# From tissue to single cell : the neural border and neural crest gene regulatory network

Aleksandr Kotov

THÈSE DE DOCTORAT

NNT : 2023UPASL016

université PARIS-SACLAY

# From tissue to single cell: the neural border and neural crest gene regulatory network

*Réseau régulateur de la bordure neurale et de la crête neurale, du tissu aux cellules uniques*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n° 577 Structure et dynamique des systèmes vivants (SDSV)
Spécialité de doctorat : Biologie du développement
Graduate School : Sciences de la vie et de la Santé. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **Signalisation, Radiobiologie et Cancer** (Université Paris-Saclay, Inserm, CNRS), sous la direction d'**Anne-Hélène MONSORO-BURQ**, Professeure, et le co-encadrement de **Thomas WALTER**, Professeur

**Thèse soutenue à Paris-Saclay, le 14 mars 2023, par**

**Aleksandr KOTOV**

## Composition du Jury

Membres du jury avec voix délibérative

| | |
|---|---|
| **Sophie CREUZET**<br>HDR, Directrice de recherche, Institut NeuroPSI | Présidente |
| **Christelle GOLZIO**<br>HDR, Directrice de recherche, Université de Strasbourg | Rapporteur & Examinatrice |
| **Franck PICARD**<br>HDR, Directeur de recherche, École Normale Supérieure de Lyon | Rapporteur & Examinateur |
| **Jean-François RIOU**<br>HDR, Directeur de recherche, Institut de Biologie Paris-Seine | Examinateur |
| **Jean-Loup DUBAND**<br>HDR, Directeur de recherche, Mondor Institute for Biomedical Research | Examinateur |
| **Antoine ZALC**<br>Directeur de recherche, Institut Cochin | Examinateur |

université PARIS-SACLAY

ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

**Titre :** Réseau régulateur de la bordure neurale et de la crête neurale, du tissu aux cellules uniques
**Mots clés :** crête neurale, scRNA-seq, réseau de régulation génique, évolution de l'état cellulaire
**Résumé**: La crête neurale est un groupe de cellules qui apparaît au cours du développement embryonnaire des vertébrés et donne naissance à un large éventail de types cellulaires, tels que les neurones, le osteoblastes craniofaciaux et les mélanocytes. La crête neurale se forme à partir d'un domaine ectodermique qui forme également les progéniteurs des organes sensoriels céphaliques, les placodes. Dans cette thèse de doctorat, nous utilisons une combinaison de techniques, dont le séquençage d'ARN sur cellule unique, pour étudier les réseaux de régulation génique contrôlant la formation de la crête neurale. En analysant les profils transcriptionnels de cellules individuelles, nous avons pu identifier des sous-populations distinctes de cellules au sein des progéniteurs de la crête neurale. Nous avons également identifié des gènes spécifiquement exprimés ou enrichis dans chaque sous-population et utilisé le séquençage après immunoprécipitation de la chromatine pour étudier les sites de liaison de quelques facteurs de transcription-clés qui régulent ce réseau et en définir les gènes-cibles. De plus, nous avons étudié le rôle de deux micro-ARN dans la régulation de ce réseau. Nos résultats mettent en évidence la nature complexe et dynamique des réseaux de régulation des gènes contrôlant le développement et la fonction de la crête neurale. De plus, nous avons étudié la relation entre la crête neurale et les placodes. Nous avons également développé un outil informatique capable de prédire l'évolution des types de cellules et l'avons appliqué sur la crête neurale pour étudier son histoire évolutive. Dans l'ensemble, nos découvertes fournissent de nouvelles informations sur les mécanismes contrôlant le développement et la fonction de la crête neurale et ont des implications potentielles pour la compréhension et le traitement des neurocristopathies, défauts congénitaux ou tumeurs liés à la crête neurale.

**Title:** From tissue to single cell: the neural border and neural crest gene regulatory network
**Keywords:** neural crest, scRNA-seq, gene regulatory network, cell state evolution
**Abstract**: The neural crest is a group of cells that arises during embryonic development in vertebrates and gives rise to a diverse array of cell types, such as neurons, craniofacial osteoblasts and melanocytes. Neural crest cells emerge from an ectoderm territory, the neural border, which also forms progenitors of the head sensory organs, the cranial placodes. In this doctoral thesis, we use a combination of techniques, including single-cell RNA-sequencing, to study the gene regulatory networks driving neural crest formation (NC-GRN). By analyzing the transcriptional profiles of individual cells, we were able to identify distinct subpopulations of cells among early neural crest progenitors. We also identified genes that are specifically expressed or enriched in each subpopulation and used ChIP-sequencing for a few key transcription factors, in order to define their binding sites and target genes, within the NC-GRN. In addition, we investigated the role of micro-RNAs in regulating this network. Our findings highlight the complex and dynamic nature of the gene regulatory networks controlling the development of the neural crest. Furthermore, we studied the relationship between the neural crest and placodes. We also developed a computational tool that can predict the evolution of cell types and applied it on the neural crest to study its evolutionary history. Overall, our findings provide new insights into the mechanisms controlling the development of the neural crest and its related cell types. Our studies have potential implications for the understanding and treatment of neural crest-related disorders.

# ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my thesis supervisor, Anne-Hélène Monsoro-Burq, for her guidance and support throughout the course of this project. Her valuable advice and suggestions greatly improved the quality of this thesis and me as a specialist. Also, Igor Adameyko, Andrei Zinovyev, Thomas Walter and Leon Peshkin, thank you for your invaluable experience you shared with me, I really appreciate it.

I am also grateful to the members of my thesis jury: Franck Picard, Christelle Golzio, Sophie Creuzet, Antoine Zalc, Jean-Francois Riou, and Jean-Loup Duband, for their time and expertise in evaluating my work.

I would like to thank my colleagues and friends at Curie Institute (Subham, Laura, Chenxi, Vincent, and many others) and from NEUcrest network (Ira, Nico, Merin, Amy, Marco, Haris and others) for their support and camaraderie throughout the years.

I am also grateful to my family for their love and encouragement. Special thanks go to my wife, Anya, for her unwavering support and patience during the long hours of writing and research.

Finally, I would like to acknowledge the support of NEUcrest project for providing the financial support that made this research possible and a great manager of this project, Daria, who helped us a lot.

Thank you all.

# ABBREVIATIONS

ATAC-Seq Assay for Transposase-Accessible Chromatin using sequencing

AP Anetrior-Posterior

AUC Area under curve

BMP Bone morphogenetic protein

ChIP-Seq chromatin immunoprecipitation sequencing

DNA Deoxyribonucleic acid

DV Dorso-Ventral

FGF: Fibroblast Growth Factor

EMT Epithelial-mesenchymal transition

ENSp: Enteric Nervous System progenitors

GO Gene Ontology

GOF Gain of Function

GRN Gene regulatory network

ISH In Situ Hybridization

MMPs Matrix metalloproteinases

MO Morpholino oligonucleotides

NB Neural border

NC Neural crest

NP Neural plate

PAX Paired box

PC Placodes

RA: Retinoic Acid

RNA-Seq ribonucleic acid sequencing

ROC Receiver operating characteristic

ScRNA-seq single-cell ribonucleic acid sequencing

TGF-B transforming growth factor beta

TFs Transcription Factors

WNT Wingless and Int-1

X. laevis Xenopus Laevis

X. tropicalis Xenopus tropicalis

# ABSTRACT

The neural crest is a group of cells that arises during embryonic development in vertebrates and gives rise to a diverse array of cell types, e.g., neurons, glia of the peripheral nervous system, melanocytes, and skeletal and connective tissues in the head. It originates from the neural border, a domain of the ectoderm positioned between the neural plate and the non-neural ectoderm. However, the gene regulatory networks underlying the development and function of these cells are not fully understood and there are still many unsolved problems in the field.

In this doctoral thesis, we use a combination of cutting-edge techniques, including single-cell RNA-sequencing, ChIP-sequencing, and knock-down experiments, to study the gene regulatory networks of the neural crest and the neural border (NC-GRN, NB-GRN). By analyzing the transcriptional profiles of single cells, we were able to identify distinct subpopulations of cells within poorly studied pre-migratory neural crest cells. We also were able to define scRNA-seq signature of the neural border and identify genes that are specifically expressed or enriched in each subpopulation in ectoderm and pre-migratory neural crest. We then used ChIP-seq for a few chosen transcription factors, to define their binding sites and their target genes within the NB/NC-GRN. To further identify how these transcription factors regulate their targets, we used knock-down experiments followed by RNA sequencing and constructed a transcription factor-centered GRN driving neural crest development.

In addition to studying the transcriptional networks of the neural crest and neural border, we also investigated the role of two microRNAs (miRNAs) in regulating these networks. Our findings highlight the complex and dynamic nature of the gene regulatory networks controlling the development and function of the neural crest.

Furthermore, we studied the relationship between the neural crest and placodes, a group of cells that also originates from the neural border and gives rise to sensory organs. Our findings suggest that instead of contrasting “neural border” and “non-neural vs neural" hypotheses, which are the current and competing models proposed until now, we find that these routes are not exclusive and find trajectories supporting the emergence of neural crest from either the neural border or the nascent neural ectoderm on one hand, as well as two trajectories leading to placodes from either the neural border or the non-neural ectoderm.

Last, we developed a computational tool that can predict the evolution of cell types and applied it on the neural crest. Using this tool, it is possible to study the evolutionary history of the neural crest and gain new insights into the mechanisms underlying its development and diversification.

Overall, our findings provide new insights into the mechanisms controlling the development and function of the neural crest and neural border, and have potential implications for the understanding and treatment of neural crest-related disorders.

# RÉSUMÉ ÉTENDU EN FRANÇAIS

Les cellules de la crête neurale forment une population de progéniteurs multipotents et migrateurs présents dans les embryons de vertébrés, essentiels pour le système nerveux périphérique et entérique, les structures craniofaciales, les cellules endocrines et pigmentaires, entre autres. Peu après la gastrulation, dans la partie la plus antérieure de l'embryon, les cellules NC sont induites à partir de la bordure neurale (BN) dorso-latérale, un domaine de l'ectoderme situé entre l'ectoderme non neural et l'ectoderme de la plaque neurale (Alkobtawi et Monsoro-Burq, 2020 ; Plouhinec et al., 2017). Au cours du processus de neurulation, la spécification et l'induction des cellules de la crête neurale progressent en vague de la zone antérieure à la zone postérieure de la plaque neurale. Ce processus s'accompagne de l'activation de divers programmes génétiques qui définissent les stades précoces et immatures des CN. En plus du programme général des CN, plusieurs modules moléculaires régionaux sont également activés le long de l'axe antéro-postérieur du corps, ce qui permet de définir des sous-populations de CN au potentiel spécifique (Ling et Sauka-Spengler, 2019 ; Tang et al., 2021). La façon dont ces programmes liés aux sous-populations sont interconnectés avec le module pan-CN, et comment et quand ils sont activés dans les cellules CN pré-migratoires sont peu décrits. A la fin de la neurulation, les cellules de la crête neurale quittent l'ectoderme dorsal par une transition épithélium / mésenchyme (EMT), suivie d'une migration extensive vers une variété de tissus cibles, où les cellules de la crête neurale se différencient en plus de trente types cellulaires différents, y compris les neurones et les cellules gliales périphériques et entériques, les os craniofaciaux, les chondrocytes, les adipocytes et le mésenchyme, les cellules sécrétrices de chromaffine et les cellules pigmentaires. La biologie de la crête neurale a été analysée au cours du développement et de l'évolution, ce qui a permis d'élucider des réseaux élaborés de régulation génique au cours de la dernière décennie (Monsoro-Burq et al., 2005 ; Simoes-Costa et Bronner, 2016). Ces réseaux restent toutefois incomplets et n'expliquent pas la plupart des défauts observés dans les neurocristopathies humaines (Medina-Cuadra et Monsoro-Burq, 2021).

Ce problème est sur le point d'être résolu grâce à la transcriptomique sur cellule unique, qui permet une description complète du développement de la NC, au cours des differentes étapes du développement, ainsi qu'une comparaison avec les tissus adjacents (par exemple, au niveau de la frontière neuronale). Cela permettrait de définir les trajectoires génétiques du

développement de l'arbre complet de la lignée CN. De récentes études scRNAseq sur la CN ont principalement exploré les CN après migration, en utilisant des embryons de poulets, de poissons et de souris (Artinger et Monsoro-Burq, 2021). En revanche, les cellules individuelles de la CN avant la migration ont fait l'objet d'une exploration limitée, principalement autour du stade de l'EMT et sur de petits nombres de cellules à un niveau spécifique de l'axe du corps (Tang et al., 2021 ; Zalc et al., 2021).

Outre la crête neurale, la BN donne également naissance aux placodes postérieures, à l'ectoderme non neural et à la partie dorsale du tube neural (Steventon et Mayor, 2012 ; Streit et Stern, 1999). La question est de savoir si ces quatre types de cellules proviennent d'un progéniteur précoce commun multipotent et comment les décisions relatives à leur destin sont orchestrées au niveau de la BN pendant la gastrulation reste mal comprise. Bien que la formation du territoire de la BN ait été définie par l'expression de quelques gènes pendant la gastrulation (p. ex. pax3 et pax7) (Basch et al., 2006 ; Monsoro-Burq et al., 2005 ; Plouhinec et al, 2017), le moment de la spécification de la NB à partir du reste de l'ectoderme dorsal et les circuits conduisant les décisions de destin entre les quatre destins cellulaires dérivés de la BN (NC, placodes, ectoderme non neural et tube neural dorsal) restent à établir (Groves et LaBonne, 2014 ; Maharana et Schlosser, 2018 ; Steventon et Mayor, 2012). En outre, le calendrier des décisions de lignage dans la CN prémigratoire le long de l'axe antéro-postérieur, le maintien d'une sous-population de CN multipotente et les mécanismes moléculaires pilotant chaque état de l'arbre de lignage de la CN prémigratoire restent inexplorés.

En outre, les mécanismes complexes entre les voies de signalisation et les marqueurs des frontières neurales ne sont pas encore clairement compris (Garnett et al., 2012). Les marqueurs de la BN pax3/msx1/zic1/tfap2a/hes4 se contrôlent mutuellement dans une boucle de rétroaction et ont besoin d'une signalisation WNT supplémentaire (de Crozé et al., 2011 ; Simoes-Costa et Bronner, 2016). Les marqueurs des progéniteurs de placodes de xenope eya1/six1 influencent les marqueurs de la BN et de la CN pax3 et foxd3 ainsi que les inducteurs de la BN tfap2a, msx1, dlx3, gata2, foxi1 (Maharana et Schlosser, 2018). Ensemble, ces régulations croisées élaborées stabilisent les choix de destin dans la bordure neurale, la crête neurale et les placodes. Toutes ces études ouvrent une discussion sur deux modèles différents proposés pour le développement de la bordure neurale :

1.Le modèle de "compétence binaire" proposé par Schlosser dans lequel la CN et les placodes sont produites individuellement : la CN du côté de la frontière neurale et les placodes du côté de l'ectoderme non neural (Pieper et al., 2012 ; Schlosser, 2008).

2.Le modèle "BN" propose que la bordure neurale bi-potente produise à la fois des progéniteurs de la CN et des placodes, dont les positions comparatives sont définies à des stades ultérieurs par des marqueurs spéciaux.

Outre ces questions qui restent sans réponse concernant la régulation des cellules de la crête neurale et de la bordure neurale, il est également nécessaire d'examiner d'autres mécanismes de régulation des gènes, tels que la régulation effectuée par les ARN non codants. Le rôle des miARN au cours du développement a été signalé pour la première fois en 2003, lorsque Bernstein a supprimé Dicer (qui coupe l'ARN double brin et les molécules de pré-miARN pour créer de courts fragments d'ARN double brin, y compris les miARN) chez la souris, détectant une létalité embryonnaire précoce (Bernstein et al., 2003). D'autres études ont également montré que les miARN jouent un rôle important dans de nombreux processus de développement (Antonaci et Wheeler, 2022 ; Mok et al., 2017). Les microARN sont essentiels au développement normal des animaux et sont impliqués dans différents processus biologiques. L'expression aberrante des miARN est liée à diverses maladies humaines (Hanna et al., 2019). De plus, les miARN sont sécrétés dans les fluides extracellulaires. Ces miARN ont été largement signalés comme des biomarqueurs possibles pour certaines maladies et ils agissent également comme des molécules de signalisation pour modérer les contacts entre cellules (Bhaskaran et Mohan, 2014).

Ainsi, pour répondre aux questions concernant l'hétérogénéité des décisions relatives au destin cellulaire avant et après l'induction de la CN, nous avons utilisé des transcriptomes sur cellules uniques provenant de huit stades de développement consécutifs de *Xenopus tropicalis*, comprenant 6135 cellules CN, afin de fournir un profilage développemental complet de la CN pré-migratoire. Nous avons découvert plusieurs nouvelles sous-populations de cellules CN et mis en évidence leurs trajectoires précises, résultant en huit sous-populations CN émigrant des niveaux antérieurs aux niveaux vagaux de l'axe du corps. Il est intéressant de noter que certains destins émergent beaucoup plus tôt que prévu, que la diversité des CN est maintenue lors de l'EMT et qu'une diversification supplémentaire se produit au début de la migration. Nous proposons une séquence temporelle d'événements moléculaires qui sous-tendent les états transcriptomiques successifs et les décisions relatives au destin qui soutiennent l'émergence des cellules CN de la bordure neurale pendant la gastrulation jusqu'aux premiers états migratoires au début de l'organogenèse. En outre, nous identifions des facteurs de transcription clés impliqués dans la ramification de la lignée principale et validons plusieurs prédictions de

régulation in vivo. Nous fournissons donc un vaste réseau de régulation génique décrivant l'émergence de la lignée de la crête neurale dans l'ectoderme des embryons de vertébrés.

En outre, pour étudier la décision relative à la date cellulaire dans la BN et les relations entre les principaux acteurs de la régulation, nous avons (1) utilisé des transcriptomes sur cellules uniques provenant de trois stades de développement de Xenopus tropicalis, comprenant 17138 cellules de l'ectoderme précoce pour fournir un profilage développemental complet de la BN, ainsi que (2) la combinaison de constructions de gènes inductibles et de cycloheximide (un médicament qui bloque la traduction des protéines) pour Pax3, Zic1, et Dlx3. Les analyses du connectome des facteurs de transcription et de la bifurcation ont démontré l'émergence précoce des destins de la crête neurale au stade de la plaque neurale, parallèlement à un lignage multipotent non biaisé persistant jusqu'après la transition épithéliale-mésenchymateuse. Des expériences de perturbation ont révélé des cibles uniques et partagées parmi les acteurs les plus importants de la crête neurale (étude en cours).
Ensuite, afin d'étendre nos connaissances sur la régulation de la CN par le réseau de régulation génique (RRG), nous avons sélectionné deux miARN, dont l'importance pour la crête neurale avant la migration a été démontrée précédemment. Les principaux objectifs étaient de confirmer que miR-196a et miR-219 sont effectivement nécessaires au développement de la CN en utilisant le knockdown par morpholino des microARN miR-196a et miR-219 et de commencer à déchiffrer à quel niveau les miARN sont impliqués dans le développement de la CN de Xenopus (étude en cours).
Enfin, pour en savoir plus sur l'évolution de la CN et d'autres types de cellules, nous avons développé l'outil informatique basé sur le ML scEvoNet (single-cell Evolutionary networks).

Au cours des 150 dernières années, les types cellulaires ont été étudiés en termes de développement et de morphologie, alors qu'avec l'avènement de la transcriptomique sur cellules uniques, de nombreux types cellulaires distincts et cachés ont été identifiés. Avec l'utilisation de scRNA-seq, une ère complètement nouvelle s'ouvre pour comprendre l'évolution des types cellulaires. Apparemment, les types ancestraux de cellules animales étaient polyfonctionnels (la cellule éponge possède à la fois des propriétés neurales et immunitaires), mais au cours de l'évolution, les fonctions ont été réparties entre différentes cellules (Arendt, 2008). Cette "division du travail" est l'une des clés de l'évolution des types cellulaires. Pourtant, il y a un manque d'outils computationnels scRNA-seq pour étudier l'évolution des états cellulaires. Nous présentons donc ici scEvoNet, le logiciel mis en œuvre pour étudier cette

conception à partir de données scRNA-seq. L'utilisation de cet outil pour étudier le continuum des états du transcriptome entre les stades de développement ou les espèces permettra d'expliquer la dynamique complexe des états cellulaires.

**Un "modèle de double convergence" réconciliateur décrit les trajectoires convergentes initiant les états de crête neurale et de placode.**

Pour révéler les mécanismes moléculaires qui distinguent l'induction de la crête neurale et des placodes dans la zone frontalière entre l'ectoderme neural et l'ectoderme non neural pendant la gastrulation et vérifier les deux modèles ("Compétence binaire" et "BN"), nous avons généré un ensemble de données pour toutes les cellules de l'ectoderme des stades 11 à 13 (en utilisant l'ensemble de données scRNA-seq de l'embryon entier). Un changement de destin via une séparation binaire (bifurcation) est généralement observé lorsqu'une population apparemment homogène de précurseurs se divise en deux types de cellules présentant des profils d'expression génique distincts. La BN elle-même est interprétée comme une zone de compétence générale pour le tube neural, l'ectoderme non neural, la CN et les placodes. Et bien qu'elle présente une certaine association préférentielle de destins cellulaires, qui rappelle le modèle de compétence binaire proposé par Schlosser (Pieper et al., 2012 ; Schlosser, 2008) où la CN et les placodes sont induits séparément, nous avons observé que la zone BN contribue aux CN et aux PC parallèlement aux contributions convergentes des progéniteurs de la plaque neurale et de l'ectoderme non neural.

Il est intéressant de noter que des gènes distincts sont activés pour obtenir le même état et que certains gènes sont activés avec des dynamiques d'expression différentes selon les bifurcations. Par exemple, pendant la transition BN vers CN, sox9 et c3 étaient activés avant la bifurcation, ce qui indique que ces gènes pouvaient déclencher le passage des cellules des progéniteurs BN vers un destin CN. En revanche, au cours de la transition entre l'ectoderme neural et la CN, sox9 et c3 étaient des gènes tardifs (activés après la bifurcation), tandis que foxd3 et zic1 étaient exprimés précocement. Cette observation suggère un nouveau modèle de décisions relatives au destin dans l'ectoderme en développement, où des programmes génétiques parallèles et distincts activés dans des progéniteurs distincts de l'ectoderme peuvent conduire à un état similaire, soutenant un "modèle de convergence double" pour la crête neurale et l'émergence des placodes à partir de l'ectoderme dorsal, fusionnant ainsi les modèles actuellement concurrents de "l'origine de la bordure neurale" opposée à "l'origine de l'ectoderme neural/non neural".

En outre, le gène pax3, spécifique de la zone BN, a été exprimé avant la bifurcation dans le programme génétique BN vers CN et a activé l'expression des marqueurs de la branche CN tardive sox9, sox8, zic1, pcdh8 et c3. En outre, le connectome de l'ectoderme décrit des gènes CN connectés au reste du réseau par l'intermédiaire de Pax3 et de Sox9, ce qui suggère que Sox9 pourrait jouer une fonction non encore décrite en aval de Pax3 dans l'induction de la CN et en amont des autres marqueurs de la branche tardive de la CN. Ceci est en accord avec le fait que sox9 est un gène précoce dans la branche BN→CN. La validation expérimentale a montré que la co-activation de Sox9 augmentait fortement l'expression de snail2 dans l'iCN, tandis que la déplétion de Sox9 réduisait l'activation de snail2, indiquant que Sox9 est nécessaire en aval de Pax3 pour une induction efficace de la CN par Pax3 et Zic1. Alors que Sox9 a été décrit comme essentiel pour l'induction des CN, le mécanisme d'épistasie entre les spécificateurs de BN, Sox9 et Snail2 est resté inconnu jusqu'à cette expérience.

**Une combinaison de stratégies omiques et in vivo valide de vastes ensembles de régulations génétiques conduisant à la dynamique de la diversification de la crête neurale.**
L'énigme de savoir si les groupes de la CN utilisent des biais différentiels pour naviguer dans un paysage de destin unique, ou s'ils sont confrontés à des décisions fondamentalement uniques dans des paysages distincts, peut être abordée en identifiant des bifurcations "muettes" où l'amorçage multi-lignes se produit mais où un destin n'est jamais exprimé. La transcriptomique sur cellules uniques capture des différences suffisantes entre les états transcriptionnels, ce qui peut signifier des processus de prise de décision. Bien que l'on suppose généralement que les changements transcriptionnels importants sont liés à des événements de choix du destin, ils pourraient plutôt signifier une différenciation après l'engagement, tandis que de petits changements n'affectant que quelques gènes puissent être responsables de la rupture de l'homogénéité. De nouvelles approches permettant de suivre ces événements à travers les stades de développement sont capables de définir de légers changements dans les modèles d'expression génique.

**Hétérogénéité dans la CN pré-migratoire**
L'ensemble des données ScRNA-seq de l'embryon entier avec une extraction précise des cellules de la CN nous a permis de définir la dynamique temporelle des trajectoires qui aboutissent aux 8 états de la crête neurale présents au début de la migration le long des positions axiales crânienne et vagale. Notre analyse temporelle met en évidence trois points importants

qui nous permettent d'approfondir notre compréhension de la biologie de la crête neurale. Tout d'abord, il y a des débats de longue date sur le moment des décisions relatives au destin de la CN, avant ou après l'EMT du tube neural, dans une variété de modèles animaux (Kalcheim et Kumar, 2017). Il est important de noter que nous n'avons pas détecté d'expression distinctive de marqueurs prédictifs du destin avant l'EMT (par exemple, pour les destins neuronaux, gliaux, squelettiques ou mélanocytaires). Cela suggère que, si certains progéniteurs CN étaient orientés vers un destin donné avant l'EMT, ils ne présentaient pas de signature détectable dans notre ensemble de données. Cependant, nos observations sont en accord avec plusieurs études de traçage des lignées montrant la multipotence élevée de la plupart des cellules CN lorsqu'elles sont marquées avant l'EMT (Baggiolini et al., 2015). Les premiers marqueurs de différenciation sont trouvés après l'émigration, puisque nous avons détecté une expression de type myosine dans un petit sous-ensemble de cellules suggérant l'émergence de myofibroblastes dérivés de la CN, précédemment peu décrits, peu après l'EMT. Deuxièmement, nos données confirment la diversification précoce en plusieurs états cellulaires distincts avant, pendant et après l'EMT, ce qui contraste avec la suggestion récente selon laquelle, lors de l'EMT, les progéniteurs de la CN se regrouperaient en un seul état multipotent commun (Zalc et al., 2021). Le contenu cellulaire élevé de notre ensemble de données prouve le contraire, suggérant que cette observation précédente faite sur un sous-ensemble plus petit de NC crâniennes n'a pas entièrement capturé la diversité des états pré-migratoires des CN. Une autre observation clé est la présence d'une population principale de CN non biaisée vers un état particulier, exprimant les marqueurs des cellules immatures de la crête neurale, à partir desquelles toutes les autres trajectoires émergent. Cette trajectoire cellulaire non biaisée est maintenue pendant et après l'EMT, ce qui suggère qu'une population de cellules de la crête neurale très plastique, semblable à une tige, émigre et est soumise aux signaux du microenvironnement avant de choisir son destin.

**Le connectome CN et les bifurcations multiples décrivent la logique moléculaire des décisions relatives au destin cellulaire dans le CN vagal et crânien.**

Nous élargissons le réseau de gènes impliqués dans le NC-RRG en prédisant des corrélations de gènes à partir de l'ensemble des données de la CN pour relier les facteurs de transcription à leurs cibles potentielles dans le transcriptome de la CN. Le réseau résultant de >16.000 connexions TF-cibles potentielles a été partiellement validé par la combinaison des expériences knock-down et ChIP-seq pour certains des nœuds les plus liés du réseau. En particulier, nous nous sommes concentrés sur le rôle de Pax3 et TFAP2e en milieu de neurulation, qui étaient

auparavant connus pour leurs fonctions antérieures dans l'induction de la NC. Dans les cas de Pax3 et de Tfap2e, nous constatons que les trois approches aboutissent à des listes de gènes cibles qui se chevauchent partiellement en raison de l'utilisation de paramètres différents (par stade, par expression, etc.). Ensemble, ces données fournissent un génome élargi et validé spécifique à la CN pour deux spécificateurs BN/CN clés, révélant une représentation statique du RRG dans la CN.

Pour accéder à la dynamique temporelle de la régulation de la CN, nous avons examiné l'analyse des trajectoires temporelles afin de démêler la dynamique de l'expression génique spécifique à la branche qui sous-tend les bifurcations et la diversification de l'état. Pour chaque bifurcation, nous fournissons une liste de gènes clés susceptibles de contrôler les choix de ramification. Nous avons également utilisé la modulation expérimentale de la fonction de facteurs de transcription pivots dans la crête neurale pré-migratoire (Pax3, TFAP2e) pour valider ces prédictions de bifurcation. Cela donne une vue d'ensemble de la hiérarchie des décisions moléculaires qui régissent le réseau de régulation des gènes de la crête neurale crânienne et vagale, depuis l'induction à la frontière neurale jusqu'à la migration précoce, avec une validation expérimentale.

**Rôle des miARN dans le développement de la CN**

On sait qu'un grand nombre de gènes sont impliqués dans le développement de la CN, mais tous n'ont pas été déterminés. Récemment, la science est passée de l'étude des gènes codant pour les protéines à celle des gènes non codants, par exemple les miARN. Ces ARN non codants jouent un rôle important au cours du développement de l'embryon en régulant l'expression des gènes codant pour les protéines, principalement en se liant au 3'-UTR de l'ARN messager et en supprimant la production de protéines en raison de la déstabilisation de l'ARN messager. De nombreuses analyses ont montré que la régulation des gènes par les miARN est importante pendant le développement de l'embryon (Bushati et Cohen, 2007 ; Stefani et Slack, 2008). Les embryons de poisson zèbre mutants Dicer présentent des défauts de gastrulation et de somitogenèse (Giraldez et al., 2005) et meurent (Wienholds et al., 2003). Dans la rétine, la perte de Dicer entraîne une régression rétinienne à des stades plus avancés (Decembrini et al., 2008). De multiples analyses ont montré l'implication des miARNs au cours du développement neuronal (Papagiannakopoulos et Kosik, 2009 ; Walker et Harland, 2009). Nous avons donc décidé de développer cette direction et d'élargir notre compréhension du RRG-CN en étudiant les miARN dont l'expression a déjà été démontrée dans la CN et le tissu neural chez l'embryon de Xenopus (Godden et al., 2021 ; Ward et al., 2018).

Nous avons sélectionné les miR-219 et miR-196a pour déterminer leur fonction dans le développement de la CN, nous avons conçu des MO qui ont été conçus pour éliminer le miRNA mature et être complémentaires (Flynt et al., 2017). Grâce au séquençage de l'ARN sur des tissus de CN disséqués qui avaient été traités avec des MOs pour miR-196a et miR-219, il a été révélé que les miARN peuvent avoir un rôle dans le développement des CN et des tissus environnants. Une forte réduction de l'expression des marqueurs de la CN : foxd3, twist1, sox8, sox9, sox10, et snail2 peut être observée, après miRNA-KD, plus significativement avec miR-196a KD. Les marqueurs neuraux sox2 et sox3 sont significativement enrichis. Les marqueurs de la bordure des plaques neurales sont également enrichis (gmnn, hes3, pax3, et de façon non significative zic1), ainsi que le marqueur de pluripotence pou5f3. La caractérisation fonctionnelle de Sox10, Snail2 et Pax3 ainsi que les résultats de l'ARN-seq suggèrent que les miR-196a et miR-219 KD, qui réduisent l'expression de sox10 et snail2, empêchent le maintien des cellules de la CN. L'étude souligne l'importance de comprendre les différents rôles de miR-196a et miR-219 dans le développement des cellules de la CN dans les embryons de Xenopus. Plus important encore, les données présentées dans l'étude suggèrent que ces deux miARN ont des effets distincts sur le développement des cellules de la CN, le miR-196a ayant un effet plus important sur la réduction de l'expression des marqueurs de la CN et le passage du destin cellulaire à un état multipotent de type BN, et le miR-219 contribuant à une augmentation de la multipotence sans activer le programme BNB. Cela suggère que les miR-196a et miR-219 jouent des rôles différents au cours du développement des cellules de la CN.

**Relations entre Pax3, Dlx3 et Zic1 dans la bordure neurale**

Afin d'examiner avec précision les cibles précoces immédiates des facteurs de transcription les plus importants dans la bordure neurale (Pax3, Zic1 et Dlx3), nous avons lancé un nouveau projet par micro-dissection d'explants de la bordure neurale, ce qui nous permet d'obtenir une grande spécificité spatiale. Pour la première étape du projet (voir la section Résultats 2.4.1), nous avons recherché des cibles précoces immédiates spécifiques de ces facteurs de transcription dans le NB à plusieurs stades.

Jusqu'à présent, nous avons identifié un ensemble de gènes régulés par trois facteurs de transcription. Nos résultats confirment les rôles précédemment connus de ces facteurs de transcription dans la régulation des gènes impliqués dans la formation de la crête neurale, tels que les marqueurs des cellules de la crête neurale comme snai2, twist1, et sox9, ainsi que foxd3, et tfap2b. Il est également remarquable que la majorité des gènes identifiés comme cibles de Pax3 soient régulés à la baisse en l'absence de Pax3, ce qui confirme l'idée que Pax3 joue un

rôle dans l'activation des spécificateurs de la crête neurale. Cependant, un petit nombre de gènes se sont également révélés régulés à la hausse. En outre, nous avons constaté que Pax3 et Dlx3 ont des cibles régulatrices communes, qui jouent un rôle dans la formation de la crête neurale, comme alx1 et hnf1b. Nos données suggèrent que le rôle de Pax3 dans la formation de la crête neurale est plus important aux stades précoces, car nous avons observé moins de cibles pour Pax3 aux stades ultérieurs. De plus, en utilisant une analyse précédente de l'ectoderme par scRNA-seq, nous avons trouvé que Pax3-GOF activait à la fois les programmes NENP et NE/NBNC, tandis que Dlx3-GOF activait le programme NNENNE2 mais désactivait les programmes NCNP. Dans l'ensemble, notre analyse a montré des schémas d'activation distincts pour Pax3 et Dlx3. Cette étude apporte un nouvel éclairage sur les réseaux de régulation transcriptionnelle qui contrôlent la formation de la crête neurale. En identifiant les gènes et les voies spécifiques qui sont ciblés par ces facteurs de transcription, nous pouvons acquérir une meilleure compréhension des processus moléculaires complexes impliqués dans le développement de la BN, CN, PC et d'autres tissus ectodermiques.
Les prochains objectifs du projet sont [1] l'examen de la coopération entre les spécificateurs de la BN pendant l'induction du destin CN/PC en défiant les cibles précoces immédiates des gènes en l'absence d'un autre, par exemple, les cibles de Zic1 en l'absence de Pax3/Dlx3. Et [2] en recherchant des exigences spécifiques au stade des spécificateurs de la BN en analysant le transcriptome de la BN en l'absence de l'un des gènes à l'aide de morpholinos.

**Évolution des types cellulaires avec scEvoNet**

Au cours des 150 dernières années, les types cellulaires ont été étudiés en termes de développement, de morphologie et de caractéristiques moléculaires, alors qu'avec l'avènement de la transcriptomique sur cellules uniques, de nombreux types cellulaires distincts et cachés ont été identifiés. Avec l'utilisation de scRNA-seq, une ère complètement nouvelle s'ouvre pour comprendre l'évolution des types cellulaires. Apparemment, les types ancestraux de cellules animales étaient polyfonctionnels (certaines cellules d'éponge ont à la fois des propriétés neurales et immunitaires), mais au cours de l'évolution, les fonctions ont été réparties entre différentes cellules (Arendt, 2008). Cette "division du travail" est l'une des clés de l'évolution des types de cellules. Pourtant, il existe un manque d'outils computationnels scRNA-seq pour étudier l'évolution des états cellulaires. Nous avons donc développé un outil informatique, scEvoNet, qui permet d'étudier ce concept à partir de données scRNA-seq. Tout d'abord, notre outil crée la matrice des similitudes entre les types de cellules de deux ensembles de données fournis par les utilisateurs. Deuxièmement, il construit le réseau bipartite reliant les gènes et

les différents états cellulaires, qui peut être utilisé comme source de sélection des gènes potentiellement cooptés (partagés entre les types de cellules), ainsi que des gènes qui sont importants pour le type de cellule particulier dans la comparaison inter-espèces, et des gènes qui ne sont significatifs que pour une seule espèce. En outre, il peut être utilisé pour étudier la progression et l'évolution des tumeurs en explorant les gènes et les programmes de gènes qui sont partagés entre divers types de cellules dans le microenvironnement tumoral. L'utilisation de cet outil pour étudier le continuum des états du transcriptome entre les stades de développement ou les espèces, ou entre les états normaux et pathologiques, permettra d'expliquer la dynamique complexe des états cellulaires. Les événements d'abandon, lorsqu'un gène n'est pas détecté alors qu'il est présent, et les effets de lot peuvent limiter la précision des prédictions faites à partir des données scRNA-seq. La génération d'ensembles de données provenant de plusieurs espèces ou l'utilisation de techniques de capture scRNA-seq améliorera les résultats d'outils tels que scEvoNet.

**Conclusions**

Mon travail vise à améliorer de manière significative la compréhension des mécanismes moléculaires impliqués dans la formation de la crête neurale. Bien que de nombreuses questions aient trouvé une réponse, il reste encore des domaines à explorer, comme la transition des cellules de la CN de type PN vers les cellules ENSp. En outre, l'application de scEvoNet aux données de diverses espèces, y compris les céphalochordés, les tuniciers et les vertébrés, permettrait d'obtenir des informations précieuses.

# Table of Contents

# 1. General introduction

## 1.1 Neural crest gene regulatory network

The neural crest is a group of multipotent stem cells found in vertebrates that gives rise to a wide variety of cell types, including Schwann cells, melanocytes, and craniofacial bones. Defects in neural crest development can cause multiple birth defects and syndromes. The neural crest was first identified by His in 1868 in a developing chick embryo. It is located on the "crest" of the neural tube and has been studied using fate mapping and quail-chick heterospecific grafts to understand its role in development along the anterior-posterior body axis (Le Douarin and Kalcheim, 1999). The anteriormost population is the cranial neural crest, emerging from the dorsal neural tube between the posterior part of the forebrain to the posterior hindbrain levels. The cranial neural crest gives rise to craniofacial cartilage, bone, connective tissue, smooth muscle, neurons of the trigeminal ganglion as well as head peripheral neurons and glia, and melanocytes. The more posterior population is the vagal neural crest, which arises adjoining the 1-7 somites. The vagal neural crest is crucial for the formation of the enteric nervous system (Le Douarin and Teillet, 1973). From 1 to 3 somites there is the cardiac neural crest plays an essential role in ventricular septum formation and aorticopulmonary septation. Beyond somite 7 is the trunk neural crest, which contributes to the formation of melanocytes, dorsal root, and sympathetic ganglia. Lastly, at the caudal level to somite 28 in chick embryos (and caudal to somite 24 in embryonic mice and humans), there is a sacral neural crest giving rise to the enteric nervous system as well (Figure 1).

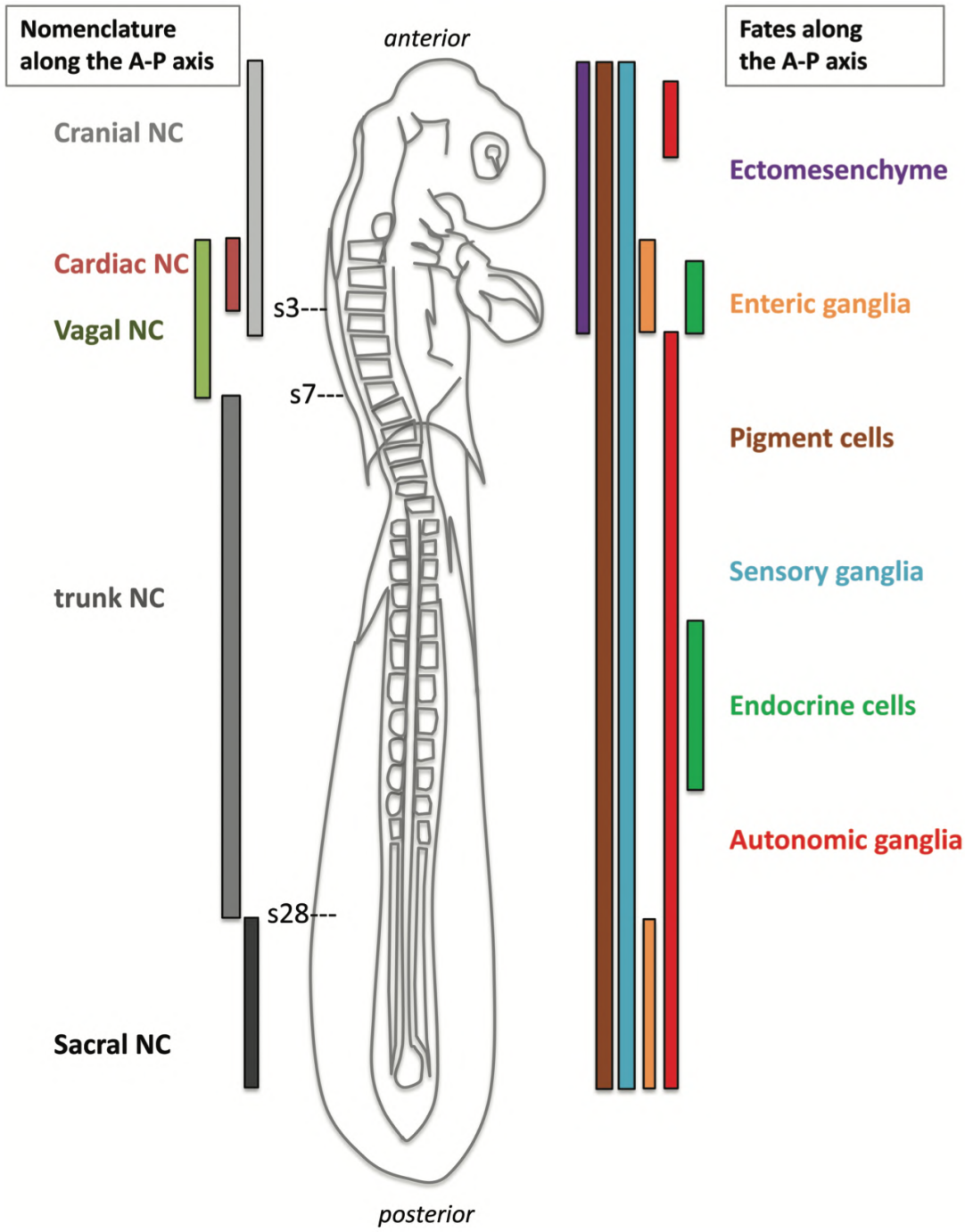



**Figure 1. Different neural crest cell populations (depending on the position along AP axis) and fates.** A scheme of a chick embryo after two days of incubation is shown. The NC is divided into four regions: the cranial, which includes the cardiac NC (from the posterior diencephalon to somite 4); the vagal NC (somites 1 to 7); the trunk NC (somites 7 to 28); and the sacral NC, which is located posterior to somite 28 (Alkobtawi and Monsoro-Burq, 2020)

Neural crest has been the focus of intense research for 150 years and currently gathers one of the most advanced gene networks in vertebrate development. Neural crest cells are induced in the ectodermal germ layer during gastrulation. They are located in the zone bordering the neural plate, in the edge between the neural plate and the nonneural ectoderm (Figure 2). Subpopulations of cells with neural crest origin can be distinctly traced throughout stages of embryonic development using experimental cell labelling or combinations of specific marker genes when they can be defined. During gastrula stage, *pax3*, and *pax7* genes label the lateral and posterior neural border, while *zic1* labels specifically the anterior NB (Seal and Monsoro-Burq, 2020). The formation and specification of the neural border into neural crest is regulated by the cooperative activity of several signaling pathways, such as WNT and BMP, as well as by TFs such as Tfap2 family, Pax3 and Pax7, Zic1, and Hes4; and this trend is consistent among vertebrate species (de Crozé et al., 2011; Hong and Saint-Jeannet, 2005; Milet et al., 2013; Monsoro-Burq et al., 2005; Plouhinec et al., 2017; Sato et al., 2005; Seal and Monsoro-Burq, 2020). WNT/BMP signals arise from lateral and posterior parts of the body axis, and while WNT/BMP inhibitors are expressed in medial areas such as the neural plate midline and the notochord (for BMP antagonists) or anterior body axis levels (for WNT antagonists). The resulting equilibrium between triggering and inhibiting signals develops a gradient of BMP and WNT activity, and the neural plate border cells arise within a location exposed to medium levels of BMP/WNT activity (Groves and LaBonne, 2014). Delicate modulations of the cell response to these external signals further refine the dynamics of these gradients (Alkobtawi et al., 2021). The Notch pathway might participate in the patterning of the neural plate border as well (Endo et al., 2002; Monsoro-Burq et al., 2005). The specific sources of all these triggers and inhibitors vary across species, but generally, the combined effort of these signals generates the neural plate border zone and activate neural plate border markers including Msx1, Tfap2a, Gbx2, Zic1, Pax3, Dlx3/5, Gata2/3, Foxi (Khudyakov and Bronner-Fraser, 2009; Meulemans and Bronner-Fraser, 2004; Monsoro-Burq et al., 2005; Nichane et al., 2008). Tfap2a appears to be a critical activator of neural plate border markers, and also necessary for the expression of Foxi genes and Gata genes (Bhat et al., 2013). The role of Tfap2a during the induction of neuroblasts and neural crest cells has been mainly studied in frog embryos. There, Tfap2a is expressed in the non-neural ectoderm at the beginning of gastrulation and is upregulated in the neuroblast region during mid-gastrulation as a result of WNT signaling.

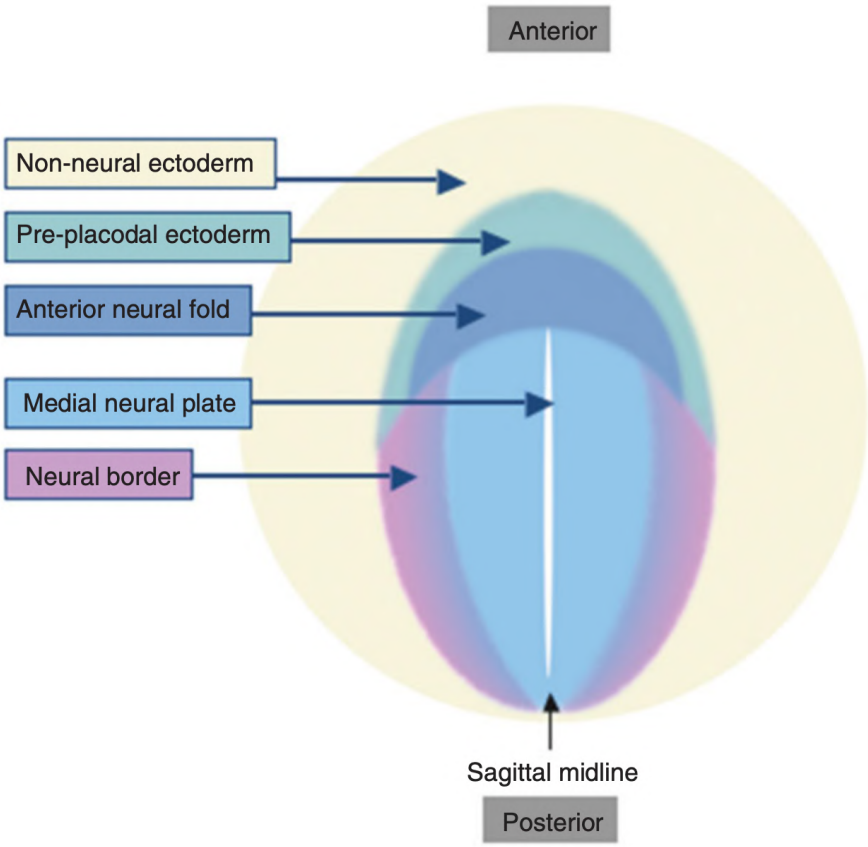



**Figure 2. Scheme of ectoderm positioning.** The patterning of the ectoderm is followed by interactions that define the neural crest (Pegoraro and Monsoro-Burq, 2013)

The function of Tfap2a is crucial at all stages of neuroblast/neural crest development, from neuroblast specification to later stages of neural crest cell development. This has been demonstrated in several studies using frog embryos (de Crozé et al., 2011; T. Luo et al., 2003). Lateral neural border genes Dlx5/6, Gata2/3 and Foxi1/3 regulate each other in numerous regulatory loops (de Crozé et al., 2011; Grocott et al., 2012; Kwon et al., 2010; Matsuo-Takasaki et al., 2005; McLarren et al., 2003). These relations "lock down" the regulation state and allow the stabilization/maintenance of expression of neural plate border markers, which are often retained in the developing progenitors through later stages of development.

The process of inducing the neural crest from the neural border zone is characterized by the activation of a group of genes known as "neural crest specifiers" (see below) by the activity of NB specifiers (see above) and signaling molecules found in the medial neural border region. These neural crest specifiers trigger a process called epithelial-to-mesenchymal transition (EMT), which allows the neural crest cells to lose their epithelial character, to move away from the ectoderm and to acquire the ability to migrate. The neural crest specifiers also cross-regulate each other, helping to maintain this unique state (Sauka-Spengler and Bronner-Fraser, 2008). The earliest NC markers which are expressed in the developing NC in the chick: FoxD3, Ets1 and Snai1/2 (Khudyakov and Bronner-Fraser, 2009). In frog embryos, Sox8 and Sox9 are among the earliest genes expressed in the neuroblast region that specify neural crest cells, preceding the expression of Snai2 and Foxd3. In contrast, Sox10 is expressed later in the pre-migratory neural crest cells (Alkobtawi et al., 2018; Hong and Saint-Jeannet, 2005; Spokony et al., 2002). Both Sox9 and Sox10 are activated by the canonical WNT pathway, and the expression of Sox10 is controlled by Sox9 and Snail2 (Hong and Saint-Jeannet, 2005; Spokony et al., 2002). FoxD3 is controlled by the neural border zone specifiers Pax3/7 and Msx1 (Monsoro-Burq et al., 2005; Simões-Costa et al., 2012). Pax3 is cooperating with Zic1 driving NC specification in frog (Bae et al., 2014; Monsoro-Burq et al., 2005; Plouhinec et al., 2014; Sato et al., 2005). For instance, in *Xenopus laevis*, Zic1 and Pax3 bind to the promoter of Snai1/2 and activate them as immediate-early targets (Plouhinec et al., 2014). Pax3 also directly regulates FoxD3 expression in the chick (Simões-Costa et al., 2012). Generally, it suggests that Pax3, Msx1, together with Zic1 are essential for the accession of NC identity, and activation of multiple NC markers (Milet et al., 2013; Monsoro-Burq et al., 2005). The importance of Msx1, Pax3 and Zic1 in neural crest program initiation is clear, but other genes in the neural border, such as Myb, Myc, and Prdm1a (Bellmeyer et al., 2003; Betancur et al., 2010; Powell et al., 2013), are also essential for the NC specification. Moreover, additional

signal input appears to be necessary for the expression of NC genes: suppressing the WNT signal effects downstream of NB induction results in the loss of NC markers Snai2 and FoxD3 (García-Castro et al., 2002; Monsoro-Burq et al., 2005; Simões-Costa and Bronner, 2015), and WNT has been demonstrated working together with Zic1/Pax3 to drive NC specification (de Crozé et al., 2011; Sato et al., 2005).

The process of neural crest specification culminates in a distinct cell state that, in cranial region, emerges located between the neural plate and the posterior placodal region. This pre-migratory NC expresses the full complement of marker genes, including FoxD3, Snai2, and Sox9, as well as some neural border genes such as Tfap2a and Zic1. After specification, NC cells undergo delamination and begin to migrate. The process of epithelial-to-mesenchymal transition (EMT), which is required for NC migration, is complex and involves the regulation of many genes. The WNT signaling pathway is thought to play a role in the regulation of EMT, specifically through the control of Snai1 or Snai2 (Yook et al., 2006). One of the main drivers of EMT in NC is the modulation by NC specifiers of the structural genes maintaining of the epithelial tissue morphology: Ecad (cadherin type 1, associated with high adhesiveness)/Ncad (cadherin type 2, associated with low adhesiveness; Figure 3). Further activation of the type 2 cadherins changes cell-cell adhesion and initiates EMT, involving cell surface modifications that lead to the revocation of adherens junctions. Transcriptional control of cadherins is the central to this process. Several neural crest specifier genes participate in cadherins regulation during EMT, with subtle species-specific variations. For example, Sox10 represses Ncad in migratory cells in chick, leading to lower levels in the neural crest than in the neural tube (Cheung et al., 2005). Similarly, overexpression of FoxD3 in chick NC leads to decreasing of expression of Ncad and increased expression of Cad7 (Cheung et al., 2005; Dottori et al., 2001). In frog embryos, the expression of cadherin1 decreases at the beginning of NC EMT, while the expression of cadherin2 is increased (Theveneau et al., 2010). In all species studied so far, Snai1 and 2 TFs function as repressors that downregulate type 1 cadherins upstream of EMT. Notably, Snai genes also represses type 2 cadherin Cad6b (Taneyhill, 2008), that should be downregulated for the NC delamination (Coles et al., 2007). Sox9 is an important gene that is associated with Snai1 in EMT and plays a role in driving EMT processes together with Snai1. (Cheung and Briscoe, 2003; Liu et al., 2013). EMT also includes the activation of the TF Twist1 mediating together with Snai genes changes in cell-cell interaction that allow for delamination and dispersion of neural crest cells (Lander et al., 2013). All this led to the loss of the polarized epithelium phenotype and the acquisition of cell motility (Bahm et al., 2017; Morrison et al.,

2017). In addition to modulating adhesion, other structural changes are required to allow the detachment and dispersion of emigrating cells. These include the degradation of the basement membrane by metalloproteases promoting cell invasion (Simões-Costa and Bronner, 2015).

