



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Integrative CUT&RUN and Single-Cell Multiome Analysis Reveals Putative
Neural Crest Enhancers and Promoters Associated with Orofacial Cleft Risk
Loci

verfasst von | submitted by

Inês Vitória Naidenka Rosa

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt |
Degree programme code as it appears on
the student record sheet:

UA 066 232

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Neuroscience

Betreut von | Supervisor:

Igor Adameyko PhD

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Prof. Dr. Igor Adameyko, for welcoming me into not just one, but both of his laboratories. Thank you for the guidance, support, and countless hours of conversation we shared - as friends, colleagues, supervisor and mentee. Thank you for the drinks we shared and created, for showing me that the words science and fun can belong in the same sentence, and for opening the world to me in ways I could not have imagined.

For a long time now, Igor's qualities and character have been a source of motivation and inspiration for the scientist I hope to become. I have often felt a remarkable sense of similarity and connection with him in many respects. To me, he personifies a distant, yet increasingly reachable vision of what the future may hold. Scientifically, many things that can deeply perturb or frustrate others bring me immense joy. In that regard, Igor's erratic, unconventional, and somehow beautifully chaotic ideas never cease to amaze me.

I believe Igor embodies the true spirit of science and what it feels like to be completely consumed by it. In a world that often feels increasingly constrained and predictable, I have come to deeply appreciate the uniqueness he brings to the scientific community. He reminds us that the greatest scientific discoveries do not always emerge from rigid plans, but often from curiosity, creativity, and the willingness to explore the unexpected.

Among the many memories I will carry with me, one of my favourites is the rare sight of Igor sitting at a cryostat, teaching me how to section tissue for the very first time. Moments like these, both scientific and personal, have shaped my journey far beyond what can be captured in this thesis. Thank you, Igor, for everything.

I was equally fortunate to have Dr. Mona Christensen as both my scientific supervisor and lab partner. Mona and I began our journeys in the Adameyko lab around the same time, starting our projects from scratch side by side. This meant that we shared an experience that many friendships do not encounter even over the course of years. Together, we celebrated both small and major victories, experienced moments of frustration, and occasionally found ourselves frustrated with each other. Our bond is an unusual one, perhaps best described as a meeting of opposites, a kind of yin and yang. Yet, I believe that learning how to understand, appreciate, and coexist with our differences is what creates the strongest relationships.

Over time, we came to realize that we are not as different as we initially thought. Together, we have learned not only about each other, but also about science, life, and the people around us. Mona is the person who taught me almost everything I have come to know during my time in the Adameyko lab, and for that I will always be deeply grateful. It is somewhat strange to think that one day Mona will no longer be by my side making amazing science, but I like to believe that regardless of where our paths take us, we will always be there for one another.

Lastly, I would like to thank Nikita Vaulin. Not only the hardworking bioinformatician who has contributed immensely to this project and many others, but also the person who helped us see the light at the end of the tunnel during difficult times. On many occasions, Nikita was the calm and level-headed presence in the room when Mona and I could not be. His ability to remain composed and find solutions where others saw only problems has been invaluable. Beyond the laboratory, Nikita has become one of the greatest friendships that the Adameyko lab has given me, and I am incredibly grateful for his support, companionship, and willingness to help whenever it was needed.

Contributions

I would also like to acknowledge the facilities and technical staff whose expertise and support made this work possible. I am grateful to the members of the Flow Cytometry and the Biomedical Sequencing core facilities at the Medical University of Vienna and CeMM, respectively, for their assistance and for providing access to the equipment and resources required for this project.

I would further like to clarify the contributions to the datasets used in this thesis. All wet-lab experiments, including tissue collection, cell sorting, library preparation, and chromatin profiling, were performed by Dr. Mona Christensen and myself. Computational processing, quality control, and initial handling of the sequencing datasets were carried out by Nikita Vaulin. The downstream analyses presented in this thesis, including integration of GWAS, scATAC-seq, CUT&RUN, and scRNA-seq datasets, candidate regulatory element prioritization, gene assignment, motif analysis, and interpretation of the results, were performed by me.

Table of Contents

Index

<i>Zusammenfassung</i>	8
<i>Abstract</i>	9
<i>1. Introduction</i>	12
1.1 The Art of Building an Organism	12
1.2 One Cell to Many	12
1.3 The Neural Crest: A Manifesto to Nonconformism	13
1.4 Forming the Neural Crest.....	13
1.5 The Many Fates of Neural Crest	14
1.6 Cranial Neural Crest: Making the Head	16
1.7 Neurocristopathies: The Cost of Nonconformity	17
1.8 Orofacial Clefts: Unfinished Sculptures	18
1.9 The Genetic Basis of Orofacial Clefts.....	19
1.10 Craniofacial Development Beyond Coding DNA	20
1.11 CUT&RUN and Single Cell ATAC Sequencing as Tools to Find Regulatory Elements	22
<i>2. Aim</i>	23
<i>3. Materials & Methods</i>	24
3.1 Mouse Husbandry, Strains and Genotyping.....	24
3.2 Embryo Staging	26
3.3 Tissue Collection: 10X Multiomics	27
3.4 Tissue Collection: CUT&RUN	27
3.5 10X Library Preparation and Sequencing.....	29

3.6 10X Data Processing.....	29
3.7 CUT&RUN Antibody Validation.....	30
3.8 CUT&RUN Library Preparation and Sequencing.....	31
3.9 CUT&RUN Data Processing.....	32
3.10 SNP Dataset Collection	32
3.11 Genomic Region Expansion and LiftOver	32
3.12 Selection of Expanded Regions.....	33
3.13 Identification of Putative Enhancers and Promoters	34
3.14 Finding the Target Genes for Identified Regulatory Elements.....	34
3.15 Distinguishing Putative Enhancers and Promoters	35
3.16 Gene Expression Analysis.....	35
3.17 Motif Analysis.....	36
4. Results	37
4.1 10x Multiome Allows Selection of Neural Crest Cells for Further Analysis	37
4.2 CUT&RUN Antibody Validation Shows Signal Specificity	38
4.3 Putative Enhancers and Promoters Identification	40
4.4 Description of Selected <i>Cis</i> -Regulatory Regions.....	42
4.5 Motif Analysis Reveals <i>de novo</i> Significantly Enriched Motifs	48
5. Discussion	49
5.1 Identification of Candidate Regulatory Elements at Orofacial Cleft Risk Loci	49
5.2 Regulatory Signatures Associated with Highest Expressed Neural Crest Genes	49
5.3 Putative Regulatory Regions Associated with Nrp1 and Adora2b	51

5.4 Additional Transcription Factors Potentially Contributing to Neural Crest Regulatory Regions Activity	53
6. Conclusion.....	55
Supplementary Data	57
References	63

Zusammenfassung

Orofaziale Spalten (OFCs) gehören zu den häufigsten angeborenen kraniofazialen Fehlbildungen und entstehen durch ein komplexes Zusammenspiel genetischer und umweltbedingter Faktoren. Obwohl genomweite Assoziationsstudien (Genome-Wide Association Studies, GWAS) zahlreiche mit dem OFC-Risiko assoziierte Genloci identifiziert haben, sind die regulatorischen Mechanismen, über die diese Varianten die kraniofaziale Entwicklung beeinflussen, bislang nur unzureichend verstanden. Da sich die meisten assoziierten Varianten in nicht-kodierenden Bereichen des Genoms befinden, ist die Identifizierung der regulatorischen Elemente und Zielgene, über die sie ihre Wirkung entfalten, eine besondere Herausforderung. Ziel dieser Studie war es, potenzielle regulatorische Elemente zu identifizieren und zu priorisieren, die mit OFC-Risikoloci in kranialen Neuralleistenzellen (cranial neural crest cells, CNCCs) assoziiert sind, sowie einen Ansatz zur Verknüpfung nicht-kodierender genetischer Varianten mit genregulatorischen Netzwerken der kraniofazialen Entwicklung zu etablieren. Mit OFCs assoziierte genetische Varianten wurden aus zwei publizierten GWAS-Datensätzen gewonnen und ihre humanen genomischen Koordinaten in entsprechende Mauskoordinaten überführt. Die Annotation neuralleistenspezifischer Muster der Chromatinzugänglichkeit sowie von Transkriptionsfaktor-Bindungsstellen in diesen OFC-assozierten Regionen erfolgte anhand von Single-Cell-ATAC-Sequenzierungsdaten (scATAC-seq) von Neuralleistenzellen sowie CUT&RUN-Profilen für vier Neuralleisten-Transkriptionsfaktoren und verschiedene Histonmodifikationsmarker. Regionen, die Transkriptionsfaktorbindung, charakteristische Histonsignaturen und Chromatinzugänglichkeit aufwiesen, wurden als relevante Kandidatenloci ausgewählt. Überlappende Neuralleisten-scATAC-seq-Peaks dienten anschließend zur Definition der Grenzen potenzieller regulatorischer Elemente. Diese wurden anhand ihrer Transkriptionsfaktor-Bindungsmuster und Histonmodifikationen charakterisiert und mithilfe von BEDTools *closest* ihren nächstgelegenen Genen zugeordnet. Die Expression der Zielgene wurde anhand von Single-Cell-RNA-Sequenzierungsdaten (scRNA-seq) aus E9.5-Neuralleistenzellen untersucht. Zusätzlich wurde eine de novo Motivanalyse durchgeführt, um weitere Transkriptionsfaktor-Bindungsmotive innerhalb der Kandidatenregionen zu identifizieren.

Die integrative Analyse identifizierte 38 genomische Regionen, die sowohl mit einem erhöhten OFC-Risiko als auch mit regulatorischer Aktivität in Neuralleistenzellen assoziiert waren, und ermöglichte deren Verknüpfung mit potenziellen Zielgenen. Vier Kandidatenregionen wurden auf Grundlage ihrer Chromatinsignaturen und Expressionsmuster für eine detaillierte Charakterisierung ausgewählt. Drei Regionen, die mit den Genen *Maib*, *Itgb1* und *Rtn3* assoziiert waren, überlappten annotierte Promotorregionen und zeigten eine Anreicherung von H3K4me3 sowie die Bindung mehrerer Neuralleisten-Transkriptionsfaktoren, was auf eine promotorassoziierte regulatorische Aktivität hinweist. Im Gegensatz dazu befand sich die mit *Pard3* assoziierte Region innerhalb eines intronischen Bereichs und wies eine H3K4me1-Anreicherung ohne nachweisbare Transkriptionsfaktorbindung auf, was mit den Eigenschaften eines potenziellen Enhancers übereinstimmt. Die de novo Motivanalyse identifizierte 46 signifikant angereicherte Motive, darunter Bindungsstellen für LEF1, MSX1, RARG und SMAD4 – Transkriptionsfaktoren, die nachweislich an der Spezifikation der Neuralleiste, der Gaumenentwicklung und der kraniofazialen Morphogenese beteiligt sind. Von den 38 identifizierten Zielgenen wurden bislang lediglich sieben mit OFCs in Verbindung gebracht, was die begrenzte funktionelle Charakterisierung vieler GWAS-assoziiierter Loci verdeutlicht. Darüber hinaus wurde keines der in dieser Studie identifizierten potenziellen cis-regulatorischen Elemente bisher im Zusammenhang mit OFC-assoziierten GWAS-Loci beschrieben. Zusammenfassend liefert diese Arbeit einen priorisierten Katalog potenzieller neuralleistenspezifischer regulatorischer Elemente, die mit OFC-Risikoloci assoziiert sind, und etabliert einen reproduzierbaren Ansatz zur Verknüpfung nicht-kodierender genetischer Varianten mit genregulatorischen Mechanismen, die der kraniofazialen Entwicklung und der Anfälligkeit für orofaziale Spalten zugrunde liegen.

Abstract

Orofacial clefts (OFCs) are among the most common congenital craniofacial malformations and arise through complex interactions between genetic and environmental factors. Although genome-wide association studies (GWAS) have identified numerous *loci* associated with OFC risk, the regulatory mechanisms through which these variants influence craniofacial development remain poorly understood. Most associated variants are located within non-coding regions of the genome, making it

challenging to identify the regulatory elements and target genes through which they act. This study aimed to identify and prioritize putative regulatory elements associated with OFC risk *loci* in cranial neural crest cells (CNCCs) and to establish a framework for linking non-coding genetic variation to gene regulatory networks involved in craniofacial development.

Genetic variants associated with OFCs were obtained from two published Genome Wide Association Studies (GWAS) datasets and human genomic coordinates were converted to the mouse genomic coordinates. The presence of neural crest-specific chromatin openness patterns and transcription factor binding in these OFC-associated regions were annotated by using neural crest cell scATAC-sequencing (scATAC-seq) data and CUT&RUN profiles for four neural crest transcription factors and histone modifications markers. Regions exhibiting transcription factor occupancy, histone signatures and chromatin accessibility were retained as *loci* of interest. Neural crest scATAC-seq peaks overlapping these *loci* were subsequently used to define the boundaries of putative regulatory elements, which were then characterized using transcription factor binding and histone modification profiles and linked to their nearest genes using BEDTools closest. The resulting regulatory elements were assigned to their nearest genes. Gene expression of the target genes was assessed using E9.5 neural crest single cell RNA sequencing (scRNA-seq) data. *De novo* motif enrichment analysis was performed to identify additional transcription factor binding signatures within the candidate regulatory regions. This integrative analysis identified 38 genomic regions associated with both, increased OFC risk and regulatory activity in neural crest cells and allowed linking them to putative target genes. Four candidate regions were selected for detailed characterization based on chromatin signatures and gene expression patterns. Three regions associated with *Mafb*, *Itgb1*, and *Rtn3* overlapped annotated promoter regions and displayed H3K4me3 enrichment together with occupancy by multiple neural crest transcription factors, consistent with promoter associated regulatory activity. In contrast, the region associated with *Pard3* was located within an intronic interval and exhibited H3K4me1 enrichment without detectable transcription factor occupancy, consistent with a putative enhancer region.

De novo motif analysis identified 46 significantly enriched motifs, including binding sites for LEF1, MSX1, RARG, and SMAD4, transcription factors with established roles in neural crest specification, palatogenesis, and craniofacial morphogenesis. Of the 38

associated genes, only seven have previously been linked to OFCs, highlighting the limited functional characterization of many GWAS-associated *loci*.

Moreover, none of the candidate *cis*-regulatory elements identified in this study have previously been described in the context of OFC-associated GWAS *loci*. Together, these findings provide a prioritized catalogue of candidate neural crest regulatory elements associated with OFC risk *loci* and establish a reproducible framework for connecting non-coding genetic variation to gene regulatory mechanisms underlying craniofacial development and susceptibility to orofacial clefts.

1. Introduction

1.1 The Art of Building an Organism

Understanding embryonic development has long been a central goal in biology^{1,2}. Questions surrounding fertilization, the emergence of robust yet flexible systems, and more recently the ethical implications of creating embryonic models, have been explored extensively^{3,4}. This persistent interest reflects a broader effort to understand the fundamental principles directing the emergence of complex multicellular life and has supported the ongoing effort to dissect the mechanisms underlying development at the most fundamental level.

A worthwhile feature of embryogenesis is the apparent paradox between conservation and diversity. In biology, processes are more often than not discussed in terms of variation, however, the core regulatory mechanisms that regulate development are surprisingly conserved across species¹. How similar genetic and molecular systems give rise to very different morphological outcomes remains a central question in developmental biology.

At the heart of this question lies the embryo. Understanding the embryo and the conserved systems that govern morphological diversity is key for progress. One of the most fundamental problems in biology is therefore how a single fertilized cell can give rise to hundreds of specialized cell types and ultimately to a fully patterned organism¹. Addressing this question requires understanding not only the genetic programs that guide development, but also how these programs are dynamically integrated with physical and spatial constraints within the embryo.

1.2 One Cell to Many

A critical step in embryogenesis is gastrulation, during which the embryo is reorganized into three primary germ layers: the ectoderm, mesoderm, and endoderm. These layers give rise to all tissues in the adult organism and lay the foundation for the body plan^{1,5,6}. Each germ layer follows distinct developmental trajectories governed by tightly regulated gene regulatory networks. These networks integrate signaling cues, transcription factor activity, and chromatin states to progressively restrict cellular potential and guide differentiation^{5,6}. In this sense, development can be viewed as an interplay between physical architecture and a genetic “grammar” that defines cellular identity.

Classical models of development, such as Waddington's epigenetic landscape, describe this process as a progressive restriction of cell fate potential. As development proceeds, cells move from a pluripotent state toward increasingly specialized identities, with less alternative outcomes available over time^{7,8}.

However, not all cell populations strictly follow this model. Certain populations retain an impressive degree of plasticity, challenging traditional views of lineage restriction and showcasing the inherent plasticity in developmental systems^{9,10}.

1.3 The Neural Crest: A Manifesto to Nonconformism

The neural crest represents one of the most striking examples of developmental plasticity. Often referred to as a "fourth germ layer," this vertebrate-specific cell population is characterized by its unusual multipotency and migratory capacity^{1,9,11}.

Since its discovery by Wilhelm His in the 19th century, the neural crest has been central to numerous hypotheses and discoveries in developmental biology¹². Its contributions span a wide range of tissues and structures, from components of the peripheral nervous system to craniofacial cartilage and bone⁹.

What makes the neural crest particularly remarkable is its ability to generate diverse cell types while undergoing extensive migration throughout the embryo. This combination of multipotency and range sets it apart from most other cell populations and positions it as a valuable system for understanding how genetic programs and physical behaviors interact during development.

1.4 Forming the Neural Crest

Neural crest cells originate at the border between the neural plate and the non-neural ectoderm, a region defined by the integration of multiple signaling pathways.

Among the key pathways involved are Bone Morphogenetic Proteins (BMP), Wnt signaling, Fibroblast Growth Factors (FGF), and retinoic acid (RA). Instead of acting independently, these signals establish gradients and combinatorial inputs that define the neural plate border region^{6,9,13,14}

BMP signaling plays a central role in this process. High levels of this signaling molecule promote epidermal fate, while low levels promote neural tissue. Neural crest cells arise at intermediate levels of BMP activity, where border-specifying transcription factors such as PAX3/7 and MSX1/2 are activated^{13–15}.

WNT signaling is essential for neural crest specification, promoting the activation of key transcription factors such as *FOXD3* and *SOX10*, which maintain neural crest identity and multipotency^{13,14,16–18}. FGF signaling contributes to cell competence and survival, while retinoic acid helps establish anterior-posterior patterning through regulation of *Hox* genes^{10,19}. These signaling pathways intersect to initiate a hierarchical gene regulatory network, progressing from border specification to neural crest specification and finally to the activation of effector genes controlling migration and differentiation. While the core components of this network are conserved across vertebrates, differences in timing and signal integration contribute to species-specific outcomes²⁰. In fact, signal intensity, timing, and spatial distribution can have significant developmental consequences.

