perforated cells, we removed cells with a
846 high percentage of MT/RB transcripts. Then, we removed gene sets which might affect the
847 normalization step. These include foreign tissue contaminants, heatshock, ribosomal, clock
848 and hemoglobin genes. For normalization, we excluded the expression of genes if their
849 expression was more than 3 percent of the total expression of the cell, because it could greatly
850 affect the resulting normalized values for all other genes (*69*). Next, we assessed the cell cycle
851 dynamics. The algorithm calculates the difference between the mean of the given list of cell
852 cycle genes and the mean of the reference genes. To build such a reference, we randomly
853 selected a set of genes that matched the distribution of the expression in the list. This way we
854 got S phase, G2M phase scores for each cell. We did not regress out cell cycle effect because
855 it had no significant influence. For normalization, we used scanpy function
856 *scanpy.pp.normilize_total* without highly expressing genes and target_sum=10^4.

857 ***Clustering and NC cells selection***

858 For each stage, we performed independent standard dimensionality reduction with
859 PCA, computing a neighborhood graph (n_pcs selected based on the amount of variance
860 explained by each PC, n_neighbors selected manually) and UMAP (*70*). For clustering, we
861 used the Leiden algorithm (*24*). For each cluster, we defined cluster-specific genes with
862 differential expression analysis (scanpy t-test_overestim_var) and selected only clusters which
863 were the most similar to NC cells using NC signatures from (*17*).

864 ***NC classifier***

865 First, we generated a dataset by labeling cells as NC based on the previous differential
866 expression analysis between clusters (*17*). For feature selection, we trained the LightGBM
867 model (*20*) on the whole embryo dataset of 177250 cells, retrieved the top 2000 important
868 features and used it for re-training the model (default setting with n_estimators=500). The
869 resulting model detected NC cells in the test dataset from the *Xenopus tropicalis* dataset with
870 accuracy 0.99 and F1 score 0.90. Only the trained dataset sample was used to get the top
871 important features and retrain the model on the top 2000 features. To test the model
872 robustness, we validated the NC classifier in the Zebrafish scRNA-seq dataset (*21*). We used a
873 sigmoid function to smoothen the technical batch effect between the datasets. Genes that were
874 not found in the Zebrafish dataset were replaced with NaN values (728 from top 2000
875 important genes were not found in the Zebrafish dataset). Finally, our model predicted NC
876 cells in the Zebrafish dataset with AUC score 0.95 and F1 score 0.66 (in comparison, the
877 random model with strategy="stratified" (https://scikit-
878 learn.org/stable/modules/generated/sklearn.dummy.DummyClassifier.html) applied on the
879 imbalanced Wagner dataset has F1 score=0.05).

880

881 ***GRN generation***

882 To predict the NC-GRN, we used GRNBoost2 (*43*), one of the best GRN predictors
883 according to recent benchmarking (*42*). Through this algorithm, for each gene in the cell-gene
884 matrix, a tree-based regression model is built to predict gene expression using the expression
885 of TFs. Each gene-specific model produces a partial GRN with regulatory associations from
886 the most predictive TFs for the gene. Then, all regulatory associations are pooled and ranked
887 by importance to complete the GRN. The *Xenopus* TF list (1417 TFs) was taken from
888 *Xenopus tropicalis* TF catalog (*45*). We analyzed the resulting network of TFs and their
889 targets using the networkx package and identified the most important nodes by calculating
890 betweenness and degree centralities with betweenness_centrality, degree_centrality network
891 functions.

892 ***Principal graph generation***

The tree analysis was carried out using the scFates package (*50*), based on ElPiGraph using the concept of elastic energy and a gradient descent-like optimization of the graph topology (*50*). However, this approach is too sensitive to build the principal graph for the whole NC dataset. Therefore, to generate the main tree, we used the PAGA algorithm (*25*). This revealed cluster-cluster relationships including the early stages where the strongest connectivity was observed. Further we used ElPiGraph to study specific branches and bifurcation points.

Branching analysis

Using ScFates we defined features significantly changing along the tree, and then using pseudotime values and differential expression analysis, determined early and late branch-specific features. For each bifurcation point of interest, we selected a set of cells related to the clusters involved in the bifurcation. The selection of parameters for building the principal tree for each point of the bifurcation was carried out using brute force approach.

Chromatin immunoprecipitation sequencing (ChIPseq)

Chromatin immunoprecipitation was performed according to (*71*). Embryos were injected in both blastomeres at the two-cell stage with tracing amounts (75 pg) of mRNA encoding either Pax3-FLAG-HA, or TFAP2a-FLAG or TFAP2e-FLAG. Injected embryos were collected at mid-neurula stage 14 (Pax3 and TFAP2a, 100 *X. laevis* embryos/condition) or at NC early migratory stage 19 (TFAP2e, 100 *X. tropicalis* embryos/condition). IP efficiency was tested on *snai2* promoter (Table S9). After sequencing, 100 bp single-end reads were aligned to *X. laevis* genome version 9.2 or *X. tropicalis* v10.0 using bowtie2. Peaks were called using MACS2. For Pax3 and TFAP2a, we selected peaks common in three replicates, for TFAP2e we used stricter MACS2 score cutoff=500. Target genes were searched with bedtools (window size = 10kb).

In vivo experiments: Xenopus laevis injections, microdissections, grafting and small RNA-seq

In vivo injections, NB/NC dissections and grafting were done as previously (*2*, *72*) using *X. laevis* embryos. For knockdown experiments, previously validated antisense morpholino oligonucleotides (MO) were used to deplete *pax3, zic1,* and *tfap2e* transcripts (GeneTools). Pax3 or Zic1 MO (20ng) (*9*, *73*) or *tfap2e* MO (20ng) (*49*). Depletion efficiency was verified using in situ hybridization on sibling embryos to verify reduction of *snai2* expression (not shown). One Pax3 morphant anterior NB explant (stage 14) or one TFAP2e-morphant NC explant (stage 17) were dissected from the injected side, in triplicate. After RNA extraction and cDNA library preparation, each individual explant was sequenced (small RNA-seq). The resulting 100 bp paired-end sequencing reads were aligned to the *X. laevis* genome version 9.2 using STAR and the count reads were analyzed using String Tie. Differentially expressed genes were selected considering log2FC and expression difference in absolute values (abs. diff. >=100 and <=500: log2FC>1.5 or log2FC<-1.5; abs. diff >=500 and <=1000: log2FC>1 or log2FC<-1; abs. diff >=1000 and <=3000: log2FC>0.5 or log2FC<-0.5; abs.diff.>3000: log2FC>0.33 or log2FC<-0.33)

iNC assay, RNA quantification and RT-qPCR

The induced neural crest assay (iNC) used co-activation of dexamethasone-inducible Pax3-GR and Zic1-GR at gastrula initiation stage 10.5, in pluripotent blastula ectoderm (animal caps dissected at blastula stage 9) (*29*, *74*). This was combined with Sox9 depletion (40 ng of *sox9* MO) (*75*), or gain-of-function (300 pg *sox9* mRNA). At the desired stage, explants were harvested and processed for RTqPCR as in (*18*). Primers listes in Table S9.

Whole-mount in situ hybridization (ISH)

Whole-mount *in situ* hybridization followed a protocol optimized for superficial structures (*76*). Embryos were imaged using a Lumar V12 Binocular microscope equipped with bright field and color cameras (Zeiss).

943
944
945 **References**

946 1. B. F. Eames, D. M. Medeiros, I. Adameyko, *Evolving Neural Crest Cells* (CRC Press,
947 2020).

948 2. J.-L. Plouhinec, S. Medina-Ruiz, C. Borday, E. Bernard, J.-P. Vert, M. B. Eisen, R. M.
949 Harland, A. H. Monsoro-Burq, A molecular atlas of the developing ectoderm defines
950 neural, neural crest, placode, and nonneural progenitor identity in vertebrates. *PLOS*
951 *Biology*. **15**, e2004045 (2017).

952 3. B. Steventon, R. Mayor, Early neural crest induction requires an initial inhibition of Wnt
953 signals. *Dev Biol*. **365**, 196–207 (2012).

954 4. A. Streit, C. D. Stern, Establishment and maintenance of the border of the neural plate in
955 the chick: involvement of FGF and BMP activity. *Mechanisms of Development*. **82**, 51–
956 66 (1999).

957 5. D. Bhattacharya, A. P. Azambuja, M. Simoes-Costa, Metabolic Reprogramming
958 Promotes Neural Crest Migration via Yap/Tead Signaling. *Dev Cell*. **53**, 199-211.e6
959 (2020).

960 6. A. L. Figueiredo, F. Maczkowiak, C. Borday, P. Pla, M. Sittewelle, C. Pegoraro, A. H.
961 Monsoro-Burq, *Development*, in press, doi:10.1242/dev.157644.

962 7. I. T. C. Ling, T. Sauka-Spengler, Early chromatin shaping predetermines multipotent
963 vagal neural crest into neural, neuronal and mesenchymal lineages. *Nat Cell Biol*. **21**,
964 1504–1517 (2019).

965 8. W. Tang, Y. Li, A. Li, M. E. Bronner, Clonal analysis and dynamic imaging identify
966 multipotency of individual Gallus gallus caudal hindbrain neural crest cells toward
967 cardiac and enteric fates. *Nat Commun*. **12**, 1894 (2021).

968 9. A.-H. Monsoro-Burq, E. Wang, R. Harland, Msx1 and Pax3 cooperate to mediate FGF8
969 and WNT signals during Xenopus neural crest induction. *Dev Cell*. **8**, 167–178 (2005).

970 10. M. Simoes-Costa, M. E. Bronner, Reprogramming of avian neural crest axial identity and
971 cell fate. *Science*. **352**, 1570–1573 (2016).

972 11. L. Medina-Cuadra, A. H. Monsoro-Burq, Xenopus, an emerging model for studying
973 pathologies of the neural crest. *Curr Top Dev Biol*. **145**, 313–348 (2021).

974 12. K. B. Artinger, A. H. Monsoro-Burq, Neural crest multipotency and specification: power
975 and limits of single cell transcriptomic approaches. *Fac Rev*. **10**, 38 (2021).

976 13. A. Zalc, R. Sinha, G. S. Gulati, D. J. Wesche, P. Daszczuk, T. Swigut, I. L. Weissman, J.
977 Wysocka, Reactivation of the pluripotency program precedes formation of the cranial
978 neural crest. *Science*. **371**, eabb4776 (2021).

979 14. M. L. Basch, M. Bronner-Fraser, M. I. García-Castro, Specification of the neural crest
980 occurs during gastrulation and requires Pax7. *Nature*. **441**, 218–222 (2006).

981 15. A. K. Groves, C. LaBonne, Setting appropriate boundaries: Fate, patterning and
982 competence at the neural plate border. *Dev Biol*. **389**, 2–12 (2014).

983 16. S. K. Maharana, G. Schlosser, A gene regulatory network underlying the formation of
984 pre-placodal ectoderm in Xenopus laevis. *BMC Biol*. **16**, 79 (2018).

985 17. J. A. Briggs, C. Weinreb, D. E. Wagner, S. Megason, L. Peshkin, M. W. Kirschner, A.
986 M. Klein, The dynamics of gene expression in vertebrate embryogenesis at single-cell
987 resolution. *Science*. **360**, eaar5780 (2018).

988 18. M. Alkobtawi, P. Pla, A. H. Monsoro-Burq, BMP signaling is enhanced intracellularly
989 by FHL3 controlling WNT-dependent spatiotemporal emergence of the neural crest. *Cell*
990 *Reports*. **35**, 109289 (2021).

991 19. N. de Crozé, F. Maczkowiak, A. H. Monsoro-Burq, Reiterative AP2a activity controls
992 sequential steps in the neural crest gene regulatory network. *Proc Natl Acad Sci U S A*.
993 **108**, 155–160 (2011).

994 20. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, in *Advances in*
995 *Neural Information Processing Systems* (Curran Associates, Inc., 2017;
996 https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-
997 Abstract.html), vol. 30.

998 21. D. E. Wagner, C. Weinreb, Z. M. Collins, J. A. Briggs, S. G. Megason, A. M. Klein,
999 Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo.
1000 *Science*. **360**, 981–987 (2018).

1001 22. C. Pegoraro, A. H. Monsoro-Burq, Signaling and transcriptional regulation in neural
1002 crest specification and migration: lessons from xenopus embryos. *Wiley Interdiscip Rev*
1003 *Dev Biol*. **2**, 247–259 (2013).

1004 23. A. Thawani, A. K. Groves, Building the Border: Development of the Chordate Neural
1005 Plate Border Region and Its Derivatives. *Front Physiol*. **11**, 608880 (2020).

1006 24. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-
1007 connected communities. *Sci Rep*. **9**, 5233 (2019).

1008 25. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression
1009 data analysis. *Genome Biology*. **19**, 15 (2018).

1010 26. S. Tani-Matsuhana, K. Inoue, Identification of regulatory elements for MafB expression
1011 in the cardiac neural crest. *Cells Dev*. **167**, 203725 (2021).

1012 27. S. E. Akbareian, N. Nagy, C. E. Steiger, J. D. Mably, S. A. Miller, R. Hotta, D. Molnar,
1013 A. M. Goldstein, Enteric neural crest-derived cells promote their migration by modifying
1014 their microenvironment through tenascin-C production. *Developmental Biology*. **382**,
1015 446–456 (2013).

1016 28. J. England, S. Loughna, Heavy and light roles: myosin in the morphogenesis of the heart.
1017 *Cell Mol Life Sci*. **70**, 1221–1239 (2013).

1018 29. C. Milet, F. Maczkowiak, D. D. Roche, A. H. Monsoro-Burq, Pax3 and Zic1 drive
1019 induction and differentiation of multipotent, migratory, and functional neural crest in
1020 Xenopus embryos. *Proc Natl Acad Sci U S A*. **110**, 5528–5533 (2013).

1021 30. E. Nitzan, S. Krispin, E. R. Pfaltzgraff, A. Klar, P. A. Labosky, C. Kalcheim, A dynamic
1022 code of dorsal neural tube genes regulates the segregation between neurogenic and
1023 melanogenic neural crest cells. *Development*. **140**, 2269–2279 (2013).

1024 31. J. A. Morrison, R. McLennan, J. M. Teddy, A. R. Scott, J. C. Kasemeier-Kulesa, M. M.
1025 Gogol, P. M. Kulesa, Single-cell reconstruction with spatial context of migrating neural
1026 crest cells and their microenvironments during vertebrate head and neck formation.
1027 *Development*. **148**, dev199468 (2021).

1028 32. P. L. Martínez-Morales, R. Diez del Corral, I. Olivera-Martínez, A. C. Quiroga, R. M.
1029 Das, J. A. Barbas, K. G. Storey, A. V. Morales, FGF and retinoic acid activity gradients
1030 control the timing of neural crest cell emigration in the trunk. *J Cell Biol*. **194**, 489–503
1031 (2011).

1032 33. R. A. S. Kjolby, R. M. Harland, Genome-wide identification of Wnt/β-catenin
1033 transcriptional targets during Xenopus gastrulation. *Dev Biol*. **426**, 165–175 (2017).

1034 34. R. K. T. Kam, W. Shi, S. O. Chan, Y. Chen, G. Xu, C. B.-S. Lau, K. P. Fung, W. Y.
1035 Chan, H. Zhao, Dhrs3 protein attenuates retinoic acid signaling and is required for early
1036 embryonic patterning. *J Biol Chem*. **288**, 31477–31487 (2013).