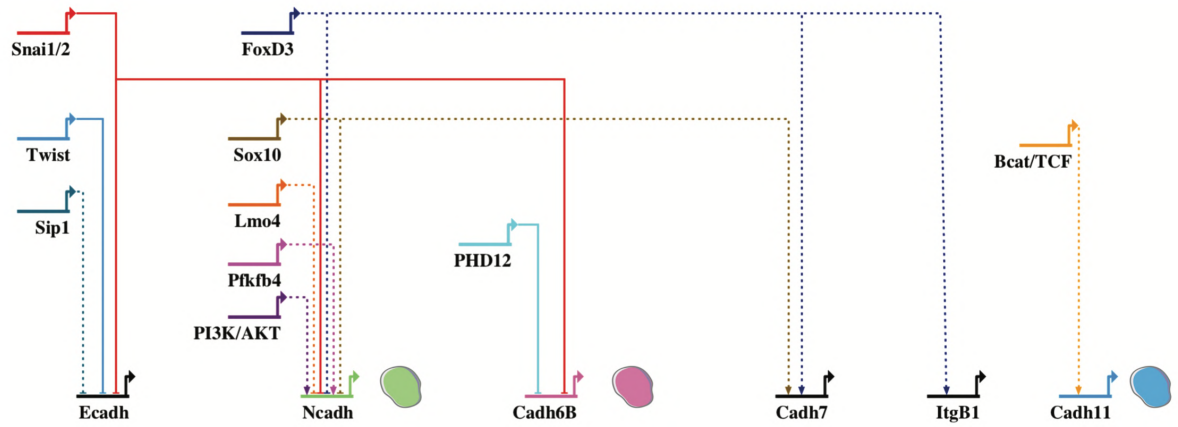



**Figure 3. The main factors that contribute to neural crest epithelial-mesenchymal transition.** TFs are summarized in a network that covers the transcriptional activity of NC specifiers and their partners on cadherins and β integrin during NC EMT. It also includes the posttranslational activity of metalloproteinases on cadherins. Direct regulation is represented by a solid line, while indirect regulation is represented by a dotted line (Alkobtawi and Monsoro-Burq, 2020).

## 1. 2 Neural crest multipotency

Numerous studies conducted over the last decades confirm the hypothesis that the neural crest is a multipotent cell population with the capacity to give rise to multiple potential derivatives. The unicellular injection of fluorescent dextran molecules into the trunk and cranial neural crest cells before their delamination first confirmed the neural crest multipotency (Bronner-Fraser and Fraser, 1988). They showed that the injected cells descendants were able to differentiate into multiple cell types such as Schwann cells, and melanocytes, hence meaning that some cells of the trunk neural crest were multipotent (Figure 4). In series of other studies, researchers cultured clones of cranial neural crest cells obtained from a single quail neural crest cell and showed that the cells displayed a broad developmental potential, confirming the results obtained by Bronner-Fraser for the whole complement of NC derivatives (Baroffio and Blot, 1992; Dupin et al., 2018). Recently, in precise vivo labeling with the Confetti system proved that individual neural crest precursors are capable of generating multiple cell types (Baggiolini et al., 2015).

Multipotency refers to the ability of stem cells to differentiate into various specific cell types. This process involves receiving and responding to certain signaling cues in order to maintain a state of "intermediate" flexibility, allowing the stem cells to differentiate into different cell types. Multipotency is broader in early development and progressively restricted. For example, frog blastula cells of the blastocoele roof are pluripotent as they can colonize the tree germ layers experimentally and provide various derivatives in those three domains (Buitrago-Delgado et al., 2015). The mechanism allowing neural crest multipotency has been matter of debate, as it defies the usual understanding of progressive restriction in potential during development: NC has a much broader potential than its progenitor tissue the ectoderm (Artinger and Monsoro-Burq, 2021). Mechanisms shared between blastula cells and NC cells have been explored. In neural crest as well as in blastula cells genes snai2 and sox5 seem responsible for maintaining a multipotency state (Buitrago-Delgado et al., 2015). In this article, the authors considered NC as a part of the ectoderm which kept some of the characteristics of the early blastula during gastrulation. This contrasts with the concept that the neural crest is a cell population that becomes multipotent through the instructive WNT, FGF, and BMP signals.

To reconstruct the principles governing NC multipotency, VENTX2/NANOG gene family, as well as OCT4 orthologs Pou5f3 genes were determined as a vertebrate-specific invention, operating downstream of NB specification, necessary for the expression of

multipotency markers in the early NC, such as Oct25/Oct4 and *tert*, and required for the formation of the ectomesenchyme lineage but not for the sensory neuronal lineage (Scerbo and Monsoro-Burq, 2020), suggesting that during specification, NC progenitors actively gain the ability to differentiate into multiple cell types, rather than retaining multipotency from the blastula stage. The presence of blastula-stage pluripotency genes may inhibit the formation of the neural crest.

In another work, Zalc et al. proposed that cranial NC cells extend their developmental possibility through temporary reacquisition of the pluripotency signature, specifically by Oct4 reactivation which is needed for the following formation of ectomesenchyme (Zalc et al., 2021).

Hovland et al. found that the OCT4-SOX2 dimer is necessary to establish an epigenomic signature specific to the neural crest, which is lost as the cells commit to their fate. The targets of the OCT4-SOX2 dimer in the neural crest are distinct from those of embryonic stem cells, suggesting that the dimer has context-specific functions. The binding of OCT4-SOX2 to neural crest enhancers requires the presence of the TFAP2A, which interacts with the dimer. Thus, Yamanaka factors are being adapted for use in multipotent cells to control chromatin organization and determine their potential for development (Hovland et al., 2022).

WNT is a central modulator of neural crest uniqueness, working reiteratively during the formation and differentiation of the NC (Raible and Ragland, 2005). NC cells are supposed to lose multipotency after their delamination from the neural tube and become cell type-restricted relying on the path of their migration. Ventrally migrating NC cells are limited to neural and glial cell fates, and dorsolaterally migrating NC cells are limited to a melanocyte fate (Le Douarin and Kalcheim, 1999). Numeral studies have tried to define when and where NC cells lose their multipotency and become restricted in their fate. Clonal analysis of NC cells showed half of the initial NC cells exist as a cell type-restricted cells, yielding clones of just one cell type (Henion and Weston, 1997), while the other half includes bipotent precursors (Lahav et al., 1998). A population of early NC cells which express the receptor tyrosine kinase Kit only express melanocyte cell type markers in rodents. Moreover, avian Kit+ NC cells always gave rise to clones comprising only melanocytes (Henion and Weston, 1997; R. Luo et al., 2003; Wilson et al., 2004). This indicates that the expression of Kit is an early determinant of cell type fate in dorsolaterally migrating NC cells. Consistent with this idea, none of the ventrally migrating NC cells were found to express Kit (Wilson et al., 2004) or differentiate into

melanocytes even under conditions that preferred the development of melanocytes (Reedy et al., 1998). These studies suggested that migratory NC cells might comprise a population of fate-restricted precursors that give rise to single or dual cell types. Bhattacharya et al. proposed the mechanism through which Wnt, pluripotency factor Lin28a and let-7 miRNAs control the deployment and the following silencing of the multipotency program in a position-dependent manner. The increased levels of the Lin28a inhibit let-7, protecting the NC GRN from repression by let-7. As NC cells migrate out from the Wnt source, Lin28a levels are decreased, resulting in an expression growth of let-7 and succeeding repression of multipotency factors (Bhattacharya et al., 2018)

The molecular principles underlying NC multipotency stay incompletely understood. NC express multiple markers of multipotency: c-Myc, Sox2, Oct4, and Snai1, Sox5, which are also involved in multipotency in blastula cells (Buitrago-Delgado et al., 2015). Based on these shared gene sets, it was proposed that the NC cells keep a multipotent condition comparable to blastula multipotent cells, instead of re-acquiring another multipotency program after neurulation. Nevertheless, this gene set is not only discovered in multipotent cells and therefore cannot be assumed as a “multipotency signature” accounting for the unique ability of NC to form multiple derivatives (Briggs et al., 2018).

In chick embryos, earlier in development at the open neural plate stage, a small fraction of cells found at the neural/nonneural border co-express markers of neural plate, NC, and placode progenitors, in variable proportion (Roellig et al., 2017). In terms of differentiation, the functional meaning of such gene co-expression is yet to be defined. Models of "multilineage priming" or "oscillatory gene expressions" are being discussed. In the "multilineage priming" model, a stem cell is able to differentiate into multiple cell types through a process called "priming," which involves the activation of certain genes and the suppression of others. This process allows a stem cell to become primed for differentiation into a particular cell type, but it does not necessarily mean that the cell will differentiate into that cell type. In the "oscillatory gene expression" model, stem cells regulate their pluripotency and self-renewal through the cyclical activation and suppression of certain genes. This process is regulated by a number of factors, including the presence of signaling molecules and the activity of transcription factors (Hu et al., 1997; Imayoshi et al., 2013). Robert Kelsh et al. proposed a model in which adult pigment stem cells undergo a "progressive fate restriction" process, in which they progressively lose the ability to differentiate into certain cell types as they differentiate into others. This process is thought to be regulated by a combination of intrinsic and extrinsic factors, including

the activation of certain genes and the presence of signaling molecules (Kelsh et al., 2017). Also, Sox10 appears as a key element in the NC multipotency GRN, as it is essential for the progenitor induction in multiple NC lineages and because its prolonged-expression limits the commitment of several lineages while maintaining immature progenitors (Kim et al., 2003).

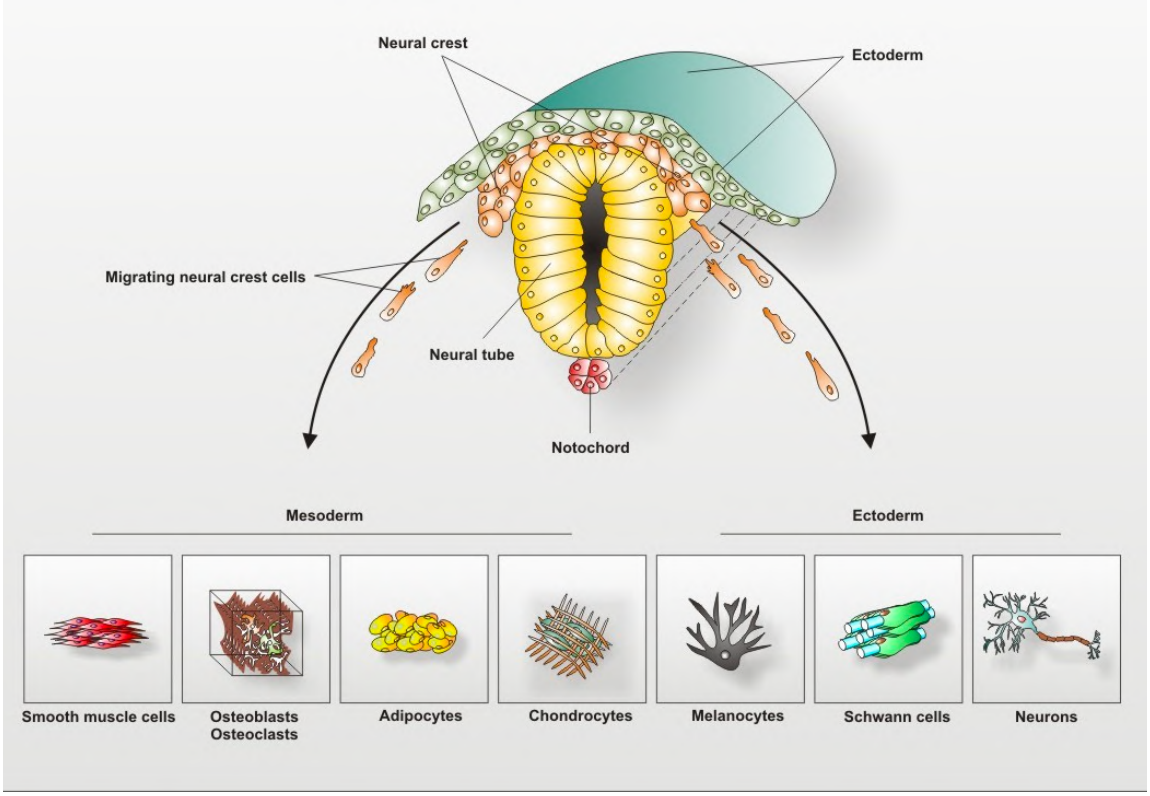



**Figure 4. Cell types derived from neural crest cells in adults.** Neural crest cells migrate out of the developing neural tube and differentiate into multiple cell types, including peripheral neurons, adipocytes, smooth muscle cells, and Schwann cells ("Neural Crest-Derived Stem Cells - Bielefeld University," n.d.)

Thus, the current data suggests that NC cells exhibit a certain degree of heterogeneity in their ability to differentiate before and during migration from the neural folds. While some NC cells may lose their early multipotency, multipotent cells can still be found in subsets of NC cells during migration in both avian and mammalian species. Additionally, NC-derived cells isolated at post-migratory stages still contain oligopotent progenitors, such as those that give rise to Schwann cells and pigment cells (reviewed by Dupin and Coelho-Aguiar, 2013; Dupin and Sommer, 2012).

## 1.3 NC origin and evolution

The origin of vertebrates about 500 million years ago was a milestone in animal evolution, which has since led to the diversification of millions of species. One of the most important embryological features common to all vertebrates is the neural crest and the cranial placodes. Early chordates probably did not have a "bona fide" neural crest cell population delaminating from the central nervous system but maybe a few neural crest-like cells (Jeffery, 2007). Most likely, the first chordates presumably inherited the neural plate border involved in the formation of the medial-lateral pattern of the embryonic neuroectoderm and the production of PNS neurons, as well as the ability to generate sensory neurons from the ventral epidermis (Zhao et al., 2019).

With the lineage evolution leading to the tunicates (protochordates) and vertebrates, we see the earliest evidence of the neural crest (Figure 5). While protochordate neural tube structures are relatively simple, they contain the basic structure (differentiating into four regions: the fore-midbrain, midbrain–hindbrain boundary, hindbrain, and spinal cord; dorsoventral patterning, but some signaling systems may be lost) and many of the genetic mechanisms that are also found in vertebrates. These elements have been preserved throughout the lineage leading to these organisms (Holland and Chen, 2001). A considerable number of genes are expressed in similar patterns in the neural crest of vertebrates and the dorsal midline epidermis of protochordates, suggesting a connection between these two structures in the evolutionary history of these organisms: Pax3/7, Bmp family, and Dlx genes (Franz, 1993; Holland et al., 1996; Liem et al., 1995; Panopoulou et al., 1998; Robinson and Mahon, 1994; Wada et al., 1996). These patterns are most noticeable in the dorsal epidermis of ascidians, which give rise to epidermal and sensory neuronal cells. This suggests that the evolutionary origins of the neural crest may lie in the dorsal midline epidermis, possibly including the dorsal neural tube. If the hypothesis of the origin of the neural crest from the protochordate dorsal mid-epidermis is accurate, then still pronounced differences remain between the neural crest of vertebrates and its “signs” in protochordates; for example, such key characteristics of the neural crest as the ability to migrate and pluripotency. Therefore, there is an assumption that the evolution of the neural crest of vertebrates does not consist in the origin of a new cell type, but in the acquisition of new cellular properties (Wada, 2001). In addition to the previously mentioned migration and pluripotency, the neural crest has acquired another important new characteristic: the transfer of anteroposterior positional information (Trainor and Krumlauf,

2000). Because the NC contributes to skeletal tissues or cranial ganglia that exhibit a distinct pattern of innervation consistent with their rhombomeric origin, the anteroposterior information of the neural crest is significant for the development of craniofacial structure in vertebrates facilitating predatory lifestyle by generation of novel vertebrate head structures (Engleka and Epstein, 2016; Muñoz and Trainor, 2015). Regarding the appearance of the migration property, the epidermal cells of the dorsal midline of Amphioxus show migration (Holland et al., 1996), but these cells do not stratify, and the migration occurs as a "sheet" of cells. Therefore, delamination before migration is unique to the vertebrate neural crest. For this process, the switching of cell adhesion molecules is essential: cdh2 and cdh6 are expressed in pre-migrating neural crest cells, and these cell adhesion molecules must be downregulated for neural crest cells to delaminate (Taneyhill, 2008).

Using knowledge of NC GRN, several attempts were made to determine the evolutionary rudiments of NC among invertebrate chordates, the tunicates, and, the cephalochordates which were assumed as the closest living invertebrate relatives of the vertebrates until recently. The main approach was to analyze the expression of neural crest markers during stages of neurulation when neural crest cell delamination occurs in vertebrates (Pasini et al., 2006; Stone and Hall, 2004). Nevertheless, Bronner-Fraser and colleagues rather than trying to describe NC by one or more putatively definitive genes, they attempted to find the whole GRN underpinning NC induction, differentiation and migration (Sauka-Spengler et al., 2007). Using this definition, they have aimed to compare theoretical GRNs between vertebrates and invertebrate chordates, attempting to find evidence of the evolutionary emergence of neural crest and the assembly of its gene regulatory network (Meulemans and Bronner-Fraser, 2004; Sauka-Spengler et al., 2007). Both gene-based and gene-network-based approaches for NC among invertebrate chordates have shown that Bmp2/4 is expressed in the non-neural ectoderm and that Pax3/7, msx1/2, Zic, and snail are expressed at the neural plate border. Yet, expressions of Id, Ets, Twist, SoxE, and FoxD, were not seen at this position. Clearly, the majority of genes involved in NC GRN are expressed in mesodermal, rather than neural plate derivatives. These outcomes indicate that, although parts of the genetic program that controls NC in development were present in invertebrate chordates, it was not until the emergence of NC at the establishment of the vertebrate lineage that the GRN was ultimately established (Simões-Costa and Bronner, 2015). Therefore, an evolutionary rudiment of NC/NB cells exists in invertebrate chordates, though it was just after the evolution of the vertebrates that the other essential genes required for NC production were recruited into this program

(Martinez-Morales et al., 2007). Therefore, it was defined that NC cells are an exclusively vertebrate characteristic.

Overall, the evolution of NC cells is a complex and multifaceted process that is still not fully understood. Further research is needed to better understand the mechanisms underlying the evolution of NC cells and their role in the development of vertebrates. It is necessary to reassess this process as a series of acquisitions of novel and unique cell properties, such as pluripotency and migration. By doing so, we can better understand the complex history of the neural crest and its development in different organisms. These new cellular properties are essential to the evolution of the finely patterned craniofacial structures that are a defining feature of vertebrates. One limitation is that there are relatively few NC-specific gene expression datasets available, which limits our ability to study the evolution of NC cells and the mechanisms that underlie their development. Thus, to better understand the gene expression patterns of NC cells and how they have evolved, it would be helpful to generate new scRNA-seq datasets from a variety of species (from cephalochordates to tunicates and vertebrates). These datasets could be used to identify NC-specific gene expression patterns and to compare NC gene expression across different species.

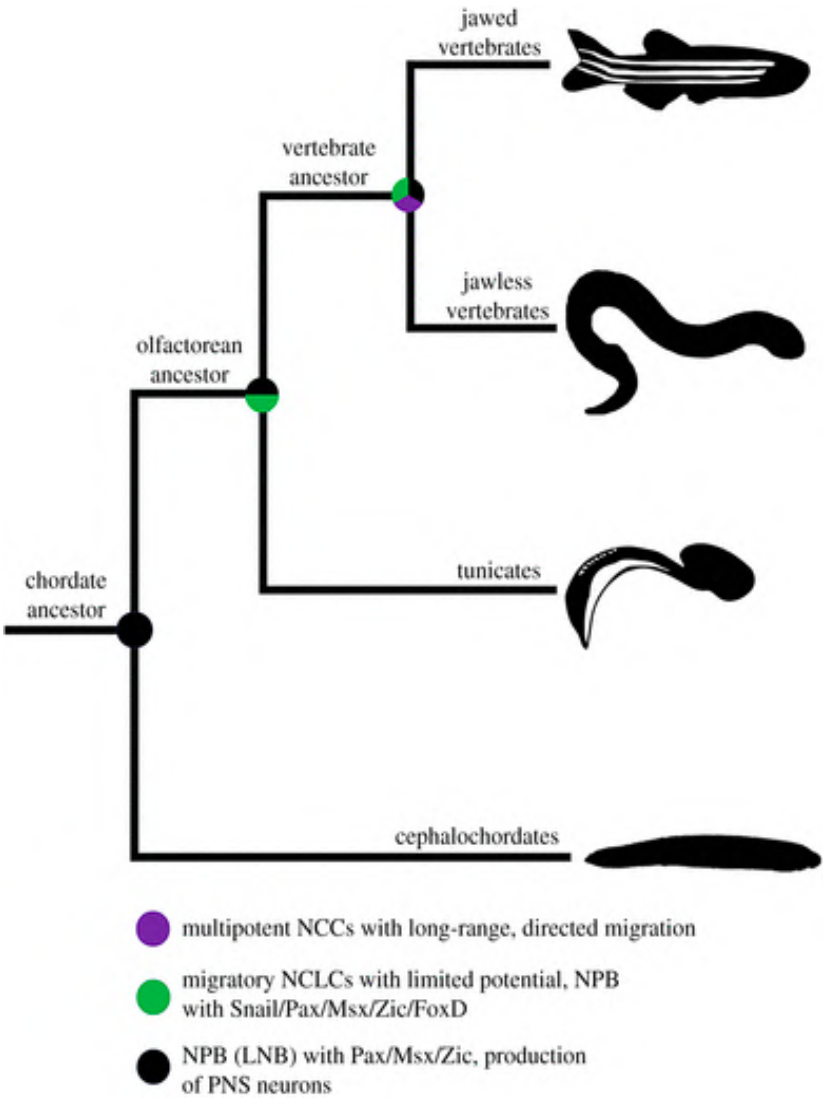



**Figure 5: A model proposing the evolution of neural crest cells through successive stages, traced back to their ancestral progenitors within the chordate phylogeny.** The patterning of the neuroectoderm into a peripheral nervous system (PNS) and the generation of cells that can migrate throughout the embryo predate the emergence of chordates and vertebrates, and are therefore is not unique to neural crest, suggesting that early chordates likely inherited these capabilities from earlier bilaterian ancestors. It is these capabilities (a neural border with migratory cells) that may provide a developmental foundation for the evolution of neural crest-like cells in the ancestors of tunicates and vertebrates, following their separation from the cephalochordate lineage approximately 600 million years ago (York and McCauley, 2020)

## 1.4 Gene regulatory networks discovery

Gene regulatory networks (GRNs) play a crucial role in development by controlling the activity of genes.

GRNs are hierarchical: Davidson has proposed that GRNs are hierarchical in nature, with certain genes acting as "master regulators" that control the expression of other genes. These master regulators are thought to act at the top of the hierarchy and to play a key role in orchestrating the expression of downstream genes.

GRNs are modular: Davidson has also suggested that GRNs are modular, with different modules of genes acting as functional units that are responsible for specific developmental processes. These modules are thought to be relatively autonomous, with each module acting independently to control the expression of a specific set of genes.

GRNs are dynamic: GRNs are dynamic and responsive to changes in their environment, including both extrinsic factors such as signaling molecules and intrinsic factors such as the activity of other genes. Davidson has proposed that this dynamic nature of GRNs allows them to respond to changes in their environment and to adjust gene expression in response to these changes (Davidson, 2010, 2009; Erwin and Davidson, 2009).

These networks are made up of transcription factors (TFs) that bind to specific DNA sequences in target genes and regulate their activity. The binding of these transcription factors can be influenced by a number of factors, such as the presence of coactivator or corepressor proteins, the overall structure of the DNA, and chemical modifications to the DNA (Blitz et al., 2017).

The construction of such networks requires the identification of the involved regulatory genes and the characterization of their temporal and spatial patterns of expression. Different types of interaction (activation/repression) between TFs or signals can be investigated using perturbation analysis, in which the function of each gene is specifically blocked. The resulting expression changes are then integrated to reveal direct and non-direct relationships and uncover the structure of the GRN and epistatic relationships. Modulating the expression of regulatory genes can have various effects on an organism, depending on the importance of the network linkage that is disrupted. If the function of a key node is affected, it can potentially disrupt critical processes in embryo development and cause significant changes to the phenotype.

However, in some cases, the changes to the phenotype may be minimal if the affected gene has a less significant role in the network (Li and Davidson, 2009). To establish a GRN by interpreting perturbation data it is essential to follow certain logic rules (Li and Davidson, 2009):

- Temporal restriction: (1) direct transcriptional activation requires simultaneous expression of TF, and the target gene or TF should be expressed before the target gene;
- Spatial restriction: (1) TF and target gene should be expressed in the same domain; (2) direct transcriptional repression requires that the repressor and its targets do not overlap in the same region;
- Parsimonious topology: (1) the direct target of the transcription factor is influenced more strongly than the indirect; (2) the interaction is considered indirect if the identified direct links are enough to explain the data;
- Special linkages: double negative control logic is needed when tampering with a canonical repressor causes a decrease in target gene expression.

The same rules are used when building a regulatory gene network based on other types of data.

One of the most accurate and by far the one of the most common types of disturbance used in the construction of GRN was the introduction of morpholine-substituted antisense oligonucleotides (MASO). Although in certain circumstances the gene expression is also controlled in other ways. Morpholine oligonucleotides block the access of other molecules to small (about twenty-five nucleotides long) specific sequences due to complementary pairing with the corresponding RNA. Morpholine oligonucleotides serve as research tools in reverse genetics for gene knockdown. From the very beginning of the development of the GRN, the MASO blockade of gene expression has been the main tool (Levine and Davidson, 2005; Monsoro-Burq et al., 2005; Sinner et al., 2006). MASO embryonic toxicity is low, and it can be inserted by microinjection. However, morpholinos may have unintended effects on the expression of other genes, and that this can cause "artifacts" or misleading results in research studies (Gerety and Wilkinson, 2011). To minimize possible artifacts, it is important to use multiple experimental approaches and consider multiple lines of evidence. This can help to validate the results of morpholino studies and increase confidence in the conclusions that are drawn (Eisen and Smith, 2008).

One example is the work of Schlosser & Maharana where was used microinjections of mRNAs and Morpholino antisense oligonucleotides (Mos) into embryos of *Xenopus laevis* to study how early ectodermal TFs FoxI1a, Ventx2, Msx1, TFAP2, and Zic1, Pax3 and Hes4 affect the establishing of pre-placodal ectoderm, neural crest, and neural plate regions during gastrulation (Maharana and Schlosser, 2018). Starting with MO-mediated knockdown of early ectoderm TFs they investigated that expression of Six1 and Eya1 in pre-placodal ectoderm was suppressed in a high proportion of embryos after knockdown of both dorsally limited TF Zic1, Pax3, and Hairy2b, and ventrally limited TF AP2, Vent2, and FoxI1a. This indicated that all TFs are required to establish the expression of Six1 and Eya1 in the pre-placodal area. After exploring all significant early ectodermal TFs, together with the results of some previous studies, they have generated a GRN regarding the formation of pre-placodal area and NC territories in neural borders and have proposed a new model for neural border development. Overall, the paper provides new insights into the genetic and regulatory mechanisms that underlie the formation of pre-placodal ectoderm in *Xenopus* and may help to inform future research on this topic. Another limitation of the MO is that there is no information about direct or indirect TF-target relationships.

The direct and indirect TF-target interaction might be available using publicly available databases e.g., RcisTarget (Aibar et al., 2017). RcisTarget determines TF binding motifs that are over-represented in a certain gene list. It picks DNA motifs that are greatly over-represented in the region of the transcription start site of genes in the gene set. This is achieved using a database that contains genome-wide rankings of species for each motive. Motives that are then annotated for TF and those that have a high enrichment index are kept. Finally, for each motif and set of genes, it predicts candidate genes and consequently defines direct links TF-target. The limitation of the approach is that current TF motifs databases are available only for human, mice, and drosophila and it heavily relies on particular cellular context. Therefore, in order to do this analysis for other model organisms, the scientist will most likely need to update the current evolutionarily closest database with data obtained from e.g., the species-specific Chromatin Immunoprecipitation (ChIP-seq) analysis. Such experimental methods together with sequencing can determine direct targets of transcriptional regulators. Yet, ChIP-seq requires optimization for each individual TF and the use of antibodies that can identify the native protein or a labeled version of it. This can be a technical challenge, particularly for TFs where the label or tag interferes with function, for species that are not easily transformable, or for tissues that are limited in availability (Park, 2009). Since the global transcript levels are

comparatively easy to measure in most species and tissues, several approaches have been created to identify relationships between regulators and their targets by examining the changes in transcription levels across many samples (e.g., Bonneau et al., 2006; Margolin et al., 2006). These approaches provide a first approximation of regulatory interactions that are used to direct further experimental validation. The hypothesis of these approaches is that the regulatory association between a TF and targets can be determined from a similarity between the RNA levels of the regulator TF and targets. With the adequate variation in expression, the targets of a given TF can be predicted based on correlated changes in expression. Initial strategies developed to do this focused on the correlation between TF and targets, considering that activators are positively associated and repressors are negatively associated with their targets' expression levels (except that one TF can be activator or repressor according to different cell contexts).

Additions to this simple idea have contained pre-clustering of the data, modifying regression analysis, including training classifier models, and integrating prior biological knowledge or extra experimental data. Each of them has enhanced the capacity to determine relationships between TF and targets, even in sparse/noisy data (Huynh-Thu et al., 2010; Marbach et al., 2012; Margolin et al., 2006). In 2010, the DREAM5 challenge assessed the capacity of different methods to identify GRNs from gene expression data (Marbach et al., 2012). One of the top-performing strategies was GENIE3 (Huynh-Thu et al., 2010). This approach uses the machine-learning abilities of random forest to specify targets for specified TFs (Breiman, 2001). Other successfully implemented methods include Inferelator, ARACNE, CSI, and SVM (Bonneau et al., 2006; Margolin et al., 2006; Penfold et al., 2012; Qian et al., 2003). Typical to these methods is the use of transcript levels to assess the association between a TF and its targets. Experiments conducted in time series can produce more kinetic knowledge useful for associating TFs and potential targets. Many methods were developed that take benefit from the more information obtainable from time series data as reviewed in Bar-Joseph et al., 2012 (Bar-Joseph et al., 2012). Yet, the steady-state transcript level as estimated by most high-throughput transcriptional assays such as RNA-Seq is a measure of both: transcriptional activity and mRNA stability. Thus, the association between expression levels alone may not provide a direct estimation of transcriptional regulation as it can be confounded by the RNA stability of the target. Additionally confusing the identification of TF relationships is the fact that one gene can be controlled by different TFs in response to different triggers. Also, the

epigenetic regulations that have delayed action (poised enhancers) are ignored in this approach e.g. JmjD2a (Strobl-Mazzulla et al., 2010).

Single-cell expression data is particularly advantageous for computing GRN since, unlike bulk RNA data, it does not confuse bio-signals by averaging all cells in one sample. However, this data presents several challenges, including significant cellular heterogeneity, low density, differences in sequencing depth among cells, and the influence of the cell cycle. These characteristics make it difficult to accurately analyze and interpret the data. Nevertheless, the development of GRNs inference methods has prompted the development of comprehensive comparative frameworks. Several unsupervised/self-supervised approaches have been suggested to predict single-cell GRN. SCODE (Matsumoto et al., 2017) uses standard differential equations to rebuild GRN. This approach utilizes linear depends on a specific relational expression which can be evaluated using linear regression. GENIE3 (Huynh-Thu et al., 2010) or GRNBoost2 (Moerman et al., 2019) which is the first step of SCENIC (Aibar et al., 2017; Van de Sande et al., 2020), use tree-based regressions. This method involves fitting a tree-shaped model to the data, with the regulators from the provided list serving as the branches. The significance of each gene in the model is used to determine the weight of its interactions with the target gene's pattern. By combining the weighted relations for all of the genes, they are able to construct regulatory networks (Figure 6). PIDC utilizes ideas from information theory: for every pair of genes, given a third gene, the method splits the pairwise mutual information between the first and second genes into a duplicative and distinctive component. It calculates the proportion between the distinctive component and the mutual information. The aggregate of this proportion over all genes is the unique contribution between two genes. This approach then operates per-gene thresholds to determine the most significant relations for each gene. Another algorithm, DeepSEM (Shu et al., 2021) presents a structural equation model with a β-VAE framework to predict the regulatory connections between genes in a GRN. A recent test of twelve GRN methods demonstrated that the algorithms struggled to predict real GRNs and suggested that poor performance was due to insufficient resolution in the scRNA-seq data (Pratapa et al., 2020).

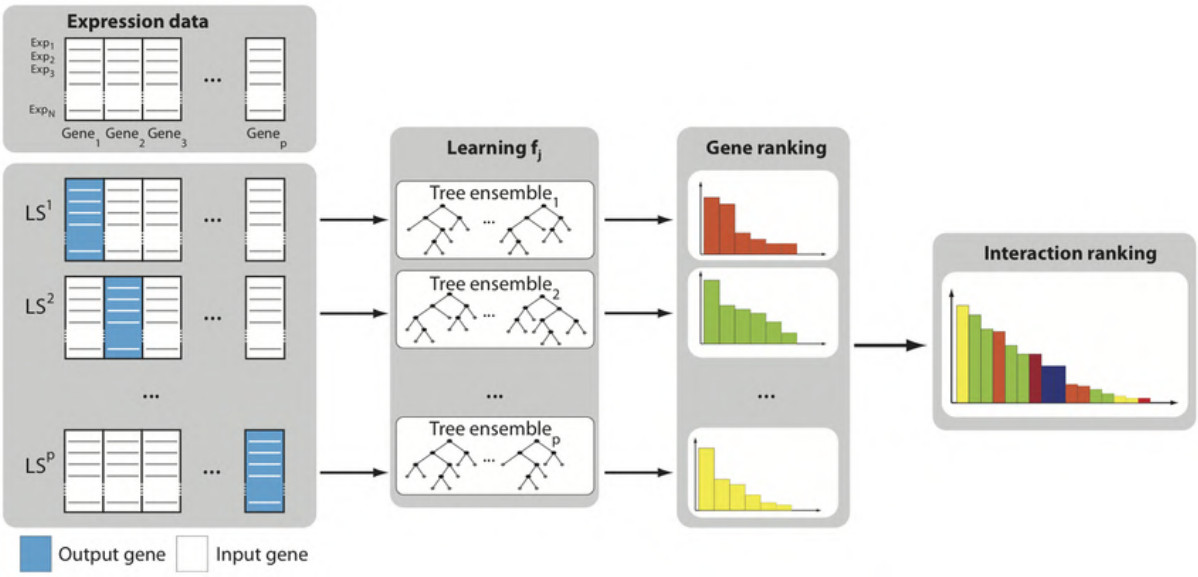



**Figure 6. GENIE3/GRNboost2 algorithm.** A model is created for each gene by using expression levels of all other genes as input values and expression levels of that specific gene (G) as the output values. From these models, a function is learned and a ranking is calculated for all genes, with the exception of gene K. The gene rankings are then combined to determine the overall ranking of all regulatory links (Moerman et al., 2019).

The process of generating gene regulatory networks typically involves three main steps. First, the raw data is preprocessed in order to make it more amenable to analysis. This can include techniques like variational autoencoders, which convert the data into a more manageable form. Next, relevant features are extracted from the preprocessed data in order to identify patterns and trends. Finally, these patterns are used to build a generalized model of gene regulation, which can be used to make predictions or gain further insight into the underlying biological processes (Eraslan et al., 2019). Computing the underlying pattern of gene regulation is basically searching for the right GRN among all possible GRNs that best fit the real data (e.g., using random forest). Although substantial advancement has been made, deriving a network of regulatory interactions between genes still stays a challenge. For scRNA-seq data, GRN reconstruction approaches on bulk RNA data are not directly applicable. As the biological value of the sample changes from the average of several cells in the bulk data to the value of one cell, the shape of the gene expression data also changes. For now, since the approaches developed for single-cell transcriptomics usually demand a large number of time points for GRN generation, it is usually appropriate for a small number of genes. Adding numerous genes to a network generation may demand the development of algorithm considering many additional regulatory relations between added genes: as the number of genes extends, the number of edges and the need for input data can increase dramatically.

Another approach which can be applied for GRN generation, scFates, is a computational tool that is developed to predict cell fate in single-cell RNA-sequencing (scRNA-seq) data. It is based on the idea that the genes that are expressed in a cell at a given time can provide clues about the cell's developmental potential and its potential to differentiate into different cell types. scFates uses machine learning algorithms to analyze scRNA-seq data and predict the cell fate of individual cells based on their gene expression patterns. It can be used to study the developmental potential of cells in different tissues and at different stages of development, and to identify the genes and pathways that are involved in the regulation of cell fate (Faure et al., 2023).

The development of transcriptome analysis has facilitated genome-wide studies of transcription patterns in early embryos. Several recent studies have presented transcriptomes either using whole embryos or with more defined spatial regions by combining embryo sections and spatial reconstruction (Blitz et al., 2017; Plouhinec et al., 2017). However, given the heterogeneous nature of tissues, the bulk tissue networks mostly describe the average activity of all cell populations, in which the most common cell types dominate. Therefore, scRNA-seq

is a more convenient source for obtaining cell type-specific GRNs. Recently, a large number of works have been published in the field of predicting GRN during the development of an embryo (Briggs et al., 2018; Kotov et al., 2022; Soldatov et al., 2019; Wagner et al., 2018; Williams et al., 2019). In one of them Briggs et al. profiled 137000 single-cell transcriptomes of Xenopus tropicalis in ten time points from stage 8 to stage 22 (from transcription start to early organogenesis). The main objective of the work was to identify new cell types at different stages of development of the frog embryo, as well as visualize and explain the transitions between one cell type to another. To do this, the researchers invited a large number of specialists in the field of developmental frog biology to annotate all the cell types. As a result, over 10 time points, each cell was connected to a similar cell from a different time point, thereby obtaining developmental transitions. In Xenopus, it was also observed that about 50% of transcription factors were reused, indicating a degree of conservatism in their usage. This work is interesting due to the high number of cells in the dataset, which can be used as additional data in predicting gene regulatory networks since the average depth and the number of genes and transcripts per cell in the data are relatively insufficient.

Another article by Soldatov et al. (Soldatov et al., 2019) focused on the later stages of mouse embryonic development (E8.5-E10.5 stages) and extracted Wnt1 and Sox10 positive neural crest cells. Using the SmartSeq2 protocol they managed to get a deep coverage (7000+ genes per cell) but only a few cells (around 200). They used the PAGODA method (Fan et al., 2016) to cluster cells. RNA velocity was applied to get the transcriptional dynamics of cells (La Manno et al., 2018). Two distinct subpopulations in pre-migratory neural crest cells were identified. The earliest population before EMT included cells that have not yet begun to delaminate from the neural tube. It is marked by the peak expression of neural plate border specifiers previously associated with NC, such as Zic1, Msx1. The second subpopulation is marked by activation of the key EMT gene Snai1 and the absence of Atoh1 and is accompanied by sequential temporary activation of a set of genes, including Dlx5, Pak3, Pdgfra, and Hapln. The fate determination of neural crest cells involved a series of sequential binary decisions involving the coactivation and gradual commitment to specific programs, and is achieved through the synchronization of relevant programs and repression of competing ones.

Another work is the profiling of transcriptional and chromatin dynamics of specifically extracted neural crest cells. Encouraged by the lack of unbiased representation of vertebrate neural crest GRN, Williams et al. developed an integrative approach that allows a depiction and interrogation of NC GRN. First, they did transcriptional profiling in NC cells. The main

difference from previous work is a more precise selection of NC (Citrine-positive) but only 300 cells for 5-6ss (somite-stage or HH9) and 600 cells for 8-10ss (HH10) for bulk RNA-seq and around 3000 cells for scRNA-seq. Transcriptional profiling included a comprehensive combination of SmartSeq2, 10X (only for 6-7ss), and bulk RNA sequencing. For bulk RNA-seq, they applied a widely used WGCNA (Langfelder and Horvath, 2008). This method allows to construction the gene networks and detect gene modules or hubs based on hierarchical clustering in a matrix of co-expressed genes. Using this method, they detected previously known clusters with some divergence. To further improve the analysis, Williams et al. used ATAC-seq and Capture-C (Davies et al., 2016) to identify the NC-specific topologically associating domain and five enhancers within this domain. Comparing this NC-specific and non-NC epigenetic information revealed chromatin accessibility changes in different cell types reflect such changes in tissue-specific enhancer activity. Additionally, they performed ChIP-seq in neural crest cells for one of the super-enhancers (SE) H3K27Ac, a cluster of elements within a topologically associating domain. Using the rank ordering they got a sorted list of super-enhancers (SEs). Focused on the Sox10 SE locus (one of the top in the list of SEs and an important gene in NC development) researchers identified a cluster of elements near the Sox10 domain. All of the detected enhancers of the Sox10 SE acted in conjunction and were critical for Sox10 expression during NC EMT. Together with results of functional contribution of enhancers using CRISPR-mediated knockout, ChIP-seq results helped to understand that some of the enhancers (enh-99) might be essential for initiation, but not for maintenance of neural crest program. Furthermore, some stem cell regulatory elements (enh-89) can be reused in neural crest cells, and some (enh-84, enh-85, enh-87) are essential for NC robustness. The early epigenomic heterogeneity was also observed on a transcriptional level which is confirmed by SmartSeq2 and 10X Chromium scRNA profiles (137 and 3091 cells). Finally, to generate NC GRN, Williams et al. integrated scRNA co-expression data and direct TF inputs (obtained through enhancers motifs search). After integration, they focused only on k-CI3 and k-CI1-related GRNs. As a result, they revealed that k-CI3 enhancers manage the early establishment of the neural progenitor state between NC and non-NC cells.

Therefore, in order to gain a more comprehensive understanding of gene regulatory networks, it is necessary to use a variety of analytical techniques and data sources (e.g., bulk RNA-seq, scRNA-seq, ChiP-seq, Capture-C, ATAC-seq, scATAC-seq, and Perturbation-seq). Moreover, the results can be improved using the same data, for example, by examining enhancers for other genes (in the article Sox10 SE is only explored). Selecting the most accurate

algorithms for predicting relationships between genes is also crucial, as demonstrated by recent benchmarking studies (Holland et al., 2020; Pratapa et al., 2020), where Pratapa et al. compared methods on the three different types of networks: synthetic, literature-curated Boolean models and transcriptional regulatory networks, and Holland et al. applied bulk-based methods to scRNA-seq datasets (Figure 7).


**Figure 7: GRN predictors benchmarking.** This summary table compares the properties and performance of various GRN inference algorithms. The table includes information about the methodology and input requirements of each algorithm, whether the method requires time-ordered cells, and whether the inferred edges are directed and signed (indication activating/inhibiting role). The table also provides summary results for the performance of each algorithm on various measures of stability, as well as the running time and memory usage of each method. The results in the "Pseudotime" column are only relevant for the seven methods that require this input (Pratapa et al., 2020).

## 1.5 Research question and PhD project

Understanding the fundamental biology of the neural crest is important because dysregulation of the gene regulatory networks that control neural crest development can lead to various disorders and diseases, including craniofacial abnormalities and neurocristopathies. Our enhanced knowledge of the NC gene regulatory network (NC GRN) can be used to improve the understanding of human diseases. These regulatory networks are biological computational modules that carry out decision-making processes on a molecular level.

There is still much that is not known about how the fate of NC cells is initiated during the patterning of the dorsal ectoderm. This includes how NC acquire their regional diversification during the early pre-migratory stages. Also, it is unknown how the NC is heterogeneous at early pre-migratory stages, and what are the key genes that trigger the heterogeneity. So further research is needed to identify and characterize these key genes and their roles in NC heterogeneity and cell fate decisions. Additionally, the origins and evolution of the neural crest from an NC-like state are not well understood.

Thus, my PhD project primary concern is the focusing on the understanding of the gene regulatory networks that control the development of particular subpopulations of neural crest cells. Our focus is on the gastrula pre-migratory stages of the NC, during which we have studied NC fate in comparison to placodal, neural plate, and non-neural ectoderm fates, as well as different NC fates including various vagal and cranial subpopulations. Our goal is to understand how NC cells differentiate into these different fates and how this process, known as branching, is regulated at the molecular level. Specifically, we are interested in identifying and characterizing the key genes that are involved in this process and determining how they contribute to the development of different NC subpopulations and fates. In addition to our main focus, we are also interested in expanding upon the current knowledge about known key regulators such as Pax3 and Tfap2e and specific miRNAs, which have been shown to be important for the NC.

To do this, I have used a variety of available approaches, including computational predictions and results from experimental approaches generated in the team. Therefore, we attempted to achieve the following goals:

- Develop a comprehensive strategy for predicting and validating gene regulatory networks

- Implement this approach on the neural crest at early and poorly studied pre-migratory stages and provide the tree of the pre-migratory NC cells
- Develop an approach to study the cell type molecular profile transition. Main ambition would be providing the tool that will predict the cell type evolution.

# 2. Results

## 2.1 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**, Mansour Alkobtawi, Subham Seal, Vincent Kappès, Sofia Medina Ruiz, Hugo Arbès, Richard Harland, Leonid Peshkin, Anne H. Monsoro-Burq

doi: https://doi.org/10.1101/2022.03.23.485460

### 2.1.1 Overview of key findings

Despite decades of study using various cellular and molecular approaches at the tissue level or on small groups of cells, the induction of neural crest cell fate during dorsal ectoderm patterning with respect to adjacent cell types (placodes, nonneural ectoderm, and neural plate) remains unclear. Furthermore, the mechanisms underlying the regional diversification of neural crest cells during the early pre-migratory stage are not well understood. Previous analyses of single-cell transcriptomes have largely focused on subsets of neural crest cells after their emigration from the neural tube (Soldatov et al., 2019; Williams et al., 2019) or pre-migratory NC but on limited anterior-posterior domain or limited developmental period (Lukoseviciute et al., 2018; Zalc et al., 2021). The crucial steps of pre-migratory neural crest diversification (along the whole cranial and vagal areas) have not been addressed with those approaches yet. In this work, we have used a combination of scRNA-seq predictions and experimental validation with the aim to greatly expanded our knowledge of the early stages of neural crest development. So, the main questions were:

What are the key regulatory genes that control the development and differentiation of neural crest cells?

How do the gene regulatory networks control neural crest development change as the cells progress through different stages of development from induction to migration?

What is the relationship between the neural crest cells and the other cell types that arise from the neural border zone (e.g., placodes, non-neural ectoderm, and neural tube)?

### 2.1.2 Temporal characterization of pan-NC signature

In the first part of our study, we used libraries from a previous screen, re-sequenced in order to improve quality of the data (genes/counts per cell). Then, from the whole embryo dataset, we generated a single-cell transcriptomic analysis of 6135 neural crest (NC) cells for early gastrulation to early migration stages (Figure 1A of the paper 1). First, we sought to identify pan-NC markers based on their NC-specific expression and low variation across the NC population. We identified *tfap2b, c9, c3*, and sox8 as highly expressed throughout NC during neurulation, closely followed by *snai2* and *sox10,* with specific pan-NC expression (Figures 1B-C of the paper 1). Stage-specific expression variations suggested that *snai2* is the best marker for labeling early NC cells at stages 13-16, followed by *tfap2b* from stage 14 to 20, and *c3, c9*, and *sox8* between stages 14 to 18. These findings provide sc-transcriptomics insights into the timing of *Xenopus* NC specification and regionalization during embryonic development, confirming known NC markers e.g., *tfap2b/sox9* and describing poorly described previously e.g., *c9/c3*.

### 2.1.3 In silico validation of NC cells selection

We sought to confirm the accuracy of the dataset of neural crest cells by creating a classifier that could accurately identify these cells using their gene expression signature. Thus, we validated the NC selection method by designing a LightGBM (Ke et al., 2017) NC classifier and testing it on another vertebrate organism, 14 hpf zebrafish (Wagner et al., 2018), to evaluate its accuracy for NC annotation. The result of 0.95 for the AUC score and 0.66 for F1 score was remarkable since 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; Figure 1 figure supplement 1 of the paper). This result confirmed the accuracy and robustness of our NC cells selection criteria. This part of work was also used as basis for a development of the computational tool scEvoNet (Kotov et al., 2022; see the last chapter of the Results section)

### 2.1.4 High heterogeneity in NC cells during NC induction

To identify subclusters within the neural crest cells, we determined the optimal number of clusters by incrementally increasing their number and evaluating the presence of biologically meaningful gene expression patterns. We uncovered the highly diverse transcriptomes

underlying pre-migratory NC cell biology along a cranial to vagal segment of the body axis. We defined 16 clearly distinct states with specific gene signatures, compared to 8 states previously described in (Briggs et al., 2018; Figure 2A of the paper 1). Ten of these clusters displayed an explicit mesenchymal NC signature expressing various levels of *itga4/integrin-a4*, *vim/vimentin*, and *fn1/fibronectin* (Figure 2 figure supplement 1 of the paper 1). Additionally, we uncovered two previously undescribed clusters: a muscle-like NC subpopulation, expressing actin and myosin-like genes (*actnc1+, myl1+* cells), and 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 (Figure 2 figure supplement 2 of the paper 1). Unexpectedly, as early as at gastrula stages 12-13 we also detected three NC cell states clearly distinct from the main early NC state: one vagal *hnf1b+* cluster, two cranial *cyp26c1+*, and *rpe65+* clusters. Thus, the study identified 16 distinct states of neural crest cells with specific gene signatures, including two previously undescribed clusters, and found that some neural crest populations at gastrula stages 12-13 are distinct from the main early neural crest state.

### 2.1.5 Connectome analysis

Exploring the GRN in neural crest development can provide insight into how these networks control the formation and function of neural crest cells, which are important for the development of various tissues and organs in the body. Thus, by applying the GRNBoost2 on selected NC cells we retrieved a network of over 16 thousand potential TF-targets connections with a median of 22 connections per TF (Table 1 in the paper 1). Among the most connected genes for each stage, we retrieved the previously known neural border specifiers *pax3* and *zic1* at NC induction stages, consistent with their demonstrated role at the heart of the NC-GRN (Plouhinec et al., 2014), as well as other known important for NC development genes *zic3, olig4, sox9*, and the anterior NC marker *dmbx1* (Hernandez-Lagunas et al., 2011; Nakata et al., 1997; Simoes-Costa and Bronner, 2016; Spokony et al., 2002). At pre-migratory stages, the prominent nodes included the NC specifiers *tfap2b, sox10* and *snai2*, the anterior NC markers rpe65 and alx1, and the hindbrain hox gene hoxb3. At migration stages, *tfap2e, mycn, dlx2 and egr2* (*krox20*) displayed both high expression and many connections. Thus, by analyzing the largest available dataset on neural crest development, we have constructed a comprehensive gene regulatory network that reveals previously unidentified mechanisms of regulation in neural crest development.

In order to directly validate this network experimentally, we used ChIP-seq and Morpholino knock-down experimental validation (Figure 4 in the paper 1). For knockdown experiments, previously validated antisense morpholino oligonucleotides were used to deplete Pax3, or TFAP2e transcripts (Hong et al., 2014; Monsoro-Burq et al., 2005). Each explant was sequenced individually (RNA-seq). This revealed that expression of 1333 transcripts were decreased in Pax3 morphant NB, confirming that Pax3 is essential to activate a large NB/NC gene set. We verified direct Pax3-binding targets in vivo using ChIP-seq on mid-neurula stage embryos and identified 657 potential targets in the whole embryo, of which 475 were expressed in the SC NC dataset. Moreover, 80 of the ChIP-seq-validated targets were predicted by GRN-Boost2 modeling including *psmd4, psen2, sp7, notch1,* and *hnf1b*. Using a similar approach, we confirmed TFAP2e as an important regulator at early EMT stages. Among 848 targets of TFAP2e predicted from scRNA-seq, 99 showed changed expression after TFAP2e knockdown in NC in vivo (e.g., *tfap2b, sox10, sncaip*). Moreover, using ChIP-seq for TFAP2e, we identified 642 targets expressed in the NC dataset, among 805 targets for the whole embryo, including top-scored *rmb20, pim1, arl5b and tfap2a*. In both Pax3 and TFAP2e cases, we find that the three approaches result in partially overlapping gene target lists due to their use of different parameters (stage-wise, expression-wise, etc.). The validation results using ChIP-seq and Morpholino knockdown experiments confirm the importance of Pax3 and TFAP2e in regulating gene expression during neural crest development, and expand our knowledge about the actual target genes of these transcription factors. The use of multiple approaches also reveals overlapping target genes, highlighting the utility of this approach.

### 2.1.6 NC branching analysis

Next, we applied branching analysis. Branching analysis of scRNA-seq data involves identifying patterns of gene expression in individual cells that correspond to specific route, and is an important tool for understanding the underlying regulatory mechanisms of cellular differentiation and development. This approach is different from traditional correlation expression analysis, which examines the relationship between gene expression levels across a population of cells, because it allows for a more detailed and nuanced understanding of gene expression patterns within individual cells in the context of time. Through branching analysis, we uncovered gene programs governing the bifurcation, consisting of ‘early’ genes activated before bifurcation and ‘late’ genes with continued expression in each branch (late): *nrp2, alk, rnd1, adam19* (early) and *shisa3*, *frdm6* (late) for the cranial branch, and *mafb*, *klb*, *mdk* (early)

and *prdm1*, *cfl*, *hoxd3* (late) for the vagal branch. For the cardiac NC cluster, early-enriched transcripts were *mafb*, *mycn*, *prdm1*, *nolc1* and *eef1d*, late ones were *egr2*, *hoxd3*, *epha4* and *abtb2*. For the ENSp cluster, early markers were *olig4* and *fbn2*. In order to test experimentally if 'early' factors affected the expression of 'late' genes, we examined which late ENSp genes may be targets of the early TF Pax3, through ChIP-seq and depletion analyses. Some genes were bound by Pax3 (*cfb*, *bmp5*), while others showed increased (*krt8*) or decreased (*cldn1*, *hoxb6*, *hoxa7*) expression without evidence for Pax3 binding. Similar analysis on cranial NC clusters showed that branch 10-specific genes were *rpe65*, *dmbx1*, *rgr*, *lmx1a*, and *zfhx4*, while late gene programs included *alx1* and *bmper*. Interestingly, *pax3* was expressed early in the cranial bifurcation from 7 towards 10 and 11; expressed in cluster 7 cells, *pax3* was specifically enriched in *rpe65*$^{+}$ cluster 10. The Pax3 ChIP-seq and MO datasets revealed that rpe65+ branch-specific transcripts *comt*, *slc23a2* and *rpe65* were decreased after pax3 depletion while others, *bmper*, *f10, sema3a*, *dmbx1*, *slc16a12* and *kif26a*, were bound by Pax3 in vivo. TFAP2e depletion and ChIP-seq validated that TFAP2e depletion reduced the expression of nine hox-dlx2+branch-specific genes (e.g., early gene *mef2c*, and late genes *dlx2*, *mmp14*, *vim*), and that TFAP2e bound four other genes in cluster 11 signature (*c9*, *vim*, *mmp14* and *mycn*). Here, we identified specific genes and transcription factors that play a role in the bifurcation process during vagal and cranial NC development, and found that certain "early" transcription factors can affect the expression of "late" genes in the process. This contributes to a deeper understanding of the genetic mechanisms underlying bifurcations during NC development.