1.5 The Many Fates of Neural Crest

Following specification, neural crest cells undergo an epithelial-to-mesenchymal transition (EMT), enabling them to delaminate from the neural tube and migrate throughout the embryo (Fig. 1).

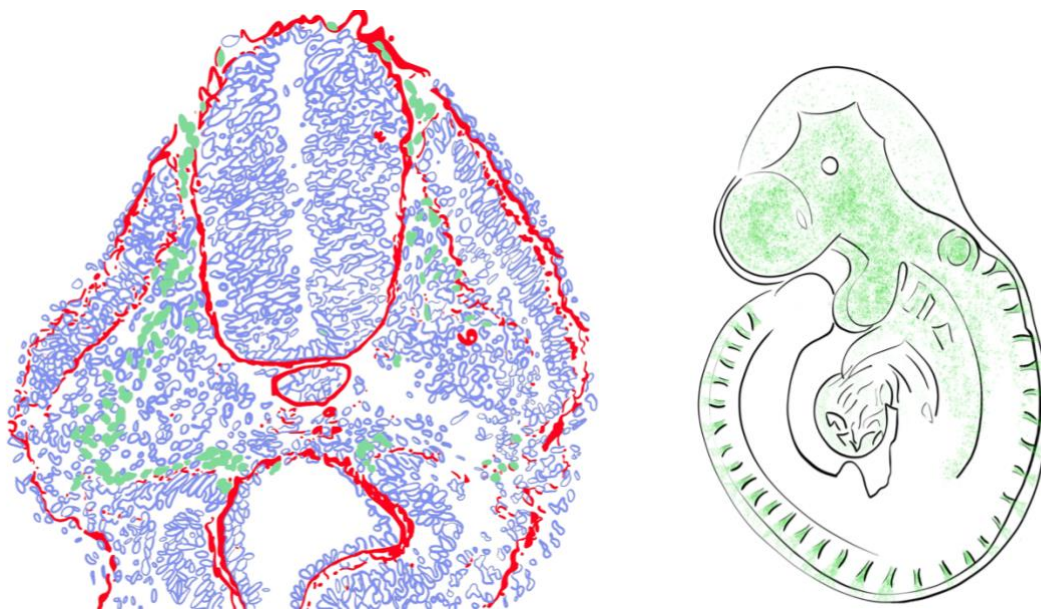


Figure 1-Illustration of an E9.5 mouse neural tube with delaminating neural crest cells cross-section and of neural crest cell migration in the E9.5 mouse embryo body axis. On the left, in red is represented Agrin positive extracellular matrix. In green is represented Sox10 positive neural crest cells, delaminating from the neural tube. In blue is represented DAPI labeled nuclei. On the right, spatio-temporal distribution of migrating neural crest cells at E9.5 stage, showcasing NCC migration potential.

This transition involves the downregulation of adhesion molecules such as E-cadherin and N-cadherin, mediated by transcription factors including SNAIL and SLUG, allowing neural crest cells to detach from the matrix^{11,21,22}.

Once migratory, neural crest cells exhibit remarkable plasticity, contributing to a wide range of tissues. Their ultimate fate is influenced in part by their position along the anterior–posterior axis, giving rise to distinct subpopulations:

Cranial neural crest cells contribute to craniofacial structures, including bone, cartilage, and teeth. Vagal and cardiac neural crest populations contribute to components of the peripheral nervous system and cardiovascular system. Trunk neural crest cells give rise to melanocytes and sympathetic ganglia, while sacral neural crest cells contribute to the pelvic ganglion and the enteric nervous system^{1,6,9–11,11,14,23,24} (Fig.2).



Figure 2-Representation of spatial distribution and lineage differentiation of neural crest derivatives in the E14.5 mouse embryo. This illustration depicts the distributed regions of neural crest derivatives after migration phase is completed. Distinct colours represent the five different neural crest subpopulations: in blue cranial neural crest, in green vagal neural crest, in pink cardiac neural crest, in yellow trunk neural crest and in purple sacral neural crest.

1.6 Cranial Neural Crest: Making the Head

In the hindbrain, segmentation into rhombomeres establishes positional identity through Hox gene expression. This organization guides cranial neural crest cells into distinct migratory streams, ensuring that different populations contribute to specific structures without inappropriate mixing. Among neural crest populations, cranial neural crest cells stand out due to their particularly broad plasticity, challenging classical models of developmental restriction and shaping of the vertebrate's head^{6,9,25,26}.

Although they originate from the ectoderm, they are capable of generating ectomesenchymal derivatives such as bone and cartilage, which in the rest of the body are of mesodermal origin.

This ability was crucial for the success of vertebrates and has been linked to the “new head” hypothesis. This model proposes that the emergence of cranial neural crest cells allowed the evolution of complex craniofacial structures, such as the jaw and teeth, allowing a predatory lifestyle instead of passive filter feeding (Fig.3). This transition opened new ecological niches and is considered a fundamental step in vertebrate evolution^{26,27}.

From a developmental perspective cranial neural crest cells are characterized by their ability to retain and even expand their developmental potential during key stages of development. This plasticity is associated with dynamic changes in chromatin accessibility, allowing these cells to access gene regulatory programs that are normally restricted in other contexts. During migration, cranial neural crest cells undergo a series of lineage decisions, giving rise to sensory, autonomic, and mesenchymal derivatives.

Transcription factors such as TWIST1 have been implicated in enabling the acquisition of ectomesenchymal potential, although the precise mechanisms underlying this process remain to be fully understood^{28,29}.

While this flexibility is essential for the formation of complex craniofacial structures, it can also introduce vulnerability. Systems that rely on tightly coordinated signaling, precise timing, and gene regulation are dependent on multiple sequential checkpoints and stages, which can make them particularly sensitive to perturbation.

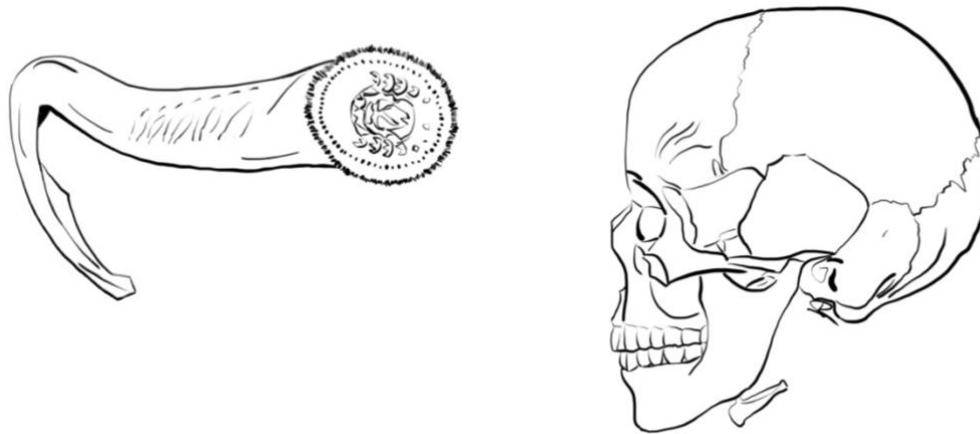


Figure 3-Evolution of vertebrate craniofacial structures driven by cranial neural crest cells. This illustration depicts the evolutionary and developmental consequences of cranial neural crest emergence. On the left, the lamprey is shown as a representative jawless vertebrate exhibiting a simplified craniofacial organization and filter-feeding mode of nutrition. On the right, a human skull illustrates the complexity of vertebrate craniofacial structures, including jaws and teeth, largely derived from cranial neural crest contributions. Together, these images highlight the role of cranial neural crest cells in enabling the transition from simple head architectures to highly specialized craniofacial systems.

1.7 Neurocristopathies: The Cost of Nonconformity

Disruptions in neural crest development can lead to a diverse group of disorders collectively known as neurocristopathies. These conditions arise from defects in neural crest induction, migration, or differentiation and often affect multiple organ systems^{30–32}. Given the broad contribution of neural crest cells, the resulting phenotypes can be highly variable, ranging from pigmentary disorders to severe craniofacial malformations and defects in the peripheral nervous system. Additionally, the boundaries between developmental stages are not strictly defined, and defects in one process often influence others. As a result, neurocristopathies frequently involve overlapping mechanisms, complicating both classification and analysis. Together, they represent a prevalent category for human birth defects, infant morbidity and long-term disability.

Among these pathologies, craniofacial abnormalities are particularly prominent, reflecting the central role of cranial neural crest cells in shaping facial structures.

Moreover, historical records support the idea that these conditions have been part of the human phenotypic spectrum, with documentation spanning Ancient Egypt, pre-Hispanic South American cultures, and the Roman Empire^{33–35}.

1.8 Orofacial Clefts: Unfinished Sculptures

Orofacial clefts represent one of the most common congenital craniofacial malformations. Studies of these malformations have helped us understand how disruptions in embryonic development translate into structural defects. Clinically, they present as cleft lip, cleft palate, or cleft lip and palate (Fig.4).

Additionally, a key distinction is made between syndromic and non-syndromic forms. Syndromic clefts occur as part of broader genetic disorders and are often accompanied by additional abnormalities, whereas non-syndromic clefts arise as isolated conditions without other evident phenotypes. Altogether approximately 70% of cleft lip with or without cleft palate are non-syndromic³⁶⁻³⁹.

From a developmental perspective, orofacial clefts result from failures in coordinated morphogenetic processes that shape the developing face. Facial development involves the growth, patterning, and fusion of five facial prominences: the frontonasal prominence, paired maxillary prominences, and paired mandibular prominences. These structures are composed largely of cranial neural crest-derived mesenchyme that migrates into the developing face during early embryogenesis^{6,10}.

The frontonasal prominence gives rise to the medial and lateral nasal prominences, which contribute to the nose and upper lip. Formation of the upper lip requires the fusion of the paired maxillary prominences with the medial nasal prominences. Disruption of growth or fusion of these prominences can result in unilateral or bilateral cleft lip^{40,41}.

The palatal shelves arise as bilateral outgrowths of the maxillary prominences and are predominantly of neural crest derived. Following vertical growth alongside the tongue and elevation, the shelves fuse with each other and the nasal septum at the midline. Failure of palatal shelf growth, elevation, or fusion can lead to cleft palate.

Although monogenic syndromes, chromosomal abnormalities, environmental factors and prenatal exposure to teratogens can be accounted for orofacial clefts aetiology, non-syndromic clefts, in particular, are characterized by genetic heterogeneity and do not follow simple Mendelian inheritance patterns^{36,37,39,42}.

Variants in genes such as *Irf6* have been identified as contributors to orofacial clefts, nonetheless much of the underlying regulatory structure remains to be fully comprehended^{43,44}. Mouse models carrying loss-of-function and patient-derived *Irf6* mutations have helped understand its role in palatogenesis. These models have shown consistently cleft palate and epithelial abnormalities, including impaired differentiation, excessive proliferation, and oral adhesions that disrupt palatal shelves.

Current evidence implicates IRF6 to be part of a broader orofacial gene regulatory network (GRN) that controls the balance between proliferation and differentiation of the epithelial tissues lining the developing facial prominences and palatal shelves during craniofacial morphogenesis^{36,45}. The disruption of these processes can result in orofacial clefting. IRF6 is known to interact with pathways including TGF- β and Notch, while additional factors, such as TFAP2A, are proposed components of this regulatory network^{46,47}. Altogether, this suggests that OFCs arise through the disruption of interconnected developmental pathways rather than defects in a single gene alone.

As a result, understanding the genetic basis of non-syndromic clefts remains challenging, as it has become clear that these phenotypes do not arise solely from the disrupted gene functions, but also from perturbations in the precise control of gene expression during critical stages of craniofacial development.



Figure 4-Classification of orofacial clefts. Top row illustrates lip morphology, including normal anatomy, unilateral cleft lip, and bilateral cleft lip. Bottom row illustrates palate morphology, including normal anatomy, unilateral cleft palate, and bilateral cleft palate.

1.9 The Genetic Basis of Orofacial Clefts

To understand the regulatory logic of orofacial clefts, early studies relied primarily on family-based linkage analyses and candidate gene approaches⁴⁸. Linkage studies examined the co-inheritance of genetic markers and disease phenotypes within affected families, allowing identification of broad chromosomal regions likely to contain disease-associated genes^{49,50}. Candidate gene studies, in turn, focused on genes with known roles in craniofacial development and embryogenesis^{49,51}. Through these approaches, several genes involved in craniofacial morphogenesis were identified, including *Irf6*, which remains one of the strongest genetic contributors to orofacial clefts.

With the emergence of genome-wide association studies (GWAS), prior assumptions regarding disease biology were no longer required. Instead, genetic variants could be

mapped across the entire genome, allowing the identification of genomic regions associated with disease susceptibility in large populations⁴⁹.

GWAS primarily investigate single nucleotide polymorphisms (SNPs), which are single-base pair variations which occur naturally throughout the genome. Although most SNPs do not directly influence biological functions or cause disease, they can be useful genetic markers, since neighboring variants tend to be inherited together, a phenomenon known as linkage disequilibrium (LD)⁵². Therefore, a SNP associated with disease risk may not itself be causal but may indicate the presence of a nearby functional variant within the same genomic region⁵³.

By comparing SNP frequencies between affected individuals and unaffected individuals, GWAS can identify *loci* associated with increased disease susceptibility. GWAS on orofacial cleft affected individuals has led to the identification of numerous risk *loci*, including regions near genes such as *Irf6*, *Arhgap29*, *Pax7*, *Vax1*, *Mafb*, and *Ntn*⁴⁹. However, translating these associations into biological mechanisms remains challenging, as many risk-associated variants occur within non-coding regions of the genome. As a result, the functional regulatory elements and target genes through which these variants act remain to be fully understood.

1.10 Craniofacial Development Beyond Coding DNA

While mutations in protein-coding genes have been linked to severe craniofacial abnormalities, they can only explain a fraction of both pathological and normal variation in facial morphology⁵⁴. It is widely accepted that cellular development, morphology and function are governed by precise patterns of gene expression. These patterns are established by the coordinated action of genomic regulatory elements known as enhancers or *cis*-regulatory modules⁵⁵. Over three decades after the initial discovery of enhancers⁵⁶, many of their properties have been elucidated. However, despite major efforts, we only have an incomplete picture of enhancers in animal genomes.

Access to genetic information in eukaryotes is regulated through chromatin, a nucleoprotein complex that packs DNA into repeating nucleosomal units and establishes a repressive ground state for transcription and other DNA-templated processes. Chromatin organization is shaped by a combination of molecular features, including DNA methylation, post-translational histone modifications, transcription factor binding, and the activity of *cis*-regulatory elements, such as promoters and enhancers, that together

determine the accessibility of genomic regions to the transcriptional machinery. Different combinations of these features define distinct functional configurations known as chromatin states, which vary across genomic *loci*, change dynamically during cell differentiation, and can be propagated across cell divisions. By controlling the accessibility of regulatory DNA, chromatin states constitute a key layer of gene expression regulation, linking the epigenetic landscape of a cell to its transcriptional output.

In eukaryotic genomes, protein-coding genes are embedded within a regulatory architecture that includes the core promoter, gene body, and distal regulatory elements. The promoter is located immediately upstream of the transcription start site and serves as the binding platform for RNA polymerase II and general transcription factors, which together form the transcriptional machinery required to initiate gene expression. However, promoter activity is modulated by additional regulatory DNA elements, called enhancers, that integrate developmental and environmental signals to ensure context-specific gene expression. Enhancers are non-coding regulatory DNA elements that can be located upstream, downstream, and at great distances from their target genes, often ranging from tens to hundreds of kilobase pairs in length^{57,58}. They function by recruiting transcription factors and co-regulatory proteins, which in turn enable the formation of chromatin loops that bring enhancers into physical proximity with gene promoters. This spatial organization allows transcription factors bound at enhancers to modulate gene expression with high specificity and precision, creating an active chromatin state⁵⁹.

Slight genetic variations, such as SNPs, in particular in non-coding regulatory regions, can change enhancer activity. SNPs can alter transcription factor binding sites and affect the enhancer affinity to the promoter, resulting in significant changes in gene expression without disrupting the actual gene sequence. In craniofacial development, changes like these are particularly relevant since precise spatiotemporal control of gene expression is required for the proper behaviour of cranial neural crest cells⁵⁸. Even subtle changes in the regulatory programs can lead to modified cell migration, proliferation and tissue fusion. This type of variation is now thought to account for a big portion of non-syndromic orofacial clefts, rather than the classical view of protein coding mutations. SNPs located nearby or at craniofacial enhancers could be responsible for the disruption of critical developmental milestones, leading to incomplete fusion of the maxillary prominences with nasal prominences or of the palatal shelves. Nonetheless, and despite the identification of orofacial clefts risk *loci*, linking non-coding variants to their target

regulatory elements and downstream genes remains a major challenge. This is largely due to the incomplete knowledge of the chromatin landscape and transcription factor occupancy in embryonic cell populations, such as cranial neural crest cells.

1.11 CUT&RUN and Single Cell ATAC Sequencing as Tools to Find Regulatory Elements

Cleavage Under Targets and Release Using Nuclease (CUT&RUN) provides an effective solution to linking non-coding variants to their target regulatory elements, as a next-generation chromatin profiling method that overcomes many limitations of others, such as ChIP-seq^{60,61}. In CUT&RUN, antibodies are used to tether a fusion protein consisting of Protein A (or Protein A/G) and micrococcal nuclease (MNase) to the target protein in intact nuclei. Controlled activation of MNase cleaves DNA adjacent to the binding site, releasing the associated DNA fragments directly into the supernatant. These fragments can then be purified and sequenced. Since CUT&RUN is performed under native conditions and avoids crosslinking, it produces exceptionally low background and high-resolution profiles, even from small numbers of cells^{60,61}. This makes it particularly well-suited for studying small developmental cell populations, where methods such as ChIP-seq are not as achievable. Furthermore, the precise cleavage pattern of CUT&RUN allows not only the mapping of binding sites but also to infer features of the local chromatin environment. Single-cell ATAC sequencing (scATAC-seq) complements CUT&RUN by profiling chromatin accessibility at single-cell resolution, enabling the identification of open chromatin regions that are accessible for transcription factor binding and likely to represent active regulatory elements such as enhancers and promoters⁶². Combining these methods in the mouse neural crest allows the direct identification of enhancers and promoters that are accessible and bound by neural crest transcription factors during critical steps of development. Integrating these chromatin landscapes with GWAS-associated SNP regions enables the prioritization of high-confidence promoters and putative enhancers that may be disrupted by non-coding genetic variation. Thus, this strategy can provide a mechanistic framework for linking risk *loci* associated with orofacial clefts to altered gene regulatory programs rather than protein-coding sequence disruption.

