

**DISRUPTION OF NEURAL CREST ENHANCER  
LANDSCAPES AS AN ETIOLOGICAL MECHANISM FOR  
HUMAN NEUROCRISTOPATHIES**

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Michaela Bartusel

aus Koblenz

Köln

2019

Berichtersteller/in: Prof. Dr. Brunhilde Wirth

Prof. Dr. Mirka Uhlířova

Tag der letzten mündlichen Prüfung: 06.12.2019

# TABLE OF CONTENTS

<b>Table of Contents</b> .....	<b>3</b>
<b>Kurzzusammenfassung</b> .....	<b>7</b>
<b>Abstract</b> .....	<b>9</b>
<b>1 Introduction</b> .....	<b>11</b>
1.1 Transcriptional regulation of the genome .....	11
1.1.1 Enhancers as key cis-regulatory elements .....	12
1.1.2 Enhancers in human disease .....	13
1.2 Neural Crest – the Fourth Germ Layer .....	15
1.2.1 Neural crest development – a multistep process .....	15
1.2.2 Cranial NCC and their regulation during formation of the vertebrate head .....	16
1.2.3 Models to study NCC .....	17
1.3 Orofacial Clefts .....	19
1.4 Genome Wide Association Studies (GWAS) .....	20
1.5 Novel candidate genes for non-syndromic OFC .....	23
1.6 Branchiooculofacial Syndrome (BOFS) and TFAP2A .....	26
1.7 Aims .....	29
<b>2 Material &amp; Methods</b> .....	<b>30</b>
2.1 Equipment .....	30
2.2 Chemicals and Reagents .....	31
2.3 Tissue Culture Procedures .....	34
2.3.1 Cell lines .....	35
2.3.2 Tissue culture media and reagents .....	35
2.3.3 hiPSC culture .....	37
2.3.4 Derivation of patient-specific hiPSC .....	37
2.3.5 Transfection of hiPSC .....	38
2.3.6 Generation of clonal hiPSC lines .....	38
2.3.7 hNCC differentiation and short term maintenance .....	39
2.3.8 Differentiation of NCC derived lineages .....	39
2.3.9 Alcian Blue staining .....	40
2.3.10 Scratch Assay .....	40
2.3.11 CFSE Proliferation Assays .....	41

## Table of Contents

2.3.12 Cell proliferation growth curves.....	42
2.4 Clinical and genetic characterization of the BOFS patient inversion .....	42
2.4.1 Patient description .....	42
2.4.2 Characterization of chromosomal abnormalities.....	43
2.4.3 Targeted Locus Amplification (TLA).....	44
2.5 General molecular biology methods.....	46
2.5.1 Genomic DNA extraction .....	46
2.5.2 PCR for genotyping.....	47
2.5.3 RNA extraction .....	48
2.5.4 cDNA synthesis .....	48
2.5.5 RT-qPCR.....	48
2.5.6 Allele-specific analysis of TFAP2A expression .....	49
2.5.7 Chemical Competent Cells.....	49
2.5.8 Molecular cloning.....	50
2.6 Enhancer reporter assays.....	51
2.6.1 Enhancer cloning .....	51
2.6.2 In vitro reporter assays.....	52
2.6.3 In vivo reporter assays .....	53
2.7 Immunological Methods .....	53
2.7.1 Chromatin Immunoprecipitation (ChIP).....	53
2.7.2 Immunofluorescence (IF) assays in cultured cells .....	55
2.7.3 Flow Cytometry .....	56
2.7.4 Western Blot .....	57
2.8 Genetic engineering by Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR/Cas9) .....	59
2.8.1 Design of guide RNAs (gRNA) .....	59
2.8.2 Cloning of gRNA into the Cas9-Vector.....	60
2.8.3 Delivery of pX330A-Cas9-gRNA vector into target cells .....	60
2.8.4 Deletion detection by PCR.....	61
2.8.5 Rescue experiments .....	63
2.8.6 pGEM-T cloning .....	64
2.9 Chicken Embryo Procedures .....	64
2.9.1 Chicken strain .....	64
2.9.2 Incubation and determination of developmental stage .....	64
2.9.3 Isolation of whole embryos, facial prominences, or neural tube sections.....	64
2.9.4 Electroporation.....	65
2.9.5 Whole Mount in situ Hybridization (WISH) .....	65

2.9.6	Overexpression of human FAM49A .....	67
2.10	Genomic methods.....	67
2.10.1	Circular Chromatin Conformation Capture (4C).....	67
2.10.2	ChIP-seq.....	70
2.10.3	ATAC-seq .....	70
2.10.4	RNA-seq.....	71
2.10.5	Single cell RNA-seq (scRNA-seq).....	72
2.11	Computational Analyses of next generation sequencing data .....	72
2.11.1	4C-seq.....	72
2.11.2	ChIP-seq.....	73
2.11.3	ATAC-seq .....	74
2.11.4	RNA-seq.....	74
2.11.5	scRNA-seq.....	75
2.11.6	Classification of GWAS SNPs.....	75
2.11.7	Data Availability.....	76
2.11.8	Public genomic datasets used .....	76
2.11.9	Software and Algorithms.....	77
<b>3</b>	<b>Results .....</b>	<b>79</b>
3.1	Disruption of the TFAP2A regulatory domain causes branchiooculo-facial syndrome (BOFS) and illuminates pathomechanisms for other human neurocristopathies .....	79
3.1.1	A unique BOFS patient with a de novo heterozygous inversion in chr6 that does not disrupt the TFAP2A coding sequence.....	80
3.1.2	Use of a robust in vitro differentiation system to model the pathological mechanisms of human neurocristopathies .....	82
3.1.3	Epigenomic and topological characterization of the TFAP2A locus regulatory landscape in hNCC.....	85
3.1.4	The TFAP2A-TAD contains regulatory elements that are essential for TFAP2A expression in hNCC.....	90
3.1.5	The 89 Mb heterozygous inversion leads to monoallelic and reduced TFAP2A levels in patient hNCC .	98
3.1.6	Monoallelic TFAP2A expression in patient hNCC is caused by a loss of physical interactions between the inverted TFAP2A allele and its cognate hNCC enhancers .....	107
3.1.7	The patient inversion does not lead to enhancer adoption or ectopic gains in gene expression in hNCC.....	112
3.1.8	BOFS patient hNCC display a defective transcriptome and reduced migratory capacity .....	116
3.1.9	Gene downregulation in BOFS patient hNCC is directly mediated by reduced TFAP2A binding to active hNCC enhancers .....	120

## Table of Contents

3.2	Molecular and functional characterization of a locus contributing to non-syndromic orofacial cleft identifies DDX1 as a novel regulator of facial development .....	127
3.2.1	Comparison between OFC risk-loci and hNCC enhancers as a strategy to identify novel genes implicated in facial development .....	127
3.2.2	Linking the candidate enhancer to its putative target genes based on 3D chromatin architecture .....	132
3.2.3	Functional characterization of the 2p24.2 enhancer element in chicken embryos and human neural crest cells confirms its regulatory activity .....	133
3.2.4	DDX1 and MYCN, but not FAM49A, are expressed in the developing facial tissues involved in OFC ...	138
3.2.5	In vivo loss of function assays reveal an essential role for DDX1 in facial development .....	140
3.2.6	The cellular phenotype of DDX1 knockout cells suggests a proliferation defect .....	146
3.2.7	DDX1 <sup>+/-</sup> hNCC display a defective transcriptome that suggests a role for DDX1 in cell cycle control ...	156
3.2.8	DDX1 binds to promoters and genes bodies of expressed genes suggesting involvement in a transcription coupled repair mechanism .....	157
<b>4</b>	<b>Discussion and future directions .....</b>	<b>161</b>
4.1	Insights into the pathomechanism of a syndromic neurocristopathy with implications for non-syndromic forms .....	161
4.2	Research models to investigate human neurocristopathies .....	164
4.3	Enhancer-gene specificity requires more than physical proximity .....	166
4.4	Beyond GWAS: possibilities and limitations .....	168
4.5	Taking environmental factors into account .....	171
4.6	Enhancer redundancy as a widespread mechanism conferring transcriptional robustness .....	172
4.7	DDX1 as a novel regulator of facial development and a new candidate gene for molecular genetic testing in patients with OFC .....	174
4.8	Potential function of DDX1 during NC development .....	176
4.9	Conclusion and significance .....	178
	<b>References .....</b>	<b>179</b>
	<b>Figure Index .....</b>	<b>198</b>
	<b>Table index .....</b>	<b>201</b>
	<b>List of Abbreviations .....</b>	<b>202</b>
	<b>Acknowledgements .....</b>	<b>204</b>
	<b>Eidesstattliche Erklärung .....</b>	<b>205</b>
	<b>Lebenslauf .....</b>	<b>206</b>

## KURZZUSAMMENFASSUNG

Die embryonale Entwicklung der menschlichen Gesichtszüge ist ein hochkomplexer Mechanismus, der eine sehr genaue räumliche und zeitliche Regulierung der Genexpression während der Entwicklung der Neuralleiste (NL) erfordert. Neuralleistenzellen (NLZ) sind ein transienter embryonaler Zelltyp mit breitem Differenzierungspotential, der zur Bildung und Morphogenese mehrerer Gewebe, einschließlich großer Teile des Gesichts, beiträgt. Wie jeder andere Zelltyp besitzen NLZ einen charakteristischen Satz von Enhancern, die durch Steuerung der Expression spezifischer Gene die zelluläre Identität definieren. Eine Beeinträchtigung dieser Regulation kann zu Missbildungen, z. B. einer Lippen-Kiefer-Gaumen-Spalte (LKGS), führen, die häufig als Neurokristopathien bezeichnet werden und sowohl für die Betroffenen als auch für die Gesellschaft eine schwere Belastung darstellen. Das zentrale Ziel dieser Arbeit ist es, zu verstehen, wie genetische oder strukturelle Störungen der Enhancer-Aktivität während der NL-Entwicklung zu humanen Neurokristopathien führen können. Das gewonnene Wissen soll langfristig dazu dienen, eine frühzeitige Erkennung zu gewährleisten und mögliche therapeutische Ansätze aufzuzeigen.

Im Rahmen dieser Arbeit haben wir den Pathomechanismus sowohl einer syndromischen (hier: Branchiookulofaziales Syndrom (BOFS)) als auch einer nicht-syndromischen (hier: LKGS) Neurokristopathie untersucht, indem wir *in vitro* und *in vivo* Entwicklungsmodelle mit gentechnischen Ansätzen und mehreren genomischen Methoden kombinierten.

Zunächst beschreiben wir einen Patienten mit BOFS, der im Gegensatz zu bekannten Fällen keine Mutationen innerhalb von *TFAP2A*, einem NL-Master-Regulator, aufweist. Stattdessen trägt der Patient eine *de novo* heterozygote Inversion, bei der eine der Bruchstellen 40 kb von *TFAP2A* entfernt liegt. Unter Verwendung patientenspezifischer humaninduzierter pluripotenter Stammzellen (hiPSC) und eines *in vitro* Differenzierungssystems in NLZ konnten wir zeigen, dass die Inversion zu einem Verlust der physischen Interaktion zwischen dem invertierten *TFAP2A*-Allel und seinen zugehörigen Enhancern führt. Dies wiederum führt zu monoallelischer und unzureichender *TFAP2A*-Expression in menschlichen NLZ. Insgesamt bietet dieser erste Teil einen aussagekräftigen Ansatz, um die pathologischen Mechanismen von Strukturvarianten zu untersuchen, die die 3D-Genomorganisation stören können, die aber aus verschiedenen Gründen (z. B. eingeschränkter Zugang zu relevantem Patientenmaterial, Unterschiede in der Gendosisempfindlichkeit zwischen Mäusen und Menschen, Schwierigkeiten bei der Rekapitulation bestimmter Strukturvarianten) *in vivo* nicht richtig bewertet werden können.

Zweitens haben wir zuvor generierte NLZ-Enhancer-Karten mit LKGS-Risikostellen kombiniert, die durch genomweite Assoziationsstudien (GWAS) identifiziert wurden. Dadurch konnten wir einen hochkonservierten Enhancer (Enh2p24.2) als potenziellen Kandidaten für genetische Varianten ausmachen, die zur Entwicklung von LKGS beitragen. GWAS verbinden genetische Polymorphismen mit quantitativen Merkmalen und komplexen Störungen. Die meisten krankheitsassoziierten Polymorphismen treten jedoch in nicht-kodierenden Regionen des menschlichen Genoms auf, weshalb die ursächliche Relevanz dieser genetischen Varianten nicht einfach mit einem Gen in Verbindung gebracht werden kann. Interessanterweise treten Polymorphismen, die mit einer LKGS assoziiert sind, überdurchschnittlich häufig in NLZ-Enhancern auf. Daher nehmen wir an, dass diese Polymorphismen zur Ursache der Störung beitragen, indem sie NLZ-Enhancer und folglich die Expression relevanter Gene verändern. Unter Verwendung von Enh2p24.2 als Ausgangspunkt für 4C-seq Experimente konnten wir zwei Gene, *MYCN* und *DDX1*, als potenzielle Ziele des Enhancers identifizieren. Mithilfe von *in vitro* und *in vivo* NLZ-Entwicklungsmodellen konnten wir außerdem zeigen, dass beide Gene für eine normale Gesichtsentwicklung unerlässlich sind. Während *MYCN* kein überraschender Kandidat für die Entwicklung von LKGS war, könnte die Identifizierung von *DDX1* als neuer Regulator der Gesichtsentwicklung neue Einblicke in die molekularen Prozesse (z. B. die transkriptionsgekoppelte DNS Reparatur) liefern, die mit LKGS und möglicherweise anderen humanen Neurokristopathien (z. B. Neuroblastom) zusammenhängen.

## ABSTRACT

The embryonic development of the human facial features is a highly complex mechanism which requires very exact spatial and temporal regulation of gene expression during neural crest (NC) development. NC cells (NCC) are a transient embryonic cell type with wide differentiation potential that contributes to the formation and morphogenesis of multiple tissues and organs, including many parts of the face. Just like any other cell type, NCC possess a characteristic set of enhancers that, by controlling the expression of specific genes, define cellular identity. Impairment of this regulation can lead to craniofacial malformations, such as orofacial cleft (OFC), which are frequently referred to as neurocristopathies and that represent a heavy burden on both the affected individuals and society. Understanding how genetic or structural disruption of enhancer activity during NC development can lead to human neurocristopathies is the central goal of this work. In the long term, the gained knowledge should serve to enable early detection and show potential therapeutic approaches.

Here we investigate the pathomechanism of both syndromic (i.e. Branchiooculofacial Syndrome (BOFS)) and non-syndromic (i.e. OFC) neurocristopathies, by combining *in vitro* and *in vivo* NC developmental models with genetic engineering approaches and multiple genomic methods.

First, we describe a unique patient with BOFS, who, in contrast to previously reported cases, does not present a heterozygous mutations within *TFAP2A*, a NC master regulator. Instead, the patient carries a *de novo* heterozygous 89 Mb inversion in which one of the breakpoints is located 40 kb downstream of *TFAP2A*. We first showed that this inversion separates *TFAP2A* from enhancers that are located within the same large topologically associating domain (TAD) and that are essential for *TFAP2A* expression in NCC. Importantly, using patient-specific human induced pluripotent stem cells (hiPSC) and a robust *in vitro* differentiation system towards NCC, we then showed that the inversion causes a loss of physical interactions between the inverted *TFAP2A* allele and its cognate enhancers, leading to *TFAP2A* monoallelic and haploinsufficient expression in human NCC. Overall, this first part provides a powerful approach to investigate the pathological mechanisms of structural variants predicted to disrupt 3D genome organization of gene regulatory landscapes and that, due to various reasons (i.e. limited access to relevant patient material, differences in gene dosage sensitivity between mice and humans, difficulties in recapitulating certain structural variants), cannot be properly evaluated *in vivo*.

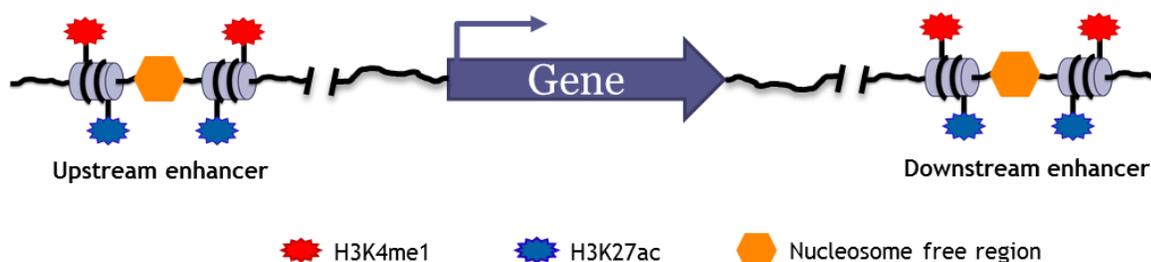
Second, we combined previously generated hNCC enhancer maps with OFC risk-loci identified through genome-wide association studies (GWAS) and, as a result, we revealed a highly conserved enhancer (i.e. Enh2p24.2) as a potential candidate harboring genetic variants involved in OFC. GWAS link common single nucleotide polymorphisms (SNPs) with quantitative traits and complex disorders. However, most disease-associated SNPs occur in non-coding regions of the human genome and consequently, the etiological relevance of these genetic variants cannot be easily connected to a gene. Nevertheless, accumulating evidences suggest that these disease-associated SNPs may contribute to human disease susceptibility by altering enhancers. Interestingly, SNPs associated with OFC are overrepresented in NCC enhancers. Therefore, we hypothesize that SNPs associated with OFC contribute to the etiology of the disorder by altering NCC enhancers and, consequently, the expression of relevant genes. Using Enh2p24.2 as a bait in circularized chromosome conformation capture sequencing (4C-seq) experiments, we identified two distally located genes, *MYCN* and *DDX1*, as its potential targets. Using *in vitro* and *in vivo* NCC developmental models, we then demonstrated that both genes are essential for normal facial development. While *MYCN* was not a surprising candidate to be involved in the etiology of OFC, the identification of *DDX1* as a novel regulator of facial development might provide new insights into the molecular processes (e.g. transcription-coupled DNA repair) implicated in OFC and, potentially, other human neurocristopathies (e.g. neuroblastoma).

# 1 INTRODUCTION

## 1.1 Transcriptional regulation of the genome

Transcriptional regulation happens at several genomic levels: Firstly, distal *cis*-regulatory elements are bound by transcription factors (TFs) and co-factors in order to open the underlying chromatin (Calo and Wysocka, 2013; Roeder, 2005; Spitz and Furlong, 2012). Secondly, these distal *cis*-regulatory elements need to interact with the promoters of their target genes in order to induce their expression, which involves the formation of chromatin loops. In addition, complex 3D organization of the genome leads to the formation of large regulatory domains, which ensure that interactions between distal *cis*-regulatory elements and genes within the nucleus occur in a specific manner (Dixon et al., 2012; Krijger and de Laat, 2016). Each of the regulatory layers is described in more detail below.

Recent advances in transcriptomic, epigenetic and chromosome conformation capture (3C)-based methods have dramatically improved the functional annotation of the human genome (Dixon et al., 2015; Dunham et al., 2012; Roadmap Epigenomics et al., 2015; Siggens and Ekwall, 2014). This led to the appreciation that a significant fraction of the non-coding human genome is filled with *cis*-regulatory elements that are essential for the establishment and maintenance of cell type-specific gene expression programs (Dunham et al., 2012; Heintzman et al., 2009; Rada-Iglesias et al., 2011; Roadmap Epigenomics et al., 2015). Furthermore, different classes of *cis*-regulatory elements (enhancers, promoters, silencers, insulators) can be identified based on specific combinations of epigenetic modifications, including histone acetylation, histone methylation or binding of certain TFs and co-factors (Calo and Wysocka, 2013; Chen et al., 2016; Rada-Iglesias et al., 2012).



**Figure 1.1.1 Simplified schematic illustration of enhancers and their specific epigenetic signatures.** Chromatin marks characterizing active enhancers are shown in red (histone-3-lysine-4-monomethylation, H3K4me1), blue (histone-3-lysine-27-acetylation, H3K27ac) and orange (nucleosome free region). Their distal position in relation to the gene they regulate is displayed by spacers in the DNA (vertical black lines).

### 1.1.1 Enhancers as key *cis*-regulatory elements

Here we focus on the role of enhancers, which are distal *cis*-regulatory sequences that contain clusters of TF binding sites and regulate the transcription of a gene by physical interaction with its promoter over long distances (even more than 1 Mb) and in an orientation-independent manner (de Laat and Duboule, 2013; Spitz and Furlong, 2012; Yanez-Cuna et al., 2013). Enhancers are characterized by displaying an open, nucleosome free chromatin that is accessible to TFs and co-factors (Figure 1.1.1). Moreover, based on the presence of specific epigenetic modifications, such as histone acetylation or methylation, enhancers can be classified into poised, primed and active (Calo and Wysocka, 2013). The combination of these different epigenetic features can be used to predict the location of enhancers within the genome as well as their regulatory state (Shlyueva et al., 2014). The repertoire of active enhancers generally varies across tissues and developmental stages and, thus, enhancers are largely cell-type specific (Chen et al., 2016; Rada-Iglesias et al., 2012; Rada-Iglesias et al., 2011). This is in full agreement with the major role of enhancers in controlling the establishment of cell-type and developmental-stage specific gene expression programs.

Furthermore, there is also evidence that many genes, especially those with major cell identity regulatory functions, display complex regulatory landscapes in which multiple enhancers robustly control the expression of a single gene (Hnisz et al., 2013).

The elucidation of the 3D architecture of nuclear chromatin revealed that enhancers physically interact with the promoter of their target genes in order to regulate their expression (Cruz-Molina et al., 2017; Deng et al., 2012; Kolovos et al., 2016; Ren et al., 2017; Tolhuis et al., 2002; Weintraub et al., 2017). Thereby, the linearly closest gene to a given enhancer can but does not always have to be its main target (Lettice et al., 2003; Sanyal et al., 2012; Smemo et al., 2014). Moreover, enhancer-gene contacts mainly occur within large genomic regions called topologically associating domains (TADs). TADs represent regulatory units that not only facilitate the communication between enhancers and their target genes but also provide insulation by preventing the establishment of unwanted enhancer-gene interactions (Dixon et al., 2012; Downen et al., 2014; Lupianez et al., 2015; Nora et al., 2012; Schwarzer et al., 2017). TADs are believed to be formed through a loop extrusion mechanism, whereby cohesin-dependent loops become anchored at convergent CTCF sites. This results in the formation of TAD boundaries that insulate genes within a TAD from interacting with enhancers located in neighboring TADs. Consequently, smaller and dynamic loops connecting enhancers with their target genes frequently occur within TADs (Dixon et al., 2012; Fudenberg et al., 2016; Haarhuis et al., 2017; Ji et al., 2016; Nuebler

et al., 2018; Pombo and Dillon, 2015; Ren et al., 2017; Schwarzer et al., 2017). Although the physical communication between an enhancer and a gene is not necessarily functional in terms of gene expression, it provides a strong indication towards linking enhancers with their most likely target genes.

### 1.1.2 Enhancers in human disease

An increasing number of studies revealed that the disruption of enhancer activity or of the communication between genes and enhancers can lead to human disease (Smith and Shilatifard, 2014; Spielmann et al., 2018). Additionally, most of the genetic variants associated with complex human disorders overlap enhancers and, in some cases, such variants can alter enhancer activity and lead to quantitative changes in gene expression (Bauer et al., 2013; Maurano et al., 2012; Smemo et al., 2014; Soldner et al., 2016).

Moreover, disruption of TAD organization by structural variants (SVs) (i.e. deletions, duplications, inversions, translocations) alters the physical interaction between genes and enhancers and ultimately leads to human disease by causing gains and/or losses in gene expression (i.e. gain-of-function SV and loss-of-function SV, respectively) (Matharu and Ahituv, 2015; Smith and Shilatifard, 2014; Spielmann et al., 2018; Weischenfeldt et al., 2013). Depending on the type of SV, these changes in gene expression can occur through several pathomechanisms. On one hand, duplications and deletion extending beyond TAD boundaries can lead to the formation of novel TADs (i.e. Neo-TADs) or the fusion of adjacent TADs, respectively. In both cases, this can lead to the occurrence of ectopic enhancer-gene interactions (i.e. enhancer adoption or hijacking) and pathological increases in gene expression (Franke et al., 2016; Groschel et al., 2014; Hnisz et al., 2016; Lupianez et al., 2015; Northcott et al., 2014; Peifer et al., 2015; Vicente-Garcia et al., 2017). On the other hand, translocations and inversions that span TAD boundaries can lead to shuffling of TADs that are usually spatially separated, which can then lead to gains and/or losses in gene expression. In this case, pathological gains in gene expression are also caused by enhancer adoption or hijacking, while gene silencing and haploinsufficiency might occur through position effect variegation (PEV) and/or loss of endogenous enhancer-gene interactions (i.e. enhancer disconnection) (Birnbaum et al., 2012a; Birnbaum et al., 2012b; Goubau et al., 2013; Kleinjan et al., 2001; Kleinjan and van Heyningen, 2005; Lettice et al., 2002; Mehrjouy et al., 2018; Redin et al., 2017; Roessler et al., 1997; Zepeda-Mendoza et al., 2018).

On the basis of the preceding mechanisms caused by SVs and taking advantage of the fact that TADs are largely invariant across different cell types and species (Ghavi-Helm et al., 2014), recently, computational approaches to calculate the pathological consequences of human SVs have been developed (Bianco et al., 2018; Dixon et al., 2012; Peric-Hupkes et al., 2010; Weischenfeldt et al., 2017). However, enhancer activity is cell-type and developmental-stage specific, while the 3D organization of the genome, including TADs, might also be more variable than previously thought (Bonev et al., 2017). Furthermore, our understanding of the rules governing enhancer-gene communication and enhancer responsiveness are still rudimentary (Arnold et al., 2017; Hanssen et al., 2017; Lyu et al., 2018; Ren et al., 2017; van Arensbergen et al., 2014). Therefore, the computational predictions need to be functionally validated in the relevant patient material (i.e. specific embryonic cell types or tissues), which, unfortunately, and due to both ethical and technical reasons, is frequently not available.

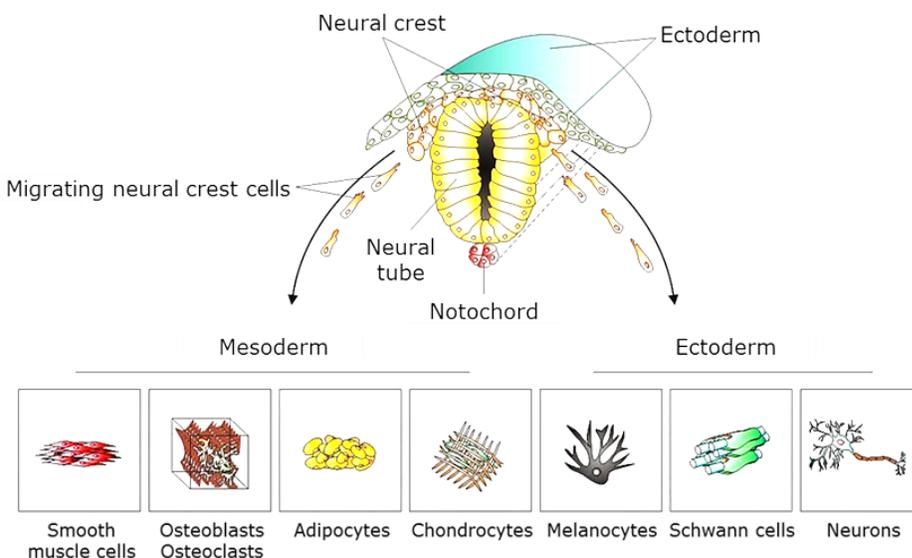
One way to overcome these limitations was recently presented by engineering of several human gain-of-function SVs in mouse embryos using CRISPR/Cas9 technology and recapitulating the molecular pathology (Franke et al., 2016; Kraft et al., 2015; Lupianez et al., 2015). Nevertheless, this strategy might not be applicable for loss-of-function SVs because of differences in gene dosage sensitivity between mice and humans that are often observed for developmental genes implicated in human congenital disorders (i.e. humans are haploinsufficient for many developmental genes while mice are not) (Andersen et al., 2014; Bartha et al., 2018; Bedell et al., 1997; Moon, 2008; Wilkie, 2003). In these cases, particular loss-of-function SVs should best be studied in the appropriate human embryonic tissue, which can be, as stated above, technically and ethically challenging. Moreover, despite the possibility to use CRISPR/Cas9 engineering to recapitulate SVs *in vitro* and *in vivo*, the generation of certain variants is rather inefficient, mostly because of their large sizes or the type of re-arrangement (e.g. translocations). Consequently, and due to the lack of appropriate experimental pipelines, the exact long-range regulatory mechanisms by which certain SV cause human disease remain largely hypothetical and mostly supported by correlative observations (Birnbaum et al., 2012b; Kleinjan and van Heyningen, 2005; Mehrjouy et al., 2018; Redin et al., 2017; Zepeda-Mendoza et al., 2018).

In principle, the use of patient derived hiPSC, containing the disease-causative SV, in combination with an appropriate *in vitro* differentiation protocol towards the desired cell type, could be used to shed light on the functional mechanism of otherwise non-tractable SVs.

## 1.2 Neural Crest – the Fourth Germ Layer

### 1.2.1 Neural crest development – a multistep process

The facial structures of humans are probably the feature that best defines our individual appearance. They are our gates to communication with others and harbor the majority of our senses. Evolutionarily, the head of vertebrates in general and of humans in particular acquired unique abilities (i.e. new sensory organs and active predatory behavior) (Gans and Northcutt, 1983; Meulemans and Bronner-Fraser, 2004) and therefore attained a new cell type, the neural crest (NC) (Baker and Bronner-Fraser, 1997b; Le Douarin and Kalcheim, 1999).



**Figure 1.2.1 Illustration of the multipotency of embryonic neural crest cells (NCC).**

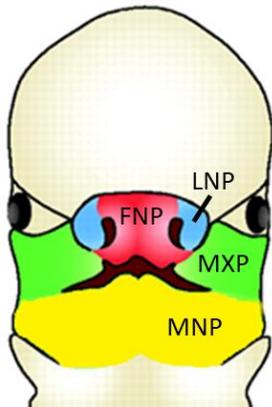
The neural crest arises between the newly formed ectoderm and the neural tube during the development of the vertebrate embryo. While the neural tube gives rise to the central nervous system (brain and spinal cord), NCC migrate away from the neural tube and have the ability to differentiate into a number of mesodermal and ectodermal cell types that are components of many different adult tissues. Modified from Kaltschmidt et al. (2012).

The formation of this transient, embryonic cell population specifies in the neural plate border during early embryogenesis (Baker and Bronner-Fraser, 1997a; Knecht and Bronner-Fraser, 2002). Subsequently, NC progenitors undergo an epithelial-to-mesenchymal transition (EMT), delaminate from the dorsal part of the neural tube and migrate to different locations within the vertebrate embryo (Gammill and Bronner-Fraser, 2003; Le Douarin and Kalcheim, 1999). Once they reach their final destinations, neural crest cells (NCC) proliferate and eventually differentiate into a great variety of ectodermal and mesodermal derivatives, overall contributing to the morphogenesis and function of many organs and tissues throughout the body, including most of

the facial bones and cartilage (Figure 1.2.1) (Cordero et al., 2011; Sauka-Spengler and Bronner-Fraser, 2008).

### 1.2.2 Cranial NCC and their regulation during formation of the vertebrate head

Although NCC are generally pluripotent, there are differences depending on their anterior-posterior (i.e. cranial, cardiac, vagal, trunk and sacral NCC) and dorso-ventral identities. Anterior-posterior positional identity of NCC depends on an expression gradient of *HOX* genes along the anteroposterior axis of the embryo while the dorso-ventral positional identity is determined by the expression pattern of *DLX* genes (Santagati and Rijli, 2003). For example, NCC in the trunk have the potential to form melanocytes as well as several types of neurons and glia cells, NCC in the embryonic head region additionally form mesenchymal derivatives, such as cartilage, bone, and connective tissues (Knecht and Bronner-Fraser, 2002; Santagati and Rijli, 2003).



**Figure 1.2.2 Facial prominences of the embryonic vertebrate head.**

The frontonasal (or median nasal) prominence (FNP, red) contributes to the forehead, the middle of the nose, the philtrum of the upper lip and the primary palate. The lateral nasal prominences (LNP, blue) form the sides of the nose. The maxillary prominences (MXP, green) give rise to the upper jaw, the sides of the upper lips and the secondary palate. The mandibular prominences (MNP, yellow) form the lower jaw, the lower lips and the sides of the middle and lower face. Modified from Helms et al. (2005).

The embryonic face of vertebrates consists of seven prominences derived from cranial NCC: the single central frontonasal prominence (FNP) and paired lateral nasal (LNP), maxillary (MXP) and mandibular (MNP) prominences. The MXP later transforms into the upper jaw, the MNP into the lower jaw (Figure 1.2.2) (Helms et al., 2005). Accordingly, the secondary palate is derived from the MXP while the primary palate is a product of the FNP (Cordero et al., 2011). Transcription within the individual prominences is highly dynamic during the formation of the face (Feng et al.,

2009), indicating that a distinct network of *cis*-regulatory regions tightly controls craniofacial development (Rada-Iglesias et al., 2012; Wilderman et al., 2018).

Thus, the cranial NCC are the progenitors of many facial tissues and any disruption during the multistep process involved in regulating cranial NCC development can result in a large variety of craniofacial malformations (Cordero et al., 2011; Mishina and Snider, 2014; Snider and Mishina, 2014; Tapadia et al., 2005; Wilkie and Morriss-Kay, 2001). The abnormalities caused by defects of NC development, including abnormalities occurring in the branchial arches or the face, are termed neurocristopathies (Benish, 1975). Therefore, investigating the regulatory mechanisms involved in NCC development is a crucial step towards the identification of the molecular etiology of human neurocristopathies.

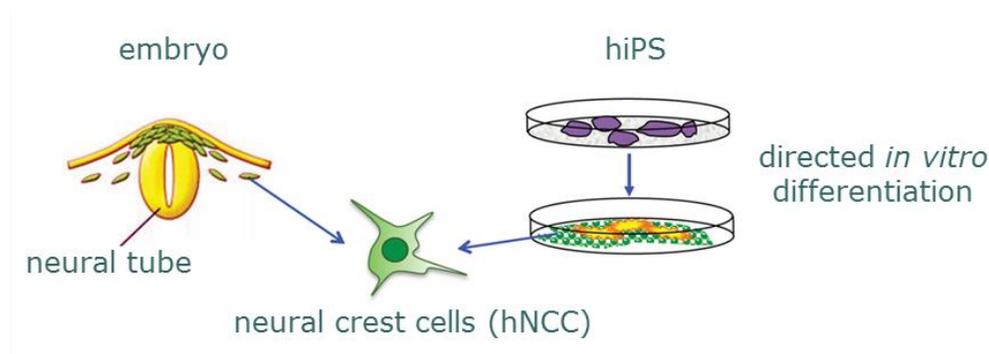
### 1.2.3 Models to study NCC

However, NCC, and thus the basis of human neurocristopathies, are difficult to study in humans due to their transient, migratory and most importantly, embryonic nature. Consequently, ethical concerns are an insurmountable barrier for human *in vivo* studies. Fortunately, the core mechanism supporting facial development are relatively uniform and conserved between diverse species (Sauka-Spengler and Bronner-Fraser, 2008). Therefore, results from model organisms provide insight into the genetic causes of human craniofacial defects. Since the NC is a novelty to vertebrates it naturally cannot be studied in invertebrates, such as yeast, worms or flies which are otherwise popular model organisms for genetic analyses. Also, mouse embryos are small and difficult to manipulate at the early developmental stages when the tissues of interest form. In contrast, chicken embryos and other avian model organisms, such as duck and quail, are quite large at these stages of development and easier to manipulate because they develop *in ovo*. Therefore, especially chicken embryos have historically been used to study the NC in great detail (Abramyan and Richman, 2018; Le Douarin, 2004). What makes avian model organisms so attractive is not only the accessibility of the embryos but also the fact that early neural development of chicken embryos bears closer resemblance to humans than rodents, such as mice, do (Barriga et al., 2015; Gammill and Bronner-Fraser, 2003).

Nevertheless, cell-type specific *cis*-regulatory elements, such as those active in NCC, frequently display limited conservation (Acloque et al., 2009; Barriga et al., 2015) even between closely related species such as chimp and human (Prescott et al., 2015). Therefore, the etiological basis of human neurocristopathies should be ideally studied in human cells. However, access to or

availability of relevant patient tissue is often restricted. This limitation can, in general, be overcome by utilization of *in vitro* differentiation protocols towards specific human embryonic cell types (Adamo et al., 2015; Lee et al., 2010; Menendez et al., 2013; Okuno et al., 2017; Shi et al., 2017; Takahashi et al., 2007; Tchieu et al., 2017). For neurocristopathies several protocols for *in vitro* differentiation of human embryonic stem cells (hESC) or induced pluripotent stem cells (hiPSC) into NCC are available (Bajpai et al., 2010; Lee et al., 2010; Menendez et al., 2013). In addition, these *in vitro* differentiation systems are highly tractable, as they can be genetically modified with ease and can yield large amounts of biological material for molecular analysis.

Thus, in the work presented here, we combined *in vivo* data obtained from analysis of chicken embryos with a robust *in vitro* human NC differentiation system that faithfully recapitulates NC development (Figure 1.2.3) (Bajpai et al., 2010). This allows us to confirm any findings in the context of the human genome and investigate the molecular mechanism behind the visible phenotype. This protocol has previously been used to map the set of enhancers regulating the specific gene expression program of this unique cell type (Rada-Iglesias et al., 2012) as well as to elucidate the etiological mechanism of human neurocristopathies, such as CHARGE syndrome (Bajpai et al., 2010). Furthermore, this differentiation system could be in principle combined with the derivation of patient-specific hiPSC, which is emerging as a powerful strategy to model human disease, especially when animal models are either not optimal or not easily available (Barrell et al., 2019; Okuno et al., 2017).



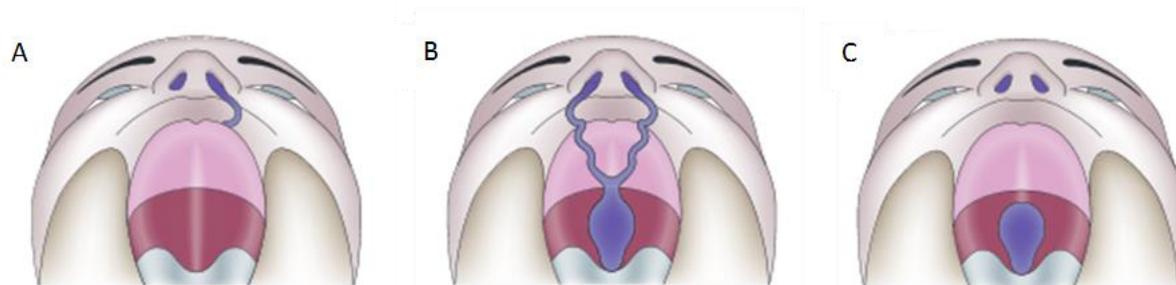
**Figure 1.2.3** Schematic illustration of parallels between neural crest differentiation in the embryo and directed *in vitro* differentiation.

Modified from Rada-Iglesias et al. (2013).

### 1.3 Orofacial Clefts

As an example for a neurocristopathy here we focus on orofacial cleft (OFC). OFC is one of the most common craniofacial birth defects among humans. One of about 700 babies world-wide is born with an OFC. There are ethnic differences in the prevalence of OFC ranging from one case in 500 individuals of Asian or Native American origin over one in 1000 in European populations to one case in 2500 individuals of African ancestry (Dixon et al., 2011; Rahimov et al., 2012).

OFC cases can be divided, based on epidemiological and embryological features, in cleft lip with and without cleft palate (CL/P) and cleft palate only (CPO) (Figure 1.3.1) (Dixon et al., 2011). CL/P further displays a range of phenotypic severity from notches or grooves in vermillion or philtrum to complete unilateral cleft lip and bilateral cleft lip and palate. OFC also includes microforms, represented e.g. by small defects of the lip and/or alveolar arch, ridges above the lip and defects of the superior orbicularis oris muscle, which are only detectable by ultrasound (Jugessur et al., 2009; Neiswanger et al., 2007).



**Figure 1.3.1 Illustration of orofacial clefting.**

Unilateral cleft lip (A), bilateral cleft lip and palate (B), cleft palate only (C). Clefts can affect only the lip (A), lip, primary palate (pink) and secondary palate (red) (B) or only the secondary palate (C). Modified after Dixon et al. (2011)

In human embryos, OFC appears during early embryonic development, between the third and fifth week of pregnancy, when the facial prominences fuse and any disturbance can lead to defective fusion (Dixon et al., 2011; Etchevers et al., 2006; Thomas et al., 2008; Wyszynski, 2002). Closure failure can happen due to mis-regulated proliferation or abnormal cell death during advanced stages of the fusion event but can also have its origin in earlier stages during the specification, EMT or migration of neural crest cells (Cordero et al., 2011; Etchevers et al., 2006; Mossey et al., 2009).

OFC can lead to mild or severe deformations of the face, either as part of a syndrome together with other malformations or as an isolated cleft. Genetically, non-syndromic OFC is a highly

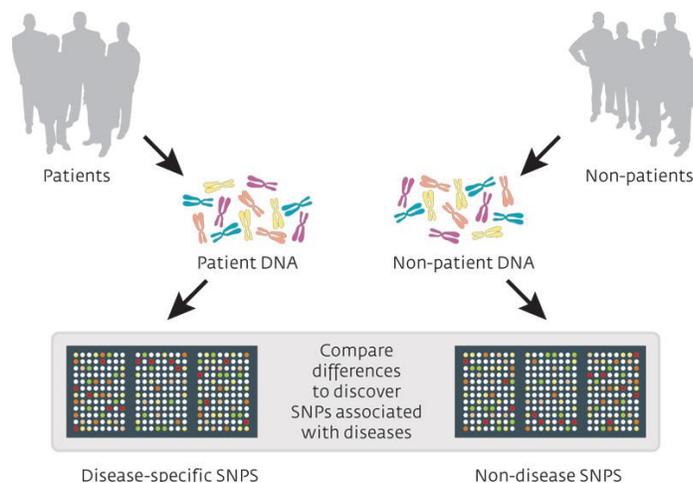
heterogeneous disorder that involves a wide range of genes/loci. While the causal gene for syndromic clefts is often known (e.g. *IRF6* for Van der Woude syndrome (OMIM #119300), *TFAP2A* for branchiooculofacial syndrome (OMIM #113620), *TP63* for EEC syndrome (OMIM #604292), *TBX22* for ankyloglossia (OMIM #303400) and *CDH7* for CHARGE syndrome (OMIM #214800)), the etiology of the non-syndromic form of OFC is, in general, much more challenging to unravel, because it involves complex genetics as well as environmental factors (Dixon et al., 2011). Additionally, the majority (70%) of patients with CL/P are non-syndromic cases, which makes the search for the etiological basis more complicated (Calzolari et al., 2007; Mossey et al., 2009).

Even though, nowadays, OFCs are routinely surgically corrected, it still represents an immense physiological, psychological and financial burden for the affected individuals, their families, and society (Christensen et al., 2004; Wehby and Cassell, 2010; Wyszynski, 2002). Therefore, a more profound knowledge of the etiology of this disorder is crucial. Understanding of the OFC-causative genes and the pathways involved would not only improve genetic counselling but may ultimately also enable targeted prenatal repair of OFC and other congenital disorders. For example, efforts have been made to correct CL/P in murine embryos *ex utero* by increasing expression of *Wnt9b* in *Pbx*-deficient mice, because *Wnt9b* acts downstream of *Pbx* within the Wnt signaling pathway and therefore was partly able to rescue the *Pbx* deficit (Dong et al., 2017). On this account, several approaches have been used to identify new candidate loci for non-syndromic OFC.

### 1.4 Genome Wide Association Studies (GWAS)

In the past, candidate regions/genes for OFC, as with many other diseases, were mostly discovered using linkage studies in affected families. This way, *IRF6* was found to be not only a candidate for Van der Woude Syndrome (Kondo et al., 2002; Murray et al., 1990) but also for non-syndromic OFC (Butali et al., 2014; de Lima et al., 2009; Leslie et al., 2013; Zuccherro et al., 2004). Similarly, *FOXE1* was identified as an OFC candidate gene by a combination of linkage and association studies (Marazita et al., 2004; Moreno et al., 2009). Still, the complexity of diseases such as non-syndromic OFC require more extensive studies to elucidate their genetic basis. Therefore, genome-wide association studies (GWAS), which have also been performed for numerous other complex traits and diseases like diabetes, asthma or schizophrenia, were used to find new risk loci for OFC. These GWAS identify common single nucleotide polymorphisms

(SNPs) that are overrepresented in a group of patients with a specific disease, e.g. isolated OFC, in comparison to matched healthy controls (Figure 1.4.1). The aim of GWAS is to map risk loci through the detection of associations between genotype frequency and trait status (McCarthy et al., 2008).



**Figure 1.4.1 Schematic description of genome-wide association studies (GWAS).**

Selected single nucleotide polymorphisms are genotyped in a representative group of people with a specific phenotype or a disease (patients) and in a matching population or control group of healthy individuals (non-patients).

This way, GWAS have been producing a large number of candidate SNPs associated with a wide range of quantitative traits and complex human diseases, including about 40 new risk loci for OFC (Beaty et al., 2010; Beaty et al., 2013; Birnbaum et al., 2009; Carlson et al., 2017; Grant et al., 2009; Haaland et al., 2018; Leslie et al., 2017; Leslie et al., 2016; Ludwig et al., 2017; Ludwig et al., 2012; Mangold et al., 2010; Sun et al., 2015; Yu et al., 2017).

Additionally, these OFC GWAS confirmed the previously identified candidate genes *IRF6* and *FOXE1* as significantly associated with OFC (Beaty et al., 2010; Birnbaum et al., 2009; Grant et al., 2009; Leslie et al., 2016; Ludwig et al., 2014; Ludwig et al., 2012; Mangold et al., 2010).

However, finding the disease-causative SNP on the basis of pure association of common non-coding SNPs, as obtained by GWAS, is not straight forward. Firstly, disease-associated common SNPs identified by GWAS are likely to have a small effect size with reduced penetrance. In addition, the SNPs reported as disease-associated by GWAS are typically part of larger haplotypes in which several SNPs are in linkage disequilibrium (LD) and, thus, equally likely to be causal for the investigated disease/trait. Nevertheless, among the SNPs in LD only a few may have functional biological effect, termed “driver variants”, while other SNPs of the same haplotype are merely “passenger variants” (Thieme and Ludwig, 2017). Finally, it is also possible that the

causative variants are rare ones and, thus, cannot be found among the common SNPs reported in GWAS.

To remedy these problems, targeted re-sequencing and exome sequencing have become popular tools to screen risk regions for rare mutations in patients and have successfully identified a few new candidate genes for OFC (i.e. *ARHGAP29* (Leslie et al., 2012; Liu et al., 2017a), *CDH1* (Brito et al., 2015; Bureau et al., 2014; Ittiwut et al., 2016), *GRHL* (Peyrard-Janvid et al., 2014)).

Another major limitation of GWAS is that the majority (80-95%) of disease-associated SNPs occur in non-coding regions of the DNA (Hindorff et al., 2009; Maurano et al., 2012; Welter et al., 2014). In contrast to coding-mutations, the pathological consequences of non-coding genetic variants are much more difficult to predict and interpret. Importantly, most of the non-coding variants reported in GWAS lie within putative *cis*-regulatory elements, such as enhancers, identified through epigenomic profiling. The SNPs occurring within enhancers might alter the activity of these regulatory elements, which can lead to quantitative changes in gene expression and, thus, to increase of human disease susceptibility (Bauer et al., 2013; Dickel et al., 2013; Huang et al., 2014; Leslie et al., 2015; Lettice et al., 2003; Lidral et al., 2015; Maurano et al., 2012; Monteiro and Freedman, 2013; Pasquali et al., 2014; Sakabe et al., 2012; Smemo et al., 2014; Soldner et al., 2016; Spieler et al., 2014; Visel et al., 2009). However, non-coding SNPs that strongly affect enhancer activity and gene expression are probably rare and the expression of many genes is believed to be controlled by multiple and partially redundant enhancers, which increases transcriptional robustness (Hnisz et al., 2013; Osterwalder et al., 2018). Therefore, changing the activity of a single enhancer due to non-coding variants might minimally alter gene expression under normal conditions. Consequently, most disease-associated variants within enhancers are predicted to influence the expression of their target genes in a moderate and quantitative manner, which subtly alters the risk for a disease but only leads to a pathological phenotype in the presence of additional genetic and environmental factors (Dixon et al. 2011). Therefore, disease-associated SNPs reported in GWAS that occur within enhancers have been used in several studies as a starting point to uncover novel genes implicated in human complex disorders without specifically aiming to identify the disease causative variants that might occur within the enhancer (Smemo et al., 2014; Uslu et al., 2014).

## 1.5 Novel candidate genes for non-syndromic OFC

Based on the previous concepts, in this thesis we have used an OFC-risk haplotype overlapping with an hNCC active enhancer to uncover three potential novel genes implicated in the etiology of OFC. We briefly describe below what is already known about these genes.

### **DDX1 – an RNA helicase with manifold functions**

DDX1 is a member of the DEAD-box protein family, the largest group of RNA helicases characterized by a couple of conserved motifs, including the signature motif, D(asp)-E(glu)-A(ala)-D(asp), which is involved in ATP hydrolysis, RNA binding and unwinding (Linder et al., 1989). DEAD-box proteins have been implicated in all aspects of RNA metabolism, including RNA processing, RNA stability, RNA transport and RNA remodeling (Li et al., 2008). They modulate RNA secondary structures during processes such as ribosome biogenesis, nuclear and mitochondrial splicing, mRNA export and decay as well as translation initiation (Rocak and Linder, 2004).

The *DDX1* gene is located at 2p24 and was originally identified in neuroblastomas and retinoblastomas where it was found to be amplified and overexpressed (Godbout et al., 1998; Godbout and Squire, 1993), usually together with its neighboring gene *MYCN* (Defferrari et al., 2007; George et al., 1996; Manohar et al., 1995; Squire et al., 1995; Weber et al., 2004).

DDX1 is ubiquitously expressed within the developing embryo, however with spatial and temporal variation, the highest levels occurring in proliferating and neuroectodermal cells, which suggests a specific role for DDX1 in a subset of differentiated cells and a more general role in undifferentiated cells (Godbout et al., 2002; Godbout and Squire, 1993).

The DDX1 protein plays a role in 3'end pre-mRNA processing (Bleoo et al., 2001), in alternative and stress-induced splicing (Germain et al., 2015; Zhong et al., 2018), and in regulation of translation (Li et al., 2018). DDX1 is indirectly involved in the regulation of mRNA decay, through binding of the decay-promoting KSRP protein (Chou et al., 2013) and also in transcription regulation (Ishaq et al., 2009; Tanaka et al., 2009), tRNA maturation and miRNA processing (Han et al., 2014; Popow et al., 2014; Sunden et al., 2007).

DDX1 is used by viruses for replication (Fang et al., 2004; Xu et al., 2010) and nuclear export (Fang et al., 2005; Robertson-Anderson et al., 2011). DDX1 is involved in tumor progression

(Balko and Arteaga, 2011; Germain et al., 2011; Tanaka et al., 2009) and has recently been identified as a molecular player in post-transcriptional RNA G-quadruplex conversion into R-loops during IgH class switch recombination (CSR) (Ribeiro de Almeida et al., 2018). R-loops are RNA-DNA hybrids between nascent RNA and its template DNA, formed either co-transcriptionally or post-transcriptionally, with DDX1 being only involved in the latter (Ribeiro de Almeida et al., 2018). Destabilization of R-loops are suggested to lead to an increase of DNA double strand breaks (DSBs) and general genome instability (Li et al., 2016).

Depending on the cell type and its biological function within a specific cell type, DDX1 can be localized in the nucleus and/or the cytoplasm. In the nucleus, DDX1 has been reported to form foci that have been associated with structures involved in RNA processing (Bleoo et al., 2001; Li et al., 2006). Moreover, upon ionizing radiation DDX1 is recruited to DNA DSBs and co-localize with DNA damage response proteins (i.e.  $\gamma$ H2X and pATM), suggesting that DDX1 is required for early DNA damage response at transcriptionally active regions of the genome (Li et al., 2008). Further evidence points towards a role for DDX1 to ensure efficient DSB repair and cell survival by removal of RNA from accumulating R-loops at DSB (Li et al., 2016).

Additionally, DDX1 was found in stress granules in the cytoplasm (Hildebrandt et al., 2019; Onishi et al., 2008; Ozeki et al., 2019). In contrast, DDX1 co-localized with granules involved in neuronal RNA transport, which shuttle molecules between the cytoplasm and the nucleus and are therefore located in both compartments (Miller et al., 2009; Perez-Gonzalez et al., 2014). Similarly, viruses use DDX1 for RNA transport (Fang et al., 2004; Kanai et al., 2004) and in cancer cells, overexpressing *DDX1*, the protein is also located both in the cytoplasm and in the nucleus (Godbout et al., 2007). Together with DDX21 and DHX36, DDX1 acts as a viral sensor in the cytosol, which can translocate to the mitochondria (Zhang et al., 2011).

*In vitro*, *DDX1* has been shown to unwind both RNA-RNA and RNA-DNA duplexes and digest single stranded RNA, though this ability to unwind double stranded RNA are most likely dependent on other components of a complex containing DDX1 (Chen et al., 2002; Li et al., 2008).

While other DEAD-box protein family members are proposed to be involved in germ cell development (DDX4), organ differentiation (DDX5), spermatogenesis (DDX25) and visual system development (DDX41) (Abdelhaleem et al., 2003), DDX1 function during development is largely uncharacterized, mostly because *Ddx1* null<sup>-/-</sup> mice die very early during embryogenesis (~E3.5) (Hildebrandt et al., 2015). A recent study proposes that this early lethality might be due to the role of DDX1 in the degradation of maternal RNA at very early stages of embryonic development

(1-2-cell stage) (Hildebrandt et al., 2019). Furthermore, while knockdowns of DDX1 have been studied (Li et al., 2008; Ribeiro de Almeida et al., 2018), there have been no reports of complete human *DDX1* knockout cell lines, further supporting an essential role for DDX1 in cell viability. Only in a *Drosophila melanogaster* study, Ddx1 knockout flies were viable but much smaller than controls and loss of Ddx1 disrupted spermatogenesis (Germain et al., 2015).

Finally, while the molecular function of DDX1 seems to be highly diverse, there are no facial phenotypes connected to *DDX1*, and its role during embryogenesis, especially in craniofacial development, is largely unknown.

### **MYCN - a transcription factor with connections to NC-related development**

*MYCN* encodes a transcription factor of the MYC family which belongs to the basic-helix-loop-helix-zipper (bHLHZ) class. Besides *MYCN*, this family of proto-oncogenes also includes *MYCL* and *MYC* and regulates expression of genes involved in a number of cellular processes, including proliferation, differentiation, apoptosis and energy metabolism (Henriksson and Luscher, 1996).

In some cases, *MYCN* and *MYC* can partially compensate for each other in certain cellular processes (i.e. differentiation and cell growth), because they regulate the same set of genes (Charron et al., 1992; Malynn et al., 2000; Sawai et al., 1993), while in other cases they appear to have different, cell type-specific functions (Stanton et al., 1992). Also, while *MYC* is highly and ubiquitously expressed in embryonic and adult proliferating cells, *MYCN* expression is more restricted to certain tissues and significant expression could be detected solely during embryonic development (Zimmerman et al., 1986). *N-myc* null mice die during early embryonic development (E10.5) (Charron et al., 1992; Stanton et al., 1992).

Although *MYCN* is primarily expressed in normal developing embryos, it was first identified in neuroblastomas, an NC-derived pediatric tumor that develops in the peripheral nervous system. Neuroblastomas can be caused by the genomic amplification of *MYCN* and its consequent overexpression (Kohl et al., 1983; Schwab et al., 1983). Subsequently, amplification of *MYCN* has been associated with a variety of tumors and is generally considered as a proto-oncogene (Wei et al., 2008). *MYCN* overexpression is associated with poor prognosis/bad outcome in neuroblastomas (Brodeur et al., 1984; Kohl et al., 1984).

Functionally, *MYCN* has been shown to promote proliferation in neuroblastomas and other cancers while *in vitro* knockdown of *MYCN* resulted in decreased cell proliferation (Kramer et al.,

2016; Liu et al., 2016; Liu et al., 2017b; Liu et al., 2019; Swartling et al., 2012; Zeid et al., 2018). Specifically in the NC, MYCN was found to be critical for cell growth and cell cycle progression (Zhang et al., 2016).

Heterozygous mutations in or deletions of the *MYCN* gene leading to haploinsufficiency were also found in patients with Feingold syndrome (OMIM#164280), a rare autosomal dominant syndrome, which is characterized by a constellation of phenotypes such as microcephaly, facial and hand abnormalities, tracheoesophageal fistula, duodenal atresia, and developmental delay (Burnside et al., 2018; Celli et al., 2003; Chen et al., 2012; Feingold et al., 1997; Marcelis et al., 2008; Tetzlas et al., 2006; van Bokhoven et al., 2005). The facial phenotypes did not explicitly include OFC but rather other NC-related malformations such as micrognathia and dental malocclusion (Tetzlas et al., 2006).

Additionally, *MYC*, the most important paralog of *MYCN*, is well known for its role in the neural crest (Bellmeyer et al., 2003; Kerosuo and Bronner, 2016; Wei et al., 2007). Most notably, an NCC-specific long-range enhancer regulating *MYC* expression was previously found to be required for normal facial development (Uslu et al., 2014).

### **FAM49A – a largely uncharacterized gene**

Not much is known about *FAM49A*. Its full name is “family with sequence similarity 49 member A” and therefore belongs to the FAM49 protein family, whose members have largely unknown functions. *FAM49A* encodes a conserved protein of 323 amino acids that is broadly expressed in the brain of mice and zebrafish and in human brain and thyroid but very little in hNCC (Fagerberg et al., 2014; Howe et al., 2013; Richardson et al., 2010). Two independent GWAS and a replication study proposed that the 2p24 OFC-risk locus analyzed in this work might involve *FAM49A* because the risk-haplotype partly overlaps with the coding region of *FAM49A* (Chen et al., 2018; Leslie et al., 2016; Yu et al., 2017).

## **1.6 Branchiooculofacial Syndrome (BOFS) and TFAP2A**

BOFS is a rare autosomal dominant congenital syndrome displaying incomplete penetrance and variable expressivity and is characterized by branchial, ocular, ear and facial abnormalities

(Fujimoto et al., 1987; Lin et al., 1993; Milunsky et al., 2008; Milunsky et al., 2011). Many of these abnormalities are likely to be caused by developmental defects affecting the neural crest, which is why BOFS is considered as a neurocristopathy (Bennaceur et al., 1998; Lin et al., 1993). In contrast to the high genetic heterogeneity of non-syndromic orofacial cleft, all previously known BOFS cases carry heterozygous mutations in or deletions of the *TFAP2A* gene, which encodes for a transcription factor considered as a NC master regulator (Aliferis et al., 2011; Brewer et al., 2004; Galliani et al., 2012; Gestri et al., 2009; LeBlanc et al., 2013; Li et al., 2013; Lim et al., 2005; Lin et al., 1993; Milunsky et al., 2008; Milunsky et al., 2011; Murray et al., 2013; Reiber et al., 2010; Stoetzel et al., 2009; Tekin et al., 2009).

The vast majority of BOFS cases are due to loss of function missense mutations within the gene, indicating that *TFAP2A* is a dosage sensitive gene during human embryogenesis (Brewer et al., 2004; Dooley et al., 2019; Gammill and Bronner-Fraser, 2003; Lin et al., 1993; Milunsky et al., 2008; Milunsky et al., 2011; Sauka-Spengler and Bronner-Fraser, 2008). For example, in all reported cases by Milunsky et al. (2011), the only BOFS family carrying a heterozygous deletion of the *TFAP2A* gene was the only family not displaying the typical CL/P, indicating that the complete absence of one copy of the gene may have a milder phenotypic effect (excluding CL/P from the spectrum) than mutations within the gene that result in antimorphic (i.e. dominant negative) alleles. In agreement with this, LeBlanc et al. (2013) reported a patient with an extremely mild form of BOFS, likewise lacking one copy of the *TFAP2A* gene. Although these limited number of reports suggest that there might be a genotype-phenotype connection in BOFS that explains its variable expressivity, the authors also indicate that additional studies are necessary to determine if these observations are consistent (Milunsky et al., 2008; Milunsky et al., 2011), especially in light of a family with more severe expressivity of BOFS (including CL/P) that, nevertheless, carries a heterozygous deletion of *TFAP2A* (Gestri et al., 2009). Alternatively, the variable expressivity of BOFS may be explained by possible additional genetic and environmental risk factors that interact with *TFAP2A*.

90% of mutations within the *TFAP2A* gene have been found to be accumulating in a highly conserved hotspot region around exons four to six, which are, notably, almost free of any naturally occurring SNPs (Milunsky et al., 2011; Reiber et al., 2010). This indicates a prevalence of this region for *de novo* mutations, supported by the observation that several mutations are recurrent, and/or the special importance of this specific region, which encodes the *TFAP2A* DNA binding domain. Consequently, variants in other regions of the gene might be less likely to affect *TFAP2A* function and, thus, to lead to a phenotype.

The *TFAP2A* gene is located on human chromosome 6p24 (Gaynor et al., 1991; Williamson et al., 1996) and encodes the 52 kDa transcription factor TFAP2A (or AP2- $\alpha$ ). TFAP2A binds to the consensus sequence CCAGGC (de Croze et al., 2011; Luo et al., 2003; Nikitina et al., 2008) by its helix-span-helix motif (Mitchell et al., 1987; Mohibullah et al., 1999; Williams et al., 1988), just like the other two AP2-family members AP2- $\beta$  and AP2- $\gamma$  (Eckert et al., 2005; Williamson et al., 1996). *TFAP2A* is essential for early stages of NC differentiation, particularly for specification at the neural plate border and participates in the BMP and WNT pathways, which are both crucial in NC development (Barrallo-Gimeno and Nieto, 2005; Knight et al., 2005; Wang et al., 2011).

In human and mouse embryos, *TFAP2A* is expressed in the lens, neural retina, nasal processes, and epithelial lining of the oral cavity and palatal shelves (Gestri et al., 2009). Furthermore, *TFAP2A* has been found to be especially important for regulation in the development of lens vesicles in the eye, facial prominences, cranial closure and limb buds (Schorle et al., 1996; West-Mays et al., 1999). In accordance with this, TFAP2A actively regulates transcription of genes involved in embryogenesis of the neural tube, body wall, limbs, eye, ear and face (Ahituv et al., 2004; Nelson and Williams, 2004; Schorle et al., 1996; Zhang et al., 1996). In NC precursor cells, TFAP2A binds to distal enhancers co-occupied by p300 and H3K27ac and to promoters of key regulatory genes involved in craniofacial development co-bound by NR2F1/2 (Miranda et al., 2018; Rada-Iglesias et al., 2012).

Interestingly, *Tfap2a*-null mice die during embryonic development or at birth due to severe defects of NCC derivatives, including extreme dysmorphogenesis of the face. In contrast, a conditional knockout of *Tfap2a* in NCC results in a milder, BOFS-like phenotype (Brewer et al., 2004; Schorle et al., 1996; Zhang et al., 1996). Importantly, the morphology of *Tfap2a*<sup>+/-</sup> mice does not seem to be altered, indicating that a heterozygous disruption of *Tfap2a* does not lead to haploinsufficiency in mice (Zhang et al., 1996). Therefore, to enlighten the exact etiological mechanism of BOFS, it seems to be necessary to study the role of TFAP2A in a human embryonic context.

Finally, a role for *TFAP2A* in tumorigenesis has also been proposed, attributing TFAP2A with a tumor-suppressor activity by direct interaction with p53 and suppression of MYC (Gaubatz et al., 1995; Heimberger et al., 2005; Jean et al., 1998; McPherson et al., 2002; Orso et al., 2007; Stabach et al., 2006).

## 1.7 Aims

The overall aim of this thesis was to uncover how the genetic or structural disruption of enhancer activity during neural crest development can lead to human neurocristopathies. To achieve this goal, I combined *in vitro* and *in vivo* developmental models with genetic engineering approaches and multiple genomic methods.

In the first part of this work, we investigated whether structural variants can cause congenital diseases, such as BOFS, by disrupting enhancer landscapes. More specifically, we characterized a large inversion that moves the *TFAP2A* gene far away from numerous enhancers usually located within the same TAD as *TFAP2A*, we determined the importance of these enhancers for NCC development, examined the interaction between the *TFAP2A* promoter and its cognate enhancers, investigate the consequences of the inversion for *TFAP2A* expression and assessed the phenotypic consequences during NC development.

In the second part of the thesis, our goal was to shed light on the molecular basis for the etiology of non-syndromic OFC and, in doing so, identify new genes implicated in facial development. Non-coding variants associated with complex developmental abnormalities such as non-syndromic OFC are predicted to influence the expression of their target genes in a quantitative and moderate manner, which together with additional genetic and environmental factors might lead to a pathological phenotype. Alternatively, more dramatic changes in the expression or function of the target genes themselves could independently lead to craniofacial abnormalities (e.g. *TFAP2A* in BOFS). Here we used disease-associated SNPs as a starting point to identify NCC active enhancers potentially linked to novel genes implicated in the etiology of non-syndromic OFC. More specifically, we characterized an OFC risk locus at 2p24.2 identified by GWAS and overlapping with an active hNCC enhancer. Firstly, we determined the function of the enhancer itself and then investigated the biological role of its putative target gene(s), focusing preferentially on *DDX1* as it represents a potentially novel regulator of NC and facial development.

## 2 MATERIAL & METHODS

### 2.1 Equipment

Table 2.1 Equipment

Equipment	Brand	Specification
Automated cell counter	Biorad	TC20
Bacteria Incubator, Ecotron	Infors HT	s00120638
Bioruptor plus	Diagenode	B01020001
ChemiDoc MP Imaging System	Biorad	1708265
Centrifuge, microtubes, cooling	Hermle	Z 216 MK
Centrifuge, universal, cooling	Hermle	Z 326 K
Centrifuge, universal, cooling	Hermle	Z 383 K
Centrifuge, high performance	Beckman Coulter	Avanti J-E
Egg Incubator	Hemel Brutgeräte	Thermo-de-Luxe 200
LightCycler 480 II	Roche	5662
Microscope, confocal	Olympus	Fluoview FV 1000
Microscope, fluorescent	Olympus	IX 83
Microscope, fluorescent	Nikon	ECLIPSE TS100
Microscope, inverted	Leica	DMILED 376977
Microscope, Stereomicroscope	Olympus	SZX16
Mini Gel Tank	Invitrogen	A25977
Nanodrop	Thermo Scientific	F673
Plate Reader, Tristar	Berthold Technologies	LB941
Power supply	Biorad	041BR110323
Power supply	Biorad	043BR500041
Shaker	Skyline	12DE117
Spectrophotometer UV/Vis	Beckman Coulter	DU730
Spin Concentrator	Eppendorf	5301
Thermo Block	Ditabis	980052301
Thermocycler c1000 Touch	Biorad	ct024292
Thermocycler Tprofessional	Biometra	TRIO46
Tissue Culture Incubator	Sanyo	8070263
Tissue Culture Hood	Kojair	22198
Waterbath	Memmert	325741
Western Blotting Chamber, Mini-PROTEAN® Tetra Cell	Biorad	1658004

## 2.2 Chemicals and Reagents

Table 2.2 Chemicals

Chemical	Brand	Catalog Number
<b>2-Mercaptoethanol</b>	Gibco	31350010
<b>Acetone</b>	Roth	5025
<b>Adenosine triphosphate (ATP)</b>	Sigma Aldrich	A2383
<b>Agarose</b>	Life Technologies	16500100
<b>Alcian Blue 8GX</b>	Sigma Aldrich	A3157
<b>Aquatex Permanent Aqueous Mounting Agent</b>	Sigma-Aldrich	108562
<b>Bovine Serum Albumin (BSA)</b>	Roth	3737
<b>Bromophenol Blue</b>	Sigma Aldrich	B0126
<b>CAA (2-Chloroacetamide)</b>	Sigma-Aldrich	C0267
<b>Chloroform</b>	Sigma Aldrich	366919
<b>DAPI (4',6-Diamidin-2-phenylindol)</b>	Sigma Aldrich	D9542
<b>DTT (Dithiothreitol)</b>	Roth	6908
<b>DMSO (Dimethyl Sulfoxide)</b>	Sigma Aldrich	D2650
<b>Ethanol</b>	Roth	5054
<b>EDTA (Ethylendiaminetetraacetic acid)</b>	Roth	8043
<b>EGTA (Ethylene glycol-bis(<math>\beta</math>-aminoethylether)-N,N,N',N'-tetraacetic acid)</b>	Roth	3054
<b>Fast Green FCF</b>	Sigma Aldrich	F7258
<b>Fluoromount-G mounting medium</b>	SouthernBiotech	0100-01
<b>Formaldehyde Solution 37%</b>	Sigma Aldrich	252549
<b>Glycerol</b>	Roth	3783
<b>HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)</b>	Roth	HN78
<b>Hydrochloric acid (HCl)</b>	Roth	281.1
<b>IGEPAL CA-630</b>	Sigma Aldrich	I8896
<b>Isopropanol</b>	Roth	9866
<b>Lithium Chloride (LiCl)</b>	Roth	3739
<b>Magnesium Chloride (MgCl<sub>2</sub>)</b>	Roth	KK36
<b>Methanol</b>	Sigma Aldrich	494437
<b>Orange G dye</b>	Sigma Aldrich	O3756
<b>Neutral Red Dye</b>	Sigma Aldrich	N7005
<b>N-Lauroylsarcosine</b>	Sigma Aldrich	61743
<b>Non-Fat Milk Powder</b>	AppliChem	A0830
<b>NP-40</b>	Sigma Aldrich	I3021
<b>Nuclear Fast Red Solution</b>	Sigma Aldrich	N3020

<b>Phenol-Chlorophorm Isoamyl Alcohol (25:24:1)</b>	Sigma Aldrich	P2069
<b>Phosphate buffered saline (PBS)</b>	Sigma Aldrich	D8537
<b>Potassium Chloride (KCl)</b>	Roth	HN02
<b>Sodium Acetate (C<sub>2</sub>H<sub>3</sub>NaO<sub>2</sub>)</b>	Roth	6773
<b>Sodium Chloride (NaCl)</b>	Roth	3957
<b>Sodium (Na)-Deoxycholate</b>	Sigma Aldrich	D678
<b>Sodium dodecyl sulfate (SDS)</b>	Roth	183
<b>Triton X-100</b>	Roth	3051
<b>Trizma Base (TRIS)</b>	Sigma Aldrich	T1503
<b>Tween-20</b>	Roth	9127

Table 2.3 Molecular Biology Reagents and Enzymes

<b>Reagent/Enzyme</b>	<b>Brand</b>	<b>Catalog Number</b>
<b>AccuPrime™ GC-Rich DNA Polymerase</b>	Invitrogen	12337016
<b>Agencourt AMPureXP beads</b>	Beckmann Coulter	A63881
<b>Ampicillin</b>	Roth	HP62
<b>BamHI-HF</b>	NEB	R3136L
<b>BbsI</b>	NEB	R0539L
<b>BCIP (5-Bromo-4-chloro-3-indolyl phosphate p-toluidine salt)</b>	VWR	0885
<b>Bis-Tris Plus Gels, 8%</b>	Invitrogen	NW00080
<b>Bis-Tris Gels 4-12%</b>	Biorad	3450124
<b>Blocking Powder</b>	Roche	1 096 176
<b>BstBI</b>	NEB	R0519L
<b>Deionized Formamide</b>	VWR	0606
<b>CHAPS (3-[(3-Cholamidopropyl) dimethylammonio]-1-propansulfonat)</b>	VWR	0465
<b>DIG-11-UTP</b>	Roche	11209256910
<b>DNA ladder Gene Ruler 1 kb plus (75-20.000 bp)</b>	Peqlab	25-2240
<b>dNTP mix</b>	Promega	U1515
<b>DpnII</b>	NEB	R0543M
<b>Dynabeads M-280 sheep anti-rabbit</b>	Invitrogen	11203D
<b>Dynabeads Protein G</b>	Invitrogen	1004D
<b>Dynabeads MyOne Silane</b>	Invitrogen	37002D
<b>ECL Western Blotting Substrate</b>	GE Healthcare	RPN2109
<b>EcoRI-HF</b>	NEB	R3101S
<b>Expand Long Template PCR system</b>	Roche	11681842001
<b>Exo-SAP-IT</b>	Affymetrix	78201
<b>FastAP</b>	Thermo Scientific	EF0651

<b>Fab fragment</b>	Roche	11093274910
<b>Heparin</b>	Sigma Aldrich	H3400
<b>IPTG (Isopropyl <math>\beta</math>-D-thiogalactoside)</b>	Sigma Aldrich	I6758
<b>KpnI-HF</b>	NEB	R3142S
<b>LB-Agar</b>	Roth	X965
<b>LB-Medium (Lennox)</b>	Roth	X964
<b>Midori Green Advance</b>	Biozym	617004
<b>NBT (Nitro blue tetrazolium chloride)</b>	VWR	VWRV0329
<b>NheI-HF</b>	NEB	R3131L
<b>Nitrocellulose Western Blotting Membrane, 0.45 <math>\mu</math>m</b>	GE Healthcare	GE10600003
<b>NlaIII</b>	NEB	R0125L
<b>NTP set</b>	Roche	11277057001
<b>One Taq DNA Polymerase</b>	NEB	M0509S
<b>ORA qPCR Green ROX L Mix</b>	HighQu	QPD0105
<b>Platinum Taq DNA Polymerase</b>	Invitrogen	10966034
<b>Power SYBR Green 2x Master Mix</b>	Applied Biosystems	4368577
<b>Protease Inhibitor Cocktail, Complete Ultra</b>	Roche	5892953001
<b>Proteinase K</b>	Sigma Aldrich	P2308
<b>PVDF Membrane</b>	Biorad	1620184
<b>Q5 High-Fidelity Polymerase 2x Master Mix</b>	NEB	M0492S
<b>Rabbit TrueBlot HRP</b>	Rockland	18-8816-33
<b>RLT Lysis Buffer</b>	Qiagen	79216
<b>RNase A</b>	Peqlab	12-RA-03
<b>RNase I (Ambion)</b>	Invitrogen	AM2294
<b>RNase Inhibitor, murine</b>	NEB	M0314L
<b>SDS MOPS Running Buffer (20x)</b>	Invitrogen	NP0001
<b>SeaKem GTG Agarose</b>	Lonza	50074
<b>SpeI-HF</b>	NEB	R3133S
<b>SYBR Safe</b>	Invitrogen	S33102
<b>T4 Ligase (1U/<math>\mu</math>l)</b>	Invitrogen	15224-041
<b>T4 Ligase</b>	NEB	M0202M
<b>T4 PNK</b>	NEB	M0201L
<b>T4 RNA ligase</b>	NEB	M0437M
<b>T7 RNA Polymerase</b>	Roche	1088176700
<b>TD buffer</b>	Illumina	15027866
<b>TDE1 (transposase)</b>	Illumina	15027868
<b>X-Gal (5-Bromo-4-chloro-3-indolyl <math>\beta</math>-D-galactoside)</b>	Sigma-Aldrich	B4252
<b>Yeast RNA</b>	Calbiochem	55714

Table 2.4 Commercial Kits

Kit	Brand	Catalog Number
<b>AffinityScript qPCR cDNA Synthesis Kit</b>	Agilent	600559
<b>CellTrace CFSE Cell Proliferation Kit</b>	Invitrogen	C34554
<b>Chromium Single Cell 3' Library &amp; Gel Bead Kit v2</b>	10x Genomics	PN-120237
<b>Cyto Tune-iPS Sendai Kit</b>	Invitrogen	A16517
<b>Expand Long Template PCR System</b>	Roche	11681842001
<b>Human Genome CGH Microarray 244K</b>	Agilent	G4411B
<b>InnuPREP RNA Mini Kit</b>	Analytic Jena	845-KS-2040250
<b>MinElute PCR Purification Kit</b>	Qiagen	28004
<b>Nextera XT DNA Library Preparation Kit</b>	Illumina	FC-131-1024
<b>NucleoSpin PCR- and Gel purification Kit</b>	Macherey-Nagel	740609
<b>NucleoSpin Plasmid</b>	Macherey-Nagel	740588
<b>NucleoBond Xtra Maxi</b>	Macherey-Nagel	740414
<b>pGEM-T Vector System I</b>	Promega	A3600
<b>Pierce BCA Protein Assay Kit</b>	Thermo Scientific	23227
<b>ProtoScript II First Strand cDNA Synthesis Kit</b>	NEB	E6560L
<b>QIAquick PCR purification Kit</b>	Qiagen	28104
<b>RNA Clean and concentrator-5</b>	Zymo	R1016
<b>TruSeq RNA Library Prep Kit v2</b>	Illumina	RS-122-2001
<b>Turbo DNase Kit</b>	Invitrogen	AM1907M

Recipes of protocol-specific buffers and antibodies can be found in the respective protocol sections.

### 2.3 Tissue Culture Procedures

Cells were cultured at 37°C in a humid incubator with 5% CO<sub>2</sub> and treated under a laminar flow cell culture hood to guarantee sterile conditions. Moreover, antibiotics and antimycotics were added to the media and they were sterile filtered before use.

### 2.3.1 Cell lines

Wildtype hiPSC line B (WT#1) was kindly provided by the Center for Physiology and Pathophysiology at the Institute for Neurophysiology, Medical Faculty, University of Cologne, Germany and wildtype hiPSC lines S24 (WT#2) and T12 (WT#3) by the Pediatric Endocrinology and Diabetology laboratory at the Department of Pediatrics, University Clinic Carl Gustav Carus, TU Dresden, Germany.

Clonal hiPSC lines with CRISPR/Cas9 mediated deletions were generated from one of the previous WT hiPSC lines. Heterozygous SNPs occurring within the parental WT hiPSC line were used to confirm the presence of heterozygous deletions following CRISPR/Cas9 genetic engineering.

As described in Laugsch et al. (2019), patient-specific hiPSC were derived from fibroblasts. Use of the BOFS patient's tissue for this study was approved by his parents and the ethics committee at the University of Southampton - where the patient was enrolled and sampled - and happened in accordance with the Declaration of Helsinki regarding Ethical Principles for Medical Research Involving Human Subjects.

