

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE BIOCÊNCIAS

CAMILA MANSO MUSSO

Investigação do mecanismo patogênico responsável pela
síndrome Richieri-Costa-Pereira

*Investigating the pathogenic mechanism underlying Richieri-
Costa-Pereira syndrome*

São Paulo
2019

CAMILA MANSO MUSSO

Investigação do mecanismo patogênico responsável pela
síndrome Richieri-Costa-Pereira

*Investigating the pathogenic mechanism underlying Richieri-
Costa-Pereira syndrome*

Tese apresentada ao Instituto de Biociências da
Universidade de São Paulo, para a obtenção de
Título de Doutor em Ciências, na Área de
Biologia/Genética.

Orientadora: Profa. Dra. Maria Rita dos Santos e
Passos-Bueno

São Paulo
2019

FICHA CATALOGRÁFICA

Manso Musso, Camila

Investigação do mecanismo patogênico responsável pela síndrome Richieri-Costa-Pereira / Camila Manso Musso; orientadora Maria Rita dos Santos e Passos-Bueno. -- São Paulo, 2019.

196 f.

Tese (Doutorado) - Instituto de Biociências da Universidade de São Paulo, Departamento de Genética e Biologia Evolutiva.

1. Síndrome Richieri-Costa-Pereira . 2. iPSCs. 3. Células de crista neural . 4. Modelagem in vitro . 5. Modelo animal . I. dos Santos e Passos-Bueno, Maria Rita , orient. II. Título.

Comissão Julgadora

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Profa. Dra. Maria Rita dos Santos e Passos-Bueno
Orientadora

À memória de Raúl Marcel González Garcia, amigo e professor.
A ele, que despertou em mim a paixão pelas Ciências Biológicas,
plantando a semente da profissional que me tornei.

“A vida é assim:
esquenta e esfria,
aperta e daí afrouxa,
sossega e depois
desinquieta.
O que ela quer da gente
é coragem.”

Guimarães Rosa

AGRADECIMENTOS

Agradeço a Deus pelas bênçãos. Ao meu pai Eder agradeço pelo apoio, sábios conselhos, paciência e cuidado, e à minha mãe Maurília pelo colo, dedicação e exemplo de perseverança e doçura. A eles, toda a minha gratidão por compartilharem dos meus sonhos e pelo esforço a fim de me oferecer as melhores oportunidades sempre com um sorriso no rosto. Muito obrigada por serem a maior fonte de amor puro e incondicional.

Agradeço aos meus irmãos Mariane e Murilo pelos abraços apertados e ao Boris por ser a melhor companhia. Ao meu avô Manso (em memória) e à minha Avó Landa pelas orações e gentilezas. Agradeço à minha família pela torcida constante e ao Luiz pelo carinho e incentivo. Aos amigos Belinha e Lucas, agradeço pela alegria de viver e conviver.

Aos meus queridos amigos de Baependi, Ana Rita, Amanda, Bia, Daiane, Gabi, Gui, Lari, Rafael e Yasmin, e às amigas Annie e Erika, agradeço pelas boas vibrações e pelo aconchego dos encontros, mensagens e conversas. Muito obrigada por tanto amor!

Deixo também meu agradecimento a todos os professores e orientadores que tive nesta longa jornada, do jardim de infância ao doutorado, por compartilharem comigo o conhecimento e por acreditarem em mim. Agradeço em especial à Dra. Maria Rita pela orientação, dedicação e pelas oportunidades. Agradeço à Dra. Debra Silver por me receber em seu laboratório na Duke University e a todos os amigos de Durham, em especial Aaron, Andy, Bianca, Fernando, Ileana, Laura, Mariah e Vanessa por me acolherem de braços abertos.

Não poderia deixar de agradecer à Naila, Simone e a todos os funcionários do CEGH-CEL, técnicos, médicos assim como aos pacientes e familiares que colaboraram com esse trabalho.

E finalmente agradeço aos amigos do laboratório 200 e vizinhos por todas as risadas e agradável convivência. Em especial, agradeço aos amigos Àgatha, Carol, Clarice, Cláudia, Dani, Duda, Elisa, Juninha, Letícia, Luciano, Pri, Pontinho, Sofia e Vanessa por serem pessoas especiais e tornarem o laboratório um lugar gostoso de se trabalhar. Muito obrigada por vocês estarem presentes tanto nos momentos de alegria quanto nos momentos de desânimo dessa montanha russa chamada doutorado. Vocês fizeram essa aventura infinitamente mais fácil e divertida. Um

agradecimento super especial ao Gerson que tanto me orientou e ensinou, tornando esse trabalho possível.

Agradeço com todo o meu amor a todos que de alguma forma contribuíram para a minha formação profissional e para a finalização desse ciclo. Que sorte a minha ter vocês!

NOTAS

Esta tese de doutorado compreende um trabalho inédito desenvolvido entre 2015 e 2019 no Laboratório de Genética do Desenvolvimento, gerido pela Dra. Maria Rita Passos-Bueno, do Centro de Estudos do Genoma Humano e Células Tronco, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil.

A tese foi redigida no modelo de artigos/capítulos, no idioma inglês. Dois artigos publicados (capítulos II e III) e dois artigos não publicados (capítulos IV e V) foram incluídos no corpo principal da tese. Esses capítulos estão precedidos por uma introdução geral (capítulos I) e seguidos de uma discussão geral e conclusões (capítulos VI). Publicações em co-autoria não relacionadas ao tema principal encontram-se sumarizadas nos apêndices.

Este trabalho foi importante para estabelecer uma colaboração internacional com a Dra. Debra Silver da Duke University, Carolina do Norte, EUA. O desenvolvimento da pesquisa envolveu um período de doutorado sanduíche, durante o qual alguns experimentos foram realizados no laboratório estrangeiro.

O projeto que resultou na presente tese foi cadastrado na Plataforma Brasil e contou com o parecer substanciado do Comitê de Ética em Pesquisa do Instituto de Biociências da Universidade de São Paulo (número 1.463.852/2016).

Este trabalho recebeu apoio financeiro da CAPES, CNPq e Company of Biologists.

NOTES

This doctoral thesis comprises an original work conceived and developed between 2015 and 2019 at the Laboratory of Developmental Genetics headed by Dr. Maria Rita Passos-Bueno within the Human Genome and Stem Cell Research Center, Institute of Biosciences, University of São Paulo, São Paulo, Brazil.

The thesis was organized into scientific articles that are presented as the core chapters. Two published articles (chapters II and III) and two unpublished articles (chapters IV and V) were included in the main body of the thesis. These chapters are preceded by a general introduction (chapter I) and followed by the general discussion and conclusions (chapter VI). Additional co-authored publications not related to the main subject of the thesis were assigned to an appendix section.

This work was important for establishing an international collaboration with Dr. Debra Silver from Duke University, North Carolina, USA. The development of this doctoral research involved an internship period, during which several experiments were carried out in the Silver lab.

The project that resulted in this thesis was registered on the Plataforma Brasil and approved by the Ethics Committee of the Institute of Biosciences at University of São Paulo (number 1.463.852/2016).

This work was financed by CAPES, CNPq and Company of Biologists.

LIST OF ABBREVIATIONS

A3SS	Alternative 3' splice sites
A5SS	Alternative 5' splice sites
ANOVA	Analysis of variance
AS	Alternative splicing
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CEGH-CEL	Centro de Estudos do Genoma Humano e Células Tronco
CEPID	Centro de Pesquisa, Inovação e Difusão
CFM	Craniofacial malformations
CNCC	Cranial neural crest cells
CNPEM	Centro Nacional de Pesquisa em Energia e Materiais
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CNS	Central nervous system
CNV	Copy number variations
CRISPR	Clustered regularly interspaced short palindromic repeats
DAVID	Database for Annotation, Visualization and Integrated Discovery
DEG	Differentially expressed genes
DEP	Differentially expressed proteins
DNA	Deoxyribonucleic acid
E	Embryonic stage
ECM	Extracellular matrix
EEP	Exclusively expressed proteins
EJC	Exon junction complex
EMT	Epithelial to mesenchymal transition
ENS	Enteric nervous system
ESC	Embryonic stem cells
FA	Focal adhesion
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FDR	False discovery rate
Fig.	Figure
g	Relative centrifuge force
GO	Gene ontology
GSEA	Gene set enrichment analysis
hiPSCs	Human induced pluripotent stem cells
hNCCs	Human neural crest cells
hnRNP	Heterogeneous nuclear ribonucleoproteins
HRAC	Hospital de Reabilitação de Anomalias Craniofaciais
IACUC	Institutional Animal Care and Use Committee
IB	Instituto de Biociências
IGV	Integrative Genomics Viewer

iNCCs	iPSC-derived neural crest cells
iPSCs	Induced pluripotent stem cells
JAX	The Jackson Laboratory
KEGG	Kyoto Encyclopedia of Genes and Genomes
LCMS	Liquid chromatography mass spectrometry
LFQ	Label-free quantification
LNBio	Laboratório Nacional de Biociências
MC	Meckel's cartilage
min	Minutes
mNCCs	Mouse neural crest cells
mRNA	Messenger RNA
MXE	Mutually exclusive exons
n	Sample size
NBC	Neural plate border-like cells
NCCs	Neural crest cells
NCP	Neurocristopathies
NIH	National Institutes of Health
NMD	Nonsense-mediated RNA decay
nMSCs	NCC-derived mesenchymal stem cells
NPB	Neural plate border
NSCs	Neural stem cells
nt	Nucleotides
OMIM	Online mendelian inheritance in man
PCR	Polymerase chain reaction
PNS	Peripheral nervous system
RCPS	Richieri-Costa-Pereira syndrome
RI	Retained introns
RNA	Ribonucleic acid
RNAi	RNA interference
rRNA	Ribosomal RNA
RT	Room temperature
RT-qPCR	Real-time quantitative polymerase chain reaction
SE	Skipped exons
SEM	Standard error of the mean
SNP	Single nucleotide polymorphism
SRCP	Richieri-Costa-Pereira syndrome
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TALEN	Transcription activator-like effector nuclease
USP	Universidade de São Paulo
UTR	Untranslated region
ZFN	Zinc finger nuclease

INDEX

CHAPTER I

General Introduction	12
Objectives	32

CHAPTER II

Complexity of the 5' Untranslated Region of EIF4A3, a Critical Factor for Craniofacial and Neural Development

Main Text.....	33
Supplementary Material.....	52

CHAPTER III

EIF4A3 Deficient Human iPSCs and Mouse Models Demonstrate Neural Crest Defects that Underlie Richieri-Costa-Pereira Syndrome

Main Text.....	56
Supplementary Material.....	91

CHAPTER IV

RNA Sequencing of EIF4A3 Deficient Neural Crest Cells Reveals Novel Pathways Functionally Linked to RCPS Pathogenesis

Main Text.....	106
Supplementary Material.....	132

CHAPTER V

Recapitulation of Neural Crest Specification and EMT via Induction from Neural Plate Border-Like Cells

Main Text.....	138
Supplementary Material.....	163

CHAPTER VI

General Discussion	167
Conclusions.....	178

CHAPTER VII

Abstract 179

Resumo..... 181

APPENDIX

Additional Publications..... 183

CHAPTER I

GENERAL INTRODUCTION

1. Vertebrate Craniofacial Development

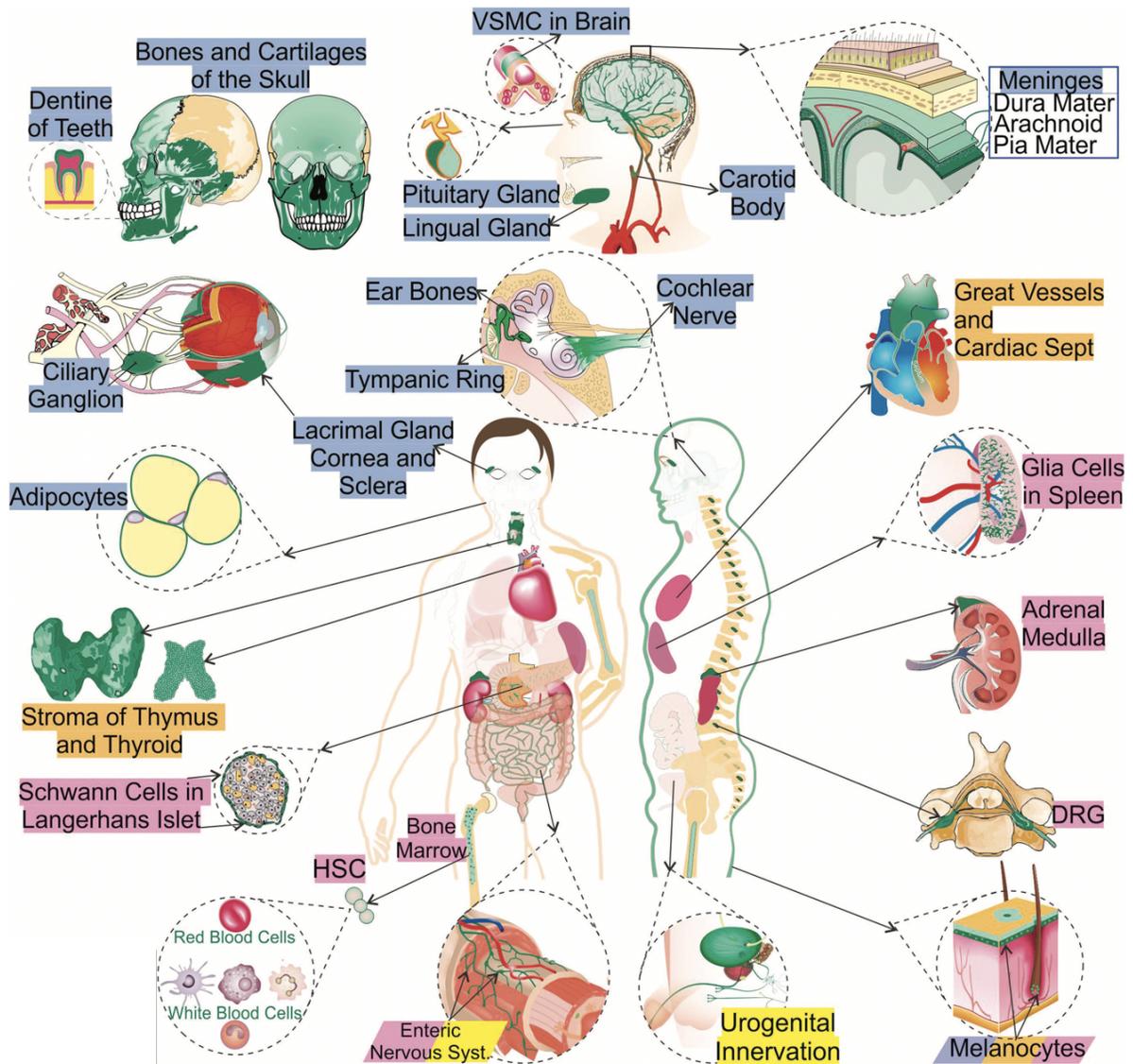
A defining feature of the vertebrate lineage is the presence of a head that is both supported and protected by a robust skeleton (Kaucka and Adameyko 2019; Jandzik et al. 2015). Development of the head structure involved the acquisition of neural crest cells and ectodermal placodes, which was a crucial step and powerful propellant for vertebrate evolution (Gans and Northcutt 1983; Horie et al. 2018).

The head represents the most complex part of the body and its morphogenesis requires a sequence of spatio-temporally coordinated events. Craniofacial development encompasses an umpteen number of signaling molecules mainly belonging to the Wnt, fibroblast growth factor (FGF) and bone morphogenetic protein (BMP) pathways, which precisely coordinate the assembly of components from distinct sources including the ectoderm, endoderm and mesoderm as well as neural crest cells (Adameyko and Fried 2016).

1.1 Neural Crest Cells

Neural crest cells (NCCs) comprise a transient, multipotent population of migratory cells that originates at the border of the neural plate in the developing embryo (Bronner and LeDouarin 2012). They contribute to the formation of different systems, including the peripheral nervous system (PNS), enteric nervous system (ENS), and cardiovascular system while also playing a major role in the development of forebrain and midbrain structures. NCCs are capable

of giving rise to cell types as diverse as neurons, glia, melanocytes, endocrine cells, muscle, bone, cartilage and connective tissues (Fig. 1). Because of their striking plasticity, unique properties and importance in development, the neural crest has been dubbed the “fourth germ layer” of the vertebrate embryo (Cordero et al. 2011; Trainor 2013, Zhang et al. 2014).



NC Domains: Cranial NCC Cardiac NCC Trunk NCC Sacral NCC

Figure 1. Schematic representation of the wide diversity of neural crest derivatives. The tissue, cell and organ names are highlighted with different colors according to its domain origin. Adapted from Vega-Lopez et al. 2018.

During late gastrulation and primary neurulation, the neural crest arises between the neural ectoderm that generates the central nervous system (CNS) and the adjacent non-neural ectoderm that generates the epidermis, a region known as neural plate border (NPB). The edges of the neural plate move upward to form the paired neural folds, which are brought together and fuse dorsally to form the neural tube in secondary neurulation. The neural crest cells reside at the dorsalmost portion of this hollow tube (Scott 2010).

After induction and specification, NCCs undergo a process designated as the epithelial to mesenchymal transition (EMT) and leaves the neuroepithelium territory through a delaminating process. The EMT orchestrates the transition from an epithelial to a mesenchymal cell phenotype, which includes enhanced migratory capacity, invasiveness and elevated resistance to apoptosis. This biological process encompasses a series of events regulated by a tightly controlled network requiring sequential activation and/or repression by molecular effectors (Theveneau and Mayor 2012).

After EMT, NCCs migrate along specific routes throughout the embryo. Once they arrive at their destination, they undergo intense proliferation and differentiation as derivative structures begin to take shape (Theveneau and Mayor 2012).

1.2 Cranial Neural Crest Cells

The NCCs can be divided into four domains i.e. cranial, cardiac, trunk, and vagal/sacral based on their axial level. The NCCs' fates vary depending upon their domain, where they migrate to and settle. Only the cranial or cephalic neural crest cells (CNCCs) have the potential to differentiate into cartilage and bone (Scott 2010).

There are differences in NCC development along the rostrocaudal axis of the embryo as well as between species. In mammals, the CNCCs delaminate all at once when the neural tube is still open, while in birds it coincides with the fusion of neural folds. Nevertheless, the trunk NCCs delaminate progressively after the neural tube closure in all

animals. After delamination, they migrate as a continuous wave moving into distinct streams (Bronner and LeDouarin 2012; Theveneau and Mayor 2012).

CNCCs can be further subdivided into forebrain-, midbrain- and hindbrain-derived populations. The rostral CNCCs originate the frontonasal process that generates the bones of the face, whereas the posterior cells migrate into the pharyngeal arches to produce the craniofacial mesenchyme. The cells that colonize the first (mandibular) pharyngeal arch give rise to the jawbones and middle ear bones. On the other hand, the cells from the second, third, fourth and sixth pharyngeal arches generate hyoid cartilage of the neck, thymus, parathyroid and thyroid glands (Cordero et al. 2011; Santagati and Rijli 2003) (Fig. 2B and C).

NCCs have specific segment identities according to their position along the anterior-posterior axis specified by a combination of *Hox* genes (Parker et al. 2018). However, CNCCs from the forebrain, midbrain and anterior hindbrain (rhombomeres 1 and 2) do not express any *Hox* genes (Fig. 2A). The patterning of CNCCs arising from *Hox*-negative regions are mainly controlled by distal-less (*Dlx*) genes (Depew et al. 2002).

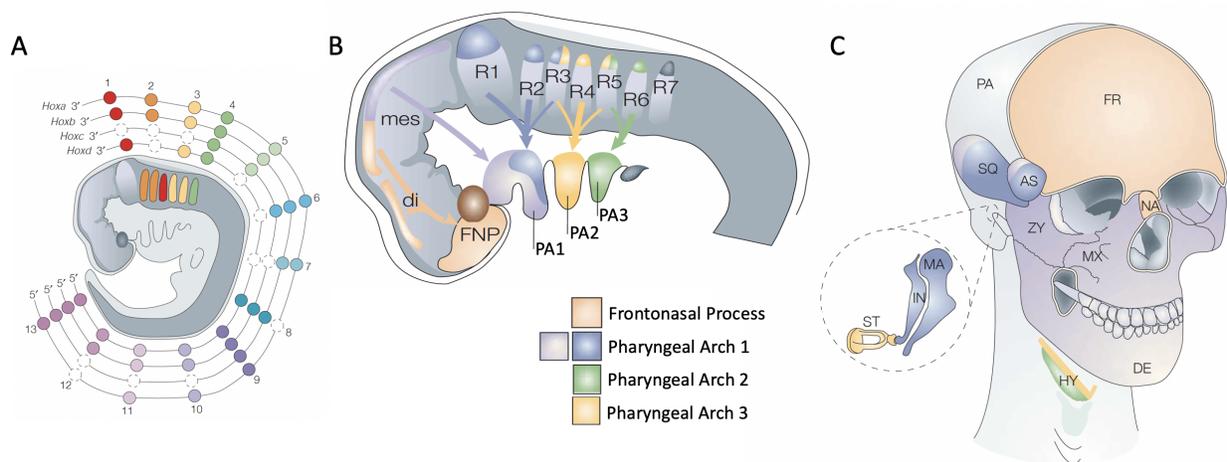


Figure 2. Scheme of the fate of cranial neural crest cells (CNCCs). **(A)** Representation of the Homeobox genes by color code, showing the four clusters and 13 paralogous groups present in humans. The physical position of each *Hox* in the cluster correlates with its expression along the anteroposterior axis of the embryo. **(B)** The CNCC migratory pathways and subsequent colonization of the frontonasal process and pharyngeal arches are shown in the embryo. **(C)** The skull indicates comparative contributions of CNCC populations to cranial skeletal elements. AS, alisphenoid; DE, dentary; di, diencephalon; FNP, frontonasal process; FR, frontal; HY, hyoid; IN, incus; MA, malleus; mes, mesencephalon; MX, maxillary; NA, nasal; PA,

parietal; PA1-P3, pharyngeal arches 1-3; R1-R7, rhombomeres 1-7; SQ, squamosal; ST, stapes; ZY, zygomatic. Adapted from Santagati and Rijli 2003.

Regarding the development of the mandible, after formation of the mandibular arch the CNCC-derived mesenchyme undergoes condensation. Subsequently the mesenchymal cells differentiate into chondrocytes and organize into a bilateral rod-shaped cartilage, i.e. Meckel's cartilage (MC), surrounded by the perichondrium. This cartilage comprises a primordial anlage of the skeletal element of the mandible, functioning as a morphogenetic template (Amano et al. 2010). Mesenchymal cells close to the MC start differentiating into osteoblasts and are surrounded by an osteogenic front. The MC lengthens ventromedially and dorsolaterally on both sides of the mandibular arch and fuses at the distal tip originating the mandibular symphysis. Therefore, formation of the bones of the mandible occurs in two different ways: the most distal region undergoes endochondral ossification in the symphyseal region, while the middle part undergoes intramembranous ossification. Whereas the proximal ends of Meckel's cartilage persist and form the basis of the middle ear bones, the intermediate part is degraded. In addition to the MC, mandibular development also involves other secondary cartilages e.g. coronoid and condylar cartilages (Parada and Chai 2015; Radlanski et al. 2003; Logjes et al. 2018) (Fig. 3).

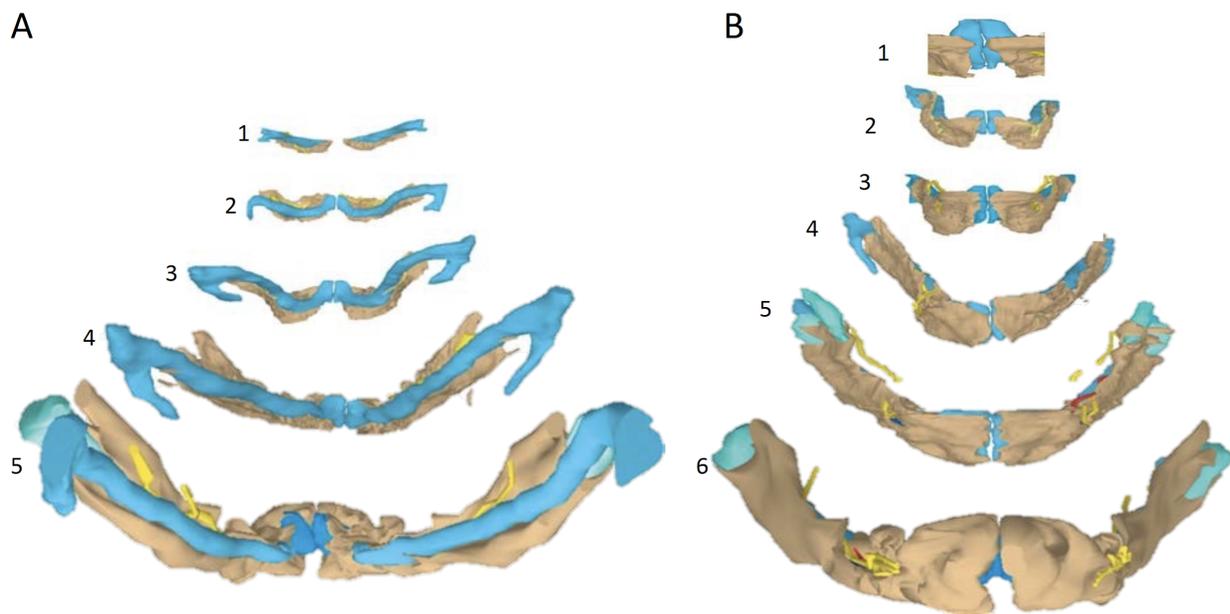


Figure 3. 3D reconstructions of human mandible at different stages of development. The mandible of human embryos is shown in dorsal views (**A**) and anterior views (**B**). (**A**) Figures I-V are arranged according to maturation at the same scales. CRL (crown-rump-length) sizes: **I** 22mm, **II** 19mm, **III** 25mm, **IV** 66mm, **V** 117mm. (**B**) Figures I-V are arranged according to maturation at the same scales, except **I**, which shows a close-up view of the symphyseal region. CRL sizes: **I** 21mm, **II** 25mm, **III** 24mm, **IV** 66mm, **V** 76mm, **VI** 117mm. Bone is represented in brown, Meckel's cartilage in blue, cartilaginous core of condylar process in light blue, cartilaginous core of coronoid process in turquoise, inferior alveolar and lingual nerves in yellow, and inferior alveolar artery in red. Adapted from Radlanski et al. 2003.

Anomalies in any step of NCC development, such as neural crest specification, delamination, migration, proliferation, survival and differentiation are often associated with the pathophysiology of a variety of disorders known as neurocristopathies, which frequently encompass craniofacial abnormalities (Vega-Lopez et al. 2018; Knecht and Bronner-Fraser 2002).

A tightly controlled spatial and temporal signaling network is required for the multistep embryonic development, yet the molecular cascades that control human craniofacial development have until recently been little understood. The identification and functional studies of mutations that cause craniofacial malformations (CFMs) represent a promising approach for understanding craniofacial morphogenesis and have given significant insight into how the head develops. It is noteworthy that the intricate series of molecular signals orchestrating NCC development are distinct across organisms. However, there is a clear conservation of most

mechanisms leading to NCC formation and the transcriptional regulators of these processes largely overlap (e.g. *Snail*, *Twist* and *FoxD3* genes) (Barriga et al. 2015).

2. Craniofacial Malformations

As a highly social species, the human face is in many ways the most powerful and influential region of our anatomy. The face not only conveys our emotions to others, but also constitutes a fundamental aspect of any individual's identity. Indeed, no physical feature evokes more profound reactions when disease disfigures it than the face (Jack and Schyns 2015; Helms et al. 2005).

Craniofacial malformations comprise a heterogeneous group of disorders that collectively represent over one-third of all congenital birth defects. They arise from disturbances in craniofacial morphogenesis during embryonic development, affecting tissues of the skull and face (Gorlin et al. 2001).

CFMs can compromise a wide variety of craniofacial tissues at different levels of severity and can occur as isolated events or in association with other clinical manifestations (Gorlin et al. 2001). The most common craniofacial malformations, affecting 1:700 live-births worldwide, are the orofacial clefts, characterized by the discontinuity of the structures forming the lip and/or palate (Schutte and Murray 1999; Mossey et al. 2009). Another example are the craniosynostosis, a condition that involves premature ossification of cranial sutures resulting in abnormal skull shape as well as cranial growth restriction and increased intracranial pressure (Khanna et al. 2011).

These disturbances lead to functional impairment such as airway obstruction, disorders of feeding, hearing, speech and psychosocial difficulties. Accordingly, CFMs demand extensive treatment from several distinct health professionals entailing a significant burden to both patients and the health care system overall (Hamm and Robin 2015).

Craniofacial malformations are caused by genetic and/or environmental factors including chromosomal alterations and maternal alcohol intake, smoking and nutritional deficits (Jones

2011; Brito et al. 2012; Sabbagh et al. 2015). CFMs are genetically and clinically heterogeneous and both monogenic (Mendelian) and multifactorial inheritances have been reported (Twigg and Wilkie 2015). We would like to highlight among monogenic forms the subject of this thesis, Richieri-Costa-Pereira syndrome, which is described in detail below.

3. Richieri Costa-Pereira Syndrome

Richieri-Costa-Pereira syndrome (RCPS; MIM #268305) is a rare acrofacial dysostosis initially described in five unrelated Brazilian females in 1992 (Richieri-Costa and Pereira 1992). In the subsequent years, the first male patients (Richieri-Costa and Pereira 1993) as well as additional cases have been described mainly in Brazilian individuals. To date, 40 individuals have been described in the literature with RCPS, including two non-Brazilian patients (Tabith Júnior and Gonçalves 1996; Walter-Nicolet et al. 1999; Richieri-Costa and Brandão-Almeida 1997; Golbert et al. 2007; Graziadio et al. 2009; Tabith and Bento-Gonçalves 2003; Guion-Almeida and Richieri-Costa 1998; Favaro et al. 2011; Vendramini et al. 2004; Bertola et al. 2018; Raskin et al. 2013). The fact that the majority of patients are from São Paulo state and a notable number of them are from an isolated geographical area named Vale do Ribeira have suggested a founder effect for this disorder (Favaro et al. 2011).

RCPS presents a wide phenotypic spectrum and is characterized by several craniofacial and limb abnormalities, including microstomia, Robin sequence (micrognathia, glossoptosis and cleft palate), abnormal fusion of the mandible, absence of central lower incisors, epiglottis hypoplasia/agenesis, laryngeal and ear anomalies, hypoplastic thumbs, mesomelic shortening of upper and lower limbs, clubfeet, learning and speech impairments, and microcephaly, which has been recently included (Bertola et al. 2018) (Fig. 4). The phenotype expressivity is variable, ranging from only mild defects in fusion of the mandible to severe cases with agenesis of the mandible and craniofacial pansynostosis (Raskin et al. 2013). RCPS affects multiple structures derived from neural crest cells (NCCs), suggesting that these cells are compromised during embryonic development.

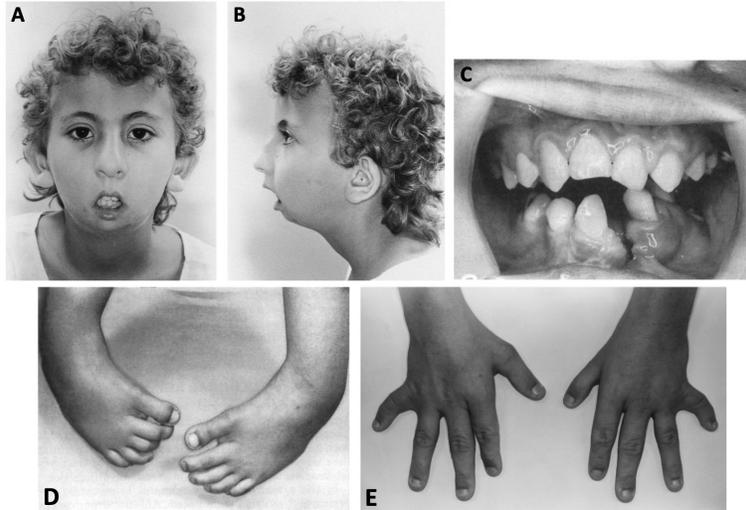


Figure 4. RCPS patients illustrating typical facial features of the syndrome, including microstomia, micrognathia (**A**, **B**) and cleft of lower alveolar ridge (**C**), as well as limb defects, including clubfeet (**D**) and hand anomalies (**E**). Patients reported by Richieri-Costa and Pereira 1993; Richieri-Costa Pereira and Brandão-Almeida 1997; Tabith and Bento-Gonçalves 2003.

3.1 Etiology

In 2014, our group identified the causative mutation of this syndrome by homozygosity mapping and targeted sequencing. RCPS is mainly caused by an increased number of repeat motifs in the 5'UTR of the *EIF4A3*, which encodes a core component of the exon junction complex (EJC). The *EIF4A3* 5'UTR is complex, presenting multiple allele patterns varying in size and organization of three types of motifs: (1) CA-18nt, TCGGCAGCGGCAGCGAGG; (2) CACA-20nt, TCGGCAGCGGCACAGCGAGG and (3) CGCA-20nt, TCGGCAGCGGCGCAGCGAGG (Favaro et al. 2014).

RCPS patients show a distinct allelic pattern, determined not only by the larger number of motifs but also by the presence of a unique motif containing an A-to-G substitution. The most prevalent allelic pattern among controls is an initial CACA-20nt repeated 2 to 9 times, followed by one CA-18nt, one CACA-20nt, and a final CA-18nt. However, RCPS patients present an initial CACA-20nt, followed by 12 to 13 repeats of CGCA-20nt, another CACA-20nt, and one final CA-

18nt (Favaro et al. 2014). Therefore, control alleles carry up to 12 repeats whereas disease-associated alleles carry 15 or 16 repeats (Fig. 5).

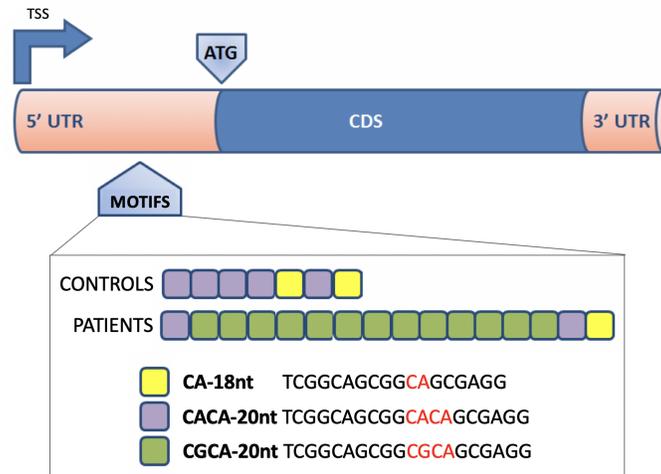


Figure 5. Cartoon of the repeat motifs within the *EIF4A3* 5'UTR, ending 43 bases upstream the first ATG. The different colors represent the motif types: CA-18nt, CACA-20nt and CGCA-20nt. TSS, transcription start site; UTR, untranslated region; CDS, coding sequence. Adapted from Alvizi 2017.

The most part of affected individuals are homozygous for the 16-repeat allele. However, a compound heterozygote patient was reported by Favaro et al. carrying a 14-repeat allele *in trans* with a likely pathogenic non-synonymous point mutation (c.809A>G; p.Asp270Gly), who interestingly presents a milder phenotype than 16-repeat allele homozygotes (Favaro et al. 2014).