2. Aim

The regulatory mechanisms linking genetic variation to craniofacial development are not fully understood, particularly in the context of non-coding variation associated with orofacial clefts. While genome-wide association studies (GWAS) have identified multiple risk *loci*, the functional regulatory elements and target genes by which these variants act are largely unknown. To address this, I aim to identify and prioritize putative regulatory elements, such as enhancers and promoters, active in cranial neural crest cells that overlap with orofacial cleft-associated genetic variants.

I map human GWAS *loci* associated with orofacial clefts to the mouse genome and integrate the regions of interest with single-cell chromatin accessibility data from cranial neural crest cells. I characterize these regions using transcription factor binding and histone modification data obtained from CUT&RUN assays to identify putative regulatory elements, including both enhancers and promoter regions, associated with neural crest cell development.

I prioritize the putative regulatory elements overlapping orofacial cleft-associated *loci* and link them to putative target genes using genomic proximity and neural crest gene expression profiles derived from scRNA-seq data at E9.5. The final outcome is a prioritized catalogue of neural crest putative regulatory elements associated with orofacial cleft risk loci, and a reproducible framework for linking pathology-associated non-coding variation to gene regulatory networks in craniofacial development.

3. Materials & Methods

3.1 Mouse Husbandry, Strains and Genotyping

All mouse work was performed under a permit approved by the Ethical Committee for Advice and Assessment of Research Projects on Animals of the Medical University of Vienna (Ethik-Kommission der MedUni Wien zur Beratung und Begutachtung von Forschungsprojekten am Tier) adhering to Austrian and European regulations and guidelines for animal experimentation. Additionally, all animal experiments were approved by the Austrian Federal Ministry for Education, Science and Research under the permit GZ 66.009/0018-WF/V/3b/2017. All mouse lines were kept on a C57BL/6JRj (Janvier Labs) background and housed under standard conditions at the Core Facility Laboratory Animal Breeding and Husbandry (CFL) of the Medical University of Vienna. *Wnt1^{Cre}* mouse line was used to drive neural crest-specific recombination and *Mesp1^{Cre}* mouse line was used to drive mesoderm-specific recombination. These Cre lines were mated with R26RtdTomato mouse line, which carries a Cre-inducible tdTomato fluorescent reporter expressed upon Cre-mediated recombination. This mouse models allowed permanent lineage tracing of labeled cells and their descendants, and acquiring embryos with fluorescent marker expression in neural crest and mesoderm lineages, respectively. *Wnt1^{Cre}* mice were received from The Jackson Laboratory, stock number 009107 (full strain name: B6.Cg-H2az2 ^{Tg(Wnt1-cre)}11Rth ^{Tg(Wnt1-GAL4)}11Rth/J), R26RtdTomato mice were also received from The Jackson Laboratory, stock number 007914 (full strain name: B6.Cg-Gt(ROSA)26Sor^{tm14(CAG-tdTomato)}Hze/J) and *Mesp1^{Cre}* mice (full name *Mesp1-IRES2-Cre*) were received from Dr. Kenneth Chien's laboratory (Karolinska Institute, Sweden)

Prior to selection of the males and females used to produce the *Wnt1^{Cre}*; R26RTdTomato and *Mesp1^{Cre}*; R26RTdTomato heterozygous embryos, genotyping by PCR, genomic DNA was extracted from biopsy samples using a rapid alkaline lysis (Truett et. al, 2000). Briefly, tissue biopsies were centrifuged to ensure samples were positioned at the bottom of the tube: Each sample was resuspended in 100 μ L of alkaline lysis buffer (25 mM NaOH, 0,2 mM disodium EDTA, pH ~12) and incubated at 95° for 20-60 minutes. Following incubation, samples were centrifuged to collect condensation and minimize cross-contamination. DNA was neutralized by adding 100 μ L of 40 mM Tris-HCl (pH 5.5). Samples were vortexed and stored at 4 °C until use. For PCR, 1 μ L of DNA extract

was used per 20 μ L PCR reaction. PCR reactions were performed using a commercial 2 $\times$ PCR master mix (Promega, M7123). Each reaction had a final volume of 20 μ L, prepared as follows:

Table 1-PCR master mix used in genotyping.

Component	Volume per reaction
2X Promega Green Buffer	10 μ l
Forward Primer	0,3 μ l
Reverse Primer	0,3 μ l
Nuclease-free Water	8,4 μ l
Template DNA	1 μ l
Total Volume	20 μ l

A master mix containing all components except template DNA was prepared for each sample, including one positive control, one negative control, and one additional reaction to account for pipetting error. The master mix was mixed thoroughly by vortexing and briefly centrifuged prior to aliquoting.

Aliquots of 19 μ L master mix were dispensed into PCR tubes, followed by the addition of 1 μ L of DNA sample. PCR amplification was performed under standard cycling conditions consisting of an initial denaturation at 95 $^{\circ}$ C for 2 min, followed by 34 cycles of denaturation at 95 $^{\circ}$ C for 1 min, annealing at 58–60 $^{\circ}$ C (depending on primer set) for 30–60 s, and extension at 72 $^{\circ}$ C for 1 min, with a final extension at 72 $^{\circ}$ C for 5 min.

Primers used for genotyping were obtained from The Jackson Laboratory and are listed below:

Table 2-Primer sequences used for genotyping.

Rosa26tdTomato

Primer Name	Sequence (5' $\rightarrow$ 3')	T _m ($^{\circ}$ C)
RosaTomato TG F	CTG TTC CTG TAC GGC ATG G	62.6
RosaTomato TG R	GGC ATT AAA GCA GCG TAT CC	61.6
RosaTomato WT F	AAG GGA GCT GCA GTG GAG TA	65.3
RosaTomato WT R	CCG AAA ATC TGT GGG AAG TC	60.9

Unspecific Cre Recombinase

Primer Name	Sequence (5' → 3')	Tm (°C)
Unspecific Cre F	GCG GTC TGG CAG TAA AAA CTA TC	64,6
Unspecific Cre R	GTG AAA CAG CAT TGC TGT CAC TT	62,9

Mesp1 Cre Recombinase

Primer Name	Sequence (5' → 3')
Mesp1 WT F	GTT GAT TGT GGC GCT GTT C
Mesp1 WT R	AGC TTG CAT GAT CTC CGG TAT TGA A
Mesp1 Cre F	AGC GAT GGA TTT CCG TCT CTG G
Mesp1 Cre R	AGC TTG CAT GAT CTC CGG TAT TGA A

PCR products were separated by agarose gel electrophoresis using 1% (w/v) agarose gels prepared in 1× TAE buffer. Gels were cast by dissolving agarose in buffer, followed by addition of GelRed nucleic acid stain (10000X Water, Sigma-Aldrich Handels GmbH), and poured into casting trays with appropriate combs. After solidification, samples were loaded and run in 1× TAE buffer, alongside 5 µL of GeneRuler DNA Ladder (1 Kb ready-to-use 50 ng, LifeTech Austria).

3.2 Embryo Staging

To ensure collected embryos corresponded to the desired developmental stages, we compared their morphology to staging criteria as described in Qiu et al., 2024 . Somite number was counted and additional morphological characteristics, such as the degree of hind limb bud formation, were assessed. Fluorescent marker expression in neural crest and mesoderm lineages were detected using a stereomicroscope equipped with a fluorescent light source (Leica M165 FC) - Fig.5.

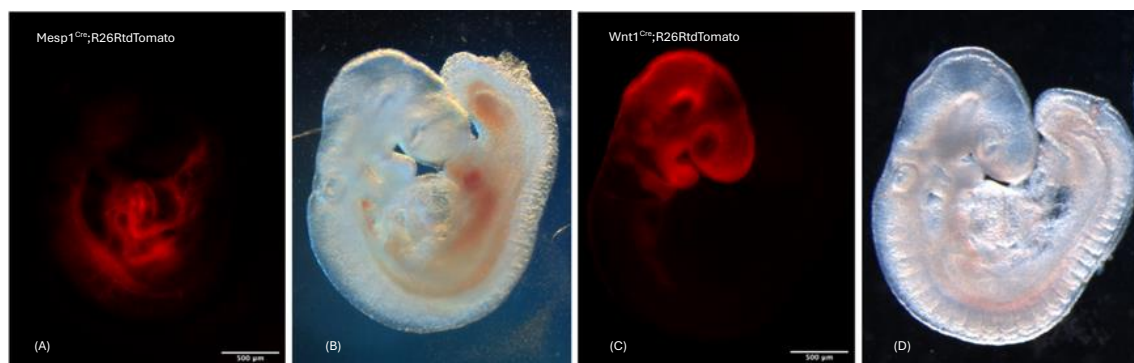


Figure 5-Fluorescent marker expression in neural crest and mesoderm lineages

Embryonic day 9 (E9) mouse embryos with lineage-specific fluorescent reporter expression. (A) *Mesp1*-driven reporter expression visualized by red fluorescence in mesodermal tissues. (B) Corresponding brightfield image of the same embryo shown in (A). (C) *Wnt1*-driven reporter expression marking neural crest cells, visualized by fluorescence. (D) Corresponding brightfield image of the same embryo shown in (C). Scale bar = 500 μ m.

3.3 Tissue Collection: 10X Multiomics

Wnt1^{Cre};R26RtdTomato and *Mesp1*^{Cre};R26RtdTomato embryos at embryonic days (E) 9.5 were used for embryonic cell collection. While this study primarily focuses on *Wnt1* derived neural crest cells, *Mesp1* derived cells were processed in parallel as part of the generation of a broader single-cell multiomic reference dataset that is being used in ongoing studies and was used here for gene expression analyses.

For all experiments, following time-mating, the day of vaginal plug detection was considered E0.5. Mice were sacrificed using isoflurane followed by cervical dislocation. Embryo heads were separated from the trunks at the level of the otic placode using a stereomicroscope equipped with a fluorescent light source (Leica M165 FC), and *Wnt1*^{Cre};R26RtdTomato and *Mesp1*^{Cre};R26RtdTomato heads were then separately incubated in 0,05% Trypsin-EDTA (Sigma-Aldrich Handels GmbH) at 37°C for 15 minutes while rocking (300 RPM), with gentle pipetting every 5 minutes to aid dissociation. The enzymatic reactions were stopped by adding ice-cold 10% Fetal Bovine Serum (Thermo Fischer Scientific) in Dulbecco's Phosphate-Buffered Saline without calcium and magnesium (1X Gibco, Fischer Scientific GmbH). Cells were pelleted by centrifugation at 4°C (300-500xG, 5 minutes) and resuspended in 0,5% Bovine Serum Albumin (Sigma-Aldrich Handels GmbH) in Phosphate-Buffered Saline. The cellular suspensions were filtered separately through a 40 μ m-pore size nylon cell strainer (Falcon, Corning, USA). TOMATO positive cells were sorted using a fluorescence-activated cell sorter equipped with a 561 nm laser (ARIA Fusion).

3.4 Tissue Collection: CUT&RUN

TOMATO positive E9.5 *Wnt1*^{Cre};R26RtdTomato embryos were used for CUT&RUN WHEN targeting histone modifications: H3K4me3, H3K27me3 and H3K4me1. These histone modifications are hallmarks of active chromatin, repressed chromatin, and active enhancers, respectively^{64,65}. Due to the relatively low abundance of transcription factor binding events and the relatively small number of neural crest cells in early embryos, it was not feasible to generate clear transcription factor CUT&RUN profiles from sorted

Wnt1^{Cre};R26RTomato cells alone. To overcome this, we profiled several transcription factors known to be expressed in cranial neural crest cells (SOX10, TFAP2A, ETS1, and FOXD3) using wild type C57BL/6Jrj embryonic heads (Fig.6). At E9.5, SOX10, FOXD3, TFAP2A, and ETS1 are expressed in migrating cranial neural crest cells. SOX10 is additionally detected in premigratory neural crest cells within the dorsal neural tube, the developing cranial ganglia, and the otic vesicle¹⁶. TFAP2A is also expressed in the surface ectoderm, whereas ETS1 is detected in the dorsal neural tube and developing cranial mesenchyme^{66,67}. FOXD3 expression is largely restricted to premigratory and migratory neural crest populations^{18,68}.

By focusing on genomic regions bound by multiple transcription factors and integrating this with neural crest-specific histone modification data and ATAC-sequencing data, we were able to identify candidate regulatory regions active in cranial neural crest cells.

Tissue containing cranial neural crest cells were isolated from Wnt1^{Cre};R26RtdTomato mouse embryos, using a stereomicroscope equipped with a fluorescent light source (Leica M165 FC). Embryo heads were separated from the trunks at the level of the otic placode, and most forebrain and midbrain tissues were removed to reduce non-neural crest cell contribution to the samples. Heads were then incubated in 0,05% Trypsin-EDTA (Sigma-Aldrich Handels GmbH) at 37°C for 15 minutes, with gentle pipetting every 5 minutes to aid dissociation. The enzymatic reaction was stopped by adding ice-cold 10% Fetal Bovine Serum (Thermo Fischer Scientific) in Dulbecco's Phosphate-Buffered Saline without calcium and magnesium (1X Gibco, Fischer Scientific GmbH). Cells were pelleted by centrifugation at 4°C (300-500xG, 5 minutes) and resuspended in 0,5% Bovine Serum Albumin (Sigma-Aldrich Handels GmbH) in Phosphate-Buffered Saline. The cellular suspension was filtered through a 40 µm-pore size nylon cell strainer (Falcon, Corning, USA) and collected. TOMATO positive cells were sorted using a fluorescence-activated cell sorter equipped with a 561 nm laser (ARIA Fusion).

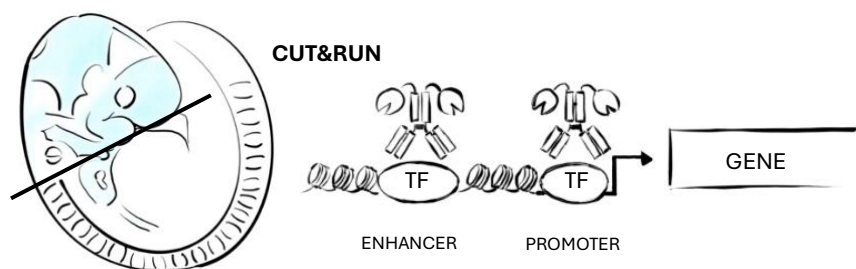


Figure 6-Neural crest profiled transcription factors using CUT&RUN.

Overview of the experimental workflow using E9 wild type C57BL/6JRj embryonic heads to profile transcription factors Sox10, Tfap2a, Ets1 and Foxd3. The embryo head is shown with cranial neural crest cells highlighted in blue. Heads were dissected below the otic placode (black line) and processed for CUT&RUN analysis.

3.5 10X Library Preparation and Sequencing

10x Genomics library preparation and sequencing were performed by the Biomedical Sequencing Facility (BSF) at the Centre for Molecular Medicine (CeMM), Vienna, Austria.

3.6 10X Data Processing

10X Multiome reads processing, and alignment was carried out in the Biomedical Sequencing Facility (BSF) at the Centre for Molecular Medicine (CeMM), Vienna, Austria using CellReanger ARC pipeline v. 2.1.0 (10X Genomics, 2023). Downstream analysis was performed by one of our laboratory's bioinformatician using Scanpy v. 1.12.1⁶⁹ and pyCisTarget v.1.1⁷⁰. Cells were filtered to contain 200 - 6000 genes and less than 10% of mitochondrial counts. Genes expressed in less than 3 cells were discarded. Doublet detection was done using sc.pp.scrublet with sample as a batch key and an automatic threshold was applied. Cells were also classified as doublets and filtered out if they co-expressed both male-associated (Ddx3y, Eif2s3y, Kdm5d, Uty) and female-associated (Xist, Tsix) genes. Gene counts were log1pCP10K-normalised.

Wnt1^{Cre};R26RTomato and Mesp1^{Cre};R26RTomato lineages were annotated separately based on principal component analysis (PCA) and kNN-graph construction on 2000 highly variable genes. We used 10 PCs with 30 neighbours for Wnt1^{Cre};R26RTomato lineage and 40 PCs 15 neighbours for Mesp1^{Cre};R26RTomato lineage. Clusters characterized by an enrichment of mitochondrial genes were discarded from subsequent analyses. Final dataset visualization was produced using UMAP on a joint kNN-graph (10 PCs with 50 neighbours). Identified populations were used to create pseudo-bulk ATAC-seq openness profiles using export_pseudobulk function from pycisTopic. Open chromatin peaks were mapped using MACS2 via peak_calling function from pycisTopic with default parameters.

3.7 CUT&RUN Antibody Validation

To validate the spatial expression and specificity of SOX10, TFAP2A, ETS1, and FOXD3 antibodies in embryonic neural crest cells, E9.5 mouse embryos were cryosectioned and analyzed by immunohistochemistry. (Fig. 7, Results & Discussion) Antibodies for CUT&RUN were selected based on neural crest specificity. All four transcription factors are known canonical neural crest transcription factors, highly expressed and active in cranial neural crest cells at E9.5 in mouse embryos, orchestrating neural crest cells delamination and EMT. To further verify this, antibody validation was performed to ensure robust and specific detection of target proteins *in situ*, which is crucial for the efficiency of CUT&RUN experiments⁶¹.

E9.5 wild type C57BL/6JRj embryos were collected in ice-cold phosphate-buffered saline. They were then fixed in 4% paraformaldehyde (prepared from 16% stock, Thermo Fisher Scientific) at 4 °C for 1.5–2 h with gentle agitation. After fixation, embryos were washed three times in Phosphate-Buffered Saline (1X Gibco, Fischer Scientific GmbH) and incubated in 30% sucrose at 4 °C for 24 h. For embedding, embryos were incubated in a 1:1 mixture of 30% sucrose and Optimal Cutting Temperature (OCT) compound for 2h at room temperature, followed by incubation in 100% OCT for an additional 2h. Samples were then embedded in OCT and stored at -20 °C before sectioning. Cryosections were cut at a thickness of 16 µm.

Prior to immunostaining, sections underwent antigen retrieval using Dako Antigen Retrieval Buffer (Agilent Technologies, USA), followed by permeabilization and blocking in Normal Donkey Serum (NDS, Jackson ImmunoResearch). Sections were incubated overnight at 4 °C with primary antibodies against Sox10, Tfap2a, Ets1, and Foxd3 diluted in PBST (Phosphate-Buffered Saline + 0,1% Tween) and Normal Donkey Serum.

Following washes, sections were incubated with fluorescently conjugated secondary antibodies. Nuclei were stained with DAPI (Thermo Fisher Scientific, Invitrogen). After staining, sections were mounted using Mowiol mounting medium (Millipore) and coverslipped to preserve fluorescence.

A complete list of primary and secondary antibodies is provided in Tables 3 and 4.

Immunostained sections were imaged using a Zeiss LSM 780 confocal microscope with a 10× objective to assess signal specificity and cellular localization.

Table 3-Antibody list for primary antibodies used for validation.