### 2.3.2 Tissue culture media and reagents

Table 2.5 Tissue culture reagents and commercial media

Reagent	Brand	Catalog Number
<b>Accutase</b>	Sigma Aldrich	A6964
<b>Antimycotic/Antibiotic solution</b>	Hyclone	SV30079.01
<b>B27 Supplement (50x)</b>	Gibco	17504044
<b>BSA (Bovine Serum Albumin)</b>	Gemini Bio-Products	GEM 700-104-PJ
<b>DMEM-F12, GlutaMAX</b>	Gibco	10565018
<b>DMSO (Dimethyl Sulfoxide)</b>	Sigma Aldrich	D2650
<b>Doxycycline Hyclate</b>	Sigma Aldrich	D9891
<b>EGF Supplement</b>	Peprtech	AF-100-15
<b>bFGF</b>	Peprtech	100-18B
<b>FuGene HD Transfection Reagent</b>	Promega	E2313
<b>G418 (Neomycin)</b>	AppliChem	A6798
<b>Geltrex LDEV-Free, hESC-Qualified</b>	Gibco	A1413302
<b>GlutaMAX Supplement</b>	Gibco	35050061
<b>Insulin</b>	Gemini Bio-Products	700-1121P

<b>KnockOut DMEM</b>	Gibco	10829018
<b>KnockOut Serum Replacement (KSR)</b>	Invitrogen	10828028
<b>Leukemia Inhibitory Factor (LIF)</b>	Sigma-Aldrich	ESG1106
<b>MesenCult ACF Chondrogenic Differentiation Medium</b>	StemCell Technologies	05455
<b>mTeSR1</b>	StemCell Technologies	85850
<b>N2-Supplement (100x)</b>	Gibco	17502048
<b>Neurobasal Medium</b>	Gibco	21103049
<b>Puromycin</b>	AppliChem	A2856
<b>StemMACS iPS-Brew XF</b>	Miltenyi Biotec	130-104-368
<b>Thiazovivin</b>	Axon Medchem	Axon1535
<b>Versene</b>	Gibco	15040066

Table 2.6 Composition of Tissue Culture Media

Medium	Component	Final Concentration
<b>hiPSC Freezing Medium (2x)</b>	mTeSR1 or Brew	60%
	DMSO	10%
	KSR	30%
<b>NCC differentiation medium</b>	Neurobasal	0.5x
	DMEM F12, GlutaMAX	0.5x
	N2 Supplement	0.5x
	B27 Supplement	0.5x
	GlutaMAX	0.5x
	Antimycotic/Antibiotic solution	1x
	bFGF	20 ng/ml
	EGF	20 ng/ml
	Insulin	5 µg/ml
<b>NCC maintenance medium</b>	Neurobasal	0.5x
	DMEM F12, GlutaMAX	0.5x
	N2 Supplement	0.5x
	B27 Supplement	0.5x
	GlutaMAX	0.5x
	Antimycotic/Antibiotic solution	1x
	bFGF	20 ng/ml
	EGF	20 ng/ml
	BSA	1 mg/ml
<b>SMC differentiation medium</b>	DMEM F12, GlutaMAX	90%
	FBS	10%

<b>Neuron/Glia differentiation medium</b>	DMEM F12, GlutaMAX	1x
	B27 Supplement	1x
	Glutamine	2 mM
	LIF	50 ng/ml
	FBS	1%

### 2.3.3 hiPSC culture

hiPSC were typically cultured on plates coated with Geltrex using StemMACS iPS-Brew XF medium supplemented with Antimycotic/Antibiotic solution. When cells were freshly transfected, they were cultured in mTeSR1 medium supplemented with Antimycotic/Antibiotic solution (see 2.3.5).

Confluent hiPSC were passaged with Versene or Accutase using a 1:5 to 1:20 dilution. ROCK inhibitor Thiazovivin was added to the medium for the first 24 hours after splitting (and thawing). hiPSC were frozen in StemMACS iPS-Brew XF mixed with freezing medium (2x) (see Table 2.6) and stored at -150°C.

### 2.3.4 Derivation of patient-specific hiPSC

As described in Laugsch et al. (2019), patient-derived fibroblasts – provided by the Wessex Regional Genetics Laboratory, Salisbury, UK – were reprogrammed into hiPSC using standard Yamanaka factors (OSKM) by non-integrating Cyto Tune-iPS Sendai Kit (Seki et al., 2012). hiPSC were then validated for pluripotent marks (OCT4, NANOG) by immunofluorescent staining (see 2.7.2). Furthermore, cytogenetic analysis (G-banding) was performed to confirm a normal karyotype, except for the known 89 Mb inversion of chromosome 6. Subsequently, the inversion was confirmed in hiPSC by PCR (see 2.5.2) using primers listed in Table 2.7.

Table 2.7 PCR primers for 89 Mb inversion genotyping

Primer	Sequence 5'-3'
<b>BOFS P1</b>	CTCTCTTTGCCCTCTGTGGA
<b>BOFS P2</b>	GAGAGGCGACCCGGAAGTGA
<b>BOFS P3</b>	ACGGAAAAGCGGGGACTGTG
<b>BOFS P4</b>	GGGCAGTTATGTGGAGGAGA

### **2.3.5 Transfection of hiPSC**

DNA was delivered into hiPSC by non-liposomal transfection with FuGene HD Transfection Reagent (Promega) during passaging. Briefly, a total of 1 µg plasmid DNA was mixed with 100 µl of mTeSR1 medium without supplements and 4 µl of transfection reagent, followed by 15 minutes of incubation at room temperature (RT) with occasional flicking of the tube.

Meanwhile, hiPSC were detached from the tissue culture plates using Versene and approximately 350.000 cells were re-suspended in 1 ml of mTeSR1 with Thiazovivin. Cells were then mixed with each transfection reaction and transferred to one well of a 12-well plate.

After 24 hours, medium was changed and, if applicable, GFP expression examined by fluorescent microscopy. Once GFP was detectable, transfected cells were selected by adding antibiotics (puromycin or neomycin/G418) to the medium either for 24-48 hours (for transient transfections, e.g. for CRISPR/Cas9 vectors see section 2.8.3) or permanently (when using integrative PiggyBAC vectors, e.g. for reporter assays and rescue experiments see sections 2.6.2 and 2.8.5, respectively).

Once hiPSC had recovered from the antibiotic selection the medium was gradually changed from mTeSR1 to StemMACS iPS-Brew XF medium.

### **2.3.6 Generation of clonal hiPSC lines**

To generate clonal hiPSC lines (e.g. after CRISPR/Cas9 engineering), the targeted hiPSC population was split several times with Accutase at high density and eventually, when cells had adapted to single cell splitting, seeded in Brew with Thiazovivin in 96-well plates at a density of 1 cell per well. Cells were grown for about two weeks and medium was changed every three to four days until colonies were big enough to be split. During splitting with Accutase into 48-well plates, half of the cells were taken for DNA extraction to test the genotype of individual clones by PCR and subsequent Sanger sequencing (see methods section 2.8.4). Once the presence of desired deletions and clonality were confirmed, cells were expanded and used for downstream analysis.

### 2.3.7 hNCC differentiation and short term maintenance

Differentiation of hiPSC into hNCC was performed as previously described (Bajpai et al., 2010; Prescott et al., 2015; Rada-Iglesias et al., 2012). Briefly, hiPSC were grown in 6-well plates until confluent, then treated with 2 mg/ml collagenase in KO DMEM medium at 37°C for 1-2 hours until colonies detached. Cell aggregates were then carefully washed with PBS and finally plated in Petri dishes in NCC differentiation medium (see Table 2.6).

Embryoid body (EB) formed about twelve hours later and spontaneously attached to the surface of the Petri dish after six to eight days. Attached EBs then gave rise to hNCC outgrowths. Medium was changed every two to three days.

On day 11 of differentiation, hNCCd11 were either harvested for downstream analysis or split for short term maintenance. For short term maintenance, any remaining floating EBs were removed by aspiration and hNCC were dissociated with Accutase and seeded (50.000 cells per cm<sup>2</sup>) on a fibronectin (5 mg/ml) coated cell culture dish in hNCC maintenance medium (see Table 2.6), in which Insulin was replaced by BSA (Prescott et al., 2015). After another passage with Accutase, hNCCp2 were harvested for downstream analysis.

### 2.3.8 Differentiation of NCC derived lineages

To test the differentiation capacity of hNCCp2, three different protocols were used as described in Laugsch et al. (2019).

#### 2.3.8.1 Smooth Muscle cells

For differentiation into smooth muscle cells (SMC), hNCCp2 were cultured in SMC differentiation medium (see Table 2.6) for seven days with medium change every 2-3 days. Then, cells were analyzed by immunofluorescence (see methods section 2.7.2) with an antibody against the smooth muscle marker SM22-alpha (Abcam).

#### 2.3.8.2 Neurons/glia cells

For differentiation into neurons/glia cells, hNCCp2 were cultured in neurons/glia cell differentiation medium (see Table 2.6) for 14 days with medium change ever 2-3 days. Then, cells were analyzed by immunofluorescence (see methods section 2.7.2) with an antibody against the neural marker TUJ1 (Abcam).

### 2.3.8.3 Chondrocytes

For differentiation into chondrocytes,  $5 \times 10^5$  hNCCp2 were pelleted by centrifugation and then cultured at the bottom of a 15 ml polypropylene tubes with 0.5 ml of chondrocyte differentiation medium (see Table 2.6). Three days later, pellets had rounded up and 0.5 ml of medium was added. Hereafter, half of the medium was changed and tubes were flicked to keep cells from attaching to the surface of the tube every three days, up to a total of 21 days.

On day 21 chondrocytes were fixed with 10% formalin for 30 minutes at RT and embedded in paraffin for subsequent Alcian Blue staining (see 2.3.9) to identify cells that had successfully differentiated into chondrocytes.

### 2.3.9 Alcian Blue staining

As described in Laugsch et al. (2019), chondrocytes were fixed in 10% formalin for 30 minutes at RT and then embedded in paraffin. Afterwards,  $8 \mu\text{m}$  sections of the embedded tissue were made with a microtome and placed on microscope slides. Sections were then washed with xylene twice for 5 minutes to deparaffinize them. Subsequently, washes with decreasing amounts of ethanol (100%-70%) and a final rinse under running tap water for 5 minutes each were performed to rehydrate the sections.

Eventually, sections were incubated in 1% Alcian Blue solution (pH 2.5) in acetic acid for 30 minutes. The slides were then rinsed under running tap water, counterstained with Nuclear Fast Red solution for 5 minutes at RT and rinsed again. Subsequently, sections were dehydrated once again by washes with increasing amounts of ethanol (50%-100%) and a final wash with acetone. Slides were then mounted using Aquatex Permanent Aqueous Mounting Agent.

Sections were imaged using a bright field stereomicroscope (Olympus SZX16).

### 2.3.10 Scratch Assay

Initially, scratch assays were performed with silicon inserts, as described in Laugsch et al. (2019).  $1.5 \times 10^5$  hNCC were seeded in every well of a fibronectin-coated 24-well plate with silicon culture inserts (ibidi). The inserts limit the growth of the cells to certain areas while cell-free gaps are created. 24 hours after passaging, inserts were carefully removed, medium changed and

migration of hNCC into the cell-free gap was monitored for the following 24 hours with an EVOS FL Auto Cell Imaging System and Onstage Incubator Life, taking photos of the process every two hours.

Additional scratch assays were performed manually in 6-well plates to confirm the results of the initial scratch assays and for further quantification. Here, one scratch per well was made with a plastic pipette tip and migration was monitored at 0 h, 10 h and 22 h by bright field microscopy.

Representative images of the gaps were analyzed using KNIME, OMERO and Photoshop CS2.

For quantification of the differences in cell migration between WT and BOFS hNCCp2, the gaps/scratch areas were measured using FIJI (ImageJ) (Schindelin et al., 2012) ROI tools. The borders of the gaps were manually outlined in images taken at 0 h, 4 h and 10 h and the size of the areas were normalized by division of the area at 4 h and 10 h by the area at 0 h. Student's t-test was applied to evaluate the statistical significance of the difference between WT and BOFS hNCC regarding their capacity to close the cell-free gap.

### **2.3.11 CFSE Proliferation Assays**

As described in Laugsch et al. (2019), proliferation assays were performed using the CellTrace CFSE Cell Proliferation Kit. Briefly,  $5 \times 10^6$  of WT and BOFS hNCCp2 were labelled with 10  $\mu$ M Carboxyfluorescein Succinimidyl Ester (CFSE) in 1 ml PBS by incubation at 37°C for 10 minutes. As negative controls, wells were treated equally in 1 ml PBS without dye. The reaction was stopped by adding 5 ml of ice-cold hNCC maintenance medium and incubating the mixture for 5 minutes on ice. Labelled cells were then centrifuged, washed twice with fresh medium and finally seeded on a fibronectin coated 6-well plate. Samples were collected every 24 hours for 4 days and analyzed using flow cytometry with 488 nm excitation.

Cell Proliferation is inversely proportional to the Mean Fluorescent Intensity (MFI) of CFSE. Therefore, we measured the average MFI of each sample and calculated the Fluorescent Dilution Factor (FDF) as a measure of the proliferation rate according to Griessinger et al. (2016) by dividing the MFI measured on day 1 by the MFI measured on days 2, 3 and 4.

### 2.3.12 Cell proliferation growth curves

10<sup>5</sup> cells (hiPSC or hNCCp2; WT, DDX1<sup>+/-</sup>, DDX1<sup>-/-</sup>::TetON-*DDX1* or MYCN<sup>-/-</sup>) were seeded on a fibronectin coated plate and cultured with (100 ng/ml) or without DOX. For each sample and condition three wells were maintained for each measurement. Samples were collected every 24 hours for eight days and counted on an automated cell counter (Biorad). Measurements were done in four replicates per well.

## 2.4 Clinical and genetic characterization of the BOFS patient inversion

### 2.4.1 Patient description

As described in Laugsch et al. (2019), the patient is a 17 year-old son of non-consanguineous healthy parents (Figure 3.1.1 and Table 2.8). At birth, by normal delivery at 38 weeks gestation, he weight 3.71 kg and measured 35 cm head circumference. At the age of two, there were concerns about his development, especially about his speech and locomotion skills.

The facial features of the patient were dysmorphic with long palpebral fissures, a flattened nasal tip, full lips, a short philtrum and upturned ear lobes as well as a high arched palate and a sector of heterochromia in his left iris. There were no hairy patches and no hemangiomas or unusual patches of skin on his neck, as sometimes observed in BOFS patients. At the age of four, the patient's speech and language difficulties had intensified, he had begun to squint, suffered from severe sensorineural hearing loss in his right ear and fluctuating conductive hearing loss secondary to middle ear effusion in his left ear. In addition, he had palatopharyngoplasty surgery for velopharyngeal insufficiency.

He reported significant difficulties with self-organization and social interactions but did not meet the ICD 10 criteria for diagnosis of autism when assessed at seven years of age. In summary, the patient meets the criteria to be classified as a BOFS case, although his phenotype is mild. Especially the syndromic cleft lip is only present as a microform.

His clinical features contributing to the complex characteristics of BOFS are listed in Table 2.8.

Table 2.8 Clinical features of the BOFS patient

Reported BOFS feature	Present Y/N	Comment about patient
<b>Branchial anomaly</b>	N	
<b>Ocular anomaly</b>	Y	Alternating convergent squint and pooling of tears left eye
<b>Characteristic facial features</b>	Y	
<b>Malformed pinnae</b>	N	
<b>Upturned ear lobes</b>	Y	See Figure 3.1.1
<b>Thick nasal tip</b>	Y	
<b>Cleft lip</b>	Y, microform	Unusual vermillion border with prominent philtral pillars
<b>Cleft palate</b>	N	Pharyngoplasty for decreased palatal function
<b>Ectodermal anomalies</b>	Y	Presence of primary dentition aged 17
<b>Hearing loss</b>	Y	Right-sided severe sensorineural hearing loss and fluctuating left-sided conductive hearing loss secondary to middle ear effusion (hearing aid used)
<b>Abnormal vestibular function</b>	Y	Abnormal test. Poor balance in the dark. No dizziness. MRI Normal
<b>Kidney anomalies</b>	N	Normal USS
<b>Supernumerary nipples</b>	Y	x 2

## 2.4.2 Characterization of chromosomal abnormalities

As described in Laugsch et al. (2019), initial genome analysis was performed by array-CGH with the Agilent 244K oligonucleotide array according to the manufacturer's instructions.

The inversion breakpoints were mapped in metaphase chromosomes by two-colored FISH using probes from BAC, PAC and fosmid constructs according to standard protocols. Briefly, the 6q16.2 breakpoint was mapped to a genomic region containing no known genes with BAC RP11-390H11 and the 6p24.3 breakpoint was mapped to a region containing *TFAP2A* with PAC RP1-290I10. The location of the 2<sup>nd</sup> breakpoint was then further refined with fosmid WI2-506N5 to exclude *TFAP2A* from the immediate breakpoint region.

Nucleotide resolution of the inversion breakpoints was achieved by paired-end sequencing using 100 bp reads on a single lane of the Illumina HiSeq2500.

### 2.4.3 Targeted Locus Amplification (TLA)

As described in Laugsch et al. (2019), TLA was performed as a service from Cergentis (<https://www.cergentis.com>) according to their standard protocol (de Vree et al., 2014). Briefly, patient-derived hiPSC were crosslinked, fragmented and re-ligated and then analyzed with four sets of primers (Table 2.9)

Table 2.9 Primers used for TLA

Primer	Sequence 5'-3'
<b>1A</b>	fwd: AGTATGTGTGAGTGTGTGGG rev: TCGCAGATGTTAGTTCACAGT
<b>1B</b>	fwd: CCATCTCTAGGGTTTGAGGTATT rev: TCTTTGCAGAGAATATAGCCAGA
<b>2</b>	fwd: AGGGCCAGTTCCTTCTCTTG rev: CGAAAGTCTGTGCCCTTCGTT
<b>3</b>	fwd: AACATTGCACTCTTATAGCTGA rev: GGCACTTCAGAATATACTTGC

Primer-sets 1A and 1B were used to amplify both *TFAP2A* alleles in order to (i) phase rs1675414, a heterozygous SNP in the BOFS patient cells, with respect to the inversion, (ii) identify additional heterozygous SNPs in the BOFS patient cells and (iii) to phase them with respect to the inversion. Primer-sets 2 and 3 were designed upstream of the 6p24 and downstream of the 6q16 inversion breakpoints, respectively, in order to confirm and refine the position of the two inversion breakpoints. The primer sets were used in individual TLA amplifications. PCR products were purified, Illumina NGS libraries prepared according to the manufacturer's protocol and sequenced on an Illumina sequencer.

Reads were mapped to the human genome (hg19) using BWA-SW (Li and Durbin, 2009), allowing partial mapping which is optimally suited for identifying reads that span breakpoints.

Firstly, the TLA data was used to confirm and refine the breakpoints of the inversion identified in our patient:

6p24 breakpoint: *chr6:99,103,875 (rev)* fused to ***chr6:10,355,280 (fwd)***, with 3bp of homologous sequence at the breakpoint.

CCTCTGAGCTGCAGTTGACAACGTTACAGTTGGTAGCTCTTACAGTGGCCTAGCTCCTGGG  
 CTGTTTGAGGTAAAATGATATGAAATGGATTGAAGCAAGAAGGTAACAAGCATGATTCCC  
**AGCATTGTGGGGGTAACCTTGTACTTTTT**

6q16 breakpoint: *chr6:99,103,872 (fwd)* fused to **chr6:10,355,274 (rev)** with a 7bp insertion

CAAAATCTCAGTGTCTTCAAACAACAGATTTATTTCTTATTACATTATCTGTCCATTATAGG  
 GCAGATGAGCCCTATAATGTATTTGGTAAGCACTTTAGAAGCCTTTAAAATAGTACTCTAA  
**GGCAGGTAACGTGGTGTGCCTCTAGT**

Next, the TLA data was used to identify heterozygous SNPs and link them to the inversion allele in the BOFS patient. Therefore, SNP calling was performed 100kb upstream and downstream of the inversion breakpoints and heterozygous SNPs were selected at positions with  $\geq 30X$  coverage and with a minor allele frequency of 20-80 %. Assigning SNP alleles to either the wild-type (WT) or inversion (INV) allele was by usage of different primer-sets amplifying different alleles. The linkage analyses on the heterozygous SNPs was performed in four sub-regions:

- i. SNP-region 1 is the region upstream of the 6p24.3 breakpoint. In this region, SNPs were called in the data of primer-set 2 that amplified both alleles of this region. This resulted in the detection of 10 heterozygous SNPs in region 1. Linkage of these SNPs to the inversion was performed using the data of primer-sets 1A and 1B, knowing that these two primer-sets only have coverage of the WT allele in this region. All 10 SNPs were indeed homozygous in the 1A and 1B data set, and both 1A and 1B showed the same linkage results for all positions. So, all 10 SNPs were phased relative to the inversion.
- ii. SNP-region 2 is the region downstream of the 6p24.3 breakpoint. In this region, SNPs were called in the data of primer-set 1A and 1 that each amplified both alleles in this region. In total, 79 heterozygous SNPs were detected in region 2. Linkage of these SNPs to the inversion was determined using the data of primer-sets 2 and 3, knowing that primer-set 2 shows coverage on the WT allele, and primer-set 3 shows coverage on the INV allele. In total 71 heterozygous SNPs were phased with the inversion in region 2. Among these is the rs1675414 SNP at position chr6:10,412,188, for which the TLA data showed that the “T” allele is found on the INV chr6, whereas the “C” allele is found on the WT chr6. This information is useful later for allele-specific analysis of *TFAP2A* expression (2.5.6).
- iii. SNP region 3 is the region upstream of the 6q16.2 breakpoint. In this region, SNPs were called in the data of primer-set 2 and 3. There were no SNPs that met the coverage

requirement in both primer-sets, so no SNPs could be linked to either WT or INV in region 3.

- iv. SNP region 4 is the region downstream of the 6q16.2 breakpoint. In this region, SNPs were called in the data of primer-set 3 that amplified both alleles of this region. 55 heterozygous SNPs were detected in region 4. Linkage of these SNPs to the inversion was performed using the data of primer-sets 1a and 1b, knowing that these two primer-sets only have coverage of the INV allele in this region. In total, 52 heterozygous SNPs were phased relative to the INV in region 4.

Altogether, 133 heterozygous SNPs have been specifically linked to either the WT or the INV allele in our BOFS patient cells (see Laugsch et al. (2019) Table S2 for details).

## 2.5 General molecular biology methods

### 2.5.1 Genomic DNA extraction

Extraction of DNA for genotyping by PCR with the Platinum Taq Polymerase System was achieved by adding QuickExtract DNA Extraction Solution (see Table 2.10 for details) to the sample, incubating the DNA at 65°C for 6 minutes and heating it up to 98°C for 2 minutes. As soon as the tube was cooled back down to RT, DNA was ready for PCR. Volumes of QuickExtract Solution varied depending on the sample size, typically using 1 ml for 3 million cells and scaling down accordingly for less cells.

Table 2.10 Composition of QuickExtract DNA Extraction Solution

Name	Component	Final Concentration
<b>QuickExtract DNA Extraction Solution</b>	KCl	50 mM
	TRIS pH 8.3	10 mM
	MgCl <sub>2</sub>	2.5 M
	NP40	0.45%
	Tween20	0.45%
	Proteinase K	0.4 mg/ml

## 2.5.2 PCR for genotyping

PCR was performed using the Platinum Taq DNA Polymerase according to the manufacturer's instructions with Promega dNTP mix. For a 25  $\mu$ l reaction 1-2  $\mu$ l of DNA in QuickExtract Solution was used as input.

A list of primers used for PCR can be found in each corresponding section where PCR was applied (Table 2.7, Table 2.14, Table 2.28). Primers were designed with Primer3 for a melting temperature of 60°C. PCRs were performed using the parameters listed in Table 2.11.

Table 2.11 PCR cycling program

Step	Temperature	Time
<b>Initial denaturation</b>	94°C	2 minutes
<b>39 cycles</b>	<b>Denature</b>	94°C
	<b>Anneal</b>	62°C
	<b>Extend</b>	72°C
<b>Final extension</b>	72°C	3 minutes
<b>Hold</b>	12°C	indefinitely

Afterwards, PCR products were separated on a 1-2% agarose gel, depending on the size of the product. The agarose gel was prepared and run with 1x TAE buffer and stained with Midori Green Advanced DNA dye. PCR products were loaded with 6x Orange G solution and run alongside the Gene Ruler 1 kb Plus DNA ladder to determine the size of the PCR products. Gels were visualized and pictures taken with the ChemiDoc MP Imaging System (Biorad).

Table 2.12 Buffers and Solutions for Agarose Gel Electrophoresis

Solution/Buffer	Component	Final Concentration
<b>TAE Buffer (50x)</b>	Tris pH 8.6	2 M
	Acetate	1 M
	EDTA	50 mM
<b>Orange G (6x)</b>	TAE Buffer	6x
	Glycerol	60% (v/v)
	Orange G dye	0.15% (w/v)

### 2.5.3 RNA extraction

Total RNA was isolated from cultured cells, usually from one confluent well of a 6-well plate (ca. 1-3 million cells), using the innuPREP RNA Mini Kit according to the manufacturer's instructions. hiPSC or hNCC were lysed immediately in the dish with 400 µl of RL lysis buffer, scraped off the surface of the cell culture plate with a cells scraper (Sarstedt) and frozen at -20°C until column purification. Extracted RNA was kept at -80°C.

### 2.5.4 cDNA synthesis

cDNA was generated by reverse transcription of total RNA using the ProtoScript II First Strand cDNA Synthesis Kit (NEB) according to the user manual. For general gene expression analysis by RT-qPCR oligo-dT primers were used to amplify the mRNA-specific poly-A tail.

### 2.5.5 RT-qPCR

Real-time quantitative PCR was performed on the Light Cycler 480 II (Roche) with the ORA qPCR Green ROX L Mix (HighQu). Gene expression fold changes were calculated by the delta-delta Ct method. *ACTB*, *EEF2* and *GAPDH* were used as housekeeping genes. A list of primer sequences used for RT-qPCR can be found in Table 2.13.

Table 2.13 Primer Sequences for RT-qPCR

Gene	Primer Sequence (5' -> 3') forward	Primer Sequence (5' -> 3') reverse
<b>TFAP2A</b>	GCCTCTCGCTCCTCAGCTCC	CGTTGGCAGCTTTACGTCTCCC
<b>FAM20</b>	GAGATTGCTGCCTTCCACCT	CTCCAGAGCTTCTTGTCCCCG
<b>GDF6</b>	CCCACGAGTACATGCTGTCA	GAGGAGTGTGCGAGAGATCG
<b>EGF</b>	AGTGGCTGAAATCATGGTGTCA	TCCATCCCCAGCAAATCCTTT
<b>JAG1</b>	CGCGTGACCTGTGATGACTA	ACCTGGGAGTTTGCAAGACC
<b>FREM2</b>	TCGTCTGAATTGGGCATGGA	GCAGTTCTGTCTCTTGTGCC
<b>KMT2D</b>	TGGGGATCTTCATGCTCAGC	TGGAAAGCTGCCTGTCAACA
<b>HYLS1</b>	CAGAATCTGCGGTGCTCTG	CCATTGCTTCCCCTACACTGT
<b>CDH2</b>	CAGCTCCACCTCGAAGTACA	CGTTGGCAGCTTTACGTCTC
<b>EEF2</b>	CTATCTGCCCGTCAACGAGT	GATCTGCCAGTGGTCAAACA
<b>ACTB</b>	TCAAGATCATTGCTCCTCCTGAG	ACATCGCTGGAAGGTGGACA
<b>GAPDH</b>	GCACCGTCAAGGCTGAGAAC	AGGGATCTCGCTCCTGGAA
<b>NANOG</b>	ACAACCTGGCCGAAGAATAGCA	GGTTCCCAGTCGGGTTCCAC

<b>OCT4</b>	GGGTTTTTGGGATTAAGTTCTTCA	GCCCCCACCCTTTGTGTT
<b>DDX1</b>	TGGCTTTGTTGCTCTTTCCAAG	CGGAGCATTGGGGAGAAACT
<b>MYCN</b>	CACAAGGCCCTCAGTACCTC	TTGGTGTGGAGGAGGAACG
<b>FAM49A</b>	TTGGGTTGGTGTGTAGGGTG	AAAGCTTGAATGCGGTGTG

### 2.5.6 Allele-specific analysis of TFAP2A expression

As described in Laugsch et al. (2019), total RNA, isolated as described above (2.5.2), was treated with Turbo DNase (Invitrogen) to remove any remaining DNA. cDNA was then synthesized with the ProtoScript II First Strand cDNA Synthesis Kit (NEB) using random primers for amplification of both nascent and processed RNA. Reference DNA was extracted as described above (2.5.1). The heterozygous SNP rs1675414, identified by TLA (2.4.3), and located within the first exon of *TFAP2A*, was then amplified by PCR (2.5.2) with primers listed in Table 2.14. Subsequently, PCR products were purified using NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) and subject to Sanger sequencing (Microsynth Seqlab) with the reverse primer.

Table 2.14 Primer Sequences for TFAP2A SNP genotyping

Primer	Sequence 5'-3'
<b>TFAP2A_rs1675414_fwd</b>	TTGCTACTGAGACTCCCG
<b>TFAP2A_rs1675414_rev</b>	AACTTGAACCACCGATTTCGC

Quantitative allele ratio analysis was performed using the PeakPicker software (Ge et al., 2005). The rs1675414 SNP was genotyped in genomic DNA (gDNA) and cDNA from hNCC (hNCCd11 and hNCCp2) derived from two WT and three BOFS patient hiPSC lines, respectively. For each cell line, five genotyping measurements were performed in gDNA and cDNA obtained from at least two independent hNCC differentiations.

### 2.5.7 Chemical Competent Cells

To make chemically competent *E. coli* cells, we used the Top10 strain, generously gifted by the Kurian Lab, CMMC, Cologne. Bacteria were grown in 3 ml LB medium without antibiotics at 37°C (shaking) for 7-8 hours during the day. All 3 ml were then transferred into 250 ml LB-medium without antibiotics and grown overnight at 37°C while shaking.

The next day, 10 ml of the overnight culture were transferred into 500 ml fresh LB-medium without antibiotics and grown for about 2 hours at 37°C (shaking) until OD<sub>600</sub> reached a value between 0.3 and 0.4. Cells were then pelleted (10 minutes, 4°C, 5000 rpm), re-suspended in a total of 250 ml cold 0.1M CaCl<sub>2</sub> (sterile filtered) and incubated on ice for 20 minutes. After another round of centrifugation as described above, the supernatant was discarded and cells were re-suspended in a total of 5 ml cold CaCl<sub>2</sub>. Finally, Glycerol was added to a final concentration of 10% and *E. coli* were aliquoted and snap frozen in liquid N<sub>2</sub> to be stored at -80°C.

Efficiency of competent cells was tested by transformation with a plasmid stock dilution of 10 pg/μl, 100 pg/μl and 1000 pg/μl according to the transformation protocol (see 2.5.8).

## 2.5.8 Molecular cloning

Cloning was generally performed by amplification of genomic DNA or cDNA by PCR with proofreading Q5 High-Fidelity DNA Polymerase (NEB) and primers with added restriction sites. Full length *MYCN* cDNA was amplified with AccuPrime GC-rich Polymerase (Invitrogen), due to its high GC-content. PCR products and vectors were treated with restriction enzymes overnight at 37°C, followed by column purification with the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) and ligation with T4 ligase (NEB) overnight at 4°C.

A list of empty plasmid vectors used for various applications within this work can be found in Table 2.15. Additionally, the pGEM-T vector system I (Promega) was used for cloning of PCR products without added restriction sites but with A-overhangs (see 2.8.5).

The primers and restriction enzymes used for generating all of the vectors described in this work are listed in the corresponding methods sections (2.6.1, 2.8.2, 2.8.5, 2.9.6).

Table 2.15 Plasmids

Name	Application	Source	Order Number
<b>KA0717_pPB-hCMV*1-cHA-IRESVenus</b>	cDNA Expression in vitro	Addgene	124168
<b>KA0637_pPBCAG-rtTAM2-IN</b>	rtTA transactivator	Addgene	124166
<b>Super PiggyBAC Transposase expression vector</b>	Integration of PiggyBAC vector into genome	Systems Biosciences	PB200PA-1
<b>PiggyBac-CMV-MCS-EF1-RedPuro</b>	a) empty: electroporation control <i>in vivo</i>	Systems Bioscience	PB514B-2

	b) with cDNA: cDNA expression <i>in vivo</i>		
<b>PiggyBAC enhancer GFP Neo</b>	<i>In vitro</i> reporter	Buecker et al. (2014)	N/A
<b><math>\beta</math>-Globin-GFP</b>	<i>In vivo</i> reporter	Bergsland et al. (2011)	gifted
<b>pX330A-hCas9-long-chimeric-gRNA-GFP</b>	CRISPR/Cas9	Kurian Lab, CMMC, Cologne	gifted

## Transformation of *E.coli*

Chemical competent cells were thawed on ice for 30 minutes. 3  $\mu$ l of a 10  $\mu$ l ligation was added to 30  $\mu$ l of bacteria in a 1.5 ml tube and incubated on ice for 15 minutes. Heat shock was performed at 37°C for 1 minute followed by incubation on ice for 1 minute. Subsequently, 1 ml of LB medium was added to the reaction and cells were incubated at 37°C for 1 hour with constant shaking at 900 rpm. Afterwards, the tube was centrifuged at 1000 rpm for 1 minute at RT, the supernatant was removed up to 100  $\mu$ l and the bacteria were re-suspended in the remaining LB medium in order to be spread on an LB-Agar plate with appropriate antibiotic selection (e.g. Ampicillin). Plates were then incubated at 37°C overnight (14-18 hours) and colonies were picked for colony PCR or inoculation of LB-medium overnight culture for DNA Mini Preps.

## Plasmid DNA extraction from bacteria

Plasmid DNA for transfection of hiPSC and electroporation of chicken embryos was extracted using the NucleoSpin Plasmid or the NucleoBond Xtra Maxi Kit (Macherey-Nagel), respectively, following the manufacturer's protocol. For transfection of hiPSC (see 2.3.5), a total of 1  $\mu$ g plasmid DNA was used in a maximal volume of 5  $\mu$ l. For electroporation of chicken embryos (see 2.9.4), plasmid DNA was used at a concentration of 10 to 18  $\mu$ g/ $\mu$ l. Therefore, DNA was solved in as little H<sub>2</sub>O as possible, usually 100-300  $\mu$ l, and concentrated with a spin concentrator (Eppendorf) whenever necessary.

## 2.6 Enhancer reporter assays

### 2.6.1 Enhancer cloning

As referred to in Laugsch et al. (2019), for *in vivo* reporter assays in chicken embryos, a previously described  $\beta$ -Globin-GFP enhancer reporter system (Bergsland et al., 2011) was used while for *in vitro* reporter assays in hNCC, a previously described PiggyBAC transposon-based enhancer

reporter system (Buecker et al., 2014) was used. Enhancer sequences (Enh100, Enh105, Enh480, and Enh2p24.2) were amplified from human genomic DNA extracted from WT hiPSC using the Q5 High-Fidelity DNA Polymerase (NEB) with primers listed in Table 2.16. Restricted and purified PCR products (see Table 2.16 for respective restriction enzymes) were cloned into the corresponding vectors with T4 ligase (NEB) as described in section 2.5.8.

**Table 2.16** Primer sequence for enhancer reporter assays

Enhancer	Sequence 5'-3'	Restriction site	Product size
<b>Enh100 for <math>\beta</math>-Globin</b>	Fwd: AAAAAAGGTACCAGTTCTTCTCTTTTCCTCCCTT	KpnI	1955
	Rev: AAAAAAGGATCCGATATTTCTCCTGCCCTGC	BamHI	
<b>Enh100 for PiggyBAC</b>	Fwd: AAAAAAGGATCCAGTTCTTCTCTTTTCCTCCCTT	BamHI	1955
	Rev: AAAAAATTCGAAGATATTTCTCCTGCCCTGC	BstBI	
<b>Enh105 for <math>\beta</math>-Globin</b>	Fwd: AAAAAAGGTACCTGGCAATTTTAGAAGCGGCA	KpnI	450
	Rev: AAAAAAGCTAGCACACATCGAGACCATTTTCATGC	NheI	
<b>Enh105 for PiggyBAC</b>	Fwd: AAAAAAGGATCCTGGCAATTTTAGAAGCGGCA	BamHI	450
	Rev: AAAAAAGAATTCACACATCGAGACCATTTTCATGC	EcoRI	
<b>Enh480 for <math>\beta</math>-Globin</b>	Fwd: AAAAAAGGTACCTCCAACATCATCATGTGCA	KpnI	1269
	Rev: AAAAAAGGATCCGGACACCAGCTTATATCCTGT	BamHI	
<b>Enh480 for PiggyBAC</b>	Fwd: AAAAAAGGATCCTCCAACATCATCATGTGCA	BamHI	1269
	Rev: AAAAAATTCGAAGGACACCAGCTTATATCCTGT	BstBI	
<b>Enh2p24.2 for <math>\beta</math>-Globin</b>	Fwd: AAAAAAAGGTACCAGCAGTTCCACCATACACAC	KpnI	2820
	Rev: AAAAAAAGGATCCTTGAGACGTGGTGAAACTCC	BamHI	
<b>Enh2p24.2 for PiggyBAC</b>	Fwd: AAAAAAAGGATCCAGCAGTTCCACCATACACAC	BamHI	2820
	Rev: AAAAAAATTCGAATTGAGACGTGGTGAAACTCC	BstBI	

### 2.6.2 *In vitro* reporter assays

WT hiPSC were co-transfected with the PiggyBAC enhancer reporter vector and a vector expressing Super PiggyBAC Transposase (see Table 2.15) as described in section 2.3.5 with FuGene HD transfection reagent (Promega). A negative control cell line was created by co-transfection with the empty PiggyBAC enhancer reporter vector and the Super PiggyBAC Transposase expression vector. 24 hours after transfection, hiPSC were constantly treated with 0.2 mg/ml G418/Neomycin to select cells that had stably incorporated the PiggyBAC enhancer reporter vector. Surviving cells were subsequently expanded and, when confluent, differentiated into hNCC as described in section 2.3.7. GFP levels were assessed using a Nikon ECLIPSE TS100 fluorescent microscope.

### 2.6.3 In vivo reporter assays

Chicken embryos of stage HH9-10 were electroporated with  $\beta$ -Globin-GFP enhancer reporter vector as described in section 2.9.4 together with an empty PiggyBac-CMV-MCS-EF1-RedPuro vector constitutively overexpressing RFP as an electroporation control. After electroporation, eggs were sealed with tape and incubated up to stage HH14-16 and HH20-23. GFP and RFP levels were assessed using an Olympus SZX16 stereomicroscope with EXFO X-site series 120PC Q for fluorescence illumination.

## 2.7 Immunological Methods

### 2.7.1 Chromatin Immunoprecipitation (ChIP)

ChIPs were performed in hNCCp2 and chicken frontonasal prominences (FNPs) as previously described (Laugsch et al., 2019; Rada-Iglesias et al., 2012; Rehimí et al., 2017). Briefly, cells or tissue were crosslinked with 1% formaldehyde for 10 minutes at RT and the reaction was quenched with 0.125 M glycine. For TFAP2A and DDX1 ChIPs approximately 20 million hNCCp2 were used per reaction, while for histone ChIPs about seven million hNCCp2 were used instead. H3K4me2 ChIP in chicken FNPs was performed with 15 FNPs from HH24 embryos. hNCC were lysed in three sequential lysis buffers (LB1, LB2, LB3) while FNPs were only lysed in LB3. Chromatin of all samples was sonicated in LB3 using the Bioruptor (Diagenode), hNCC in 1.5 ml LB3 for 25 cycles in 15 ml falcons and FNPs in 300  $\mu$ l LB3 for eleven cycles in 1.5 ml tubes, to an average fragment size of 500-2000 bp. Antibodies were then added to the fragmented chromatin in the indicated amounts (see Table 2.17) and incubated overnight at 4°C while 10% of sonicated chromatin was kept as input DNA. TFAP2A and DDX1 ChIPs were performed in duplicates, while all histone ChIPs were performed as single experiments. The next day, 100  $\mu$ l of protein G Dynabeads (Thermo Fisher) magnetic beads were washed with ChIP blocking solution (0.5% BSA in PBS), added to the ChIP reactions and incubated for 4 hours at 4°C to bind the antibodies. Afterwards, the magnetic beads were washed with RIPA wash buffer and bound chromatin was eluted with elution buffer at 65°C. Chromatin was then de-crosslinked at 65°C overnight and subsequently incubated with RNase A and Proteinase K. DNA was purified with phenol-chloroform-isoamyl alcohol (25:24:1), followed by ethanol precipitation. ChIP DNA was finally dissolved in H<sub>2</sub>O.

Table 2.17 Antibodies for ChIP

Antigen	Manufacturer	Catalog Number	Amount
<b>TFAP2A</b>	Santa Cruz Biotechnologies	sc-12726	10 µg
<b>DDX1</b>	Genetex	GTX105205	10 µg
<b>DDX1</b>	Proteintech	11357-1-AP	10 µg
<b>H3K9me3</b>	Active Motif	39161	5 µg
<b>H3K27ac</b>	Active Motif	39133	5 µg
<b>H3K27me3</b>	Active Motif	39155	5 µg
<b>H3K4me2</b>	Active Motif	39141	5 µg

Table 2.18 Buffers for ChIP-seq

Buffer	Component	Final concentration
<b>Lysis Buffer 1 (LB1)</b>	HEPES-KOH pH 7.5	50 mM
	NaCl	140 mM
	EDTA	1 mM
	Glycerol	10% (v/v)
	NP-40	0.5% (v/v)
	Triton-X-100	0.25% (v/v)
<b>Lysis Buffer 2 (LB2)</b>	TRIS-HCl pH 8.0	10 mM
	NaCl	200 mM
	EDTA	1 mM
	EGTA	0.5 mM
<b>Lysis Buffer 3 (LB3)</b>	TRIS-HCl pH 8.0	10 mM
	NaCl	100 mM
	EDTA	1 mM
	EGTA	0.5 mM
	Na-Deoxycholate	0.1% (v/v)
	N-Lauroylsarcosine	0.5% (v/v)
<b>RIPA wash buffer</b>	HEPES-KOH pH 7.5	50 mM
	LiCl	500 mM
	EDTA	1 mM
	NP-40	1% (v/v)
	Na-Deoxycholate	0.7% (v/v)
<b>TE Buffer</b>	TRIS-HCl pH 8.0	10 mM
	EDTA	1 mM
<b>Elution buffer</b>	TRIS-HCl pH 8.0	50 mM
	EDTA	10 mM
	SDS	1%

## 2.7.2 Immunofluorescence (IF) assays in cultured cells

hiPSC and hNCC were grown on coverslips coated with Geltrex or fibronectin, respectively. As described in Laugsch et al. (2019), confluent cells were fixed with 3.7% formaldehyde in PBS for 15 minutes at RT and washed twice with PBS. Fixed cells were then stored in PBS at 4°C until further use or immediately permeabilized for 10 minutes at RT with IF Permeabilization Buffer (see Table 2.21). Afterwards, cells were blocked for one hour at RT with IF Blocking Buffer (see Table 2.21) and then incubated with antibody at the recommended dilution in IF Antibody Dilution Buffer (see Table 2.19 and Table 2.21) at 4°C overnight. After three washes with IF Wash Buffer (see Table 2.21), appropriate secondary antibodies (Table 2.20) matching the respective primary antibody were diluted in IF Antibody Dilution Buffer and incubated for 30 minutes at 37°C. Cells were again washed three times with IF Wash Buffer and incubated with 0.1 µM DAPI for 10 minutes at RT. Following another wash with IF Wash Buffer and one with PBS, the cells were permanently mounted on slides with FluoromountG for subsequent analysis by fluorescent microscopy (Olympus IX 81).

Table 2.19 Primary antibodies for Immunofluorescence

Antigen	Antibody	Manufacturer	Catalog Number
<b>OCT4</b>	Polyclonal goat anti Oct-3/4	Santa Cruz Technologies	sc-8628
<b>NANOG</b>	Polyclonal rabbit anti-Nanog	Santa Cruz Technologies	sc-33759
<b>TFAP2A</b>	Monoclonal mouse anti-TFAP2A	Santa Cruz Technologies	sc-12726
<b>SOX2</b>	Polyclonal goat anti-SOX2	Santa Cruz Technologies	sc-17320
<b>TRA-1-60</b>	Monoclonal mouse anti-TRA-1-60	Santa Cruz Technologies	sc-21705
<b>NR2F1</b>	Monoclonal mouse anti-COUP-TFI	Perseus Proteomics	H8132
<b>NR2F2</b>	Monoclonal mouse anti-COUP-TFII	Perseus Proteomics	H7147
<b>TWIST1</b>	Polyclonal rabbit anti-TWIST1	Invitrogen	PA5-49688
<b>SOX9</b>	Polyclonal rabbit anti-SOX9	Millipore	AB5535
<b>SM22-alpha</b>	Polyclonal rabbit anti-SM22-alpha	Abcam	ab14106
<b>TUJ1</b>	Polyclonal rabbit anti-beta Tubulin II	Abcam	ab18207
<b>DDX1</b>	Polyclonal rabbit anti-DDX1	Genetex	GTX105205

Table 2.20 Secondary Antibodies for Immunofluorescence

Antibody	Manufacturer	Catalog Number
<b>Alexa Fluor 488 goat anti-rabbit IgG (H+L)</b>	Invitrogen	A-11008
<b>Alexa Fluor 488 goat anti-mouse IgG (H+L)</b>	Invitrogen	A-11001
<b>Alexa Fluor 594 goat anti-mouse IgG (H+L)</b>	Invitrogen	A11032
<b>Alexa Fluor 594 goat anti-rabbit IgG (H+L)</b>	Invitrogen	R-37117
<b>Alexa Fluor 594 donkey anti-goat IgG (H+L)</b>	Invitrogen	A-11058

Table 2.21 Composition of Buffers for Immunofluorescence

Buffer	Component	Final concentration
<b>IF Permeabilization buffer</b>	Triton X-100	0.5% (v/v)
	BSA	5% (w/v)
	PBS	1x
<b>IF Blocking solution</b>	BSA	5% (w/v)
	PBS	1x
<b>IF Antibody dilution buffer</b>	BSA	1% (w/v)
	PBS	1x
<b>IF Wash Buffer</b>	Tween20	0.05% (v/v)
	PBS	1x

### 2.7.3 Flow Cytometry

For expression analysis of the NCC surface marker p75 (=NGFR), hNCC were detached from the dish with Accutase, re-suspended in FACS Wash Buffer (see Table 2.24) and stained with an antibody against NGFR conjugated to PerCP for 1 hour at 4°C. Afterwards, cell were washed with FACS Wash Buffer and measured on a FACS Calibur together with an unstained control.

Expression of nuclear proteins (TFAP2A, SOX9, NR2F1) was analyzed in fixed cells, as described in Laugsch et al. (2019). Briefly, hNCC were treated with 2% formaldehyde in PBS for 15 minutes at RT, permeabilized with FACS Permeabilization Buffer (see Table 2.24) for 30 minutes at RT and blocked with FACS Blocking Buffer (see Table 2.24) for another 30 minutes at RT. Primary antibodies were then diluted in FACS Blocking Buffer and incubated with the cells for 1 hour at RT. Afterwards, cells were washed with FACS Wash Buffer and incubated with an appropriate secondary antibody, matching the respective primary antibody, diluted in FACS Blocking Buffer for 30 minutes at RT. Finally, cells were washed with FACS Wash Buffer and measured on a FACS Calibur together with an unstained control.

Table 2.22 Primary/Conjugated Antibodies for Flow Cytometry

Antigen	Antibody	Manufacturer	Catalog Number
<b>P75/NGFR</b>	PerCP/Cy5.5 anti-human CD271 (NGFR)	Biologend	345111
<b>TFAP2A</b>	Monoclonal mouse anti-TFAP2A	Santa Cruz Technologies	sc-12726
<b>SOX9</b>	Polyclonal rabbit anti-SOX9	Millipore	AB5535
<b>NR2F1</b>	Monoclonal mouse anti-COUP-TF I/NR2F1	R&D Systems	PP-H8132-00

Table 2.23 Secondary Antibodies for Flow Cytometry

Antibody	Manufacturer	Catalog Number
<b>Mouse F(ab)2 IgG (H+L) PerCP-conjugated Antibody</b>	R&D Systems	F0114
<b>Rabbit IgG APC-conjugated Antibody</b>	R&D Systems	F0111
<b>DyLight™ 488 Donkey anti-rabbit IgG Antibody</b>	Biologend	406404

Table 2.24 Composition of Buffers for Flow Cytometry

Buffer	Component	Final concentration
<b>FACS Permeabilization buffer</b>	Triton X-100	0.1% (v/v)
	BSA	5% (w/v)
	PBS	1x
<b>FACS Blocking solution</b>	BSA	5% (w/v)
	PBS	1x
<b>FACS Wash Buffer</b>	BSA	0.5% (w/v)
	PBS	1x

## 2.7.4 Western Blot

Proteins were extracted from cultured cells by adding 1 ml ice cold RIPA buffer (see Table 2.25), supplemented with 1x protease inhibitor, to  $10^7$  cells and incubating both for 30 minutes at 4°C, rotating. Subsequently, the extract was centrifuged (30 minutes, 4°C, 14.000 g) to accumulate the DNA at the bottom of the tube, and DNA was removed with a gel loading tip. Protein concentration was determined using the Pierce BCA Protein Assay Kit on a plate reader (Berthold Technologies).

For SDS-PAGE, 20-40 µg of protein extract were mixed with 6x Laemmli buffer (see Table 2.25) and then heated to 95°C for 10 minutes. Denatured proteins were separated on a precast Bolt 8% Bis-Tris Gel (Invitrogen) with NuPage MOPS SDS Running Buffer (Invitrogen) and transferred to a 0.45 µm nitrocellulose western blotting membrane by wet transfer with Transfer Buffer (see Table 2.25)

Afterwards, the membrane was blocked with Blocking Buffer (see Table 2.25) for one hour at RT and subsequently incubated with primary antibodies (see Table 2.26) diluted in Blocking Buffer at 4°C overnight. After three washes with TBS-T, Blocking Buffer with appropriate secondary antibody (see Table 2.26) was added to the membrane and incubated for one hour at RT. HRP was then detected using the chemiluminescent ECL Western Blotting Substrate with the ChemiDoc MP (Biorad).

Table 2.25 Composition of Buffers for Western Blot

Buffer	Component	Concentration
<b>RIPA Buffer for Protein Extraction</b>	TRIS-HCl pH 8.0	50 mM
	NaCl	150 mM
	NP-40	1.0% (v/v)
	Na-Deoxycholate	0.5% (v/v)
	SDS	0.1% (v/v)
<b>Laemmli Buffer (6x)</b>	TRIS-HCl pH 6.8	375 mM
	Glycerol	60% (v/v)
	DTT	600 mM
	Bromphenol Blue	0.06% (w/v)
	SDS	12% (v/v)
<b>Transfer Buffer for Western</b>	TRIS Base	25mM
	Glycine	190 mM
	Methanol	20% (v/v)
<b>Blocking Buffer for Western</b>	TBS	1x
	Tween20	1% (v/v)
	Non-Fat Milk Powder	5% (w/v)
<b>TBS-T</b>	TBS	1x
	Tween20	1% (v/v)

Table 2.26 Antibodies for Western Blot

Antigen	Antibody	Brand	Catalog Number
<b>HA-tag</b>	Rabbit polyclonal anti HA-tag	Abcam	ab9110
<b>DDX1</b>	Rabbit polyclonal anti-DDX1	Genetex	GTX105205
<b>MYCN</b>	Mouse monoclonal anti-n-MYCN (NMYC-1)	Genetex	GTX20057
<b><math>\beta</math>-Tubulin</b>	Mouse monoclonal anti- $\beta$ -Tubulin	Sigma-Aldrich	T0198
<b>TBP</b>	Mouse monoclonal TATA binding protein TBP [1TBP18]	Abcam	ab818
<b>Rabbit IgG (2ndary)</b>	Goat anti-rabbit IgG-HRP	Invitrogen	10037282
<b>Mouse IgG (2ndary)</b>	Rabbit anti mouse IgG-HRP	Invitrogen	10670264

## 2.8 Genetic engineering by Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR/Cas9)

CRISPR/Cas9 was used to engineer genetic deletion in hiPSC and in chicken embryos. The design of guide RNAs and the cloning, as well as the detection of deletions are the same in both the *in vitro* and the *in vivo* approach and will be outlined in this section. Delivery of the plasmids, though, was different. A detailed description of the nonliposomal transfection and electroporation methods used in each case can be found in the Tissue Culture Procedures (2.3.5) and the Chicken Embryo Procedures (2.9.4) sections, respectively.

### 2.8.1 Design of guide RNAs (gRNA)

gRNAs were designed with the Benchling CRISPR guide design tool ([www.benchling.com](http://www.benchling.com)) using template sequences that flank the region to be deleted. Usually, gRNAs were designed, cloned and delivered in pairs, one on either side of the region of interest, to create two DSBs between which the deletion should be introduced. In case of the MYCN and FAM49A targeting in neural crest of chicken embryos, three gRNAs were delivered simultaneously to create as many possible combinations of disruptive re-arrangements either by indels at individual gRNA sites or by deletions and inversions between pairs of gRNAs.

A list of all gRNAs used in this work can be found in Table 2.27.

Table 2.27 Guide RNA sequences for CRISPR/Cas9

Name	Sequence	Species
EnhTFAP2A L2	TGTGCTGAAAGGGTCGTCGCTGG	human
EnhTFAP2A L5	GGATAACTTGCTTACCCCACTGG	human
EnhTFAP2A 34	ATTCTAAACATTCCCCGCACTGG	human
EnhTFAP2A 39	CTTTGAAAGATTATCTGCCTAGG	human
Enh2p24.2 L	AAGGGATCTGGCATCGGCAAAGG	human
Enh2p24.2 R	GGATACATGGGCATTCATATTGG	human
DDX1 L	CGACATCAAACATGGGTGAAGGG	human
DDX1 R	TATGTGTATATTGAAGCTAGAGG	human
MYCN L	AGGCCTCAACTCTTTAAGTGTGG	human
MYCN R	TTTGGTAGTATTCGTCCCATTGG	human
DDX1 ex5	ACTATTA <del>AA</del> ACTGGTGGGGCAGG	chicken
DDX1 ex24	TGTATGCAGTAGTCGTGGAAAGG	chicken

<b>MYCN 1</b>	CAGTGTCCCTCATGTCCATCGGGG	chicken
<b>MYCN 3</b>	CCACCGAACGTGGTTCAGCATGG	chicken
<b>MYCN 9</b>	ACACCCATTTGCTCCGTACCCGG	chicken
<b>FAM49A 1</b>	GAAGTCTTTAAGTCACCCAGAGG	chicken
<b>FAM49A 5</b>	AGATTGGCATGTCAAGAGTGGGG	chicken
<b>FAM49A 10</b>	TCAACTAACAGGTAGCATGCAGG	chicken

### 2.8.2 Cloning of gRNA into the Cas9-Vector

Single stranded gRNA oligos with orientation-specific BbsI overhangs (Fwd: CACC & Rev: AAAC) were annealed and ligated into the pX330A-Cas9 vector, previously linearized with BbsI (NEB). *E. coli* were then transformed with the ligation reaction mix as described in section 2.5.8. Positive bacteria clones were then validated by colony PCR with the corresponding forward oligo and the reverse primer pX335\_seq\_R (5'-AACGCCAATAGGGACTTTCC-3'), binding to the vector, using Platinum Taq Polymerase (Invitrogen). Resulting PCR products were column purified (NucleoSpin PCR and Gel Purification Kit, Macherey-Nagel) and Sanger sequenced (Microsynth SeqLab) with the reverse primer to ensure that the correct gRNA sequence was inserted into the vector. Plasmids were subsequently isolated by Mini Prep (NucleoSpin Plasmid, Macherey-Nagel).

### 2.8.3 Delivery of pX330A-Cas9-gRNA vector into target cells

#### In vitro

Transfection of hiPSC was performed as described in section 2.3.5 with pairs of pX330A-Cas9 vectors containing gRNAs upstream and downstream of the region to be deleted. For deletion of part of the *TFAP2A* regulatory domain, two rounds of transfection were necessary to obtain homozygous deletions (EnhTFAP2A L2 & L5 in the first round, EnhTFAP2A 34 & 39 in the second round, see Table 2.27). After the first round of transfection and confirmation of the deletion at the population level (see 2.8.4), clonal lines were generated (see 2.3.6) and a heterozygous line was subject to the second round of transfection to generate a homozygous deletion.

For generating the deletions of *DDX1* and *MYCN* genes, a single round of transfection was sufficient to yield clones with homozygous deletions. In the case of *DDX1*, though, the deletion was not introduced in WT hiPSC, but in TetON-*DDX1* hiPSC in which an exogenous copy of *DDX1* was stably integrated and inducibly expressed upon addition of Doxycycline (DOX) (see

2.8.5). This inducible *DDX1* expression was used to rescue the otherwise lethal phenotype resulting from the homozygous deletion of *DDX1*.

### In vivo

Electroporation of chicken embryos HH9-10 was performed as described in 2.9.4 with sets of two or three pX330A-Cas9 vectors containing gRNAs that target *DDX1*, *MYCN* or *FAM49A*. After electroporation, eggs were sealed with tape and incubated up to stage HH14-16 and HH20-23. Embryos of stage HH14-16 were isolated from the egg and the targeted neural tube was isolated as described in 2.9.3. Subsequently, DNA was extracted from the neural tube as described in 2.5.1 and used for detection of the intended deletions by PCR (see 2.8.4).

The electroporated embryos were imaged at HH20-23 with an Olympus SZX16 stereomicroscope and pictures were processed by Adobe Photoshop.

## 2.8.4 Deletion detection by PCR

Transfected cell populations or isolated neural tube of chicken embryos were collected and DNA was extracted as described in section 2.5.1. Subsequently, genotyping PCRs were performed as described in section 2.5.2 with primers listed in Table 2.28. (“gg” marking chicken specific primer sequences) and the deletions were confirmed by Sanger sequencing of the PCR products (Microsynth Seqlab). Primers were usually designed in a way that naturally occurring heterozygous SNPs could be used to identify clonality in heterozygous cell lines. Consequently, these cell lines (enhTFAP2A) were subjected to a second round of transfection to generate homozygous deletions.

Table 2.28 Primers for detection of deletions introduced by CRISPR/Cas9

Primer	Sequence 5'-3'
enhTFAP2A P1	CATTCCCGCAGCTTTGATTCTTTCA
enhTFAP2A P2	AGTTCTCTGCATTTCCCGCAGC
enhTFAP2A P3	AACGAGTGTAGCAGGGAACA
enhTFAP2A P4	TGCACAGAACACGACTGCATGA
enhTFAP2A P5	TGCTCTCTGCCTTGGGCTGAT
enhTFAP2A P6	TGACGTCAGCCATTTTGCAA
enhTFAP2A P7	TTTGCCCTCCACTGTGACACTCTC
enhTFAP2A P8	GAGAGGCGACCCGGAAGTGA
enhTFAP2A P9	ACGGAAAAGCGGGGACTGTG

<b>Enh2p24.2 I</b>	ACACAGAAGACAGGAGAGCC
<b>enh2p24.2 II</b>	CCATTCCCTGTGGCCTGATA
<b>enh2p24.2 III</b>	GAGGCACCACAGTCTTGCTC
<b>enh2p24.2 IV</b>	GCACAAGACAGGGCTCAAAA
<b>enh2p.24.2 SNP F</b>	CCCATCTTTGCCAGCTAAC
<b>enh2p.24.2 SNP R</b>	GTCTAGGGCAGGATCAGCAA
<b>DDX1 I</b>	CTACAGATAGCACAGCACGC
<b>DDX1 II</b>	CGGTATCAGGGTCACAGCTA
<b>DDX1 III</b>	TGGCTACCTTCCTAACCAGC
<b>DDX1 IV</b>	CACTGGCACCATCACATCAG
<b>DDX1 SNP F</b>	TGGTAAAGTCATGGGCCACT
<b>DDX1 SNP R</b>	CTGGAGGAGCTGCTTTTCTT
<b>MYCNex3 I</b>	AGGAGAGAGGGGAAGAATGC
<b>MYCNex3 II</b>	CTTCCCCATACCCCAACTGT
<b>MYCNex3 III</b>	TGGCATTACACTCTGTCCCT
<b>MYCNex3 IV</b>	ACCGAGTACAACCTGACACT
<b>ggDDX1 ex5 F</b>	CCCATGGGTGTTGCAACTGA
<b>ggDDX1 ex5 R</b>	GCATCCCAGTTCCTTCCACA
<b>ggDDX1 ex24 F</b>	GCCTGTATTGGAAAATTGCTGC
<b>ggDDX1 ex24 R</b>	ACACCACAGTCTTTTCAGAGGA
<b>ggFAM49A 1F</b>	ACATCACAGCAGGCAAAGAA
<b>ggFAM49A 10R</b>	AGCTTTTAATCCTTGGGGCT

Transfected cell populations and neural tubes of electroporated chicken embryo always contained a mix of non-targeted cells and targeted cells carrying different rearrangements: small indels at the individual gRNA sites, inversions between pairs of gRNAs or the desired deletions. *In vitro*, clonal lines with CRISPR/Cas9 mediated deletions were obtained as described in section 2.3.6. *In vivo*, the presence of abundant genetic alterations leading to the knockout of the targeted gene, were confirmed by PCR amplification (see Table 2.28. for primers, “gg” marking chicken specific sequences) of the entire targeted region, subsequent cloning of the PCR product into the pGEM-T vector and Sanger sequencing (see 2.8.6).

## 2.8.5 Rescue experiments

### In vitro

For the *in vitro* rescue experiments a PiggyBAC vector system (KA0717+KA0637+Supertransposase) was used that stably integrated an exogenous version of *DDX1* into the genome. *DDX1* was amplified from cDNA generated from hNCC RNA with primers listed in Table 2.29 and cloned into the KA0717 pPB-hCMV\*1-cHA-IRESVenus vector with the indicated restriction enzymes and adding a Kozak sequence (GCCACCATGG) to the forward primer as described in section 2.5.8. Subsequently, WT hiPSC were transfected with (KA0717+KA0637+Supertransposase, see Table 2.15 for details) in molar ratio of 10:1:3 as described in section 2.3.5. After selection with G418/neomycin, the targeted cell population was expanded and subject to CRISPR/Cas9 engineering with gRNAs designed for deletion of the endogenous copy of *DDX1* while the exogenous copy was slightly overexpressed by addition of 100 ng/ml DOX to the medium.

### In vivo

Rescue experiments were also performed *in vivo* to ascertain knockout specificity by overexpressing human genes in chicken embryos that could not be targeted by the chicken specific gRNAs. Cloning was performed as previously described (see 2.5.8). Briefly, human *DDX1* and *MYCN* full length cDNAs, amplified from hNCC derived cDNA with primers listed in Table 2.29, were cloned into the PiggyBac-CMV-MCS-EF1-RedPuro vector with restriction enzymes indicated in the same table. Electroporation of chicken embryos, as described in section 2.9.4, was performed with a plasmid mix of pX330A vectors with gRNA for CRISPR/Cas9 mediated knockout of the endogenous gene copy (chicken *DDX1* or *MYCN*) and the PiggyBAC vector constitutively overexpressing the corresponding human gene (human *DDX1* or *MYCN*, respectively).

Table 2.29 Primer sequence for rescue experiments

Enhancer	Sequence 5'-3'	Restriction site	Product size
<b>DDX1 for KA0717</b>	Fwd: AAAAACTAGTGCCACCATGGCGGCCTTCTCCGAG Rev: AAAAACTAGTGAAGGTTCTGAACAGCTG	SpeI SpeI	2220
<b>DDX1 for PiggyBAC-Red</b>	Fwd: AAAAAGCTAGCGCCACCATGGCGGCCTTCTCCGAG Rev: AAAAAGGATCCTCAGAAGGTTCTGAACAG	NheI BamHI	2220
<b>MYCN for PiggyBAC-Red</b>	Fwd: AAAAAGCTAGCGCCACCATGCCGAGCTGCTCCACG Rev: AAAAAGAATTCCTAGCAAGTCCGAGCGTG	NheI EcoRI	1392

### **2.8.6 pGEM-T cloning**

PCR products obtained to detect *DDX1* deletion in neural tube of chicken embryos (see 2.8.4) were cloned into the pGEM-T vector system I (Promega) according to the manufacturer's instructions. The vector was then transformed into *E. coli* as described in section 2.5.8. with additional blue-white selection using X-Gal and IPTG according to the user manual. Subsequently, plasmid DNA was extracted from 15 white colonies followed by Sanger sequencing (Microsynth Seqlab).

## **2.9 Chicken Embryo Procedures**

### **2.9.1 Chicken strain**

Fertilized eggs of the chicken breed White Leghorn (*Gallus gallus domesticus*) were obtained from a local breeder (LSL Rhein-Main).

### **2.9.2 Incubation and determination of developmental stage**

Fresh, fertilized eggs, were incubated at 37°C and 80% humidity in a poultry egg incubator (Thermos-de-Lux) until the embryos reach the desired developmental stage. Developmental stages were determined according to the Hamburg Hamilton (HH) staging system (Hamburger and Hamilton, 1992). Eggs were windowed and embryos were made visible by staining with a drop of Neutral Red dye or injection of ink into the egg yolk underneath the embryo for contrast. When microsurgical procedures of the embryo were implemented within the egg (e.g. for CRISPR/Cas9 experiments (2.8.3) or enhancer reporter assays (2.6.3)), eggs were sealed afterwards and re-incubated until the desired developmental stage.

### **2.9.3 Isolation of whole embryos, facial prominences, or neural tube sections**

Embryos of the desired developmental stage were isolated from the egg with scissors and forceps, transferred to a 5 ml petri dish with 1x PBS and all supportive extraembryonic membranes were removed. For whole mount *in situ* hybridization whole embryos were used. Facial

prominences (frontonasal prominence; FNP) were isolated for Chromatin Conformation Capture (4C, see 2.10) or Chromatin Immunoprecipitation (ChIP, see 2.7.1) with surgical scissors, pooled in a 1.5 ml tube, flash frozen in liquid nitrogen and stored at -80°C until further use. Neural tubes were isolated for DNA extraction after CRISPR/Cas9 targeting, first by roughly exposing the targeted area with scissors and then incubating the tissue briefly in trypsin (1-2 minutes) at RT with a subsequent wash in PBS. Afterwards, the surrounding tissue could be removed with forceps, leaving the neural tube ready for DNA extraction.

#### **2.9.4 Electroporation**

Electroporation was performed as previously described by Rehimí et al. (2016). Briefly, eggs of stage HH9-10 were windowed and the extraembryonic membrane was opened at the site of electroporation. Plasmids of interest (enhancer reporter vector, see 2.6 or CRISPR/Cas9 vector with or without rescue vector, see 2.8) were mixed with an electroporation control plasmid (PiggyBac-CMV-MCS-EF1-RedPuro, see 2.5.8) and Fast Green solution (Sigma-Aldrich) at a ratio of 1:1:1 and microinjected into the frontal neural tube of the chicken embryos. The neural tube was then electroporated by five square pulses of 20 V within 20 ms width using the Intracel TSS20 OVODYNE Electroporator.

#### **2.9.5 Whole Mount *in situ* Hybridization (WISH)**

Gene expression patterns in chicken embryos were analyzed by whole mount *in situ* hybridization with digoxigenin-labeled riboprobes according to the WISH protocol from the Gallus gallus expression *in situ* hybridization analysis (GEISHA) website (Acloque et al., 2008; Streit and Stern, 2001). Briefly, embryos of the desired state were freshly collected in PBS and fixed with 4% paraformaldehyde overnight. Embryos were then dehydrated in 100% methanol overnight and subsequently rehydrated through a graded methanol series into PBT. Embryos were then digested with Proteinase K and fixed again in 4% paraformaldehyde. Until hybridization, embryos were stored in prehybridization solution (prehyb) at -20°C for several days. In the meantime, DIG-labelled antisense probes were amplified from cDNA of HH10 chicken embryo heads by PCR with an NTP-DIG mix (Roche, #11277057001 & #11209256910) and T7 RNA Polymerase (Roche, #1088176700). The resulting PCR products were gel purified and denatured at 95°C for two minutes. Primers for probe amplification are listed in Table 2.30. The probes were then added to

fresh prehyb at a concentration of 1000 ng/ml and incubated with the embryos at 65°C for about 48 hours. After several washed with SSC and KTBT buffer embryos were blocked in PBT with 20% sheep serum for 3 hours at RT. For antibody detection embryos were subsequently incubated with anti-DIG-AP Fab fragment (Roche, #11093274910) in 20% sheep serum (1:3000) at 4°C overnight. Then, embryos were rinsed in PBT and NTMT buffer prior to the color reaction. This was performed in the dark with NBT and BCIP in NTMT buffer at RT until general background color began to appear. Staining was stopped by several washes in NTMT and KTBT buffer and a final wash in PBT. Background staining was removed through a graded methanol series to 100% methanol. Photos of the stained embryos were taken with an inverted bright field microscope (Leica).

Table 2.30 Primers for *in situ* hybridization

Gene	Sequence 5'-3'	Annealing temperature	Product size
<b>DDX1</b>	TTATAAAAGCTTGCGGCCGCAGAATATCGGGCAATCAAGGAACACAA GCTCTAGAAATTAACCCTCACTAAAGGACCCCTTTCCACGACTACTG	63°C	418 bp
<b>TFAP2A</b>	TTATAAAAGCTTGCGGCCGCAGAATATCCTACAGCCTCAACTCCCTG GCTCTAGAAATTAACCCTCACTAAAGGTGTACTTGGAGGTGGAGCTG	63°C	447 bp
<b>MYCN</b>	TTATAAAAGCTTGCGGCCGCAGAATATCCCTTCCCGTCAACAAG GCTCTAGAAATTAACCCTCACTAAAGGACTTCTGCTCCTCTGCCTGA	64°C	661 bp

Table 2.31 Composition of Buffers for WISH

Buffer	Component	Final Concentration
<b>SSC pH 7.0 (20x)</b>	NaCl	3.0 M
	Sodium Citrate	0.3 M
<b>Prehyb</b>	Deionized Formamide	50%
	SSC	5x
	Blocking Powder	2%
	Triton X-100	0.1%
	CHAPS	0.1%
	Yeast RNA	50 µg/ml
	EDTA pH 8.0	5 mM
<b>KTBT</b>	Heparin	50 µg/ml
	Tris-HCl (pH 7.5)	50 mM
	NaCl	150 mM
	KCl	10 mM
	Triton X-10	1%

<b>NTMT</b>	NaCl	100 mM
	Tris (pH 9.5)	100 mM
	MgCl <sub>2</sub>	50 mM
	Triton X-100	0.1%

## 2.9.6 Overexpression of human *FAM49A*

*FAM49A* overexpression experiments were performed *in vivo* to rule out any possible contribution of *FAM49A* misexpression to the OFC phenotype. Cloning was performed as previously described (see 2.5.8). Briefly, the human *FAM49A* transcript sequence, amplified from hNCC derived cDNA with primers listed in Table 2.32Table 2.29, was cloned into the PiggyBac-CMV-MCS-EF1-RedPuro vector with restriction enzymes indicated in the same table. Electroporation of chicken embryos, as described in section 2.9.4, was performed with the PiggyBAC vector constitutively overexpressing an endogenous gene copy of human *FAM49A* or the empty PiggyBAC vector for the negative control.

Table 2.32 Primer sequence for overexpression of human *FAM49A* in chicken embryos

Enhancer	Sequence 5'-3'	Restriction site	Product size
<b>FAM49A for PiggyBAC-Red</b>	Fwd: AAAAAGCTAGCGCCACCATGGGAAACCTGCTCAAA	NheI	972 bp
	Rev: AAAAAGAATTCCTACTGAAGCATTGCTCG	EcoRI	

## 2.10 Genomic methods

### 2.10.1 Circular Chromatin Conformation Capture (4C)

Circular Chromatin Conformation Capture (4C) assays were performed as previously described (Stadhouders et al., 2013) with slight modifications. Briefly, 2-3x10<sup>7</sup> cells or 25 FNP of stage HH24 chicken embryos were crosslinked with 1% formaldehyde for 10 minutes, quenched with 0.125 M glycine and afterwards washed with PBS. Cells were re-suspended in lysis buffer (see Table 2.33), incubated for 15 minutes on ice, and centrifuged for 5 minutes at 4000 g and 4°C. Then, nuclei were re-suspended in 1.2x CutSmart restriction buffer (NEB), SDS was added to a

final concentration of 0.3%, followed by incubation at 37°C shaking at 900 rpm for 1 hour. Subsequently, Triton X-100 was added to a final concentration of 2% followed by another incubation at 37°C shaking at 900 rpm for 1 hour. Finally, 600 U of NlaIII was added to digest the chromatin overnight at 37°C shaking at 900 rpm.

The next day, the restriction enzyme was inactivated by addition of SDS to a final concentration of 1.6% and incubation at 65°C for 20 minutes. Next, digested chromatin was transferred to a 50 ml falcon and mixed with ligation buffer (see Table 2.33) and Triton X-100 (1% final concentration) to a total of 7 ml. After 1 hour incubation at 37°C, 100 U of highly concentrated T4 DNA ligase (Invitrogen) was added to the sample and incubated at 16°C for 4 hours. Following RNase A treatment, Proteinase K was added and chromatin was de-crosslinked at 65°C overnight.

DNA was purified with phenol-chloroform-isoamyl alcohol (25:24:1), precipitated with ethanol, and finally dissolved in H<sub>2</sub>O. At this point, digestion and ligation efficiencies were evaluated by agarose gel electrophoresis followed by a second round of digestion with 50 U DpnII in DpnII-specific restriction buffer (NEB) overnight at 37°C.

DNA was again purified with phenol-chloroform-isoamyl alcohol (25:24:1), precipitated with ethanol, and dissolved in water. Afterwards, a second ligation was performed with 200 U of T4 DNA ligase (Invitrogen) in a total of 14 ml ligation buffer (see Table 2.33) at 16°C overnight.

A final round of phenol-chloroform-isoamyl alcohol (25:24:1) extraction and ethanol precipitation was followed by column purification with the QIAquick Kit (Qiagen) and efficiencies of second digestion and ligation was once again evaluated by agarose gel electrophoresis

Next, the resulting 4C libraries were amplified by inverse PCR using the Expand Long Template PCR system (Roche) with 32 cycles and primers designed as previously described (Stadhouders et al., 2013) (see Table 2.34). Illumina adaptors P5 (5'-AATGATACGGCGACCACCGAACACTC-TTCCCTACACGACGCTCTTCCGATC-3') and P7 (5'-CAAGCAGAAGACGGCATAACGA-3') were added to the primers close to the NlaIII and the DpnII restriction site, respectively, with barcodes for multiplex sequencing (e.g. CGATGT, ATCACG, TCGATG). Purified PCR products were eventually sequenced on a HiSeq2500 platform generating reads of either 74 or 100 bp in length.

Table 2.33 Composition of Buffers for 4C

Buffer	Component	Final concentration
<b>Lysis Buffer for 4C</b>	TRIS-HCl pH 8.0	10 mM
	NaCl	10 mM
	NP-40	0.2%
	Protease Inhibitor Cocktail	1X
<b>Ligation Buffer for 4C</b>	TRIS-HCl pH 7.5	50 mM
	MgCl <sub>2</sub>	10 mM
	ATP	1 mM
	DTT	1 mM

Table 2.34 Primers for inverted PCR (4C) with Illumina adaptors

Viewpoint	Primer	Sequence 5'-3'
<b>TFAP2A Prom</b>	P5 R A (P1)	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTCGATGTGCTGTGCCTAAGAATGGGG
<b>TFAP2A Prom</b>	P5 R B (P1)	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTATCACGGCTGTGCCTAAGAATGGGG
<b>TFAP2A Prom</b>	P7 F (P2)	CAAGCAGAAGACGGCATAACGATCGATGGAGAGGGAGGGTCAAGCT C
<b>TFAP2A +100kb</b>	P5 F A	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTCGATGTCAAAGAGGCACGCACTGG
<b>TFAP2A +100kb</b>	P5 F B	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTATCACGCAAAGAGGCACGCACTGG
<b>TFAP2A +100kb</b>	P7 R	CAAGCAGAAGACGGCATAACGATCGATGAGCCTGATGAGCTCTAGT GT
<b>TFAP2A +480kb</b>	P5 F A	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTCGATGTAGTCTGGTCAGGTCAAAGGG
<b>TFAP2A +480kb</b>	P5 F B	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTATCACGAGTCTGGTCAGGTCAAAGGG
<b>TFAP2A +480kb</b>	P7 R	CAAGCAGAAGACGGCATAACGATCGATGCCAGTGCGGGGAATGTTT AG
<b>TFAP2A chicken</b>	P5 F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTCGATGTCTTCAGCGTGGAGGAATAAT
<b>TFAP2A chicken</b>	P7 R	CAAGCAGAAGACGGCATAACGATGGATGGCAGGATAAAATATG
<b>TFAP2A WT allele</b>	P5 F (P4)	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTTCGATGGGATTGAAGCAAGAAGGTAAC
<b>TFAP2A BOFS allele</b>	P5 F (P5)	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTTCGATGCCAAGGAAATAGGAGTAAGGT
<b>TFAP2A allele spec</b>	P7 R (P3)	CAAGCAGAAGACGGCATAACGATCGATGGGTAAGCACTTTAGAAGC CT

<b>DDX1 Prom</b>	P5 R	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTTTCTGGACGCCGCCTTC
<b>DDX1 Prom</b>	P7 F	CAAGCAGAAGACGGCATAACGAGCACATGTCTGATGGACCCGCGTT G
<b>MYCN Prom</b>	P7 R	CAAGCAGAAGACGGCATAACGACGTATCCCCTCCAACTCTCCAAAA
<b>MYCN Prom</b>	P5 F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTTTCCCAGAAAGAATAGCATG
<b>FAM49A Prom</b>	P5 R	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTACAGTGCCAAAATGCTGGTCCCA
<b>FAM49A Prom</b>	P7 F	CAAGCAGAAGACGGCATAACGAACAGTGCACAAAGCCGGACCTCTG
<b>Enh2p24.2</b>	P7 R	CAAGCAGAAGACGGCATAACGAATCGTGGTGAGATGAACAGGGTCT
<b>Enh2p24.2</b>	P5 F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTTCCACTTTTGATTGTTCCC
<b>Enh2p24.2 chicken</b>	P5 R	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT CCGATCTCGATGTCTTCTGTACAGAGCTGAAGAA
<b>Enh2p24.2 chicken</b>	P7 F	CAAGCAGAAGACGGCATAACGATGTTTAAGGACAGGCATTTG

### 2.10.2 ChIP-seq

As described in Laugsch et al. (2019). ChIP-seq libraries were generated according to a modified TruSeq protocol (Illumina) from ChIP and input DNA (see 2.7.1) by the Cologne Center for Genomics. Libraries from hNCCp2 ChIPs were sequenced with a 1 x 50-bp protocol on a HiSeq2500 or HiSeq4000 sequencer (Illumina) and ChIPs from chicken FNP were sequenced with a 2 x 100-bp protocol on a HiSeq2500 sequencer.