EIF4A3 transcript abundance is 30-40 % lower in affected individuals than in controls (Favaro et al. 2014). However, the functional effects of the 5'UTR motifs on gene expression and whether both the number and sequence of the motifs are involved in *EIF4A3* downregulation remain to be uncovered. In addition, the origin of the pathogenic alleles (e.g. meiotic instability or unequal crossing-over events) is still unknown. Insights into the origin and effect of these complex alleles will contribute to a better understanding of regulatory features of 5'UTR regions and their role in craniofacial development.

Despite the fact that these repeat motifs are CG rich and their increased number in the disease-associated alleles results in a gain of 37 CpG sites, they are not hypermethylated in RCPS

patients compared to controls (Alvizi 2017). Therefore, DNA hypermethylation does not explain *EIF4A3* downregulation in RCPS. Further investigation of potential transcriptional or post-transcriptional mechanisms is needed to unveil how these motifs interfere with *EIF4A3* expression.

4. The Exon Junction Complex

The EJC plays a central role in gene expression control (Le Hir et al. 2016). It contributes at many levels to mRNA metabolism, including splicing, intra-cellular mRNA localization and nuclear export, mRNA stability control by nonsense-mediated decay (NMD), translation efficiency and rRNA biogenesis (Wang et al. 2014; Hachet and Ephrussi 2004; Chazal et al. 2013; Gehring et al. 2003; Le Hir et al. 2001).

Pre-mRNAs are often associated with nuclear proteins, collectively referred to as heterogeneous nuclear ribonucleoproteins (hnRNP). mRNA processing, including capping, splicing, and polyadenylation, are accompanied by profound changes in the composition and arrangements of these protein complexes (Dreyfuss et al. 2002).

In particular, the exon junction complex (EJC) is required by the splicing machinery and is assembled 20 to 24nt upstream of spliced junctions onto mature spliced mRNAs. EJC binds to RNA without sequence specificity and acts as a link between nuclear splicing and cytoplasmic processes, as it remains associated on mRNAs when transported to the cytoplasm (Le Hir et al. 2016). This multiprotein complex is organized around four central proteins that constitute the EJC core: eIF4A3, MAGOH, Y14 (also known as RBM8A) and MLN51. MAGOH and Y14 form a tight heterodimer and contact the two globular domains (RecA1 and RecA2) of eIF4A3. The MLN51 is a long protein that surrounds eIF4A3, binds to MAGOH and contacts one nucleotide, stabilizing the complex (Fig. 6). This tetrameric core serves as a platform for multiple peripheral factors during different post-transcriptional processes (Andreou and Klostermeier 2013).

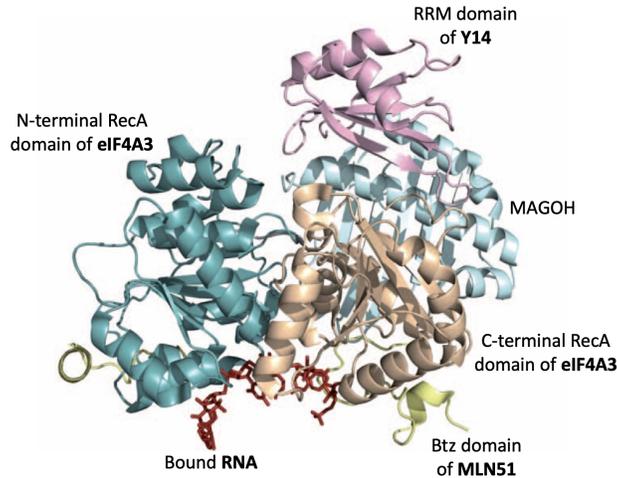


Figure 6. Organization of the domains of the exon junction complex (EJC) core elements. The N- and C-terminal RecA-domain of eIF4A3 are depicted in teal and brown, respectively, the RRM domain of Y14 in pink, MAGOH in cyan, the Btz of MLN51 in yellow, and the RNA is represented in red. Adapted from Andreou and Klostermeier 2013.

4.1 Eukaryotic Translation Initiation Factor 4A3

There are three paralogues of eIF4A in mammals: eIF4A1, eIF4A2 and eIF4A3. Despite sharing 60% sequence identity, the eukaryotic translation initiation factor 4A3 (eIF4A3) is functionally distinct from the others, is localized mostly in the nucleus, and is the only one incorporated into EJC. eIF4A1 and eIF4A2 are both cytoplasmic proteins and present higher structural and sequence similarity, sharing 90% identity to each other. However, they are not functionally redundant *in vivo*, and are involved in translation by the eIF4F translation initiation complex (Lu et al. 2014).

eIF4A3 is a member of the DEAD-box protein family of RNA helicases, which is characterized by a number of conserved motifs on a helicase core formed by two RecA-domains. The helicase core can unwind short RNA duplexes and also comprise an RNA-stimulated ATPase, acting as a nucleotide-dependent molecular switch with different affinities for RNA. In other words, in the absence of RNA or ATP, the helicase core adopts an open conformation, and binding

of RNA and ATP triggers a transition to a close conformation. It retains the closed conformation upon ATP hydrolysis, returning to an open state after phosphate release which in turn leads to release of the RNA strand. According, eIF4A3 serves as an RNA clamp (Ballut et al. 2005; Andreou and Klostermeier 2013).

4.2 EJC Components and Developmental Disorders

EJC components are mutated in several pathologies, such as in thrombocytopenia with absent radii (TAR) syndrome, for example, which is caused by *RBM8A* mutations and can present with microcephaly or Autism spectrum disorder (ASD) (Addington et al. 2011; Laumonnier et al. 2010; Nguyen et al. 2012; Tarpey et al. 2007; Albers et al. 2012). Copy number variations (CNVs) of either *EIF4A3* or *RBM8A* are associated with intellectual disability and brain malformations (Nguyen et al. 2013). Furthermore, haploinsufficiency for each of the EJC components (*Eif4a3*, *Magoh* and *Rbm8a*) in neural stem cells (NSCs) leads to aberrant neurogenesis and causes p53-mediated microcephaly in mice (Mao et al. 2016). Altogether, these genetic studies indicate disruption of *EIF4A3* impairs neurodevelopment.

It is known that biallelic hypomorphic mutations in *EIF4A3* leads to RCPS (Favaro et al. 2014), postulating *EIF4A3* as an essential element during craniofacial development. However, the pathogenetic mechanisms responsible for this syndrome are completely unknown. Ultimately, the requirements of *EIF4A3* for NCC functions is still obscure, and the molecular mechanisms by which *EIF4A3* mutation causes RCPS remain unclear.

5. Investigating the Pathogenetic Mechanisms of CFMs

The understanding of the molecular and cellular mechanisms governing craniofacial morphogenesis is critical to clarify how these CFMs arise. To date, several studies in animal and

cellular models have contributed to our understanding of normal or abnormal (disease-associated) craniofacial development.

5.1 Cellular Models

The use of primary cell cultures cannot cover the study of conditions in which embryonic tissues are primarily affected, such as in craniofacial malformations. In addition, the use of embryonic stem cells (ESCs) raises major concerns and controversies, mainly sharp ethical and political considerations, which have hindered their use for research and clinical applications (Lo and Parham 2009). Therefore, the advent of an alternative technology that allowed for reprogramming of somatic cells back to a pluripotent state was a remarkable event in the history of biological research. Indeed, the development of induced pluripotent stem cell (iPSC) method resulted in the 2012 Nobel Prize in Physiology or Medicine being awarded to its pioneer Shinya Yamanaka. iPSCs are derived from mature cells that have been genetically reprogrammed to an embryonic stem cell-like state and subsequently can be differentiated into multiple cell types (Takahashi et al. 2007).

Since the emergence of iPSCs, enormous progress has been made in stem cell biology. Reprogrammed cells have been widely used for drug discovery and testing, cell therapy development, and regenerative medicine (Shi et al. 2017). Furthermore, this powerful tool offers great promise for understanding basic mechanisms of human development and presents opportunities for modeling, and potentially treating human diseases. An increasing number of novel pathological mechanisms have been elucidated by using patient-derived iPSCs for disease modeling (Bellin et al. 2012; Ishiy 2016; Kobayashi 2016). Cellular models offer a unique approach to study cellular phenotypes and molecular mechanisms underlying human disorders, in particular those in which is difficult to obtain the tissue of interest such as developmental disorders. Assessing cells from patients carrying disease-specific genomic mutations has been a promising instrument to conduct functional studies.

iPSCs can be differentiated into disease-relevant cell types, such as neural crest cells in order to study craniofacial disorders. Recent protocols allow for the efficient generation of iPSC-derived NCCs by Wnt pathway activation and Smad blockade with the use of small molecules, producing homogeneous NCC populations (Fukuta et al. 2014; Menendez et al. 2013). However, the available methods for generation NCCs lack a NPB stage and subsequently do not recapitulate important steps involved in NCC development such as neural crest specification and EMT.

5.2 Animal Models

Many different vertebrates including amphibians, birds, teleosts and mice have been used to study craniofacial development, mainly by gene targeting/loss of function strategies. Targeting specific genes by using RNA interference (RNAi), anti-sense morpholinos or dominant-negative constructs enable assessment of their role in embryonic development (Barriga et al. 2015). In recent years, novel approaches have been developed as technical advances in genome editing, including CRISPR/Cas9 system, Zinc Finger Nuclease (ZFN) and transcription activator-like effector nuclease (TALEN) technologies, have opened new possibilities for the study of embryonic development as well as disease modeling (Wang and Sun 2019).

Over the years, one approach to control gene expression that has become increasingly widespread is the Cre/loxP system. It is a simple and powerful tool which allows researchers to create a variety of transgenic mice (McLellan et al. 2017). Briefly, the *cre* gene encodes a site-specific DNA recombinase, which recognizes *loxP* sequences and recombines these sites leading to the excision of the DNA between them. Cre/loxP recombination stands as an important strategy to investigate gene function *in vivo* and has greatly expanded the versatility with which biologic questions can be addressed in mouse models, particularly because the desired gene modification can be restricted to specific cell types or developmental stages using a spatial and/or temporal control of gene expression (conditional mouse mutants) (Kühn and Torres 2002).

Since mice closely reflect many aspects of human development, they have served as the major mammalian model system and have been extensively explored to study gene function in

craniofacial development as well as the etiology of disorders. Indeed, many mouse models have led to relevant discoveries regarding the pathogenetic mechanisms responsible for CFMs (Chai and Maxson 2006; Sakai and Trainor 2009; Fantauzzo and Soriano 2014; Mao et al. 2016). However, some discrepancies in NCC development across species should be taken into consideration (Barriga et al. 2015). A valid approach is the combination of human cell culture and mouse model systems, which may provide a paradigm to study craniofacial development and CFM pathogenesis, address reproducibility and provide validation to data.

REFERENCES

- Adameyko, I. and Fried, K. 2016. The nervous system orchestrates and integrates craniofacial development: A review. *Frontiers in physiology* 7, p. 49.
- Addington, A.M., Gauthier, J., Piton, A., et al. 2011. A novel frameshift mutation in UPF3B identified in brothers affected with childhood onset schizophrenia and autism spectrum disorders. *Molecular Psychiatry* 16(3), pp. 238–239.
- Albers, C.A., Paul, D.S., Schulze, H., et al. 2012. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nature Genetics* 44(4), pp. 435–9, S1.
- Alvizi, L. 2017. Mecanismos genéticos e epigenéticos na etiologia das fissuras orofaciais.
- Amano, O., Doi, T., Yamada, T., et al. 2010. Meckel's cartilage: discovery, embryology and evolution. *Journal of Oral Biosciences* 52(2), pp. 125–135.
- Andreou, A.Z. and Klostermeier, D. 2013. The DEAD-box helicase eIF4A: paradigm or the odd one out? *RNA Biology* 10(1), pp. 19–32.
- Ballut, L., Marchadier, B., Bagnuet, A., Tomasetto, C., Séraphin, B. and Le Hir, H. 2005. The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nature Structural & Molecular Biology* 12(10), pp. 861–869.
- 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(9), pp. 1555–1560.
- Bellin, M., Marchetto, M.C., Gage, F.H. and Mummery, C.L. 2012. Induced pluripotent stem cells: the new patient? *Nature Reviews. Molecular Cell Biology* 13(11), pp. 713–726.
- Bertola, D.R., Hsia, G., Alvizi, L., et al. 2018. Richieri-Costa-Pereira syndrome: Expanding its phenotypic and genotypic spectrum. *Clinical Genetics* 93(4), pp. 800–811.
- Brito, L.A., Meira, J.G.C., Kobayashi, G.S. and Passos-Bueno, M.R. 2012. Genetics and management of the patient with orofacial cleft. *Plastic surgery international* 2012, p. 782821.
- Bronner, M.E. and LeDouarin, N.M. 2012. Development and evolution of the neural crest: an overview. *Developmental Biology* 366(1), pp. 2–9.
- Chai, Y. and Maxson, R.E. 2006. Recent advances in craniofacial morphogenesis. *Developmental Dynamics* 235(9), pp. 2353–2375.
- Chazal, P.-E., Dagueuet, E., Wendling, C., et al. 2013. EJC core component MLN51 interacts with eIF3 and activates translation. *Proceedings of the National Academy of Sciences of the United States of America* 110(15), pp. 5903–5908.
- 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. *American Journal of Medical Genetics. Part A* 155A(2), pp. 270–279.
- Depew, M.J., Lufkin, T. and Rubenstein, J.L.R. 2002. Specification of jaw subdivisions by Dlx genes. *Science* 298(5592), pp. 381–385.
- Dreyfuss, G., Kim, V.N. and Kataoka, N. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nature Reviews. Molecular Cell Biology* 3(3), pp. 195–205.
- Fantauzzo, K.A. and Soriano, P. 2014. PI3K-mediated PDGFR α signaling regulates survival and proliferation in skeletal development through p53-dependent intracellular pathways. *Genes & Development* 28(9), pp. 1005–1017.
- Favaro, F.P., Alvizi, L., Zechi-Ceide, R.M., et al. 2014. A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. *American Journal of Human Genetics* 94(1), pp. 120–128.
- Favaro, F.P., Zechi-Ceide, R.M., Alvarez, C.W., et al. 2011. Richieri-Costa-Pereira syndrome: a unique

acrofacial dysostosis type. An overview of the Brazilian cases. *American Journal of Medical Genetics. Part A* 155A(2), pp. 322–331.

Fukuta, M., Nakai, Y., Kirino, K., et al. 2014. Derivation of mesenchymal stromal cells from pluripotent stem cells through a neural crest lineage using small molecule compounds with defined media. *Plos One* 9(12), p. e112291.

Gans, C. and Northcutt, R.G. 1983. Neural crest and the origin of vertebrates: a new head. *Science* 220(4594), pp. 268–273.

Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W. and Kulozik, A.E. 2003. Y14 and hUpf3b form an NMD-activating complex. *Molecular Cell* 11(4), pp. 939–949.

Golbert, M.B., Dewes, L.O., Philipsen, V.R., Wachholz, R.S., Deutschendorf, C. and Leite, J.C.L. 2007. New clinical findings in the Richieri-Costa/Pereira type of acrofacial dysostosis. *Clinical dysmorphology* 16(2), pp. 85–88.

Gorlin, R.J., Cohen, M.M. (Meyer M. and Hennekam, R.C.M. 2001. *Syndromes of the head and neck*. 4th ed. Oxford [England]: Oxford University Press.

Graziadio, C., Rosa, R.F.M., Zen, P.R.G., Flores, J.A.M. and Paskulin, G.A. 2009. Richieri-Costa and Pereira form of acrofacial dysostosis: first description of an adult with mesomelic shortness of the lower limbs. *American Journal of Medical Genetics. Part A* 149A(12), pp. 2886–2888.

Guion-Almeida, M.L. and Richieri-Costa, A. 1998. Brief clinical report: autosomal recessive short stature, Robin sequence, cleft mandible, pre/postaxial limb anomalies and clubfeet report of a patient with polydactyly of the halluces. *The Brazilian Journal of Dysmorphology and Speech-Hearing Disorders*.

Hachet, O. and Ephrussi, A. 2004. Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. *Nature* 428(6986), pp. 959–963.

Hamm, J.A. and Robin, N.H. 2015. Newborn craniofacial malformations: orofacial clefting and craniosynostosis. *Clinics in Perinatology* 42(2), p. 321–36, viii.

Helms, J.A., Cordero, D. and Tapadia, M.D. 2005. New insights into craniofacial morphogenesis. *Development* 132(5), pp. 851–861.

Horie, R., Hazbun, A., Chen, K., Cao, C., Levine, M. and Horie, T. 2018. Shared evolutionary origin of vertebrate neural crest and cranial placodes. *Nature* 560(7717), pp. 228–232.

Ishiy F. A. A. 2016. Estudo de marcadores moleculares no processo de diferenciação osteogênica de células-tronco mesenquimais.

Jack, R.E. and Schyngs, P.G. 2015. The human face as a dynamic tool for social communication. *Current Biology* 25(14), pp. R621-34.

Jandzik, D., Garnett, A.T., Square, T.A., Cattell, M.V., Yu, J.-K. and Medeiros, D.M. 2015. Evolution of the new vertebrate head by co-option of an ancient chordate skeletal tissue. *Nature* 518(7540), pp. 534–537.

Jones, K.L. 2011. The effects of alcohol on fetal development. *Birth Defects Research. Part C, Embryo Today: Reviews* 93(1), pp. 3–11.

Kaucka, M. and Adameyko, I. 2019. Evolution and development of the cartilaginous skull: From a lancelet towards a human face. *Seminars in Cell & Developmental Biology* 91, pp. 2–12.

Khanna, P.C., Thapa, M.M., Iyer, R.S. and Prasad, S.S. 2011. Pictorial essay: The many faces of craniosynostosis. *The Indian journal of radiology & imaging* 21(1), pp. 49–56.

Knecht, A.K. and Bronner-Fraser, M. 2002. Induction of the neural crest: a multigene process. *Nature Reviews. Genetics* 3(6), pp. 453–461.

Kobayashi, G. S. 2016. Investigação da etiologia de malformações craniofaciais com uso de células derivadas de crista neural.

Kühn, R. and Torres, R.M. 2002. Cre/loxP recombination system and gene targeting. *Methods in Molecular Biology* 180, pp. 175–204.

Laumonier, F., Shoubridge, C., Antar, C., et al. 2010. Mutations of the UPF3B gene, which encodes a protein widely expressed in neurons, are associated with nonspecific mental retardation with or without

autism. *Molecular Psychiatry* 15(7), pp. 767–776.

Le Hir, H., Gatfield, D., Izaurralde, E. and Moore, M.J. 2001. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *The EMBO Journal* 20(17), pp. 4987–4997.

Le Hir, H., Saulière, J. and Wang, Z. 2016. The exon junction complex as a node of post-transcriptional networks. *Nature Reviews. Molecular Cell Biology* 17(1), pp. 41–54.

Lo, B. and Parham, L. 2009. Ethical issues in stem cell research. *Endocrine Reviews* 30(3), pp. 204–213.

Logies, R.J.H., Breugem, C.C., Van Haaften, G., et al. 2018. The ontogeny of Robin sequence. *American Journal of Medical Genetics. Part A* 176(6), pp. 1349–1368.

Lu, W.-T., Wilczynska, A., Smith, E. and Bushell, M. 2014. The diverse roles of the eIF4A family: you are the company you keep. *Biochemical Society Transactions* 42(1), pp. 166–172.

Mao, H., McMahan, J.J., Tsai, Y.-H., Wang, Z. and Silver, D.L. 2016. Haploinsufficiency for Core Exon Junction Complex Components Disrupts Embryonic Neurogenesis and Causes p53-Mediated Microcephaly. *PLoS Genetics* 12(9), p. e1006282.

McLellan, M.A., Rosenthal, N.A. and Pinto, A.R. 2017. Cre-loxP-Mediated Recombination: General Principles and Experimental Considerations. *Current protocols in mouse biology* 7(1), pp. 1–12.

Menendez, L., Kulik, M.J., Page, A.T., et al. 2013. Directed differentiation of human pluripotent cells to neural crest stem cells. *Nature Protocols* 8(1), pp. 203–212.

Mossey, P.A., Little, J., Munger, R.G., Dixon, M.J. and Shaw, W.C. 2009. Cleft lip and palate. *The Lancet* 374(9703), pp. 1773–1785.

Nguyen, L.S., Jolly, L., Shoubridge, C., et al. 2012. Transcriptome profiling of UPF3B/NMD-deficient lymphoblastoid cells from patients with various forms of intellectual disability. *Molecular Psychiatry* 17(11), pp. 1103–1115.

Nguyen, L.S., Kim, H.-G., Rosenfeld, J.A., et al. 2013. Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders. *Human Molecular Genetics* 22(9), pp. 1816–1825.

Parada, C. and Chai, Y. 2015. Mandible and tongue development. *Current Topics in Developmental Biology* 115, pp. 31–58.

Parker, H.J., Pushel, I. and Krumlauf, R. 2018. Coupling the roles of Hox genes to regulatory networks patterning cranial neural crest. *Developmental Biology* 444 Suppl 1, pp. S67–S78.

Radlanski, R.J., Renz, H. and Klarkowski, M.C. 2003. Prenatal development of the human mandible. 3D reconstructions, morphometry and bone remodelling pattern, sizes 12–117 mm CRL. *Anatomy and embryology* 207(3), pp. 221–232.

Raskin, S., Souza, M., Medeiros, M.C., Manfron, M. and Chong E Silva, D.C. 2013. Richieri-costa and Pereira syndrome: severe phenotype. *American Journal of Medical Genetics. Part A* 161A(8), pp. 1999–2003.

Richieri-Costa, A. and Brandão-Almeida, I.L. 1997. Short stature, Robin sequence, cleft mandible, pre/postaxial hand anomalies, and clubfoot: another affected Brazilian patient born to consanguineous parents. *American Journal of Medical Genetics* 71(2), pp. 233–235.

Richieri-Costa, A. and Pereira, S.C. 1993. Autosomal recessive short stature, Robin sequence, cleft mandible, pre/postaxial hand anomalies, and clubfeet in male patients. *American Journal of Medical Genetics* 47(5), pp. 707–709.

Richieri-Costa, A. and Pereira, S.C. 1992. Short stature, Robin sequence, cleft mandible, pre/postaxial hand anomalies, and clubfoot: a new autosomal recessive syndrome. *American Journal of Medical Genetics* 42(5), pp. 681–687.

Sabbagh, H.J., Hassan, M.H.A., Innes, N.P.T., Elkodary, H.M., Little, J. and Mossey, P.A. 2015. Passive smoking in the etiology of non-syndromic orofacial clefts: a systematic review and meta-analysis. *Plos One* 10(3), p. e0116963.

Sakai, D. and Trainor, P.A. 2009. Treacher Collins syndrome: unmasking the role of Tcof1/treacle. *The*

International Journal of Biochemistry & Cell Biology 41(6), pp. 1229–1232.

Santagati, F. and Rijli, F.M. 2003. Cranial neural crest and the building of the vertebrate head. *Nature Reviews. Neuroscience* 4(10), pp. 806–818.

Schutte, B.C. and Murray, J.C. 1999. The many faces and factors of orofacial clefts. *Human Molecular Genetics* 8(10), pp. 1853–1859.

Scott, G.F. 2010. *Developmental Biology*. ed. 9.

Shi, Y., Inoue, H., Wu, J.C. and Yamanaka, S. 2017. Induced pluripotent stem cell technology: a decade of progress. *Nature Reviews. Drug Discovery* 16(2), pp. 115–130.

Tabith, A. and Bento-Gonçalves, C.G. de A. 2003. Laryngeal malformation in the Richieri-Costa-Pereira acrofacial dysostosis: description of two new patients. *American Journal of Medical Genetics. Part A* 122A(2), pp. 133–138.

Tabith Júnior, A. and Gonçalves, C.G. 1996. Laryngeal malformations in the Richieri-Costa and Pereira form of acrofacial dysostosis. *American Journal of Medical Genetics* 66(4), pp. 399–402.

Takahashi, K., Okita, K., Nakagawa, M. and Yamanaka, S. 2007. Induction of pluripotent stem cells from fibroblast cultures. *Nature Protocols* 2(12), pp. 3081–3089.

Tarpey, P.S., Raymond, F.L., Nguyen, L.S., et al. 2007. Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. *Nature Genetics* 39(9), pp. 1127–1133.

Theveneau, E. and Mayor, R. 2012. Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. *Developmental Biology* 366(1), pp. 34–54.

Trainor, P.A. 2013. *Neural Crest Cells: Evolution, Development and Disease*. ed. 1.

Twigg, S.R.F. and Wilkie, A.O.M. 2015. New insights into craniofacial malformations. *Human Molecular Genetics* 24(R1), pp. R50-9.

Vega-Lopez, G.A., Cerrizuela, S., Tribulo, C. and Aybar, M.J. 2018. Neurocristopathies: New insights 150 years after the neural crest discovery. *Developmental Biology* 444 Suppl 1, pp. S110–S143.

Vendramini, S., Netto, C. and Jr, A. 2004. Audiological findings in patients with the Richieri-Costa-Pereira syndrome.

Walter-Nicolet, E., Coëslie, A., Joriot, S., Kacet, N., Moerman, A. and Manouvrier-Hanu, S. 1999. The Richieri-Costa and Pereira form of acrofacial dysostosis: first case in a non-Brazilian infant. *American Journal of Medical Genetics* 87(5), pp. 430–433.

Wang, C. and Sun, H. 2019. [Progress in gene knockout mice]. *Sheng Wu Gong Cheng Xue Bao = Chinese Journal of Biotechnology* 35(5), pp. 784–794.

Wang, Z., Murigneux, V. and Le Hir, H. 2014. Transcriptome-wide modulation of splicing by the exon junction complex. *Genome Biology* 15(12), p. 551.

Zhang, D., Ighaniyan, S., Stathopoulos, L., et al. 2014. The neural crest: a versatile organ system. *Birth Defects Research. Part C, Embryo Today: Reviews* 102(3), pp. 275–298.

OBJECTIVES

The main objective of this study was to investigate the cellular and molecular pathogenetic mechanisms by which *EIF4A3* mutations cause Richieri-Costa-Pereira syndrome (RCPS). We carried out genomic and functional studies to test the following hypothesis: **(1)** the RCPS-associated allelic patterns arose by unequal crossing-over events; **(2)** the increased number of motifs in the *EIF4A3* 5'UTR decrease gene expression; **(3)** the decreased expression of *EIF4A3* impairs NCC functions; **(4)** the *EIF4A3* deficiency in NCCs disrupts splicing and/or expression of transcripts related to craniofacial development.

The specific objectives were as follows:

- a)** To investigate the origin of the pathogenic alleles associated with RCPS;
- b)** To evaluate the functional effects of the 5'UTR repeat motifs on gene expression;
- c)** To identify dysfunctions related to RCPS in *EIF4A3* depleted hiPSC-derived NCCs;
- d)** To access differentially expressed genes/proteins and pathways by studying the transcriptome/proteome of patient-derived and control NCCs;
- e)** To validate human transcriptomic data using NCCs from conditional *Eif4a3* haploinsufficient mouse embryos;

In parallel with the aforementioned main objective, we also aimed to develop a simple and efficient *in vitro* method for NCC generation to recapitulate developmental events involved in NCC specification and epithelial to mesenchymal transition (EMT).

The specific objective was:

- f)** To derivate hiPSC-derived NCCs through a neural plate border (NPB)-like stage.

CHAPTER II

Complexity of the 5' Untranslated Region of *EIF4A3*, a Critical Factor for Craniofacial and Neural Development

Gabriella S. P. Hsia^{1†}, Camila M. Musso^{1†}, Lucas Alvizi¹, Luciano A. Brito¹, Gerson S. Kobayashi¹, Rita C. M. Pavanello¹, Mayana Zatz¹, Alice Gardham², Emma Wakeling², Roseli M. Zechi-Ceide³, Debora Bertola^{1,4}, Maria Rita Passos-Bueno^{1*}

¹Centro de Estudos do Genoma Humano e Células Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

²North East Thames Genetics Service, Great Ormond Street Hospital, London, United Kingdom

³Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo, Bauru, Brazil

⁴Instituto da Criança, Hospital das Clínicas da FMUSP, Universidade de São Paulo, São Paulo, Brazil

†Co-first authors

***Correspondent author**

Keywords: acrofacial dysostosis, non-coding region, haplotype, expansion, crossing-over

This chapter was published in *Frontiers in Genetics*, 2018

Hsia, G.S.P., Musso, C.M., Alvizi, L., et al. 2018. Complexity of the 5' untranslated region of *EIF4A3*, a critical factor for craniofacial and neural development. *Frontiers in genetics* 9, p. 149.

<https://www.ncbi.nlm.nih.gov/pubmed/29922329>

ABSTRACT

Repeats in coding and non-coding regions have increasingly been associated with many human genetic disorders, such as Richieri-Costa-Pereira syndrome (RCPS). RCPS, mostly characterized by midline cleft mandible, Robin sequence and limb defects, is an autosomal-recessive acrofacial dysostosis mainly reported in Brazilian patients. This disorder is caused by decreased levels of *EIF4A3*, mostly due to an increased number of repeats at the *EIF4A3* 5'UTR. *EIF4A3* 5'UTR alleles are CG-rich and vary in size and organization of three types of motifs. An exclusive allelic pattern was identified among affected individuals, in which the CGCA-motif is the most prevalent, herein referred as "disease-associated CGCA-20nt motif." The origin of the pathogenic alleles containing the disease-associated motif, as well as the functional effects of the 5'UTR motifs on *EIF4A3* expression, to date, are entirely unknown. Here, we characterized 43 different *EIF4A3* 5'UTR alleles in a cohort of 380 unaffected individuals. We identified eight heterozygous unaffected individuals harboring the disease-associated CGCA-20nt motif and our haplotype analyses indicate that there are more than one haplotype associated with RCPS. The combined analysis of number, motif organization and haplotypic diversity, as well as the observation of two apparently distinct haplotypes associated with the disease-associated CGCA-20nt motif, suggest that the RCPS alleles might have arisen from independent unequal crossing-over events between ancient alleles at least twice. Moreover, we have shown that the number and sequence of motifs in the 5'UTR region is associated with *EIF4A3* repression, which is not mediated by CpG methylation. In conclusion, this study has shown that the large number of repeats in *EIF4A3* does not represent a dynamic mutation and RCPS can arise in any population harboring alleles with the CGCA-20nt motif. We also provided further evidence that *EIF4A3* 5'UTR is a regulatory region and the size and sequence type of the repeats at 5'UTR may contribute to clinical variability in RCPS.

INTRODUCTION

Over two-thirds of the human genome is comprised by repetitive elements (de Koning et al., 2011), which have been increasingly associated with functional regulatory roles. Consequently, a variety of human genetic disorders are caused by repeats in coding and non-coding sequences (Cummings and Zoghbi, 2000; Gatchel and Zoghbi, 2005; Mirkin, 2007; La Spada and Taylor, 2010; McMurray, 2010; DeJesus-Hernandez et al., 2011; Renton et al., 2011; Usdin et al., 2015; Haeusler et al., 2016). Most of these diseases are caused by unstable dynamic mutations that usually increase in size during meiotic divisions and have been associated with neurologic disorders (Virtaneva et al., 1997; Cummings and Zoghbi, 2000; Renton et al., 2011; Haeusler et al., 2016). However, poly-A repeats in *HOXD13*, the causative mechanism of a nonneurological condition, synpolydactyly, represents an exception, in which the most likely mechanism leading to increased poly-A tracts are errors in DNA replication (Muragaki et al., 1996; Warren, 1997; Brown and Brown, 2004).

We have shown that an increased number of repeats at 5'UTR of *EIF4A3* causes Richieri-Costa-Pereira syndrome (RCPS; OMIM #268305), a rare autosomal-recessive disorder affecting craniofacial and limb development, mainly described in Brazilian patients (Favaro et al., 2011, 2014; Bertola et al., 2017). RCPS individuals show a distinctive allelic pattern, determined not only by the larger number of repeats (>14 as compared to up to 12 repeats in controls), but also by the presence of a unique motif containing G instead of A nucleotide (the 'disease-associated CGCA-20nt motif') (Favaro et al., 2014). As the origin of the RCPS disease alleles remains unknown, characterizing the 5'UTR of *EIF4A3* in a populational level could give us clues on the mechanisms that originate the *EIF4A3* pathogenic alleles (e.g., meiotic instability or unequal crossing-over events), in addition to providing insights on the chance of RCPS arising in other populations.

We and others have shown that *EIF4A3* downregulation in cellular and animal models leads to defective neural crest cell migration/differentiation and neural stem cell apoptosis during embryonic development, paralleling RCPS cranioskeletal defects and microcephaly,

respectively (Mao et al., 2016; Miller et al., 2017). However, the molecular mechanism responsible for *EIF4A3* downregulation remains entirely unknown.

Therefore, this work was undertaken to investigate the origin of the pathogenic alleles containing the disease-associated CGCA-20nt motif, as well as to evaluate the functional effects of the 5'UTR motifs on *EIF4A3* expression. Insights into the origin and effect of these complex alleles will contribute to a better understanding of regulatory features of 5'UTR regions and their role in craniofacial and neural development.

MATERIALS AND METHODS

Ethics Approval Statement

The protocol was approved by the Ethics Committee of Instituto de Biociências at Universidade de São Paulo, Brazil (accession number 1.463.852). All individuals donated biological samples after providing signed informed consent.

DNA Samples

To characterize *EIF4A3* 5'UTR, 380 DNA samples from unaffected individuals unrelated to RCPS families were selected from the biorepository of CEGH-CEL. For haplotype analysis, 13 additional samples were used, 12 are also from CEGH-CEL (four unaffected individuals without CGCA-20nt motifs and four with CGCA-20nt motifs; four Brazilian RCPS patients bearing different allelic structures) and one sample of a RCPS patient, from the United Kingdom, was sent for diagnosis purposes from North East Thames Genetics Service.

In order to evaluate the effects of the motifs on *EIF4A3* expression, six DNA samples carrying *EIF4A3* alleles with distinct number of motifs were selected from the biorepository of Centro de Estudos do Genoma Humano e Células Tronco (CEGH-CEL) and used for luciferase reporter assay.

For methylation assessment, we used DNA samples of RCPS patients (n = 6; homozygous for the 16 repeats allele) from Hospital de Reabilitação de Anomalias Craniofaciais da

Universidade de São Paulo (HRAC-USP) and unaffected individuals (n = 7; homozygous for the eight repeats allele) from CEGH-CEL.

All samples were extracted from peripheral blood using the Gentra Systems Autopure LS (AutoGen) according to the manufacturer's protocol.

***EIF4A3* 5'UTR Characterization and Haplotype Analysis**

Sanger sequencing of *EIF4A3* 5'UTR (NM_014740.3) and five flanking SNPs (rs11150824, rs2289534, rs3829612, rs10782008, and rs12943620) were performed using BigDye[®] Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) and the ABI 3730 DNA Analyzer (Applied Biosystems). Sequences were analyzed using Sequencher 5.1 (Gene Codes Corporation) and Mixed Sequences Reader (Chang et al., 2012) software. The last one allowed us to better discriminate the alleles. In order to validate our analysis and obtain more reliability in our data, we randomly selected 30 heterozygous samples with different alleles' structure to sequence each allele separately, using the Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare) followed by Sanger sequencing.

To facilitate results interpretation, we referred SNPs markers according to their genomic relative position on the annotated plus strand (GRCh38/hg38): SNP1 (rs11150824) – SNP2 (rs2289534) – *EIF4A3* – SNP3 (rs3829612) – SNP4 (rs10782008) – SNP5 (rs12943620). Primers were designed using Primer-Blast1 (Untergasser et al., 2012) and are described in Supplementary Table S1. Linkage disequilibrium and haplotypes were inferred using Haploview software (Barrett et al., 2005).

Luciferase Assay

Although the larger alleles (15 or 16 repeats) have been associated with decreased *EIF4A3* expression, the causal relationship between number and/or pattern of repeats and gene downregulation is still unknown. Therefore, we investigated the role of the 5'UTR motifs on *EIF4A3* expression by luciferase assay.