Target	Primary Antibody (Cat. #)	Company	Host
Sox10	#AF2864	Bio-Techne Ireland Limited	Goat
Sox10	#ab155279	Abcam	Rabbit
Tfap2a	#ab108311	Abcam	Rabbit
Ets1	# CST14069	Cell Signaling Technology	Rabbit
Foxd3	#ab64807	Abcam	Rabbit

Table 4-Antibody list for secondary antibodies used for validation.

Target	Secondary Antibody (Cat. #)	Company	Host
Anti-Rabbit	Alexa Fluor 488	Thermo Fisher Scientific	Donkey
Anti-Mouse	Alexa Fluor 488	Thermo Fisher Scientific	Donkey
Anti-Goat	Alexa Fluor 647	Thermo Fisher Scientific	Donkey

3.8 CUT&RUN Library Preparation and Sequencing

CUT&RUN experiments were carried out following CUTANA™ CUT&RUN kit instructions (Epicpypher, 14-1048). The Wnt1^{Cre};R26RTomato positive cell samples consisted of 30 000 cells per sample, and the wild type samples consisted of up to 300,000 cells per sample. Antibodies used were anti-SOX10 (Abcam, ab155279), anti-TFAP2A (Abcam, ab108311), anti-ETS1 (Cell Signaling Technology, CST14069/Active Motif am39850), anti-FOXD3 (#ab64807), anti-H3K4me3 (Epicpypher, 13-0060), anti-H3K27me3 (Epicpypher, 13-0055), and anti-H3K4me1 (Epicpypher, 13-0057). Libraries were prepared following CUTANA™ CUT&RUN Library Prep kit instructions (Epicpypher, 14-1001). Indexing PCR was run for 14 cycles for high cell number samples (over 100K cells per sample) and for 16 cycles for low cell number samples (under 10K cells per sample). Libraries were sequenced using Illumina Novaseq6000 (v1.5).

3.9 CUT&RUN Data Processing

Raw CUT&RUN reads were processed using nf-core cutandrun pipeline v 3.2.2 (nextflow⁷¹, nf-core⁷²) by a bioinformatician in the laboratory. Briefly, reads were trimmed using Trimgalore and quality control was performed using FastQC (Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and MultiQC⁷³. Reads were aligned with Bowtie2⁷⁴ with parameters --end-to-end --very-sensitive --no-mixed --no-discordant --minins 10 --maxins 700 --dovetail to the mouse mm39 reference genome as well as to the E. coli K12-MG1655 genome for the spike-in normalisation. Reads were filtered with SAMtools (-f 0x2 -q 30)⁷⁵ and reads corresponding to mm39 Excluderanges v. 0.99.8⁷⁶ were excluded with BEDtools intersect⁷⁷. Duplicated reads were removed using Picard MarkDuplicates (Broad Institute. *Picard Toolkit*). CUT&RUN coverage was computed using BAMCoverage from deepTools⁷⁸. Peak calling was performed using MACS2⁷⁹ for the Tfap2a and Sox10 samples, and SEACR⁸⁰ for the remaining samples.

3.10 SNP Dataset Collection

Genetic variants associated with orofacial clefts were obtained from two published Genome-Wide-Association Studies. The first dataset was derived from Leslie et al. 2017 and the second was obtained from Dack et al. 2025. For the Dack et al. dataset, all 27 reported SNP regions were included for analysis. For the Leslie et al. dataset, SNPs were filtered based on genome-wide significance ($p\text{-value} < 5 \times 10^{-8}$). Two separate SNP lists were then generated, each corresponding to one study. These two studies were selected for their comprehensive coverage of associated risk *loci*, and for their focus on non-syndromic variants, which are characterized by genetic heterogeneity and do not follow simple Mendelian inheritance patterns.

3.11 Genomic Region Expansion and LiftOver

To capture the regulatory landscape surrounding GWAS-associated variants, each SNP coordinate was extended by ± 50 kilobase pairs because the SNPs reported in association studies are not necessarily the causal variants underlying disease risk. Instead, the functional variant can reside in the nearby region, that due to linkage disequilibrium is inherited together with the SNP. Therefore, regulatory elements associated with orofacial cleft susceptibility do not necessarily overlap the reported SNP directly. Expanding each

locus by 50 kilobase pairs allowed the inclusion of nearby candidate regulatory regions, including enhancers and promoters, that may contain functional variants associated with the observed GWAS signal. This window size was selected as a heuristic approach to prioritize proximal regulatory elements while maintaining a manageable search space for downstream analyses⁸³.

The resulting human genomic coordinates (GRCh38/hg38) were then converted to the mouse reference genome assembly (GRCm39/mm39) using the UCSC Genome Browser LiftOver tool (UCSC Genome Browser, accessed March 2025) and default parameters. Conversion to mm39 was necessary to ensure coordinate compatibility with the CUT&RUN and 10X Multiome datasets generated in this study. Regions that mapped partially or were split across multiple *loci* during cross-species conversion were kept in their entirety to avoid loss of potentially relevant regulatory sequences. All successfully converted regions, including fully mapped, partially mapped and split intervals mouse genomic regions were combined into a single BED file for use in downstream analysis.

3.12 Selection of Expanded Regions

To assess the regulatory potential of each expanded SNP-associated region, overlaps with CUT&RUN and neural crest scATAC-seq datasets were quantified using sequential BEDTools intersect -c operations. CUT&RUN data comprised binding profiles for four neural crest specific transcription factors: SOX10, TFAP2A, FOXD3, ETS1 and three histone modification marks: H3K4me1, H3K4me3, and H3K27me3. Each dataset was intersected independently against the expanded SNP regions, appending a count column per track while leaving the original genomic coordinates unchanged. The resulting annotated file was filtered using command awk to retain only regions with a summed transcription factor overlap count of ≥ 1 , ensuring that each candidate region harbored at least one transcription factor binding event. Histone modification and chromatin accessibility counts were retained as annotations for downstream prioritization. All intersections were additionally manually verified by using the Integrative Genomics Viewer (IGV) web application (; Robinson et al., 2023). Genomic regions displaying consistent overlap between SNP-associated *loci* and CUT&RUN/neural crest scATAC-seq signals were prioritized as candidate regulatory regions containing putative enhancers potentially involved in craniofacial development and neural crest associated gene regulation.

3.13 Identification of Putative Enhancers and Promoters

To define precise enhancer and promoter boundaries within the candidate regions, neural crest cell scATAC-seq peaks were intersected against the filtered SNP-associated regions using BEDTools intersect -u, retaining only those open chromatin peaks that fell within at least one candidate region. This approach was chosen over collapsing coordinates through the intersection of multiple CUT&RUN tracks, which would have created artificially narrow intervals inconsistent with any single biological unit, given the intrinsic size differences between transcription factor binding profiles, typically 200 to 500 base pairs, and histone modification domains, which can span several kilobases^{84,85}. Open chromatin regions identified by scATAC-seq directly reflect nucleosome eviction and chromatin accessibility, representing the physical prerequisite for enhancer activity and therefore the biologically meaningful unit for regulatory element boundary definition.

The resulting ATAC-based intervals were subsequently re-annotated with CUT&RUN peak counts for all transcription factors (SOX10, TFAP2A, ETS1, AND FOXD3) and histone modification marks (H3K4me1, H3K27me3, and H3K4me3) using sequential BEDTools intersect -c operations, producing a fully annotated putative enhancer file. This step was key to define enhancer boundaries, as the ATAC peak coordinates were used as the primary coordinate framework for each regulatory region rather than the broader expanded SNP windows.

3.14 Finding the Target Genes for Identified Regulatory Elements

Each regulatory element was then assigned to its nearest gene using BEDTools command closest against a GRCm39/mm39 gene annotation file, providing a regulatory region-to-gene mapping for downstream analysis. Genomic regions with simultaneous chromatin accessibility, transcription factor occupancy, and histone modification marks within expanded SNP-associated *loci* were defined as high-confidence (*cis*-regulatory regions, including both enhancers and promoters), representing regions potentially involved in craniofacial development and neural crest gene regulatory network, as well as OFC development.

Following identification of putative *cis*-regulatory regions, each candidate region was associated with its nearest gene based on genomic proximity using BEDTools closest against a sorted GRCm39/mm39 gene annotation file. While this approach does

not account for distal chromatin interactions, in the absence of interaction mapping data such as Hi-C, proximity-based gene assignment represents the most tractable and widely adopted strategy for enhancer-to-gene linking, consistent with approaches used in comparable regulatory genomics studies. All assignments were additionally verified in the IGV web application.

3.15 Distinguishing Putative Enhancers and Promoters

Putative *cis*-regulatory regions were manually inspected using IGV web application to understand their genomic context relative to their assigned nearest gene. Regions overlapping annotated promoter regions were classified as promoters. Regions that did not overlap promoter-associated annotations but exhibited chromatin accessibility, transcription factor occupancy, and histone modification patterns consistent with regulatory activity were classified as putative enhancers. These enhancer candidates could occur within intronic or intergenic and upstream or downstream of their nearest genes.

Histone modification profiles were used to support this classification. Enrichment of H3K4me3 was considered indicative of promoter-associated chromatin, whereas H3K4me1 enrichment in the absence of promoter overlap was taken as support of enhancer activity.

3.16 Gene Expression Analysis

To assess the biological relevance of the candidate *cis*-regulatory region-associated genes, expression profiles of the putative target genes were examined using the scRNA-seq dataset generated in this study. The four candidate genes with the highest mean expression in Wnt1^{cre}:R26RtdTomato positive cells in the 10x Multiome dataset were selected for characterization. Expression plots and heatmaps were generated in Scanpy to visualize gene expression across cell populations, allowing assessment of enrichment or specificity within neural crest-derived cell types. Genes displaying enriched or specific expression in neural crest populations were prioritized as strong candidate targets of the identified regulatory elements.

Two additional candidate enhancers were selected on the basis of their chromatin signatures: both regions displayed overlap with all three histone modification marks examined (H3K4me1, H3K4me3, and H3K27me3) indicative of a highly active regulatory state, as well as binding of three of the four transcription factors profiled

(Sox10, Tfap2a, and Ets1). Expression plots and heatmaps for the associated genes were generated in Scanpy to further characterize their activity across neural crest cell populations. Altogether, this integrative approach linked putative *cis*-regulatory regions to candidate target genes expressed in neural crest cells, supporting their potential role in craniofacial development and susceptibility to orofacial clefts.

3.17 Motif Analysis

While the binding regions of four transcription factors (SOX10, TFAP2A, ETS1 and FOXD3) were experimentally studied, enhancers and promoters typically contain binding sites for multiple transcription factors. Additionally, several putative regulatory regions did not show detectable FOXD3 occupancy, despite displaying chromatin accessibility and regulatory histone signatures, suggesting that additional transcriptional factors may contribute to their activity. Therefore, a *de novo* motif analysis was performed to identify additional transcription factor binding signatures within the regulatory regions, providing a broader and more comprehensive view of the regulatory networks potentially acting at orofacial cleft-associated *loci*.

De novo motif enrichment analysis was performed using HOMER (findMotifsGenome.pl) on the putative *cis*-regulatory regions identified through overlap between SNP-associated *loci* and CUT&RUN/scATAC-seq signals. These regions were analyzed against the GRCm39/mm39 reference genome using HOMER default parameters. Enriched sequence motifs were identified relative to matched genomic background regions and ranked according to enrichment significance.

4. Results

4.1 10x Multiome Allows Selection of Neural Crest Cells for Further Analysis

To characterize the transcriptional and chromatin accessibility landscape of both neural crest and mesodermal cell populations at E9.5, single-cell multiomic profiling (10x Multiome, combining scRNA-seq and scATAC-seq) was performed on TOMATO-positive cells isolated from *Wnt1^{Cre};R26RtdTomato* and *Mespt1^{Cre};R26RtdTomato* positive embryonic heads at E9.5. While this study primarily focuses on *Wnt1* derived neural crest cells, *Mespt1* derived cells were processed in parallel as part of the generation of a broader single-cell multiomic reference dataset that is being used in ongoing studies and was used here for gene expression analyses. Following quality control filtering and dimensionality reduction, data from both lines were integrated and transcriptionally distinct cell populations were visualized in a combined UMAP. Cell type annotations were assigned based on the expression of known marker genes (Supplementary Table 1), revealing 13 transcriptionally distinct clusters containing neural crest and mesodermal cell populations present at E9.5 (Fig.7)



Figure 7 - UMAP visualization of single-cell RNA sequencing data from E9.5 mouse embryos.

Uniform Manifold Approximation and Projection (UMAP) representation of integrated scRNA-seq data generated using the 10x Genomics platform. Each point represents an individual cell, grouped according to transcriptional similarity. Distinct clusters, shown in different colours, correspond to transcriptionally defined cell populations identified during embryonic development at E9.5. Cell identities were assigned based on established marker gene expression profiles.

Chromatin accessibility profiles generated from the same cells were analyzed in parallel using the scATAC-seq component of the 10x Multiome dataset. Following quality control

filtering and dimensionality reduction, chromatin accessibility data from both *Wnt1^{Cre};R26RTomato* and *Mespt1^{Cre};R26RtdTomato* embryos were integrated and visualized using UMAP (Fig. 8). The resulting accessibility landscape recapitulated the major cell populations identified in the transcriptomic analysis, allowing chromatin accessibility profiles to be assigned to the corresponding transcriptionally defined cell clusters.



Figure 8 - UMAP visualization of chromatin accessibility ATAC sequencing data from E9.5 mouse embryos.

Uniform Manifold Approximation and Projection (UMAP) representation of integrated scATAC-seq data generated using the 10x Genomics platform. Each point represents an individual cell, grouped according to transcriptional similarity. Distinct clusters, shown in different colours, correspond to transcriptionally defined cell populations identified during embryonic development at E9.5. Cell identities were assigned based on established marker gene expression profiles.

Neural crest cell clusters (pre-migratory and migratory neural crest) were subsequently selected for downstream analyses. Chromatin accessibility profiles derived from these clusters were used for the identification of putative enhancers and promoters through the definition of regulatory element boundaries, while the matched scRNA-seq data provided the transcriptional reference for downstream gene expression analyses, including candidate gene prioritization based on mean expression values.

4.2 CUT&RUN Antibody Validation Shows Signal Specificity

Immunohistochemical staining of E9.5 mouse embryo cryosections confirmed the specificity of SOX10, TFAP2A, and ETS1 antibodies in cranial neural crest cells. The SOX10 #AF2864 antibody, previously validated in the laboratory, served as a neural crest reference control and is shown in the red channel across all rows. The green channel displays the signal of each antibody under validation, with DAPI nuclear in the blue channel and a merged overlay in the final column. All images show cranial neural crest

cells actively delaminating from the neural tube, consistent with the expected spatiotemporal expression of these transcription factors at E9.5. SOX10, TFAP2A, and ETS1 antibodies produced specific and robust signals overlapping with the SOX10-positive neural crest cells. FOXD3, however, showed poor signal intensity. Despite this, all four antibodies were used for CUT&RUN experiments, given the well established role of FOXD3 as a canonical neural crest transcription factor at this developmental stage.

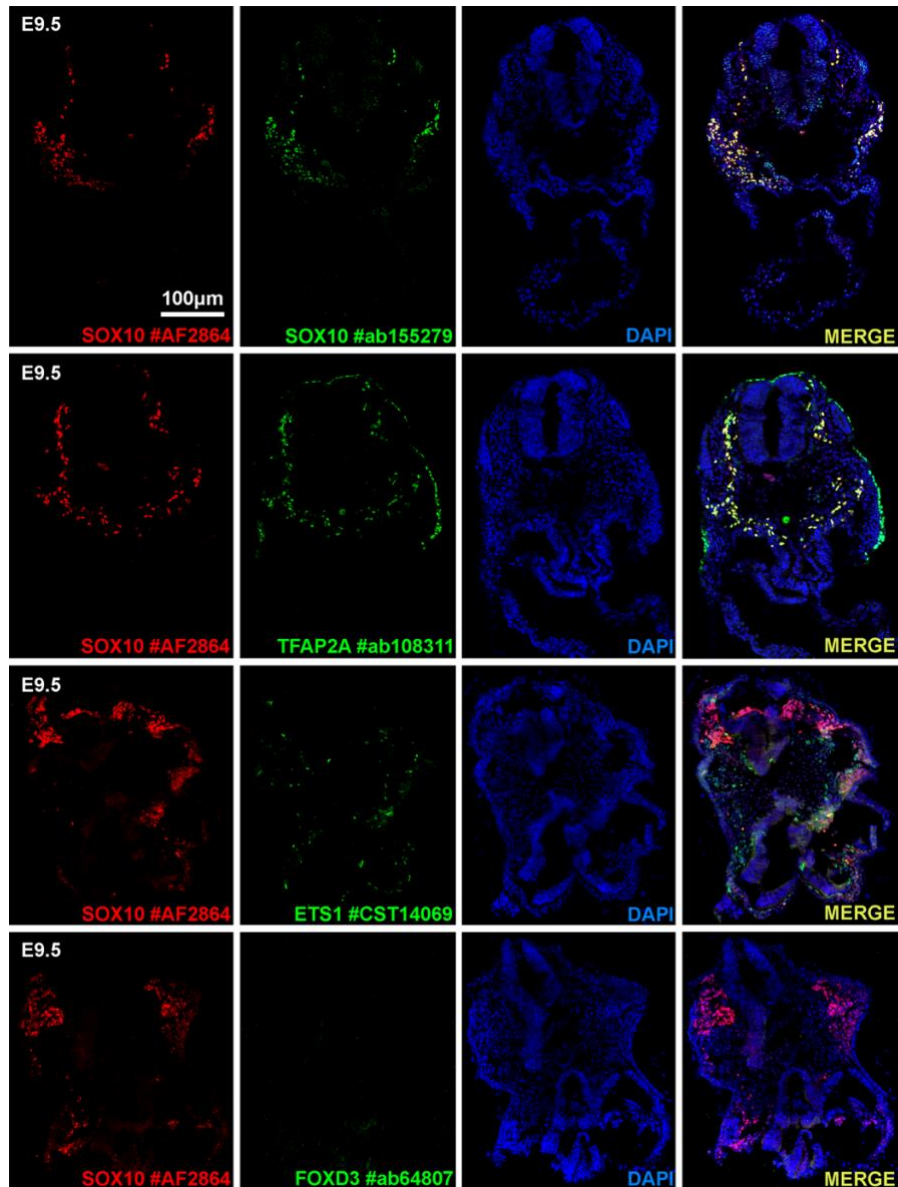


Figure 7 – Immunohistochemical validation of neural crest specific transcription factor antibodies used in CUT&RUN experiments.