### 2.10.3 ATAC-seq

Assay for Transposase Accessible Chromatin coupled to high-throughput sequencing (ATAC-seq) was performed in chicken facial tissue and hNCC as previously described (Buenrostro et al., 2013) with slight modifications, also mentioned in Laugsch et al. (2019). Briefly, two FNPs isolated from chicken embryos of stage HH23 were disaggregated with 0.125% (w/v) collagenase in PBS with 10% BSA for 30 minutes at 37°C while shaking at 600 rpm and carefully re-suspended for a single cell suspension. hNCC were disintegrated into a single cells suspension by treatment with

Accutase. Approximately 50.000 single cells were then lysed in 500 µl ice cold lysis buffer substituted with 1x protease inhibitor for 15 minutes on ice and subsequently centrifuged at 4000 g for 10 minutes at 4°C. Nuclei were re-suspended and incubated for 30 minutes at 37°C in 25 µl 2x TD buffer, 10 µl TDE1 (transposase), and 15 µl nuclease-free H<sub>2</sub>O to add illumine adaptor to open chromatin regions. Afterwards, the samples were purified with the MinElute PCR Purification Kit, whole genome sequencing libraries were processed with the NexteraXT Kit according to the manufacturer's instructions sequenced on a HiSeq2500 sequencer (Illumina).

**Table 2.35 Composition of the ATAC-seq Lysis Buffer**

Buffer	Component	Final Concentration
<b>ATAC-seq Lysis Buffer</b>	TRIS-HCL pH 7.4	10 mM
	NaCl	10 mM
	MgCl <sub>2</sub>	3 mM
	IGEPAL CA-630	0.1%

#### 2.10.4 RNA-seq

For RNA-seq total RNA, extracted as described in section 2.5.3, was treated with the Turbo DNase Kit (Invitrogen) according to the manufacturer's instructions. Then, strand-specific libraries were prepared with the TruSeq v2 RNA sample preparation Kit (Illumina) for poly-A RNA as previously described (Laugsch et al., 2019; Respuela et al., 2016) and sequenced with a 2 x 75 bp protocol on a HiSeq2500 or a HiSeq4000 sequencer platform (Illumina).

The following RNA-seq libraries were prepared:

- a) mRNA of hNCC d11: enhTFAP2A 0.4 Mb deletion (four samples: two of clone #1, two of clone #2) vs. WT (three samples of cell line B, the parental cell line)
- b) mRNA of hNCC p2: BOFS (three samples from three clones) vs. WT (two samples: one of cell line B, one of cell line S24)
- c) mRNA of hNCC p2: heterozygous DDX1 deletion (three samples from three clones) vs. WT (three samples of cell line B, the parental cell line)

### **2.10.5 Single cell RNA-seq (scRNA-seq)**

Sequencing of RNA from single cells was performed as previously described (Laugsch et al., 2019). Briefly, hNCCp2 of WT and BOFS were dissociated with Accutase into a single cell suspension, then centrifuged at 300 g for 5 minutes and washed with 1 ml of 0.04% BSA in PBS and centrifuged again. Afterwards, cells were counted, diluted to the desired concentration in 0.04% BSA in PBS, passed through a cell strainer, and counted again.

cDNA synthesis and library preparation with unique molecular identifiers (UMIs) was performed according to the Chromium Single Cell 3' Reagent Kit (10x Genomics) protocol. Briefly, cells were partitioned in nanoliter-scale Gel Bead in Emulsion (GEM), containing a single cell, a single Gel Bead, and reverse transcription reagents. The latter consists of a reaction master mix and primers with an Illumina R1 sequence, a 16 bp 10x barcode unique for each cell, a 10 bp UMI, and a poly dT sequence.

Full-length barcoded cDNA was then generated from poly-adenylated mRNA. GEMs were broken and the pooled fractions were recovered to then expose the cDNA to enzymatic fragmentation and size selection. Afterwards, libraries were constructed by adding P5, P7, a sample index and an Illumina R2 primer. The P5 and P7 primers were the used for Illumina bridge amplification and the R1 and R2 primers for paired-end Illumina sequencing.

## **2.11 Computational Analyses of next generation sequencing data**

### **2.11.1 4C-seq**

As described in Laugsch et al. (2019), 4C-seq reads were assigned to samples based on the first 10 bases of the read. Next, primer sequences were removed from the reads and the remaining sequence was trimmed to 36 bp. These 36 bases were then aligned to the human (hg19) or the chicken (galGal3) reference genome using Bowtie (Langmead et al., 2009). Subsequently, the resulting reads were analyzed with R3C-seq (Thongjuea et al., 2013) to generate RPM normalized bedgraph files for downstream visualization and analysis.

### 2.11.2 ChIP-seq

ChIP-seq sequencing reads were mapped to the human genome (hg19 assembly) or the chicken genome (galGal4 assembly) by BWA (Li and Durbin, 2009) or Bowtie (Langmead et al., 2009) and duplicate reads were discarded using SAMtools (Li et al., 2009). The resulting BAM files of the TFAP2A ChIPs were then analysed by MACS2 (Zhang et al., 2008) with the following settings:  $q \leq 10^{-5}$ ; Fold-enrichment  $\geq 5$ ; Broad Region Calling OFF. This was to determine genomic regions with significantly enriched binding of TFAP2A compared to the corresponding genomic DNA input.

To identify regions differentially bound by TFAP2A in WT and BOFS hNCCp2, we first considered all peaks ( $n=56630$ ) defined in both duplicated of WT and BOFS TFAP2A ChIP-seq experiments. Then we determined statistically differential TFAP2A binding by DiffBind (Ross-Innes et al., 2012) with the following settings:  $FDR \leq 10^{-2}$ ; Fold  $\geq 2$  (here refers to difference in mean read concentrations between WT and BOFS TFAP2A ChIP-seq signals). TFAP2A peaks were subsequently classified as “Unchanged”, “BOFS High”, and “BOFS low”.

Motif analysis in “Unchanged”, “BOFS High”, and “BOFS Low” TFAP2A peaks was performed with the MEME-ChIP software (part of the MEME suite) using standard settings (Machanic and Bailey, 2011).

TFAP2A and DDX1 ChIP-seq experiments were performed in two biological replicates. Pearson correlation coefficients between TFAP2A experiments were determined by the *bamCorrelate* tool of *deepTools* (*bins* mode with a bin size of 10 kb across the genome) (Ramirez et al., 2014).

To visualize ChIP-seq signal profiles, corresponding BAM files were transformed into bedgraph and bigwig files using the *bamCoverage* tool of *deepTools*. In the process signal was normalized as RPGS (reads per genomic content; 1x depth of coverage) with 5 bp bins. Normalized bigwig files were then also used to generate heat map signal profiles around the center of TFAP2A, DDX1 and H3K27ac peaks with the *computeMatrix* and *plotHeatmap* tools in *deepTools*. Furthermore, using the *computeMatrix* and *plotProfile* tools in *deepTools*, normalized DDX1 signals were plotted against the gene bodies of groups of genes with different levels of expression.

GREAT (McLean et al., 2010) was used to assign TFAP2A peaks to their putative target genes using the *Basal plus extension* association rule: each gene was assigned to a basal regulatory domain located a minimum of 5 kb upstream and 1 kb downstream of the TSS (regardless of any other nearby genes). The gene regulatory domain was then extended in both directions to the

nearest gene's basal domain but no more than 250 kb in one direction. Once every gene got assigned to a regulatory domain, each genomic region was associated with all genes whose regulatory domain it overlapped.

Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) was performed according to the instructions of the tool using the *GSEAPreranked* option with 1000 iterations. The gene sets used for the GSEA consisted of those genes associated with TFAP2A peaks according to GREAT as described above, grouped into BOFS Low peaks (3531 genes) and BOFS High peaks (1047 genes). GSEAPreranked analysis was performed for each set against all human genes (Ensembl Genes used in GREAT) ranked according to their expression FC between BOFS and WT hNCCp2 as obtained by the RNA-seq experiments.

### 2.11.3 ATAC-seq

ATAC-seq sequencing reads were mapped to the human genome (hg19 assembly) or the chicken genome (galGal4 assembly) by BWA (Li and Durbin, 2009) and duplicate reads were discarded.

For visualization of ATAC-seq signal profiles, the resulting BAM files were transformed into bedgraph and bigwig files, normalizing signals as RPGS with 5 bp bins according to *deepTools*.

### 2.11.4 RNA-seq

RNA-seq reads were mapped strand-specifically using TopHat2.1.1 (Kim et al., 2013) and the GENCODE 18 gene annotation set (Harrow et al., 2012). Only uniquely mapped reads were maintained.

Gene expression was normalized to fragments per kilobase of exon model per million reads mapped (FPKMs) using Cufflinks (Trapnell et al., 2010) and the GENCODE 18 gene annotation set. Subsequently, the FPKMs obtained per replicate were averaged.

Genes differentially expressed between groups of samples as described above (2.10.4), were functionally annotated using Enrichr (Kuleshov et al., 2016) and DAVID (Huang da et al., 2009). For the Enrichr analysis, *MGI*, *Human Phenotype* and *Biological Process* ontologies were considered, reflecting those categories with *Enrichr Combined Scores* >10 or the 50 functional categories with the highest *Enrichr Combined Score*. For the DAVID analysis, the *GO\_BP\_FAT*

ontology was considered, reflecting categories with p-values  $< 10^{-5}$  or the 50 functional categories with the lowest p-value.

### 2.11.5 scRNA-seq

UMIs (unique molecular identifiers) were counted with cellranger-2.1.0 (Zheng et al., 2017) using default settings on hg19. Initial dimensional reduction, clustering, and visualization were performed with cellranger-2.1.0 and cellrangerRkit\_2.0.0 within R-3.4.0. Next, counts were aggregated into a single matrix by default normalization (“normalize=mapped”) with cellranger-2.1.0. This matrix was then further processed with monocle\_2.6.4 (Qiu et al., 2017) within R-3.4.0 and dropouts were corrected using Rmagic\_1.3.0 (van Dijk et al., 2018) within R-3.4.0 for correlation analysis.

### 2.11.6 Classification of GWAS SNPs

#### Definition of haplotypes

The GWAS Catalog (Welter et al., 2014) currently reports 206 SNPs associated (significance  $p < 10 \times 10^{-6}$ ) with OFC from independent GWAS and meta-analyses. All these SNPs are located in a total of 40 risk loci reported to be associated with Orofacial Clefts (OFC). The present project is based on SNPs from a total of six OFC GWAS reported before 2014 (Beaty et al., 2010; Beaty et al., 2011; Birnbaum et al., 2009; Grant et al., 2009; Ludwig et al., 2012; Mangold et al., 2010), and additional, up to then unpublished OFC associated SNPs, kindly provided by the Institute of Human Genetics at the University of Bonn (later published in Ludwig et al. (2017)). With these SNPs we employed the HaploREG tool (Ward and Kellis, 2012) to identify all SNPs in linkage disequilibrium (LD) (i.e. follower SNPs) with the associated common SNPs reported in the GWAS (i.e. leader SNP). For every haplotype, consisting of a leader SNP and usually several follower SNPs, boundaries were defined as the position of the first and last reported SNP within the haplotype.

#### Co-localization analysis

Co-localization analysis for OFC-associated SNPs with tissue-specific epigenetic datasets was performed as previously described Ludwig et al. (2017). Briefly, haplotypes defined above were overlapped with previously published ChIP-seq data of hNCCd11 (Rada-Iglesias et al., 2012). In

these datasets, active enhancer were identified based on different combinations of epigenetic marks. To define these enhancer maps, hNCC ChIP-seq datasets were first re-analyzed with MACS2 (Zhang et al., 2008) using the following criteria:

- p300 and TFAP2A peaks were identified using  $q=10^4$
- H3K4me1: peaks were identified using broad settings and  $q=10^4$
- H3K27ac: peaks were identified using broad settings and  $q=10^6$ .

Then, distal and proximal active enhancers were defined as regions with p300 and TFAP2A peaks, each peak extended 1 kb up- and downstream, overlapping H3K27ac and H3K4me1 peaks.

### 2.11.7 Data Availability

ChIP-seq, ATAC-seq, 4C-seq, RNA-seq, scRNA-seq, datasets generated for the first part of this thesis, which is published in Laugsch et al. (2019), can be accessed under accession number GSE108522 on the Gene Expression Omnibus (GEO) repository.

Genomic datasets generated for the second part of this thesis (e.g. further ChIP-seq, 4C-seq, RNA-seq) have not been published yet but will also become available on the GEO repository.

### 2.11.8 Public genomic datasets used

Publically available ChIP-seq datasets used in this study have been published in Rada-Iglesias et al. (2012); Rada-Iglesias et al. (2011) and are available on the GEO repository under the following accession numbers:

- Chicken embryo HH20 FNP:
  - o H3K27ac (GSM933349)
- hESC:
  - o p300 (GSM602291)
  - o H3K27me3 (GSM714806)
  - o H3K27ac (GSM602294)
  - o H3K4me1 (GSM602295)
  - o H3K4me3 (GSM602296)

- hNCC day11
  - p300 (GSM714804)
  - H3K27me3 (GSM714806)
  - H3K27ac(GSM714807)
  - H3K4me1 (GSM714808)
  - H3K4me3 (GSM714809)

Publically available Hi-C data generated in hESC by Dixon et al. (2015) were visualized using the 3D Genome Browser (Wang et al., 2018).

## 2.11.9 Software and Algorithms

Table 2.36 Software and Algorithms for Analysis of Next Generation Sequencing Data

Software/Algorithm	Source	Identifier
<b>3D Genome Browser</b>	(Wang et al., 2018)	<a href="http://3dgenome.org/">http://3dgenome.org/</a>
<b>BD FACSDiva</b>	BD Biosciences	<a href="http://www.bdbiosciences.com/us/instruments/research/software/flow-cytometry-acquisition/c/2046613">http://www.bdbiosciences.com/us/instruments/research/software/flow-cytometry-acquisition/c/2046613</a>
<b>Bowtie</b>	Langmead et al. (2009)	<a href="http://bowtie-bio.sourceforge.net/">http://bowtie-bio.sourceforge.net/</a>
<b>BWA</b>	Li and Durbin (2009)	<a href="http://bio-bwa.sourceforge.net/">http://bio-bwa.sourceforge.net/</a>
<b>cellranger-2.1.0</b>	Zheng et al. (2017)	<a href="https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger">https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger</a>
<b>Cufflinks</b>	Trapnell et al. (2010)	<a href="https://github.com/cole-trapnell-lab/cufflinks">https://github.com/cole-trapnell-lab/cufflinks</a>
<b>DAVID</b>	Huang da et al. (2009)	<a href="https://david.ncifcrf.gov/">https://david.ncifcrf.gov/</a>
<b>deepTools</b>	Ramirez et al. (2014)	<a href="http://deeptools.ie-freiburg.mpg.de/">http://deeptools.ie-freiburg.mpg.de/</a>
<b>DESeq2</b>	Anders and Huber (2010)	<a href="https://bioconductor.org/packages/release/bioc/html/DESeq2.html">https://bioconductor.org/packages/release/bioc/html/DESeq2.html</a>
<b>DiffBind</b>	Ross-Innes et al. (2012)	<a href="https://bioconductor.org/packages/release/bioc/html/DiffBind.html">https://bioconductor.org/packages/release/bioc/html/DiffBind.html</a>
<b>Enrichr</b>	Kuleshov et al. (2016)	<a href="http://amp.pharm.mssm.edu/Enrichr/">http://amp.pharm.mssm.edu/Enrichr/</a>
<b>FastQC</b>	Babraham Bioinformatics	<a href="http://www.bioinformatics.babraham.ac.uk/projects/fastqc/">http://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
<b>FIJI (ImageJ)</b>	Schindelin et al. (2012)	<a href="https://fiji.sc/">https://fiji.sc/</a>
<b>FlowJo 7.6.5.</b>	FlowJo, LLC	<a href="https://flowjo.com/">https://flowjo.com/</a>
<b>Galaxy</b>	Afgan et al. (2016)	<a href="http://main.g2.bx.psu.edu/">http://main.g2.bx.psu.edu/</a>
<b>Gene Set Enrichment Analysis (GSEA)</b>	Subramanian et al. (2005)	<a href="http://software.broadinstitute.org/gsea/index.jsp">http://software.broadinstitute.org/gsea/index.jsp</a>
<b>GREAT</b>	McLean et al. (2010)	<a href="http://bejerano.stanford.edu/great/public/html/">http://bejerano.stanford.edu/great/public/html/</a>

## Material & Methods

<b>HaploREG</b>	(Ward and Kellis, 2012)	<a href="https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php">https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php</a>
<b>KNIME</b>	KNIME AG, Zurich, Switzerland	<a href="https://www.knime.com/">https://www.knime.com/</a>
<b>MACS2</b>	Zhang et al. (2008)	<a href="http://liulab.dfci.harvard.edu/MACS/">http://liulab.dfci.harvard.edu/MACS/</a>
<b>MEME-ChIP software (part of the MEME suite)</b>	Machanick and Bailey (2011)	<a href="http://meme-suite.org/tools/meme-chip">http://meme-suite.org/tools/meme-chip</a>
<b>monocle_2.6.4</b>	Qiu et al. (2017)	<a href="http://cole-trapnell-lab.github.io/monocle-release/">http://cole-trapnell-lab.github.io/monocle-release/</a>
<b>OMERO</b>	University of Dundee & Open Microscopy Environment	<a href="https://www.openmicroscopy.org/">https://www.openmicroscopy.org/</a>
<b>PeakPicker2</b>	Ge et al. (2005)	N/A
<b>Primer3</b>	Untergasser et al. (2012)	<a href="http://primer3.ut.ee/">http://primer3.ut.ee/</a>
<b>R3C-seq</b>	Thongjuea et al. (2013)	<a href="http://bioconductor.org/packages/release/bioc/html/r3Cseq.html">http://bioconductor.org/packages/release/bioc/html/r3Cseq.html</a>
<b>Rmagic_1.3.0</b>	van Dijk et al. (2018)	<a href="https://github.com/KrishnaswamyLab/MAGIC">https://github.com/KrishnaswamyLab/MAGIC</a>
<b>SAMtools</b>	Li et al. (2009)	<a href="http://samtools.sourceforge.net">http://samtools.sourceforge.net</a>
<b>TopHat2</b>	Kim et al. (2013)	<a href="https://ccb.jhu.edu/software/tophat/index.shtml">https://ccb.jhu.edu/software/tophat/index.shtml</a>

## 3 RESULTS

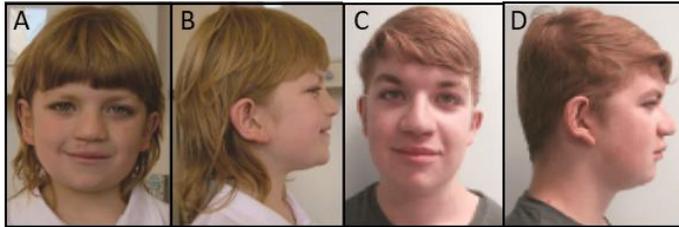
### 3.1 Disruption of the TFAP2A regulatory domain causes branchiooculo-facial syndrome (BOFS) and illuminates pathomechanisms for other human neurocristopathies

Remark: The first part of the results section, namely “Disruption of the TFAP2A regulatory domain causes branchiooculo-facial syndrome (BOFS) and illuminates pathomechanisms for other human neurocristopathies” has recently been published in *Cell Stem Cell* as “Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs” (Laugsch et al., 2019). Magdalena Laugsch and I contributed equally to this work with the help of many others:

Patient diagnosis and clinical characterization as well as skin fibroblast collection was performed by Katherine Lachlan (Table 2.8, Figure 3.1.1). Initial characterization of chromosomal abnormalities was performed by Julia Baptista (Figure 3.1.2). Targeted Locus Amplification was performed by *Cergentis* (Figure 3.1.3, Figure 3.1.27D). *In vivo* reporter assays (Figure 3.1.15) and Alcian Blue staining (Figure 3.1.31) were performed by Rizwan Rehim. CRSPR/Cas9 engineering of the 0.4 Mb TFAP2A enhancer deletion in WT hiPSC was performed by Agathi Karaolidou (Figure 3.1.16, Figure 3.1.17) and consecutive characterization by Hafiza Alirzayeva (Figure 3.1.18, Figure 3.1.19, Figure 3.1.20, Figure 3.1.21). Magdalena Laugsch derived the BOFS patient-specific hiPSC from fibroblasts and characterized them (Figure 3.1.24, Figure 3.1.25, Figure 3.1.26), performed some of the flow cytometry (Figure 3.1.7, Figure 3.1.29, Figure 3.1.30) and immunofluorescence experiments (Figure 3.1.6, Figure 3.1.8, Figure 3.1.35). Magdalena performed the proliferation (Figure 3.1.51) and scratch assays with help from Peter Zentis (Figure 3.1.49, Figure 3.1.50). Bioinformatic analysis of next generation sequencing data from ChIP-seq, ATAC-seq, 4C-seq or RNA-seq samples I had prepared were performed by Giuliano Crispantu, Petros Kolovos, Milos Nikolic, Tore Bleckwehl and Alvaro Rada-Iglesias (Figure 3.1.9 to Figure 3.1.13, Figure 3.1.22 Figure 3.1.23, Figure 3.1.36 to Figure 3.1.41, Figure 3.1.44 to Figure 3.1.47, Figure 3.1.51, Figure 3.1.52, Figure 3.1.53, Figure 3.1.54 to Figure 3.1.59). All other experiments were performed by me.

### 3.1.1 A unique BOFS patient with a *de novo* heterozygous inversion in chr6 that does not disrupt the *TFAP2A* coding sequence

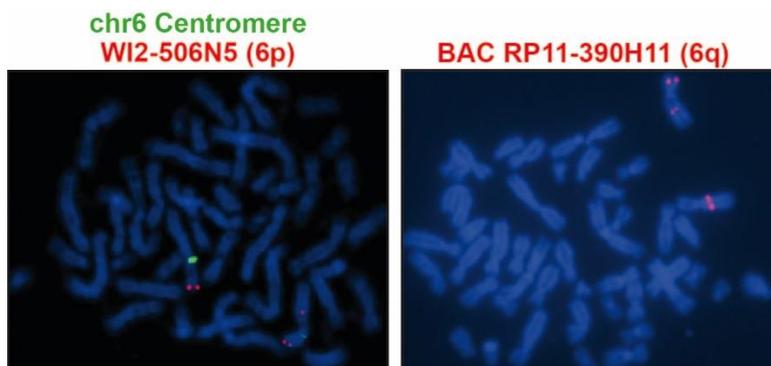
The patient described here and in Laugsch et al. (2019) has a partial BOFS phenotype, including hearing and ocular abnormalities as well as dysmorphic facial features (see methods section 2.4.1, Table 2.8 and Figure 3.1.1).



**Figure 3.1.1 Facial appearance of the patient.**

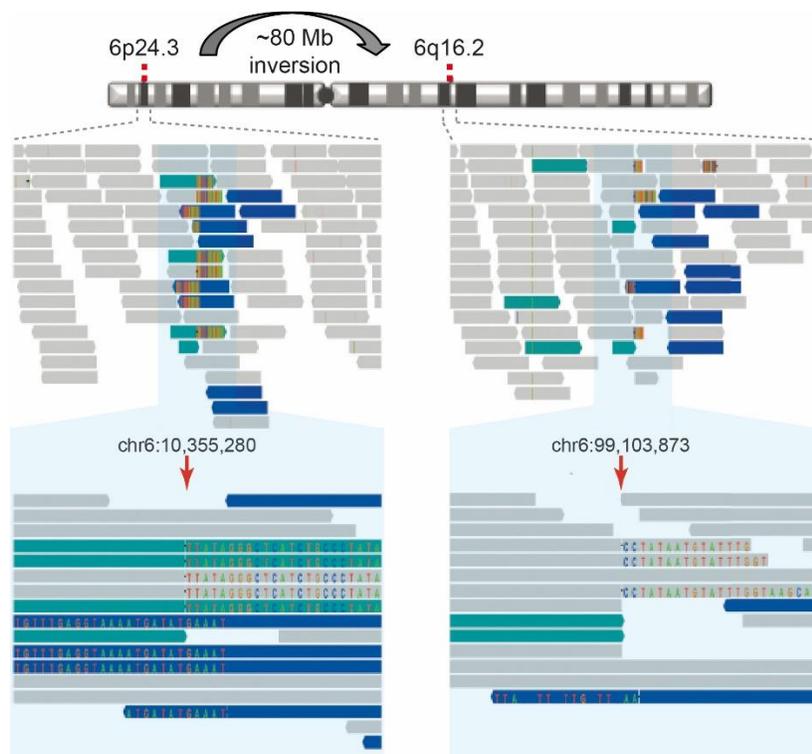
Shown at 10 (A, B) and 17 (C, D) years of age, front (A,C) and lateral (B,D) view. Published in Laugsch et al. (2019), Figure 1A.

Although BOFS is usually caused by point mutations or deletions within *TFAP2A*, sequencing of this gene and array CGH analyses in this patient did not reveal any coding mutations or unbalanced chromosomal abnormalities, respectively (data not shown). Cytogenetic analysis, though, revealed an 89Mb *de novo* heterozygous inversion in chromosome 6 ( $46,XY\ inv(6)(p24.3;q16.2)$ ). Using two-color DNA FISH (Figure 3.1.2) and paired-end whole genome sequencing (Figure 3.1.3), the 6p24.3 and 6q16.2 inversion breakpoints were mapped to positions chr6:10355280 and chr6:99103873 (hg19), respectively.



**Figure 3.1.2 Fluorescence in situ hybridization (FISH) experiments.**

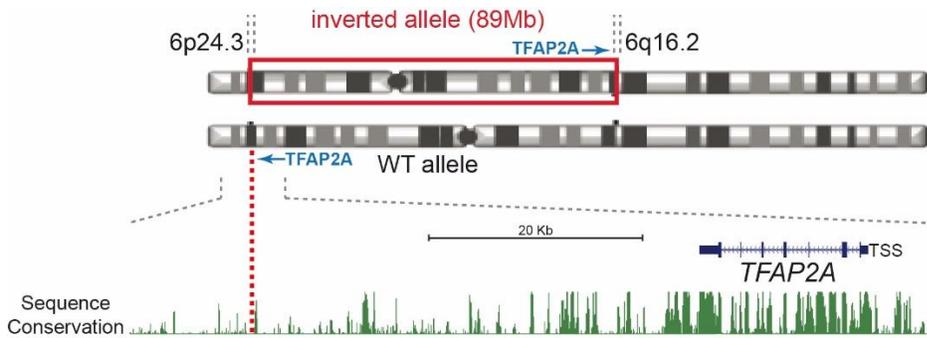
Left panel: A centromeric probe for chromosome 6 (chr6) is labelled in green. Fosmid WI2-506N5 (red) shows the expected signal at 6p on both chr6. An additional signal observed on one of the chr6 is located on 6q, indicating that this fosmid spans the inversion breakpoint. Right panel: BAC RP11-390H11 shows the expected signal on 6q on both chr6. The additional signal observed on one of the chr6 is located on 6p, indicating that this probe spans the inversion breakpoint. Published in Laugsch et al. (2019), Figure S1b.



**Figure 3.1.3 Whole-genome sequencing results around the patient's inversion breakpoints.**

Paired sequencing reads of the BOFS patient are shown around the 6p24.3 and the 6q12.6 inversion breakpoints. Reads mapping in the expected orientation are shown in grey; reads in which both members of the pair map either in forward or in reverse direction are shown in green and blue, respectively. These green and blue reads originate from DNA fragments spanning the inversion breakpoint. They also display multi-colored sequences representing nucleotides that do not match the reference genome in this position and therefore enable the identification of the inversion breakpoints (red arrows). Published in Laugsch et al. (2019), Figure S1a.

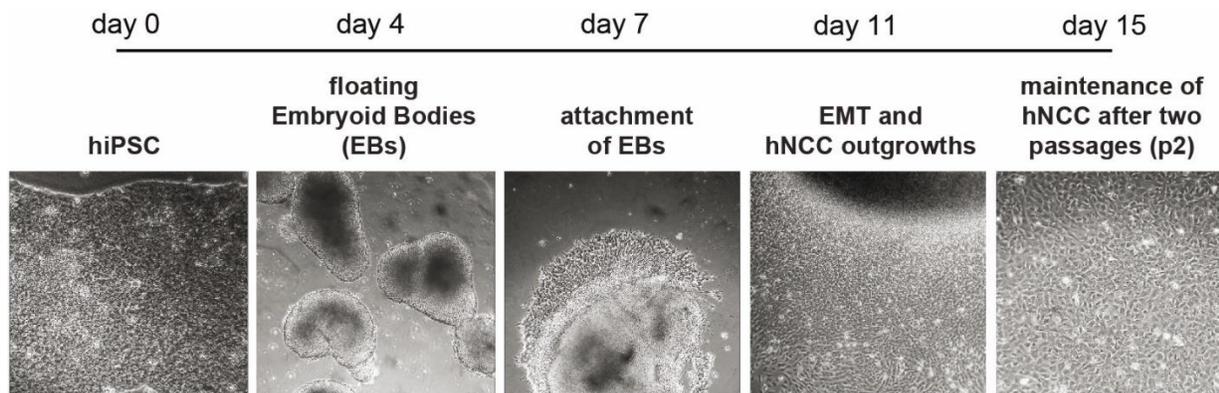
Remarkably, although the identified 6p24.3 inversion breakpoint did not disrupt the coding region of *TFAP2A*, it was localized ~40kb downstream of *TFAP2A* (Figure 3.1.4), which indicates that the inversion could have a regulatory effect on the expression of *TFAP2A* and, consequently, cause BOFS (Kleinjan and van Heyningen, 2005; Spielmann et al., 2018). This could also explain the partial BOFS phenotype found in our patient, since BOFS cases caused by deletions predicted to lead to haploinsufficient *TFAP2A* expression typically show milder phenotypes than patients harboring *TFAP2A* coding mutations predicted to create antimorphic alleles (Li et al., 2013; Milunsky et al., 2008; Milunsky et al., 2011).



**Figure 3.1.4 Schematic view of the chr6 heterozygous inversion in the BOFS patient.** The 6p24.3 inversion breakpoint is located approximately 40 kb downstream of *TFAP2A*. Published in Laugsch et al. (2019), Figure 1B.

### 3.1.2 Use of a robust *in vitro* differentiation system to model the pathological mechanisms of human neurocristopathies

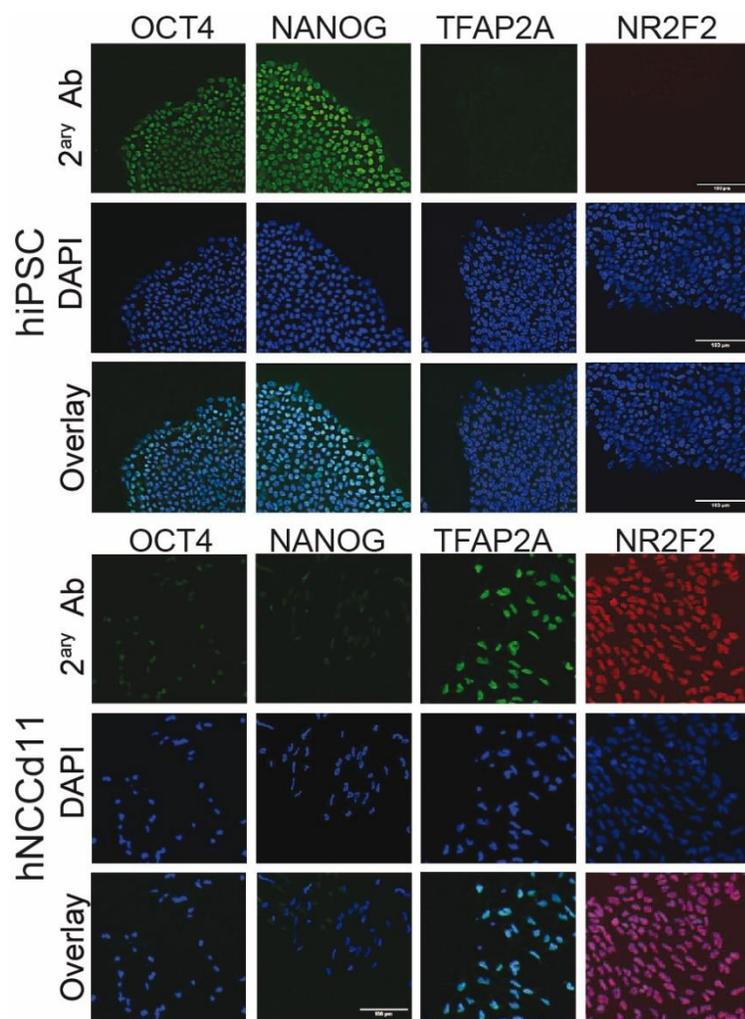
Potential pathological consequences of the heterozygous 89 Mb inversion specific to our BOFS patient were examined using a previously described *in vitro* differentiation system which allows robust differentiation of hESC/hiPSC into hNCC (Figure 3.1.5) (Bajpai et al., 2010; Prescott et al., 2015; Rada-Iglesias et al., 2012).



**Figure 3.1.5 The *in vitro* differentiation system for hNCC.** The previously described protocol (Bajpai et al., 2010; Rada-Iglesias et al., 2012) enables differentiation of hESC into hNCC and was successfully used in several WT hiPSC (here WT-B). On day 11 of differentiation (hNCCd11), hNCC were expanded for 2 passages (hNCCp2) as described in Prescott et al. (2015). Published in Laugsch et al. (2019), Figure S1c.

According to this protocol, hESC/hiPSC are first treated to form floating embryoid bodies with neuroectodermal identity, which then spontaneously attach to the surface of the culture dish and produce migratory and proliferative cranial NCC. After 11 days of differentiation, a mix of differentiated cells can be obtained (referred to as hNCCd11), consisting of hNCC and, to a lesser

extent, neuroectodermal progenitors, as shown by immunofluorescence labelling with pluripotency markers as well as neural crest markers in hiPSC and hNCCd11 (Figure 3.1.6) and flow cytometry for major neural crest markers in EBs and the mixed population on day 11 (Figure 3.1.7).

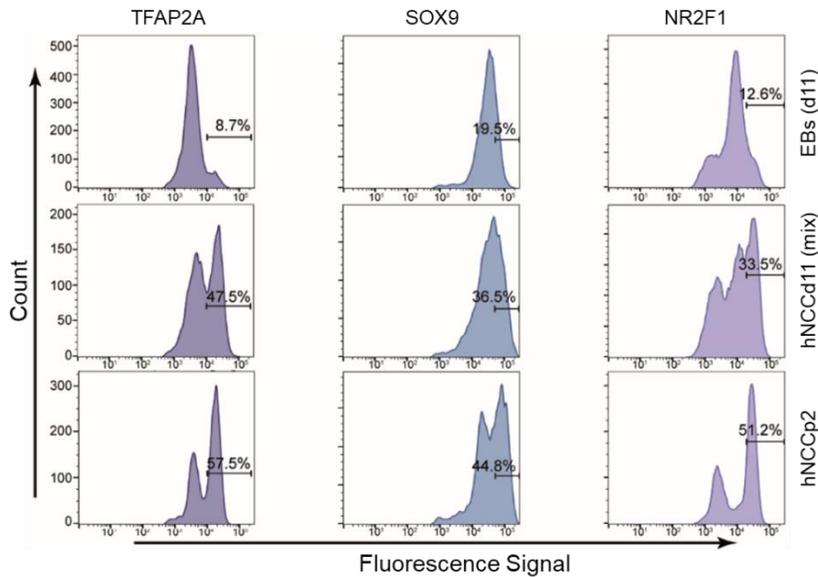


**Figure 3.1.6 Immunofluorescence for pluripotency and neural crest markers**

hiPSC and hNCCd11 were labeled for pluripotency (OCT4, NANOG) and NC (TFAP2A, NR2F2) markers using Alexa Fluor 488 (green) and 594 (red) secondary antibodies. Nuclei were stained with DAPI (blue). Published in Laugsch et al. (2019), Figure S1d.

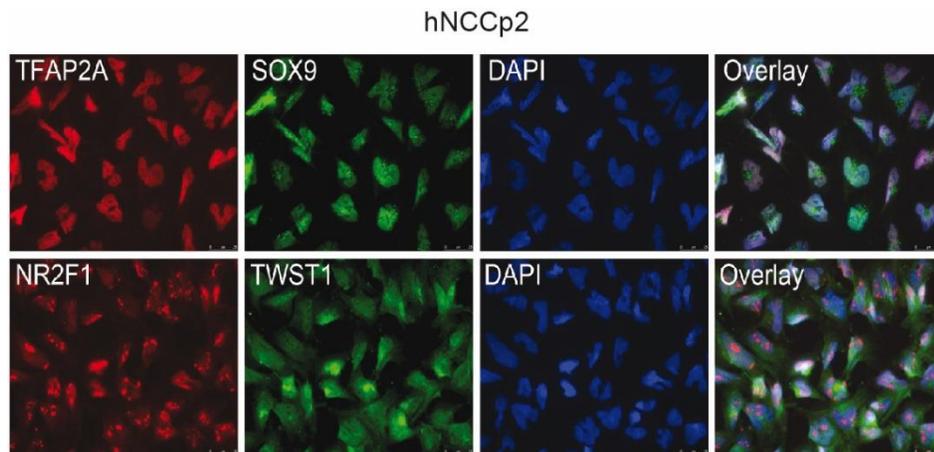
On day 11 of differentiation, hNCCd11 can be expanded (Prescott et al., 2015), which leads to a more pure and homogeneous hNCC population after two passages (referred to as hNCCp2), in which neuroectodermal progenitors are almost completely eliminated, as shown by flow cytometry (Figure 3.1.7) and immunofluorescence (Figure 3.1.8) for major neural crest markers in pure hNCCp2.

## Results



**Figure 3.1.7 Expression of major neural crest markers analyzed by FACS.**

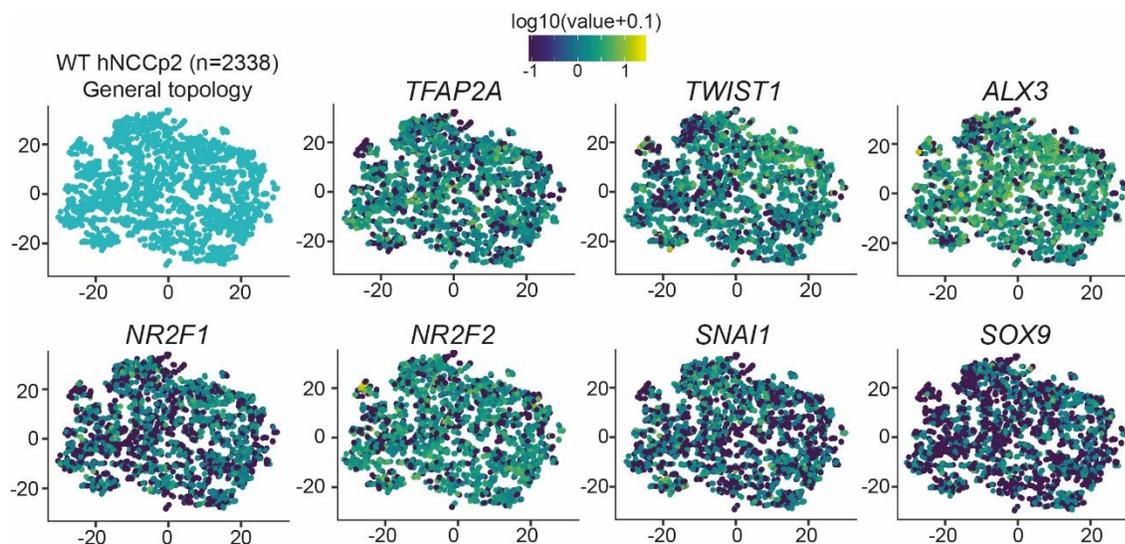
Flow cytometry was used to analyze expression of major NC markers upon differentiation of WT hiPSC into hNCC. Embryoid bodies that were still floating on day 11 of differentiation were analyzed separately from the mixed population of attached EBs and hNCC, and hNCC after 2 passages (hNCCp2). Published in Laugsch et al. (2019), Figure S1e.



**Figure 3.1.8 Immunofluorescence for neural crest markers in hNCCp2**

hNCCp2 were labeled with primary antibodies against major NC markers (TFAP2A, SOX2, NR2F1, NR2F2) and Alexa Fluor 488 (green) and 594 (red) secondary antibodies. Nuclei were stained with DAPI (blue). Published in Laugsch et al. (2019), Figure S1f.

Further characterization of the hNCCp2 by single-cell RNA whole genome sequencing (scRNA-seq) confirmed that they represent a fairly homogeneous cell population in which multiple cranial NCC and craniofacial ectomesenchyme markers are widely expressed (Figure 3.1.9).

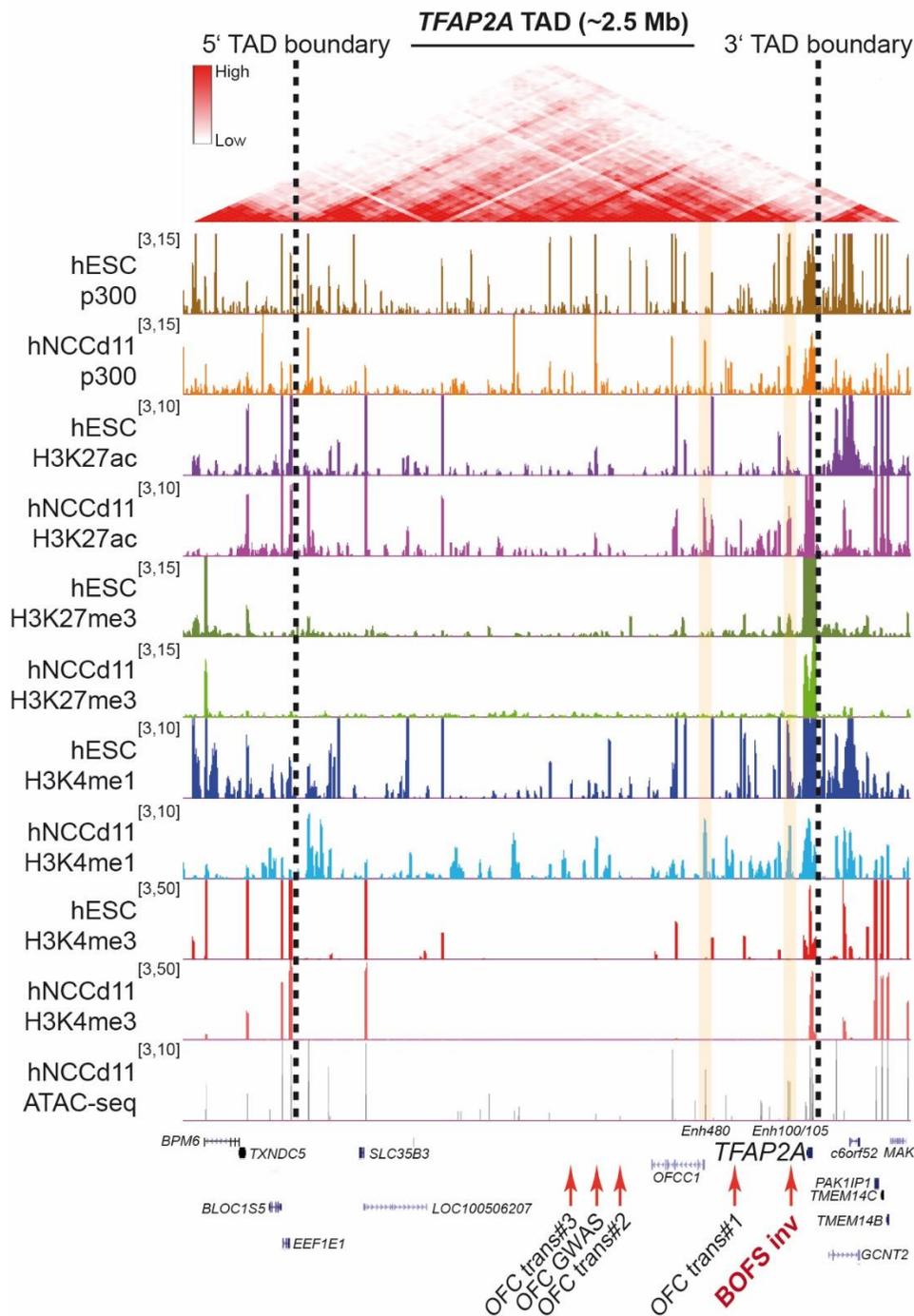


**Figure 3.1.9 Single cell RNA sequencing in WT hNCCp2**

scRNA-seq data were generated from 2338 WT hNCCp2. t-distributed Stochastic Neighbor Embedding (tSNE) visualization of all analyzed hNCC did not reveal the presence of distinct cell populations (General topology plot). Expression of selected ectomesenchyme and cranial NC markers is shown for all analyzed hNCC. Published in Laugsch et al. (2019), Figure S1g.

### 3.1.3 Epigenomic and topological characterization of the *TFAP2A* locus regulatory landscape in hNCC

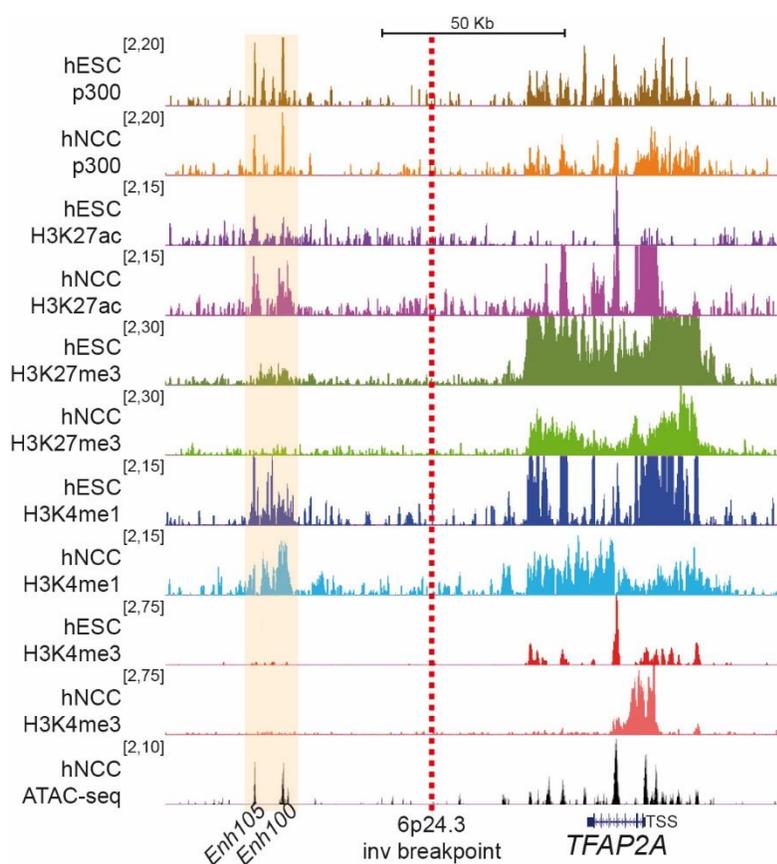
Using the *in vitro* differentiation system, we first evaluated the *TFAP2A* regulatory landscape in hNCC by combining previously generated epigenomic maps from hESC and hNCCd11 (Rada-Iglesias et al., 2012; Rada-Iglesias et al., 2011) with newly generated ATAC-seq data from hNCCd11 and publically available Hi-C data from hESC (Dixon et al., 2015), visualized with the 3D Genome Browser (Wang et al., 2018) (Figure 3.1.10, Figure 3.1.11).



**Figure 3.1.10 Epigenetic profile of hESC and hNCCd11 around the *TFAP2A* locus.**

Publicly available Hi-C data from hESC (Dixon et al., 2015) were combined with previously published ChIP-seq data for p300 and various histone modifications from hESC and hNCCd11 (Rada-Iglesias et al., 2012; Rada-Iglesias et al., 2011) and with newly generated ATAC-seq data from hNCCd11. ChIP-seq and ATAC-seq signals were normalized as RPGC (reads per genomic content) using five bp bins. *TFAP2A* is located at the 3' end of a large topologically associated domain (TAD), whose boundaries are shown as black dashed lines. Selected hNCC active enhancers located within the *TFAP2A*-TAD are highlighted in yellow: Enh480 represents an hNCC active enhancer located ca. 480 kb downstream of *TFAP2A*; Enh100/105 denotes a couple of hNCC-active enhancers located ca. 100 and 105 kb downstream of *TFAP2A*. The red arrows represent the locations of the inversion breakpoint in our BOFS patient, the translocation breakpoints identified in three different OFC patients (Davies et al., 2004) as well as a risk-haplotype associated with OFC by a GWAS (Yu et al., 2017). Published in Laugsch et al. (2019), Figure 1C.

According to these data, *TFAP2A* is located at the 3' end of a large (>2.5 Mb) gene-poor TAD (hereafter referred to as *TFAP2A*-TAD) which harbors a large number of enhancers that are active in hNCCd11 (i.e. high p300, H3K27ac, H3K4me1 and ATAC-seq). We further noticed that some of these active hNCC enhancers were initially poised in hESC (i.e. marked with p300, H3K27me3, H3K4me1) (Rada-Iglesias et al., 2011), e.g. Enh100 and Enh105, while others became active *de novo* in hNCC, e.g. Enh480. Interestingly, the 6p24.3 BOFS inversion breakpoint is situated between *TFAP2A* and most of the active hNCC enhancers within the *TFAP2A*-TAD (Figure 3.1.11). In addition, translocation breakpoints reported in three different patients with non-syndromic orofacial clefting (OFC) (Davies et al., 2004) as well as a risk-haplotype that was previously connected to OFC by a GWAS (Yu et al., 2017) were also located within the *TFAP2A*-TAD (see Figure 3.1.10). Hence, it is probable that this region contains regulatory information which, when disrupted, can lead to neurocristopathies as a result of changes in *TFAP2A* expression.

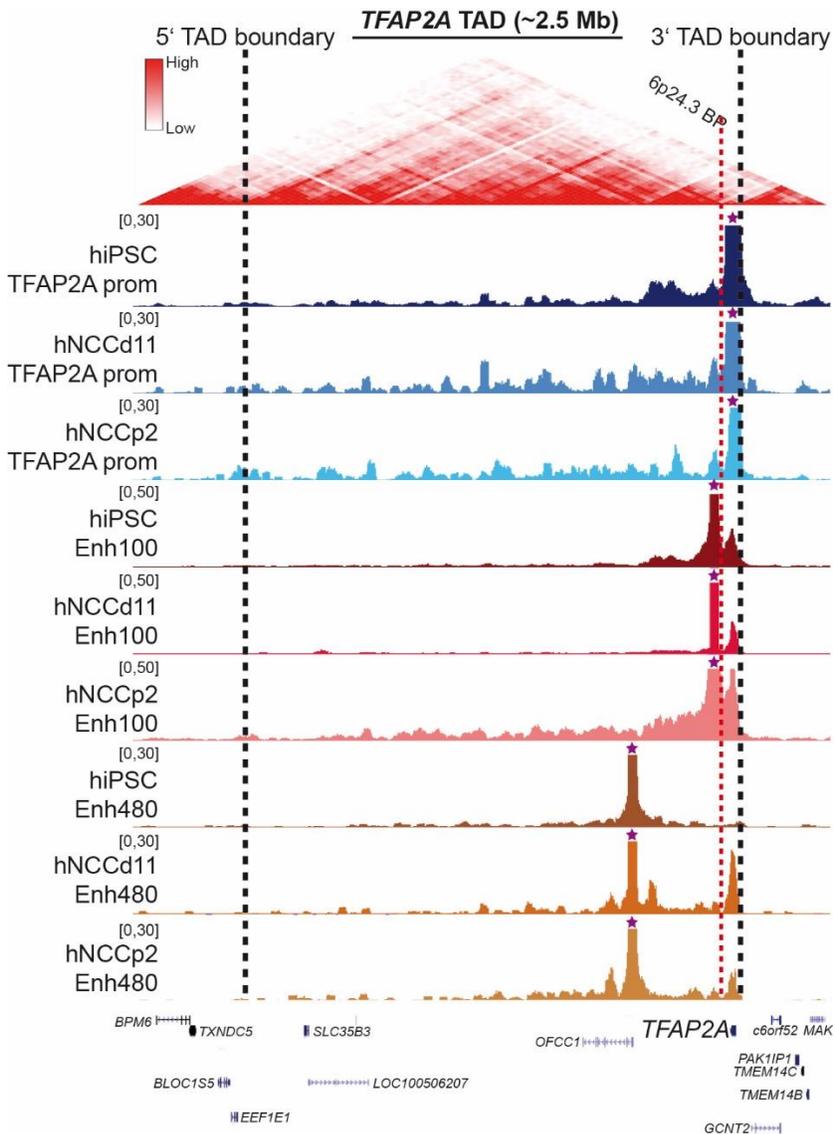


**Figure 3.1.11 Close-up view of the epigenetic profile.**

Zoom-in view of the epigenomic data shown in the previous figure, centered on the 6p24.3 inversion breakpoint identified in our BOFS patient (red dashed line). ChIP-seq and ATAC-seq signals were normalized as RPGC (reads per genomic content) using five bp bins. Published in Laugsch et al. (2019), Figure 1D.

## Results

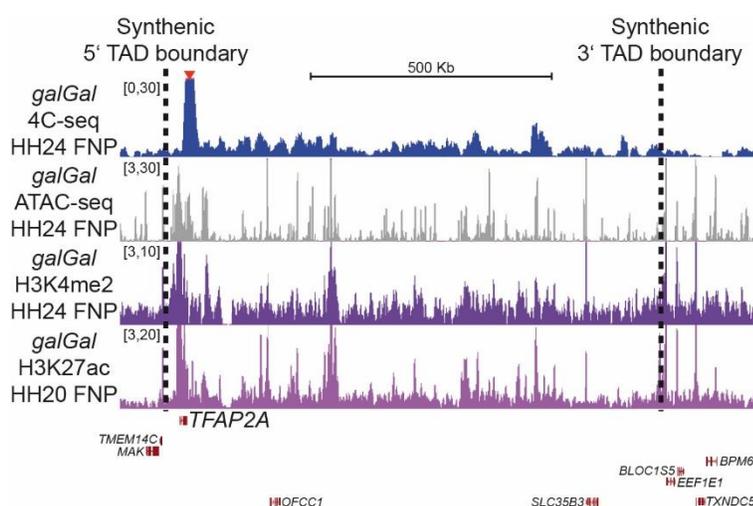
To gather additional evidences supporting the existence and functional relevance of the *TFAP2A*-TAD, 4C-seq experiments were performed (Stadhouders et al., 2013) in WT hiPSC, hNCCd11 and hNCCp2 (Figure 3.1.12). By using the *TFAP2A* promoter as a viewpoint we recapitulated the Hi-C data and confirmed the boundaries of the *TFAP2A*-TAD in hNCC. We also used two enhancers (i.e. Enh480 and Enh100) as viewpoints that were active in hNCC and located downstream of *TFAP2A*. These revealed strong and highly specific interactions between the individual enhancers and the *TFAP2A* promoter.



**Figure 3.1.12** 4C-seq profiles around the *TFAP2A* locus.

Publicly available Hi-C data from hESC (Dixon et al., 2015) used to identify the large TAD, in which *TFAP2A* is located. 4C-seq data were generated from WT hiPSC, hNCCd11, and hNCCp2 using the *TFAP2A* promoter, Enh100 and Enh480 as viewpoints (purple stars). 4C-seq signals were normalized as RPMs (reads per million) and a smoothing window of eight pixels was applied for visualization. The red dashed line marks the location of the 6p24.3 inversion breakpoint. Published in Lausch et al. (2019), Figure 1E.

Interestingly, interaction between *TFAP2A* and Enh100, bound by H3K27me3 in hiPSC (Figure 3.1.10, Figure 3.1.11), was also detected in hiPSC while this was not the case for interaction between *TFAP2A* and Enh480, which was not bound by H3K27me3 in hiPSC and became activated *de novo* in hNCC (Figure 3.1.10). This observation is consistent with recent work from our laboratory according to which poised enhancers that are marked by H3K27me3 physically interact with their target genes in pluripotent cells already (Cruz-Molina et al., 2017). Additionally, while both Enh100 and Enh480 interacted similarly with *TFAP2A* in hNCCd11, these interactions became particularly prominent in hNCCp2 for Enh100 but not for Enh480 (Figure 3.1.12), which suggest that these enhancers might be dynamically used during hNCC development.



**Figure 3.1.13 4C-seq, ATAC-seq and ChIP-seq profiles from chicken embryonic FNP.**

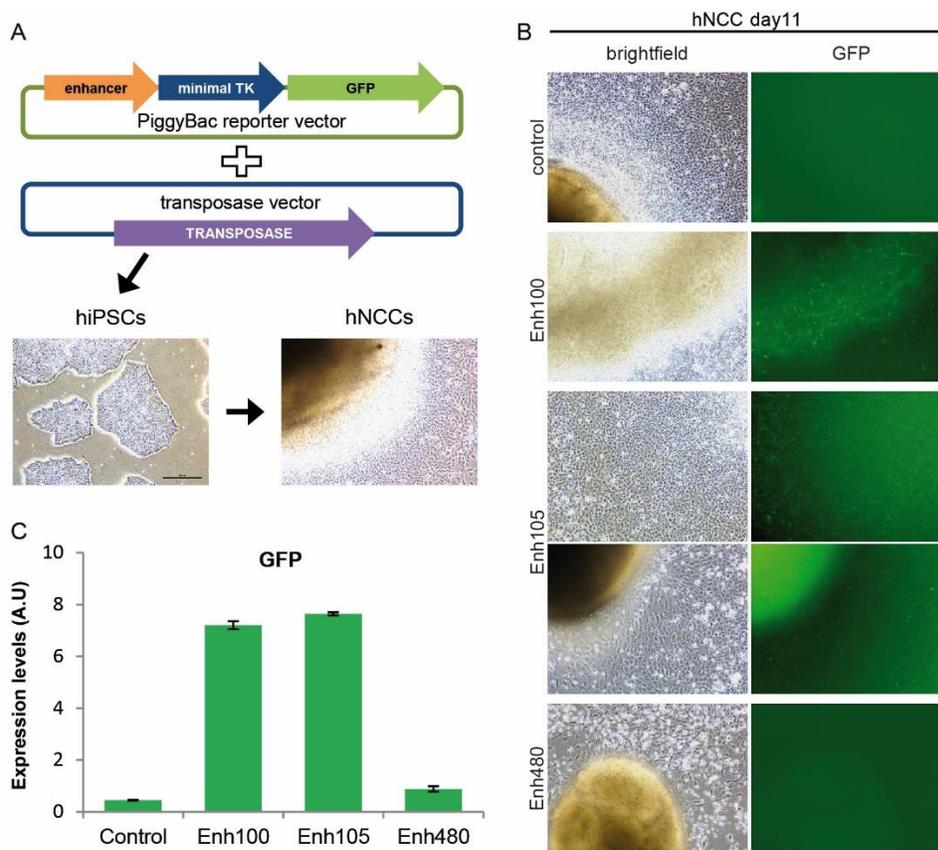
Data is shown around the *TFAP2A* locus in the chicken genome. Black dashed lines show the approximate location of the synthetic boundaries of the human *TFAP2A*-TAD. 4C-seq was performed using a region at the 3' end of *TFAP2A* as viewpoint (red arrow). 4C-seq, ATAC-seq and H3K4me2 ChIP-seq were generated from stage HH24 chicken embryos, H3K27ac ChIP-seq was previously generated from stage HH20 chicken embryos (Rada-Iglesias et al., 2012). 4C-seq signals were normalized as RPMs (reads per million) and a smoothing window of eight pixels was applied for visualization. ChIP-seq and ATAC-seq signals were normalized as RPGC (reads per genomic content) using five bp bins. Published in Lausch et al. (2019), Figure 1F.

To further support the functional relevance of the *TFAP2A*-TAD we used chicken embryonal frontonasal prominences (FNP) to study the *TFAP2A* locus in NCC-derived facial mesenchyme (Figure 3.1.13). In the process, a 4C-seq experiment with the viewpoint located at the chicken *TFAP2A* gene indicated, that the *TFAP2A*-TAD is evolutionarily conserved and its synthetic boundaries are similar to those in human NCC. Data from ATAC-seq and ChIP-seq for H3K27ac and H3K4me2 in the FNP further identified many active enhancers within the chicken *TFAP2A*-TAD. Overall, the *TFAP2A* regulatory domain displays highly conserved epigenomic and topological features *in vivo*, which supports its physiological relevance (Dixon et al., 2012; Gomez-Marin et al., 2015).

### 3.1.4 The *TFAP2A*-TAD contains regulatory elements that are essential for *TFAP2A* expression in hNCC

Next, we evaluated whether, within the *TFAP2A*-TAD, the region 5' of the 6p24.3 inversion breakpoint, found in our BOFS patient, contains *cis*-regulatory information required for correct *TFAP2A* expression in hNCC. First, we employed *in vitro* and *in vivo* reporter assays to examine the activity of the previously mentioned enhancers (Enh100, Enh105, Enh480) during NCC development.

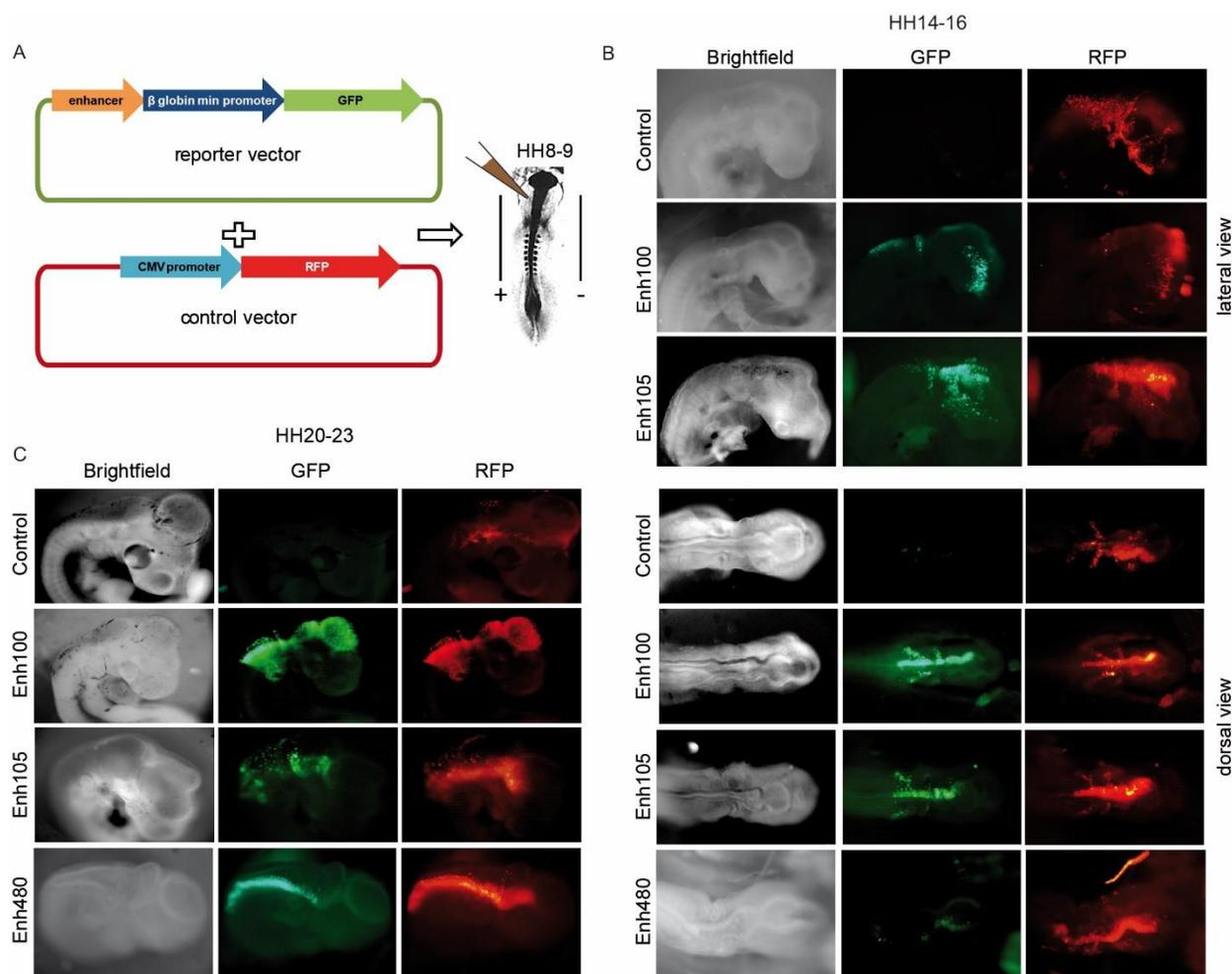
The *in vitro* reporter system revealed that Enh100 and Enh105 were able to strongly drive GFP expression in hNCC outgrowths, while the capacity of Enh480 to activate GFP expression was considerably lower (Figure 3.1.14).



**Figure 3.1.14** *In vitro* reporter assay for *TFAP2A* enhancers in hNCC.

(A) Schematic diagram illustrating the *in vitro* reporter assay performed in hNCC. Human enhancer sequences (Enh100, Enh105, Enh480) were cloned into a PiggyBAC reporter vector in which a minimal promoter (min TK promoter) drives GFP expression when enhanced by an active enhancer. The resulting reporter vectors and a transposase-expressing vector were co-transfected into WT hiPSC. Following antibiotic selection, a population of hiPSC in which the reporter vector had been integrated into the human genome was differentiated into hNCC. GFP signal was imaged on day 11 of differentiation. (B) Representative GFP signals observed in hNCC differentiated from hiPSC with integrated reporter vectors that were either empty (control) or contained sequences of Enh100, Enh105 or Enh480. (C) GFP levels measured by RT-qPCR in the hNCC shown in (B). Expression values were normalized to three housekeeping genes (*ACTB*, *EEF2*, *GAPDH*). Error bars represent standard deviation from three technical replicates. Published in Laugsch et al. (2019), Figure S2a-c.

Accordingly, the *in vivo* reporters performed in chicken embryos displayed strong and reproducible GFP expression in the dorsal neural tube, migrating cranial NCC and facial mesenchyme for Enh100 and Enh105 while weaker expression was observed for Enh480 and only within the dorsal neural tube (Figure 3.1.15). Together with the previous 4C-seq results (Figure 3.1.12), these reporter assays suggest that while Enh480 might contribute to *TFAP2A* expression during NCC specification in the dorsal neural tube, Enh100 and Enh105 might play a more prevalent role in controlling *TFAP2A* expression across different stages of NCC development.

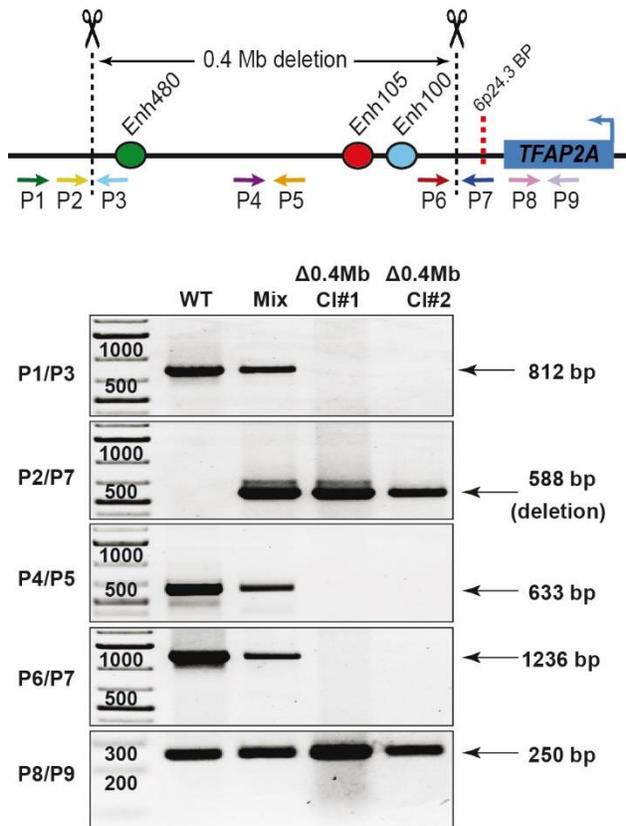


**Figure 3.1.15** *In vivo* reporter assays for *TFAP2A* enhancers in chicken embryos.

(A) Schematic diagram illustrating the *in vivo* reporter assay performed in chicken embryos. Human enhancer sequences (Enh100, Enh105, Enh480) were cloned into a reporter vector in which a minimal promoter (min  $\beta$ -globin promoter) drives GFP expression in the presence of an active enhancer. Another vector, in which a strong promoter (CMV promoter) constitutively drives RFP expression, was used as a positive control (control vector) for electroporation. Both vectors were co-electroporated into the anterior neural tube of stage HH9-10 chicken embryos to evaluate GFP and RFP expression at later developmental stages. (B-C) Representative images of (B) stage HH14-16 and (C) stage HH20-23 chicken embryos co-electroporated with the control vector (RFP) and the reporter vector (GFP) (either empty (control) or containing one of the human enhancer sequences (Enh100, Enh105, Enh480)). (B) For stage HH14-16 two embryos are shown for each enhancer: one in lateral (upper panel) and one in dorsal view (lower panel). (C) Embryos of stage HH20-23 are shown in lateral (control, Enh100, Enh105) or dorsal (Enh480) view. Published in Laugsch et al. (2019), Figure 2A-C and S2d+e.

## Results

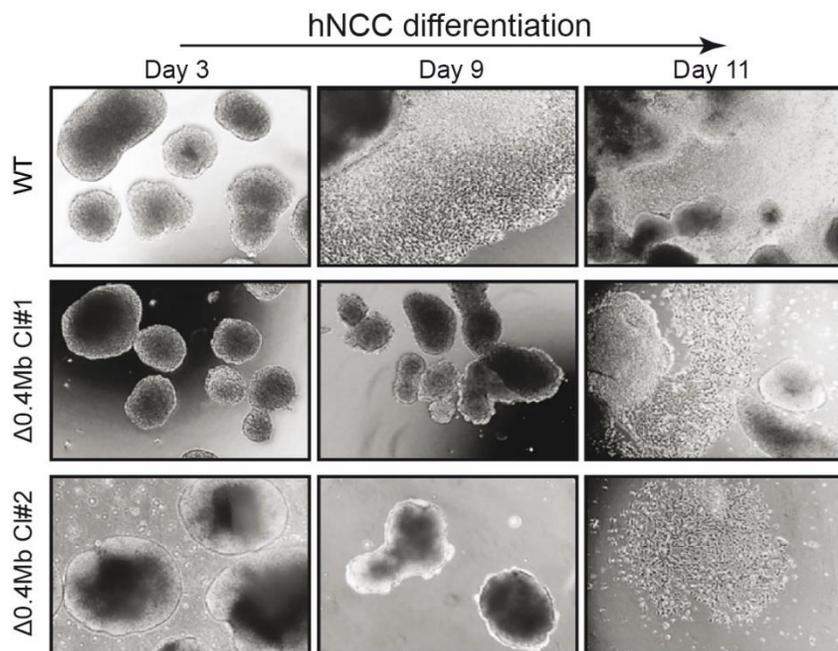
To evaluate the functional relevance of the previous enhancers more directly, we used CRISPR/Cas9 to engineer a deletion of 0.4 Mb in WT hiPSC, spanning several hNCC active enhancers, including Enh100/105 and Enh480, and located 5' of the 6p24.3 inversion breakpoint (Figure 3.1.16).



**Figure 3.1.16 CRISPR/Cas9 mediated 0.4 Mb deletion within the *TFAP2A*-TAD.**

Upper panel: Schematic diagram illustrating the CRISPR/Cas9 strategy used to engineer a 0.4 Mb deletion within the *TFAP2A*-TAD together with the location of the PCR primers employed to determine the presence of WT and 0.4 Mb deletion alleles. WT hiPSC were co-transfected with a pair of vectors containing the Cas9 nuclease gene for overexpression with a guide RNA (gRNA) targeted to regions located 5' of Enh480 and 3' of Enh100, respectively (black dashed lines). The relative location of the 6p24.3 inversion breakpoint is denoted by a red dashed line. Different primer pairs were specific for the WT allele (P1/P3, P4/P5, P6/P7), the 0.4 Mb deletion (P2/P7) or for both (P8/P9). Lower panel: Genotyping results obtained by using the PCR primers described above are shown for WT hiPSC, a mixed population of CRISPR/Cas9 targeted hiPSC, and two hiPSC clonal lines homozygous for the 0.4 Mb deletion ( $\Delta 0.4\text{Mb}$  Cl#1,  $\Delta 0.4\text{Mb}$  Cl#2). Published in Laugsch et al. (2019), Figure S2f+g.

When we assessed the hNCC differentiation potential of two different clonal hiPSC lines homozygous for the 0.4 Mb deletion ( $\Delta 0.4\text{Mb}$  Cl#1,  $\Delta 0.4\text{Mb}$  Cl#2), they both revealed impaired attachment of floating EBs to the surface of the culture plate and less hNCC outgrowths on day 11 of the *in vitro* differentiation compared to the WT isogenic control (Figure 3.1.17), indicating compromised and/or delayed hNCC specification.

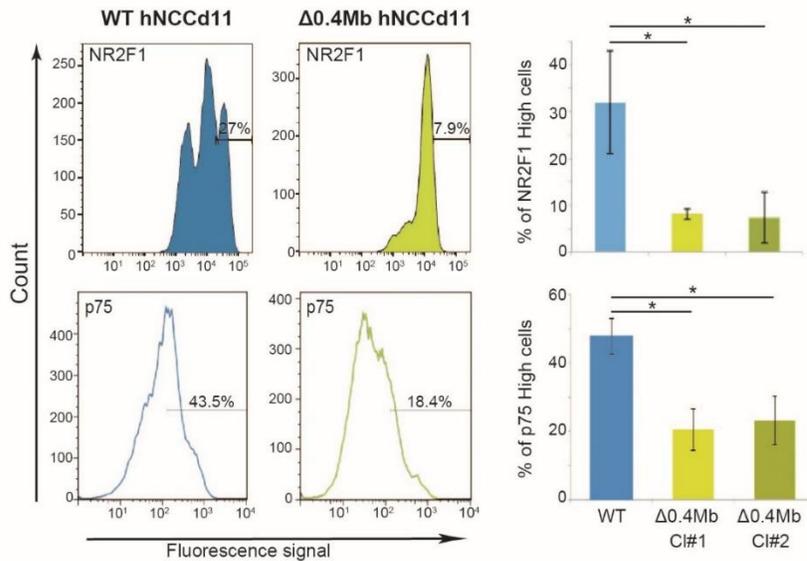


**Figure 3.1.17 Compromised hNCC differentiation of hiPSC carrying a homozygous 0.4 Mb deletion.**

Representative images illustrating the compromised hNCC differentiation capacity of two different clonal hiPSC lines ( $\Delta 0.4\text{Mb C}\#1$ ,  $\Delta 0.4\text{Mb C}\#2$ ) with a homozygous deletion of 0.4 Mb ( $\Delta 0.4\text{Mb}$ ) within the *TFAP2A-TAD* in comparison with their parental WT hiPSC line. The  $\Delta 0.4\text{Mb}$  hiPSC lines formed morphologically normal EBs (Day 3), but showed a delay in EB attachment to the plate surface and in the emergence of hNCC outgrowths (Day 9), which ultimately resulted in reduced hNCC numbers (Day 11). Published in Laugsch et al. (2019), Figure 2E.

Correspondingly, the  $\Delta 0.4\text{Mb}$  hiPSC produced fewer hNCC as quantified by flow cytometry using a couple of hNCC markers (i.e. NR2F1, p75) (Better et al., 2010; Jamal et al., 2018; Rada-Iglesias et al., 2012) (Figure 3.1.18).

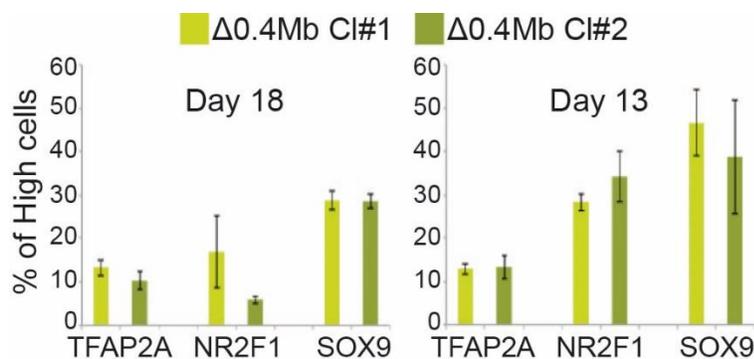
## Results



**Figure 3.1.18 FACS analysis of neural crest markers in hNCCd11 derived from WT and 0.4 Mb deletion hiPSC lines.**

Two  $\Delta 0.4\text{Mb}$  clones and their isogenic WT hiPSC line were differentiated into hNCC and processed on day 11 of differentiation (hNCCd11). The % of cells expressing high levels of NR2F1 and p75 was quantified by flow cytometry. Representative experiments in WT and  $\Delta 0.4\text{Mb}$  hNCCd11 are shown in the left and middle panels, respectively. The average result from six biological replicates of WT and four biological replicates for each  $\Delta 0.4\text{Mb}$  clone are presented in the right panels. Error bars represent standard deviations. P-values were calculated using Student's t-test (\*p-value < 0.01; ns = not significant). Published in Laugsch et al. (2019), Figure 2F and S2h.