Sequences of interest of unaffected and RCPS individuals were amplified by PCR (primer sequences in Supplementary Table S1), purified using the Illustra GFX PCR DNA and Gel Band

Purification kit (GE Healthcare) and cloned into the pGL4.24[luc2P/minP] vector (Promega), upstream of a minimal promoter and the luc2P gene. Sanger sequencing confirmed all constructs: pGL4.24 vectors carrying control alleles with 4 repeats (3 CACA-20-nt and 1 CA-18-nt), 7 repeats (4 CACA-20-nt, 1 CA-18-nt, 1 CACA-20-nt, and 1 CA-18-nt), 10 repeats (7 CACA-20-nt, 1 CA-18-nt, 1 CACA-20-nt, and 1 CA-18-nt) and 12 repeats (1 CACA-20-nt, 10 CGCA-20nt, and 1 CA-18-nt); pGL4.24 vectors carrying pathogenic alleles with 14 repeats (2 CACA-20nt, 10 CGCA-20nt, 1 CACA-20nt, and 1 CA-18nt) and 16 repeats (1 CACA-20nt, 13 CGCA-20nt, 1 CACA-20nt, and 1 CA-18nt). In order to investigate the individual effect of each motif, differing in the composition of central nucleotides, we also constructed vectors carrying sequences with only 1 CA-18-nt, 1 CACA-20-nt, and 1 CGCA-20nt, synthesized by Integrated DNA Technologies (IDT).

Human embryonic kidney (HEK) 293T cells, cultured in high glucose DMEM supplemented with 1% penicillin/streptomycin and 10% fetal bovine serum (FBS) (all provided by Life Technologies), were plated in 96-well plates 24 h prior to transfection (2×10^4 cells/well). Transient transfections were performed in triplicate by using TurboFectin 8.0 (OriGene) according to the manufacturer's instruction. Cells were cotransfected with 180 ng of the pGL4.24 constructs and 20 ng of the pRL-SV40 vector (Promega) containing the Renilla luciferase gene, used as a transfection control. The plasmid pLuc generated from pGL3-control template was used as positive control (Soltys et al., 2013). Forty-eight hours after DNA transfection, luciferase activity was measured with the Dual-Glo R Luciferase Assay System in a GloMax Multi 96-microplate Luminometer (Promega). Firefly luminescence results were normalized by Renilla luminescence and the relative luciferase activity was determined. Statistical analyses were performed with one-way ANOVA and Tukey post hoc test. Significance was set at $p < 0.05$.

Methylation Assay

Indeed, methylated CpGs are involved in gene repression especially when occurring at promoter/5'UTR (Feil and Fraga, 2012; Schübeler, 2015). Since *EIF4A3* 5'UTR motifs are CG-rich and the disease CGCA-20nt allele shows increased number of CpGs compared to control alleles, we evaluated DNA hypermethylation as a plausible mechanism behind *EIF4A3* downregulation in RCPS patients.

One microgram of genomic DNA from each sample were submitted to bisulfite conversion using EpiTect Bisulfite Conversion Kit (QIAGEN). Bisulfite converted DNA was subsequently used for PCR, in which primers were designed with MethPrimer2 (Li and Dahiya, 2002) and are shown in Supplementary Table S1. Amplicons were checked by agarose electrophoresis and cloned using the TOPO TA Cloning Kit for sequencing (Thermo Fisher Scientific). Sanger sequencing was carried out for 10 clones per sample using the BigDye[®] Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) and the ABI 3730 DNA Analyzer (Applied Biosystems). Sequencing files were then analyzed for methylation quantification using the online tool BISMA – Bisulfite Sequencing DNA Methylation Analysis (Rohde et al., 2010) with lower threshold conversion rate at 95%, lower threshold sequence identity at 90%, upper threshold of N-sites at cytosine positions at 20% and per threshold gaps allowed at 20% as filtering parameters. Methylation values were computed and differences between groups tested using Fisher's Exact Test. Significance was set at $p < 0.05$.

RESULTS

The *EIF4A3* 5'UTR is characterized by the presence of 18- or 20-nucleotide-long motifs differing in the composition of central nucleotides, namely CA-18nt (TCGGCAGCGGCAGCGAGG), CACA-20nt (TCGGCAGCGGCACAGCGAGG), and CGCA-20nt (TCGGCAGCGGCAGCGAGG) (Favaro et al., 2014). Unaffected individuals have 3–12 repeats composed mostly by CA-18nt and CACA-20nt motifs, while RCPS patients have 14–16 repeats, with a higher number of the CGCA-20nt motif (Favaro et al., 2014). The motif CGCA-20nt will be herein referred as 'disease-associated CGCA-20nt motif.'

Unaffected Individuals Are Mostly Heterozygous and May Also Present CGCA-20nt Motifs

Sanger sequencing of *EIF4A3* 5'UTR in 380 unaffected individuals revealed 43 different alleles (Figure 1) in heterozygosity in 85% of individuals. The total number of repeats per allele varied from 2 to 17, and the most common alleles contained 7 (25.46%) or 8 (23.21%) repeats

(Supplementary Figure S1). The number and organization of the CA-18nt or CACA-20nt motifs varied between the alleles. We identified not only alleles containing a single motif, but also complex alleles with multiple organizations of the CA-18nt or CACA-20nt motifs (Figure 1). For example, one allele exclusively constituted by the CA-18nt motif was found. As for the other alleles, the number of the first CACA-20nt motifs varied from 2 to 8 with or without different combinations of CA-18nt motifs (Figure 1). The most common allele contained seven repeats and was constituted by four repeats of CACA-20nt motif followed by 1 CA-18nt, 1 CACA-20nt, and 1 final CA-18nt. We also identified two atypical alleles comprising 12- or 20-nucleotide-long sequences inserted between motifs, which do not align to any viral sequences or neighboring genes (data not shown). Finally, the disease-associated CGCA-20nt motif, originally found only among RCPS patients, was identified here in eight heterozygous control individuals, within alleles with 11 or more repeats (Figure 1). The largest allele (17 repeats) was found in heterozygosis in one control individual and contained 14 CGCA-20nt motifs. This allele must be pathogenic when in homozygosis.

Haplotype Analysis Suggests That the RCPS Allele Originated More Than Once

To understand the origin of the alleles containing the disease-associated CGCA-20nt motif, we initially characterized the haplotypes of 13 samples: five affected individuals (four Brazilian and one from the United Kingdom) and eight control individuals (Supplementary Table S2), using five SNPs flanking *EIF4A3* and spanning 519 kb (Figure 2A). We observed a weak linkage disequilibrium in this block ($D' < 0.48$).

The Brazilian RCPS patients showed at least two different haplotypes. The haplotype associated with the pathogenic 14-repeat allele is different from the one associated with the 16-repeat allele (Figure 2B), which in turn were not observed in any of the tested unaffected individuals (Figure 2C). The fact that the two affected alleles with 14 and 16 repeats are embedded within different haplotypes, undetected in unaffected individuals, suggests a distinct origin for these two different-sized pathogenic alleles in our population. Next, we analyzed the homozygous (16 repeats) UK RCPS sample, which showed a different haplotype from those observed in the Brazilian RCPS (Figure 2B). However, one UK RCPS haplotype is similar to a haplotype observed in three of the unaffected individuals carrying the disease-associated CGCA-20nt motif (Figures 2B, D). Besides, the alleles in these control individuals also show similarities in motif organization in relation to the allele in the UK RCPS patient, suggesting that these alleles possibly share a common origin (Figure 2E and Supplementary Table S2).

Transcriptional Activity Is Inversely Correlated to Motif Number at *EIF4A3* 5'UTR

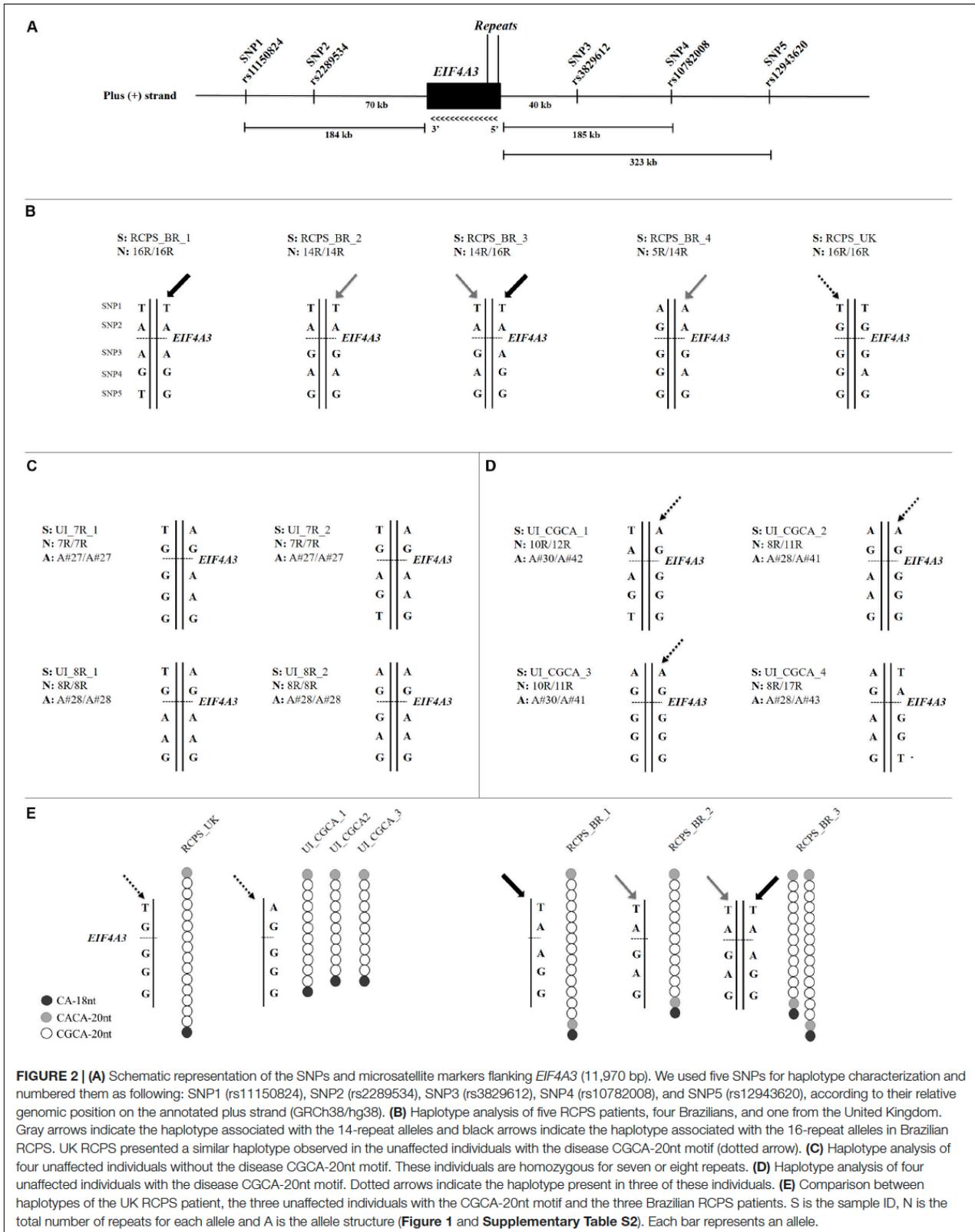
To clarify the functional role of the allelic structure at *EIF4A3* 5'UTR, we generated constructs varying in size and composition (Figure 3), and carried out luciferase reporter assays. Inverse correlation between number of motifs and luciferase activity was observed ($n = 4$ independent experiments; $p < 0.05$; Figure 3A). Further, by investigating each motif individually, there was a discrete reduction in expression for the disease-associated CGCA-20nt, albeit not statistically significant ($n = 3$ independent experiments; Figure 3B).

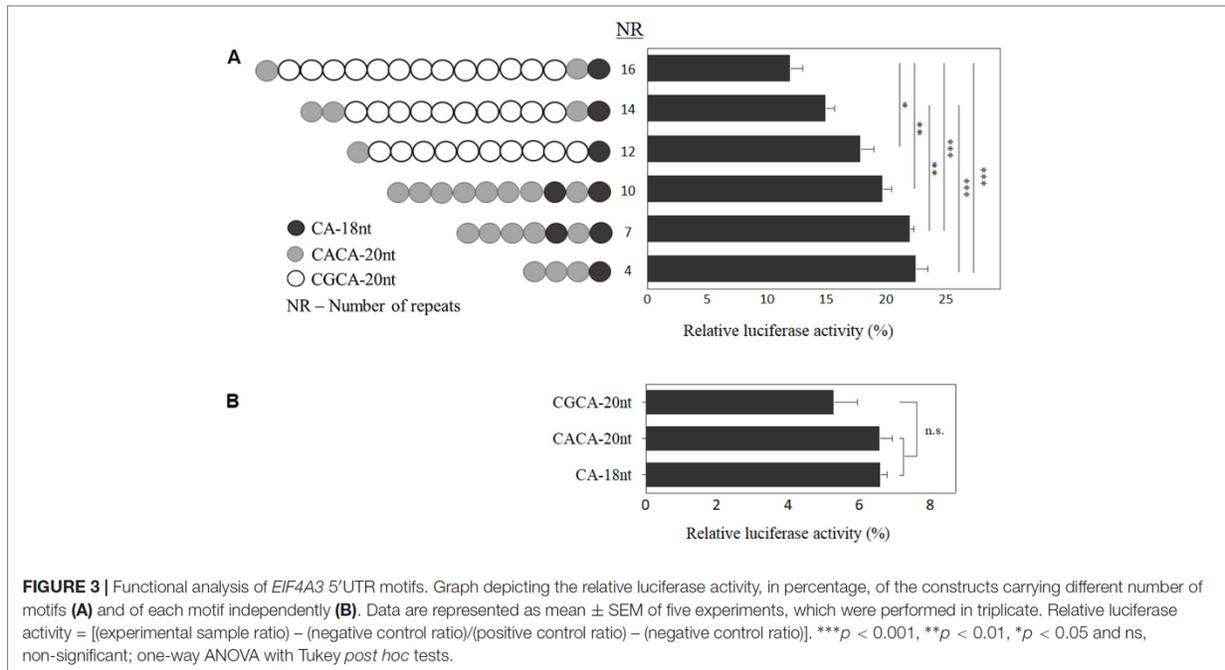
In order to address whether the motif sequence plays a role in *EIF4A3* gene expression, based on data shown in Figure 3A, we calculated the effect per motif type on luciferase activity. Comparing the luciferase activity of alleles with similar composition of motifs, carrying 4 and 10

repeats (22.45% and 19.69%, respectively), we observed a decrease of 2.76% in expression, which represents a reduction of about 0.46% per CACA-20nt motif added in the allele structure. On the other hand, between the alleles carrying 12 and 16 repeats (17.8% and 11.96%, respectively), with comparable allelic structure, the difference was 5.84%, which means a reduction in luciferase activity of 1.46% per CGCA-20nt motif added (three times higher). Based on these results, we suggest that both size and allele sequence play a role in gene regulation.

5'UTR Hypermethylation Is Not Responsible for *EIF4A3* Downregulation

We inspected methylation levels at the 5'UTR of *EIF4A3*, as the increased number of repeats in RCPS patients leads to gain of 37 CpG sites. We observed that both RCPS and controls did not show abundant methylation of this region (1.7% and 2.8% of methylated CpGs, respectively), with a discrete reduction of methylation in RCPS ($p < 0.05$) (Supplementary Table S3). There was no evident methylation variation at any specific CpG between RCPS and controls, as both presented low methylation levels.





DISCUSSION

Expansions at non-coding regions have been extensively described in neurological disorders, and characterization of these regions have greatly contributed to the understanding of novel regulatory mechanisms (Gatchel and Zoghbi, 2005; Mirkin, 2007; La Spada and Taylor, 2010; McMurray, 2010; Russo et al., 2015; Usdin et al., 2015; Haeusler et al., 2016). Despite the great advances in genome sequence analysis, DNA of repetitive regions is still difficult to be sequenced. In fact, the *EIF4A3* 5'UTR is not covered in GnomAD database, which reinforces the importance of characterizing this region through Sanger sequencing. In this study, Sanger sequencing analysis of *EIF4A3* 5'UTR in 380 unaffected individuals revealed 43 different alleles, with the most common alleles containing seven or eight repeats. Some of these alleles presented only one type of motif (CA-18nt or CACA-20nt) with different total number of repeats, while others presented a visible combination of these two common motifs, suggesting that these alleles may have originated through unequal crossing-over events. These results show its polymorphic nature and confirm the structural complexity, and uniqueness of this region, which

was not comparable to any gene in which dynamic pathogenic expansions at non-coding regions had been reported (Brook et al., 1992; Campuzano et al., 1996; Moseley et al., 2006; Daughters et al., 2009; Galloway and Nelson, 2009; Sato et al., 2009; DeJesus-Hernandez et al., 2011; Kobayashi et al., 2011). In this enlarged cohort, 1% of the alleles (8/760) harbors the disease-associated CGCA-20nt motif. Interestingly, the largest alleles in this cohort (≥ 11 repeats), including one with 17 repeats, contained the disease CGCA- 20nt motif. These results suggest that RCPS could occur in any population containing alleles with the CGCA-20nt motif. Indeed, one of the patients here included is from United Kingdom (Bertola et al., 2017).

The haplotype analyses were performed in order to get insights on the origin of the pathogenic alleles with increased number of repeats. Results revealed that the pathogenic alleles with 14 and 16 repeats of Brazilian patients have distinct origins, which in turn are different from the haplotypes of the UK RCPS patient. These results suggest that the pathogenic alleles have arisen more than once. It is of note that the UK RCPS patient shares a common haplotype and also a similar motif structure with three unaffected individuals carrying the disease-associated CGCA-20nt motif, suggesting a common ancestral among them. It is possible that, similarly to the alleles in the control population, the affected alleles may have arisen also through unequal crossing. This hypothesis is also supported by the observation that the number of repeats at *EIF4A3* 5'UTR seems to be stable across generations (Favaro et al., 2014). This phenomenon is more comparable to the one observed in synpolydactyly, in which the poly-A at the 3'end of *HOXD13* might have originated by unequal crossing over as it is quite stable when transmitted across generations (Warren, 1997), as opposed to dynamic mutations observed in neurological conditions, which have arisen only once (Virtaneva et al., 1997; Gatchel and Zoghbi, 2005; Mirkin, 2007; La Spada and Taylor, 2010; McMurray, 2010; DeJesus-Hernandez et al., 2011; Renton et al., 2011; Usdin et al., 2015; Haeusler et al., 2016).

Next, we demonstrated that the number and allele sequence of motifs at 5'UTR is involved in *EIF4A3* expression. These results thus suggest a potential cis-acting regulatory mechanism for these motifs on gene expression and confirm our previous finding that the affected alleles were associated with *EIF4A3* downregulation in cells from different tissues from

RCPS patients, including peripheral blood (Favaro et al., 2014; Miller et al., 2017). DNA hypermethylation is not the mechanism for *EIF4A3* downregulation in RCPS patients, as RCPS and control blood samples showed similar methylation levels. Hypomethylation, as observed at *EIF4A3* 5'UTR, is consistent with transcriptional activity (Baubec and Schübeler, 2014; Schübeler, 2015). There could be alternative mechanisms by which the increased number of motifs could repress *EIF4A3* expression: the disease CGCA-20nt could create binding sites for repressor proteins and act as a silencer; or more complex mechanisms, including post-translational events, could be involved. Further functional studies are needed to pinpoint the exact molecular mechanism underlying *EIF4A3* downregulation.

The inverse correlation between number of repeats and *EIF4A3* expression levels suggests that the structure of the 5'UTR *EIF4A3* may modulate the phenotype. Indeed, a broader phenotypic spectrum within RCPS has been observed, ranging from individuals with severe phenotype (homozygous for the 16-repeat allele or heterozygous for the 15- and 16-repeat alleles) to individuals with less severe skeletal involvements, harboring smaller number of motifs (homozygous for the 14- repeat allele or compound heterozygous for 14 repeats and a point mutation (c.809A>G) (Favaro et al., 2014; Bertola et al., 2017). Despite limitations in sample size, these results suggest that RCPS phenotypic variability depends upon the number and allele sequence of repeats at the 5'UTR of *EIF4A3*.

CONCLUSION

In summary, we provide evidence supporting that the *EIF4A3* 5'UTR is highly polymorphic, comprising at least 43 different alleles, which may have originated through extensive recombination in this region. Haplotype analysis in control and in affected individuals suggests that there are more than one haplotype associated with the disease and RCPS alleles might have originated from unequal crossing-over events. We also provided further evidence that *EIF4A3* 5'UTR is a regulatory region, and that *EIF4A3* downregulation in RCPS is not mediated by CpG methylation. Moreover, our findings provide insights to explain clinical variability in RCPS.

AUTHOR CONTRIBUTIONS

DB, RP, and MZ provided the Brazilian DNA samples. AG and EW provided DNA sample and clinical information from the UK RCPS patient. GH, CM, and MP-B conceived and designed the study. GH performed the *EIF4A3* 5'UTR characterization and haplotype analyses. LB assisted with haplotype and linkage disequilibrium analyses. CM carried out the luciferase assay and LA performed the methylation assay. GH, CM, LA, LB, GK, and MP-B discussed and interpreted the results and wrote the manuscript. GH and CM have a major contribution in writing the manuscript. All authors contributed to the final version and approved the manuscript.

FUNDING

This work was funded by Centro de Pesquisa, Inovação e Difusão (CEPID, 2013/08028-1), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 2015/21781-6), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, PQ10/2011), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, 1470032).

ACKNOWLEDGMENTS

We thank Dr. Meire Aguenta and Dr. Antonia Cerqueira (IBUSP) for technical support. We also thank Dr. Marcelo Nobrega and Dr. Débora Sobreira (University of Chicago, United States) who donated the sequences for constructs used in the luciferase assay.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

- Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457
- Baubec, T., and Schübeler, D. (2014). Genomic patterns and context specific interpretation of DNA methylation. *Curr. Opin. Genet. Dev.* 25, 85–92. doi: 10.1016/j.gde.2013.11.015
- Bertola, D. R., Hsia, G., Alvizi, L., Gardham, A., Wakeling, E. L., Yamamoto, G. L., et al. (2017). Richieri-Costa-Pereira syndrome: expanding its phenotypic and genotypic spectrum. *Clin. Genet.* 93, 800–811. doi: 10.1111/cge.13169
- Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., et al. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 30 end of a transcript encoding a protein kinase family member. *Cell* 68, 799–808. doi: 10.1016/0092-8674(92)90154-5
- Brown, L. Y., and Brown, S. A. (2004). Alanine tracts: the expanding story of human illness and trinucleotide repeats. *Trends Genet.* 20, 51–58. doi: 10.1016/j.tig.2003.11.002
- Campuzano, V., Montermini, L., Moltò, M. D., Pianese, L., Cossée, M., Cavalcanti, F., et al. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271, 1423–1427. doi: 10.1126/science.271.5254.1423
- Chang, C.-T., Tsai, C.-N., Tang, C. Y., Chen, C.-H., Lian, J.-H., Hu, C.-Y., et al. (2012). Mixed sequence reader: a program for analyzing DNA sequences with heterozygous base calling. *ScientificWorldJournal* 2012:365104. doi: 10.1100/2012/365104
- Cummings, C. J., and Zoghbi, H. Y. (2000). Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.* 9, 909–916. doi: 10.1093/hmg/9.6.909
- Daughters, R. S., Tuttle, D. L., Gao, W., Ikeda, Y., Moseley, M. L., Ebner, T. J., et al. (2009). RNA gain-of-function in spinocerebellar ataxia type 8. *PLoS Genet.* 5:e1000600. doi: 10.1371/journal.pgen.1000600
- de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7:e1002384. doi: 10.1371/journal.pgen.1002384
- DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245–256. doi: 10.1016/j.neuron.2011.09.011
- Favaro, F. P., Alvizi, L., Zechi-Ceide, R. M., Bertola, D., Felix, T. M., de Souza, J., et al. (2014). A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. *Am. J. Hum. Genet.* 94, 120–128. doi: 10.1016/j.ajhg.2013.11.020
- Favaro, F. P., Zechi-Ceide, R. M., Alvarez, C. W., Maximino, L. P., Antunes, L. F., Richieri-Costa, A., et al. (2011). Richieri-Costa-Pereira syndrome: a unique acrofacial dysostosis type. An overview of the Brazilian cases. *Am. J. Med. Genet. A* 155A, 322–331. doi: 10.1002/ajmg.a.33806
- Feil, R., and Fraga, M. F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* 13, 97–109. doi: 10.1038/nrg3142
- Galloway, J. N., and Nelson, D. L. (2009). Evidence for RNA-mediated toxicity in the fragile X-associated tremor/ataxia syndrome. *Future Neurol.* 4:785. doi: 10.2217/fnl.09.44
- Gatchel, J. R., and Zoghbi, H. Y. (2005). Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.* 6, 743–755. doi: 10.1038/nrg1691
- Haeusler, A. R., Donnelly, C. J., and Rothstein, J. D. (2016). The expanding biology of the C9orf72 nucleotide repeat expansion in neurodegenerative disease. *Nat. Rev. Neurosci.* 17, 383–395. doi: 10.1038/nrn.2016.38

Kobayashi, H., Abe, K., Matsuura, T., Ikeda, Y., Hitomi, T., Akechi, Y., et al. (2011). Expansion of intronic GGCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. *Am. J. Hum. Genet.* 89, 121–130. doi: 10.1016/j.ajhg.2011.05.015

La Spada, A. R., and Taylor, J. P. (2010). Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* 11, 247–258. doi: 10.1038/nrg2748

Li, L.-C., and Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18, 1427–1431. doi: 10.1093/bioinformatics/18.11.1427

Mao, H., McMahon, J. J., Tsai, Y. H., Wang, Z., and Silver, D. L. (2016). Haploinsufficiency for core exon junction complex components disrupts embryonic neurogenesis and causes p53-mediated microcephaly. *PLoS Genet.* 12:e1006282. doi: 10.1371/journal.pgen.1006282

McMurray, C. T. (2010). Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet.* 11, 786–799. doi: 10.1038/nrg2828

Miller, E. E., Kobayashi, G. S., Musso, C. M., Allen, M., Ishiy, F. A. A., de Caires, L. C., et al. (2017). EIF4A3 deficient human iPSCs and mouse models demonstrate neural crest defects that underlie Richieri-Costa-Pereira syndrome. *Hum. Mol. Genet.* 26, 2177–2191. doi: 10.1093/hmg/ddx078

Mirkin, S. M. (2007). Expandable DNA repeats and human disease. *Nature* 447, 932–940. doi: 10.1038/nature05977

Moseley, M. L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A. K., Daughters, R. S., et al. (2006). Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat. Genet.* 38, 758–769. doi: 10.1038/ng1827

Muragaki, Y., Mundlos, S., Upton, J., and Olsen, B. R. (1996). Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* 272, 548–551. doi: 10.1126/science.272.5261.548

Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., et al. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72, 257–268. doi: 10.1016/j.neuron.2011.09.010

Rohde, C., Zhang, Y., Reinhardt, R., and Jeltsch, A. (2010). BISMA—fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics* 11:230. doi: 10.1186/1471-2105-11-230

Russo, A., Pacchierotti, F., Cimini, D., Ganem, N. J., Genescà, A., Natarajan, A. T., et al. (2015). Genomic instability: crossing pathways at the origin of structural and numerical chromosome changes. *Environ. Mol. Mutagen.* 56, 563–580. doi: 10.1002/em.21945

Sato, N., Amino, T., Kobayashi, K., Asakawa, S., Ishiguro, T., Tsunemi, T., et al. (2009). Spinocerebellar ataxia type 31 is associated with “inserted” pentanucleotide repeats containing (TGGAA)_n. *Am. J. Hum. Genet.* 85, 544–557. doi: 10.1016/j.ajhg.2009.09.019

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature* 517, 321–326. doi: 10.1038/nature14192

Soltys, D. T., Rocha, C. R., Lerner, L. K., de Souza, T. A., Munford, V., Cabral, F., et al. (2013). Novel XPG (ERCC5) mutations affect DNA repair and cell survival after ultraviolet but not oxidative stress. *Hum. Mutat.* 34, 481–489. doi: 10.1002/humu.22259

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40:e115. doi: 10.1093/nar/gks596

Usdin, K., House, N. C., and Freudenreich, C. H. (2015). Repeat instability during DNA repair: insights from model systems. *Crit. Rev. Biochem. Mol. Biol.* 50, 142–167. doi: 10.3109/10409238.2014.999192

Virtaneva, K., D’Amato, E., Miao, J., Koskiniemi, M., Norio, R., Avanzini, G., et al. (1997). Unstable minisatellite expansion causing recessively inherited myoclonus epilepsy, EPM1. *Nat. Genet.* 15, 393–396. doi: 10.1038/ng0497-393

Warren, S. T. (1997). Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13. *Science* 275, 408–409. doi: 10.1126/science.275.5298.408

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00149/full#supplementary-material>

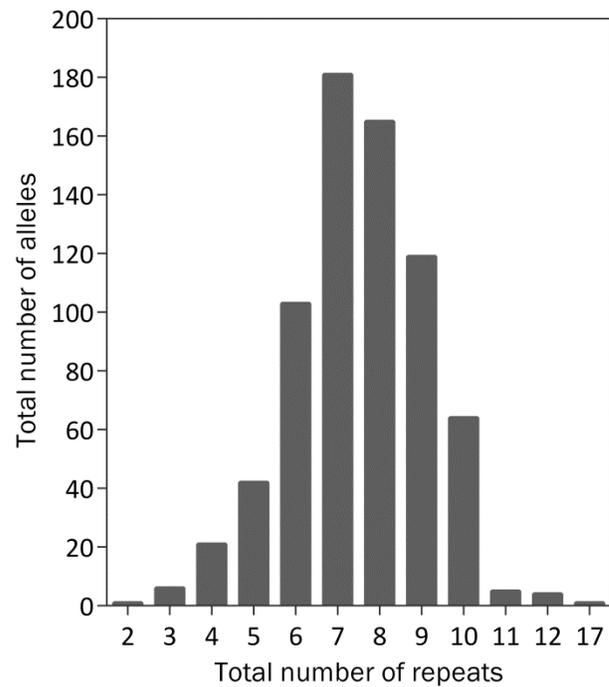


Figure S1. Distribution of number of alleles per total number of repeats identified in the 380 samples of unaffected individuals. This analysis did not take in consideration the motifs' structure.

Table S1. Primer sequences

Target	Forward strand (5' → 3')	Reverse strand (5' → 3')
<i>EIF4A3</i> Luciferase assay	AAAGGTACCACGCCCAGTTCCTTTCAC	AAAAAGCTTGAACGTGGGGGTCACATC
<i>EIF4A3</i> Methylation analysis	GYGAGAGTAGAAATATTTTATTTTTT	TCATATCTTCCTCTTTAAACAACC
<i>EIF4A3</i> Characterization analysis	ACGCCCAGTTCCTTTCAC	GAACGTGGGGGTCACATC
rs11150824	TGAGCAAAATATGGCAGGAAGTC	AGAATGCCCAGCGGTTGAC
rs2289534	ACAGAGGGTGACCTTGATCC	AAAGGGACACCTTTACCTCCC
rs3829612	GAAGTCTGCACGTCGGGGT	AGTCTCATGTTTTCCGGTTCCC
rs10782008	TGACTTCTGACACTTTCTTTTCGC	AGCTGGAAGATGTGGCTGTC
rs12943620	GACGCACAAAACAGATCATCGC	ACATCTGGGACCGCAAGCTG

Table S2. List of the 13 samples used in haplotype analysis and the description of motifs' structure reported in each allele

Sample ID	Sample description	Number of repeats (R)	Alleles' structure	
UI_7R_1	Unaffected individual	7R/7R	A#27	A#27
UI_7R_2	Unaffected Individual	7R/7R	A#27	A#27
UI_8R_1	Unaffected Individual	8R/8R	A#28	A#28
UI_8R_2	Unaffected Individual	8R/8R	A#28	A#28
UI_CGCA_1	Unaffected Individual with CGCA-20nt	10R/12R	A#30	A#42
UI_CGCA_2	Unaffected Individual with CGCA-20nt	8R/11R	A#28	A#41
UI_CGCA_3	Unaffected Individual with CGCA-20nt	10R/11R	A#30	A#41
UI_CGCA_4	Unaffected Individual with CGCA-20nt	8R/17R	A#28	A#43
RCPS_BR_1	RCPS Brazilian	16R/16R	1 CACA-20nt + 13 CGCA-20nt + 1 CACA-20nt + 1 CA-18nt	1 CACA-20nt + 13 CGCA-20nt + 1 CACA-20nt + 1 CA-18nt
RCPS_BR_2	RCPS Brazilian	14R/14R	1 CACA-20nt + 11 CGCA-20nt + 1 CACA-20nt + 1 CA-18nt	1 CACA-20nt + 11 CGCA-20nt + 1 CACA-20nt + 1 CA-18nt
RCPS_BR_3	RCPS Brazilian	14R/16R	1 CACA-20nt + 11 CGCA-20nt + 1 CACA-20nt + 1 CA-18nt	1 CACA-20nt + 13 CGCA-20nt + 1 CACA-20nt + 1 CA-18nt
RCPS_BR_4	RCPS Brazilian	5R/14R	4 CACA-20nt + 1 CA-18nt	2 CACA-20nt + 2 CGCA-20nt + 1 CACA-20nt + 3 CGCA-20nt + 1 CACA-20nt + 3 CGCA-20nt + 2 CACA-20nt
RCPS_UK	RCPS London	16R/16R	1 CACA-20nt + 14 CGCA-20nt + 1 CA-18nt	1 CACA-20nt + 14 CGCA-20nt + 1 CA-18nt

Table S3. Methylation levels in the analyzed regions in RCPS patients and control individuals; $p=0.0461$, Fisher's Exact Test

	RCPS		Controls	
	Number of CpGs	%	Number of CpGs	%
Methylated CpGs	40	1.7	29	2.8
Non-methylated CpGs	2352	98.3	1011	97.2
Total	2392		1040	

CHAPTER III

***EIF4A3* Deficient Human iPSCs and Mouse Models Demonstrate Neural Crest Defects that Underlie Richieri-Costa-Pereira Syndrome**

Emily E. Miller^{1†}, Gerson S. Kobayashi^{2†}, Camila M. Musso², Miranda Allen¹, Felipe A. A. Ishiy², Luiz C. de Caires Jr², Ernesto Goulart², Karina Griesi-Oliveira², Roseli M. Zechi-Ceide³, Antonio Richieri-Costa³, Debora R. Bertola², Maria Rita Passos-Bueno^{2*}, Debra L. Silver^{1,4,5,6*}

¹Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, USA

²Centro de Estudos do Genoma Humano e Células Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

³Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo, Bauru, Brazil

⁴Department of Neurobiology, ⁵Department of Cell Biology and ⁶Duke Institute for Brain Sciences, Duke University Medical Center, Durham, USA

†Co-first authors

*Corresponding author

Keywords: phenotype, congenital abnormality, mutation, cartilage, chondrogenesis, clavicle, craniofacial abnormalities, embryo, micrognathism, neural crest, osteogenesis, mandible, Meckel's diverticulum, mice

This chapter was published in *Human Molecular Genetics*, 2017

Miller, E.E., Kobayashi, G.S., Musso, C.M., et al. 2017. *EIF4A3* deficient human iPSCs and mouse models demonstrate neural crest defects that underlie Richieri-Costa-Pereira syndrome. *Human Molecular Genetics* 26(12), pp. 2177–2191.