### 2.1.7 The NB zone signature

The NB zone is a region of cells in the developing neural tube that is characterized by the expression of certain transcripts, including Pax3. However, the NB zone does not have a single cell RNA signature. Here we identified a detailed signature for the *tfap2a*$^{+}$*zic1*$^{+}$cells (NB zone). Several transcripts were enriched from stage 11: *tfap2c*, *pax3*, *sox9*, *hes1*, *gmnn* and *myc*. We examined whether Pax3, previously shown to appear as early as stage 10.5 in vivo (de Crozé et al., 2011), triggers the expression of NB zone signature genes in vivo; in Pax3 morphant NB transcriptome, we found decreased expression of the other NB genes *sox9*, *axin2*, *zic3* and *zic1*, while early NE and NNE marker *lhx5.1* was increased, confirming its key role as an activator in early NB. Thus, the role of Pax3 in the NB zone is to activate the expression of genes that are important for the development of the NB zone.

### 2.1.8 Ectoderm branching analysis

The development of neural crest and placode cells is a complex process that is not fully understood. Two models have been proposed to explain this process: the "binary competence" model, in which NC and PC cells are formed independently on opposite sides of the neural border, and the "neural border" (NB) model, which proposes that both NC and PC progenitors are generated from a bi-potent neural border, and the final fates of these cells are determined by specific markers at later stages of development. Here, we explored three different developmental routes that can lead to the formation of NC and PC cells, and examined the genes that are involved in these processes. Three developmental routes were considered: a) NE → NC; b) NB zone → PC and NC; c) NNE → PC; these possibilities are consistent with current models of NC and PC formation, the “neural border zone model” (route b) and the “neural vs non-neural” model (routes a & c). We found that distinct genes were activated to obtain the same state and some genes were activated with different expression dynamics relative to different bifurcations. For example, during the NB→NC transition, *sox9* and *c3* were activated before bifurcation suggesting that they could play a part in the fate decision network from NB progenitors. In contrast, during the NE→NC gene program, *sox9* and *c3* were late genes while *foxd3* and *zic1* were expressed early. Last, NB zone-specific gene *pax3* was expressed prior to bifurcation in the NB→NC gene program and activated expression of late NC branch markers *sox9*, *sox8*, *zic1*, *pcdh8*, and *c3*. Moreover, the Ectoderm connectome described NC genes connected to the rest of the network through Pax3 and Sox9, suggesting that Sox9 might play a yet undescribed function downstream of Pax3 in NC induction and upstream of the other late NC-branch markers. We confirmed that Sox9 enhances NC induction downstream of Pax3 and Zic1 by combining Sox9 depletion or gain-of-function in the induced-neural crest assay (iNC is Pax3/Zic1-based NC induction from pluripotent ectoderm cells).

### 2.1.9 Conclusions

In this work, we exploit the resolution of high-density single-cell transcriptomes collected from 8 frog developmental stages to unravel the emergence of the neural crest lineage from the ectoderm during gastrulation, followed by diversification of neural crest progenitors during neurulation and upon EMT. Modeling gene transcription dynamics around each cell state allows the inference of the underlying molecular networks. We selected several important nodes for large-scale and in vivo experimental validation. This study highlights the previously unknown temporal sequence of states in pre-migratory neural crest development. Firstly, we

characterize neural crest activation either from a transient neural border state, or from a neural plate state during mid-gastrulation, and suggest a model that reconciles current debates upon multiple possible routes leading from immature ectoderm to neural crest and placodes. Secondly, we delineate the early and later neural crest transcriptome trajectories during neurulation and define key regulators of branching, leading to eight transitional states and eight early migration states.

Further research could be conducted to validate the findings of this study through large-scale and in vivo experiments, using the identified key regulatory nodes as a starting point. This could help to confirm the importance of these nodes in the development of the neural crest lineage and provide additional insights into the underlying molecular mechanisms. It would be interesting to extend this research to other organisms, such as mice or humans, to determine whether the developmental pathways and regulatory mechanisms identified in this study are conserved across species.

Publication 1

# 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**, Mansour Alkobtawi, Subham Seal, Vincent Kappès, Sofia Medina Ruiz, Hugo Arbès, Richard Harland, Leonid Peshkin, Anne H. Monsoro-Burq

doi: https://doi.org/10.1101/2022.03.23.485460

# 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 in the frog *Xenopus tropicalis*. During gastrulation, a transient neural border zone state precedes the choice between neural crest and placodes which includes multiple and converging gene programs. Transcription factor connectome and bifurcation analyses demonstrate the early emergence of neural crest fates at the 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, 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 (Eames et al., 2020; Plouhinec et al., 2017). In addition to the NC, the NB territory also gives rise to posterior placodes, non-neural ectoderm and the dorsal part of the neural tube (Steventon and Mayor, 2012; Streit and Stern, 1999). 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) (Bhattacharya et al., 2020; Figueiredo et al., 2017). In addition to this pan-NC program, several regional molecular modules are activated along the anterior-posterior body axis and define subpopulations with specific potential (Ling and Sauka-Spengler, 2019; Tang et al., 2021). 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.

NC biology has been scrutinized during development and evolution, leading to the elucidation of elaborate gene regulatory networks (GRNs) during the last decade (Monsoro-Burq et al., 2005; Simoes-Costa and Bronner, 2016). These networks, however, remain incomplete and do not account for most of the defects observed in human neurocristopathies (Medina-Cuadra and Monsoro-Burq, 2021). 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. Most of the recent SC studies on NC cells have mainly explored NC after emigration (Artinger and Monsoro-Burq, 2021; Supplementary File 1 - Table S1). 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 (Ling and Sauka-Spengler, 2019; Zalc et al., 2021). Earlier on,

formation of the NB territory has been defined by expression of a few genes during gastrulation (e.g. *pax3* and *pax7)* (Basch et al., 2006; Monsoro-Burq et al., 2005; Plouhinec et al., 2017), however 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 (Groves and LaBonne, 2014; Maharana and Schlosser, 2018; Steventon and Mayor, 2012). Furthermore, the timing of lineage decisions in the pre-migratory NC along the anterior-posterior axis, the maintenance of a multipotent NC subpopulation, 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. During neurulation, we define several new premigratory NC subpopulations and highlight the transcriptomic trajectories between these early states and eight NC subpopulations emigrating from anterior to vagal levels of the body axis. Interestingly, we find that distinct signatures for prospective NC fates emerge much earlier than previously anticipated, that NC state diversity is maintained upon EMT and that further diversification occurs at the onset of migration. During gastrulation, we explore neural border development and its specification into the NC and placodes. At each stage, we propose a temporal sequence of molecular events underlying these successive transcriptomic states and identify key candidate transcription factors involved in the branching between states. Importantly, we validate several regulatory predictions using transcriptomes and *in vivo* approaches. Last, we propose a model of "dual convergence" for parallel transcriptomic routes driving neural border specification. We therefore provide an extensive gene regulatory network describing the emergence of the neural crest from the ectoderm of vertebrate embryos.

## Results

### Defining the diversity of premigratory neural crest states.

Using deeper re-sequencing of single cell (SC) series from whole *X. tropicalis* embryos (Briggs et al., 2018), taken at 8 consecutive developmental stages, followed by updated genome annotation and alignment, we have scrutinized 6135 neural crest (NC) cells from early gastrulation to early migration stages. The larger cell number of our new dataset allowed greater 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).

# Kotov et al., Figure 1

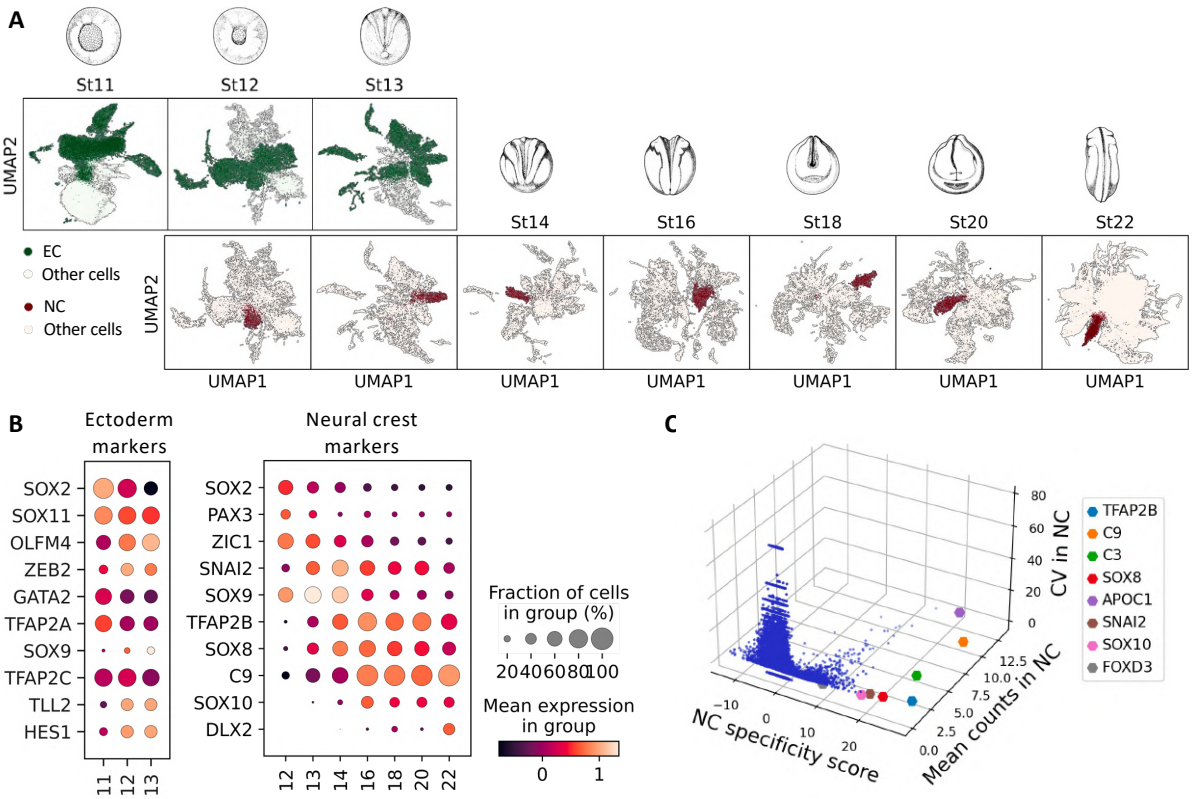



**Figure 1. Cell selection for neural crest, ectoderm and neural border. (A)** Ectoderm (EC, stages 11-13, green) and NC cells (stages 12-22, brown) were selected from a whole embryo SC transcriptome dataset of 177250 cells. **(B)** Dotplots for well-referenced gene expressions used to identify EC and NC at each stage. Dot size represents the number of cells expressing the gene, color represents the average expression level. Neural border was defined as stage 11-13 *tfap2a*$^{+}$, *zic1*$^{+}$ cells. **(C)** 3D scatter-plot of NC score specificity (z-scores), mean gene expression levels (counts) and coefficient of variation (CV) in NC cells, defining a few highly expressed pan-NC genes during neurulation. Additional validation of NC cell selection was done using a binary classifier depicted in Figure 1 - figure supplement 1.

# Kotov et al., Figure 1 - figure supplement 1

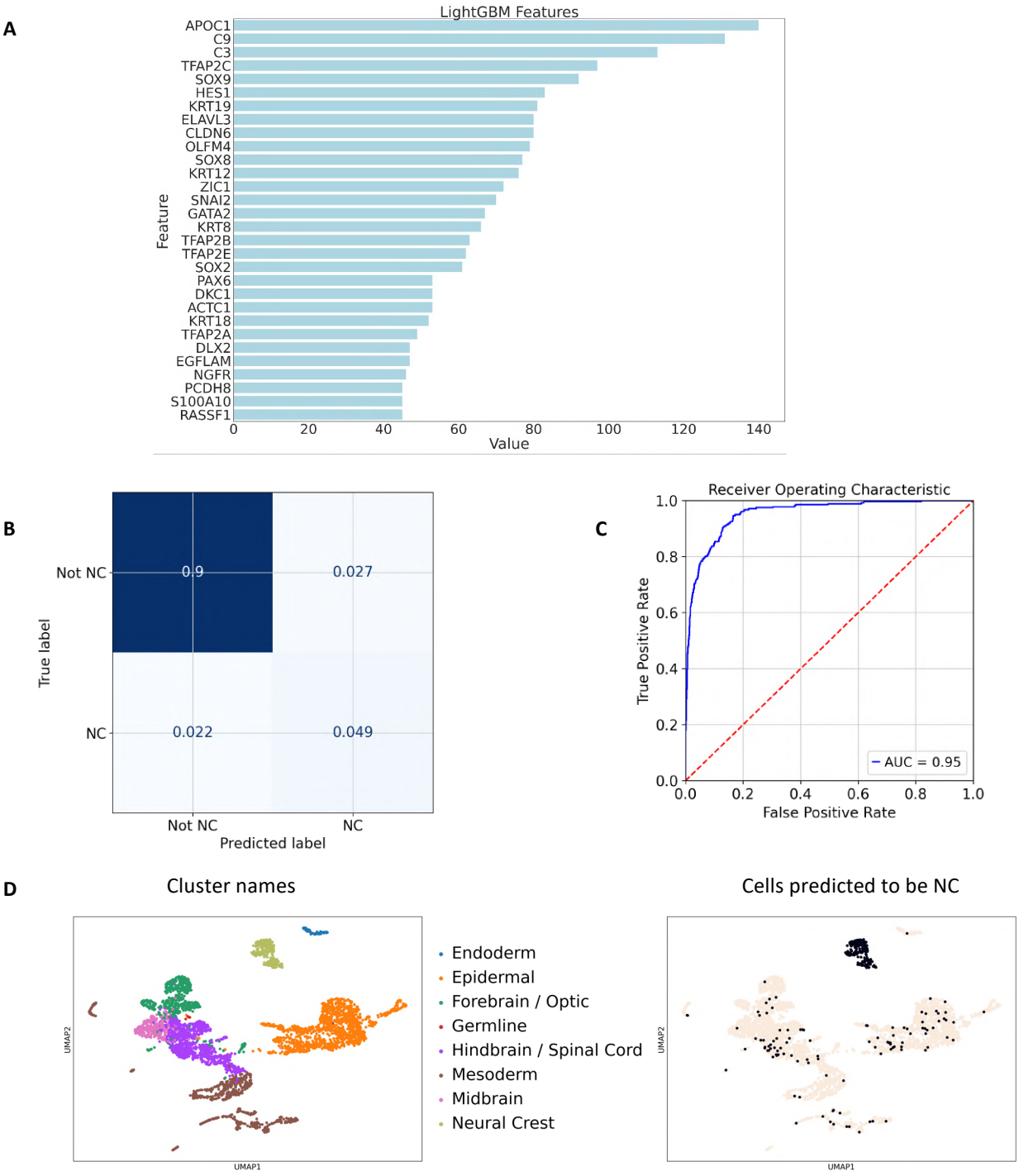



**Figure 1 - figure supplement 1**. **A NC classifier based on Xenopus NC selection retrieves NC cells in other whole embryo datasets.** A binary classifier was trained on the frog NC/non-NC cells (see methods). Using the whole embryo data matrix of gene expression, we selected the top 10% important features for NC detection. In order to validate it in a whole embryo SC Zebrafish dataset (Wagner et al., 2018), we removed genes present in frog data that were absent in the Zebrafish dataset: Using the subselected top genes, we re-built the model using frog data and obtained the following results: 0.99 accuracy and 0.90 f1 score on a test dataset. When applied to Zebrafish dataset, we obtained also very high scores: the accuracy 0.95 and f1 score 0.66 for 14 stage cells, indicating good initial NC cell selection. (A) Top features (gene) for NC classification using LightGBM model. X axis value is the relative importance of the certain gene for the NC classifier (B) Confusion matrix indicates the true/false positive and the true/false negative scores for the results obtained on the Zebrafish dataset for stage 14. (C) ROC curve for the binary NC classifier indicates high AUC score of 0.95, meaning high recovery of true positives by the model. (D) UMAP for Zebrafish stage 14 hpf cells (left) with model-predicted NC cells. The model accurately identifies the cluster of interest (NC-khaki) despite the technical and biological batch effects between Xenopus and Zebrafish datasets.

Through unsupervised Leiden clustering of the whole embryo dataset at each stage (a total of 177250 cells) (Traag et al., 2019), we identified NC clusters by expression of well-established NC genes from induction to migration (gene-supervised approach, Figure 1A-B; Supplementary File 1 - Table S2). From the annotated frog dataset, signatures of NC/non-NC cells were then used to train a classifier which faithfully detected NC cells when applied to a zebrafish whole embryo dataset. This test indicated that the initial gene-supervised analysis had retrieved *bona fide* NC cells (details in Supplementary Materials; Figure 1 – figure supplement 1). Based on their NC-specific expression and low variation across the NC populations, we found that *tfap2b, c9, c3* and *sox8* are highly expressed throughout NC during neurulation, closely followed by *snai2* and *sox10* (Figure 1C). Stage-specific expression variations suggested that the early NC is best labeled by *snai2* at stages 13-16, followed by *tfap2b* from stage 14 to 20, and *c3, c9, sox8* between stages 14 to 18. Collectively, these genes thus define a canonical "pan-NC" signature.

From late gastrulation to emigration stages, reclustering defined sixteen different NC states, compared to 8 described previously (Briggs et al., 2018) (Figure 2A, B). Partition-based graph abstraction analysis (PAGA) (Wolf et al., 2018) showed that most NC clusters are highly interconnected with stronger connectivity between early (1-4), vagal (8, 9, 12 and 13), and cranial (10, 7 and 11) clusters (Figure 2C). Each cluster's characteristics are described in detail in supplementary text, only new general features are described here (Supplementary file 1, Table S3). First, we identified 10 clusters expressing various levels of *itga4*/integrin-a4, *vim*/vimentin or *fn1*/fibronectin, an explicit mesenchymal NC signature (Figure 2 – figure supplement 1A)**.** These clusters mainly included stage 18 to 22 cells (Figure 2 – figure supplement 1B), stages when the cranialmost NC undergoes EMT and early migration. Using *hox* gene expression, we positioned each cluster along the body axis (cranialmost clusters were *hox*-negative while vagal clusters expressed a range of anterior-to-posterior *hox* genes; Figure 2D). Next, we queried whether all NC cells adopted a similar “stem-like” state upon EMT as proposed recently (Zalc et al., 2021). Instead, we observed high diversity across the mesenchymal clusters: for example, clusters 10, 13, and 15 all undergo a transition to a mesenchymal $vim^{+}$ state but 10 and 15 expressed the ectomesodermal marker *twi1*, while 13 instead highly expressed *tnc* (Figure 2B, E). Clearly, in the large dataset considered, the diversity of states was maintained as cells transitioned from pre-EMT stages to EMT and early migration. Among the mesenchymal clusters, cluster 16 was the only one enriched simultaneously for late pan-NC markers (*tfap2b, sox10, c9*) and differentiation markers, muscle genes *actc1* and *myl;* (Figure 2E, Figure 2 – figure supplement 2). This NC dataset does reveal other cranial or vagal cluster expressing determination or differentiation markers. This observation matches recent lineage tracing in

# Kotov et al., Figure 2

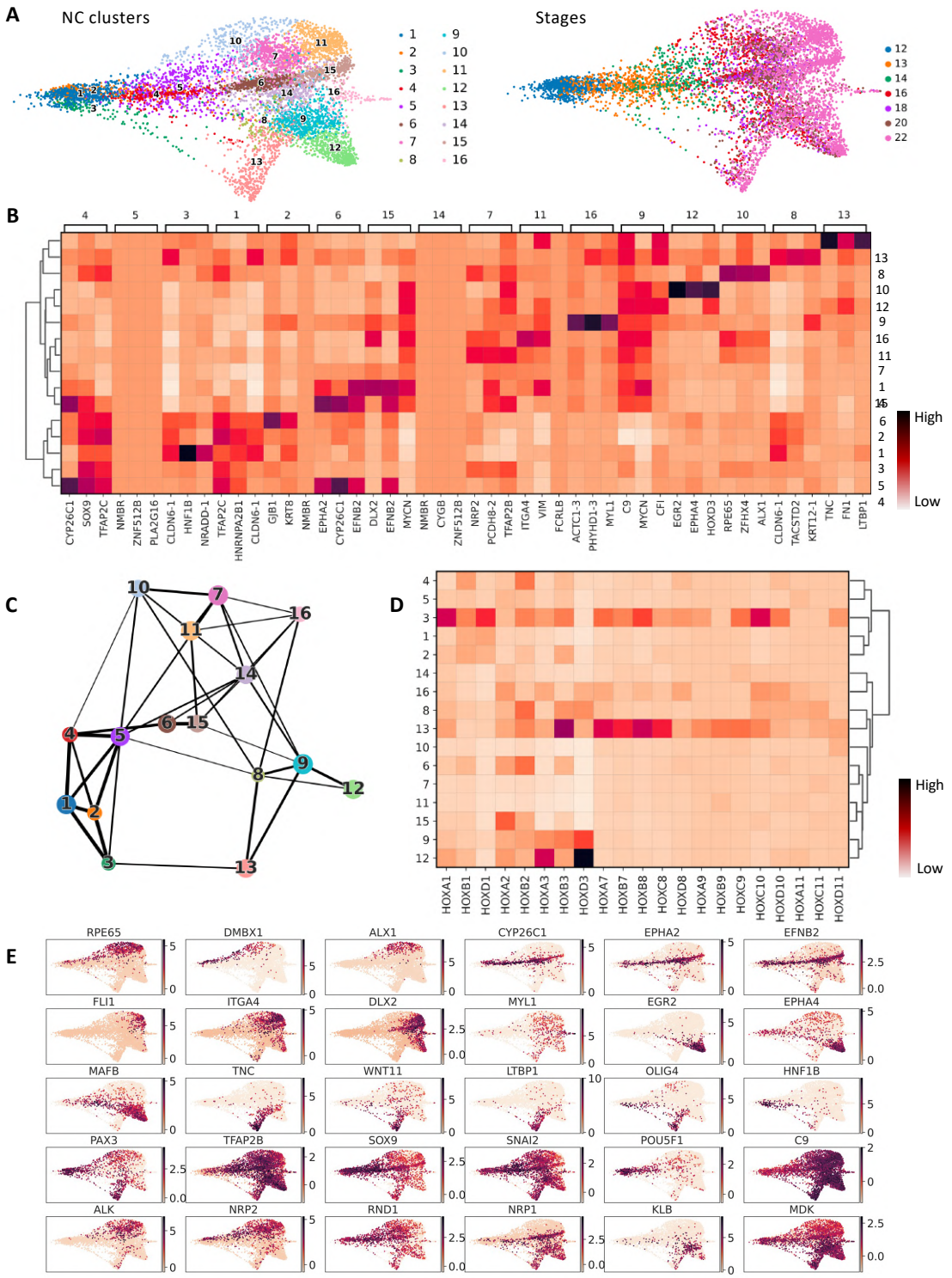



**Figure 2. Premigratory neural crest transcriptome heterogeneity. (A**) Leiden clustering revealed 16 distinct states (clusters) before and during EMT (developmental stages 12-22). 1- Early unbiased / tfap2c$^{+}$; 2- Early unbiased / gjb1$^{+}$; 3- Premigratory sacral / hnf1b$^{+}$; 4- Premigratory 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 ENSp / tnc$^{+}$; 14- EMT unbiased; 15- Migratory late R4 / efnb2$^{+}$; 16- Migratory muscle-like /myl1$^{+}$ **(B)** Top-3 enriched genes for each cluster (top and bottom lines), with their expression in the other clusters and hierarchical clustering between cluster. **(C)** PAGA estimates cluster connectivity where line thickness increases with stronger connections**. (D)** Hox gene signature of each cluster used to approximate their position along the antero-posterior body axis. **(E)** Expression of key cluster-specific genes, including *rpe65, dmbx1, alx1* (cluster 10), *cyp26c1, epha2, efnb2* (clusters 4/6/15), *fli1, itga4, dlx2* (cluster 11), *egr2, epha4, mafb* (cluster 12), *tnc, wnt11, ltbp1* (cluster 13), early *olig4, hnf1b* (cluster 3*)* and muscle-like NC specific *myl1* (cluster 16). Genes expressed broadly in NC cells define a "canonical NC" signature: early *pax3, tfap2b, sox9, snai2, and c9.* Multipotency-related genes are present mostly until mid-neurula stage *(pou5f1). Alk, nrp2, rnd1* and *nrp1, klb, mdk* are early specifiers of cranial and vagal NC respectively. Detailed cluster characteristics can be found in Supplementary File Table S3 and Text, and in Figure 2 - figure supplements 1-5.

# Kotov et al., Figure 2 - figure supplement 1

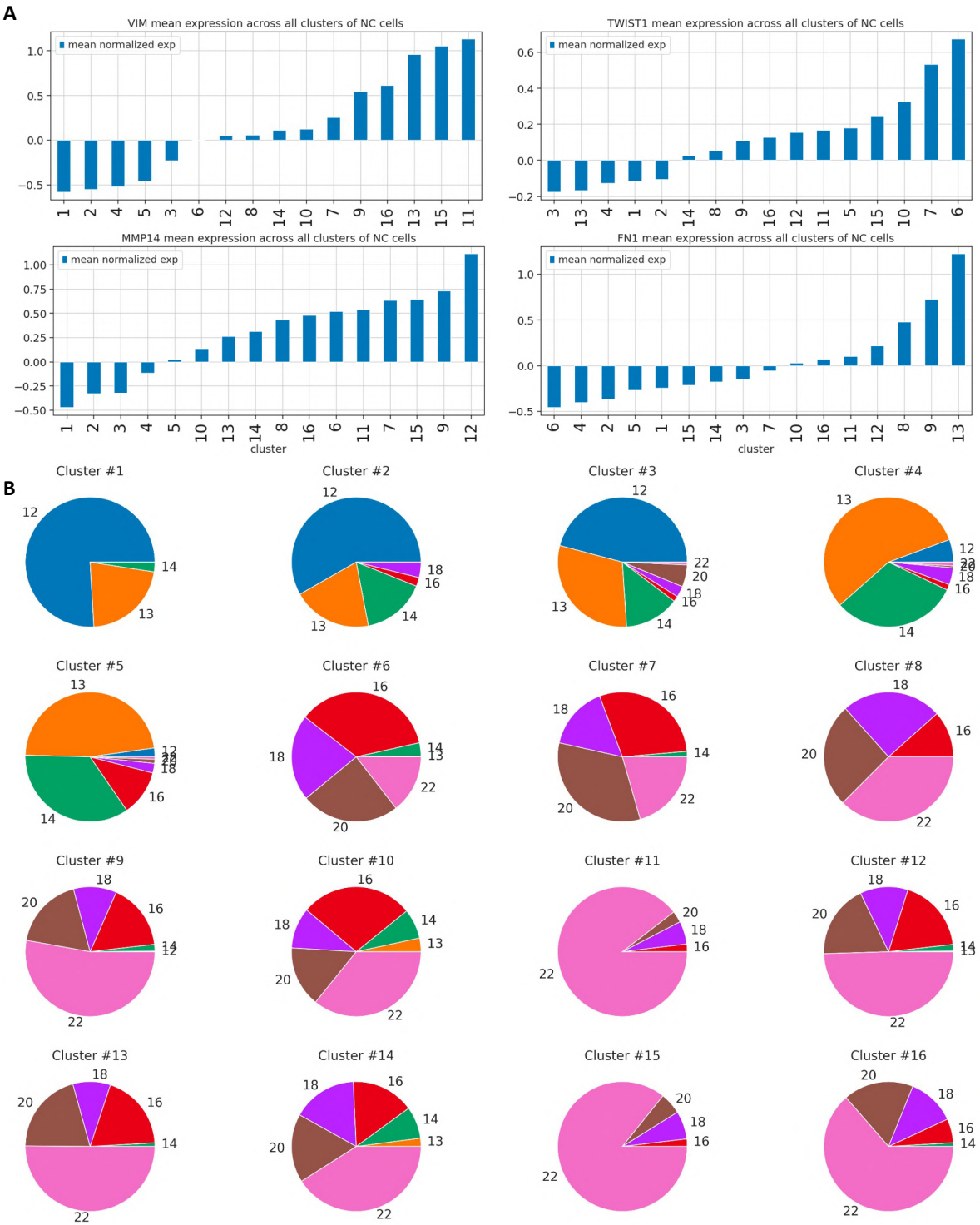



**Figure 2 - figure supplement 1. Developmental stages and expression of EMT/migration markers per cluster.**
(A) Each cluster presents various levels of expression of EMT and migration regulators. Bar plots with mean normalized expression for each cluster for vim, twist1, mmp14 and fn1. (B) Pieplots represent the proportion of cells at each stage in each cluster. Some clusters include cells of 6 successive stages, e.g. cluster 10.

## Kotov et al., Figure 2 - figure supplement 2

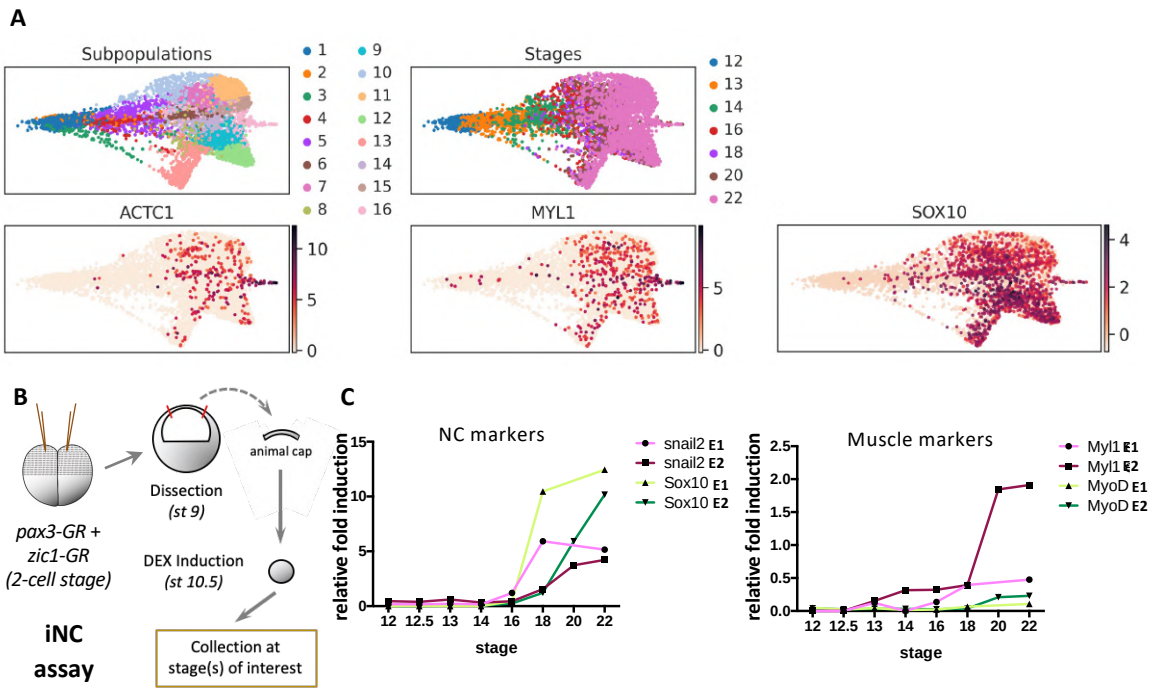



**Figure 2 – figure supplement 2. Identification of a muscle-like cluster.** (A) At early migration stage (20-22), cluster 16 is strongly enriched for myosin-like genes as well as bona fide NC markers, such as sox10. (B, C) Direct NC induction in ectoderm explants grown in vitro (iNC assay) followed by RTqPCR from induction to emigration stages, also identifies the late activation of myoD and myl1 expression at stage 20. This validates the appearence of this muscle-like NC subpopulation among other NC derivatives.

# Kotov et al., Figure 2 - figure supplement 3

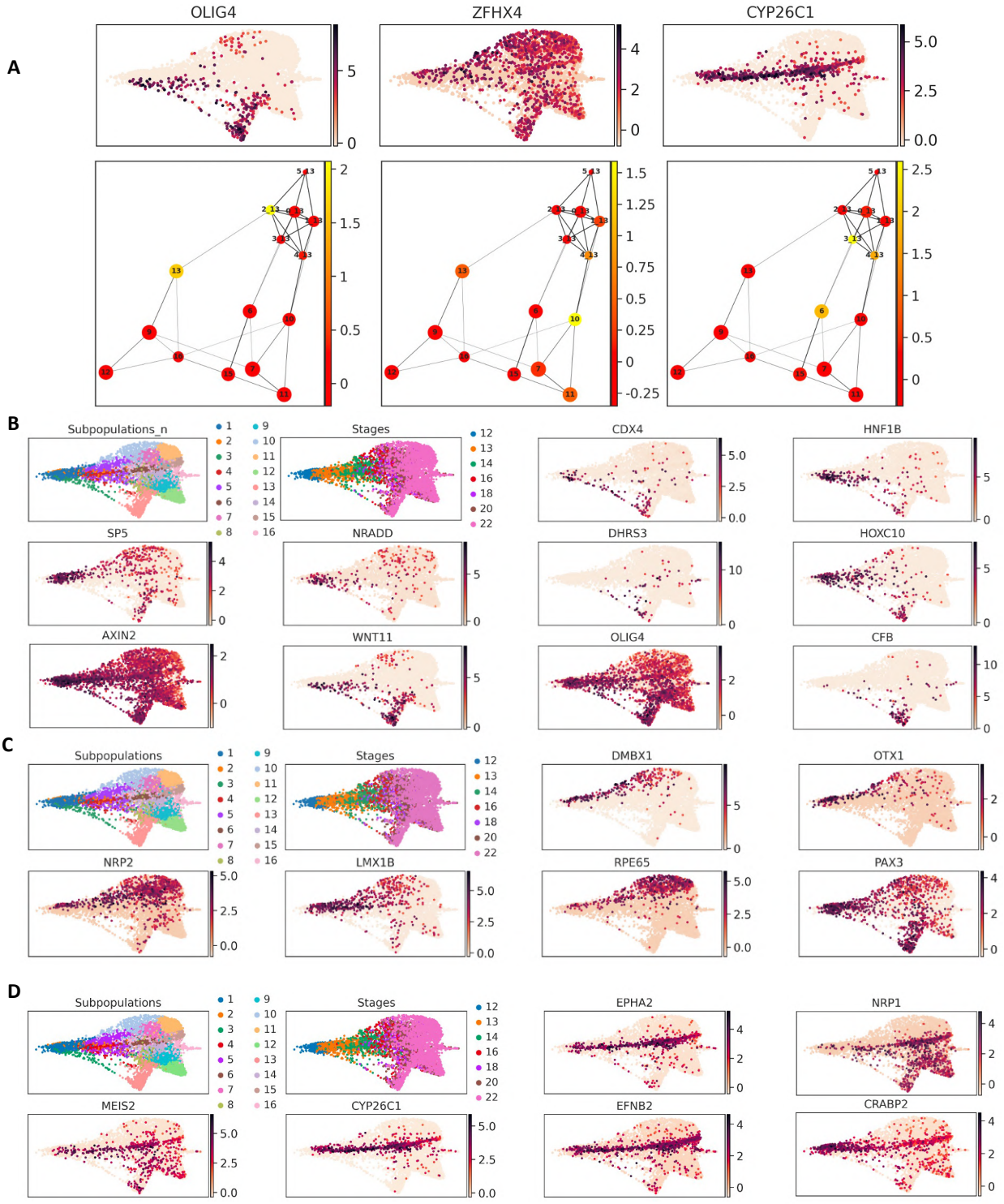



**Figure 2 - figure supplement 3. Early stage biased NC clusters.** (A) From the end of gastrulation, three previously undescribed early trends emerge. Clustering of stage 13 cells only, followed by PAGA analysis identified their similarities with the stage 20-22 clusters: an olig4-enriched subpopulation linked to ENSp cluster 13; a zfhx4-enriched cluster related to rpe65+ cluster 10 and a third stage 13 subpopulation expressing cyp26c1 linked to the cranial NC branch (clusters 4, 6, 15). (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. (C) UMAP plots of genes specific for cranial rpe65+hox- NC subpopulation specific genes (D) UMAP plots of genes specific for rhombomere 4, cranial NC subpopulations.

## Kotov et al., Figure 2 - figure supplement 4

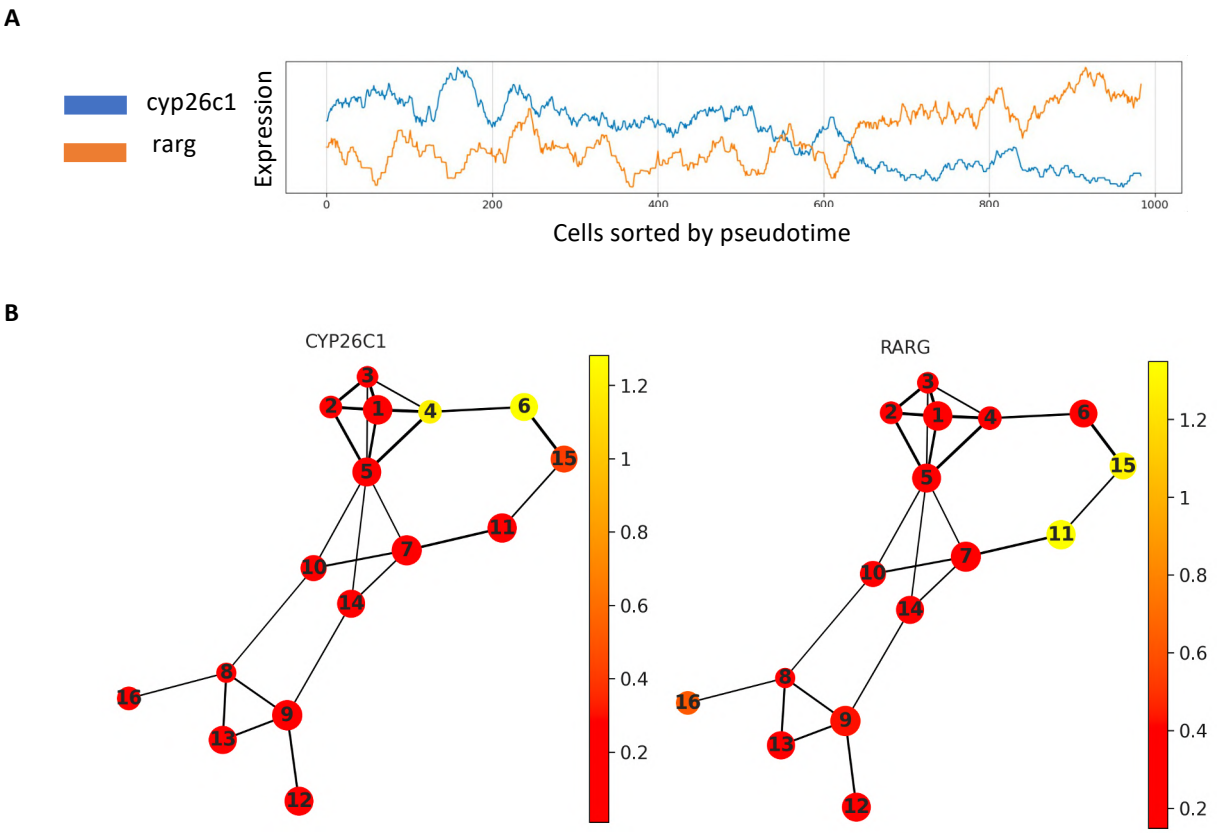



**Figure 2 - figure supplement 4. Early posterior enteric cluster 4 and related late clusters 6, 15.** (A) Line Plot for expression dynamics for cluster 4, 6, 15, representing the relationships between rarg and cyp26c1. Cells within clusters are sorted according to pseudotime. (B) PAGA plots with rarg and cyp26c1 mean expression.

## Kotov et al., Figure 2 - figure supplement 5

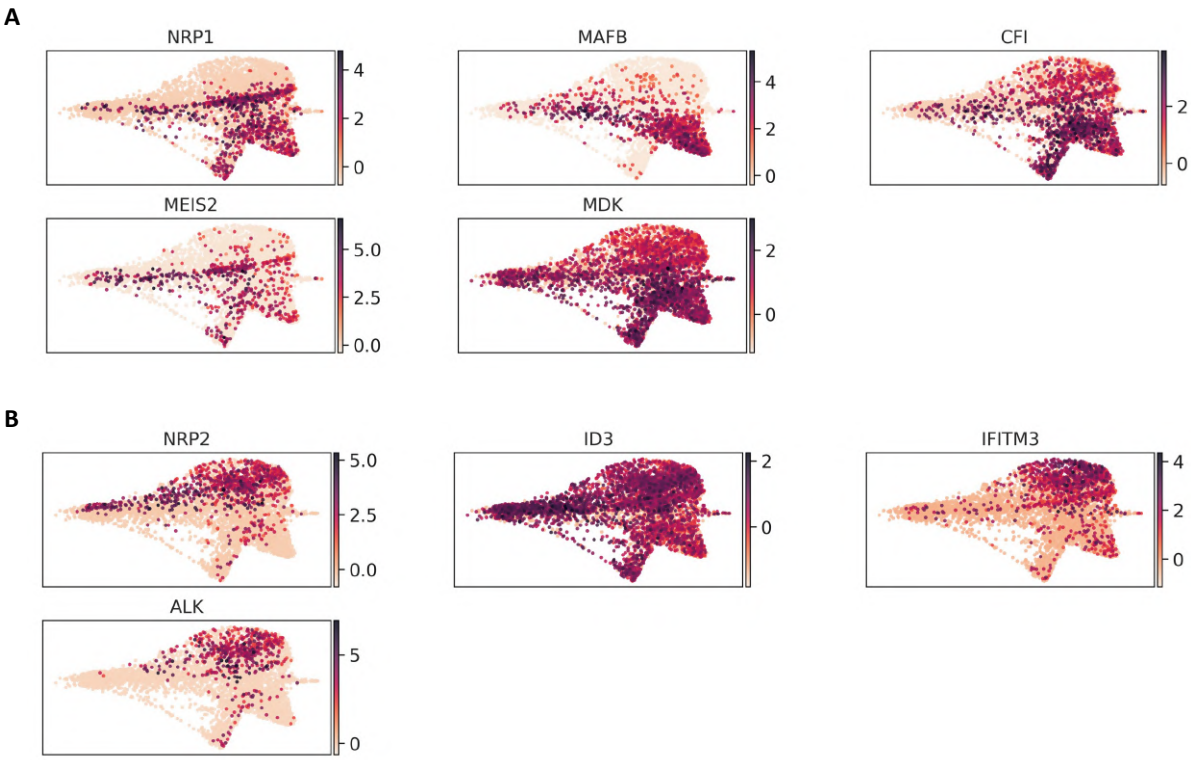



**Figure 2 - figure supplement 5. Early stage, bipotent NC clusters – cranial vs vagal lineages.**
UMAP plots for vagal (clusters 9, 12 and 13) and cranial (clusters 7, 10 and 11) specifiers.

cranial NC and supports that most NC cell fates are determined post-EMT (Baggiolini et al., 2015; Morrison et al., 2021).

Next, we were able to define multiple progenitor states linked to the later clusters defined above, and consequently delineate transcriptional dynamics of cranial, vagal and enteric nervous system progenitors between late gastrulation and EMT. Late cranial clusters 10 (*nrp2*$^{+}$*rpe65*$^{+}$*hox*$^{-}$) and 11 (*nrp2*$^{+}$*dlx2*$^{+}$*hox*$^{-}$) were related to cluster 7 (*nrp2*$^{+}$ *hox*$^{-}$) while late vagal clusters 12 (*mafb*$^{+}$*epha4*$^{+}$) and 13 (*tnc*$^{+}$*wnt11*$^{+}$*ltbp1*$^{+}$) were linked to cluster 9 (*nrp1*$^{+}$). Cluster 13 was also linked to the early cluster 3 (*hnf1b*$^{+}$). In turn, clusters 7 and 9 were related to the unbiased cluster 5 (*tfap2c*$^{+}$*sox9*$^{+}$). The *nrp1*$^{+}$ late cranial cluster 15 was linked to clusters 6 and 4, all of which specifically expressed *cyp26c1, epha2 and efnb2.* Cluster 16 (*myl1*$^{+}$, Figure 2 - figure supplement 2) was related to unbiased cluster 14, which was in turn linked to unbiased clusters 5, 1 (*c9*$^{+}$*sox9*$^{+}$) and 2 (*gjb1*$^{+}$) (Figure 2 – figure supplements 3-5; Figure 3).

Interestingly, most clusters gathered cells from multiple stages, indicating that a given transcriptome state can be reached by different cells over long developmental periods (Figure 2 – figure supplement 1B). For example, cranialmost cells of cluster 10 are generated over the entire duration of neurulation from neural plate stage 13 to migration stage 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, cranial clusters 11 and 15, two migratory clusters highly expressing *vimentin (vim*$^{+}$*),* are suddenly generated at post-EMT stage 22. This could reflect an abrupt step of mesenchymalisation in the epithelial-to-mesenchymal transition process of cranial NC, reflected in their transcriptomes.

Moreover, we detected cells with an immature pan-NC signature throughout neurulation and EMT, representing a long-lasting potentially “multipotent” and "stem-like" population (1, 5, 14; Figure 3). These unbiaised progenitors formed 72% of all NC cells during induction (gastrula and neural plate stages 12-14), 15% at pre-migratory and EMT stages 16-18, and 9% among tailbud stage 22 NC cells (Figure 3C). This dataset thus identifies the stem-like and partially biased premigratory NC populations; details the temporal dynamics of multipotency genes and the sequential appearence of regional modules; and characterizes the progressive activation of mesenchymal markers across a diversity of cell states. In sum, we uncover the highly diverse transcriptomes underlying premigratory NC cell biology along a large, cranial to vagal segment of the body axis.

## Kotov et al., Figure 3

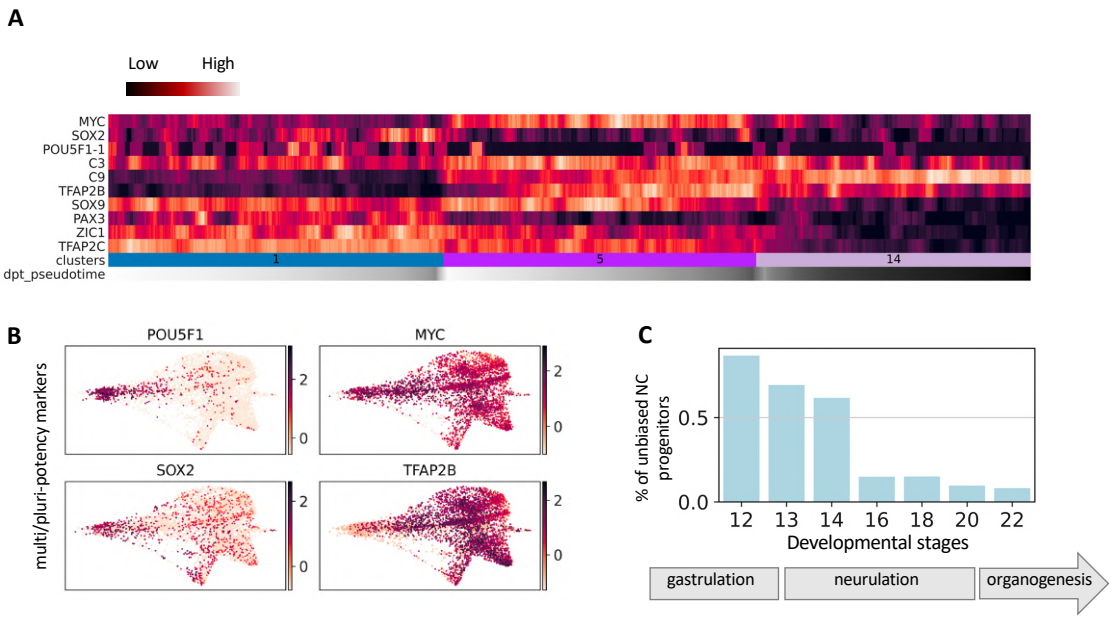



**Figure 3. Characteristics of the unbiased NC clusters.**
**(A)** Expression of pluripotency and "stem" neural crest genes (each bar is a group of 20 cells) in unbiased clusters 1, 5, 14. Pseudotime represents how far the cell has advanced along a given transcriptional path. **(B)** UMAP plots depicting expression of pluripotency-related markers. **(C)** Proportions of the unbiased cells at each developmental stage of the NC dataset.

## Connectome analysis by intersecting sc-transcriptomics, morpholino RNA-seq and ChIP-seq.

To expand the network of genes involved in the NC-GRN, we used an input list of 1417 transcription factors (TFs) (Blitz et al., 2017) and applied GRNboost2 on the NC dataset to link transcription factors to their potential targets in the NC transcriptome. We retrieved a network of 16978 potential TF-targets with a median of 22 connections per TF. Among the most connected genes for each stage (Table 1), we retrieved the previously known neural border specifiers *pax3* and *zic1* at NC induction stages, as well as *zic3*, *olig4*, *sox9*, and the anterior NC marker *dmbx1*. At pre-migratory stages, the prominent nodes included the NC specifiers *tfap2b, sox10 and snai2*, the anterior NC markers *rpe65* and *alx1*, and the hindbrain hox gene *hoxb3*. At migration stages, *tfap2e*, *mycn*, *dlx2* and *egr2* (*krox20)* displayed both high expression and many connections.

**Table 1.** Highly connected genes in the neural crest network predicted by GRNboost2.

| 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 |

In order to experimentally test the predicted network, we sequenced NB/NC microdissected ectoderm after *in vivo* knockdown of selected nodes *pax3* and *tfap2e*, using previously validated antisense morpholino oligonucleotides (MO) (Figure 4A, Hong et al., 2014; Monsoro-Burq et al., 2005). Expression of 1333 transcripts were decreased in Pax3 morphant NB, confirming that Pax3 is essential to activate a large NB/NC gene set ( Figure 4A, Supplementary file 1 – Table S4). We verified direct Pax3-binding targets *in vivo* using ChIP-seq on mid-neurula stage embryos (Figure 4B) and identified 657 potential targets in the whole embryo, of which 475 were expressed in the SC NC dataset (Supplementary file 1 – Table S6), including known direct targets, *e.g. cxcr4* and *prtg* (Plouhinec et al., 2014; Xu et al., 2018). Moreover, 80 of the ChIPseq-validated targets were predicted by GRN-Boost2 modelling including *psmd4*, *psen2*, *sp7*, *notch1*, *hnf1b*. In sum, we provide here a genome-wide Pax3 NC-GRN with three complementary approaches: co-expression predictions, Pax3 depletion and Pax3 chromatin-binding (Figure 4C). Using a similar approach, we confirmed TFAP2e as an important regulator at early EMT stages. Among 848 targets of TFAP2e predicted from scRNA-seq, 99 showed changed expression after TFAP2e knockdown in NC *in vivo* (e.g., *tfap2b, sox10, sncaip;* Figure 4 – figure supplements 1-2; Supplementary file 1 – Table S5). Moreover, using ChIP-seq for TFAP2e, we identified 642 targets expressed in the NC dataset, among 805 targets for the whole embryo, including top-scored *rmb20, pim1, arl5b* and *tfap2a* (Supplementary file 1 – Table S7). In both Pax3 and Tfap2e cases, we find that the three approaches result in partially overlapping gene target lists due to their use of different parameters (stage-wise, expression-wise, etc; Figure 4C, see Materials and Methods). Together, these data provide an enlarged and validated NC-specific genome-wide connectome for two key NB/NC specifiers.

# Kotov et al., Figure 4

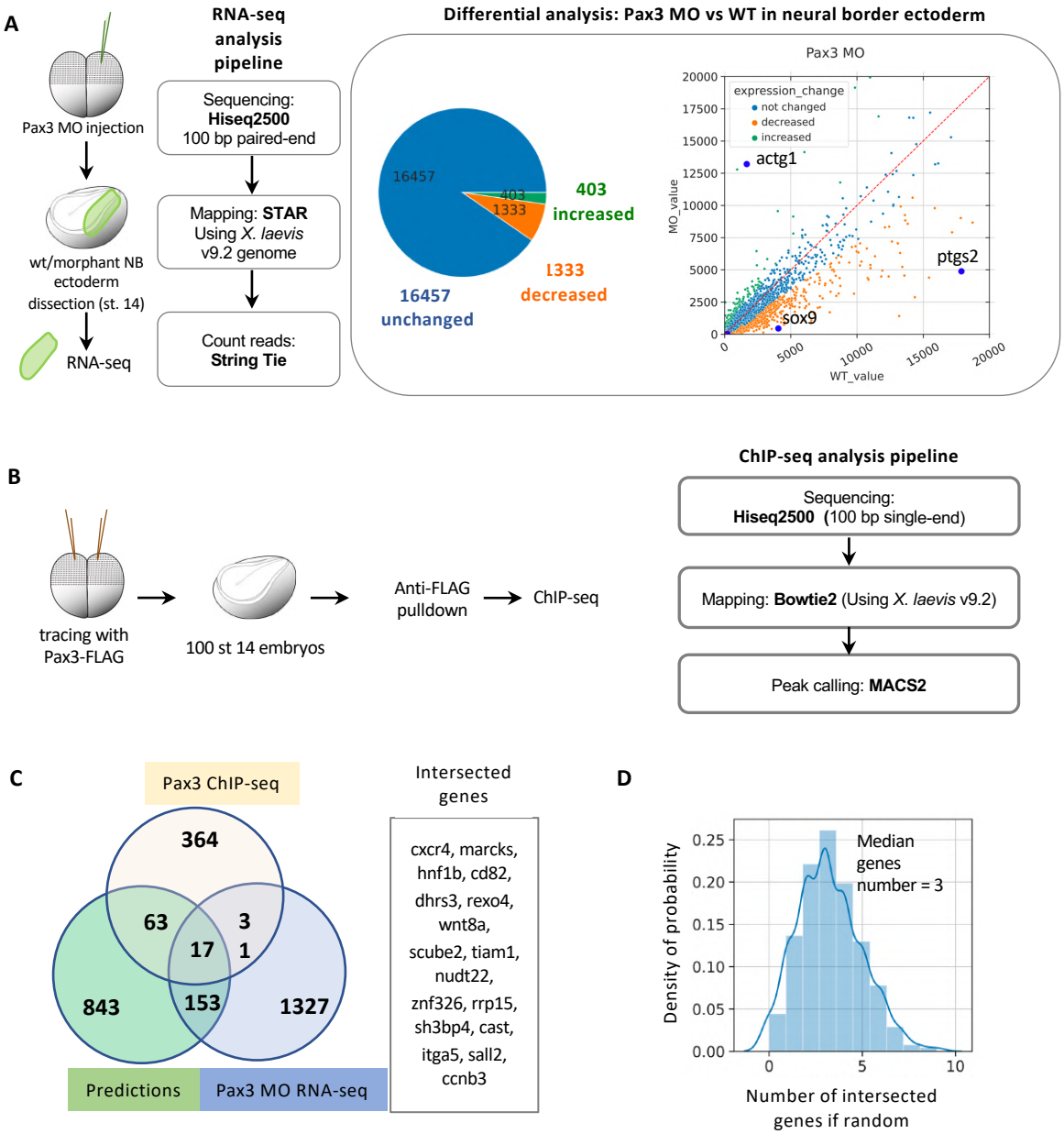



**Figure 4. Pax3 and TFAP2e connectome generation and validation.** (A) RNA-sequencing was done on microdissected NB ectoderm (stage 14), either wild-type (wt) or following Pax3 depletion in vivo (Pax3 MO). Using standard differential analysis, 1333 genes were found downregulated after Pax3 depletion, including sox9, olig4, hnf1b, msx1 (Table S4). (B) Tracing amounts of Pax3-FLAG expression were used to pull-down Pax3-bound chromatin in vivo without phenotypic modification, followed by standard ChIP-seq analysis pipeline, identifying 657 direct target genes among which 475 were expressed in the NC dataset. (C) Venn diagram compares Pax3 target genes validated by ChIP-seq, MO-RNA-seq and GRN-boost2 modeling. The 17 genes linked to Pax3 by all three methods are listed here (full lists in Supplementary file 1 – Tables S4 and S6). (D) Random sampling (bootstrap) was used to test the average number of genes present in the intersection by the three methods if random chance was applied to these datasets. At random, an average of 3 genes would be found in the intersection compared to 17 found here (C). Similar approach was applied to TFAP2e in Figure 4 - figure supplement 1, 2 and Table S5.