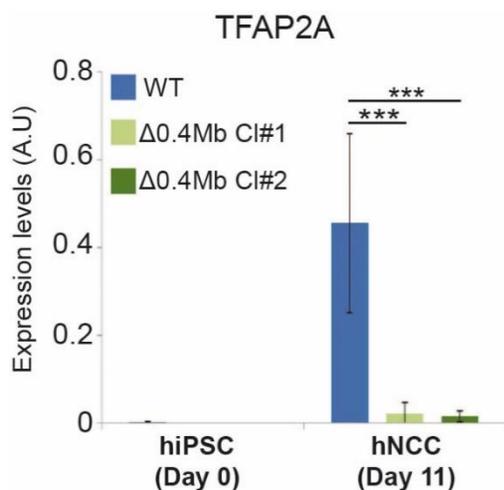
To judge if the 0.4 Mb deletion irreversibly impaired or, alternatively, only delayed hNCC specification,  $\Delta 0.4\text{Mb}$  hiPSC were retained for 13 to 18 days under hNCC differentiation conditions. Although in some cases hNCC specification seemed to be severely compromised (Figure 3.1.19 left panel), others showed increasing attachment of floating EBs with time and subsequent appearance of hNCC outgrowths (Figure 3.1.19 right panel, compare to Figure 3.1.17), indicating a loss of robustness in the hNCC differentiation process. Overall, these data suggest that the 0.4 Mb deletion delays but does not fully abrogate hNCC specification.



**Figure 3.1.19 hNCC differentiation of  $\Delta 0.4\text{Mb}$  clones for up to 18 days evaluated by flow cytometry.**

Two  $\Delta 0.4\text{Mb}$  clones ( $\Delta 0.4\text{Mb}$  Cl#1,  $\Delta 0.4\text{Mb}$  Cl#2) were maintained in hNCC differentiation conditions for up to 18 days (left panel) or until hNCC outgrowths emerged on day 13 (right panel). The % of cells representing high levels of TFAP2A, NR2F1 and SOX9 was quantified by FACS. The following number (n) of biological replicated were measured: Day18 TFAP2A and SOX9 Cl#1 n=3, Cl#2 n=3, NR2F1 Cl#1 n=2, Cl#2 n=2, Day 13 TFAP2A, NR2F1, and SOX9 Cl#1 n=4, Cl#2 n=4. Error bars represent standard deviation. Published in Laugsch et al. (2019), Figure S2i.

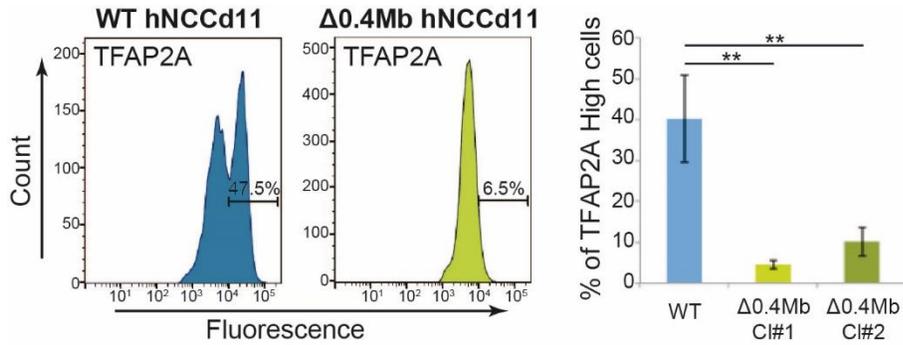
Notably, the 0.4 Mb deletion dramatically reduced *TFAP2A* induction during hNCC formation, as shown on the RNA and protein level by RT-qPCR (Figure 3.1.20) and flow cytometry (Figure 3.1.21), respectively. *TFAP2A* levels remained low even upon prolonged maintenance of  $\Delta 0.4\text{Mb}$  clones in hNCC differentiation conditions (compare Figure 3.1.19 to Figure 3.1.21).



**Figure 3.1.20 *TFAP2A* expression levels measured by RT-qPCR in WT and 0.4 Mb deletion.**

RNA of two  $\Delta 0.4\text{Mb}$  clones ( $\Delta 0.4\text{Mb}$  Cl#1,  $\Delta 0.4\text{Mb}$  Cl#2) and their isogenic WT hiPSC line was analyzed in hiPSC (day 0) and on day 11 of hNCC differentiation (Day 11). Expression values of *TFAP2A* were normalized to three housekeeping genes (ACTB, *EEF2*, *GAPDH*). The error bars represent standard deviation from twelve measurements in WT (technical triplicates in four different biological replicates) and six measurements in each of the two  $\Delta 0.4\text{Mb}$  clones (technical triplicates in four different biological replicates), both in hiPSC and in hNCCd11. P-values were calculated using Student's t-test (\*\*\*)p-value<0.0001). Published in Laugsch et al. (2019), Figure 2G.

## Results



**Figure 3.1.21 FACS analysis for TFAP2A expression in hNCCd11 derived from WT and 0.4 Mb deletion hiPSC lines.**

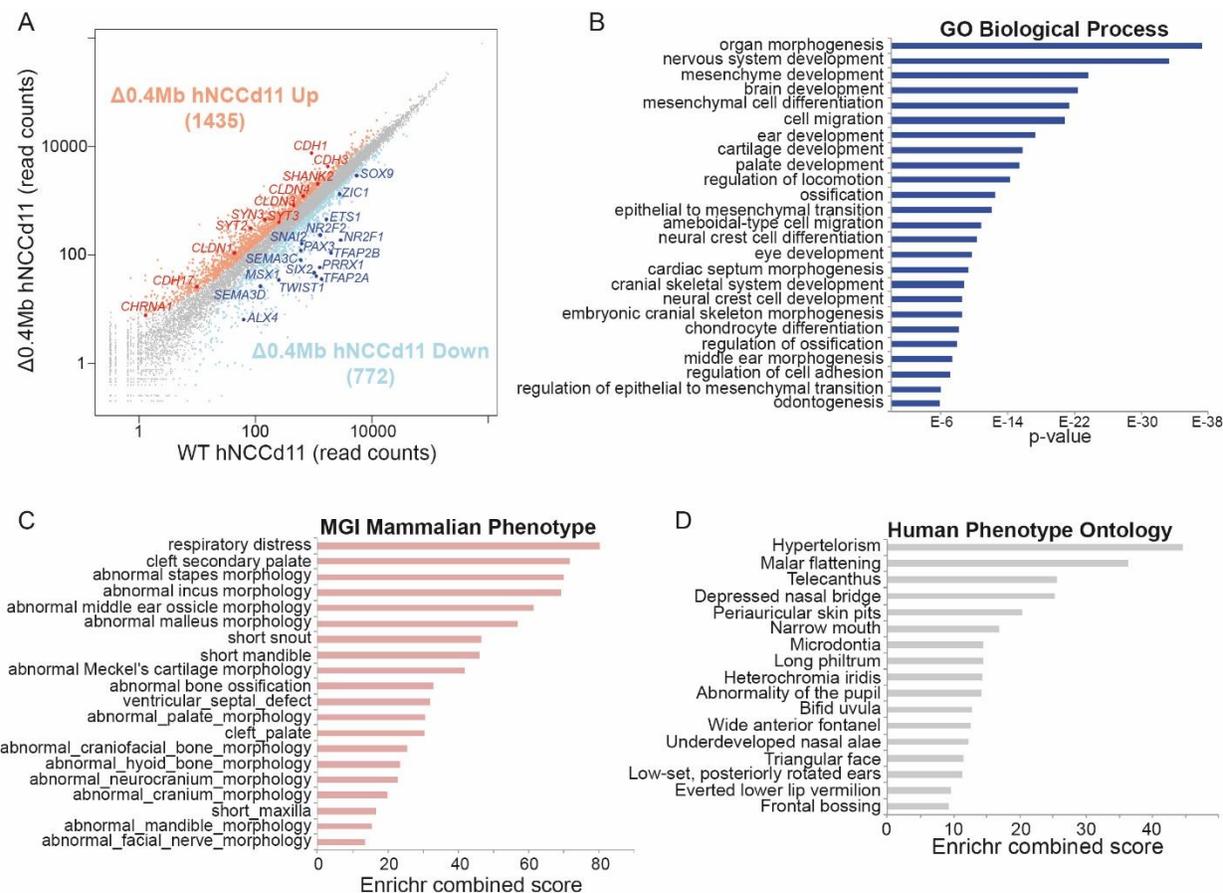
Two  $\Delta 0.4\text{Mb}$  clones and their isogenic WT hiPSC line were differentiated into hNCC and processed on day 11 of differentiation (hNCCd11). The % of cells expressing high levels of TFAP2A was quantified by flow cytometry. Representative experiments in WT and  $\Delta 0.4\text{Mb}$  hNCCd11 are shown in the left and middle panels, respectively. The average result from six biological replicates of WT and four biological replicates for each  $\Delta 0.4\text{Mb}$  clone are presented in the right panels. Error bars represent standard deviations. P-values were calculated using Student's t-test (\*\*p-value < 0.001). Published in Laugsch et al. (2019), Figure 2H.

For a more general evaluation of how this strong decrease of TFAP2A levels could disturb the hNCC differentiation process, we performed RNA-seq experiments on day 11 of the *in vitro* differentiation in WT and 0.4 Mb deletion cells.

Differential gene expression analysis between WT and  $\Delta 0.4\text{Mb}$  hNCCd11 showed that a many genes were dysregulated in the  $\Delta 0.4\text{Mb}$  cells (772 downregulated and 1435 upregulated). These included, as anticipated, a severe downregulation of *TFAP2A* (Figure 3.1.22A). Notably, none of the genes directly flanking the 0.4 Mb deletion and *TFAP2A* displayed differential expression between WT and  $\Delta 0.4\text{Mb}$  hNCCd11 (Figure 3.1.23). Together with the previous 4C-seq data (Figure 3.1.12), these results strongly support that the 0.4 Mb deleted region comprised enhancers which are specifically regulating *TFAP2A* expression.

Furthermore, among those genes downregulated in the  $\Delta 0.4\text{Mb}$  cells (Figure 3.1.22A) there were major regulators of the hNCC differentiation process involved in hNCC specification (e.g. *ZIC1*, *MSX1*, *PAX3*), epithelial to mesenchymal transition (EMT) (e.g. *TWIST1*, *SNAI2*, *PRRX1*), migration (*SEMA3D*, *SEMA3C*), or craniofacial morphogenesis (e.g. *ALX4*, *SIX2*) (Figure 3.1.22B-D).

On the other hand, among those genes upregulated in the  $\Delta 0.4\text{Mb}$  cells (Figure 3.1.22A) there were genes involved in the maintenance of epithelial identity and EMT inhibition (e.g. *CDH1*, *CLDN3*, *CLDN4*).

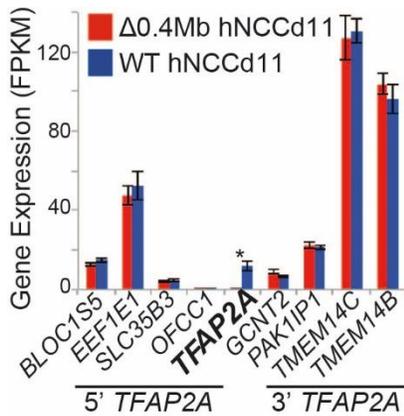


**Figure 3.1.22 RNA-seq in hNCCd11 derived from WT and  $\Delta 0.4\text{Mb}$  clones**

RNA samples were derived in duplicates for each  $\Delta 0.4\text{Mb}$  clone ( $\Delta 0.4\text{Mb Cl\#1}$ ,  $\Delta 0.4\text{Mb Cl\#2}$ ) and in triplicates for the parental WT hNCCd11. **(A)** Human genes were plotted according to the average normalized RNA-seq read counts in WT hNCCd11 and  $\Delta 0.4\text{Mb}$  hNCCd11. Genes considered as significantly up- or downregulated in  $\Delta 0.4\text{Mb}$  hNCC compared to WT hNCC are shown in red and blue, respectively. **(B-D)** Significantly downregulated genes in  $\Delta 0.4\text{Mb}$  hNCCd11 were functionally annotated according to the following terms: **(B)** “Gene Ontology Biological Process”, selected terms among those with a p-value  $< 10^{-5}$  are shown; **(C)** “MGI Mammalian Phenotype”, selected terms among the 50 most significant overrepresented ones are shown; and **(D)** “Human Phenotype Ontology”, selected terms with an Enrichr combined score  $> 10$  are shown. Published in Laugsch et al. (2019), Figure 2I+J and S2k+l.

Overall, our results confirm, in agreement with previous reports in both mouse and human cells (Brewer et al., 2004; Tchieu et al., 2017), that high levels of *TFAP2A* are important but not essential for hNCC specification. Most importantly, our data demonstrate that the *TFAP2A*-TAD and more specifically the 0.4 Mb deleted region contains *cis*-regulatory elements that are necessary for correct *TFAP2A* expression during the hNCC differentiation process.

## Results



**Figure 3.1.23** Expression levels of *TFAP2A* and immediately flanking genes measured by RNA-seq in WT and 0.4 Mb deletion hNCCd11.

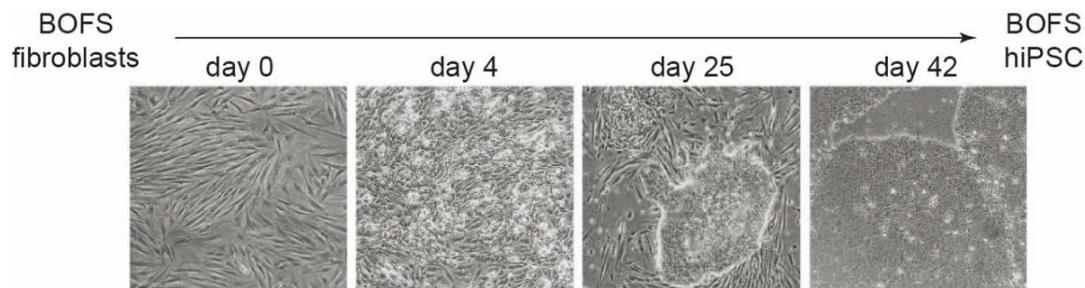
Average expression levels are shown as FPKM in hNCCd11. Error bars represent standard deviation for measurement of 3 biological replicates for WT and 4 biological replicates for the 0.4 Mb deletion. \*Considered as significantly differentially expressed. Published in Laugsch et al. (2019), Figure S2j.

### 3.1.5 The 89 Mb heterozygous inversion leads to monoallelic and reduced *TFAP2A* levels in patient hNCC

Based on the extensive characterization of the *TFAP2A* regulatory domain described above, we hypothesized that the 89 Mb heterozygous inversion in the BOFS patient could disturb the interaction between the inverted *TFAP2A* allele and enhancers located on the other side of the breakpoint. Since these enhancers were shown to be crucial for *TFAP2A* expression in hNCC, the impaired interaction with their target gene could lead to reduced *TFAP2A* levels in the patient hNCC which would ultimately cause BOFS.

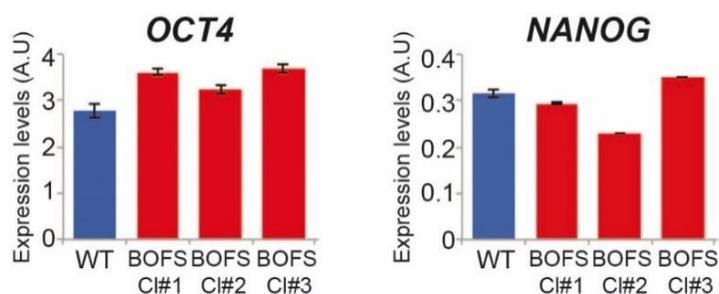
Alternatively or even in addition, the patient inversion could place the inverted *TFAP2A* allele within a heterochromatic environment which could lead to silencing due to position effect variegation (PEV) (Kleinjan and van Heyningen, 2005; Wallrath and Elgin, 1995).

To test these hypotheses and to discriminate between the potential pathomechanisms, we reprogrammed patient fibroblasts into hiPSC using a non-integrating Sendai virus vector system (Seki et al., 2012) (Figure 3.1.24). Pluripotency of the generated hiPSC clones was confirmed by RT-qPCR (Figure 3.1.25) and IF (Figure 3.1.26).



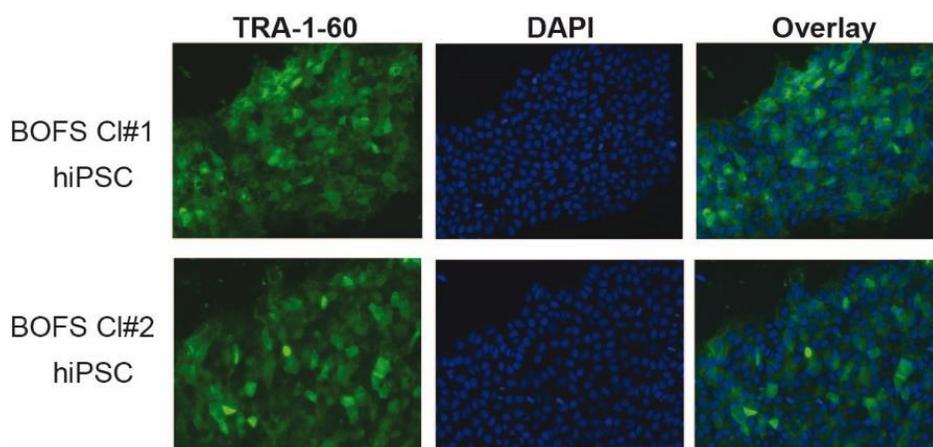
**Figure 3.1.24 Reprogramming of BOFS patient-derived fibroblast into hiPSC.**

Fibroblasts taken from the BOFS patient with the 89 Mb inversion were reprogrammed into hiPSC. Images of cellular states observed at different days during the reprogramming process are shown. Published in Laugsch et al. (2019), Figure 3A.



**Figure 3.1.25 RT-qPCR for major pluripotency markers in BOFS hiPSC.**

Expression levels of *OCT4* and *NANOG* were measured by RT-qPCR in one WT (WT-B) and three different hiPSC clones (BOFS Cl#1, BOFS Cl#2, BOFS Cl#3) derived from BOFS patient fibroblasts. Expression values were normalized to two housekeeping genes (*EEF2* and *GAPDH*). The error bars represent standard deviations from three technical replicates. Published in Laugsch et al. (2019), Figure S3c.

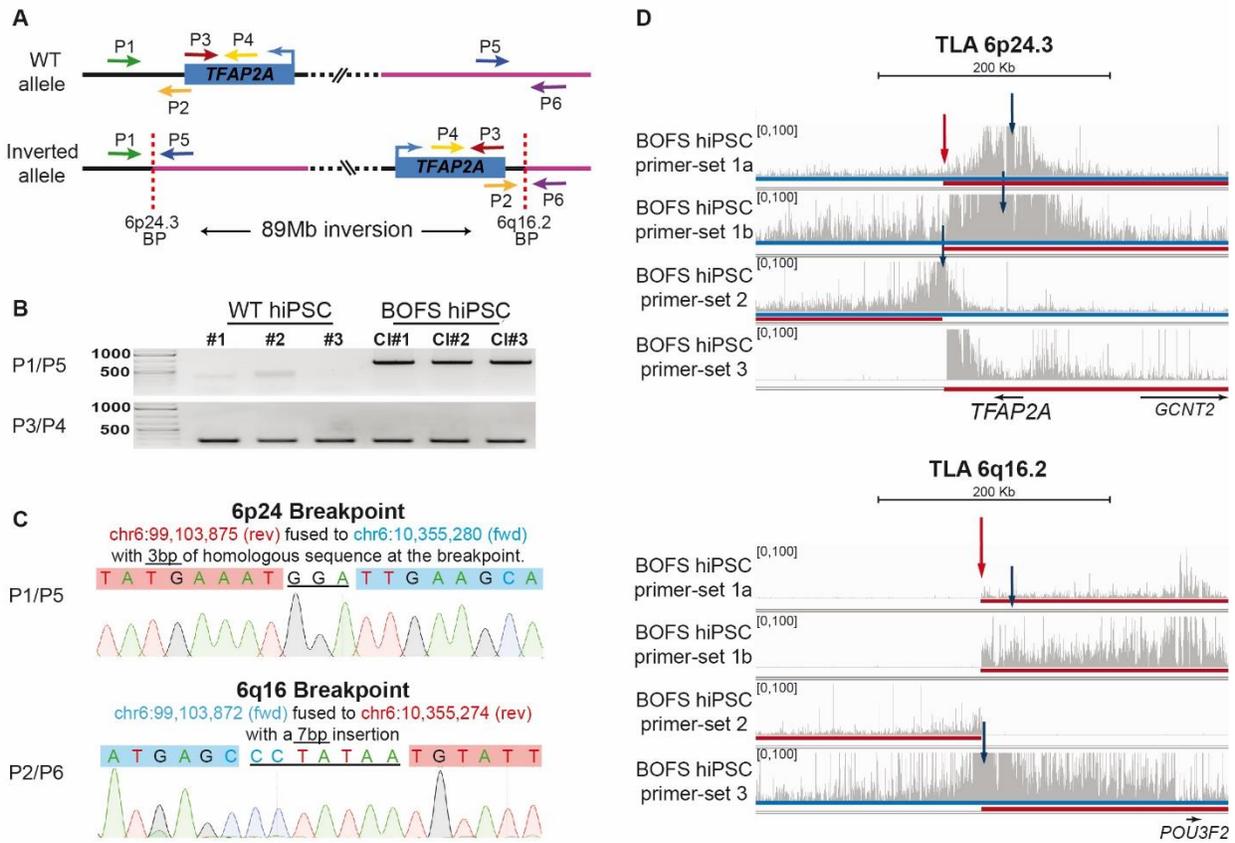


**Figure 3.1.26 Immunofluorescence for the human pluripotency marker TRA-1-60.**

IF experiments were performed in two BOFS hiPSC lines (BOFS Cl#1 and BOFS Cl#2). TRA-1-60 proteins are visualized in green by Alexa Fluor 488 2ndary antibody. Nuclei were stained with DAPI (blue). Published in Laugsch et al. (2019), Figure S3d.

## Results

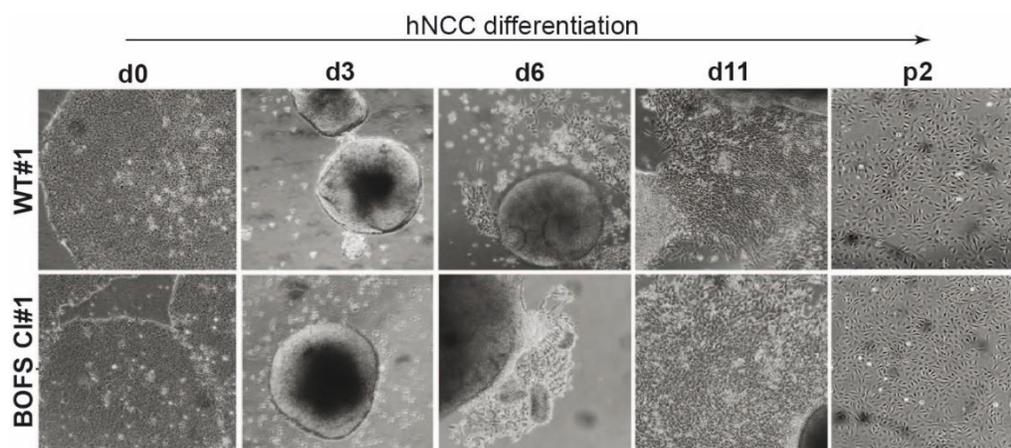
The existence of the 89 Mb heterozygous inversion in the patient-specific hiPSC was verified by PCR and Sanger sequencing (Figure 3.1.27A-C) and by Targeted Locus Amplification (TLA) (Figure 3.1.27D), which also enabled us to further refine the locations of the inversion breakpoints.



**Figure 3.1.27 Confirmation of the 89 Mb inversion by PCR, Sanger sequencing and TLA.**

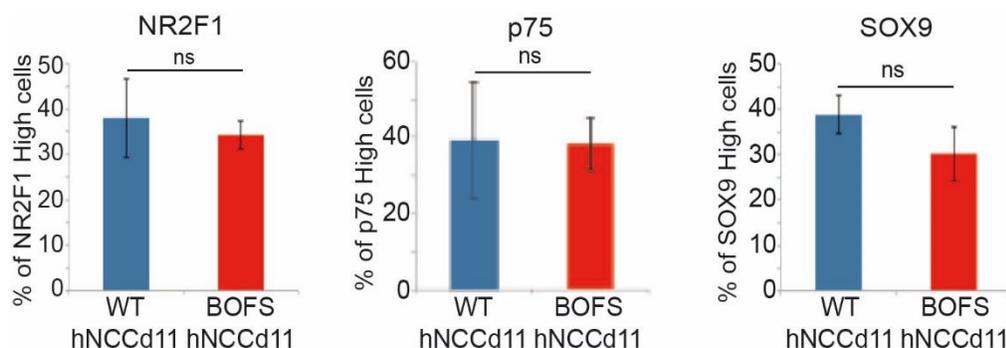
(A) Schematic view of the primer locations used to confirm the presence of the heterozygous 89 Mb inversion in hiPSC lines derived from the BOFS patient. The P1/P5 and the P2/P6 primer combinations are specific for the inversion allele, while the P3/P4 primer combination amplifies both the WT and the inversion alleles. (B) Genotyping results obtained by using the PCR primer pairs P1/P5 and P3/P4 (as described in (A)) are shown for three WT hiPSC lines (WT#1, WT#2, WT#3) and three patient-specific hiPSC clonal lines (BOFS Cl#1, BOFS Cl#2, BOFS Cl#3) derived from the BOFS patient. (C) Sanger sequencing of the PCR products obtained with the primer combinations P1/P5 and P2/P6 in the BOFS hiPSC confirmed the inversion breakpoints. (D) Targeted Locus Amplification (TLA) profiles (de Vree et al., 2014) are shown around the 6p24.3 (upper panel) and the 6q16.2 (lower panel) inversion breakpoints. Data were generated in hiPSC derived from our BOFS patient (BOFS Cl#1). The results obtained with different primer sets (1a, 1b, 2 and 3, see methods section 2.4.3 for more details), are shown from top to bottom. The dark blue arrows point at the locations of the primers, while the red arrows indicate the position of the breakpoints. The blue horizontal lines mark genomic regions where the WT allele contributes to the coverage. The red horizontal lines mark the regions where the inverted allele contributes to the coverage. Published in Laugsch et al. (2019), Figure S3a+b and S3e+f.

Strikingly, during the differentiation of several patient hiPSC lines into hNCC, we did not observe any remarkable differences in hNCC specification or maintenance compared to several control WT hiPSC lines (Figure 3.1.28). This lack of significant effect was confirmed by flow cytometry for major NC markers on day 11 of differentiation (NR2F1, p75, SOX9) (Figure 3.1.29) and after two passages (NR2F1, SOX9) (Figure 3.1.30).



**Figure 3.1.28 Differentiation of WT and BOFS hiPSC into hNCC.**

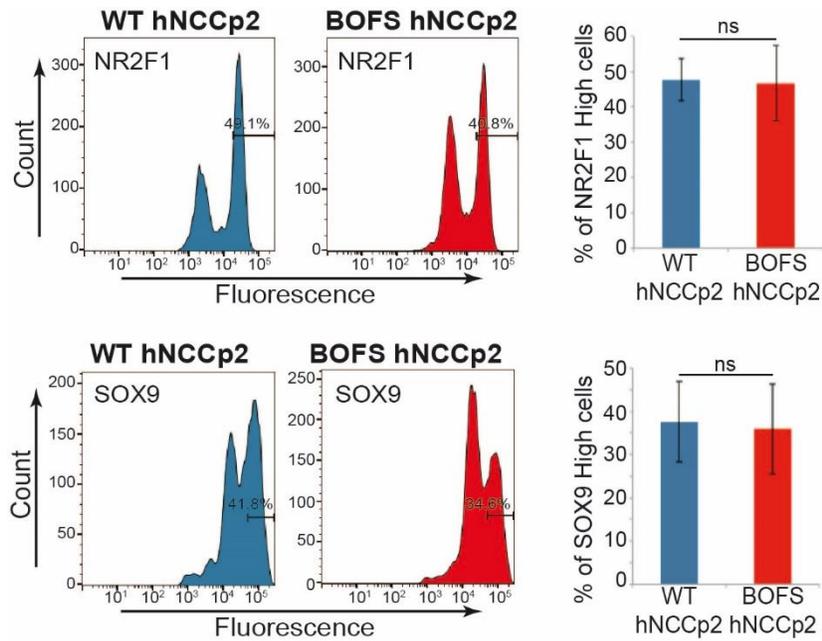
Representative images for all tested cell lines, here shown for WT#1 and BOFS CI#1, taken at different days during the hNCC differentiation process (d0-d11) and after 2 passages (p2). Published in Laugsch et al. (2019), Figure 3B.



**Figure 3.1.29 FACS analysis for NC markers in hNCCd11.**

WT and BOFS patient cells on day 11 of differentiation (hNCCd11) were analyzed by flow cytometry. The % of cells expressing high levels of NR2F1, p75 or SOX9 was quantified. The following number (n) of biological replicates were measured for the indicated proteins: NR2F1, SOX9 WT n=7 (three of WT#1, two of WT#2, two of WT#3), BOFS n=8 (four of BOFS CI#1, four of BOFS CI#2); p75 WT n=8 (four of WT#1, two of WT#2, two of WT#3), BOFS n=14 (six of BOFS CI#1, two of BOFS CI#2, six of BOFS CI#3). Average results are shown in the right panels. Error bars represent standard deviations (\*\*p-value<0.0001; ns=not significant). Published in Laugsch et al. (2019), Figure 3C and S3g.

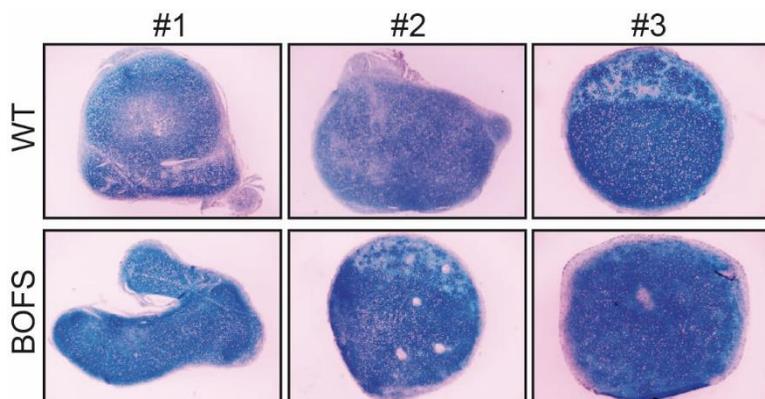
## Results



**Figure 3.1.30 FACS analysis for NC markers in hNCCp2.**

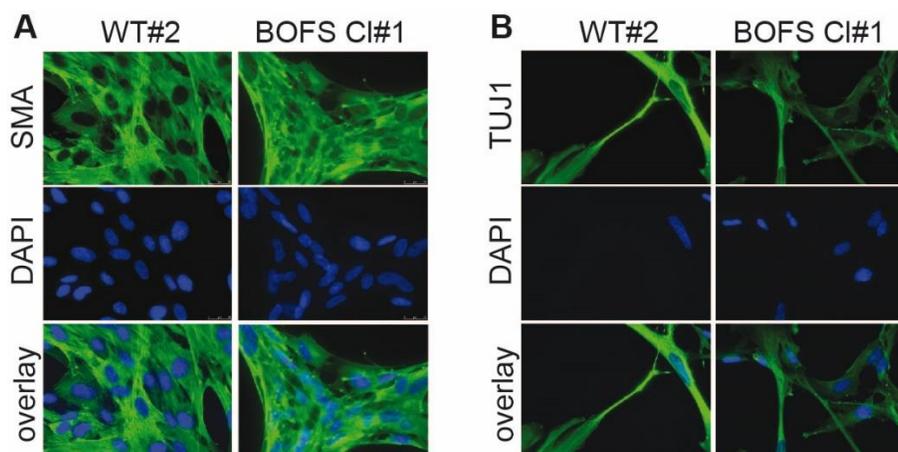
hNCC derived from WT and BOFS patient cells were passaged twice (hNCCp2) and the % of cells expressing high levels of NR2F1 or SOX9 was quantified by flow cytometry. Representative experiments in WT hNCCp2 and BOFS hNCCp2 are shown in the left and middle panels, respectively. The average results of ten biological replicates for the WT cells (four of WT#1, two of WT#2, four of WT#3) and five biological replicates for BOFS patient cells (three of BOFS Cl#1, two of BOFS Cl#3) are shown in the right panels for each protein. Error bars represent standard deviations (\*\*p-value<0.0001; ns=not significant). Published in Laugsch et al. (2019), Figure 3D and S3h.

Moreover, hNCCp2 derived from WT and BOFS patient hiPSC could be differentiated into cartilage (Figure 3.1.31), smooth muscle cells (Figure 3.1.32B), and neurons (Figure 3.1.32B) with similar efficiencies, which demonstrated that the patient hNCC keep their multipotency and differentiation potential.



**Figure 3.1.31 Alcian Blue staining of chondrocytes derived from hNCCp2.**

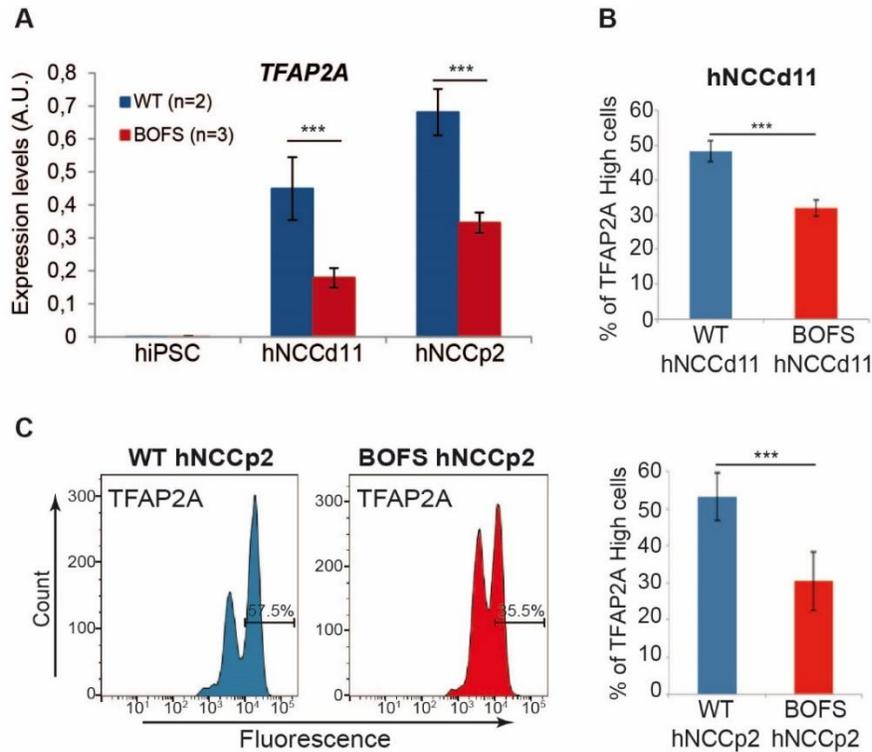
WT (#1, #2, #3) and BOFS (Cl#1, Cl#2, Cl#3) hNCCp2 were differentiated into cartilage and chondrocytes were stained with Alcian Blue. Published in Laugsch et al. (2019), Figure S3i.



**Figure 3.1.32 Immunofluorescence for smooth muscle and neuronal markers.**

WT and BOFS hNCCp2 were differentiated into (A) smooth muscle cells, characterized by expressing SMA (shown in green) or into (B) neurons, characterized by expressing TUJ1 (shown in green). Nuclei were stained with DAPI (blue). Representative images of both differentiations are shown for WT#2 and BOFS Cl#1. Similar results were obtained for additional WT and BOFS cell lines. Published in Laugsch et al. (2019), Figure S3j+k.

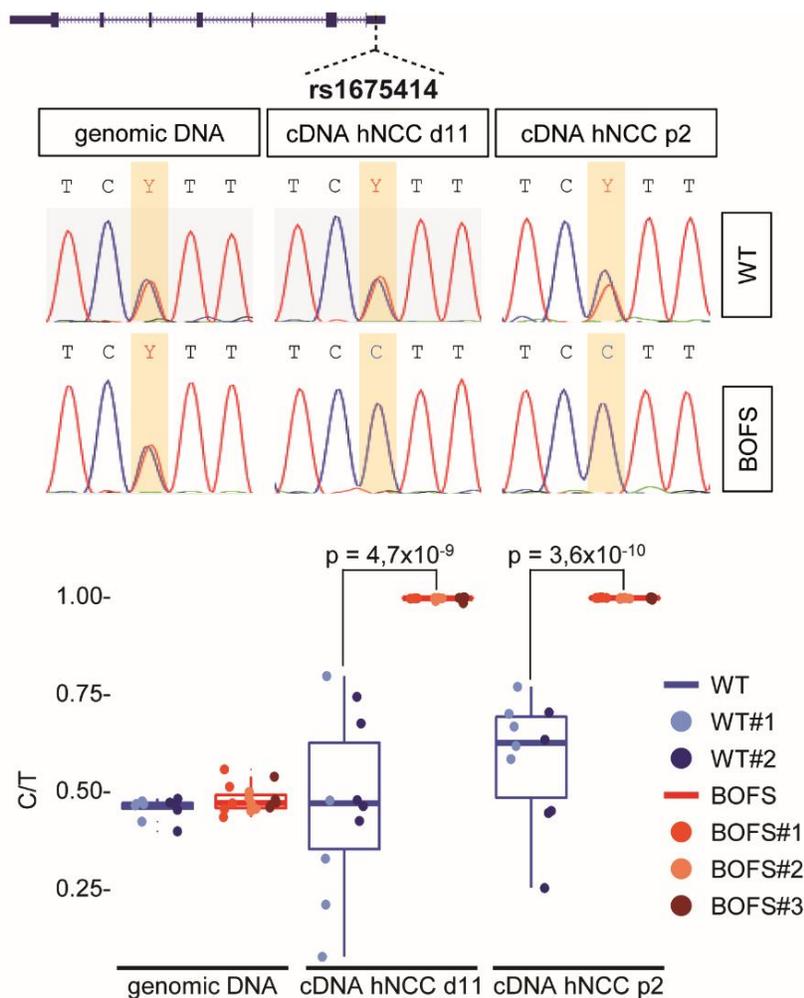
However, hNCC derived from BOFS patient hiPSC consistently expressed lower levels of *TFAP2A*, reduced to approximately half the amount measured in WT hNCC. This reduction was observed both at the RNA and on the protein levels as measured by RT-qPCR and FACS, respectively (Figure 3.1.33).



**Figure 3.1.33 Reduced *TFAP2A* expression levels in BOFS hNCC evaluated by RT-qPCR and FACS.** (A) Transcription levels of *TFAP2A* were analyzed by RT-qPCR in hiPSC, hNCCd11, and hNCCp2 of two WT cell lines (WT#1, WT#2) and three BOFS clonal lines (BOFS Cl#1, BOFS Cl#2, BOFS Cl#3). Expression values of *TFAP2A* were measured in five technical replicates for each cell line and normalized to three housekeeping genes (*ACTB*, *EEF2*, *GAPDH*). (B) WT and BOFS patient cells on day 11 of differentiation (hNCCd11) were analyzed by flow cytometry. The % of cells expressing high levels of *TFAP2A* protein was quantified. The average result from seven biological replicates of WT (three of WT#1, two of WT#2, two of WT#3) and eight biological replicates of BOFS (four of BOFS Cl#1, four of BOFS Cl#2) are presented. (C) hNCC derived from WT and BOFS patient cells were passaged twice (hNCCp2) and % of cells expressing high levels of *TFAP2A* was quantified by flow cytometry. Representative experiments in WT hNCCp2 and BOFS hNCCp2 are shown in the left and middle panels, respectively. The average results of ten biological replicates for the WT cells (four of WT#1, two of WT#2, four of WT#3) and five biological replicates for BOFS patient cells (three of BOFS Cl#1, two of BOFS Cl#3) are presented in the right panel. Error bars represent standard deviations. P-values were calculated using Student's t-test (\*\*\*)p-value<0.0001). Published in Laugsch et al. (2019), Figure 3C+E+G.

To answer if the reduced expression levels of *TFAP2A* in BOFS patient hNCC was caused by the heterozygous 89 Mb inversion, we took advantage of a SNP (rs1675414) identified within the first exon of *TFAP2A* that was heterozygous (C/T) in both the BOFS patient and two of the WT hiPSC lines. Hence, this SNP made it possible to investigate *TFAP2A* expression in an allele-specific manner (Figure 3.1.34). Additionally, the previously described TLA data (de Vree et al., 2014) enabled us to phase this and other heterozygous SNPs in the BOFS patient in relation to the inversion. The TLA data showed that the C and T alleles for the rs1675414 SNP were located within the WT allele and inverted *TFAP2A* allele, respectively (see methods section 2.4.3 for more details). Next, we quantitatively genotyped the rs1675414 SNP in gDNA and cDNA from both BOFS patient and WT hNCC. Remarkably, while clear bi-allelic *TFAP2A* expression was detected

in WT hNCC, the patient hNCC solely expressed the WT *TFAP2A* allele (i.e. C for rs1675414) (Figure 3.1.34, Figure 3.1.46).

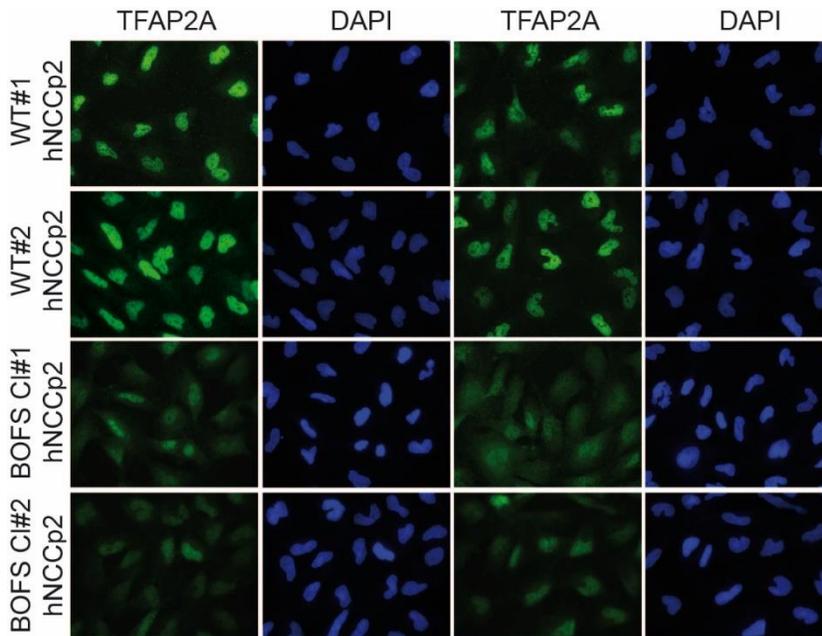


**Figure 3.1.34 Allele-specific expression analysis of *TFAP2A* by quantitative genotyping of rs1675414.**

Transcription levels of *TFAP2A* were measured in an allele-specific manner by quantitative genotyping of a SNP (rs1675414) located within the first exon of *TFAP2A* that was heterozygous (C/T) in the BOFS patient cells and two different WT hiPSC lines (WT#1, WT#3). The rs1675414 SNP was amplified and subsequently Sanger sequenced in genomic DNA (gDNA), and cDNA from hNCC (hNCCd11 and hNCCp2) derived from two WT (WT#1, WT#3) and three BOFS (Cl#1, Cl#2, Cl#3) hiPSC lines, respectively. A representative Chromatogram illustrating the genotyping of rs1675414 in gDNA and cDNA from both WT and BOFS hNCC is shown in the upper part. The overall quantification of the genotyping results is shown as boxplots in the lower part. For each cell line, five genotyping measurements were performed in gDNA and cDNA obtained from at least two independent hNCC differentiations. P-values were calculated using Student's t-test. Published in Laugsch et al. (2019), Figure 3H.

## Results

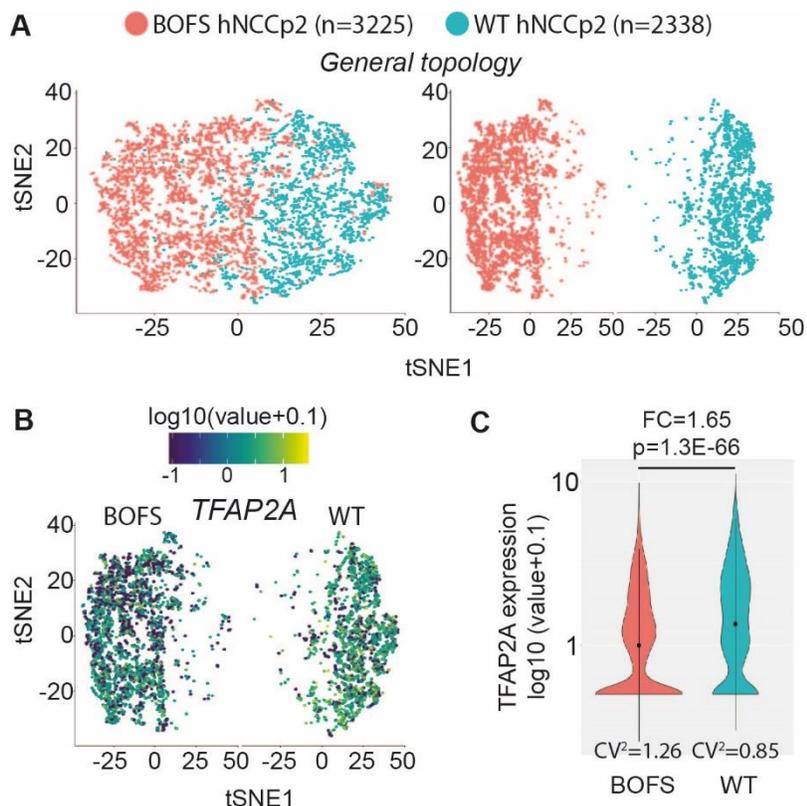
Notably, though we had observed that the gross *TFAP2A* expression in patient hNCCp2 was reduced to approximately half the amount measured in WT hNCCp2 (Figure 3.1.33), the immunofluorescence (Figure 3.1.35) and FACS (Figure 3.1.33) analyses showed that there was heterogeneity in *TFAP2A* protein levels among single cells, especially in patient hNCCp2.



**Figure 3.1.35 Immunofluorescence for *TFAP2A* in WT and BOFS hNCCp2.**

*TFAP2A* protein levels (shown in green) were evaluated by IF in hNCCp2 derived from two WT (#1, #2) and two BOFS (CI#1, CI#2) hiPSC lines. Nuclei were stained with DAPI (blue). Two representative images are displayed for each cell line. Published in Lausch et al. (2019), Figure 3F.

To investigate whether this variability in *TFAP2A* expression was also evident on the transcriptional level, we extended the previous scRNA-seq performed in WT hNCCp2 (Figure 3.1.9) to the analysis of BOFS hNCCp2 (Figure 3.1.36). Bulk analysis of the scRNA-seq data confirmed that *TFAP2A* levels were reduced close to 2-fold in the BOFS hNCCp2 (Figure 3.1.36C). However, when single-cell data was considered, it became evident that the reduced *TFAP2A* levels in BOFS hNCCp2 were the consequence of more variable *TFAP2A* expression (Figure 3.1.36B-C). Namely, there was a significant increase in the amount of BOFS hNCCp2 in which *TFAP2A* was expressed at low levels, while a few BOFS hNCCp2 still displayed high *TFAP2A* expression levels similar to those of WT hNCCp2. Thus, monoallelic *TFAP2A* expression does not only result in reduced but also more heterogeneous *TFAP2A* expression in the BOFS patient hNCC.



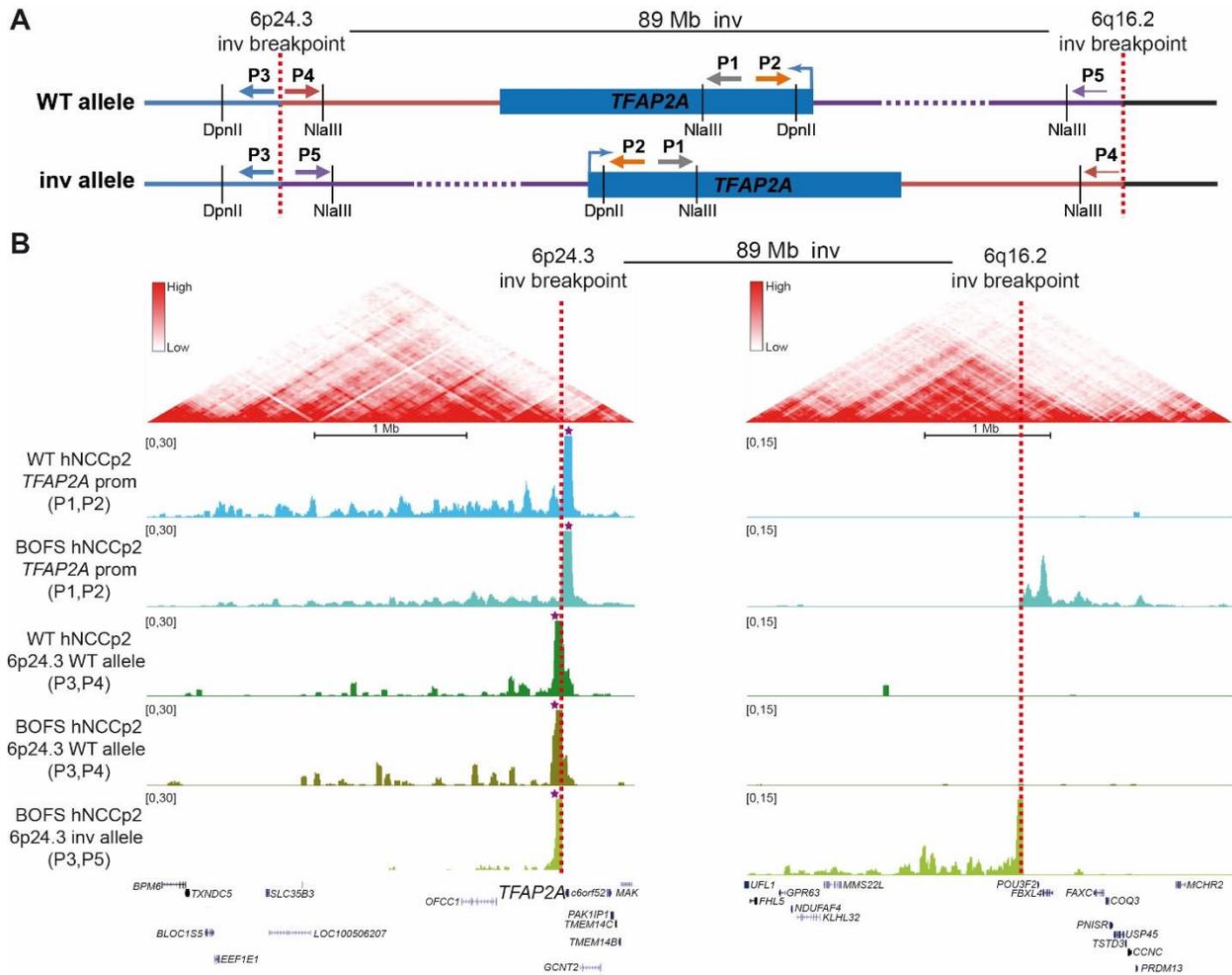
**Figure 3.1.36 Single cell RNA sequencing in WT and BOFS hNCCp2.**

(A) scRNA-seq data were generated from 2338 WT hNCCp2 and 3225 BOFS hNCCp2. WT (blue) and BOFS (red) hNCCp2 were visualized by t-distributed Stochastic Neighbor Embedding (tSNE) on the same plot (left panel) or separate plots (right panel), but using the same tSNE parameters. (B) tSNE plot (same as in the right panel of A) showing the distribution of *TFAP2A* expression in all analyzed WT and BOFS hNCCp2. (C) scRNA-seq expression data for *TFAP2A* in WT and BOFS hNCCp2 presented as violin plots. FC: fold-change in the average expression of *TFAP2A* between WT and BOFS. CV<sup>2</sup>: squared coefficient of variation. P-value was calculated using the likelihood ratio test from Monocle (Qiu et al., 2017). Published in Laugsch et al. (2019), Figure 3I-K.

### 3.1.6 Monoallelic *TFAP2A* expression in patient hNCC is caused by a loss of physical interactions between the inverted *TFAP2A* allele and its cognate hNCC enhancers

To assess whether the heterozygous 89 Mb inversion leads to monoallelic *TFAP2A* expression by physically separating the inverted *TFAP2A* allele from its enhancers, we implemented additional 4C-seq experiments in WT and BOFS hNCCp2 (Figure 3.1.37). With the *TFAP2A* promoter as a viewpoint, BOFS hNCCp2 displayed considerable interactions not only within the *TFAP2A*-TAD, as observed for the WT hNCCp2, but also 3' of the 6q16.2 inversion breakpoint (Figure 3.1.37, Figure 3.1.38).

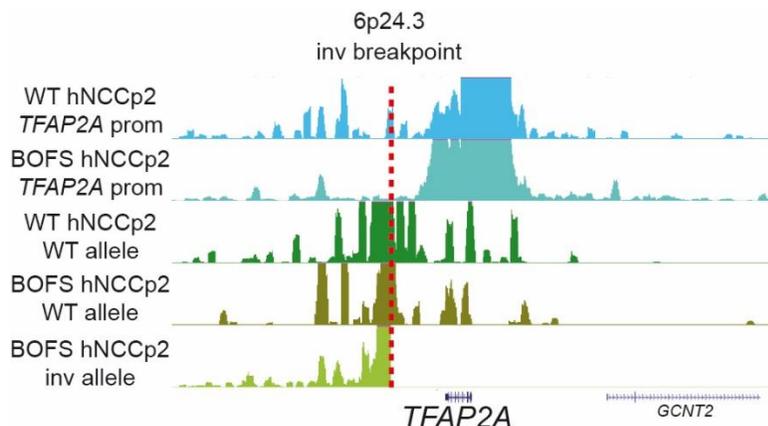
## Results



**Figure 3.1.37 Allele-specific 4C-seq experiments in BOFS hNCCp2.**

(A) Schematic diagram showing the location of primers used in the 4C-seq experiments performed in patient hNCC. The P1/P2 primer pair (“*TFAP2A* promoter” viewpoint) does not distinguish between the WT and inverted alleles. The P3/P4 and the P3/P5 primer pairs (“6p24.3 BP” viewpoint) are specific for the WT and the inverted alleles, respectively. (BP=breakpoint) (B) 4C-seq profiles generated in WT hNCCp2 and BOFS hNCCp2 (derived from WT#1 and BOFS Cl#1 hiPS, respectively) are shown around the 6p24.3 and the 6q16.2 inversion breakpoints. 4C-seq experiments were performed using the *TFAP2A* promoter and the 6p24.3 BP as viewpoints with the primer pairs described in (A). 4C-seq signals were normalized as RPMs (reads per million) and a smoothing window of eight pixels was applied for visualization. Published in Laugsch et al. (2019), Figure 4A+B.

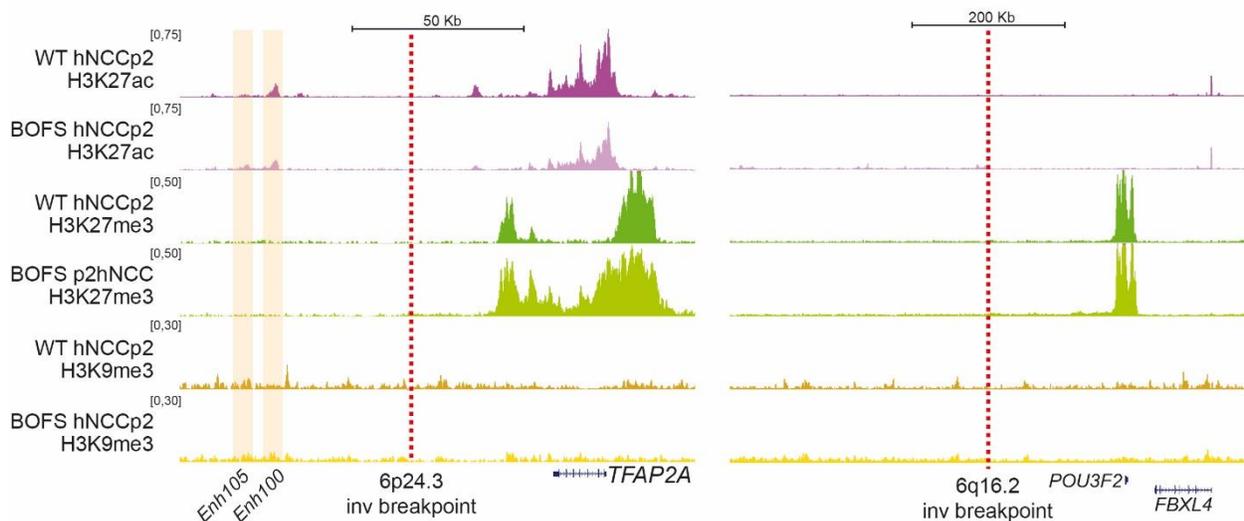
These results suggest that the inverted *TFAP2A* allele engages into new contacts within a shuffled TAD, a phenomenon previously described in Spielmann et al. (2018), 89 MB away from its endogenous location. Yet, in these 4C-seq standard experiments it is impossible to differentiate between the WT and the inverted allele of the BOFS patient. Hence, we performed allele-specific 4C-seq experiments in the patient hNCC. Here, the viewpoint was the restriction fragment containing the 6p24.3 breakpoint and we used primer combinations that were specific for either WT or inverted alleles (Figure 3.1.37, Figure 3.1.38). This strategy was previously described by Franke et al. (2016).



**Figure 3.1.38 Close-up view of the 4C-seq profiles centered on the 6p24.3 inversion breakpoint.**

Zoom-in of the 4C-seq profiles displayed in Figure 3.1.37B around the 6p24.3 inversion breakpoint (red dashed line). Published in Laugsch et al. (2019), Figure 4C.

With the WT allele-specific primer combination we detected physical interactions both 5' and 3' of the 6p24.3 breakpoint, including clear contacts with the *TFAP2A* gene, as observed in WT cells. In contrast, the primer combination specific for the inverted allele only displayed interactions 5' of the 6p24.3 and 6q16.2 inversion breakpoints, but no contacts were observed with the *TFAP2A* gene located 3' of the 6p24.3 breakpoint. These results conclusively demonstrate physical disconnection between the inverted *TFAP2A* allele and its hNCC enhancers.



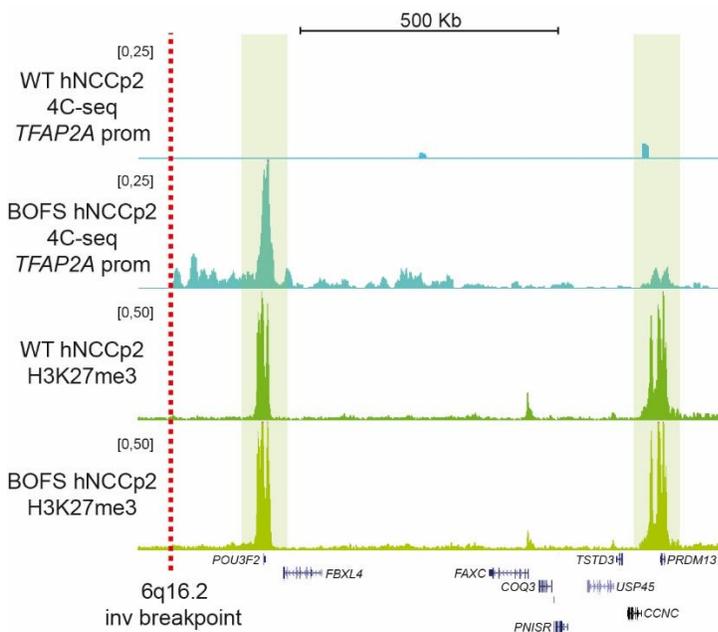
**Figure 3.1.39 ChIP-seq profiles for histone modifications in WT and BOFS hNCC around the inversion breakpoints.**

ChIP-seq data were generated for H3K27ac, H3K27me3, and H3K9me3 in WT and BOFS hNCCp2, derived from WT#1 and BOFS Cl#1 hiPSC, respectively. Profiles are shown around the 6p24.3 (red dashed line, left panel) and the 6q16.2 (red dashed line, right panel) BOFS inversion breakpoints. ChIP-seq signals were normalized as RPGC (reads per genomic content) using five bp bins. Published in Laugsch et al. (2019), Figure 4D and S4a.

## Results

As mentioned above (see results section 3.1.5), one non-mutually exclusive mechanism whereby the patient inversion could cause monoallelic *TFAP2A* expression may be by positional effect variegation (PEV) (Kleinjan and van Heyningen, 2005; Wallrath and Elgin, 1995). This would result in gene silencing by placing the inverted *TFAP2A* allele into a heterochromatic environment.

To examine this hypothesis, we performed ChIP-seq for H3K27ac, H3K27me3 and H3K9me3 in WT and BOFS hNCCp2 (Figure 3.1.39, Figure 3.1.40, Figure 3.1.41). The results of the H3K27ac and the H3K27me3 ChIP-seq matched the previously performed gene expression analysis (see Figure 3.1.33), i.e. within the *TFAP2A* locus the ChIP-seq signals for H3K27ac, a histone modification marking active regions, was higher in WT than in BOFS patient hNCC, while the signals for H3K27me3, a histone modification marking inactive regions, was lower in WT hNCCp2 (Figure 3.1.39, Figure 3.1.41).



**Figure 3.1.40 4C-seq profiles aligned with H3K27me3 ChIP-seq profiles illustrate new physical interactions at H3K27me3 bound locations.**

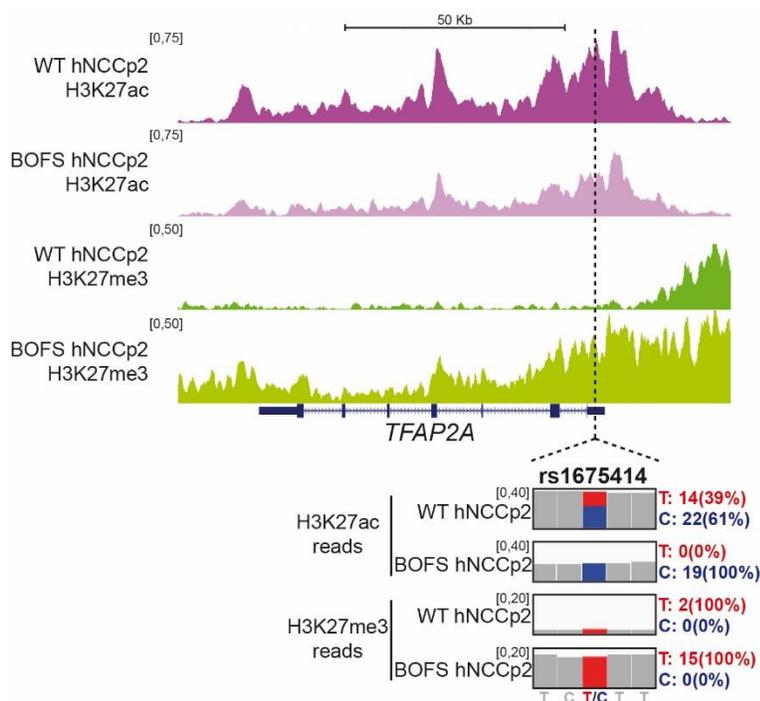
4C-seq profiles generated in WT and BOFS hNCCp2 (derived from WT#1 and BOFS Cl#1 hiPSC, respectively) using the *TFAP2A* promoter as a viewpoint are shown around the 6p24.3 BOFS patient inversion breakpoint together with the H3K27me3 ChIP-seq profiles generated from hNCCp2 of the same WT and BOFS cell lines. 4C-seq signals were normalized as RPMs (reads per million) and a smoothing window of eight pixels was applied for visualization. ChIP-seq signals were normalized as RPGC (reads per genomic content) using five bp bins. Published in Laugsch et al. (2019), Figure S4b.

Notably, when we used the previously described heterozygous SNP rs1675414 for allele-specific analysis of the ChIP-seq data we found that in the patient hNCC, the inverted and the WT allele of *TFAP2A* were solely enriched in H3K27me3 and H3K27ac, respectively (Figure 3.1.41). Similar results were obtained for other heterozygous SNPs identified by TLA in the BOFS patient cells

and located proximal to *TFAP2A* (data not shown), supporting our observation that the *TFAP2A* WT allele is actively expressed while the *TFAP2A* inverted allele is silent.

In contrast, the levels of H3K9me3, a heterochromatic histone modification mechanistically implicated in PEV (Elgin and Reuter, 2013; Tchasovnikarova et al., 2015), were very similar in WT and in BOFS patient hNCC and, anyway, rather low within the *TFAP2A* gene (Figure 3.1.39). This indicates that the inverted *TFAP2A* allele does not become heterochromatized (i.e. marked with H3K9me3). Instead, the inverted *TFAP2A* allele seems to keep the original H3K27me3 mark that is already present in pluripotent cells (Figure 3.1.39, Figure 3.1.11) and expression is never induced due to the lack of communication with the relevant hNCC enhancers.

Additionally, alignment of 4C-seq and H3K27me3 ChIP-seq profiles revealed that the inverted H3K27me3-marked *TFAP2A* allele establishes interactions with other H3K27me3-marked promoters (e.g. *POU3F2*) within its new TAD (Figure 3.1.40). This is in agreement with a previously described mechanism (Cruz-Molina et al., 2017; Schoenfelder et al., 2015), whereby H3K27me3-marked regions within the same TAD tend to engage into very strong physical interactions.

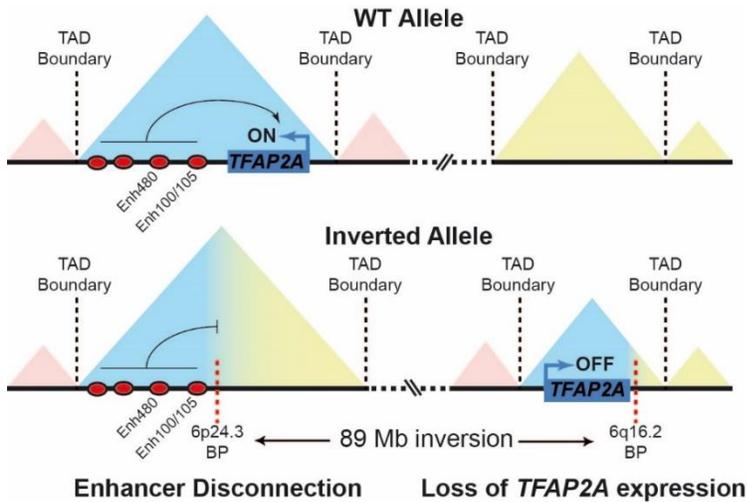


**Figure 3.1.41 ChIP-seq profiles and quantification of read numbers at the rs1675414 SNP**

H3K27ac and H3K27me2 ChIP-seq profiles generated in WT and BOFS hNCCp2 (derived from WT#1 and BOFS Cl#1 hiPSC, respectively) are displayed around *TFAP2A* (upper part). H3K27ac and H3K27me3 levels within *TFAP2A* were analyzed in an allele-specific manner by quantifying the number of ChIP-seq reads spanning the rs1675414 heterozygous SNP (black dashed line) and mapping to either the C or the T allele (lower part). ChIP-seq signals were normalized as RPGC (reads per genomic content) using five bp bins. Published in Laugsch et al. (2019), Figure S4c.

## Results

Overall, our results strongly suggest that rather than PEV, the loss of communication between *TFAP2A* gene and its cognate enhancer causes monoallelic *TFAP2A* expression which again leads to lower *TFAP2A* levels in the patient hNCC (Figure 3.1.42).

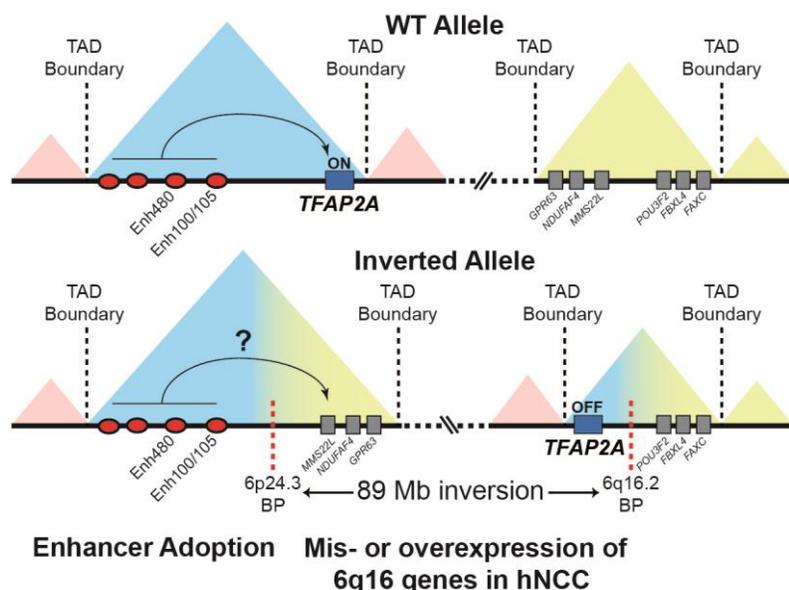


**Figure 3.1.42 Schematic diagram of how enhancer disconnection may cause loss of *TFAP2A* expression.**

Representation of how the heterozygous 89 Mb inversion identified in our BOFS patient leads to a loss of physical interaction between the inverted *TFAP2A* allele and its cognate enhancers (e.g. Enh100, Enh105, Enh480) and how this could subsequently cause a loss of *TFAP2A* expression. Published in Laugsch et al. (2019), Figure 4E.

### 3.1.7 The patient inversion does not lead to enhancer adoption or ectopic gains in gene expression in hNCC

In addition to physically disconnecting the inverted *TFAP2A* allele from its cognate hNCC enhancers, the patient inversion moves genes from the 6q16.2 locus (genes: GPR63, NDUFAF4, MMS22L) into the proximity of those active hNCC enhancers that originally would have regulate *TFAP2A* expression (Figure 3.1.43). This rearrangements creates a shuffled TAD, a previously described phenomenon (Spielmann et al., 2018), in which these genes from the 6q16.2 locus could be activated by the *TFAP2A* enhancers and consequently misexpressed through an enhancer adoption mechanism. In principle, such gene misexpression could also contribute to at least some of the BOFS phenotypes in our patient.

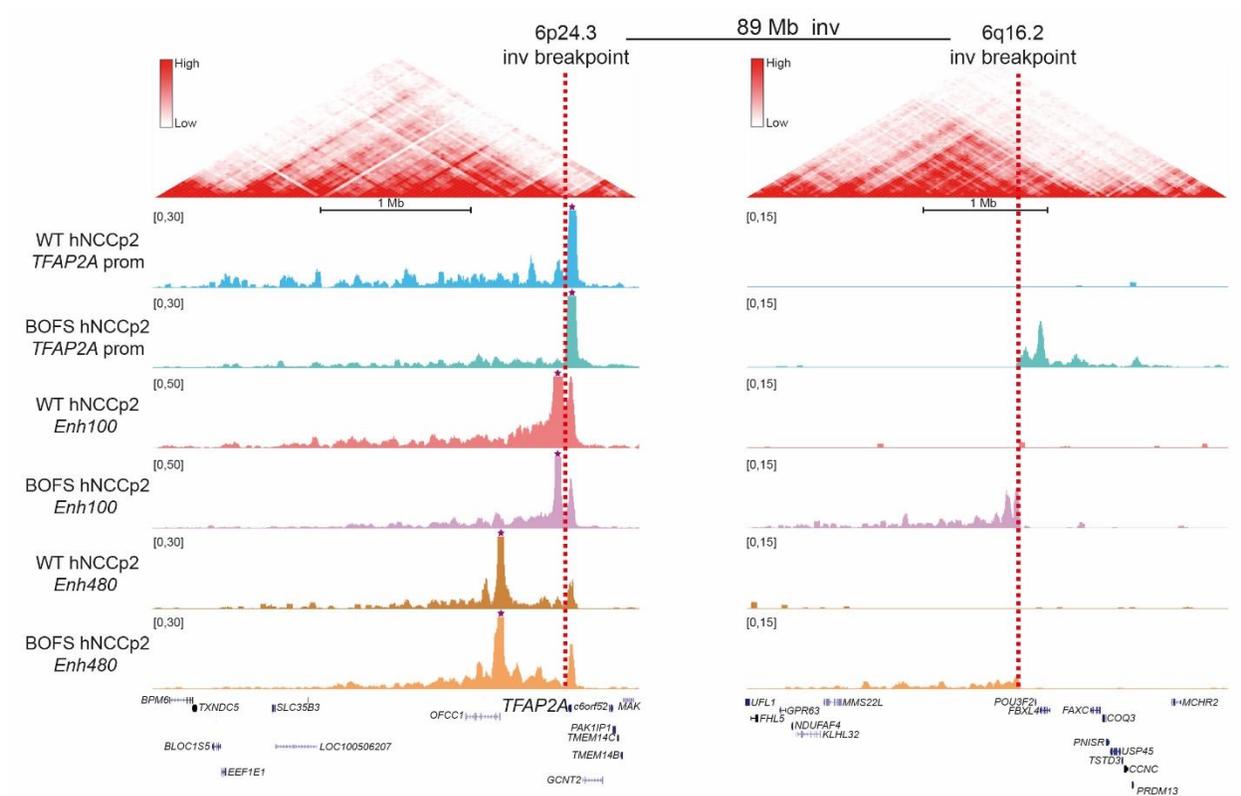


**Figure 3.1.43 Schematic diagram of how enhancer adoption may cause mis- or overexpression of 6q16 genes.** Representation of how the heterozygous 89 Mb inversion identified in our BOFS patient could theoretically lead to an enhancer adoption mechanism whereby genes originally located within the 6q16.2 locus (i.e. *MMS22L*, *NDUFAF4*, and *GPR63*) are placed within a shuffled TAD together with the *TFAP2A* hNCC active enhancers (e.g. *Enh100*, *Enh105*, *Enh480*). This enhancer adoption mechanism could, in principle, result in the mis- or overexpression of some of the 6q16.2 genes in the patient hNCC. Published in Laugsch et al. (2019), Figure 5A.

To test this hypothesis, we complemented our previous 4C-seq data in WT hNCCp2 with additional 4C-seq experiments in the patient hNCCp2 using the *Enh100* and the *Enh480* as viewpoints (Figure 3.1.44). As expected, in BOFS patient hNCC these enhancers show broad interactions within the *TFAP2A*-TAD, representing the WT *TFAP2A* allele, but also display contacts with areas on the other side (5') of the 6q16.2 inversion breakpoint (Figure 3.1.44).

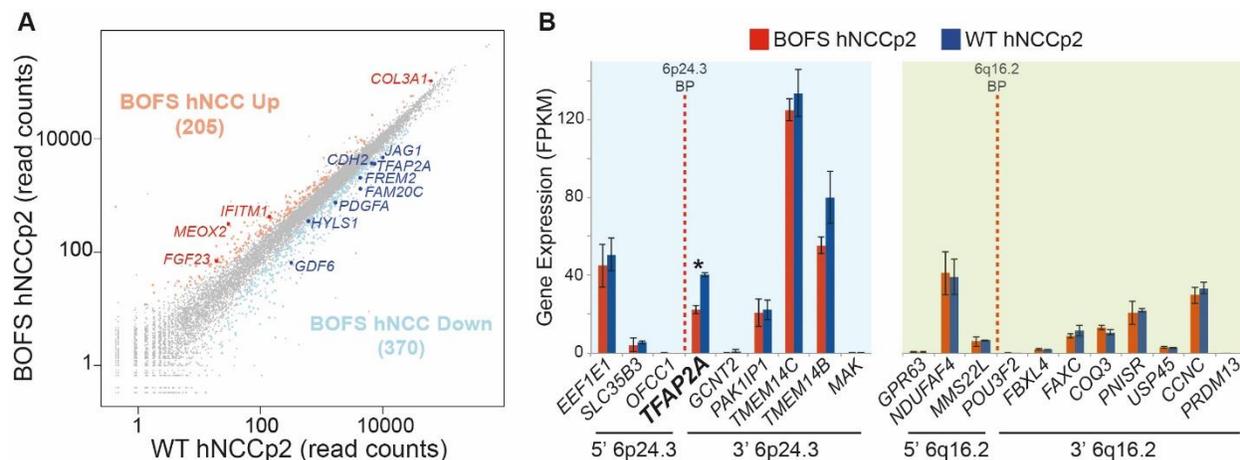
Nevertheless, these interaction signals rapidly drop with increasing linear distance and we could not detect specific communication between the *TFAP2A* enhancers and any of the 6q16.2 genes. In conclusion, the results of the 4C-seq experiments contradict the notion of an enhancer adoption mechanism and make it unlikely that the 6q16.2 genes are misexpressed in the patient hNCC.

## Results



**Figure 3.1.44** 4C-seq profiles of WT and BOFS hNCCp2 around the 6p24.3 and the 6q16.2 inversion breakpoints. 4C-seq data was generated from WT#1 and BOFS Cl#1 hiPSC derived hNCCp2 using the *TFAP2A* promoter, Enh100 and Enh480 viewpoints. 4C-seq signals were normalized as RPMs (reads per million) and a smoothing window of eight pixels was applied for visualization. Published in Laugsch et al. (2019), Figure 5B.

To test this prediction, we performed differential gene expression analysis using RNA-seq data generated in WT and BOFS hNCCp2 (Figure 3.1.45A). Besides *TFAP2A*, a couple of hundred genes showed differential gene expression in the WT and BOFS hNCCp2. Yet, when we considered the genes immediately flanking the 6p24.3 and the 6q16.2 inversion breakpoints, none of them (including *MMS22L*, *NDUFAF4* or *GPR63*), except *TFAP2A*, were differentially expressed (Figure 3.1.45B). Ultimately, this confirms that although the patient inversion places the 6q16.2 genes and the *TFAP2A* enhancers within a shuffled TAD, this does not result in an enhancer adoption mechanism.