Cryosections (16 µm) of E9.5 wild type C57BL/6JRj embryos were immunostained for SOX10, TFAP2A, ETS1, and FOXD3. Each row shows the SOX10 #AF2864 signal (red) as a neural crest reference control, the target transcription factor signal (green), DAPI nuclear counterstain (blue), and a merged channel overlay. Rows 1-4: SOX10 #AF2864 (red) co-stained with SOX10 #ab155279, TFAP2A #ab108311, ETS1 #CST14069, and FOXD3 #ab64807 (green), respectively. Co-localization of the target signal with the

SOX10-positive neural crest cells in the merged panels confirms antibody specificity and spatial expression in cranial neural crest cells at E9.5. Scale bar=100 μ m.

4.3 Putative Enhancers and Promoters Identification

From the 100kb SNP-associated regions (see methods section “Genomic Region Expansion and LiftOver”), 156 regions were selected based on binding of neural crest specific transcription factors (SOX10, TFAP2A, ETS1 and FOXD3), presence of histone modification marks (H3K4me3, H3K4me1, H3K27me3), as well as chromatin accessibility signal (table provided upon request). Following genomic intersection and definition of precise boundaries using scATAC peak coordinates, a refined set of putative enhancers and promoters was obtained (table provided upon request) and paired with the respective nearest target genes.

In total, 65 enhancer/promoter-gene pairs were identified, of which 43 genes were expressed in neural crest clusters of 10X Multiome at E9.5 (Fig. 7). Five genes corresponding to Gm or Rik annotations were excluded from downstream analyses. Gm annotations typically represent computationally predicted genes, whereas Rik annotations originate from RIKEN cDNA sequencing projects and often correspond to transcripts with limited functional characterization (Mouse Genomics, 2024). Due to the lack of established gene functions and limited annotation available for these *loci*, their potential roles in craniofacial development and orofacial cleft susceptibility could not be reliably assessed, and they were therefore excluded from further analysis. This resulted in a final set of 38 enhancer/promoter-gene pairs summarized in Table 5, containing genomic coordinates of putative regulatory elements, transcription factor binding profiles (SOX10, TFAP2A, FOXD3, ETS1), histone modification signals (H3K4me1, H3K27me3, H3K4me3), gene annotations, strand orientation, and distances to the nearest genes.

Table 5 – Final set of 38 regulatory elements-genes pairs in E9.5 mouse neural crest scRNA-seq data and respective information. Genomic coordinates (Chr, Start, End) correspond to the mouse reference genome assembly GRCm39/mm39. Regulatory elements were defined from neural crest scATAC-seq peaks overlapping OFC-associated GWAS loci and linked to their nearest genes. Columns corresponding to Sox10, Tfap2a, Foxd3, Ets1, H3K4me1, H3K27me3, and H3K4me3 indicate the number of overlapping CUT&RUN peaks identified for the respective transcription factor or histone modification mark. A value of 0 indicates no overlap, whereas values ≥ 1 indicate one or more overlapping peaks. Gene column indicates the nearest gene assigned to each regulatory element, and the Strand column indicates gene orientation on the reference genome, with “+” indicating the forward strand and “-” indicating the reverse strand. Highlighted in grey, the four genes with the highest mean expression values in neural crest cell clusters at E9.5, their respective regulatory regions and features.

Chr	Start	End	Sox10	Tfap2a	FoxD3	Ets1	H3K4me1	H3K27me3	H3K4me3	Gene	Strand
chr1	75497752	75498854	1	1	0	1	1	0	1	Stk11ip	+
chr1	75522380	75523806	0	0	0	0	1	0	1	Slc4a3	+
chr1	192533351	192533704	0	1	0	0	1	1	0	Sertad4	-
chr1	192717298	192718782	1	1	0	1	0	1	1	Syt14	-
chr1	192812030	192813018	0	0	0	1	0	0	1	Utp25	-
chr1	192835079	192835897	0	0	0	0	1	1	0	Irf6	+
chr1	192871713	192871996	0	0	0	1	1	0	0	Traf3ip3	-
chr1	192907863	192908019	0	0	0	0	1	0	0	Hsd11b1	-
chr1	192954667	192955910	0	0	0	0	1	1	0	G0s2	-
chr11	62139080	62141040	0	1	0	1	1	1	1	Adora2b	+
chr11	62161452	62162105	0	0	0	0	1	0	0	Zswim7	-
chr11	62171480	62172953	0	1	0	1	0	0	1	Ttc19	+
chr11	68203983	68204258	0	0	0	0	1	0	0	Ntn1	-
chr11	68322671	68323587	0	0	0	1	0	1	0	Pik3r5	+
chr12	12401946	12402064	0	0	0	0	0	0	0	Cyria	+
chr19	7459137	7461184	1	1	0	1	0	0	1	Rtn3	-
chr19	7471236	7472386	0	0	0	1	0	0	1	Ati3	+
chr19	59121660	59122152	0	0	0	1	1	1	0	Vax1	-
chr2	160207879	160210632	1	1	0	1	2	1	1	Mafb	-
chr3	121811754	121811928	0	1	0	1	1	0	0	Arhgap29	+
chr3	121839815	121840321	0	1	0	0	1	1	0	Abca4	+
chr4	46352978	46353664	0	0	0	1	1	1	0	Foxe1	+
chr4	46371089	46371402	0	0	0	0	1	1	0	Trmo	-
chr4	124743401	124745347	1	1	0	1	2	0	1	Mtf1	+
chr4	124743401	124745347	1	1	0	1	2	0	1	Yrdc	+
chr4	124755290	124756524	0	0	0	1	0	1	1	Maneal	-
chr4	139538168	139539097	0	1	0	0	1	1	0	Pax7	-
chr4	139548140	139552606	1	1	0	1	1	1	0	Pax7	-
chr6	32292176	32292299	0	0	0	0	0	0	0	Plxna4	-
chr8	88888171	88888309	0	0	0	0	0	0	0	Heatr3	+
chr8	88951803	88952934	0	0	0	0	1	0	0	Tent4b	+
chr8	89011409	89011600	0	0	0	0	1	0	0	Adcy7	+
chr8	128225316	128226007	0	0	0	0	1	0	0	Pard3	+

chr8	129083675	129086762	1	1	0	1	1	1	1	Nrp1	+
chr8	129218195	129218648	0	0	0	0	1	0	0	Nrp1	+
chr8	129411543	129413072	1	1	0	0	0	0	1	Itgb1	+
chr8	129601941	129602223	0	0	0	0	0	0	0	Ccdc7a	-
chr8	129960284	129961191	0	0	0	1	0	0	1	Zfp1006	-

In summary, we identified 13 putative enhancer regions associated with genes *Sertad4*, *Traf31p3*, *Hsd11b1*, *Zswim7*, *Ntn1*, *Vax1*, *Foxe1*, *Trmo*, *Pax7*, *Tent4b*, *Adecy7*, *Pard3* and *Nrp1* genes, and 25 promoter regions.

4.4 Description of Selected *Cis*-Regulatory Regions

To further describe these pairs, from Table 5, the four genes with the highest mean expression values (*Itgb1*, *Rtn3*, *Mafb*, and *Pard3*) in neural crest cell clusters at E9.5 and their respective *cis*-regulatory regions were selected for downstream analysis (Table 6). This selection was corroborated by UMAP visualization and heatmap clustering (Fig. 10). The transcription factor binding profiles and histone signatures of the respective regulatory regions were subsequently assessed to understand which type of regulatory element was associated with each gene.

Table 6 - Four highest expressed genes associated with orofacial cleft SNP regions identified in mouse neural crest scRNA-seq data at E9.5

Gene name	Expression mean value
<i>Itgb1</i>	0.957643
<i>Rtn3</i>	0.831788
<i>Mafb</i>	0.658041
<i>Pard3</i>	0.628575

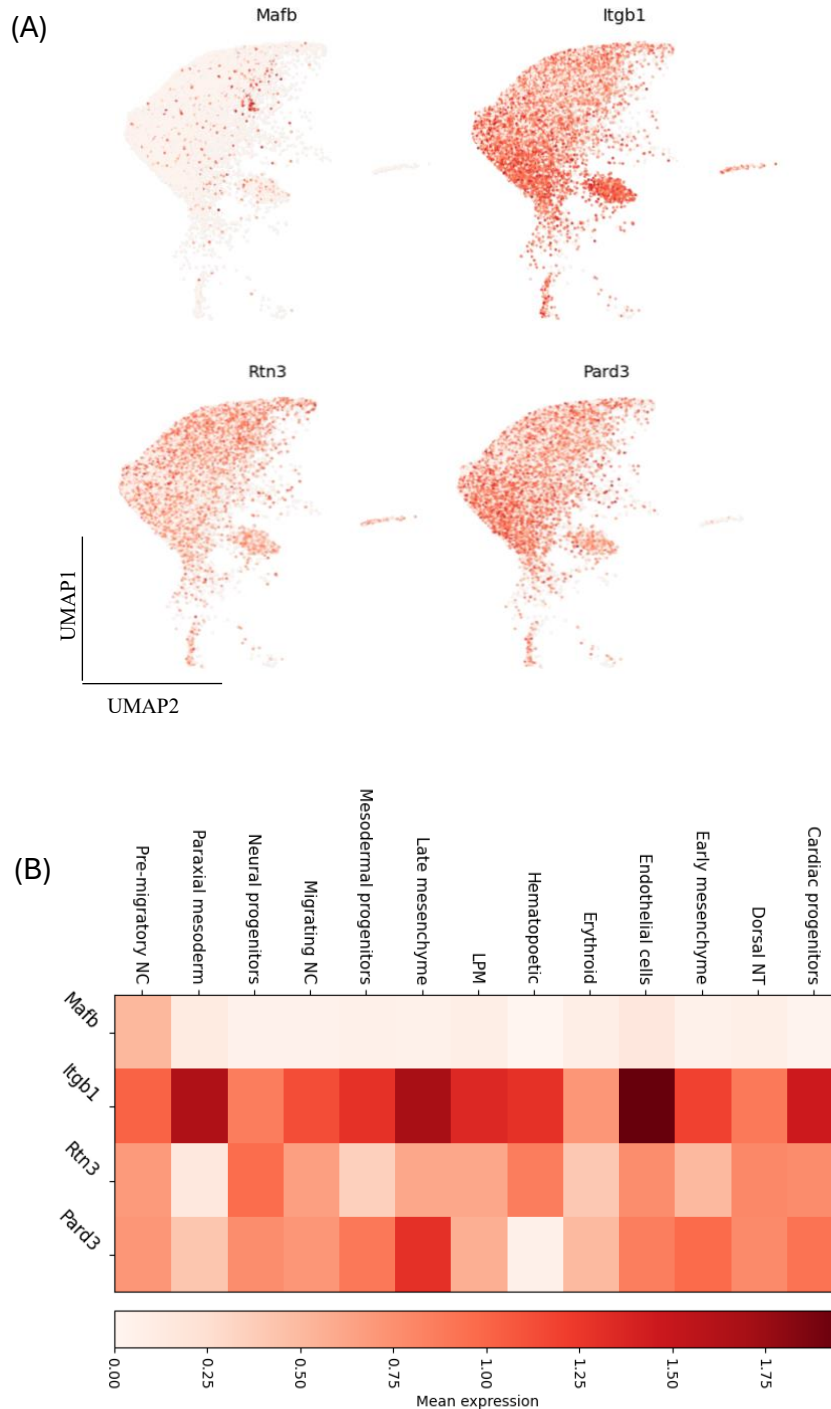


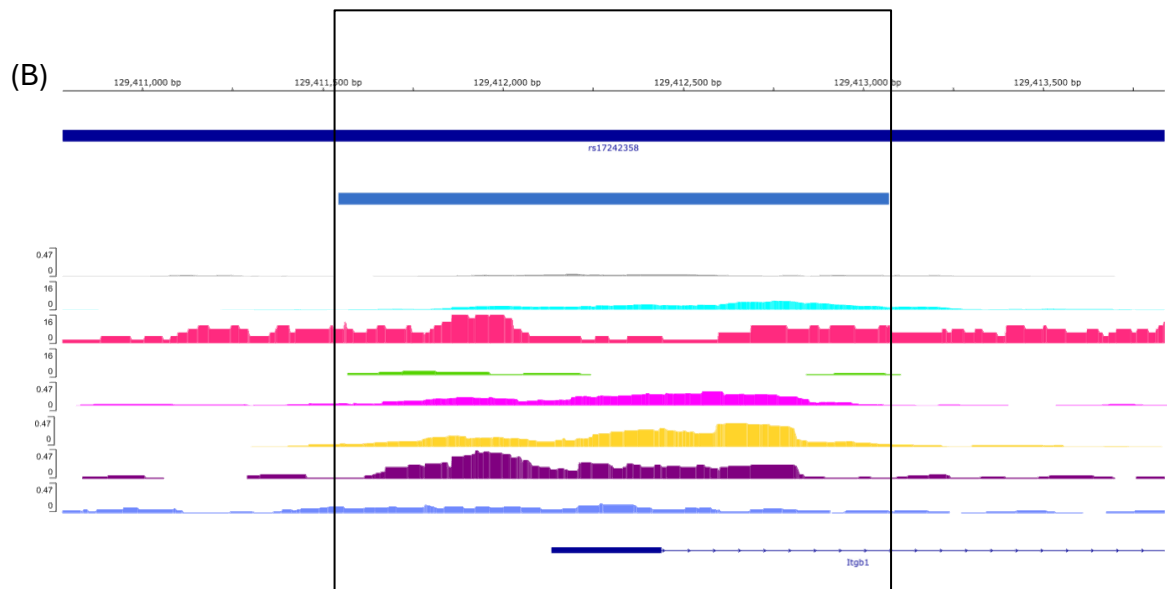
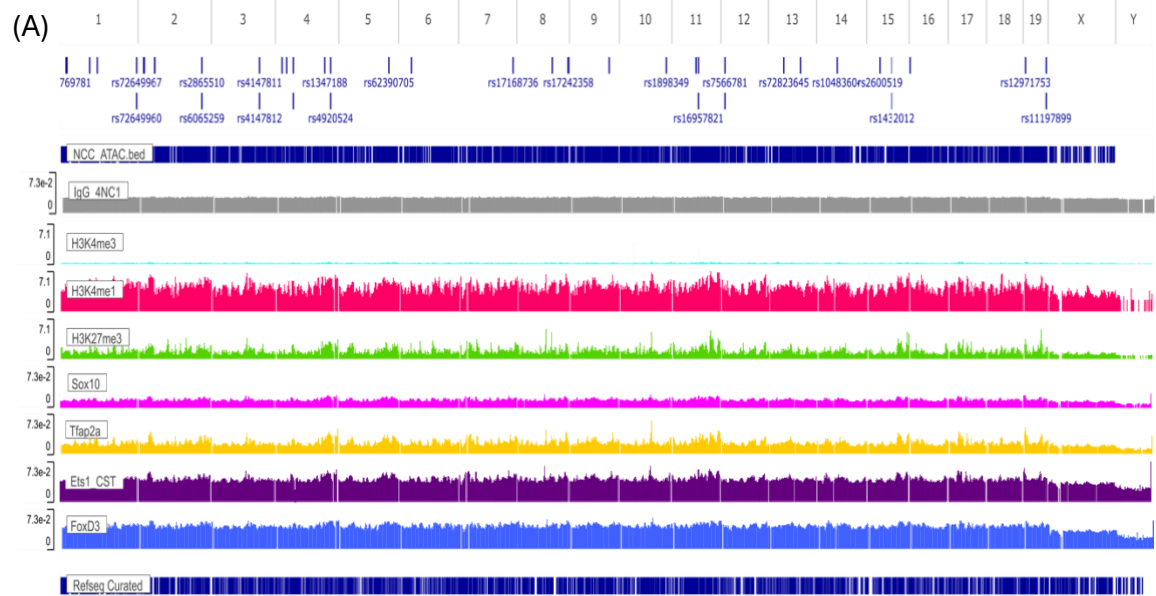
Figure 10 - UMAP and Heatmap visualization of four highest expressed genes present in E9.5 mouse scRNA-seq data.

(A) UMAP representations of integrated scRNA-seq data generated using the 10x Genomics platform. Each point represents an individual cell. Individual UMAPs display the expression patterns of four selected genes showing the highest expression in neural crest cells at E9.5 across all cells from E9.5 mouse embryos. Cells are colored according to normalized gene expression levels, with higher intensity indicating increased expression. (B) Heatmap with mean expression levels of four selected genes with highest expression in neural crest cells at E9.5 across transcriptionally clustered cell populations. Expression values are represented by color intensity, with higher intensity indicating increased relative expression.

Expression of all four genes was detected in one or more neural crest cell clusters, however, their expression was not restricted to these populations and was also observed across other embryonic cell populations, namely in *Mesp1^{Cre};R26RtdTomato* derived cells. The regulatory features of the genomic regions associated with these genes are summarized in Table 5 and highlighted in grey.

Among the four candidate regulatory regions, the region associated with *Mafb* displayed the most complex regulatory profile. Binding was detected for three of the four transcription factors profiled (SOX10, TFAP2A, and ETS1), together with two H3K4me1 peaks, one H3K27me3 peak, and one H3K4me3 peak. Manual inspection using IGV confirmed that this region overlapped an annotated promoter region.

In comparison, the regulatory region associated with *Itgb1* exhibited a more restricted profile, with binding detected for SOX10 and TFAP2A and enrichment for a single H3K4me3 peak. Similarly, the region associated with *Rtn3* displayed H3K4me3 enrichment together with SOX10, TFAP2A, and ETS1 occupancy. Manual inspection in IGV showed that both regions also overlapped annotated promoter regions (Fig. 12 B, C). In contrast, the regulatory region associated with *Pard3* showed enrichment for H3K4me1 but lacked detectable binding by any of the profiled transcription factors. Unlike the other candidate regions, IGV visualization indicated that this region was located within an intronic region rather than overlapping an annotated promoter (Fig.12 D). Based on its genomic location and chromatin profile, this region was classified as a putative enhancer. Overall, three of the four candidate regulatory regions were associated with annotated promoter regions and displayed H3K4me3 enrichment, whereas the *Pard3* associated region exhibited an intronic localization together with an H3K4me1 signal, indicative of a putative enhancer region. The genomic context and chromatin profiles of all four candidate regulatory regions are shown in Figure 12.