1037 35. A. J. Burns, P. J. Pasricha, H. M. Young, Enteric neural crest-derived cells and neural
1038 stem cells: biology and therapeutic potential. *Neurogastroenterol Motil*. **16 Suppl 1**, 3–7
1039 (2004).

1040 36. A. Collazo, M. Bronner-Fraser, S. E. Fraser, Vital dye labelling of Xenopus laevis trunk
1041 neural crest reveals multipotency and novel pathways of migration. *Development*. **118**,
1042 363–376 (1993).

1043 37. V. Sasselli, V. Pachnis, A. J. Burns, The enteric nervous system. *Dev Biol*. **366**, 64–73
1044 (2012).

1045 38. N. Nagy, T. Kovacs, R. Stavely, V. Halasy, A. Soos, E. Szocs, R. Hotta, H. Graham, A.
1046 M. Goldstein, *Development*, in press, doi:10.1242/dev.199825.

1047 39. P. M. Helbling, C. T. Tran, A. W. Brändli, Requirement for EphA receptor signaling in
1048 the segregation of Xenopus third and fourth arch neural crest cells. *Mech Dev*. **78**, 63–79
1049 (1998).

1050 40. M. Addison, Q. Xu, J. Cayuso, D. G. Wilkinson, Cell Identity Switching Regulated by
1051 Retinoic Acid Signaling Maintains Homogeneous Segments in the Hindbrain. *Dev Cell*.
1052 **45**, 606-620.e3 (2018).

1053 41. R. Lumb, S. Wiszniak, S. Kabbara, M. Scherer, N. Harvey, Q. Schwarz, Neuropilins
1054 define distinct populations of neural crest cells. *Neural Dev*. **9**, 24 (2014).

1055 42. A. Pratapa, A. P. Jalihal, J. N. Law, A. Bharadwaj, T. M. Murali, Benchmarking
1056 algorithms for gene regulatory network inference from single-cell transcriptomic data.
1057 *Nat Methods*. **17**, 147–154 (2020).

1058 43. T. Moerman, S. Aibar Santos, C. Bravo González-Blas, J. Simm, Y. Moreau, J. Aerts, S.
1059 Aerts, GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory
1060 networks. *Bioinformatics*. **35**, 2159–2161 (2019).

1061 44. C. Niehrs, N. Pollet, Synexpression groups in eukaryotes. *Nature*. **402**, 483–487 (1999).

1062 45. I. L. Blitz, K. D. Paraiso, I. Patrushev, W. T. Y. Chiu, K. W. Y. Cho, M. J. Gilchrist, A
1063 catalog of Xenopus tropicalis transcription factors and their regional expression in the
1064 early gastrula stage embryo. *Dev Biol*. **426**, 409–417 (2017).

1065 46. J.-L. Plouhinec, D. D. Roche, C. Pegoraro, A.-L. Figueiredo, F. Maczkowiak, L. J.
1066 Brunet, C. Milet, J.-P. Vert, N. Pollet, R. M. Harland, A. H. Monsoro-Burq, Pax3 and
1067 Zic1 trigger the early neural crest gene regulatory network by the direct activation of
1068 multiple key neural crest specifiers. *Dev Biol*. **386**, 461–472 (2014).

1069 47. M. Uehara, K. Yashiro, S. Mamiya, J. Nishino, P. Chambon, P. Dolle, Y. Sakai,
1070 CYP26A1 and CYP26C1 cooperatively regulate anterior-posterior patterning of the
1071 developing brain and the production of migratory cranial neural crest cells in the mouse.
1072 *Dev Biol*. **302**, 399–411 (2007).

1073 48. M. Xu, Y. Li, J. Du, H. Lin, S. Cao, Z. Mao, R. Wu, M. Liu, Y. Liu, Q. Yin, PAX3
1074 Promotes Cell Migration and CXCR4 Gene Expression in Neural Crest Cells. *J Mol*
1075 *Neurosci*. **64**, 1–8 (2018).

1076 49. C.-S. Hong, A. Devotta, Y.-H. Lee, B.-Y. Park, J.-P. Saint-Jeannet, Transcription factor
1077 AP2 epsilon (Tfap2e) regulates neural crest specification in Xenopus. *Dev Neurobiol*. **74**,
1078 894–906 (2014).

1079 50. L. Albergante, E. Mirkes, J. Bac, H. Chen, A. Martin, L. Faure, E. Barillot, L. Pinello, A.
1080 Gorban, A. Zinovyev, Robust and Scalable Learning of Complex Intrinsic Dataset
1081 Geometry via ElPiGraph. *Entropy*. **22**, 296 (2020).

1082 51. N. Le Douarin, C. Kalcheim, *The Neural Crest* (Cambridge University Press, Cambridge,
1083 ed. 2, 1999; https://www.cambridge.org/core/books/neural-
1084 crest/CC1993A22EDB67E40992CC2FD3771536), *Developmental and Cell Biology*
1085 *Series*.

1086 52. S. G. Gonzalez Malagon, K. J. Liu, ALK and GSK3: Shared Features of Neuroblastoma
1087 and Neural Crest Cells. *J Exp Neurosci*. **12**, 1179069518792499 (2018).

1088 53. S. Seal, A. H. Monsoro-Burq, Insights Into the Early Gene Regulatory Network
1089 Controlling Neural Crest and Placode Fate Choices at the Neural Border. *Front Physiol*.
1090 **11**, 608812 (2020).

1091 54. D. Roellig, J. Tan-Cabugao, S. Esaian, M. E. Bronner, Dynamic transcriptional signature
1092 and cell fate analysis reveals plasticity of individual neural plate border cells. *Elife*. **6**,
1093 e21620 (2017).

1094 55. C. R. T. Exner, A. Y. Kim, S. M. Mardjuki, R. M. Harland, sall1 and sall4 repress
1095 pou5f3 family expression to allow neural patterning, differentiation, and morphogenesis
1096 in Xenopus laevis. *Developmental Biology*. **425**, 33–43 (2017).

1097 56. M. Haas, J. L. Gómez Vázquez, D. I. Sun, H. T. Tran, M. Brislinger, A. Tasca, O.
1098 Shomroni, K. Vleminckx, P. Walentek, ΔN-Tp63 Mediates Wnt/β-Catenin-Induced
1099 Inhibition of Differentiation in Basal Stem Cells of Mucociliary Epithelia. *Cell Rep*. **28**,
1100 3338-3352.e6 (2019).

1101 57. C. Kalcheim, D. Kumar, Cell fate decisions during neural crest ontogeny. *Int J Dev Biol*.
1102 **61**, 195–203 (2017).

1103 58. A. Baggiolini, S. Varum, J. M. Mateos, D. Bettosini, N. John, M. Bonalli, U. Ziegler, L.
1104 Dimou, H. Clevers, R. Furrer, L. Sommer, Premigratory and migratory neural crest cells
1105 are multipotent in vivo. *Cell Stem Cell*. **16**, 314–322 (2015).

1106 59. J. Faber, P. D. Nieuwkoop, *Normal Table of Xenopus Laevis (Daudin): A Systematical &*
1107 *Chronological Survey of the Development from the Fertilized Egg till the End of*
1108 *Metamorphosis* (Garland Science, 2020).

1109 60. V. Savova, E. J. Pearl, E. Boke, A. Nag, I. Adzhubei, M. E. Horb, L. Peshkin,
1110 Transcriptomic insights into genetic diversity of protein-coding genes in X. laevis. *Dev*
1111 *Biol*. **424**, 181–188 (2017).

1112 61. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient
1113 alignment of short DNA sequences to the human genome. *Genome Biology*. **10**, R25
1114 (2009).

1115 62. A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A.
1116 Weitz, M. W. Kirschner, Droplet barcoding for single cell transcriptomics applied to
1117 embryonic stem cells. *Cell*. **161**, 1187–1201 (2015).

1118 63. G. Baruzzo, K. E. Hayer, E. J. Kim, B. Di Camillo, G. A. FitzGerald, G. R. Grant,
1119 Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat Methods*. **14**,
1120 135–139 (2017).

1121 64. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M.
1122 Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*.
1123 **29**, 15–21 (2013).

1124 65. V. Petukhov, J. Guo, N. Baryawno, N. Severe, D. T. Scadden, M. G. Samsonova, P. V.
1125 Kharchenko, dropEst: pipeline for accurate estimation of molecular counts in droplet-
1126 based single-cell RNA-seq experiments. *Genome Biology*. **19**, 78 (2018).

1127 66. M. Gao, M. Ling, X. Tang, S. Wang, X. Xiao, Y. Qiao, W. Yang, R. Yu, Comparison of
1128 high-throughput single-cell RNA sequencing data processing pipelines. *Briefings in*
1129 *Bioinformatics*. **22**, bbaa116 (2021).

1130 67. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M.
1131 Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell*.
1132 **177**, 1888-1902.e21 (2019).

1133 68. A. T. L. Lun, D. J. McCarthy, J. C. Marioni, A step-by-step workflow for low-level
1134 analysis of single-cell RNA-seq data with Bioconductor. *F1000Res*. **5**, 2122 (2016).

1135 69. C. Weinreb, S. Wolock, B. K. Tusi, M. Socolovsky, A. M. Klein, Fundamental limits on
1136 dynamic inference from single-cell snapshots. *PNAS*. **115**, E2467–E2476 (2018).

1137 70. M. Jacomy, T. Venturini, S. Heymann, M. Bastian, ForceAtlas2, a Continuous Graph
1138 Layout Algorithm for Handy Network Visualization Designed for the Gephi Software.
1139 *PLOS ONE*. **9**, e98679 (2014).

1140 71. A. E. Wills, R. Gupta, E. Chuong, J. C. Baker, Chromatin immunoprecipitation and deep
1141 sequencing in Xenopus tropicalis and Xenopus laevis. *Methods*. **66**, 410–421 (2014).

1142 72. C. Milet, A. H. Monsoro-Burq, Dissection of Xenopus laevis neural crest for in vitro
1143 explant culture or in vivo transplantation. *J Vis Exp* (2014), doi:10.3791/51118.

1144 73. T. Sato, N. Sasai, Y. Sasai, Neural crest determination by co-activation of Pax3 and Zic1
1145 genes in Xenopus ectoderm. *Development*. **132**, 2355–2363 (2005).

74. C.-S. Hong, J.-P. Saint-Jeannet, The Activity of Pax3 and Zic1 Regulates Three Distinct Cell Fates at the Neural Plate Border. *MBoC*. **18**, 2192–2202 (2007).

75. R. F. Spokony, Y. Aoki, N. Saint-Germain, E. Magner-Fink, J.-P. Saint-Jeannet, The transcription factor Sox9 is required for cranial neural crest development in Xenopus. *Development*. **129**, 421–432 (2002).

76. Monsoro-Burq AH. A rapid protocol for whole-mount in situ hybridization on Xenopus embryos. CSH Protoc. 2007 Aug 1;2007:pdb.prot4809. doi: 10.1101/pdb.prot4809. PubMed PMID: 21357147.

Supplementary references

77. M. Nichane, X. Ren, J. Souopgui, E. J. Bellefroid, Hairy2 functions through both DNA-binding and non DNA-binding mechanisms at the neural plate border in Xenopus. *Developmental Biology*. **322**, 368–380 (2008).

Acknowledgements

The authors are grateful to Drs. Andrej Zinoviev, Gerhard Schlosser, Igor Adameyko and Thomas Walter for insightful scientific discussions and comments on the manuscript. We thank C. Lantoine for animal husbandry, Q. Thuillier for technical assistance. We thank J. Briggs for help in single cell sequencing. We thank J.L. Plouhinec for initial analysis of small RNAseq. High-throughput sequencing, except for single cell sequencing, has been performed by the ICGex NGS platform of the Institut Curie supported by the grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the Agence Nationale de la Recherche ("Investissements d'Avenir" program), by the Canceropole Ile-de-France and by the SiRIC-Curie program -SiRIC Grant « INCa-DGOS-4654 ».

Funding:
This project receives funding from
European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860635, NEUcrest ITN (AHMB)
Agence Nationale pour la Recherche (ANR-15-CE13-0012-01; AHMB)
Agence Nationale pour la Recherche ANR-21-CE13-0028; AHMB)
Institut Universitaire de France (AHMB)
National Institutes of Health NICHD award R01HD073104 (LP)
National Institutes of Health NIH R01 GM42341 (RMH)
National Institutes of Health NIH R35GM127069 (RMH)

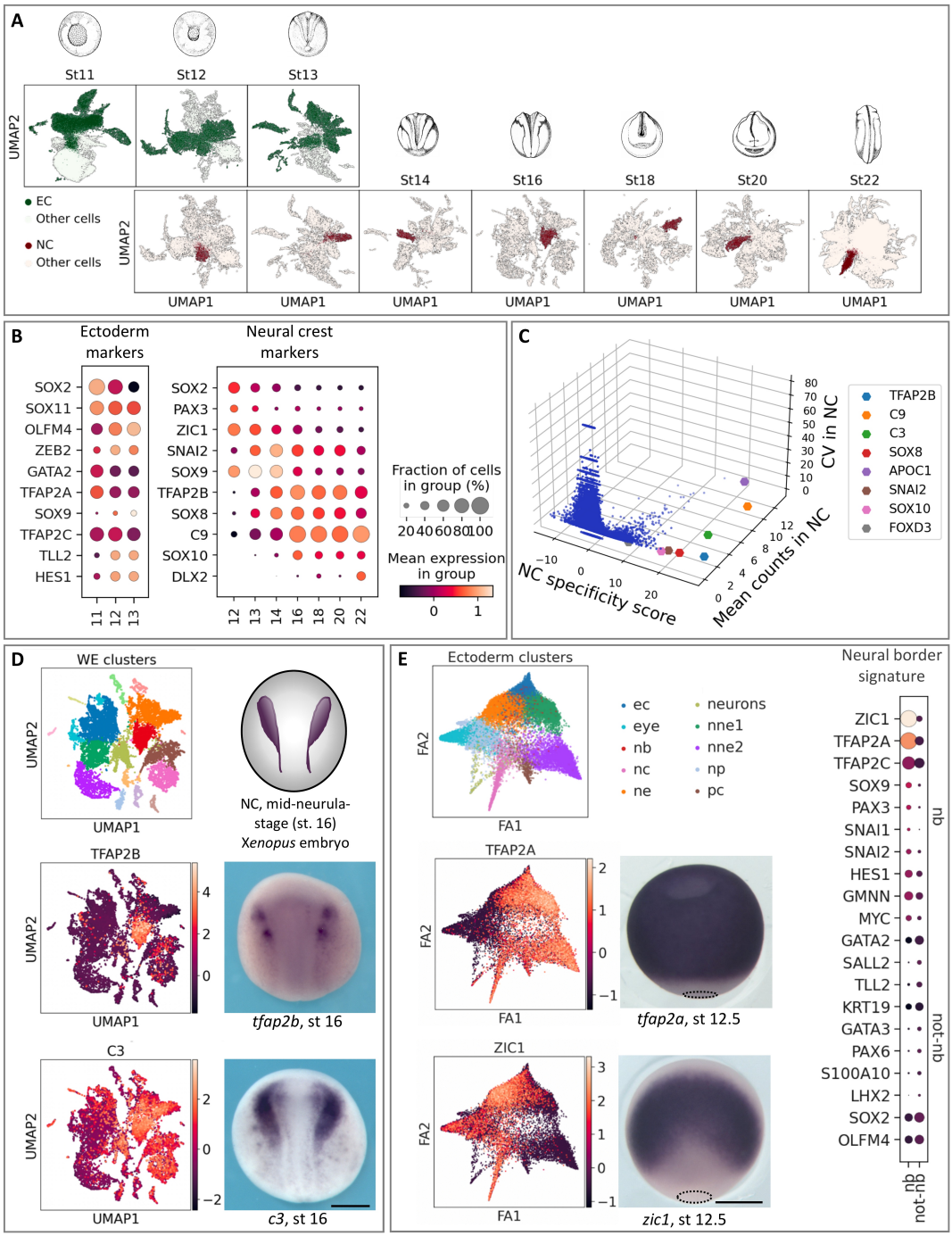
Author contributions:
Conceptualization: AK, SS, LP, AHMB
Methodology: AK, MA, SS, VK, SMR, HA, LP, AHMB
Investigation: AK, MA, SS, VK, HA, LP, AHMB
Visualization: AK, MA, SS, VK, AHMB
Supervision: LP, RMH, AHMB
Writing—original draft: AK, SS, AHMB
Writing—review & editing: AK, SS, RMH, AHMB

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials. Accession numbers to the datasets, together with their description are under NCBI Gene Expression Omnibus # GSE198494.