<https://doi.org/10.1093/hmg/ddx078>

ABSTRACT

Biallelic loss-of-function mutations in the RNA-binding protein EIF4A3 cause Richieri-Costa-Pereira syndrome (RCPS), an autosomal recessive condition mainly characterized by craniofacial and limb malformations. However, the pathogenic cellular mechanisms responsible for this syndrome are entirely unknown. Here, we used two complementary approaches, patient-derived induced pluripotent stem cells (iPSCs) and conditional *Eif4a3* mouse models, to demonstrate that defective neural crest cell (NCC) development explains RCPS craniofacial abnormalities. RCPS iPSCs have decreased migratory capacity, a distinct phenotype relative to other craniofacial disorders. *Eif4a3* haploinsufficient embryos presented altered mandibular process fusion and micrognathia, thus recapitulating the most penetrant phenotypes of the syndrome. These defects were evident in either ubiquitous or NCC-specific *Eif4a3* haploinsufficient animals, demonstrating an autonomous requirement of *Eif4a3* in NCCs. Notably, RCPS NCC-derived mesenchymal stem-like cells (nMSCs) showed premature bone differentiation, a phenotype paralleled by premature clavicle ossification in *Eif4a3* haploinsufficient embryos. Likewise, nMSCs presented compromised *in vitro* chondrogenesis, and Meckel's cartilage was underdeveloped *in vivo*. These findings indicate novel and essential requirements of *EIF4A3* for NCC migration and osteochondrogenic differentiation during craniofacial development. Altogether, complementary use of iPSCs and mouse models pinpoint unique cellular mechanisms by which *EIF4A3* mutation causes RCPS, and provide a paradigm to study craniofacial disorders.

INTRODUCTION

Richieri-Costa–Pereira syndrome (RCPS; OMIM #268305) is a rare autosomal-recessive acrofacial dysostosis in which mandible development, particularly the distal and medial regions, may be severely affected. It is characterized by midline mandibular cleft usually associated with failure of mandibular symphyseal fusion, Robin sequence, laryngeal abnormalities, radial and tibial defects, among other clinical findings (1). Phenotype expressivity is variable, ranging from only mild defects in mandibular fusion or laryngeal clefts to more severe clinical manifestations associated with agenesis of the mandible (1-3).

RCPS is mainly caused by non-coding expansions in the 5'UTR of the gene *EIF4A3*. This region is characterized by multiple allelic patterns that vary in size and organization of repeat motifs. The most prevalent allelic pattern among normal individuals has 5-12 repeats, whilst the majority of affected individuals harbor 14–16 repeats. In addition, a missense mutation has been found in trans to an expansion allele in one RCPS patient (3). These alterations are believed to cause partial loss of function of *EIF4A3*, as 30–40% reduction of *EIF4A3* mRNA expression has been reported in RCPS patients' lymphocytes and adult mesenchymal cells.

EIF4A3 encodes a DEAD-box helicase that is a core component of the RNA-binding exon junction complex (EJC), which controls post-transcriptional events, including alternative splicing, non-sense-mediated mRNA decay, translation initiation and RNA localization (4). *EIF4A3* binding to mRNA is stabilized by two additional core components, MAGOH and RBM8A. *Eif4a3* knockdown in zebrafish results in embryos with alterations in craniofacial cartilage/bone development and clefting of the lower jaw (3). *Eif4a3* is also required in *Xenopus* for embryonic development including formation of the peripheral nervous system and melanocytes (5,6). In mice, conditional haploinsufficiency of *Eif4a3* or other EJC components in the brain impairs neural progenitor proliferation, differentiation and survival (7). Although a relationship between *EIF4A3* loss of function and RCPS has been established, the requirement of *Eif4a3* for mammalian embryonic craniofacial development and the pathogenic cellular mechanism responsible for the syndrome is entirely unknown.

The craniofacial structures compromised in RCPS are suggestive of disturbances in neural crest or neural crest-derived tissue development. Neural crest cells (NCCs) are a transient cell population originating from the neuroectoderm located at the neural plate border during neurulation. The foremost segment of the neuraxis gives rise to cranial NCCs, which migrate and populate the mesenchyme of the developing pharyngeal arches. Mandibular arches fuse to form a mesenchymal mandible. The cranial NCC-derived mesenchyme then undergoes proliferation and differentiation, generating most of the cranium (including Meckel's cartilage, followed by the mandible), ear components and larynx, amongst others (8–10). NCC defects in apoptosis, proliferation, and cell migration underlie some craniofacial disorders, including Treacher Collins syndrome and Nager syndrome, raising the question as to their potential role in RCPS pathology (11-13). Furthermore, NCC mesenchymal differentiation could also play a role in RCPS (14). In order to understand the etiology of RCPS it is critical to define which, if any, of these cellular mechanisms are relevant.

In this study, we addressed these gaps, using two novel models, induced pluripotent stem cells (iPSCs) and mouse mutants, with a particular focus on mandible development. Owing to the natural limitations in studying human embryos, integration of these approaches offers invaluable insight into human craniofacial malformations; NCCs and their derivatives can be generated from patients and screened for disease-relevant phenotypes (15,16), which can be further examined in mice. We found that RCPS patient-derived iPSCs differentiated toward a NCC lineage exhibited autonomous defects in cell migration and those differentiated into mesenchymal derivatives were prone to premature ossification and altered chondrogenesis. *Eif4a3* haploinsufficiency in mice caused altered fusion of mandibular processes, impaired development of Meckel's cartilage, and premature skeletal ossification. These defects were caused by an autonomous requirement of *Eif4a3* in NCC and resulted in significant loss of mandibular structures, thus modelling common malformations of RCPS. Together, complementary use of iPSCs and mouse models pinpoint developmental mechanisms by which *EIF4A3* mutation causes RCPS.

RESULTS

Derivation of iPSC cultures

We first set out to establish iPSC lines from RCPS patients. Two of the RCPS patients (F8417-1 and F8417-2) were homozygous for the 16-repeat allele, while a third patient (F6099-1) had a 14-repeat allele in trans with the missense mutation p.Asp270Gly (3); controls were homozygous for the 6-repeat allele (F9048-1) and heterozygous for 7-repeats/6-repeats (F7405-1 and F8799-1) (Supplementary Material, Table S1; see Materials and Methods). Information regarding sample usage in experiments is shown in Supplementary Material, Table S1. After reprogramming, all iPSC lines displayed pluripotent stem cell-like morphology and positive staining for pluripotency markers OCT3/4 and SSEA-4 (Supplementary Material, Fig. S1A), as well as expression of *OCT3/4*, *NANOG* and *ALP* transcripts (Supplementary Material, Fig. S1B-D). Further, iPSCs were able to generate teratoma-containing tissues from all three germ layers *in vivo* (Supplementary Material, Fig. S1F-H). All iPSCs showed no detectable signs of aneuploidy or genomic integration of the episomal vectors (Supplementary Material, Fig. S1E and I).

NCCs can be derived from RCPS iPSCs and exhibit migration defects

iPSC-derived NCCs (iNCCs) were induced from RCPS and control iPSCs under a methodology based on TGF- β /Activin pathway blockade and WNT pathway activation without the presence of caudalization factors (17,18), which has been shown to favor cranial NCC formation (19). After differentiation, iNCC populations were positively stained for NCC markers p75 and HNK1 with proportions of double-positive cells ranging between 84.7 and 97.6% in both control and patients' cells (Supplementary Material, Fig. S2A-H). Furthermore, RT-qPCR assays showed upregulation of NCC markers *PAX3*, *SOX10*, *ZIC1* and *TFAP2A*, and downregulation of the pluripotency marker *OCT3/4*, when compared with the originating iPSCs (Supplementary Material, Fig. S2I-M). Moreover, RCPS iNCCs showed reduction of ~70% in *EIF4A3* mRNA and ~50% in protein expression compared with control iNCCs (Fig. 1A and B).

Alterations involving proliferation, apoptosis, or migration of NCCs or their cellular derivatives are considered to be underlying mechanisms in a number of NCC-related diseases (9,20). To screen for cellular phenotypes associated with RCPS, we examined cell cycle, apoptosis and migration of RCPS iNCCs, compared with controls. No significant differences in cell cycle distribution or basal apoptosis were detected between patients and controls (Supplementary Material, Fig. S3A-G). Wound-healing migration assays, however, revealed significantly decreased migratory capacity in RCPS iNCCs when compared with iNCCs from control individuals (Fig. 1C and D). This phenotypic assessment reveals delayed NCC migration may contribute to early defects of RCPS.

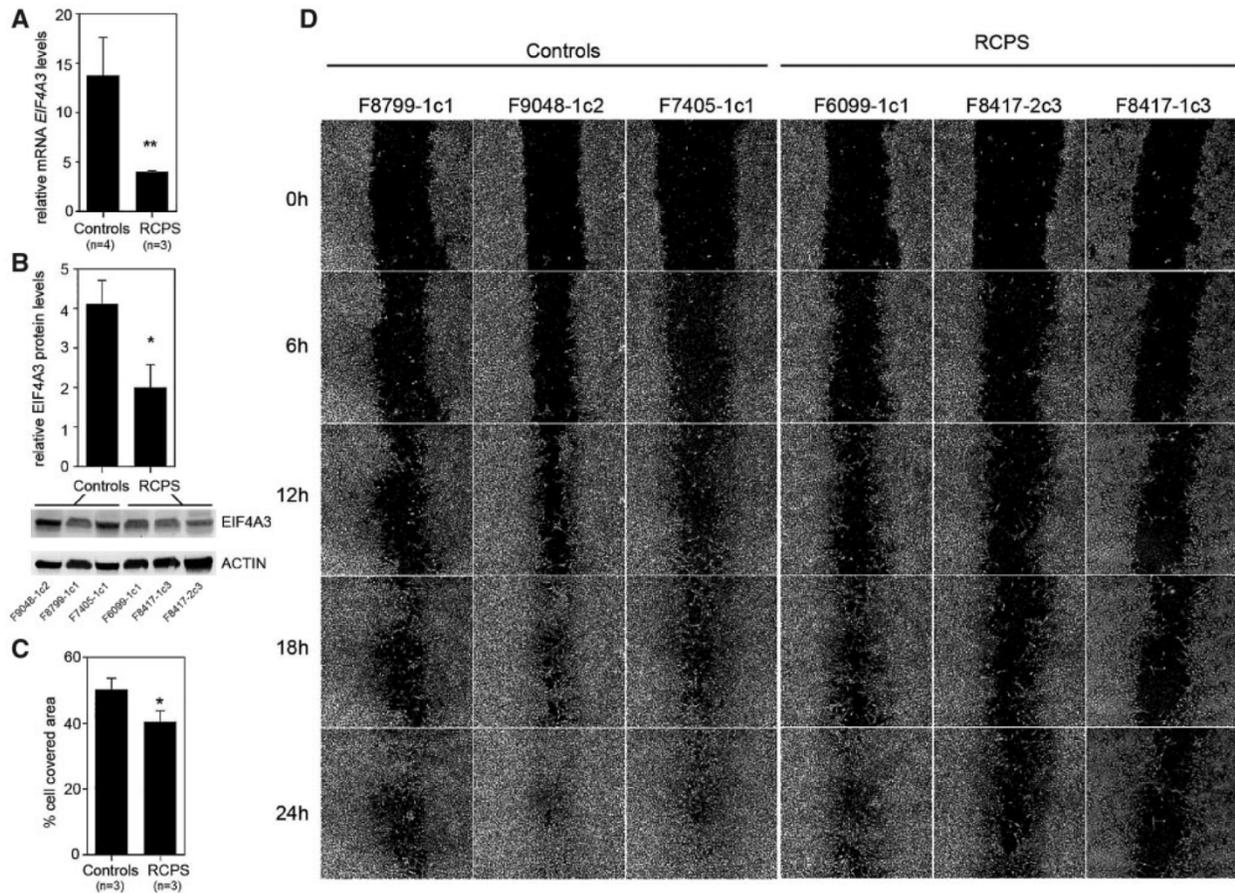


Figure 1. Reduced cell migration in RCPS iNCCs. **(A, B)** RT-qPCR assessment of *EIF4A3* mRNA expression **(A)** and protein levels **(B)**, with western blot assay showing relative EIF4A3/b-actin protein level, in iPSCs from controls and RCPS patients. **(C)** Bar graph depicting the rate of cell migration [cell-covered area (%)], at 24 h; data shown are representative of two independent assays and three independent measurements

in each. (D) Representative phase-contrast micrographs acquired immediately after wounding and at 6, 12, 18 and 24h afterwards; All values represent mean \pm SEM, and sample numbers (n) are indicated *P-value <0.05, **P-value <0.01; Student's t-test.

Mesenchymal stem-like cells can be derived from iNCCs and exhibit differentiation defects

One of the hallmarks of the cranial neural crest is the ability to generate mesenchymal precursors that give rise to the craniofacial skeleton and other structures during embryonic development (9,21). Therefore, to screen for cellular phenotypes associated with the RCPS phenotype, we generated NCC-derived mesenchymal stem-like cells (nMSCs) from RCPS and control iNCCs. nMSCs showed significant transcriptional upregulation of mesenchymal marker *ENG* (CD105) and downregulation of neural crest marker *TFAP2A* in comparison to iNCCs (Supplementary Material, Fig. S4A and B). Moreover, nMSCs exhibited typical and homogeneous mesenchymal immunophenotype, with positive staining (>82%) for mesenchymal markers CD73, CD166 and CD90, and negative staining (<5%) for endothelial marker CD31 (Supplementary Material, Fig. S4C); accordingly, nMSCs could be differentiated into mesenchymal derivatives *in vitro* (Supplementary Material, Fig. S4D). These results are in agreement with the expected mesenchymal stem cell phenotype seen in previous work by our group (22-24). Finally, a discrete, significant reduction (~15%) of *EIF4A3* mRNA expression was observed in RCPS nMSCs in comparison to controls (Supplementary Material, Fig. S5A).

EIF4A3* deficiency is associated with premature osteogenic differentiation *in vitro

Next, we hypothesized that the craniofacial phenotypes seen in RCPS could also be caused by disturbances in NCC mesenchymal differentiation, leading to alterations in cartilage formation and/or intramembranous ossification, particularly during mandible development. We therefore investigated chondrogenic and osteogenic differentiation in nMSCs from RCPS and control individuals. Since differences in cell density may influence results of *in vitro* chondrogenic and osteogenic differentiation, all nMSCs were first subjected to proliferation and apoptosis assays. No significant differences in either parameter were detected between RCPS and control nMSCs (Supplementary Material, Fig. S5B and C).

We performed chondrogenic differentiation and then assessed mRNA expression of key chondrogenesis markers, *SOX9* and *ACAN*. In relation to controls, RCPS nMSCs showed augmented expression of early chondrogenesis marker *SOX9* and lower expression of late chondrogenesis marker *ACAN*, after 9 days of differentiation (Fig. 2A and B). During osteogenic differentiation, alkaline phosphatase (ALP) enzymatic activity was significantly increased in RCPS cells after 9 days of osteoinduction, and Alizarin Red staining revealed greater matrix mineralization evident as brown precipitate after 21 days, in comparison to controls (Fig. 2C-E). Next, we assessed expression of key genes involved in osteogenesis during the initial 6 days of differentiation. Compared with controls, at day 6 of osteoinduction, RCPS cells exhibited upregulation of transcription factor *RUNX2*, as well as upregulation of early osteoblast marker *COL1A1*, and late-stage osteoblast marker *BGLAP*. At day 6 of differentiation, ALP expression was higher in RCPS samples, but did not reach statistical significance (Fig. 2F-I). These findings suggest that RCPS nMSCs show premature osteogenic differentiation in comparison to controls. Altogether, use of iNCCs and nMSCs point to defective NCC migration and dysregulation of osteo/chondrogenesis in RCPS pathology.

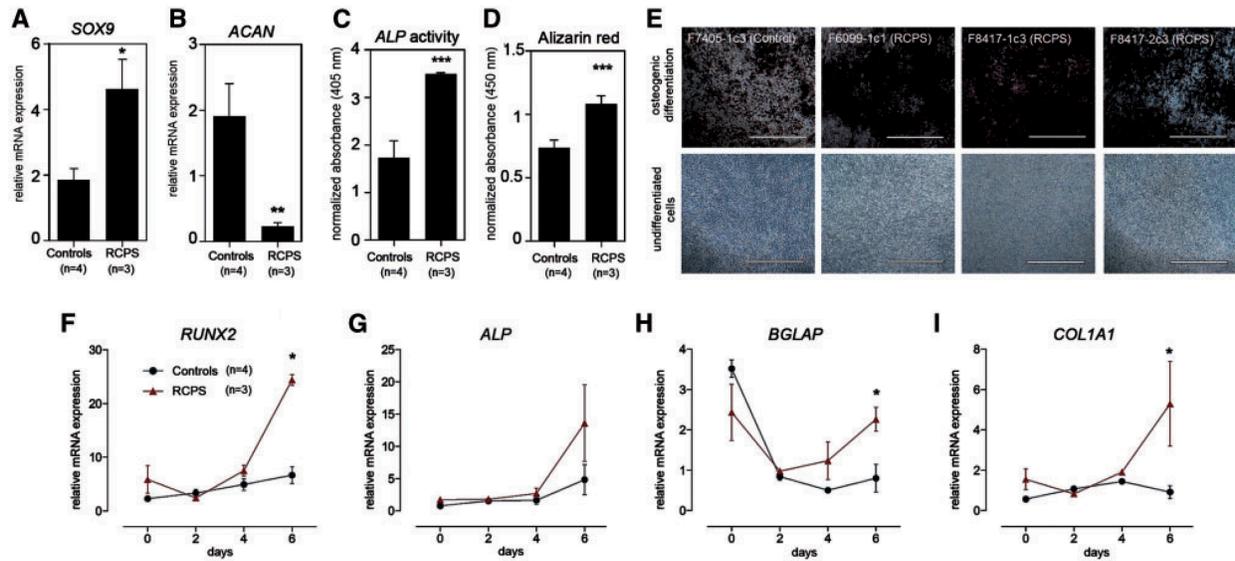


Figure 2. Alterations in osteogenic potential in RCPS nMSCs. **(A, B)** RT-qPCR assessment of *SOX9* **(A)**, *ACAN* **(B)** in nMSCs from controls and RCPS patients. **(C)** Quantification of alkaline phosphatase (ALP) enzymatic activity, after 9 days, and **(D)** alizarin red after 21 days of osteoinduction, in RCPS cells in comparison to controls. Measurements from differentiated cells were normalized to paired, undifferentiated negative staining controls. **(E)** Representative alizarin red staining micrographs showing matrix mineralization (in dark brown) of RCPS samples vs. one representative control (osteogenic differentiation); micrographs are shown paired to respective negative controls (undifferentiated cells). **(F-I)** Transcriptional profile of osteogenesis markers (*RUNX2*, *ALP*, *BGLAP* and *COL1A1*) during the initial 6 days of osteoinduction. All values represent mean \pm SEM, and sample numbers (n) are indicated. **(A, B)** Two-way ANOVA with Bonferroni post-tests; **(F-I)** Student's t-test; *P-value <0.05, **P-value <0.01, ***P-value <0.001, Scale bars, **(E)** 1000 μ m.

***Eif4a3* germline haploinsufficient embryos are developmentally delayed with craniofacial abnormalities**

To complement analyses of human iPSCs, we next investigated *Eif4a3* function in craniofacial development using mouse models. We generated germline *Eif4a3* haploinsufficient mice, by crossing our previously described *Eif4a3*^{lox/+} mice (7) to ubiquitous *CMV-Cre* lines. To examine craniofacial development, embryos were collected at embryonic days (E) 10.5, E11.5 and E12.5 (Table 1). Heterozygous animals were recovered at normal Mendelian ratio at both E10.5 and E11.5; however, lethality was noted by E12.5 (Table 1). Thirty percent of E10.5 *Cmv-Cre;Eif4a3*^{lox/+} embryos (n=8) exhibited marked developmental delay relative to control, as evidenced by reduced body size and delayed eye and limb development (defined as type A

mutants with 2 mandibular arches) (Fig. 3A, B and K). Eleven percent of E10.5 *Cmv-Cre; Eif4a3^{lox/+}* embryos (n=3) showed developmental delay and severe facial malformations (defined as type B mutants with 1 apparent mandibular arch) (Supplementary Material, Fig. S6A-C).

Relative to E10.5, E11.5 *Cmv-Cre; Eif4a3^{lox/+}* embryos showed a more penetrant and wider spectrum of phenotypes including developmental delay (reduced body size, reduced eye pigmentation, delayed limb development), neural tube kinking or delayed closure, and facial malformations (Fig. 3C–J). In marked contrast to control *Cmv-Cre* littermates, 74% of *Cmv-Cre; Eif4a3^{lox/+}* embryos were developmentally delayed (type A) (Fig. 3F, H and K). Six percent were especially severe with developmental delay and holoprosencephaly, showing a single mandible formation which lacks two distinct mandibular arches, and has no apparent brain septation (type B) (Fig. 3K; Supplementary Material, Fig. S6D–F). EIF4A3 protein levels were reduced in *Cmv-Cre; Eif4a3^{lox/+}* embryos by ~50% compared with control (Fig. 3L). As further evidence of a role for EIF4A3 in mandible development, we confirmed EIF4A3 expression in the developing mandible (Supplementary Material, Fig. S6G and H). These findings indicate that reduced EIF4A3 impairs embryonic development and may model mandible malformations of RCPS syndrome.

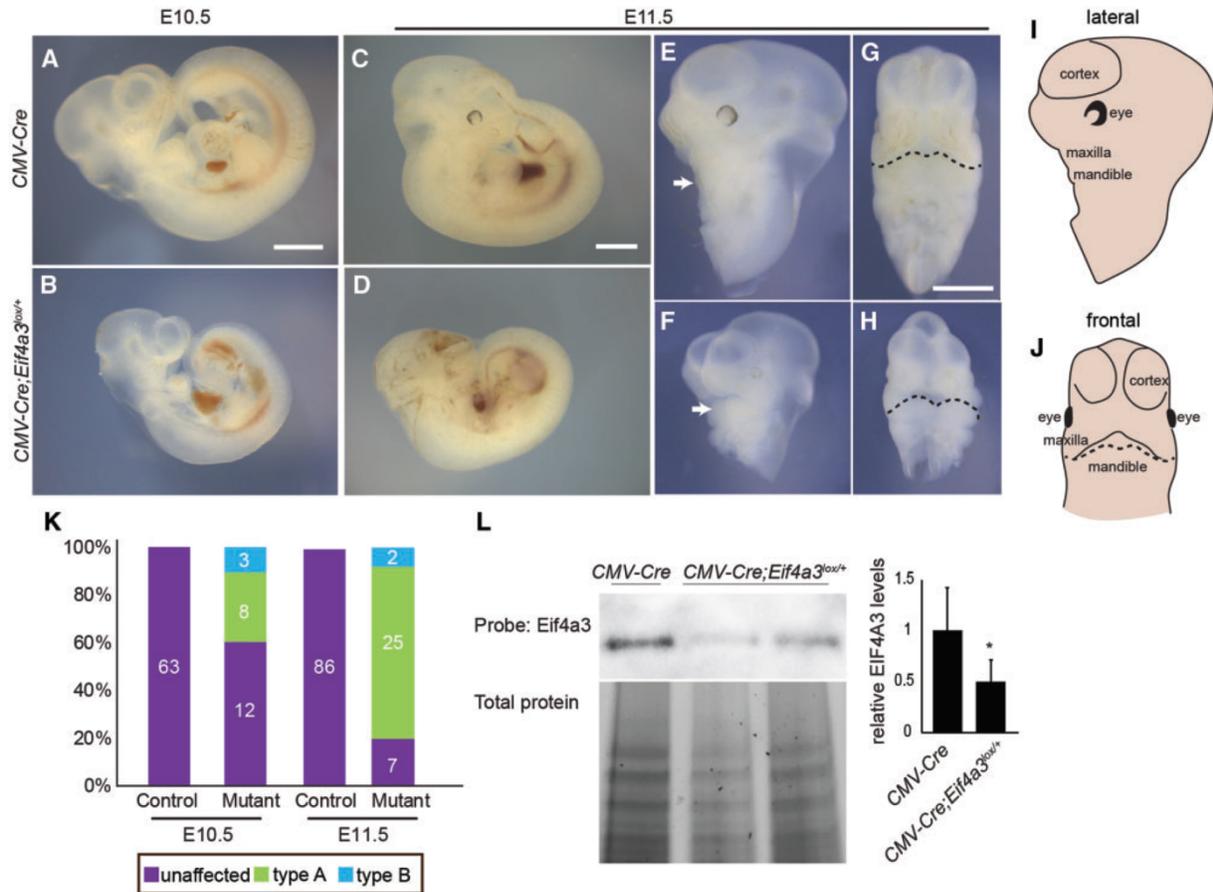


Figure 3. Ubiquitous Cre-mediated *Eif4a3* haploinsufficiency causes developmental delay and craniofacial defects. Whole mount embryos of indicated genotypes at E10.5 (**A, B**) and E11.5 (**C-H**). Both lateral views (**E, F**) and frontal views (**G, H**) of head only are shown for *CMV-Cre* (**A, C, E, G**) and *Cmv-Cre; Eif4a3^{lox/+}* (**B, D, F, H**). *Cmv-Cre; Eif4a3^{lox/+}* embryos shown are type A (developmental delay with two mandibular prominences) (**F, H**). Mandibles are indicated by arrows (**E, F**) and a dotted line (**G, H**). (**I, J**) Schematics with relevant labels for facial features in the lateral (**I**) and frontal views (**J**). (**K**) Graph depicting quantification of types A and B (more severe with developmental delay, craniofacial abnormality and single mandibular prominence) for indicated ages and genotypes. Total number of embryos examined in each category is indicated. (**L**) Western blot analysis of whole body from indicated genotypes probed for EIF4A3 and total protein. Relative to control and following normalization for total protein levels, *Cmv-Cre; Eif4a3^{lox/+}* embryos have 50% reduction of EIF4A3 (right). n=5 controls, n=5 mutants. Error bars, SD, P<0.05, Student's t-test. Scales bars, 1mm.

We next examined fusion of the mandibular processes in type A E10.5 and E11.5 tissue sections (Fig. 4A-F). At E10.5, when these processes are just beginning to fuse (25), intermandibular groove depth was equivalent between control and *Cmv-Cre; Eif4a3^{lox/+}* mutants

(285 and 281 μm , respectively) (Fig. 4A-C). By E11.5, mandibular processes have normally begun to fuse, as evidenced by groove depth of 87 μm in control. In contrast, intermandibular grooves of E11.5 *Cmv-Cre; Eif4a3^{lox/+}* embryos were significantly deeper, on average 170 μm (Fig. 4D-F). This phenotype was evident across multiple stage-matched mutant and control embryos, suggesting that developmental delay is insufficient to explain delayed fusion. These results indicate that *Eif4a3* is required between E10.5 and E11.5 for mandibular process outgrowth and fusion.

We next investigated potential causes of mandibular process fusion defects. Mitosis of mesenchymal progenitors that generate cartilage is important for mandibular development (26). Quantification of mitotic cells within mandibular processes, using phospho-histone 3 (PH3) staining, showed no significant differences between E10.5 control and *Cmv-Cre; Eif4a3^{lox/+}* embryos (Fig. 4G-I). However, we observed a slight, but not quite significant, increase in mitosis in E11.5 mutants (Fig. 4J-L). This suggests mitosis delays could contribute to mandible developmental defects but is unlikely the primary cause. Restriction of apoptosis has also been shown to be critically important for mandibular development, as evidenced in *Twsg1* mutants (27). Terminal deoxynucleotidyl transferase UTP nick end labeling (TUNEL) staining of E10.5 whole mount embryos showed evidence of apoptosis only in the most severe type B E10.5 *Cmv-Cre; Eif4a3^{lox/+}* mutants (Supplementary Material, Fig. S7A-I). These results exclude massive apoptosis as a main cause of the mandibular fusion defects in this mouse model.

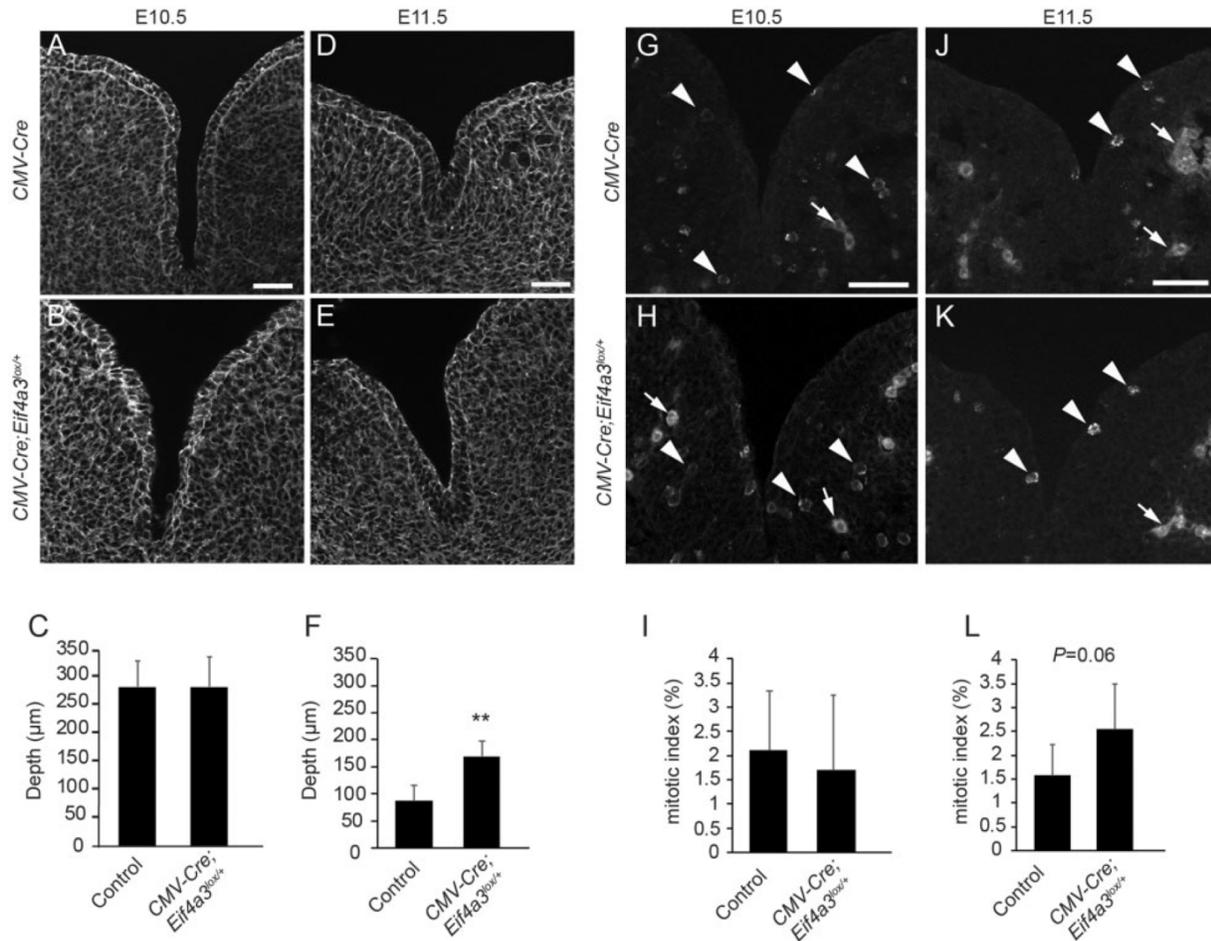


Figure 4. Mandibular formation defects in ubiquitous *Eif4a3* haploinsufficient embryos. Images depicting phalloidin staining to highlight cell architecture of mandibular prominences in *CMV-Cre* (A, D) and type A *Cmv-Cre; Eif4a3^{lox/+}* embryos (B, E), at E10.5 (A, B, n=6 controls, n=3 mutants) and E11.5 (D, E n=4 controls, n=5 mutants). (C, F) Graphs depicting average quantification of mandible clefts at E10.5 (C) and at E11.5 (F) for indicated genotypes. Images depicting PH3 staining in E10.5 *CMV-Cre* (G) and type A *Cmv-Cre; Eif4a3^{lox/+}* embryos (H) at E10.5 (G, H, n=6 controls, n=4 mutants) and E11.5 (J, K, n=4 controls, n=5 mutants). Arrowheads point to mitotic figures while arrows point to background vasculature. (I, L) Graphs depicting average quantification of mitotic index at E10.5 (I) and at E11.5 (L) for indicated genotypes. Error bars, SD, **P<0.005, Student's t-test, scale bars, 50 μm.

Neural crest specific haploinsufficiency of *Eif4a3* disrupts craniofacial development

We further probed the mechanism by which *Eif4a3* deficiency impairs mandible development by asking what is the cell autonomous requirement of *Eif4a3* during fusion of the mandibular processes. Given defects seen in patient-derived iNCCs, we hypothesized that *Eif4a3* functions in NCCs *in vivo*. To deplete *Eif4a3* specifically in NCCs, we generated *Wnt1-Cre;*

Eif4a3^{lox/+} embryos. Compared with control (n=26), E11.5 *Wnt1-Cre; Eif4a3*^{lox/+} embryos exhibited severe and 100% penetrant craniofacial malformations (n=10) (Fig. 5A-I). The overall limb development of mutant embryos was relatively normal. Although their body size was slightly reduced, other features of developmental progression such as formation of hair follicles were intact. This argues that unlike *Cmv-Cre; Eif4a3*^{lox/+} embryos there is no significant developmental delay in *Wnt1-Cre; Eif4a3*^{lox/+} embryos. 100% of *Wnt1-Cre; Eif4a3*^{lox/+} embryos exhibited micrognathia (Fig. 5B, C, E, F, H and I). Similar to *Cmv-Cre; Eif4a3*^{lox/+} embryos, *Wnt1-Cre; Eif4a3*^{lox/+} embryos (n=3) showed deeper intermandibular grooves than control (n=6) (Fig. 5J, K and M). Some mutant embryos (n=4) also had more shallow grooves than control, highlighting a spectrum of phenotypes associated with *Eif4a3* haploinsufficiency (Fig. 5J, L and M). Increased apoptosis (CC3+ staining) was primarily evident in lateral facial structures of *Wnt1-Cre; Eif4a3*^{lox/+} embryos, to a slightly greater extent than in *Cmv-Cre; Eif4a3*^{lox/+} embryos (Supplementary Material, Fig. S8A-I). This suggests that apoptosis may contribute in part to mandible groove defects, but is likely not the only cause. Importantly, together with the iNCC results, these data argue that *Eif4a3* expression in NCCs and its derivatives contributes to RCPS pathogenesis.