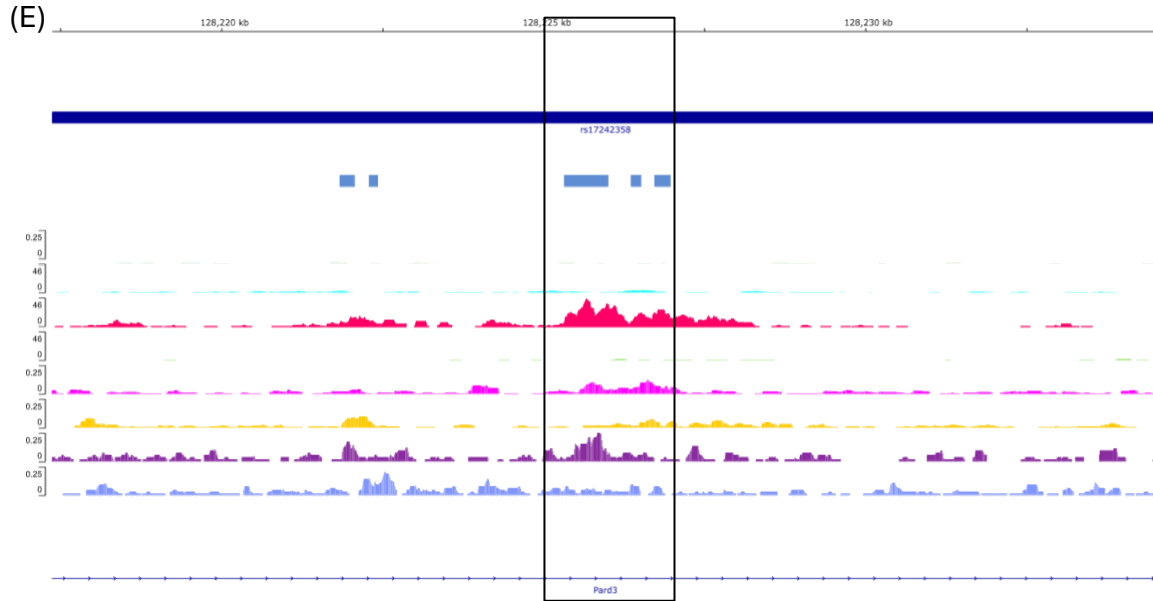


Figure 12 - Chromatin profiles of candidate regulatory regions associated with craniofacial GWAS loci in neural crest cells.

(A) IGV visualization showing an overview of all four candidate regulatory regions across their respective chromosomal contexts. Tracks displayed from top to bottom: chromosomal coordinates, expanded SNP-associated regions, neural crest cell scATAC-seq signal, IgG control CUT&RUN histone modification tracks (H3K4me3, H3K4me1, and H3K27me3), CUT&RUN transcription factor occupancy tracks (SOX10, TFAP2A, ETS1, and FOXD3), and RefSeq curated gene annotations.

(B–E) IGV visualizations showing chromatin profiles of individual candidate regulatory regions associated with *Itgb1* (B), *Rtn3* (C), *Mafb* (D), and *Pard3* (E). For each region, tracks are displayed in the same order as in (A). Expanded SNP-associated regions are highlighted to indicate the boundaries of each candidate locus. The *Pard3*-associated region (E) is located within an intronic region and displays H3K4me1 enrichment without detectable transcription factor occupancy, consistent with a putative enhancer. The remaining regions (B–D) overlap annotated promoters and show H3K4me3 enrichment.

Additionally, two regulatory region-gene pairs lacking established associations with orofacial clefts were selected for further characterization based on their chromatin signatures: both regions displayed overlap with all three histone modification marks examined (H3K4me1, H3K4me3, and H3K27me3) indicative of a highly active regulatory state, as well as multiple binding of the transcription factors profiled (SOX10, TFAP2A, and ETS1). These findings can be seen as multiple lines of evidence for the activity of these regions during neural crest development. In addition, these putative target genes have not been associated with OFCs before, making them possible new candidate genes behind these pathologies.

The *Nrpl* gene was associated with two distinct regulatory regions: one promoter and one putative enhancer. The promoter region displayed a complex regulatory profile, with binding detected for three of the four transcription factors profiled (SOX10, TFAP2A, and ETS1), together with one H3K4me1 peak, one H3K27me3 peak, and one H3K4me3

peak. In contrast, no transcription factor binding was detected within the putative enhancer region, which exhibited a single H3K4me1 peak. Manual inspection in IGV confirmed that these regions overlapped an annotated promoter region and an intronic region, respectively.

The *Adora2b* gene was associated with a single regulatory region displaying a promoter-like chromatin signature. This region showed enrichment for one H3K4me1 peak, one H3K27me3 peak, and one H3K4me3 peak, together with binding by the transcription factors TFAP2A and ETS1. IGV visualization confirmed this region overlapped an annotated promoter region.

4.5 Motif Analysis Reveals *de novo* Significantly Enriched Motifs

Although several putative enhancers and promoters showed occupancy for SOX10, TFAP2A, and ETS1, FOXD3 binding was not observed within any of our regions. In addition, while the five analyzed candidate promoter regions showed binding of several of the studied transcription factors, the putative enhancers of *Pard3* and *Nrpl* did not. Enhancers and promoters typically contain binding sites for multiple transcription factors, and additional potential transcription factors were sought for using motif analysis.

Motif analysis was conducted against 96902 background sequences and identified 46 significantly enriched motifs (Supplementary Table 2). Among the enriched motifs were transcription factors with established roles in development and neural regulation, including Hoxd13, Pax8, Sox9, and Foxa1. Several of these factors have previously been implicated in embryonic patterning, neural crest development, and craniofacial morphogenesis^{1,13,14,16,23}.

From the full motif set, four candidates were prioritized for further description based on their relevance to neural crest biology and craniofacial development.

Table 8 - *De novo* enriched transcription factor motifs prioritized for neural crest and craniofacial regulatory relevance.

Motif	TF name	<i>p</i> -value	% of target regions
	Lef1	1e-38	19.85%
	Msx1	1e-29	5.82%
	RARG	1e-47	3.63%
	SMAD4	1e-48	1.81%

5. Discussion

5.1 Identification of Candidate Regulatory Elements at Orofacial Cleft Risk Loci

The regulatory mechanisms through which non-coding genetic variation contributes to orofacial cleft susceptibility remain to be fully understood. While GWAS have identified numerous risk *loci* associated with cleft lip and palate, identifying and linking functional regulatory elements to target genes underlying these associations remains challenging. Here, the integration of neural crest cell chromatin accessibility, transcription factor occupancy, and histone modification signatures allowed the identification of 38 candidate regulatory regions associated with genes expressed during craniofacial development.

Of the 38 associated genes identified, only 7 (*Irf6*, *Arhgap29*, *Foxe1*, *Pax7*, *Vax1*, *Mafb*, and *Ntn1*) have strong, previously established links to orofacial clefts through GWAS^{37,39,51,86}. Importantly, and to my knowledge, none of the putative enhancer or promoter regions identified in this study have been previously described in the context of orofacial cleft-associated risk *loci*. These findings, therefore, provide a set of candidate regulatory elements that may contribute to the interpretation of non-coding genetic variation associated with craniofacial development and orofacial clefts.

5.2 Regulatory Signatures Associated with Highest Expressed Neural Crest Genes

Among the candidate regulatory regions identified, the *Mafb* associated region displayed the most complex regulatory signature. This region demonstrated occupancy by SOX10, TFAP2A, and ETS1 together with enrichment for H3K4me1, H3K4me3, and H3K27me3. Although manual inspection using IGV confirmed this region overlapped an annotated promoter region, the simultaneous presence of H3K4me3 and H3K27me3 could be indicative of a bivalent chromatin configuration, a regulatory state capable of maintaining developmental genes in a transcriptionally poised condition while keeping responsiveness to future activation signals during embryogenesis⁸⁷. The additional presence of H3K4me1 may further indicate enhancer-associated chromatin features occurring in close proximity to the promoter region of *Mafb*.

The complex chromatin landscape observed at this locus is consistent with the established developmental functions of *Mafb*. Previous studies have demonstrated roles for *Mafb* in

cellular differentiation, hindbrain segmentation, cartilage formation, and neural development^{88,89}. *Mafb* has also been shown to interact with members of the ETS transcription factor family, including ETS1⁹⁰, supporting the biological relevance of the ETS1 occupancy detected within this study. Furthermore, *Mafb* expression has been detected in epithelial tissues surrounding the palatal shelves during palatal fusion, suggesting a role in craniofacial morphogenesis. Multiple genetic variants located near the *Mafb* locus have previously been associated with non-syndromic cleft lip and palate, further supporting the relevance of the candidate regulatory region identified here^{81,86}.

In comparison to the *Mafb* locus, the remaining candidate regulatory regions showed more restricted and context-dependent chromatin signatures. The region associated with *Itgb1* displayed a promoter regulatory profile characterized by occupancy of SOX10 and TFAP2A together with a single H3K4me3 peak. IGV inspection confirmed that this signal overlapped an annotated promoter region. The biological relevance of *Itgb1* as a candidate gene associated with cranial neural crest regulatory networks is supported by its established roles in cell adhesion, polarity, and extracellular matrix assembly, processes fundamental to embryogenesis. Integrin family members contribute broadly to tissue morphogenesis, immune function, and wound repair, and changes in their expression are known to affect cell migration. Prior studies have implicated *Itgb1* in cancer cell migration, and its increased expression has been reported in ovarian and breast cancers. The combination of these adhesive and migratory functions with the chromatin occupancy observed here supports *Itgb1* as a biologically plausible candidate gene whose regulatory control during craniofacial morphogenesis requires further investigation.

Similarly, the regulatory region linked to *Rtn3* exhibited a promoter-associated chromatin signature marked by H3K4me3 enrichment and co-occupancy of SOX10, TFAP2A, and ETS1. As with *Itgb1*, IGV visualization confirmed overlap with an annotated promoter region, supporting the interpretation that this locus represents a transcriptionally active or poised promoter element. *Rtn3* encodes a reticulon family protein involved in shaping tubular endoplasmic reticulum and has been implicated in synaptic organization, with altered expression reported in neurodegenerative disease contexts, including Alzheimer's disease. Although *Rtn3* has not been extensively characterized in craniofacial development, its identification within our embryonic neural crest dataset is of interest. The regulatory region identified here may therefore reflect a previously uncharacterized transcriptional input during palatogenesis.

In contrast, the regulatory region associated with *Pard3* showed a different type of regulation. This locus displayed H3K4me1 signature but lacked detectable binding by SOX10, TFAP2A, or ETS1, and showed no evidence of H3K4me3 enrichment. IGV analysis indicated that the region is within an intronic sequence rather than overlapping a promoter region, consistent with classification as a putative enhancer element. *Pard3* is a key regulator of cell polarity, epithelial junction formation, and polarized tissue architecture. Functional studies in zebrafish have shown that introduction of a truncating *Pard3* variant leads to ethmoid plate dysplasia, characterized by partial loss of the median ethmoid plate and failure to form a smooth upper edge, recapitulating features of a cleft-like phenotype. These findings indicate that loss of *Pard3* function disrupts craniofacial development, providing experimental support for the biological relevance of the candidate regulatory region identified at this locus.

Moreover, the identification of candidate regulatory regions associated with several established orofacial cleft genes, including *Irf6*, *Arhgap29*, *Foxe1*, *Pax7*, *Vax1*, and *Ntn1*, further supports the prioritization strategy used in this study. Importantly, while these genes have previously been linked to orofacial cleft susceptibility through GWAS, the analysis performed provides additional regulatory resolution by identifying putative cis-regulatory elements, transcription factor occupancy, and histone modification profiles within their genomic landscapes in cranial neural crest cells. These features are not captured by GWAS alone and represent novel mechanistic insight into their regulation during craniofacial development. These findings suggest that integrating chromatin accessibility, transcription factor occupancy, and histone modification data can not only successfully recover biologically relevant *loci* implicated in craniofacial development but also prioritize previously uncharacterized genes with potential roles in craniofacial development.

5.3 Putative Regulatory Regions Associated with *Nrp1* and *Adora2b*

To explore whether potentially relevant regulatory signatures could also be identified at *loci* not previously associated with orofacial clefts, two additional candidate genes, *Nrp1* and *Adora2b*, were selected for further characterization.

The *Nrp1* locus was associated with two distinct regulatory regions: one overlapping the *Nrp1* promoter and a second classified as a putative enhancer. The promoter-associated region demonstrated occupancy by SOX10, TFAP2A, and ETS1 together with

enrichment for H3K4me1, H3K4me3, and H3K27me3. Similar to the *Mafb* associated region, the coexistence of H3K4me3 and H3K27me3 may reflect a bivalent chromatin configuration, potentially maintaining the *locus* in a poised transcriptional state during neural crest development. Previous studies have demonstrated an essential role for Neuropilin-1 (*Nrp1*) in cranial neural crest migration, stream organization, and cranial nervous system patterning⁹¹. Disruption of *Nrp* mediated signaling results in abnormal neural crest migration and defects in craniofacial development^{91,92}. In addition, *Nrp* signaling contributes to developmental structures whose disruption can result in craniofacial abnormalities. Together, these observations support the biological plausibility of *Nrp1* as a candidate gene associated with cranial neural crest regulatory networks.

In contrast, the developmental role of *Adora2b* in craniofacial morphogenesis remains poorly characterized. *Adora2b* encodes an adenosine receptor highly involved in various tumor types, modulating inflammatory signaling and enhancing cell proliferation and migration⁹³, processes that are of particular relevance during embryonic tissue patterning. These functional parallels are of interest in the context of neural crest biology, as neural crest cells similarly rely on tightly regulated migratory and proliferative programs during craniofacial development. This raises the possibility that signaling pathways such as *Adora2b* mediated responses may constitute evolutionarily conserved mechanisms that are differentially regulated in development and disease. This is in line with hypotheses proposing similarities or even conservation between EMT in neural crest and EMT-like processes in cancer progression^{94,95}. In this context, the identified promoter region for *Adora2b* could represent a potential regulatory element contributing to neural crest cell behaviour. The candidate promoter-associated region identified for *Adora2b* displayed occupancy by TFAP2A and ETS1 together with enrichment for H3K4me1, H3K4me3, and H3K27me3. Although direct evidence linking *Adora2b* to craniofacial development is limited, this gene has been implicated in the regulation of cellular proliferation, migration, and inflammatory signaling pathways. Consequently, the identification of a candidate regulatory region associated with *Adora2b* raises the possibility that this gene may participate in developmental pathways not yet explored in the context of orofacial clefts. However, functional validation will be required to establish any direct role in craniofacial development or disease susceptibility.

5.4 Additional Transcription Factors Potentially Contributing to Neural Crest Regulatory Regions Activity

Unexpectedly, none of the regulatory regions identified in this study displayed detectable FOXD3 occupancy. Given the established role of FOXD3 in neural crest specification, maintenance of multipotency, and regulation of neural crest progenitor fate decisions, this observation was relevant^{17,18,68}. While the absence of FOXD3 binding could reflect a technical limitation of the CUT&RUN dataset, FOXD3 enrichment was indeed detectable within the genome browser tracks despite the comparatively weak signal obtained with the antibody. A more likely explanation may relate to the biological context of the candidate regions themselves. During neural crest development, FOXD3 expression must be downregulated as cells undergo EMT transition and commit to mesenchymal lineages, as continued FOXD3 activity prevents normal mesenchymal differentiation. Since the majority of the candidate regions identified in this study are associated with genes involved in mesenchymal structures and craniofacial development, the absence of FOXD3 occupancy may reflect the transcriptional state of these regions in differentiating neural crest cells, rather than a lack of regulatory activity. These findings suggest that the candidate regulatory regions identified may be regulated through alternative transcriptional mechanisms. To explore this possibility, *de novo* motif enrichment analysis was performed to identify additional transcription factors potentially contributing to enhancer activity. Several biologically relevant candidates emerged, including LEF1, MSX1, RARG, and SMAD4. LEF1 is a downstream effector of the WNT/ β -catenin pathway, that regulates gene expression associated with EMT, a process essential for craniofacial development and neural crest cell migration⁹⁶. Similarly, MSX1 is a critical transcriptional regulator activated during neural crest specification downstream of BMP signaling and is required for palatogenesis and odontogenesis. Both human and mouse studies have linked MSX1 dysfunction to orofacial clefting and tooth agenesis^{97–100}. In addition, retinoic acid (RA) signaling through RARG has a key role in craniofacial morphogenesis, as both excessive and deficient RA levels disrupt palate development and can result in cleft palate formation. Moreover, RA signaling interacts with the TGF- β pathway, highlighting the interconnected nature of craniofacial gene regulatory networks^{101–103}. In line with these findings, SMAD4-mediated TGF- β signaling regulates *Irf6* expression in the palate⁴⁶. Together, these findings support the concept that enhancer and promoter activity at orofacial cleft-associated *loci* is likely governed by complex combinatorial transcription factor networks rather than a limited

set of neural crest regulators. The motif enrichment results therefore provide additional candidate regulatory mechanisms that may contribute to craniofacial gene regulation and warrant future investigation.

Several limitations should be considered when interpreting the findings of this study. Candidate target genes were assigned using a nearest-gene approach, which may not always reflect the true regulatory interactions occurring *in vivo*. Enhancers can regulate genes over considerable genomic distances and may bypass nearby genes to interact with more distal targets. Additionally, this study relied on computational prioritization and chromatin-based annotations without direct experimental validation of enhancer activity. Therefore, the regulatory functions of these candidate regions remain putative. In spite of these limitations, the integration of single-cell chromatin accessibility, transcription factor occupancy, and histone modification datasets provided a robust framework for prioritizing candidate regulatory elements associated with orofacial cleft risk *loci*. Future studies including reporter assays, CRISPR-based perturbation approaches, and additional transcription factor profiling are required to validate the candidate enhancers and promoters identified here and to establish their contribution to craniofacial development and orofacial cleft susceptibility.

6. Conclusion

Understanding how non-coding genetic variation contributes to orofacial cleft (OFC) development remains a major challenge in craniofacial biology. While genome-wide association studies have identified numerous OFC-associated *loci*, the regulatory elements and target genes through which these variants exert their effects are often unknown. In this study, an integrative framework combining GWAS data, single cell multiome datasets, and CUT&RUN chromatin profiling was used to identify and prioritize candidate regulatory elements active in cranial neural crest cells, a key cell population involved in craniofacial development.

By integrating chromatin accessibility, transcription factor occupancy, and histone modification profiles, 38 candidate *cis*-regulatory elements associated with OFC risk *loci* were identified and linked to putative target genes. Detailed characterization of selected regions revealed regulatory signatures consistent with both promoter and enhancer associated activity, demonstrating the advantage of combining multiple epigenomic datasets to refine candidate regulatory regions within non-coding GWAS intervals. Furthermore, *de novo* motif analysis uncovered enrichment of transcription factor binding motifs with established roles in neural crest biology and craniofacial morphogenesis, including LEF1, MSX1, RARG, and SMAD4, suggesting additional regulatory pathways that may contribute to OFC susceptibility.

Importantly, most genes associated with the identified candidate elements have not previously been implicated in OFCs, highlighting the gap that remains between genetic association studies and functional understanding of disease mechanisms. The regulatory elements identified here therefore represent valuable candidates for future experimental validation and may provide new insights into the gene regulatory networks governing craniofacial development.

Overall, this work establishes a reproducible strategy for connecting non-coding disease-associated variants to candidate *cis*-regulatory elements and target genes in a cell type specific context.