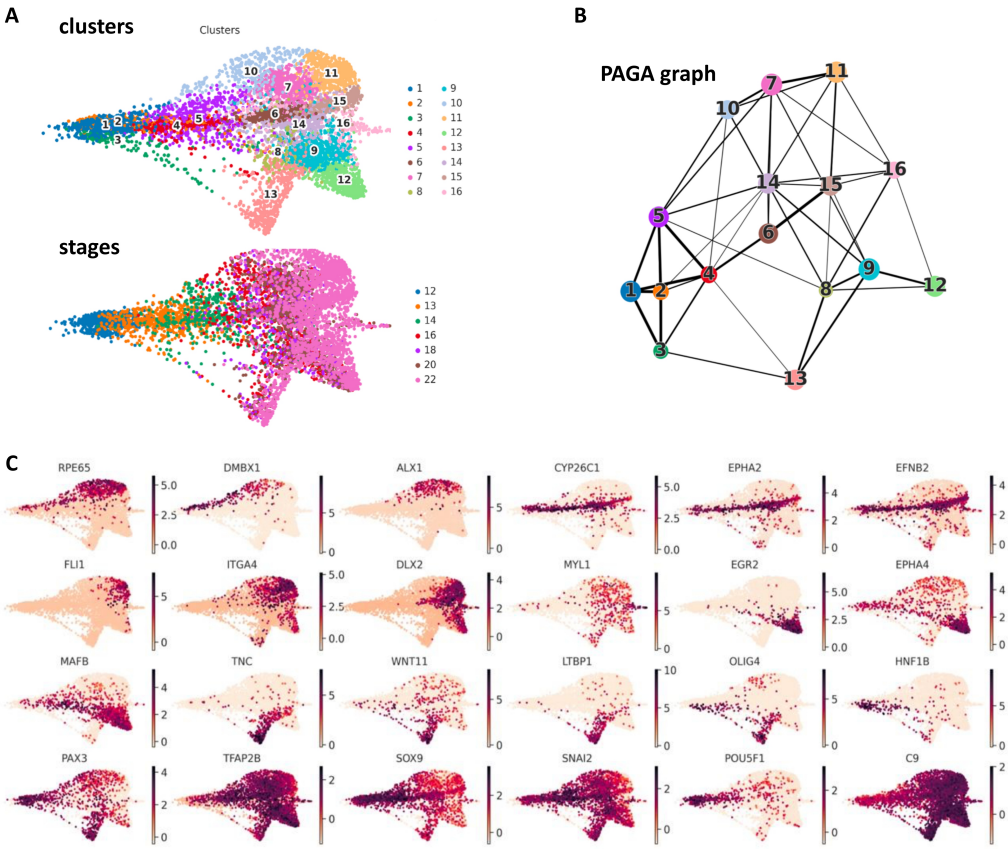
Figures and Tables

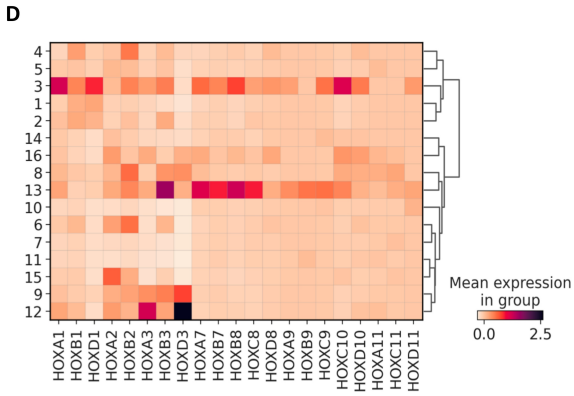
Kotov et al., Figure 1



1203 **Fig. 1. Cell selection for neural crest, ectoderm and neural border. (A**) Ectoderm
1204 (EC, stages 11-13, green) and NC cells (stages 12-22, brown) were selected
1205 from a whole embryo SC transcriptome dataset of 177250 cells. (**B)** Dotplots
1206 for well-referenced gene expressions used to identify EC and NC at each stage.
1207 Dot size represents the number of cells expressing the gene, color represents
1208 the average expression level. (**C)** 3D scatter-plot of NC score specificity (z-
1209 scores), mean gene expression levels and coefficient of variation (CV) in NC
1210 cells, defining a few highly expressed pan-NC genes during neurulation. (**D)**
1211 Mid-neurula stage 16 whole embryo UMAP and ISH show expression of
1212 *tfap2b,* specifically in NC cells, and of *c3*, in NC cells and also neural and non
1213 neural ectoderm. (**E)** NB signature was defined by differential expression
1214 analysis of cells highly co-expressing *tfap2a* and *zic1* (nb, UMAP and ISH)
1215 compared to the other ectoderm cells (not-nb): in addition to *tfap2a* and *zic1*,
1216 the genes enriched in the NB zone are *sox9, pax3, snai1/2, hes1, gmnn and*
1217 *myc*. Scale bar, 500 µm.

Kotov et al., Figure 2



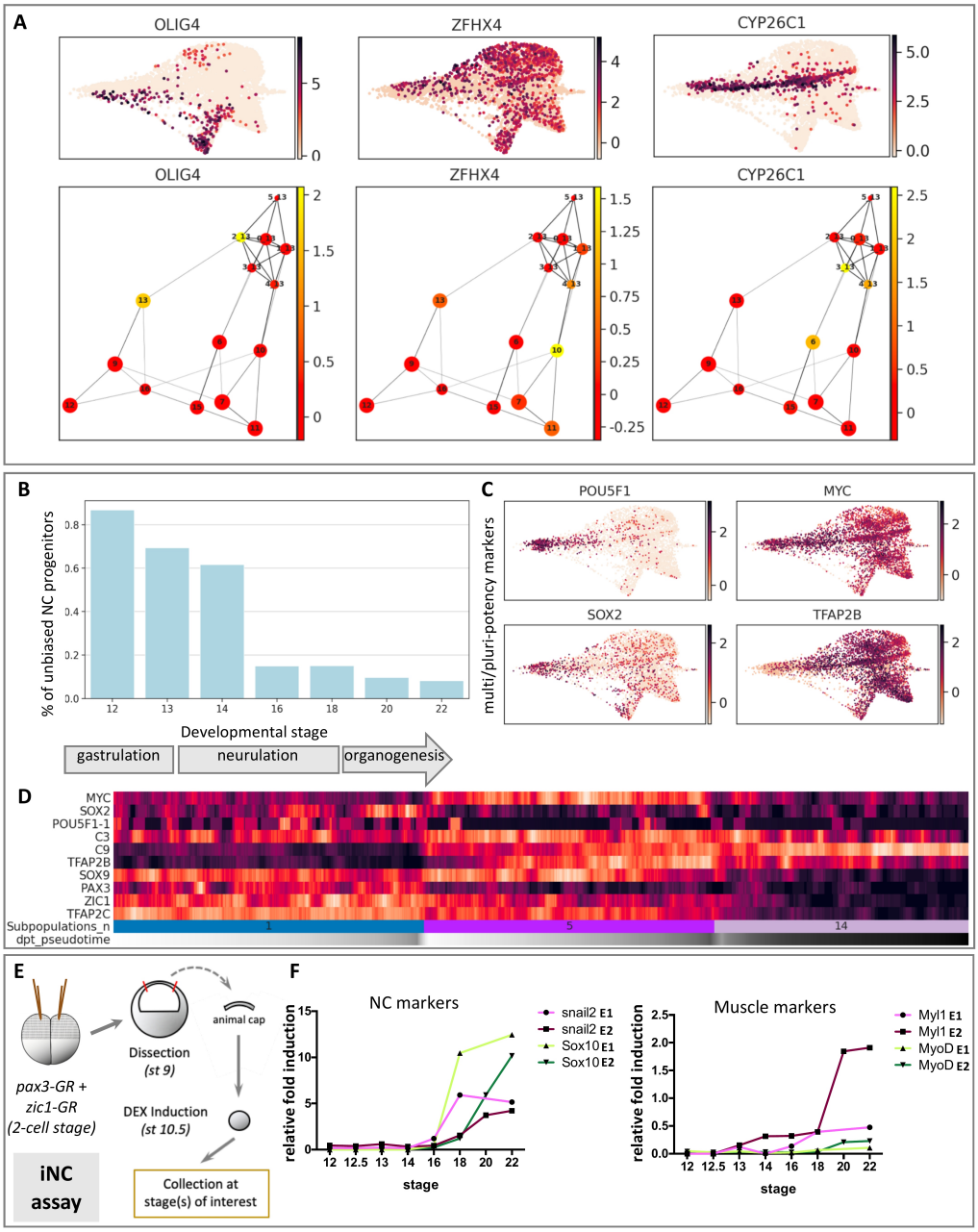


E

1- Early unbiased / *tfap2c*
2- Early unbiased / *gjb1*
3- Premigratory sacral / *hnf1b*
4- Premigratory early R4 /*cyp26c1*
5- Premigratory unbiased
6- Premigratory late R4 / *cyp26c1*
7- Migratory bipotent cranial / *nrp2*
8- Migratory rhombencephalic / *hoxb2*
9- Migratory bipotent vagal / *hoxd3*
10- Premigratory Early cranial / *alx1*
11- Migratory cranial / *itga4*
12- EMT Cardiac / *egr2*
13- Migratory ENS-p / *tnc*
14- EMT unbiased
15- Migratory late R4 / *efnb2*
16- Migratory muscle-like /*myl1*

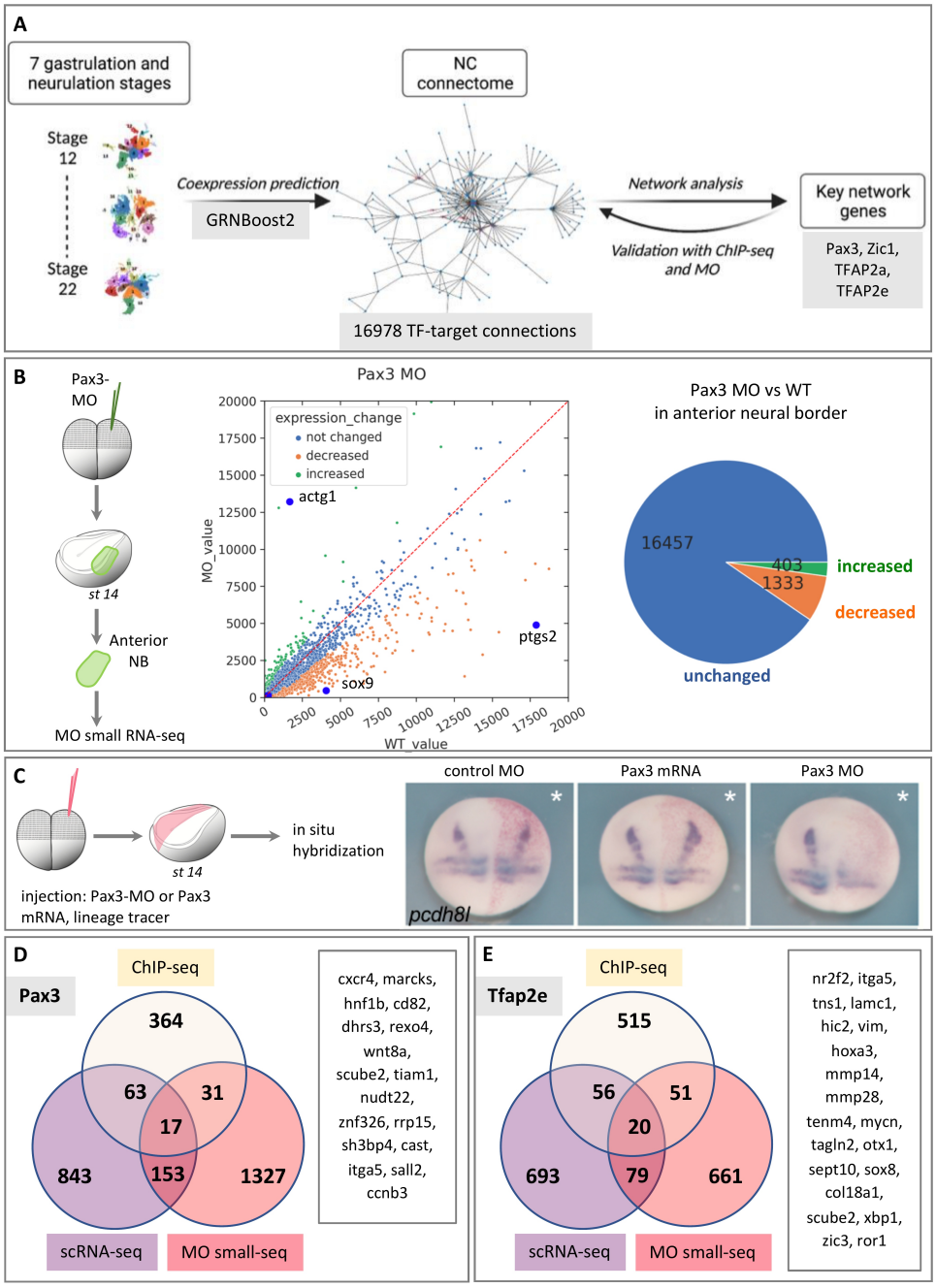
1219 **Fig. 2. Premigratory neural crest transcriptome heterogeneity. (A**) Forced-directed
1220 graph for the NC dataset (Leiden clustering) revealed 16 distinct states
1221 (clusters) before and during EMT (developmental stages 12-22). **(B)** Principle
1222 Graph analysis (PAGA) estimates clusters connectivity which is strong during
1223 NC induction stages 12-13, yet showing early NC heterogeneity, with early
1224 emergence of *cyp26c1*$^{+}$ clusters #4/6/15 and ENSp cluster #3. Line thickness
1225 increases with higher connections **(C)** Expression of key cluster-specific genes,
1226 including *rpe65, dmbx1, alx1* (#10), *cyp26c1, epha2, efnb2* (#4/6/15), *fli1,*
1227 *itga4, dlx2* (#11), *egr2, epha4, mafb* (#12), *tnc, wnt11, ltbp1* (#13), early *olig4,*
1228 *hnf1b* (#3*)* and muscle-like NC specific *myl1* (#16). Genes expressed broadly in
1229 NC cells define a "canonical NC" signature: early *pax3, tfap2b, sox9, snai2,*
1230 *and c9.* Multipotency-related genes are present mostly until mid-neurula stage
1231 *(pou5f1).* **(D)** The hox gene signature of each cluster provides its approximate
1232 position along the antero-posterior body axis. **(E)** Each cluster's main
1233 expression characteristics used to name them.