## Kotov et al., Figure 4 - figure supplement 1

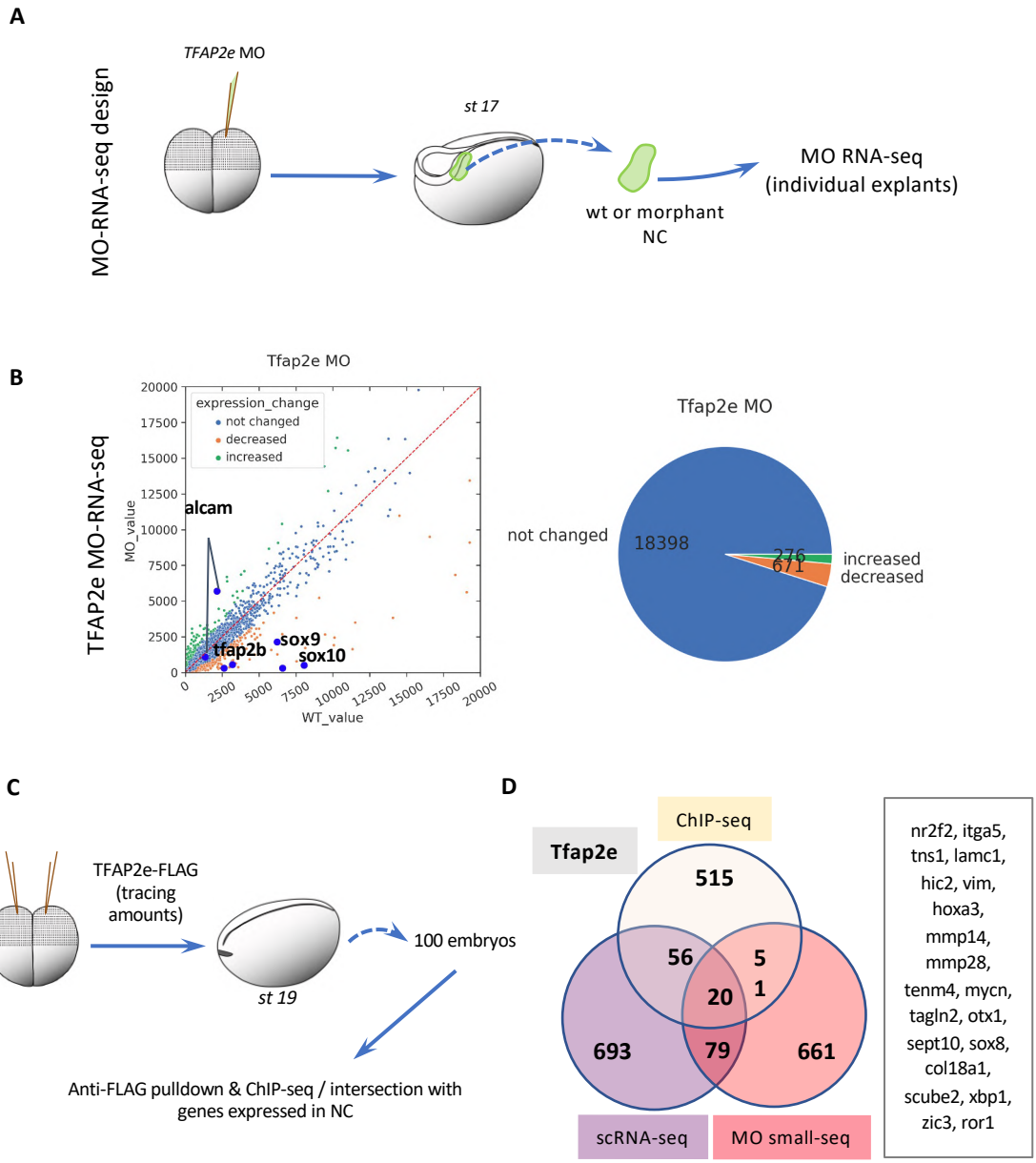



**Figure 4 - figure supplement 1. TFAP2e connectome generation and validation**
**(A)** Design of RNA-seq on microdissected NC explants after depletion of TFAP2e. **(B)** General statistics for differential analysis after RNA-seq of NC depleted for TFAP2e. **(C)** Design of ChIP-seq experiments for TFAP2e. **(D)** Venn diagram of target genes validated by ChIP-seq and transcriptome of TFAP2e-depleted NC, compared to predictions using GRNBoost2. The 20 genes linked to TFAP2e by all three methods are listed, which represents a significant increased compared to a "at random" situation. Among them we find genes important for NC EMT and migration *mmp14, mmp28, itga5* and *vim*.

## Kotov et al., Figure 4 - figure supplement 2

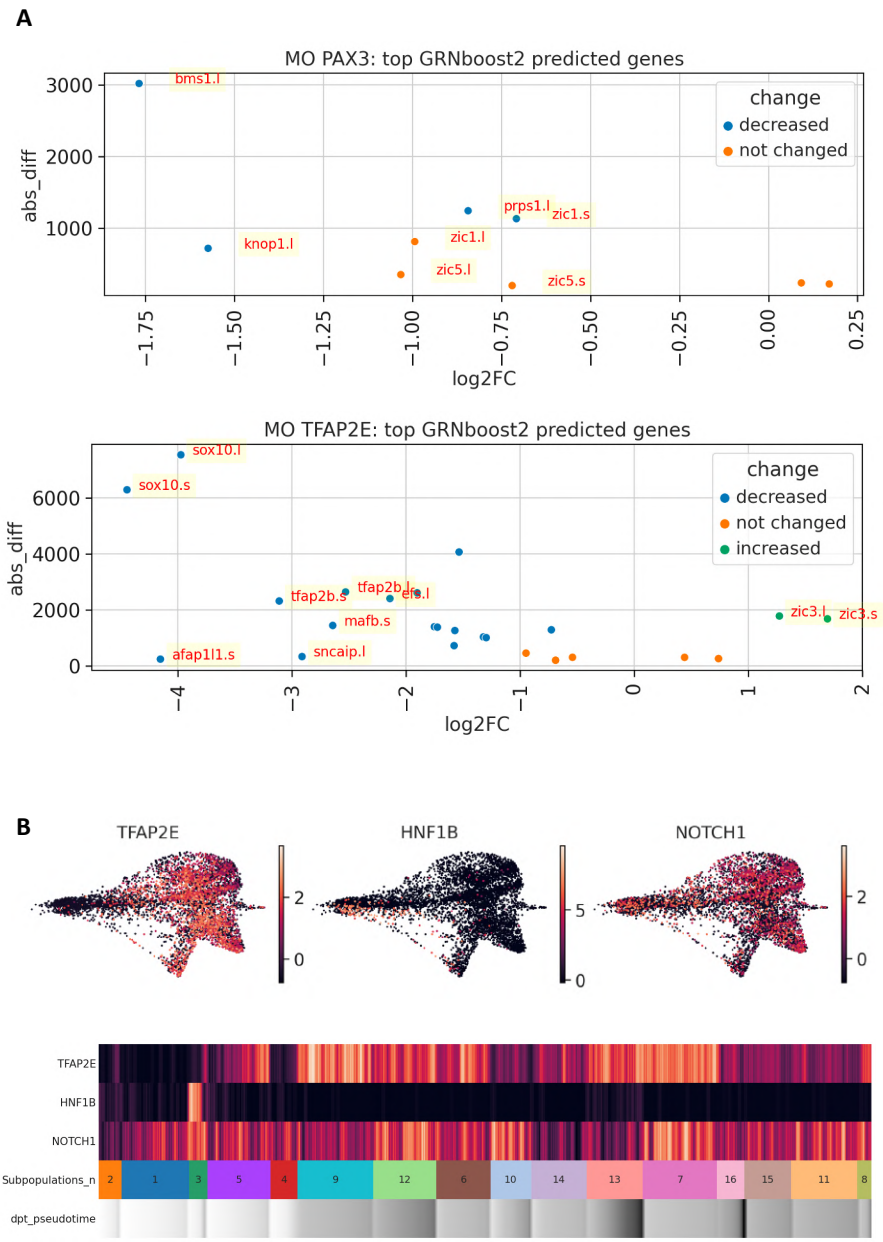



**Figure 4 - figure supplement 2**. **Validation of Pax3 and TFAP2e connectomes.** (A) The expression levels of the genes top-scored with GRNboost2 for Pax3 or TFAP2e were tested after Pax3 or TFAP2e depletion in NB/NC *in vivo*. Several of them were strongly decreased after Pax3 or TFAP2e depletion (e.g. *bms1* or *sox10* respectively). Others were unchanged meaning that, despite similar transcriptome dynamics at the single cell level (evaluated by GRNBoost2), there was no major functional regulation by either Pax3 or TFAP2e *in vivo* at the considered stage (unless unknown compensatory mechanisms were to apply). (B) Conversely, we identified some targets of TFAP2e, confirmed with ChIP-seq and MO, which were not predicted by GRNboost2 (e.g. *hnf1b* and *notch1*). This may result from functional regulations creating different expression patterns or different temporal dynamics, or else expression in adjacent cells (non cell-autonomous regulations), which are all parameters for which GRNBoost2 would not identify linkage. For example, neither *hnf1b* nor *notch1* were predicted as TFAP2e targets: *hnf1b* displays a very different expression pattern than *TFAP2e,* in the NC dataset, while *notch1,* despite a closely-related expression pattern at the cluster level, may in fact display distinct cell-to-cell expression at the single cell level.

## Branching towards biased premigratory neural crest subpopulations is controlled by key transcription factors.

To explore the temporal dynamics of TF expression that may specify decision points in the development of pre-migratory NC, we used tree inference (ElPiGraph) and advanced pseudotime downstream analysis focused on fate biasing using scFates (Albergante et al., 2020). We thus explored potential branch-specific transcriptional regulation from the calculated pseudotime, in order to determine not only the gene-to-gene dependency but also the temporal order in which their functions may be accomplished. ElPiGraph approximates datasets with complex topologies to build the graph structure, with the limitation that it cannot be applied to large datasets with many potential branches. Thus, using the principal graph constructed with PAGA (Figure 2B), we sub-selected cells around three main bifurcation points in the NC lineage tree and then applied scFates branching analysis (Figure 5A-F).

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

Cranial NC cells emerge from the neural tube anterior to the otic vesicle while vagal NC cells form from the hindbrain region adjacent to somites 1–7 (Le Douarin and Kalcheim, 1999). Our data indicated that the first cells biased towards vagal or cranial populations arose from the unbiased cluster 5 around early neural fold stage 14. Although we did not see a separation of cluster 5 into two populations at the chosen level of clustering, we still observe an early internal predisposition marked by the expression of early cranial (*nrp2)* and vagal (*mafb*) markers in sub-regions of cluster 5: cluster 5 cells that highly express *nrp2* were 7 times closer to the cranial state, while cells that highly express *mafb* were 1.5 times closer to the vagal state (Figure 5 – figure supplement 1A and B). Through branching analysis, we uncovered gene programs governing the bifurcation, consisting of 'early' genes activated before bifurcation and 'late' genes with continued expression in each branch (late): *nrp2*, *alk*, *rnd1*, *adam19* (early) and *shisa3*, *frdm6* (late) for the cranial branch, and *mafb*, *klb*, *mdk* (early) and *prdm1*, *cf1*, *hoxd3* (late) for the vagal branch (Figure 5A-B). By high specificity and early expression, *nrp2 and mafb* were the best early predictors of branching between the cranial and vagal populations (Figure 2E). These results match *in vivo* analyses on expression of key NC regulators, such as Mafb during cardiac NC specification or Alk in cranial NC migration (Gonzalez Malagon and Liu, 2018; Tani-Matsuhana and Inoue, 2021).

# Kotov et al., Figure 5

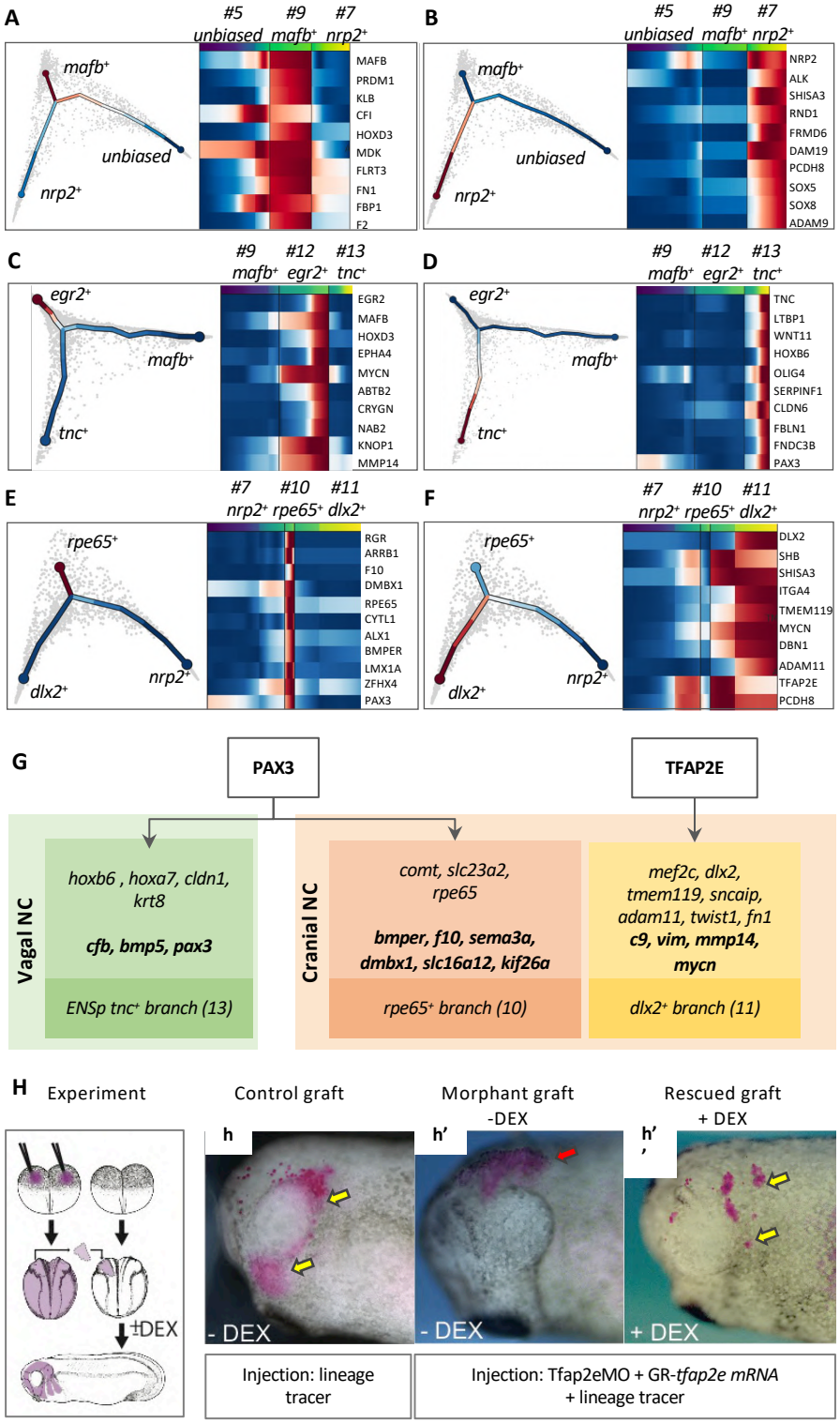
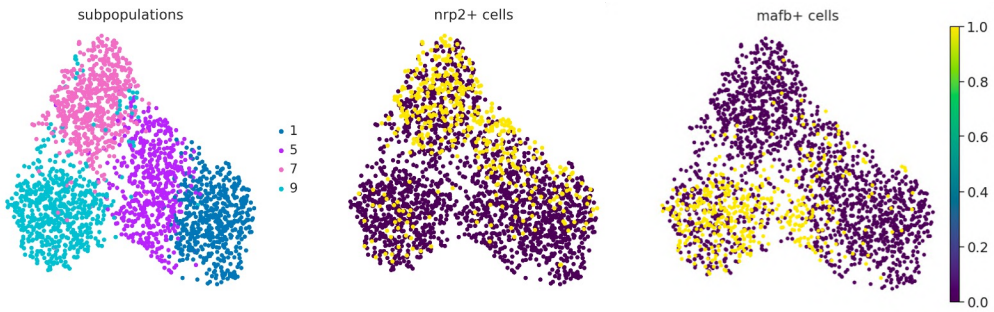



**Figure 5. NC branching analysis.** Transcriptomes of cells subselected around a chosen bifurcation point were analyzed using tree inference and pseudotime downstream analysis, yielding gene programs accompanying each trajectory. Gene programs for bifurcation of (A-B) premigratory unbiased cluster 5 into cluster 9 (A) and cluster 7 (B); (C-D) of migratory bipotent vagal cluster 9 into clusters 12 (C) and 13 (D); (E-F) of migratory bipotent cranial cluster 7 into clusters 10 (E) and 11 (F). (G) MO-mediated depletion and ChIP-seq was used to test Pax3 function in either vagal branching (cluster 12 vs 13) or cranial branching (cluster 11 vs 10) in vivo. Branch-specific genes with modified expression after Pax3 depletion (standard letters) or bound by Pax3 (in bold letters) are indicated. Using a similar approach, TFAP2e function in cranial branching (cluster 10 vs 11) was validated in vivo. (H) To test TFAP2e function in cranial NC migration, premigratory NC (traced in pink) was grafted into wild-type host embryos. In comparison to control (wt) NC active migration (h, yellow arows), TFAP2e morphant NC cells do not emigrate towards craniofacial areas and remain at the graft site (h', red arrow). In contrast, reactivation of TFAP2e-GR in morphant cells upon EMT stage restores NC cell migration (h'', yellow arrows).

# Kotov et al., Figure 5 - figure supplement 1

**A**




**B**

| | 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 |

**C**

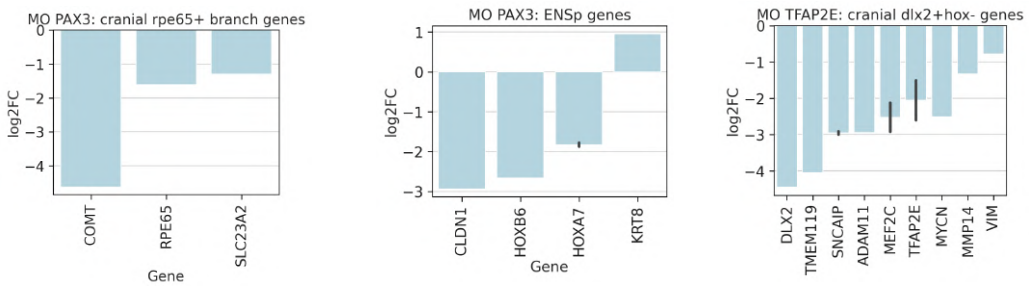



**Figure 5 - figure supplement 1. Branching analysis for NC dataset.**
**(A)** In order to identify potential early bias in the globally homogeneous cluster 5, cells located around the bifurcation between clusters 1 (dark blue), 5 (purple), 7 (pink), and 9 (cyan) were sub-selected. Cells expressing *nrp2* or *mafb* above 80 percentile were highlighted. This revealed that cluster 5 presented a minor predisposition for either *mafb* or *nrp2* expression. **(B)** Tables depicting cluster-cluster similarities calculated with PAGA. PAGA revealed that *nrp2*$^{+}$ cells of cluster 5 were 7 times more similar to the cranial state (cluster 7) than to the vagal state (cluster 9). Conversely *mafb*$^{+}$ cells of unbiased cluster 5 were 1.5 times closer to the vagal state than to the cranial. **(C)** To test branching analysis predictions, we used *in vivo* depletion of either Pax3 or TFAP2e, and tested the changes in expression of the late genes in Pax3- or TFAP2e-related branch.

***Vagal to enteric split and cranial subdivisions at neural fold stages.***

From the apparently homogeneous clusters 7 and 9, branching analysis predicted candidates regulating subsequent bifurcation. For the cardiac NC cluster 12, early-enriched transcripts were *mafb, mycn, prdm1, nolc1* and *eef1d*, late ones being *egr2, hoxd3, epha4* and *abtb2*. For the ENSp cluster 13, early markers were *olig4* and *fbn2*. Due to its significantly increased expression in later ENSp cells, *pax3* was identified as a late actor, but the expression-pseudotime heatmap showed that *pax3* was already expressed prior to branching in ENSp progenitors and increased afterward (Figures 2D and 5D). Therefore, we also assigned *pax3* to the early ENSp branch. Later ENSp gene program consists of *tnc*, *ltbp1*, *wnt11*, and *hoxb6*. In order to test experimentally if 'early' factors affected expression of 'late' genes, we examined which late ENSp genes may be targets of the early TF Pax3, through ChIP-seq and depletion analyses. Some genes were bound by Pax3 (*cfb*, *bmp5)*, while others showed increased (*krt8)* or decreased (*cldn1, hoxb6*, hoxa7) expression without evidence for Pax3 binding (Figure 5G, Figure 5 – figure supplement 1C).

Similar analysis on cranial NC clusters 7-10-11 showed that branch 10-specific genes were *rpe65, dmbx1, rgr, lmx1a,* and *zfhx4*, while late gene programs included *alx1* and *bmper* (Figure 5E-F). Interestingly, *pax3* was expressed early in the cranial bifurcation from 7 towards 10 and 11; expressed in cluster 7 cells, *pax3* was specifically enriched in *rpe65*$^{+}$ cluster 10. The Pax3 ChIP-seq and MO datasets revealed that *rpe65*$^{+}$ branch-specific transcripts *comt, slc23a2* and *rpe65* were decreased after *pax3* depletion while others, *bmper, f10, sema3a, dmbx1, slc16a12* and *kif26a*, were bound by Pax3 *in vivo* (Figure 5G). On the other hand, *tfap2e* expression initiated before bifurcation and was enriched in cluster 11 relative to cluster 10. Using TFAP2e depletion and ChIP-seq, we also validated that TFAP2e depletion reduced the expression of nine *hox*$^{-}$*dlx2*$^{+}$ branch-specific genes (e.g. early gene *mef2c*, and late genes *dlx2, mmp14, vim*), and that TFAP2e bound four other genes in cluster 11 signature (*c9, vim, mmp14 and mycn*, Figure 5G, Figure 5 – figure supplement 1C). While TFAP2e is essential for NC induction (Hong et al., 2014), direct regulation of *vim* and *mmp14* suggested that TFAP2e may also control cranial NC EMT and migration. We tested this later role *in vivo* using low-level depletion of TFAP2e which still allowed initial NC induction. Specifically, a pre-EMT NC explant co-injected with low-levels of TFAP2e MO, mRNA encoding a dexamethasone-inducible TFAP2e and a lineage tracer, was grafted into a wild-type control embryo prior to EMT (stage 17, Figure 5H). Morphant NC remained at the grafted site while wild-type cells efficiently populated the craniofacial areas (Figures 5Hh, h'). Importantly, when TFAP2e was re-activated in morphant NC at EMT stage, cell migration was restored and lineage-traced cells were found along NC cranial migration routes (Figure 5Hh"). Together, these results validate

branching analysis predictions and demonstrated that TFAP2e regulates expression of EMT effectors and cranial NC migration *in vivo*.

In conclusion, using computational approaches we have defined and analyzed three main bifurcation points in the premigratory NC dataset: from unbiased to vagal and cranial NC, from vagal NC to cardiac and ENSp fates, and from early cranial to either *rpe65*$^+$ or *dlx2*$^+$ cranial NC. For each branch, we defined specific gene programs, including early actors predicted to trigger specific states. Lastly, we validated the link between numerous branch-specific genes and two of the early regulators, Pax3 and TFAP2e.

**Coexisting neural border, ventral and dorsal ectoderm gene programs specify neural crest and placodes.**

***Neural border cell signature displays enriched ventral and dorsal gene expressions which are regulated by Pax3***

To unveil the molecular mechanisms that distinguish NC and placode induction at the neural-nonneural ectoderm border zone during gastrulation, we collected data for all ectoderm cells from stages 11-13 and identified 10 cell clusters (Figures 1A and 6A). The NB zone is an ectodermal area located in-between *sox2*$^+$ neuroepithelium (NE) and *epidermal keratin*$^+$ nonneural ectoderm (NNE), co-expresses *tfap2a, zic1* and *pax3* and gives rise to both NC and placodes (Figure 6A, B and Figure 6 - figure supplement 1B) (Seal and Monsoro-Burq, 2020). The early developmental dynamics of this ectodermal area has not yet been described at the single-cell level, and it remains unknown if NB cells resemble adjacent progenitors or if they exhibit a specific gene signature. On the force-directed graph plot, *tfap2a*$^+$*zic1*$^+$ cells (zone 3) did not appear as an individual cluster, rather as cells spread out amongst clusters 1, 2, 5 and 6 (Figure 6A, Figure 6 – figure supplement 1A). While we observed low cell density specifically for the NB zone between stages 11 and 12, we excluded a stage-related sampling issue since NE/NNE areas presented normal density. Rather this could relate to a faster transcriptional transition of NB cells compared to NNE cells: if early ectoderm cells transit through the NB state quickly before switching towards NC or placode states, fewer cells would be captured. Alternatively, such a plot could be obtained if the NB zone was highly heterogeneous and contained a mosaic of states (NB, NNE or NE).

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

| 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 |

We then identified a detailed signature for the *tfap2a*$^{+}$*zic1*$^{+}$ cells (zone 3). Several transcripts were enriched from stage 11: *tfap2c*, *pax3*, *sox9*, *hes1*, *gmnn* and *myc* (Figure 6 – figure supplement 1D). Most of these transcripts also formed main nodes in the whole Ectoderm connectome (10085 gene connections, Table 2). We examined whether Pax3, previously shown to appear as early as stage 10.5 *in vivo* (de Crozé et al., 2011), triggers expression of NB zone signature genes *in vivo*; in Pax3 morphant NB transcriptome, we found decreased expression of the other NB genes *sox9*, *axin2*, *zic3* and *zic1*, while early NE and NNE marker *lhx5.1* was increased (Figure 6 – figure supplement 1E). These results verify and consolidate Pax3 as a major activator for the early NB signature.

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

Based on the force-directed graph, we hypothesized three possible developmental routes leading to NC and placodes from stage 11 ectoderm cells: a) NE → NC; b) NB zone → PC and

# Kotov et al., Figure 6

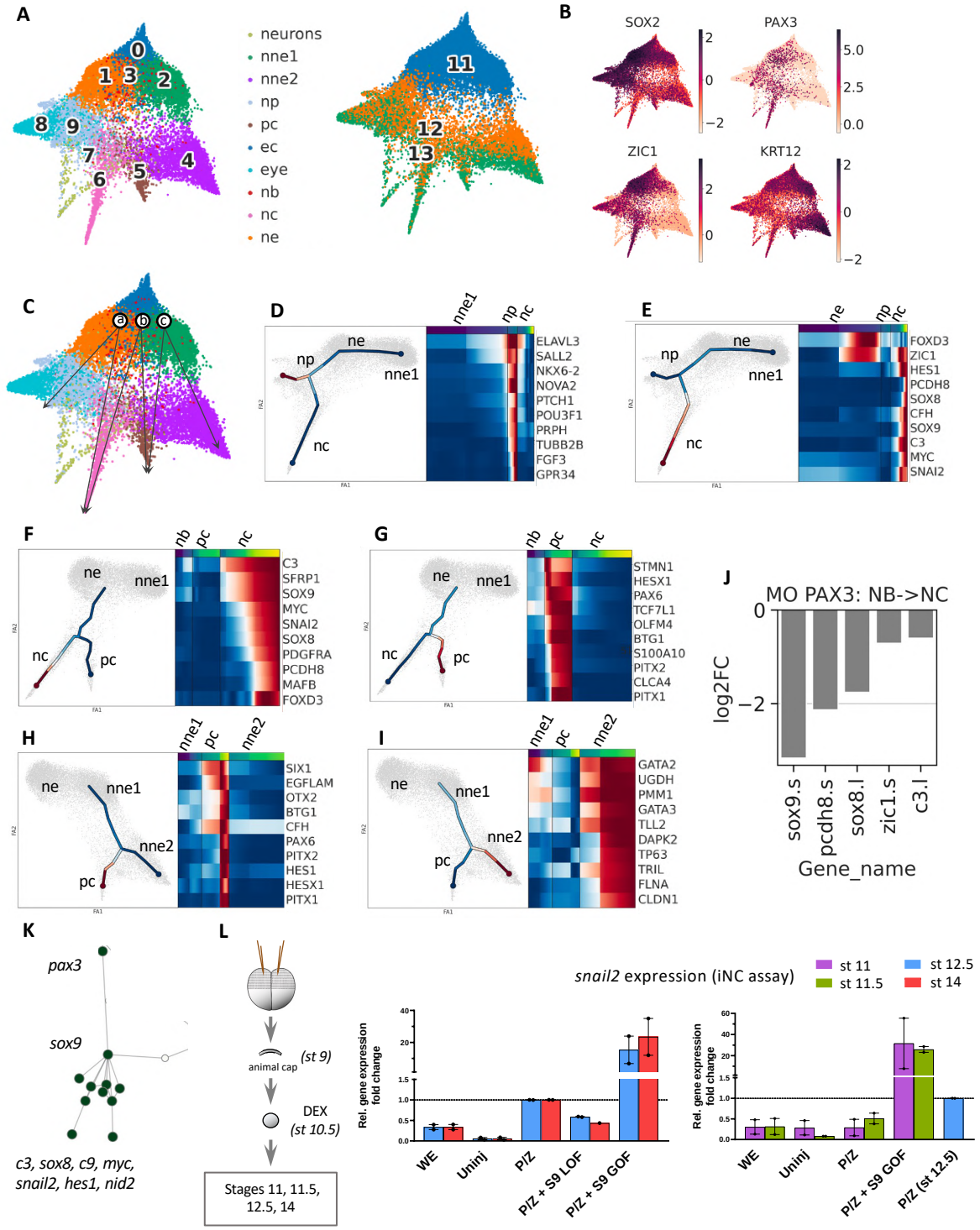



**Figure 6. Ectoderm branching and validation of Sox9 early role in NC induction.** (A) Forced-directed graphs for the Ectoderm dataset, with cluster numbers and gastrula stages indicated. (B) UMAP plots for genes marking the 3 major ectoderm areas: sox2 is enriched in the nascent neural ectoderm while krt12 (epidermal keratin 12) is highest in the non-neural ectoderm. Unlike the other areas, the NB zone does not display a specific gene signature and is usually depicted by pax3 expression or the overlapping expressions of ventral (tfap2a) and dorsal (zic1) genes (red cells in A, located across dorsal and ventral clusters). (C) Based on the forced-directed graph we hypothesized several paths for NC and PC development at stages 11-13, considering two possible origins for both populations (NP and NB for NC, and NNE and NB for PC). (D) Early NNE (NNE1)->NP gene program, (E) NE->NC gene program, (F) NB->NC gene program, (G) NB->PC gene program, (H) NNE1->PC gene program, (I) NNE1->later NNE (NNE2) gene program (J) Connectome and scFates branching analysis revealed that Pax3 was an important node in the predicted GRN and an early gene for branch NB->NC. In vivo, Pax3 depletion impacted expression of several NB->NC branch-specific genes, including sox9, c3 and foxd3. (K) Analysis of the Ectoderm connectome suggested a novel epistasis relationship between pax3, sox9 and other downstream NC specifiers. (L) In iNC assay, Sox9 acted downstream of Pax3 and was 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, activating Sox9 was 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 6 - figure supplement 1

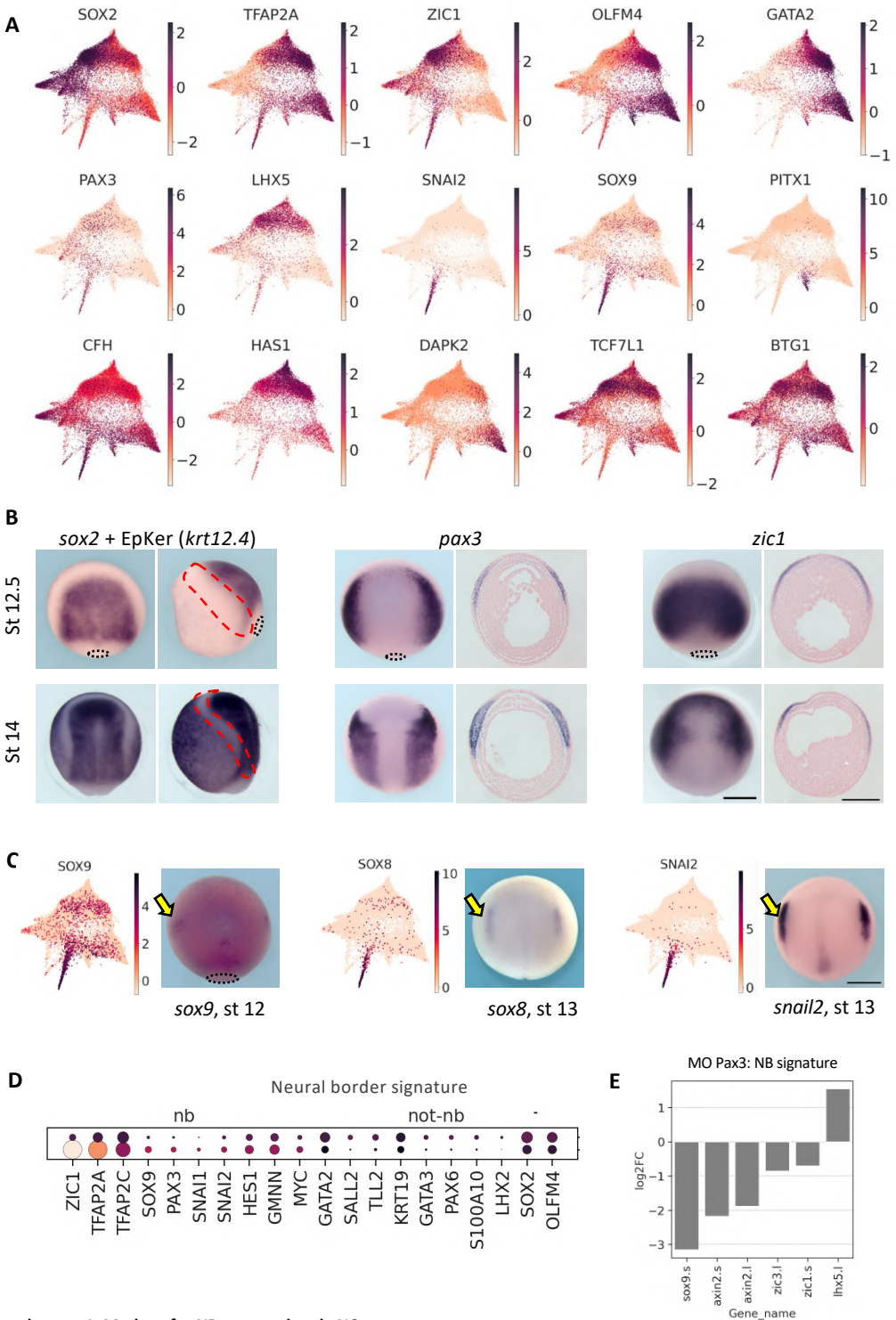



**Figure 6 - figure supplement 1. Markers for NB zone and early NC.**
**(A)** Ectoderm UMAP plots for stages 11, 12 and 13 representing gene expression of the main markers for ventral, dorsal and neural border, across the gastrula stage ectoderm. **(B)** In situ hybridization patterns for genes marking the 3 major ectoderm areas at stages 12.5 (late gastrula) and 14 (neurula). On dorsal and lateral views, *sox2* marks the neural (dorsal) ectoderm while *krt12* (epidermal keratin 12) marks the non-neural (ventral) ectoderm. Unlike the other derivatives, the NB zone does not yet have a large and specific gene marker list and is usually depicted by *pax3* expression or the overlapping expressions of dorsal and ventral genes like *zic1* and *tfap2a* respectively (dorsal views and transverse histological sections after whole mount *in situ* hybridization). Importantly, the neural border area (red dotted line) can also be defined as a zone devoid of both *sox2* and *keratin* expression as early as stage 12.5. **(C)** Ectoderm 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 $\mu$m. **(D)** Heatmap identify new genes enriched in the neural border defining an enlarged gene signature. **(C)** Neural border genes expression affected after Pax3 depletion *in vivo*.

NC; c) NNE → PC (Figure 6C); these possibilities are consistent with current models of NC and PC formation, the "neural border zone model" (route b) and the "neural vs non-neural" model (routes a & c) (Figure 6C) (Maharana and Schlosser, 2018; Roellig et al., 2017; Seal and Monsoro-Burq, 2020). Interestingly using branching analysis, no direct route was found between early NNE (nne1) and NC, or between NE and PC, confirming biological features, such as the partial neuralization needed for NC induction or the close relationship between placodes and NNE (Alkobtawi et al., 2021; Briggs et al., 2018). Since ElPiGraph cannot be applied on the whole dataset, we applied branching analysis to each potential route and compared the resulting gene programs underlying NC vs PC fate decisions. For route (a), branching towards NP from the NE state involved early genes *elav3 and sall2* and late genes *nkx6* and *tubb2b*, consistent with current knowledge on neural plate induction (*sall2, nkx6*) and primary neurogenesis (*elav3, tubb2b*) (Figure 6D, Exner et al., 2017). Route (a) branching towards NC involved early genes *zic1* and *foxd3* and late genes *c3, myc, sox8, sox9, pcdh8* and *snai2* (Figure 6E). In comparison, NC cells emerging from the NB zone by route (b) expressed *c3* and *sox9* before bifurcation, while NC markers *foxd3, sox8, pcdh8, snail2* were enriched after splitting (Figure 6F). Similarly, we explored both proposed routes of placodal development: route (b) from NB zone showed early enrichment of *tcf7l1, hesx1* and late for *stmn1, pax6, pitx1/2* (Figure 6G) while route (c) from NNE exhibited early enrichment of placode specifiers *six1, otx2*, and late expression of placode markers *egflam, pax6, pitx1/2*, Figure 6H). Last, the route (c) branching for NNE formation confirmed known developmental dynamics with early enrichment for *gata2* and late for ectoderm stem cell marker *tp63* and epithelial cells *cldn1* (Figure 6I) (Haas et al., 2019). This hierarchy of gene expression, some likely to respond to external signals as well, along the different branches thus opens avenues to further elaborate each of the NC- and PC-GRNs.

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

Interestingly, according to the route studied, we found that (i) distinct genes were activated to obtain the same state, and that (ii) some genes were activated with different expression dynamics relative to different bifurcations. For example, during the NB→NC transition (route b, Figure 6F), *sox9* and *c3* were activated early (before bifurcation) suggesting that they could play a part in the fate decision network from NB progenitors. In contrast, during the NE→NC gene program (Figure 6E) *sox9* and *c3* were late genes while *foxd3* and *zic1* were expressed early. This observation suggested a new model of fate decisions in the developing ectoderm, where parallel and distinct genetic programs activated in distinct ectoderm progenitors may lead to a similar state.

Last, NB zone-specific gene *pax3* was expressed prior to bifurcation in the route (b), NB→NC gene program, and activated expression of late NC branch markers *sox9, sox8, zic1, pcdh8* and *c3* (Figure 6J). Moreover, the Ectoderm connectome described NC genes connected to the rest of the network through Pax3 and Sox9 (Figure 6K), suggesting that Sox9 might play a yet undescribed function downstream of Pax3 in NC induction and upstream of the other late NC-branch markers. This agreed with *sox9* being an early gene in the NB->NC branch. We tested the epistasis relationships between Pax3 and Sox9 in NC induction by combining Sox9 depletion or gain-of-function in the induced-neural crest assay (iNC is Pax3/Zic1-based NC induction from pluripotent ectoderm cells, Figure 6L). Early NC marker *Snail2* expression starts from gastrula stage 12.5 both *in vivo* and in iNC, and increases at neurula stage 14 (Figure 6 – figure supplement 1C). At both stages, co-activation of Sox9 strongly increased *snail2* expression in iNC, while Sox9 depletion reduced *snail2* activation (Figure 6L), indicating that Sox9 is required for efficient NC induction by Pax3 and Zic1. Interestingly, when iNC explants were analyzed prior to the normal onset of *snail2* expression, at mid-gastrula stage 11/11.5, Sox9 activation highly increased *snail2* expression, suggesting that Sox9 synergizes with Pax3 and Zic1 at the onset of NC induction.

In conclusion, we have defined a new transcriptional signature for the incompletely described Neural Border zone, established a global Ectoderm connectome and validated experimentally NC-related nodes, in particular highlighting how Sox9 enhances NC induction downstream of Pax3 and Zic1. We characterized three different transcriptional programs branching from neural, neural border and non-neural ectoderm progenitors towards NC and PC, and propose a model in which multiple co-existing paths lead to the early NC or placode states in gastrula-stage ectoderm.

## Discussion

In this work, we exploit the resolution of high-density single cell transcriptomes collected from 8 frog developmental stages to unravel the emergence of the neural crest lineage from the ectoderm during gastrulation, followed by diversification of neural crest progenitors during neurulation and upon EMT. Modeling gene transcription dynamics around each cell state allows the inference of the underlying molecular networks. We selected several important nodes for large-scale and *in vivo* experimental validation (Figure 7A, B). This study highlights the previously unknown temporal sequence of states in premigratory neural crest development. Firstly, we characterize neural crest activation either from a transient neural border state, or from a neural plate state during mid-gastrulation, and suggest a model that reconciles current debates upon multiple possible routes leading from immature ectoderm to neural crest and

placodes (Figure 7C). Secondly, we delineate the early and later neural crest transcriptome trajectories during neurulation and define key regulators of branching, leading to eight transitional states and eight early migration states (Figure 7D).

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

The molecular signature of the neural border ectoderm has been overtly simple, with only Pax3 (in frog and fish) or Pax7 (in chick) as relatively specific markers for this domain during gastrulation and early neurulation stages (Basch et al., 2006; Monsoro-Burq et al., 2005). Here, we have characterized the neural border ectoderm state by two features: the lower level of expression of genes expressed by adjacent neural (dorsal) and non-neural (ventral) cells (*sox2*, *lhx2*, *sall2*, *gata2*, *ker19*) and the increased expression of a large gene list including *zic1*, *tfap2a/c*, *pax3*, *sox9*, *hes1*, *cmyc*. We find that this state seems more transient than other ectoderm states, suggesting that these fate decisions occur quickly. In frog embryos, the end of gastrulation is clearly defined by blastopore closure, and this allows more precise exploration of timing compared to organisms with simultaneous gastrulation and neurulation such as chick embryos. In frog, fate choices in the dorsal ectoderm happen during the second half of gastrulation (between stage 11 and 12.5). As the neural plate forms and the blastopore closes (stage 13), fate decisions are clearly established between neural, non-neural, placode and neural crest with robust molecular signatures (Figure 6C). Modeling neural crest emergence from the neural border cell state confirmed the central role for Pax3 and suggested novel epistatic relationships between Pax3 and Sox9 upstream of the definitive neural crest state, defined by *snail2* expression. Importantly, we propose a novel model for the transcriptional pattern of decisions between the four main ectoderm fates, neural, non-neural, placodal, neural crest. Instead of contrasting “neural border” and “non-neural vs neural" hypotheses, we find that these routes are not exclusive and find trajectories supporting the emergence of neural crest from either the neural border or the nascent neural ectoderm on one hand, as well as two trajectories leading to placodes from either the neural border or the non-neural ectoderm. In each case, the gene programs underlying those alternative trajectories involve a subset of common genes and a few specific factors (Figure 6). For example, specific expression of *tcf7l1* and *stmn1* is found in placodes arising from NB zone, compared to the NNE route (Figure 6G, H). For neural crest, early *sox9* expression in the NB route contrasts with post-bifurcation expression in the NP route (Figure 6E, F). Thus, our SC transcriptome modeling reconciles and combines previously alternatives in a "Dual convergence Model" of neural crest and placode patterning, together with specific gene signatures ripe for future functional exploration (Figure 7C).

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

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

Our temporal analysis highlights three important points deepening our understanding of NC biology. Firstly, there have been long standing debates about the timing of NC fate decisions, prior or after EMT from the neural tube, in a variety of animal models (Kalcheim and Kumar, 2017). Importantly, we did not detect distinctive expression of predictive fate markers before EMT (e.g. for neuronal, glial skeletogenic or melanocyte fates). This suggested that, if some NC progenitors were biased towards a given fate prior to EMT, they did not exhibit a detectable signature in our dataset. However, our observations are in agreement with several lineage tracing studies showing the high multipotency of most NC cells when marked prior to EMT (Baggiolini et al., 2015). The first differentiation markers are found after emigration, as we detected myosin-like expression in a small subset of cells suggesting the emergence of previously poorly described NC-derived myofibroblasts shortly after EMT (Figure 2 – figure supplement 2). It also supports that our dataset is sensitive enough to detect other fate-specific markers if they were expressed at the end of neurulation. Secondly, our data support the early diversification into several distinct cell states prior, during and after EMT, contrasting with the

**Kotov et al., Figure 7**

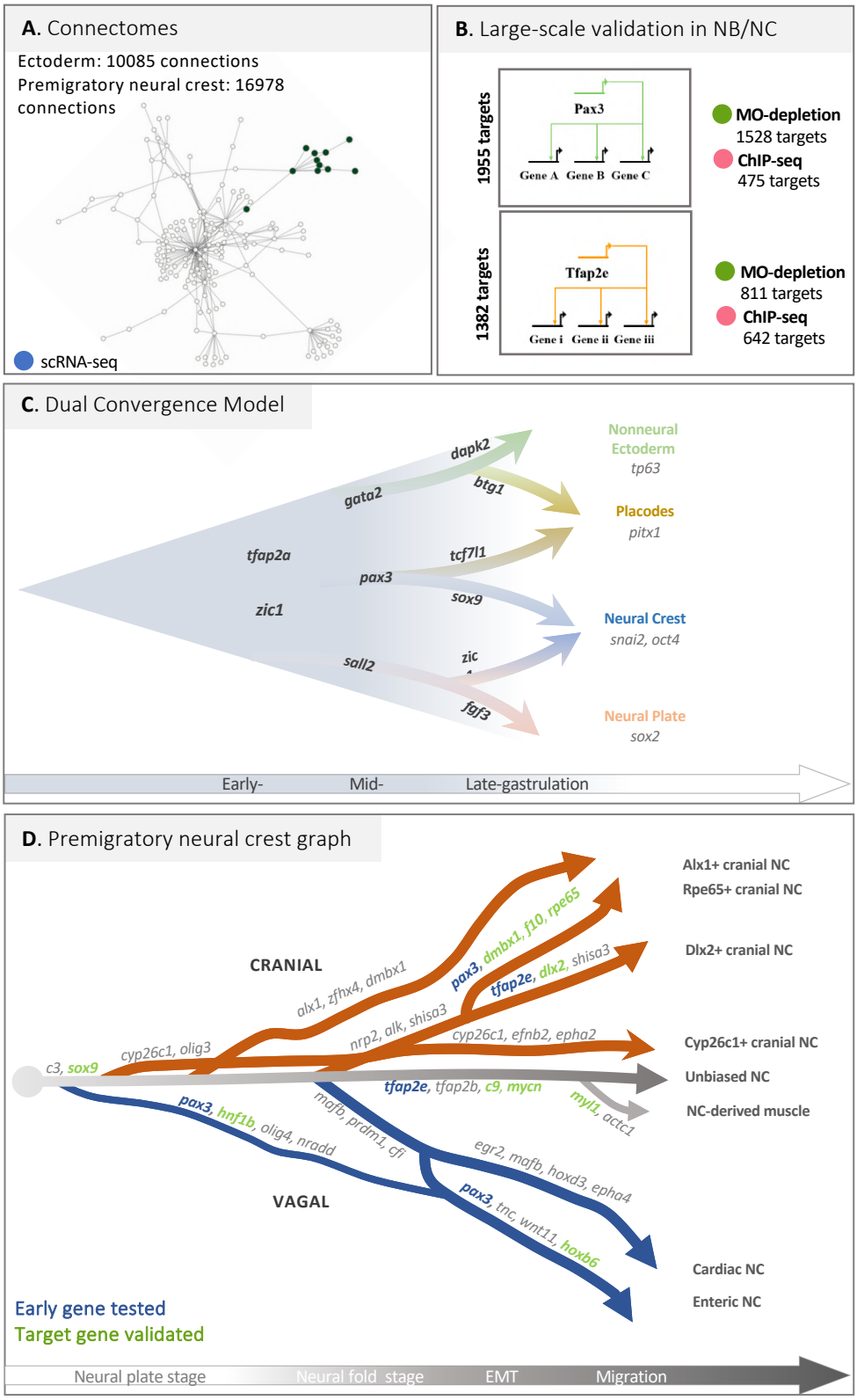



**Figure 7. Neural crest GRN and developmental trancriptome trajectories.** (A) GRNBoost2-predicted connectomes for Ectoderm (gastrula stage) and Neural crest (neurulation stage) can be queried at https://github.com/Qotov/neucrest_grn. (B) RNA-seq on dissected neural border ectoderm/neural crest explants and ChIP-seq provide large-scale GRNs linked to major nodes Pax3 and TFAP2e. (C) At gastrula stage (time line indicated), the neural border cells expressing tfap2a, zic1 and pax3 present trajectories towards placodes and neural crest. Those trajectories converge with a trajectory from neural plate towards neural crest and another one from non-neural ectoderm towards placodes. Branching analysis highlights a gene signature underlying those transcriptome transitions (the top gene is indicated here). (D) Transcriptomic tree of the neural crest cell states, from the end of gastrulation, during neurulation and upon epithelial-mesenchymal transition, ending at early migratory stage (time line indicated). Gene signatures supporting each trajectories are summarized. When Pax3 or Tfap2e were specifically expressed prior to branching (indicated in blue), their function on the expression of other genes in the branch-specific signature was tested in vivo. Genes with expression modulated after Pax3 or TFAP2e, manipulation in vivo and regulated are indicated in green.

recent suggestion that upon EMT the NC progenitors would regroup into a single common multipotent state (Zalc et al., 2021). The high cell content of our dataset proves otherwise, suggesting that this previous observation made on a smaller subset of cranial NC did not fully capture the diversity of pre-migratory NC states. Lastly, temporal trajectory analysis unravels the branch-specific dynamics of gene expression underlying bifurcations and state diversification. For each bifurcation, we provide a list of key genes likely to control branching choices (Figures 5 and 6). We further validate these predictions in several instances, by experimental modulation of pivotal transcription factors function in the premigratory neural crest (Pax3, TFAP2e), followed by *in vivo* or deep sequencing analysis. In sum, our study provides a comprehensive view of the hierarchy of molecular decisions driving the cranial and vagal neural crest gene regulatory network from induction at the neural border to early migration, with unprecedented resolution and deep learning-aided experimental validation. We propose a new "Dual Convergence Model" for neural crest and placode lineage emergence, and provide a detailed roadmap of the main molecular events in the premigratory and early migrating NC-GRN. Using a dedicated interactive network visualization interface, any gene of interest can be queried. Moreover, the detailed sequence of cell states provided here will prove an essential reference for monitoring induction of neural crest derivatives, for example from patient-derived induced pluripotent stem cells, when reliable specification protocols preferably recapitulate the steps of embryonic development.

## Materials ad Methods

### *Experimental Design.*

Single cell transcriptomes from developing *X. tropicalis* embryos were scrutinized for NC development using machine-learning tools to infer the gene regulatory network (GRN) and the gene programs underlying branching of fates. These predictions were largely validated *in vivo* using micro-manipulations in *X. laevis* embryos followed by RNA-seq or ChIPseq. Detailed material and methods are given in Supplementary File 1.

### *Single cell sequencing and processing.*

No new materials were collected for this study. Instead, we re-sequenced the SC RNA libraries for developmental stages NF11 to NF22 (Faber and Nieuwkoop, 2020) used in (Briggs et al., 2018) using NovaSeq S2. All datasets are deposited under NCBI Gene Expression Omnibus number GSE198494. For SC analysis, we used the *X.tropicalis* v.10 genome assembly, gene models v. 10.7, together with STAR aligner and the DropEst pipeline. After filtration by counts and genes numbers (>200 genes; >300 counts), we gathered a dataset of 177250 cells. In the cells of interest (Ectoderm and NC cells), mean counts number was 1778, and mean gene number was 1035. scRNA-seq postprocessing was done using Scanpy, a

comprehensive scRNA pipeline that is functionally similar to Seurat (Stuart et al., 2019, see references and details in Supplementary File 1).