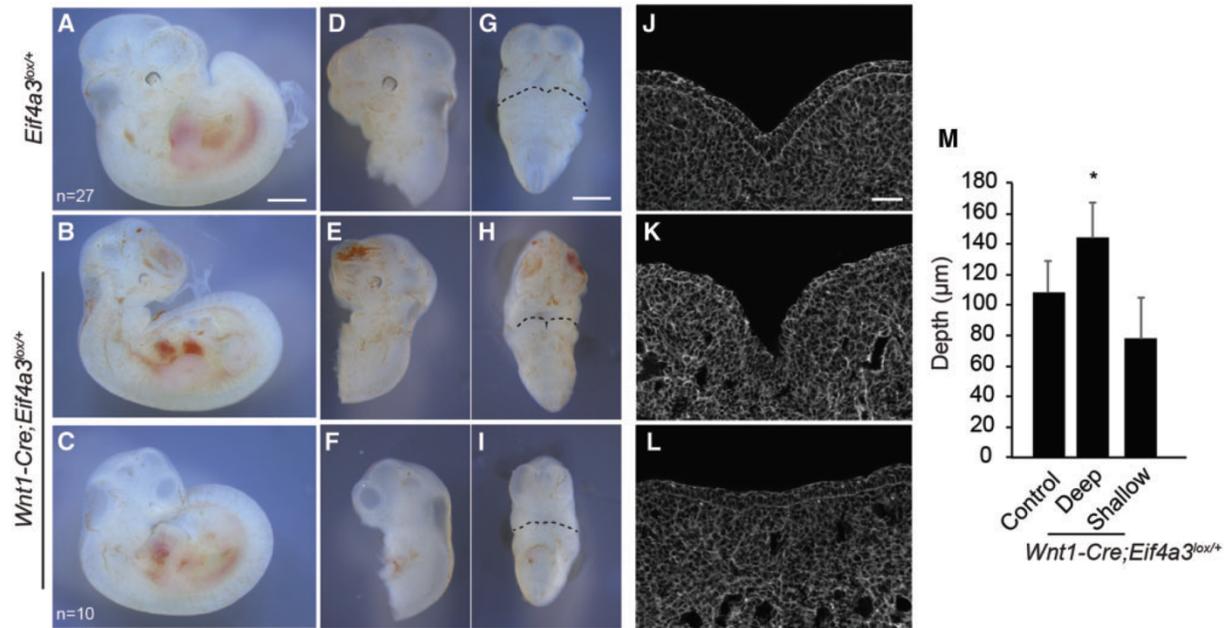


Figure 5. *Wnt1-Cre* mediated haploinsufficiency of *Eif4a3* in NCCs causes craniofacial defects. Whole mount embryos of indicated genotypes at E11.5 showing lateral views (A-F) and frontal views (G-I) shown for *Eif4a3*^{lox/+} (A, D, G, n=27 embryos) and *Wnt1-Cre; Eif4a3*^{lox/+} (B, C, E, F, H, I, n=10 embryos total for B, C). Images of mandibular processes stained for phalloidin for *Eif4a3*^{lox/+} (J) or *Wnt1-Cre; Eif4a3*^{lox/+} (K, L). Less severe (B, E, H, K) and more severe (C, F, I, L) *Wnt1-Cre; Eif4a3*^{lox/+} embryos are depicted. (M) Graph depicting quantification of mandible grooves for indicated genotypes, representing embryos with deeper grooves (n=3), those with shallow grooves (n=4) and controls (n=6). Mean values shown for all embryos. Error bars, SD, P<0.05, Student's t-test. Scales bars, (A-I) 1mm, (J-L) 50 µm.

Given that patient nMSCs exhibited dysregulated chondrogenesis and precocious osteogenic differentiation, we next monitored formation of Meckel's cartilage and clavicle ossification *in vivo* (Fig. 6). Compared with control, the head structures of E14.5 *Wnt1-Cre; Eif4a3*^{lox/+} embryos were markedly dysmorphic, missing ears, face and skull. E14.5 mutant embryos were also slightly reduced in body size, but had relatively normal limbs, similar to E11.5 embryos (Fig. 6A and B). Meckel's cartilage serves as the template for the future mandible (28). In skeletal preparations of E14.5 control embryos (n=25), Meckel's cartilage was evident as a V-shaped structure, as previously described (Fig. 6C and E). In contrast, this structure was markedly absent or reduced in E14.5 *Wnt1-Cre; Eif4a3*^{lox/+} embryos (n=4) (Fig. 6D and F). Owing to loss of this structure, to examine bone ossification we focused on the clavicle, which is also affected in a subset of RCPS patients (2,29,30) and shares a NCC origin with the mandible (31). Alizarin red

staining of the clavicle revealed ossification of only 25% of control embryos (5/20) compared with 100% of E14.5 *Wnt1-Cre; Eif4a3^{lox/+}* embryos (4/4) (Fig. 6C–G; Supplementary Material, Fig. S9A–D). This indicates that NCC-specific haploinsufficiency of *Eif4a3* disrupts formation of Meckel’s cartilage and causes premature clavicle ossification. This parallels observations made with human RCPS nMSCs, and provides a mechanistic explanation for mandibular and clavicular anomalies seen in patients.

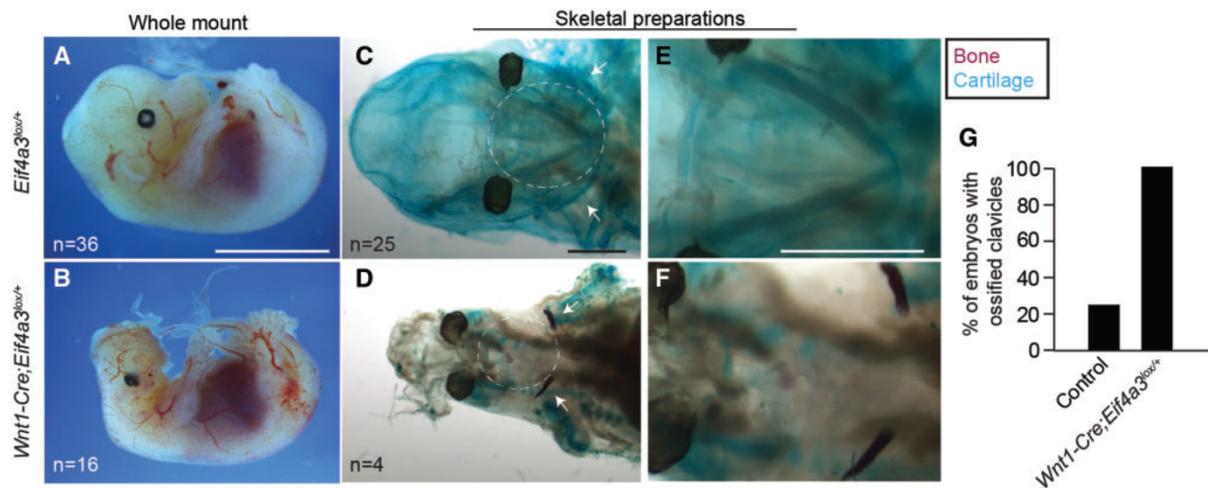


Figure 6. *Wnt1-Cre* mediated haploinsufficiency of *Eif4a3* disrupts Meckel’s cartilage development and causes premature clavicle ossification. (A, B) Whole mount E14.5 *Eif4a3^{lox/+}* (A) and *Wnt1-Cre; Eif4a3^{lox/+}* embryos (B) with number of embryos analyzed indicated. Skeletal preparations of E14.5 *Eif4a3^{lox/+}* (C, E) and *Wnt1-Cre; Eif4a3^{lox/+}* embryos (D, F) showing low magnification and higher magnification images. Note cartilaginous and well-formed Meckel’s cartilage (blue staining, white circle) in control but underdeveloped in mutants (C-F). Note controls lack clavicle ossification (purple, white arrows) which is present in mutant clavicles (C-F). (G) Graph depicting quantification of fraction of embryos with ossified clavicles. Scales bars, (A, B) 5mm, (C-F) 1mm.

Finally, we assessed the impact of NCC-specific *Eif4a3* haploinsufficiency upon craniofacial structures at the end of gestation. Compared with control (n=8), E18.5 *Wnt1-Cre; Eif4a3^{lox/+}* embryos exhibited severely hypoplastic mandibles (n=5) (Fig. 7A and B). We observed additional phenotypes associated with defective NCC development, including reduced eyelid closure, loss of outer ear structures, and in some cases exencephaly. Hematoxylin and eosin (H&E) analyses of control mandibles demarcated the tongue, dental papilla, dental epithelium,

nasal cavities, whisker follicles and maxillary structures (Fig. 7C and E). Conversely, no defined mandibular structures or nasal passages were evident in mutant embryos, although tongues were present without clefts, similar to RCPS patients (Fig. 7D and F; Supplementary Material, Fig. S10A-H). Taken together, these findings reveal an essential requirement of *Eif4a3* in NCCs and their mesenchymal derivatives during craniofacial development.

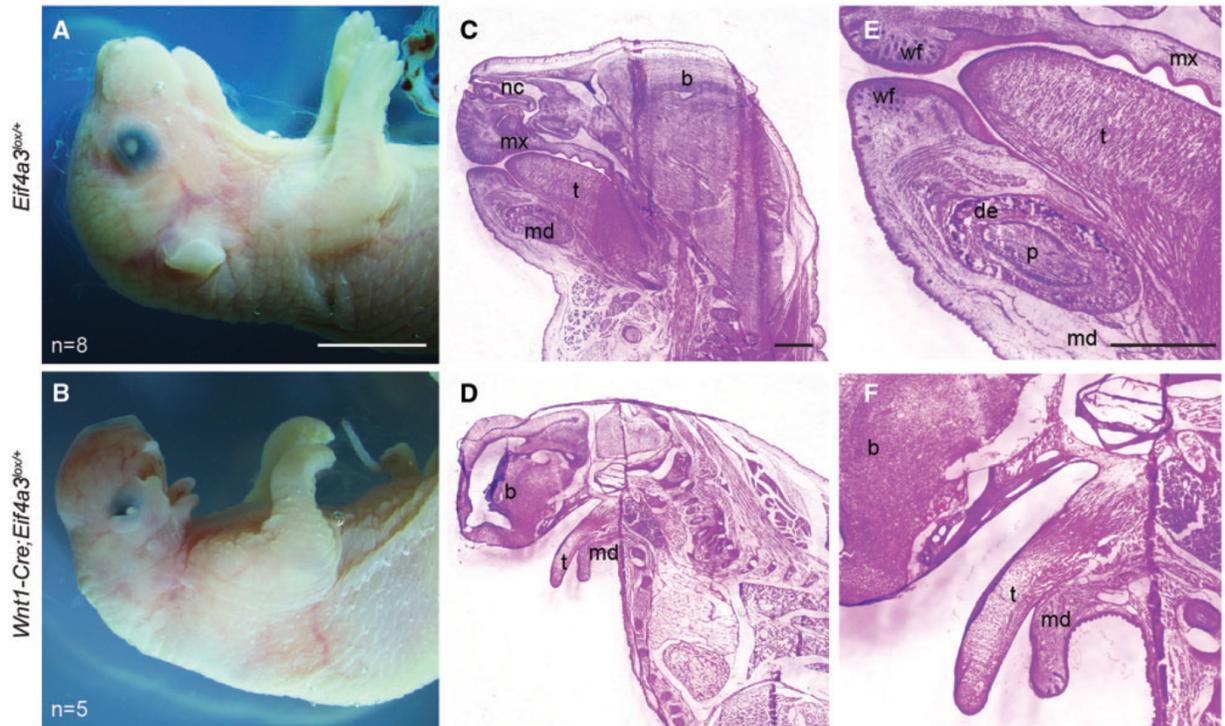


Figure 7. *Wnt1-Cre* mediated *Eif4a3* haploinsufficiency causes underdevelopment of the mandible and craniofacial defects. (A, B) E18.5 whole mount *Eif4a3*^{lox/+} (A) and *Wnt1-Cre; Eif4a3*^{lox/+} (B) embryos with number of embryos listed. H&E staining of sagittal tissue sections from *Eif4a3*^{lox/+} (C, E) and *Wnt1-Cre; Eif4a3*^{lox/+} embryos (D, F) showing low and 3x high magnification images of the same embryos. t, tongue; p, dental papilla; b, brain; de, dental epithelium; nc, nasal cavity; md, mandible; wf, whisker follicle; mx, maxilla. Scales bars, (A, B) 5mm, (C-F) 1mm.

DISCUSSION

RCPS is a craniofacial syndrome caused by loss-of-function mutations in the RNA-binding protein EIF4A3. The cellular mechanisms by which reduced EIF4A3 causes the phenotypic specificity of this syndrome have been entirely unknown. Here, we employed two complementary approaches to gain fundamental new insights into the underlying pathology of this developmental syndrome. Patient-derived iPSCs highlight delayed NCC migration and precocious bone differentiation in RCPS pathology. *Eif4a3* haploinsufficient mice recapitulate mandibular and craniofacial abnormalities associated with RCPS, and establish *in vivo* requirements for *Eif4a3* in NCC development and differentiation. Altogether this points to specific and converging alterations of NCC as an underlying cause of RCPS.

One major unsolved problem concerning craniofacial development and disease is the existence of a number of clinically discernible syndromes caused by mutations in genes that exert basic and ubiquitous functions. Genetic mutations affecting mRNA splicing and ribosome biogenesis, for example, manifest tissue-specific phenotypes restricted to craniofacial structures not only in RCPS but in additional disorders such as Treacher Collins syndrome (11,32-34). Elucidating the pathogenetic mechanism behind RCPS will aid in clarifying the etiology of these other craniofacial syndromes and understanding developmental pathways critical for craniofacial morphogenesis.

Generation of RCPS patient-derived iPSCs and a new mouse model

Here, to elucidate the pathogenetic mechanism responsible for RCPS, we generated patient-specific, iPSC-derived iNCCs and nMSCs. All iPSCs showed hallmarks of pluripotent stem cells, including high expression of pluripotency-associated genes, no aneuploidies and *in vivo* teratoma formation, showing that reduced *EIF4A3* in RCPS patients' cells does not negatively impact iPSC generation. Moreover, iPSCs were efficiently differentiated into iNCCs exhibiting typical p75(NTR)/HNK1 double-positive staining and high expression of NCC markers. The neural

crest identity of all iNCCs was further confirmed through generation of homogeneous nMSC populations with mesenchymal immunophenotype and ability to give rise to mesenchymal derivatives *in vitro*, as previously reported for other iNCC lines (17,18,35). Taken together, our strategy successfully yielded disease-relevant cell types to study RCPS *in vitro*.

We also generated two new *Eif4a3* mouse models that recapitulate several anomalies seen in RCPS patients, including abnormal mandible development, microstomia, micrognathia, ear anomalies and clavicle defects. RCPS is caused by recessive hypomorphic *EIF4A3* alleles, resulting in 15-70% reduction in *EIF4A3* levels depending on cell type (3). It is noteworthy that haploinsufficiency in mice reduces *EIF4A3* levels by 50% yet results in comparable phenotypes to humans whose *EIF4A3* levels are more varied. These potential differences in susceptibility to disease phenotypes could be because that humans are genetically diverse, whereas mice are inbred. In addition, some of the phenotypic variability in the *Eif4a3* Cre-lox mouse models could be due to Cre efficiency. The use of *Wnt1*-Cre enabled us to investigate how *Eif4a3* depletion affects craniofacial development; however, this Cre is not active in the developing limb (36). In the future, it will be of interest to employ other Cre drivers to recapitulate additional relevant RCPS phenotypes. Regardless, given that the *Eif4fa3* haploinsufficient mice present with the most penetrant phenotypes of RCPS, these models are valuable for gaining mechanistic insights into RCPS pathology. Beyond this, the *Eif4a3* haploinsufficient mice also demonstrate new biological functions for *Eif4a3* in NCC and craniofacial development. These roles are likely conserved given requirements for *Magoh* and *Eif4a3* in development of NCC-derived cell types in mice and *Xenopus*, respectively (5,37).

Neural crest defects underlying RCPS

The use of patient-derived iPSCs pinpoints defective NCC migration as one key cellular mechanism to explain RCPS etiology. We speculate *EIF4A3* depletion delays NCC migration at early developmental stages, causing defective mandibular fusion and formation, as seen in *Eif4a3*-deficient embryos. This may be due to reduced mesenchymal progenitors populating the pharyngeal arches and/or impaired cartilage and bone formation. Although migration defects are common in mouse models with micrognathia, they have not been observed in animal models for

the craniofacial disorders ribosomopathies (12,38) or spliceopathies (11,13), despite the clinical overlap of RCPS with these conditions. This highlights an interesting and potentially unique role for NCC migration in RCPS etiology.

Although some NCC-related disorders have been attributed to increased cell death or altered proliferation (13,38,39), neither RCPS iNCCs and nMSCs nor mouse models showed massive alterations in either process. This suggests that aberrant apoptosis and proliferation are not a major cause of RCPS, however the possibility remains that low levels of apoptosis and proliferation delays could impair outgrowth and fusion of mandibular arches and thus contribute to RCPS pathogenesis. Indeed, apoptosis was detectable in especially severe *Cmv-Cre* and *Wnt1-Cre Eif4a3* mouse mutants. Additionally, mitotic index was increased, albeit not quite significantly, in *Eif4a3* haploinsufficient mandibular processes, consistent with known roles for *EIF4A3* and its binding partners in mitosis (40,41,42).

Both *in vitro* and *in vivo* experiments also demonstrate that dysregulation of chondrogenesis and osteogenesis is highly relevant for RCPS. Paralleling observations with RCPS-derived nMSCs, NCC-specific depletion of *Eif4a3* disrupts formation of Meckel's cartilage and causes premature clavicle ossification. Development of the cranioskeletal elements affected in RCPS relies upon proper differentiation of NCC and their progeny residing in the craniofacial mesenchyme. In particular, mandible morphogenesis is thought to depend upon proper formation of Meckel's cartilage before intramembranous ossification (14,43). Altogether this suggests a mechanistic explanation for mandibular and clavicular anomalies seen in RCPS patients.

***Eif4a3* deficiency and phenotypic variability**

Both iPSC and mouse models established here recapitulate deficiencies in mandibular development, amongst the most relevant craniofacial alterations of RCPS. A large spectrum of phenotypic variability in development of mandibular processes is observed in *Eif4a3* haploinsufficient mice. This parallels RCPS, where there is great clinical variability amongst patients, even within the same family (1,3). Several RCPS patients have died in early infancy owing to severe micrognathia associated with severe respiratory distress or required surgical

interventions for survival (1,2,44). A fraction of embryos with either ubiquitous or NCC-specific *Eif4a3* depletion presented with especially severe phenotypes not so far described in humans, including holoprosencephaly or exencephaly. Likewise, *Eif4a3* haploinsufficiency in neural progenitors causes severe microcephaly, also not reported in patients (7). These phenotypic differences could be explained by inbred genetic backgrounds of mice, exposing phenotypes masked in humans. Alternatively, brain malformations in humans could be embryonic lethal. Consistent with this possibility, spontaneous abortions have been reported in RCPS families [reviewed in (1)].

***Eif4a3* mechanisms and RCPS etiology**

Going forward it will be important to build upon the cellular mechanisms defined here, to further understand how *EIF4A3* functions in NCCs and in particular how it influences early ossification. Predictions can be made from studies in the developing brain, where haploinsufficiency of *Eif4a3* and other EJC components impair neural progenitor proliferation, leading to precocious neuronal differentiation (7). In the case of the EJC component, MAGOH, a prolonged mitosis of neural stem cells is sufficient to trigger altered neuronal differentiation in these mutants (41). Perhaps *Eif4a3* has parallel requirements in mitosis of NCCs, which leads to precocious ossification. In the brain, *Eif4a3* haploinsufficiency also triggers aberrant p53 activation and altered expression of ribosomal components. Dysregulation of ribosome biogenesis and p53 signaling are linked to many craniofacial disorders, including Treacher Collins syndrome (12). Given this connection, along with parallel roles for *Eif4a3* in bone differentiation, it is fascinating to consider if *EIF4A3* regulates ribosome biogenesis and p53 signaling in NCCs, and if this contributes to RCPS pathology. Moreover, craniofacial disorders similar to RCPS can be caused by mutation of spliceosomal components (11). Given its canonical role in the EJC, it will also be of interest to elucidate if *EIF4A3* influences NCC via splicing regulation. For example, *EIF4A3* may bind and regulate levels and/or splicing of pro-chondrogenic transcripts required for ossification. Investigation of transcriptome-wide RNA regulation and splicing control in *Eif4a3* mutants will provide valuable insights into how *EIF4A3* influences chondrogenesis at a molecular level.

In summary, we integrate iPSC technology with mouse models to study a craniofacial syndrome, establishing a foundation for characterizing biological mechanisms responsible for craniofacial disorders. Using patient-derived iPSCs we could recapitulate NCC migration and mesenchyme differentiation *in vitro*, and show both are compromised in *EIF4A3* depleted cells. Likewise, *Eif4a3* deficient mouse models exposed defective mandible process formation and bone ossification *in vivo*. Altogether, this points to critical cellular mechanisms within NCCs which help explain aberrant mandible development in RCPS patients. Going forward, the models established here are valuable for further elucidating how *EIF4A3* mutations causes RCPS, and more broadly, for establishing an experimental framework to study craniofacial disorders.

MATERIALS AND METHODS

Dermal fibroblast and erythroblast cultures

Skin biopsies were extracted from the lower posterior flank region with the use of 2-mm punch instruments and stored in DMEM/high glucose (Life Technologies) supplemented with 100 mg/ml of Normocin (Invivogen), at 4 °C. Within 24 h, biopsies were rinsed with PBS and incubated in a Dispase solution (1 mg/ml; Sigma-Aldrich) in DMEM/high glucose, overnight at 4°C. Specimens were then rinsed two times with PBS, their epidermis was removed and the dermis was minced with a scalpel and seeded into 25cm² culture flasks containing fibroblast medium consisting of DMEM/high glucose supplemented with 1% penicillin–streptomycin and 10% fetal bovine serum (FBS, Life Technologies). CD71⁺ cells from whole blood samples were reprogrammed according to Okita et al. (45). Briefly, peripheral blood mono-nuclear cells (PBMCs) were isolated by adding 2 mM EDTA in PBS to whole blood samples, and after centrifugation at 1000g for 10 min, transferring the diluted blood to a Leucosep tube pre-filled with Ficoll-Paque (GE Healthcare). Then, the tube was centrifuged at 400g for 30 min at 18°C, the plasma was discarded, and the cells from white ring collected. PBMCs were cultured in Stem Span Medium (Stem Cell) supplemented with 50 ng/ml of recombinant human stem cell factor (SCF; R&D), 500 U/ml of erythropoietin (EPO; R&D), 1 mM dexamethasone (Sigma), 40 ng/ml of IGF-1,

10 ng/ml of interleukin 3 and 10 mg/ml gentamicin (Miltenyi Biotec) in a 24-well ultra-low attachment plate (Corning). After 4 days in culture, magnetic sorting of CD71+ cells was performed with the use of microbeads (Miltenyi Biotec).

Generation of iPSCs

Dermal fibroblast cultures were derived from skin punch biopsies from three RCPS patients (F8417-1, F8417-2 and F6099-1), previously genotyped elsewhere (3), and from two controls (F9048-1 and F7405-1). Erythroblast cultures were derived from peripheral blood collection from one control (F8799-1). Fibroblasts and erythroblasts were reprogrammed with the use of episomal vectors (pCXLE-hOCT3/4-shP53-F, addgene plasmid 27077; pCXLE-hSK, addgene plasmid 27078; pCXLE-hUL, addgene plasmid 27080), as described (45), in an Amaxa Nucleofector II (program U-020 for fibroblasts and T-016 for erythroblasts) with either NHDF (fibroblasts) or CD34+ (erythroblasts) nucleofector kits (Lonza), according to the manufacturer's recommendations. Two days after nucleoporation, cells were co-cultivated with irradiated murine embryonic fibroblasts (Millipore) in embryonic stem cell medium (DMEM/F12 supplemented with 2 mM GlutaMAX-I, 0.1 mM non-essential amino acids, 100 mM 2-mercaptoethanol, 20% of knockout serum replacement, 0.1% Gentamicin, 10 ng/ml of bFGF (all provided by Life Technologies), 2 mM SB431542, 0.5 mM valproic acid sodium salt, 0.25 mM sodium butyrate (Sigma-Aldrich), 0.5 mM PD0325901 and 2 mM Thiazovivin (Miltenyi Biotec). Typical iPSC colonies were transferred to 60-mm Matrigel (BD-Biosciences)-coated plates and expanded in iPSC medium consisted of Essential 8™ Medium (Life Technologies) supplemented with 100 mg/ml of Normocin (Invivogen).

Total DNA was extracted from iPSC cultures with the use of NucleoSpin Tissue (Macherey-Nagel), following supplier's instructions. For PCR reactions, primers targeting the OriP gene present in the backbone of the episomal vectors were used (Supplementary Material, Table S2), according to recommendations provided elsewhere (Epi5 Episomal iPSC Reprogramming Kit, Life Technologies). Multiplex ligation-dependent probe amplification (MLPA) analysis was performed with subtelomeric kits (P036 and P070; MRC-Holland) to detect chromosomal imbalances, as previously described (46). One of the control iPSCs used in this study (F7405-1) had been

generated with retroviral transduction and has already been described and characterized elsewhere (47).

Differentiation of iNCCs from iPSCs

Procedures for iNCC derivation were based on previously published methodology (17,18). Before differentiation, iPSC colonies were adapted to single-cell passaging by rinsing cells with PBS followed by dissociation with Accutase (Life Technologies) for up to 5 min at room temperature (RT), followed by centrifugation at 200g for 4 min and seeding onto 60-mm Matrigel-coated petri dishes. After two subcultures, single cells were seeded onto 60-mm Matrigel-coated dishes at 1×10^4 cells/cm². Two days post-seeding, medium was changed to iNCC differentiation medium, composed of Essential 6TM Medium (Life Technologies) supplemented with 8 ng/ml bFGF (Life Technologies), 20 mM SB431542 (Sigma-Aldrich), 1 mM CHIR99021 (Sigma-Aldrich) and 100 mg/ml Normocin; differentiation medium was changed every single day. After ~2–4 days, neural crest-like cells were seen detaching from colony borders. The cells were split before reaching confluence, using Accutase, at RT, until differentiated cells detached. Cell suspensions were centrifuged at 200g for 4 min and re-seeded into new 60-mm Matrigel-coated dishes in fresh iNCC differentiation medium. With this method, passaging was performed whenever necessary, for 15 days, after which morphologically homogeneous iNCC cultures were obtained. Differentiated iNCCs were cultivated for up to eight passages in iNCC differentiation medium, replenished daily. In all procedures involving single-cell passaging, media were supplemented with 5 mM Rock inhibitor (Sigma-Aldrich) upon seeding and maintained for 24 h; after about 4 days of iNCC differentiation, Rock inhibitor was no longer needed to maintain cell viability.

Wound healing assay

For the wound healing assay (48), iNCCs were seeded at 5×10^5 cells/cm² into non-coated 24-well plates (Corning), in iNCC medium. When cells reached 90–100% confluence, the monolayer was scratched in a straight line with a p200 pipette tip. Debris was removed and the edge of the scratch was smoothed by washing two times with DPBS (Life Technologies). Then,

the culture medium was replaced and cell migration was monitored. Cells were seeded in quadruplicates and three fields were selected in each well, which were photographed every 30 min during 24 h, using the High Content Imaging InCell Analyzer 2200 (GE Healthcare) at the CEFAP facility, Institute of Biomedical Sciences, University of São Paulo, Brazil. All samples were assessed simultaneously in two independent experiments. The percentage of the wound covered by migrating cells after 24h was quantified in RCPS and controls iNCCs, on ImageJ. Differentiation of nMSCs from iNCCs nMSC populations were obtained through culturing of iNCCs with mesenchymal stem cell medium, as previously described (17). In brief, iNCCs were seeded at 2×10^4 cells/cm² onto noncoated 60-mm tissue culture dishes in nMSC medium (DMEM/ F12 supplemented with 10% FBS, 2 mM GlutaMAX-I, 0.1 mM non-essential aminoacids and 100 mg/ml Normocin. Cells were differentiated for 6 days and passaged with TrypLE™ Express (Life Technologies) when needed. nMSC cultures were expanded in nMSC medium for up to 6 passages, with medium changes every 3 days.

nMSC chondrogenic, osteogenic and adipogenic differentiation

For the osteogenic induction, cells were seeded in 12-well plates (Corning) (10^4 cells/cm²), in triplicates. After 3 days, medium was replaced with osteogenic induction medium (StemPro Osteogenesis Kit, Life Technologies); in parallel, negative controls were cultivated in nMSC medium. Differentiation and nMSC media were changed every 2–3 days. After 9 days, ALP activity was quantified through incubation with phosphatase substrate (Sigma-Aldrich), and the resulting p-nitrophenol was quantified colorimetrically using a Multiskan EX ELISA plate reader (Thermo Scientific) at 405 nm; after 21 days, extracellular matrix mineralization was assessed through alizarin red staining at 450 nm. In both assays, absorbance data were normalized by subtracting from undifferentiated, negative controls. For chondrogenesis, 1×10^5 cells/well were plated into 6-well plates and after 3 days nMSC growth medium was replaced with chondrogenic medium (StemPro Chondrogenesis Kit; Life Technologies). To quantify chondrogenic markers, total RNA was extracted after 9 days of differentiation. After 21 days, pellets were fixed and frozen in Tissue-Tek O.C.T (Sakura) and then, cut in 5- μ m cryosections which were fixed with 4% paraformaldehyde and stained with Alcian Blue 0.1% in 0.1 M HCl. Adipogenic differentiation was

performed on 1×10^4 cells/cm² seeded into 6-well plates. After cells achieved 80% confluence, nMSC growth medium was replaced with adipogenesis medium (StemPro Adipogenesis Kit, Life Technologies). Cells were differentiated for 21 days, after which Oil red staining was performed. Cells were washed with PBS, fixed in 4% paraformaldehyde and stained with 0.5% Oil Red in isopropanol. Pictures were taken under an Axiovision microscope (Zeiss).

Apoptosis, proliferation and cell cycle assays *in vitro*

For the apoptosis assays, a total of 10^5 cells/well was seeded into 6-well culture plates (Corning). On the next day, apoptotic activity was measured with a kit based on Annexin V and 7-AAD staining (Guava Nexin Reagent; Millipore), following the manufacturer's instructions. Cells treated with 10 mM H₂O₂ for 30 min were used as staining controls. Subpopulations were ascertained in a Guava flow cytometer (EMD Millipore) as follows: non-apoptotic cells-Annexin V(-) and 7-AAD(-); early apoptotic cells-Annexin V(+) and 7-AAD(-); late-stage apoptotic and dead cells-Annexin V(+) and 7-AAD(+). Cell proliferation was assessed with the use of an XTT assay (Cell Proliferation Kit II; Roche), following supplier's instructions. Briefly, cells were seeded into 96-well culture plates at 2×10^3 cells/well, in quadruplicates. To quantify metabolically active cells, medium was changed to DMEM/F12 without phenol red (Life Technologies), a solution of XTT was added, and cells were incubated at 37 °C for 3h. Immediately following incubation, plates were colorimetrically assessed in a microplate spectrophotometer (Epoch; BioTek) at 450 nm. To determine the percentage of cells in G₀/G₁, S and G₂/M phases based on DNA content, a cell cycle assay was performed, using the Guava Cell Cycle Reagent (Millipore), in a Guava easyCyte flow cytometer (Millipore), according to the manufacturer's instructions.

Flow cytometry

To assess the immunophenotype of iNCCs, cells were detached with Accutase, and washed two times with two volumes of blocking solution (4% BSA in PBS). iNCCs were incubated with the conjugated antibodies in blocking solution in the dark for 1h at 4°C, washed two times with PBS, and fixed in 1% paraformaldehyde/PBS. The following antibodies were used: IgM K FITC Mouse Anti-Human CD57 (anti-HNK1; BD Pharmingen 561906), IgG1 K Alexa Fluor 647 Mouse

Anti-Human CD271 (anti-p75NTR; BD Pharmingen 560877), FITC Mouse IgM K isotype control (BD Pharmingen 555583) and Alexa Fluor 647 Mouse IgG1 K isotype control (BD Pharmingen 557714). Antibody concentrations followed the manufacturer's recommendations. A minimum of 5000 events were acquired in a FACS Aria II flow cytometer (BD Biosciences) and analyzed on FlowJo (X 10.0.7r2). nMSCs were dissociated with TrypLE Express and washed two times with two volumes of blocking solution (1% BSA in PBS without Ca²⁺ and Mg²⁺). nMSCs were incubated in the aforementioned conditions with the following conjugated antibodies: FITC Mouse Anti-Human CD31 (BD Pharmingen 555445), APC Mouse Anti-Human CD73 (BD Pharmingen 560847), PE Mouse Anti-Human CD90 (BD Pharmingen 555596), PE Mouse Anti-Human CD166 (BD Pharmingen 559263) and FITC, PE and APC Mouse IgG1 K Isotype Controls (BD Pharmingen 555748, 554681 and 555749, respectively). Antibody concentrations followed the manufacturer's recommendations. At least 5000 events were acquired in FACS Aria II equipment and analyzed on FlowJo (X 10.0.7r2).

Real-time quantitative PCR (RT-qPCR)

Total RNA was obtained from cell samples with the use of Nucleospin RNA II extraction kit (Macherey-Nagel) following the manufacturer's recommendations. Briefly, 1 mg of total RNA was converted into cDNA using Superscript IV (Life Technologies) and oligo-dT primers according to the manufacturer's recommendations. RT-qPCR reactions were performed with 2x Fast SYBR Green PCR Master Mix (Life Technologies) and 50-400nM of each primer. Fluorescence was detected using the 7500 Fast Real-Time PCR System (Life Technologies), under standard temperature protocol. Primer pairs were either designed with Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>; date last accessed March 10, 2017) or retrieved from Primer Bank (<http://pga.mgh.harvard.edu/primerbank/>; date last accessed March 10, 2017) primers are listed in Supplementary Material, Table S2 and their amplification efficiencies (E) were determined by serial cDNA dilutions log₁₀-plotted against Ct values, in which $E=10^{-1/\text{slope}}$. Gene expression was assessed relative to a calibrator cDNA (ΔCt). NormFinder (49) was used to determine the most stable endogenous control (among *ACTB*, *TBP*, *HMBS*, *GAPDH* and *HPRT1*), and calculate normalization factors ($E^{-\Delta\text{Ct}}$) for each sample. The final relative

expression values were determined based on a previous method (50) by dividing $E^{-\Delta Ct}$ of target genes by $E^{-\Delta Ct}$ the endogenous control. All relative expression values were log-transformed for analysis and graphed in linear scale, unless stated otherwise. Primers were supplied by Exxtend.

Western blot analyses

Cells were grown until 80% confluence in complete culture medium. Total protein lysates were obtained using RIPA buffer (Sigma) supplemented with protease and phosphatase inhibitor cocktails (Sigma-Aldrich). Protein concentration was determined using a bicinchoninic acid (BCA) assay kit according to the manufacturer's instructions (Thermo Fisher Scientific). Cell lysates were denatured by adding Tris-glycine SDS sample buffer (Life Technologies) containing 0.3 M 2-mercaptoethanol (Sigma). Then, 20 mg of total cell lysates were separated by SDS-PAGE and dry-transferred to nitrocellulose membranes with the iBlot system (Life Technologies) according to the manufacturer's recommendations. Membranes were blocked in buffer TBS-T (1 M Tris-HCl pH 7.5, 5M NaCl, 0.1% Tween-20) with 5% BSA under constant stirring for 1h at RT, washed four times with TBS-T, and incubated with Rabbit anti-EIF4A3 (1:2000 dilution in TBS-T/BSA; sc-33632; Santa Cruz) primary antibody, overnight at 4°C, under constant stirring. Detection was performed using anti-rabbit IgG, HRP-linked Antibody (1:2500 dilution in TBS-T/BSA; #7074; Cell Signaling) under constant stirring for 1h at RT, followed by incubation with ECL Prime (GE Healthcare) and capture in the Image Quant LAS 4000 Mini (GE Healthcare). β -Actin was used as a loading control (1:30 000 dilution in TBS-T/BSA; ab49900, Abcam). Densitometric analyses were performed on the Image Quant TL 8.1 software (GE Healthcare). EIF4AIII protein levels were quantified and normalized to the corresponding β -actin levels.

Embryos were lysed in RIPA lysis buffer (Thermo Scientific) with protease inhibitors (Pierce). BCA assays were performed (Pierce) and 20 mg of lysate was run on 4–20% gradient precast TGX gels (Biorad), activated following Biorad's standard protocol, then transferred onto PVDF membranes (Biorad). Membranes were blocked with 5% milk then blotted with rabbit anti-EIF4A3 1:50 (sc67369, Santa Cruz) and protein levels normalized to total protein as determined by TGX activation using UV. Imaging was performed with a Biorad ChemiLab XRS+ machine with ImageLab software.