Kotov et al., Figure 3



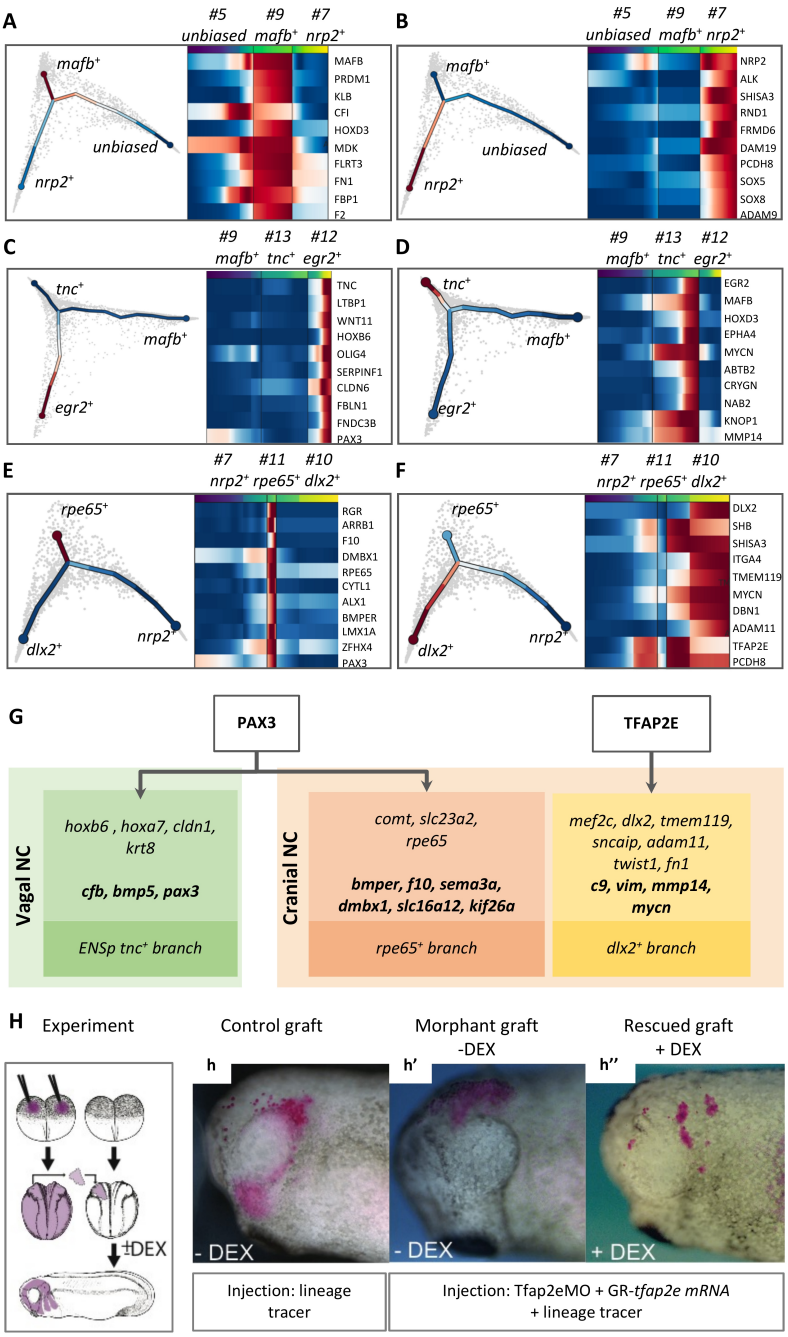
1235 **Fig. 3. Emergence of heterogeneity from unbiased NC during NC induction. (A)**
1236 From the end of gastrulation, three previously undescribed early trends emerge.
1237 Clustering of stage 13 cells only, followed by PAGA analysis identified their
1238 similarities with the stage 20-22 clusters: an *olig4*-enriched subpopulation
1239 linked to ENSp cluster #13; a *zfhx4*-enriched cluster related to *rpe65*$^{+}$ cluster
1240 #11 and a third stage 13 subpopulation expressing *cyp26c1* linked to the cranial
1241 NC branch #4, #6, #15. **(B, C, D)** Until mid-neurulation (st. 14), most cranial
1242 and vagal NC cells expressed a canonical NC stem cell-like signature (*sox9,*
1243 *tfap2c, snai2)* together with high expression of pluripotency markers (*e.g.*
1244 *pou5f1, cmyc,* and *sox2).* No indication of bias toward a specific fate was
1245 detected in clusters #1, #5 and EMT-stage cluster #14. (B) Expression patterns
1246 in pseudotime for unbiased clusters #1, #5, #14; B) Proportions of the unbiased
1247 cells in the NC dataset; C) UMAP plots. **(E, F)** At early migration stage (20-
1248 22), cluster 16 expresses myosin-like genes. Direct NC induction in ectoderm
1249 explants *in vitro* (iNC assay) followed by RTqPCR, identifies *myoD* and *myl1*
1250 expression at a similar stage.

Kotov et al., Figure 4



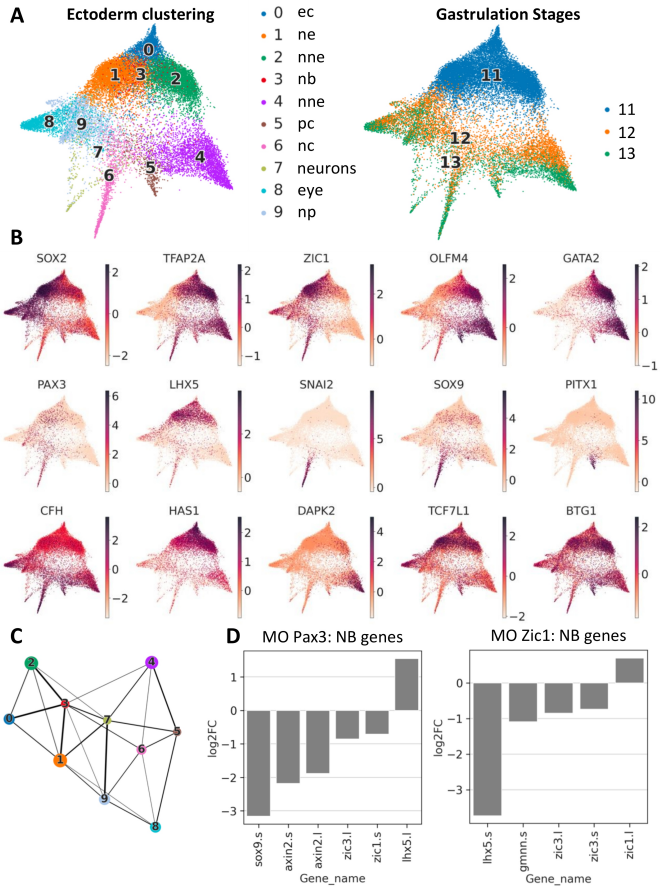
1252 **Fig. 4. NC connectome generation and validation. (A)** NC cells from all stages were
1253 used to build a large-scale connectome with GRNBoost2. ChIP-seq and MO
1254 knockdown small RNA-seq were used to validate a significant number of links
1255 predicted for *pax3, tfap2e, tfap2a and zic1*. **(B)** Overview of *pax3* MO analysis
1256 in NB/NC. Most affected genes were decreased indicating that Pax3 mostly
1257 displays transcriptional activation action in NC. **(C)** *Pcdh8l,* a target of Pax3 is
1258 regulated by Pax3 gain-of-function (mRNA) and loss-of-function (MO) *in vivo*.
1259 ISH, Injected side is traced in red. **(D, E)** Venn diagrams of target genes
1260 validated by ChIP-seq and MO-small-seq for *pax3* (D) and *tfap2e* (E). Genes
1261 linked to *pax3* and *tfap2e* by the three methods are listed.

Kotov et al., Figure 5



1263 **Fig. 5. NC branching analysis.** Cells sub-selected around a chosen bifurcation point
1264 were analyzed using tree inference, advanced pseudotime downstream analysis,
1265 and scFates. Gene programs for bifurcation of premigratory unbiased cluster #5
1266 into clusters #7 **(A)** and #9 **(B);** of migratory bipotent vagal cluster #9 into
1267 clusters #12 **(C)** and #13 **(D)** and of migratory bipotent cranial cluster #7 into
1268 clusters #11 **(E)** and #10 **(F). (G)** MO-mediated depletion and ChIP-seq
1269 validate Pax3 function in both vagal (clusters #12 vs #13) and cranial (#11 vs
1270 #10) branching, and TFAP2e function in cranial (#10 vs #11) branching. **(H)**
1271 Premigratory NC (pink) was grafted into host embryos. In comparison to
1272 control (h), TFAP2e morphant NC (h') does not emigrate towards craniofacial
1273 areas. Activation of TFAP2e-GR in morphant cells upon EMT stage restores
1274 migration (h").

Kotov et al., Figure 6



1276 **Fig. 6. Ectoderm dataset analysis. (A)** Leiden clustering using 17138 early ectoderm
1277 cells from 11, 12, 13 NF stages. *0) Ectoderm stage 11; 1) neural ectoderm; 2)*
1278 *non-neural ectoderm; 3) NB zone; 4) non-neural ectoderm stage 13; 5)*
1279 *placodes; 6) neural crest; 7) neurons; 8) eye primordium; 9) neural plate.* **(B)**
1280 UMAP plots of key genes in the Ectoderm dataset. **(C)** PAGA plot for
1281 Ectoderm dataset reveals a central role of the NB zone (3) between ne, nne, nc
1282 and pc. **(D)** The NB signature is affected by Pax3 or Zic1 depletion.

Kotov et al., Figure 7

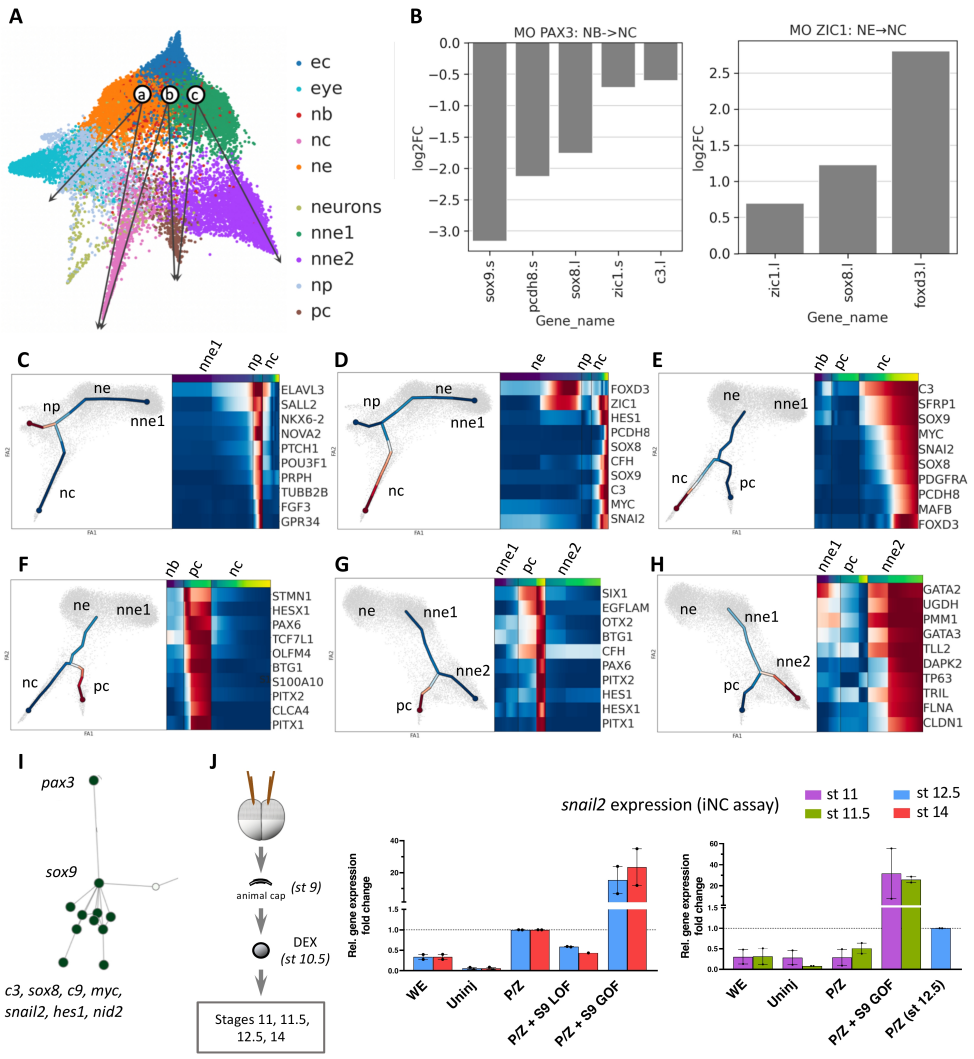
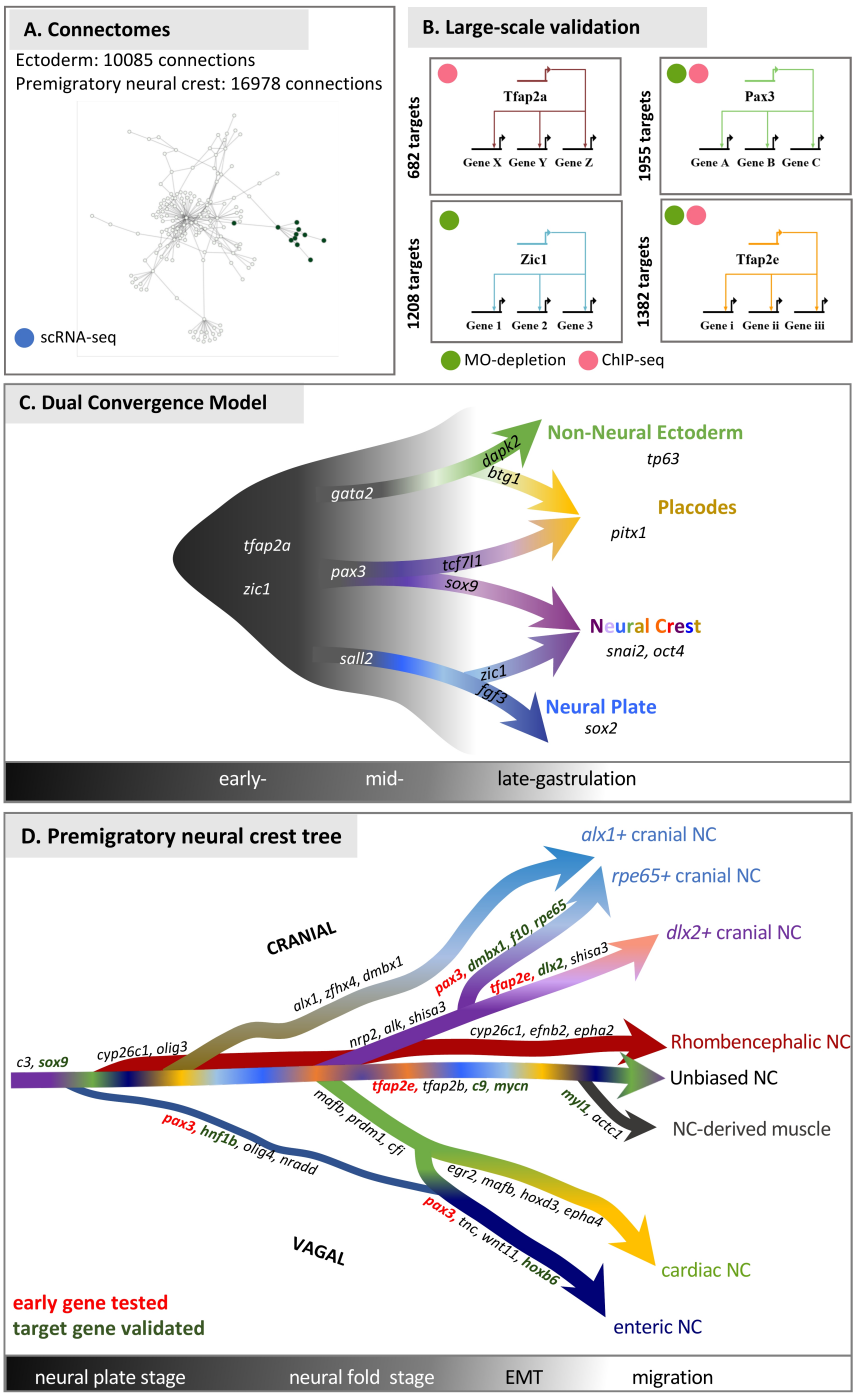