***Clustering and NC cells selection***

For each stage, we obtained independent standard dimensionality reduction with PCA, computing a neighborhood graph and UMAP (Jacomy et al., 2014) followed by clustering, (Leiden algorithm, Traag et al., 2019). For each cluster, we defined cluster-specific genes with differential expression analysis (scanpy t-test_overestim_var) and selected only clusters which were the most similar to NC cells using NC signatures from (Briggs et al., 2018). For NC subclustering, 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).

***GRN generation***

Using the temporal dynamics of gene expression, we used GRNBoost2 to infer genetic co-regulation, starting from a list of TFs (Blitz et al., 2017). GRNBoost2 retrieves the gene regulatory network (GRN) from the expression data-matrix (Moerman et al., 2019). In the resulting network of TFs and their targets we identified the most important nodes by calculating betweenness and degree centralities.

***Principal graph generation and branching analysis.***

The tree analysis was carried out using the scFates package (Albergante et al., 2020). 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 (Wolf et al., 2018). 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. 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 (Wills et al., 2014) after injection of tracing amounts (75 pg) of mRNA encoding either Pax3-FLAG-HA or TFAP2e-FLAG). After sequencing, 100 bp single-end reads were aligned to *X. laevis* genome version 9.2 (Pax3) or *X. tropicalis* v10.0 (TFAP2e) using Bowtie2. Peaks were called using MACS2.

For Pax3 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 RNA-seq***

*In vivo* injections, NB/NC dissections and grafting were done as previously (Milet and Monsoro-Burq, 2014; Plouhinec et al., 2017) using *X. laevis* embryos. For knockdown experiments, previously validated antisense morpholino oligonucleotides (MO) were used to deplete *pax3,* or *TFAP2e* transcripts (Monsoro-Burq et al., 2005; Hong et al., 2014). One Pax3 morphant anterior NB explant (stage 14) or one TFAP2e-morphant NC explant (stage 17) were dissected from the injected side, in triplicate. Each explant was sequenced individually (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.***

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) (Hong and Saint-Jeannet, 2007; Milet et al., 2013). This was combined with Sox9 depletion (40 ng of *sox9* MO) (Spokony et al., 2002), or gain-of-function (300 pg *sox9* mRNA). At the desired stage, explants were harvested and processed for RTqPCR as in (Alkobtawi et al., 2021). Primers are listed in Table S8.

***Whole-mount in situ hybridization (ISH)***

Whole-mount *in situ* hybridization followed a protocol optimized for NC (Monsoro-Burq, 2007). Embryos were imaged using a Lumar V12 Binocular microscope equipped with bright field and color cameras (Zeiss).

**References**

Albergante L, Mirkes E, Bac J, Chen H, Martin A, Faure L, Barillot E, Pinello L, Gorban A, Zinovyev A. 2020. Robust and Scalable Learning of Complex Intrinsic Dataset Geometry via ElPiGraph. *Entropy* **22**:296. doi:10.3390/e22030296

Alkobtawi M, Pla P, Monsoro-Burq AH. 2021. BMP signaling is enhanced intracellularly by FHL3 controlling WNT-dependent spatiotemporal emergence of the neural crest. *Cell Reports* **35**:109289. doi:10.1016/j.celrep.2021.109289

572 Artinger KB, Monsoro-Burq AH. 2021. Neural crest multipotency and specification: power
573 and limits of single cell transcriptomic approaches. *Fac Rev* **10**:38. doi:10.12703/r/10-
574 38
575 Baggiolini A, Varum S, Mateos JM, Bettosini D, John N, Bonalli M, Ziegler U, Dimou L,
576 Clevers H, Furrer R, Sommer L. 2015. Premigratory and migratory neural crest cells
577 are multipotent in vivo. *Cell Stem Cell* **16**:314–322. doi:10.1016/j.stem.2015.02.017
578 Basch ML, Bronner-Fraser M, García-Castro MI. 2006. Specification of the neural crest
579 occurs during gastrulation and requires Pax7. *Nature* **441**:218–222.
580 doi:10.1038/nature04684
581 Bhattacharya D, Azambuja AP, Simoes-Costa M. 2020. Metabolic Reprogramming Promotes
582 Neural Crest Migration via Yap/Tead Signaling. *Dev Cell* **53**:199-211.e6.
583 doi:10.1016/j.devcel.2020.03.005
584 Blitz IL, Paraiso KD, Patrushev I, Chiu WTY, Cho KWY, Gilchrist MJ. 2017. A catalog of
585 Xenopus tropicalis transcription factors and their regional expression in the early
586 gastrula stage embryo. *Dev Biol* **426**:409–417. doi:10.1016/j.ydbio.2016.07.002
587 Briggs JA, Weinreb C, Wagner DE, Megason S, Peshkin L, Kirschner MW, Klein AM. 2018.
588 The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution.
589 *Science* **360**:eaar5780. doi:10.1126/science.aar5780
590 de Crozé N, Maczkowiak F, Monsoro-Burq AH. 2011. Reiterative AP2a activity controls
591 sequential steps in the neural crest gene regulatory network. *Proc Natl Acad Sci U S A*
592 **108**:155–160. doi:10.1073/pnas.1010740107
593 Eames BF, Medeiros DM, Adameyko I. 2020. Evolving Neural Crest Cells. CRC Press.
594 Exner CRT, Kim AY, Mardjuki SM, Harland RM. 2017. sall1 and sall4 repress pou5f3 family
595 expression to allow neural patterning, differentiation, and morphogenesis in Xenopus
596 laevis. *Developmental Biology* **425**:33–43. doi:10.1016/j.ydbio.2017.03.015
597 Faber J, Nieuwkoop PD. 2020. Normal Table of Xenopus Laevis (Daudin): A Systematical &
598 Chronological Survey of the Development from the Fertilized Egg till the End of
599 Metamorphosis. Garland Science.
600 Figueiredo AL, Maczkowiak F, Borday C, Pla P, Sittewelle M, Pegoraro C, Monsoro-Burq
601 AH. 2017. PFKFB4 control of Akt signaling is essential for premigratory and
602 migratory neural crest formation. *Development* dev.157644. doi:10.1242/dev.157644
603 Gonzalez Malagon SG, Liu KJ. 2018. ALK and GSK3: Shared Features of Neuroblastoma and
604 Neural Crest Cells. *J Exp Neurosci* **12**:1179069518792499.
605 doi:10.1177/1179069518792499

606 Groves AK, LaBonne C. 2014. Setting appropriate boundaries: Fate, patterning and
607 competence at the neural plate border. *Dev Biol* **389**:2–12.
608 doi:10.1016/j.ydbio.2013.11.027
609 Haas M, Gómez Vázquez JL, Sun DI, Tran HT, Brislinger M, Tasca A, Shomroni O,
610 Vleminckx K, Walentek P. 2019. ΔN-Tp63 Mediates Wnt/β-Catenin-Induced
611 Inhibition of Differentiation in Basal Stem Cells of Mucociliary Epithelia. *Cell Rep*
612 **28**:3338-3352.e6. doi:10.1016/j.celrep.2019.08.063
613 Hong C-S, Devotta A, Lee Y-H, Park B-Y, Saint-Jeannet J-P. 2014. Transcription factor AP2
614 epsilon (Tfap2e) regulates neural crest specification in Xenopus. *Dev Neurobiol*
615 **74**:894–906. doi:10.1002/dneu.22173
616 Hong C-S, Saint-Jeannet J-P. 2007. The Activity of Pax3 and Zic1 Regulates Three Distinct
617 Cell Fates at the Neural Plate Border. *MBoC* **18**:2192–2202. doi:10.1091/mbc.e06-11-
618 1047
619 Jacomy M, Venturini T, Heymann S, Bastian M. 2014. ForceAtlas2, a Continuous Graph
620 Layout Algorithm for Handy Network Visualization Designed for the Gephi Software.
621 *PLOS ONE* **9**:e98679. doi:10.1371/journal.pone.0098679
622 Kalcheim C, Kumar D. 2017. Cell fate decisions during neural crest ontogeny. *Int J Dev Biol*
623 **61**:195–203. doi:10.1387/ijdb.160196ck
624 Le Douarin N, Kalcheim C. 1999. The Neural Crest, 2nd ed, Developmental and Cell Biology
625 Series. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511897948
626 Ling ITC, Sauka-Spengler T. 2019. Early chromatin shaping predetermines multipotent vagal
627 neural crest into neural, neuronal and mesenchymal lineages. *Nat Cell Biol* **21**:1504–
628 1517. doi:10.1038/s41556-019-0428-9
629 Maharana SK, Schlosser G. 2018. A gene regulatory network underlying the formation of pre-
630 placodal ectoderm in Xenopus laevis. *BMC Biol* **16**:79. doi:10.1186/s12915-018-0540-
631 5
632 Medina-Cuadra L, Monsoro-Burq AH. 2021. Xenopus, an emerging model for studying
633 pathologies of the neural crest. *Curr Top Dev Biol* **145**:313–348.
634 doi:10.1016/bs.ctdb.2021.03.002
635 Milet C, Maczkowiak F, Roche DD, Monsoro-Burq AH. 2013. Pax3 and Zic1 drive induction
636 and differentiation of multipotent, migratory, and functional neural crest in Xenopus
637 embryos. *Proc Natl Acad Sci U S A* **110**:5528–5533. doi:10.1073/pnas.1219124110
638 Milet C, Monsoro-Burq AH. 2014. Dissection of Xenopus laevis neural crest for in vitro
639 explant culture or in vivo transplantation. *J Vis Exp*. doi:10.3791/51118

Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, Aerts S. 2019. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**:2159–2161. doi:10.1093/bioinformatics/bty916

Monsoro-Burq AH. 2007. A Rapid Protocol for Whole-Mount In Situ Hybridization on Xenopus Embryos. *Cold Spring Harb Protoc* **2007**:pdb.prot4809. doi:10.1101/pdb.prot4809

Monsoro-Burq A-H, Wang E, Harland R. 2005. Msx1 and Pax3 cooperate to mediate FGF8 and WNT signals during Xenopus neural crest induction. *Dev Cell* **8**:167–178. doi:10.1016/j.devcel.2004.12.017

Morrison JA, McLennan R, Teddy JM, Scott AR, Kasemeier-Kulesa JC, Gogol MM, Kulesa PM. 2021. Single-cell reconstruction with spatial context of migrating neural crest cells and their microenvironments during vertebrate head and neck formation. *Development* **148**:dev199468. doi:10.1242/dev.199468

Plouhinec J-L, Medina-Ruiz S, Borday C, Bernard E, Vert J-P, Eisen MB, Harland RM, Monsoro-Burq AH. 2017. A molecular atlas of the developing ectoderm defines neural, neural crest, placode, and nonneural progenitor identity in vertebrates. *PLOS Biology* **15**:e2004045. doi:10.1371/journal.pbio.2004045

Plouhinec J-L, Roche DD, Pegoraro C, Figueiredo A-L, Maczkowiak F, Brunet LJ, Milet C, Vert J-P, Pollet N, Harland RM, Monsoro-Burq AH. 2014. Pax3 and Zic1 trigger the early neural crest gene regulatory network by the direct activation of multiple key neural crest specifiers. *Dev Biol* **386**:461–472. doi:10.1016/j.ydbio.2013.12.010

Roellig D, Tan-Cabugao J, Esaian S, Bronner ME. 2017. Dynamic transcriptional signature and cell fate analysis reveals plasticity of individual neural plate border cells. *Elife* **6**:e21620. doi:10.7554/eLife.21620

Sato T, Sasai N, Sasai Y. 2005. Neural crest determination by co-activation of Pax3 and Zic1 genes in Xenopus ectoderm. *Development* **132**:2355–2363. doi:10.1242/dev.01823

Seal S, Monsoro-Burq AH. 2020. Insights Into the Early Gene Regulatory Network Controlling Neural Crest and Placode Fate Choices at the Neural Border. *Front Physiol* **11**:608812. doi:10.3389/fphys.2020.608812

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

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

Steventon B, Mayor R. 2012. Early neural crest induction requires an initial inhibition of Wnt signals. *Dev Biol* **365**:196–207. doi:10.1016/j.ydbio.2012.02.029

Streit A, Stern CD. 1999. Establishment and maintenance of the border of the neural plate in the chick: involvement of FGF and BMP activity. *Mechanisms of Development* **82**:51–66. doi:10.1016/S0925-4773(99)00013-1

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell* **177**:1888-1902.e21. doi:10.1016/j.cell.2019.05.031

Tang W, Li Y, Li A, Bronner ME. 2021. Clonal analysis and dynamic imaging identify multipotency of individual Gallus gallus caudal hindbrain neural crest cells toward cardiac and enteric fates. *Nat Commun* **12**:1894. doi:10.1038/s41467-021-22146-8

Tani-Matsuhana S, Inoue K. 2021. Identification of regulatory elements for MafB expression in the cardiac neural crest. *Cells Dev* **167**:203725. doi:10.1016/j.cdev.2021.203725

Traag VA, Waltman L, van Eck NJ. 2019. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**:5233. doi:10.1038/s41598-019-41695-z

Uehara M, Yashiro K, Mamiya S, Nishino J, Chambon P, Dolle P, Sakai Y. 2007. CYP26A1 and CYP26C1 cooperatively regulate anterior-posterior patterning of the developing brain and the production of migratory cranial neural crest cells in the mouse. *Dev Biol* **302**:399–411. doi:10.1016/j.ydbio.2006.09.045

Wills AE, Gupta R, Chuong E, Baker JC. 2014. Chromatin immunoprecipitation and deep sequencing in Xenopus tropicalis and Xenopus laevis. *Methods* **66**:410–421. doi:10.1016/j.ymeth.2013.09.010

Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**:15. doi:10.1186/s13059-017-1382-0

Xu M, Li Y, Du J, Lin H, Cao S, Mao Z, Wu R, Liu M, Liu Y, Yin Q. 2018. PAX3 Promotes Cell Migration and CXCR4 Gene Expression in Neural Crest Cells. *J Mol Neurosci* **64**:1–8. doi:10.1007/s12031-017-0995-9

Zalc A, Sinha R, Gulati GS, Wesche DJ, Daszczuk P, Swigut T, Weissman IL, Wysocka J. 2021. Reactivation of the pluripotency program precedes formation of the cranial neural crest. *Science* **371**:eabb4776. doi:10.1126/science.abb4776

**Author contributions:**

Conceptualization: AK, SS, LP, AHMB
Methodology: AK, MA, SS, VK, SMR, HA, LP, AHMB
Investigation: AK, MA, SS, VK, SMR, 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.

**Funding:**

This project received 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)

**Acknowledgements**

The authors are grateful to Drs. A. Zinoviev, G. Schlosser, I. Adameyko and T. Walter for insightful scientific discussions and comments on the manuscript. We thank C. Lantoine for animal husbandry, Q. Thuillier for technical assistance, and present and past members of the Monsoro-Burq lab for their support. We thank J. Briggs for help in single cell sequencing. We thank J.L. Plouhinec for preliminary RNAseq analysis. High-throughput sequencing, except for single cell sequencing, used 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 ».

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

**List of Supplementary Materials**

- Figure 1 – figure supplement 1
- Figure 2 – figure supplements 1, 2, 3, 4 and 5
- Figure 4 – figure supplements 1 and 2
- Figure 5 – figure supplement 1
- Figure 6 – figure supplement 1
- Supplementary File 1
    a. Tables S1, S2 and S3
    b. Detailed description of NC clusters
    c. Tables S4, S5, S6, S7 and S8
    d. Supplementary Matdrials and Methods

## 2.2. Neural border and neural crest gene regulatory network: unpublished data

In this chapter, we present our unpublished results related to the paper above. First, we will describe the improved bioinformatics pipeline used for reads alignment and further analysis in our study. This approach differs from the one previously described in the Briggs et al. paper. Next, we will present the results of our TFAP2a ChIP-seq experiments and discuss the identified targets and their potential functional implications. We will also report on the results of our analysis of gene expression changes following Zic1 MO-depletion and RNA-seq.

### 2.2.1 A new pipeline description and dataset benchmarking

In order to process new scRNA-seq data and reprocess previously published data (Briggs et al, 2018) we decided to select a different aligner and scRNA pipeline due to several factors. The first one is that Bowtie (Langmead et al., 2009) which is used in the InDrops pipeline (Klein et al., 2015) was not designed to align transcripts to the genome, and therefore it is not splice-aware. An RNA-seq read aligner that takes splicing into account should prioritize aligning reads to exons rather than introns, and should attempt to identify and align to downstream exons if possible. Using Bowtie to align reads to a transcriptome raises concerns about selecting the appropriate transcript. Therefore, STAR has been chosen, because it is one of the most effective RNA-seq aligners and has a higher accuracy for both correct and incorrect alignments compared to other aligners (Baruzzo et al., 2017). Secondly, STAR is fast and automatically handles preprocessing tasks such as removing low-quality bases and adapters (Dobin et al., 2013). However, due to bugs in the pipeline (which has not been updated in the past 4 years) and the difficulty of integrating another aligner into the pipeline, we have chosen to replace InDrops with the more recent DropEst pipeline (Petukhov et al., 2018). Recent benchmarking showed that DropEst exceeds other options in terms of sensitivity and efficiency (running time and memory use) (Gao et al., 2021). As a result, we got two different versions of the pipeline which use dropEst and STAR for versions v2 and v3 of Indrops platform. For v2 Indrops we run dropTag which generates tagged fastq files for alignment, then we run STAR alignment and dropEst, which estimates counts matrix. For v3 Indrops we firstly use Indrops for demultiplexing and very soft filtering, then we generate tagged files with our script (similar to dropTag without considering reads base quality for barcodes) and then alignment and quantification as in the v2 version (Figure 8A).

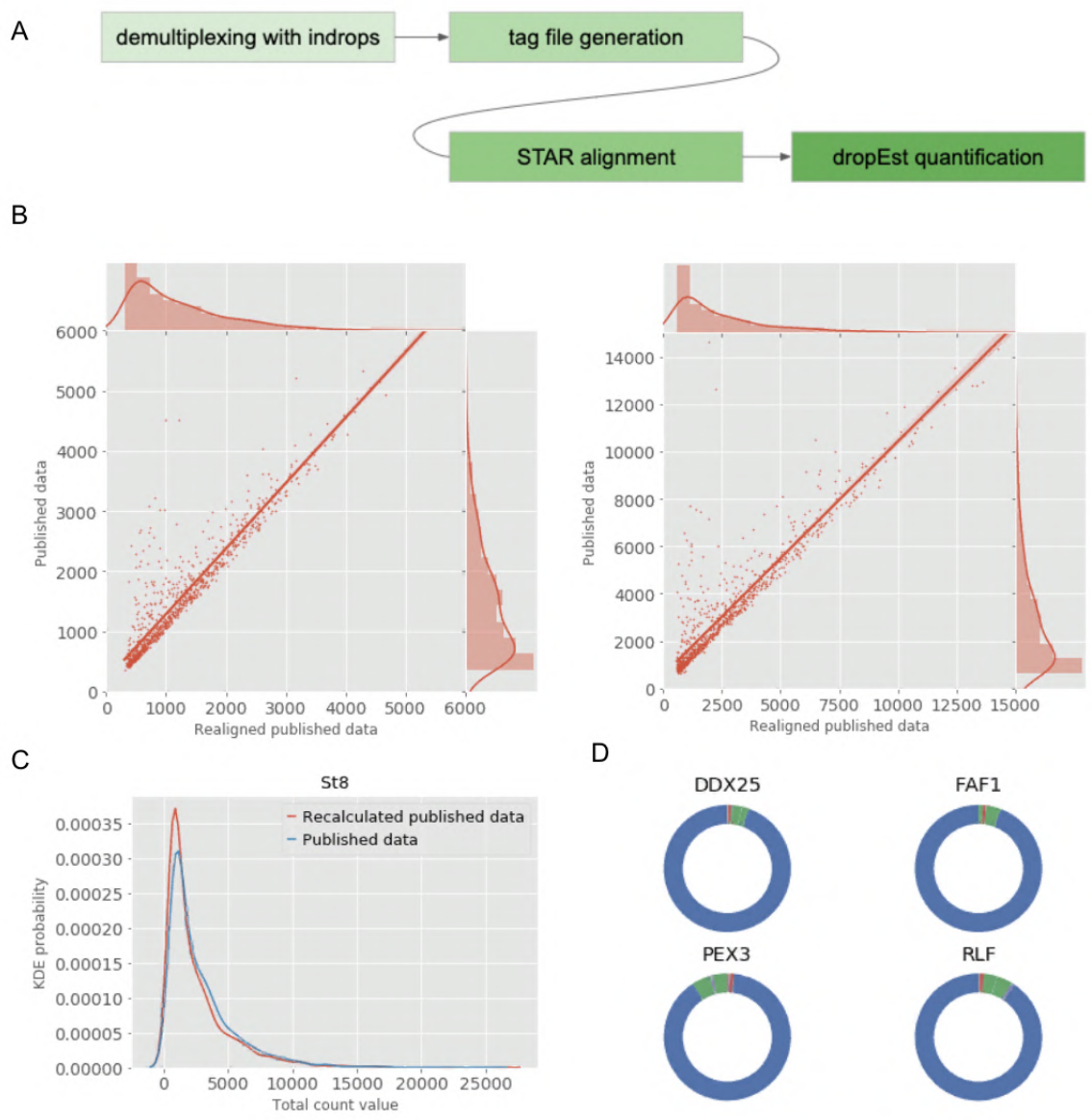



**Figure 8. ScRNAseq data preprocessing and comparison to published data**. A. New pipeline schema (for Indrops v3 data); B. Cell to cell comparison of total counts and genes number for Indrops and our pipeline; C. Total counts KDE in two datasets; D. Expression change for some selected genes which are in common for 964 cells. Blue is for 0 UMIs (no change), green is for 1 UMI.

### 2.2.2 Published data recalculation

To benchmark our pipeline which better suits the RNA-seq data, we took published count matrices from GSE113074 (which are counted using inDrops) and compared it with the same data but processed with our pipeline. We evaluate it in terms of the number of cells, genes per cell, and total counts per cell. After filtration with min_counts=650, min_genes=300 and min_cells=3 we got a total of 1069 and 1160 cells respectively for our pipeline and Indrops for stage NF-8 v2 Indrops data from GEO. From Figure 8B we can see that the results obtained with the new pipeline are quite similar to the published matrices, which meets our criteria in terms of sensitivity. From Figure 8C we see that the probability of a certain total count in two datasets is very similar. Thus, we see that our new pipeline is only slightly inferior to the Indrops pipeline in terms of the number of cells, genes and UMIs. We also checked dynamics in expression change for certain genes. We observed that 90-95% of cells did not change the expression of four genes. A similar situation is observed with the data from stage 12, but with some fragmentation. After the same filtration, we get 7916 cells for our pipeline and 7959 for published Indrops data.

### 2.2.3 Additional experimental validation

From the analysis of the whole Ectoderm connectome (10085 gene connections) and found TFAP2a among the most significant ones. Tfap2a plays a role in neural crest induction by participating in the balance of Bmp and Wnt signaling in the neural plate border and is necessary and sufficient for this process to proceed (Wang et al., 2011). Thus, we aimed to expand our understanding of the role of Tfap2a in neural crest induction and identify the direct targets of this important transcription factor. Using ChIP-seq we retrieved 682 direct targets of TFAP2a expressed in the NC dataset, among the 1393 targets bound by TFAP2a in the whole embryo, including NB signature genes *pax3*, *sox9*, *myc* and *tfap2a* itself suggesting the key role in the ectodermal cell types development during gastrulation (Figure 9; Table 2). Also, we depleted selected gene products in vivo using previously validated antisense morpholino oligonucleotides (MO) designed against Zic1 (Figure 10). The differential analysis found decreased expression of 1103 genes in Zic1 morphant NB (Figure 9). This confirmed that this NB specifier is essential to activate a large NB/NC gene signature, and provided the most complete list to date of Zic1 indirect targets in the pre-migratory neural crest (Table 1). We found that GRNBoost2 predicted 1136 genes linked to Zic1, of which 188 were significantly changed in Zic1 MO cells.

| 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 |

| 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 |

**Tables 1 and 2. Table 1.** Top decreased genes in Zic-MO samples with the absolute expression change > 1000 (on the left); **Table 2.** Top MACS2 scored direct targets of Tfap2a.

Interestingly, *cyp26a1* was connected to *zic1* in the predicted network and was significantly decreased in Zic1 MO. This supports the previous characterization of *cyp26c1* as an immediate-early target of Pax3 and Zic1 (Plouhinec et al., 2014), and may relate to potential roles of Zic1 and Pax3 in anteroposterior patterning through regulation of retinoid signaling (Uehara et al., 2007). Overall, the experimental validation has not only validated the accuracy of the bioinformatic predictions, but has also significantly expanded our understanding of the roles of TFAP2a and Zic1 during neural crest induction and development. The experimental results have identified 105 direct targets of TFAP2a and 188 indirect targets of Zic1 (some of which are known), providing new insights into the mechanisms underlying neural crest induction and development.

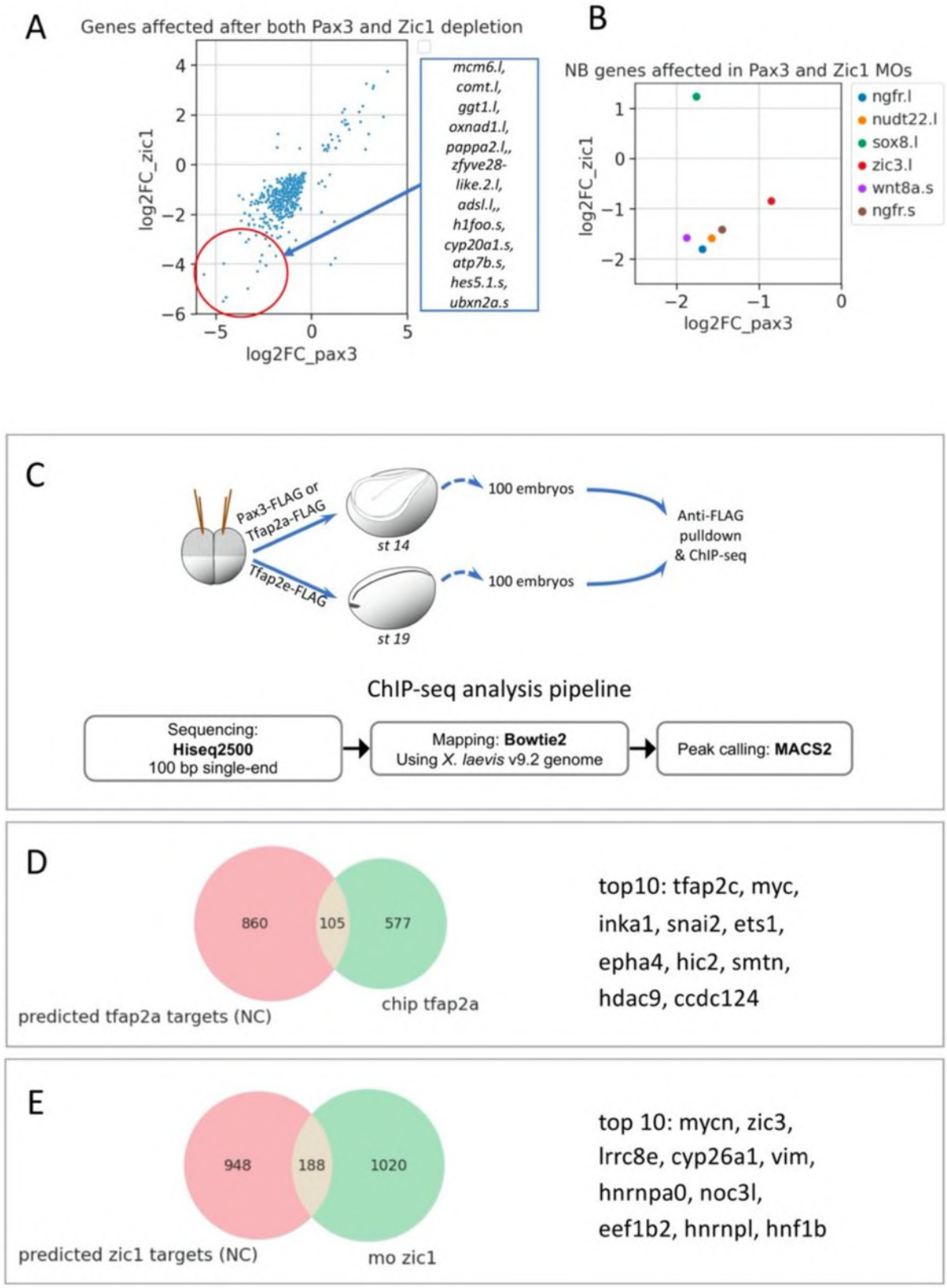



**Figure 9. Direct gene regulation by Tfap2a, Zic1, and Pax3 during the NC induction.** In order to reveal direct targets and validate scRNA-seq prediction, we performed ChIP-seq for

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.

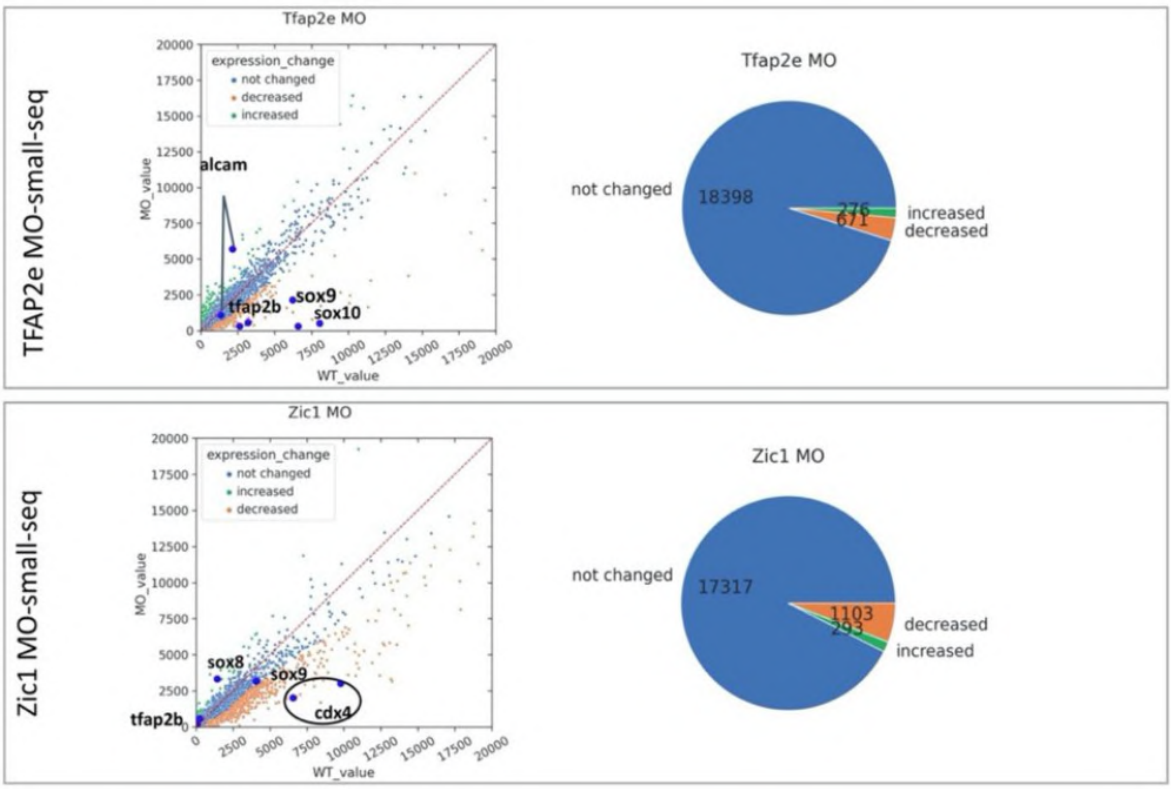



**Figure 10. MO experiments overview.** We performed MO knockdown experiments for the main genes from the predicted networks for the NC dataset. The figure depicts general statistics for changed and unchanged genes after Tfap2e, and Zic1 depletion compared to controls

# 2.3 Investigating the role of miRNAs miR-196a and miR-219 in the development of neural crest in *Xenopus*

Alice M. Godden, Marco Antonaci, Nicole Ward, Méghane Sittewelle, **Aleksandr Kotov**, Anne-Hélène Monsoro-Burq, Grant N. Wheeler

*In progress.*

## 2.3.1 Introduction

MicroRNAs (miRNAs) are short, single-stranded, non-coding RNAs, approximately 22 nucleotides long (Alberti and Cochella, 2017; Lee et al., 1993). miRNAs can be found in intronic regions of the genome, and as such are processed from the introns. They can also be found as independent genes while a number of miRNAs can be transcribed in one transcript and then processed individually. MicroRNAs are highly conserved and are abundant between species with many orthologues (Bartel, 2004).

MiRNAs have been found to have roles in the development of invertebrates like the worm and fruit fly (Chandra et al., 2017) and vertebrate tissues, including chick, mouse, frog and fish (Ahmed et al., 2015; Mok et al., 2017; Ward et al., 2018). MiRNAs have been shown to be involved in the regulation of the NC specification, migration (Figure 11) (Weiner, 2018). Dysregulation of miRNA expression has been linked to various neurocristopathies, suggesting that miRNAs may be potential therapeutic targets for these disorders (Bachetti et al., 2021; Du et al., 2020; Evsen et al., 2020; Schoen et al., 2017). Neurocristopathies are developmental congenital disorders where there is aberrant NC migration, specification, or differentiation (Gouignard et al., 2016; Ward et al., 2018). Neurocristopathies include DiGeorge syndrome, Waardenburg syndrome, cranio-fronto-nasal dysplasia (cleft palate), and some cancers such as neuroblastoma and melanoma (Gouignard et al., 2016; Ward et al., 2018). Therefore, studying the role of miRNAs in NC development may provide new insights into the regulatory mechanisms underlying NC biology and may identify potential targets for the treatment of neurocristopathies. In addition, understanding the role of miRNAs in NC development may also reveal new pathways that could be targeted to promote NC-based regenerative therapies. Overall, the study of miRNAs in NC development has the potential to improve our understanding of the complex regulatory networks controlling NC biology and may lead to the development of new therapeutic strategies for NC-related disorders.

Gessert et al. found that several miRNAs, including miR-130a, miR-219, miR-23b, miR200b, miR-96, and miR-196a, play a role in NC and eye development in Xenopus laevis (Gessert et al., 2010). Using a morpholino approach, they demonstrated that miR-130a, miR-219, and miR-23b are necessary for normal eye development, while the knockdown of miR-200b, miR-96, and miR-196a leads to craniofacial defects and other NC-associated phenotypes. Ward et al. studied miRNAs using Xenopus laevis embryonic organoids (also known as "animal caps") induced to become NC or neural tissue. Small-RNA sequencing and differential analysis were performed to identify miRNAs specifically expressed in NC-induced animal caps. The most abundant miRNAs detected were miR-219, miR-196a, miR-218-2, miR-10b, miR-204a, miR-130b/c, miR-23, and miR-24, with miR-219 being the most enriched miRNA in NC-induced animal caps, followed by miR-196a (Ward et al., 2018). The purpose of this research was to confirm if miR-196a and miR-219 play a necessary role in the development of *Xenopus* NC, using MO-mediated KD to knock down these microRNAs. Additionally, this study aimed to investigate the level at which miR-196a and miR-219 influence NC development.

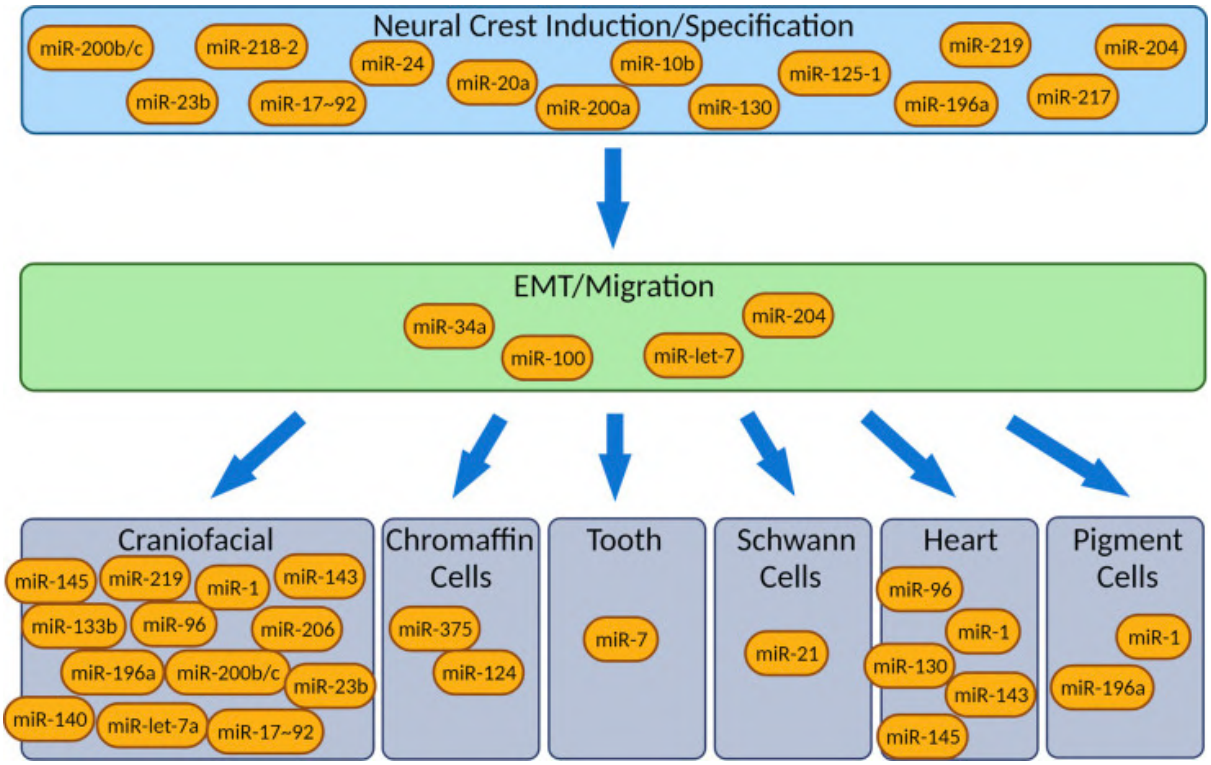



**Figure 11. MiRNAs during NC development** (Antonaci and Wheeler, 2022). The roles of different microRNAs in the development of NC, from induction to differentiation.

### 2.3.2 Results

Previously our colleagues have shown miR-219 and miR-196a to be expressed in NC and neural tissue in the Xenopus embryos (Godden et al., 2021; Ward et al., 2018). To determine their function in NC development we designed MO that was designed to knock down the mature miRNA (and be complementary to it) (Flynt et al., 2017). To profile and investigate the role of miRNAs in the development of Xenopus NC and NPB, whole-mount in situ hybridization experiments were carried out on miRNA KD embryos (Figure 12). Loss of miR-219 led to the strong loss of NC expression of Snail2 and Sox10, an expansion of NPB marker Pax3 in the superficial ectoderm, and a loss of Xhe2, an HG marker (develops from ectoderm) (Figure 12). Loss of miR-196a led to similar phenotypes, with loss of NC with stronger loss of Sox10 compared to miR-219 KD, and a subtle reduction in Pax3 expression.

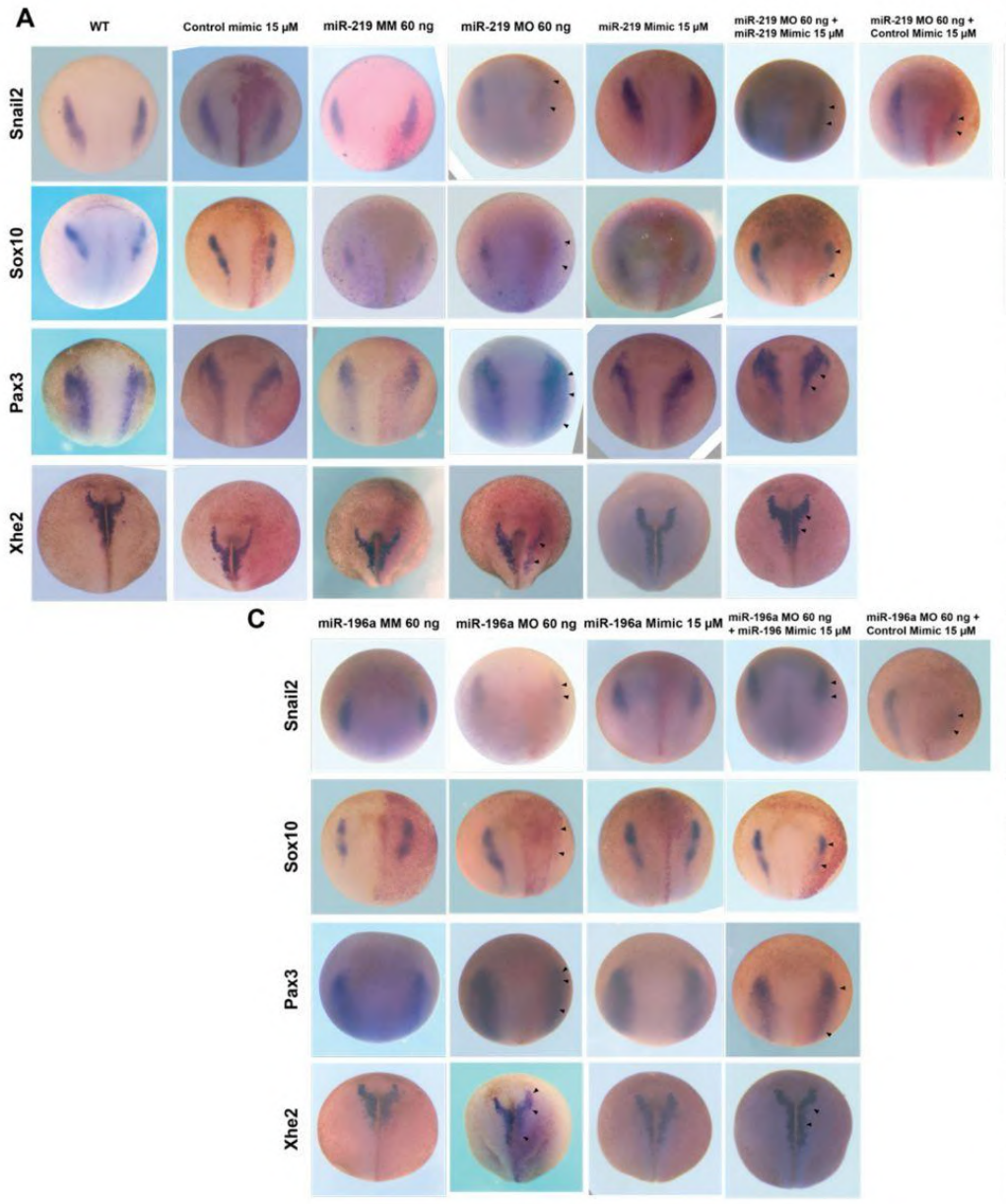



**Figure 12. Whole-mount in situ hybridization for miRNA MO samples (196a and 219).** The effects of the miRNAs-219/196a on Snai2, Sox10, Pax3 and Xhe2.

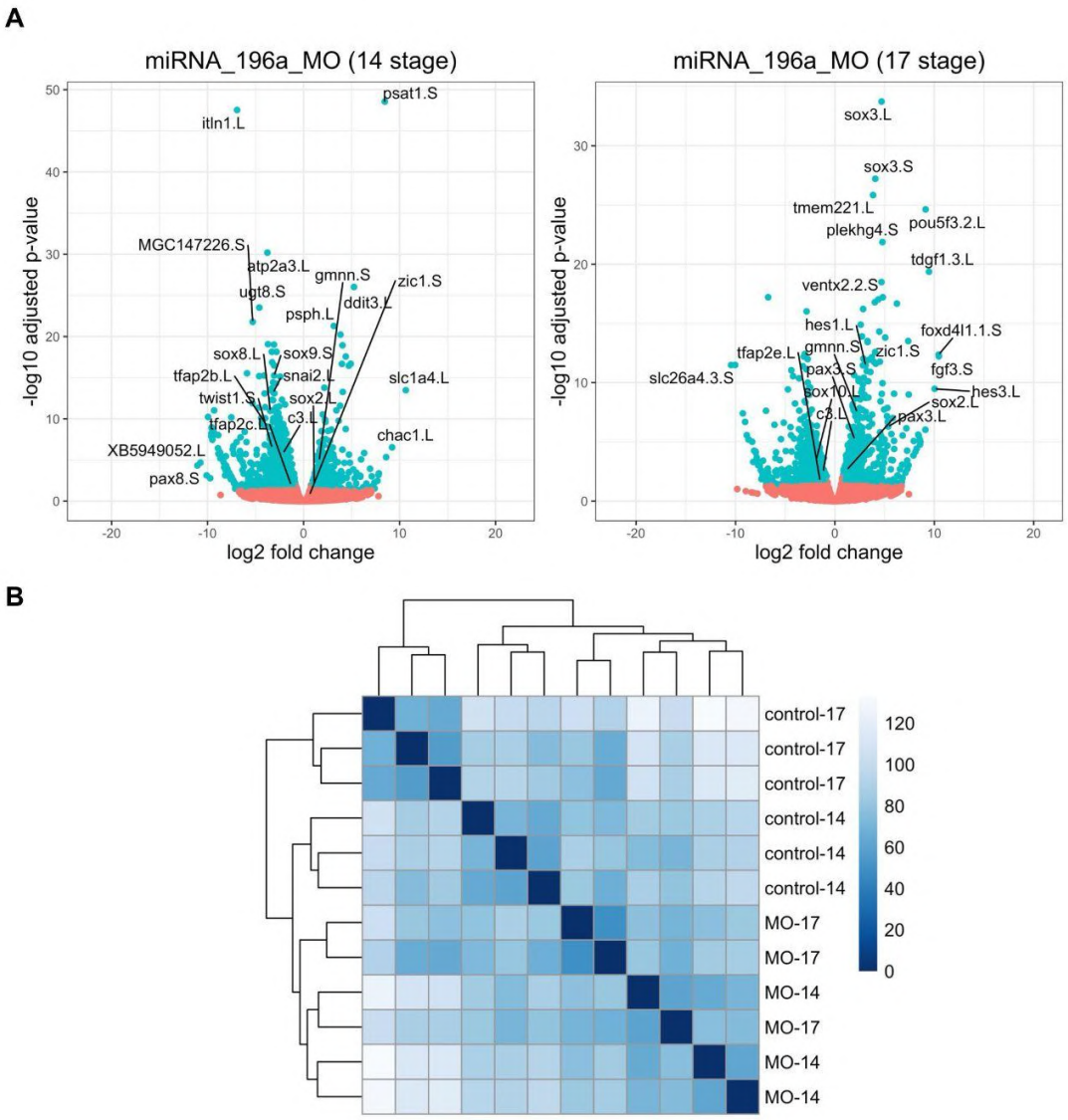



**Figure 13. MiRNA 196 top DE genes and selected NC-related genes.** A. Volcano plots for miRNA_196a-MO for stage 14 and stage 17; B. Matrix of distances between different samples.

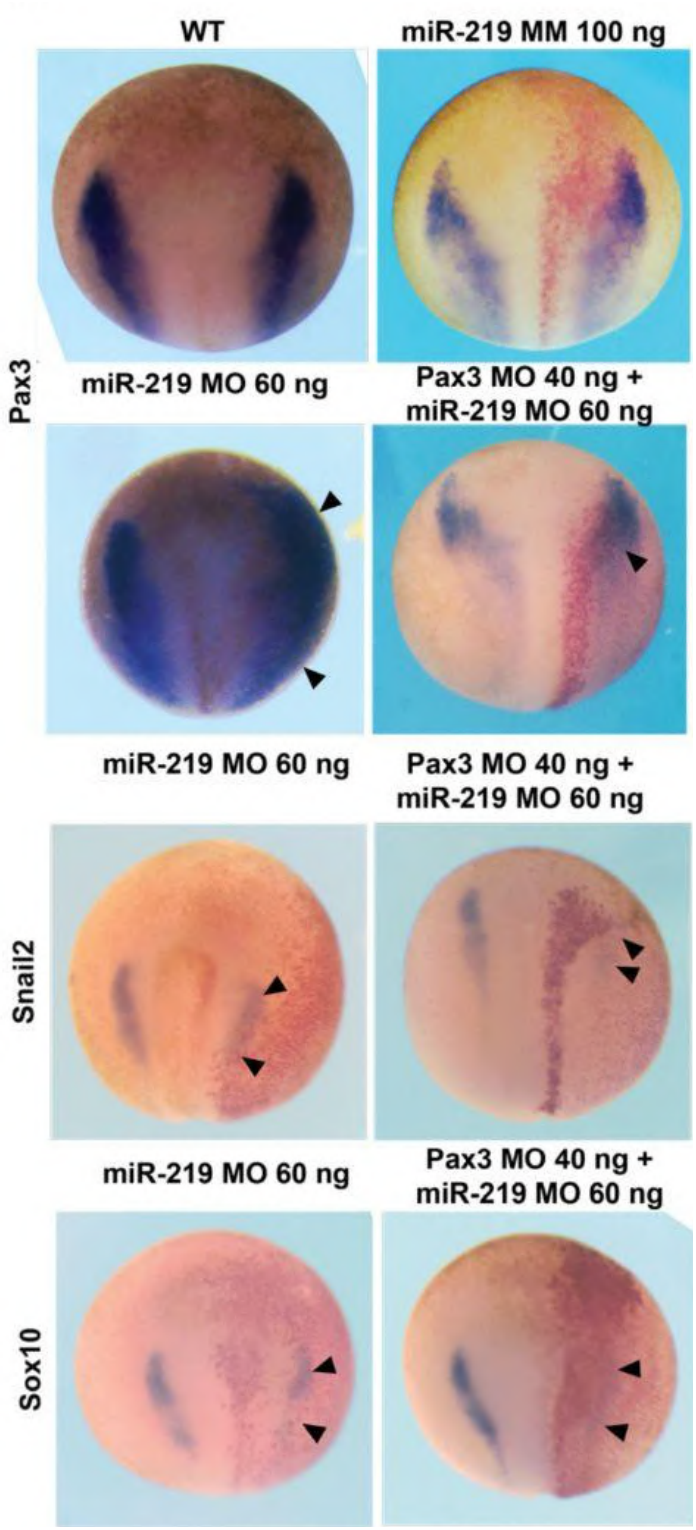



**Figure 14. Characterization of Sox10, Snail2, and Pax3.** The effect of Pax3MO+miRNA-219MO.

Through RNA-sequencing on dissected NC tissue that had been treated with MOs for miR-196a and miR-219, it was revealed that miRNAs may have a role in the development of NC and surrounding tissues. A strong reduction in expression of NC markers: *foxd3*, *twist1*, *sox8*, *sox9*, *sox10*, and *snail2* can be seen, following miRNA-KD, more significantly with miR-196a KD. Other genes that were knocked down were *pax8* genes and *itln1* and others (see Figures 13, 15). Genes that are enriched are more likely to be direct miRNA targets as miRNAs function to bind to complementary 3' untranslated regions and post-transcriptionally repress gene expression (Rooda et al., 2020). Neural markers *sox2*, and *sox3* were significantly enriched. Neural plate border markers were also enriched (*gmnn*, *hes3*, *pax3*, and not significantly *zic1*) and also pluripotency marker *pou5f3* was enriched. MiRNA expression in Xenopus has been characterized previously (Ahmed et al., 2015). Specifically, miR-219 and miR-196a have been shown in the lab to be expressed in developing Xenopus NC (Ward et al., 2018). Using an enrichment map we observed a cluster of Nc-related GO terms (Figure 17).

Functional characterization of Sox10, Snail2, and Pax3 were chosen following MO-mediated miRNA KD, and rescue for miR-196a and miR-219. Xenopus laevis embryos were injected at the 4-cell stage of development into one dorsal blastomere with 300 pg lacZ plus MO, miRNA mimic, or combination. For all markers, no phenotype was seen following injection of embryos with control miRNA mimic (cel-miR-39-3p), miR-196a mimic, miR-219 mimic, and mismatch (MM) MO. Here we report the first-time synthetic miRNA mimics have been used to rescue phenotypes seen by the whole mount in situ hybridization. Sox10 depletion is known to lead to a rise in NC cell apoptosis and a reduction in cell proliferation (Honoré et al., 2003). Snail2 is required for the induction of NC; and is anti-apoptotic (Klymkowsky et al., 2010). This suggests that miR-196a and miR-219 KD, which reduce the expression of Sox10 and Snail2, prevent the proliferation and maintenance of a pool of NC cells.