Before beginning this work, I often pictured developmental transitions and cell fate decisions as relatively simple on/off events. What I came to appreciate is how many layers of regulation are involved, and how fate decisions are rarely strict or fixed outcomes. Instead, they emerge from the interaction of multiple regulatory mechanisms acting across different levels of organization.

One of the concepts that fascinated me the most throughout this project was cellular plasticity. The fact that cells originating from the same lineage can adopt such diverse fates raises fundamental questions about how identity is established, maintained, and altered. The more I learn about gene regulation, the less development appears as a rigid genetic program and the more it resembles a dynamic and responsive system, shaped by chromatin accessibility, transcription factor activity, regulatory interactions, and cellular context.

Additionally, I believe approaches such as the one presented here may contribute to a larger goal shared by many in the scientific community: building a more comprehensive understanding of gene regulatory networks and the non-coding regulatory elements that shape them. While this study focused on OFCs, similar strategies could help uncover regulatory mechanisms underlying other craniofacial disorders, neurocristopathies, and potentially many other biological conditions. Each newly characterized regulatory element helps understand a part of the genome that has remained largely unexplored, bringing us closer to understanding how complex biological systems are organized.

Perhaps the most important lesson I take from this work is that biological systems are simultaneously robust and fragile. The same plasticity that allows life to generate diversity, adapt, and evolve can also create vulnerabilities when regulatory networks are perturbed. We often think of development as a perfectly orchestrated process, but studying it closely reveals how much of its apparent order emerges from countless interacting components whose relationships we are only beginning to understand.

At the same time, this work raises new questions. Which of the candidate regulatory elements identified here play direct functional roles during craniofacial development? How do they interact with one another within larger gene regulatory networks? Are the regulatory principles observed in cranial neural crest cells unique to this system, or do they represent broader mechanisms shared across other developmental contexts and diseases? More fundamentally, is there a common regulatory logic that governs how cells establish and maintain their identities, or has evolution produced multiple paths capable of reaching similar outcomes?

In that sense, this project has left me with an even greater sense of curiosity. Every newly identified regulatory element, every connection between a regulatory element and its target gene, reveals not only what we know, but also how much remains unexplored.

Supplementary Data

Supplementary table 1-Cell type annotations were assigned based on the expression of known marker genes

Dataset	Cell Type	Marker Gene
Mesp1	Mesenchyme	Tbx1
Mesp1	Mesenchyme	Pitx2
Mesp1	Mesenchyme	Col1a1
Mesp1	Mesenchyme	Pdgfra
Mesp1	Mes.	Twist1
Mesp1	Mes.	Prrx1
Mesp1	Mes.	Alx1
Mesp1	Early M.	Tbx2
Mesp1	Early M.	Tbx3
Mesp1	Early M.	Sox8
Mesp1	Late M.	Lum
Mesp1	Late M.	Col1a1
Mesp1	Late M.	Itga4
Mesp1	Late M.	Col1a2
Mesp1	Late M.	Col3a2
Mesp1	Lateral_Plate_Mesoderm	Prrx1
Mesp1	Lateral_Plate_Mesoderm	Foxf1
Mesp1	Lateral_Plate_Mesoderm	Hand1
Mesp1	Lateral_Plate_Mesoderm	Hand2
Mesp1	Lateral_Plate_Mesoderm	Bmp4
Mesp1	Endothelial	Pecam1
Mesp1	Endothelial	Cdh5
Mesp1	Endothelial	Kdr
Mesp1	Endothelial	Emcn
Mesp1	Hematopoietic	Ptprc
Mesp1	Hematopoietic	Spi1
Mesp1	Hematopoietic	Runx1
Mesp1	Hematopoietic	Itgam
Mesp1	Erythroid	Gata1
Mesp1	Erythroid	Hba-a1
Mesp1	Erythroid	Hbb-y
Mesp1	Erythroid	Alas2
Mesp1	Cardiac	Nkx2-5

Mesp1	Cardiac	Gata4
Mesp1	Cardiac	Tnnt2
Mesp1	Cardiac	Myl7
Mesp1	Cardiac	Isl1
Mesp1	Smooth_Muscle	Acta2
Mesp1	Smooth_Muscle	Tagln
Mesp1	Smooth_Muscle	Myh11
Mesp1	Smooth_Muscle	Cnn1
Mesp1	Paraxial	Tbx6
Mesp1	Paraxial	Msgn1
Mesp1	Paraxial	Tcf15
Mesp1	Paraxial	Meox1
Mesp1	QC	mt-Cytb
Mesp1	QC	mt-Co2
Mesp1	Melanocytes	Mitf
Mesp1	Melanocytes	Dct
Mesp1	Melanocytes	Pmel
Mesp1	Melanocytes	Tyr
Wnt1	Roof_plate	Lmx1a
Wnt1	Roof_plate	Msx1
Wnt1	Roof_plate	Msx2
Wnt1	Roof_plate	Wnt6
Wnt1	Dorsal_NT	Sox2
Wnt1	Dorsal_NT	Sox3
Wnt1	Dorsal_NT	Olig3
Wnt1	Dorsal_NT	En1
Wnt1	Dorsal_NT	Pax5
Wnt1	PreMig_NC	Wnt1
Wnt1	PreMig_NC	Pax3
Wnt1	PreMig_NC	Msx1
Wnt1	PreMig_NC	Zic1
Wnt1	PreMig_NC	Bmp6
Wnt1	Delaminating_NC	Sox10
Wnt1	Delaminating_NC	Foxd3
Wnt1	Delaminating_NC	Tfap2a
Wnt1	Delaminating_NC	Tfap2b
Wnt1	Mig_NC	Crabp1
Wnt1	Mig_NC	Sox10
Wnt1	Mig_NC	Tfap2b

Wnt1	Mig_NC	Dlx5
Wnt1	Mig_NC	Pdgfra
Wnt1	Mig_NC	Hoxa2
Wnt1	Mig_NC	Hoxb2
Wnt1	Mesenchyme	Twist1
Wnt1	Mesenchyme	Prrx1
Wnt1	Mesenchyme	Alx1
Wnt1	Mesenchyme	Dlk1
Wnt1	Early_M	Tbx2
Wnt1	Early_M	Tbx3
Wnt1	Late_M	Col1a1
Wnt1	Late_M	Col3a1
Wnt1	Late_M	Col1a2
Wnt1	Late_M	Lum
Wnt1	CNS_neurons	Tubb3
Wnt1	CNS_neurons	Elavl3
Wnt1	CNS_neurons	Map2
Wnt1	SCP	Plp1
Wnt1	SCP	Fabp7
Wnt1	SCP	Erbp3
Wnt1	SCP	Mpz
Wnt1	Mel	Mitf
Wnt1	Mel	Dct
Wnt1	Mel	Pmel
Wnt1	Surface_ecto	Epcam
Wnt1	Surface_ecto	Krt8
Wnt1	Surface_ecto	Krt18
Wnt1	Surface_ecto	Wnt6
Wnt1	Vascular	Pecam1
Wnt1	Vascular	Pdgfrb
Wnt1	Vascular	Acta2
Wnt1	Proliferation	Mki67
Wnt1	Proliferation	Top2a
Wnt1	QC	mt-Cytb
Wnt1	QC	mt-Co2

Supplementary table 2 -Significantly enriched transcription factor binding motifs identified by HOMER de novo motif analysis of candidate regulatory elements

Homer de novo Motif Results (homer_output//)

[Non-redundant Motif File of Results](#)

[Known Motif Enrichment Results](#)

[Gene Ontology Enrichment Results](#)

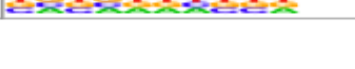
If Homer is having trouble matching a motif to a known motif, try copy/pasting the matrix file into [STAMP](#)

More information on motif finding results: [HOMER | Description of Results | Tips](#)

Total target sequences = 1048

Total background sequences = 96902

* - possible false positive

Rank	Motif	P-value	log P-value	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1		1e-107	-2.484e+02	6.49%	0.07%	435.3bp (147.6bp)	Hoxd13/MA0909.4/Jaspar(0.724) More Information Similar Motifs Found	motif file (matrix)
2		1e-93	-2.147e+02	5.92%	0.07%	89.4bp (146.1bp)	Plagl1/MA1615.2/Jaspar(0.628) More Information Similar Motifs Found	motif file (matrix)
3		1e-89	-2.064e+02	8.21%	0.31%	213.8bp (166.1bp)	Pax8(Paired,Homeobox)/Thyroid-Pax8-ChIP-Seq(GSE26938)/Homer(0.619) More Information Similar Motifs Found	motif file (matrix)
4		1e-84	-1.954e+02	4.29%	0.02%	286.4bp (162.5bp)	NFAT(RHD)/Jurkat-NFATC1-ChIP-Seq(Jolma_et_al)/Homer(0.683) More Information Similar Motifs Found	motif file (matrix)
5		1e-77	-1.777e+02	6.49%	0.19%	227.9bp (152.0bp)	SOX15/MA1152.2/Jaspar(0.663) More Information Similar Motifs Found	motif file (matrix)
6		1e-73	-1.693e+02	2.96%	0.01%	180.8bp (190.4bp)	Rfx5(HTH)/GM12878-Rfx5-ChIP-Seq(GSE31477)/Homer(0.691) More Information Similar Motifs Found	motif file (matrix)
7		1e-73	-1.687e+02	2.58%	0.00%	10.1bp (100.8bp)	SOX9/MA0077.2/Jaspar(0.654) More Information Similar Motifs Found	motif file (matrix)
8		1e-73	-1.683e+02	8.02%	0.45%	172.9bp (159.2bp)	PKNOX1/MA0782.3/Jaspar(0.795) More Information Similar Motifs Found	motif file (matrix)
9		1e-71	-1.639e+02	2.29%	0.00%	659.9bp (0.0bp)	CEBPG/MA1636.2/Jaspar(0.700) More Information Similar Motifs Found	motif file (matrix)
10		1e-69	-1.593e+02	3.34%	0.01%	101.9bp (177.5bp)	FOXA1/MA0148.5/Jaspar(0.672) More Information Similar Motifs Found	motif file (matrix)
11		1e-68	-1.582e+02	12.12%	1.57%	426.9bp (164.4bp)	Hmgal/MA2124.1/Jaspar(0.661) More Information Similar Motifs Found	motif file (matrix)
12		1e-64	-1.483e+02	4.58%	0.09%	27.6bp (175.8bp)	PB0040.1_Lef1_1/Jaspar(0.627) More Information Similar Motifs Found	motif file (matrix)
13		1e-59	-1.377e+02	2.48%	0.01%	31.9bp (105.5bp)	ZNF324/MA1977.2/Jaspar(0.710) More Information Similar Motifs Found	motif file (matrix)
14		1e-59	-1.375e+02	2.96%	0.01%	3.9bp (120.9bp)	bZIP-IRF(bZIP-IRF)/Th17-BatF-ChIP-Seq(GSE39756)/Homer(0.661) More Information Similar Motifs Found	motif file (matrix)
15		1e-58	-1.348e+02	8.30%	0.76%	502.8bp (162.1bp)	ZBED2/MA1971.2/Jaspar(0.588) More Information Similar Motifs Found	motif file (matrix)
16		1e-58	-1.347e+02	4.48%	0.11%	70.2bp (159.4bp)	PB0081.1_Tcf1_1/Jaspar(0.656) More Information Similar Motifs Found	motif file (matrix)
17		1e-58	-1.336e+02	4.29%	0.09%	129.3bp (165.8bp)	MSANTD3/MA1523.2/Jaspar(0.795) More Information Similar Motifs Found	motif file (matrix)
18		1e-57	-1.330e+02	2.48%	0.01%	75.4bp (97.2bp)	MF0005.1_Forkhead_class/Jaspar(0.728) More Information Similar Motifs Found	motif file (matrix)
19		1e-56	-1.310e+02	2.19%	0.00%	117.1bp (96.8bp)	Duxbl(Homeobox)/NIH3T3-Duxbl-HA-ChIP-Seq(GSE119782)/Homer(0.665) More Information Similar Motifs Found	motif file (matrix)
20		1e-54	-1.257e+02	1.81%	0.00%	0.0bp (56.2bp)	PB0115.1_Ehf_2/Jaspar(0.664) More Information Similar Motifs Found	motif file (matrix)

21		1e-54	-1.257e+02	1.81%	0.00%	0.0bp (0.0bp)	PBX2/MA1113.3/Jaspar(0.634) More Information Similar Motifs Found	motif file (matrix)
22		1e-54	-1.243e+02	2.19%	0.00%	10.6bp (194.9bp)	PB0105.1_Arid3a_2/Jaspar(0.614) More Information Similar Motifs Found	motif file (matrix)
23		1e-53	-1.230e+02	5.06%	0.21%	360.8bp (180.3bp)	TFAP2A/MA0810.2/Jaspar(0.597) More Information Similar Motifs Found	motif file (matrix)
24		1e-52	-1.210e+02	15.94%	3.88%	442.3bp (167.6bp)	PBX2(Homeobox)/K562-PBX2-ChIP-Seq(Encode)/Homer(0.793) More Information Similar Motifs Found	motif file (matrix)
25		1e-51	-1.194e+02	1.91%	0.00%	3.7bp (19.1bp)	MYRF(MYRF)/CFPAC1-MYRF-ChIP-Seq(GSE145627)/Homer(0.603) More Information Similar Motifs Found	motif file (matrix)
26		1e-48	-1.126e+02	1.81%	0.00%	0.0bp (148.5bp)	ZNF467(Zf)/HEK293-ZNF467.GFP-ChIP-Seq(GSE58341)/Homer(0.596) More Information Similar Motifs Found	motif file (matrix)
27		1e-48	-1.126e+02	1.81%	0.00%	0.0bp (48.5bp)	SD0003.1_at_AC_acceptor/Jaspar(0.644) More Information Similar Motifs Found	motif file (matrix)
28		1e-48	-1.126e+02	1.81%	0.00%	0.0bp (68.6bp)	Smad4/MA1153.2/Jaspar(0.617) More Information Similar Motifs Found	motif file (matrix)
29		1e-47	-1.099e+02	3.63%	0.08%	31.5bp (137.7bp)	RARG/MA1553.2/Jaspar(0.808) More Information Similar Motifs Found	motif file (matrix)
30		1e-45	-1.044e+02	3.91%	0.13%	48.1bp (150.3bp)	Hnf6b(Homeobox)/LNCaP-Hnf6b-ChIP-Seq(GSE106305)/Homer(0.655) More Information Similar Motifs Found	motif file (matrix)
31		1e-44	-1.028e+02	6.87%	0.72%	303.3bp (163.4bp)	ETV5::DRGX/MA1944.2/Jaspar(0.664) More Information Similar Motifs Found	motif file (matrix)
32		1e-44	-1.025e+02	8.87%	1.35%	498.6bp (163.5bp)	IKZF1/MA1508.2/Jaspar(0.890) More Information Similar Motifs Found	motif file (matrix)
33		1e-40	-9.260e+01	2.00%	0.01%	33.9bp (177.9bp)	POL013.1_MED-1/Jaspar(0.753) More Information Similar Motifs Found	motif file (matrix)
34		1e-39	-9.180e+01	9.35%	1.73%	435.3bp (163.1bp)	POU5F1/MA1115.2/Jaspar(0.682) More Information Similar Motifs Found	motif file (matrix)
35		1e-38	-8.865e+01	19.85%	7.32%	531.0bp (165.2bp)	Lef1/MA0768.3/Jaspar(0.755) More Information Similar Motifs Found	motif file (matrix)
36		1e-37	-8.567e+01	1.43%	0.00%	0.0bp (93.3bp)	GMEB2/MA0862.1/Jaspar(0.658) More Information Similar Motifs Found	motif file (matrix)
37		1e-36	-8.508e+01	2.77%	0.06%	4.0bp (174.3bp)	ZNF768(Zf)/Raji-ZNF768-ChIP-Seq(GSE111879)/Homer(0.612) More Information Similar Motifs Found	motif file (matrix)
38		1e-36	-8.404e+01	1.81%	0.01%	0.0bp (114.3bp)	ZSCAN21/MA2336.1/Jaspar(0.655) More Information Similar Motifs Found	motif file (matrix)
39		1e-36	-8.350e+01	13.07%	3.67%	506.1bp (167.6bp)	PH0044.1_Homez/Jaspar(0.687) More Information Similar Motifs Found	motif file (matrix)
40		1e-29	-6.850e+01	5.82%	0.87%	902.7bp (161.4bp)	PH0106.1_Msx1/Jaspar(0.797) More Information Similar Motifs Found	motif file (matrix)
41		1e-26	-6.017e+01	1.81%	0.03%	0.0bp (190.3bp)	Hoxa10(Homeobox)/ChickenMSG-Hoxa10.Flag-ChIP-Seq(GSE86088)/Homer(0.676) More Information Similar Motifs Found	motif file (matrix)
42		1e-24	-5.668e+01	5.25%	0.87%	902.8bp (172.2bp)	Ahr::Arnt/MA0006.2/Jaspar(0.776) More Information Similar Motifs Found	motif file (matrix)
43		1e-20	-4.719e+01	2.86%	0.26%	82.6bp (151.8bp)	PtflA/MA1618.2/Jaspar(0.741) More Information Similar Motifs Found	motif file (matrix)
44		1e-15	-3.556e+01	15.17%	7.70%	337.8bp (174.6bp)	EWSR1-FLI1/MA0149.1/Jaspar(0.676) More Information Similar Motifs Found	motif file (matrix)

19/05/26, 13:06

homer_output// - Homer de novo Motif Results

45 *		1e-8	-1.975e+01	5.15%	2.08%	109.3bp (163.4bp)	SOX10/MA0442.3/Jaspar(0.858) More Information Similar Motifs Found	motif file (matrix)
46 *		1e0	-9.728e-01	14.12%	13.75%	396.7bp (176.5bp)	ZNF384/MA1125.2/Jaspar(0.974) More Information Similar Motifs Found	motif file (matrix)

References

1. Gilbert, S. F. & Gilbert, S. F. *Developmental Biology*. (Sinauer Associates, 2000).
2. Weisblat, D. A. Embryonic development as a quasi-historical process.
3. Wellmann, J. History of Embryology: Visualizations Through Series and Animation. in *The Palgrave Handbook of the History of Human Sciences* 259–290 (Palgrave Macmillan, Singapore, 2022). doi:10.1007/978-981-16-7255-2_22.
4. Rossant, J. Why study human embryo development? *Developmental Biology* **509**, 43–50 (2024).
5. Rossant J, Tam PPL. Exploring early human embryo development | Science. (2018)
6. Rinkoff, S. & Adlard, R. E. Embryology, Craniofacial Growth, And Development. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2026).
7. Waddington, C. H. *The Strategy Of The Genes*. (1957).
8. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: A Landscape Takes Shape. *Cell* **128**, 635–638 (2007).
9. Mayor, R. & Theveneau, E. The neural crest. *Development* **140**, 2247–2251 (2013).
10. Marta N. Shahbazi. Mechanisms of human embryo development: from cell fate to tissue shape and back *Development* 15 (2020); 147 (14): dev190629. doi: <https://doi.org/10.1242/dev.190629>
11. Dupin, E., Creuzet, S. & Douarin, N. M. L. The Contribution of the Neural Crest to the Vertebrate Body. in *Madame Curie Bioscience Database [Internet]* (Landes Bioscience, 2013).
12. His, Wilhelm, His, W. & Edinburgh, R. C. of P. of. *Untersuchungen Über Die Erste Anlage Des Wirbelthierleibes : Die Erste Entwicklung Des Hühnchens Im Ei*. (F.C.W. Vogel, Leipzig, 1868). doi:10.5962/bhl.title.104964.
13. Sauka-Spengler, T., Bronner-Fraser, M. A gene regulatory network orchestrates neural crest formation *Nat Rev Mol Cell Biol* **9**, 557–568 (2008)
14. Simões-Costa, M. & Bronner, M. E. Establishing neural crest identity: a gene regulatory recipe. *Development* **142**, 242–257 (2015).
15. Khudyakov, J. & Bronner-Fraser, M. Comprehensive spatiotemporal analysis of early chick neural crest network genes. *Dev Dyn* **238**, 716–723 (2009).