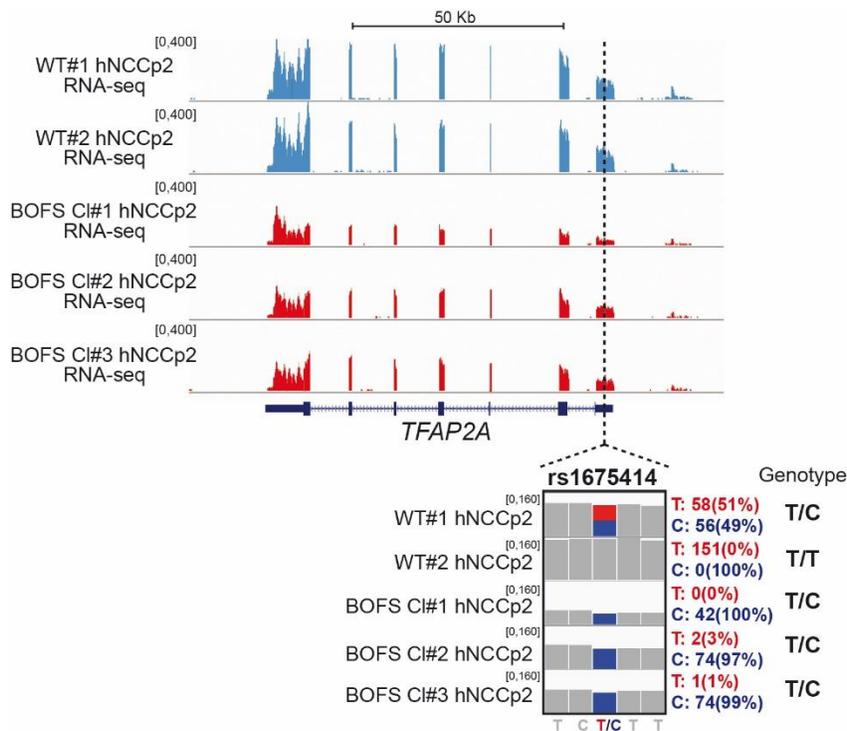
**Figure 3.1.45 Differential gene expression analysis by RNA-seq in WT and BOFS hNCCp2.**

Experiments were performed in hNCCp2 derived from 2 WT hiPSC lines (WT#1, WT#2) and 3 BOFS hiPSC clonal lines (BOFS Cl#1, BOFS Cl#2, BOFS Cl#3). **(A)** Human genes were plotted according to the average normalized RNA-seq read counts in WT hNCCp2 and BOFS hNCCp2. Genes considered as significantly up- or downregulated in BOFS hNCC compared to WT hNCC are shown in red and blue, respectively. **(B)** Average expression levels measured by RNA-seq are shown in FPKM in WT and BOFS hNCCp2 for genes immediately flanking the 6p24.3 and the 6q16.2 breakpoints. Error bars represent standard deviations. \*Genes considered as differentially expressed. Published in Laugsch et al. (2019), Figure 5C+D.

Of note, the previous RNA-seq data was also used to analyze *TFAP2A* expression in an allele-specific manner by quantifying the number of RNA-seq reads spanning the SNP rs1675414 and mapping them to either the C or the T alleles. This confirmed the previously observed reduced and monoallelic *TFAP2A* expression in the patient hNCCp2 (Figure 3.1.46).

Overall, our data clearly show that in our patient, BOFS is caused by *TFAP2A* haploinsufficiency rather than by the ectopic expression of other genes located in the proximity of the inversion breakpoints.

## Results

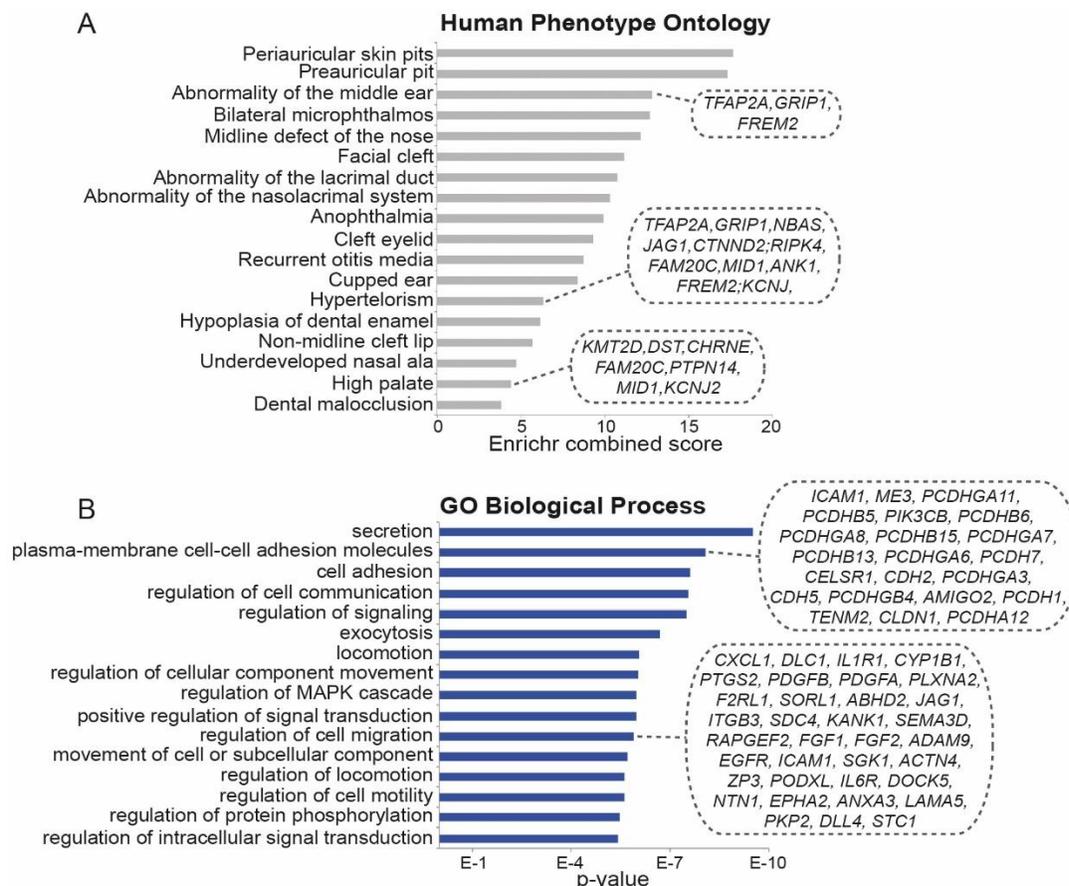


**Figure 3.1.46 Allele-specific analysis of *TFAP2A* expression using RNA-seq data of WT and BOFS hNCCp2.** RNA-seq profiles generated from WT (WT#1, WT#2) and BOFS (BOFS Cl#1, BOFS Cl#2, BOFS Cl#3) hNCCp2 derived from hiPSC are shown at the *TFAP2A* locus (upper part). Allele-specific expression of *TFAP2A* was evaluated by quantifying the number of RNA-seq reads spanning the SNP rs1675414 and mapping them to either the C or the T alleles (lower part). All cell lines evaluated are heterozygous for the evaluated SNP (genotype C/T), except for WT#1, which is homozygous (genotype T/T). Published in Laugsch et al. (2019), Figure S5.

### 3.1.8 BOFS patient hNCC display a defective transcriptome and reduced migratory capacity

Having elucidated the etiological mechanism for BOFS in our patient, we then aimed to find out more about the molecular basis of the underlying *TFAP2A* haploinsufficiency. As an initial step we went back to the previous RNA-seq data generated in WT and BOFS hNCCp2 (Figure 3.1.45).

When we functionally annotated the differentially expressed genes *in silico* we found that those genes downregulated in BOFS compared to WT hNCCp2 were particularly interesting because they were significantly enriched in genes with major roles in craniofacial morphogenesis. These genes included, for example JAG1, FAM20C, GREM2, and GRIP1 which, when mutated, lead to phenotypes typically included within the BOFS spectrum (e.g. periauricular skin pits, abnormality of the middle ear, hypertelorism, high palate) (Hamosh et al., 2002; Humphreys et al., 2012; Wang et al., 2015) (Figure 3.1.47A).



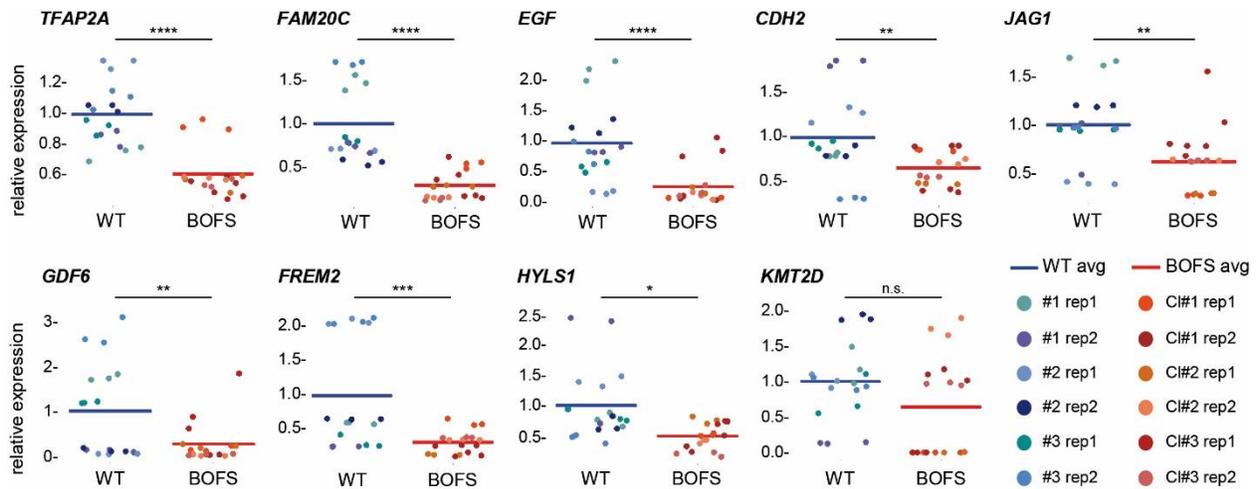
**Figure 3.1.47 Functional annotation of differentially expressed genes between WT and BOFS hNCC.**

Genes considered as significantly downregulated in BOFS hNCCp2 compared to WT hNCCp2 were functionally annotated according to (A) "Human Phenotype" and (B) "Gene Ontology Biological Processes" terms using Enrichr or DAVID *in silico* tools, respectively (Huang da et al., 2009; Kuleshov et al., 2016). Selected terms among the 50 most significantly overrepresented ones are shown. Genes involved in (A) phenotypes included in the BOFS spectrum and such possibly involved in (B) hNCC migration are highlighted. Published in Laugsch et al. (2019), Figure 6A+B.

Moreover, another set of genes downregulated in BOFS hNCCp2 were significantly enriched in members of signaling pathways and in cell adhesion molecules that have previously been implicated in the migration of various cell types, including NCC (e.g. PCDH85, PCDH86, PCDHGA8, PCDH7, CDH2, PCDH1, CLDN1, PDGFP, PDGFA, PLXNA2, JAG1, ITGB3, SEMA3D) (Figure 3.1.47B).

The downregulation in BOFS hNCCp2 compared to WT hNCCp2 of several of these genes involved in BOFS phenotypes and NCC migration was validated by RT-qPCR (Figure 3.1.48).

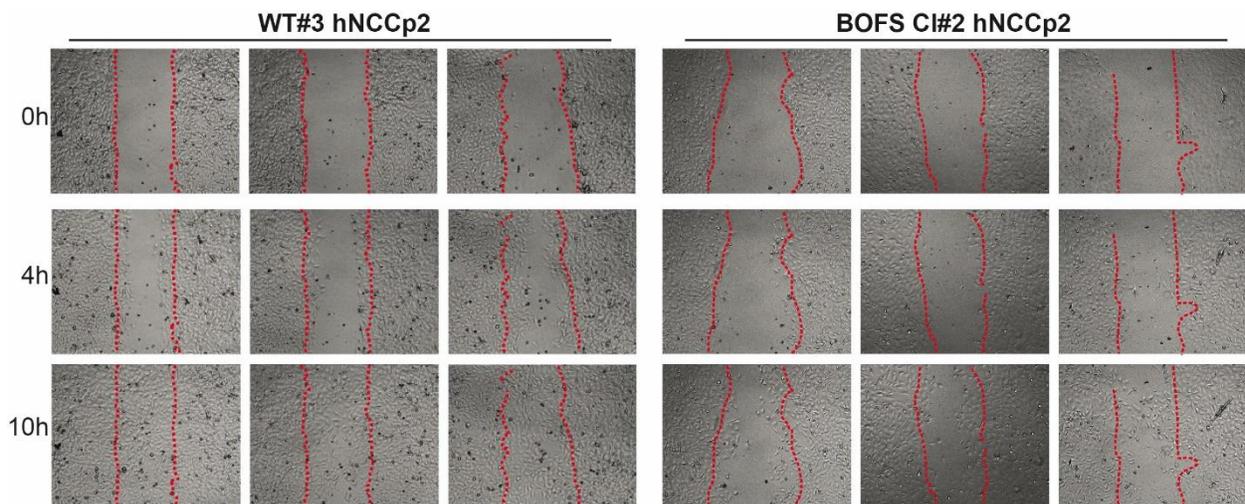
## Results



**Figure 3.1.48 RT-qPCR validation of selected downregulated genes in BOFS hNCCp2.**

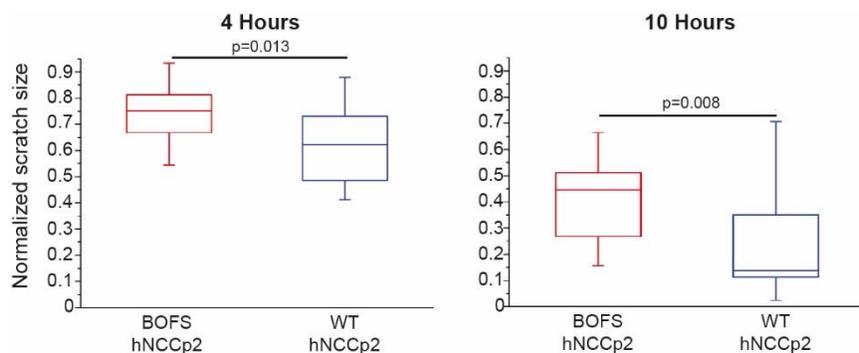
Expression levels for a subset of genes considered as downregulated in BOFS hNCCp2 compared to WT hNCCp2 according to RNA-seq experiments were measured by RT-qPCR in hNCCp2 derived from 3 different WT (WT#1, WT#2, WT#3) and 3 different BOFS patient clonal (BOFS Cl#1, BOFS Cl#2, BOFS Cl#3) hiPSC lines. Expression values were normalized to three housekeeping genes (ACTB, EFF2, GAPDH). The average expression for each gene was calculated from 18 measurements for WT cells and 18 measurements for BOFS cells (technical triplicates from biological duplicates in each of the 3 WT and 3 BOFS hiPSC lines, respectively). Published in Laugsch et al. (2019), Figure S6a.

Because the *in silico* analysis hinted toward a migratory defect in BOFS hNCC, we performed *in vitro* migration assays in both WT and BOFS hNCCp2, which showed a moderate but consistent decrease in the migration capacity of BOFS hNCCp2 (Figure 3.1.49, Figure 3.1.50).



**Figure 3.1.49 Scratch assays with WT and BOFS hNCCp2.**

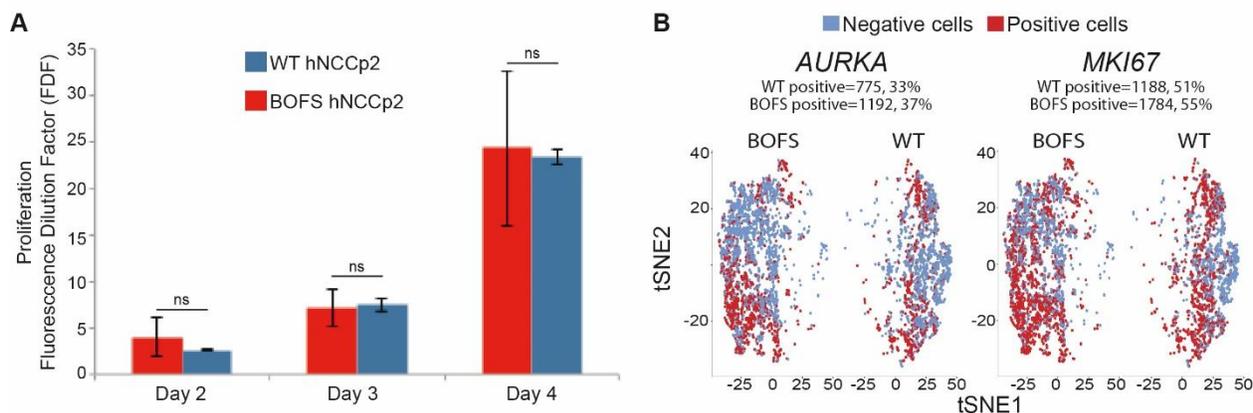
Scratch assays were performed using hNCCp2 derived from WT and BOFS hiPSC using cell culture inserts (see 2.3.10 for more details). Cell migration was monitored at the indicated times (0h, 4h, 10h). Representative images of WT#3 (left) and BOFS Cl#2 (right) hNCCp2 are shown from a total of 21 biological replicates in three WT cell lines (WT#1, WT#2, WT#3) and 22 biological replicates in three BOFS cell lines (BOFS Cl#1, BOFS Cl#2, BOFS Cl#3). Published in Laugsch et al. (2019), Figure 6C.



**Figure 3.1.50 Quantification of the scratch assays in WT and BOFS hNCCp2**

The scratch assays described in Figure 3.1.49 were quantified by measuring the outline of the scratch areas in images acquired at 0h, 4h and 10h. For each assay the area determined at 4h and 10 h were normalized to the scratch area measured at 0h. These normalized scratch sizes are shown as box plots for measurements taken at 4h (left) and 10h (right). P-values were calculated using Student's t-test for measurements in 21 biological replicates in WT hNCCp2 and 22 biological replicates in BOFS hNCCp2. Published in Laugsch et al. (2019), Figure 6D.

Importantly, the migration deficiency in BOFS hNCC was not accompanied by altered proliferation, which was similar in WT and BOFS hNCCp2 as shown by a CFSE proliferation assay (Figure 3.1.51A) and evaluation of the scRNA-seq data for two major proliferation markers (i.e. AURKA and MKI67) (Figure 3.1.51B).



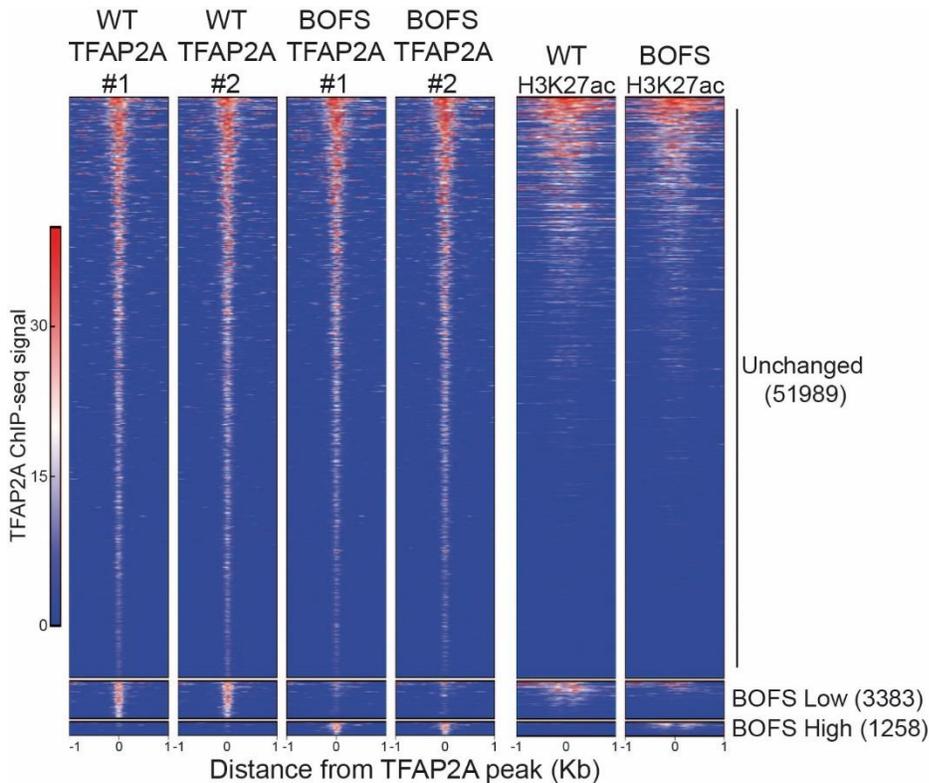
**Figure 3.1.51 Proliferation assay and expression of proliferation markers in WT and BOFS hNCCp2.**

(A) Carboxyfluorescein succinimidyl ester (CFSE) proliferation assays (see 2.3.11 for details) were performed in WT hNCCp2 (7 biological replicates: WT#2 n=3, WT#3 n=4) and BOFS hNCCp2 (6 biological replicates: CI#1 n=4, CI#3 n=2). Proliferation rates were calculated as a Fluorescein Dilution Factor (FDF) by dividing the Mean Fluorescein Intensity (MFI) measured on Day 1 by the MFI measured on Day 2, 3, and 4 for each biological replicate. P-values were calculated using Student's t-test (ns = not significant). (B) scRNA-seq data were generated as described in Figure 3.1.36 for BOFS and WT hNCCp2. The tSNE plots show the expression distribution of two major proliferation markers (i.e. AURKA, MKI67) for all analyzed BOFS and WT cells, indicating for each cell whether these genes are expressed (red; UMI $\geq$ 1) or inactive (blue; UMI=0). The number and percentage of cells expressing the respective marker is presented for WT and BOFS hNCCp2. Published in Laugsch et al. (2019), Figure S6c+d.

Overall, our data demonstrate that the heterozygous inversion found in the BOFS patient causes modest but relevant transcriptional and phenotypic defects (e.g. migratory ability) in hNCC that are in agreement with the typical craniofacial abnormalities associated with BOFS.

### 3.1.9 Gene downregulation in BOFS patient hNCC is directly mediated by reduced TFAP2A binding to active hNCC enhancers

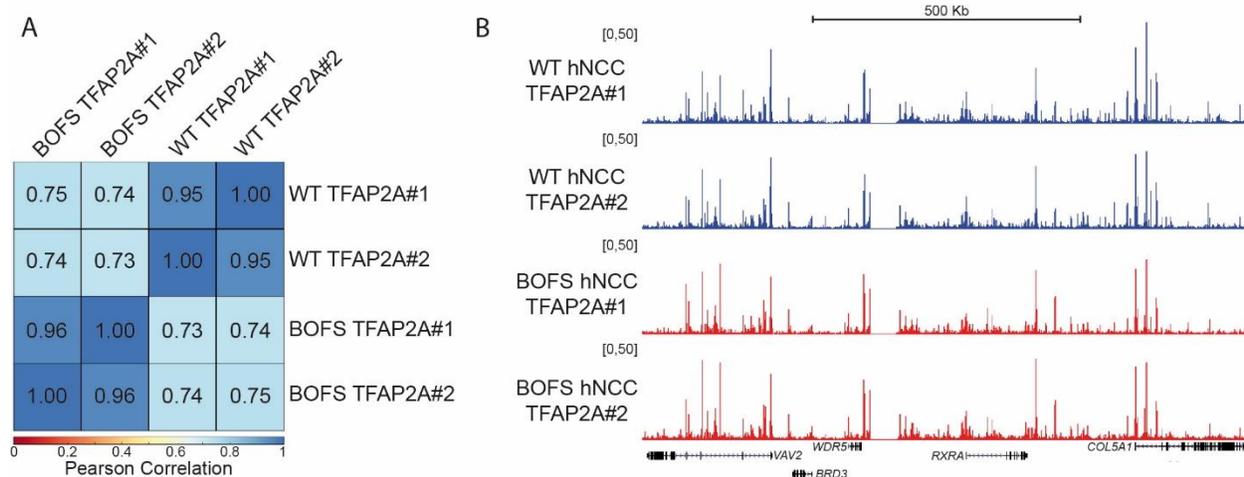
Finally, to test whether the previous gene expression differences could be directly mediated by TFAP2A binding, we performed ChIP-seq experiments for TFAP2A in both WT and BOFS hNCCp2 (Figure 3.1.52, Figure 3.1.53).



**Figure 3.1.52 Heat maps of all TFAP2A and H3K27ac ChIP-seq peaks in WT or BOFS hNCCp2.**

ChIP-seq data for TFAP2A (biological duplicates) and H3K27ac was generated from hNCCp2 derived from WT#1 and BOFS Cl#1 hiPSC. A total of 56630 TFAP2A ChIP-seq peaks were identified in WT and BOFS hNCCp2 and clustered into 3 groups, according to the differential TFAP2A binding between WT and BOFS hNCCp2: (1) unchanged, (2) lower in BOFS hNCCp2 and (3) higher in BOFS hNCCp2 (see methods section 2.11.1 for more details). ChIP-seq signals for TFAP2A and H3K27ac around all these TFAP2A peaks are visualized as heat maps. Published in Laugsch et al. (2019), Figure 7A.

Despite the reduced TFAP2A levels in BOFS compared to WT hNCCp2, in general, the binding pattern of TFAP2A was very similar in WT and BOFS hNCCp2, and most TFAP2A binding sites did not change between WT and BOFS hNCCp2 (Figure 3.1.52). This observation was confirmed in a Pearson Correlation Coefficient analysis (Figure 3.1.53A) and can be easily appreciated upon visual inspection of individual loci (Figure 3.1.53B).

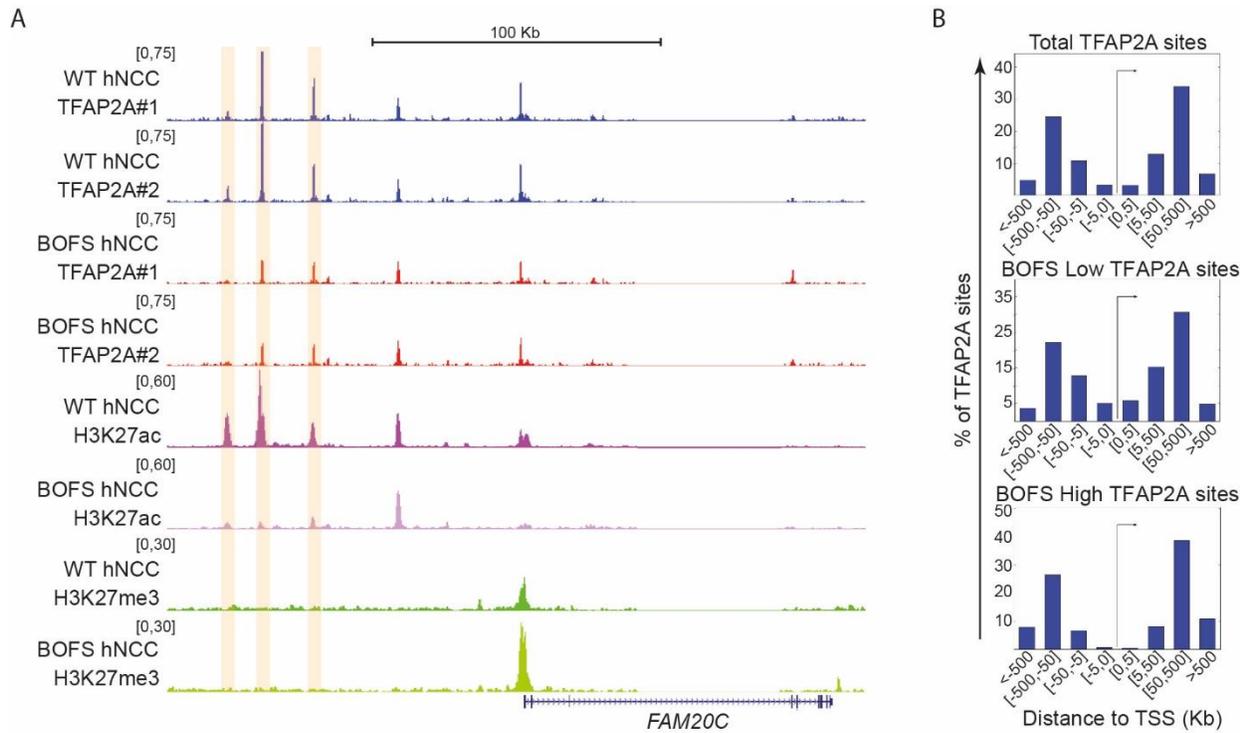


**Figure 3.1.53 Similar binding pattern of TFAP2A ChIP-seq peaks in WT and BOFS hNCC.**

**(A)** Pearson correlation coefficient between TFAP2A ChIP-seq experiments performed in WT and BOFS hNCCp2 as biological replicates. **(B)** ChIP-seq profiles for TFAP2A in WT hNCCp2 (biological duplicates: WT#1 and WT#2) and BOFS hNCCp2 (biological duplicates: BOFS Cl#1, BOFS Cl#2) are shown at a representative locus (i.e. BRD3/WDR5/RXR4 locus) for similar TFAP2A binding pattern and signal intensity in WT and BOFS hNCC. Published in Laugsch et al. (2019), Figure S7a+b.

Yet, at a couple of thousand sites TFAP2A binding differed between WT and BOFS hNCCp2, revealing nearly 3 times more sites (3383 vs. 1258) with reduced TFAP2A binding in BOFS compared to WT (BOFS Low vs BOFS High) (Figure 3.1.52).

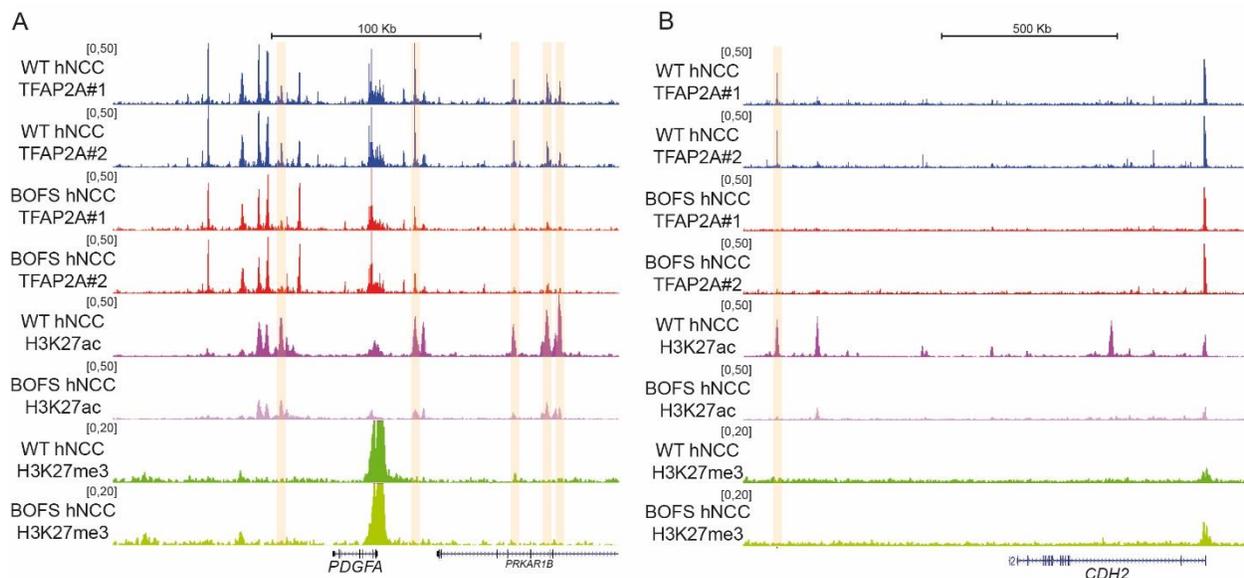
## Results



**Figure 3.1.54 ChIP-seq profiles of TFAP2A and histone marks at enhancers and distance to nearest TSS in WT and BOFS hNCC.**

**(A)** ChIP seq profiles for TFAP2A, H3K27ac and H3K27me3 in WT#1 and BOFS CI#1 hNCCp2 are shown at a representative locus (i.e. FAM20C) that contains three (highlighted) active hNCC enhancers displaying reduced TFAP2A binding and H3K27ac levels in BOFS hNCCp2 compared to WT hNCCp2. **(B)** Distance from TFAP2A binding sites in total number of TFAP2A sites (upper panel), BOFS Low TFAP2A sites (middle panel) and BOFS High TFAP2A sites (lower panel) to the transcription start site (TSS) of nearby human genes. Published in Laugsch et al. (2019), Figure 7B and S7e.

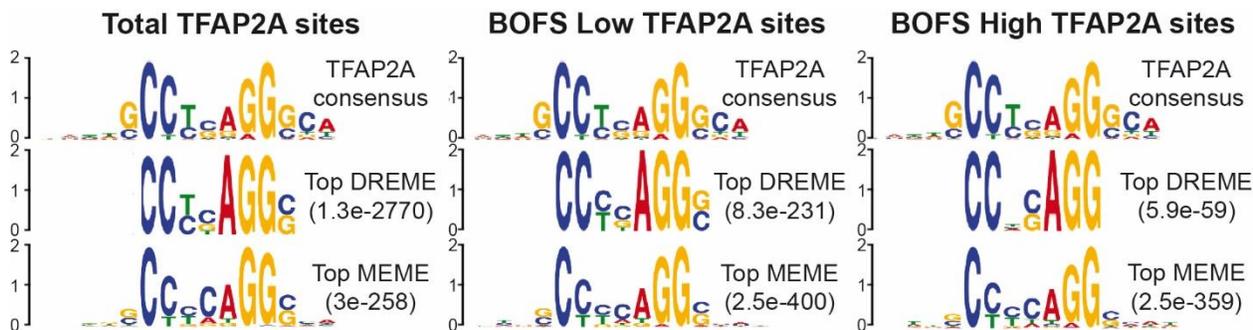
Furthermore, our ChIP-seq data indicate that sites with lower TFAP2A binding in BOFS hNCCp2 (BOFS Low) tend to occur within distal enhancers which also displayed reduced H3K27ac levels in BOFS hNCCp2 (Figure 3.1.52, Figure 3.1.54, Figure 3.1.55).



**Figure 3.1.55 Additional loci displaying differential TFAP2A binding and H3K27ac levels between WT and BOFS hNCCp2.**

ChIP-seq profiles for TFAP2A, H3K27ac and H3K27me3 in WT#1 and BOFS Cl#1 hNCCp2 are shown at the (A) PDGFA and the (B) CDH2 loci, representing additional enhancers to the ones in Figure 3.1.54 with differential TFAP2A binding between WT and BOFS hNCCp2. Active hNCC enhancers displaying lower TFAP2A binding and H3K27ac levels in BOFS hNCCp2 compared to WT hNCCp2 are highlighted. Published in Laugsch et al. (2019), Figure S7c+d.

One potential molecular mechanism that could explain the cases where TFAP2A is differentially bound in WT and BOFS hNCCp2 is by differences in the underlying TFAP2A binding motifs. To test this possibility, we performed an *in silico* motif analysis using both DREME and MEME algorithms (Figure 3.1.56). However, the results seem to be very similar in BOFS Low and BOFS High TFAP2A bound sites and therefore cannot explain the differential binding pattern in these regions.

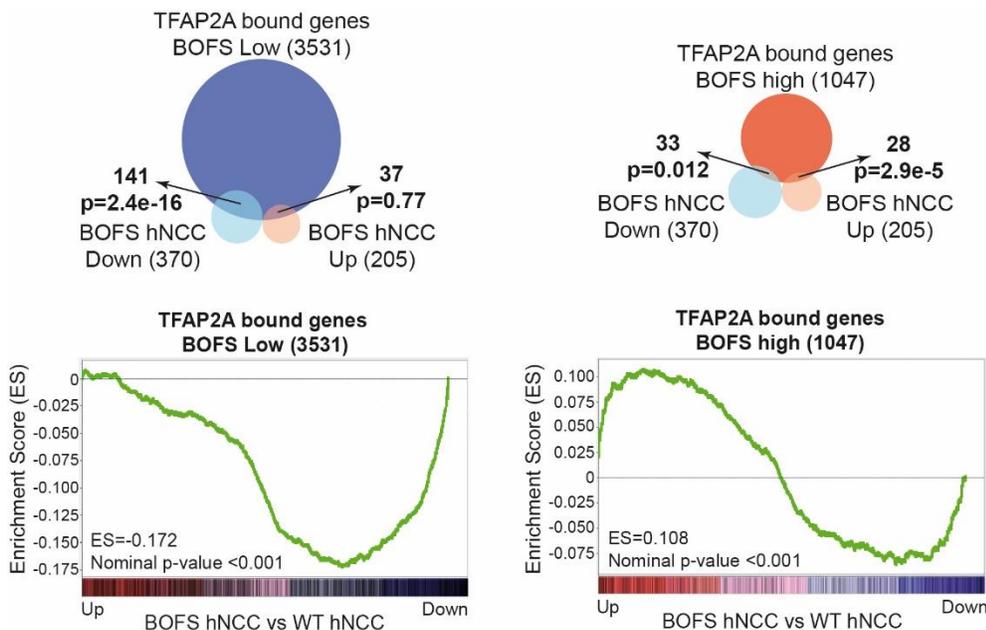


**Figure 3.1.56 Motif analysis of differentially bound TFAP2A sites.**

Top overrepresented TFAP2A binding motifs identified by DREME and MEME motif binding algorithms (Machanic and Bailey, 2011) for “Total”, “BOFS Low”, and “BOFS High” TFAP2A bound regions. Published in Laugsch et al. (2019), Figure S7f.

## Results

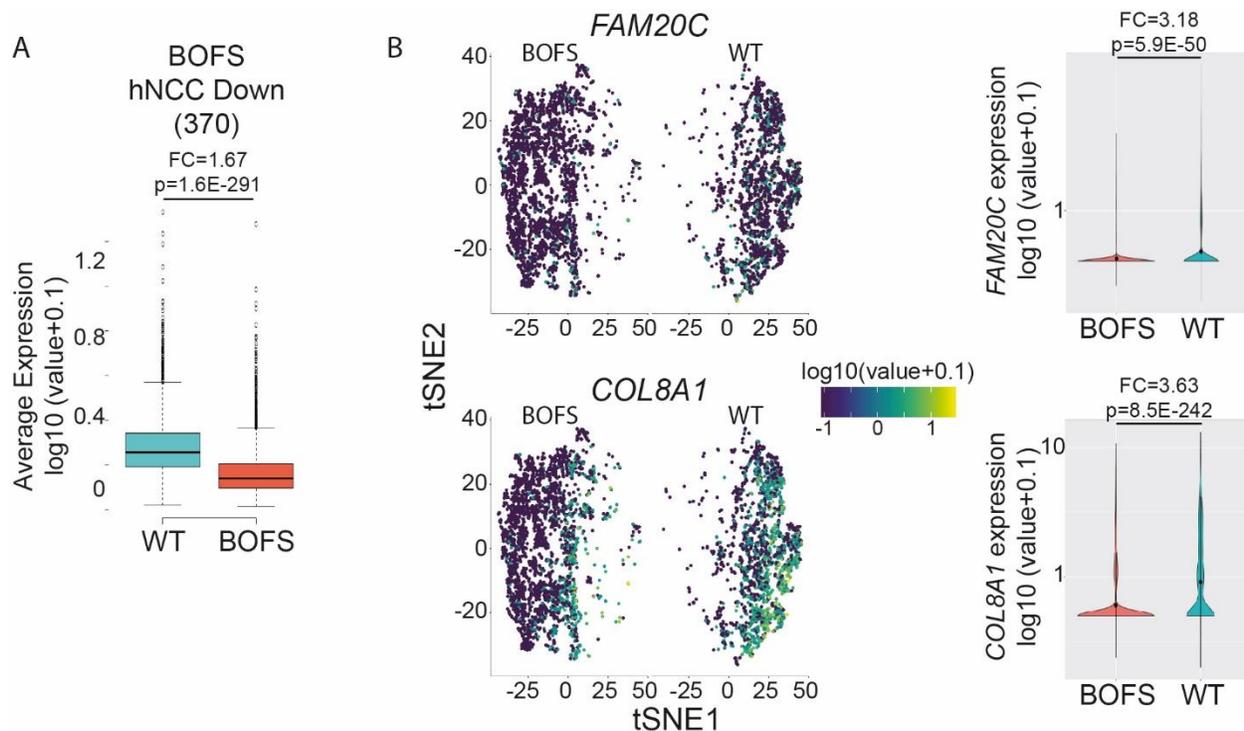
Next, we linked those sites differentially bound by TFAP2A to the closest genes using GREAT (McLean et al., 2010), and discovered a highly significant overlap ( $p=2.4e^{-16}$ ) between genes showing lower TFAP2A ChIP-seq signals in BOFS than in WT hNCCp2 (BOFS Low) and genes with lower expression levels in BOFS compared to WT hNCC p2 (BOFS hNCC Down) (Figure 3.1.57, left panels). In contrast, genes linked to TFAP2A sites with higher binding in BOFS hNCCp2 (BOFS High) displayed considerably milder overlaps with both genes up- and downregulated in BOFS hNCCp2 (Figure 3.1.57, right panels).



**Figure 3.1.57 Relationship between TFAP2A binding and transcriptional changes in BOFS hNCC.**

Venn diagrams (upper part) showing the overlaps between genes linked to differentially bound TFAP2A sites (“BOFS Low”: dark blue; “BOFS High”: red) and genes downregulated (light blue) or upregulated (light red) in BOFS hNCCp2. p-values were calculated using heterogeometric tests. Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) (lower part) for genes linked to “BOFS Low” (left) or “BOFS High” (right) TFAP2A sites was performed with respect to the global transcriptional changes observed between WT and BOFS hNCCp2. ES = enrichment score. Published in Laugsch et al. (2019), Figure 7C and S7g.

The direct regulatory function of TFAP2A on the expression of those genes downregulated in BOFS hNCCp2 was further supported by analysis of the previously described scRNA-seq data generated in WT and BOFS hNCCp2 (Figure 3.1.58, Figure 3.1.59). Firstly, the downregulation of this set of genes in the BOFS hNCCp2 was confirmed by analyzing the scRNA-seq data of WT and BOFS hNCCp2 in bulk (Figure 3.1.58).

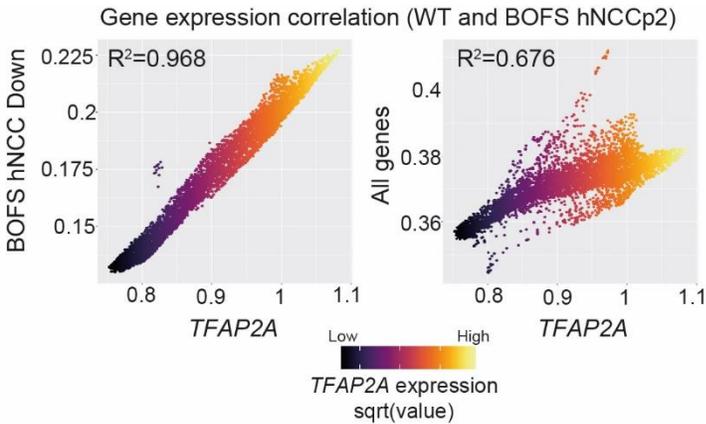


**Figure 3.1.58 Bulk RNA-seq analysis of scRNA-seq data.**

**(A)** According to bulk RNA-seq using the scRNA-seq data generated in WT and BOFS hNCCp2, average expression levels of genes downregulated in BOFS compared to WT hNCCp2 are displayed as box plots. FC = Fold change in the average expression between WT and BOFS hNCCp2. p-value was calculated using Student's t-test. **(B)** Expression of two genes (*FAM20C* and *COL8A1*) considered as "BOFS Low" according to TFAP2A ChIP-seq and transcriptionally downregulated in BOFS hNCCp2 according to bulk RNA-seq analysis is shown as tSNE plots (left panels) and violin plots (right panels) using the scRNA-seq data generated in WT and BOFS hNCCp2. FC = Fold change in the average expression between WT and BOFS hNCCp2. p-values were calculated using likelihood ratio tests. Published in Laugsch et al. (2019), Figure 7D+E.

Most importantly, we detected a strong positive correlation between the average expression of the BOFS downregulated genes and the TFAP2A levels within single hNCCp2 (Figure 3.1.59, left panel), while this correlation was much weaker between TFAP2A levels and transcription levels of all genes expressed in hNCCp2 (Figure 3.1.59, right panel):

## Results



**Figure 3.1.59 Correlation between the expression levels of *TFAP2A* and the genes downregulated in patient hNCCp2 within single cells.**

Every single hNCCp2 analyzed by scRNA-seq is displayed as a dot in which the expression of *TFAP2A* (colored according to the levels of *TFAP2A* expression, x-axis) is compared to the average expression of those genes considered as downregulated in BOFS hNCCp2 (left panel, BOFS hNCC Down according to bulk RNA-seq, y-axis) or to all genes (right panel, x-axis). The scRNA-seq data for all analyzed WT and BOFS hNCCp2 was used after library normalization, square root transformation and dropout correction using MAGIC (van Dijk et al., 2018).  $R^2$  = squared Pearson correlation coefficient. Published in Laugsch et al. (2019), Figure 7F)

Overall, our findings strongly suggest that the differences in genes expression between WT and BOFS hNCCp2 and, more specifically, the downregulation of genes in BOFS hNCC are a direct consequence of reduced *TFAP2A* binding to important enhancers in the BOFS hNCC. More generally, these results support a universal mechanism in which lower *TFAP2A* activity and/or levels in BOFS patient hNCC cause a change in transcription, and consequently disturb the morphogenetic features of hNCC at a pathological level.

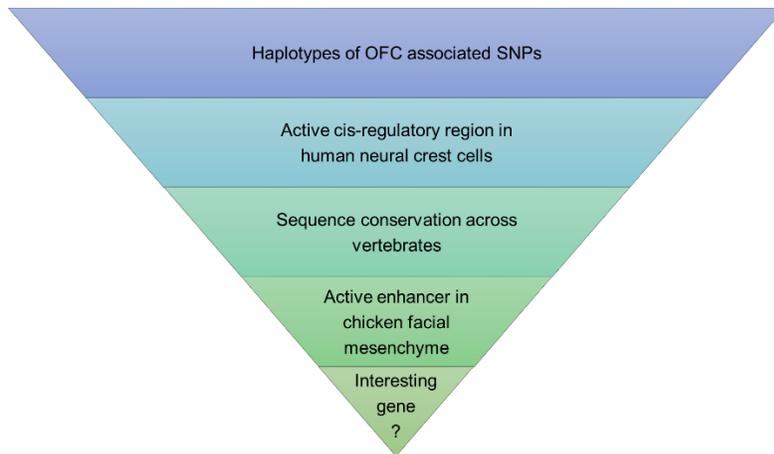
### **3.2 Molecular and functional characterization of a locus contributing to non-syndromic orofacial cleft identifies DDX1 as a novel regulator of facial development**

Remark: The following part “Molecular and functional characterization of a locus contributing to non-syndromic orofacial cleft identifies DDX1 as a novel regulator of facial development” has not been published yet. This project is the result of combined efforts of a number of people. Rizwan Rehimi performed all electroporation of chicken embryos for the CRISPR/Cas9 experiments and documented the embryos (Figure 3.2.12, Figure 3.2.15, Figure 3.2.16 and Figure 3.2.17). Milos Nikolic did the computational analysis necessary for the overlap of OFC risk-haplotypes with hNCC enhancers described in section 2.11.6 (Classification of GWAS SNPs) and incorporated in Figure 3.2.1 which led to the selection of the candidate region described in Figure 3.2.2. Giuliano Crispantu processed the RNA-seq data incorporated into Figure 3.2.30 and Figure 3.2.33. Victor Sanchez analyzed NGS data presented in Figure 3.2.32 and Figure 3.2.33. All other experiments were performed by me.

#### **3.2.1 Comparison between OFC risk-loci and hNCC enhancers as a strategy to identify novel genes implicated in facial development**

Based on the knowledge that OFC-associated SNPs are overrepresented in hNCC *cis*-regulatory elements (Ludwig et al., 2017) and that such SNPs can influence the expression of the *cis*-regulatory element's target genes (Spitz and Furlong, 2012; Uslu et al., 2014; Yanez-Cuna et al., 2013) we aimed at identifying novel regulators of facial development.

We started our approach (see Figure 3.2.1) by comparing risk-haplotypes harboring SNPs reported to be associated with OFC in at least one GWAS (Beaty et al., 2010; Beaty et al., 2011; Birnbaum et al., 2009; Grant et al., 2009; Ludwig et al., 2012; Mangold et al., 2010) with hNCC *cis*-regulatory modules (i.e. distal and proximal active enhancers) as described in Ludwig et al. (2017) and based on epigenomic data generated by Rada-Iglesias et al. (2012). We further selected only those loci with sequence conservation across vertebrates and then those which also had active enhancer marks in chicken facial mesenchyme. These two criteria should ensure that the selected loci and the respective enhancers are evolutionarily conserved and, thus, more likely to be of functional relevance during facial development.

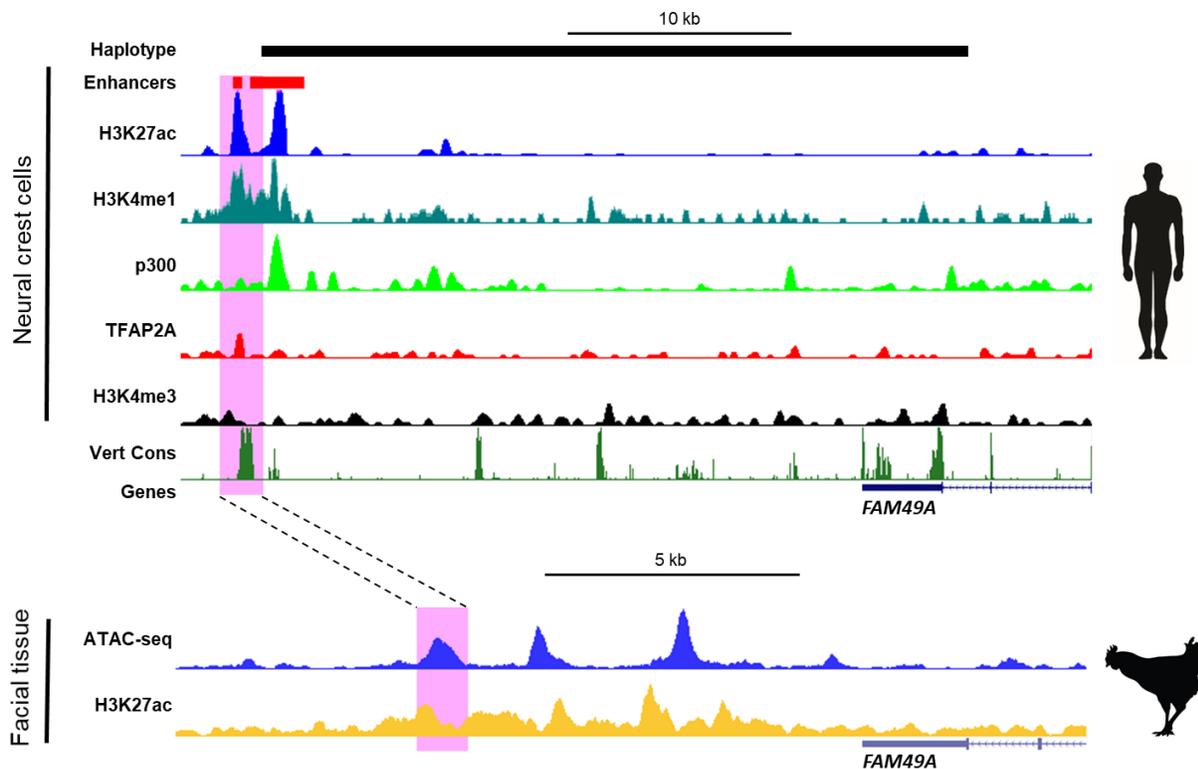


**Figure 3.2.1 Diagram of selection criteria for candidate enhancers.**

The five criteria stated above were employed to select a candidate locus for functional analysis. Of all reported OFC associated SNPs, 83 were located within 35 haplotypes intersecting hNCC active enhancers. Among these enhancers, ten were conserved at both the sequence and functional level. We then selected our candidate locus based on the genes that were potentially controlled by one of those conserved enhancers.

The described selection criteria resulted in an assortment of ten enhancers with both sequence and functional conservation between chicken and human and that contain SNPs associated with orofacial clefts in human. Among these ten enhancers, we identified active NCC enhancers overlapping haplotypes associated with OFC within well-studied susceptibility loci of OFC, for example 6p24.3 (OFC1, OMIM #119530) and 8q24.21 (OFC12, OMIM #812858), overall supporting the potential of our approach.

Here we focus on a new locus on chromosome 2 (2p24.2), which has not been characterized before, especially not in connection with OFC. Yet, the locus is particularly interesting because it spans an evolutionarily conserved active NCC enhancer (Enh2p24.2) of about 2 kb harboring four orofacial cleft associated SNPs. The enhancer displays active chromatin marks in both hNCC (Rada-Iglesias et al., 2012) and in facial tissue of chicken embryos and, interestingly, it is located upstream of the neuronal gene *FAM49A*, which has not been previously implicated in facial development (Figure 3.2.2).



**Figure 3.2.2. Epigenetic profile of human neural crest cells and chicken facial tissue at the 2p24.2 locus.**

The location of the haplotype comprising all OFC associated SNPs of this region is shown together with previously published ChIP-seq data for p300, TFAP2A and various histone modifications from hNCCd11 (Rada-Iglesias et al., 2012). In addition, the corresponding genomic region in the chicken genome is also shown, including previously published H3K27ac ChIP-seq data (Rada-Iglesias et al., 2012) and newly generated ATAC-seq from the facial mesenchyme of chicken embryos. ChIP-seq and ATAC-seq signals were normalized as RPGC (reads per genomic content) using five bp bins. The human candidate enhancer (Enh2p24.2) and its chicken homolog are highlighted in pink. The closest gene to the candidate enhancers is *FAM49A*, whose 3'UTR also partially overlaps with the OFC-risk haplotype.

Importantly, SNPs in this 2p24.2 risk locus were first suggested to be nominally associated with OFC in a GWAS by Ludwig and colleagues (2012). This association was then confirmed to be highly significant ( $p=4.22 \times 10^{-8}$  and  $p=5.83 \times 10^{-22}$ ) in two independent GWAS (Leslie et al., 2016; Yu et al., 2017) (Table 3.1), thus highlighting its etiological relevance. Nevertheless, the functionality and potential pathomechanism of this risk region remains unknown and barely characterized.

## Results

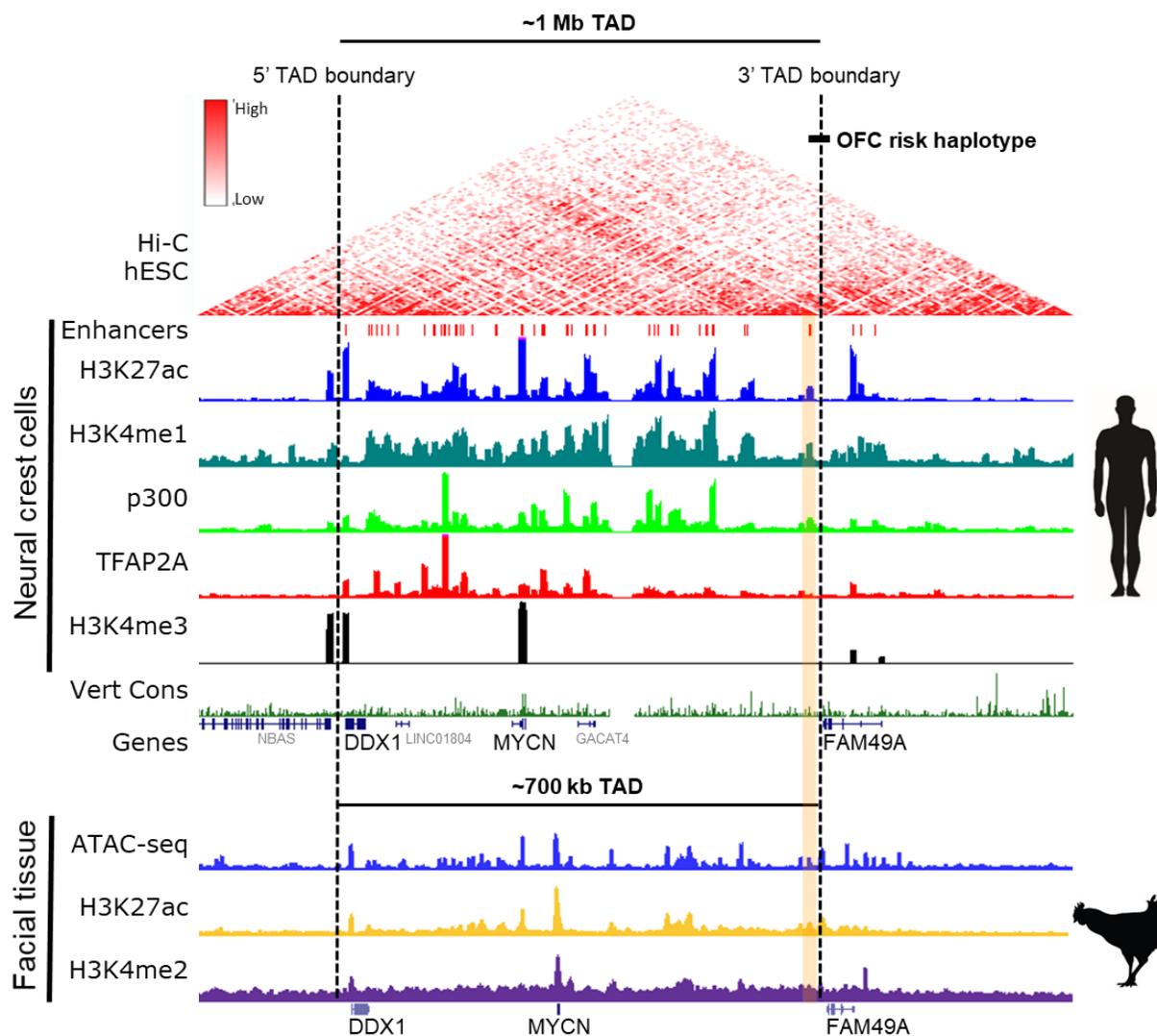
**Table 3.1** Reported OFC associated GWAS SNPs within the 2p24.2 locus.

dbSNP ID	Allele (ref/risk)	Position on chr2 (hg38)	# of SNPs in LD ( $r^2=1$ )	LD with rs7552	source	reported p-value
<b>rs4832647</b>	A/G	16524128	2	0.81	Ludwig et al. (2012)	$9.40 \times 10^{-6}$
<b>rs4497881</b>	C/A	16525623	2	0.93	Ludwig et al. (2012)	$1.77 \times 10^{-6}$
<b>rs4441471</b>	A/G	16534140	21	0.98	Ludwig et al. (2012)	$7.70 \times 10^{-7}$
<b>rs15653</b>	G/A	16550242	9	1.00	Ludwig et al. (2012)	$5.35 \times 10^{-6}$
<b>rs7566780</b>	A/G	16548089	0	0.85	Leslie et al. (2016)	$4 \times 10^{-9}$
<b>rs4832465</b>	T/G	16527280	21	0.98	Leslie et al. (2016)	$7.42 \times 10^{-7}$
<b>rs4832468</b>	C/T	16529297	21	0.98	Leslie et al. (2016)	$1.36 \times 10^{-7}$
<b>rs7552</b>	A/G	16552660	9	x	Leslie et al. (2016) Yu et al. (2017)	$4.22 \times 10^{-8}$ $5.83 \times 10^{-22}$

All OFC associated SNPs on 2p24.2 are in high LD ( $r^2 > 0.8$ ) with the most significant SNP rs7552 (Table 3.1). The haplotype comprising all associated SNPs on this locus stretches over a region of about 30 kb (chr2:16523421-16554164(hg38)). The 3' end of this region includes rs7552 and overlaps the 5' UTR of the *FAM49A* gene (see Figure 3.2.2).

Remarkably, according to publically available Hi-C data from hESC (Dixon et al., 2015), the Enh2p24.2 is located at the 3' end of a 1 Mb TAD whose 3' border is located right before the *FAM49A* gene. This suggests that Enh2p24.2 might not control *FAM49A* but genes located within its own TAD instead (i.e. DDX1 and MYCN) (see Figure 3.2.3).

The TAD harboring DDX1, MYCN and the candidate enhancer Enh2p24.2 displays an extremely high density of neural crest enhancers, while the immediately flanking TADs are basically enhancer free (Figure 3.2.3). This suggests that DDX1 and/or MYCN expression are tightly regulated in hNCC and, thus, likely to be functionally relevant in this cell type.

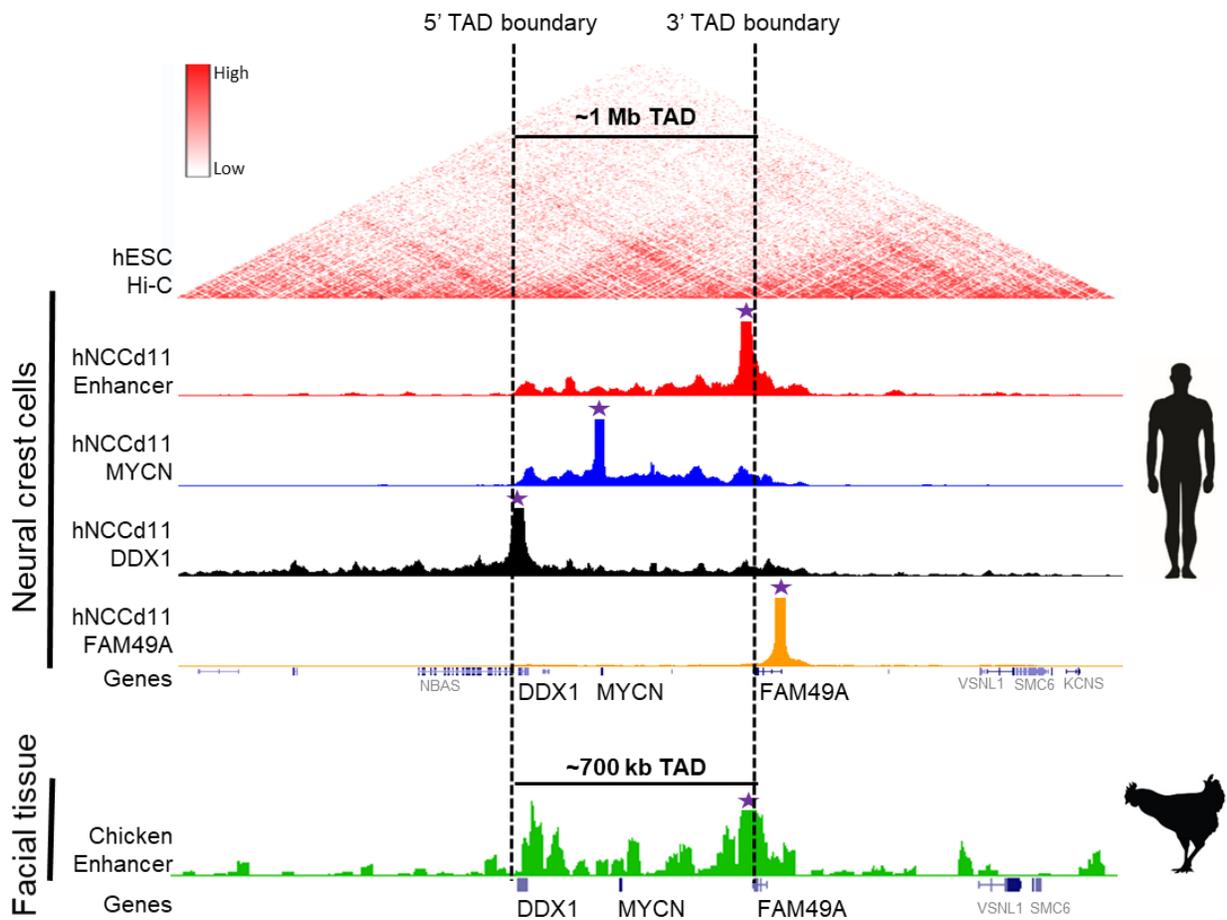


**Figure 3.2.3 Epigenetic profile of hNCC and chicken facial tissue of the TAD enclosing the Enh2p24.2 and both *MYCN* and *DDX1* but not *FAM49A*.**

Publicly available Hi-C data from hESC (Dixon et al., 2015) were combined with previously published ChIP-seq data for p300, TFAP2A and various histone modifications from hNCCd11 (Rada-Iglesias et al., 2012). For the corresponding homologous region in the chicken genome previously published H3K27ac ChIP-seq data (Rada-Iglesias et al., 2012) and newly generated H3K4me2 ChIP-seq and ATAC-seq data from the facial mesenchyme from chicken embryos are shown. ChIP-seq and ATAC-seq signals were normalized as RPGC (reads per genomic content) using five bp bins. The human candidate enhancer (Enh2p24.2) and its chicken homolog are highlighted in yellow and the location of the OFC risk haplotype is outlined in black in the upper part of the figure.

### 3.2.2 Linking the candidate enhancer to its putative target genes based on 3D chromatin architecture

Although the Hi-C data mentioned above already gave some insights into the 3D organization of the locus of interest, both the *Enh2p24.2* and *FAM49A* are located close to a TAD border. Consequently, it was difficult to discern whether they were in the same or different TADs. Additionally, the publically available Hi-C data was generated from hESC, while we were interested in uncovering the topological and regulatory architecture of the locus of interest in hNCC. Although TADs are considered to be generally conserved between cell types, differences, especially at the TAD borders, have also been reported (Bonev et al., 2017).



**Figure 3.2.4** 4C-seq profiles at the TAD enclosing *DDX1*, *MYCN* and *Enh2p24.2*.

Publically available Hi-C data from hESC (Dixon et al., 2015) were used to identify the TAD in which the *Enh2p24.2*, *MYCN* and *DDX1* are located. 4C-seq data were generated in WT hNCCp2 using the *Enh2p24.2* and the promoter of *MYCN*, *DDX1*, and *FAM49A* as viewpoints. Moreover, 4C-seq data from chicken facial tissue of HH24 stage embryos was also generated using the *Enh2p24.2* chicken homolog as a viewpoint (purple stars). 4C-seq signals were normalized as RPMs (reads per million) and a smoothing window of eight pixels was applied for visualization.

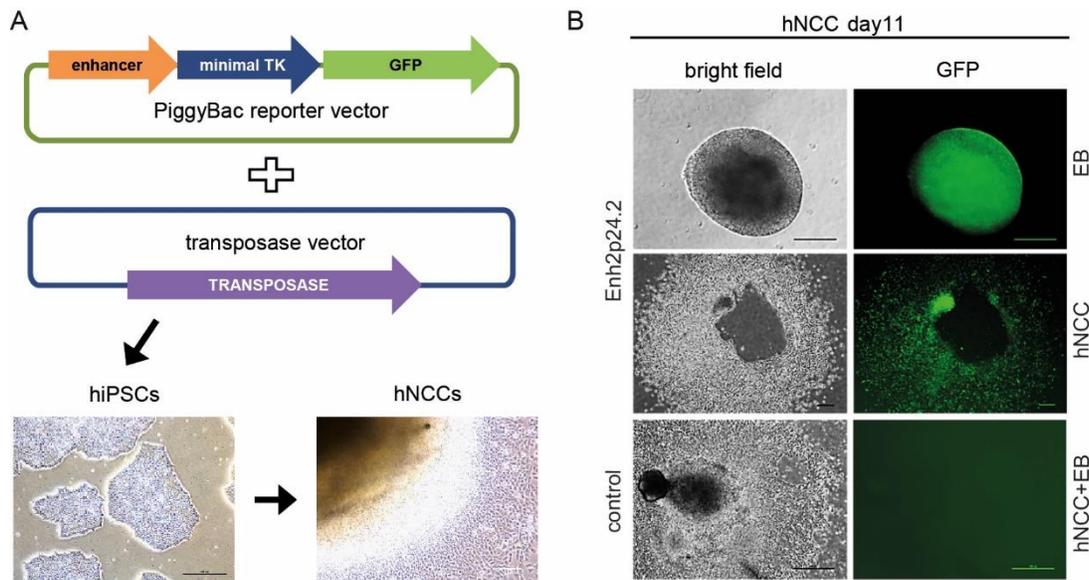
Thus, to more conclusively determine which genes interact with the Enh2p24.2, we performed 4C-seq experiments (Stadhouders et al., 2013) in WT hNCCp2 using the enhancer Enh2p24.2 as well as the promoters of MYCN, DDX1 and FAM49A as viewpoints (Figure 3.2.4). The 4C-seq profiles confirmed that the Enh2p24.2, MYCN and DDX1 interacted within the TAD identified by Hi-C. FAM49A, on the other hand appeared to be located within a different and small domain, insulated from Enh2p24.2 (Figure 3.2.4). Additionally, this regulatory topology seems to be evolutionarily conserved, since we observed a similar 4C-seq interaction profile in the facial tissue of chicken embryos, which is i.a. derived from NCC (Figure 3.2.4). All this suggests that the candidate enhancer physically interacts and potentially controls *DDX1* and/or *MYCN* rather than *FAM49A*.

### **3.2.3 Functional characterization of the 2p24.2 enhancer element in chicken embryos and human neural crest cells confirms its regulatory activity**

To assess whether the Enh2p24.2 displays enhancer activity in the neural crest and/or its derived facial tissue, we performed *in vivo* and *in vivo* reporter assays.

The *in vitro* reporter assay confirmed that Enh2p24.2 possesses enhancer activity, as strong GFP expression was observed in both neuroectodermal progenitors and hNCC outgrowths (Figure 3.2.5). This suggests that Enh2p24.2 displays enhancer activity during hNCC specification, delamination and migration.

## Results

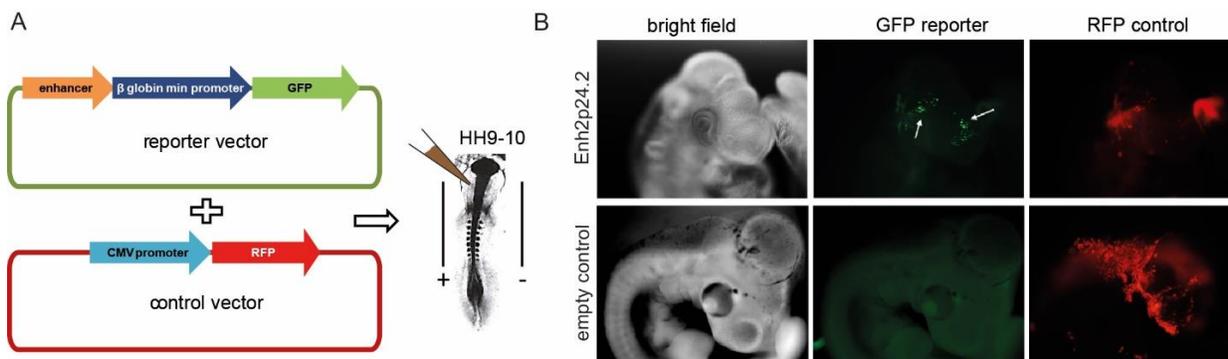


**Figure 3.2.5** *In vitro* reporter assay for the Enh2p24.2 in hNCC.

(A) Schematic diagram illustrating the *in vitro* reporter assay performed in hNCC. The human Enh2p24.2 sequence was cloned into a PiggyBAC reporter vector in which a minimal promoter (min TK promoter) drives GFP expression in the presence of an active enhancer. The resulting reporter vectors and a transposase-expressing vector were co-transfected into WT hiPSC. Following antibiotic selection, a population of hiPSC in which the reporter vector had been stable integrated into the human genome, was differentiated into hNCC. GFP signal was imaged on day 1 of differentiation. (B) Representative GFP signal observed in hNCC differentiated from hiPSC with integrated reporter vectors that were either empty (control) or contained the Enh2p24.2 sequence.

Complimentary results were obtained by the *in vivo* reporter assay, where GFP expression was induced by the human Enh2p24.2 sequence, specifically in NCC within the craniofacial region of stage HH20-23 chicken embryos (Figure 3.2.6). These observations suggest a regulatory role for Enh2p24.2 at later stages of NCC development, e.g. once the NCC arrive at their destination after migration from the neural tube.

Overall, the previous reporter assays confirmed the regulatory activity of the candidate enhancer both in human and in chicken NCC. Consequently, it is likely that the OFC risk region is indeed relevant to facial development and that the risk variants within Enh2p24.2 might affect the expression of the enhancer target genes.



**Figure 3.2.6** *In vivo* reporter assay for the Enh2p24.2 in chicken embryos.

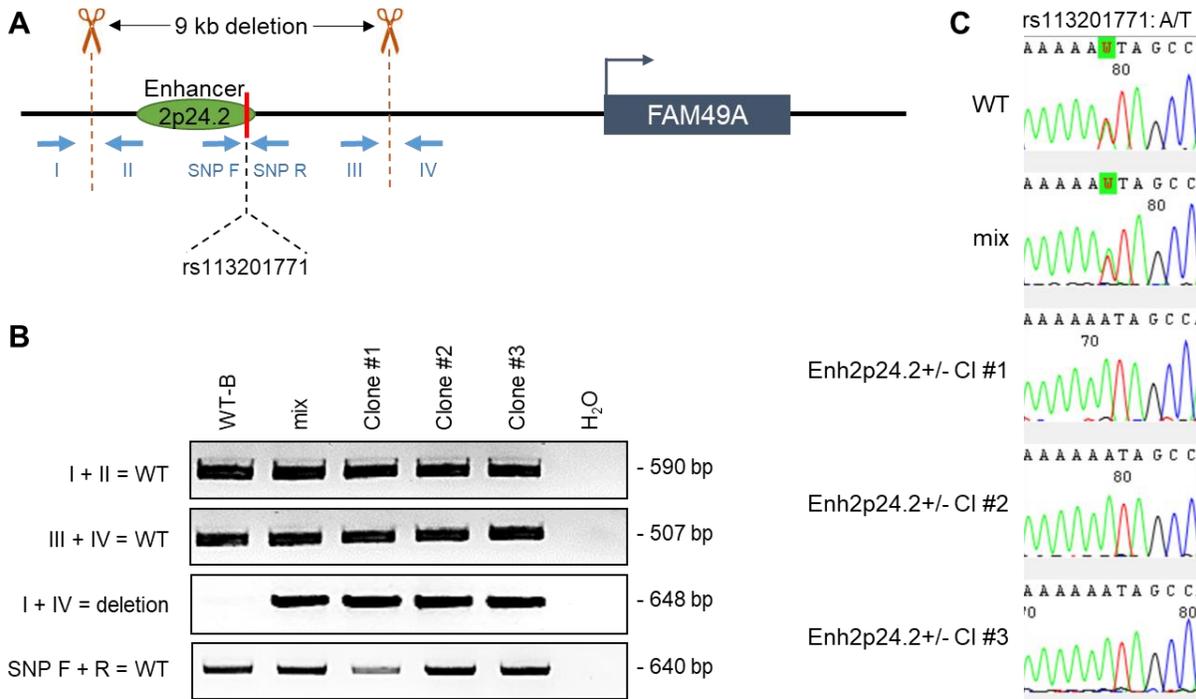
(A) Schematic diagram illustrating the *in vivo* reporter assay performed in chicken embryos. The human enhancer sequence of the Enh2p24.2 was cloned into a reporter vector in which a minimal promoter (min  $\beta$ -globin promoter) drives GFP expression in the presence of an active enhancer. A second vector in which a strong promoter (CMV promoter) drives RFP expression was used as a positive control for successful electroporation (control vector/RFP control). Both vectors were co-electroporated into the anterior neural tube of stage HH9-10 chicken embryos to evaluate GFP and RFP expression at later stages. (B) Representative images of stage HH20-23 chicken embryos co-electroporated with the control vector (RFP control) and the reporter vector (GFP reporter), either containing the human enhancer sequence or empty (empty control). Chicken embryos are shown in lateral view.

To more conclusively test the functional relevance of the enhancer Enh2p24.2, we deleted it in WT hiPSC using CRISPR/Cas9 engineering. Briefly, we used a pair of guide RNAs (gRNAs) located 3' and 5' of the enhancer region to create a deletion of about 9 kb, spanning the approximately 2 kb long Enh2p24.2 as well as its flanking sequences (Figure 3.2.7).

Notably, despite several attempts, we were only able to obtain clones with a heterozygous deletion of the enhancer (Figure 3.2.7). This suggests an essential role for some of the regulatory elements located within the deleted region, either the enhancer itself or other so far unidentified components.

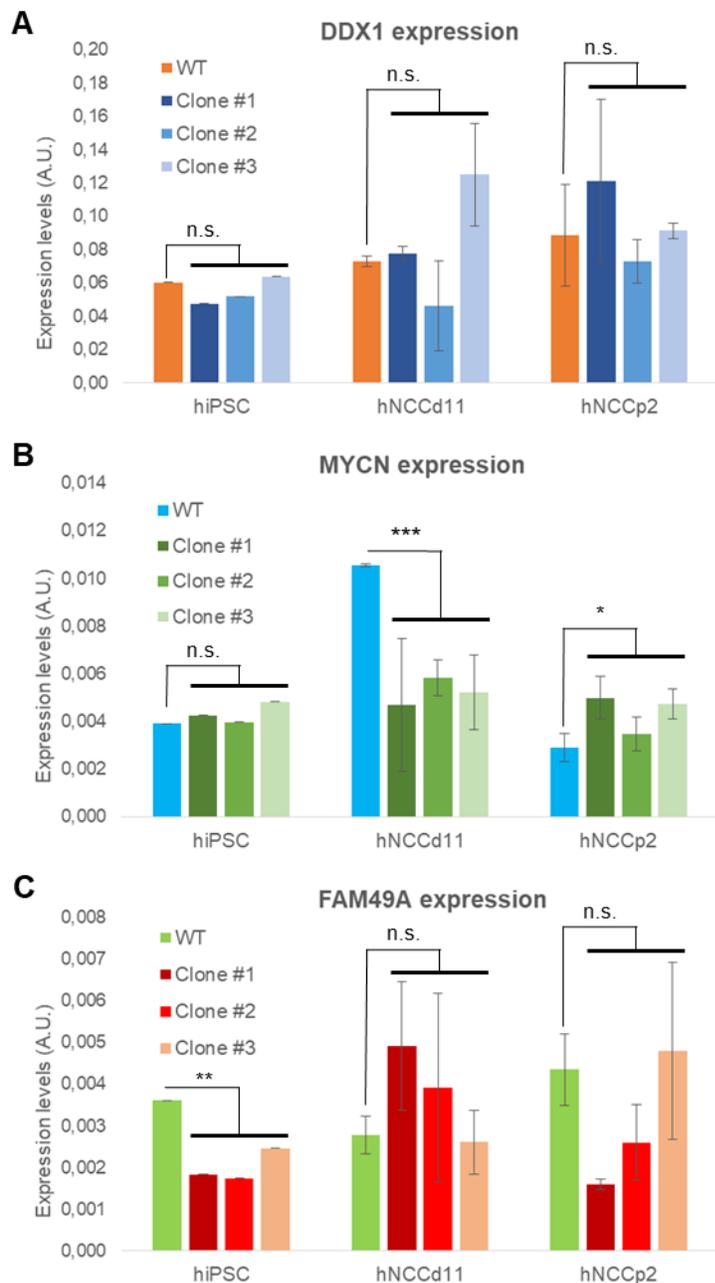
The heterozygous clones (Enh2p24.2<sup>+/-</sup> CI #1, CI #2, CI #3), though, did not show any obvious phenotype. The survival of these three different Enh2p24.2<sup>+/-</sup> hiPSC clonal lines or their hNCC differentiation potential did not differ from that of their parental WT hiPSC line (data not shown).