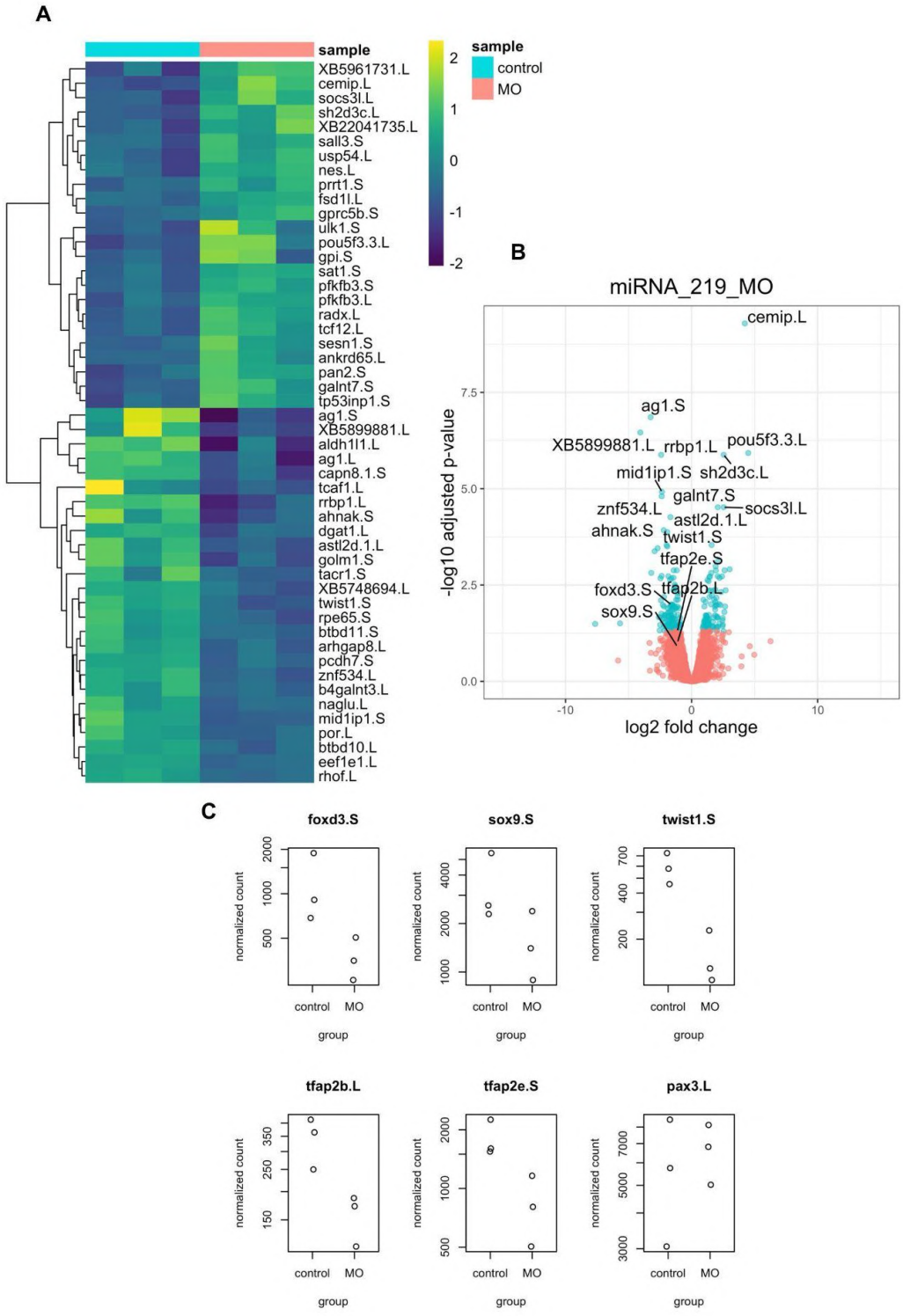



**Figure 15. Result for DE analysis for 219 miRNA MO.** A. Heatmap for the most DE genes; B. Volcano plot for miRNA 219 MO sample; C. NC markers expression change.

MiR-196a KD led to a strong expansion of Pax3 expression (Figure 13). Misexpression of Pax3 can also lead to a reduction in Snail2 expression (Hong and Saint-Jeannet, 2007). The expanded phenotype could be indicative of miR-196a targeting something upstream of NC and neural plate development, potentially placodal. Potentially, upstream Wnt, FGF of BMP signalling could be affected leading to the expansion of Pax3 (Pla and Monsoro-Burq, 2018).

To further assess the impact of miR-219 MO KD on Pax3 expression a Pax3 MO was used to rescue the expanded Pax3 phenotype generated following miR-219 MO KD. Pax3 MO dose-response impact on Pax3 and Snail2 expression data can be seen in Figure 16. Pax3 MO used in this project is as described in (Monsoro-Burq et al., 2005). Pax3 MO reduces the amount of Pax3 expressed, and miR-219 MO increases Pax3 expression, as indicated by black arrows. When used together the expanded Pax3 phenotype is rescued. To test for rescue specificity Pax3 MO and miR-219 MO were trialed to assess the impact on Sox10 and Snail2. All these markers show reduced expression upon miR-219 MO KD, however, when used in conjunction with Pax3 MO this only led to a stronger reduction in gene expression (Figure 14). This suggests that the Pax3 MO rescue effect on expanded Pax3 phenotype is specific. Furthermore, it suggests that miR-219 is potentially acting upstream of the neural plate border in the development of Xenopus NC.

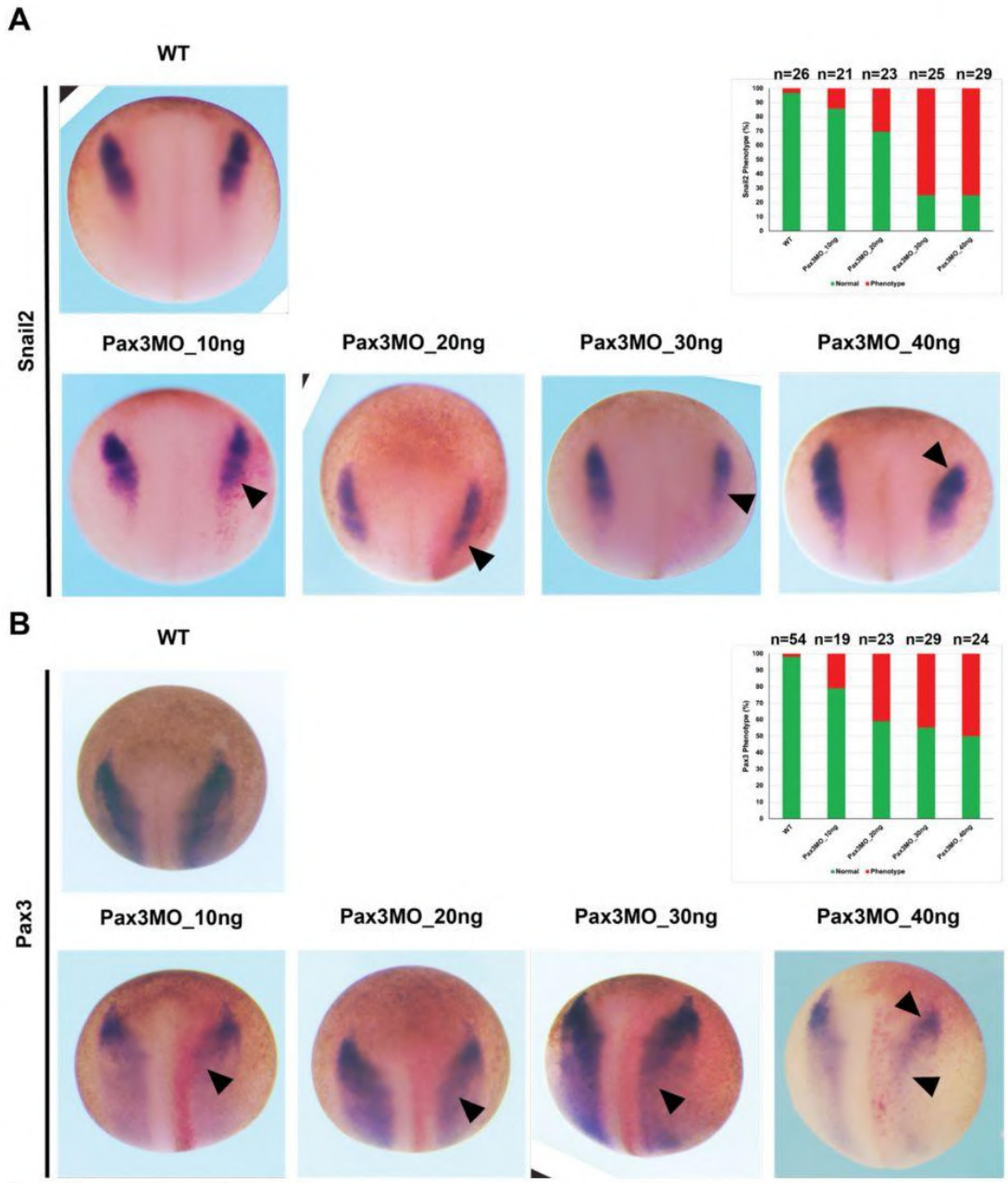



**Figure 16. MO dose-response impact on Pax3 and Snail2 expression data.**

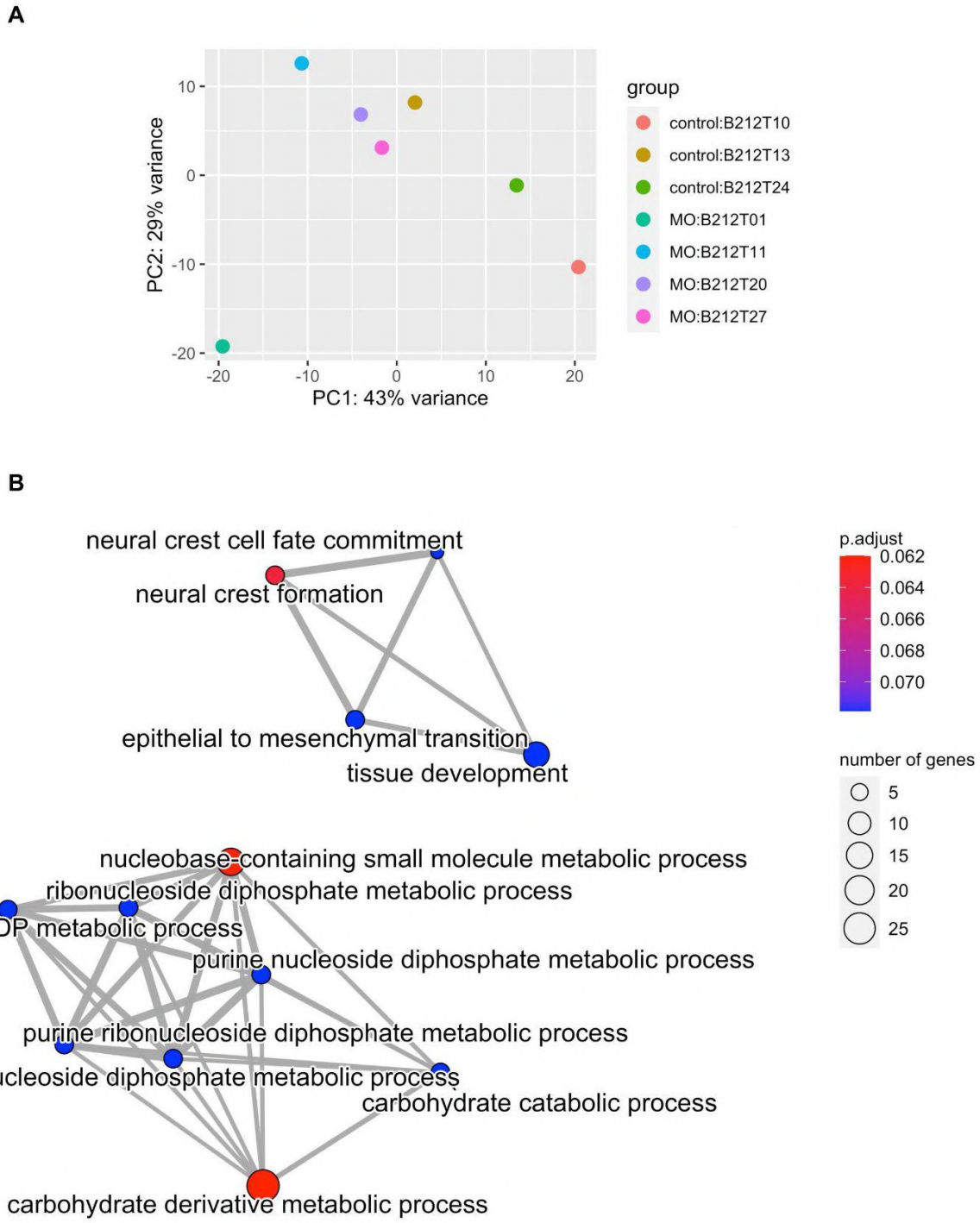



**Figure 17. MiRNA 219 MO PCA and Enrichment map.** A. PCA for 219 MO sample, for the postprocessing we removed one outlier; B. Enrichment map organizes enriched GO terms into a network with edges connecting overlapping gene sets. In this way, we obtained a cluster of NC-related terms: NC formation, commitment, EMT, and tissue development.

Sox2 increase in miR-196a KD (Figure 13) could contribute towards the reduction in NC marker expression seen in Figure 12; previous work in avian models suggests Sox2 misexpression can inhibit NC formation (Wakamatsu et al., 2004). Other work implies murine Sox2 is implicated in the regulation of EMT; significant as this is imperative for the development of cranial features (Mandalos et al., 2014). NC cells are essential in the development of the eye and the lens (Grocott et al., 2020). Previous work in the lab has shown small eye phenotypes in miRNA knockout embryos (Godden et al., 2021); this supports the disruption to Pax6 phenotype observed in miR-219 KD where the top arrow is pointing to the lens field that appears to be reduced in expression (Figure 18).

### 2.3.3 Discussion

We have presented novel work into developing miRNA rescue experiments and functional characterization of miR-196a and miR-219 in Xenopus NC development. The loss of miR-219 and miR-196a in Xenopus embryos resulted in a reduction of NC expression for Snail2 and Sox10, as well as an expansion of the NB marker Pax3 in the ectoderm and a loss of Xhe2. RNA-sequencing experiments on NC tissue treated with MOs for miR-219 and miR-196a revealed a decrease in NC marker expression, with a stronger effect seen with miR-196a KD. Rescue experiments using synthetic miRNA mimics showed that miR-196a and miR-219 are required for the proliferation and maintenance of NC cells

The expanded *pax3/zic1/gmnn/hes/sox2/3* expression at stage 17 together with increased expression of multipotency markers *pou5f3* and *ventx2* following miR-196a KD may suggest that miR-196a targets the NC development and may switch cell fate into more multipotent NB-like state (Figure 19). The KD of miRNA-219 led also to increasing of multipotency factors *pou5f3*, but without impact on the NB markers, suggesting that it this contributes to an increase in multipotency without turning on the NB program. From the data presented it is hypothesized that the miRNAs miR-196a and miR-219 are having differing roles in the development of NC, affecting different induction and specification events. It is possible that the miRNAs may be targeting or directly or indirectly Wnt signaling and other events at the development of the neural plate border. Finally, we showed both miR-219 and miR-196a are required for maintenance of NC cells in *Xenopus* embryos.

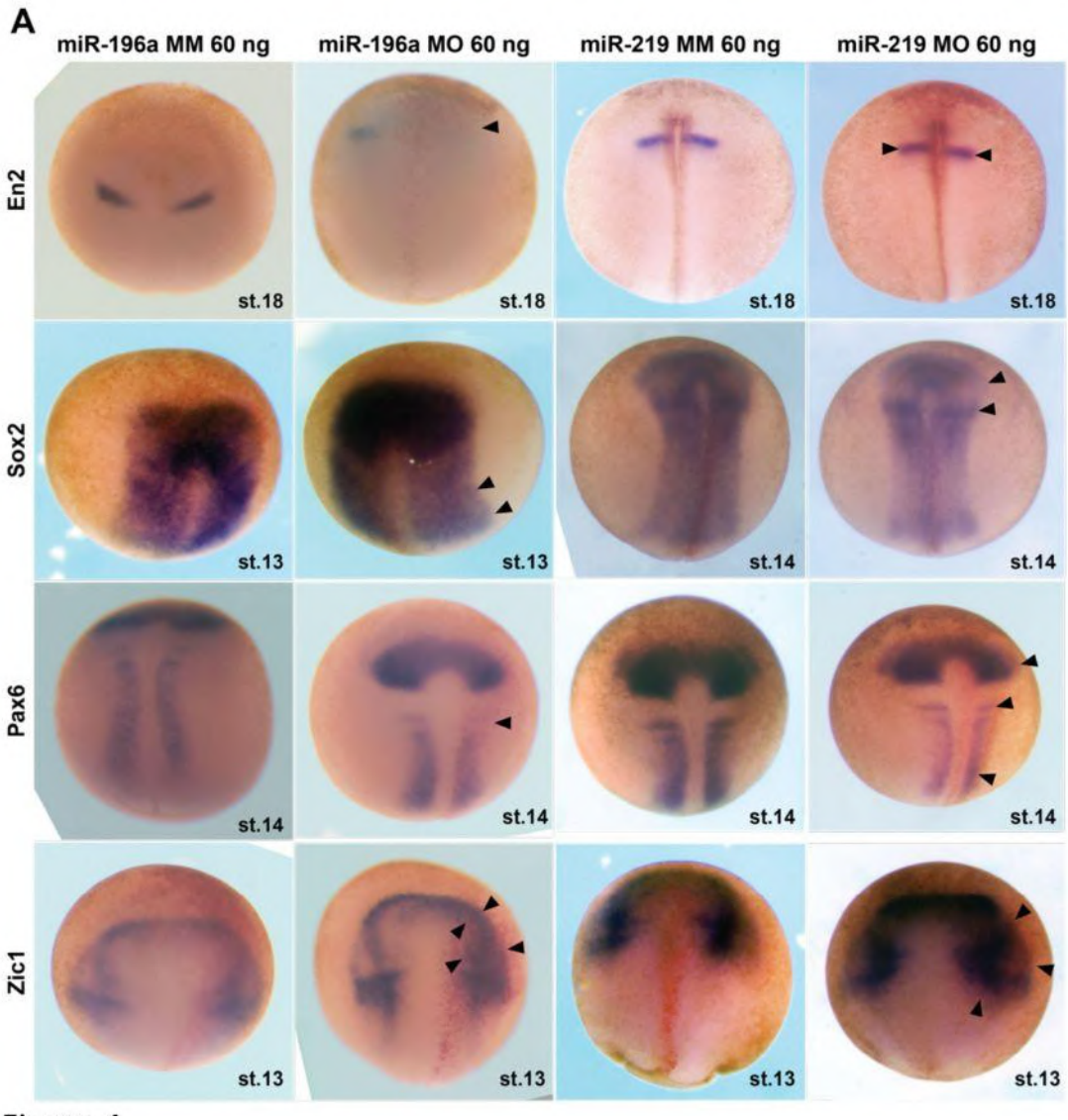



**Figure 18. In situ images for 196a MO.** The effects of miRNA-196aMO on Sox2, Pax6 and Zic1.

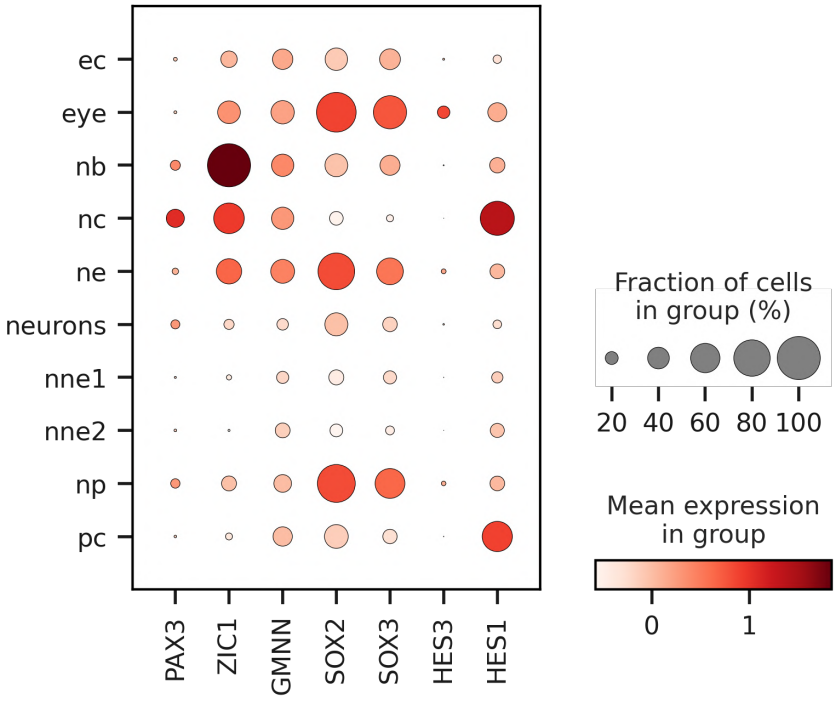



**Figure 19. Gene expression for up-regulated genes in MiRNA-196a KD in the Ectoderm dataset, scRNA-seq data from (**Kotov et al., 2022**).** Ec – ectoderm, eye – eye primordium, nb – neural border, nc – neural crest, ne – neural ectoderm, nne – non-neural ectoderm, np – neural plate, pc – placodes.

## 2.3.4 Materials and methods

***Xenopus husbandry***

All experiments were carried out in accordance with relevant laws and institutional guidelines at the University of East Anglia, with full ethical review and approval, compliant to UK Home Office regulations. To obtain *Xenopus laevis* embryos, females were primed with 100 units of PMSG and induced with 500 units of human chorionic gonadotrophin. Eggs were collected manually and fertilized in vitro. Embryos were de-jellied in 2% L-cysteine, incubated at 18°C and microinjected in 3% Ficoll into 1 cell at the 2-cell stage in the animal pole with 5 nL of enhancer reporter plasmid at 400 ng/μL or GFP capped RNA as control. Embryos were left to develop at 23°C. Embryo staging is according to Nieuwkoop and Faber normal table of Xenopus development. GFP/LacZ capped RNA for injections was prepared using the SP6 kit, 50 pg was injected per embryo.

***Embryo injection***

Embryos were injected using a 10 nL calibrated needle. MO dose was optimized to 60 ng for miRNAs; MO and lacZ were injected at 4 cell stage of embryo development into the right dorsal blastomere. miR-219 mimic was used from (Qiagen, 339173 YM0047076-ADA, MIMAT0000276); I-miR-219a-5p miRCURY LNA miRNA Mimic, compatible with xtr-miR-219 sequence: 5'UGAUUGUCCAAACGCAAUUCU. For miR-196a: (Qiagen, 339173 YM00470616-ADA, MIMAT0000226); I-miR-196a-5p compatible with xtr-miR-196a sequence: 5' UAGGUAGUUUCAUGUUGUUGGG. A negative control miRNA mimic recommended by Qiagen was used (Qiagen, 331973 YM00479902-ADA); Negative control (cel-miR-39-3p), sequence 5'UCACCGGGUGUAAAUCAGCUUG. To rescue expanded Pax3 phenotypes Pax3 MO (Table 1) was optimized at 40 ng to provide a reduction in Pax3 expression. This was co-injected with miR-219 MO 60 ng. As this isn't possible in one injection Mos were made up so Pax3 MO final concentration was 20 ng and miR-219 MO was 30 ng, two injections into the embryo at 4 cell stage into 2 blastomeres then gave a concentration of Pax3 MO 40 ng and miR-219 MO 60 ng. For miR-219 MO MM two injections of 50 ng gave a final dose of 100 ng MO.

### *RNA sequencing*

For RNA-sequencing embryos were injected into one blastomere at 4 cell stage with one of four MO's (miR-219, miR-219MM, miR-196a or miR-196aMM) into one dorsal blastomere to target neural and NC tissue in one side of the embryo only. Embryos were left to develop until stage 14. One group underwent WISH to check NC genes were knocked down (data not shown) and the other group underwent NC dissections and RNA was extracted. Three replicates were collected for each condition (MO, MM, and non-injected control). RNA samples underwent quality control using a Bioanalyzer and q-RT-PCR was used to further validate the KD of NC-specific genes in MO-injected samples. Samples were then processed to Illumina sequencing. RNA sequencing was carried out on miR-196a and miR-219 control, MM MO and MO-treated samples all of which contained three replicates each (except miR-219 MO which contained four). All samples were tested prior to sequences using PCR (data not shown) to detect expected gene expression. RNA samples were processed for library preparation and 50bp paired-end sequencing on the HiSeq High Output run mode PE100 for sequencing.

### *Data analysis*

Reads were mapped to the X. laevis v10.1 genome assembly using STAR (v.2.7.3a) (Dobin et al., 2013). Differential expression analysis was carried out using DESeq2 (v.1.32.0) (Love et al., 2014) in R (v.4.1.1). Genes with an adjusted p-value below 0.05-0.15 were considered significant and were reported by the workflow. The gene model used in the DE bioinformatic analysis was X. laevis (NCBI v10.1). For GO enrichment analysis of a DE genes we used ClusterProfiler (v.4.0.5).

# 2.4 Relationships of Pax3, Dlx3 and Zic1 during gastrulation in the ectoderm

Subham Seal, Haris Bin Fida, **Aleksandr Kotov**, Gerhard Schlosser, Anne-Helene Monsoro-Burq

*In progress.*

## 2.4.1 Introduction

The neural border (NB), being situated in the plane of the ectoderm, sandwiched between the neural (NE) and the non-neural (NNE) ectoderm, does not have defined boundaries at early stages. It is defined as the intersection of multiple genes expression patterns, such as Pax3, Zic1, Dlx3, Msx1, Tfap2a, and others – critical TFs for the establishment of the NC/PC formation which are called NB specifiers because they play a vital role in the initial formation of the neural border zone (de Crozé et al., 2011; Luo et al., 2001; Mizuseki et al., 1998; Monsoro-Burq et al., 2005; Sato et al., 2005). Although their individual functions are rather well-studied, very few studies have explored the positive or negative cooperation between the different genes that define NE, NB, and NNE (Seal and Monsoro-Burq, 2020). We thus have planned to study the targets of important NB genes Pax3, Zic1, and Dlx3, and the cooperation between them during NB fate specification.

Our main aims are to look at the transcriptional targets of each of these NB specifiers and whether there is any dependency between the activity of these genes when they are co-expressed in vivo. For this purpose, we combined two different paradigms: gain-of-function/knockdown of the selected genes and global translation inhibition. The gain-of-function experiments involve increasing the amount of expressed gene or protein while knockdown entails a decrease in their amounts. Generally, in our Xenopus system, we use mRNA for gain-of-function and morpholino for knockdown. To add a temporal control for the gain-of-function, we used mRNAs containing GR (glucocorticoid receptor) constructs. These mRNAs produce fusion proteins that remain in the cytosol but are nuclearly imported upon addition of dexamethasone, a steroid hormone analog (Kolm and Sive, 1995). We used a moderate gain-of-function to limit the drastic consequences of gene overexpression and off-target effects. Alternatively, morpholino-mediated depletion involves the use of small, stable oligomers called morpholinos to specifically target and inhibit the expression of a particular gene. The antisense morpholinos bind to single-stranded mRNA and prevent their translation. Our motivation for the second paradigm, inhibition of global protein translation using

cycloheximide, was to avoid serial transcriptional events, due for example to the secondary activity of a transcription factor itself a target of one of our factors of interest (i.e. avoiding indirect target). Under this condition, our transcription factors of interest can only interact with the cofactors already present and control transcription of targets in a stage-specific manner: this allows us to precisely look at the immediate-early targets of these transcription factors. Finally, we selected three stages of interest: stage 11 for NB induction, stage 12.5 for NC/placode induction, and stage 14 for fully specified NB.

We injected 10 nl of the corresponding mix into both cells of 2-celled embryos with either mRNA-GR or MO at the following concentrations:

- Pax3-GR – 30 pg/10 nl
- Zic1-GR – 70 pg/10 nl
- Dlx3-GR – 150 pg/10 nl
- each MO – 20 ng/10 nl

At our stage of interest, we micro-dissected the neural border ectoderm (for high spatial specificity) and subjected it to one of the following treatments (Figure 20):

- U – (for mRNA-GR injections) DMSO for 2h30
- C – (for mRNA-GR injections) Cycloheximide for 30m, followed by Cycloheximide + EtOH for 2h;
- D – (for mRNA-GR injections) Cycloheximide for 30m, then Cycloheximide + Dexamethasone for 2h
- M – (for MO injections) Immediately lysed
- N – (for MO injections) ¾ NAM for 2h30.

To decrease the effects of embryo-to-embryo developmental variability, we dissected the NB from left and right sides of the same embryo and subjected them to paired conditions (for example, C and D or M and N). After the treatment step, the dissected tissue was collected individually and lysed in Trizol. We collected 6 sets of samples for each condition and after conducting quality checks (Bioanalyzer and RT-qPCR), the samples were sent for sequencing.

This set of experiments fulfill the following aims:

- A. To identify the temporal repertoire of immediate-early targets of Pax3, Zic1, and Dlx3 in the NB ectoderm at key stages of its development (stages 11, 12.5, and 14): inspect for stage-specific targets by differential analysis.

- B. The NB transcriptome in absence of one of the genes (using morpholinos), at stage 12.5: scan for stage-specific requirements of NB specifiers.

Further, to test the cooperation between the NB genes, we injected the embryos with a combination of mRNA-GR of one gene and MO targeting another: for example, a combination of Pax3-GR mRNA and Zic1 MO. From these embryos, the micro-dissected neural border ectoderm was subjected to paradigms U, C or D. This fulfills our final aim C.

- C. To identify dependency between NB specifiers by finding the immediate-early targets of each gene in the absence of another one during initial specification (for example, Pax3 targets after depletion of Zic1 or Dlx3 compared to the target list defined in A and B), at stage 12.5, then define the cooperation between NB specifiers during NC/PC fate induction.

The sequencing was performed through Nova-seq Illumina sequencing. The sequencing results will be combined with the previous datasets (Briggs et al., 2018; Kotov et al., 2022) to achieve greater statistical significance. Currently, we are analyzing the data for the second and third aims. While aim C expand the results from aim A and B, it will also help us to study how the different NB transcription factors cooperate. Based on the analyses, we will validate a few of the novel genes identified as important during NB development. Reads were mapped to the X. laevis v10.1 genome assembly using STAR (v.2.7.3a) (Dobin et al., 2013). Differential expression analysis was carried out using DESeq2 (v.1.32.0) (Love et al., 2014) in R (v.4.1.1). Genes with an adjusted p-value below 0.05 were considered significant and were reported by the workflow. The gene model used in the DE bioinformatic analysis was X. laevis (NCBI v10.1). For GO enrichment analysis of a DE genes we used ClusterProfiler (v.4.0.5).

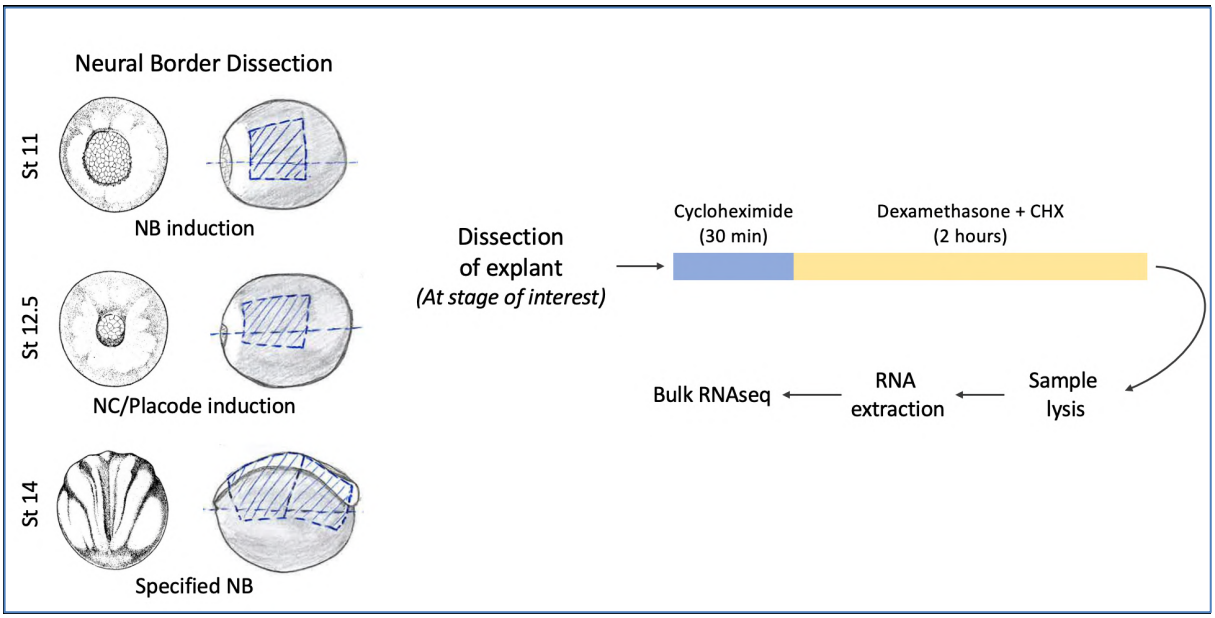



**Figure 20. Design of the experiment.**

### 2.4.2 First results: immediate-early targets of NB genes for stage 11

Using differential expression analysis, we found that the largest number of target genes was observed in Pax3 (293), then Dlx3 (77), and Zic1 (12) (Figures 21, 23), supporting data from (ratio of targets by quantity) (Plouhinec et al., 2014). Expectedly, among Pax3 targets, were found NC markers *snai2*, *twist1*, *sox9*, *foxd3*, *tfap2b* which validates our experimental paradigm. NC formation and EMT GO terms were also linked to the DE list of genes. Most of the genes were upregulated which confirms major activating role of Pax3. Only a few genes were downregulated, among them ionocyte marker *foxi1* which labels immature non-neural ectoderm (Mir et al., 2007). Later on, it is maintained in differentiated cells of the epidermis. Ionocyte is important for the regulation of electrolyte balance in various tissues and they are found in various tissues, including the skin, gills, and gut (Jänicke et al., 2007). The upregulation of *foxi1* may therefore be associated with a fate shift into more non-neural ectoderm fates.

Next, we compared first results of Pax3-GOF and Dlx3-GOF with results of branching analysis in ectoderm dataset in our recent work (Kotov et al., 2022). First, we observed that Pax3-GOF leads to activation of both programs: NE→NP (e.g., *nkx6-2, sox2*) and NE/NB→NC (e.g., *foxd3* and *snai2*) (Figure 22). Furthermore, we observed that Dlx3-GOF activates NNE→NNE2 program, but drastically (all genes are decreased) deactivates NC/NP programs: e.g., *c9*, *cfb* for NC program, and *sall2*, *maneal*, *zeb2* (Figure 24). For PC program, we noted "mixed" regulation: *c9* (which is both in NC and PC programs) was deactivated, while *egflam* was upregulated.

For Dlx3 GO term analysis revealed that Wnt signaling and head development terms are related to the DE list of genes, confirming the work of Sumiyama et al (Sumiyama et al., 2002). Pax3 and Dlx3 both affect 24 genes in common: for example, cranial NC marker *alx1* (downregulated) and marker for early enteric nervous system progenitors and vagal NC *hnf1b* (Kotov et al., 2022) (Figure 25). At stages 12 and 14, for Pax3 we defined fewer targets 43 and 27 respectively. We defined *foxa1, prdm13, tbx18, prdm8 and gsx1* are targets across all 3 stages for Pax3. Collectively those first results support that our experimental paradigm is correct and identifies the respective actions of key transcription factors in the fate choices at the neural border. Because we just received the last series of sequencing results, we will soon be able to explore the entire dataset, in the hope to provide a large landscape of how the neural border derivatives are defined during embryogenesis.

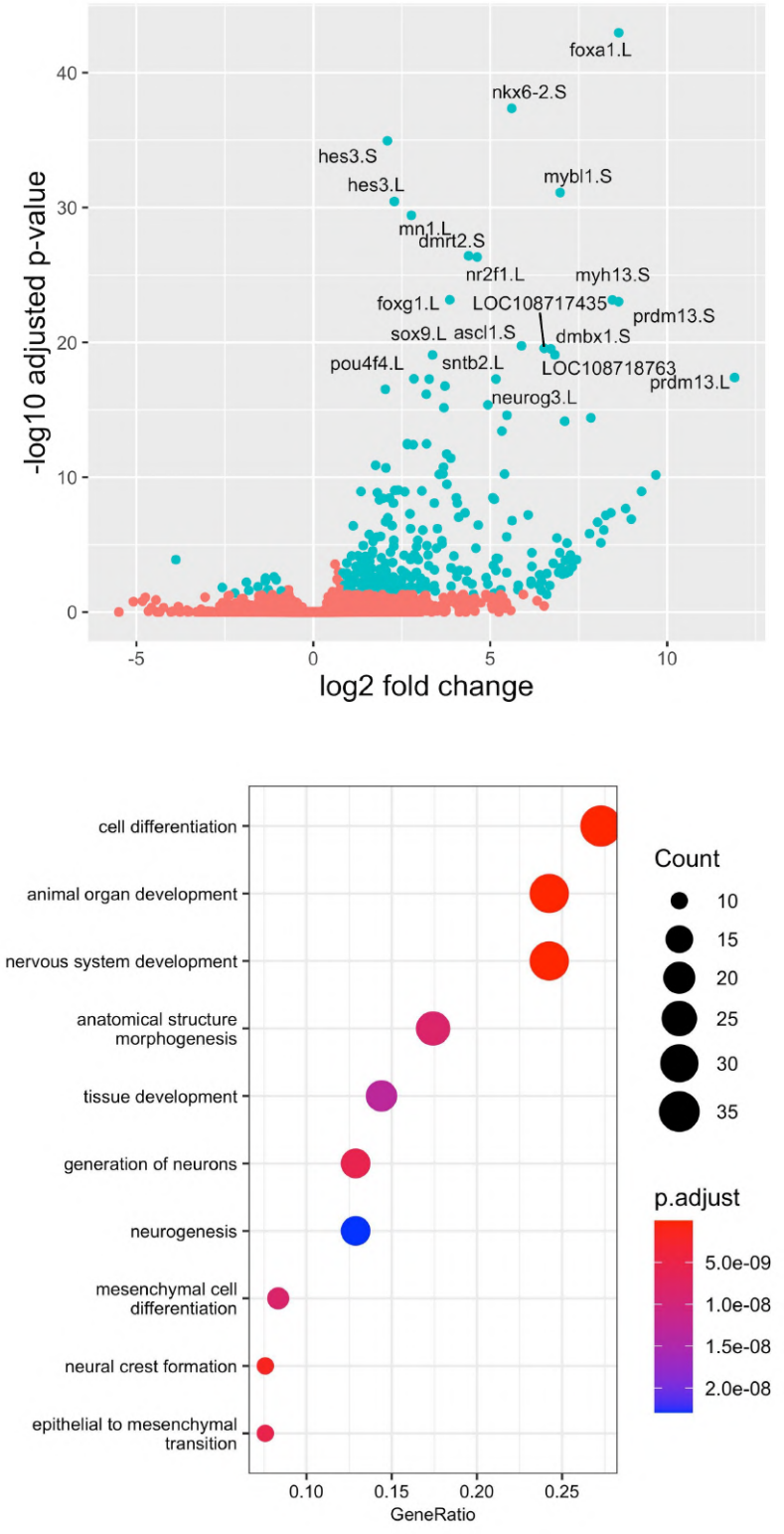



**Figure 21. Pax3 top DE targets and related GO terms.**

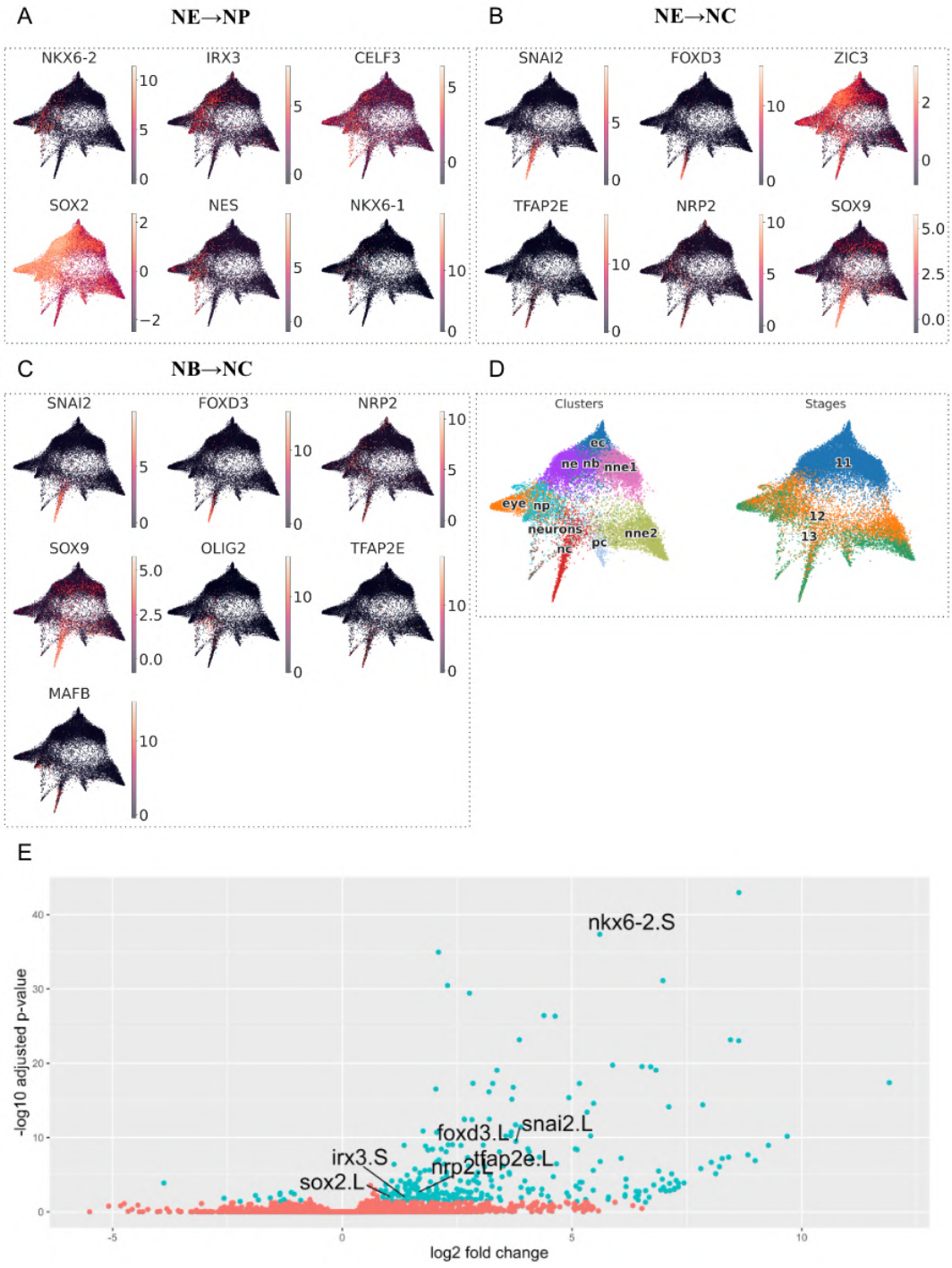



**Figure 22. Branches-related genes that are DE in Pax3-GOF, scRNA-seq data from** (Kotov et al., 2022). A. NE→NP branch genes; B. NE→NC branch genes; C. NB→NC branch genes; D. Annotation for dataset: clusters and stages; E. Expression levels for branches-related genes in Pax3-GOF sample

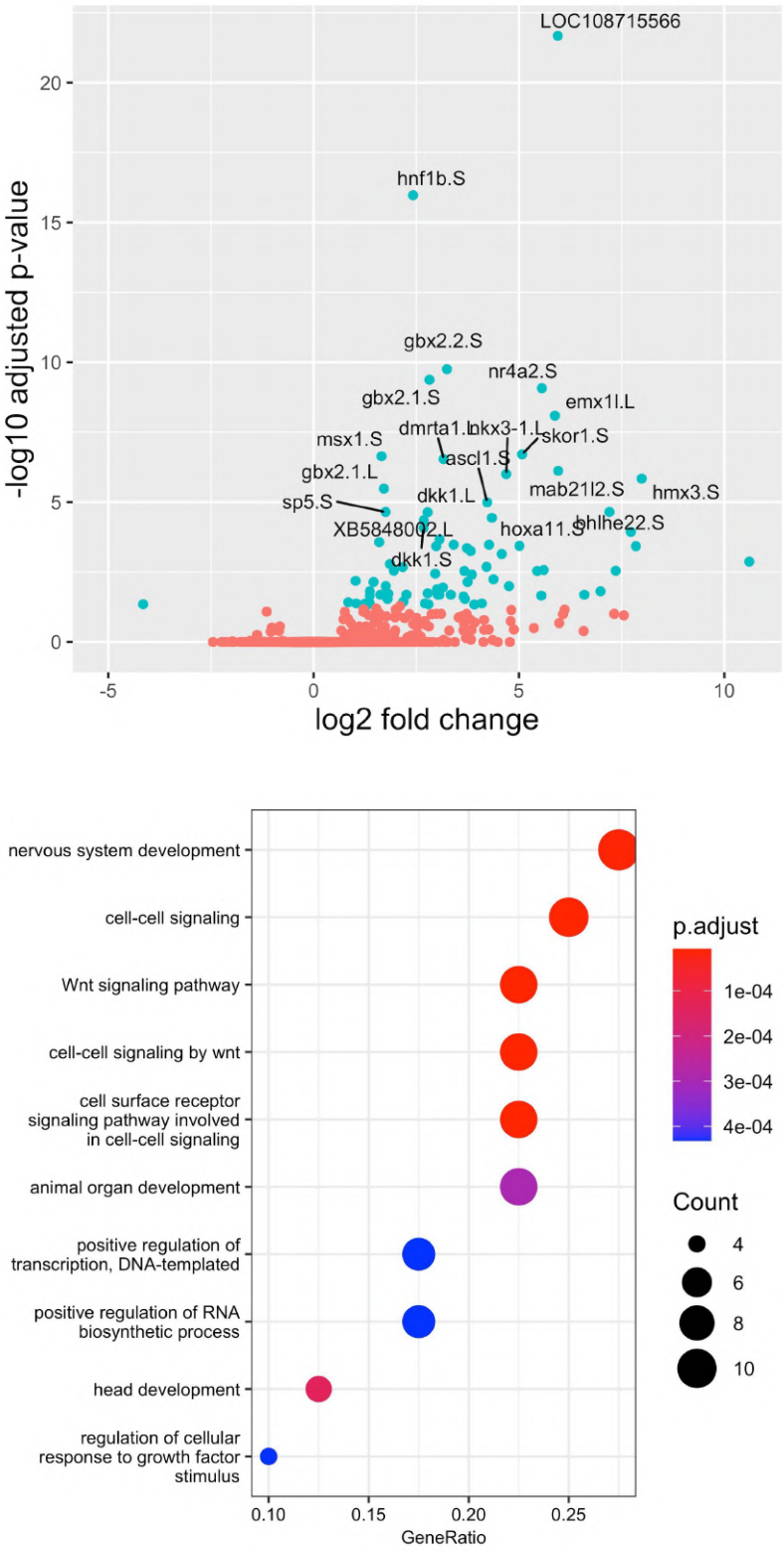



**Figure 23. Dlx3 top DE targets and related GO terms.**

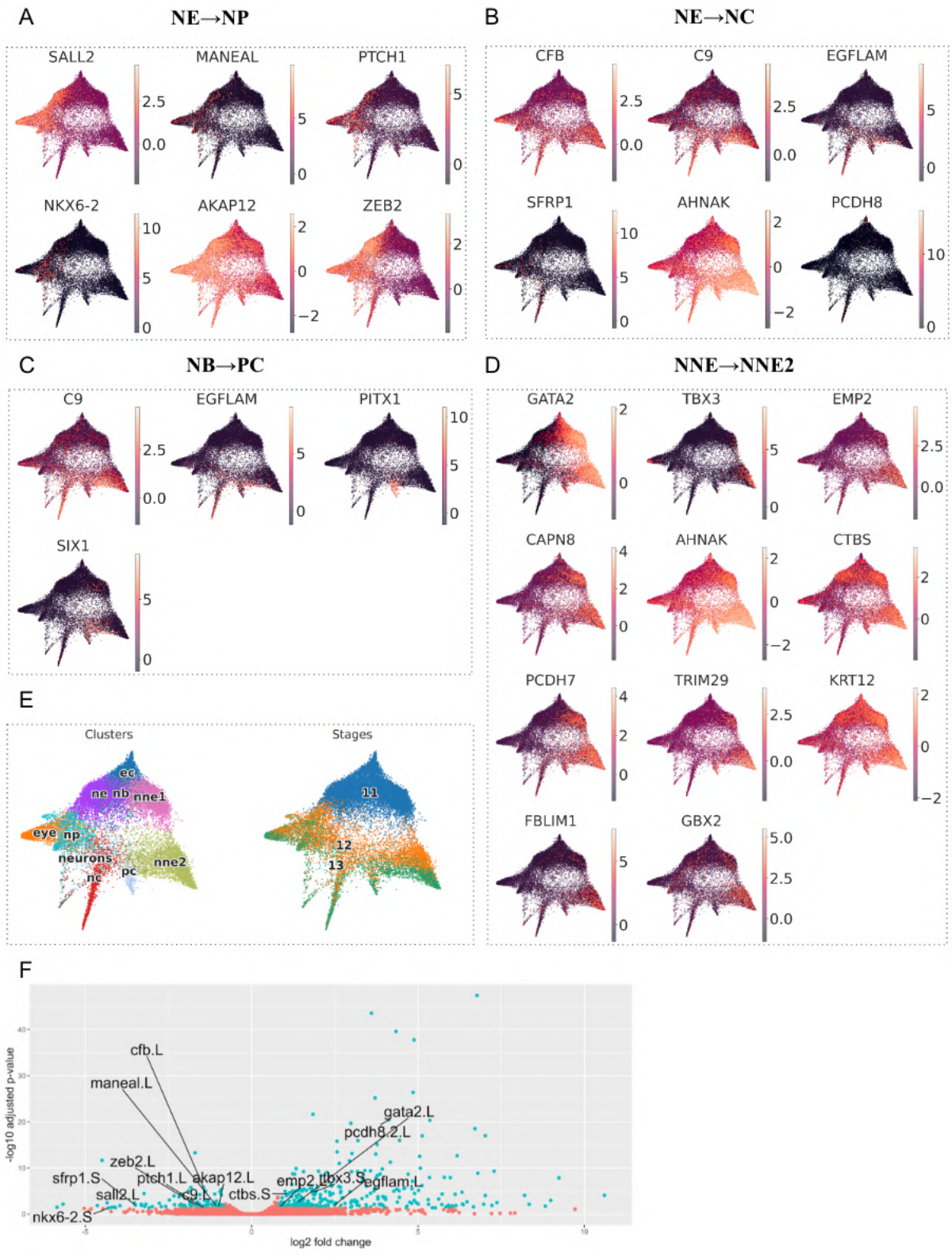



**Figure 24. Branches-related genes that are DE in Dlx3-GOF, scRNA-seq data from** (Kotov et al., 2022). A. NE→NP branch genes; B. NE→NC branch genes; C. NNE→PC branch genes; D. NNE→NNE2(late) branch genes; E. Annotation for dataset: clusters and stages; F. Expression levels for branches-related genes in Dlx3-GOF sample

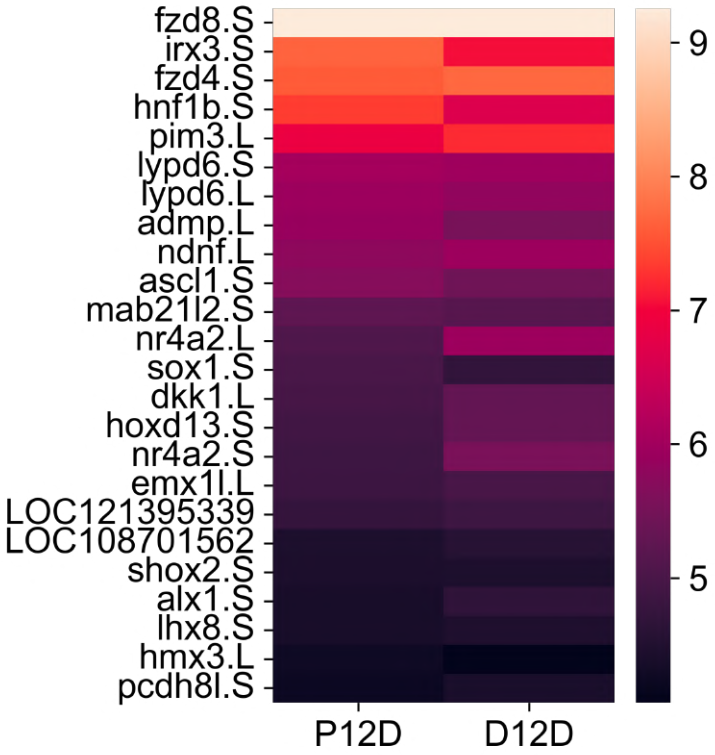



**Figure 25. Expression levels for targets both Dlx3 and Pax3.** Y-axis is in expression units

## 2.5 Computational tool for the prediction of the cell type evolution

**Aleksandr Kotov**, Andrei Zinovyev, Anne-Helene Monsoro-Burq

doi: https://doi.org/10.1101/2022.12.07.519467

### 2.5.1 Overview of key findings

Numerous computational methods now exist for analyzing single-cell data and identifying cell states. These methods mostly rely on the expression of genes considered as markers for a given cell state. Yet, there is a lack of scRNA-seq computational tools to study cell states' evolution, particularly how they change their molecular profiles. Here we present scEvoNet, a Python tool for predicting cell type evolution in cross-species or disease-related (e.g., cancer) scRNA-seq datasets. ScEvoNet builds the confusion matrix of cell states and a bipartite network connecting genes and cell states. Utilizing this framework and exploring the continuum of transcriptome states between developmental stages and cell types will help explain cell state dynamics.

### 2.5.2 The application architecture

The workflow of scEvoNet is illustrated in Figure 1A in paper 2. ScEvoNet takes (1) an expression matrix, and (2) a list of cell labels as input data per organism/time point of interest. For each cell type provided by a user, scEvoNet generates an LGBM binary classifier (one cell type vs all other cells). Next, scEvoNet uses each cell type model to predict cells from both datasets. This way we get a confusion matrix with cell type to cell type comparison. In the next step, scEvoNet builds a network where the nodes are cell types or genes so that cell types can only connect to genes. This strategy is similar to GRNboost2 (Moerman et al., 2019) which outperformed many other tools in a recent benchmarking study (Pratapa et al., 2020). GRNboost2 generates a gene-gene network similarly, whereas scEvoNet extends it to all cell types in two datasets. Furthermore, scEvoNet implements next the shortest path search in order to generate a subnetwork of interest.

### 2.5.3 Validation of the scEvoNet in *Mus musculus* scRNA-seq dataset

For the input data, we used whole embryo scRNA-seq datasets for the *Xenopus tropicalis,* a non-amniote tetrapod vertebrate, at an early developmental stage (late gastrulation stage 12, and neurulation stages 13 (neural plate) and 14 (neural fold)) and *Mus musculus,* a

mammal, at a similar developmental stage (late gastrulation stage 8.25). We validated how one particular NC-based classifier trained on the *Xenopus* dataset will able to recognize NC cells in the mouse dataset and obtained the result of a 0.89 AUC score (Figure 2 in paper 2). AUC stands for "area under the curve," and in this context it refers to the area under the curve of a receiver operating characteristic (ROC) curve. An ROC curve is a graphical plot that compares the true positive rate and false positive rate of a binary classifier, and the AUC score is a metric that quantifies the performance of the classifier. AUC scores can range from 0 to 1, with higher scores indicating better performance. Our results suggest that the classifier was able to accurately recognize NC cells in the mouse dataset.