Immunofluorescence and H and E staining

Cells were fixed in 4% paraformaldehyde for 20 min at RT, followed by permeabilization with PBS 0.2% Triton X-100, for 30 min, at 4°C. Blocking was carried out with PBS 5% BSA for 1h at RT, followed by incubation with primary antibodies in blocking solution overnight, at 4°C, under constant stirring. After washing two times with PBS, cells were incubated with secondary antibodies in blocking solution at 4°C for 1h, in the dark, followed by two PBS washes and counterstaining with DAPI solution (Life Technologies) for 2 min at RT. After a final PBS wash, cells were analyzed with a fluorescent microscope (Axiovision; Zeiss). The primary Anti-OCT4 antibody (ab19857; Abcam), secondary Goat anti-Rabbit IgG (H.L) antibody—Alexa Fluor 546 conjugate (A-11010; Life Technologies), primary Anti-SSEA4 antibody [MC813] (ab16287; Abcam) and secondary Goat anti-Mouse IgG (H.L) antibody—Alexa Fluor 488 conjugate (A-11001; Life Technologies) were used. Antibody concentrations followed the manufacturer's recommendations.

Embryos were fixed overnight in 4% paraformaldehyde (PFA) at 4°C, followed by submersion in 10% sucrose, then 20% sucrose until sinking, then embedded in NEG50 and kept at -80°C. Face cryostat sections (20 µm) were prepared and either stored at -80°C or prepared for staining. Sections were permeabilized with 0.25% Triton X-100 for 10 min and blocked with either 5% NGS or MOM block reagent (Vector laboratories) for 1h at RT. Primary antibodies were used at the following dilutions: rhodamine phalloidin 1:100 (R415, Molecular Probes), PH3 1:200 (06-570, Millipore), EIF4A3 1:500 (A302-980A, Bethyl), CC3 (9661, Cell Signaling) for 2h at RT or overnight at 4°C. Sections were counterstained with Hoescht for 15 min at RT with secondary antibodies (Thermo Scientific) which were used at 1:200. Standard H&E protocols were used, in brief: sections were dehydrated in Xylene for 1.5 min, Flex100 48 s (Thermo Scientific), Flex 95 24 s (Thermo Scientific), running water for 30 s, then put through a series of dyes and enhancers, hematoxylin (Ricca Chemical Company) 1.5 min, water 1 min, clarifier 30 s (Thermo Scientific), running water 30 s, bluing reagent (Thermo Scientific) 24 s, running water 30 s, Flex95 24 s, Eosin Y 24 s (Harleco), then again dehydrated in Flex100 48 s x2, Xylene 48 s x2 before mounting. Depth was calculated by drawing a straight line from the two highest points of each arch and measuring downward into the deepest part of the cleft. Mitotic index was calculated by counting the

number of mitotic figures (based on both PH3 staining and DNA morphology) and dividing by the total number of cells as determined by Hoescht staining.

Mouse husbandry and genetics

Plug dates were defined as embryonic day (E) 0.5 on the morning the plug was identified. *Cmv-Cre* mice and *Wnt1-Cre2* mice were from Jackson labs (36). As *Cmv-Cre* is X-linked, we generated embryos for all analyses using both *Cmv-Cre* females and males. No phenotypic differences were observed in both types of crosses. Animals reported in Table 1 were generated using *Cmv-Cre*^{Tg/+} females crossed to *Eif4a3*^{lox/+} males. *Eif4a3*^{lox/lox} mice were generated as previously described (7). Genotyping primers and conditions are listed in Supplementary Material, Table S3.

TUNEL assay

The ApoTag Fluorescein *In Situ* Apoptosis Detection Kit (Millipore) was utilized for whole mount TUNEL staining. Standard protocols were slightly modified as follows: embryos were fixed in 1% PFA for 20 min at 4°C, washed in 1x PBS and forebrain and heart were pierced with a needle. Embryos were dehydrated in a series of Ethanol washes (25% 10 min, 50% 10 min, 70% 10 min, 95% 10 min x2, 100% 5 min x3) and then rehydrated in reverse order. Embryos were pretreated with Proteinase K (20 mg/ml) for 15 min at RT then washed with 1x PBS 2 min x2. Embryos were put into equilibration buffer for 5 min at RT, then working strength TdT Enzyme for 2h at 37°C in a humidified chamber. The reaction was completed by putting the embryos into STOP/WASH buffer for 40 min at 37 °C. The embryos were then washed 3 min x3 in 2 mM levamisole, then 1 min x3 in 1x PBS. Anti-digoxigenin conjugate was used at working strength for 40 min RT in the dark then the embryos were washed 2 min x4 in 1x PBS and then imaged.

Skeletal preparations

E14.5 embryos were dissected and fixed in 100% EtOH overnight at 4°C, then 100% acetone overnight at 4°C. Alcian blue staining solution stock was prepared as follows: 5 ml 0.4%

Alcian Blue 8 GX in 70% EtOH, 5 ml glacial acetic acid, 70 ml 95% EtOH, 20 ml water, then to 10 ml of staining solution 100 μ L of 0.5% Alizarin Red was added. Bone only skeletal preps were performed the same way but without 0.4% Alcian Blue 8GX. Embryos were rocked overnight in staining solution at RT then destained with 1% KOH for approximately 2h. Embryos were then put through an increasing glycerol gradient then stored in 50% glycerol.

Statistics

All iPSC experiments were performed in triplicate, unless stated otherwise herein. Statistical comparisons were performed on the GraphPad Prism software (v6.0). Values were represented as means \pm standard error of the mean (SEM). The level of statistical significance was set at $P < 0.05$. Mouse experiments were represented as means \pm standard deviation (SD). All P-values were represented as * < 0.05 , ** < 0.005 . Significance for all experiments was performed using a standard Student's two-tailed t-test with the exception of the mouse lethality table (Table 1) which was done using Chi-square analysis.

Ethics statement

The experimental procedures involving samples from human subjects were approved by the Ethics Committee of Instituto de Biociências at Universidade de São Paulo, Brazil (accession number 1.463.852). Subjects donated biological samples only after providing signed informed consent. All animal work was approved by Duke University IACUC.

ACKNOWLEDGEMENTS

We thank members of the Silver lab for helpful discussions, Dr. Hiro Matsunami for sharing *Wnt1*-Cre mice, Dr. Yong-Hui Jiang for sharing *Cmv*-Cre mice. We thank Naila Lourenço, Andressa Morales, Patrícia Semedo-Kuriki for technical support in MLPA and flow cytometry analysis, and Gabriella Hsia for *EIF4A3* molecular diagnosis. We also thank students of the Passos-Bueno lab for support and discussions.

FUNDING

Funding for research from NIH (R01NS0830897) to D.L.S.; NSF Graduate Research Fellowship (DGF 1106401) to E.E.M.; FAPESP/CEPID (2013/08028-1) and CNPq (305405/2011-5) to M.R.P.B. and CNPq research fellowship (143094/2011-4) to G.S.K.

Conflict of Interest statement. None declared.

REFERENCES

1. Favaro, F.P., Zechi-Ceide, R.M., Alvarez, C.W., Maximino, L.P., Antunes, L.F.B.B., Richieri-Costa, A. and Guion-Almeida, M.L. (2011) Richieri-Costa–Pereira syndrome: a unique acrofacial dysostosis type. An overview of the Brazilian cases. *Am. J. Med. Genet. A*, 155, 322–331.
2. Raskin, S., Souza, M., Medeiros, M.C., Manfron, M. and Chong e Silva, D.C. (2013) Richieri-Costa and Pereira syndrome: Severe phenotype. *Am. J. Med. Genet. A*, 161, 1999–2003.
3. Favaro, F.P., Alvizi, L., Zechi-Ceide, R.M., Bertola, D., Felix, T.M., de Souza, J., Raskin, S., Twigg, S.R.F., Weiner, A.M.J., Armas, P., et al. (2014) A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. *Am. J. Hum. Genet.*, 94, 120–128.
4. Le Hir, H.E., Sauliere, J. and Wang, Z. (2016) The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell. Biol.*, 17, 41–54.
5. Haremakei, T. and Weinstein, D.C. (2012) Eif4a3 is required for accurate splicing of the *Xenopus laevis* ryanodine receptor pre-mRNA. *Dev. Biol.*, 372, 103–110.
6. Haremakei, T., Sridharan, J., Dvora, S. and Weinstein, D.C. (2010) Regulation of vertebrate embryogenesis by the exon junction complex core component Eif4a3. *Dev. Dyn.*, 239, 1977–1987.
7. Mao, H., McMahon, J.J., Tsai, Y.-H., Wang, Z. and Silver, D.L. (2016) Haploinsufficiency for core exon junction complex components disrupts embryonic neurogenesis and causes p53-mediated microcephaly. *PLoS Genet.*, 12, e1006282–e1006227.
8. Maclean, G., Dollé, P. and Petkovich, M. (2009) Genetic disruption of CYP26B1 severely affects development of neural crest derived head structures, but does not compromise hindbrain patterning. *Dev. Dyn.*, 238, 732–745.
9. Cordero, D.R., Brugmann, S., Chu, Y., Bajpai, R., Jame, M. and Helms, J.A. (2010) Cranial neural crest cells on the move: Their roles in craniofacial development. *Am. J. Med. Genet. A*, 155, 270–279.
10. Gong, S.-G. (2014) Cranial neural crest—migratory cell behavior and regulatory networks. *Exp. Cell. Res.*, 325, 90–95.
11. Lehalle, D., Wiczorek, D., Zechi-Ceide, R.M., Passos-Bueno, M.R., Lyonnet, S., Amiel, J. and Gordon, C.T. (2015) A review of craniofacial disorders caused by spliceosomal defects. *Clin. Genet.*, 88, 405–415.
12. Jones, N.C., Lynn, M.L., Gaudenz, K., Sakai, D., Aoto, K., Rey, J.-P., Glynn, E.F., Ellington, L., Du, C., Dixon, J., et al. (2008) Prevention of the neurocristopathy Treacher Collins syndrome through inhibition of p53 function. *Nat. Med.*, 14, 125–133.
13. Devotta, A., Juraver-Geslin, H., Gonzalez, J.A., Hong, C.-S. and Saint-Jeannet, J.-P. (2016) Sf3b4-depleted *Xenopus* embryos—a model to study the pathogenesis of craniofacial defects in Nager syndrome. *Dev. Biol.*, 415, 371–382.
14. Chai, Y., Jiang, X., Ito, Y., Bringas, P., Han, J., Rowitch, D.H., Soriano, P., McMahon, A.P. and Sucov, H.M. (2000) Fate of the mammalian cranial neural crest during tooth and mandibular morphogenesis. *Development*, 127, 1671–1679.
15. Noisa, P. and Raivio, T. (2014) Neural crest cells: from developmental biology to clinical interventions. *Birth. Defect. Res. C*, 102, 263–274.
16. Neben, C.L., Roberts, R.R., Dipple, K.M., Merrill, A.E. and Klein, O.D. (2016) Modeling craniofacial and skeletal congenital birth defects to advance therapies. *Hum. Mol. Genet.*, 25, R86–R93.
17. 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.
18. Fukuta, M., Nakai, Y., Kirino, K., Nakagawa, M., Sekiguchi, K., Nagata, S., Matsumoto, Y., Yamamoto, T., Umeda, K., Heike, T., et al. (2014) Derivation of mesenchymal stromal cells from pluripotent stem cells

through a neural crest lineage using small molecule compounds with defined media. *PLoS ONE*, 9, e112291–e112225.

19. Huang, M., Miller, M.L., McHenry, L.K., Zheng, T., Zhen, Q., Ilkhanizadeh, S., Conklin, B.R., Bronner, M.E. and Weiss, W.A. (2016) Generating trunk neural crest from human pluripotent stem cells. *Sci. Rep.*, 6, 19727.

20. Passos-Bueno, M.R., Ornelas, C.C. and Fanganiello, R.D. (2009) Syndromes of the first and second pharyngeal arches: a review. *Am. J. Med. Genet. A*, 149A, 1853–1859.

21. Bronner, M.E. and LeDouarin, N.M. (2012) Development and evolution of the neural crest: an overview. *Dev. Biol.*, 366, 2–9.

22. Kobayashi, G.S., Alvizi, L., Sunaga, D.Y., Francis-West, P., Kuta, A., Almada, B.V.P., Ferreira, S.G., de Andrade-Lima, L.C., Bueno, D.F., Raposo-Amaral, C.E., et al. (2013) Susceptibility to DNA damage as a molecular mechanism for non-syndromic cleft lip and palate. *PLoS ONE*, 8, e65677–e65611.

23. Griesi-Oliveira, K., Sunaga, D.Y., Alvizi, L., Vadasz, E. and Passos-Bueno, M.R. (2013) Stem cells as a good tool to investigate dysregulated biological systems in autism spectrum disorders. *Autism Res.*, 6, 354–361.

24. Yeh, E., Atique, R., Ishiy, F.A.A., Fanganiello, R.D., Alonso, N., Matushita, H., da Rocha, K.M. and Passos-Bueno, M.R. (2011) FGFR2 mutation confers a less drastic gain of function in mesenchymal stem cells than in fibroblasts. *Stem Cell Rev. Rep.*, 8, 685–695.

25. Chai, Y., Sasano, Y., Bringas, P., Mayo, M., Kaartinen, V., Heisterkamp, N., Groffen, J., Slavkin, H. and Shuler, C. (1997) Characterization of the fate of midline epithelial cells during the fusion of mandibular prominences in vivo. *Dev. Dyn.*, 208, 526–535.

26. Jeong, J., Mao, J., Tenzen, T., Kottmann, A.H. and McMahon, A.P. (2004) Hedgehog signaling in the neural crest cells regulates the patterning and growth of facial primordia. *Genes Dev.*, 18, 937–951.

27. MacKenzie, B., Wolff, R., Lowe, N., Billington, C.J., Jr, Peterson, A., Schmidt, B., Graf, D., Mina, M., Gopalakrishnan, R. and Petryk, A. (2009) Twisted gastrulation limits apoptosis in the distal region of the mandibular arch in mice. *Dev. Biol.*, 328, 13–23.

28. Ramaesh, T. and Bard, J.B.L. (2003) The growth and morphogenesis of the early mouse mandible: a quantitative analysis. *J. Anat.*, 203, 213–222.

29. Golbert, M.B., Dewes, L.O., Philipsen, V.R., Wachholz, R.S., Deutschendorf, C. and Leite, J.C.L. (2007) New clinical findings in the Richieri-Costa/Pereira type of acrofacial dysostosis. *Clin. Dysmorphol.*, 16, 85–88.

30. Ogando, P.B., Pires, F., Krummenauer, R.C.P., Collares, M.V.M. and Lubianca Neto, J.F. (2011) [Laryngeal malformations in the Richieri Costa and Pereira syndrome with airway obstruction]. *Braz. J. Otorhinolaryngol.*, 77, 138.

31. Matsuoka, T., Ahlberg, P.E., Kessar, N. and Iannarelli, P. (2005) Neural crest origins of the neck and shoulder. *Nature*, 436, 347–355.

32. Armistead, J. and Triggs-Raine, B. (2014) Diverse diseases from a ubiquitous process: the ribosomopathy paradox. *FEBS Lett.*, 588, 1491–1500.

33. Danilova, N. and Gazda, H.T. (2015) Ribosomopathies: how a common root can cause a tree of pathologies. *Dis. Models Mech.*, 8, 1013–1026.

34. Yelick, P.C. and Trainor, P.A. (2015) Ribosomopathies: global process, tissue specific defects. *Rare Dis.*, 3, e1025185.

35. Menendez, L., Yatskievych, T.A., Antin, P.B. and Dalton, S. (2011) Wnt signaling and a Smad pathway blockade direct the differentiation of human pluripotent stem cells to multipotent neural crest cells. *Proc. Natl. Acad. Sci.*, 108, 19240–19245.

36. Lewis, A.E., Vasudevan, H.N., O'Neill, A.K., Soriano, P. and Bush, J.O. (2013) The widely used Wnt1-Cre transgene causes developmental phenotypes by ectopic activation of Wnt signaling. *Dev. Biol.*, 379, 229–234.

37. Silver, D.L., Leeds, K.E., Hwang, H.-W., Miller, E.E. and Pavan, W.J. (2013) The EJC component Magoh regulates proliferation and expansion of neural crest-derived melanocytes. *Dev. Biol.*, 375, 172–181.
38. Dixon, J., Jones, N.C., Sandell, L.L., Jayasinghe, S.M., Crane, J., Rey, J.-P., Dixon, M.J. and Trainor, P.A. (2006) Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proc. Natl. Acad. Sci.*, 103, 13403–13408.
39. Brugmann, S.A., Allen, N.C., James, A.W., Mekonnen, Z., Madan, E. and Helms, J.A. (2010) A primary cilia-dependent etiology for midline facial disorders. *Hum. Mol. Genet.*, 19, 1577–1592.
40. Silver, D.L., Watkins-Chow, D.E., Schreck, K.C., Pierfelice, T.J., Larson, D.M., Burnett, A.J., Liaw, H.-J., Myung, K., Walsh, C.A., Gaiano, N., et al. (2010) The exon junction complex component Magoh controls brain size by regulating neural stem cell division. *Nat. Neurosci.*, 13, 551–558.
41. Pilaz, L.-J., McMahon, J.J., Miller, E.E., Lennox, A.L., Suzuki, A., Salmon, E. and Silver, D.L. (2016) Prolonged mitosis of neural progenitors alters cell fate in the developing brain. *Neuron*, 89, 83–99.
42. Mao, H., Pilaz, L.J., McMahon, J.J., Golzio, C., Wu, D., Shi, L., Katsanis, N. and Silver, D.L. (2015) Rbm8a haploinsufficiency disrupts embryonic cortical development resulting in microcephaly. *J. Neurosci.*, 35, 7003–7018.
43. Parada, C. and Chai, Y. (2015) Mandible and Tongue Development. *Curr. Top. Dev. Biol.*, 115, 31–58.
44. Souza, J., dal Vesco, K., Tonocchi, R., Closs-Ono, M.C., Passos-Bueno, M.R. and da Silva-Freitas, R. (2011) The Richieri-Costa and Pereira syndrome: report of two Brazilian siblings and review of literature. *Am. J. Med. Genet. A*, 155A, 1173–1177.
45. Okita, K., Yamakawa, T., Matsumura, Y., Sato, Y., Amano, N., Watanabe, A., Goshima, N. and Yamanaka, S. (2013) An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells*, 31, 458–466.
46. Jehee, F.S., Takamori, J.T., Medeiros, P.F.V., Pordeus, A.C.B., Latini, F.R.M., Bertola, D.R., Kim, C.A. and Passos-Bueno, M.R. (2011) Using a combination of MLPA kits to detect chromosomal imbalances in patients with multiple congenital anomalies and mental retardation is a valuable choice for developing countries. *Eur. J. Med. Genet.*, 54, e425–e432.
47. Ishiy, F.A.A., Fanganiello, R.D., Griesi-Oliveira, K., Suzuki, A.M., Kobayashi, G.S., Morales, A.G., Capelo, L.P. and Passos-Bueno, M.R. (2015) Improvement of in vitro osteogenic potential through differentiation of induced pluripotent stem cells from human exfoliated dental tissue towards mesenchymal-like stem cells. *Stem Cells Int.*, 2015, 1–9.
48. Liang, C.-C., Park, A.Y. and Guan, J.-L. (2007) In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nat. Protoc.*, 2, 329–333.
49. Andersen, C.L., Jensen, J.L. and Ørntoft, T.F. (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.*, 64, 5245–5250.
50. Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, 29, e45.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online

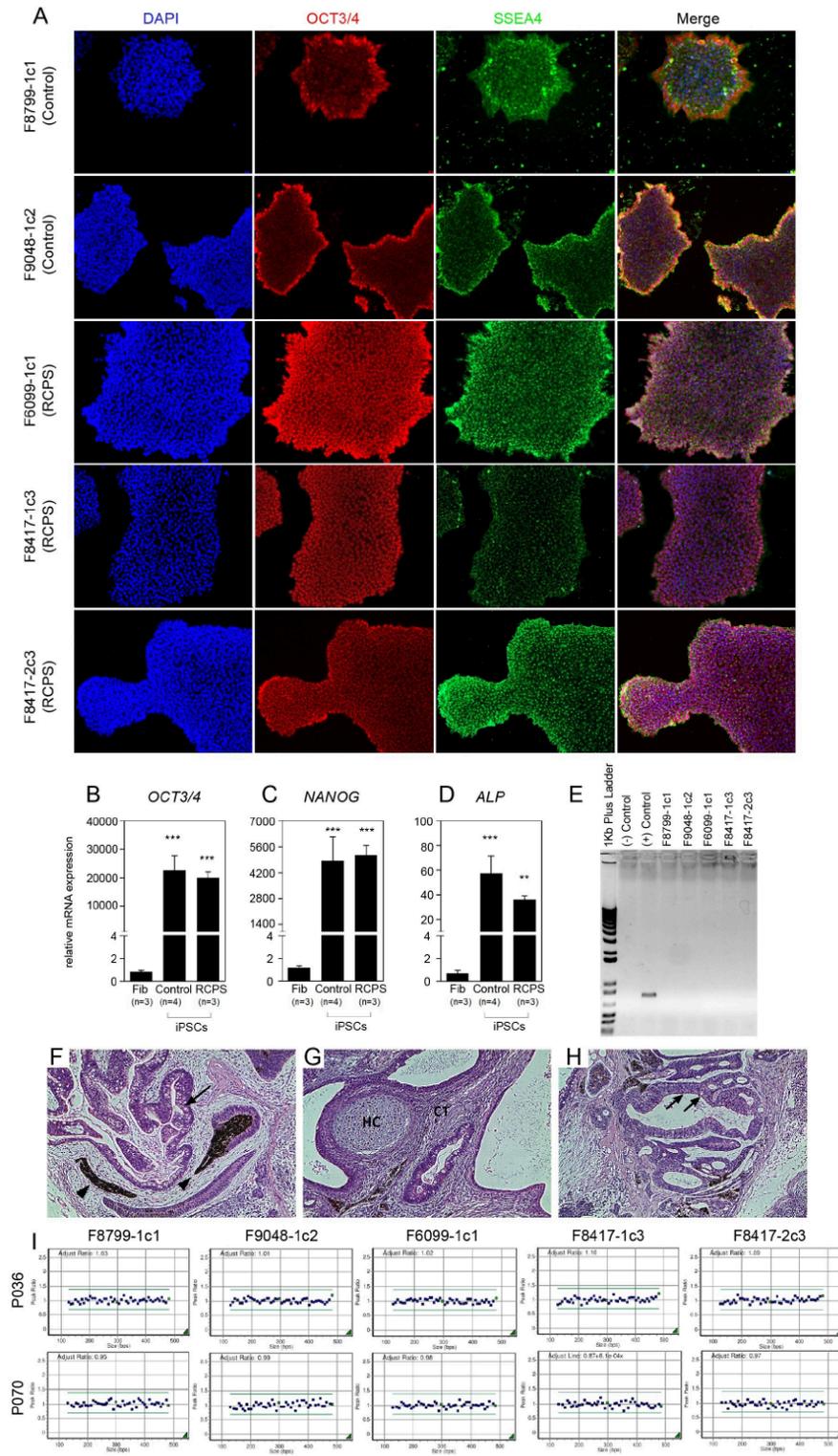


Figure S1. Characterization of induced pluripotent stem cells (iPSCs). **(A)** Immunofluorescence staining of pluripotency markers OCT3/4 (red) and SSEA-4 (green) in iPSCs (10x magnification); DAPI nuclear staining is in blue. **(B-D)** RT-qPCR analysis showing transcriptional upregulation of pluripotency associated genes *OCT3/4*, *NANOG*, and *ALP* in controls and RCPS patients (with n=number of biological samples). Graphs were plotted relative to expression data of adult fibroblasts. **(E)** End-point PCR for vector backbone gene OriP; DNA from one iPSC line in which genomic integration had been detected was used as positive control. **(F-H)** Representative images of teratoma generated by subcutaneous injection of iPSCs into nude mice; hematoxylin and eosin tissue staining depict tissue originating from the three germ layers. **(F)** Arrowheads: pigmented epithelium with melanin granules (ectoderm); arrow: gastrointestinal epithelium with goblet cells (endoderm). **(G)** HC: hyaline cartilage and CT: connective tissue (mesoderm). **(H)** Arrows: ciliated respiratory epithelium (ectoderm). **(I)** MLPA analysis with peak ratios for subtelomeric probes (blue dots) and control probes (green dots) using P036 and P070 kits. Data are represented as mean \pm SEM, (***) p-value < 0.001, (**) p-value < 0.01, Student's t-test.

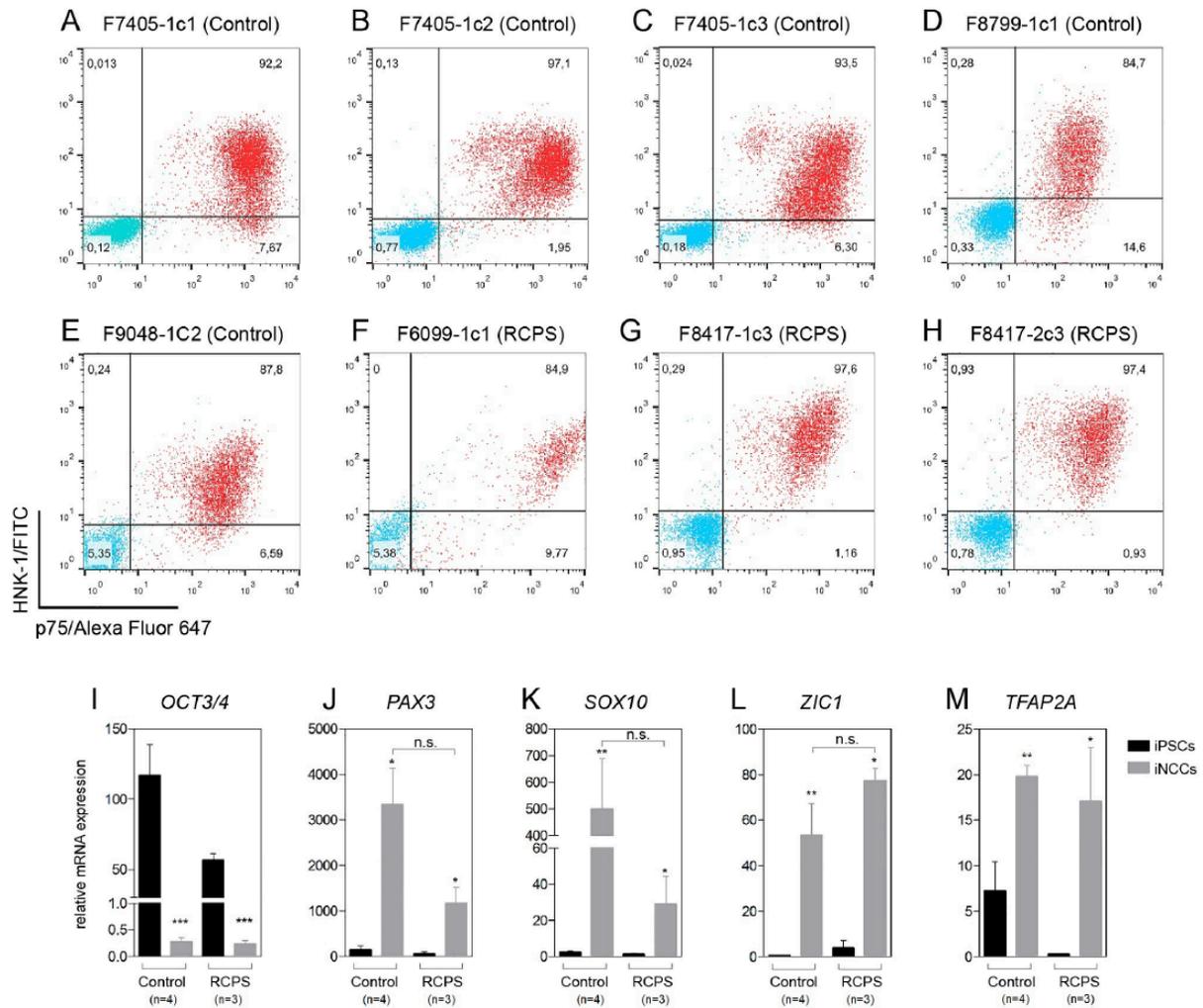


Figure S2. Characterization of iPSC-derived neural crest cells (iNCCs). **(A-H)** Biparametric flow cytometry dot plots for HNK-1/FITC and p75/Alexa Fluor 647 expression in RCPS and control cells. Values in upper right quadrants represent p75+/HNK-1+ events. **(I-M)** RT-qPCR assessment of neural crest markers (*PAX3*, *SOX10*, *ZIC1* and *TFAP2A*) and pluripotency marker (*OCT3/4*) in iPSCs (black) and iNCCs (gray) from controls and RCPS patients with n=number of biological samples; values represent mean \pm SEM, (***) p-value < 0.001, (**) p-value < 0.01, (*) p-value < 0.05, (n.s.) not significant, Student's t-test.

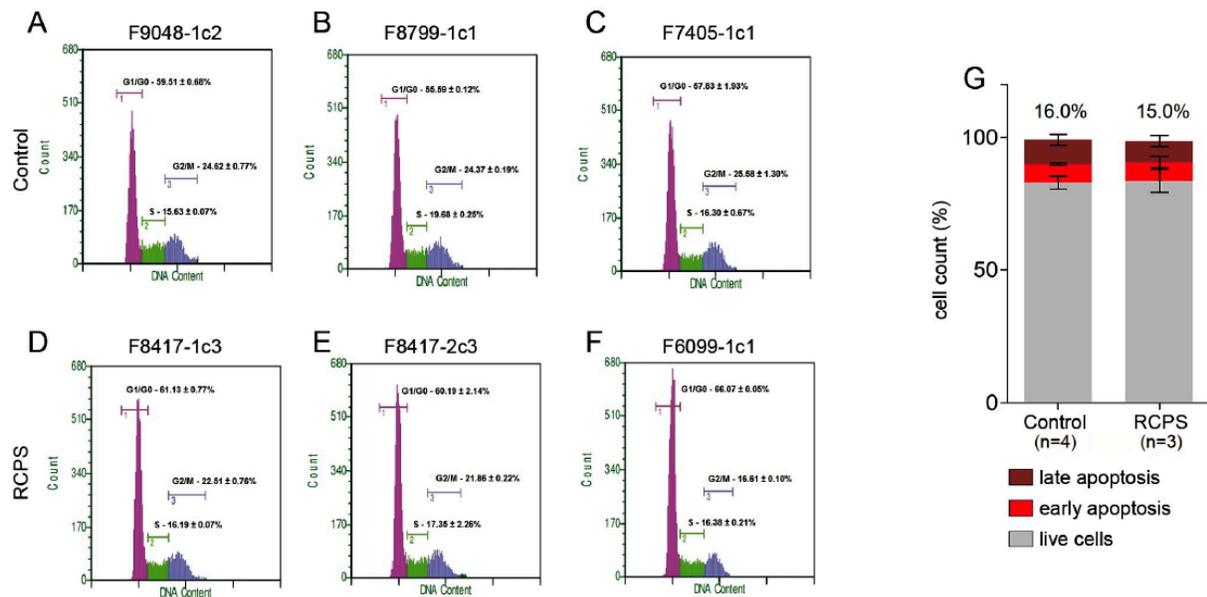


Figure S3. Cell cycle and apoptosis analysis of iNCCs. (A-F) Flow cytometry graphs depicting percentage of cells in the G1/G0, S, or G2/M phase based on DNA content (PI), of iNCCs from RCPS patients and control subjects. (G) 7-AAD/Annexin V flow cytometry results showing the fraction of live cells and cells undergoing early and late apoptosis in iNCCs; the percentages represent the sum of early and late apoptotic events; n=number of biological samples. Data are represented as mean ± SEM.

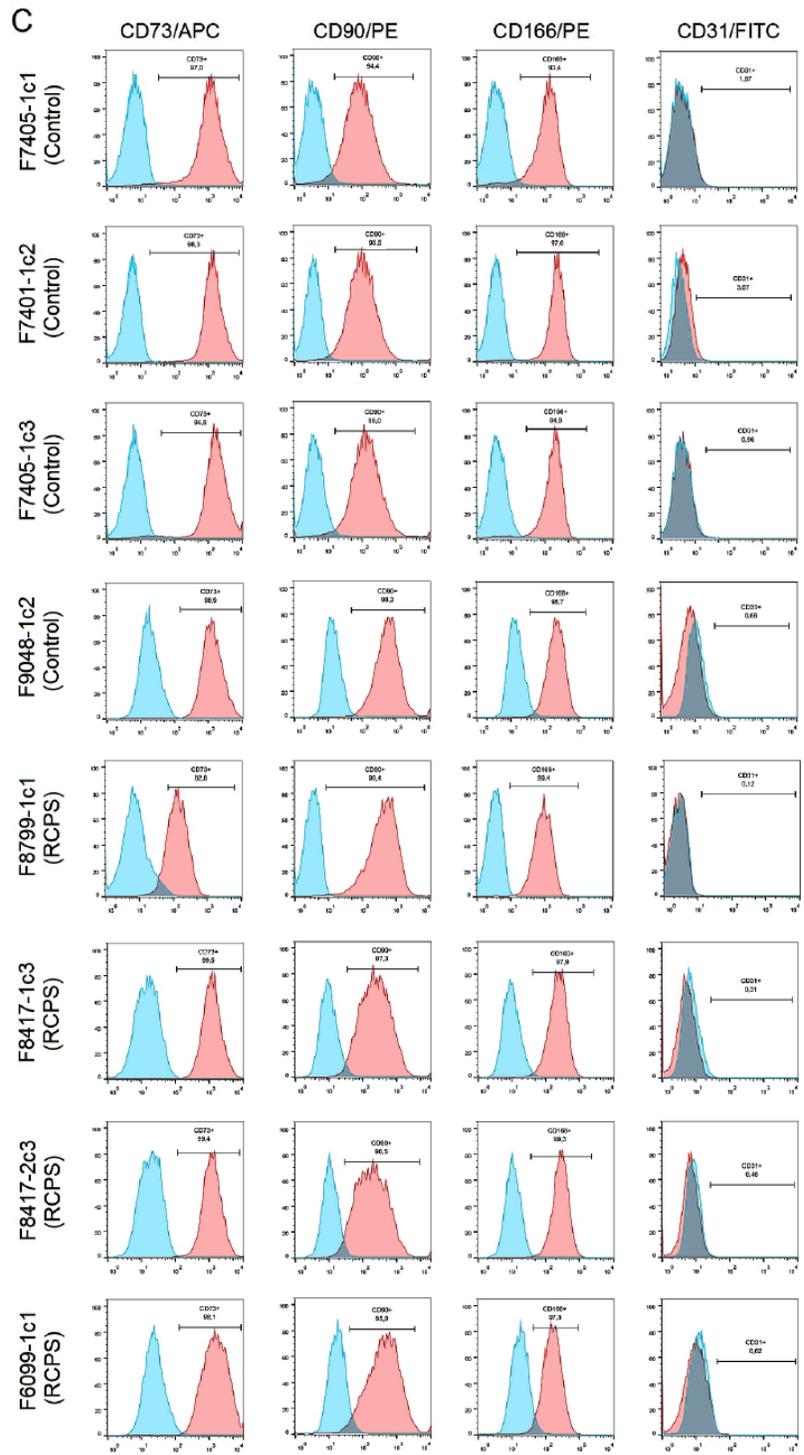
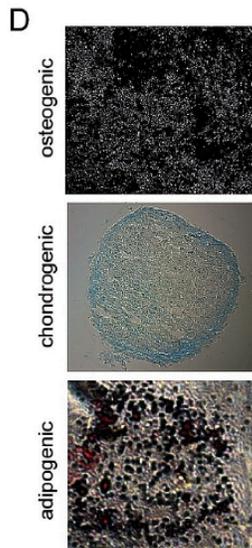
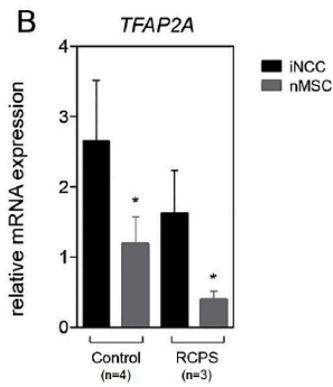
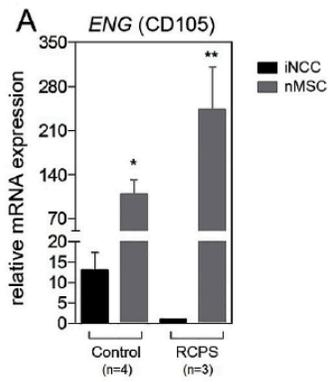


Figure S4. Characterization of nMSCs. **(A, B)** RT-qPCR analysis showing upregulation of mesenchymal marker *ENG* (CD105) **(A)** and downregulation of neural crest marker *TFAP2A* **(B)** in nMSCs compared to iNCCs with n=number of biological samples. **(C)** Flow cytometry immunophenotype profile of nMSCs showing positive staining for mesenchymal markers CD73, CD90 and CD166, and negative staining for endothelial marker CD31. Histograms represent event count (y-axis) vs. fluorescence (x-axis). Experimental data (red) were plotted in overlay with data from isotype controls (blue). **(D)** Representative images of osteogenic, chondrogenic, and adipogenic differentiation of nMSCs, detected with Alizarin Red staining (10x magnification), Alcian Blue staining (10x magnification), and Oil Red staining (40x magnification), respectively. Data are represented as mean \pm SEM. (**) p-value < 0.01, (*) p-value < 0.05, Student's t-test.