Fig. 7. Ectoderm branching and validation of Sox9 early role in NC induction. (A) Branching analysis retrieves the possible gene programs for NC and PC development at stages 11-13, considering two sources of origin for both populations: NP and NB for NC, and NNE and NB for PC. It revealed that Zic1 is an important (connectome) and early gene for branch NE->NC, while Pax3 is early gene for branch NB->NC. Pax3 depletion impacts expression of several branch-specific genes, including *sox9, foxd3,* as well as *zic1* while Zic1 depletion increases *sox8* and *foxd3*. **(C)** NNE1->NP gene program, **(D)** NE->NC gene program, **(E)** NB->NC gene program, **(F)** NB->PC gene program, (**G)** NNE1->PC gene program, **(H) NNE1->**NNE2 gene program..**(I)** NB connectome analysis suggested a novel epistasis relationship between *pax3*, *sox9* and other downstream NC specifiers. **(J)** In iNC assay, Sox9 acts downstream of Pax3 and is essential for activating the downstream NC program. Additionally, at gastrula stages, a time point at which *pax3*/*zic1* activation does not yet induce *snail2* expression in iNC, adding Sox9 activation is sufficient to obtain high levels of *snail2* precociously. RT-qPCR analysis showing relative *snail2* expression fold change in iNC at late gastrula and early neurula stages. WE - whole embryo; Uninj - uninjected animal caps; P/Z - *pax3*-GR + *zic1*-GR iNC; S9 LOF - *sox9* loss-of-function (LOF) or gain-of-function (GOF).

Kotov et al., Figure 8



1305 **Fig. 8. Neural crest GRN and developmental trajectories. (A)** GRNBoost2-
1306 predicted connectomes for Ectoderm (gastrula stage) and Neural crest
1307 (neurulation stage). **(B)** RNA-seq on dissected neural border/neural crest
1308 explants or ChIP-seq provide large-scale GRNs linked to Pax3, Zic1, TFAP2a,
1309 and TFAP2e. **(C)** At gastrula stage (time line indicated), the neural border cells
1310 expressing *tfap2a, zic1*and *pax3* present trajectories towards placodes and
1311 neural crest. Those trajectories converge with a trajectory from neural plate
1312 towards neural crest and another one from non-neural ectoderm towards
1313 placodes. Branching analysis highlights a gene signature underlying those
1314 transcriptome transitions (the top gene is indicated here). (D) Lineage tree of
1315 the neural crest cells, from the end of gastrulation, during neurulation and upon
1316 epithelial-mesenchymal transition, ending at early migratory stage (time line
1317 indicated). Gene signatures supporting each trajectory is summarized. The
1318 function of transcription factors expressed prior branching (Pax3, Tfap2e, red)
1319 was validated on expression of genes expressed after branching (green).

1320 **Table 1. Highly connected genes in the GRNboost2 neural crest network.**
1321

Gene	Degree centrality	Betweeness centrality	Max. expression stage	NC specificity
TFAP2C	28.028933	21.510673	12	6.59
ZIC3	16.455696	7.885226	12	3.73
ZIC1	8.137432	2.351524	12	9.34
ZIC2	7.233273	2.026896	12	3.62
DMBX1	3.616637	11.876445	12	1.57
PAX3	3.435805	2.211830	12	5.93
SOX9	11.573237	8.575830	13	14.07
OLIG4	4.882459	2.642370	13	4.54
MAFB	7.956600	4.174946	14	09.08
SNAI2	5.244123	0.927599	14	20.03
SOX8	4.520796	2.071516	14	20.60
RPE65	1.446655	0.534404	14	6.49
TFAP2B	8.499096	4.135991	16	25.47
HOXB3	3.978300	0.959282	16	2.35
SOX10	3.797468	0.705758	16	15.40
NR6A1	3.797468	0.706433	16	9.18
ZFHX4	2.169982	0.791908	16	4.83
ALX1	2.169982	1.627017	16	3.69
EPHA2	1.989150	1.403078	18	3.77
E2F3	1.446655	2.256467	18	0.68
TFAP2E	5.786618	1.850976	20	15.21
EGR2	5.424955	2.505897	20	5.96
MYCN	17.359855	12.737248	22	8.81
DLX2	12.839060	8.413243	22	6.88
HOXD3	8.137432	3.417475	22	4.96
EEF1D	4.701627	0.795235	22	9.55
ZBTB16	3.074141	1.013930	22	2.80
TFEC	2.169982	0.562422	22	0.10
VIM	2.169982	0.879581	22	2.43

1322

1325 **Table 2. Highly connected genes in the GRNboost2 ectoderm network**

1326

Gene	Degree centrality	Betweeness centrality	Mean nc/pc/nb specificity	NC specificity	PC specificity	NB specificity
TFAP2C	70.319635	6.203872	8.778883	17.587187	3.636364	5.113099
TFAP2A	67.123288	5.770185	11.238043	4.836522	-0.884711	29.762320
ZIC1	47.488584	1.674848	15.645578	18.120411	-8.226690	37.043015
HES1	30.136986	0.815641	13.273738	27.994427	11.076355	0.750431
SOX9	24.657534	0.463349	11.848502	38.367329	-4.275626	1.453803
MYC	17.808219	0.340629	9.269785	22.733620	4.699314	0.376421
SNAI2	17.579909	0.192783	7.027262	27.004408	-3.017278	-2.905344
PITX1	16.438356	0.115279	5.463268	-3.656815	21.928877	-1.882259
PAX3	14.611872	0.172905	5.509155	15.614752	-3.379381	4.292094
SOX8	12.100457	0.094409	5.176289	19.021635	-2.535636	-0.957131
C3	9.817352	0.019518	12.694010	36.559040	6.013386	-4.490397

1327

From neural border to migratory stage: A comprehensive single cell roadmap of the timing and regulatory logic driving cranial and vagal neural crest emergence

Supplementary Materials

Aleksandr Kotov, Mansour Alkobtawi†, Subham Seal†, Vincent Kappès, Sofia Medina, Hugo Arbès, Richard Harland, Leonid Peshkin, and Anne H. Monsoro-Burq*

*Corresponding author. Email: anne-helene.monsoro-burq@curie.fr

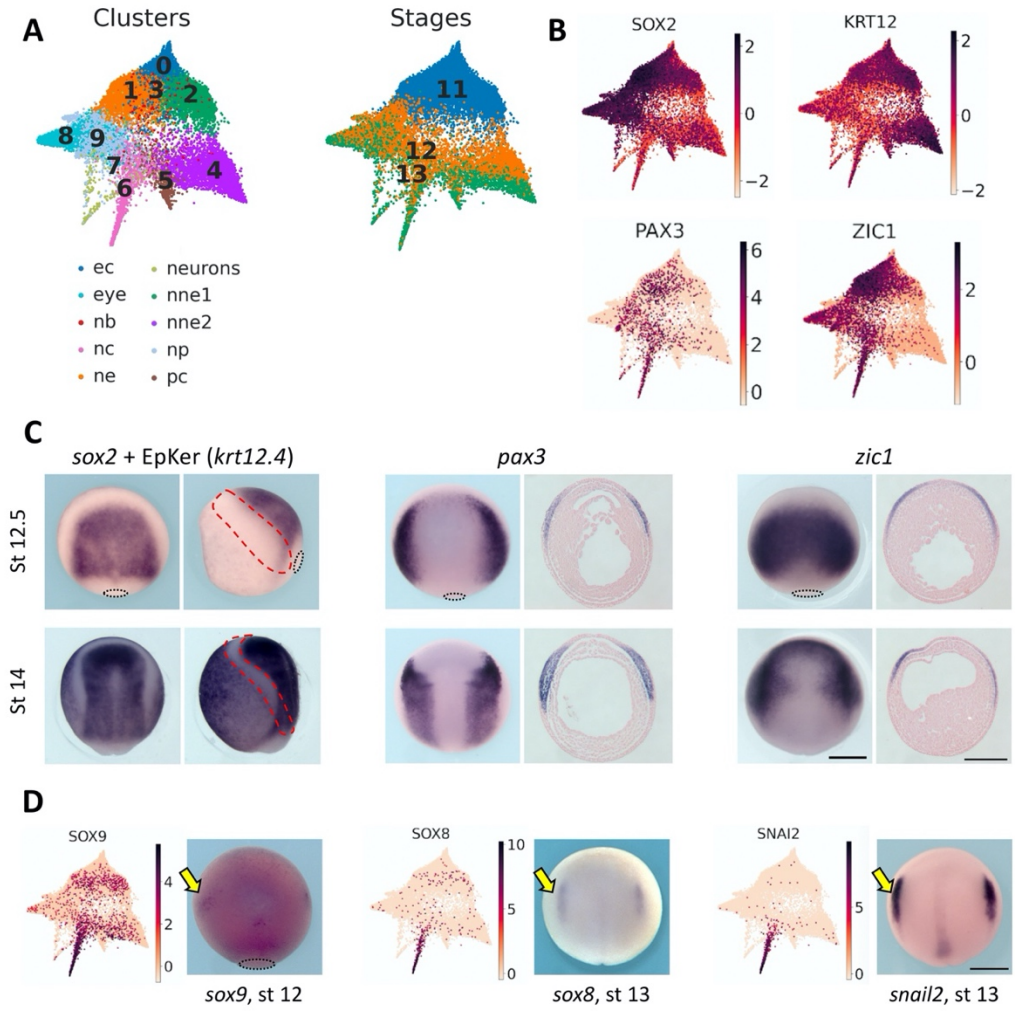


Fig. S1. **Markers for NB zone and early NC.**
(A) Ectoderm UMAP plots for stages 11,12 and 13 depicting different clusters and their corresponding stage. (B) Gene markers for the 3 major ectoderm derivatives. *sox2* marks the neural ectoderm while *krt12* (epidermal keratin 12) marks the non-neural ectoderm. Unlike the other derivatives, the NB zone does not yet have a strictly specific gene marker and is usually depicted by *pax3* expression or the overlapping expressions of genes like *tfap2a* and *zic1*. (C) ISH images of embryos at stage 12.5 (late gastrula) and stage 14 (mid neurula) depicting the gene markers for the ectoderm derivatives mentioned in (B). The red dotted region depicts the neural border zone, devoid of both *sox2* and *krt12* expression. Scale bar, 500 μm. (D) EC UMAP and corresponding ISH images depict the earliest stages of detection of 3 well-known early NC markers. *Sox9* is initiated at the prospective NC zone at stage 12. *Sox8* and *snail2* are detected at the early NC at stage 13. Yellow arrows point to the NC region. Scale bar, 500 μm.

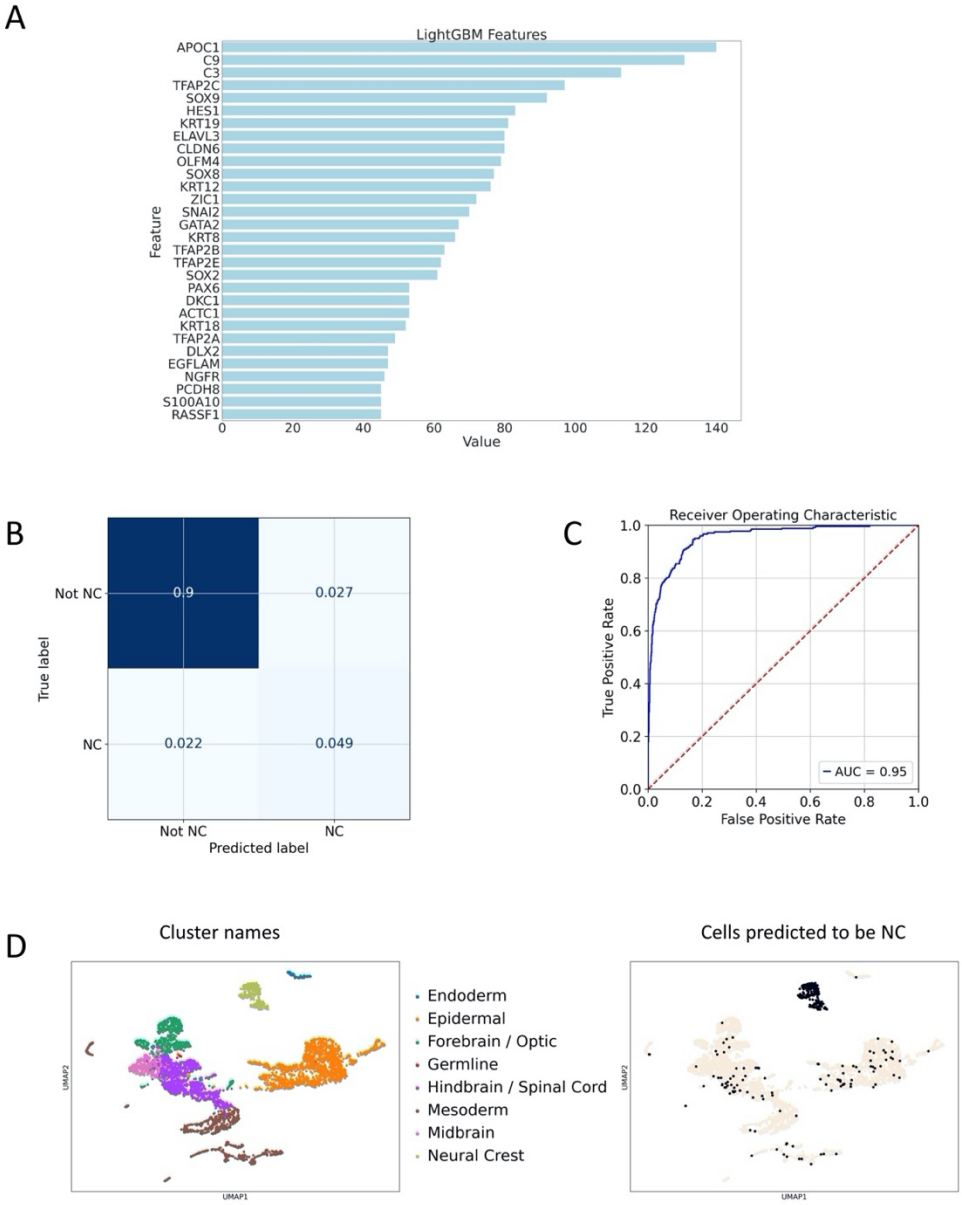


Fig. S2. **LightGBM NC classifier.**
LightGBM-based binary classificator for the NC cells. Using the whole embryo data matrix of gene expression, we have selected the top 10% LightGBM model important features for NC detection. In order to validate it in the Zebrafish dataset (21), we only kept genes that were in the Zebrafish dataset. Using only top features, we re-built the model (test results: 0.99 accuracy and 0.90 f1 score) and validated it in Zebrafish dataset with the accuracy 0.95 and f1 score 0.66 in 14 stage cells. **(A)** Top features for NC classification using LightGBM model. **(B)** Confusion matrix for validation in Zebrafish dataset for stage 14. **(C)** ROC curve for binary NC classifier predicted scores. **(D)** UMAP for Zebrafish 14 stage cells with model predicted cells as NC. Prediction accuracy is enough for detection of cluster of interest (NC) despite of technical and biological batch effect between Xenopus and Zebrafish datasets.