## Results



**Figure 3.2.7 Generation of a heterozygous deletion of the candidate enhancer Enh2p24.2 using CRISPR/Cas9.** (A) Schematic diagram illustrating the CRISPR/Cas9 strategy used to engineer a 9 kb deletion spanning the enhancer Enh2p24.2 in WT hiPSC. The location of the pair of gRNAs used to generate the deletion as well as the PCR primers used to determine the presence of WT and deletion alleles are shown as red scissors and blue arrows, respectively. The different primer pairs were specific either for the WT (I+II, III+IV, SNP F+R) or the deletion (I+IV) allele. (B) Genotyping results obtained by PCR using the primers described above are shown for WT hiPSC, a mixed population of CRISPR/Cas9 targeted hiPSC, and three hiPSC clonal lines heterozygous for the Enh2p24.2 deletion (Enh2p24.2+/- Cl #1, #2, #3). (C) Chromatogram illustrating the Sanger sequencing results of the previously described SNP F+R PCR products at the location of the rs113201771 SNP. This SNP is heterozygous (A/T) in the WT parental hiPSC line and was therefore used to confirm that the Enh2p24.2 deletions were heterozygous in the previous clonal lines.

To investigate whether the deletion of one Enh2p24.2 allele had any influence on the expression of its putative target genes, we performed RT-qPCR for *DDX1*, *MYCN*, and *FAM49A* in hiPSC, hNCCd11 and hNCCp2 (Figure 3.2.8). While transcription of *DDX1* was not clearly altered by the heterozygous deletion of Enh2p24.2, levels of *MYCN* were lower in hNCCd11 and, surprisingly, *FAM49A* expression was reduced in hiPSC. These observations indicate that the deleted region might regulate expression of *FAM49A* and *MYCN*, but only in certain cell types (hiPSC and hNCCd11, respectively) at different stages of embryonic development. Furthermore, in the case of *DDX1*, it is possible that, under normal conditions, the function of the missing Enh2p24.2 is compensated by any of the numerous other enhancers within its TAD (compare to Figure 3.2.3).



**Figure 3.2.8 Expression levels of the putative target genes of *Enh2p24.2* measured by RT-qPCR in WT cells and cells with a heterozygous enhancer deletion.**

Gene expression was measured by RT-qPCR using RNA collected from three *Enh2p24.2*<sup>+/-</sup> hiPSC clones and their isogenic WT hiPSC as well as upon differentiation into hNCCd11, and hNCCp2. *DDX1*, *MYCN* and *FAM49A* expression values were normalized to three housekeeping genes (*ACTB*, *EEF2*, *GAPDH*). The error bars represent standard deviation from three measurements in hiPSC (technical triplicates per sample), six measurements for in hNCCd11 (technical triplicates in two biological replicates per sample) and nine measurements in hNCCp2 (technical triplicates in three different biological replicates per sample). P-values were calculated using Student's t-test (n.s. = not significant, \*p-value  $\leq$  0.05, \*\*p-value  $\leq$  0.01).

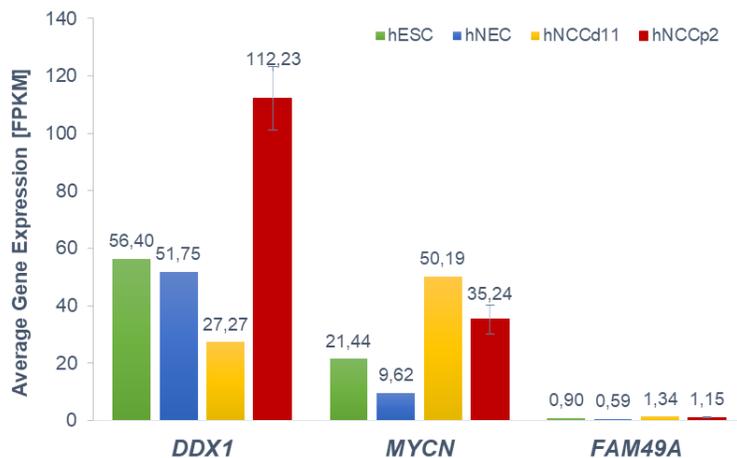
Due to the impossibility to generate hiPSC with homotygous deletions of *Enh2p24.2*, the impact of this enhancer on the expression of its putative target genes could not be unambiguously

assessed. Additional experiments such as allele-specific expression analysis could help clarifying the function of the enhancer. Still, our results show that the OFC-risk region 2p24.2 contains an enhancer with regulatory activity in neural crest cells and thus supports the potential significance of this locus for NCC-related developmental processes and craniofacial abnormalities like OFC.

### 3.2.4 *DDX1* and *MYCN*, but not *FAM49A*, are expressed in the developing facial tissues involved in OFC

The 2p24.2 OFC-risk locus is situated within an evolutionarily conserved regulatory domain that displays high genetic, epigenetic and topological similarities between distant species (i.e. human and chicken). In order to gather additional evidences that could support the etiological relevance of the locus in OFC, we then investigated the expression profile of the nearby genes (i.e. *DDX1*, *MYCN* and *FAM49A*).

According to previous RNA-seq data generated in WT hESC, neuroectoderm (hNEC) and hNCCd11, (Rada-Iglesias et al., 2012; Rada-Iglesias et al., 2011) and hNCCp2 (Laugsch et al., 2019), *DDX1* and *MYCN* are expressed in considerable amounts both in hESC and hNCC while *FAM49A* is lowly expressed in all cell types considered (Figure 3.2.9).

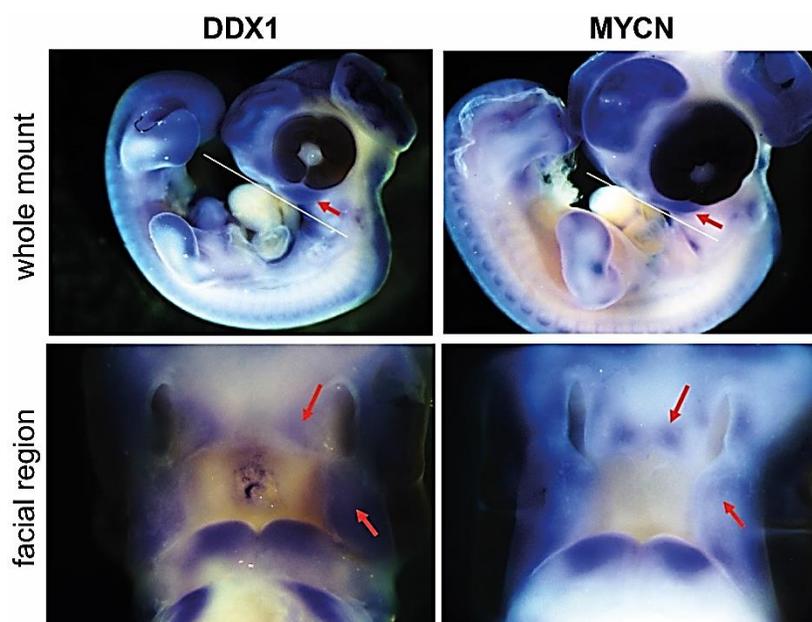


**Figure 3.2.9** Transcription levels of *DDX1*, *MYCN* and *FAM49A* in human ESC, NEC and NCC.

Expression levels measured by RNA-seq are shown as FPKM in WT human embryonic stem cells (hESC), neuroectoderm (hNEC) and neural crest cells (hNCC) for *DDX1*, *MYCN* and *FAM49A*. hESC and hNEC data are taken from Rada-Iglesias et al. (2011), hNCCd11 data from Rada-Iglesias et al. (2012) and hNCCp2 data from Laugsch et al. (2019). Error bars for hNCCp2 represent standard deviations from two biological replicates while the remaining RNA-seq datasets were generated as single replicates.

To test if the chicken homologues of the human *DDX1* and *MYCN* were also expressed in chicken embryos, and more importantly, to determine the precise expression pattern of these two genes, we performed *in situ* hybridization in stage HH25 chicken embryos (Figure 3.2.10).

In accordance with the regulatory landscape of both *DDX1* and *MYCN* in the chicken facial mesenchyme, which suggests that these two genes were likely to be active (Figure 3.2.3), *in situ* hybridization in chicken embryos confirmed that *DDX1* and *MYCN* were specifically expressed in distinct parts of the developing face (Figure 3.2.10). Both the central frontonasal prominence (FNP) and the paired maxillary prominences (MXP) express *DDX1* as well as *MYCN*. Importantly, these two prominences correspond to the human tissues that fail to fuse when OFC occurs, thus further supporting the relevance of these two genes in the development of the face and in OFC etiology.



**Figure 3.2.10 Expression pattern of *DDX1* and *MYCN* in chicken embryos.**

Whole mount *in situ* hybridization was performed in stage HH25 chicken embryos with probes against *DDX1* and *MYCN* RNAs. Embryos are presented as whole mount images in lateral and ventral views. The ventral images are focused on the facial region as indicated by a white line in the lateral view images. Red arrows indicate regions with considerable gene expression in those facial tissues implicated in OFC.

In contrast, despite extensive efforts to perform *in situ* hybridization experiments for *FAM49A*, it was not even possible to generate hybridization probes for this gene due to its overall low expression during the relevant embryonic stages (data not shown). This further suggests that *FAM49A* is neither involved facial development nor in OFC.

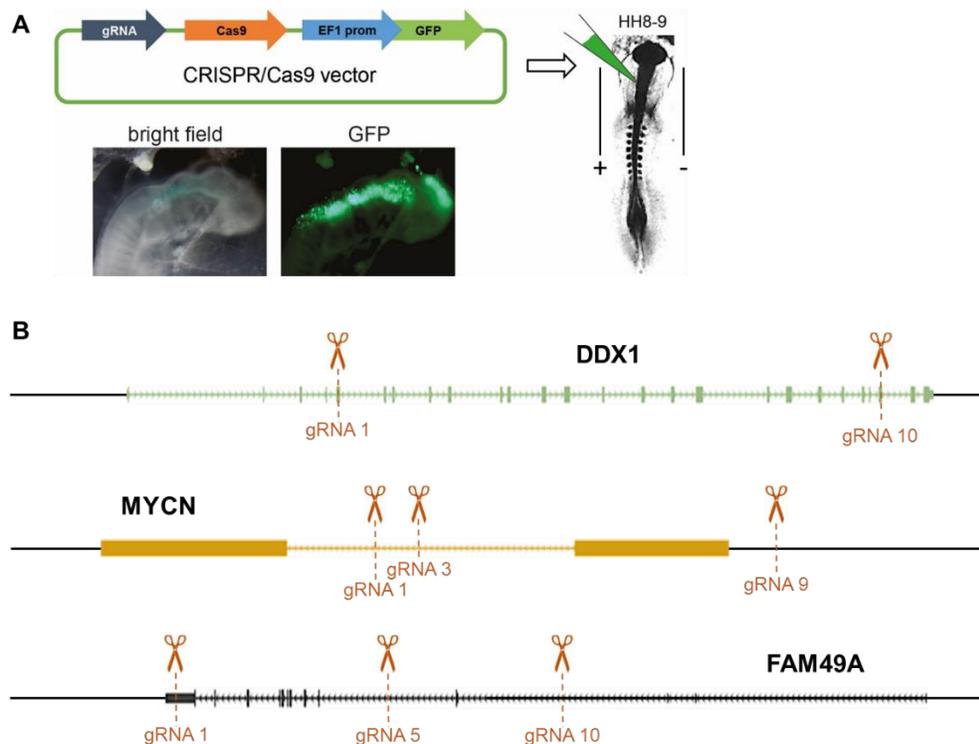
Mutations within *MYCN* have been reported in patients with Feingold syndrome (OMIM#164280), an autosomal dominant disorder characterized by a number of malformations including facial abnormalities (Burnside et al., 2018; Chen et al., 2012; Marcelis et al., 2008; Teszas et al., 2006; van Bokhoven et al., 2005). Furthermore, *MYCN* has been described as a proto-oncogene in neuroblastoma, an NC-derived pediatric tumor (Brodeur et al., 1984; Kohl et al., 1984; Kohl et al., 1983; Kramer et al., 2016; Liu et al., 2019; Schwab et al., 1983; Zeid et al., 2018). Therefore *MYCN* is considered to be important for facial development and, thus, represents an obvious candidate gene whose misexpression due to genetic variants within the 2p24.2 risk haplotype could contribute to OFC.

*DDX1*, on the other hand, has not been previously implicated in either facial development or OFC. Therefore, in order to provide novel insights into neural crest and craniofacial development as well as into the etiology of OFC, we decided to preferentially focus on *DDX1*

### **3.2.5 *In vivo* loss of function assays reveal an essential role for *DDX1* in facial development**

To directly assess how crucial the candidate genes (*DDX1*, *MYCN*) are for facial development, and to more conclusively exclude *FAM49A* as a player in OFC, we performed CRISPR/Cas9 in chicken embryos to generate mosaic gene knockouts specifically in the cranial neural crest.

Briefly, we unilaterally electroporated the NC progenitors located in the anterior-dorsal neural tube of stage HH9-10 chicken embryos with vectors expressing Cas9-GFP and gRNAs targeting *DDX1*, *MYCN*, or *FAM49A*. The gRNAs were designed to generate, upon NHEJ repair, internal loss-of-function deletions within each of the candidate genes (Figure 3.2.11).

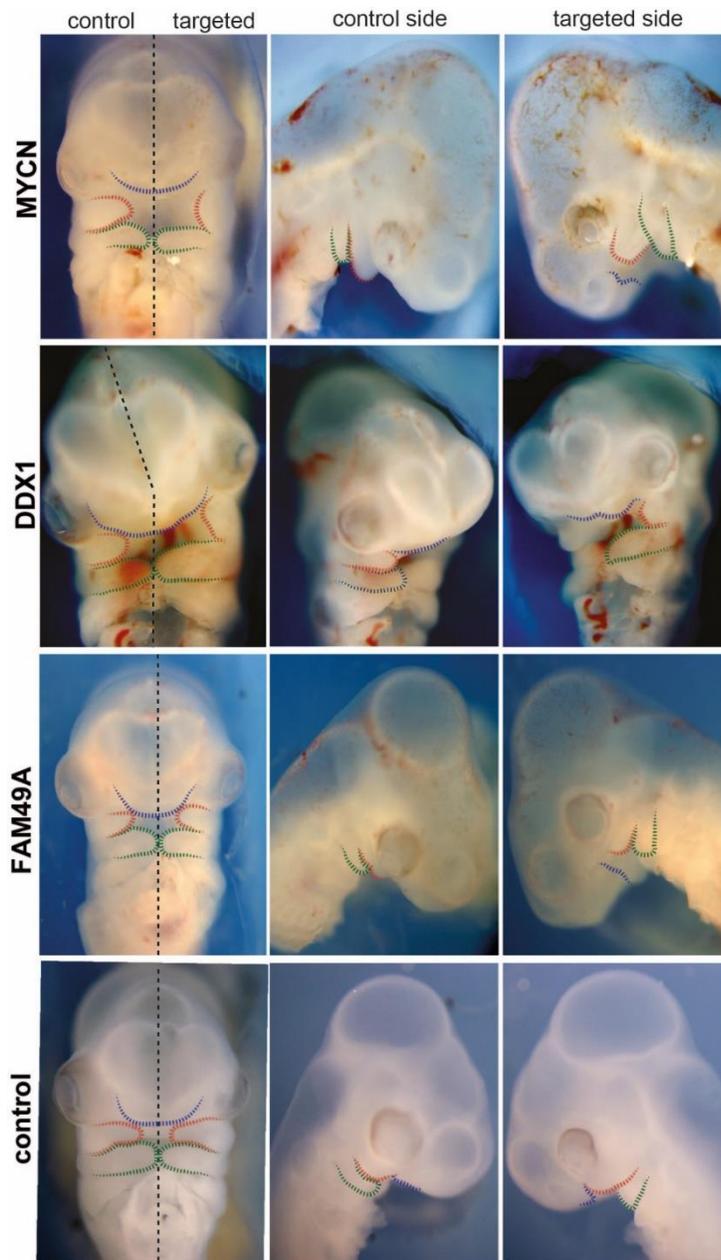


**Figure 3.2.11** Schematic diagram depicting the CRISPR/Cas9 strategy used to target the candidate genes *in vivo*.

**(A)** Schematic diagram illustrating the *in vivo* CRISPR/Cas9 experiments performed in chicken embryos. The anterior-dorsal neural tube of stage HH9-10 chicken embryos, where NCC progenitors reside, were unilaterally electroporated with pX330A-Cas9 vectors expressing Cas9 and gRNA against the candidate genes. These vectors also express GFP, allowing the electroporated cells to be transiently tracked. **(B)** Chicken *DDX1* was targeted with a pair of gRNAs designed within exons that can cause a disruption of the gene by creating small indels at individual gRNA sites within or by deletion/inversion of the entire region between the two gRNAs. *MYCN* and *FAM49A* were targeted with three gRNAs each in order to create as many possible combinations of disruptive re-arrangements (deletions and inversions) between pairs of gRNAs.

Strikingly, CRISPR/Cas9 targeting of *DDX1* and *MYCN* within the neural crest had a substantial and highly reproducible effect on the electroporated side of the face, while the control side was unaffected (Figure 3.2.12). Most notably, the MXP and the periocular mesenchyme were clearly reduced in the electroporated side, thus, resulting in an early orofacial cleft phenotype. In contrast, targeting of *FAM49A* or electroporation of chicken embryos with the empty Cas9 vector (i.e. without gRNAs) had no effect on facial development (Figure 3.2.12).

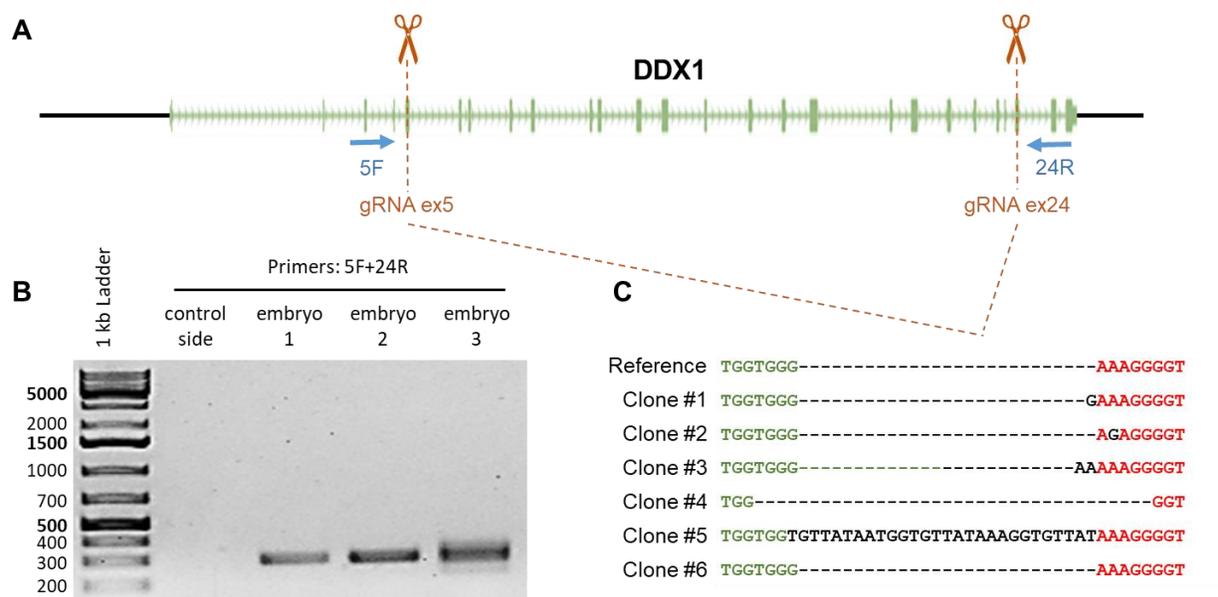
## Results



**Figure 3.2.12 Genetic engineering of chicken neural crest cells by CRISPR/Cas9 shows a distinct facial phenotype in MYCN and DDX1 targeted embryos.**

Representative images of stage HH20-23 chicken embryos after unilateral targeting of Cas9 with gRNAs that direct the nuclease to *DDX1*, *MYCN* or *FAM49A* or without any gRNA (control). The embryos are displayed in ventral view (left panels) showing both the untargeted control and the targeted side next to each other as well as in lateral view (middle and right panels) of the control and targeted sides, respectively. The frontonasal prominence (FNP) is outlined in blue, maxillary prominences (MXP) in red and the mandibular prominences (MNP) in green.

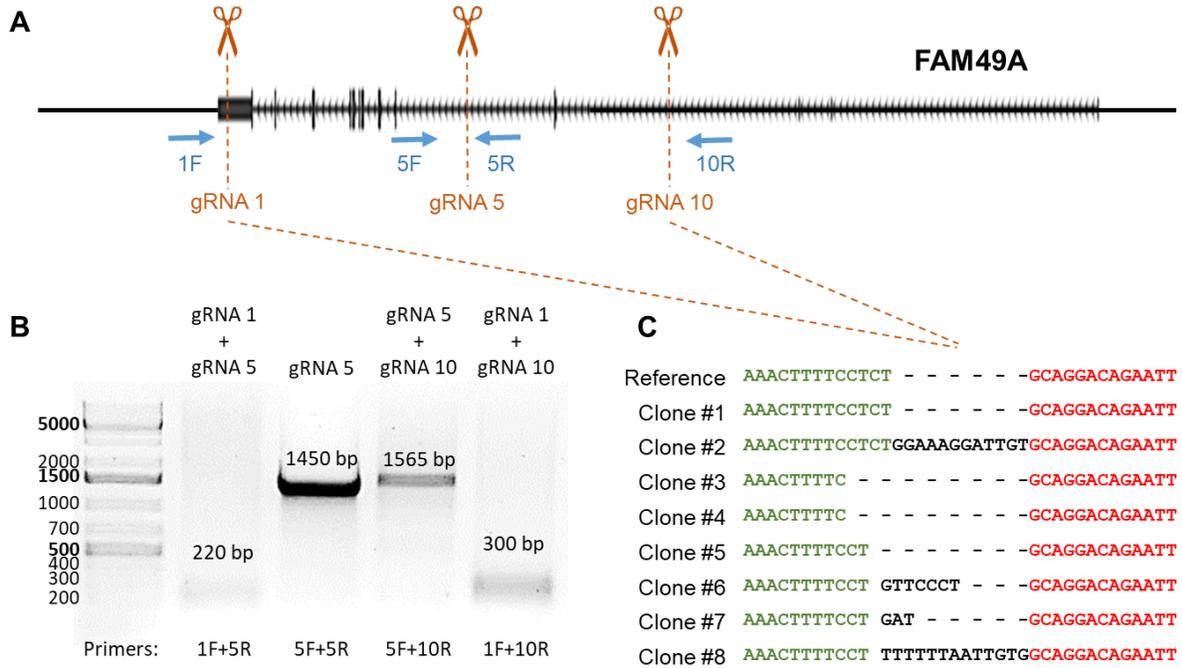
The presence of the expected deletions in *DDX1* and *FAM49A* was detected by PCR followed by cloning into the pGEM-T vector and Sanger sequencing (Figure 3.2.13, Figure 3.2.14). This proved that *FAM49A* was in fact deleted in the targeted embryos but the deletion did not have any phenotypic effect during early facial development.



**Figure 3.2.13 CRISPR/Cas9 mediated deletions within the chicken *DDX1* gene.**

**(A)** Schematic diagram of the chicken *DDX1* gene with the location of the gRNAs and the PCR primers used to detect the presence of deletions within a population of targeted cells. DNA extracted from CRISPR/Cas9 targeted neural tubes or the control side of stage HH14-16 embryos was amplified by PCR using the primers above. **(B)** The PCR primers specifically amplified the desired deletions but not the much larger WT sequence. **(C)** The resulting PCR products were cloned into the pGEM-T vector (Promega). Several clones of the pGEM-T vector containing the *DDX1* deletion PCR product were Sanger sequenced and aligned to the expected sequence of the chicken genome. Green letters represent nucleotides downstream of gRNA 1, red letters represent nucleotides upstream of gRNA 10, black letters and X represent insertions and deletions, respectively. Black dashes connect neighboring nucleotides in when other clones display additional nucleotides in this location.

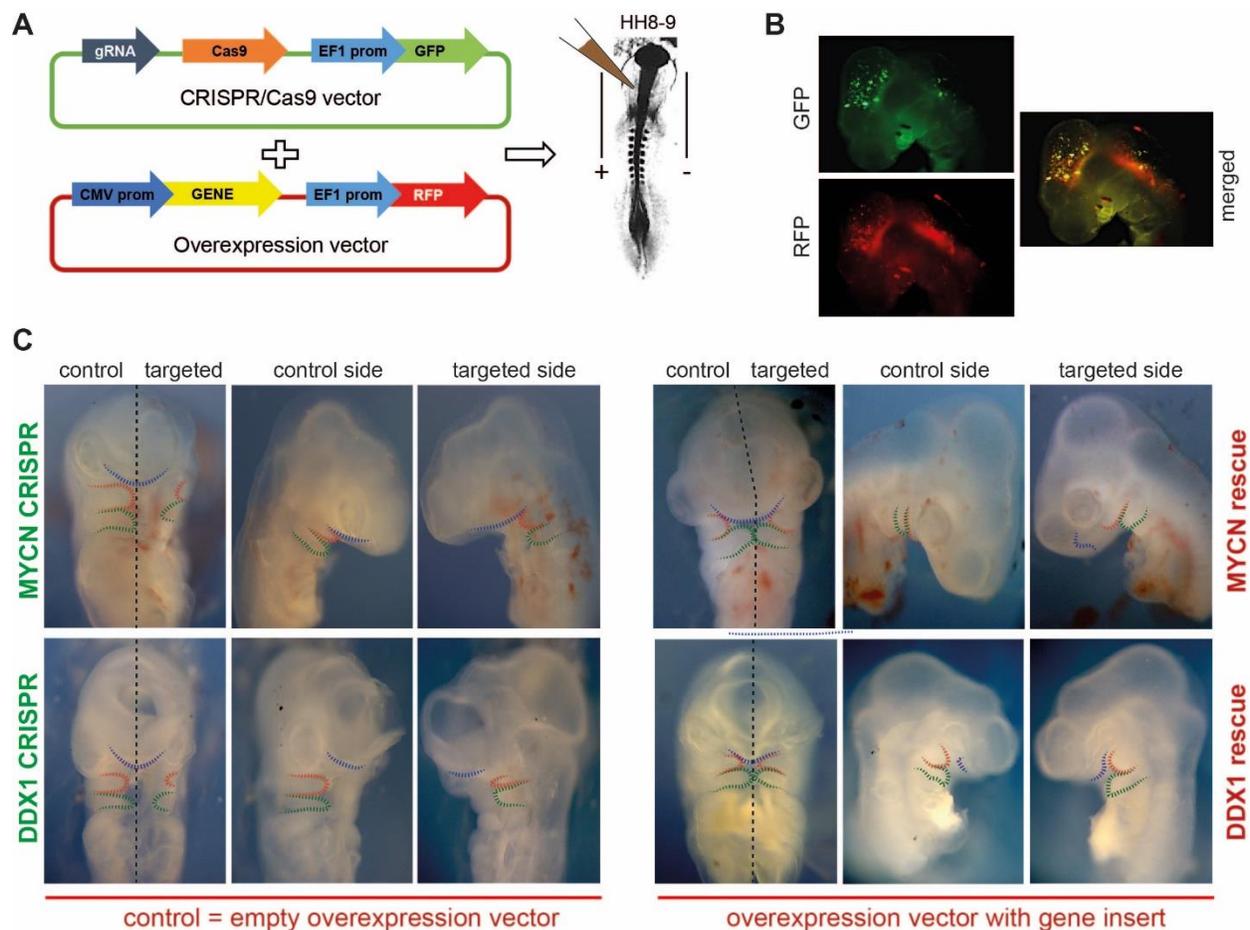
## Results



**Figure 3.2.14 CRISPR/Cas9 mediated deletion within the chicken *FAM49A* gene.**

(A) Schematic diagram of the chicken *FAM49A* gene with the location of the gRNAs and the PCR primers used to detect the presence of deletions within a population of targeted cells. DNA extracted from CRISPR/Cas9 targeted neural tubes of stage HH14-16 embryos was amplified by PCR using the primers above. (B) Specific combinations of the primers described above (1F+5R, 5F+10R, 1F+10R) were used to genotype any possible deletions between pairs of gRNAs, while the primer combination 5F+5R demonstrated the presence of WT alleles within the cell population. (C) The PCR product of the 1F+10R primer combination, amplifying the largest possible deletion was cloned into the pGEM-T vector (Promega). Several clones of the pGEM-T vector containing the *FAM49A* deletion PCR product were Sanger sequenced and aligned to the expected sequence of the chicken genome. Green letters represent nucleotides downstream of gRNA 1, red letters represent nucleotides upstream of gRNA 10, black letters and X represent insertions and deletions, respectively. Black dashes connect neighboring nucleotides in when other clones display additional nucleotides in this location.

Despite extensive efforts, it was not possible to detect deletions within *MYCN*, probably due to the high content of repetitive and GC-rich sequences in this locus. To demonstrate that the observed facial defects were caused by the specific loss of either *DDX1* or *MYCN* function, we performed rescue experiments. Briefly, Chicken embryos were co-electroporated with Cas9 vectors targeting either *DDX1* or *MYCN* together with a vector overexpressing the corresponding human homolog, which could not be targeted by the chicken-specific gRNAs (Figure 3.2.15).



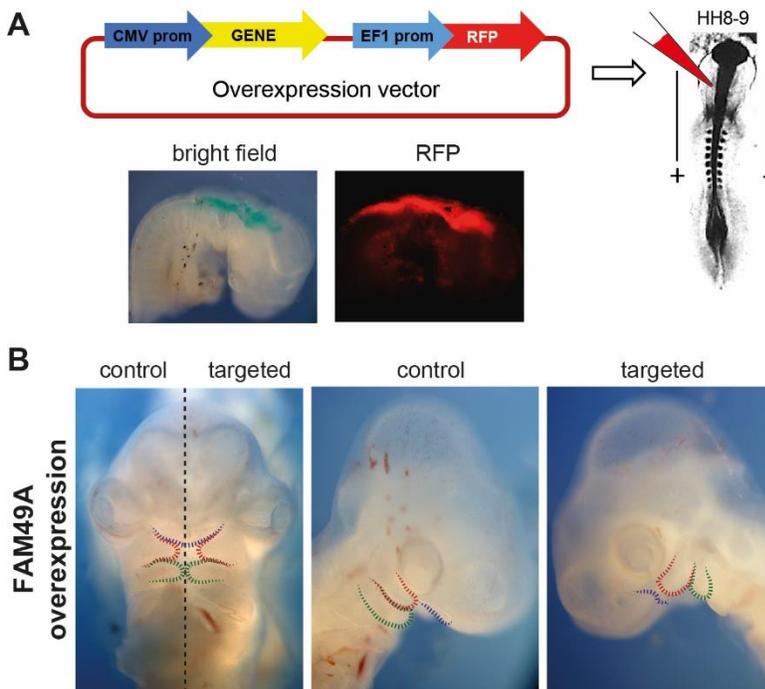
**Figure 3.2.15** Rescue experiments for *MYCN* and *DDX1* in CRISPR/Cas9 engineered chicken embryos.

(A) Schematic diagram of rescue experiments in which vectors to inactivate genes of interest in chicken embryos were combined with the simultaneous overexpression of the corresponding human homolog. The anterior-dorsal neural tube of stage HH9-10 chicken embryos, where NCC progenitors reside, were unilaterally co-electroporated with the pX330A-Cas9 vector and a vector expressing the human gene (*MYCN* or *DDX1*). (B) Representative images showing the fluorescent signal in electroporated neural crest cells within developing chicken embryos. While the CRISPR/Cas9 vector expresses GFP, the overexpression vector expresses RFP. (C) Representative images of stage HH20-23 chicken embryos after unilateral CRISPR/Cas9 engineering of *MYCN* or *DDX1* either together with the empty overexpression vector (control, left panel) or with the same vector expressing the human *MYCN* or *DDX1* genes, respectively. The embryos are displayed in ventral view (left panels) showing both the untargeted control and the targeted side next to each other and in lateral view (middle and right panels) for the control and targeted sides separately. The frontonasal prominence (FNP) is outlined in blue, maxillary prominences (MXP) in red and the mandibular prominences (MNP) in green.

The previous loss of function experiments support an important role for both *MYCN* and *DDX1* during craniofacial development. In contrast, *FAM49A* does not seem to be involved in this process, at least under normal conditions (Figure 3.2.9, Figure 3.2.12). Alternatively, OFC could also be caused by the ectopic expression of a particular gene in the wrong cellular context, i.e. *FAM49A* expression in NCC. To test this possibility, we performed gain of function experiments for *FAM49A* in chicken embryos. However, following electroporation of the anterior-dorsal neural

## Results

tube of stage HH9-10 chicken embryos with a vector overexpressing FAM49A, no obvious facial phenotypes were observed (Figure 3.2.16).



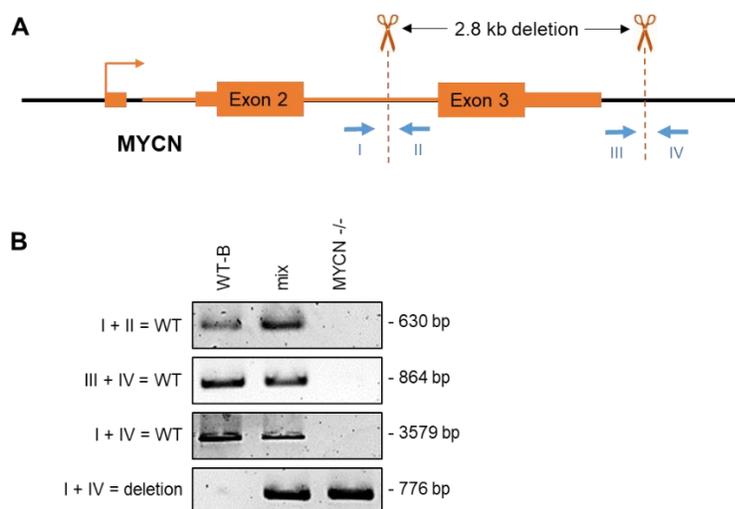
**Figure 3.2.16 Overexpression of *FAM49A* in neural crest cells of chicken embryos**

(A) Schematic diagram illustrating the *in vivo* overexpression experiments performed in chicken embryos. The anterior neural tube of stage HH9-10 chicken embryos, where NCC progenitors reside, were unilaterally electroporated with the PiggyBAC vector containing the human *FAM49A* sequence, which also expresses RFP (B) Representative images of stage HH20-23 chicken embryos after unilateral overexpression of *FAM49A*. The embryo is displayed in ventral view (left panel) showing both the untargeted control and the targeted side next to each other and in lateral view (middle and right panels) of control and targeted sides separately. The FNP is outlined in blue, the MXP in red and the MNP in green.

These *in vivo* loss- and gain-of-function experiments conclusively exclude *FAM49A* as a major regulator of facial development, while supporting an important role for *MYCN* and *DDX1* in this process. Furthermore, these experiments suggest that the misexpression of *MYCN* and/or *DDX1*, for example due to variants within the 2p24.2 risk locus, could be causally involved in OFC.

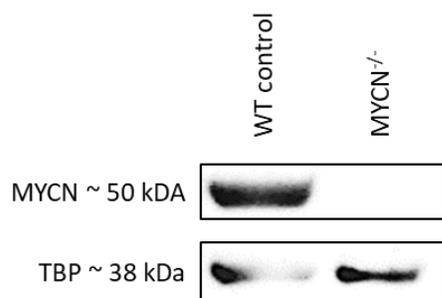
### 3.2.6 The cellular phenotype of *DDX1* knockout cells suggests a proliferation defect

To investigate the cellular mechanism/s behind the phenotypes observed upon loss of *DDX1* and *MYCN* function *in vivo*, we performed similar CRISPR/Cas9-mediated genome engineering experiments in our hiPSC *in vitro* differentiation system (Figure 3.2.17, Figure 3.2.19).



**Figure 3.2.17 Generation of MYCN<sup>-/-</sup> hiPSC using CRISPR/Cas9.**

**(A)** Schematic diagram illustrating the CRISPR/Cas9 strategy used to engineer a 2.8 kb deletion within the MYCN gene with a pair of gRNA together with the location of the PCR primers used to determine the presence of WT and deletion alleles. WT hiPSC were co-transfected with a pair of vectors expressing the Cas9 nuclease and gRNAs targeted to regions located 5' and 3' of the third exon of MYCN, respectively. The herewith deleted exon contains most of the functional domains of MYCN and therefore effectively knocks out the function of the protein. Different primer pairs were specific for the WT allele (I+II, III+IV) and the deletion (I+IV). **(B)** Genotyping results obtained by PCR using the primers described above are shown for WT hiPSC, a mixed population of CRISPR/Cas9 targeted hiPSC, and a hiPSC clonal line homozygous for the MYCN deletion.



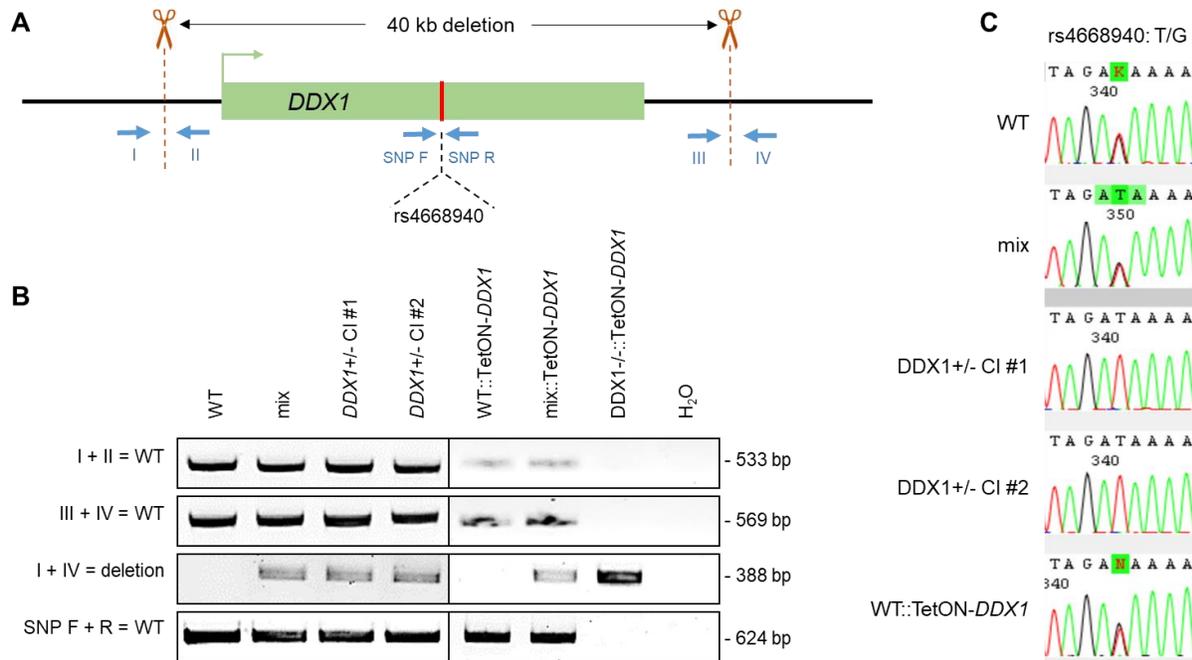
**Figure 3.2.18 Western blot analysis of MYCN in MYCN<sup>-/-</sup> hiPSC.**

**(A)** Proteins were extracted from MYCN<sup>-/-</sup> hiPSC and its isogenic WT control. Equal amounts of protein were loaded and immunoblotted against MYCN. TBP was used as a loading control.

While we were seemingly able to create MYCN<sup>-/-</sup> hiPSC (Figure 3.2.17, Figure 3.2.18), generation of a DDX1 deletion proved to be considerably more difficult. Despite great efforts, we initially only obtained heterozygous hiPSC clones, suggesting that DDX1 is essential for the survival of hiPSC (Figure 3.2.19). This is in full agreement with the early embryonic lethality (E3.5) of DDX1<sup>-/-</sup> mice *in vivo* and unsuccessful efforts to generate homozygous DDX1 deletions in other human cells *in vitro* (Hildebrandt et al., 2015; Ribeiro de Almeida et al., 2018). Hence, we engineered the homozygous DDX1 deletion on hiPSC in which exogenous DDX1 could be inducibly

## Results

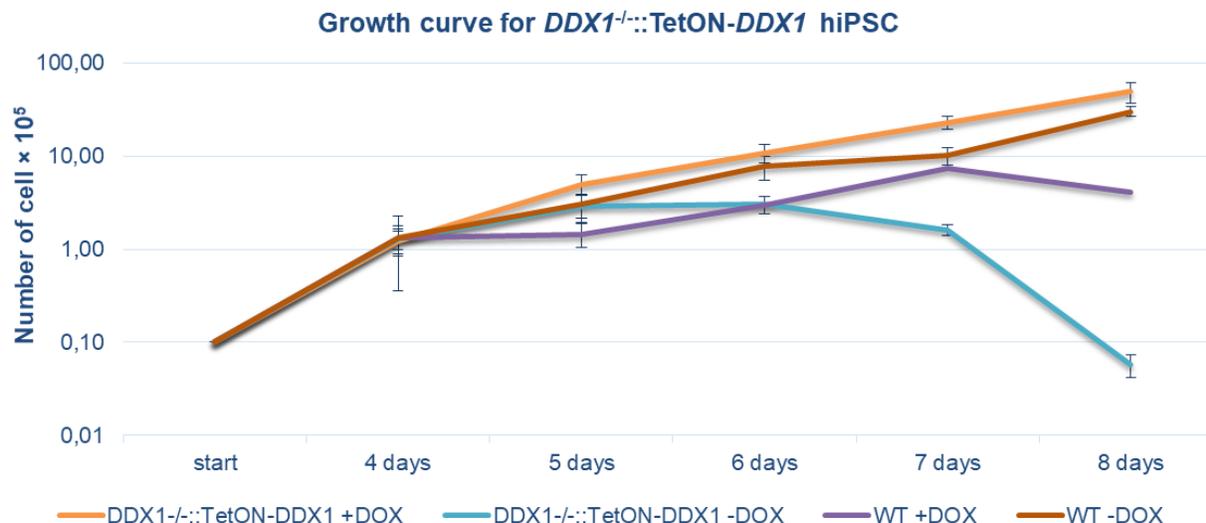
overexpressed upon addition of Doxycycline (DOX) (TetON-*DDX1* hiPSC) (Figure 3.2.19). This strategy enabled us to rescue the lethal phenotype observed in hiPSC and investigate the effects of *DDX1* loss-of-function during NC differentiation.



**Figure 3.2.19 CRISPR/Cas9 mediated deletion of the human *DDX1* gene.**

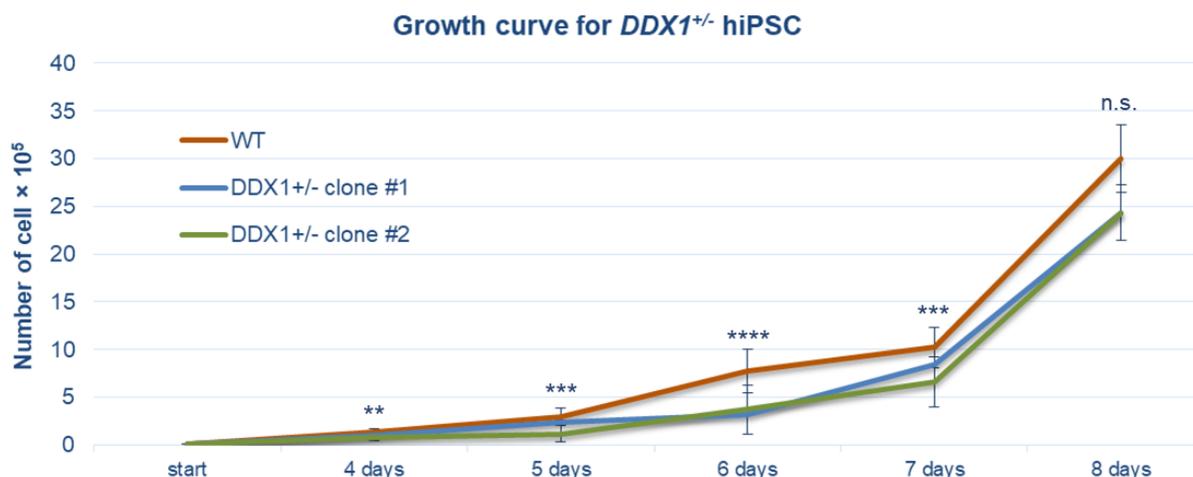
(A) Schematic diagram illustrating the CRISPR/Cas9 strategy used to engineer a 40 kb deletion of *DDX1* with a pair of gRNAs together with the location of the PCR primers used to determine the presence of WT and deletion alleles. WT hiPSC were co-transfected with a pair of vectors expressing the *Cas9* nuclease and gRNAs targeted to regions located 5' and 3' of *DDX1*, respectively. Different primer pairs were specific for the WT allele (I+II, III+IV, SNP F+R) and the deletion (I+IV). (B) Genotyping results obtained by PCR using the primers described above are shown for WT hiPSC, a mixed population of CRISPR/Cas9 targeted hiPSC, and two hiPSC clonal lines heterozygous for the *DDX1* deletion (*DDX1*<sup>+/-</sup> CI #1, *DDX1*<sup>+/-</sup> CI #2). Additionally, the rescue parental line, carrying a doxycycline inducible exogenous copy of the *DDX1* gene (WT::TetON-*DDX1*), a mixed population of the CRISPR/Cas9 targeted hiPSC (mix::TetON-*DDX1*) and a hiPSC clonal line homozygous for the *DDX1* deletion (*DDX1*<sup>-/-</sup>::TetON-*DDX1*) were genotyped by PCR. (C) Chromatogram illustrating the Sanger sequencing results of the previously described SNP F+R PCR products at the location of the rs4668940 SNP, which is heterozygous (T/G) in the WT parental hiPSC line and was therefore used to confirm the genotype of heterozygous clonal lines.

Once *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC were obtained, we first confirmed whether *DDX1* was indeed essential for the survival of pluripotent cells. In agreement with our previous results, approximately a week after removal of DOX, we observed massive cell death in *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC. To record this effect in a quantitative manner, we generated a growth curve that tracked cell numbers over the course of eight days and demonstrated the lethality of the *DDX1* knockout in hiPSC (Figure 3.2.20).



**Figure 3.2.20** Growth curve showing the number of cells over time for *DDX1*<sup>-/-</sup>::TetON-*DDX1* and WT hiPSC. The *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC line and its isogenic WT hiPSC line were both cultured with and without DOX (100 ng/ml). DOX was removed from the *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC 24 hours prior to the start of the experiment. Initially,  $0.1 \times 10^5$  cell per well were seeded per well and the number of cells was counted every 24 hours starting four days after splitting. The error bars represent standard deviation from twelve measurements (four technical replicates from three biological replicates) performed for each sample and time point.

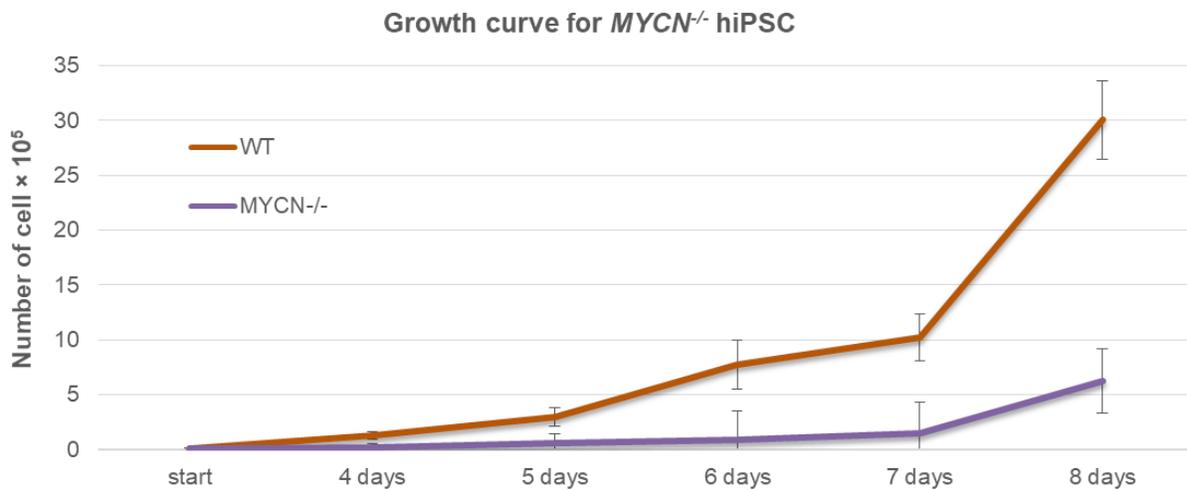
In contrast, the hiPSC with a heterozygous *DDX1* deletion (*DDX1*<sup>+/-</sup> clone #1 and clone #2) were able to survive and only showed a mild decrease in their proliferation rate (Figure 3.2.21).



**Figure 3.2.21** Growth curve showing number of cells over time for *DDX1*<sup>+/-</sup> and WT hiPSC. Two *DDX1*<sup>+/-</sup> hiPSC clonal lines and their isogenic WT hiPSC line were seeded at a density of  $0.1 \times 10^5$  cells per well. The number of cells was counted every 24 hours starting four days after splitting. The error bars represent standard deviation from twelve measurements (four technical replicates from three biological replicates) for each sample and time point. p-values were calculated using Student's t-test (n.s. = not significant, \*\*p-value $\leq$ 0.01, \*\*\*p-value $\leq$ 0.001, \*\*\*\*p-value $\leq$ 0.0001) upon comparison of both *DDX1*<sup>+/-</sup> hiPSC clonal lines taken together against the WT hiPSC.

## Results

Finally, the *MYCN*<sup>-/-</sup> hiPSC, displayed extremely slow proliferation but no major signs of cell death (Figure 3.2.22). This is not surprising, because *MYCN* is considered as a proto-oncogene that promotes proliferation in neuroblastomas and other cancers (Kramer et al., 2016; Liu et al., 2016; Liu et al., 2017b; Liu et al., 2019; Zeid et al., 2018). Therefore, a proliferation defect in the *MYCN* knockout hiPSC was expected. Similar proliferation defects due to decreased *MYCN* levels during neural crest development (e.g. in facial mesenchyme progenitors) could certainly contribute to the emergence of an OFC phenotype. In support of the hypothesis, *MYCN* haploinsufficiency causes Feingold syndrome, which is characterized, among other symptoms, by craniofacial abnormalities (Burnside et al., 2018; Chen et al., 2012; Marcelis et al., 2008; Teszas et al., 2006; van Bokhoven et al., 2005).



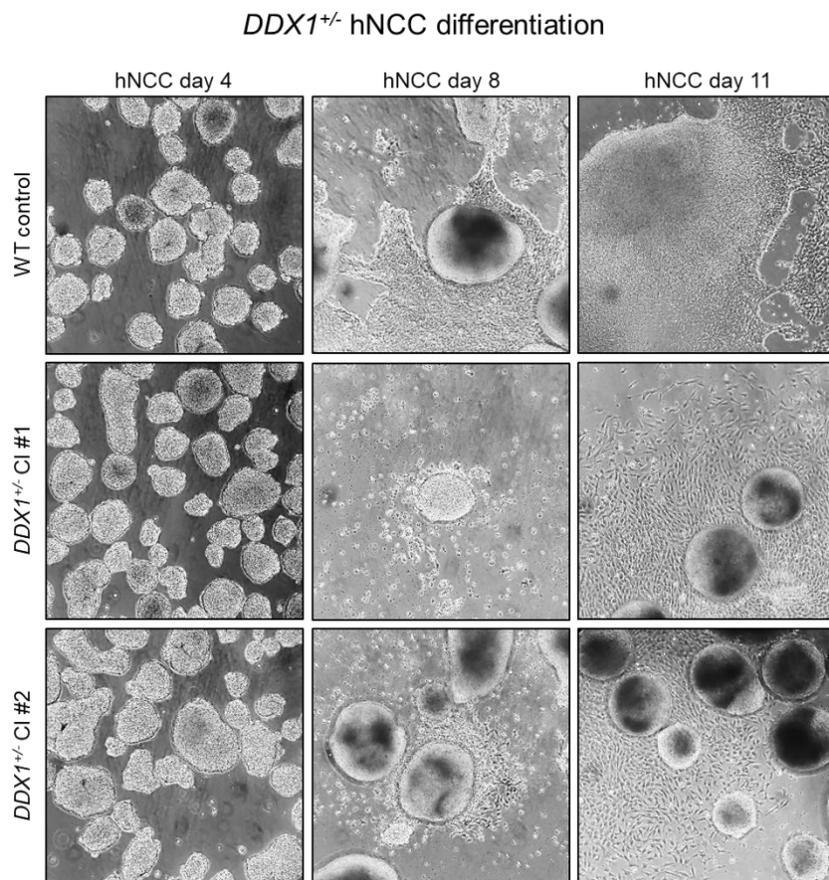
**Figure 3.2.22 Growth curve showing number of cells over time for *MYCN*<sup>-/-</sup> and WT hiPSC.**

The *MYCN*<sup>-/-</sup> hiPSC clonal line and its isogenic WT hiPSC line were seeded at a density of  $0.1 \times 10^5$  cells per well. The number of cells was counted every 24 hours starting four days after splitting. The error bars represent standard deviation from twelve measurements (four technical replicates from three biological replicates) for each sample and time point. Differences in cell numbers between *MYCN*<sup>-/-</sup> and WT hiPSC were highly significant at all times ( $p\text{-value} \leq 2.6 \times 10^{-7}$ , calculated by Student's t-test).

In contrast to the expected relevance of *MYCN* misexpression in the etiology of OFC, much less is known about the role of *DDX1* during facial development. However, the observed phenotype upon *DDX1 in vivo* knockout was similar to the one observed in the *MYCN* knockout. Therefore, we decided to further investigate the molecular basis of the observed phenotype upon loss of *DDX1* function.

First, we assessed the NC differentiation potential of the *DDX1*<sup>+/-</sup> and *DDX1*<sup>-/-</sup> knockout hiPSC lines. The *DDX1*<sup>+/-</sup> hiPSC formed EBs normally though in comparison to WT hiPSC seemed to

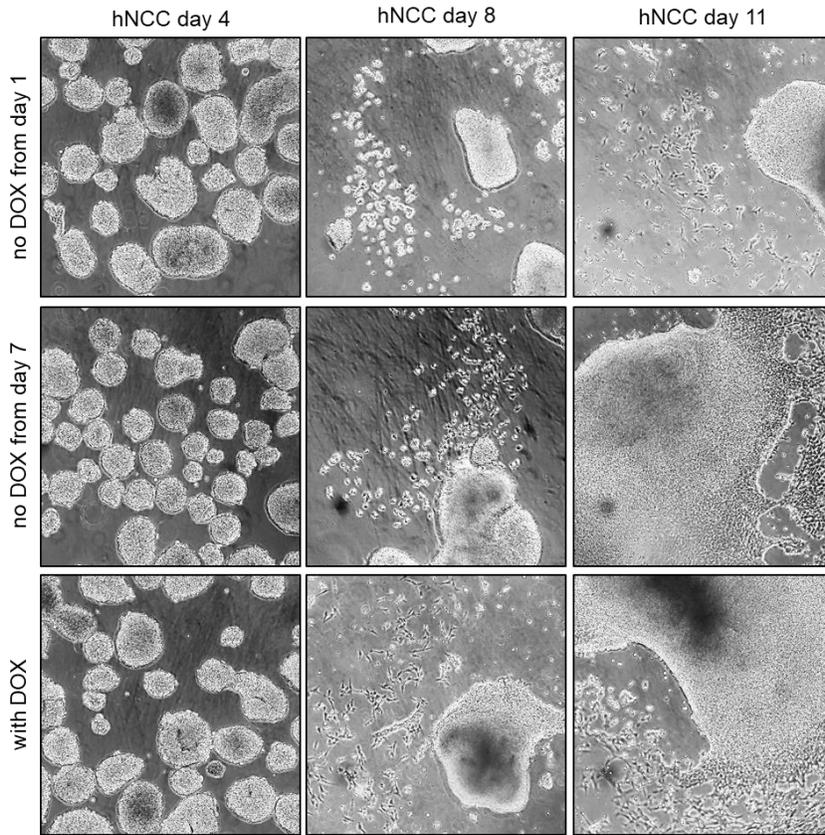
showed reduced numbers of hNCC outgrowths (Figure 3.2.23), although this observation has yet to be quantified.



**Figure 3.2.23 Compromised hNCC differentiation of hiPSC carrying a heterozygous *DDX1* deletion.**

Representative images illustrating the compromised hNCC differentiation capacity of two different clonal hiPSC lines (*DDX1*<sup>+/-</sup> Cl#1 and Cl#2) with a heterozygous deletion *DDX1* in comparison with their parental WT hiPSC line. The *DDX1*<sup>+/-</sup> hiPSC lines formed morphologically normal EBs (Day 4), but showed a delay in EB attachment to the plate surface and in the emergence of hNCC outgrowths (Day 8), which ultimately resulted in reduced hNCC numbers (Day 11).

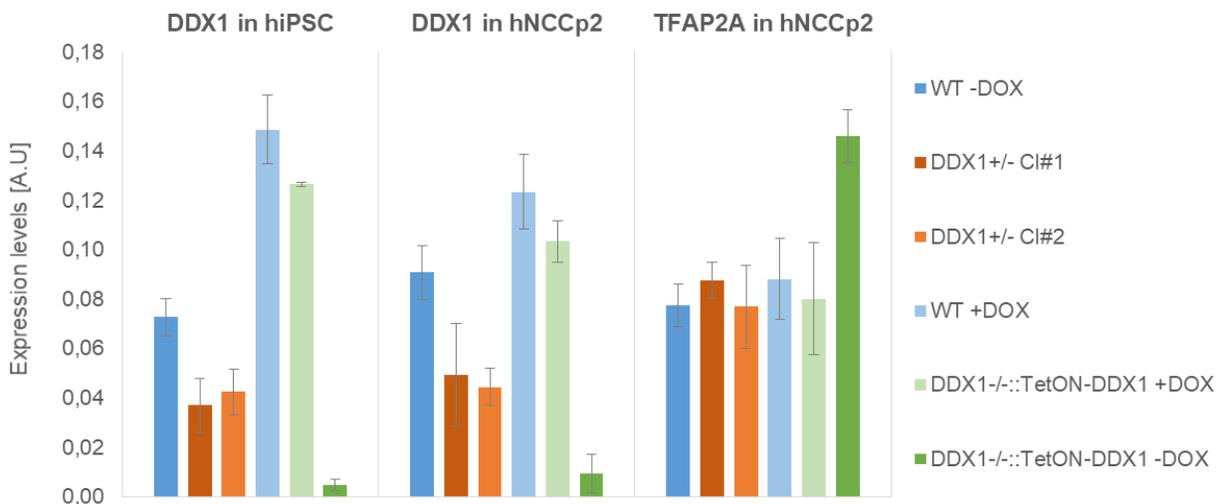
On the other hand, the *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC did not show any obvious differentiation defects when treated with DOX during the entire differentiation process or when DOX was removed on day 7. Though, when DOX was removed at the beginning of the differentiation process (day 1), EBs formed normally but failed to attach and only gave rise to very few hNCC outgrowths (Figure 3.2.24).

*DDX1*<sup>-/-</sup>::TetON-*DDX1* hNCC differentiation

**Figure 3.2.24 hNCC differentiation of *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC with and without DOX.**

Representative images illustrating the compromised NC differentiation capacity of *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC when DOX was removed at the beginning (day 1) of the differentiation process. Under these conditions, *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC formed morphologically normal EBs (day 4) but showed a delay in EB attachment to the plate surface and in the emergence of hNCC outgrowths (day 8), which ultimately resulted in reduced hNCC numbers (day 11). In contrast, *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC differentiated into hNCC without any obvious defects, when treated with DOX until day 7 or throughout the entire differentiation process.

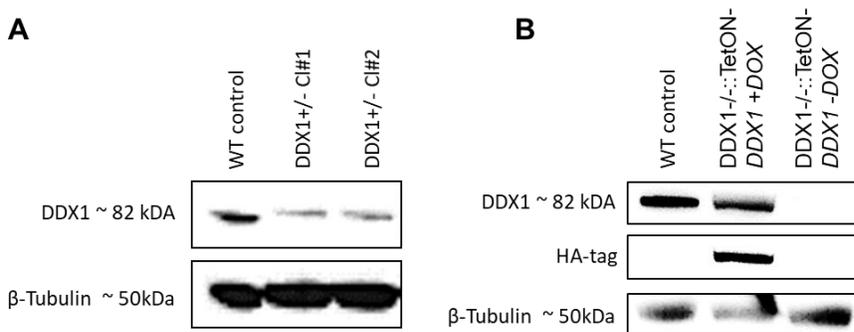
Notably, *DDX1* expression levels were decreased approximately 2-fold in *DDX1*<sup>+/-</sup> cells and extremely diminished in *DDX1*<sup>-/-</sup>::TetON-*DDX1* upon removal of DOX both in hiPSC and in hNCCp2 after NC differentiation had been performed in presence of DOX. Nevertheless both cell lines were able to form hNCC, as measured by levels of *TFAP2A* expression (Figure 3.2.25).



**Figure 3.2.25 DDX1 and TFAP2A expression levels measured by RT-qPCR in cells with DDX1 deletions.**

RNA from two *DDX1*<sup>+/-</sup> clones and their isogenic WT hiPSC line, as well as a *DDX1*<sup>-/-::TetON-DDX1</sup> clonal line was analyzed in hiPSC and hNCCp2. *DDX1*<sup>-/-::TetON-DDX1</sup> hiPSC were differentiated into hNCC in the presence of DOX and then propagated with and without DOX for 4 days. *DDX1* expression was measured in both hiPSC and hNCCp2, *TFAP2A* expression exclusively in hNCCp2. Expression values were normalized to three housekeeping genes (*ACTB*, *EEF2*, *GAPDH*). The error bars represent standard deviation from nine measurements for *DDX1* in hiPSC (technical triplicates in three biological replicates per sample), twelve measurements for *DDX1* in hNCCp2 (technical triplicates in four biological replicates per sample) and nine measurements for *TFAP2A* in hNCCp2 (technical triplicates in three different biological replicates per sample). P-values were calculated using Student's t-test (n.s. = not significant, \*p-value ≤ 0.05, \*\*p-value ≤ 0.01, \*\*\*\*p-value ≤ 0.0001).

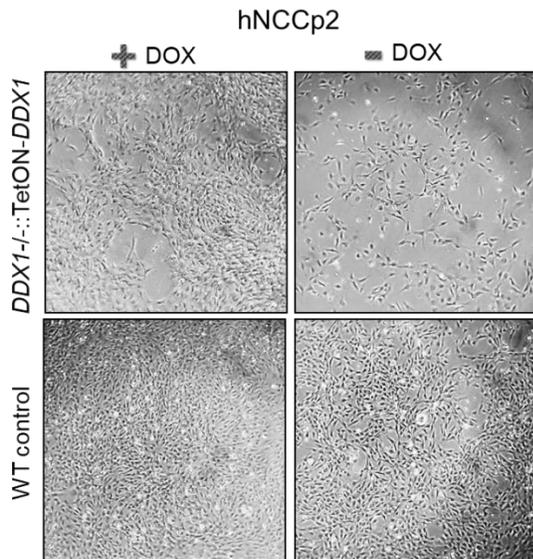
Similarly, in hNCCp2 the DDX1 protein was slightly reduced in the heterozygous *DDX1* deletion clones (*DDX1*<sup>+/-</sup> clone #1 and clone #2) compared to the WT and completely absent in the inducible knockout (*DDX1*<sup>-/-::TetON-DDX1</sup> -DOX) (Figure 3.2.26).



**Figure 3.2.26 Western blot analysis of DDX1 in hNCCp2 derived from *DDX1*<sup>+/-</sup> and *DDX1*<sup>-/-::TetON-DDX1</sup> hiPSC. (A)** Proteins were extracted from hNCCp2 of two heterozygous *DDX1*<sup>+/-</sup> lines and their isogenic WT control. Equal amounts were loaded and immunoblotted against DDX1.  $\beta$ -Tubulin was used as a loading control. **(B)** Proteins were extracted from WT hNCCp2 and hNCCp2 derived from *DDX1*<sup>-/-::TetON-DDX1</sup> hiPSC, which had been differentiated in the presence of DOX and then propagated with DOX or without DOX for 4 days. Equal amounts of each sample were loaded and immunoblotted against DDX1 and the HA-tag added to the exogenous copy of DDX1.  $\beta$ -Tubulin was used as a loading control.

## Results

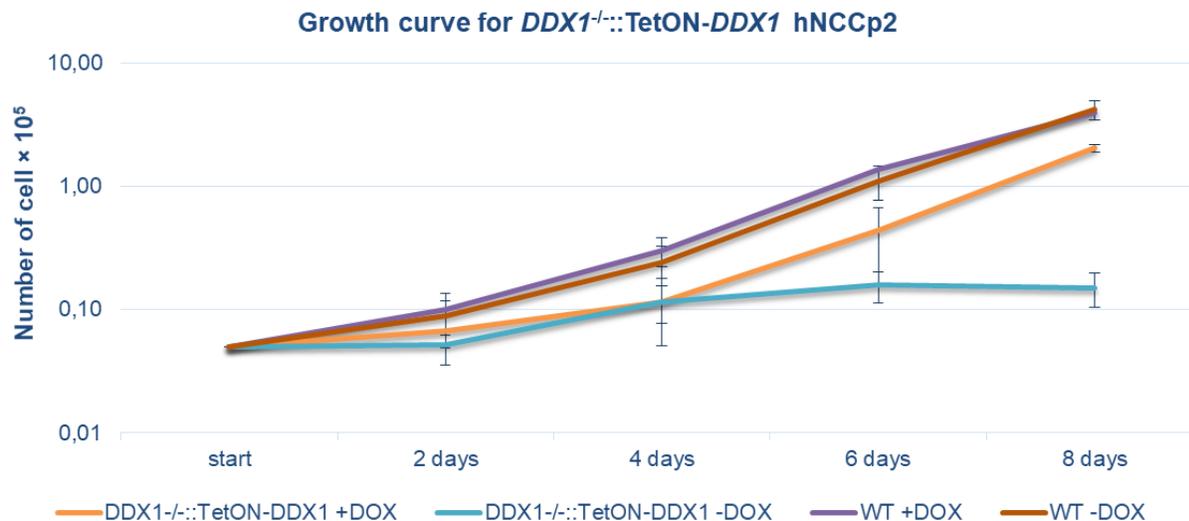
Interestingly, during propagation of *DDX1*<sup>-/-</sup>::TetON-*DDX1* hNCCp2 we noticed a reduced proliferation rate when DOX was removed once the differentiation process had been completed (Figure 3.2.27).



**Figure 3.2.27 Representative photos of *DDX1*<sup>-/-</sup>::TetON-*DDX1* and WT hNCCp2**

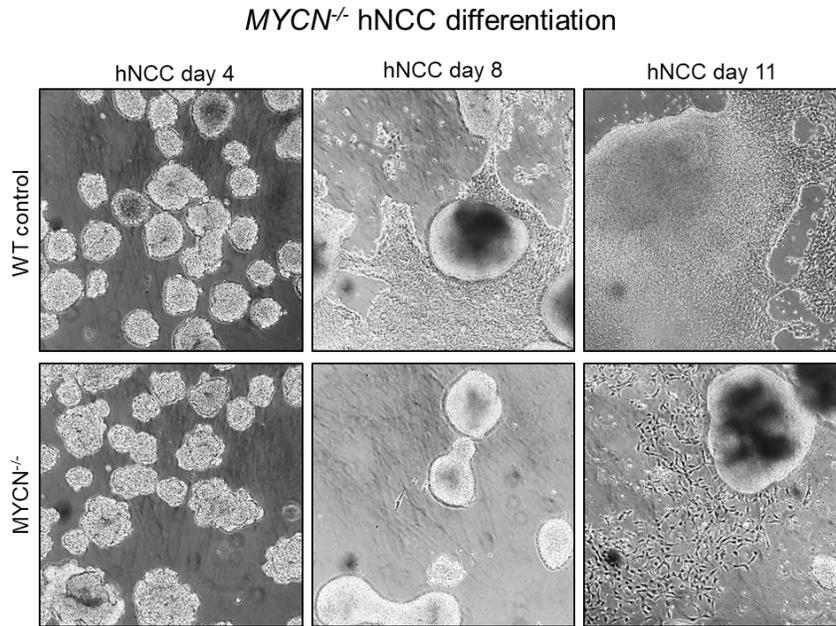
The *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC were differentiated into hNCCd11 in presence of DOX to stimulate exogenous *DDX1* expression during the differentiation process. WT hiPSC were differentiated in absence of DOX. Once hNCC were passaged into hNCCp2 they were seeded at equal density and hNCCp2 were grown for 6 days under conditions with and without DOX.

Thus, we performed proliferation assays in which we counted the number of hNCCp2 every 48 hours at five time points (Figure 3.2.28). Unlike the effects observed in hiPSC, removal of DOX in *DDX1*<sup>-/-</sup>::TetON-*DDX1* hNCCp2 did not result in visible cell death but rather a strong reduction in their growth (Figure 3.2.28), eventually entering a non-proliferative state that can be stable maintained for several weeks (data not shown) and whose nature still needs to be determined (e.g. senescence, differentiation, etc.). Furthermore, the proliferation defect observed in hNCC upon loss of *DDX1*, similar to the one observed for *MYCN*<sup>-/-</sup> hiPSC, is a plausible explanation for the OFC-like phenotype observed in the *in vivo* knockout experiments in chicken embryos. The molecular basis of this proliferation defect in hNCCp2 upon loss of *DDX1* remains to be elucidated and is currently being investigated.



**Figure 3.2.28 Growth curve showing the number of cells over time for *DDX1*<sup>-/-</sup>::TetON-*DDX1* and WT hNCCp2.** *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC were differentiated into hNCC in the presence of DOX. *DDX1*<sup>-/-</sup>::TetON-*DDX1* hNCCp2 and its isogenic WT hNCCp2 were both cultured with and without DOX (100 ng/ml). DOX was removed from the *DDX1*<sup>-/-</sup>::TetON-*DDX1* hiPSC 24 hours prior to the start of the experiment. Initially,  $0.1 \times 10^5$  cell per well were seeded per well and the number of cells was counted every 48 hours at five time points. The error bars represent standard deviation from twelve measurements (four technical replicates from three biological replicates) performed for each sample and time point.

MYCN has previously been reported to be critical for cell growth and cell cycle progression in hNCC (Zhang et al., 2016). Therefore we planned to test the proliferation rate of *MYCN*<sup>-/-</sup> hNCCp2, as well. However, during the NC differentiation process, we observed a strongly reduced differentiation capacity displayed by impaired attachment of floating EBs and much less hNCC outgrowths on day 11 of the *in vitro* differentiation compared to the WT isogenic control (Figure 3.2.29). Consequently, it would be best to assess possible *MYCN*<sup>-/-</sup> hNCC proliferation defects in an inducible loss-of-function system like the one we used for *DDX1*. This way we would be able to differentiate *MYCN*<sup>-/-</sup> hiPSC into hNCC without impaired differentiation and, subsequently, investigate the effects of MYCN loss upon removal of DOX in hNCCp2.

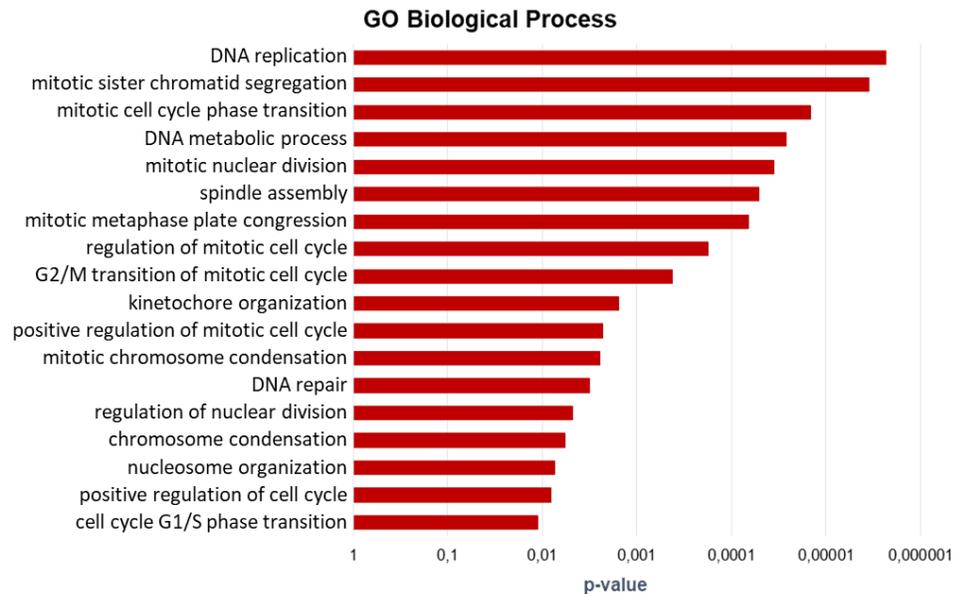


**Figure 3.2.29 Compromised hNCC differentiation of *MYCN*<sup>-/-</sup> hiPSC.**

Representative images illustrating the compromised hNCC differentiation capacity of *MYCN*<sup>-/-</sup> hiPSC in comparison with its parental WT hiPSC line. The *MYCN*<sup>-/-</sup> hiPSC formed morphologically normal EBs (day 4), but showed a delay in EB attachment to the plate surface and in the emergence of hNCC outgrowths (day 8), which ultimately resulted in reduced hNCC numbers (day 11).

### 3.2.7 *DDX1*<sup>+/-</sup> hNCC display a defective transcriptome that suggests a role for *DDX1* in cell cycle control

To gain some general insights into how reduction in *DDX1* could affect the cellular properties of hNCC, we performed RNA-seq experiments in WT and *DDX1*<sup>+/-</sup>hNCCp2 (Figure 3.2.30). 657 genes were differentially expressed (FC=1.25, p≤0.01) between all three WT and all three *DDX1*<sup>+/-</sup> clones (366 up, 291 down). Functional annotation of the 99 most significantly upregulated (p≤0.001) genes in the *DDX1*<sup>+/-</sup> hNCCp2, revealed a striking enrichment of genes involved in cell proliferation/mitosis (Figure 3.2.30) (e.g. cell cycle (CDC20, CDC25B, CDC25C, CCNB2, ESPL1), DNA replication and G1 to S phase transition (MCM2, MCM5, MCM7) and DNA repair (CCNB2, H2AFX, CDC25C)).



**Figure 3.2.30 Functional annotation of genes significantly upregulated in DDX<sup>+/-</sup> hNCCp2 compared to WT hNCCp2.**

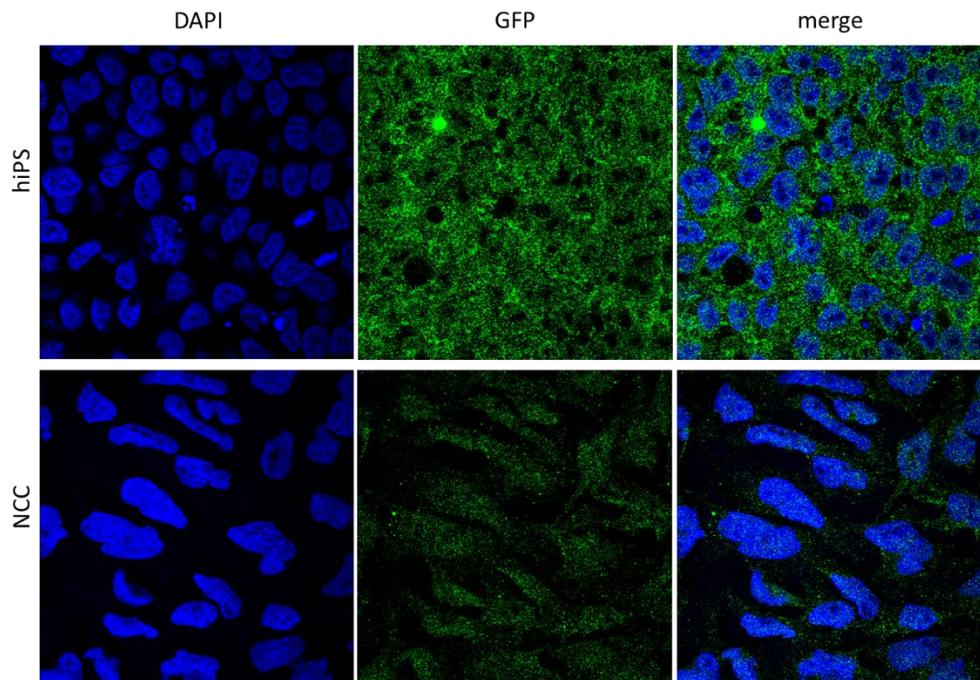
RNA-seq was performed in hNCCp2 derived from three different *DDX1*<sup>+/-</sup> hiPSC clones and from their isogenic WT hiPSC (as biological triplicates). Significantly upregulated genes ( $p < 0.001$ ) with a fold change  $> 1.25$  were functionally annotated to “Gene Ontology Biological Process”. Selected terms among the 30 most significantly overrepresented ones are shown.