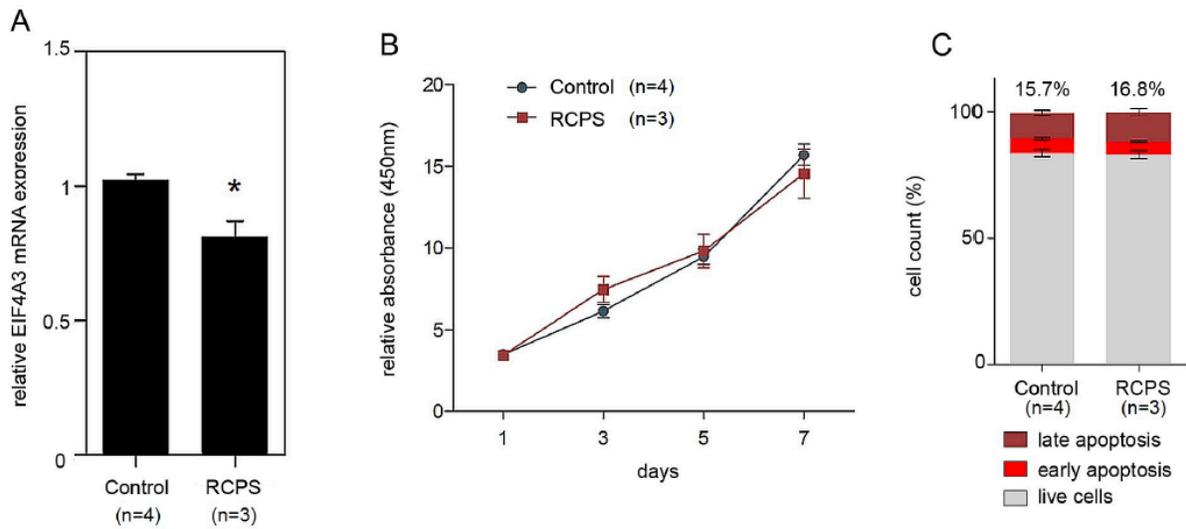


Figure S5. Proliferation and apoptosis analysis of nMSCs. **(A)** RT-qPCR assessment of *EIF4A3* mRNA expression in nMSCs from controls and RCPS patients. **(B)** XTT assay depicting the proliferation profile of control and RCPS nMSCs. **(C)** 7-AAD/Annexin V flow cytometry results showing the fraction of live cells and cells undergoing early and late apoptosis in nMSCs; the percentages represent the sum of early and late apoptotic events. Values represent mean \pm SEM, n=number of biological samples. (*) p-value < 0.05, Student's t-test.

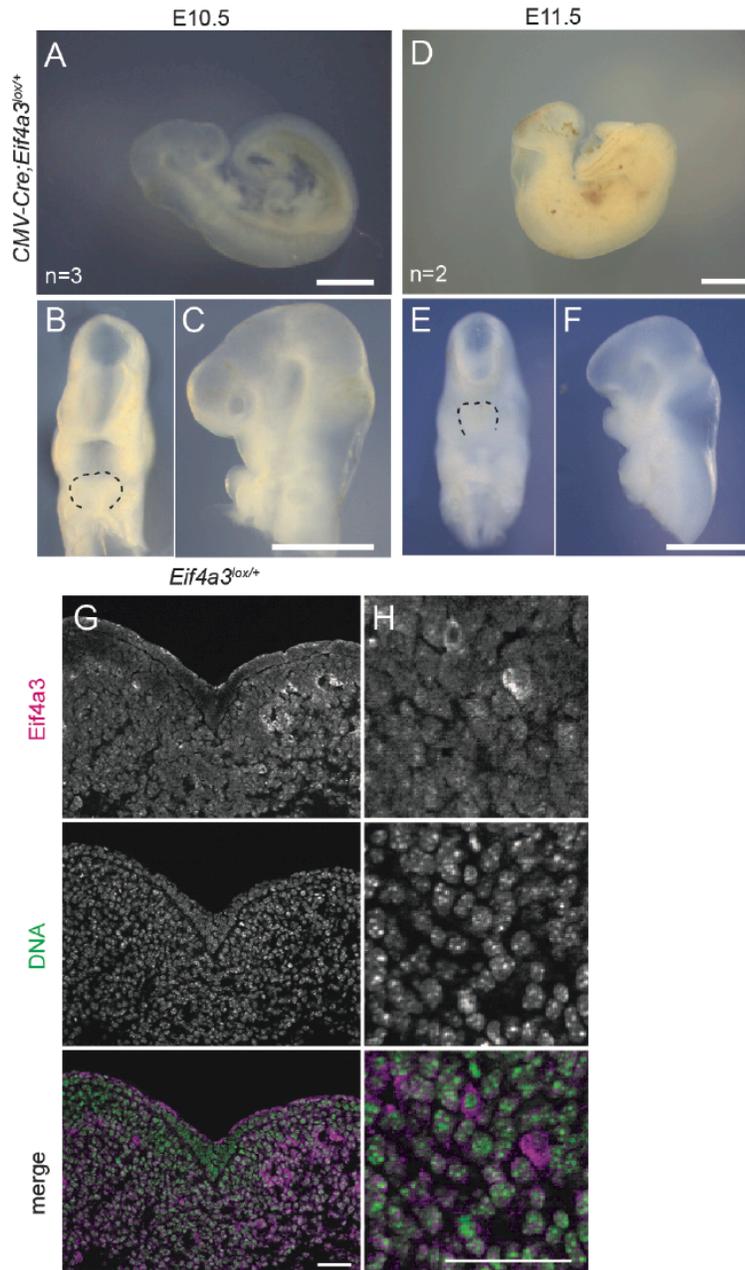


Figure S6. Phenotypic characterization of *Cmv-Cre; Eif4a3^{lox/+}* embryos and *EIF4A3* expression. (**A-C**) E10.5 *Cmv-Cre; Eif4a3^{lox/+}* mutant type B embryos exhibit reduced body size and severe craniofacial abnormalities resembling holoprosencephaly as outlines with a black dotted line in (**B**). (**D-F**) E11.5 type B *Cmv-Cre; Eif4a3* mutant embryos are developmentally delayed and exhibit holoprosencephaly and only one mandibular arch as outlined with a black dotted line in (**E**). (**G-H**) Sections of E11.5 mandible in *Eif4a3^{lox/+}* embryos show EIF4A3 expression is primarily nuclear at 1x (**G**) and 3.77x zoom (**H**). n=number of embryos. Scale bars, 1 mm (**A-E**); 50 μ m (**G-H**).

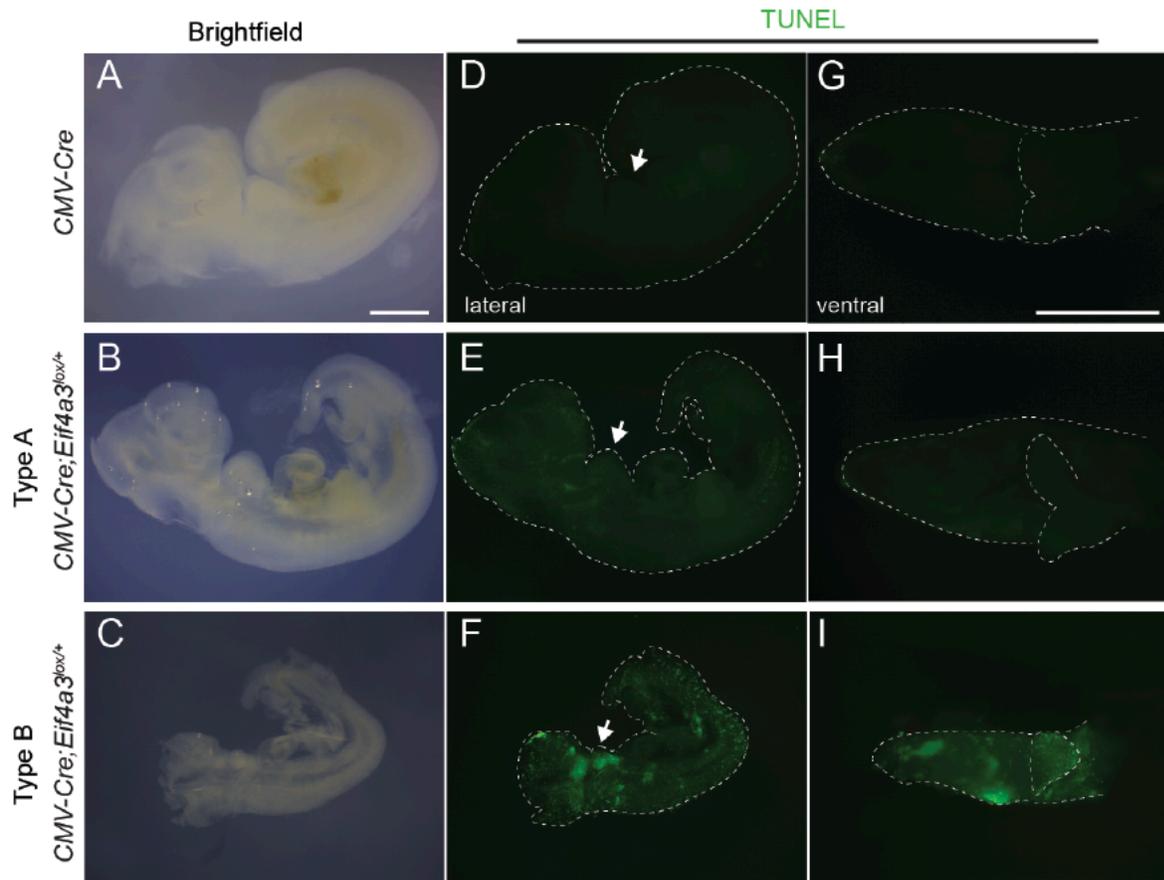


Figure S7. *Cmv-Cre; Eif4a3^{lox/+}* Type B mutants show cell death throughout the body and in developing face. (A-C) Representative brightfield images of *Cmv-Cre* (A), type A *Cmv-Cre; Eif4a3^{lox/+}* (B) and type B *Cmv-Cre; Eif4a3^{lox/+}* embryos (C). (D-I) The same embryos depicted in A-C with TUNEL staining to highlight cell death with lateral views (D-F) and ventral views (G-I). Dotted lines demarcate the whole embryo (D-F) and mandible and heads (G-I) and white arrows point to the developing mandible (D-F). Scale bars, 1 mm.

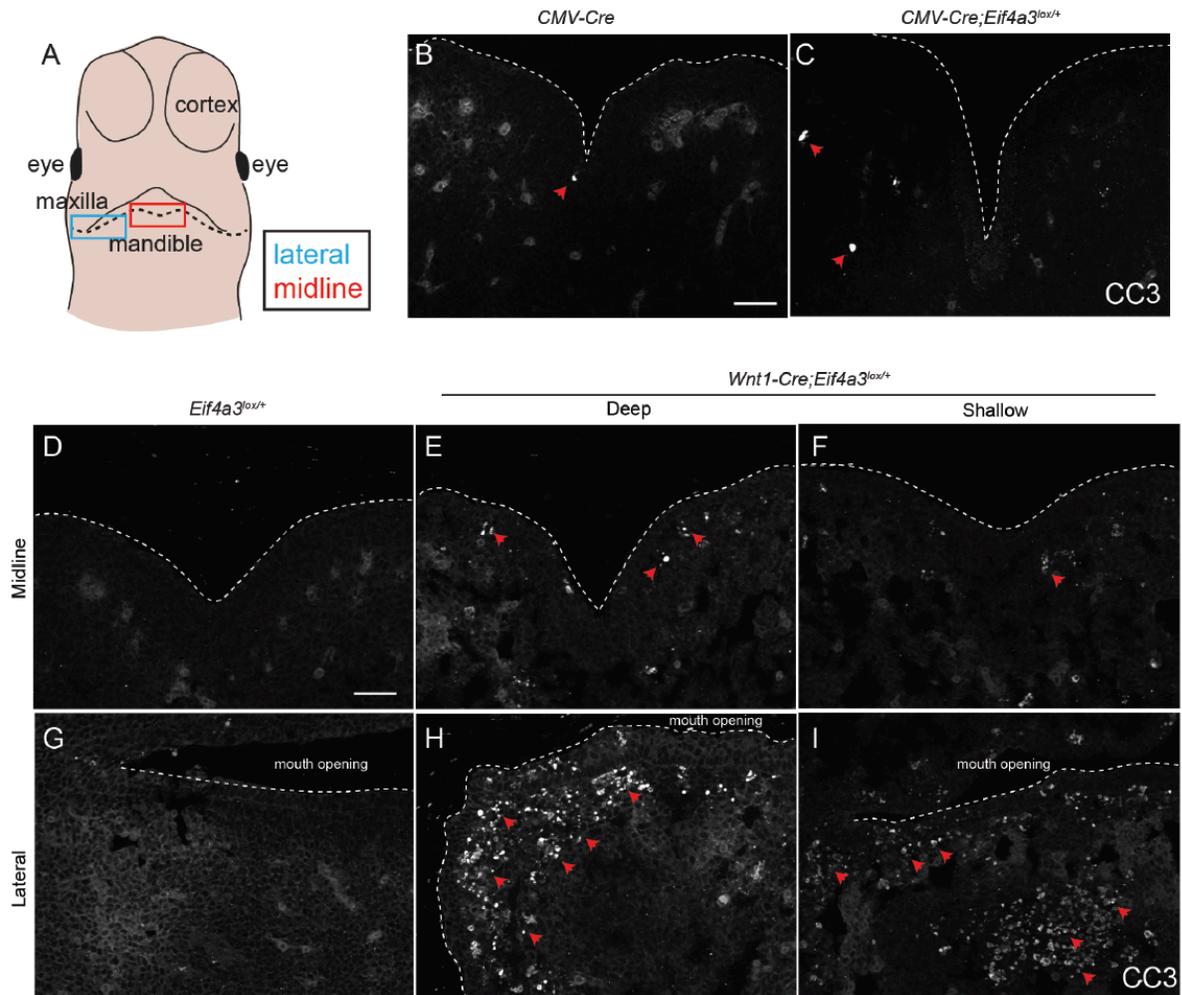


Figure S8. *Cmv-Cre; Eif4a3^{lox/+}* and *Wnt1-Cre; Eif4a3^{lox/+}* embryos show few CC3-positive cells in the mandible midline. (A) Schematic representing a frontal view of an embryo with midline regions indicated for (B-F) and lateral regions indicated for (G-I). (B, C) Images of CC3 staining of the mandible midline in E11.5 *Cmv-Cre* control (B) and type A *Cmv-Cre; Eif4a3^{lox/+}* embryos (C), showing little cell death in either genotype. (D-I) Images of CC3 staining of the mandible midline (D-F) and lateral mandible (G-I) in E11.5 *Eif4a3^{lox/+}* control (D, G) and *Wnt1-Cre; Eif4a3^{lox/+}* embryos (E, F, H, I). Note there is little cell death at midline but in the lateral mandible, away from the groove defects, there is apparent CC3 staining and cell death. Mouth openings are labeled to help orient the image. Red arrows point to some cells undergoing apoptosis. Scale bars, 50 μ m.

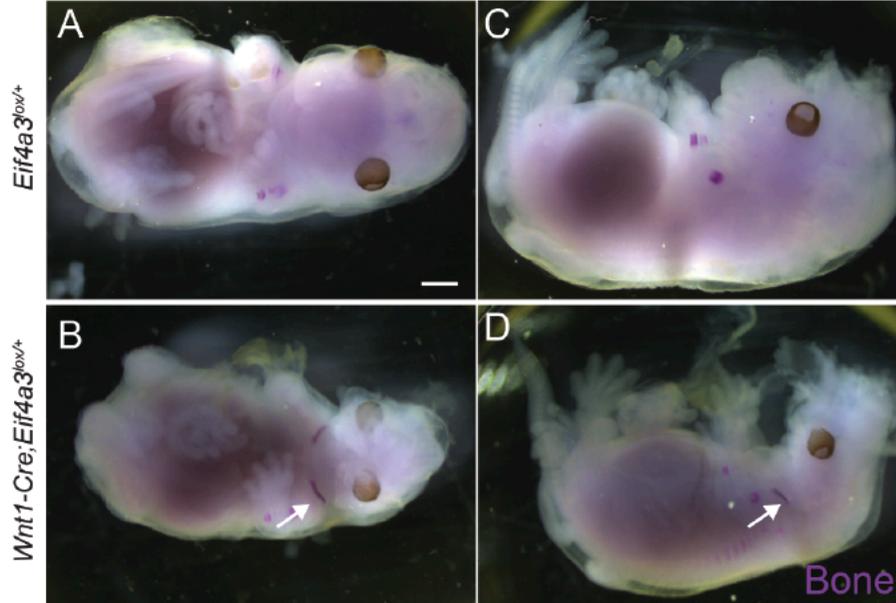


Figure S9. Skeletal preparations depicting clavicle ossification in E14.5 *Wnt1-Cre; Eif4a3^{lox/+}* embryos. (A-D) Whole mount E14.5 *Eif4a3^{lox/+}* control embryos (A, C) and littermate *Wnt1-Cre; Eif4a3^{lox/+}* embryos (B, D) stained with Alizarin red. Note presumptive bone in the developing limb of control and conversely ossification occurring in limbs, clavicles (arrows), and ribs of mutant. Scale bars, 1 mm.

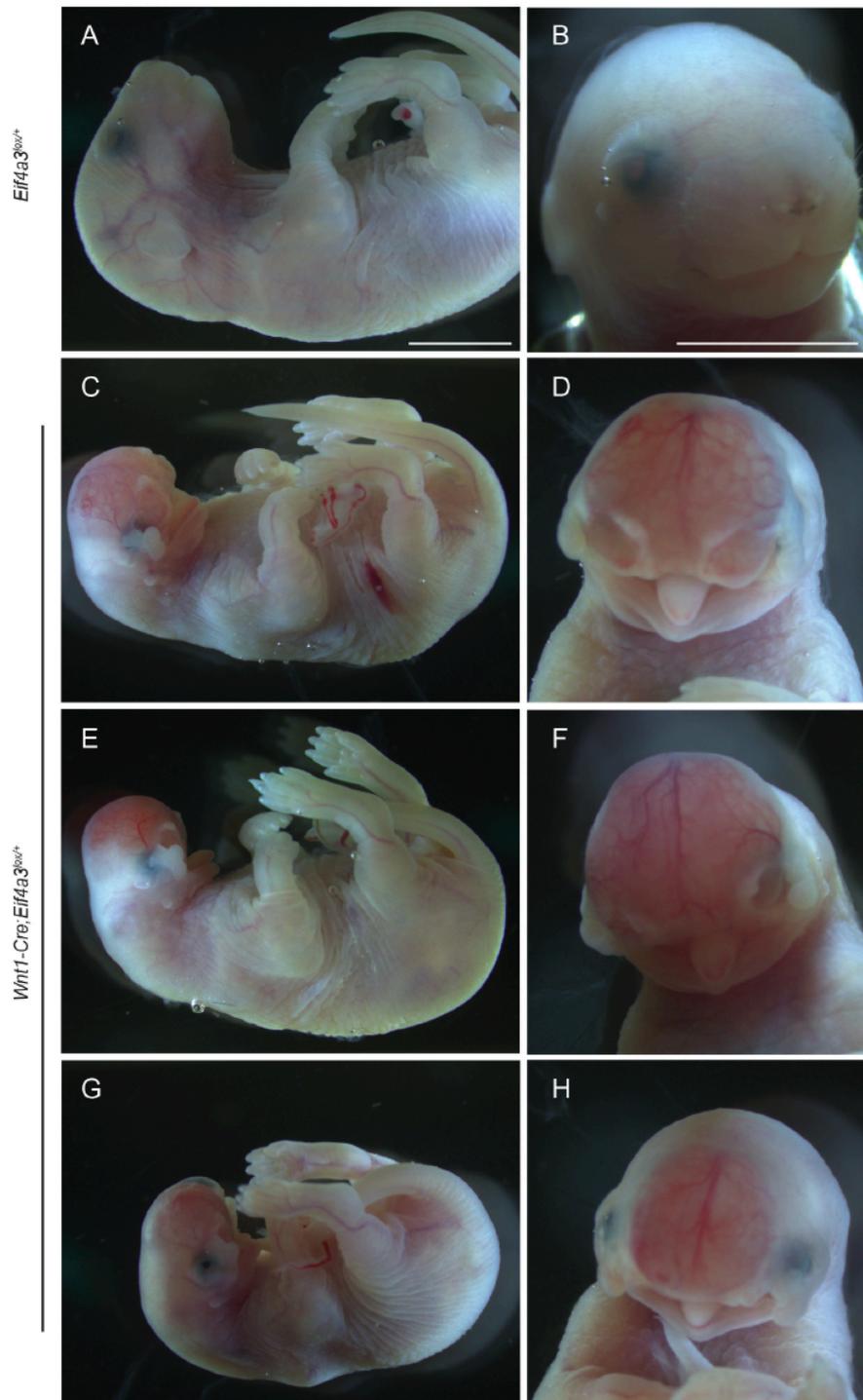


Figure S10. E18.5 *Wnt1-Cre; Eif4a3^{lox/+}* embryos have severe craniofacial defects but develop a non clefted tongue. (A-H) Whole mount images of lateral view (A, C, E, G) and ventral view of face (B, D, F, H) of *Eif4a3^{lox/+}* control embryo (A, B) and three different *Wnt1-Cre; Eif4a3^{lox/+}* mutant embryos (C-H). Scale bars, 5 mm.

Table S1. Cell samples

ID/clone	Subject	Cell type	Age*	Gender	Phenotype	Genotype <i>EIF4A3</i> **	Reprogramming method	Assays						
								<i>EIF4A3</i> RT-qPCR	Western blot	Wound healing	Mesenchymal differentiation	Apoptosis	Cell cycle	XTT Proliferation
F7405-1 c1	F7405-1	fibroblasts	27	male	control	7 repeats 6 repeats	retroviral	X	X	X	X	X	X	X
F7405-1 c2								X			X	X		X
F7405-1 c3								X			X	X		X
F9048-1 c2	F9048-1	fibroblasts	20	male	control	6 repeats 6 repeats	episomal	X	X	X	X	X	X	X
F8799-1 c1	F8799-1	erythroblasts	20	female	control	7 repeats 6 repeats	episomal		X	X			X	
F8417-1 c3	F8417-1	fibroblasts	11	female	RCPS	16 repeats 16 repeats	episomal	X	X	X	X	X	X	X
F8417-2 c3	F8417-2	fibroblasts	9	male	RCPS	16 repeats 16 repeats	episomal	X	X	X	X	X	X	X
F6099-1 c1	F6099-1	fibroblasts	10	male	RCPS	14 repeats c.809A>G	episomal	X	X	X	X	X	X	X

*age upon tissue collection

**please refer to Favaro et al (2014) for additional information

Table S2. Primer sequences used in RT-qPCR

Target	Forward primer (5'-3')	Reverse Primer (5'-3')
<i>ACTB</i>	TGAAGTGTGACGTGGACATC	GGAGGAGCAATGATCTTGAT
<i>GAPDH</i>	ATCACCATCTTCCAGGAGCG	GGGCAGAGATGATGACCCTTT
<i>HPRT1</i>	CCTGGCGTCGTGATTAGTGAT	AGACGTTTCAGTCCTGTCCATAA
<i>HMBS</i>	AGCTTGCTCGCATAACAGACG	AGCTCCTTGGTAAACAGGCTT
<i>TBP</i>	GTGACCCAGCATCACTGTTTC	GCAAACCAGAAACCCTTGCG
<i>OCT3/4</i>	GTGGTCAGCCAACTCGTCA	CCAAAAACCCTGGCACAACCT
<i>OriP</i>	TTCCACGAGGGTAGTGAACC	TCGGGGGTGTTAGAGACAAC
<i>NANOG</i>	TGGACACTGGCTGAATCCTTC	CGTTGATTAGGCTCCAACCAT
<i>EIF4A3</i>	GGAGATCAGGTCGATACGGC	GATCAGCAACGTTTCATCGGC
<i>PAX3</i>	AAGCCCAAGCAGGTGACAAC	CTCGGATTTCAGCTGAAC
<i>ZIC1</i>	AAGGTCCACGAATCCTCCTC	TTGTGGTCGGGTTGTCTG
<i>TFAP2A</i>	CTCCGCCATCCCTATTAACAAG	GACCCGGAAGTGAACAGAAGA
<i>SOX10</i>	CCTCACAGATCGCCTACACC	CATATAGGAGAAGGCCGAGTAGA
<i>ENG</i>	TGCACTTGGCCTACAATTCCA	AGCTGCCCCACTCAAGGATCT
<i>RUNX2</i>	AGTGGACGAGGCAAGAGTTTC	GTTCCCGAGGTCCATCTACTG
<i>ALP</i>	GATACAAGCACTCCCACTTCATCTG	CTGTTTCAGCTCGTACTGCATGTC
<i>COL1A1</i>	GGGCCAAGACGAAGACAT	CAACACCCTTGCCGTTGTCTG
<i>BGLAP</i>	GGCGCTACCTGTATCAATGG	GTGGTCAGCCAACTCGTCA

Table S3. Mouse genotyping primers

Target	Forward Primer (5'-3')	Reverse primers (5'-3)	Conditions
<i>Eif4a3^{lox}</i>	CTTGCAGTTGTCTTT CTGCGG	AAGGGTTATTGAATAT GATCGGAATTG	98°C X 30s (1X); 98°C X 10s, 65°C X 20s, 72°C X 20s (35X); 72°C X 2min (1X) with Q5 DNA polymerase (NEB)
<i>Cmv-Cre</i>	GCATTACCGGTCGAT GCAACGAGTGATG	GAGTGAACGAACCCT GGTCGAAATCAGTG	94°C X 3 min (1X); 94°C X 30s, 61°C X 30s, 72°C X 30s (35X); 72°C X 7min (1X)
FABP (internal control for Cre)	TGGACAGGACTGGA CCTCTGCTTTCCTA	TAGAGCTTTGCCACAT CACAGGTCATTCA	Same as <i>Cmv-Cre</i>
Wnt- Jackson Lab: 16773 and 16774 transgene	CAGCGCCGCAACTA TAAGAG	CATCGACCGGTAATGC AG	94°C X 2 min (1X); 94°C X 20s, 65°C X 15s, 68°C X 10s (10X); 94°C X 20s, 60°C X 15s, 72°C X 10s (28X); 72°C X 1 min (1X)
Wnt- Jackson Lab oIMR8744 and oIMR8745	CAAATGTTGCTTGTC TGGTG	GTCAGTCGAGTGCACA GTTT	Same as <i>Wnt1</i> transgene

CHAPTER IV

RNA Sequencing of *EIF4A3* Deficient Neural Crest Cells Reveals Novel Pathways Functionally Linked to RCPS Pathogenesis

Camila M. Musso¹, Gerson S. Kobayashi¹, Karina O. Griesi², Giovanna Pontillo-Guimarães¹, Gabriella S.P. Hsia¹, Emily E. Miller³, Luiz C. de Caires Jr¹, Ernesto Goulart¹, Cibele Masotti⁴, Pedro A.F. Galante⁴, Flavia V. Winck⁵, Debra L. Silver³, Maria Rita Passos-Bueno^{1*}

¹Centro de Estudos do Genoma Humano e Células Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

²Hospital Israelita Albert Einstein, São Paulo, Brazil

³Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, USA

⁴Centro de Oncologia Molecular, Instituto Sírio-Libanês de Ensino e Pesquisa, Hospital Sírio-Libanês, São Paulo, Brazil

⁵Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

*Correspondent author

Keywords: iPSCs, Richieri-Costa-Pereira syndrome, pathogenetic mechanisms, phenotype, migration defects, craniofacial abnormalities, disease modeling, mice, ribosomes

ABSTRACT

Richieri-Costa-Pereira syndrome (RCPS) is an autosomal-recessive condition mainly characterized by craniofacial and limb abnormalities. This disorder is caused by hypomorphic biallelic mutations in the *EIF4A3* 5'UTR, which is a core component of the exon junction complex (EJC). We have previously shown that *EIF4A3* deficiency leads to defective neural crest cell (NCC) migration and dysregulated osteogenic/chondrogenic differentiation of NCC mesenchymal derivatives. Yet despite the requirement of *EIF4A3* for NCC and craniofacial development, the pathogenetic molecular mechanisms responsible for the impaired NCC functions are entirely unknown. Here, we employed the use of patient iPSC-derived NCCs combined with conditional *Eif4a3* haploinsufficient mouse model to uncover altered pathways underlying RCPS pathogenesis. Transcriptomic and proteomic analyses indicated that reduced levels of *EIF4A3* alter expression and/or splicing of cell-extracellular matrix (ECM) interaction components as well as ribosomal and neurogenesis-related genes. Altogether, this study associates aberrant splicing of riboproteins and defective cell-adhesion with NCC migration/differentiation defects in RCPS etiology.

INTRODUCTION

Richieri-Costa-Pereira syndrome (RCPS; OMIM #268305) is a rare autosomal-recessive acrofacial dysostosis reported mostly in Brazilian patients. RCPS is characterized by several craniofacial abnormalities, including midline cleft of the mandible, microstomia, Robin sequence, absence of central lower incisors, laryngeal and ear anomalies along with limb defects and learning impairments (Favaro et al. 2011), as well as microcephaly which has recently been included in the phenotypic spectrum of this disorder (Bertola et al. 2018).

RCPS is mainly caused by an increased number of repeat motifs within the *EIF4A3* 5'UTR, which in turn results in 15-70 % reduction in *EIF4A3* expression depending on cell type (Favaro et al. 2014; Hsia et al. 2018; Miller et al. 2017). The RNA-binding protein eIF4A3 is a core component of the exon junction complex (EJC), which is involved in various aspects of mRNA metabolism including splicing, mRNA export and subcellular localization, nonsense-mediated RNA decay (NMD), and translational control (Palacios et al. 2004; Le Hir et al. 2016).

We have demonstrated that *EIF4A3* deficiency impairs neural crest cell (NCC) development, a finding that is consistent with the clinical features of RCPS patients, which involve malformation of craniofacial structures originated from NCCs. Patient-derived NCCs presented decreased migratory capacity and defective osteogenic/chondrogenic differentiation of NCC mesenchymal derivatives (Miller et al. 2017). However, the molecular mechanisms underlying these impaired cell functions remain unclear.

Numerous disorders beyond RCPS, known collectively as neurocristopathies (Bolande 1974) comprise congenital malformations caused by disruption of neural crest cell development e.g. Diamond Blackfan Anemia, Frontonasal Dysplasia and Auriculo-Condylar syndrome (Vega-Lopez et al. 2018). NCCs are a transient multipotent stem cell population originating from the neural plate border and give rise to the peripheral neurons, glia, muscle, bone, cartilage and connective tissues that generate most of the cranium (Cordero et al. 2011; Trainor 2013).

Here, in order to expose candidate targets and pathways controlled by *EIF4A3* at the onset of NCC development, we assessed transcriptome sequencing of two complementary

models, RCPS patient-derived iPSCs and conditional mouse mutants. A better understanding of the pathogenetic mechanisms by which *EIF4A3* mutations impair NCC development will not only shed light on RCPS pathophysiology but will also clarify the molecular underpinnings that regulate craniofacial development.

RESULTS

Transcriptional changes in RCPS hNCCs

iPSCs were differentiated towards NCCs by using a combination of WNT activation and SMAD inhibition without the presence of caudalization factors (Miller et al. 2017; Menendez et al. 2013; Fukuta et al. 2014), as well as without regulation of cranial positioning (Suga et al. 2017). Although it has been shown that this method favors cranial NCC formation (Huang et al. 2016), the human iPSC-derived neural crest cell (hNCCs) are representative of a mixed population comprising all NCC domains.

Information regarding all human samples used in this study is depicted in table S1 and both the iPSC and NCC lineages were extensively characterized elsewhere (Miller et al. 2017), with the exception of the novel F11167-1 cl1 line (Fig. S1; see Material and Methods). Briefly, iPSCs displayed positive staining for pluripotency markers OCT3/4 and SSEA-4, did not present aneuploidies, and demonstrated the ability to generate teratomas *in vivo*, as well as were successfully differentiated in homogenous NCC populations with high expression of NCC markers and p75/HNK1 double-positive staining.

Comparative transcriptome analysis between patient-derived and control NCCs revealed alterations in expression of 121 transcripts (FDR, $q < 0.1$) amongst the 13,915 detectable coding and noncoding transcripts expressed in NCCs, corresponding to 0.9 % of the total. Out of the 121 differentially expressed transcripts (DEGs), 92 (76 %) are coding and the remaining are noncoding and pseudogenes (Table S4).

Hierarchical clustering of the DEGs revealed segregation of control and RCPS patient-derived samples, as shown in the heatmap of figure 1A. A higher number of transcripts (77) were

upregulated in RCPS patients (Fig. 1B and 1C; Table S4). Several identified DEGs (*CCDC80*, *CAV1*, *CELF2*, *LFNG*, *SOX1*, *POSTN*, *COL4A1* and *COL4A2*) were validated by RT-qPCR assays, which showed similar trends to RNA-seq data (Fig. 1D).

***EIF4A3* deficiency disrupts expression of cell adhesion to ECM components in hNCCs**

Next, we assessed molecular pathways that are altered in response to decreased *EIF4A3* expression. GSEA using the DAVID KEGG analysis demonstrated a significant enrichment ($p < 0.1$) of extracellular matrix- (ECM) receptor interaction and focal adhesion (FA) pathways (Fig. 1E; Table S5), which are known to be important during cell migration and wound healing.

Closer inspection of the DEGs within each GO category revealed enrichment mainly of extracellular matrix and plasma membrane components, related to cell projection and ECM structural organization ($p < 0.05$) (Fig. 1E).