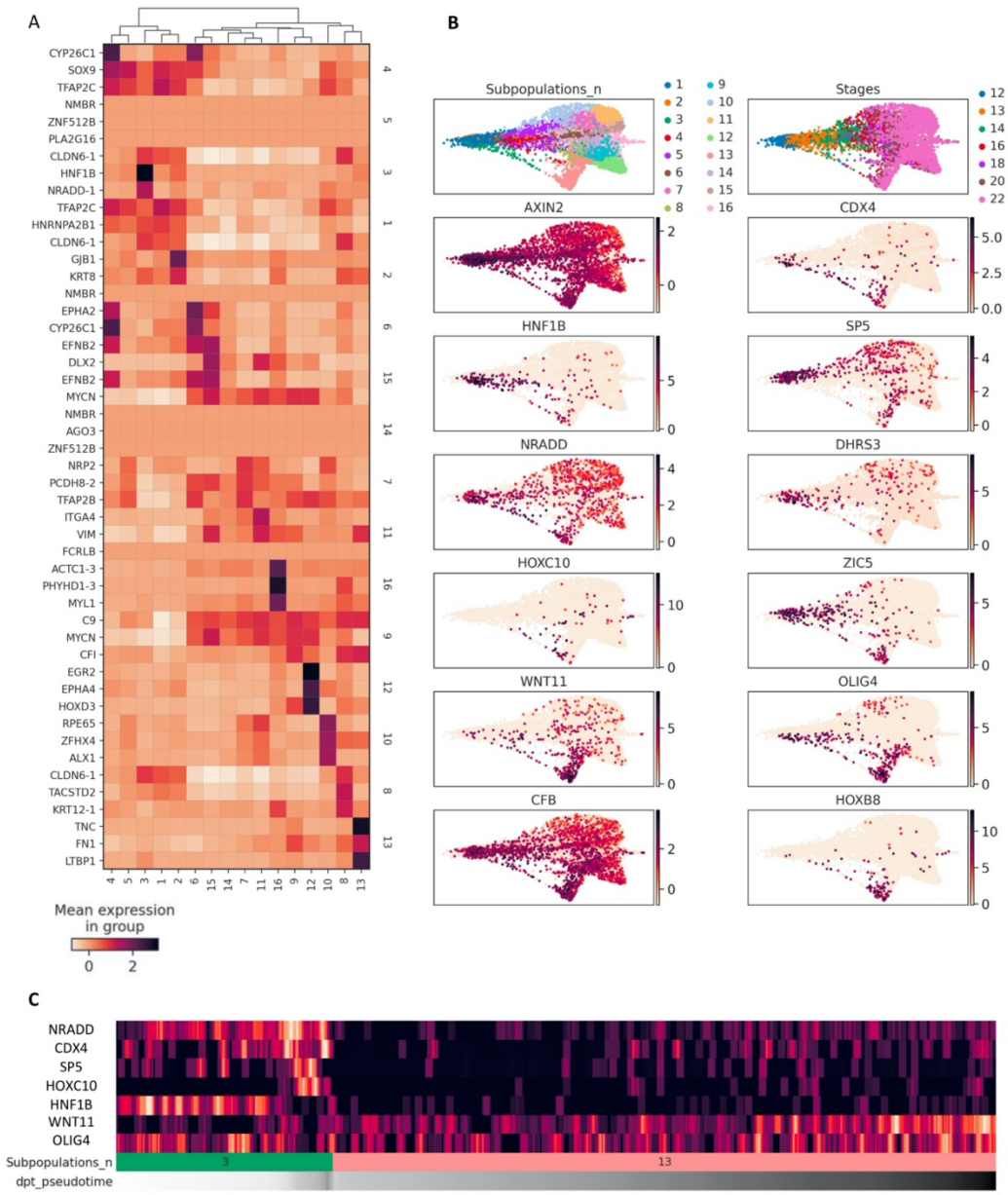


Fig. S3. **Cluster-specific genes.**
(A) Matrixplot with the top 3 specific genes per cluster for the Neural Crest dataset. **(B)** UMAP plots for the early-biased NC subpopulation expressing *hnf1b*, *cdx4*, *nradd* and posterior hox genes. In addition to cluster-specific *hnf1b* and *olig4*, *dhrs3* is co-expressed with *zic5* which specifically stays active during the whole "sacral" trajectory from stage 12 to late ENSp. Generally, *dhrs3* is expressed in cells that do not express *cyp26c1*. *Dhrs3* thus participates since stage 12 to a strong demarcation, earlier than posterior HOX expression appearing later. **(C)** Expression of cluster 3 specific genes along the pseudotime.

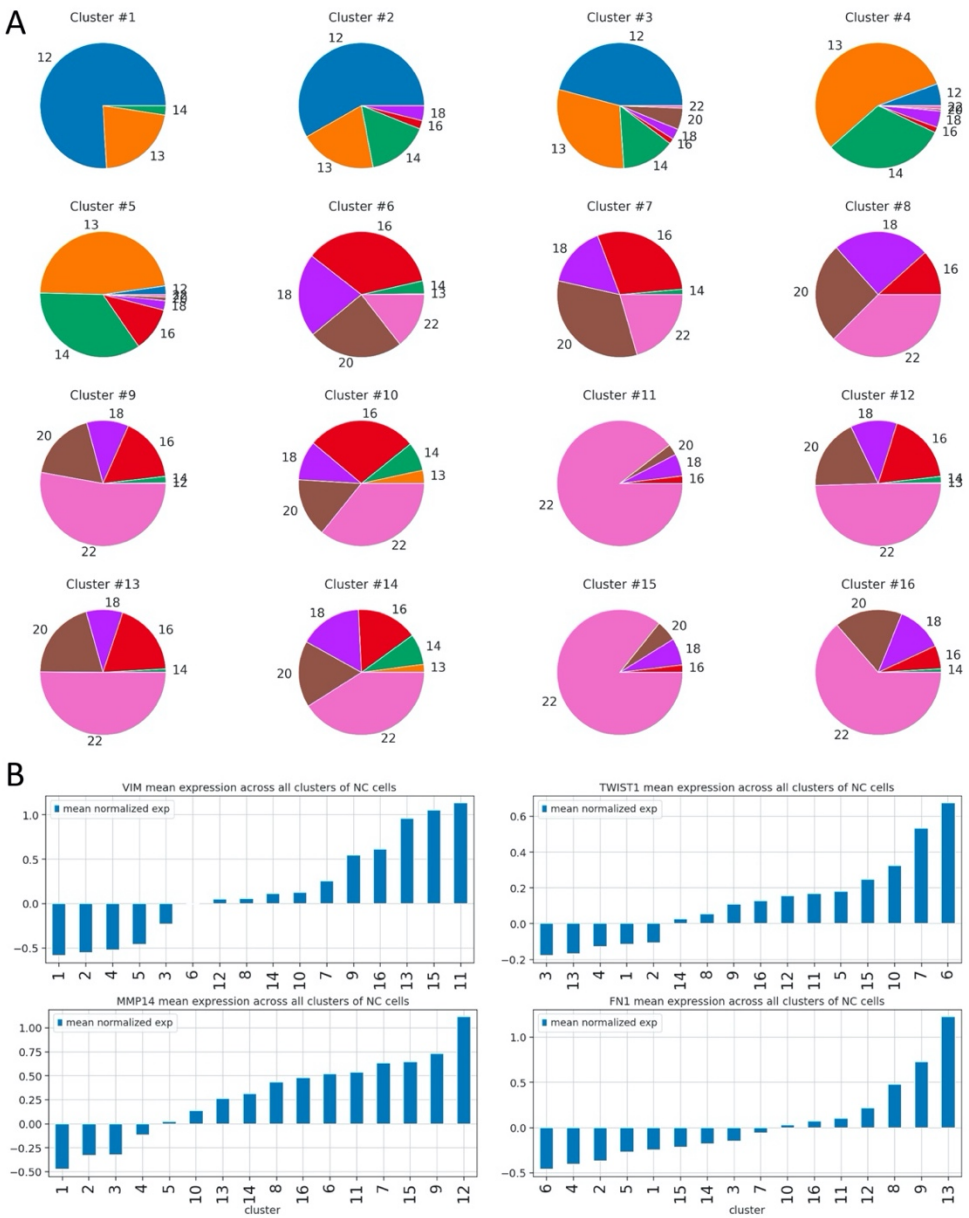


Fig. S4. **Developmental stages and expression of EMT markers per cluster.**
(A) Pieplots represent the proportion of cells at each stage in each cluster. Some clusters include cells of 6 stages, e.g. cluster 10. **(B)** Each cluster presents various levels of expression of EMT and migration regulators. Bar plots with mean normalized expression for each cluster for *vim, twist1, mmp14* and *fn1*.

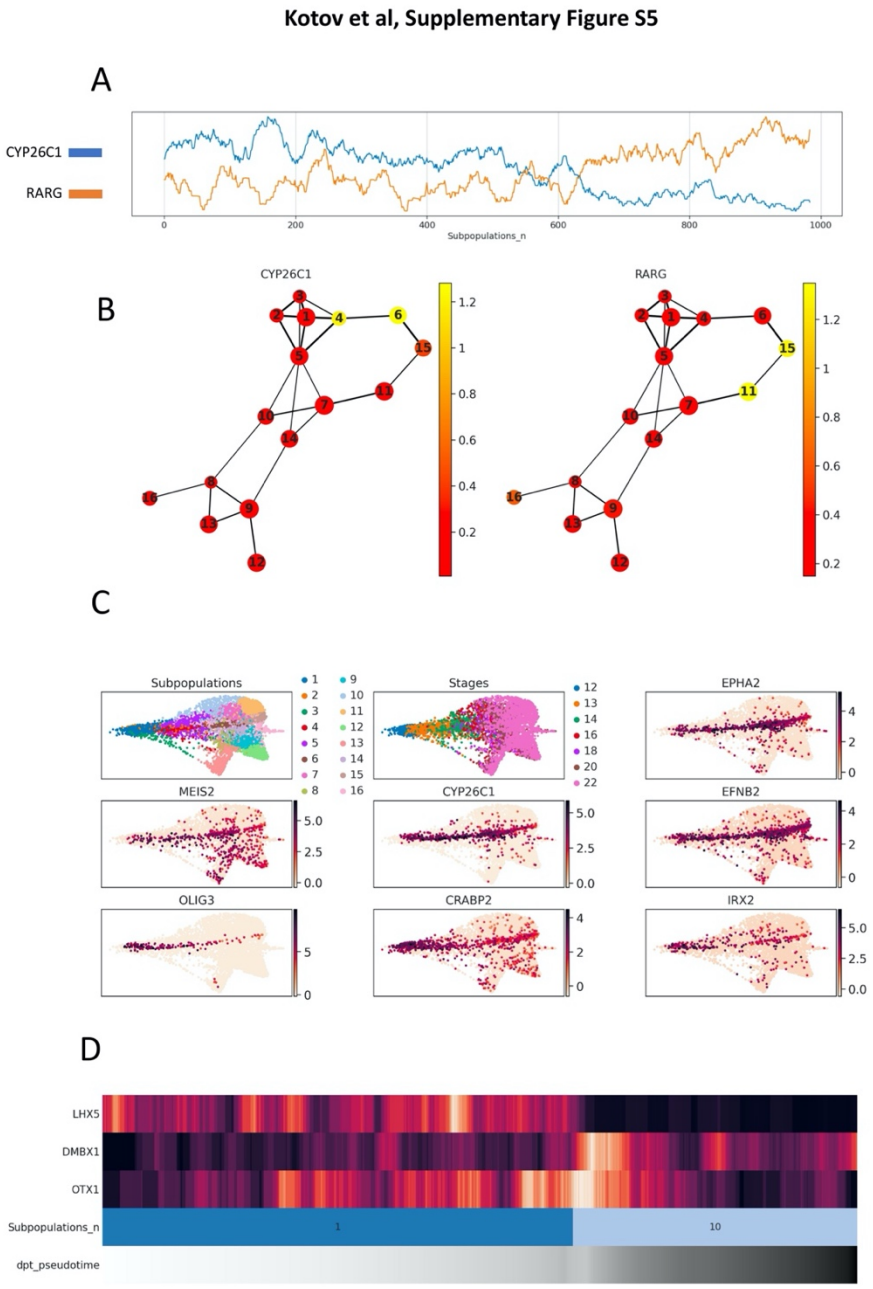


Fig. S5. **Early posterior enteric cluster #4 and related late clusters #6, #15.**
(A) Line Plot for expression dynamics for cluster 4, 6, 15, representing the relationships between *rarg* and *cyp26c1*. Cells within clusters are sorted according to pseudotime. **(B)** PAGA plot with *rarg* and *cyp26c1* mean expression. **(C)** Key specific genes for RA stream (clusters 4, 6 and 15). In the middle RA-dependent cluster #6, we observed increased expression of *crabp2*, which delivers all-trans retinoic acid to retinoic acid receptors (RARs, RXRs). Interestingly, during NC induction *crabp2* is expressed homogeneously in unbiased NC and early RA-dependent CNC, but later the *crabp2* expression remains only in the *cyp26c1*$^{+}$ stream. **(D)** Expression heatmap for early cranial cluster 10 markers *dmbx1* and *otx1* and multipotent cluster 1 marker *lhx5*.