Overall, even though these results are only based on the analysis of *DDX1*<sup>+/-</sup> hNCC, the RNA-seq data point towards a role for DDX1 in cell cycle control and, possibly, in DNA repair. These results further support the observed lethality in hiPSC and the strong proliferation defect in hNCCp2 upon removal of DOX in *DDX1*<sup>-/-::Tet-ON-*DDX1*</sup> cells. Yet, additional experiments are currently undergoing in order to confirm these hypotheses as correct.

### 3.2.8 DDX1 binds to promoters and genes bodies of expressed genes suggesting involvement in a transcription coupled repair mechanism

To further understand how DDX1 could be involved in cell cycle control and/or DNA repair, we then examined the subcellular localization of the DDX1 protein by immunoprecipitation in hiPSC and hNCCp2 (Figure 3.2.31). We observed nuclear and cytoplasmic DDX1 expression in hiPSC while in hNCCp2, DDX1 appeared to be preferentially nuclear.

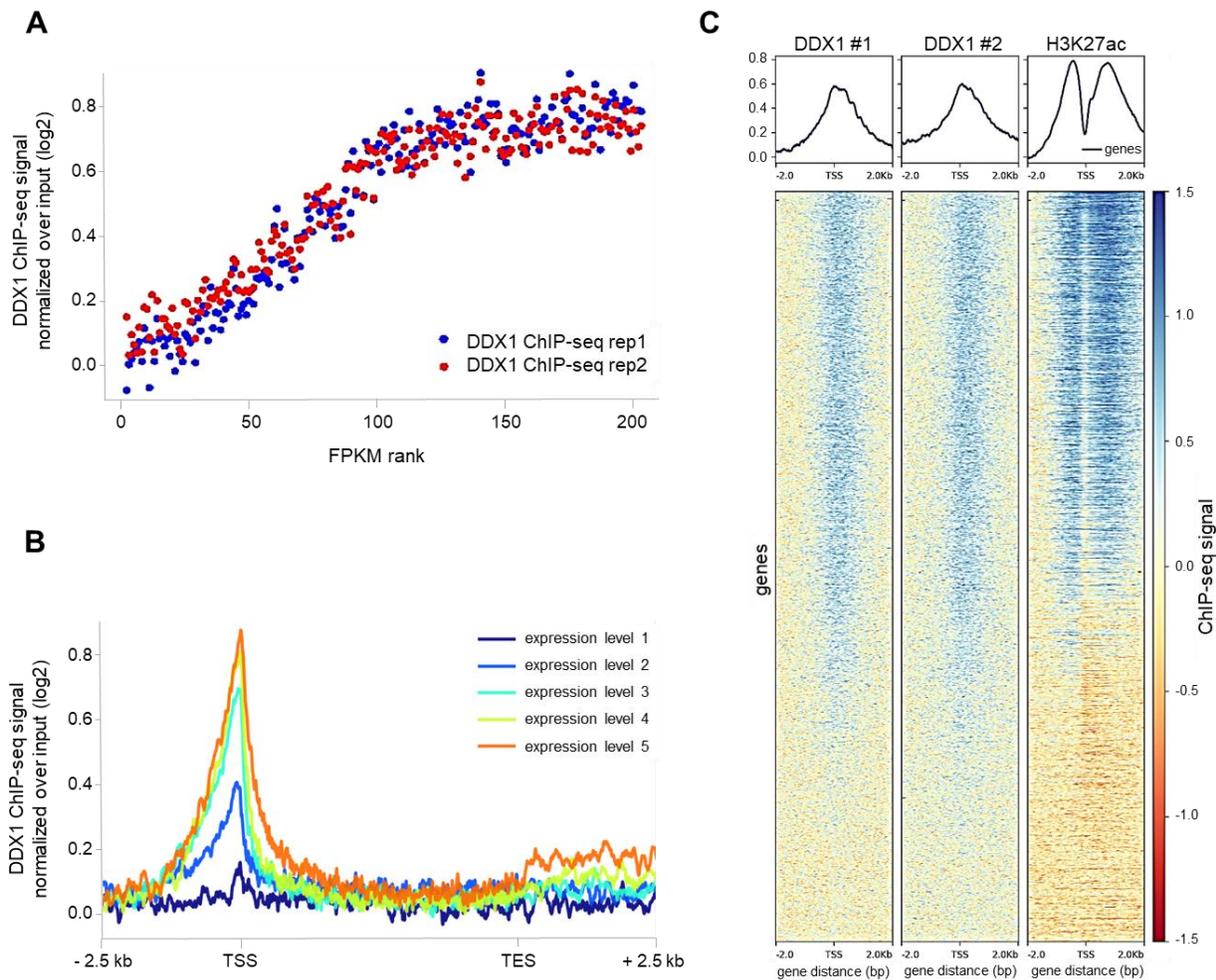
## Results



**Figure 3.2.31 Immunofluorescence for DDX1 in hiPSC and hNCC.**

IF experiments were performed in WT hiPSC and hNCCp2. Nuclei were stained with DAPI (blue). Images were taken at 60x magnitude.

The previous observations suggest that, particularly in hNCC, DDX1 can be associated to chromatin and could control gene expression by binding, directly or indirectly, to DNA. Therefore, we then decided to perform ChIP-seq experiments for DDX1 in WT hNCCp2 (Figure 3.2.32, Figure 3.2.33).

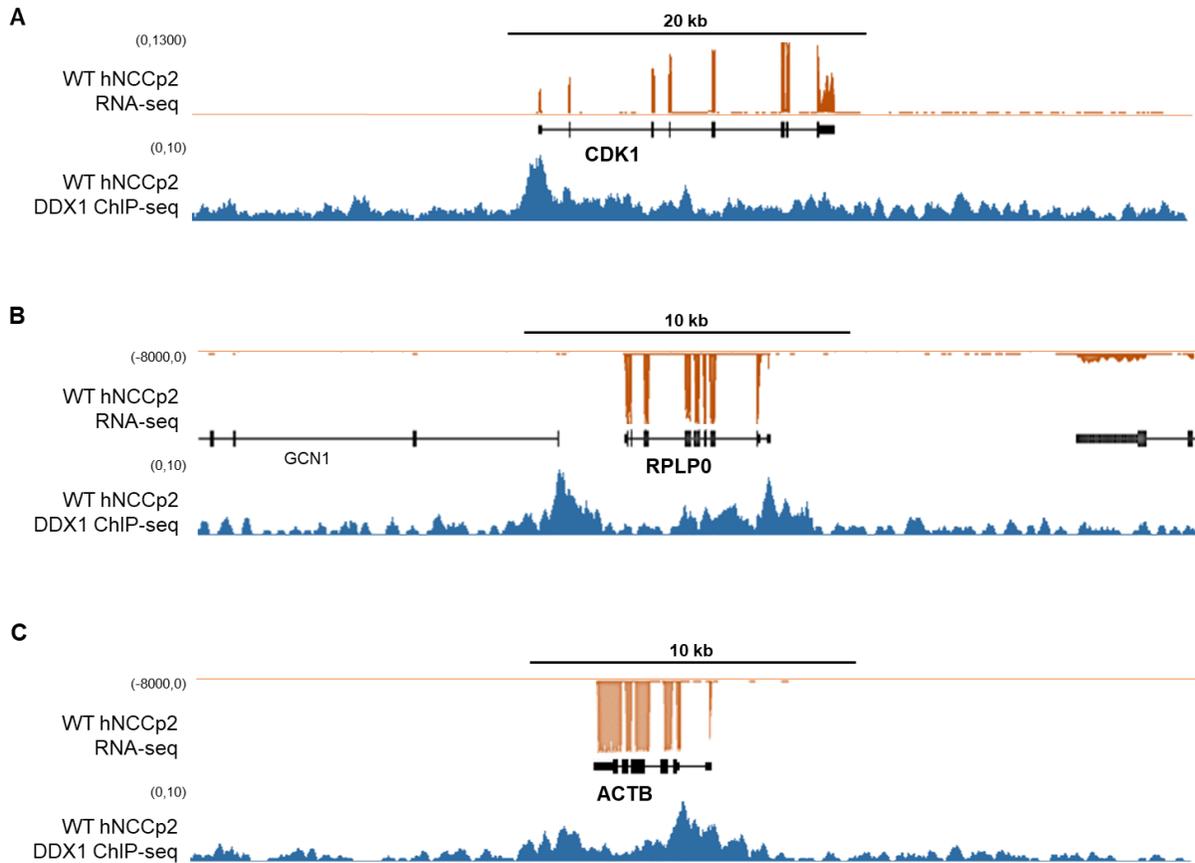


**Figure 3.2.32 Correlation between DDX1 binding to promoters and expression of the corresponding genes.** **(A)** All coding genes were ordered in increasing value of FPKMs in WT hNCCp2 and bundled into groups of 100 genes. Each group was then classified to values from 1 to 200 and ranked in increasing order of FPKMs (x-axis). These values were then plotted against the average DDX1 ChIP-seq signals around the promoter ( $\pm$  500 bp) of the corresponding genes in each group. **(B)** Average DDX1 ChIP-seq profiles across the gene bodies of genes (scaled to a length of 5 kb) with different expression levels. All genes expressed in WT hNCCp2, were divided in 5 groups according to their expression levels: expression level 1 refers to the top 20% of genes with the highest expression level in WT hNCCp2, level 5 to the bottom 20% of genes with the lowest expression level. Genes with expression levels in between are grouped accordingly in the remaining quantiles. DDX1 ChIP-seq signal normalized over input was then plotted along the assigned gene bodies extended 2.5 kb up- and downstream of the transcription start site (TSS) and the transcription end site (TES). **(C)** Coordinates of all protein coding genes were sorted according to their expression in decreasing order of FPKMs in WT hNCCp2. ChIP-seq data for DDX1 (biological duplicates) and H3K27ac was generated from WT hNCCp2. Individual ChIP-seq signals for DDX1 and H3K27ac are visualized as heat maps and average ChIP signals as plots around the TSS  $\pm$  2 kb.

Analysis of the average DDX1 ChIP-seq signal over the gene body of all human coding genes revealed that DDX1 binding shows a strong positive correlation with high gene expression levels, especially upstream of the TSS and downstream of the TES (Figure 3.2.32, Figure 3.2.33). This binding pattern of DDX1 strongly resembles that of RNA Pol II suggesting that DDX1 could somehow be coupled to the RNA Pol II machinery (Figure 3.2.32), perhaps in transcription

## Results

coupled repair. However, whether DDX1 plays an important role in this or any other RNA Pol II-associated process remains to be elucidated.



**Figure 3.2.33 DDX1 ChIP-seq profiles and RNA-seq read counts at three example loci.**

ChIP-seq profiles for DDX1 in WT hNCCp2 are shown together with expression levels from RNA-seq in WT hNCCp2 at three representative genes: (A) CDK1, (B) RPLP0 and (C) ACTB.

## 4 DISCUSSION AND FUTURE DIRECTIONS

### 4.1 Insights into the pathomechanism of a syndromic neurocristopathy with implications for non-syndromic forms

Here we provide an extensive characterization of the mechanism whereby a *de novo* heterozygous inversion causes BOFS, a syndromic form of OFC, without directly altering the *TFAP2A* coding sequence. Using a combination of patient-derived hiPSC, *in vitro* differentiation systems and genomic and genetic engineering approaches, we show that the inversion causes a loss of physical interactions between the inverted *TFAP2A* allele and its cognate enhancers, ultimately leading to monoallelic and haploinsufficient *TFAP2A* expression in hNCC.

All previously known cases of BOFS are caused by heterozygous mutations or deletions within the coding sequence of *TFAP2A* that result in antimorphic, hypomorphic or null alleles, overall indicating the *TFAP2A* is a dosage-sensitive gene during human embryogenesis (Aliferis et al., 2011; Galliani et al., 2012; Gestri et al., 2009; LeBlanc et al., 2013; Li et al., 2013; Lin et al., 1993; Milunsky et al., 2008; Milunsky et al., 2011; Murray et al., 2013).

Here, we describe a novel etiological mechanism for BOFS in which *TFAP2A* haploinsufficiency is caused by a heterozygous inversion that disrupts the *TFAP2A*-TAD, leading to a loss of enhancer-gene interaction and monoallelic *TFAP2A* expression in hNCC. Of course, it is possible that the disruption of the *TFAP2A* regulatory domain could have similar effects in other human cell types (e.g. epidermis) that can also contribute to the emergence of the BOFS phenotype during embryogenesis (Park and Kim, 2009). However, our approach, using an *in vitro* hNCC differentiation system, conclusively shows, that although the moderate yet relevant expression changes caused by the heterozygous inversion do not affect hNCC specification, they do disrupt an important hNCC property, i.e. migration (Okuno et al., 2017). Consequently, these migratory defects could disrupt the formation of craniofacial structures and ultimately cause at least some of the phenotypes observed in BOFS patients.

Notably, single cell RNA-seq revealed that the monoallelic expression of *TFAP2A* did not lead to a 2-fold reduction of *TFAP2A* transcript in every single cell but rather lead to a more heterogeneous expression of *TFAP2A*. Namely, a large number of patient hNCC barely expressed *TFAP2A* while others displayed normal *TFAP2A* transcript levels. This observation could be a consequence of less frequent transcriptional bursts (Fukaya et al., 2016), due to the

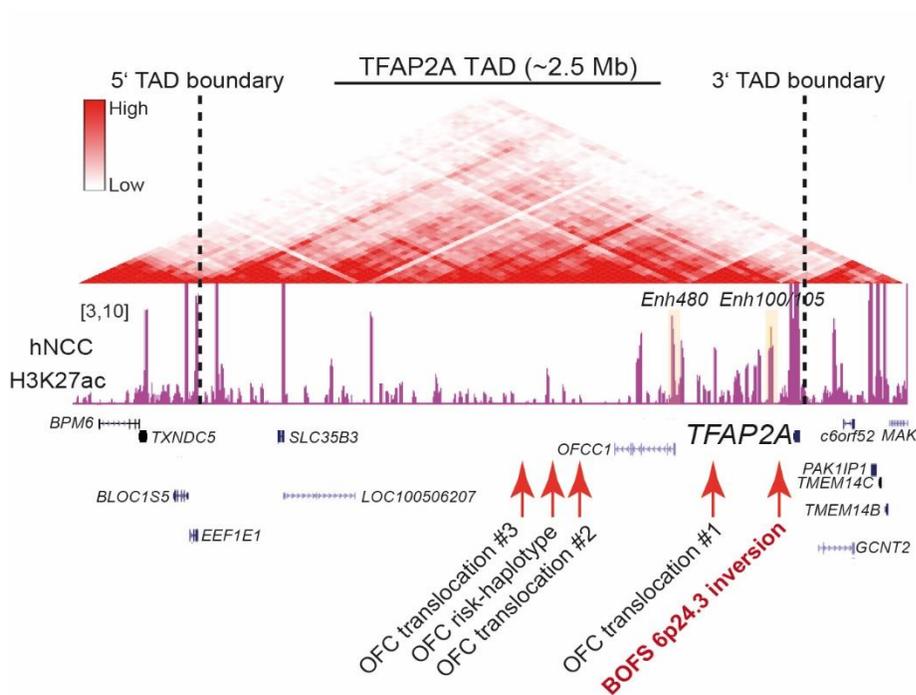
fact that only one *TFAP2A* allele is functional, ultimately resulting in more cells with lower levels of *TFAP2A* expression. Despite the possibility that some of the cells expressing reduced levels of *TFAP2A* could appear due to technical reasons (i.e. scRNA-seq dropouts), we made similar observations by immunofluorescence staining of *TFAP2A* protein (Figure 3.1.35). Therefore, our data suggest that haploinsufficiency for *TFAP2A* and perhaps for other developmental genes may arise due to not only lower gene expression levels but also higher transcriptional heterogeneity (Antolovic et al., 2017; Mohammed et al., 2017). The partial and mild phenotype of our BOFS patient can be explained by the overall moderate transcriptional changes observed in the patient-specific hiPSC derived hNCC, which is also in agreement with mild phenotypes observed in other BOFS patients with *TFAP2A* null alleles (Milunsky et al., 2008; Milunsky et al., 2011).

In contrast, BOFS patients with mutations within *TFAP2A* that are predicted to cause antimorphic alleles usually show more severe phenotypes within the BOFS spectrum (Li et al., 2013; Milunsky et al., 2011). Accordingly, we were also able to show that more extreme expressional changes (here the consequence of a homozygous deletion of an important regulatory region of *TFAP2A* leading to stronger reduction of *TFAP2A* expression) can already affect hNCC specification and therefore could possibly cause much more severe craniofacial phenotypes (Brewer et al., 2004; Schorle et al., 1996; Tchieu et al., 2017; Zhang et al., 1996). Overall, these results confirm previous models suggested to explain the variable phenotypes observed in BOFS patients, whereby the varying activity of *TFAP2A* determines expressivity (Li et al., 2013).

In a study by Feng et al. (2008) a conserved intronic enhancer of *TFAP2A* was identified to regulate *TFAP2A* expression in the face and limb mesenchyme. However, the loss of this regulatory element had a greater effect on *TFAP2A* gene expression in the limb than in the face, indicating that there may be additional enhancers that specifically regulate *TFAP2A* expression in the face. This is in accordance with our results, which suggest that the distal enhancers located 5' of the inversion breakpoint of our BOFS patient are specific to the face and the cranial NCC.

Our extensive characterization of the *TFAP2A*-TAD also provides major insights into the molecular mechanism of other neurocristopathies, including non-syndromic OFC. For example, three OFC patients with translocation breakpoints in the 6p24 region, where the *TFAP2A*-TAD is located, have previously been reported (Figure 4.1.1) (Davies et al., 2004). The authors hypothesize that these translocation breakpoints could affect the expression and/or function of *OFCC1*, a poorly characterized gene located ca 350 kb downstream of *TFAP2A* (Davies et al., 2004). However, according to our RNA-seq (Figure 3.1.45) this gene is barely expressed in hNCC and *Ofcc1* null mice do not show any craniofacial phenotype (Ohnishi et al., 2011). All these

findings argue against an involvement of *OFCC1* in the etiology of OFC. Instead, we hypothesize that these three translocation breakpoints, all located 5' of the 6p24 inversion breakpoint of our BOFS patient, also disrupt the *TFAP2A*-TAD and therefore lead to misregulation of *TFAP2A* expression in hNCC. However, in these translocations, a larger fraction of the *TFAP2A* regulatory domain is spared and left in proximity of the gene, possibly leading to less profound changes in *TFAP2A* expression. This could explain why an isolated OFC is observed in these patients, which can be considered as a milder phenotype compared to the more severe abnormalities associated with BOFS. Additionally, a recent GWAS identified a susceptibility locus for non-syndromic OFC also located within the *TFAP2A*-TAD, 5' of the 6p24 inversion breakpoint (Figure 4.1.1) (Yu et al., 2017). In this case, SNPs within this OFC risk-haplotype might disrupt one or more hNCC-active enhancers and thereby affect *TFAP2A* expression.



**Figure 4.1.1 Schematic diagram of the *TFAP2A*-TAD.**

Red arrows indicate locations of translocation breakpoints identified in three different OFC patients (Davies et al., 2004) and a risk-haplotype associated with OFC according to GWAS (Yu et al., 2017).

## 4.2 Research models to investigate human neurocristopathies

There are a large number of human congenital disorders, including BOFS, which are caused by heterozygous mutations or structural variants (SVs) within developmental genes (Hamosh et al., 2002). Some cases of gain-of-function SVs have been recapitulated in mouse embryos by using CRISPR/Cas9 genetic engineering to then uncover their mechanism of action and phenotypic consequences *in vivo* (Franke et al., 2016; Kraft et al., 2015; Lupianez et al., 2015). However, this approach might not be generally applicable to investigate SVs that, due to their large size (e.g. the 89 Mb inversion in our BOFS patient) or type of genetic rearrangement (e.g. translocation), might not be efficiently engineered with the currently available CRISPR/Cas9-based tools. Similarly, animal models might not be always suitable to investigate the long-range pathologically effects of balanced SV (i.e. inversions and translocations) that can theoretically result in both gain and loss of function events by a TAD shuffling mechanism (Spielmann et al., 2018). In the case of loss-of-function pathomechanisms, these might not be always faithfully recapitulated in mice due to the frequent differences in gene dosage sensitivity between mice and humans. While humans are often haploinsufficient for a large number of developmental genes, heterozygous mice for those same genes are frequently unaffected and only homozygous null animals show disease-like phenotypes, typically with increased severity (Andersen et al., 2014; Bartha et al., 2018; Bedell et al., 1997; Kleinjan and van Heyningen, 2005). For example, *Tfap2a* null mice display a severe BOFS-like phenotype while *Tfap2a*<sup>+/-</sup> animals appear morphologically normal and, thus, are not haploinsufficient (Brewer et al., 2004; Schorle et al., 1996; Zhang et al., 1996). How these differences in gene dosage sensitivity arise, is currently unknown. However, genetic background, gene-environmental interactions and distinct regulatory networks are all likely to contribute.

The differences between mice and humans with regard to *TFAP2A* haploinsufficiency implies the need to optimally study BOFS and similar congenital disorders in the relevant embryonic tissues/cell types. However, the cell types involved in congenital disorders, specifically the NC in BOFS, are largely intractable and extremely difficult to directly isolate from the embryo because of its migratory and transient nature, especially in humans. As a result, most balanced SVs causing congenital disorders have only been investigated on the basis of correlative observations while the exact long-range regulatory mechanism remains elusive and is mostly speculative (Birnbaum et al., 2012b; Kleinjan and van Heyningen, 2005; Mehrjouy et al., 2018; Redin et al., 2017; Zepeda-Mendoza et al., 2018). However, the establishment of hiPSC technology and the improvement of *in vitro* differentiation protocols provide new possibilities to model and explore

human diseases (Adamo et al., 2015; Bajpai et al., 2010; Okuno et al., 2017; Shi et al., 2017; Takahashi et al., 2007; Tchieu et al., 2017). Therefore, our methodology of combining patient-derived hiPSC carrying a large balanced structural variant, such as the one described in Laugsch et al. (2019), with appropriate *in vitro* differentiations systems is an extremely useful approach to systematically evaluate the molecular consequences of certain types of SVs.

Moreover, while quite a few studies have investigated the role of *TFAP2A* in NC (Aliferis et al., 2011; Brewer et al., 2004; Galliani et al., 2012; Gammill and Bronner-Fraser, 2003; Gestri et al., 2009; LeBlanc et al., 2013; Li et al., 2013; Lin and Milunsky, 2011; Milunsky et al., 2008, 2011; Murray et al., 2013; Sauka-Spengler and Bronner-Fraser, 2008), these previous studies were typically performed using heterologous systems that do not recapitulate NCC development (Li et al., 2013). We propose that our hNCC differentiation system could be used, for example, to more thoroughly assess the model of variable expressivity observed in BOFS patients and investigate how different antimorphic, hypomorphic or null alleles of *TFAP2A* can affect different steps of NCC development, such as specification, EMT, migration, proliferation and differentiation.

Finally, although the combination of patient-derived hiPSC with NCC differentiation is not new (Barrell et al., 2019; Okuno et al., 2017) our work illustrates how combining this system with CRISPR/Cas9 engineering and genomic sequencing technologies can be a powerful approach to investigate the molecular basis of human congenital diseases that are caused by the structural or genetic disruption of regulatory landscapes. These approaches might be particularly useful in cases where animal models do not fully recapitulate the human phenotype and/or when the relevant cell type (i.e. NCC) is not easily accessible.

Another useful application of the *in vitro* hNCC system described here is the approach of finding new candidate genes implicated in complex disorders according to GWAS. Genes and diseases can be linked by investigating the effects of a disease-associated SNP through their locations within enhancers that are specifically active in disease-relevant tissue (Chen et al., 2014; Huang and Ovcharenko, 2015; Li et al., 2017; Nikolic et al., 2017; Schwessinger et al., 2017). One selection criteria for meaningful enhancers is their evolutionary conservation both on the sequence and functional levels. However, relevant enhancers are not always evolutionarily conserved (Blow et al., 2010). In addition, disease-associated SNPs or any potentially causative SNPs are, to the best of our knowledge, not conserved. Therefore, when aiming at discovering disease causative SNPs, especially in combination with functional studies in disease relevant cell types that are not easily accessible *in vivo* (i.e. NCC), the combination of genomic engineering by CRISPR/Cas9 with an *in vitro* NCC differentiation system may prove very useful.

Notably, a recent study identified a new candidate gene for OFC, *MRPL53*, a subunit of mitochondrial ribosomes that interacts with MYC. This gene was identified by studying regulatory variants in mesenchymal cells derived from lip muscle tissue discarded during routine corrective surgeries (Masotti et al., 2018). The lip muscle indeed represents a tissue that is affected by the formation of OFC and might therefore appear to be reasonable candidate material for studies of regulatory variants in a disease relevant cell type. Still, material from myectomies of affected patients consists of those cells that made it to the final differentiation stage despite aberrations during prenatal development. We understand that studies investigating the pathomechanism leading to a congenital disorder like OFC must therefore be performed during the developmental process that is disturbed and ultimately results in the malformation, e.g. in precursor cells of facial muscles, i.e. NCC.

Finally, an alternative *in vivo* model that can be used to replace the mouse model, particularly in NC studies, is the chicken embryo. Chicken embryos have historically been used to study the NC because compared to mice they are quite large at the early developmental stages when the NC forms. Furthermore, avian models are very well accessible and therefore easier to manipulate than mice because they develop *in ovo* (Abramyan and Richman, 2018; Le Douarin, 2004). These properties make it possible to generate transgenic animals in a very short time. And although a limitation of this model is the chimeric nature of such transgenes, targeted/conditional knockouts can be produced efficiently. While the application of the CRISPR/Cas9 technology *in vivo* in general and in chicken embryos in particular is not completely novel (Dickinson et al., 2013; Gagnon et al., 2014; Gandhi et al., 2017; Ren et al., 2014; Stolfi et al., 2014), we were not only able to utilize this genomic engineering tool to efficiently knockout several genes in the neural tube/crest but could also observe facial phenotypes that closely resembled OFC (Figure 3.2.12). Thus, the chicken embryo represents a suitable *in vivo* model to study NC related phenotypes.

### 4.3 Enhancer-gene specificity requires more than physical proximity

Structural variants (SVs) such as insertions, deletions, inversions or duplications are commonly found in human genomes and can cause human disease. Recently, the disruption of TAD boundaries through SVs emerged as a novel etiological mechanism for cancer and congenital disorders (Franke et al., 2016; Hnisz et al., 2016; Lupianez et al., 2015; Peifer et al., 2015; Vicente-Garcia et al., 2017).

Balanced SVs, such as translocations or inversions, do not change the total amount of DNA and the respective breakpoints are often located in gene free regions. Still, their occurrence can lead to cellular abnormalities by shuffling enhancers and genes in newly formed TADs (Franke et al., 2016; Kraft et al., 2019; Lupianez et al., 2015; Peifer et al., 2015; Vicente-Garcia et al., 2017). Disruption of TAD boundaries, which coincide with CTCF binding sites that, together with other factors, function as insulators between adjacent regulatory domains seems to be the crucial mechanism causing pathological gene expression changes in cancer and congenital disorders (Hnisz et al., 2016; Lupianez et al., 2015). The consequences of TAD disruption are often pathological gains of gene expression due to the physical communication of enhancers with inactive genes that become ectopically expressed (enhancer adoption or enhancer hijacking) (Franke et al., 2016; Lupianez et al., 2015; Peifer et al., 2015; Vicente-Garcia et al., 2017).

Alternatively, there are other structural variants that are believed to cause human disease through positional effect variegation (PEV) and/or loss of endogenous enhancer-gene interaction (e.g. enhancer disconnection) that can lead to gene silencing (Kleinjan et al., 2001; Kleinjan and van Heyningen, 2005; Lettice et al., 2003; Roessler et al., 1997). Still, in most of these cases the exact pathomechanism has not been fully elucidated, mainly because of the difficulties in accessing the relevant human cell type/tissue (Kleinjan and van Heyningen, 2005). Our approach of combining the hNCC *in vitro* differentiation system with genomic sequencing tools, however, made it possible to exclude PEV as a potential etiological mechanism and determine a loss of enhancer-promoter contacts as the cause for the TFAP2A haploinsufficiency.

Interestingly, although the inversion of the BOFS patient described in Laugsch et al. (2019) moves novel genes into a shuffled TAD together with the *TFAP2A* hNCC enhancers, this did not result in enhancer adoption, in contrast to what has been described for other SVs implicated in human diseases (Franke et al., 2016; Lupianez et al., 2015; Peifer et al., 2015; Vicente-Garcia et al., 2017). Instead, the inversion disconnects one *TFAP2A* allele from its cognate enhancers leading to monoallelic and haploinsufficient *TFAP2A* expression in patient hNCC. This observation shows that simply placing enhancers and genes within the same TAD is not always sufficient to lead to ectopic gains in gene expression. Instead, it suggests that while in cases of an enhancer adoption mechanism the properties of the respective enhancer fits those of the gene promoter (Franke et al., 2016; Lupianez et al., 2015; Peifer et al., 2015; Vicente-Garcia et al., 2017), in other cases additional genetic and/or epigenetic mechanisms might be necessary to mediate productive enhancer-promoter communication (Arnold et al., 2017; van Arensbergen et al., 2014). Enhancer-promoter interactions are known to be mediated by Cohesin-dependent loops that are anchored by a combination of transcription factors and co-factor binding both to the enhancer and the

promoter (e.g. YY1, Polycomb, Mediator) as well as CTCF-CTCF interactions that stabilize a subset of enhancer-gene loops (Bonev et al., 2017; Cruz-Molina et al., 2017; Ji et al., 2016; Kagey et al., 2010; Ren et al., 2017; Weintraub et al., 2017). One possibility is that enhancer-promoter specificity requires that both elements bind the same TFs (Bonev et al., 2017; Thormann et al., 2018). In addition, core promoters might differ in their enhancer responsiveness depending on the type of core promoter elements they present as well as the co-activators that are bound by them (Arnold et al., 2017; Haberle et al., 2019). CTCF and Cohesin, on the other hand, are essential for the establishment of the loops involved in TAD boundary formation, but their absence does not entail major transcriptional changes (Rao et al., 2017; Schwarzer et al., 2017), and thus, they are unlikely to define the compatibility between genes and enhancers (Hanssen et al., 2017; Kragesteen et al., 2018; Nora et al., 2012). Thus, the mechanism/s that determine whether a SV leads to enhancer adoption or not are still not fully understood but we hypothesize that they involve genetic elements yet to be identified, which topologically dictate enhancer-promoter compatibility.

#### 4.4 Beyond GWAS: possibilities and limitations

Besides the extensive characterization of the BOFS patient with the heterozygous inversion disrupting the *TFAP2A*-TAD, this work also provides new insights into non-syndromic OFC by utilizing disease-associated SNPs within enhancers to find new candidate genes involved in this common congenital abnormality. The disease-associated SNPs used as a basis for our approach were acquired from publically available GWAS. GWAS have been instrumental in the discovery of new possible causes for diseases with complex etiology such as diabetes, schizophrenia, and also OFC. This way, thousands of genetic variants contributing to the risk of a disease phenotype have been identified, including a number of SNPs highly associated with orofacial clefts in particular (Beaty et al., 2010; Beaty et al., 2013; Birnbaum et al., 2009; Carlson et al., 2017; Grant et al., 2009; Haaland et al., 2018; Leslie et al., 2017; Leslie et al., 2016; Ludwig et al., 2017; Ludwig et al., 2012; Mangold et al., 2010; Sun et al., 2015; Yu et al., 2017).

Nonetheless, in most cases there are no direct functional evidences to support whether the candidate SNPs or haplotypes are causally involved in the investigated disease. Moreover, our capacity to identify novel disease-associated variants largely exceeds the rate at which the underlying mechanism can be experimentally determined. Consequently, there is a huge number

of genetic variants that have been associated with a trait but not analyzed yet in terms of their potential etiological mechanism. Thus, data coming from GWAS represents a great and still largely unexplored resource for mechanistic studies (Vockley et al., 2017). Since disease-associated variants mostly occur within non-coding sequences, narrowing down the regulatory regions that might be disrupted by disease-associated variants and that might have downstream consequences on gene expression can be a time consuming task when performed experimentally (e.g. deletions by CRISPR/Cas9 genetic engineering). Recently, genome-wide CRISPR/Cas9 screens are being developed, which might streamline this process (Qi et al., 2017). Generally, increasing efforts are being made to transfer genetic findings into functionally important discoveries, emerging as the new field of functional genomics (Thieme and Ludwig, 2017).

Several methods and computational tools have been developed to identify the disease-causal variants occurring in regulatory sequences. (Ameur et al., 2009; Bochdanovits et al., 2014; Boyle et al., 2012; Chen et al., 2014; Coetzee et al., 2015; Hormozdiari et al., 2015; Hormozdiari et al., 2014; Huang and Ovcharenko, 2015; Khurana et al., 2013; Kichaev et al., 2014; Kilpinen et al., 2013; Lee et al., 2015; Li et al., 2017; Schaub et al., 2012; Schwessinger et al., 2017; Ward and Kellis, 2012). In these approaches, TF binding sites within the candidate regulatory sequences are computationally predicted since genetic variants that alter such binding sites are more likely to be causally involved in the investigated phenotype. Notably, in a number of recent studies, extensive functional analyses were used to successfully pinpoint non-coding variants causally involved in several human complex disorders (Fogarty et al., 2014; Fortini et al., 2014; Guenther et al., 2014; He et al., 2015; Huang et al., 2014; Leslie et al., 2015; Pan et al., 2017; Pasquali et al., 2014; Spieler et al., 2014; Visser et al., 2015). However, the identification of these causative variants is still a very laborious and non-trivial task. For OFC specifically, so far only two functional regulatory variants influencing expression of genes implicated in OFC could be identified with certainty (Leslie et al., 2015; Uslu et al., 2014).

In the case of *Enh2p24.2*, initial allele specific analysis of common variants within the enhancer region (rs11893260, rs4832647, rs6735186, rs10182776) could not identify any of these common variants as causative (data not shown). This observation is in line with the hypothesis that the genetic variants causally implicated in complex diseases are frequently rare ones that co-segregate with the risk-haplotypes but that cannot be easily identified due to their low frequency. GWAS are designed on the common disease-common variant hypothesis (CD-CV). However, the alleles identified through GWAS are typically not causative but rather in linkage disequilibrium (LD) with the true causal variants and may only account for a fraction of the heritability of complex traits. Hence, the focus is being shifted to common disease-rare variant (CD-RV) hypothesis,

which assumes that rare variants have larger effects on the phenotype (Marian, 2011). This hypothesis is supported by evidences suggesting that rare mutations with severe effects are responsible for a substantial portion of complex human diseases. This CD-RV variant hypothesis could also explain the phenotypic heterogeneity of common diseases, which might be caused by a large collection of individually rare or even private mutations with different functional consequences (McClellan & King 2010). Alternatively, “outside variants” in weak LD with GWAS risk SNPs have been proposed to interact with risk SNPs, consequently influencing target gene transcription and ultimately clinical risk (Corradin et al., 2016).

Yet, when using GWAS, which are usually based on populations of a certain ethnicity, one has to take into consideration that there are ethnic differences in the prevalence of OFC (Dixon et al., 2011; Rahimov et al., 2012). Therefore, we hypothesize that certain risk loci identified on the basis of genetic variants that are prevalent in a specific population are more likely to harbor rare causative SNPs in individuals of the same ethnicity. The candidate locus we investigated (2p24) was reported as associated with OFC with the highest significance in a population with Chinese origin (Chen et al., 2018; Yu et al., 2017) while the significance in Brazilian and European populations was considerably lower (Leslie et al., 2016; Ludwig et al., 2012) (see Table 3.1). Consequently, a search for rare variants within this region in a group of OFC patients with European ancestry might be less successful than in a group of Chinese OFC patients.

Overall, once a risk locus has been identified by GWAS, it seems more likely to find evidences supporting the involvement of novel genes in the investigated disease/trait rather than the causing variant/s themselves. Therefore, in this work we used SNPs associated with OFC as a starting point for the investigation of enhancers and their target genes, which we hypothesize, might not have been previously implicated in OFC and/or craniofacial development. Although here we focus on OFC, our strategy of combining GWAS data with enhancer and 3D chromatin maps should be applicable to elucidate the etiology of other complex diseases if an appropriate *in vitro* differentiation system is available.

## 4.5 Taking environmental factors into account

The etiology of complex diseases is characterized by a combination of genetic and environmental risk factors. Therefore, the effect of a single genetic variant, even a rare one, may have minor effects in absence of environmental risk factors. For example, gene expression changes caused by genetic variants implicated in OFC may not alone cause the emergence of any abnormality, but could lead to the OFC phenotype in combination with maternal exposure to teratogens (e.g. cigarette smoke, alcohol) or dietary deficiencies (e.g. folic acid, vitamins, zinc) (Hutson et al., 2017; Mossey et al., 2009; Stuppia et al., 2011). Interestingly, even paternal exposure to agricultural chemicals has been associated, albeit inconsistently, with OFC (Garcia, 1998; Gordon and Shy, 1981).

There have also been several attempts to correlate variants within specific susceptibility genes with certain environmental risk factors. These studies could represent a valuable personalized approach for primary prevention of OFC for at-risk mothers. Detection of certain genetic risk-variants in pregnant women could make it possible to recommend these women to avoid exposure of specific environmental agents (Mossey et al., 2009; Stuppia et al., 2011).

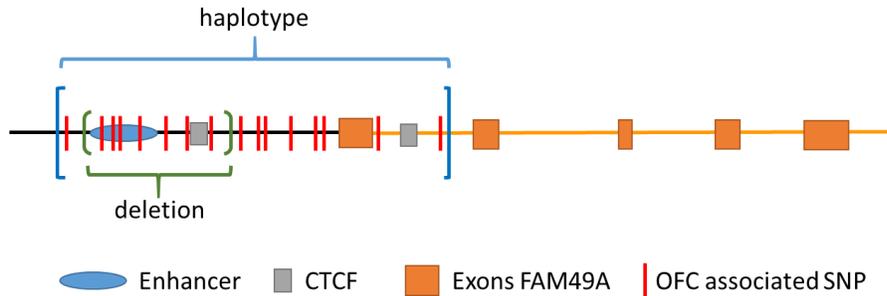
Prenatal alcohol exposure (PAE) itself is the greatest known cause of developmental disability (Smith et al., 2014b). The thereby caused fetal alcohol spectrum disorder (FASD) has a wide range of phenotypic manifestations including craniofacial deficits (Smith et al., 2014b). Investigating the gene-environmental interactions between OFC-risk genetic variants and common teratogens, such as alcohol, can lead to new insights into the relationship between susceptibility loci and environmental factors (Smith et al., 2014a).

Therefore, when looking for disease-causative variants, one may have to evaluate the combined effect of different genetic variants and teratogens. In principle, these type of studies could be performed with our *in vitro* differentiation system, using CRISPR/Cas9 to engineer the variants, and adding the teratogen in different doses to the cell culture medium. However, although different concentrations of the teratogen can easily be tested, this type of *in vitro* differentiation systems cannot recapitulate how different teratogens might be processed in the body of the mother. In any case, initial and potentially relevant results obtained with these *in vitro* systems can then be confirmed using *in vivo* models that can better recapitulate certain aspects of human embryogenesis.

#### 4.6 Enhancer redundancy as a widespread mechanism conferring transcriptional robustness

The expression of many developmental genes is controlled by multiple and redundant enhancers which provide gene expression and phenotypic robustness against mutations within individual enhancers as well as against environmental stress (Osterwalder et al., 2018). However, in other cases, some highly-conserved cell type-specific enhancers seem to have such important function and the expression of their target gene is affected by the deletion of either individual or pairs of enhancers (Dickel et al., 2018). Additionally, even genetic variants within single enhancers altering individual TF binding have been shown to affect target gene expression and cause abnormal phenotypes (Fogarty et al., 2014; Fortini et al., 2014; Guenther et al., 2014; He et al., 2015; Huang et al., 2014; Leslie et al., 2015; Pan et al., 2017; Pasquali et al., 2014; Spieler et al., 2014; Uslu et al., 2014; Visser et al., 2015). The Enh2p24.2 is predicted to control the expression of *MYCN* and/or *DDX1* in the cranial NC and facial tissues. This enhancer is located together with these two genes, in a large TAD that contains dozens of additional enhancers. It is likely that all these enhancers, including Enh2p24.2, control in an at least partially redundant manner the expression of *MYCN* and *DDX1*.

Our initial allele-specific analysis to investigate the functional relevance of Enh2p24.2 suggests a less robust gene expression control when one enhancer allele is deleted (data not shown). However, these results should be taken with caution, since few biological replicates have been analyzed so far. Unambiguous evidence supporting the functional role of Enh2p24.2 will require considerable higher number of replicates and a rigorous statistical analysis of the data. This more extensive characterization will clarify if this specific enhancer is essential for the expression of any of the target genes and, thus, for NC and craniofacial development. However, because the entire *MYCN-DDX1*-TAD is full of regulatory elements (see Figure 3.2.3), it is very likely that, despite the high conservation of Enh2p24.2 there is some kind of enhancer redundancy that may ensure phenotypic robustness and that makes every enhancer individually dispensable (Osterwalder et al., 2018). Alternatively, loss of a single, weakly active enhancer may have a very mild effect, in case enhancer additivity applies (Bothma et al., 2015). Therefore, the disruption of Enh2p24.2 might only lead to gene expression and phenotypic changes under suboptimal conditions, e.g. environmental risk factors (e.g. alcohol) and/or additional genetic variants in other loci.



**Figure 4.6.1 Schematic illustration of the OFC associated haplotype at 2p24.2.**

OFC-associated SNPs are stretched across a region of 30 kb (chr2:16523421-16554164(hg38)), the 3' end of the haplotype, including the highly significant OFC associated SNP rs7552, coinciding with the 5' UTR of the *FAM49A* gene while the 5' end of the haplotype overlaps with the conserved enhancer Enh2p24.2. The 9 kb enhancer deletion generated in this work comprises not only the 3 kb enhancer Enh2p24.2 but also a CTCF binding site located 3' of the enhancer.

Another interesting factor of the 2p24.2 OFC locus is that the risk haplotype not only overlaps with the Enh2p24.2 NCC-active enhancer and the *FAM49A* gene, but also with two CTCF binding sites which, according to our 4C-seq experiments (Figure 3.2.4), most likely represent the 3' boundary of the *DDX1-MYCN-TAD* (Figure 4.6.1). Disruption of the CTCF binding site possibly leading to ectopic expression or genes in the neighboring TAD may also be a potential pathomechanism for OFC at the 2p24.2 locus. However, there are several reasons that argue against this gain-of-function mechanism as relevant for OFC etiology: On one hand the presence of another CTCF site within the last intron of *FAM49A* might ensure that a robust TAD boundary for the *DDX1-MYCN-TAD* is still established (see Figure 4.6.1). Moreover, additional CTCF binding sites further upstream of *FAM49A* (not shown) might place *FAM49A* into its own separated micro-TAD, excluding genes located further upstream of *FAM49A* from being ectopically activated by enhancers within the *DDX1-MYCN-TAD*. On the other hand, our *FAM49A* overexpression experiments *in vivo* argue against the ectopic expression of this gene in the NC and/or craniofacial tissues to be relevant for OFC (Figure 3.2.16). In contrast, it is possible that the deleted CTCF might be important for facilitating the physical proximity between Enh2p24.2 and *DDX1/MYCN*. Ultimately, the contribution of rare SNPs within the CTCF binding site located between the En2p24.2 and the *FAM49A* gene remains uncertain but offers an interesting platform for further studies of the disease-associated haplotype that lead us to *DDX1* as a novel regulator of facial development.

#### 4.7 DDX1 as a novel regulator of facial development and a new candidate gene for molecular genetic testing in patients with OFC

Yu et al. (2017) and Chen et al. (2018) recently proposed *FAM49A* as the gene through which the 2p24.2 risk locus could contribute to OFC. However, our current data does not support neither *FAM49A* as the target of the Enh2p24.2 enhancer nor as a major regulator of facial development. Therefore, *FAM49A* is unlikely to be involved in OFC etiology.

The main potential target genes of Enh2p24.2, the ATP-dependent RNA helicase DDX1 and the transcription factor MYCN, have so far not been directly connected to the OFC phenotype. Therefore, both genes could provide novel insights into the molecular basis of OFC. Still, mutations within MYCN have been identified in patients with Feingold syndrome, which includes facial aberrations (Burnside et al., 2018; Chen et al., 2012; Marcelis et al., 2008; Teszas et al., 2006; van Bokhoven et al., 2005). MYCN has previously been found to be critical for cell growth and cell cycle progression in hNCC (Zhang et al., 2016). Also, MYCN has a major role as an oncogene in neuroblastomas, an NC-derived pediatric tumor (Brodeur et al., 1984; Kohl et al., 1984; Kohl et al., 1983; Kramer et al., 2016; Liu et al., 2019; Schwab et al., 1983; Zeid et al., 2018). Altogether, a role for MYCN in facial development and OFC is not surprising and already supported by the current literature.

*MYCN* and *DDX1* have been reported to be co-amplified in cases of neuroblastomas (Defferrari et al., 2007; Weber et al., 2004) and nephroblastomas, an NC unrelated pediatric tumor (Fievet et al., 2013; Micale et al., 2016). In contrast, microdeletions identified in cases of Feingold syndrome affected both *MYCN* and *DDX1* (Burnside et al., 2018; Chen et al., 2012). The duplications and deletions affecting both genes are not surprising, given the close physical proximity of the two genes, yet they could potentially hint towards a combined effect of both genes. This hypothesis is supported by our observation that the NCC-specific knockout of either of the two genes individually causes a similar phenotype (i.e. OFC). We speculate that by affecting the expression of both *MYCN* and *DDX1*, the 2p24.2 causative variants could contribute to OFC even if the expression changes for each gene individually are moderate.

Similar to our analysis of DDX1, the cellular function of MYCN could, due to its known function as a transcription factor, be examined in more detail by chromatin immunoprecipitation sequencing (ChIP-seq) in hNCC. This might lead to novel insights regarding the possible MYCN downstream target genes in hNCC, which could help to understand the function of this transcription factor during development. Furthermore, comparison of MYCN binding and H3K27ac occupancy in WT

and *MYCN*<sup>-/-</sup> hNCC could clarify the relationship between both marks, since loss of the mouse *MYCN* homolog *N-myc* has been shown to lead to global reduction of histone acetylation through downregulation of a histone acetyltransferase (HAT), but the exact location of these events have not been determined (Knoepfler et al., 2006).

For *DDX1* on the other hand, no involvement in facial development has been reported so far. However, *DDX21*, another RNA helicases belonging to the same family, has been found to play a role in ribosomopathies (Calo et al., 2018), a group of congenital malformations with heterozygous mutations in general regulators of ribosome biogenesis (Yelick and Trainor, 2015). Interestingly, although ribosomes are required across all tissues, the defects associated with ribosomopathies are manifested in a tissue-selective manner, e.g. for *DDX21* in NC-derived craniofacial tissues (Calo et al., 2018; Yelick and Trainor, 2015). This is similar to what may be happening for *DDX1*, which according to full KO appears to have a vital role for cell survival (Hildebrandt et al., 2015; Ribeiro de Almeida et al., 2018) but that could also be involved in tissue-specific defects (e.g. OFC) when its levels are partially reduced or when its expression is diminished in specific cellular contexts (e.g. NCC).

Although the exact molecular mechanism through which *DDX1* misregulation contributes to OFC still needs to be determined, our results strongly indicate that *DDX1* represents a novel regulator of facial development. Therefore, the *DDX1* gene could be screened for coding mutations in patients with OFC. Until now, *DDX1* had not been considered as a marker for OFC and, consequently, our findings could expand the spectrum of genes that should be screened in OFC patients and also considered in human genetic counseling for expectant mothers. Chiefly, our collaborators (Kerstin Ludwig, Bonn) have access to DNA of OFC patients and recently performed whole exome sequencing in OFC patients. Alterations within the *DDX1* gene itself would ultimately confirm our strategy to identify new genes involved in important developmental processes and complex congenital diseases. Still, OFC remains a complex phenotype where even a coding mutation may not be a definite explanation for the observed phenotype, especially when the mutation is heterozygous and, thus, the expected molecular consequences are mild. It is therefore possible that a reduction in *DDX1* function (due to coding or non-coding mutations) might only lead to an OFC when combined with additional genetic or environmental influences.

Again, it is important to consider that a certain risk locus identified on the basis of genetic variants that are prevalent in a specific population may not harbor variants in a group of patients with different ethnicity to the group of patient this risk-locus was identified in. Our candidate locus 2p24 leading us to the *DDX1* gene was found to be highly significantly associated with OFC in a

Chinese population (Chen et al., 2018; Yu et al., 2017). Therefore, the *DDX1* gene or its regulatory landscape might preferentially carry variants that can potentially cause OFC in individuals of Chinese origin.

#### 4.8 Potential function of *DDX1* during NC development

The phenotype we observed in the *in vitro* knockout of *DDX1* (Figure 3.2.28), suggests a proliferation defect in hNCCp2 which, according to functional annotation of up-regulated genes in *DDX1*<sup>+/-</sup> clones might be a consequence of impaired cell cycle control (Figure 3.2.30). This could occur due to an increase in DNA damage, as previous work by Li et al. (2008) indicates that *DDX1* co-localizes with  $\gamma$ H2AX and could be implicated in DNA repair. A similar phenotype to the one we observed in hNCCp2 was previously reported in colorectal cancer cells, where CRISPR/Cas9 mediated knockout of *DDX1* lead to a slower growth rate and ultimately to loss of tumorigenicity (Tanaka et al., 2018). Chiefly, impaired proliferation in mesenchymal progenitors during palate formation is considered to be critical for OFC, thus supporting the relevance of *DDX1* in the etiology of this common abnormality as well as in palatogenesis (Meng et al., 2009).

In order to investigate whether the growth defect that we have observed in *DDX1*<sup>+/-</sup> hNCCp2 is due to impaired proliferation, these cells could be stained for KI-67, a widely used proliferation marker (Scholzen and Gerdes, 2000). In addition, we could determine the cell cycle profile of hNCCp2, which would tell us more about the distribution of cells within the different stages of the cell cycle and reveal a potential cell cycle arrest. Alternatively, the reduced growth that we observed in *DDX1*<sup>+/-</sup> hNCCp2 could also be due to an increase in the cells that die or that become senescent. To examine these possibilities, staining of Annexin V or  $\beta$ -galactosidase could be used to quantify the number of apoptotic and senescent cells, respectively. Furthermore, if the reason for the defective proliferation/growth of *DDX1*<sup>+/-</sup> hNCCp2 is an impairment of DNA damage repair, regardless of whether this results in cell cycle arrest and/or apoptosis, this could be shown by detection of DNA damage by immunostaining of  $\gamma$ H2AX (Li et al., 2008). The binding pattern of *DDX1* throughout the gene bodies of highly transcribed genes observed by ChIP-seq (Figure 3.2.32, Figure 3.2.33) closely resembles that previously observed for RNA Pol II (Descostes et al., 2014; Harlen and Churchman, 2017). This indicates that *DDX1* binding might somehow be coupled to the RNA Pol II transcription machinery and, thus possibly involved in transcription-coupled repair.

Further mechanistic analysis of DDX1 may reveal how its loss leads to the growth defects observed in hNCCp2 and to an OFC-like phenotype *in vivo*. There are a couple of experimental procedures that may help us to uncover the exact mechanism by which DDX1 is involved in hNCC development:

(i) If we confirm that *DDX1*<sup>-/-</sup> hNCCp2 display increased DNA damage, we could investigate whether DDX1 is directly involved in DNA repair by performing DDX1 immunoprecipitation and subsequent mass spectrometry to identify proteins interacting with DDX1. This could reveal whether DDX1 interacts with proteins involved in DNA repair and also provide insights into the particular DNA repair pathway in which it might be involved (e.g. NHEJ, NER, etc.).

(ii) Due to its proposed interaction with RNA, the molecular function of DDX1 could be further investigated by performing cross-linking immunoprecipitation sequencing (CLIP-seq) experiments to identify the RNA binding sequences bound by DDX1 in hNCCp2. This might tell us more about the cellular function of DDX1 in hNCC and its role during embryonic development. For example, according to our ChIP-seq profiles and if, as we speculate, DDX1 is involved in transcription-coupled repair, CLIP experiments might show general DDX1 binding to the transcripts of highly expressed genes.

(iii) Since *MYCN* and *DDX1* have both been found to be overexpressed in neuroblastomas and it has been shown that both genes are often co-expressed (De Preter et al., 2002; Defferrari et al., 2007; Kaneko et al., 2007; Manohar et al., 1995), there is a possibility that both genes are part of the same pathway. If this is the case, it would be interesting to test if the *DDX1*<sup>-/-</sup> phenotype could also be rescued by overexpression of *MYCN* or if the *MYCN*<sup>-/-</sup> phenotype could be rescued by overexpression of *DDX1*. Intriguingly, and in accordance with this hypothesis, both *MYCN* and *DDX1* have recently been implicated in a DNA damage response mechanism, specifically DNA double strand break repair (Hu et al., 2019; Li et al., 2016). Moreover, *MYCN* was found to be involved in a novel *MYCN*-PARP-DNA damage response pathway (Zhang et al., 2018), which points towards base or nucleotide excision repair. Taken together, there is increasing evidence that the molecular mechanism by which both *MYCN* and *DDX1* might be involved in the etiology of OFC could be connected to a DNA repair mechanism but the exact role of and the connection between both genes still remains to be elucidated. And while *MYCN* is known to regulate facial development, the identification of *DDX1* as a novel regulator in this context opens up new avenues towards a better understanding of these complicated processes.

## 4.9 Conclusion and significance

The formation of the vertebrate face requires a network of tightly controlled genes whose expression has to be spatially and temporarily regulated. While the occurrence of common SNPs contributes to the phenotypic variation of our faces (and to many other individual traits), disturbance of crucial regulatory regions by rare SNPs, e.g. by preventing TF-binding or even binding of a 3D structural protein (e.g. CTCF, Cohesin), or SVs can result in quantitatively larger gene expression changes that disrupt facial development and lead to malformations. Here we investigated the pathomechanism of both a syndromic form of OFC (i.e. BOFS) and the non-syndromic form of OFC.

First, we conclusively show that the heterozygous 89 Mb inversion found in our BOFS patient causes the physical disconnection between the inverted *TFAP2A* allele and its cognate enhancers, ultimately leading to monoallelic and haploinsufficient *TFAP2A* expression in human NCC. Furthermore, we show that the silencing of the inverted *TFAP2A* allele does not involve PEV and that, at least in hNCC, this inversion does not result in enhancer adoption or ectopic gains in gene expression. Overall, our work provides a powerful approach to investigate the pathomechanisms of structural variants predicted to disrupt 3D genome organization and that, due to various reasons (i.e. limited access to relevant patient material, differences in gene dosage sensitivity between mice and humans, difficulties in recapitulating certain structural variants), cannot be properly evaluated *in vivo*.

Second, we combined hNCC enhancer maps with OFC risk-loci and identified the highly conserved enhancer Enh2p24.2 as a potential candidate harboring genetic variants involved in OFC. Using Enh2p24.2 as a discovery bait, we identified *MYCN* and *DDX1* as its targets, which we then showed to be essential for normal facial development. While *MYCN* was not a surprising candidate to be involved in the etiology of OFC, the identification of *DDX1* as a novel regulator of facial development might provide new insights into the molecular processes (e.g. transcription-coupled DNA repair) implicated in OFC and, potentially, other human neurocristopathies (e.g. neuroblastoma).

## REFERENCES

- Abdelhaleem, M., Maltais, L., and Wain, H. (2003). The human DDX and DHX gene families of putative RNA helicases. *Genomics* *81*, 618-622.
- Abramyan, J., and Richman, J.M. (2018). Craniofacial development: discoveries made in the chicken embryo. *Int J Dev Biol* *62*, 97-107.
- Acloque, H., Adams, M.S., Fishwick, K., Bronner-Fraser, M., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. *J Clin Invest* *119*, 1438-1449.
- Acloque, H., Wilkinson, D.G., and Nieto, M.A. (2008). In situ hybridization analysis of chick embryos in whole-mount and tissue sections. *Methods Cell Biol* *87*, 169-185.
- Adamo, A., Atashpaz, S., Germain, P.L., Zanella, M., D'Agostino, G., Albertin, V., Chenoweth, J., Micale, L., Fusco, C., Unger, C., *et al.* (2015). 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. *Nat Genet* *47*, 132-141.
- Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., *et al.* (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* *44*, W3-W10.
- Ahituv, N., Rubin, E.M., and Nobrega, M.A. (2004). Exploiting human--fish genome comparisons for deciphering gene regulation. *Hum Mol Genet* *13 Spec No 2*, R261-266.
- Aliferis, K., Stoetzel, C., Pelletier, V., Helle, S., Angioi-Duprez, K., Vigneron, J., Leheup, B., Marion, V., and Dollfus, H. (2011). A novel TFAP2A mutation in familial Branchio-Oculo-Facial Syndrome with predominant ocular phenotype. *Ophthalmic Genet* *32*, 250-255.
- Ameur, A., Rada-Iglesias, A., Komorowski, J., and Wadelius, C. (2009). Identification of candidate regulatory SNPs by combination of transcription-factor-binding site prediction, SNP genotyping and haploChIP. *Nucleic Acids Res* *37*, e85.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* *11*, R106.
- Andersen, T.A., Troelsen Kde, L., and Larsen, L.A. (2014). Of mice and men: molecular genetics of congenital heart disease. *Cell Mol Life Sci* *71*, 1327-1352.
- Antolovic, V., Miermont, A., Corrigan, A.M., and Chubb, J.R. (2017). Generation of Single-Cell Transcript Variability by Repression. *Curr Biol* *27*, 1811-1817 e1813.
- Arnold, C.D., Zabidi, M.A., Pagani, M., Rath, M., Schernhuber, K., Kazmar, T., and Stark, A. (2017). Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. *Nat Biotechnol* *35*, 136-144.
- Bajpai, R., Chen, D.A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C.P., Zhao, Y., Swigut, T., and Wysocka, J. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature* *463*, 958-962.
- Baker, C.V., and Bronner-Fraser, M. (1997a). The origins of the neural crest. Part I: embryonic induction. *Mech Dev* *69*, 3-11.
- Baker, C.V., and Bronner-Fraser, M. (1997b). The origins of the neural crest. Part II: an evolutionary perspective. *Mech Dev* *69*, 13-29.
- Balko, J.M., and Arteaga, C.L. (2011). Dead-box or black-box: is DDX1 a potential biomarker in breast cancer? *Breast Cancer Res Treat* *127*, 65-67.
- Barrallo-Gimeno, A., and Nieto, M.A. (2005). The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development* *132*, 3151-3161.
- Barrell, W.B., Griffin, J.N., Harvey, J.L., HipSci, C., Danovi, D., Beales, P., Grigoriadis, A.E., and Liu, K.J. (2019). Induction of Neural Crest Stem Cells From Bardet-Biedl Syndrome Patient Derived hiPSCs. *Front Mol Neurosci* *12*, 139.
- Barriga, E.H., Trainor, P.A., Bronner, M., and Mayor, R. (2015). Animal models for studying neural crest development: is the mouse different? *Development* *142*, 1555-1560.
- Bartha, I., di Iulio, J., Venter, J.C., and Telenti, A. (2018). Human gene essentiality. *Nat Rev Genet* *19*, 51-62.
- Bauer, D.E., Kamran, S.C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M.C., Smith, E.C., Pinello, L., *et al.* (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* *342*, 253-257.

## References

- Beaty, T.H., Murray, J.C., Marazita, M.L., Munger, R.G., Ruczinski, I., Hetmanski, J.B., Liang, K.Y., Wu, T., Murray, T., Fallin, M.D., *et al.* (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nat Genet* 42, 525-529.
- Beaty, T.H., Ruczinski, I., Murray, J.C., Marazita, M.L., Munger, R.G., Hetmanski, J.B., Murray, T., Redett, R.J., Fallin, M.D., Liang, K.Y., *et al.* (2011). Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. *Genet Epidemiol* 35, 469-478.
- Beaty, T.H., Taub, M.A., Scott, A.F., Murray, J.C., Marazita, M.L., Schwender, H., Parker, M.M., Hetmanski, J.B., Balakrishnan, P., Mansilla, M.A., *et al.* (2013). Confirming genes influencing risk to cleft lip with/without cleft palate in a case-parent trio study. *Hum Genet* 132, 771-781.
- Bedell, M.A., Largaespada, D.A., Jenkins, N.A., and Copeland, N.G. (1997). Mouse models of human disease. Part II: recent progress and future directions. *Genes Dev* 11, 11-43.
- Bellmeyer, A., Krase, J., Lindgren, J., and LaBonne, C. (2003). The protooncogene c-myc is an essential regulator of neural crest formation in xenopus. *Dev Cell* 4, 827-839.
- Benish, B.M. (1975). Letter: "The neurocristopathies: a unifying concept of disease arising in neural crest development". *Hum Pathol* 6, 128.
- Bennaceur, S., Buisson, T., Bertolus, C., and Couly, G. (1998). Branchio-oculo-facial syndrome with cleft lip and bilateral dermal thymus. *Cleft Palate Craniofac J* 35, 454-459.
- Bergsland, M., Ramskold, D., Zaouter, C., Klum, S., Sandberg, R., and Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev* 25, 2453-2464.
- Bettters, E., Liu, Y., Kjaeldgaard, A., Sundstrom, E., and Garcia-Castro, M.I. (2010). Analysis of early human neural crest development. *Dev Biol* 344, 578-592.
- Bianco, S., Lupianez, D.G., Chiariello, A.M., Annunziatella, C., Kraft, K., Schopflin, R., Wittler, L., Andrey, G., Vingron, M., Pombo, A., *et al.* (2018). Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat Genet* 50, 662-667.
- Birnbaum, R.Y., Clowney, E.J., Agamy, O., Kim, M.J., Zhao, J., Yamanaka, T., Pappalardo, Z., Clarke, S.L., Wenger, A.M., Nguyen, L., *et al.* (2012a). Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res* 22, 1059-1068.
- Birnbaum, R.Y., Everman, D.B., Murphy, K.K., Gurrieri, F., Schwartz, C.E., and Ahituv, N. (2012b). Functional characterization of tissue-specific enhancers in the DLX5/6 locus. *Hum Mol Genet* 21, 4930-4938.
- Birnbaum, S., Ludwig, K.U., Reutter, H., Herms, S., Steffens, M., Rubini, M., Baluardo, C., Ferrian, M., Almeida de Assis, N., Alblas, M.A., *et al.* (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat Genet* 41, 473-477.
- Bleoo, S., Sun, X., Hendzel, M.J., Rowe, J.M., Packer, M., and Godbout, R. (2001). Association of human DEAD box protein DDX1 with a cleavage stimulation factor involved in 3'-end processing of pre-mRNA. *Mol Biol Cell* 12, 3046-3059.
- Blow, M.J., McCulley, D.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., *et al.* (2010). ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet* 42, 806-810.
- Bochdanovits, Z., Simon-Sanchez, J., Jonker, M., Hoogendijk, W.J., van der Vaart, A., and Heutink, P. (2014). Accurate prediction of a minimal region around a genetic association signal that contains the causal variant. *Eur J Hum Genet* 22, 238-242.
- Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A., *et al.* (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171, 557-572 e524.
- Bothma, J.P., Garcia, H.G., Ng, S., Perry, M.W., Gregor, T., and Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *Elife* 4.
- Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., *et al.* (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 22, 1790-1797.
- Brewer, S., Feng, W., Huang, J., Sullivan, S., and Williams, T. (2004). Wnt1-Cre-mediated deletion of AP-2alpha causes multiple neural crest-related defects. *Dev Biol* 267, 135-152.
- Brito, L.A., Yamamoto, G.L., Melo, S., Malcher, C., Ferreira, S.G., Figueiredo, J., Alvizi, L., Kobayashi, G.S., Naslavsky, M.S., Alonso, N., *et al.* (2015). Rare Variants in the Epithelial Cadherin Gene Underlying the Genetic Etiology of Nonsyndromic Cleft Lip with or without Cleft Palate. *Hum Mutat* 36, 1029-1033.

- Brodeur, G.M., Seeger, R.C., Schwab, M., Varmus, H.E., and Bishop, J.M. (1984). Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* 224, 1121-1124.
- Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* 14, 838-853.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10, 1213-1218.
- Bureau, A., Parker, M.M., Ruczinski, I., Taub, M.A., Marazita, M.L., Murray, J.C., Mangold, E., Noethen, M.M., Ludwig, K.U., Hetmanski, J.B., *et al.* (2014). Whole exome sequencing of distant relatives in multiplex families implicates rare variants in candidate genes for oral clefts. *Genetics* 197, 1039-1044.
- Burnside, R.D., Molinari, S., Botti, C., Brooks, S.S., Chung, W.K., Mehta, L., Schwartz, S., and Papenhausen, P. (2018). Features of Feingold syndrome 1 dominate in subjects with 2p deletions including MYCN. *Am J Med Genet A* 176, 1956-1963.
- Butali, A., Mossey, P.A., Adeyemo, W.L., Eshete, M.A., Gaines, L.A., Even, D., Braimah, R.O., Aregbesola, B.S., Rigdon, J.V., Emeka, C.I., *et al.* (2014). Novel IRF6 mutations in families with Van Der Woude syndrome and popliteal pterygium syndrome from sub-Saharan Africa. *Mol Genet Genomic Med* 2, 254-260.
- Calo, E., Gu, B., Bowen, M.E., Aryan, F., Zalc, A., Liang, J., Flynn, R.A., Swigut, T., Chang, H.Y., Attardi, L.D., *et al.* (2018). Tissue-selective effects of nucleolar stress and rDNA damage in developmental disorders. *Nature* 554, 112-117.
- Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? *Mol Cell* 49, 825-837.
- Calzolari, E., Pierini, A., Astolfi, G., Bianchi, F., Neville, A.J., and Rivieri, F. (2007). Associated anomalies in multi-malformed infants with cleft lip and palate: An epidemiologic study of nearly 6 million births in 23 EUROCAT registries. *Am J Med Genet A* 143A, 528-537.
- Carlson, J.C., Standley, J., Petrin, A., Shaffer, J.R., Butali, A., Buxo, C.J., Castilla, E., Christensen, K., Deleyiannis, F.W., Hecht, J.T., *et al.* (2017). Identification of 16q21 as a modifier of nonsyndromic orofacial cleft phenotypes. *Genet Epidemiol* 41, 887-897.
- Celli, J., van Bokhoven, H., and Brunner, H.G. (2003). Feingold syndrome: clinical review and genetic mapping. *Am J Med Genet A* 122A, 294-300.
- Charron, J., Malynn, B.A., Fisher, P., Stewart, V., Jeannotte, L., Goff, S.P., Robertson, E.J., and Alt, F.W. (1992). Embryonic lethality in mice homozygous for a targeted disruption of the N-myc gene. *Genes Dev* 6, 2248-2257.
- Chen, C., Guo, Q., Shi, J., Jiao, X., Lv, K., Liu, X., Jiang, Y., Hui, X., and Song, T. (2018). Genetic variants of MGMT, RHPN2, and FAM49A contributed to susceptibility of nonsyndromic orofacial clefts in a Chinese population. *J Oral Pathol Med* 47, 796-801.
- Chen, C.P., Lin, S.P., Chern, S.R., Wu, P.S., Chang, S.D., Ng, S.H., Liu, Y.P., Su, J.W., and Wang, W. (2012). A de novo 4.4-Mb microdeletion in 2p24.3 --> p24.2 in a girl with bilateral hearing impairment, microcephaly, digit abnormalities and Feingold syndrome. *Eur J Med Genet* 55, 666-669.
- Chen, C.Y., Chang, I.S., Hsiung, C.A., and Wasserman, W.W. (2014). On the identification of potential regulatory variants within genome wide association candidate SNP sets. *BMC Med Genomics* 7, 34.
- Chen, H.C., Lin, W.C., Tsay, Y.G., Lee, S.C., and Chang, C.J. (2002). An RNA helicase, DDX1, interacting with poly(A) RNA and heterogeneous nuclear ribonucleoprotein K. *J Biol Chem* 277, 40403-40409.
- Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martin, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., *et al.* (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* 167, 1398-1414 e1324.
- Chou, C.F., Lin, W.J., Lin, C.C., Luber, C.A., Godbout, R., Mann, M., and Chen, C.Y. (2013). DEAD box protein DDX1 regulates cytoplasmic localization of KSRP. *PLoS One* 8, e73752.
- Christensen, K., Juel, K., Herskind, A.M., and Murray, J.C. (2004). Long term follow up study of survival associated with cleft lip and palate at birth. *BMJ* 328, 1405.
- Coetzee, S.G., Coetzee, G.A., and Hazelett, D.J. (2015). motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* 31, 3847-3849.
- Cordero, D.R., Brugmann, S., Chu, Y., Bajpai, R., Jame, M., and Helms, J.A. (2011). Cranial neural crest cells on the move: their roles in craniofacial development. *Am J Med Genet A* 155A, 270-279.

## References

- Corradin, O., Cohen, A.J., Luppino, J.M., Bayles, I.M., Schumacher, F.R., and Scacheri, P.C. (2016). Modeling disease risk through analysis of physical interactions between genetic variants within chromatin regulatory circuitry. *Nat Genet* **48**, 1313-1320.
- Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., van Ijcken, W.F.J., Grosveld, F., Frommolt, P., Bazzi, H., *et al.* (2017). PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell* **20**, 689-705 e689.
- Davies, S.J., Wise, C., Venkatesh, B., Mirza, G., Jefferson, A., Volpi, E.V., and Ragoussis, J. (2004). Mapping of three translocation breakpoints associated with orofacial clefting within 6p24 and identification of new transcripts within the region. *Cytogenet Genome Res* **105**, 47-53.
- de Croze, N., Maczkowiak, F., and Monsoro-Burg, A.H. (2011). Reiterative AP2a activity controls sequential steps in the neural crest gene regulatory network. *Proc Natl Acad Sci U S A* **108**, 155-160.
- de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499-506.
- de Lima, R.L., Hoper, S.A., Ghassibe, M., Cooper, M.E., Rorick, N.K., Kondo, S., Katz, L., Marazita, M.L., Compton, J., Bale, S., *et al.* (2009). Prevalence and nonrandom distribution of exonic mutations in interferon regulatory factor 6 in 307 families with Van der Woude syndrome and 37 families with popliteal pterygium syndrome. *Genet Med* **11**, 241-247.
- De Preter, K., Speleman, F., Combaret, V., Lunec, J., Laureys, G., Eussen, B.H., Francotte, N., Board, J., Pearson, A.D., De Paepe, A., *et al.* (2002). Quantification of MYCN, DDX1, and NAG gene copy number in neuroblastoma using a real-time quantitative PCR assay. *Mod Pathol* **15**, 159-166.
- de Vree, P.J., de Wit, E., Yilmaz, M., van de Heijning, M., Klous, P., Verstegen, M.J., Wan, Y., Teunissen, H., Krijger, P.H., Geeven, G., *et al.* (2014). Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat Biotechnol* **32**, 1019-1025.
- Defferrari, R., Tonini, G.P., Conte, M., Papio, F., Sementa, A.R., Valent, A., Schena, F., Perri, P., and Mazzocco, K. (2007). Concomitant DDX1 and MYCN gain in neuroblastoma. *Cancer Lett* **256**, 56-63.
- Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233-1244.
- Descostes, N., Heidemann, M., Spinelli, L., Schuller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., *et al.* (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *Elife* **3**, e02105.
- Dickel, D.E., Visel, A., and Pennacchio, L.A. (2013). Functional anatomy of distant-acting mammalian enhancers. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120359.
- Dickel, D.E., Ypsilanti, A.R., Pla, R., Zhu, Y., Barozzi, I., Mannion, B.J., Khin, Y.S., Fukuda-Yuzawa, Y., Plajzer-Frick, I., Pickle, C.S., *et al.* (2018). Ultraconserved Enhancers Are Required for Normal Development. *Cell* **172**, 491-499 e415.
- Dickinson, D.J., Ward, J.D., Reiner, D.J., and Goldstein, B. (2013). Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nat Methods* **10**, 1028-1034.
- Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., *et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380.
- Dixon, M.J., Marazita, M.L., Beaty, T.H., and Murray, J.C. (2011). Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet* **12**, 167-178.
- Dong, X., Landford, W.N., Hart, J., Risolino, M., Kaymakcalan, O., Jin, J., Toyoda, Y., Ferretti, E., Selleri, L., and Spector, J.A. (2017). Toward Microsurgical Correction of Cleft Lip Ex Utero through Restoration of Craniofacial Developmental Programs. *Plast Reconstr Surg* **140**, 75-85.
- Dooley, C.M., Wali, N., Sealy, I.M., White, R.J., Stemple, D.L., Collins, J.E., and Busch-Nentwich, E.M. (2019). The gene regulatory basis of genetic compensation during neural crest induction. *PLoS Genet* **15**, e1008213.
- Downen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schujijs, J., Lee, T.I., Zhao, K., *et al.* (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374-387.
- Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., *et al.* (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.

- Eckert, D., Buhl, S., Weber, S., Jager, R., and Schorle, H. (2005). The AP-2 family of transcription factors. *Genome Biol* 6, 246.
- Elgin, S.C., and Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb Perspect Biol* 5, a017780.
- Etchevers, H.C., Amiel, J., and Lyonnet, S. (2006). Molecular bases of human neurocristopathies. *Adv Exp Med Biol* 589, 213-234.
- Fagerberg, L., Hallstrom, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., *et al.* (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* 13, 397-406.
- Fang, J., Acheampong, E., Dave, R., Wang, F., Mukhtar, M., and Pomerantz, R.J. (2005). The RNA helicase DDX1 is involved in restricted HIV-1 Rev function in human astrocytes. *Virology* 336, 299-307.
- Fang, J., Kubota, S., Yang, B., Zhou, N., Zhang, H., Godbout, R., and Pomerantz, R.J. (2004). A DEAD box protein facilitates HIV-1 replication as a cellular co-factor of Rev. *Virology* 330, 471-480.
- Feingold, M., Hall, B.D., Lacassie, Y., and Martinez-Frias, M.L. (1997). Syndrome of microcephaly, facial and hand abnormalities, tracheoesophageal fistula, duodenal atresia, and developmental delay. *Am J Med Genet* 69, 245-249.
- Feng, W., Huang, J., Zhang, J., and Williams, T. (2008). Identification and analysis of a conserved Tcfap2a intronic enhancer element required for expression in facial and limb bud mesenchyme. *Mol Cell Biol* 28, 315-325.
- Feng, W., Leach, S.M., Tipney, H., Phang, T., Geraci, M., Spritz, R.A., Hunter, L.E., and Williams, T. (2009). Spatial and temporal analysis of gene expression during growth and fusion of the mouse facial prominences. *PLoS One* 4, e8066.
- Fievet, A., Belaud-Rotureau, M.A., Dugay, F., Abadie, C., Henry, C., Taque, S., Andrieux, J., Guyetant, S., Robert, M., Dubourg, C., *et al.* (2013). Involvement of germline DDX1-MYCN duplication in inherited nephroblastoma. *Eur J Med Genet* 56, 643-647.
- Fogarty, M.P., Cannon, M.E., Vadlamudi, S., Gaulton, K.J., and Mohlke, K.L. (2014). Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. *PLoS Genet* 10, e1004633.
- Fortini, B.K., Tring, S., Plummer, S.J., Edlund, C.K., Moreno, V., Bresalier, R.S., Barry, E.L., Church, T.R., Figueiredo, J.C., and Casey, G. (2014). Multiple functional risk variants in a SMAD7 enhancer implicate a colorectal cancer risk haplotype. *PLoS One* 9, e111914.
- Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schopflin, R., Kraft, K., Kempfer, R., Jerkovic, I., Chan, W.L., *et al.* (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 538, 265-269.
- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* 15, 2038-2049.
- Fujimoto, A., Lipson, M., Lacro, R.V., Shinno, N.W., Boelter, W.D., Jones, K.L., and Wilson, M.G. (1987). New autosomal dominant branchio-oculo-facial syndrome. *Am J Med Genet* 27, 943-951.
- Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell* 166, 358-368.
- Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Akhmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., and Schier, A.F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* 9, e98186.
- Galliani, E., Burglen, L., Kadlub, N., Just, W., Sznajder, Y., de Villemeur, T.B., Soupre, V., Picard, A., and Vazquez, M.P. (2012). Craniofacial phenotype in the branchio-oculo-facial syndrome: four case reports. *Cleft Palate Craniofac J* 49, 357-364.
- Gammill, L.S., and Bronner-Fraser, M. (2003). Neural crest specification: migrating into genomics. *Nat Rev Neurosci* 4, 795-805.
- Gandhi, S., Piacentino, M.L., Vieceli, F.M., and Bronner, M.E. (2017). Optimization of CRISPR/Cas9 genome editing for loss-of-function in the early chick embryo. *Dev Biol* 432, 86-97.
- Gans, C., and Northcutt, R.G. (1983). Neural crest and the origin of vertebrates: a new head. *Science* 220, 268-273.
- Garcia, A.M. (1998). Occupational exposure to pesticides and congenital malformations: a review of mechanisms, methods, and results. *Am J Ind Med* 33, 232-240.