### 2.5.4 scEvoNet can effectively find similar cell types and potentially co-opted genes

Next, using published cluster annotations, we run scEvoNet on mouse and frog datasets and obtained the confusion matrix. To identify genes that are highly conserved and cell type-specific in the evolution of the NC we selected a sub-network that consists of the shortest paths from *Xenopus* NC to mouse NC with the top 3 close cell types according to the confusion matrix. We determined several groups of genes differently related to cell types within this subnetwork. The first group includes genes that are associated between NC and a closely-related cell type in only one organism (e.g., NC and neural plate in frog: *sox2, sox3, snai2, hes1, zic1*; NC and midhindbrain in mice: *gadd45a, mdk, ptn*), which could be the result of divergence of function and can be studied using scRNA-seq of the ancestor organisms. The second group consists of genes that are characteristic for NC both in *Xenopus* and *Mus musculus* (*pax3, tfap2c, tfap2a, tfap2b, sox9*), which are known confirmed markers of NC or their progenitors. The third group includes genes that are correlated between the frog NC and mouse NC through the mouse NC-related cell types (*mafb, cldn6, tfap2a* for the neural plate; *zic3, tfap2a* for the midhindbrain).

Next, we applied scEvoNet to a human breast cancer metastasis dataset (Xu et al., 2021). We found 14 genes that are directly connected to both cell types, among them *malat1*, levels of which inversely correlate with breast cancer progression and metastatic capacity (Kim et al., 2018), and *b2m* an important marker involved in carcinogenesis, invasion, and metastasis (Liu et al., 2015). Among this list of genes directly connecting two tumor cell types are several mitochondrial genes which may indirectly support the importance of mitochondrial genes in cancer metastasis (Beadnell et al., 2018) but also might be the consequence of the sample or

data processing artifact. Next, we explored what other genes from the close cell types might be involved in tumor evolution. To do this, we selected a subnetwork with all the genes related to cell types of interest and 5 similar cell types according to the confusion matrix we obtained earlier. As a result, we determined two cancer cell types (metastatic and primary) that have a common network neighbor lymph node B cells through genes *hmgb1* and *b2m*. It was shown previously that exosomal *hmgb1* promotes hepatocellular carcinoma immune evasion by stimulating TIM-1+ regulatory B cell expansion (Ye et al., 2018). Also, blockade of the *hmgb1* signaling pathway inhibits tumor growth in diffuse large B-cell lymphoma (Zhang et al., 2019). In another work, *b2m* specific B cells were defined as the most important prometastatic B cell cluster essentially contributing to distant metastasis in Clear Cell Renal Cell Carcinoma (Yang et al., 2021). *B2m* is also an important element in the immune escape mechanism since a decrease in *b2m* expression reduces the number of antigens presented on the cell surface, including tumor-related antigens, which has been shown in particular in diffuse large B-cell lymphoma (Challa-Malladi et al., 2011).

### 2.5.5 Conclusion

In this work, we presented scEvoNet, a method for analyzing the evolution of cell states from highly sparse scRNA-seq data. We show that it is applicable to studying species-to-species and tumor-to-metastasis transitions. With this tool, we re-discover a canonical gene signature that remains conserved through evolution, and also predict species-specific genes and new candidates associated with similar cell types. However, there are several limitations to the use of computational tools for predictions from scRNA-seq data. One major limitation is the issue of dropout events, which occur when a gene is not detected in a cell even though it is present at some level. This can occur for a variety of reasons, including low mRNA abundance or technical issues with the sequencing. Dropout events can significantly impact the accuracy of predictions made from scRNA-seq data, as they can mask the true expression levels of certain genes. Another limitation of scRNA-seq data is the potential for batch effects, which occur when the data being analyzed was collected under different conditions or from different sources. For example, if the scRNA-seq data was collected from different species or in different labs, there may be differences in the way the data was generated that could impact the accuracy of predictions. Nevertheless, our findings may imply either co-activation of similar gene programs from common progenitor state or the potential co-option of genes from distant cell types.

Publication 2

## scEvoNet: a gradient boosting-based method for prediction of cell state evolution

**Aleksandr Kotov**, Andrei Zinovyev, Anne-Helene Monsoro-Burq

doi: https://doi.org/10.1101/2022.12.07.519467

# scEvoNet: a gradient boosting-based method for prediction of cell state evolution

Aleksandr Kotov[1,2], Andrei Zinovyev[2,3,4], Anne-Helene Monsoro-Burq[1,2,5]*

**Affiliations**

[1]Université Paris Saclay, Faculté des Sciences d'Orsay, Orsay, France

[2]Institut Curie, PSL Research University, Paris, France

[3]INSERM, Paris, France

[4]MINES ParisTech, PSL Research University, CBIO-Centre for Computational Biology, Paris, France

[5] Institut Universitaire de France, Paris, France

* To whom correspondence should be addressed. Email: anne-helene.monsoro-burq@curie.fr

**Abstract**

- **Background**

  Exploring the function or the developmental history of cells in various organisms provides insights into a given cell type's core molecular characteristics and putative evolutionary mechanisms. Numerous computational methods now exist for analyzing single-cell data and identifying cell states. These methods mostly rely on the expression of genes considered as markers for a given cell state. Yet, there is a lack of scRNA-seq computational tools to study the evolution of cell states, particularly how cell states change their molecular profiles. This can include novel gene activation or the novel deployment of programs already existing in other cell types, known as co-option.

- **Results**

  Here we present scEvoNet, a Python tool for predicting cell type evolution in cross-species or cancer-related scRNA-seq datasets. ScEvoNet builds the confusion matrix of cell states and a bipartite network connecting genes and cell states. It allows a user to obtain a set of genes shared by the characteristic signature of two cell states even between distantly-related datasets. These genes can be used as indicators of either evolutionary divergence or co-option occurring during organism or tumor evolution. Our results on cancer and developmental datasets indicate that scEvoNet is a helpful tool for the initial screening of such genes as well as for measuring cell state similarities.

- **Conclusions**

  The scEvoNet package is implemented in Python and is freely available from https://github.com/monsoro/scEvoNet. Utilizing this framework and exploring the continuum of transcriptome states between developmental stages and species will help explain cell state dynamics.

**Keywords**

scRNA-seq, gradient boosting, evolution, gene programs, cell states, cell types, differentiation, cancer

## Background

Cells, the fundamental construction blocks of multicellular organisms, are characterized by great diversity in complex multicellular organisms. They include differentiated and function-specific cells, their stem cells for cell renewal during lifetime, and all the transitional states between these two points. In disease, cell and tissue homeostasis are altered, leading to the appearance of new pathological and dysfunctional cells. During evolution, the diversification of cell types is caused by genomic individualization relying on fundamental evolutionary principles such as functional segregation, divergence, co-option of gene modules, and *de novo* gene emergence. Co-option of gene programs is a mechanism allowing the emergence of new functions in a cell type by using existing gene networks from other cell types [1, 2]. The understanding of cell biology emanates from describing cells by their functions, their gene expression, interactions with their environment, and their lineage relationships. The emergence of single-cell RNA sequencing (scRNA-seq) began a new age of transcriptomic research, extending our understanding of cell heterogeneity and dynamics. Highly detailed atlases of cell types were produced for many tissues and organisms, in normal or pathological conditions [3–7]. Comparing those highly divergent datasets would allow asking key questions regarding the conservation of core genetic programs in poorly-related cellular contexts, the origins of cellular diversity and its evolutionary mechanisms, or the transcriptional paths leading to disease. However, data received from various biological conditions and various organisms is entangled by technical and biological batch effects which vastly complicates their comparison [8, 9]. Thus, forces shaping transcriptome dynamics remain poorly understood. Another application of scRNA-seq in evolutionary biology is accessing tumor heterogeneity and tracking its transformation as well as assessing the selective evolution of tumors during therapy or metastatic progression [10]. ScRNA-seq overcomes the constraints of classic bulk

RNA sequencing by estimating transcriptome at a single-cell level and characterizing various cell types in the tumor microenvironment. Moreover, this allows a better understanding of the molecular mechanisms facilitating tumor occurrence. Although it could potentially reveal the somatic mutations during tumor evolution, scRNA-seq data sparsity [11] often prevents mutation calling (one of the main information sources for studying tumor evolution). Still, the scRNA-seq of tumors can determine the dynamic changes in tumor heterogeneity and the transcriptional evolution of tumor cells during metastasis development [12].

Currently, there is a lack of a specific tool that uses closely or distantly-related scRNA-seq datasets as input to study the potential co-option and evolution of gene programs between different organisms during development and differentiation, or between tumor cells at different stages of tumor progression. Kun Xu et al. [12] used Monocle [13] and scVelo [14] to study transcriptome dynamics of malignant cells between the primary tumor and lymph node metastases. They also used NATMI [15] for the generation of the cell-to-cell receptor-ligand network where edges are generated based on the expression of the ligand in one cell type and of its related receptor in another cell type. However, this latter strategy is not designed to study co-option in cancer which is a crucial mechanism forcing the molecular changes that propel tumor progression [16]. In another work, Pandey et al. use scRNA-seq to study the evolutions of neuronal types by comparing cell types in larva and adult zebrafish. They utilized Random Forest to generate a model for each cell type and predict cells with each model to build a confusion matrix mapping cell types by the number of cells predicted with each model [17]. Yet, this strategy is not wrapped into the usable framework and cannot be used to extract genes that are characteristic of cell type transitions. We present scEvoNet, a method that builds a cell type-to-gene network using the Light Gradient Boosting Machine (LGBM) algorithm [18] overcoming different domain effects (different species/different datasets) and dropouts that are inherent for the scRNA-seq data [19]. This tool predicts potentially co-opted genes together with genes characteristic of each cell state during development across species. Recently we showed the ability of a similar LGBM-based classifier to detect neural crest cells in distantly-related scRNA-seq datasets [20]. Despite technical batch effects (datasets were made in different laboratories with different technologies) and biological batch effects (datasets were from two evolutionarily distant organisms and at different developmental time points), we have achieved a high AUC score of 0.95 for classifying zebrafish cells with our frog-based NC model [20]. Here we have expanded this method: scEvoNet applies to a variety of applications, e.g., between different time points during a given organism's development, between species,

and when comparing primary tumor and metastasis. We believe that scEvoNet will facilitate the study of cell state transitions in a variety of contexts and from highly divergent datasets.

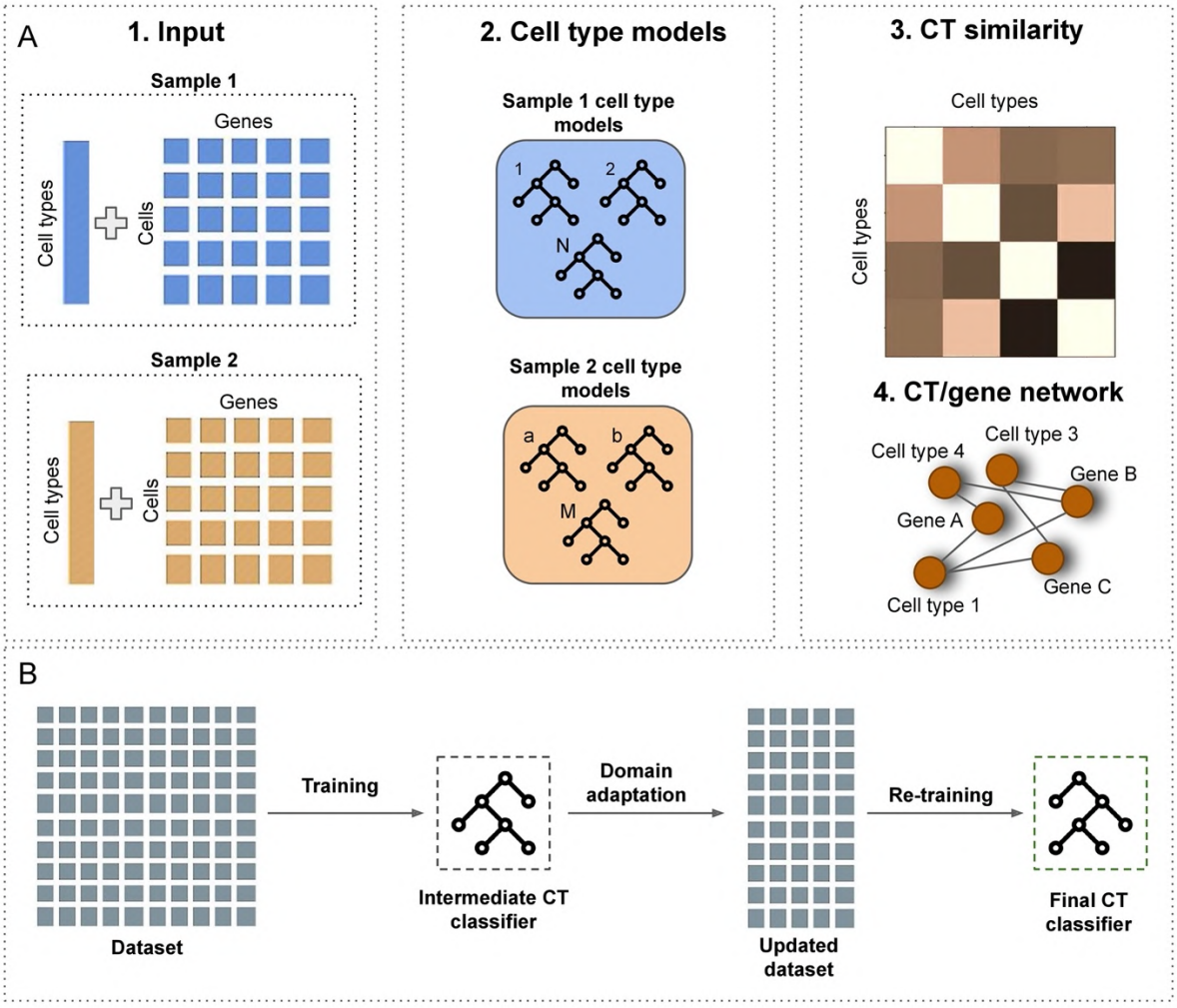



**Figure 1.** scEvoNet scheme. A) scEvoNet takes a list of clusters and a matrix of expressions for each sample as input. For each sample, it generates an object with cell type classifiers and top important features for each cluster from the provided set of clusters. In the final step, the tool builds a confusion matrix and a network of genes associated with each cell type. B) We use the LGBM algorithm to produce a classifier for each cell type. To smooth the data in order to deal with the batch effects we apply the sigmoid function and only the top important features to create the final model.

## Implementation

The workflow of scEvoNet is illustrated in Fig. 1A. scEvoNet takes (1) an expression matrix, and (2) a list of cell labels as input data per organism/time point of interest. For each cell type provided by a user, scEvoNet generates an LGBM binary classifier (one cell type vs all other cells) in two steps (Fig. 1B). Firstly, it generates a model considering all genes in the dataset. For the obtained model of the particular cell type, scEvoNet selects the top 3000

important features (cell types related genes) and uses only them to re-train the final cell type model which will be used for the generation of the cell types confusion matrix. This is part of the domain adaptation that we perform to make the resulting model less dependent on the number of genes that are missing in another dataset to which the model will be applied. Additionally, to reduce the effect of the different biological domains between datasets and to reduce the effect of the scRNA-seq data sparsity we apply a sigmoid function that smooths expression units more flexible than simple binarization, which has been shown to keep enough information for scRNA-seq data analysis [21]. For the model training, we use early stopping to avoid overfitting with 10 rounds which determines the actual number of estimators in the regressions.

Next, scEvoNet uses each cell type model to predict cells from both datasets. This way we get a confusion matrix with cell type to cell type comparison. In the next step, scEvoNet builds a network where the nodes are cell types or genes so that cell types can only connect to genes. To do this, firstly we extract top features (genes) which are important for each cell type (both positively correlated and negatively correlated) and then combine all the cell type-related top important genes into one main network with all cell types and cell type-related genes. This strategy is similar to GRNboost2 [22] which outperformed many other tools in a recent benchmarking study [23]. GRNboost2 generates a gene-gene network similarly, whereas scEvoNet extends it to all cell types in two datasets. Furthermore, scEvoNet implements a shortest path search in order to generate a subnetwork of interest. For example, to study the evolution of a particular cell type, a user might request all the shortest paths (with a selected cut-off on their length) between two cell types and scEvoNet will yield all the genes and cell types between these two using the confusion matrix as a metric of the cell types similarity. Each gene-to-cell type connection has an importance value (a score displaying how useful each feature was in the building of the boosted decision trees within the model) by which users can filter sub-networks.

## Results

First, we applied scEvoNet to identify core characteristics during the evolution of the neural crest (NC) cells using two different vertebrate organisms. The NC is a multipotent and migratory cell population unique to vertebrates and essential notably for pigment, peripheral and enteric nervous system, and craniofacial structures formation [24]. For the input data, we used whole embryo scRNA-seq datasets for the *Xenopus tropicalis,* a non-amniote tetrapod vertebrate, at an early developmental stage (late gastrulation stage 12, and neurulation stages

13 (neural plate) and 14 (neural fold)) and *Mus musculus,* a mammal, at a similar developmental stage (late gastrulation stage 8.25) [4, 25].

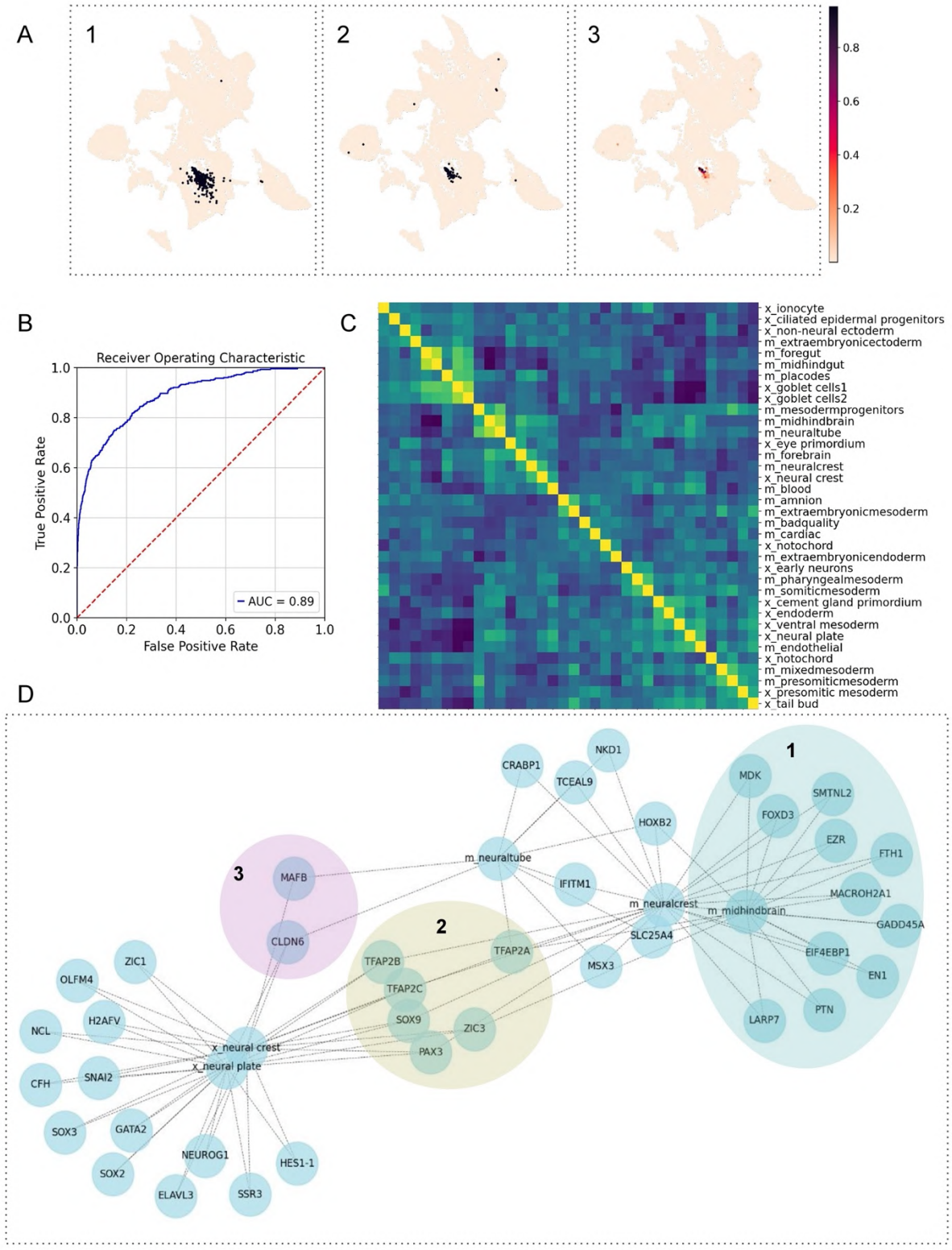



**Figure 2.** The development of the specific cell type between frog and mouse. A) First UMAP represents highlighted annotated neural crest cells in the whole embryo dataset, second UMAP represents predicted neural

crest cells with our classifier, and third UMAP represents predicted scores of our classifier. B) The AUC score for the neural crest classifier is 0.89 C) The confusion matrix for mouse and frog samples (prefix x_ is for *Xenopus*, prefix m_ is for mouse). The values in the confusion matrix are the correlations between two lists of scores for all cell type models. D) Selecting the subnetwork of 300 shortest paths from *Xenopus* neural crest to mouse neural crest shows genes that are shared with closely-related cell types, such as *mafb* or *cldn6* (group 3). It reveals two groups of genes: genes from group 1 are organism-specific genes, and genes from group 2 are important genes for the specific cell type (NC) in both organisms.

Before using scEvoNet, we validated how one NC-based classifier trained on the *Xenopus* dataset will recognize NC cells in the whole embryo mouse dataset and obtained the result of a 0.89 AUC score (Fig. 2A, B). Next, using published cluster annotations for these two whole embryo scRNA-seq datasets, we run scEvoNet and obtained the confusion matrix (Fig. 2C). In this dataset of extended complexity, the highest similarity score for frog NC remains the mouse NC. Thereafter, we built a network of cell types and related genes. To identify genes that are highly conserved and cell type-specific in the evolution of the NC we selected a sub-network that consists of the shortest paths from *Xenopus* NC to mouse NC with the top 3 close cell types according to the confusion matrix. To obtain a larger subnetwork number_of_shortest_paths=300 was used. Subsequently, we determined several groups of genes differently related to cell types within this subnetwork (Fig. 2D). The first group includes genes that are associated between NC and a closely-related cell type in only one organism (e.g. NC and neural plate in frog: *sox2, sox3, snai2, hes1, zic1*; NC and midhindbrain in mice: *gadd45a, mdk, ptn*). If this gene expression signature was the consequence of the evolutionary divergence of function, this could be studied using the scRNA-seq of the ancestor organisms. The second group consists of genes that are characteristic of NC both in *Xenopus* and *Mus musculus* (*pax3, tfap2c, tfap2a, tfap2b, sox9*): all are known markers of NC or their progenitors [26]. The third group includes genes that are associated between the frog NC and mouse NC, and shared with the mouse NC-related cell types (as defined with confusion matrix of similarities): *mafb, cldn6* for the neural plate; *zic3, tfap2a* for the midhindbrain. Thus, our tool was not only able to construct a matrix of similar cell types that can be used to study cell types similarities, but also defined three groups of genes that may have diverse roles in a cross-species transformation of the molecular profile of neural crest cells.

Next, we applied scEvoNet to a human breast cancer metastasis dataset [12]. We selected a patient with available datasets for the primary tumor and the lymph node metastases. We used a standard Scanpy [27] pipeline to obtain clusters for both matrices (primary tumor and metastasis). Marker genes from the source paper were used to annotate obtained clusters

(Fig. 3A). First, using scEvoNet, we calculated the confusion matrix (Fig. 3B). As expected, we observed high connectivity between cells of the same type from the primary tumor or metastasis in lymph nodes, e.g., B cells, plasma cells, immune cells, macrophages, dendritic cells, and tumor cells. Next, to study cancer cell evolution, we used scEvoNet to discover common cluster-specific genes between the most distant malignant clusters from the primary tumor (cluster p_cancer_cells_cox6+) and metastasis (cluster m_cancer_cells_gapdh+). We found 14 genes that were directly connected to both cell types (Fig. 3C), among them *malat1*, levels of which inversely correlate with breast cancer progression and metastatic capacity [28], and *b2m* an important marker involved in carcinogenesis, invasion, and metastasis [29]. Among this list of genes directly connecting two tumor cell types were several mitochondrial genes. Although a common hypothesis relates the expression of mitochondrial genes to sample or data processing artifacts, growing evidence supports the importance of mitochondrial genes in cancer metastasis [30]. Next, we explored what other genes from close cell types might be involved in tumor evolution. To do so, we selected a subnetwork with all the genes related to cell types of interest and 5 similar cell types according to the confusion matrix obtained earlier. As a result, we determined two cancer cell types (metastatic and primary) that have as a common network neighbor the lymph node B cells, through genes *hmgb1* and *b2m*. Interestingly, it was shown previously that exosomal *hmgb1* promotes hepatocellular carcinoma immune evasion by stimulating TIM-1+ regulatory B cell expansion [31]. Also, blockade of the *hmgb1* signaling pathway inhibits tumor growth in diffuse large B-cell lymphoma [32]. In another work, *b2m* specific B cells were defined as the most important prometastatic B cell cluster essentially contributing to distant metastasis in Clear Cell Renal Cell Carcinoma [33]. *B2m* is also an important element in the immune escape mechanism since a decrease in *b2m* expression reduces the number of antigens presented on the cell surface, including tumor-related antigens, which has been shown in particular in diffuse large B-cell lymphoma [34]. Thus, scEvoNet here provides a result supported by the literature, suggesting that users can retrieve meaningful gene candidates involved in tumor progression and immune escape in cancer.

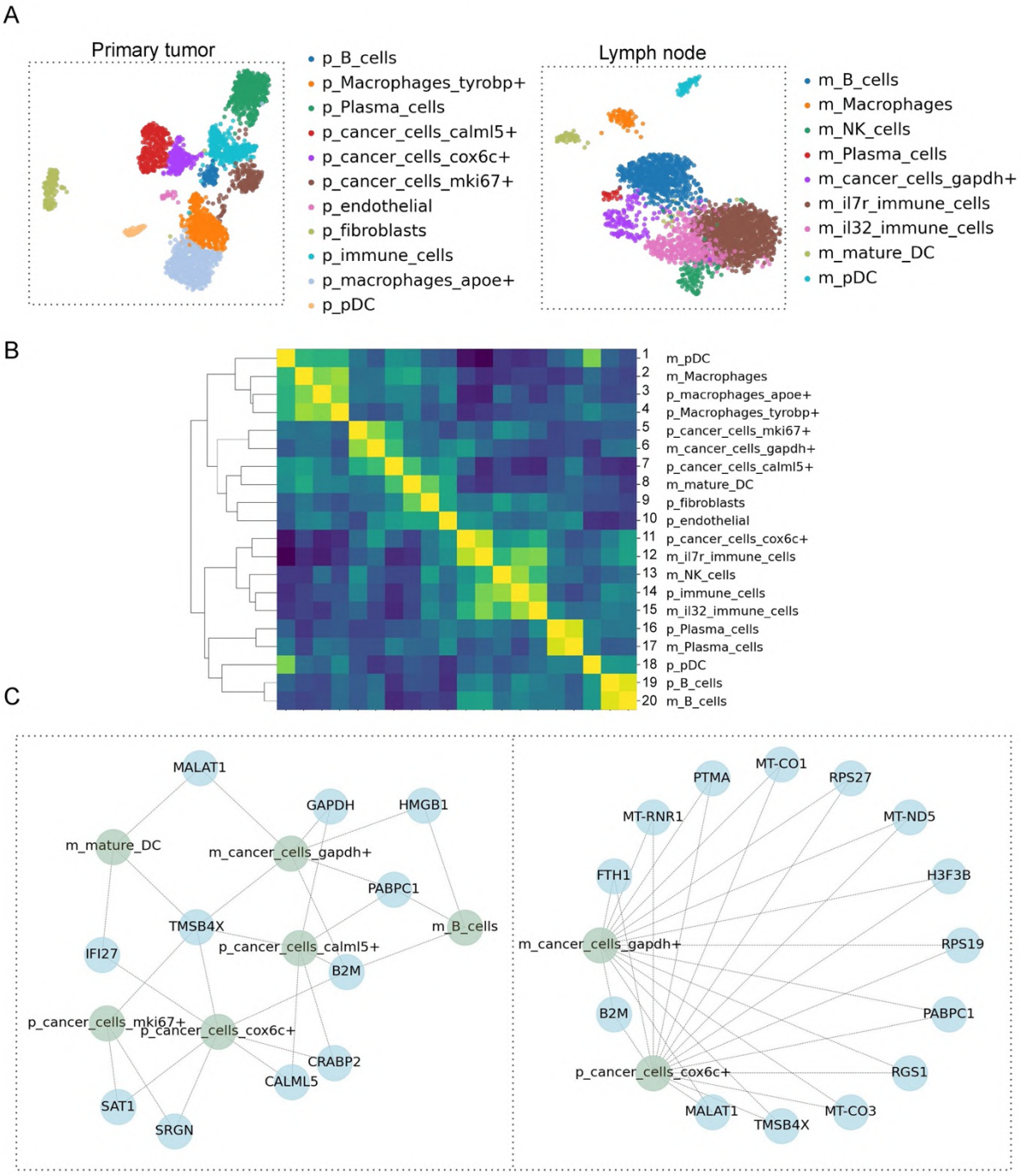



**Figure 3.** Primary tumor vs metastasis comparison. A) UMAPs for the primary human breast cancer (left) and metastasis in the lymph node (right). B) The confusion matrix shows different rates of similarity between different clusters of cancer cells in primary tumor and metastasis (the p_ prefix is for primary, and the m_ prefix is for metastasis). The values in the confusion matrix are the correlations between two lists of scores for all cell type models. C) Two subnetworks of the relation of the cluster of cancer cells in primary tumor and cancer cells in metastasis. On the left subnetwork, we show only genes related to some other cell types, on the right subnetwork we selected genes that are directly connected to clusters of interest.

## Conclusions

The evolution of cell types and gene programs is one of the main focuses of developmental biology and is crucial for a better understanding of the origin of particular functions. For the moment, there is a lack of computational tools to address this question using the abundant scRNA-seq data publicly available databases. One existing approach is not wrapped into the usable framework (e.g., R/Python package) and has only one application (cell states comparison) so it cannot be used to extract genes that are responsible for cell types transitions such as co-opted genes or genes conservatively important for several cell states [17]. In this manuscript, we present scEvoNet, a method for analyzing the evolution of cell states from highly sparse scRNA-seq data. We show that it is applicable to studying cross-species and tumor-to-metastasis transitions. With this tool, we re-discover a canonical gene signature that remains conserved through evolution, and also predict species-specific genes and new candidates associated with similar cell types. Our findings may indicate the co-option of genes or shared programs in closely related cell types. It also suggests the potential use of an immune escape mechanism in breast cancer metastasis, which has previously been shown in another cancer type. Yet, one limitation is that scEvoNet does not match gene sequences and only works with labels provided by the user, which can reduce the number of genes to be found between different cell types in cross-species comparison.

The tool is adjustable and can be utilized for an initial screening strategy. It is compatible with AnnData object format used in the Scanpy Python package [27].

## Availability and requirements

List the following:

Project name: scEvoNet

Project home: https://github.com/monsoro/scEvoNet

Operating system(s): Mac OS, Linux, Windows

Programming language: Python

License: MIT license

Any restrictions to use by non-academics: MIT license

## List of abbreviations

scRNA-seq: single-cell RNA sequencing

LGBM: Light Gradient Boosting Machine

NC: neural crest

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

We have been using public data and did not produce sequence data by ourselves.

### Competing interests

AZ is employed by EvoTec company.

### Funding

This project received funding from European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement No 860635, NEUcrest ITN (to AHMB); Agence Nationale pour la Recherche (ANR-21-CE13-0028; to AHMB) and Institut Universitaire de France (to AHMB); Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute).

### Authors' contributions

AK, AHMB, AZ conceived the project. AK designed the strategy, performed the work, did the programming, wrote and edited the manuscript. AZ reviewed the code and edited the manuscript. AHMB supervised the project and edited the manuscript.

### Acknowledgments

The authors are grateful to Drs. Igor Adameyko and Leon Peshkin for insightful scientific discussions and comments on the manuscript.

### Authors' information

-

### Footnotes

-

## References

1. Arendt D. The evolution of cell types in animals: emerging principles from molecular studies. Nat Rev Genet. 2008;9:868–82.

2. Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, et al. The origin and evolution of cell types. Nat Rev Genet. 2016;17:744–57.

3. Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science. 2018;360:981–7.

4. Briggs JA, Weinreb C, Wagner DE, Megason S, Peshkin L, Kirschner MW, et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science. 2018;360:eaar5780.

5. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, et al. Comprehensive single cell transcriptional profiling of a multicellular organism. Science. 2017;357:661–7.

6. Haber AL, Biton M, Rogel N, Herbst RH, Shekhar K, Smillie C, et al. A single-cell survey of the small intestinal epithelium. Nature. 2017;551:333–9.

7. Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, et al. Molecular Architecture of the Mouse Nervous System. Cell. 2018;174:999-1014.e22.

8. Marioni JC, Arendt D. How Single-Cell Genomics Is Changing Evolutionary and Developmental Biology. Annu Rev Cell Dev Biol. 2017;33:537–53.

9. Stuart T, Satija R. Integrative single-cell analysis. Nat Rev Genet. 2019;20:257–72.

10. Saunders NA, Simpson F, Thompson EW, Hill MM, Endo-Munoz L, Leggatt G, et al. Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives. EMBO Mol Med. 2012;4:675–84.

11. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. Genome Biology. 2020;21:31.

12. Xu K, Wang R, Xie H, Hu L, Wang C, Xu J, et al. Single-cell RNA sequencing reveals cell heterogeneity and transcriptome profile of breast cancer lymph node metastasis. Oncogenesis. 2021;10:1–12.

13. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32:381–6.

14. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol. 2020;38:1408–14.

15. Hou R, Denisenko E, Ong HT, Ramilowski JA, Forrest ARR. Predicting cell-to-cell communication networks using NATMI. Nat Commun. 2020;11:5011.

16. Billaud M, Santoro M. Is Co-option a prevailing mechanism during cancer progression? Cancer Res. 2011;71:6572–5.

17. Pandey S, Shekhar K, Regev A, Schier AF. Comprehensive Identification and Spatial Mapping of Habenular Neuronal Types Using Single-cell RNA-seq. Curr Biol. 2018;28:1052-1065.e7.

18. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2017.

19. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet. 2015;16:133–45.

20. Kotov A, Alkobtawi M, Seal S, Kappès V, Ruiz SM, Arbès H, et al. From neural border to migratory stage: A comprehensive single cell roadmap of the timing and regulatory logic driving cranial and vagal neural crest emergence. 2022;:2022.03.23.485460.

21. Qiu P. Embracing the dropouts in single-cell RNA-seq analysis. Nat Commun. 2020;11:1169.

22. Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. Bioinformatics. 2019;35:2159–61.

23. Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods. 2020;17:147–54.

24. Seal S, Monsoro-Burq AH. Insights Into the Early Gene Regulatory Network Controlling Neural Crest and Placode Fate Choices at the Neural Border. Front Physiol. 2020;11:608812.

25. Ibarra-Soria X, Jawaid W, Pijuan-Sala B, Ladopoulos V, Scialdone A, Jörg DJ, et al. Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. Nat Cell Biol. 2018;20:127–34.

26. Simões-Costa M, Bronner ME. Establishing neural crest identity: a gene regulatory recipe. Development. 2015;142:242–57.

27. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biology. 2018;19:15.

28. Kim J, Piao H-L, Kim B-J, Yao F, Han Z, Wang Y, et al. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. Nat Genet. 2018;50:1705–15.

29. Liu C, Yang Z, Li D, Liu Z, Miao X, Yang L, et al. Overexpression of B2M and loss of ALK7 expression are associated with invasion, metastasis, and poor-prognosis of the pancreatic ductal adenocarcinoma. Cancer Biomark. 2015;15:735–43.

30. Beadnell TC, Scheid AD, Vivian CJ, Welch DR. Roles of the mitochondrial genetics in cancer metastasis: Not to be ignored any longer. Cancer Metastasis Rev. 2018;37:615–32.

31. Ye L, Zhang Q, Cheng Y, Chen X, Wang G, Shi M, et al. Tumor-derived exosomal HMGB1 fosters hepatocellular carcinoma immune evasion by promoting TIM-1+ regulatory B cell expansion. Journal for ImmunoTherapy of Cancer. 2018;6:145.

32. Zhang T, Guan X-W, Gribben JG, Liu F-T, Jia L. Blockade of HMGB1 signaling pathway by ethyl pyruvate inhibits tumor growth in diffuse large B-cell lymphoma. Cell Death Dis. 2019;10:1–15.

33. Yang F, Zhao J, Luo X, Li T, Wang Z, Wei Q, et al. Transcriptome Profiling Reveals B-Lineage Cells Contribute to the Poor Prognosis and Metastasis of Clear Cell Renal Cell Carcinoma. Frontiers in Oncology. 2021;11.

34. Challa-Malladi M, Lieu YK, Califano O, Holmes A, Bhagat G, Murty VV, et al. Combined Genetic Inactivation of Beta2-Microglobulin and CD58 Reveals Frequent Escape from Immune Recognition in Diffuse Large B-cell Lymphoma. Cancer Cell. 2011;20:728–40.

# 3. Discussion and conclusion

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. Momentarily after gastrulation, in the anterior-most part of the embryo, NC cells are induced from the dorsal-lateral "neural border zone", an ectoderm domain located between the non-neural ectoderm and the neural plate ectoderm (Alkobtawi and Monsoro-Burq, 2020; Plouhinec et al., 2017). During the process of neurulation, the specification and induction of neural crest cells progresses in a wave-like manner from the anterior to posterior area of the neural plate. This process is accompanied by the activation of various gene programs that define the early and immature stages of NCs. In addition to the general NC program, several regional molecular modules are also activated along the anterior-posterior axis of the body, which help to define sub-populations of NCs with specific potential (Ling and Sauka-Spengler, 2019; Tang et al., 2021).

How these sub-populations-related programs are interconnected with the pan-NC module, and how and when they are activated in pre-migratory NC cells are poorly described. Later, at the end of neurulation, NC cells leave the dorsal ectoderm by a stereotypical epithelium-to-mesenchymal transition 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 analyzed during development and evolution, leading to the elucidation of elaborate gene regulatory networks during the last decade (Monsoro-Burq et al., 2005; Simoes-Costa and Bronner, 2016). These networks, however, remain incomplete and do not account for most of the defects observed in human neurocristopathies (Medina-Cuadra and Monsoro-Burq, 2021).

This problem is ready to be solved with single-cell transcriptomics which enable a full description of NC development, over sequential developmental stages, also in comparison, to adjacent tissues (e.g., at the neural border). That would define the developmental genetic trajectories of the complete NC lineage tree. Recent scRNAseq studies on NC cells have mainly explored NC after emigration, using chick, fish, and mouse embryos (Artinger and Monsoro-Burq, 2021). 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 (Tang et al., 2021; Zalc et al., 2021).

In addition to the neural crest, NB territory also gives rise to posterior placodes, non-neural ectoderm, and the dorsal part of the neural tube (Steventon and Mayor, 2012; Streit and Stern, 1999). 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. While the formation of the NB territory has been defined by the expression of a few genes during gastrulation (e.g., *pax3* and *pax7*) (Basch et al., 2006; Monsoro-Burq et al., 2005; Plouhinec et al., 2017), 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 (Groves and LaBonne, 2014; Maharana and Schlosser, 2018; Steventon and Mayor, 2012). 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.

Generally, studies in numerous model systems have underlined important parts of the GRN governing NB induction and NC/Placodes fate choice. Notably, the roles of the critical regulators are greatly conserved in different species (Seal and Monsoro-Burq, 2020). Yet central questions stay unrevealed and while some functional relationships start to be inducted, most direct regulations await additional validation.

Furthermore, complex mechanisms between signaling paths and NB markers are still not clearly understood (Garnett et al., 2012). The NB markers *pax3/msx1/zic1/tfap2a/hes4* control each other in a feed-forward loop and need extra WNT signaling (de Crozé et al., 2011; Simoes-Costa and Bronner, 2016). Frog placode progenitors' markers *eya1/six1* influence NB and NC markers *pax3* and *foxd3* as well as NB inducers *tfap2a*, *msx1*, *dlx3*, *gata2*, *foxi1* (Maharana and Schlosser, 2018). Together, these elaborate cross-regulation stabilize fate choices in the neural border, neural crest, and placodes. All these studies open a discussion of two different models proposed for neural border development:

1. The "Binary competence" model suggested by Schlosser in which the NC and the placodes are yielded individually: the NC on the neural side of the neural border and placodes on the non-neural ectoderm side (Pieper et al., 2012; Schlosser, 2008).

2. The “NB” model proposes the bi-potent neural border yields both NC and placode progenitors, the comparative positions of which are defined at later stages by special markers.

Recent investigations propose a mix of both models in vivo: the multipotent NB indicates co-expression of specifiers of either fate and no spatial segregation of fate-limited cells (NB model), but as development proceeds, the ability to form either NC or placodes would restrict sub-zones NB (binary competence [Briggs et al., 2018; Maharana and Schlosser, 2018; Roellig et al., 2017]). Recently was proposed another model, gradient border model, this model suggests that the fate of a cell, or the type of cell it will become, is influenced by both probability and its location in the developing embryo. According to this model, cells located in lateral NB are more likely to become cells of the epidermal or placode fates, while cells found in medial NB, are more likely to become neural crest cells. However, this model also suggests that cells in the NB are not fully committed to a specific fate, but rather are kept in an indeterminate state while they continue to express factors that could potentially lead to the development of different cell types (placodes, neural crest, non-neural ectoderm, neural ectoderm) (Thiery et al., 2022).

In addition to these questions that remain to be answered about the regulation of neural crest cells and neural border, it is also necessary to consider other gene regulatory mechanisms, such as the regulation performed by non-coding RNA. A role for miRNAs during development was first reported in 2003, when Bernstein removed Dicer (which cuts double-stranded RNA and pre-miRNA molecules to create short double-stranded RNA fragments, including miRNAs) in mice, detecting early embryonic lethality (Bernstein et al., 2003). Other studies also showed miRNAs are significant in many developmental processes (Antonaci and Wheeler, 2022; Mok et al., 2017). MicroRNAs are essential for normal animal development and are involved in different biological processes. Aberrant expression of miRNAs is linked to various human diseases (Hanna et al., 2019). Also, miRNAs are secreted into extracellular fluids. These miRNAs were widely reported as possible biomarkers for some diseases and they also act as signaling molecules to moderate cell-to-cell contacts (Bhaskaran and Mohan, 2014).

Further actions are required to fit miRNAs in GRN which manage the development of different tissues, including NC. It is essential to highlight that not all the players affected in NC development were uncovered and NC-GRN that governs is constantly being revised. Other gene regulatory mechanisms, such as the regulation managed by miRNA, need to be evaluated.

A recent review highlighted the importance of different miRNAs in NC development and neurocristopathies, including specification, migration, and differentiation (Antonachi and Wheeler, 2022). Further analysis requires to define the exact consequences of these miRNAs in NC development. Specifically, at what moment during NC development does it acts, and how is it regulated.

Thus, to answer questions regarding the heterogeneity of the cell fate decisions before and after NC induction, we used single-cell transcriptomes from eight consecutive developmental stages of *Xenopus tropicalis*, featuring 6135 NC cells, to provide comprehensive developmental profiling of 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 the 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 (submitted, deposited in BioRxiv).

Also, to study the cell date decision in NB and the relationships between main regulation actors there we (1) used single-cell transcriptomes from 3 developmental stages of *Xenopus tropicalis*, featuring 17138 early ectoderm cells to provide comprehensive developmental profiling of the NB, as well as (2) the combination of inducible gene constructs and cycloheximide (a drug that blocks protein translation) for Pax3, Zic1, and Dlx3. Transcription factor connectome and bifurcation analyses demonstrated the early emergence of neural crest fates at the neural plate stage, alongside an unbiased multipotent lineage persisting until after epithelial-mesenchymal transition. And perturbation experiments revealed shared and unique targets most important actors in NB (study in progress).

Next, to extend our knowledge about GRN regulating the NC, we selected two miRNAs, which has been previously shown to be important for the neural crest prior to migration. The main aims were to confirm that miR-196a and miR-219 are actually required

for NC development using morpholino-mediated knockdown of microRNAs miR-196a and miR-219 and start to decipher at what level the miRNAs are involved in the development of Xenopus NC (study in progress).

Lastly, to know more about the evolution of the NC and other cell types, we developed the ML-based computational tool scEvoNet (**s**ingle-**c**ell **Evo**lutionary **net**works). For the past 150 years, cell types have been studied in terms of development and morphology, whereas with the advent of single-cell transcriptomics, many distinct and hidden cell types have been identified. With the use of scRNA-seq, a completely new era is dawning in understanding the evolution of cell types. Seemingly, the ancestral types of animal cell types were poly-functional (sponge cell has both neural and immune properties), but during the evolution, functions were distributed between different cells (Arendt, 2008). This "division of labor" is one of the key ways how cell types evolve. Yet, there is a lack of scRNA-seq computational tools to study the evolution of cell states. Thus, here we present scEvoNet, the software that is implemented to study this conception from scRNA-seq data. Using this tool for studying the continuum of transcriptome states between developmental stages or species will help explain the complex dynamics of cell states (submitted, deposited in BioRxiv).

## 3.1 A reconciliatory “Model of Dual Convergence” describes the converging trajectories initiating neural crest and placode states

To reveal the molecular mechanisms that distinguish NC and placode induction at the neural-nonneural ectoderm border zone during gastrulation and check both models (“Binary competence” and “NB”), we generated dataset for all ectoderm cells from stages 11-13 (using the whole embryo scRNA-seq dataset). A fate change via a binary fate separation (bifurcation) is usually observed when an apparently homogeneous population of precursors divides into two cell types with distinct gene expression patterns. The NB itself is interpreted as a general competence zone for neural tube, nonneural ectoderm, NC, and placodes. And although it has some preferential association of cell fates, which reminds the binary competence model proposed by Schlosser (Pieper et al., 2012; Schlosser, 2008) where the NC and the placodes are induced separately, we observed that the NB zone contributes NC and PC in parallel to convergent contributions from neural plate and non-neural ectoderm progenitors.

Interestingly, we found that distinct genes were activated to obtain the same state and some genes were activated with different expression dynamics relative to different bifurcations. For example, during the NB to NC transition, *sox9* and *c3* were activated before bifurcation indicating that these genes could trigger cells from NB progenitors to a NC fate. In contrast, during the neural ectoderm to NC transition, *sox9* and *c3* were late genes (activated after bifurcation) while *foxd3* and *zic1* were expressed early. This observation suggested a new model of fate decisions in the developing ectoderm, where parallel and distinct genetic programs activated in distinct ectoderm progenitors may lead to a similar state, supporting a "Dual convergence Model" for neural crest and placode emergence from dorsal ectoderm, thus merging the currently competing models of "Neural border origin" opposed to "Neural / non-neural ectoderm origin.

Moreover, NB zone-specific gene *pax3* was expressed prior to bifurcation in the NB to NC gene program and activated expression of late NC branch markers *sox9, sox8, zic1, pcdh8,* and *c3*. Additionally, the Ectoderm connectome described NC genes connected to the rest of the network through Pax3 and Sox9, suggesting that Sox9 might play a yet undescribed function downstream of Pax3 in NC induction and upstream of the other late NC-branch markers. This agreed with *sox9* being an early gene in the NB→NC branch. Experimental validation showed that co-activation of Sox9 strongly increased *snail2* expression in iNC, while Sox9 depletion reduced *snail2* activation, indicating that Sox9 is required downstream of Pax3

for efficient NC induction by Pax3 and Zic1. While Sox9 had been described as essential for NC induction, the epistasis mechanism between NB specifiers, Sox9 and Snail2 had remained unknown until this experiment.

There are several limitations to this study. One limitation is that the data was only collected from a limited number of stages, so it is not clear how the findings would apply to other stages of development. Additionally, the study only focused on the induction of NC and placodes, so it is not clear how the findings would apply to the development of other cell types or tissues. There are several potential future directions for this research. For example, it would be interesting to investigate the role of other transcription factors in the induction of NC and placodes, and to study the development of these cell types at later stages. It would also be interesting to explore the potential clinical applications of these findings, such as in the development of therapies for developmental disorders or in regenerative medicine.

## 3.2 A combination of Omics and in vivo strategies validates large sets of gene regulations driving the dynamics of neural crest diversification

The puzzle of whether NC clusters use differential biases to navigate a single fate landscape, or are presented with fundamentally unique decisions on separate landscapes, may be approached by identifying “muted” bifurcations where multilineage priming occurs but one fate is never expressed. Single-cell transcriptomics catches sufficient differences across transcriptional states, which may mean decision-making processes. Although significant transcriptional changes are usually assumed to be linked to fate choice events, it might instead mean post-commitment differentiation, at the same time small changes affecting only several genes can be actually answerable for homogeneity breaking. New approaches for tracking such events across developmental stages are able to define slight changes in gene expression patterns.

### 3.2.1 Heterogeneity in pre-migratory NC

ScRNA-seq dataset of the whole embryo with accurate extraction of the NC cells helped us to define the temporal dynamics of trajectories that result in 8 neural crest states present upon early migration stage along the cranial and vagal axial positions. Our temporal analysis highlights three important points deepening our understanding of NC biology. Firstly, there

have been long-standing debates about the timing of NC fate decisions, prior to or after EMT from the neural tube, in a variety of animal models (Kalcheim and Kumar, 2017). Importantly, we did not detect distinctive expression of predictive fate markers before EMT (e.g., for neuronal, glial skeletogenic, or melanocyte fates). This suggested that, if some NC progenitors were biased toward a given fate prior to EMT, they did not exhibit a detectable signature in our dataset. However, our observations are in agreement with several lineages tracing studies showing the high multipotency of most NC cells when marked prior to EMT (Baggiolini et al., 2015). The first differentiation markers are found after emigration, as we detected myosin-like expression in a small subset of cells suggesting the emergence of previously poorly described NC-derived myofibroblasts shortly after EMT. Secondly, our data support the early diversification into several distinct cell states prior to, during, and after EMT, contrasting with the recent suggestion that upon EMT the NC progenitors would regroup into a single common multipotent state (Zalc et al., 2021). The high cell content of our dataset proves otherwise, suggesting that this previous observation made on a smaller subset of cranial NC did not fully capture the diversity of pre-migratory NC states. Another key observation is the presence of the main population of NC unbiased towards any particular state, expressing markers of the immature neural crest cells, from which all the other trajectories emerge. This unbiased cell trajectory is maintained during and after EMT suggesting that a very plastic, stem-like NC cell population emigrates and is subjected to the signals from the microenvironment prior to fate choices.

### 3.2.2 NC connectome and multiple bifurcations describe molecular logic of the cell fate decisions in vagal and cranial NC

We expand the network of genes involved in the NC-GRN by predicting gene correlations from the NC dataset to link transcription factors to their potential targets in the NC transcriptome. The resulting network of >16.000 potential TF-targets connections was partly validated by the combination of the knock-down experiments and ChIP-seq for some of the most linked nodes of the network. In particular, we have focused on the mid-neurulation role of Pax3 and TFAP2e, which were previously known for their earlier functions in NC induction. In both Pax3 and Tfap2e cases, we find that the three approaches result in partially overlapping gene target lists due to their use of different parameters (stage-wise, expression-wise, etc.). Together, these data provide an enlarged and validated NC-specific genome-wide connectome for two key NB/NC specifiers revealing a static depiction of the GRN in the NC.

To access the temporal dynamics of NC regulation, we examined temporal trajectory analysis unraveling the branch-specific dynamics of gene expression underlying bifurcations and state diversification. For each bifurcation, we provide a list of key genes likely to control branching choices. We further used experimental modulation of pivotal transcription factors function in the pre-migratory neural crest (Pax3, TFAP2e) to validate these bifurcation predictions. This provides a comprehensive view of the hierarchy of molecular decisions driving the cranial and vagal neural crest gene regulatory network from induction at the neural border to early migration, with experimental validation.

## 3.3. Role of miRNAs in NC development

It is known that a lot of genes are entangled in NC development, but not all of them have been determined. Recently, science focus moved from studying protein-coding genes to non-coding genes, e.g., miRNAs. These non-coding RNAs are important during embryo development regulating the expression of protein-coding genes, mostly by binding to the 3'-UTR of messenger RNA and suppressing the production of protein due to the destabilization of messenger RNA. Numerous analyses showed that miRNA-mediated gene regulation is important during embryo development (Bushati and Cohen, 2007; Stefani and Slack, 2008). Dicer mutant zebrafish embryos show defects in gastrulation and somitogenesis (Giraldez et al., 2005) and also die (Wienholds et al., 2003). In the retina, Dicer loss leads to retinal regression at later stages (Decembrini et al., 2008). Multiple analyses showed involvement of miRNAs during neural development (Papagiannakopoulos and Kosik, 2009; Walker and Harland, 2009). Therefore, we decided to develop this direction and expand our understanding of the NC GRN by studying miRNAs previously shown to be expressed in NC and neural tissue in the Xenopus embryos (Godden et al., 2021; Ward et al., 2018).

We selected miR-219 and miR-196a to determine their function in NC development we designed MOs that was designed to knock down the mature miRNA, and be complementary to it (Flynt et al., 2017). Through RNA-sequencing on dissected NC tissue that had been treated with MOs for miR-196a and miR-219, it was revealed that miRNAs may have a role in the development of NC and surrounding tissues. A strong reduction in expression of NC markers: *foxd3*, *twist1*, *sox8*, *sox9*, *sox10*, and *snail2* can be seen, following miRNA-KD, more significantly with miR-196a KD. Neural markers *sox2*, and *sox3* were significantly enriched. Neural plate border markers were also enriched (*gmnn*, *hes3*, *pax3*, and not significantly *zic1*) and also pluripotency marker *pou5f3* was enriched. Functional characterization of Sox10,

Snail2, and Pax3 together with the result of the RNA-seq suggest that miR-196a and miR-219 KD, which reduce the expression of *sox10* and *snail2*, prevent the maintenance of NC cells. The study highlights the importance of understanding the different roles of miR-196a and miR-219 in the development of NC cells in Xenopus embryos. Most importantly, the data presented in the study suggests that these two miRNAs have distinct effects on the development of NC cells, with miR-196a having a stronger effect on reducing NC marker expression and switching the cell fate into NB-like multipotent state, and miR-219 contributing to an increase in multipotency without activating the NB program. This suggests that miR-196a and miR-219 play different roles during the NC development.