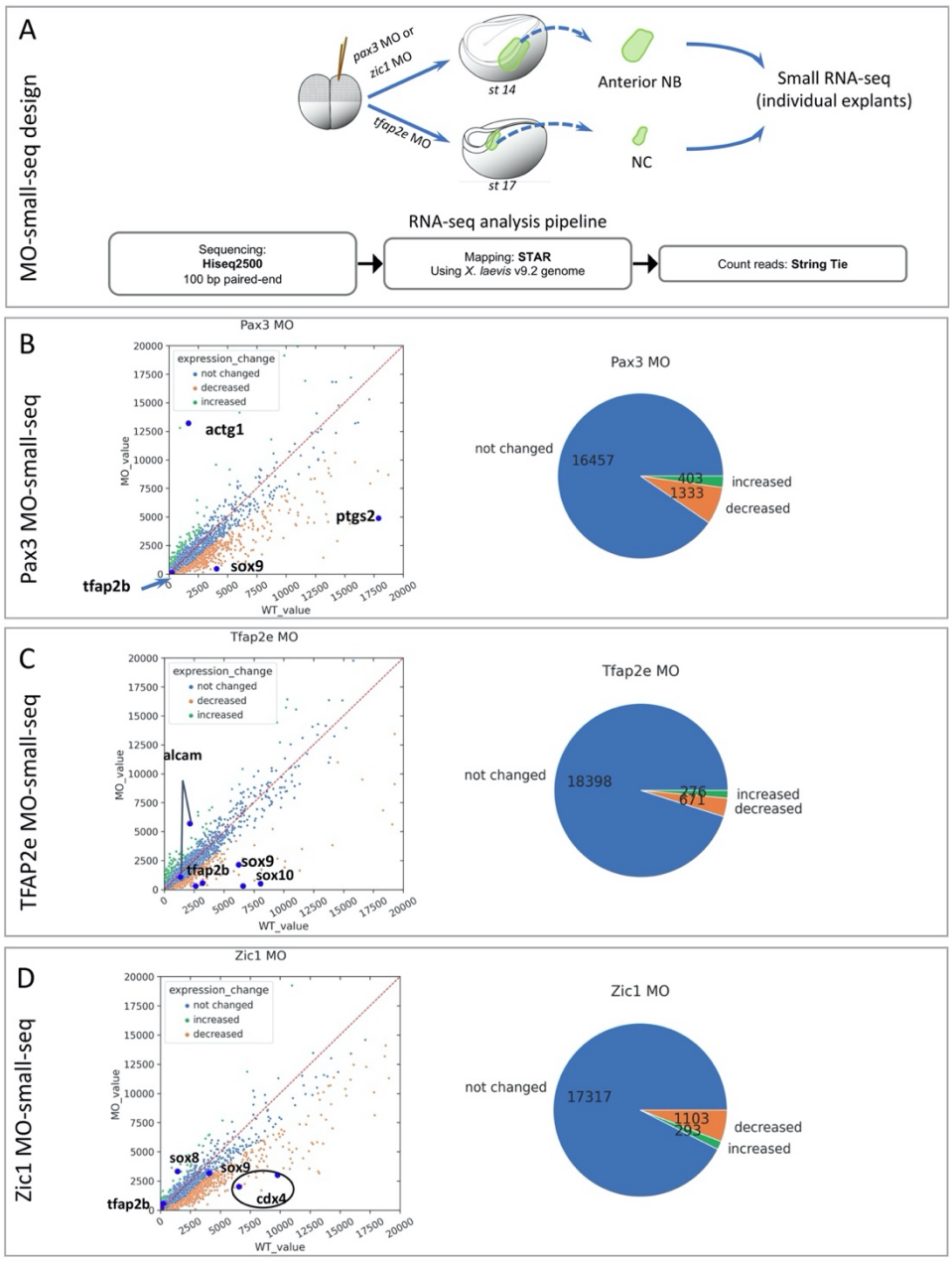


Fig. S6. **MO experiments overview.**
We performed MO knockdown experiments for the main genes from the predicted networks for NC dataset. We revealed 1333 decreased genes in comparison to 403 increased genes in *pax3* MO which confirms the mostly activating role of Pax3. For Zic1 and TFAP2e, similar proportions of increased/decreased genes are observed. **(A)** Design of MO-small-seq experiments. **(B, C, D)** General statistics for changed and unchanged genes after Pax3, Tfap2e, or Zic1 depletion compared to controls.

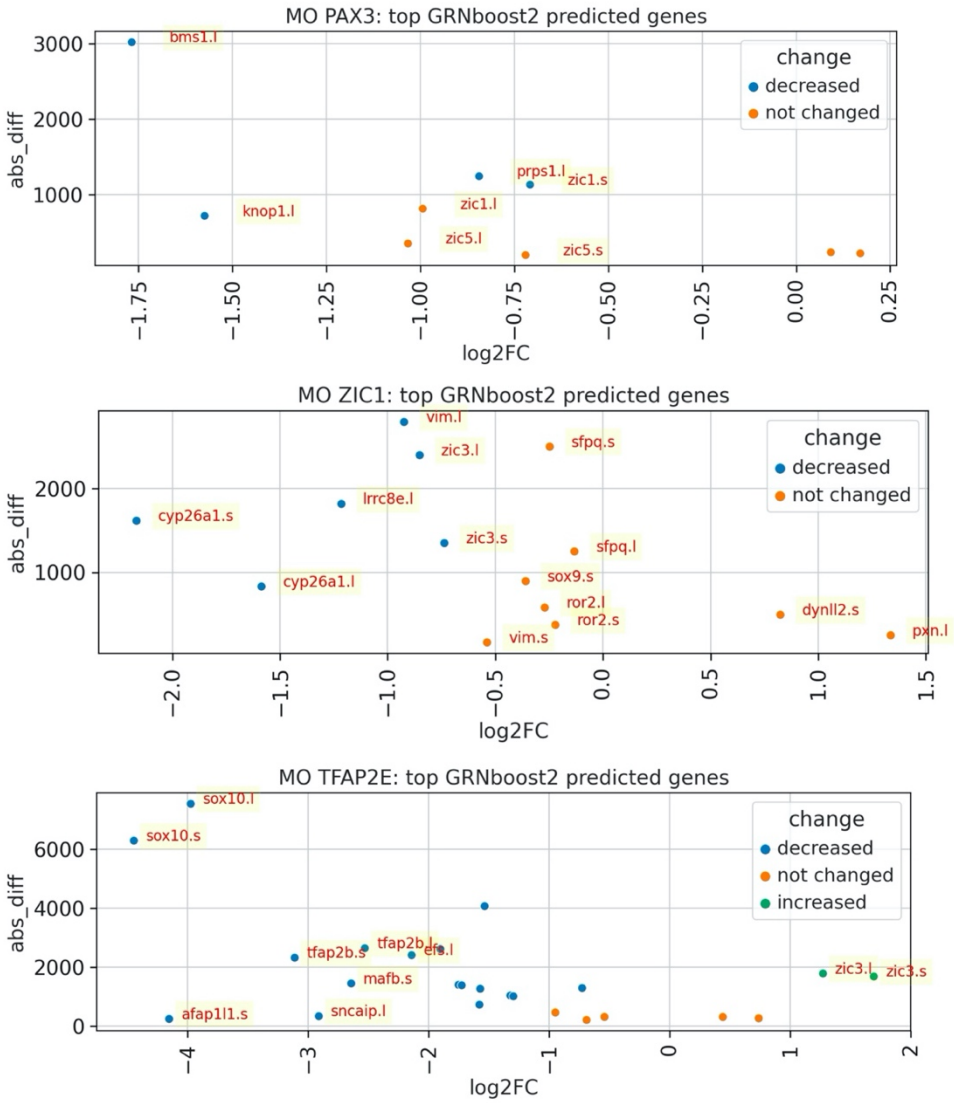


Fig. S7. **Top GRNboost2-predicted linked genes for Pax3, Tfap2e and Zic1 which also changed in MO samples.** We intersected results of the scRNA-seq predictions and MO knockdown experiments. Plots represent the top genes from the intersection for the several main actors (Pax3, Tfap2e, Zic1) from the predicted network. Among them *bms1* for Pax3, *sox10* for Tfap2e and *cyp26a1* for Zic1.

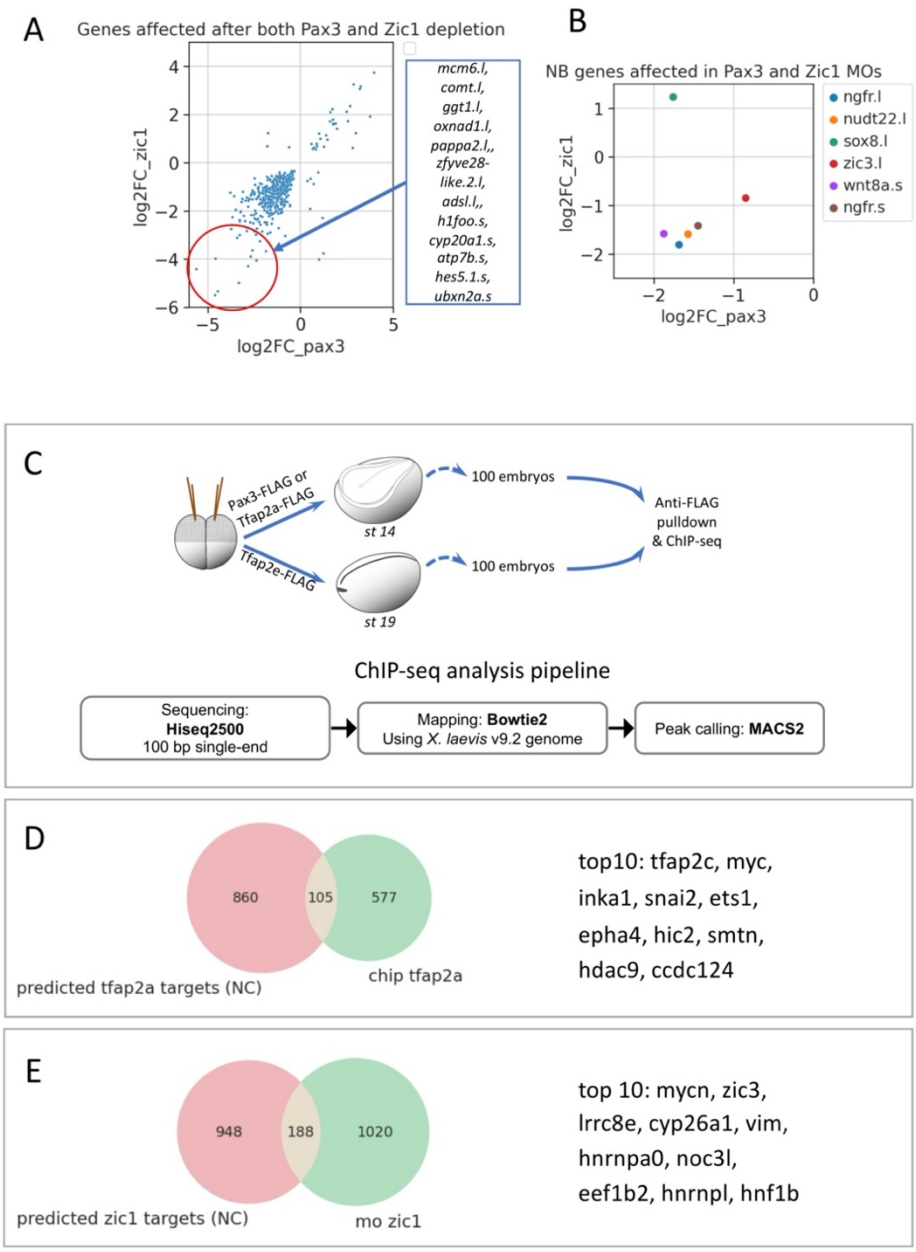
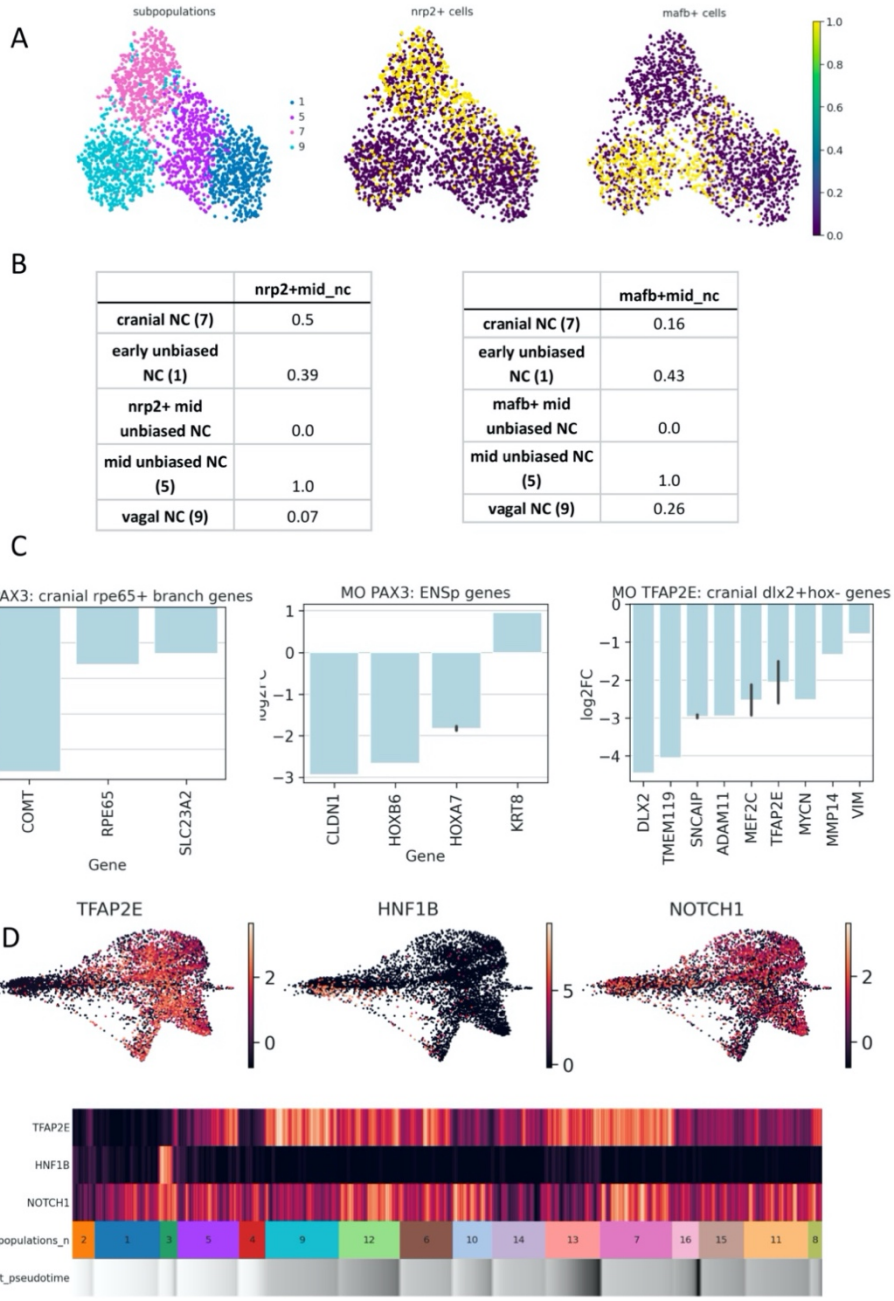


Fig. S8. **Direct gene regulation by Tfap2a, Zic1 and Pax3 during the NC induction.** In order to reveal direct targets and validate scRNA-seq prediction, we performed ChIP-seq for Tfap2a (one of the main actors in the Ectoderm network), Pax3 and Tfap2e (main actors in the NC network). ChIP-seq was performed in whole embryos (WE) then genes which are expressed in the NC cells were selected. **(A)** General overview of the most changed genes in both Pax3 and Zic1 MO samples (abs_diff>300), most of them decreased. **(B)** Genes from the NB zone signature with the lower expression level threshold. It shows that both Zic1 and Pax3 are needed for activation of *ngfr, wnt8a, sox8* and *nudt22* expression. **(C)** ChIP-seq analysis pipeline. **(D, E)** Comparison between GRNBoost2-based predictions and experimental validation for Tfap2a (ChIP-seq) and Zic1 (MO knockdown). Around 12% of genes predicted to be linked to Tfap2a are found as direct targets; while about 20% of predicted Zic1 linked genes are affected by Zic1 depletion *in vivo*.



	nrp2+mid_nc
cranial NC (7)	0.5
early unbiased NC (1)	0.39
nrp2+ mid unbiased NC	0.0
mid unbiased NC (5)	1.0
vagal NC (9)	0.07

	mafb+mid_nc
cranial NC (7)	0.16
early unbiased NC (1)	0.43
mafb+ mid unbiased NC	0.0
mid unbiased NC (5)	1.0
vagal NC (9)	0.26

Fig. S9. In-depth cluster analysis. (A-B) *Mafb* and *nrp2* expression during cell cranial and vagal predisposition. **(A)** We find early minor predisposition within cluster 5 for of *mafb* or *nrp2* expression. Cells expressing *nrp2/mafb* above 80 percentile were sub-selected around the bifurcation: clusters 1, 5, 7, 9. PAGA revealed that *nrp2*$^{+}$ cluster 5 cells are 7 times more similar to the cranial state than to the vagal stage, and *mafb*$^{+}$ cells of unbiased cluster 5 are 1.5 times closer to the vagal state than to the cranial. **(B)** Tables with cluster-cluster similarities (calculated with PAGA). **(C-D)** Pax3 and Tfap2e direct roles during bifurcations in cranial and vagal NC. **(C)** Using Pax3 and Tfap2e depletion, we evaluated the change of expression of the late genes in the Pax3 and Tfap2e related branches. **(D)** Some Tfap2e targets confirmed with ChIP-seq and MO were not predicted by GRNboost2. For example, *hnf1b* with a very different expression pattern and *notch1* with a closely related expression pattern.