## References

- Gaubatz, S., Imhof, A., Dosch, R., Werner, O., Mitchell, P., Buettner, R., and Eilers, M. (1995). Transcriptional activation by Myc is under negative control by the transcription factor AP-2. *EMBO J* *14*, 1508-1519.
- Gaynor, R.B., Muchardt, C., Xia, Y.R., Klisak, I., Mohandas, T., Sparkes, R.S., and Lusic, A.J. (1991). Localization of the gene for the DNA-binding protein AP-2 to human chromosome 6p22.3-pter. *Genomics* *10*, 1100-1102.
- Ge, B., Gurd, S., Gaudin, T., Dore, C., Lepage, P., Harmsen, E., Hudson, T.J., and Pastinen, T. (2005). Survey of allelic expression using EST mining. *Genome Res* *15*, 1584-1591.
- George, R.E., Kenyon, R.M., McGuckin, A.G., Malcolm, A.J., Pearson, A.D., and Lunec, J. (1996). Investigation of co-amplification of the candidate genes ornithine decarboxylase, ribonucleotide reductase, syndecan-1 and a DEAD box gene, DDX1, with N-myc in neuroblastoma. United Kingdom Children's Cancer Study Group. *Oncogene* *12*, 1583-1587.
- Germain, D.R., Graham, K., Glubrecht, D.D., Hugh, J.C., Mackey, J.R., and Godbout, R. (2011). DEAD box 1: a novel and independent prognostic marker for early recurrence in breast cancer. *Breast Cancer Res Treat* *127*, 53-63.
- Germain, D.R., Li, L., Hildebrandt, M.R., Simmonds, A.J., Hughes, S.C., and Godbout, R. (2015). Loss of the *Drosophila melanogaster* DEAD box protein Ddx1 leads to reduced size and aberrant gametogenesis. *Dev Biol* *407*, 232-245.
- Gestri, G., Osborne, R.J., Wyatt, A.W., Gerrelli, D., Gribble, S., Stewart, H., Fryer, A., Bunyan, D.J., Prescott, K., Collin, J.R., *et al.* (2009). Reduced TFAP2A function causes variable optic fissure closure and retinal defects and sensitizes eye development to mutations in other morphogenetic regulators. *Hum Genet* *126*, 791-803.
- Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* *512*, 96-100.
- Godbout, R., Li, L., Liu, R.Z., and Roy, K. (2007). Role of DEAD box 1 in retinoblastoma and neuroblastoma. *Future Oncol* *3*, 575-587.
- Godbout, R., Packer, M., and Bie, W. (1998). Overexpression of a DEAD box protein (DDX1) in neuroblastoma and retinoblastoma cell lines. *J Biol Chem* *273*, 21161-21168.
- Godbout, R., Packer, M., Katyal, S., and Bleoo, S. (2002). Cloning and expression analysis of the chicken DEAD box gene DDX1. *Biochim Biophys Acta* *1574*, 63-71.
- Godbout, R., and Squire, J. (1993). Amplification of a DEAD box protein gene in retinoblastoma cell lines. *Proc Natl Acad Sci U S A* *90*, 7578-7582.
- Gomez-Marin, C., Tena, J.J., Acemel, R.D., Lopez-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., Maeso, I., Beccari, L., Aneas, I., Viernas, E., *et al.* (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc Natl Acad Sci U S A* *112*, 7542-7547.
- Gordon, J.E., and Shy, C.M. (1981). Agricultural chemical use and congenital cleft lip and/or palate. *Arch Environ Health* *36*, 213-221.
- Goubau, C., Devriendt, K., Van der Aa, N., Crepel, A., Wieczorek, D., Kleefstra, T., Willemsen, M.H., Rauch, A., Tzschach, A., de Ravel, T., *et al.* (2013). Platelet defects in congenital variant of Rett syndrome patients with FOXP1 mutations or reduced expression due to a position effect at 14q12. *Eur J Hum Genet* *21*, 1349-1355.
- Grant, S.F., Wang, K., Zhang, H., Glaberson, W., Annaiah, K., Kim, C.E., Bradfield, J.P., Glessner, J.T., Thomas, K.A., Garris, M., *et al.* (2009). A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J Pediatr* *155*, 909-913.
- Griessinger, E., Anjos-Afonso, F., Vargaftig, J., Taussig, D.C., Lassailly, F., Prebet, T., Imbert, V., Nebout, M., Vey, N., Chabannon, C., *et al.* (2016). Frequency and Dynamics of Leukemia-Initiating Cells during Short-term Ex Vivo Culture Informs Outcomes in Acute Myeloid Leukemia Patients. *Cancer Res* *76*, 2082-2086.
- Groschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A.M., Erpelinck, C., van der Velden, V.H.J., Havermans, M., Avellino, R., van Lom, K., *et al.* (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* *157*, 369-381.
- Guenther, C.A., Tasic, B., Luo, L., Bedell, M.A., and Kingsley, D.M. (2014). A molecular basis for classic blond hair color in Europeans. *Nat Genet* *46*, 748-752.
- Haaland, O.A., Lie, R.T., Romanowska, J., Gjerdevik, M., Gjessing, H.K., and Jugessur, A. (2018). A Genome-Wide Search for Gene-Environment Effects in Isolated Cleft Lip with or without Cleft Palate

- Triads Points to an Interaction between Maternal Periconceptional Vitamin Use and Variants in ESRRG. *Front Genet* 9, 60.
- Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yanez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B., *et al.* (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* 169, 693-707 e614.
- Haberle, V., Arnold, C.D., Pagani, M., Rath, M., Schernhuber, K., and Stark, A. (2019). Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* 570, 122-126.
- Hamburger, V., and Hamilton, H.L. (1992). A series of normal stages in the development of the chick embryo. 1951. *Dev Dyn* 195, 231-272.
- Hamosh, A., Scott, A.F., Amberger, J., Bocchini, C., Valle, D., and McKusick, V.A. (2002). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 30, 52-55.
- Han, C., Liu, Y., Wan, G., Choi, H.J., Zhao, L., Ivan, C., He, X., Sood, A.K., Zhang, X., and Lu, X. (2014). The RNA-binding protein DDX1 promotes primary microRNA maturation and inhibits ovarian tumor progression. *Cell Rep* 8, 1447-1460.
- Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Biggs, D., Preece, C., Downes, D.J., Gosden, M., Sharpe, J.A., Sloane-Stanley, J.A., Hughes, J.R., *et al.* (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat Cell Biol* 19, 952-961.
- Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol* 18, 263-273.
- Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., *et al.* (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22, 1760-1774.
- He, H., Li, W., Liyanarachchi, S., Srinivas, M., Wang, Y., Akagi, K., Wang, Y., Wu, D., Wang, Q., Jin, V., *et al.* (2015). Multiple functional variants in long-range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer. *Proc Natl Acad Sci U S A* 112, 6128-6133.
- Heimberger, A.B., McGary, E.C., Suki, D., Ruiz, M., Wang, H., Fuller, G.N., and Bar-Eli, M. (2005). Loss of the AP-2alpha transcription factor is associated with the grade of human gliomas. *Clin Cancer Res* 11, 267-272.
- Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., *et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108-112.
- Helms, J.A., Cordero, D., and Tapadia, M.D. (2005). New insights into craniofacial morphogenesis. *Development* 132, 851-861.
- Henriksson, M., and Luscher, B. (1996). Proteins of the Myc network: essential regulators of cell growth and differentiation. *Adv Cancer Res* 68, 109-182.
- Hildebrandt, M.R., Germain, D.R., Monckton, E.A., Brun, M., and Godbout, R. (2015). Ddx1 knockout results in transgenerational wild-type lethality in mice. *Sci Rep* 5, 9829.
- Hildebrandt, M.R., Wang, Y., Li, L., Yasmin, L., Glubrecht, D.D., and Godbout, R. (2019). Cytoplasmic aggregation of DDX1 in developing embryos: Early embryonic lethality associated with Ddx1 knockout. *Dev Biol*.
- Hindorf, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106, 9362-9367.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934-947.
- Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., *et al.* (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 351, 1454-1458.
- Hormozdiari, F., Kichaev, G., Yang, W.Y., Pasaniuc, B., and Eskin, E. (2015). Identification of causal genes for complex traits. *Bioinformatics* 31, i206-213.
- Hormozdiari, F., Kostem, E., Kang, E.Y., Pasaniuc, B., and Eskin, E. (2014). Identifying causal variants at loci with multiple signals of association. *Genetics* 198, 497-508.
- Howe, D.G., Bradford, Y.M., Conlin, T., Eagle, A.E., Fashena, D., Frazer, K., Knight, J., Mani, P., Martin, R., Moxon, S.A., *et al.* (2013). ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. *Nucleic Acids Res* 41, D854-860.

## References

- Hu, X., Zheng, W., Zhu, Q., Gu, L., Du, Y., Han, Z., Zhang, X., Carter, D.R., Cheung, B.B., Qiu, A., *et al.* (2019). Increase in DNA Damage by MYCN Knockdown Through Regulating Nucleosome Organization and Chromatin State in Neuroblastoma. *Front Genet* 10, 684.
- Huang, D., and Ovcharenko, I. (2015). Identifying causal regulatory SNPs in ChIP-seq enhancers. *Nucleic Acids Res* 43, 225-236.
- Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4, 44-57.
- Huang, Q., Whittington, T., Gao, P., Lindberg, J.F., Yang, Y., Sun, J., Vaisanen, M.R., Szulkin, R., Annala, M., Yan, J., *et al.* (2014). A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. *Nat Genet* 46, 126-135.
- Humphreys, R., Zheng, W., Prince, L.S., Qu, X., Brown, C., Loomes, K., Huppert, S.S., Baldwin, S., and Goudy, S. (2012). Cranial neural crest ablation of Jagged1 recapitulates the craniofacial phenotype of Alagille syndrome patients. *Hum Mol Genet* 21, 1374-1383.
- Hutson, M.R., Keyte, A.L., Hernandez-Morales, M., Gibbs, E., Kupchinsky, Z.A., Argyridis, I., Erwin, K.N., Pegram, K., Kneifel, M., Rosenberg, P.B., *et al.* (2017). Temperature-activated ion channels in neural crest cells confer maternal fever-associated birth defects. *Sci Signal* 10.
- Ishaq, M., Ma, L., Wu, X., Mu, Y., Pan, J., Hu, J., Hu, T., Fu, Q., and Guo, D. (2009). The DEAD-box RNA helicase DDX1 interacts with RelA and enhances nuclear factor kappaB-mediated transcription. *J Cell Biochem* 106, 296-305.
- Ittiwut, R., Ittiwut, C., Siriwan, P., Chichareon, V., Suphapeetiporn, K., and Shotelersuk, V. (2016). Variants of the CDH1 (E-Cadherin) Gene Associated with Oral Clefts in the Thai Population. *Genet Test Mol Biomarkers* 20, 406-409.
- Jamal, M., Lewandowski, S.L., Lawton, M.L., Huang, G.T., and Ikonomou, L. (2018). Derivation and characterization of putative craniofacial mesenchymal progenitor cells from human induced pluripotent stem cells. *Stem Cell Res* 33, 100-109.
- Jean, D., Gershenwald, J.E., Huang, S., Luca, M., Hudson, M.J., Tainsky, M.A., and Bar-Eli, M. (1998). Loss of AP-2 results in up-regulation of MCAM/MUC18 and an increase in tumor growth and metastasis of human melanoma cells. *J Biol Chem* 273, 16501-16508.
- Ji, X., Dadon, D.B., Powell, B.E., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G., Lee, T.I., *et al.* (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* 18, 262-275.
- Jugessur, A., Farlie, P.G., and Kilpatrick, N. (2009). The genetics of isolated orofacial clefts: from genotypes to subphenotypes. *Oral Dis* 15, 437-453.
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., *et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430-435.
- Kaltschmidt, B., Kaltschmidt, C., and Widera, D. (2012). Adult craniofacial stem cells: sources and relation to the neural crest. *Stem Cell Rev Rep* 8, 658-671.
- Kanai, Y., Dohmae, N., and Hirokawa, N. (2004). Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron* 43, 513-525.
- Kaneko, S., Ohira, M., Nakamura, Y., Isogai, E., Nakagawara, A., and Kaneko, M. (2007). Relationship of DDX1 and NAG gene amplification/overexpression to the prognosis of patients with MYCN-amplified neuroblastoma. *J Cancer Res Clin Oncol* 133, 185-192.
- Kerosuo, L., and Bronner, M.E. (2016). cMyc Regulates the Size of the Premigratory Neural Crest Stem Cell Pool. *Cell Rep* 17, 2648-2659.
- Khurana, E., Fu, Y., Colonna, V., Mu, X.J., Kang, H.M., Lappalainen, T., Sboner, A., Lochovsky, L., Chen, J., Harmanci, A., *et al.* (2013). Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science* 342, 1235587.
- Kichaev, G., Yang, W.Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A.L., Kraft, P., and Pasaniuc, B. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet* 10, e1004722.
- Kilpinen, H., Waszak, S.M., Gschwind, A.R., Raghav, S.K., Witwicki, R.M., Orioli, A., Migliavacca, E., Wiederkehr, M., Gutierrez-Arcelus, M., Panousis, N.I., *et al.* (2013). Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* 342, 744-747.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14, R36.

- Kleinjan, D.A., Seawright, A., Schedl, A., Quinlan, R.A., Danes, S., and van Heyningen, V. (2001). Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. *Hum Mol Genet* 10, 2049-2059.
- Kleinjan, D.A., and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* 76, 8-32.
- Knecht, A.K., and Bronner-Fraser, M. (2002). Induction of the neural crest: a multigene process. *Nat Rev Genet* 3, 453-461.
- Knight, R.D., Javidan, Y., Zhang, T., Nelson, S., and Schilling, T.F. (2005). AP2-dependent signals from the ectoderm regulate craniofacial development in the zebrafish embryo. *Development* 132, 3127-3138.
- Knoepfler, P.S., Zhang, X.Y., Cheng, P.F., Gafken, P.R., McMahon, S.B., and Eisenman, R.N. (2006). Myc influences global chromatin structure. *EMBO J* 25, 2723-2734.
- Kohl, N.E., Gee, C.E., and Alt, F.W. (1984). Activated expression of the N-myc gene in human neuroblastomas and related tumors. *Science* 226, 1335-1337.
- Kohl, N.E., Kanda, N., Schreck, R.R., Bruns, G., Latt, S.A., Gilbert, F., and Alt, F.W. (1983). Transposition and amplification of oncogene-related sequences in human neuroblastomas. *Cell* 35, 359-367.
- Kolovos, P., Georgomanolis, T., Koeflerle, A., Larkin, J.D., Brant, L., Nikolicc, M., Gusmao, E.G., Zirkel, A., Knoch, T.A., van Ijcken, W.F., *et al.* (2016). Binding of nuclear factor kappaB to noncanonical consensus sites reveals its multimodal role during the early inflammatory response. *Genome Res* 26, 1478-1489.
- Kondo, S., Schutte, B.C., Richardson, R.J., Bjork, B.C., Knight, A.S., Watanabe, Y., Howard, E., de Lima, R.L., Daack-Hirsch, S., Sander, A., *et al.* (2002). Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nat Genet* 32, 285-289.
- Kraft, K., Geuer, S., Will, A.J., Chan, W.L., Paliou, C., Borschiwer, M., Harabula, I., Wittler, L., Franke, M., Ibrahim, D.M., *et al.* (2015). Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell Rep* 10, 833-839.
- Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schopflin, R., Markowski, J., Ibrahim, D.M., Acuna-Hidalgo, R., Despang, A., Andrey, G., *et al.* (2019). Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat Cell Biol* 21, 305-310.
- Kragesteijn, B.K., Spielmann, M., Paliou, C., Heinrich, V., Schopflin, R., Esposito, A., Annunziatella, C., Bianco, S., Chiariello, A.M., Jerkovic, I., *et al.* (2018). Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat Genet* 50, 1463-1473.
- Kramer, M., Ribeiro, D., Arsenian-Henriksson, M., Deller, T., and Rohrer, H. (2016). Proliferation and Survival of Embryonic Sympathetic Neuroblasts by MYCN and Activated ALK Signaling. *J Neurosci* 36, 10425-10439.
- Krijger, P.H., and de Laat, W. (2016). Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol* 17, 771-782.
- Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., *et al.* (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 44, W90-97.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.
- Laugsch, M., Bartusel, M., Rehimi, R., Alirzayeva, H., Karaolidou, A., Crispatzu, G., Zentis, P., Nikolic, M., Bleckwehl, T., Kolovos, P., *et al.* (2019). Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs. *Cell Stem Cell* 24, 736-752 e712.
- Le Douarin, N., and Kalcheim, C. (1999). *The neural crest*, 2nd edn (Cambridge, UK ; New York, NY, USA: Cambridge University Press).
- Le Douarin, N.M. (2004). The avian embryo as a model to study the development of the neural crest: a long and still ongoing story. *Mech Dev* 121, 1089-1102.
- LeBlanc, S.K., Yu, S., and Barnett, C.P. (2013). 6p.24 microdeletion involving TFAP2A without classic features of branchio-oculo-facial syndrome. *Am J Med Genet A* 161A, 901-904.
- Lee, D., Gorkin, D.U., Baker, M., Strober, B.J., Asoni, A.L., McCallion, A.S., and Beer, M.A. (2015). A method to predict the impact of regulatory variants from DNA sequence. *Nat Genet* 47, 955-961.
- Lee, G., Chambers, S.M., Tomishima, M.J., and Studer, L. (2010). Derivation of neural crest cells from human pluripotent stem cells. *Nat Protoc* 5, 688-701.
- Leslie, E.J., Carlson, J.C., Shaffer, J.R., Butali, A., Buxo, C.J., Castilla, E.E., Christensen, K., Deleyiannis, F.W., Leigh Field, L., Hecht, J.T., *et al.* (2017). 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.

## References

- Leslie, E.J., Carlson, J.C., Shaffer, J.R., Feingold, E., Wehby, G., Laurie, C.A., Jain, D., Laurie, C.C., Doheny, K.F., McHenry, T., *et al.* (2016). A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. *Hum Mol Genet* **25**, 2862-2872.
- Leslie, E.J., Mansilla, M.A., Biggs, L.C., Schuette, K., Bullard, S., Cooper, M., Dunwald, M., Lidral, A.C., Marazita, M.L., Beaty, T.H., *et al.* (2012). Expression and mutation analyses implicate ARHGAP29 as the etiologic gene for the cleft lip with or without cleft palate locus identified by genome-wide association on chromosome 1p22. *Birth Defects Res A Clin Mol Teratol* **94**, 934-942.
- Leslie, E.J., Standley, J., Compton, J., Bale, S., Schutte, B.C., and Murray, J.C. (2013). Comparative analysis of IRF6 variants in families with Van der Woude syndrome and popliteal pterygium syndrome using public whole-exome databases. *Genet Med* **15**, 338-344.
- Leslie, E.J., Taub, M.A., Liu, H., Steinberg, K.M., Koboldt, D.C., Zhang, Q., Carlson, J.C., Hetmanski, J.B., Wang, H., Larson, D.E., *et al.* (2015). Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. *Am J Hum Genet* **96**, 397-411.
- Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725-1735.
- Lettice, L.A., Horikoshi, T., Heaney, S.J., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Jooose, M., Akarsu, N., Oostra, B.A., Endo, N., *et al.* (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A* **99**, 7548-7553.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
- Li, H., Sheridan, R., and Williams, T. (2013). Analysis of TFAP2A mutations in Branchio-Oculo-Facial Syndrome indicates functional complexity within the AP-2alpha DNA-binding domain. *Hum Mol Genet* **22**, 3195-3206.
- Li, L., Germain, D.R., Poon, H.Y., Hildebrandt, M.R., Monckton, E.A., McDonald, D., Hendzel, M.J., and Godbout, R. (2016). DEAD Box 1 Facilitates Removal of RNA and Homologous Recombination at DNA Double-Strand Breaks. *Mol Cell Biol* **36**, 2794-2810.
- Li, L., Monckton, E.A., and Godbout, R. (2008). A role for DEAD box 1 at DNA double-strand breaks. *Mol Cell Biol* **28**, 6413-6425.
- Li, L., Roy, K., Katyal, S., Sun, X., Bleoo, S., and Godbout, R. (2006). Dynamic nature of cleavage bodies and their spatial relationship to DDX1 bodies, Cajal bodies, and gems. *Mol Biol Cell* **17**, 1126-1140.
- Li, S., Alvarez, R.V., Sharan, R., Landsman, D., and Ovcharenko, I. (2017). Quantifying deleterious effects of regulatory variants. *Nucleic Acids Res* **45**, 2307-2317.
- Li, Z., Zhou, M., Cai, Z., Liu, H., Zhong, W., Hao, Q., Cheng, D., Hu, X., Hou, J., Xu, P., *et al.* (2018). RNA-binding protein DDX1 is responsible for fatty acid-mediated repression of insulin translation. *Nucleic Acids Res* **46**, 12052-12066.
- Lidral, A.C., Liu, H., Bullard, S.A., Bonde, G., Machida, J., Visel, A., Uribe, L.M., Li, X., Amendt, B., and Cornell, R.A. (2015). A single nucleotide polymorphism associated with isolated cleft lip and palate, thyroid cancer and hypothyroidism alters the activity of an oral epithelium and thyroid enhancer near FOXE1. *Hum Mol Genet* **24**, 3895-3907.
- Lim, J.H., Booker, A.B., Luo, T., Williams, T., Furuta, Y., Lagutin, O., Oliver, G., Sargent, T.D., and Fallon, J.R. (2005). AP-2alpha selectively regulates fragile X mental retardation-1 gene transcription during embryonic development. *Hum Mol Genet* **14**, 2027-2034.
- Lin, A.E., Haldeman-Englert, C.R., and Milunsky, J.M. (1993). Branchiooculofacial Syndrome. In *GeneReviews*((R)), M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J.H. Bean, K. Stephens, and A. Amemiya, eds. (Seattle (WA)).
- Linder, P., Lasko, P.F., Ashburner, M., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J., and Slonimski, P.P. (1989). Birth of the D-E-A-D box. *Nature* **337**, 121-122.
- Liu, H., Busch, T., Eliason, S., Anand, D., Bullard, S., Gowans, L.J.J., Nidey, N., Petrin, A., Augustine-Akpan, E.A., Saadi, I., *et al.* (2017a). Exome sequencing provides additional evidence for the involvement of ARHGAP29 in Mendelian orofacial clefting and extends the phenotypic spectrum to isolated cleft palate. *Birth Defects Res* **109**, 27-37.

- Liu, K., Wang, S., Liu, Y., Gu, J., Gu, S., Xu, Z., Zhang, R., Wang, Z., Ma, H., Chen, Y., *et al.* (2016). Overexpression of MYCN promotes proliferation of non-small cell lung cancer. *Tumour Biol* 37, 12855-12866.
- Liu, L., Xu, F., Chang, C.K., He, Q., Wu, L.Y., Zhang, Z., and Li, X. (2017b). MYCN contributes to the malignant characteristics of erythroleukemia through EZH2-mediated epigenetic repression of p21. *Cell Death Dis* 8, e3126.
- Liu, Y., Liu, D., and Wan, W. (2019). MYCN-induced E2F5 promotes neuroblastoma cell proliferation through regulating cell cycle progression. *Biochem Biophys Res Commun* 511, 35-40.
- Ludwig, K.U., Bohmer, A.C., Bowes, J., Nikolic, M., Ishorst, N., Wyatt, N., Hammond, N.L., Golz, L., Thieme, F., Barth, S., *et al.* (2017). Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip +/- cleft palate and cleft palate only. *Hum Mol Genet* 26, 829-842.
- Ludwig, K.U., Bohmer, A.C., Rubini, M., Mossey, P.A., Herms, S., Nowak, S., Reutter, H., Alblas, M.A., Lippke, B., Barth, S., *et al.* (2014). Strong association of variants around FOXE1 and orofacial clefting. *J Dent Res* 93, 376-381.
- Ludwig, K.U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A., Becker, J., Herberz, R., AlChawa, T., Nasser, E., *et al.* (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nat Genet* 44, 968-971.
- Luo, T., Lee, Y.H., Saint-Jeannet, J.P., and Sargent, T.D. (2003). Induction of neural crest in *Xenopus* by transcription factor AP2alpha. *Proc Natl Acad Sci U S A* 100, 532-537.
- Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., *et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012-1025.
- Lyu, X., Rowley, M.J., and Corces, V.G. (2018). Architectural Proteins and Pluripotency Factors Cooperate to Orchestrate the Transcriptional Response of hESCs to Temperature Stress. *Mol Cell* 71, 940-955 e947.
- Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27, 1696-1697.
- Malynn, B.A., de Alboran, I.M., O'Hagan, R.C., Bronson, R., Davidson, L., DePinho, R.A., and Alt, F.W. (2000). N-myc can functionally replace c-myc in murine development, cellular growth, and differentiation. *Genes Dev* 14, 1390-1399.
- Mangold, E., Ludwig, K.U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., Reutter, H., de Assis, N.A., Chawa, T.A., Mattheisen, M., *et al.* (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat Genet* 42, 24-26.
- Manohar, C.F., Salwen, H.R., Brodeur, G.M., and Cohn, S.L. (1995). Co-amplification and concomitant high levels of expression of a DEAD box gene with MYCN in human neuroblastoma. *Genes Chromosomes Cancer* 14, 196-203.
- Marazita, M.L., Murray, J.C., Lidral, A.C., Arcos-Burgos, M., Cooper, M.E., Goldstein, T., Maher, B.S., Daack-Hirsch, S., Schultz, R., Mansilla, M.A., *et al.* (2004). Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32-35. *Am J Hum Genet* 75, 161-173.
- Marcelis, C.L., Hol, F.A., Graham, G.E., Rieu, P.N., Kellermayer, R., Meijer, R.P., Lugtenberg, D., Scheffer, H., van Bokhoven, H., Brunner, H.G., *et al.* (2008). Genotype-phenotype correlations in MYCN-related Feingold syndrome. *Hum Mutat* 29, 1125-1132.
- Masotti, C., Brito, L.A., Nica, A.C., Ludwig, K.U., Nunes, K., Savastano, C.P., Malcher, C., Ferreira, S.G., Kobayashi, G.S., Bueno, D.F., *et al.* (2018). MRPL53, a New Candidate Gene for Orofacial Clefting, Identified Using an eQTL Approach. *J Dent Res* 97, 33-40.
- Matharu, N., and Ahituv, N. (2015). Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. *PLoS Genet* 11, e1005640.
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., *et al.* (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190-1195.
- McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28, 495-501.
- McPherson, L.A., Loktev, A.V., and Weigel, R.J. (2002). Tumor suppressor activity of AP2alpha mediated through a direct interaction with p53. *J Biol Chem* 277, 45028-45033.

## References

- Mehrjouy, M.M., Fonseca, A.C.S., Ehmke, N., Paskulin, G., Novelli, A., Benedicenti, F., Mencarelli, M.A., Renieri, A., Busa, T., Missirian, C., *et al.* (2018). Regulatory variants of FOXG1 in the context of its topological domain organisation. *Eur J Hum Genet* 26, 186-196.
- Menendez, L., Kulik, M.J., Page, A.T., Park, S.S., Lauderdale, J.D., Cunningham, M.L., and Dalton, S. (2013). Directed differentiation of human pluripotent cells to neural crest stem cells. *Nat Protoc* 8, 203-212.
- Meng, L., Bian, Z., Torensma, R., and Von den Hoff, J.W. (2009). Biological mechanisms in palatogenesis and cleft palate. *J Dent Res* 88, 22-33.
- Meulemans, D., and Bronner-Fraser, M. (2004). Gene-regulatory interactions in neural crest evolution and development. *Dev Cell* 7, 291-299.
- Micale, M.A., Embrey, B.t., Macknis, J.K., Harper, C.E., and Aughton, D.J. (2016). Constitutional 560.49 kb chromosome 2p24.3 duplication including the MYCN gene identified by SNP chromosome microarray analysis in a child with multiple congenital anomalies and bilateral Wilms tumor. *Eur J Med Genet* 59, 618-623.
- Miller, L.C., Blandford, V., McAdam, R., Sanchez-Carbente, M.R., Badeaux, F., DesGroseillers, L., and Sossin, W.S. (2009). Combinations of DEAD box proteins distinguish distinct types of RNA: protein complexes in neurons. *Mol Cell Neurosci* 40, 485-495.
- Milunsky, J.M., Maher, T.A., Zhao, G., Roberts, A.E., Stalker, H.J., Zori, R.T., Burch, M.N., Clemens, M., Mulliken, J.B., Smith, R., *et al.* (2008). TFAP2A mutations result in branchio-oculo-facial syndrome. *Am J Hum Genet* 82, 1171-1177.
- Milunsky, J.M., Maher, T.M., Zhao, G., Wang, Z., Mulliken, J.B., Chitayat, D., Clemens, M., Stalker, H.J., Bauer, M., Burch, M., *et al.* (2011). Genotype-phenotype analysis of the branchio-oculo-facial syndrome. *Am J Med Genet A* 155A, 22-32.
- Miranda, P., Enkhmandakh, B., and Bayarsaihan, D. (2018). TFII-I and AP2alpha Co-Occupy the Promoters of Key Regulatory Genes Associated with Craniofacial Development. *Cleft Palate Craniofac J* 55, 865-870.
- Mishina, Y., and Snider, T.N. (2014). Neural crest cell signaling pathways critical to cranial bone development and pathology. *Exp Cell Res* 325, 138-147.
- Mitchell, P.J., Wang, C., and Tjian, R. (1987). Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. *Cell* 50, 847-861.
- Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., *et al.* (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. *Cell Rep* 20, 1215-1228.
- Mohibullah, N., Donner, A., Ippolito, J.A., and Williams, T. (1999). SELEX and missing phosphate contact analyses reveal flexibility within the AP-2[alpha] protein: DNA binding complex. *Nucleic Acids Res* 27, 2760-2769.
- Monteiro, A.N., and Freedman, M.L. (2013). Lessons from postgenome-wide association studies: functional analysis of cancer predisposition loci. *J Intern Med* 274, 414-424.
- Moon, A. (2008). Mouse models of congenital cardiovascular disease. *Curr Top Dev Biol* 84, 171-248.
- Moreno, L.M., Mansilla, M.A., Bullard, S.A., Cooper, M.E., Busch, T.D., Machida, J., Johnson, M.K., Brauer, D., Krahn, K., Daack-Hirsch, S., *et al.* (2009). FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. *Hum Mol Genet* 18, 4879-4896.
- Mossey, P.A., Little, J., Munger, R.G., Dixon, M.J., and Shaw, W.C. (2009). Cleft lip and palate. *Lancet* 374, 1773-1785.
- Murray, B., Wagle, R., Amat-Alarcon, N., Wilkens, A., Stephens, P., Zackai, E.H., Goldmuntz, E., Calkins, H., Deardorff, M.A., and Judge, D.P. (2013). A family with a complex clinical presentation characterized by arrhythmogenic right ventricular dysplasia/cardiomyopathy and features of branchio-oculo-facial syndrome. *Am J Med Genet A* 161A, 371-376.
- Murray, J.C., Nishimura, D.Y., Buetow, K.H., Ardinger, H.H., Spence, M.A., Sparkes, R.S., Falk, R.E., Falk, P.M., Gardner, R.J., Harkness, E.M., *et al.* (1990). Linkage of an autosomal dominant clefting syndrome (Van der Woude) to loci on chromosome 1q. *Am J Hum Genet* 46, 486-491.
- Neiswanger, K., Weinberg, S.M., Rogers, C.R., Brandon, C.A., Cooper, M.E., Bardi, K.M., Deleyiannis, F.W., Resick, J.M., Bowen, A., Mooney, M.P., *et al.* (2007). Orbicularis oris muscle defects as an expanded phenotypic feature in nonsyndromic cleft lip with or without cleft palate. *Am J Med Genet A* 143A, 1143-1149.
- Nelson, D.K., and Williams, T. (2004). Frontonasal process-specific disruption of AP-2alpha results in postnatal midfacial hypoplasia, vascular anomalies, and nasal cavity defects. *Dev Biol* 267, 72-92.

- Nikitina, N., Sauka-Spengler, T., and Bronner-Fraser, M. (2008). Dissecting early regulatory relationships in the lamprey neural crest gene network. *Proc Natl Acad Sci U S A* 105, 20083-20088.
- Nikolic, M., Papantonis, A., and Rada-Iglesias, A. (2017). GARLIC: a bioinformatic toolkit for aetiologically connecting diseases and cell type-specific regulatory maps. *Hum Mol Genet* 26, 742-752.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., *et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381-385.
- Northcott, P.A., Lee, C., Zichner, T., Stutz, A.M., Erkek, S., Kawauchi, D., Shih, D.J., Hovestadt, V., Zapatka, M., Sturm, D., *et al.* (2014). Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* 511, 428-434.
- Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L.A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci U S A* 115, E6697-E6706.
- Ohnishi, T., Yamada, K., Watanabe, A., Ohba, H., Sakaguchi, T., Honma, Y., Iwayama, Y., Toyota, T., Maekawa, M., Watanabe, K., *et al.* (2011). Ablation of Mrds1/Ofcc1 induces hyper-gamma-glutamyl transpeptidaseemia without abnormal head development and schizophrenia-relevant behaviors in mice. *PLoS One* 6, e29499.
- Okuno, H., Renault Mihara, F., Ohta, S., Fukuda, K., Kurosawa, K., Akamatsu, W., Sanosaka, T., Kohyama, J., Hayashi, K., Nakajima, K., *et al.* (2017). CHARGE syndrome modeling using patient-iPSCs reveals defective migration of neural crest cells harboring CHD7 mutations. *Elife* 6.
- Onishi, H., Kino, Y., Morita, T., Futai, E., Sasagawa, N., and Ishiura, S. (2008). MBNL1 associates with YB-1 in cytoplasmic stress granules. *J Neurosci Res* 86, 1994-2002.
- Orso, F., Fassetta, M., Penna, E., Solero, A., De Filippo, K., Sismondi, P., De Bortoli, M., and Taverna, D. (2007). The AP-2alpha transcription factor regulates tumor cell migration and apoptosis. *Adv Exp Med Biol* 604, 87-95.
- Osterwalder, M., Barozzi, I., Tissieres, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., *et al.* (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239-243.
- Ozeki, K., Sugiyama, M., Akter, K.A., Nishiwaki, K., Asano-Inami, E., and Senga, T. (2019). FAM98A is localized to stress granules and associates with multiple stress granule-localized proteins. *Mol Cell Biochem* 451, 107-115.
- Pan, G., Ameer, A., Enroth, S., Bysani, M., Nord, H., Cavalli, M., Essand, M., Gyllensten, U., and Wadelius, C. (2017). PATZ1 down-regulates FADS1 by binding to rs174557 and is opposed by SP1/SREBP1c. *Nucleic Acids Res* 45, 2408-2422.
- Park, M.Y., and Kim, Y.C. (2009). A Case of Branchio-oculo-facial Syndrome. *Ann Dermatol* 21, 288-290.
- Pasquali, L., Gaulton, K.J., Rodriguez-Segui, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, I., Tena, J.J., Moran, I., Gomez-Marin, C., van de Bunt, M., *et al.* (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* 46, 136-143.
- Peifer, M., Hertwig, F., Roels, F., Dreidax, D., Gartlgruber, M., Menon, R., Kramer, A., Roncaioli, J.L., Sand, F., Heuckmann, J.M., *et al.* (2015). Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* 526, 700-704.
- Perez-Gonzalez, A., Pazo, A., Navajas, R., Ciordia, S., Rodriguez-Frandsen, A., and Nieto, A. (2014). hCLE/C14orf166 associates with DDX1-HSPC117-FAM98B in a novel transcription-dependent shuttling RNA-transporting complex. *PLoS One* 9, e90957.
- Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., *et al.* (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* 38, 603-613.
- Peyrard-Janvid, M., Leslie, E.J., Kousa, Y.A., Smith, T.L., Dunnwald, M., Magnusson, M., Lentz, B.A., Unneberg, P., Fransson, I., Koillinen, H.K., *et al.* (2014). Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. *Am J Hum Genet* 94, 23-32.
- Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol* 16, 245-257.
- Popow, J., Jurkin, J., Schleiffer, A., and Martinez, J. (2014). Analysis of orthologous groups reveals archease and DDX1 as tRNA splicing factors. *Nature* 511, 104-107.
- Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* 163, 68-83.

## References

- Qi, X., Zhang, J., Zhao, Y., Chen, T., Xiang, Y., Hui, J., Cai, D., Liu, Y., Xia, L., Yu, T., *et al.* (2017). The applications of CRISPR screen in functional genomics. *Brief Funct Genomics* *16*, 34-37.
- Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.A., and Trapnell, C. (2017). Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* *14*, 309-315.
- Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S.A., Swigut, T., and Wysocka, J. (2012). Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* *11*, 633-648.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* *470*, 279-283.
- Rada-Iglesias, A., Prescott, S.L., and Wysocka, J. (2013). Human genetic variation within neural crest enhancers: molecular and phenotypic implications. *Philos T R Soc B* *368*.
- Rahimov, F., Jugessur, A., and Murray, J.C. (2012). Genetics of nonsyndromic orofacial clefts. *Cleft Palate Craniofac J* *49*, 73-91.
- Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* *42*, W187-191.
- Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., *et al.* (2017). Cohesin Loss Eliminates All Loop Domains. *Cell* *171*, 305-320 e324.
- Redin, C., Brand, H., Collins, R.L., Kammin, T., Mitchell, E., Hodge, J.C., Hanscom, C., Pillalamarri, V., Seabra, C.M., Abbott, M.A., *et al.* (2017). The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet* *49*, 36-45.
- Rehimi, R., Bartusel, M., Solinas, F., Altmuller, J., and Rada-Iglesias, A. (2017). Chromatin Immunoprecipitation (ChIP) Protocol for Low-abundance Embryonic Samples. *J Vis Exp*.
- Rehimi, R., Nikolic, M., Cruz-Molina, S., Tebartz, C., Frommolt, P., Mahabir, E., Clement-Ziza, M., and Rada-Iglesias, A. (2016). Epigenomics-Based Identification of Major Cell Identity Regulators within Heterogeneous Cell Populations. *Cell Rep* *17*, 3062-3076.
- Reiber, J., Sznajder, Y., Posteguillo, E.G., Muller, D., Lyonnet, S., Baumann, C., and Just, W. (2010). Additional clinical and molecular analyses of TFAP2A in patients with the branchio-oculo-facial syndrome. *Am J Med Genet A* *152A*, 994-999.
- Ren, G., Jin, W., Cui, K., Rodrigez, J., Hu, G., Zhang, Z., Larson, D.R., and Zhao, K. (2017). CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. *Mol Cell* *67*, 1049-1058 e1046.
- Ren, X., Yang, Z., Xu, J., Sun, J., Mao, D., Hu, Y., Yang, S.J., Qiao, H.H., Wang, X., Hu, Q., *et al.* (2014). Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in *Drosophila*. *Cell Rep* *9*, 1151-1162.
- Respuela, P., Nikolic, M., Tan, M., Frommolt, P., Zhao, Y., Wysocka, J., and Rada-Iglesias, A. (2016). Foxd3 Promotes Exit from Naive Pluripotency through Enhancer Decommissioning and Inhibits Germline Specification. *Cell Stem Cell* *18*, 118-133.
- Ribeiro de Almeida, C., Dhir, S., Dhir, A., Moghaddam, A.E., Sattentau, Q., Meinhart, A., and Proudfoot, N.J. (2018). RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. *Mol Cell* *70*, 650-662 e658.
- Richardson, L., Venkataraman, S., Stevenson, P., Yang, Y., Burton, N., Rao, J., Fisher, M., Baldock, R.A., Davidson, D.R., and Christiansen, J.H. (2010). EMAGE mouse embryo spatial gene expression database: 2010 update. *Nucleic Acids Res* *38*, D703-709.
- Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., *et al.* (2015). Integrative analysis of 111 reference human epigenomes. *Nature* *518*, 317-330.
- Robertson-Anderson, R.M., Wang, J., Edgcomb, S.P., Carmel, A.B., Williamson, J.R., and Millar, D.P. (2011). Single-molecule studies reveal that DEAD box protein DDX1 promotes oligomerization of HIV-1 Rev on the Rev response element. *J Mol Biol* *410*, 959-971.
- Rocak, S., and Linder, P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. *Nat Rev Mol Cell Biol* *5*, 232-241.
- Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett* *579*, 909-915.
- Roessler, E., Ward, D.E., Gaudenz, K., Belloni, E., Scherer, S.W., Donnai, D., Siegel-Bartelt, J., Tsui, L.C., and Muenke, M. (1997). Cytogenetic rearrangements involving the loss of the Sonic Hedgehog gene at 7q36 cause holoprosencephaly. *Hum Genet* *100*, 172-181.

- Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., *et al.* (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393.
- Sakabe, N.J., Savic, D., and Nobrega, M.A. (2012). Transcriptional enhancers in development and disease. *Genome Biol* **13**, 238.
- Santagati, F., and Rijli, F.M. (2003). Cranial neural crest and the building of the vertebrate head. *Nat Rev Neurosci* **4**, 806-818.
- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* **489**, 109-113.
- Sauka-Spengler, T., and Bronner-Fraser, M. (2008). A gene regulatory network orchestrates neural crest formation. *Nat Rev Mol Cell Biol* **9**, 557-568.
- Sawai, S., Shimono, A., Wakamatsu, Y., Palmes, C., Hanaoka, K., and Kondoh, H. (1993). Defects of embryonic organogenesis resulting from targeted disruption of the N-myc gene in the mouse. *Development* **117**, 1445-1455.
- Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., and Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. *Genome Res* **22**, 1748-1759.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682.
- Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., *et al.* (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat Genet* **47**, 1179-1186.
- Scholzen, T., and Gerdes, J. (2000). The Ki-67 protein: from the known and the unknown. *J Cell Physiol* **182**, 311-322.
- Schorle, H., Meier, P., Buchert, M., Jaenisch, R., and Mitchell, P.J. (1996). Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature* **381**, 235-238.
- Schwab, M., Alitalo, K., Klempnauer, K.H., Varmus, H.E., Bishop, J.M., Gilbert, F., Brodeur, G., Goldstein, M., and Trent, J. (1983). Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* **305**, 245-248.
- Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., C, H.H., Mirny, L., *et al.* (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56.
- Schwessinger, R., Suci, M.C., McGowan, S.J., Telenius, J., Taylor, S., Higgs, D.R., and Hughes, J.R. (2017). Sasquatch: predicting the impact of regulatory SNPs on transcription factor binding from cell- and tissue-specific DNase footprints. *Genome Res* **27**, 1730-1742.
- Seki, T., Yuasa, S., and Fukuda, K. (2012). Generation of induced pluripotent stem cells from a small amount of human peripheral blood using a combination of activated T cells and Sendai virus. *Nat Protoc* **7**, 718-728.
- Shi, Y., Inoue, H., Wu, J.C., and Yamanaka, S. (2017). Induced pluripotent stem cell technology: a decade of progress. *Nat Rev Drug Discov* **16**, 115-130.
- Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15**, 272-286.
- Siggins, L., and Ekwall, K. (2014). Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. *J Intern Med* **276**, 201-214.
- Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gomez-Marin, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., *et al.* (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371-375.
- Smith, E., and Shilatifard, A. (2014). Enhancer biology and enhanceropathies. *Nat Struct Mol Biol* **21**, 210-219.
- Smith, S.M., Garic, A., Berres, M.E., and Flentke, G.R. (2014a). Genomic factors that shape craniofacial outcome and neural crest vulnerability in FASD. *Front Genet* **5**, 224.
- Smith, S.M., Garic, A., Flentke, G.R., and Berres, M.E. (2014b). Neural crest development in fetal alcohol syndrome. *Birth Defects Res C Embryo Today* **102**, 210-220.
- Snider, T.N., and Mishina, Y. (2014). Cranial neural crest cell contribution to craniofacial formation, pathology, and future directions in tissue engineering. *Birth Defects Res C Embryo Today* **102**, 324-332.

## References

- Soldner, F., Stelzer, Y., Shivalila, C.S., Abraham, B.J., Latourelle, J.C., Barrasa, M.I., Goldmann, J., Myers, R.H., Young, R.A., and Jaenisch, R. (2016). Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* 533, 95-99.
- Spieler, D., Kaffe, M., Knauf, F., Bessa, J., Tena, J.J., Giesert, F., Schormair, B., Tilch, E., Lee, H., Horsch, M., *et al.* (2014). Restless legs syndrome-associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon. *Genome Res* 24, 592-603.
- Spielmann, M., Lupianez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. *Nat Rev Genet* 19, 453-467.
- Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* 13, 613-626.
- Squire, J.A., Thorner, P.S., Weitzman, S., Maggi, J.D., Dirks, P., Doyle, J., Hale, M., and Godbout, R. (1995). Co-amplification of MYCN and a DEAD box gene (DDX1) in primary neuroblastoma. *Oncogene* 10, 1417-1422.
- Stabach, P.R., Thiyagarajan, M.M., Woodfield, G.W., and Weigel, R.J. (2006). AP2alpha alters the transcriptional activity and stability of p53. *Oncogene* 25, 2148-2159.
- Stadhouders, R., Kolovos, P., Brouwer, R., Zuin, J., van den Heuvel, A., Kockx, C., Palstra, R.J., Wendt, K.S., Grosveld, F., van Ijcken, W., *et al.* (2013). Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc* 8, 509-524.
- Stanton, B.R., Perkins, A.S., Tessarollo, L., Sassoon, D.A., and Parada, L.F. (1992). Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop. *Genes Dev* 6, 2235-2247.
- Stoetzel, C., Riehm, S., Bennouna Greene, V., Pelletier, V., Vigneron, J., Leheup, B., Marion, V., Helle, S., Danse, J.M., Thibault, C., *et al.* (2009). Confirmation of TFAP2A gene involvement in branchio-oculo-facial syndrome (BOFS) and report of temporal bone anomalies. *Am J Med Genet A* 149A, 2141-2146.
- Stolfi, A., Gandhi, S., Salek, F., and Christiaen, L. (2014). Tissue-specific genome editing in Ciona embryos by CRISPR/Cas9. *Development* 141, 4115-4120.
- Streit, A., and Stern, C.D. (2001). Combined whole-mount in situ hybridization and immunohistochemistry in avian embryos. *Methods* 23, 339-344.
- Stuppia, L., Capogreco, M., Marzo, G., La Rovere, D., Antonucci, I., Gatta, V., Palka, G., Mortellaro, C., and Tete, S. (2011). Genetics of syndromic and nonsyndromic cleft lip and palate. *J Craniofac Surg* 22, 1722-1726.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545-15550.
- Sun, Y., Huang, Y., Yin, A., Pan, Y., Wang, Y., Wang, C., Du, Y., Wang, M., Lan, F., Hu, Z., *et al.* (2015). Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. *Nat Commun* 6, 6414.
- Sunden, Y., Semba, S., Suzuki, T., Okada, Y., Orba, Y., Nagashima, K., Umemura, T., and Sawa, H. (2007). Identification of DDX1 as a JC virus transcriptional control region-binding protein. *Microbiol Immunol* 51, 327-337.
- Swartling, F.J., Savov, V., Persson, A.I., Chen, J., Hackett, C.S., Northcott, P.A., Grimmer, M.R., Lau, J., Chesler, L., Perry, A., *et al.* (2012). Distinct neural stem cell populations give rise to disparate brain tumors in response to N-MYC. *Cancer Cell* 21, 601-613.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131, 861-872.
- Tanaka, K., Ikeda, N., Miyashita, K., Nuriya, H., and Hara, T. (2018). DEAD box protein DDX1 promotes colorectal tumorigenesis through transcriptional activation of the LGR5 gene. *Cancer Sci* 109, 2479-2489.
- Tanaka, K., Okamoto, S., Ishikawa, Y., Tamura, H., and Hara, T. (2009). DDX1 is required for testicular tumorigenesis, partially through the transcriptional activation of 12p stem cell genes. *Oncogene* 28, 2142-2151.
- Tapadia, M.D., Cordero, D.R., and Helms, J.A. (2005). It's all in your head: new insights into craniofacial development and deformation. *J Anat* 207, 461-477.

- Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Gottgens, B., Dougan, G., Dawson, M.A., and Lehner, P.J. (2015). GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* **348**, 1481-1485.
- Tchieu, J., Zimmer, B., Fattahi, F., Amin, S., Zeltner, N., Chen, S., and Studer, L. (2017). A Modular Platform for Differentiation of Human PSCs into All Major Ectodermal Lineages. *Cell Stem Cell* **21**, 399-410 e397.
- Tekin, M., Sirmaci, A., Yuksel-Konuk, B., Fitoz, S., and Sennaroglu, L. (2009). A complex TFAP2A allele is associated with branchio-oculo-facial syndrome and inner ear malformation in a deaf child. *Am J Med Genet A* **149A**, 427-430.
- Teszas, A., Meijer, R., Scheffer, H., Gyuris, P., Kosztolanyi, G., van Bokhoven, H., and Kellermayer, R. (2006). Expanding the clinical spectrum of MYCN-related Feingold syndrome. *Am J Med Genet A* **140**, 2254-2256.
- Thieme, F., and Ludwig, K.U. (2017). The Role of Noncoding Genetic Variation in Isolated Orofacial Clefts. *J Dent Res* **96**, 1238-1247.
- Thomas, S., Thomas, M., Wincker, P., Babarit, C., Xu, P., Speer, M.C., Munnich, A., Lyonnet, S., Vekemans, M., and Etchevers, H.C. (2008). Human neural crest cells display molecular and phenotypic hallmarks of stem cells. *Hum Mol Genet* **17**, 3411-3425.
- Thongjuea, S., Stadhouders, R., Grosveld, F.G., Soler, E., and Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res* **41**, e132.
- Thormann, V., Rothkegel, M.C., Schopflin, R., Glaser, L.V., Djuric, P., Li, N., Chung, H.R., Schwahn, K., Vingron, M., and Meijding, S.H. (2018). Genomic dissection of enhancers uncovers principles of combinatorial regulation and cell type-specific wiring of enhancer-promoter contacts. *Nucleic Acids Res* **46**, 3258.
- Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453-1465.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**, e115.
- Uslu, V.V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N.A., Marioni, J.C., and Spitz, F. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat Genet* **46**, 753-758.
- van Arensbergen, J., van Steensel, B., and Bussemaker, H.J. (2014). In search of the determinants of enhancer-promoter interaction specificity. *Trends Cell Biol* **24**, 695-702.
- van Bokhoven, H., Celli, J., van Reeuwijk, J., Rinne, T., Glaudemans, B., van Beusekom, E., Rieu, P., Newbury-Ecob, R.A., Chiang, C., and Brunner, H.G. (2005). MYCN haploinsufficiency is associated with reduced brain size and intestinal atresias in Feingold syndrome. *Nat Genet* **37**, 465-467.
- van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., *et al.* (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716-729 e727.
- Vicente-Garcia, C., Villarejo-Balcells, B., Irastorza-Azcarate, I., Naranjo, S., Acemel, R.D., Tena, J.J., Rigby, P.W.J., Devos, D.P., Gomez-Skarmeta, J.L., and Carvajal, J.J. (2017). Regulatory landscape fusion in rhabdomyosarcoma through interactions between the PAX3 promoter and FOXO1 regulatory elements. *Genome Biol* **18**, 106.
- Visel, A., Akiyama, J.A., Shoukry, M., Afzal, V., Rubin, E.M., and Pennacchio, L.A. (2009). Functional autonomy of distant-acting human enhancers. *Genomics* **93**, 509-513.
- Visser, M., Palstra, R.J., and Kayser, M. (2015). Allele-specific transcriptional regulation of IRF4 in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the IRF4 promoter. *Hum Mol Genet* **24**, 2649-2661.
- Vockley, C.M., Barrera, A., and Reddy, T.E. (2017). Decoding the role of regulatory element polymorphisms in complex disease. *Curr Opin Genet Dev* **43**, 38-45.
- Wallrath, L.L., and Elgin, S.C. (1995). Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes Dev* **9**, 1263-1277.
- Wang, W.D., Melville, D.B., Montero-Balaguer, M., Hatzopoulos, A.K., and Knapik, E.W. (2011). Tfap2a and Foxd3 regulate early steps in the development of the neural crest progenitor population. *Dev Biol* **360**, 173-185.

## References

- Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., *et al.* (2018). The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* 19, 151.
- Wang, Y., Wang, Y., Liu, D., Wang, W., Zhao, H., Wang, M., and Yin, H. (2015). Cordyceps sinensis polysaccharide inhibits PDGF-BB-induced inflammation and ROS production in human mesangial cells. *Carbohydr Polym* 125, 135-145.
- Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 40, D930-934.
- Weber, A., Imisch, P., Bergmann, E., and Christiansen, H. (2004). Coamplification of DDX1 correlates with an improved survival probability in children with MYCN-amplified human neuroblastoma. *J Clin Oncol* 22, 2681-2690.
- Wehby, G.L., and Cassell, C.H. (2010). The impact of orofacial clefts on quality of life and healthcare use and costs. *Oral Dis* 16, 3-10.
- Wei, J.S., Song, Y.K., Durinck, S., Chen, Q.R., Cheuk, A.T., Tsang, P., Zhang, Q., Thiele, C.J., Slack, A., Shohet, J., *et al.* (2008). The MYCN oncogene is a direct target of miR-34a. *Oncogene* 27, 5204-5213.
- Wei, K., Chen, J., Akrami, K., Galbraith, G.C., Lopez, I.A., and Chen, F. (2007). Neural crest cell deficiency of c-myc causes skull and hearing defects. *Genesis* 45, 382-390.
- Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., *et al.* (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* 171, 1573-1588 e1528.
- Weischenfeldt, J., Dubash, T., Drainas, A.P., Mardin, B.R., Chen, Y., Stutz, A.M., Waszak, S.M., Bosco, G., Halvorsen, A.R., Raeder, B., *et al.* (2017). Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat Genet* 49, 65-74.
- Weischenfeldt, J., Symmons, O., Spitz, F., and Korbel, J.O. (2013). Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet* 14, 125-138.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorf, L., *et al.* (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 42, D1001-1006.
- West-Mays, J.A., Zhang, J., Nottoli, T., Hagopian-Donaldson, S., Libby, D., Strissel, K.J., and Williams, T. (1999). AP-2alpha transcription factor is required for early morphogenesis of the lens vesicle. *Dev Biol* 206, 46-62.
- Wilderman, A., VanOudenhove, J., Kron, J., Noonan, J.P., and Cotney, J. (2018). High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep* 23, 1581-1597.
- Wilkie, A.O. (2003). Why study human limb malformations? *J Anat* 202, 27-35.
- Wilkie, A.O., and Morriss-Kay, G.M. (2001). Genetics of craniofacial development and malformation. *Nat Rev Genet* 2, 458-468.
- Williams, T., Admon, A., Luscher, B., and Tjian, R. (1988). Cloning and expression of AP-2, a cell-type-specific transcription factor that activates inducible enhancer elements. *Genes Dev* 2, 1557-1569.
- Williamson, J.A., Boshier, J.M., Skinner, A., Sheer, D., Williams, T., and Hurst, H.C. (1996). Chromosomal mapping of the human and mouse homologues of two new members of the AP-2 family of transcription factors. *Genomics* 35, 262-264.
- Wyszynski, D.F. (2002). *Cleft lip and palate : from origin to treatment* (Oxford: Oxford University Press).
- Xu, L., Khadijah, S., Fang, S., Wang, L., Tay, F.P., and Liu, D.X. (2010). The cellular RNA helicase DDX1 interacts with coronavirus nonstructural protein 14 and enhances viral replication. *J Virol* 84, 8571-8583.
- Yanez-Cuna, J.O., Kvon, E.Z., and Stark, A. (2013). Deciphering the transcriptional cis-regulatory code. *Trends Genet* 29, 11-22.
- Yelick, P.C., and Trainor, P.A. (2015). Ribosomopathies: Global process, tissue specific defects. *Rare Dis* 3, e1025185.
- Yu, Y., Zuo, X., He, M., Gao, J., Fu, Y., Qin, C., Meng, L., Wang, W., Song, Y., Cheng, Y., *et al.* (2017). Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat Commun* 8, 14364.
- Zeid, R., Lawlor, M.A., Poon, E., Reyes, J.M., Fulciniti, M., Lopez, M.A., Scott, T.G., Nabet, B., Erb, M.A., Winter, G.E., *et al.* (2018). Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma. *Nat Genet* 50, 515-523.
- Zepeda-Mendoza, C.J., Bardon, A., Kammin, T., Harris, D.J., Cox, H., Redin, C., Ordulu, Z., Talkowski, M.E., and Morton, C.C. (2018). Phenotypic interpretation of complex chromosomal rearrangements

- informed by nucleotide-level resolution and structural organization of chromatin. *Eur J Hum Genet* 26, 374-381.
- Zhang, J., Hagopian-Donaldson, S., Serbedzija, G., Elsemore, J., Plehn-Dujowich, D., McMahon, A.P., Flavell, R.A., and Williams, T. (1996). Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2. *Nature* 381, 238-241.
- Zhang, J.T., Weng, Z.H., Tsang, K.S., Tsang, L.L., Chan, H.C., and Jiang, X.H. (2016). MycN Is Critical for the Maintenance of Human Embryonic Stem Cell-Derived Neural Crest Stem Cells. *PLoS One* 11, e0148062.
- Zhang, W., Liu, B., Wu, W., Li, L., Broom, B.M., Basourakos, S.P., Korentzelos, D., Luan, Y., Wang, J., Yang, G., *et al.* (2018). Targeting the MYCN-PARP-DNA Damage Response Pathway in Neuroendocrine Prostate Cancer. *Clin Cancer Res* 24, 696-707.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137.
- Zhang, Z., Kim, T., Bao, M., Facchinetti, V., Jung, S.Y., Ghaffari, A.A., Qin, J., Cheng, G., and Liu, Y.J. (2011). DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. *Immunity* 34, 866-878.
- Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., *et al.* (2017). Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8, 14049.
- Zhong, W., Li, Z., Zhou, M., Xu, T., and Wang, Y. (2018). DDX1 regulates alternative splicing and insulin secretion in pancreatic beta cells. *Biochem Biophys Res Commun* 500, 751-757.
- Zimmerman, K.A., Yancopoulos, G.D., Collum, R.G., Smith, R.K., Kohl, N.E., Denis, K.A., Nau, M.M., Witte, O.N., Toran-Allerand, D., Gee, C.E., *et al.* (1986). Differential expression of myc family genes during murine development. *Nature* 319, 780-783.
- Zuccherro, T.M., Cooper, M.E., Maher, B.S., Daack-Hirsch, S., Nepomuceno, B., Ribeiro, L., Caprau, D., Christensen, K., Suzuki, Y., Machida, J., *et al.* (2004). Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. *N Engl J Med* 351, 769-780.

## FIGURE INDEX

Figure 1.1.1 Simplified schematic illustration of enhancers and their specific epigenetic signatures.....	11
Figure 1.2.1 Illustration of the multipotency of embryonal neural crest cells (NCC).....	15
Figure 1.2.2 Facial prominences of the embryonic vertebrate head. ....	16
Figure 1.2.3 Schematic illustration of parallels between neural crest differentiation in the embryo and directed <i>in vitro</i> differentiation.....	18
Figure 1.3.1 Illustration of orofacial clefting. ....	19
Figure 1.4.1 Schematic description of genome-wide association studies (GWAS).....	21
Figure 3.1.1 Facial appearance of the patient. ....	80
Figure 3.1.2 Fluorescence in situ hybridization (FISH) experiments.....	80
Figure 3.1.3 Whole-genome sequencing results around the patient's inversion breakpoints. ....	81
Figure 3.1.4 Schematic view of the chr6 heterozygous inversion in the BOFS patient. ....	82
Figure 3.1.5 The <i>in vitro</i> differentiation system for hNCC. ....	82
Figure 3.1.6 Immunofluorescence for pluripotency and neural crest markers ....	83
Figure 3.1.7 Expression of major neural crest markers analyzed by FACS.....	84
Figure 3.1.8 Immunofluorescence for neural crest markers in hNCCp2 .....	84
Figure 3.1.9 Single cell RNA sequencing in WT hNCCp2 .....	85
Figure 3.1.10 Epigenetic profile of hESC and hNCCd11 around the TFAP2A locus. ....	86
Figure 3.1.11 Close-up view of the epigenetic profile. ....	87
Figure 3.1.12 4C-seq profiles around the <i>TFAP2A</i> locus. ....	88
Figure 3.1.13 4C-seq, ATAC-seq and ChIP-seq profiles from chicken embryonic FNP. ....	89
Figure 3.1.14 <i>In vitro</i> reporter assay for <i>TFAP2A</i> enhancers in hNCC. ....	90
Figure 3.1.15 <i>In vivo</i> reporter assays for <i>TFAP2A</i> enhancers in chicken embryos.....	91
Figure 3.1.16 CRISPR/Cas9 mediated 0.4 Mb deletion within the <i>TFAP2A</i> -TAD.....	92
Figure 3.1.17 Compromised hNCC differentiation of hiPSC carrying a homozygous 0.4 Mb deletion.....	93
Figure 3.1.18 FACS analysis of neural crest markers in hNCCd11 derived from WT and 0.4 Mb deletion hiPSC lines. ....	94
Figure 3.1.19 hNCC differentiation of $\Delta$ 0.4Mb clones for up to 18 days evaluated by flow cytometry. ....	95
Figure 3.1.20 <i>TFAP2A</i> expression levels measured by RT-qPCR in WT and 0.4 Mb deletion.....	95
Figure 3.1.21 FACS analysis for <i>TFAP2A</i> expression in hNCCd11 derived from WT and 0.4 Mb deletion hiPSC lines. ....	96
Figure 3.1.22 RNA-seq in hNCCd11 derived from WT and $\Delta$ 0.4Mb clones.....	97
Figure 3.1.23 Expression levels of <i>TFAP2A</i> and immediately flanking genes measured by RNA-seq in WT and 0.4 Mb deletion hNCCd11.....	98
Figure 3.1.24 Reprogramming of BOFS patient-derived fibroblast into hiPSC. ....	99
Figure 3.1.25 RT-qPCR for major pluripotency markers in BOFS hiPSC. ....	99
Figure 3.1.26 Immunofluorescence for the human pluripotency marker TRA-1-60. ....	99
Figure 3.1.27 Confirmation of the 89 Mb inversion by PCR, Sanger sequencing and TLA. ....	100
Figure 3.1.28 Differentiation of WT and BOFS hiPSC into hNCC.....	101
Figure 3.1.29 FACS analysis for NC markers in hNCCd11. ....	101

Figure 3.1.30 FACS analysis for NC markers in hNCCp2. ....	102
Figure 3.1.31 Alcian Blue staining of chondrocytes derived from hNCCp2. ....	103
Figure 3.1.32 Immunofluorescence for smooth muscle and neuronal markers. ....	103
Figure 3.1.33 Reduced <i>TFAP2A</i> expression levels in BOFS hNCC evaluated by RT-qPCR and FACS. ....	104
Figure 3.1.34 Allele-specific expression analysis of <i>TFAP2A</i> by quantitative genotyping of rs1675414. ....	105
Figure 3.1.35 Immunofluorescence for <i>TFAP2A</i> in WT and BOFS hNCCp2. ....	106
Figure 3.1.36 Single cell RNA sequencing in WT and BOFS hNCCp2. ....	107
Figure 3.1.37 Allele-specific 4C seq experiments in BOFS hNCCp2. ....	108
Figure 3.1.38 Close-up view of the 4C-seq profiles centered on the 6p24.3 inversion breakpoint. ....	109
Figure 3.1.39 ChIP-seq profiles for histone modifications in WT and BOFS hNCC around the inversion breakpoints. ....	109
Figure 3.1.40 4C-seq profiles aligned with H3K27me3 ChIP-seq profiles illustrate new physical interactions at H3K27me3 bound locations. ....	110
Figure 3.1.41 ChIP-seq profiles and quantification of read numbers at the rs1675414 SNP. ....	111
Figure 3.1.42 Schematic diagram of how enhancer disconnection may cause loss of <i>TFAP2A</i> expression. ....	112
Figure 3.1.43 Schematic diagram of how enhancer adoption may cause mis- or overexpression of 6q16 genes. ....	113
Figure 3.1.44 4C-seq profiles of WT and BOFS hNCCp2 around the 6p24.3 and the 6q16.2 inversion breakpoints. ....	114
Figure 3.1.45 Differential gene expression analysis by RNA-seq in WT and BOFS hNCCp2. ....	115
Figure 3.1.46 Allele-specific analysis of <i>TFAP2A</i> expression using RNA-seq data of WT and BOFS hNCCp2. ....	116
Figure 3.1.47 Functional annotation of differentially expressed genes between WT and BOFS hNCC. ....	117
Figure 3.1.48 RT-qPCR validation of selected downregulated genes in BOFS hNCCp2. ....	118
Figure 3.1.49 Scratch assays with WT and BOFS hNCCp2. ....	118
Figure 3.1.50 Quantification of the scratch assays in WT and BOFS hNCCp2. ....	119
Figure 3.1.51 Proliferation assay and expression of proliferation markers in WT and BOFS hNCCp2. ....	119
Figure 3.1.52 Heat maps of all <i>TFAP2A</i> and H3K27ac ChIP-seq peaks in WT or BOFS hNCCp2. ....	120
Figure 3.1.53 Similar binding pattern of <i>TFAP2A</i> ChIP-seq peaks in WT and BOFS hNCC. ....	121
Figure 3.1.54 ChIP-seq profiles of <i>TFAP2A</i> and histone marks at enhancers and distance to nearest TSS in WT and BOFS hNCC. ....	122
Figure 3.1.55 Additional loci displaying differential <i>TFAP2A</i> binding an H3K27ac levels between WT and BOFS hNCCp2. ....	123
Figure 3.1.56 Motif analysis of differentially bound <i>TFAP2A</i> sites. ....	123
Figure 3.1.57 Relationship between <i>TFAP2A</i> binding and transcriptional changes in BOFS hNCC. ....	124
Figure 3.1.58 Bulk RNA-seq analysis of scRNA-seq data. ....	125
Figure 3.1.59 Correlation between the expression levels of <i>TFAP2A</i> and the genes downregulated in patient hNCCp2 within single cells. ....	126
Figure 3.2.1 Diagram of selection criteria for candidate enhancers. ....	128
Figure 3.2.2. Epigenetic profile of human neural crest cells and chicken facial tissue at the 2p24.2 locus. ....	129
Figure 3.2.3 Epigenetic profile of hNCC and chicken facial tissue of the TAD enclosing the Enh2p24.2 and both <i>MYCN</i> and <i>DDX1</i> but not <i>FAM49A</i> . ....	131
Figure 3.2.4 4C-seq profiles at the TAD enclosing <i>DDX1</i> , <i>MYCN</i> and Enh2p24.2. ....	132
Figure 3.2.5 <i>In vitro</i> reporter assay for the Enh2p24.2 in hNCC. ....	134
Figure 3.2.6 <i>In vivo</i> reporter assay for the Enh2p24.2 in chicken embryos. ....	135
Figure 3.2.7 Generation of a heterozygous deletion of the candidate enhancer Enh2p24.2 using CRISPR/Cas9. ....	136

## Figure Index

Figure 3.2.8 Expression levels of the putative target genes of Enh2p24.2 measured by RT-qPCR in WT cells and cells with a heterozygous enhancer deletion. ....	137
Figure 3.2.9 Transcription levels of <i>DDX1</i> , <i>MYCN</i> and <i>FAM49A</i> in human ESC, NEC and NCC. ....	138
Figure 3.2.10 Expression pattern of <i>DDX1</i> and <i>MYCN</i> in chicken embryos. ....	139
Figure 3.2.11 Schematic diagram depicting the CRISPR/Cas9 strategy used to target the candidate genes <i>in vivo</i> . ....	141
Figure 3.2.12 Genetic engineering of chicken neural crest cells by CRISPR/Cas9 shows a distinct facial phenotype in <i>MYCN</i> and <i>DDX1</i> targeted embryos. ....	142
Figure 3.2.13 CRISPR/Cas9 mediated deletions within the chicken <i>DDX1</i> gene. ....	143
Figure 3.2.14 CRISPR/Cas9 mediated deletion within the chicken <i>FAM49A</i> gene. ....	144
Figure 3.2.15 Rescue experiments for <i>MYCN</i> and <i>DDX1</i> in CRISPR/Cas9 engineered chicken embryos. ....	145
Figure 3.2.16 Overexpression of <i>FAM49A</i> in neural crest cells of chicken embryos ....	146
Figure 3.2.17 Generation of <i>MYCN</i> <sup>-/-</sup> hiPSC using CRISPR/Cas9. ....	147
Figure 3.2.18 Western blot analysis of <i>MYCN</i> in <i>MYCN</i> <sup>-/-</sup> hiPSC. ....	147
Figure 3.2.19 CRISPR/Cas9 mediated deletion of the human <i>DDX1</i> gene. ....	148
Figure 3.2.20 Growth curve showing the number of cells over time for <i>DDX1</i> <sup>-/-::TetON-DDX1</sup> and WT hiPSC. ....	149
Figure 3.2.21 Growth curve showing number of cells over time for <i>DDX1</i> <sup>+/-</sup> and WT hiPSC. ....	149
Figure 3.2.22 Growth curve showing number of cells over time for <i>MYCN</i> <sup>-/-</sup> and WT hiPSC. ....	150
Figure 3.2.23 Compromised hNCC differentiation of hiPSC carrying a heterozygous <i>DDX1</i> deletion. ....	151
Figure 3.2.24 hNCC differentiation of <i>DDX1</i> <sup>-/-::TetON-DDX1</sup> hiPSC with and without DOX. ....	152
Figure 3.2.25 <i>DDX1</i> and <i>TFAP2A</i> expression levels measured by RT-qPCR in cells with <i>DDX1</i> deletions. ....	153
Figure 3.2.26 Western blot analysis of <i>DDX1</i> in hNCCp2 derived from <i>DDX1</i> <sup>+/-</sup> and <i>DDX1</i> <sup>-/-::TetON-DDX1</sup> hiPSC. ....	153
Figure 3.2.27 Representative photos of <i>DDX1</i> <sup>-/-::TetON-DDX1</sup> and WT hNCCp2 ....	154
Figure 3.2.28 Growth curve showing the number of cells over time for <i>DDX1</i> <sup>-/-::TetON-DDX1</sup> and WT hNCCp2. ....	155
Figure 3.2.29 Compromised hNCC differentiation of <i>MYCN</i> <sup>-/-</sup> hiPSC. ....	156
Figure 3.2.30 Functional annotation of genes significantly upregulated in <i>DDX</i> <sup>+/-</sup> hNCCp2 compared to WT hNCCp2. ....	157
Figure 3.2.31 Immunofluorescence for <i>DDX1</i> in hiPSC and hNCC. ....	158
Figure 3.2.32 Correlation between <i>DDX1</i> binding to promoters and expression of the corresponding genes. ....	159
Figure 3.2.33 <i>DDX1</i> ChIP-seq profiles and RNA-seq read counts at three example loci. ....	160
Figure 4.1.1 Schematic diagram of the <i>TFAP2A</i> -TAD. ....	163
Figure 4.6.1 Schematic illustration of the OFC associated haplotype at 2p24.2. ....	173

## TABLE INDEX

Table 2.1 Equipment .....	30
Table 2.2 Chemicals.....	31
Table 2.3 Molecular Biology Reagents and Enzymes .....	32
Table 2.4 Commercial Kits.....	34
Table 2.5 Tissue culture reagents and commercial media.....	35
Table 2.6 Composition of Tissue Culture Media .....	36
Table 2.7 PCR primers for 89 Mb inversion genotyping .....	37
Table 2.8 Clinical features of the BOFS patient.....	43
Table 2.9 Primers used for TLA.....	44
Table 2.10 Composition of QuickExtract DNA Extraction Solution .....	46
Table 2.11 PCR cycling program.....	47
Table 2.12 Buffers and Solutions for Agarose Gel Electrophoresis .....	47
Table 2.13 Primer Sequences for RT-qPCR.....	48
Table 2.14 Primer Sequences for TFAP2A SNP genotyping.....	49
Table 2.15 Plasmids .....	50
Table 2.16 Primer sequence for enhancer reporter assays .....	52
Table 2.17 Antibodies for ChIP .....	54
Table 2.18 Buffers for ChIP-seq .....	54
Table 2.19 Primary antibodies for Immunofluorescence.....	55
Table 2.20 Secondary Antibodies for Immunofluorescence.....	55
Table 2.21 Composition of Buffers for Immunofluorescence .....	56
Table 2.22 Primary/Conjugated Antibodies for Flow Cytometry .....	56
Table 2.23 Secondary Antibodies for Flow Cytometry .....	57
Table 2.24 Composition of Buffers for Flow Cytometry .....	57
Table 2.25 Composition of Buffers for Western Blot.....	58
Table 2.26 Antibodies for Western Blot .....	58
Table 2.27 Guide RNA sequences for CRISPR/Cas9 .....	59
Table 2.28 Primers for detection of deletions introduced by CRISPR/Cas9 .....	61
Table 2.29 Primer sequence for rescue experiments .....	63
Table 2.30 Primers for <i>in situ</i> hybridization.....	66
Table 2.31 Composition of Buffers for WISH.....	66
Table 2.32 Primer sequence for overexpression of human <i>FAM49A</i> in chicken embryos.....	67
Table 2.33 Composition of Buffers for 4C.....	69
Table 2.34 Primers for inverted PCR (4C) with Illumina adaptors .....	69
Table 2.35 Composition of the ATAC-seq Lysis Buffer.....	71
Table 2.36 Software and Algorithms for Analysis of Next Generation Sequencing Data.....	77
Table 3.1 Reported OFC associated GWAS SNPs within the 2p24.2 locus.....	130

**LIST OF ABBREVIATIONS**

3C	Chromatin conformation capture
3D	Three dimensional
4C	Circular chromatin conformation capture
ATAC	Assay for transposase accessible chromatin
ATP	Adenosinotriphosphat
BAC	Bacterial artificial chromosome
BCA	Bicinchoninic acid
bHLHZ	basic Helix-Loop-Helix-Zipper
BMP	Bone morphogenic protein
BOFS	Branchiooculofacial syndrome
bp	base pair
BSA	Bovine serum albumin
CAA	2-Chloroacetamide
CD-CV	Common disease-common variant
CD-RV	Common disease-rare variant
cDNA	complementary DNA
CFSE	Carboxyfluorescein succinimidyl ester
CGH	Comparative genomic hybridization
ChIP	Chromatin Immunoprecipitation
CL/P	Cleft lip with and without cleft palate
CLIP	Crosslinking and Immunoprecipitation
CPO	Cleft palate only
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSR	Class Switch Recombination
CTCF	CCCTC-binding factor
DAPI	4',6-Diamidin-2-phenylindol
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Desoxyribonucleic acid
dNTP	Desoxyribonucleotide triphosphate
DOX	Doxycycline
DTT	Dithiothreitol
DSB	Double strand break
EB	Embryoid body
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial to mesenchymal transition
ESC	embryonic stem cell
FACS	Fluorescence activated cell sorting
FASD	Fetal alcohol spectrum disorder
FBS	Fetal bovine serum
FC	Fold change
FD	Fluorescein dilution factor
FDR	False discovery rate
FGF	Fibroblast growth factor
FISH	Fluorescence <i>in situ</i> hybridization
FNP	Frontonasal prominence
FPKM	Fragments per kilobase of exon model per million reads mapped
fwd	forward
GEISHA	Gallus gallus expression <i>in situ</i> hybridization analysis
GEM	Gel bead in emulsion
GEO	Gene expression omnibus
GFP	Green fluorescent protein
gRNA	guide RNA
GSEA	Gene set enrichment analysis
GWAS	Genome wide association study
H3K4me1	Histone three lysine four monomethylation
H3K4me2	Histone three lysine four dimethylation
H3K4me3	Histone three lysine four trimethylation
H3K9me3	Histone three lysine nine trimethylation
H3K27ac	Histone three lysine twenty-seven acetylation

HAT	Histone acetyltransferase
hESC	Human embryonic stem cell/s
HH	Hamburger Hamilton
hNCC	human neural crest cell/s
hiPSC	human induced pluripotent stem cell/s
HRP	Horseradish peroxidase
ICD	International statistical classification of diseases
Ig	Immunoglobulin
IF	Immunofluorescence
INV	Inversion
IP	Immunoprecipitation
KO	Knockout
KSR	Knockout serum replacement
LD	Linkage disequilibrium
LIF	Leukemia inhibitory factor
LNP	Lateral nasal prominences
Mb	Megabase
MFI	Mean fluorescein intensity
mio.	million
MNP	Mandibular prominences
MRI	Magnetic resonance imaging
ms	micro seconds
MXP	Maxillary prominences
NBT	Nitro blue tetrazolium chloride
NEB	New England Biolabs
NER	Nucleotide excision repair
NC	Neural crest
NCC	Neural crest cell/s
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
OFC	Orofacial cleft
PAC	P1-derived artificial chromosome
PAE	Prenatal alcohol exposure
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PEV	Positional effect variegation
PCR	Polymerase chain reaction
PerCP	Peridinin chlorophyll protein
rev	reverse
RFP	Red fluorescent protein
RIPA	Radio immunoprecipitation assay buffer
RNA	Ribonucleic acid
RPGC	Reads per genomic content
RPM	Reads per million
rpm	Rounds per minute
RT-qPCR	Real-time quantitative PCR
RT	Room temperature
SDS	sodium dodecyl sulfate
SMC	Smooth muscle cell/s
SNP	Single nucleotide polymorphism
SV	Structural variant
TAD	Topologically associated domain
TBS	Tris-buffered saline
TF	Transcription factor
TLA	Targeted locus amplification
tSNE	t-distributed stochastic neighbor embedding
TSS	Transcription start site
TES	Transcription end site
UMI	Unique molecular identifier
UTR	Untranslated region
WISH	Whole mount <i>in situ</i> hybridization
WT	wildtype

## **ACKNOWLEDGEMENTS**

First of all, I would like to thank my supervisor, Alvaro Rada-Iglesias, who has the incomparable ability to motivate his students with his own burning enthusiasm for science. He gave me the freedom to explore the exciting world of research on my own while giving me guidance whenever I needed it. I can safely say that I would not be where I am today if Alvaro did not believe in me.

This work would not have been possible without the talent of Milos Nikolic, Giuliano Crispantu and Victor Sanchez when it comes to computational analysis as well as Rizwan Rehimi and his amazing skills with chicken embryos.

I was fortunate to closely work with Magdalena Lausch for most of my PhD and I am extremely thankful for all the things she taught me.

I would also like to extend my thanks to Samantha Brugman at the Cincinnati Children's Hospital, who gave me the unique opportunity to perform some crucial experiments in her lab.

I was privileged to work in a talented group of researchers, in particular, Rodrigo Osorno, who constantly inspired me with exciting discussions, Yulia Kargopolova and Sara de la Cruz Molina, who were always willing to help and showed me so much, and, not to forget, Tore Bleckwehl and Tomas Pachano – it was an absolute pleasure to work alongside you guys.

I enjoyed the support of my mentors, Brunhilde Wirth and Michael Nothnagel, who guided me throughout my graduate studies, and Mirka Uhlirova who kindly agreed to complete my examination board.

Last but not least, I am forever grateful for the understanding of my friends and family, who too often had to wait for me because I was still busy in the lab. I promise, this will not change.

## EIDESSTATTLICHE ERKLÄRUNG

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit, einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat, dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist, sowie dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Dr. Alvaro Rada-Iglesias, Prof. Dr. Brunhilde Wirth, Prof. Dr. Michael Nothnagel und Prof. Dr. Mirka Uhlirova betreut worden.

Köln im September 2019

Michaela Bartusel

Nachfolgende Teilpublikation liegt vor:

Magdalena Lausch\*, Michaela Bartusel\*, Rizwan Rehimi, Hafiza Alirzayeva, Agathi Karaolidou, Giuliano Crispantu, Peter Zentis, Milos Nikolic, Tore Bleckwehl, Petros Kolovos, Wilfried F.J. van Ijcken, Tomo Saric, Katrin Koehler, Peter Frommolt, Katherine Lachlan, Julia Baptista, Alvaro Rada-Iglesias (2019) „Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs“, *Cell Stem Cell*, doi:10.1016/j.stem.2019.03.004

\*Diese Autoren tragen gleichermaßen zu dieser Arbeit bei.