## 3.4 Pax3, Dlx3 and Zic1 relationships in the Neural border

In order to precisely examine the immediate-early targets of the most important transcription factors in the neural border Pax3, Zic1, and Dlx3 we have started a new project by micro-dissection neural border explants which provide us with high spatial specificity. For the first step of the project (see Results section 2.4.1) we inspected for stage-specific immediate early targets of these transcription factors in the NB at multiple stages.

Up to this point, we identified a set of genes that are regulated by three transcription factors. Our findings confirm the previously known roles of these transcription factors in the regulation of genes involved in neural crest formation, such as markers of neural crest cells like *snai2*, *twist1*, and *sox9*, as well as *foxd3*, and *tfap2b*. It is also notable that the majority of the genes identified as targets of Pax3 were found to be downregulated in the absence of Pax3, supporting the idea that Pax3 plays a role in activating neural crest specifiers. However, there were also a small number of genes found to be upregulated. Moreover, we found that Pax3 and Dlx3 have a shared regulatory targets, that were found to play a role in formation of the neural crest, like *alx1* and *hnf1b*. Furthermore, our data suggests that the role of Pax3 in neural crest formation is more prominent at early stages, as we observed fewer targets for Pax3 at later stages. Also, using previous scRNA-seq ectoderm analysis, we found that Pax3-GOF activated both the NE→NP and NE/NB→NC programs, while Dlx3-GOF activated the NNE→NNE2 program but deactivated the NC→NP programs. Overall, our analysis showed distinct activation patterns for Pax3 and Dlx3. Overall, this study provides new insight into the transcriptional regulatory networks that control the formation of the neural crest. By identifying the specific genes and pathways that are targeted by these transcription factors, we can gain a

better understanding of the complex molecular processes involved in the development of NB, NC, PC and other ectodermal tissues.

The next goals of the project are [1] examining the cooperation between NB specifiers during NC/PC fate induction by defying the immediate early targets of the genes in the absence of another e.g., Zic1 targets in the absence of Pax3/Dlx3. And [2] scanning for stage-specific requirements of NB specifiers by analyzing NB transcriptome in the absence of one of the genes using morpholinos.

## 3.5 Cell type evolution with scEvoNet

For the past 150 years, cell types have been studied in terms of development, morphology and molecular features, whereas with the advent of single-cell transcriptomics, many distinct and hidden cell types have been identified. With the use of scRNA-seq, a completely new era is dawning in understanding the evolution of cell types. Seemingly, the ancestral types of animal cells were poly-functional (some sponge cell have both neural and immune properties), but during the evolution, functions were distributed between different cells (Arendt, 2008). This “division of labor” is one of the key ways how cell types evolve. Yet, there is a lack of scRNA-seq computational tools to study the evolution of cell states. Thus, we developed a computational tool scEvoNet, a software implemented to study this concept from scRNA-seq data. Firstly, our tool creates the matrix of similarities between cell types of two datasets provided by users. Secondly, it builds the bipartite network connecting genes and different cell states, which can be used as a source for screening potentially co-opted genes (shared between cell types), as well as genes that are important for the particular cell type in cross-species comparison, and genes that are meaningful only for one specie. Moreover, it can be used to study tumor progression and evolution by exploring genes and gene programs that are shared between various cell types in the tumor microenvironment. Using this tool for studying the continuum of transcriptome states between developmental stages or species, or between normal and disease states, will help explain the complex dynamics of cell states. Dropout events, where a gene is not detected despite being present, and batch effects can limit the accuracy of predictions made from scRNA-seq data. Generating sets of datasets from multiple species or using improved scRNA-seq capture techniques will improve the results of tools like scEvoNet.

## 3.6 Conclusions

Although the neural crest cell field is highly analyzed over the years with different cell and molecular methods, we still do not know how exactly is the neural crest fate initiated during dorsal ectoderm patterning. We do not know how neural crest cells acquire their regional diversification during the pre-migratory stage. The neural crest gene regulatory network consists only of transcription factors and its few targets, and previous analysis of single-cell transcriptomes has so far mainly been focused on subsets of neural crest cells after their emigration from the neural tube (Soldatov et al., 2019; Williams et al., 2019). The crucial steps of pre-migratory neural crest diversification have not been addressed with those approaches yet.

In my thesis research, I have aimed to explore various aspects of neural crest biology. This includes exploring the different models that can explain the development of ectodermal tissues (Binary competence, NB, Gradient NB models) and examining the underlying molecular causes of heterogeneity in NC cells during the very early stages of development. Furthermore, I have aimed to identify and characterize novel subpopulations within the neural crest cell cells.

Thus, we used a combination of a single-cell dataset of a developmental series of 8 stages spanning gastrulation and neurulation, perturbation-seq for the most important TFs and miRNAs to explore how the neural crest fate is initiated at the end of gastrulation and how its diversification arises during neurulation. We showed that during the process of gastrulation, the neural crest cells are formed in the neural plate and the border region of the ectoderm through two converging pathways, paralleled by two other pathways that lead to the initiation of placodes (from the neural border region and the non-neural ectoderm). We find essential early roles for Sox9, cooperating with neural border factors Pax3 and Zic1 to activate neural crest identity. Our findings suggest a “Dual convergence Model” for neural crest and placode emergence from dorsal ectoderm, thus merging the currently competing models of “Neural border origin” opposed to “Neural / non-neural ectoderm origin”. In comparison to the gradient NB model, we found no evidence of NB cells showing an inclination towards NE and NP cell fates, which does not necessarily mean that this is impossible, but rather that it is less likely than the NC and PC fates.

Our analysis reveals an unsuspected early NC heterogeneity arising just after induction. At stages 12-13 we observed 4 different subpopulations: two cranial, one vagal (early ENSp)

and one unbiased. Subsequently, using branching analysis, we generated a new lineage tree, starting from the dorsal ectoderm and connecting to the established migration patterns of the neural crest, specifically the ENSp, cardiac NC, and migrating cranial NC. We validated key bifurcation points using established assays of neural crest genetic perturbation, either in vivo, or in explants, or after microdissection, followed by RNA-sequencing or chromatin-immunoprecipitation and sequencing. Additionally, using these experimental manipulations of four selected transcription factors, Pax3, Zic1, TFAP2a, and TFAP2e, which display reiterative functions in the neural crest tree, we validate numerous aspects of the predicted gene-regulatory network.

Our integrated approach from scRNA-seq predictions to in vivo functional validation on a large scale is an important contribution to the field, as it establishes powerful strategies to validate tissue-specific GRNs globally rather than on a gene-by-gene basis. This allowed us to provide a comprehensive GRN for neural crest and dorsal ectoderm validated sets of target genes for four pivotal regulators. In sum, our work uncovers several important features of neural crest biology, such as the persistence of an unbiased (multipotent) lineage throughout neurulation, together with the emergence of cranial and vagal cells, and the gene programs likely to sustain each nascent branch. Our research addresses significant and unresolved issues in the field of neural crest biology and lays the foundation for future experimentation.

In the second part, I have participated in two projects that use bulk RNA-seq on micro-dissected tissues to explore [1] the precise roles of two microRNAs (miRNA-219, miRNA-196a) during NC development in *Xenopus* embryos, and [2] the relationships between the essential NB transcription factors (Pax3, Dlx3, Zic1) during NB/NC patterning. The final results after all validations will highlight the importance of microRNAs in regulating gene expression during neural crest development, the distinct roles of different microRNAs in induction and specification events during neural crest development and the nature of cooperation of Pax3, Dlx3 and Zic1 in regulating gene expression during NC/PC induction.

Last, I have sought to develop a computational tool called scEvoNet, which is implemented to study the evolution of cell states from scRNA-seq data. It can be used to study tumor progression and evolution by exploring genes and gene programs that are shared between various cell types. Furthermore, this tool can also be implemented on a broader set of data, including multiple species, which could provide new insights into the development and evolutionary history of different cell types. The tool has great potential for understanding the

dynamics and evolution of cell types and largely dependent on the data quality and data availability.

Collectively, my work aims at significantly enhancing the understanding of the molecular mechanisms involved in the formation of the neural crest. While many questions have been answered, there are still areas for exploration, such as investigating the transition from NP-like NC cells to ENSp cells. Furthermore, applying scEvoNet on the data for a variety of species, including cephalochordates, tunicates, and vertebrates would provide valuable insights.

# References


Ahmed A, Ward NJ, Moxon S, Lopez-Gomollon S, Viaut C, Tomlinson ML, Patrushev I, Gilchrist MJ, Dalmay T, Dotlic D, Münsterberg AE, Wheeler GN. 2015. A Database of microRNA Expression Patterns in Xenopus laevis. *PLOS ONE* **10**:e0138313. doi:10.1371/journal.pone.0138313

Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine J-C, Geurts P, Aerts J, van den Oord J, Atak ZK, Wouters J, Aerts S. 2017. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**:1083–1086. doi:10.1038/nmeth.4463

Alberti C, Cochella L. 2017. A framework for understanding the roles of miRNAs in animal development. *Development* **144**:2548–2559. doi:10.1242/dev.146613

Alkobtawi M, Pla P, Monsoro-Burq AH. 2021. BMP signaling is enhanced intracellularly by FHL3 controlling WNT-dependent spatiotemporal emergence of the neural crest. *Cell Reports* **35**:109289. doi:10.1016/j.celrep.2021.109289

Alkobtawi M., Monsoro-Burq A. H. 2020. "Chapter 1: The neural crest, a vertebrate invention" in Evolving neural crest cells. 1st Edn. eds. Eames B. F., Medeiros D. M., Adameyko I. (Boca Raton: Imprint CRC Press), 5–66.

Alkobtawi M, Ray H, Barriga EH, Moreno M, Kerney R, Monsoro-Burq A-H, Saint-Jeannet J-P, Mayor R. 2018. Characterization of Pax3 and Sox10 transgenic Xenopus laevis embryos as tools to study neural crest development. *Dev Biol* **444 Suppl 1**:S202–S208. doi:10.1016/j.ydbio.2018.02.020

Antonaci M, Wheeler GN. 2022. MicroRNAs in neural crest development and neurocristopathies. *Biochem Soc Trans* **50**:965–974. doi:10.1042/BST20210828

Arendt D. 2008. The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet* **9**:868–882. doi:10.1038/nrg2416

Artinger KB, Monsoro-Burq AH. 2021. Neural crest multipotency and specification: power and limits of single cell transcriptomic approaches. *Fac Rev* **10**:38. doi:10.12703/r/10-38

Bachetti T, Bagnasco S, Piumelli R, Palmieri A, Ceccherini I. 2021. A Common 3′UTR Variant of the PHOX2B Gene Is Associated With Infant Life-Threatening and Sudden Death Events in the Italian Population. *Front Neurol* **12**:642735. doi:10.3389/fneur.2021.642735

Bae C-J, Park B-Y, Lee Y-H, Tobias JW, Hong C-S, Saint-Jeannet J-P. 2014. Identification of Pax3 and Zic1 targets in the developing neural crest. *Dev Biol* **386**:473–483. doi:10.1016/j.ydbio.2013.12.011

Baggiolini A, Varum S, Mateos JM, Bettosini D, John N, Bonalli M, Ziegler U, Dimou L, Clevers H, Furrer R, Sommer L. 2015. Premigratory and migratory neural crest cells are multipotent in vivo. *Cell Stem Cell* **16**:314–322. doi:10.1016/j.stem.2015.02.017

Bahm I, Barriga EH, Frolov A, Theveneau E, Frankel P, Mayor R. 2017. PDGF controls contact inhibition of locomotion by regulating N-cadherin during neural crest migration. *Development* **144**:2456–2468. doi:10.1242/dev.147926

Bar-Joseph Z, Gitter A, Simon I. 2012. Studying and modelling dynamic biological processes using time-series gene expression data. *Nat Rev Genet* **13**:552–564. doi:10.1038/nrg3244

Baroffio A, Blot M. 1992. Statistical evidence for a random commitment of pluripotent cephalic neural crest cells. *J Cell Sci* **103 ( Pt 2)**:581–587. doi:10.1242/jcs.103.2.581

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**:281–297. doi:10.1016/s0092-8674(04)00045-5

Baruzzo G, Hayer KE, Kim EJ, Di Camillo B, FitzGerald GA, Grant GR. 2017. Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat Methods* **14**:135–139. doi:10.1038/nmeth.4106

Basch ML, Bronner-Fraser M, García-Castro MI. 2006. Specification of the neural crest occurs during gastrulation and requires Pax7. *Nature* **441**:218–222. doi:10.1038/nature04684

Beadnell TC, Scheid AD, Vivian CJ, Welch DR. 2018. Roles of the mitochondrial genetics in cancer metastasis: Not to be ignored any longer. *Cancer Metastasis Rev* **37**:615–632. doi:10.1007/s10555-018-9772-7

Bellmeyer A, Krase J, Lindgren J, LaBonne C. 2003. The Protooncogene c-Myc Is an Essential Regulator of Neural Crest Formation in Xenopus. *Developmental Cell* **4**:827–839. doi:10.1016/S1534-5807(03)00160-6

Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, Mills AA, Elledge SJ, Anderson KV, Hannon GJ. 2003. Dicer is essential for mouse development. *Nat Genet* **35**:215–217. doi:10.1038/ng1253

Betancur P, Bronner-Fraser M, Sauka-Spengler T. 2010. ASSEMBLING NEURAL CREST REGULATORY CIRCUITS INTO A GENE REGULATORY NETWORK. *Annu Rev Cell Dev Biol* **26**:581–603. doi:10.1146/annurev.cellbio.042308.113245

Bhaskaran M, Mohan M. 2014. MicroRNAs: history, biogenesis, and their evolving role in animal development and disease. *Vet Pathol* **51**:759–774. doi:10.1177/0300985813502820

Bhat N, Kwon H-J, Riley BB. 2013. A gene network that coordinates preplacodal competence and neural crest specification in zebrafish. *Dev Biol* **373**:107–117. doi:10.1016/j.ydbio.2012.10.012

Bhattacharya D, Rothstein M, Azambuja AP, Simoes-Costa M. 2018. Control of neural crest multipotency by Wnt signaling and the Lin28/let-7 axis. *eLife* **7**:e40556. doi:10.7554/eLife.40556

Blitz IL, Paraiso KD, Patrushev I, Chiu WTY, Cho KWY, Gilchrist MJ. 2017. A catalog of Xenopus tropicalis transcription factors and their regional expression in the early gastrula stage embryo. *Dev Biol* **426**:409–417. doi:10.1016/j.ydbio.2016.07.002

Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, Thorsson V. 2006. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol* **7**:R36. doi:10.1186/gb-2006-7-5-r36

Breiman L. 2001. Random Forests. *Machine Learning* **45**:5–32. doi:10.1023/A:1010933404324

Briggs JA, Weinreb C, Wagner DE, Megason S, Peshkin L, Kirschner MW, Klein AM. 2018. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**:eaar5780. doi:10.1126/science.aar5780

Bronner-Fraser M, Fraser SE. 1988. Cell lineage analysis reveals multipotency of some avian neural crest cells. *Nature* **335**:161–164. doi:10.1038/335161a0

Buitrago-Delgado E, Nordin K, Rao A, Geary L, LaBonne C. 2015. Shared regulatory programs suggest retention of blastula-stage potential in neural crest cells. *Science* **348**:1332–1335. doi:10.1126/science.aaa3655

Bushati N, Cohen SM. 2007. microRNA functions. *Annu Rev Cell Dev Biol* **23**:175–205. doi:10.1146/annurev.cellbio.23.090506.123406

Challa-Malladi M, Lieu YK, Califano O, Holmes A, Bhagat G, Murty VV, Dominguez-Sola D, Pasqualucci L, Dalla-Favera R. 2011. Combined Genetic Inactivation of Beta2-Microglobulin and CD58 Reveals Frequent Escape from Immune Recognition in Diffuse Large B-cell Lymphoma. *Cancer Cell* **20**:728–740. doi:10.1016/j.ccr.2011.11.006

Chandra S, Vimal D, Sharma D, Rai V, Gupta SC, Chowdhuri DK. 2017. Role of miRNAs in development and disease: Lessons learnt from small organisms. *Life Sciences* **185**:8–14. doi:10.1016/j.lfs.2017.07.017

Cheung M, Briscoe J. 2003. Neural crest development is regulated by the transcription factor Sox9. *Development* **130**:5681–5693. doi:10.1242/dev.00808

Cheung M, Chaboissier M-C, Mynett A, Hirst E, Schedl A, Briscoe J. 2005. The transcriptional control of trunk neural crest induction, survival, and delamination. *Dev Cell* **8**:179–192. doi:10.1016/j.devcel.2004.12.010

Coles EG, Taneyhill LA, Bronner-Fraser M. 2007. A critical role for Cadherin6B in regulating avian neural crest emigration. *Developmental Biology* **312**:533–544. doi:10.1016/j.ydbio.2007.09.056

Davidson EH. 2010. Emerging properties of animal gene regulatory networks. *Nature* **468**:911–920. doi:10.1038/nature09645

Davidson EH. 2009. Network design principles from the sea urchin embryo. *Current Opinion in Genetics & Development*, Genomes and evolution **19**:535–540. doi:10.1016/j.gde.2009.10.007

Davies JOJ, Telenius JM, McGowan SJ, Roberts NA, Taylor S, Higgs DR, Hughes JR. 2016. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods* **13**:74–80. doi:10.1038/nmeth.3664

de Crozé N, Maczkowiak F, Monsoro-Burq AH. 2011. Reiterative AP2a activity controls sequential steps in the neural crest gene regulatory network. *Proc Natl Acad Sci U S A* **108**:155–160. doi:10.1073/pnas.1010740107

Decembrini S, Andreazzoli M, Barsacchi G, Cremisi F. 2008. Dicer inactivation causes heterochronic retinogenesis in Xenopus laevis. *Int J Dev Biol* **52**:1099–1103. doi:10.1387/ijdb.082646sd

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. doi:10.1093/bioinformatics/bts635

Dottori M, Gross MK, Labosky P, Goulding M. 2001. The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. *Development* **128**:4127–4138. doi:10.1242/dev.128.21.4127

Du Q, de la Morena MT, van Oers NSC. 2020. The Genetics and Epigenetics of 22q11.2 Deletion Syndrome. *Front Genet* **10**:1365. doi:10.3389/fgene.2019.01365

Dupin E, Calloni GW, Coelho-Aguiar JM, Le Douarin NM. 2018. The issue of the multipotency of the neural crest cells. *Dev Biol* **444 Suppl 1**:S47–S59. doi:10.1016/j.ydbio.2018.03.024

Dupin E, Coelho-Aguiar JM. 2013. Isolation and differentiation properties of neural crest stem cells. *Cytometry A* **83**:38–47. doi:10.1002/cyto.a.22098

Dupin E, Sommer L. 2012. Neural crest progenitors and stem cells: from early development to adulthood. *Dev Biol* **366**:83–95. doi:10.1016/j.ydbio.2012.02.035

Eisen JS, Smith JC. 2008. Controlling morpholino experiments: don't stop making antisense. *Development* **135**:1735–1743. doi:10.1242/dev.001115

Endo Y, Osumi N, Wakamatsu Y. 2002. Bimodal functions of Notch-mediated signaling are involved in neural crest formation during avian ectoderm development. *Development* **129**:863–873. doi:10.1242/dev.129.4.863

Engleka KA, Epstein JA. 2016. Neural Crest Formation and Craniofacial Development In: Erickson RP, Wynshaw-Boris AJ, editors. Epstein's Inborn Errors of Development: The Molecular Basis of Clinical Disorders of Morphogenesis. Oxford University Press. p. 0. doi:10.1093/med/9780199934522.003.0006

Eraslan G, Simon LM, Mircea M, Mueller NS, Theis FJ. 2019. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun* **10**:390. doi:10.1038/s41467-018-07931-2

Erwin DH, Davidson EH. 2009. The evolution of hierarchical gene regulatory networks. *Nat Rev Genet* **10**:141–148. doi:10.1038/nrg2499

Evsen L, Li X, Zhang S, Razin S, Doetzlhofer A. 2020. let-7 miRNAs inhibit CHD7 expression and control auditory-sensory progenitor cell behavior in the developing inner ear. *Development* **147**:dev183384. doi:10.1242/dev.183384

Faure L, Soldatov R, Kharchenko PV, Adameyko I. 2023. scFates: a scalable python package for advanced pseudotime and bifurcation analysis from single-cell data. *Bioinformatics* **39**:btac746. doi:10.1093/bioinformatics/btac746

Flynt AS, Rao M, Patton JG. 2017. Blocking Zebrafish MicroRNAs with Morpholinos. *Methods Mol Biol* **1565**:59–78. doi:10.1007/978-1-4939-6817-6_6

Franz T. 1993. The Splotch (Sp1H) and Splotch-delayed (Spd) alleles: differential phenotypic effects on neural crest and limb musculature. *Anat Embryol (Berl)* **187**:371–377. doi:10.1007/BF00185895

Gao M, Ling M, Tang X, Wang S, Xiao X, Qiao Y, Yang W, Yu R. 2021. Comparison of high-throughput single-cell RNA sequencing data processing pipelines. *Briefings in Bioinformatics* **22**:bbaa116. doi:10.1093/bib/bbaa116

García-Castro MI, Marcelle C, Bronner-Fraser M. 2002. Ectodermal Wnt function as a neural crest inducer. *Science* **297**:848–851. doi:10.1126/science.1070824

Garnett AT, Square TA, Medeiros DM. 2012. BMP, Wnt and FGF signals are integrated through evolutionarily conserved enhancers to achieve robust expression of Pax3 and Zic genes at the zebrafish neural plate border. *Development* **139**:4220–4231. doi:10.1242/dev.081497

Gerety SS, Wilkinson DG. 2011. Morpholino artifacts provide pitfalls and reveal a novel role for pro-apoptotic genes in hindbrain boundary development. *Dev Biol* **350**:279–289. doi:10.1016/j.ydbio.2010.11.030

Gessert S, Bugner V, Tecza A, Pinker M, Kühl M. 2010. FMR1/FXR1 and the miRNA pathway are required for eye and neural crest development. *Dev Biol* **341**:222–235. doi:10.1016/j.ydbio.2010.02.031

Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF. 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**:833–838. doi:10.1126/science.1109020

Godden AM, Ward NJ, van der Lee M, Abu-Daya A, Guille M, Wheeler GN. 2021. An efficient miRNA knockout approach using CRISPR-Cas9 in Xenopus. doi:10.1101/2021.08.05.454468

Gouignard N, Maccarana M, Strate I, von Stedingk K, Malmström A, Pera EM. 2016. Musculocontractural Ehlers-Danlos syndrome and neurocristopathies: dermatan sulfate is required for Xenopus neural crest cells to migrate and adhere to fibronectin. *Dis Model Mech* **9**:607–620. doi:10.1242/dmm.024661

Grocott T, Lozano-Velasco E, Mok GF, Münsterberg AE. 2020. The Pax6 master control gene initiates spontaneous retinal development via a self-organising Turing network. *Development* **147**:dev185827. doi:10.1242/dev.185827

Grocott T, Tambalo M, Streit A. 2012. The peripheral sensory nervous system in the vertebrate head: a gene regulatory perspective. *Dev Biol* **370**:3–23. doi:10.1016/j.ydbio.2012.06.028

Groves AK, LaBonne C. 2014. Setting appropriate boundaries: Fate, patterning and competence at the neural plate border. *Dev Biol* **389**:2–12. doi:10.1016/j.ydbio.2013.11.027

Hanna J, Hossain GS, Kocerha J. 2019. The Potential for microRNA Therapeutics and Clinical Research. *Front Genet* **10**:478. doi:10.3389/fgene.2019.00478

Henion PD, Weston JA. 1997. Timing and pattern of cell fate restrictions in the neural crest lineage. *Development* **124**:4351–4359. doi:10.1242/dev.124.21.4351

Hernandez-Lagunas L, Powell D, Law J, Grant K, Artinger KB. 2011. prdm1a and olig4 act downstream of Notch signaling to regulate cell fate at the neural plate border. *Dev Biol* **356**:496–505. doi:10.1016/j.ydbio.2011.06.005

Holland CH, Tanevski J, Perales-Patón J, Gleixner J, Kumar MP, Mereu E, Joughin BA, Stegle O, Lauffenburger DA, Heyn H, Szalai B, Saez-Rodriguez J. 2020. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol* **21**:36. doi:10.1186/s13059-020-1949-z

Holland ND, Chen J. 2001. Origin and early evolution of the vertebrates: new insights from advances in molecular biology, anatomy, and palaeontology. *Bioessays* **23**:142–151. doi:10.1002/1521-1878(200102)23:2<142::AID-BIES1021>3.0.CO;2-5

Holland ND, Panganiban G, Henyey EL, Holland LZ. 1996. Sequence and developmental expression of AmphiDll, an amphioxus Distal-less gene transcribed in the ectoderm, epidermis and nervous system: insights into evolution of craniate forebrain and neural crest. *Development* **122**:2911–2920. doi:10.1242/dev.122.9.2911

Hong C-S, Devotta A, Lee Y-H, Park B-Y, Saint-Jeannet J-P. 2014. Transcription factor AP2 epsilon (Tfap2e) regulates neural crest specification in Xenopus. *Dev Neurobiol* **74**:894–906. doi:10.1002/dneu.22173

Hong C-S, Saint-Jeannet J-P. 2007. The Activity of Pax3 and Zic1 Regulates Three Distinct Cell Fates at the Neural Plate Border. *MBoC* **18**:2192–2202. doi:10.1091/mbc.e06-11-1047

Hong C-S, Saint-Jeannet J-P. 2005. Sox proteins and neural crest development. *Semin Cell Dev Biol* **16**:694–703. doi:10.1016/j.semcdb.2005.06.005

Honoré SM, Aybar MJ, Mayor R. 2003. Sox10 is required for the early development of the prospective neural crest in Xenopus embryos. *Dev Biol* **260**:79–96. doi:10.1016/s0012-1606(03)00247-1

Hovland AS, Bhattacharya D, Azambuja AP, Pramio D, Copeland J, Rothstein M, Simoes-Costa M. 2022. Pluripotency factors are repurposed to shape the epigenomic landscape of neural crest cells. *Developmental Cell* **57**:2257-2272.e5. doi:10.1016/j.devcel.2022.09.006

Hu M, Krause D, Greaves M, Sharkis S, Dexter M, Heyworth C, Enver T. 1997. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev* **11**:774–785. doi:10.1101/gad.11.6.774

Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. 2010. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**:e12776. doi:10.1371/journal.pone.0012776

Imayoshi I, Isomura A, Harima Y, Kawaguchi K, Kori H, Miyachi H, Fujiwara T, Ishidate F, Kageyama R. 2013. Oscillatory control of factors determining multipotency and fate in mouse neural progenitors. *Science* **342**:1203–1208. doi:10.1126/science.1242366

Jänicke M, Carney TJ, Hammerschmidt M. 2007. Foxi3 transcription factors and Notch signaling control the formation of skin ionocytes from epidermal precursors of the zebrafish embryo. *Dev Biol* **307**:258–271. doi:10.1016/j.ydbio.2007.04.044

Jeffery WR. 2007. Chordate ancestry of the neural crest: New insights from ascidians. *Seminars in Cell & Developmental Biology* **18**:481–491. doi:10.1016/j.semcdb.2007.04.005

Kalcheim C, Kumar D. 2017. Cell fate decisions during neural crest ontogeny. *Int J Dev Biol* **61**:195–203. doi:10.1387/ijdb.160196ck

Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision TreeAdvances in Neural Information Processing Systems. Curran Associates, Inc.

Kelsh RN, Sosa KC, Owen JP, Yates CA. 2017. Zebrafish adult pigment stem cells are multipotent and form pigment cells by a progressive fate restriction process. *BioEssays* **39**:1600234. doi:10.1002/bies.201600234

Khudyakov J, Bronner-Fraser M. 2009. Comprehensive spatiotemporal analysis of early chick neural crest network genes. *Dev Dyn* **238**:716–723. doi:10.1002/dvdy.21881

Kim J, Lo L, Dormand E, Anderson DJ. 2003. SOX10 Maintains Multipotency and Inhibits Neuronal Differentiation of Neural Crest Stem Cells. *Neuron* **38**:17–31. doi:10.1016/S0896-6273(03)00163-6

Kim J, Piao H-L, Kim B-J, Yao F, Han Z, Wang Y, Xiao Z, Siverly AN, Lawhon SE, Ton BN, Lee H, Zhou Z, Gan B, Nakagawa S, Ellis MJ, Liang H, Hung M-C, You MJ, Sun Y, Ma L. 2018. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. *Nat Genet* **50**:1705–1715. doi:10.1038/s41588-018-0252-3

Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. 2015. Droplet barcoding for single cell transcriptomics applied to embryonic stem cells. *Cell* **161**:1187–1201. doi:10.1016/j.cell.2015.04.044

Klymkowsky MW, Rossi CC, Artinger KB. 2010. Mechanisms driving neural crest induction and migration in the zebrafish and Xenopus laevis. *Cell Adh Migr* **4**:595–608. doi:10.4161/cam.4.4.12962

Kolm PJ, Sive HL. 1995. Efficient Hormone-Inducible Protein Function in Xenopus laevis. *Developmental Biology* **171**:267–272. doi:10.1006/dbio.1995.1279

Kotov A, Alkobtawi M, Seal S, Kappès V, Ruiz SM, Arbès H, Harland R, Peshkin L, Monsoro-Burq AH. 2022a. From neural border to migratory stage: A comprehensive single cell roadmap of the timing and regulatory logic driving cranial and vagal neural crest emergence. doi:10.1101/2022.03.23.485460

Kotov A, Zinovyev A, Monsoro-Burq A-H. 2022b. scEvoNet: a gradient boosting-based method for prediction of cell state evolution. doi:10.1101/2022.12.07.519467

Kwon H-J, Bhat N, Sweet EM, Cornell RA, Riley BB. 2010. Identification of Early Requirements for Preplacodal Ectoderm and Sensory Organ Development. *PLoS Genet* **6**:e1001133. doi:10.1371/journal.pgen.1001133

La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, Fan J, Borm LE, Liu Z, van Bruggen D, Guo J, He X, Barker R, Sundström E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S, Kharchenko PV. 2018. RNA velocity of single cells. *Nature* **560**:494–498. doi:10.1038/s41586-018-0414-6

Lahav R, Dupin E, Lecoin L, Glavieux C, Champeval D, Ziller C, Le Douarin NM. 1998. Endothelin 3 selectively promotes survival and proliferation of neural crest-derived glial and melanocytic precursors in vitro. *Proc Natl Acad Sci U S A* **95**:14214–14219.

Lander R, Nasr T, Ochoa SD, Nordin K, Prasad MS, LaBonne C. 2013. Interactions between Twist and other core epithelial–mesenchymal transition factors are controlled by GSK3-mediated phosphorylation. *Nat Commun* **4**:1542. doi:10.1038/ncomms2543

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**:559. doi:10.1186/1471-2105-9-559

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**:R25. doi:10.1186/gb-2009-10-3-r25

Le Douarin N, Kalcheim C. 1999. The Neural Crest, 2nd ed, Developmental and Cell Biology Series. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511897948

Le Douarin NM, Teillet MA. 1973. The migration of neural crest cells to the wall of the digestive tract in avian embryo. *J Embryol Exp Morphol* **30**:31–48.

Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**:843–854. doi:10.1016/0092-8674(93)90529-y

Levine M, Davidson EH. 2005. Gene regulatory networks for development. *Proceedings of the National Academy of Sciences* **102**:4936–4942. doi:10.1073/pnas.0408031102

Li E, Davidson EH. 2009. Building developmental gene regulatory networks. *Birth Defects Research Part C: Embryo Today: Reviews* **87**:123–130. doi:10.1002/bdrc.20152

Liem KF, Tremml G, Roelink H, Jessell TM. 1995. Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell* **82**:969–979. doi:10.1016/0092-8674(95)90276-7

Ling ITC, Sauka-Spengler T. 2019. Early chromatin shaping predetermines multipotent vagal neural crest into neural, neuronal and mesenchymal lineages. *Nat Cell Biol* **21**:1504–1517. doi:10.1038/s41556-019-0428-9

Liu C, Yang Z, Li D, Liu Z, Miao X, Yang L, Zou Q, Yuan Y. 2015. Overexpression of B2M and loss of ALK7 expression are associated with invasion, metastasis, and poor-prognosis of the pancreatic ductal adenocarcinoma. *Cancer Biomark* **15**:735–743. doi:10.3233/CBM-150515

Liu Y, Jin Y, Li J, Seto E, Kuo E, Yu W, Schwartz RJ, Blazo M, Zhang SL, Peng X. 2013. Inactivation of Cdc42 in neural crest cells causes craniofacial and cardiovascular morphogenesis defects. *Dev Biol* **383**:239–252. doi:10.1016/j.ydbio.2013.09.013

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550. doi:10.1186/s13059-014-0550-8

Lukoseviciute M, Gavriouchkina D, Williams RM, Hochgreb-Hagele T, Senanayake U, Chong-Morrison V, Thongjuea S, Repapi E, Mead A, Sauka-Spengler T. 2018. From Pioneer to Repressor: Bimodal foxd3 Activity Dynamically Remodels Neural Crest Regulatory Landscape In Vivo. *Dev Cell* **47**:608-628.e6. doi:10.1016/j.devcel.2018.11.009

Luo R, Gao J, Wehrle-Haller B, Henion PD. 2003. Molecular identification of distinct neurogenic and melanogenic neural crest sublineages. *Development* **130**:321–330. doi:10.1242/dev.00213

Luo T, Lee Y-H, Saint-Jeannet J-P, Sargent TD. 2003. Induction of neural crest in Xenopus by transcription factor AP2α. *Proc Natl Acad Sci U S A* **100**:532–537. doi:10.1073/pnas.0237226100

Luo T, Matsuo-Takasaki M, Sargent TD. 2001. Distinct roles for Distal-less genes Dlx3 and Dlx5 in regulating ectodermal development in Xenopus. *Mol Reprod Dev* **60**:331–337. doi:10.1002/mrd.1095

Maharana SK, Schlosser G. 2018. A gene regulatory network underlying the formation of pre-placodal ectoderm in Xenopus laevis. *BMC Biol* **16**:79. doi:10.1186/s12915-018-0540-5

Mandalos N, Rhinn M, Granchi Z, Karampelas I, Mitsiadis T, Economides AN, Dollé P, Remboutsika E. 2014. Sox2 acts as a rheostat of epithelial to mesenchymal transition during neural crest development. *Frontiers in Physiology* **5**.

Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Kellis M, Collins JJ, Stolovitzky G. 2012. Wisdom of crowds for robust gene network inference. *Nat Methods* **9**:796–804. doi:10.1038/nmeth.2016

Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, Califano A. 2006. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a

Mammalian Cellular Context. *BMC Bioinformatics* **7**:S7. doi:10.1186/1471-2105-7-S1-S7

Martinez-Morales J-R, Henrich T, Ramialison M, Wittbrodt J. 2007. New genes in the evolution of the neural crest differentiation program. *Genome Biology* **8**:R36. doi:10.1186/gb-2007-8-3-r36

Matsumoto H, Kiryu H, Furusawa C, Ko MSH, Ko SBH, Gouda N, Hayashi T, Nikaido I. 2017. SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* **33**:2314–2321. doi:10.1093/bioinformatics/btx194

Matsuo-Takasaki M, Matsumura M, Sasai Y. 2005. An essential role of Xenopus Foxi1a for ventral specification of the cephalic ectoderm during gastrulation. *Development* **132**:3885–3894. doi:10.1242/dev.01959

McLarren KW, Litsiou A, Streit A. 2003. DLX5 positions the neural crest and preplacode region at the border of the neural plate. *Dev Biol* **259**:34–47. doi:10.1016/s0012-1606(03)00177-5

Medina-Cuadra L, Monsoro-Burq AH. 2021. Xenopus, an emerging model for studying pathologies of the neural crest. *Curr Top Dev Biol* **145**:313–348. doi:10.1016/bs.ctdb.2021.03.002

Meulemans D, Bronner-Fraser M. 2004. Gene-regulatory interactions in neural crest evolution and development. *Dev Cell* **7**:291–299. doi:10.1016/j.devcel.2004.08.007

Milet C, Maczkowiak F, Roche DD, Monsoro-Burq AH. 2013. Pax3 and Zic1 drive induction and differentiation of multipotent, migratory, and functional neural crest in Xenopus embryos. *Proc Natl Acad Sci U S A* **110**:5528–5533. doi:10.1073/pnas.1219124110

Mir A, Kofron M, Zorn AM, Bajzer M, Haque M, Heasman J, Wylie CC. FoxI1e activates ectoderm formation and controls cell position in the Xenopus blastula. Development. 2007 Feb;134(4):779-88. doi: 10.1242/dev.02768.

Mizuseki K, Kishi M, Matsui M, Nakanishi S, Sasai Y. 1998. Xenopus Zic-related-1 and Sox-2, two factors induced by chordin, have distinct activities in the initiation of neural induction. *Development* **125**:579–587. doi:10.1242/dev.125.4.579

Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, Aerts S. 2019. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**:2159–2161. doi:10.1093/bioinformatics/bty916

Mok GF, Lozano-Velasco E, Münsterberg A. 2017. microRNAs in skeletal muscle development. *Semin Cell Dev Biol* **72**:67–76. doi:10.1016/j.semcdb.2017.10.032

Monsoro-Burq A-H, Wang E, Harland R. 2005. Msx1 and Pax3 cooperate to mediate FGF8 and WNT signals during Xenopus neural crest induction. *Dev Cell* **8**:167–178. doi:10.1016/j.devcel.2004.12.017

Morrison JA, McLennan R, Wolfe LA, Gogol MM, Meier S, McKinney MC, Teddy JM, Holmes L, Semerad CL, Box AC, Li H, Hall KE, Perera AG, Kulesa PM. 2017. Single-cell transcriptome analysis of avian neural crest migration reveals signatures of invasion and molecular transitions. *Elife* **6**:e28415. doi:10.7554/eLife.28415

Muñoz WA, Trainor PA. 2015. Neural crest cell evolution: how and when did a neural crest cell become a neural crest cell. *Curr Top Dev Biol* **111**:3–26. doi:10.1016/bs.ctdb.2014.11.001

Nakata K, Nagai T, Aruga J, Mikoshiba K. 1997. Xenopus Zic3, a primary regulator both in neural and neural crest development. *Proceedings of the National Academy of Sciences* **94**:11980–11985. doi:10.1073/pnas.94.22.11980

Neural Crest-Derived Stem Cells - Bielefeld University. n.d. https://www.uni-bielefeld.de/fakultaeten/biologie/forschung/arbeitsgruppen/cellbiology/research/neural-crest-derived-stem/

Nichane M, Ren X, Souopgui J, Bellefroid EJ. 2008. Hairy2 functions through both DNA-binding and non DNA-binding mechanisms at the neural plate border in Xenopus. *Developmental Biology* **322**:368–380. doi:10.1016/j.ydbio.2008.07.026

Panopoulou GD, Clark MD, Holland LZ, Lehrach H, Holland ND. 1998. AmphiBMP2/4, an amphioxus bone morphogenetic protein closely related to Drosophila decapentaplegic and vertebrate BMP2 and BMP4: insights into evolution of dorsoventral axis specification. *Dev Dyn* **213**:130–139. doi:10.1002/(SICI)1097-0177(199809)213:1<130::AID-AJA13>3.0.CO;2-6

Papagiannakopoulos T, Kosik KS. 2009. MicroRNA-124: Micromanager of Neurogenesis. *Cell Stem Cell* **4**:375–376. doi:10.1016/j.stem.2009.04.007

Park PJ. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**:669–680. doi:10.1038/nrg2641

Pasini A, Amiel A, Rothbächer U, Roure A, Lemaire P, Darras S. 2006. Formation of the Ascidian Epidermal Sensory Neurons: Insights into the Origin of the Chordate Peripheral Nervous System. *PLoS Biology* **4**:e225. doi:10.1371/journal.pbio.0040225

Pegoraro C, Monsoro-Burq AH. 2013. Signaling and transcriptional regulation in neural crest specification and migration: lessons from xenopus embryos. *Wiley Interdiscip Rev Dev Biol* **2**:247–259. doi:10.1002/wdev.76

Penfold CA, Buchanan-Wollaston V, Denby KJ, Wild DL. 2012. Nonparametric Bayesian inference for perturbed and orthologous gene regulatory networks. *Bioinformatics* **28**:i233–i241. doi:10.1093/bioinformatics/bts222

Petukhov V, Guo J, Baryawno N, Severe N, Scadden DT, Samsonova MG, Kharchenko PV. 2018. dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. *Genome Biology* **19**:78. doi:10.1186/s13059-018-1449-6

Pieper M, Ahrens K, Rink E, Peter A, Schlosser G. 2012. Differential distribution of competence for panplacodal and neural crest induction to non-neural and neural ectoderm. *Development* **139**:1175–1187. doi:10.1242/dev.074468

Plouhinec J-L, Medina-Ruiz S, Borday C, Bernard E, Vert J-P, Eisen MB, Harland RM, Monsoro-Burq AH. 2017. A molecular atlas of the developing ectoderm defines neural, neural crest, placode, and nonneural progenitor identity in vertebrates. *PLOS Biology* **15**:e2004045. doi:10.1371/journal.pbio.2004045

Plouhinec J-L, Roche DD, Pegoraro C, Figueiredo A-L, Maczkowiak F, Brunet LJ, Milet C, Vert J-P, Pollet N, Harland RM, Monsoro-Burq AH. 2014. Pax3 and Zic1 trigger the early neural crest gene regulatory network by the direct activation of multiple key neural crest specifiers. *Dev Biol* **386**:461–472. doi:10.1016/j.ydbio.2013.12.010

Powell DR, Hernandez-Lagunas L, LaMonica K, Artinger KB. 2013. Prdm1a directly activates foxd3 and tfap2a during zebrafish neural crest specification. *Development* **140**:3445–3455. doi:10.1242/dev.096164

Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. 2020. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods* **17**:147–154. doi:10.1038/s41592-019-0690-6

Qian J, Lin J, Luscombe NM, Yu H, Gerstein M. 2003. Prediction of regulatory networks: genome-wide identification of transcription factor targets from gene expression data. *Bioinformatics* **19**:1917–1926. doi:10.1093/bioinformatics/btg347

Raible DW, Ragland JW. 2005. Reiterated Wnt and BMP signals in neural crest development. *Semin Cell Dev Biol* **16**:673–682. doi:10.1016/j.semcdb.2005.06.008

Reedy MV, Faraco CD, Erickson CA. 1998. The delayed entry of thoracic neural crest cells into the dorsolateral path is a consequence of the late emigration of melanogenic neural crest cells from the neural tube. *Dev Biol* **200**:234–246. doi:10.1006/dbio.1998.8963

Robinson GW, Mahon KA. 1994. Differential and overlapping expression domains of Dlx-2 and Dlx-3 suggest distinct roles for Distal-less homeobox genes in craniofacial development. *Mech Dev* **48**:199–215. doi:10.1016/0925-4773(94)90060-4

Roellig D, Tan-Cabugao J, Esaian S, Bronner ME. 2017. Dynamic transcriptional signature and cell fate analysis reveals plasticity of individual neural plate border cells. *Elife* **6**:e21620. doi:10.7554/eLife.21620

Rooda I, Hensen K, Kaselt B, Kasvandik S, Pook M, Kurg A, Salumets A, Velthut-Meikas A. 2020. Target prediction and validation of microRNAs expressed from FSHR and aromatase genes in human ovarian granulosa cells. *Sci Rep* **10**:2300. doi:10.1038/s41598-020-59186-x

Sato T, Sasai N, Sasai Y. 2005. Neural crest determination by co-activation of Pax3 and Zic1 genes in Xenopus ectoderm. *Development* **132**:2355–2363. doi:10.1242/dev.01823

Sauka-Spengler T, Bronner-Fraser M. 2008. A gene regulatory network orchestrates neural crest formation. *Nat Rev Mol Cell Biol* **9**:557–568. doi:10.1038/nrm2428

Sauka-Spengler T, Meulemans D, Jones M, Bronner-Fraser M. 2007. Ancient evolutionary origin of the neural crest gene regulatory network. *Dev Cell* **13**:405–420. doi:10.1016/j.devcel.2007.08.005

Scerbo P, Monsoro-Burq AH. 2020. The vertebrate-specific VENTX/NANOG gene empowers neural crest with ectomesenchyme potential. *Sci Adv* **6**:eaaz1469. doi:10.1126/sciadv.aaz1469

Schlosser G. 2008. Do vertebrate neural crest and cranial placodes have a common evolutionary origin? *BioEssays* **30**:659–672. doi:10.1002/bies.20775

Schoen C, Aschrafi A, Thonissen M, Poelmans G, Von den Hoff JW, Carels CEL. 2017. MicroRNAs in Palatogenesis and Cleft Palate. *Front Physiol* **8**:165. doi:10.3389/fphys.2017.00165

Seal S, Monsoro-Burq AH. 2020. Insights Into the Early Gene Regulatory Network Controlling Neural Crest and Placode Fate Choices at the Neural Border. *Front Physiol* **11**:608812. doi:10.3389/fphys.2020.608812

Shu H, Zhou J, Lian Q, Li H, Zhao D, Zeng J, Ma J. 2021. Modeling gene regulatory networks using neural network architectures. *Nat Comput Sci* **1**:491–501. doi:10.1038/s43588-021-00099-8

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

Simões-Costa M, Bronner ME. 2015. Establishing neural crest identity: a gene regulatory recipe. *Development* **142**:242–257. doi:10.1242/dev.105445

Simões-Costa MS, McKeown SJ, Tan-Cabugao J, Sauka-Spengler T, Bronner ME. 2012. Dynamic and Differential Regulation of Stem Cell Factor FoxD3 in the Neural Crest Is Encrypted in the Genome. *PLOS Genetics* **8**:e1003142. doi:10.1371/journal.pgen.1003142

Sinner D, Kirilenko P, Rankin S, Wei E, Howard L, Kofron M, Heasman J, Woodland HR, Zorn AM. 2006. Global analysis of the transcriptional network controlling Xenopus endoderm formation. *Development* **133**:1955–1966. doi:10.1242/dev.02358

Soldatov R, Kaucka M, Kastriti ME, Petersen J, Chontorotzea T, Englmaier L, Akkuratova N, Yang Y, Häring M, Dyachuk V, Bock C, Farlik M, Piacentino ML, Boismoreau F, Hilscher MM, Yokota C, Qian X, Nilsson M, Bronner ME, Croci L, Hsiao W-Y, Guertin DA, Brunet J-F, Consalez GG, Ernfors P, Fried K, Kharchenko PV, Adameyko I. 2019. Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* **364**:eaas9536. doi:10.1126/science.aas9536

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

Stefani G, Slack FJ. 2008. Small non-coding RNAs in animal development. *Nat Rev Mol Cell Biol* **9**:219–230. doi:10.1038/nrm2347

Steventon B, Mayor R. 2012. Early neural crest induction requires an initial inhibition of Wnt signals. *Dev Biol* **365**:196–207. doi:10.1016/j.ydbio.2012.02.029

Stone JR, Hall BK. 2004. Latent homologues for the neural crest as an evolutionary novelty. *Evol Dev* **6**:123–129. doi:10.1111/j.1525-142x.2004.04014.x

Streit A, Stern CD. 1999. Establishment and maintenance of the border of the neural plate in the chick: involvement of FGF and BMP activity. *Mechanisms of Development* **82**:51–66. doi:10.1016/S0925-4773(99)00013-1

Strobl-Mazzulla PH, Sauka-Spengler T, Bronner-Fraser M. 2010. Histone Demethylase JmjD2A Regulates Neural Crest Specification. *Developmental Cell* **19**:460–468. doi:10.1016/j.devcel.2010.08.009

Sumiyama K, Irvine SQ, Stock DW, Weiss KM, Kawasaki K, Shimizu N, Shashikant CS, Miller W, Ruddle FH. 2002. Genomic structure and functional control of the Dlx3-7 bigene cluster. *Proc Natl Acad Sci U S A* **99**:780–785. doi:10.1073/pnas.012584999

Taneyhill LA. 2008. To adhere or not to adhere: the role of Cadherins in neural crest development. *Cell Adh Migr* **2**:223–230. doi:10.4161/cam.2.4.6835

Tang W, Li Y, Li A, Bronner ME. 2021. Clonal analysis and dynamic imaging identify multipotency of individual Gallus gallus caudal hindbrain neural crest cells toward cardiac and enteric fates. *Nat Commun* **12**:1894. doi:10.1038/s41467-021-22146-8

Theveneau E, Marchant L, Kuriyama S, Gull M, Moepps B, Parsons M, Mayor R. 2010. Collective chemotaxis requires contact-dependent cell polarity. *Dev Cell* **19**:39–53. doi:10.1016/j.devcel.2010.06.012

Thiery A, Buzzi AL, Hamrud E, Cheshire C, Luscombe N, Briscoe J, Streit A. 2022. A gradient border model for cell fate decisions at the neural plate border. doi:10.1101/2022.02.15.480567

Trainor P, Krumlauf R. 2000. Plasticity in mouse neural crest cells reveals a new patterning role for cranial mesoderm. *Nat Cell Biol* **2**:96–102. doi:10.1038/35000051

Uehara M, Yashiro K, Mamiya S, Nishino J, Chambon P, Dolle P, Sakai Y. 2007. CYP26A1 and CYP26C1 cooperatively regulate anterior-posterior patterning of the developing brain and the production of migratory cranial neural crest cells in the mouse. *Dev Biol* **302**:399–411. doi:10.1016/j.ydbio.2006.09.045

Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, Seurinck R, Saelens W, Cannoodt R, Rouchon Q, Verbeiren T, De Maeyer D, Reumers J, Saeys Y, Aerts S. 2020. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15**:2247–2276. doi:10.1038/s41596-020-0336-2

Wada H. 2001. Origin and evolution of the neural crest: A hypothetical reconstruction of its evolutionary history. *Development, Growth & Differentiation* **43**:509–520. doi:10.1046/j.1440-169X.2001.00600.x

Wada H, Holland PWH, Satoh N. 1996. Origin of patterning in neural tubes. *Nature* **384**:123–123. doi:10.1038/384123a0

Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM. 2018. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**:981–987. doi:10.1126/science.aar4362

Wakamatsu Y, Endo Y, Osumi N, Weston JA. 2004. Multiple roles of Sox2, an HMG-box transcription factor in avian neural crest development. *Dev Dyn* **229**:74–86. doi:10.1002/dvdy.10498

Walker JC, Harland RM. 2009. microRNA-24a is required to repress apoptosis in the developing neural retina. *Genes Dev* **23**:1046–1051. doi:10.1101/gad.1777709

Wang W-D, Melville DB, Montero-Balaguer M, Hatzopoulos AK, Knapik EW. 2011. Tfap2a and Foxd3 regulate early steps in the development of the neural crest progenitor population. *Developmental Biology* **360**:173–185. doi:10.1016/j.ydbio.2011.09.019

Ward NJ, Green D, Higgins J, Dalmay T, Münsterberg A, Moxon S, Wheeler GN. 2018. microRNAs associated with early neural crest development in Xenopus laevis. *BMC Genomics* **19**:59. doi:10.1186/s12864-018-4436-0

Weiner AMJ. 2018. MicroRNAs and the neural crest: From induction to differentiation. *Mech Dev* **154**:98–106. doi:10.1016/j.mod.2018.05.009

Wienholds E, Koudijs MJ, van Eeden FJM, Cuppen E, Plasterk RHA. 2003. The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nat Genet* **35**:217–218. doi:10.1038/ng1251

Williams RM, Candido-Ferreira I, Repapi E, Gavriouchkina D, Senanayake U, Ling ITC, Telenius J, Taylor S, Hughes J, Sauka-Spengler T. 2019. Reconstruction of the Global Neural Crest Gene Regulatory Network In Vivo. *Developmental Cell* **51**:255-276.e7. doi:10.1016/j.devcel.2019.10.003

Wilson YM, Richards KL, Ford-Perriss ML, Panthier J-J, Murphy M. 2004. Neural crest cell lineage segregation in the mouse neural tube. *Development* **131**:6153–6162. doi:10.1242/dev.01533

Xu K, Wang R, Xie H, Hu L, Wang Cong, Xu J, Zhu C, Liu Y, Gao F, Li X, Wang Cenzhu, Huang J, Zhou W, Zhou G, Shu Y, Guan X. 2021. Single-cell RNA sequencing reveals cell heterogeneity and transcriptome profile of breast cancer lymph node metastasis. *Oncogenesis* **10**:1–12. doi:10.1038/s41389-021-00355-6

Yang F, Zhao J, Luo X, Li T, Wang Z, Wei Q, Lu H, Meng Y, Cai K, Lu L, Lu Y, Chen L, Sooranna SR, Luo L, Song J, Meng L. 2021. Transcriptome Profiling Reveals B-Lineage Cells Contribute to the Poor Prognosis and Metastasis of Clear Cell Renal Cell Carcinoma. *Frontiers in Oncology* **11**.

Ye L, Zhang Q, Cheng Y, Chen X, Wang Guoying, Shi M, Zhang T, Cao Y, Pan H, Zhang L, Wang Genshu, Deng Y, Yang Y, Chen G. 2018. Tumor-derived exosomal HMGB1 fosters hepatocellular carcinoma immune evasion by promoting TIM-1+ regulatory B cell expansion. *Journal for ImmunoTherapy of Cancer* **6**:145. doi:10.1186/s40425-018-0451-6

Yook JI, Li X-Y, Ota I, Hu C, Kim HS, Kim NH, Cha SY, Ryu JK, Choi YJ, Kim J, Fearon ER, Weiss SJ. 2006. A Wnt-Axin2-GSK3beta cascade regulates Snail1 activity in breast cancer cells. *Nat Cell Biol* **8**:1398–1406. doi:10.1038/ncb1508

York JR, McCauley DW. 2020. The origin and evolution of vertebrate neural crest cells. *Open Biol* **10**:190285. doi:10.1098/rsob.190285

Zalc A, Sinha R, Gulati GS, Wesche DJ, Daszczuk P, Swigut T, Weissman IL, Wysocka J. 2021. Reactivation of the pluripotency program precedes formation of the cranial neural crest. *Science* **371**:eabb4776. doi:10.1126/science.abb4776

Zhang T, Guan X-W, Gribben JG, Liu F-T, Jia L. 2019. Blockade of HMGB1 signaling pathway by ethyl pyruvate inhibits tumor growth in diffuse large B-cell lymphoma. *Cell Death Dis* **10**:1–15. doi:10.1038/s41419-019-1563-8

Zhao D, Chen S, Liu X. 2019. Lateral neural borders as precursors of peripheral nervous systems: A comparative view across bilaterians. *Dev Growth Differ* **61**:58–72. doi:10.1111/dgd.12585