Table S1. Current scRNAseq datasets.
We analyzed current datasets of NC cells and compared it to our new dataset with the largest number of NC cells. This allows us to use complex ML approaches to reveal gene-gene relationships.

Dataset	NC selection	Cell number	Depth	Timing	scRNA-seq technology	Reference
GSE113074	X.tropicalis WE/markers	130000+ WE, 4700 NC	Low	9 gastrula and neurula batches	Indrops v2/v3	Briggs et al, 2019
GSE112294	Zebrafish WE/markers	63520 WE, 1770 NC	Low	7 gastrula & neurula batches	Indrops v1	Wagner et al, 2018
GSE106676	Zebrafish, foxd3+ cells	96 NC	High		SmartSeq2	Lukoseviciute et al, 2018
GSE130500 GSE131688	Chick, foxd3+	137 SmarSeq2 3000 10x	High + Low	Late neurula	SmartSeq2 10x	Williams et al, 2019
GSE129114	Mouse, wnt1/sox10+	200 NC	High	Neurula	SmartSeq2	Soldatov et al, 2019
GSE162044	Mouse, wnt1+ NC	1741 NC	High	Neurula	Smart-Seq2	Zalc et al, 2021
-	X.tropicalis WE/markers	140000+ WE, 6135 NC 17138 EC	Low	7 gastrula and neurula batches	Indrop v3	This study

Table S2. **Top NC specific genes for each developmental stage.**
Using Leiden clustering and differential expression analysis we constructed two datasets: Ectoderm and NC. Table represents the most specific genes expressed in the clusters in comparison to other cells of the whole embryo.

Cell type	NF stage	Cells number	Gene markers for cluster characterisation
EC	11	9190	SOX2, OLFM4. FOXI1, GATA2, SOX11
EC	12	3500	TFAP2C, C3, GATA2, OLFM4, ZEB2
EC	13	4448	TFAP2C, SOX9, OLFM4, LHX2
NC	12	623	C3, SOX9, ZIC1, SNAI2
NC	13	579	SOX9, SNAI2, C3, ZIC1
NC	14	416	SNAI2, TFAP2B, SOX9, SOX8
NC	16	882	TFAP2B, C9, C3, SOX10
NC	18	607	TFAP2B, C9, SOX8
NC	20	839	TFAP2B, SOX8, C9
NC	22	2253	TFAP2B, C3, C9, SOX10, DLX2

Table S3, S4, S5. **Top decreased genes in pax3 and tfap2e MO.**
Top decreased genes in Pax3 (left), Zic1 (middle), and Tfap2e MO (right) samples with the absolute expression change > 1000.

Gene_name	log2FC	abs_diff
oxnad1.l	-4.471306	2796.0
pou4f1.2.l	-3.792479	1208.5
stxbp5-like.1.l	-3.573356	1134.0
loc100498469-like.l	-3.269686	1227.5
slc12a2.l	-3.269554	1988.0
gatm.l	-3.219859	11768.5
pnhd.s	-3.189243	1104.5
sox9.s	-3.160360	3605.0
hnf1b.l	-3.122515	3153.0
pygm.s	-2.895865	2087.5
olig4.l	-2.869639	1287.0
loc733561.l	-2.836421	1099.5
msx1.l	-2.815328	5604.0
hoxb1.s	-2.729814	1538.0
hnf1b.s	-2.690841	2215.5
h1foo.s	-2.681030	1277.5
hyou1.l	-2.640135	1544.0
hoxd1.l	-2.637859	1410.5
hoxd1.s	-2.594724	2475.0
msx2.l	-2.587375	3742.5
olfm4.l	-2.565518	5451.0
pnp.l	-2.480788	2625.5
dhrs3.l	-2.469848	2506.0
ppat.l	-2.468018	2416.0
vim.l	-2.450922	4834.0
prkaa1.l	-2.350497	1045.5
snrnp70.l	-2.320323	11492.0
cdx2.l	-2.289480	2395.5
sp5l.l	-2.267839	3598.5
msx1.s	-2.267784	2652.0

Gene_name	log2FC	abs_diff
pappa2.l	-8.049168	1055.5
oxnad1.l	-5.325875	2855.0
loc733561.l	-4.275842	1212.5
crx.l	-4.005001	1355.0
h1foo.s	-3.891248	1411.5
crx.s	-3.755836	1057.0
cldn6.2.s	-3.040364	3454.5
nphp3.l	-2.975995	1040.5
xetrov90006074m.l	-2.862766	2312.0
pygm.s	-2.836971	2074.0
angpt4.s	-2.716769	1243.0
bnip3.l	-2.473468	1418.5
loc100498469-like.l	-2.436440	1116.5
slc12a2.l	-2.376951	1791.0
clpx.s	-2.376474	2285.0
cnn1.s	-2.336621	6767.5
pnp.l	-2.320125	2558.0
loc100495392.s	-2.270703	1480.5
cldn6.1.s	-2.254050	1493.0
pfkfb3.s	-2.213230	1744.0
mycl.l	-2.203118	2512.5
cyp26a1.s	-2.166806	1616.0
hoxc5.s	-2.105209	2095.5
abl1-like.s	-2.099759	1061.5
gad1.2.l	-2.075370	10114.5
rrm2.1.l	-2.036311	1825.5
tmod3.l	-2.035264	1362.0
s1pr5.s	-2.010914	1297.0
pds5a.l	-2.006197	2190.5
tiam1.s	-1.998983	2126.5

Gene_name	log2FC	abs_diff
sox10.s	-4.446147	6291.5
sox10.l	-3.973992	7541.0
sox8.s	-3.623336	2485.5
loc101735110.l	-3.294872	3226.0
loc100498368-like.s	-3.288774	6777.0
tp63.l	-3.282590	1379.5
fam212a.s	-3.190155	1215.0
fam212a.l	-3.189506	2014.5
tfap2b.s	-3.111449	2319.5
loc100498368-like.l	-3.058974	2636.5
foxd3.l	-2.979076	2062.0
adam11.l	-2.943961	1329.0
sox8.l	-2.803670	9715.0
endod1.l	-2.742037	1212.0
mafb.s	-2.642183	1447.0
kal1.l	-2.578369	1817.5
tfap2b.l	-2.529612	2642.5
mycn.s	-2.511291	1086.0
twist1.l	-2.453900	2819.5
twist1.s	-2.421903	2770.0
loc100124848-like.l	-2.401117	8166.0
fn1.s	-2.305269	4979.5
kiaa1217-like.l	-2.296111	1036.5
ahnak-like.s	-2.267407	3456.0
itln1.l	-2.252100	6428.5
lima1.l	-2.242827	2533.0
loc100497845-like.l	-2.208089	1035.5
itpr3.s	-2.194760	2438.5
aim1-like.1.s	-2.168639	1307.5
cmahp.l	-2.151611	1425.5

Table S6. **Pax3 ChIP-seq targets.**
Top MACS2 scored direct targets of Pax3.

chr	start	end	score	gene
chr8L	99475396	99475464	580	ASTL3A.1
chr9_10L	349156	349253	518	ATP6V0A1
chr7L	92931087	92931289	487	HES3
Scaffold46	813190	813276	484	ASS1
chr2L	176448007	176448188	448	PRCP
chr8L	118795321	118795461	440	FLAD1
chr8L	118785517	118785631	440	PSMD4
chr7L	89910794	89910942	435	VPS13D
chr3L	130428472	130431136	417	S1PR2
chr7L	15405111	15405195	390	INPP5A.1
chr6L	86407631	86407987	374	LYRM4
chr8L	119048876	119048973	368	EFNA3
chr1L	144729397	144731409	358	TBX3
chr5S	17252239	17252405	341	ATF3
chr4L	141138945	141139207	341	ATXN7
chr6L	37364463	37364585	314	HDAC9
chr6L	124571856	124571969	312	ZFAND1
chr6L	124583418	124583619	312	CHMP4C
chr4S	38950018	38950121	307	NAE1
chr9_10L	65487732	65488499	291	SLC25A12
chr5L	51178347	51178965	288	PSEN2
chr1S	115523395	115523481	286	NRL
chr6L	39747334	39747730	286	EVX1
chr9_10L	30759439	30759550	279	TGM3L.4
chr2L	88024978	88025202	267	HPCA
chr6L	82762756	82762905	266	CDH20
chr1S	12094861	12094998	263	FGFR3
chr8S	33705393	33705520	248	FUBP3
chr8L	57452010	57452126	245	SELENOW1
chr9_10S	23803020	23803090	244	USP36

Table S7. **Tfap2e ChIP-seq targets.**
Top MACS2 scored direct targets of Tfap2e.

chr	start	end	score	gene
Chr7	2771023	2771259	1502	RBM20
Chr4	136928957	136929250	1360	PIM1
Chr10	10458652	10458795	1353	ARL5B
Chr10	10464454	10464497	1353	PLXDC1
Chr5	13534757	13535087	1326	PTPN14
Chr10	50539112	50539288	1235	ARPC1A
Chr10	50529349	50529522	1235	PRKAR1A
Chr10	15734843	15734962	1218	SMARCD2
Chr2	2687034	2687603	1212	BOC
Chr9	85342807	85342887	1199	TNS1
Chr6	90349120	90349399	1198	TFAP2A
Chr6	47572683	47572824	1180	RARB
Chr4	58757061	58757134	1176	ANKRD11
Chr4	48790141	48791762	1170	BCAR1
Chr7	18235059	18235409	1164	JMJD1C
Chr2	138639744	138639788	1159	RBMS2
Chr8	6618399	6618453	1159	NOTCH1
Chr8	142072529	142072579	1153	CCT3
Chr8	142086526	142086803	1153	GLMP
Chr2	40958194	40958575	1153	PDK3
Chr1	65662851	65663810	1145	SPRY1
Chr4	67913743	67913832	1139	KCTD15
Chr9	83910392	83910819	1130	TUBA4A
Chr1	206833365	206833506	1100	RNF165
Chr1	16309482	16311265	1095	DCTN1
Chr1	16305342	16305426	1095	ADAM19
Chr10	11004405	11005055	1091	ARHGAP23
Chr6	12882851	12882961	1084	DHX30
Chr6	12892165	12892295	1084	LIMA1
Chr4	35478920	35479123	1071	PPP1R14B

Table S8. **Tfap2a ChIP-seq targets.**
Top MACS2 scored direct targets of Tfap2a.

chr	start	end	score	gene
chr8L	118798860	118799049	476	FLAD1
chr8L	118788167	118788453	476	PSMD4
chr5L	101986969	101987669	431	MCF2L.2
chr4L	74333641	74333831	423	ROR1
chr3S	83043144	83043215	395	LRIG3
chr4L	119344147	119344785	369	GPX1
chr9_10S	95234068	95234536	364	ARHGAP17
chr2L	110483824	110487409	352	SOWAHC
chr5S	1791004	1791317	341	SERTAD2
chr2L	91858061	91858536	329	MMP28
chr2L	91863324	91863411	329	TAF15
chr5S	83018695	83019717	323	PTMA
chr1S	179165531	179165761	318	CNTFR
chr1L	151166104	151166250	316	RPLP0
chr2S	158792498	158793672	316	LRTOMT
chr1L	151152633	151153237	316	GCN1
chr3L	136390271	136390378	307	ESR-5
chr4S	13935547	13940398	303	CHRM4
chr9_10L	109917922	109918127	303	LMF1
chr4S	13930436	13930514	303	MDK
chr4S	20597644	20598583	301	NAV2
chr1S	68374163	68374214	296	BTC
chr1S	56378070	56378168	295	EGF
chr1S	54034891	54035850	293	SPRY1
chr3L	28766498	28766650	275	MAT2B
chr8S	26253740	26253836	271	VGLL4L
chr9_10L	79461487	79461535	270	CDK15
chr9_10S	45513533	45514078	268	TNS1
chr7S	25366517	25366699	268	CPN1

chr3L	125880572	125880653	265	NOTCH3

Table S9. **Primers used in this study.**

primer name	purpose	sequence	reference
Snail2 Fwd pair1	ChIP-seq IP quality check	CCATCCCAACACCTGTCGTA	This study
Snail2 Rev pair1	ChIP-seq IP quality check	GCCCACCAGTTCACGTTCAT	This study
Snail2 Fwd pair2	ChIP-seq IP quality check	GCAATGCCTCAGCCTGTGAA	This study
Snail2 Rev pair2	ChIP-seq IP quality check	CAGCGCGTACTGCAATTCATTC	This study
Odc Fwd	qRT-PCR	GCCCTTTCTCCCTTTAACGC	(29)
Odc Rev	qRT-PCR	TGGTCCCAAGGCTAAAGTTG	(29)
Snail2 Fwd	qRT-PCR	CACACGTTACCCTGCGTATG	(77)
Snail2 Rev	qRT-PCR	TCTGTCTGCGAATGCTCTGT	(77)
Sox10 Fwd	qRT-PCR	CTATTACTGACACACGACGGAGC	(17)
Sox10 Rev	qRT-PCR	ACCTCTCATCCTCTGAATCCTGC	(17)
MyoD Fwd	qRT-PCR	TACACTGACAGCCCCAATGA	(29)
MyoD Rev	qRT-PCR	TGCAGAGGAGAACAGGGACT	(29)
Myl1 Fwd	qRT-PCR	GAAACACTTGGGCTGCTTTCTT	This study
Myl1 Rev	qRT-PCR	AGCAGGTTTAGCCTCAGGTTT	This study