16. Nelms, B. L. & Labosky, P. A. Sox Genes. in *Transcriptional Control of Neural Crest Development* (Morgan & Claypool Life Sciences, 2010).
17. Teng, L., Mundell, N. A., Frist, A. Y., Wang, Q. & Labosky, P. A. Requirement for Foxd3 in the maintenance of neural crest progenitors. *Development* **135**, 1615–1624 (2008).
18. Mirella Dottori, Michael K. Gross, Patricia Labosky, Martyn Goulding; The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. *Development* 1 November 2001; 128 (21): 4127–4138. doi: <https://doi.org/10.1242/dev.128.21.4127>
19. Trainor, P. A. & Krumlauf, R. *Hox* genes, neural crest cells and branchial arch patterning. *Current Opinion in Cell Biology* **13**, 698–705 (2001).
20. Stundl, J., Rajan, A. R. D. & Bronner, M. E. Neural crest gene regulatory programs as drivers of development, diversification and disease. *Nat Rev Mol Cell Biol* **27**, 467–485 (2026).
21. Mani, S. A. *et al.* The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell* **133**, 704–715 (2008).
22. Theveneau, E. & Mayor, R. Neural crest delamination and migration: From epithelium-to-mesenchyme transition to collective cell migration. *Developmental Biology* **366**, 34–54 (2012).
23. Bronner, M. E. & LeDouarin, N. M. Development and evolution of the neural crest: an overview. *Dev Biol* **366**, 2–9 (2012).
24. Minoux, M. & Rijli, F. M. Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. *Development* **137**, 2605–2621 (2010).
25. Theveneau, E., & Mayor, R. (2011). Collective cell migration of the cephalic neural crest: the art of integrating information. *Genesis (New York, N.Y. : 2000)*, 49(4), 164–176. <https://doi.org/10.1002/dvg.20700>
26. Carl Gans,R. Glenn Northcutt. Neural Crest and the Origin of Vertebrates: A New Head.*Science***220**,268-273(1983).DOI:[10.1126/science.220.4594.268](https://doi.org/10.1126/science.220.4594.268)
27. Lamanna, F. *et al.* A lamprey neural cell type atlas illuminates the origins of the vertebrate brain. *Nat Ecol Evol* **7**, 1714–1728 (2023).
28. Ruslan Soldatov *et al.* Spatiotemporal structure of cell fate decisions in murine neural crest.*Science***364**,eaas9536(2019).DOI:[10.1126/science.aas9536](https://doi.org/10.1126/science.aas9536)
29. Krispin, S., Nitzan, E., Kassem, Y. & Kalcheim, C. Evidence for a dynamic

spatiotemporal fate map and early fate restrictions of premigratory avian neural crest. *Development* **137**, 585–595 (2010).

30. Hall, B. K. Neurocristopathies. in *The Neural Crest in Development and Evolution* (ed. Hall, B. K.) 152–167 (Springer, New York, NY, 1999). doi:10.1007/978-1-4757-3064-7_11.

31. Vega-Lopez, G. A., Cerrizuela, S., Tribulo, C. & Aybar, M. J. Neurocristopathies: New insights 150 years after the neural crest discovery. *Developmental Biology* **444**, S110–S143 (2018).

32. Paul A. Trainor, K. E. N. W. Neurocristopathies: The Etiology and Pathogenesis of Disorders Arising from Defects in Neural Crest Cell Development. in *Neural Crest Cells* 361–394 (Academic Press, 2014). doi:10.1016/B978-0-12-401730-6.00018-1.

33. Scaffidi, B. K. Celebrating Cleft Lip? Osteological and Artistic Evidence of Lip Deformity in a Trophy Head Individual from Southern Peru. *Ñawpa Pacha* **0**, 1–25 (2025).

34. Hoffman, H. & Hudgins, P. A. Head and skull base features of nine Egyptian mummies: evaluation with high-resolution CT and reformation techniques. *AJR Am J Roentgenol* **178**, 1367–1376 (2002).

35. Saman, M., Gross, J., Ovchinsky, A. & Wood-Smith, D. Cleft Lip and Palate in the Arts: A Critical Reflection. *The Cleft Palate Craniofacial Journal* **49**, 129–136 (2012).

36. Babai, A. & Irving, M. Orofacial Clefts: Genetics of Cleft Lip and Palate. *Genes* **14**, 1603 (2023).

37. Beaty, T. H., Marazita, M. L. & Leslie, E. J. Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. *Fl000Res* **5**, 2800 (2016).

38. Jaruga, A., Ksiazkiewicz, J., Kuzniarz, K. & Tylzanowski, P. Orofacial Cleft and Mandibular Prognathism—Human Genetics and Animal Models. *International Journal of Molecular Sciences* **23**, 953 (2022).

39. Kini, U. Genetics and orofacial clefts: a clinical perspective. *Br Dent J* **234**, 947–952 (2023).

40. Bush, J. O. & Jiang, R. Palatogenesis: morphogenetic and molecular mechanisms of secondary palate development. *Development* **139**, 231–243 (2012).

41. Hammond, N. L. & Dixon, M. J. Revisiting the embryogenesis of lip and palate development. *Oral Dis* **28**, 1306–1326 (2022).

42. Robinson, K., Curtis, S. W. & Leslie, E. J. The heterogeneous genetic architectures of orofacial clefts. *Trends in Genetics* **40**, 410–421 (2024).
43. Ishorst, N., Henschel, L., Thieme, F., Drichel, D., Sivalingam, S., Mehrem, S. L., Fechtner, A. C., Fazaal, J., Welzenbach, J., Heimbach, A., Maj, C., Borisov, O., Hausen, J., Raff, R., Hoischen, A., Dixon, M., Rada-Iglesias, A., Bartusel, M., Rojas-Martinez, A., Aldhorae, K., ... Mangold, E. (2023). Identification of de novo variants in nonsyndromic cleft lip with/without cleft palate patients with low polygenic risk scores. *Molecular genetics & genomic medicine*, *11*(3), e2109. <https://doi.org/10.1002/mgg3.2109>
44. Ingraham, C. R. *et al.* Abnormal skin, limb and craniofacial morphogenesis in mice deficient for interferon regulatory factor 6 (Irf6). *Nat Genet* **38**, 1335–1340 (2006).
45. Kousa, Y. A. & Schutte, B. C. Toward an orofacial gene regulatory network. *Developmental Dynamics* **245**, 220–232 (2016).
46. Iwata, J., Suzuki, A., Pelikan, R. C., Ho, T. V., Sanchez-Lara, P. A., Urata, M., Dixon, M. J., & Chai, Y. (2013). Smad4-Irf6 genetic interaction and TGFβ-mediated IRF6 signaling cascade are crucial for palatal fusion in mice. *Development (Cambridge, England)*, *140*(6), 1220–1230. <https://doi.org/10.1242/dev.089615>
47. Kousa, Y. A., Fuller, E., & Schutte, B. C. (2018). IRF6 and AP2A Interaction Regulates Epidermal Development. *The Journal of investigative dermatology*, *138*(12), 2578–2588. <https://doi.org/10.1016/j.jid.2018.05.030>
48. Zhao, H., Zhang, S., Zhu, H. & Jia, Z. The Evolution of Gene Sequencing Technologies: Unveiling Genetic Architecture of Nonsyndromic Orofacial Clefts. *Genet Res (Camb)* **2026**, 3754674 (2026).
49. Marazita, M. L. The Evolution of Human Genetic Studies of Cleft Lip and Cleft Palate. *Annu Rev Genomics Hum Genet* **13**, 263–283 (2012).
50. Dixon, M. J., Marazita, M. L., Beaty, T. H. & Murray, J. C. Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet* **12**, 167–178 (2011).
51. Reynolds, K. *et al.* Genetics and signaling mechanisms of orofacial clefts. *Birth Defects Res* **112**, 1588–1634 (2020).
52. Townsend, H. A., Rosenberger, K. J., Vanderlinden, L. A., Inamo, J. & Zhang, F. Evaluating methods for integrating single-cell data and genetics to understand inflammatory disease complexity. *Front. Immunol.* **15**, (2024).
53. Sebastiani, P., Timofeev, N., Dworkis, D. A., Perls, T. T. & Steinberg, M. H. Genome-wide association studies and the genetic dissection of complex traits. *Am J*

Hematol **84**, 504–515 (2009).

54. Attanasio, C. *et al.* Fine Tuning of Craniofacial Morphology by Distant-Acting Developmental Enhancers. *Science* **342**, 1241006 (2013).
55. Shen, Y., Yue, F., McCleary, D. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012). <https://doi.org/10.1038/nature11243>
56. Nair, S. J. *et al.* Transcriptional Enhancers at 40: Evolution of a viral DNA element to nuclear architectural structures. *Trends Genet* **38**, 1019–1047 (2022).
57. Schaffner, W. Enhancers, enhancers – from their discovery to today’s universe of transcription enhancers. *Biological Chemistry* **396**, 311–327 (2015).
58. Smith, E. & Shilatifard, A. Enhancer biology and enhanceropathies. *Nat Struct Mol Biol* **21**, 210–219 (2014).
59. Friedman, M. J., Wagner, T., Lee, H., Rosenfeld, M. G. & Oh, S. Enhancer–promoter specificity in gene transcription: molecular mechanisms and disease associations. *Exp Mol Med* **56**, 772–787 (2024).
60. Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*, 6, e21856. <https://doi.org/10.7554/eLife.21856>.
61. Meers, M. P., Bryson, T. D., Henikoff, J. G., & Henikoff, S. (2019). Improved CUT&RUN chromatin profiling tools. *eLife*, 8, e46314. <https://doi.org/10.7554/eLife.46314>
62. Buenrostro, J., Wu, B., Chang, H. & Greenleaf, W. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **109**, 21.29.1 (2015).
63. Qiu, C. *et al.* A single-cell time-lapse of mouse prenatal development from gastrula to birth. *Nature* **626**, 1084–1093 (2024).
64. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* **107**, 21931–21936 (2010).
65. Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*, 39(3), 311–318. <https://doi.org/10.1038/ng1966>
66. Rothstein, M. & Simoes-Costa, M. Heterodimerization of TFAP2 pioneer factors

drives epigenomic remodeling during neural crest specification. *Genome Res* **30**, 35–48 (2020).

67. Nelms, B. L. & Labosky, P. A. ETS Genes. in *Transcriptional Control of Neural Crest Development* (Morgan & Claypool Life Sciences, 2010).

68. Mundell, N. A. & Labosky, P. A. Neural crest stem cell multipotency requires Foxd3 to maintain neural potential and repress mesenchymal fates. *Development* **138**, 641–652 (2011).

69. Fa, W., P, A. & Fj, T. SCANPY: large-scale single-cell gene expression data analysis. *Genome biology* **19**, (2018).

70. Van de Sande, B., Flerin, C., Davie, K. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15**, 2247–2276 (2020). <https://doi.org/10.1038/s41596-020-0336-2>

71. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat Biotechnol* **35**, 316–319 (2017).

72. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38**, 276–278 (2020).

73. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

74. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

75. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

76. Amemiya, H.M., Kundaje, A. & Boyle, A.P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354 (2019). <https://doi.org/10.1038/s41598-019-45839-z>

77. Aaron R. Quinlan, Ira M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics*, Volume 26, Issue 6, (2010), Pages 841–842, <https://doi.org/10.1093/bioinformatics/btq033>

78. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).

79. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

80. Meers, M. P., Tenenbaum, D., & Henikoff, S. (2019). Peak calling by Sparse

Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics & chromatin*, 12(1), 42. <https://doi.org/10.1186/s13072-019-0287-4>

81. Leslie, E. J. *et al.* Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Hum Genet* **136**, 275–286 (2017).
82. Dack, K. *et al.* Genetic heterogeneity and homogeneity among orofacial cleft subtypes: genome-wide association studies in the cleft collective. *Hum Mol Genet* **34**, 1934–1950 (2025).
83. Zhou, G.-L. *et al.* E-GWAS: an ensemble-like GWAS strategy that provides effective control over false positive rates without decreasing true positives. *Genet Sel Evol* **55**, 46 (2023).
84. Zang, C., Schones, D. E., Zeng, C., Cui, K., Zhao, K., & Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics (Oxford, England)*, 25(15), 1952–1958. <https://doi.org/10.1093/bioinformatics/btp340>
85. Pepke, S., Wold, B., & Mortazavi, A. (2009). Computation for ChIP-seq and RNA-seq studies. *Nature methods*, 6(11 Suppl), S22–S32. <https://doi.org/10.1038/nmeth.1371>
86. Rahimov, F., Jugessur, A. & Murray, J. C. Genetics of Nonsyndromic Orofacial Clefts. *Cleft Palate Craniofac J* **49**, 73–91 (2012).
87. Kumar, D., Cinghu, S., Oldfield, A. J., Yang, P., & Jothi, R. (2021). Decoding the function of bivalent chromatin in development and cancer. *Genome research*, 31(12), 2170–2184. <https://doi.org/10.1101/gr.275736.121>
88. Huang, K., Serria, M. S., Nakabayashi, H., Nishi, S. & Sakai, M. Molecular cloning and functional characterization of the mouse *mafB* gene. *Gene* **242**, 419–426 (2000).
89. Zhang, B.-H., Sun, J.-L., Zhang, S.-D., Shi, B. & Jia, Z.-L. Targeted Resequencing Identifies Novel MAFB Variants Associated With Nonsyndromic Cleft Lip With or Without Cleft Palate. *The Cleft Palate Craniofacial Journal* **63**, 1407–1413 (2026).
90. Sieweke, M. H., Tekotte, H., Frampton, J., & Graf, T. (1996). MafB is an interaction partner and repressor of Ets-1 that inhibits erythroid differentiation. *Cell*, 85(1), 49–60. [https://doi.org/10.1016/s0092-8674\(00\)81081-8](https://doi.org/10.1016/s0092-8674(00)81081-8)
91. Schwarz, Q., Vieira, J. M., Howard, B., Eickholt, B. J. & Ruhrberg, C. Neuropilin 1 and 2 control cranial gangliogenesis and axon guidance through neural crest cells.

Development **135**, 1605–1613 (2008).

92. Gammill, L. S., Gonzalez, C. & Bronner-Fraser, M. Neuropilin 2/semaphorin 3F signaling is essential for cranial neural crest migration and trigeminal ganglion condensation. *Developmental Neurobiology* **67**, 47–56 (2007).
93. Strickland, L. N. *et al.* The resurgence of the Adora2b receptor as an immunotherapeutic target in pancreatic cancer. *Front. Immunol.* **14**, (2023).
94. Delloye-Bourgeois, C. & Castellani, V. Hijacking of Embryonic Programs by Neural Crest-Derived Neuroblastoma: From Physiological Migration to Metastatic Dissemination. *Front. Mol. Neurosci.* **12**, (2019).
95. Kulesa, P. M., Morrison, J. A. & Bailey, C. M. The Neural Crest and Cancer: A Developmental Spin on Melanoma. *Cells Tissues Organs* **198**, 12–21 (2013).
96. Blazquez, R., Rietkötter, E., Wenske, B., Wlochowicz, D., Sparrer, D., Vollmer, E., Müller, G., Seegerer, J., Sun, X., Dettmer, K., Barrantes-Freer, A., Stange, L., Utpatel, K., Bleckmann, A., Treiber, H., Bohnenberger, H., Lenz, C., Schulz, M., Reimelt, C., Hackl, C., Pukrop, T. (2020). LEF1 supports metastatic brain colonization by regulating glutathione metabolism and increasing ROS resistance in breast cancer. *International journal of cancer*, *146*(11), 3170–3183. <https://doi.org/10.1002/ijc.32742>
97. Liang, J., Von den Hoff, J., Lange, J., Ren, Y., Bian, Z., & Carels, C. E. (2016). MSX1 mutations and associated disease phenotypes: genotype-phenotype relations. *European journal of human genetics : EJHG*, *24*(12), 1663–1670. <https://doi.org/10.1038/ejhg.2016.78>
98. Vastardis, H., Karimbox, N., Guthua, S. W., Seidman, J. G., & Seidman, C. E. (1996). A human MSX1 homeodomain missense mutation causes selective tooth agenesis. *Nature genetics*, *13*(4), 417–421. <https://doi.org/10.1038/ng0896-417>
99. Jumlongras, D., Bei, M., Stimson, J. M., Wang, W. F., DePalma, S. R., Seidman, C. E., Felbor, U., Maas, R., Seidman, J. G., & Olsen, B. R. (2001). A nonsense mutation in MSX1 causes Witkop syndrome. *American journal of human genetics*, *69*(1), 67–74. <https://doi.org/10.1086/321271>
100. Satokata, I., & Maas, R. (1994). Msx1 deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nature genetics*, *6*(4), 348–356. <https://doi.org/10.1038/ng0494-348>
101. Baroni, T. *et al.* Retinoic Acid, GABA-ergic, and TGF- β Signaling Systems Are Involved in Human Cleft Palate Fibroblast Phenotype. *Mol Med* **12**, 237–245 (2006).
102. Lohnes, D. *et al.* Function of the retinoic acid receptors (RARs) during

development: (I) Craniofacial and skeletal abnormalities in RAR double mutants. *Development* **120**, 2723–2748 (1994).

103. Abbott, B. D., Harris, M. W., & Birnbaum, L. S. (1989). Etiology of retinoic acid-induced cleft palate varies with the embryonic stage. *Teratology*, *40*(6), 533–553. <https://doi.org/10.1002/tera.1420400602>.