DEGs related to regionalization and pattern specification (e.g. *KIAA1324L*, *AMER2*, *FRAT2* and the homeobox genes *MSX2*, *HOXD1*, *HOXC10* and *HOXD3*) were also observed. Additionally, one category that encompasses a notable proportion of the coding DEGs (23.14 %) is nervous system development (GO:0007399), which includes genes that play important roles in cell fate decision and neurogenesis (e.g. *SOX1*, *PAX5*, *CHURC1*, *FEZ1*, *MAP2*, *SEMA3C* and *EPHA5*).

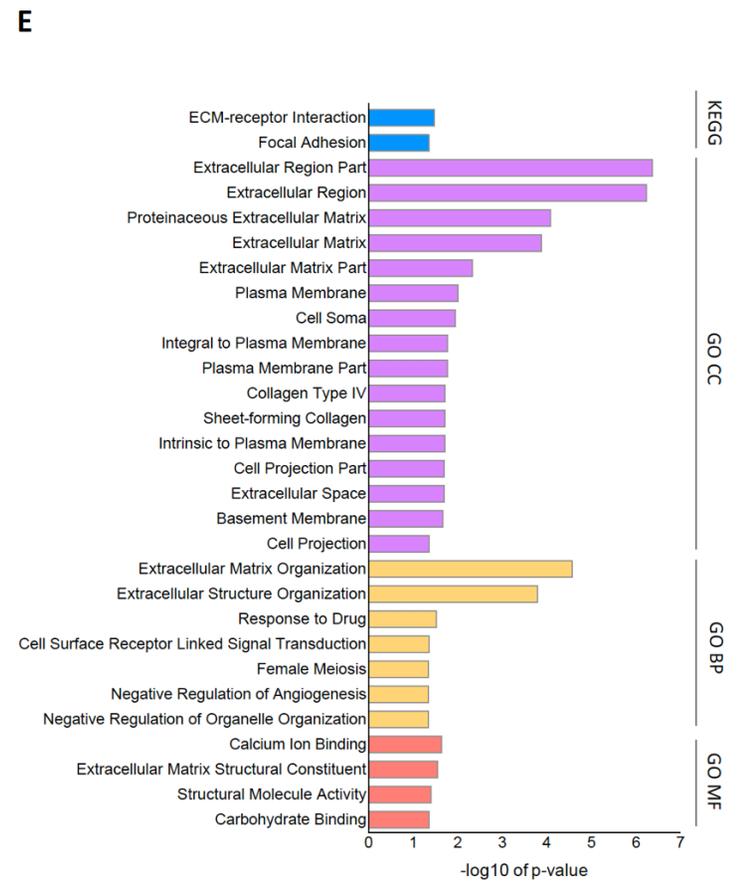
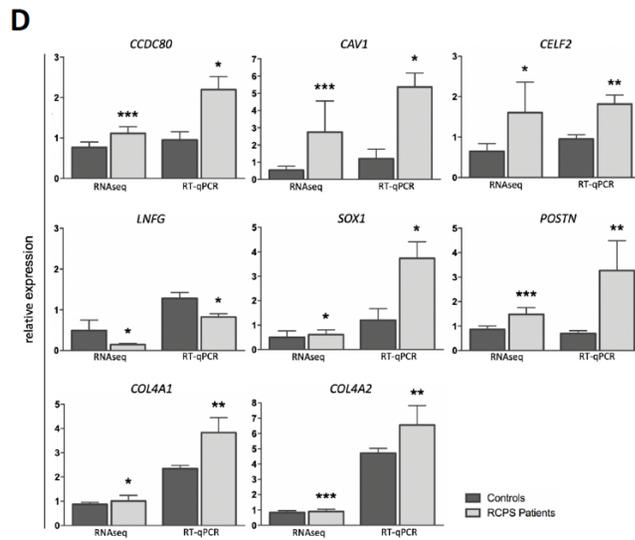
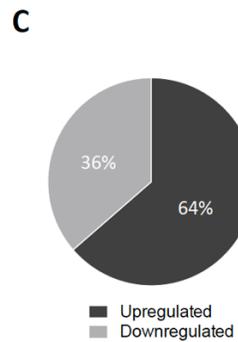
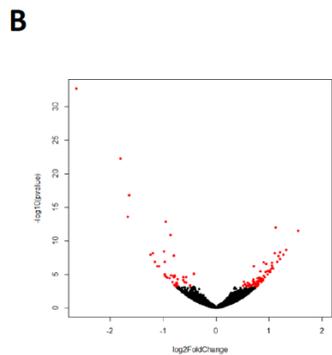
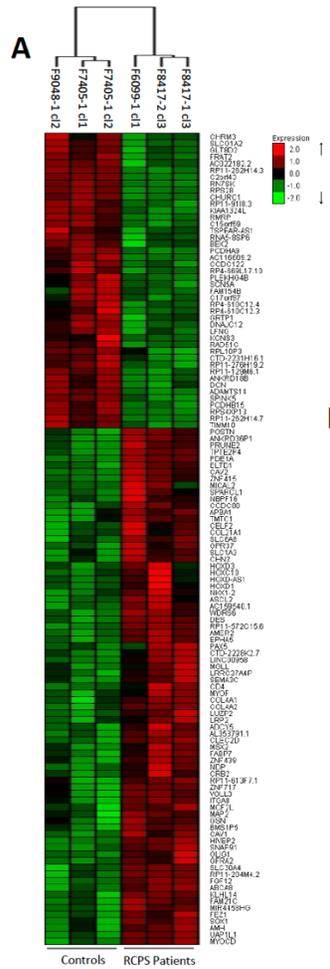


Figure 1. Transcriptome analysis of *EIF4A3* deficient human iPSC-derived neural crest cells. (A) Heatmap showing z-score normalized expression for all 121 transcripts identified as significantly different between RCPS patients and controls (FDR, $q < 0.1$), and the hierarchical clustering of the samples based on all genes using a correlation distance with complete linkage. (B) Volcano plot indicating the $-\log_{10}$ of p-value (y axis) for each transcript plotted against its \log_2 fold-change (x axis). Transcripts significantly affected are displayed in red dots. (C) Pie chart of the distribution of upregulated and downregulated DEGs. (D) RT-qPCR validation of *CCDC80*, *CAV1*, *CELF2*, *LFNG*, *SOX1*, *POSTN*, *COL4A1* and *COL4A2* compared to relative RNA-seq values. Control samples were normalized to 1.0 and data are represented as mean \pm SEM; Student's *t* test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (E) Gene set enrichment analysis of DEGs, depicting the enriched KEGG pathways ($q < 0.1$) and GO terms ($q < 0.05$) identified by DAVID; Modified Fisher Exact test (EASE Score), $-\log_{10}$ of p-values are displayed.

***EIF4A3* deficient hNCCs show aberrant splicing of riboproteins**

Considering the requirement of the EJC in splicing, we next analyzed differential splicing between controls and RCPS patients' NCCs. We identified 234 alternative splicing (AS) events (FDR, $q < 0.1$) (Table S6). These events occurred in 128 different genes, since some genes presented more than one AS event, as represented in the Venn Diagram (Fig. 2B). Out of these genes, 116 (90.6 %) are protein coding and the remaining are pseudogenes and long non-coding RNAs (Table S6).

The splicing events were mainly skipped exons (SE; 140), representing approximately 60 % of all AS events, followed by mutually exclusive exons (MXE; 44), with a lower proportion of alternative 5' or 3' splice sites (A5SS; 21 and A3SS; 15) and retained introns (RI; 14) (Fig. 2A). RT-qPCR assays validated four events localized in *FANCA*, *RPS9*, *SAT2* and *RPL9* (Fig. 2C) and one was not validated (*RPS28*).

We performed GSEA on our set of genes with significant alteration in splice variant expression. Remarkable, ribosomal transcripts made up 5.6 % of AS events (*MRPL18*, *MRPS10*, *RPL37A*, *RPL9*, *RPS2*, *RPS28*, *PRS9*). KEGG analysis revealed enrichment of ribosome and metabolic pathways (Fig. 2D). This finding was supported by GO analysis, which also detected ribosome as an enriched category in addition to biological processes related to cell cycle/division and translation (Fig. 2D; Table S7).

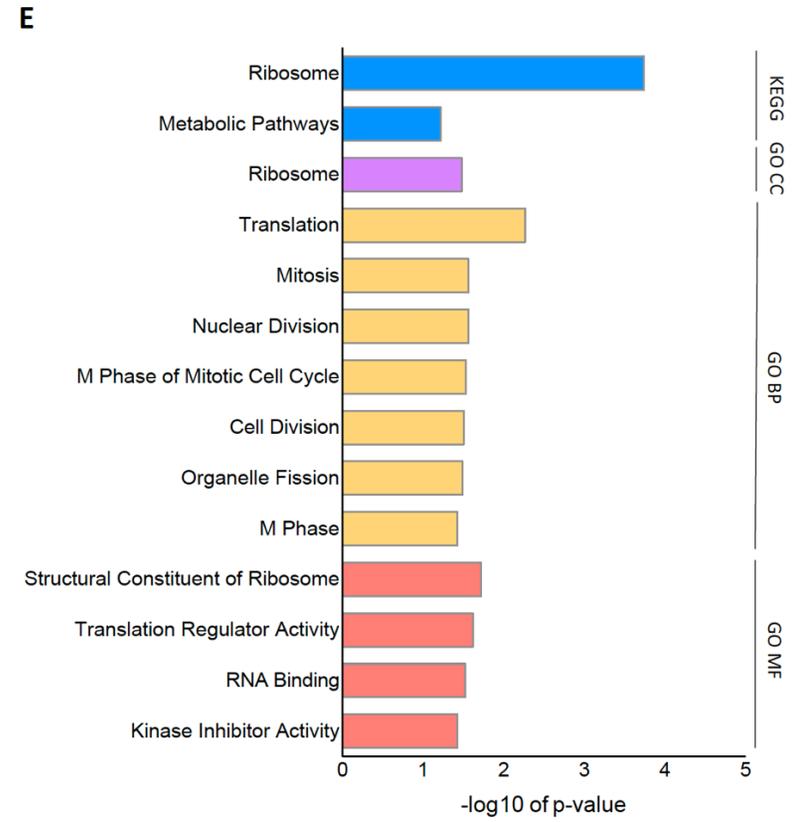
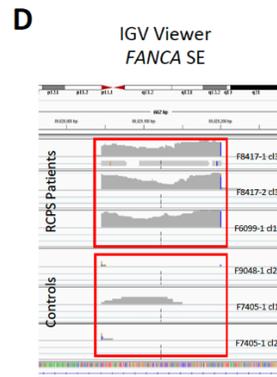
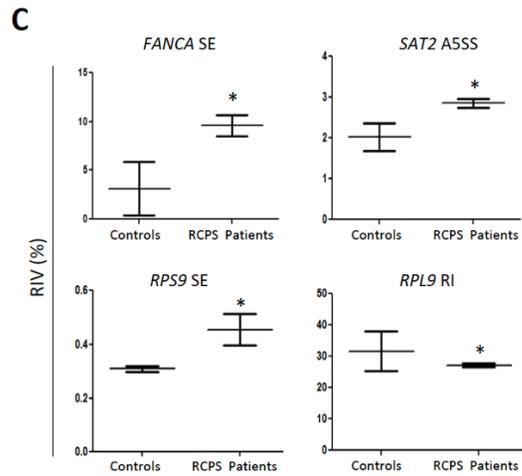
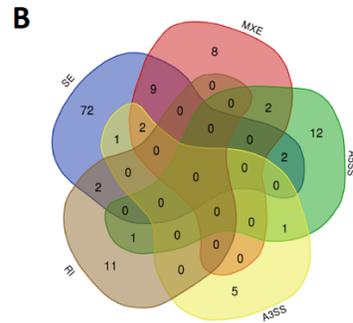
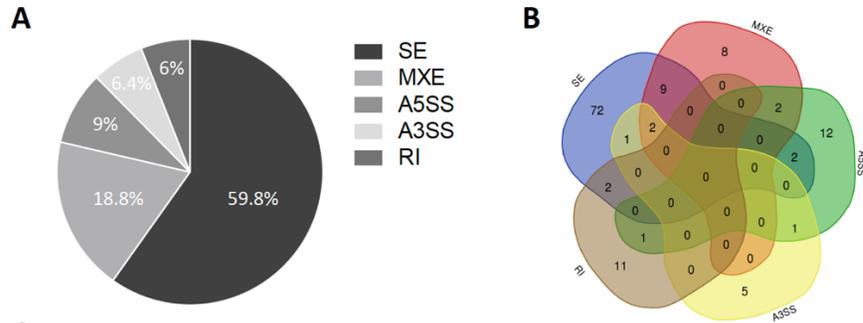


Figure 2. Splicing analysis of *EIF4A3* deficient human iPSC-derived neural crest cells. (A) Pie chart of the distribution of alternative splicing among skipped exon (SE), mutually exclusive exon (MXE), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS) and retained intron events. (B) Venn diagram illustrating the number of overlapping genes between AS events. (C) RT-qPCR validation of two SE events (*FANCA* and *RPS9*), an A5SS (*SAT2*) and a RI (*RPL9*). Graphs depicting the relative incidence of a single variant (RIV) in controls and RCPS patients. RIV is defined as the amount of a splicing variant relative to the amount of the external control; Student's *t* test, * $p < 0.05$. (D) IGV view of increased number of reads of *FANCA* exon (between 28-29 / 43 total) in RCPS patients compared to controls (red frame), confirming SE event. (E) Gene set enrichment analysis of alternatively spliced genes, showing the enriched KEGG pathways ($q < 0.1$) and GO terms ($q < 0.05$) identified by DAVID; Modified Fisher Exact test (EASE Score), $-\log_{10}$ of p-values are displayed.

***EIF4A3* deficiency influences protein levels of ribosome and proteasome components in hNCCs**

We also assessed the proteome of *EIF4A3* deficient and control NCCs using liquid chromatography mass spectrometry (LCMS)-based proteomics. A total of 625 proteins were detected. Out of the 625, 14 were found exclusively in RCPS patients and 6 were found exclusively in control individuals (Fig. 3B; Table S8) and the remaining proteins were expressed in both groups (605, 96.8%). Two exclusively expressed proteins (EEPs) also presented AS events in their corresponding genes (*RPS9* and *CTPS1*).

Ultimately, we identified 13 (2.1 %) differentially expressed proteins (DEPs), of which 9 were upregulated and 4 were downregulated in RCPS patients ($p < 0.05$). Hierarchical clustering of the DEPs revealed segregation of control and RCPS patient samples (Fig. 3A). In accordance with prior transcriptome data, the altered proteins enclose a considerable fraction of ribosomal and proteasome components (18.2 %). There was no overlap between DEGs and DEPs.

GSEA KEGG analysis of DEPs and EEPs revealed proteasome and ribosome enriched pathways and GO analysis reinforced ribonucleoprotein complex and proteasome complex as enriched terms along with biological processes mainly related to ubiquitination and regulation of ligase activity (Fig. 3D; Table S9).

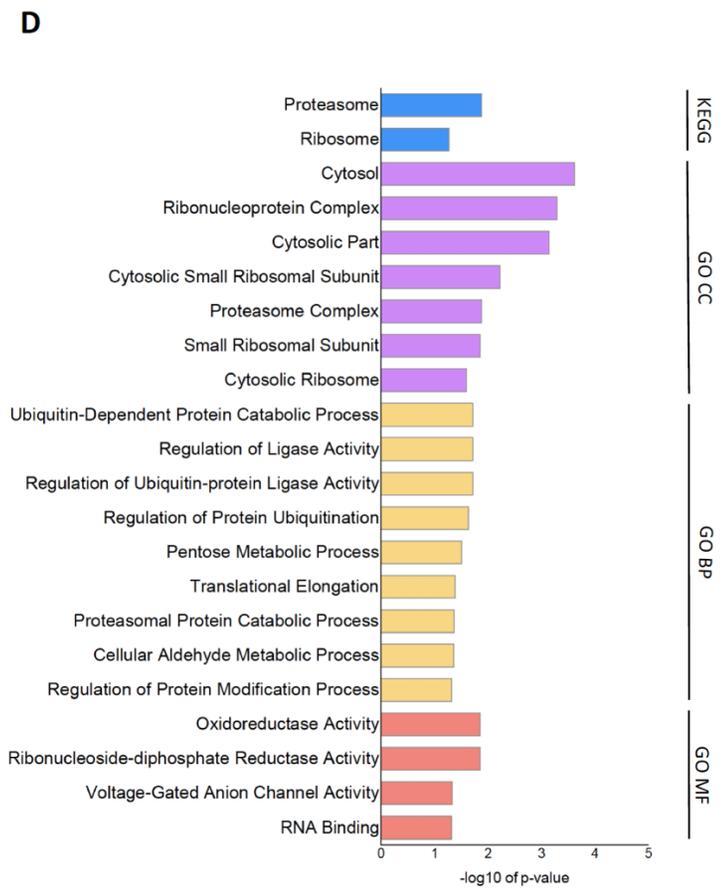
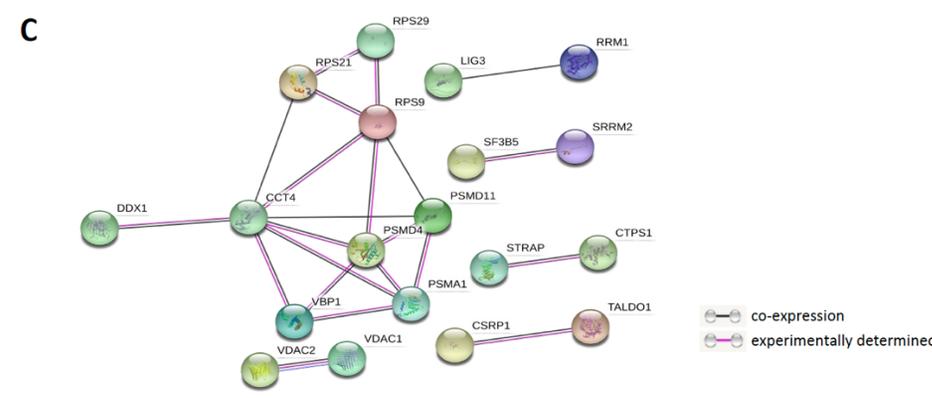
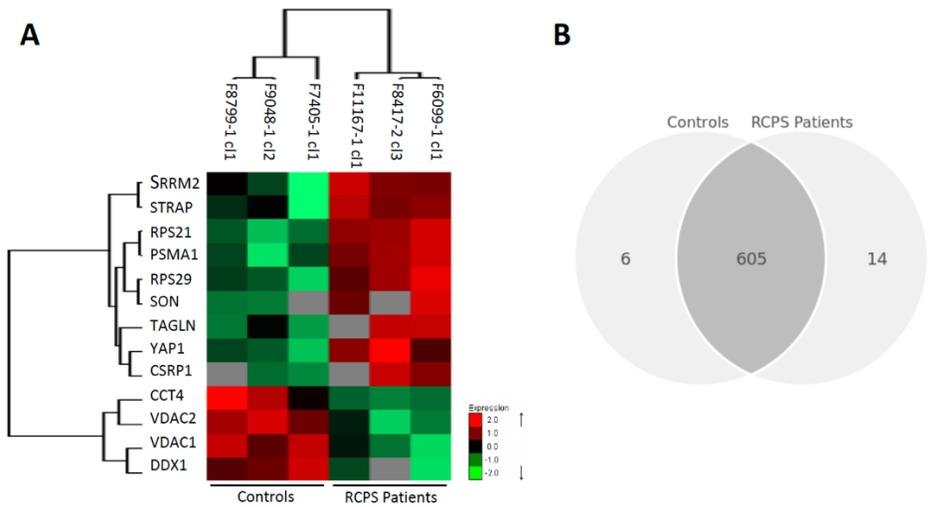


Figure 3. Proteome analysis of *EIF4A3* deficient human iPSC-derived neural crest cells. (A) Heatmap showing z-score normalized expression for all 13 proteins identified as significantly different (DEPs) between RCPS patients and controls (FDR, $q < 0.05$), and the hierarchical clustering of the samples and proteins by correlation distance with complete linkage. (B) Venn diagram illustrating the number of overlapping proteins in both RCPS and control NCCs and proteins exclusively identified in each group (EEPs). (C) STRING network analysis of DEPs and EEPs; co-expression-based associations are represented by black lines and experimentally determined by pink lines. (D) Gene set enrichment analysis of altered proteins, depicting the enriched KEGG pathways ($q < 0.1$) and GO terms ($q < 0.05$) identified by DAVID; Modified Fisher Exact test (EASE Score), $-\log_{10}$ of p-values are displayed.

***Eif4a3* haploinsufficiency in mouse NCCs**

We generated a NCC-specific conditional haploinsufficient mouse model to assess the expression profile of *Eif4a3* deficient developing NCCs. *Wnt1-Cre2* males, which express Cre recombinase in cranial and cardiac neural crest cells, were crossed to *Eif4a3*^{lox/lox} (Lewis et al. 2013) (cre.jax.org). Genotyping of genomic DNA confirmed the heterozygosity and the *Wnt1-Cre*; *Eif4a3*^{lox/+} E9.5 mice presented reduction of 40 % of *Eif4a3* expression in NCCs compared to wild type cells (*Wnt1-Cre*; *Eif4a3*^{+/+}) ($p = 0.0004$; Fig. 4B). In mice, NCC migration begins at E8.0 and the cranial neural crest cells are the first to invade the pharyngeal arches. We focused on E9.5 embryos because this embryonic stage correlates with late migration of cranial neural crest cells and early post-migration patterning/proliferation (Trainor, 2013).

Transcriptome analysis of *Eif4a3* haploinsufficient mNCCs revealed alterations in expression levels of 1,522 transcripts (FDR, $q < 0.1$), corresponding to 9.4 % of the 16,264 coding and non-coding transcripts detected in cranial neural crest cells (Table S10).

Hierarchical clustering of the mDEGs evidenced segregation of control and mutant samples, as shown in the corresponding heatmap (Fig. 4C). An equivalent proportion of transcripts were up or downregulated (713 and 809, respectively) (Fig. 4D; Table S10). Four mDEGs (*Col4a1*, *Col4a2*, *9330182L06Rik* and *Epha5*) were validated by RT-qPCR assays, which showed similar trends to RNA-seq data (Fig. 4E).

For GSEA, we focused on the 1,149 transcripts differentially expressed with high significance ($q < 0.05$). KEGG analysis revealed enrichment of several cancer pathways, p53 and MAPK signaling (Table S11) along with EMC-receptor interaction and focal adhesion, an observation that corroborates human data ($p < 0.1$) (Fig. 4F). GO analysis demonstrated

enrichment mainly of extracellular matrix and plasma membrane components, skeletal system development, organogenesis and morphogenesis and transcription regulation ($p < 0.05$) (Fig. 4F). It is noteworthy that 22.45 % of mDEGs are associated with nervous system development (GO:0007399).

It is known that haploinsufficiency for each of the core EJC components (i.e. *Magoh*, *Eif4a3* and *Rbm8a*) in neural stem cells (NSCs), using the *Emx1*-Cre conditional model to drive Cre expression in the developing brain, leads to aberrant neurogenesis and causes p53-mediated microcephaly in mice (Mao et al. 2016). Since on embryonic day 9.5 the expression of *Wnt1* can also be detected in the developing midbrain and forebrain beyond cranial NCCs (Harrison-Uy and Pleasure 2012; Lewis et al. 2013) and we dissected mouse embryos taking the entire portion upper to the second pharyngeal arch (Fig. 4), there are potentially some NSCs from telencephalon in our cell populations sorted by TdTomato/*Wnt1* expression. In this context, we checked common DEGs between *Wnt1*-Cre and *Emx1*-Cre *Eif4a3* mutants and noted an overlap of 45 transcripts, representing ~3 % of our mDEGs set. Interestingly, GSEA of these transcripts reveals p-53 signaling pathway enrichment, which includes 5 genes upregulated in both models (*Bbc3*, *Gtse1*, *Ccng1*, *Sesn2* and *Zmat3*) (Table S11).

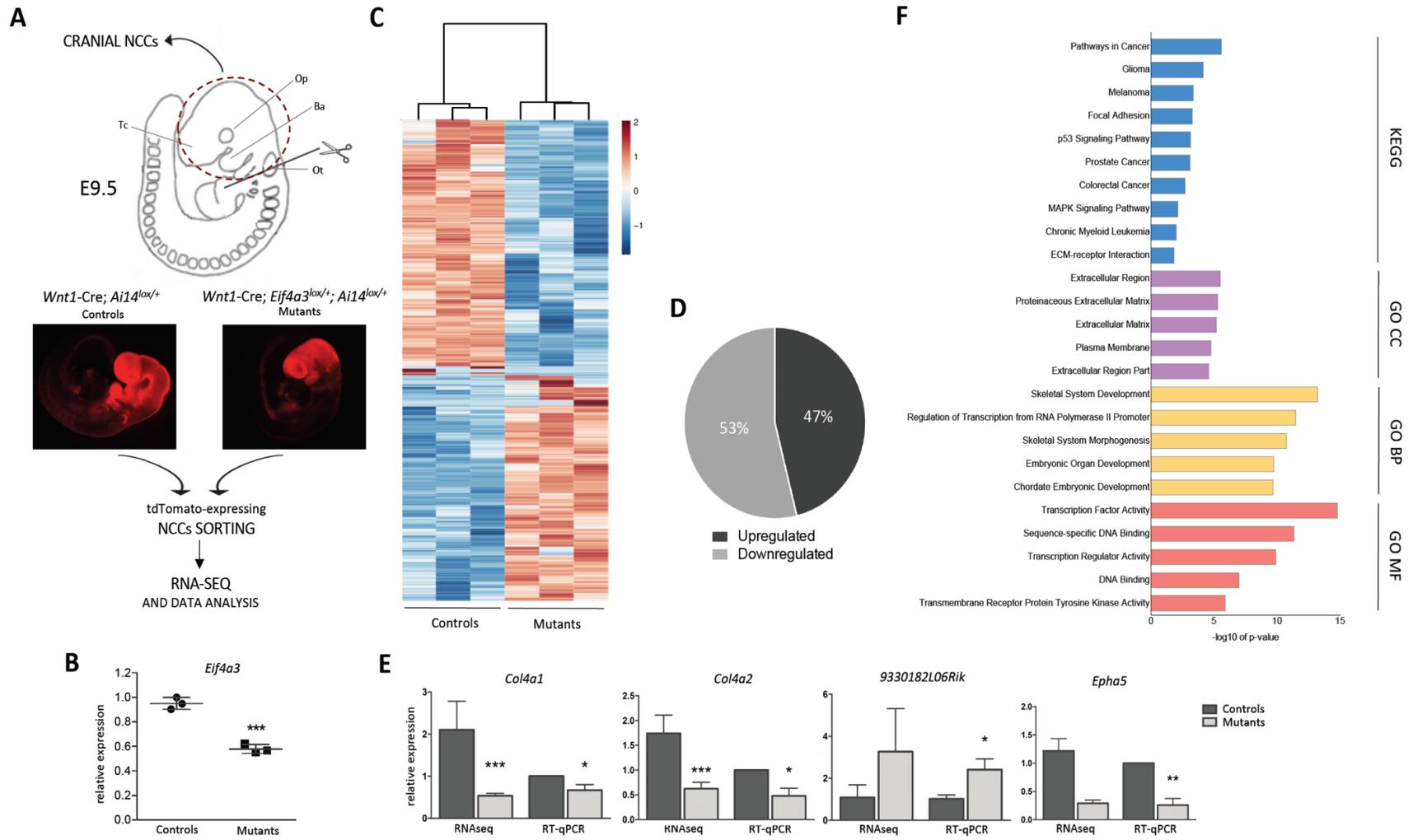


Figure 4. Transcriptome analysis of *EIF4A3* deficient mouse cranial neural crest cells. (A) Diagrammatic overview of RNA-sequencing analysis of mouse NCCs carrying the indicated genotypes. The cranial NCCs were isolated from E9.5 mouse embryos by sorting of tdTomato-expressing cells. (B) Graph depicting expression of *Eif4a3* in the NCCs from control and mutant embryos. Data are represented as mean \pm SEM; Student's *t* test, *** $p < 0.001$. (C) Heatmap showing z-score normalized expression for all 1,149 transcripts identified as significantly different between *Eif4a3* haploinsufficient and control NCCs. (FDR, $q < 0.05$), and the hierarchical clustering of the samples based on all genes using a correlation distance with complete linkage. (D) Pie chart of the distribution of upregulated and downregulated DEGs. (E) RT-qPCR validation of *Col4a1*, *Col4a2*, *9330182L06Rik* and *Epha5* compared to relative RNA-seq values. Control samples were normalized to 1.0 and data are represented as mean \pm SEM; Student's *t* test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (F) Gene set enrichment analysis of mDEGs, depicting the enriched top 10 significant KEGG pathways ($q < 0.1$) and top 5 significant GO terms for each GO category ($q < 0.05$) identified by DAVID; Modified Fisher Exact test (EASE Score), $-\log_{10}$ of *p*-values are displayed.

Additionally, the overlapping expression patterns connecting cellular and animal model were assessed. We noted 10 common DEGs between human and mouse NCCs data: *Nkx1-2*, *Hoxd3*, *Col4a1*, *Col4a2*, *Lfng*, *Pax5*, *Crb2*, *Gsn*, *Mcf2l* and *Chn2*. In addition, 9 alternatively spliced genes (*GFPT2*, *HEATR6*, *AHSA2*, *TRPS1*, *RPS9*, *IGF2BP3*, *NSL1*, *OS9* and *CPNE1*) and 5 altered proteins (*RPS9*, *LIG3*, *ISYNA1*, *VDAC2* and *SRRM2*) in human cells were included in the mouse DEG set. The only element that overlapped amongst three data sets (human AS events and exclusively expressed proteins as well as mouse DEGs) is the *RPS9*, which seems to be a central node in STRING analysis (Fig. 3C).

DISCUSSION

EIF4A3 is essentially required in NCC development, especially for migration and osteochondrogenic differentiation during craniofacial morphogenesis as we demonstrated using patient-derived NCCs and *Eif4a3* haploinsufficient mouse models (Miller et al. 2017). However, the mechanisms by which *EIF4A3* mediates NCC development are not fully known, as well as how mutations in genes that exert basic and ubiquitous functions such as *EIF4A3* cause very specific phenotypes is still poorly understood. In this study, to elucidate causes for the cellular dysfunctions behind RCPS pathogenesis and clarify how *EIF4A3* deficiency mainly compromises craniofacial development, we pinpoint targets and molecular pathways regulated by the *EIF4A3*

at the onset of craniofacial development based on transcriptome and proteome analyses of patient iPSC-derived NCCs and *Eif4a3* haploinsufficient mouse model.

Despite the fact that eIF4A3 is a core component of the EJC complex, a central regulator of mRNA metabolism, only a small fraction of transcripts was altered by decreased levels of *EIF4A3* in human NCCs (0.9 %), although considering the more drastic depletion of *Eif4a3* expression in mouse NCCs, this fraction increases (9.3 %). It is noteworthy that human samples are genetically diverse whereas mice are inbred, which exposes alterations that are normally masked in human samples. Furthermore, despite the majority of human cells are cranial NCCs, they are still representative of a mixed population of NCCs, while mouse cells are particularly cranial NCCs. Therefore, the small number of identified DEGs can be due this heterogenous background of human cells.

We have previously shown that defective NCC migratory capacity is a key cellular mechanism to explain RCPS etiology (Miller et al. 2017). Cell migration is a highly integrated process which is a crucial step in the formation of frontonasal mass and branchial arches during craniofacial development (Kurosaka and Kashina 2008). GSEA of hNCC DEGs revealed enrichment of elements of ECM-receptor interaction and focal adhesion pathways, highlighting up-regulation of a major component of caveolae (*CAV1*), an integrin (*ITGA8*) and collagen molecules (*COL4A1*, *COL4A2*). Mouse data supported these findings. The regulatory mechanisms of migration have been extensively studied and ECM interaction and focal adhesion features (e.g. size, shape and turnover dynamics) have long been associated with cell migration (Ridley et al. 2003; Kim and Wirtz 2013; Perris and Perissinotto 2000). *CAV1* has been linked to the regulation of focal adhesions and integrin-mediated actin remodeling (Goetz et al. 2008; Grande-García et al. 2007). Its overexpression is associated with decreased migratory ability in tumor cells (Zhang et al. 2000). Conversely, *CAV1*-depleted fibroblasts display increased cell migration and invasion (Lin et al. 2015). Moreover, decreased *CAV1* is associated with loss of stiffness-sensing ability, increased migratory ability and upregulation of fibrogenesis-associated *RUNX2* expression, which in turn leads to an ECM overproduction in dermal fibroblasts (Hsu et al. 2018). These alterations in cell-ECM interaction dynamic may explain cell migration disturbance underlying RCPS

pathogenesis. However, future experiments will be needed for clarifying *EIF4A3* roles in cellular adhesion regulation.

Furthermore, we have found that *EIF4A3* regulates alternative splicing events in human NCCs as also evidenced in both *Drosophila* and HeLa cells (Wang et al. 2014; Hayashi et al. 2014). In particular, we demonstrated aberrant splicing of ribosomal components. The splicing events were mainly skipped exons, representing approximately 60 %, consistent with previous reports in which *EIF4A3* depletion in cancer cells (Wang et al. 2014) as well as in mouse NSCs (Mao et al. 2016) resulted in a majority of exon skipping events.

Ribosomal components appear to be especially sensitive to *EIF4A3* deficit, since we identified aberrant splicing as well as alterations at the proteomic level in *EIF4A3* deficient NC cells. In fact, several human genetic conditions are caused by defects in ribosome biogenesis i.e. ribosomopathies (Mills and Green 2017), a diverse group of disorders encompassing heterogeneous clinical manifestations but a surprising tendency toward tissue specificity, such as Treacher Collins and Bowen-Conradi syndromes (Choismel et al. 2007; Sakai and Trainor 2009; Armistead et al. 2009), which interestingly have phenotype overlap with RCPS. Additionally, altered expression of ribosomal components was demonstrated in mouse brains deficient for each of the core EJC components (Mao et al. 2016), reinforcing that EJC may be a strong regulator of protein homeostasis machinery. Therefore, alteration of ribosomal components may contribute to RCPS pathogenesis. Another evidence for this association is a study showing that expression of human eIF4A3 restore the ribosome biogenesis defects in *Fal1p* depleted yeast cells, which is a functional orthologous to *EIF4A3* (Alexandrov et al. 2011). However, how *EIF4A3* influences ribosomal components is still unknown and may be an important question for the future.

RPS9 is a promising ribosomal gene for further investigation, being an element that overlapped amongst four data sets. *RPS9* suffered two AS events in RCPS patients' NCCs and the protein was exclusively expressed in control NCCs, as well as it was differentially expressed between control and *EIF4A3* deficient mouse NCCs. Additionally, two *RPS9* AS events were also identified in *EIF4A3* deficient mouse NSCs (Mao et al. 2016), supporting this finding. However, there are few relevant information about *RPS9* roles during embryonic development in the

literature. Regarding expression during craniofacial development, SysFACE (Systems tool for craniofacial expression-based gene discovery) tracks show that *RPS9* has highest maxilla and frontonasal process specific expression but is also expressed in the developing mandible.

EIF4A3 is super stoichiometric relative to other EJC members in mammalian cells (Singh et al. 2015; Singh et al. 2012). In line with this fact, *EIF4A3* and *RBM8A* mutations result in RCPS syndrome and TAR syndromes (Brunetti-Pierri et al. 2008; Mefford et al. 2008); Albers et al. 2012; Favaro et al. 2014), with overlapping yet diverse phenotypes. TAR is a rare condition characterized mainly by reduced number of platelets and absence of the radius and can involve microcephaly or autism spectrum disorder (ASD) (Albers et al. 2012). Both are autosomal-recessive disorder and compromise craniofacial and limb development, but with distinguishable phenotypes which are likely from distinct pathogenic mechanisms. In this context, *EIF4A3* may potentially plays EJC-independent roles and insufficient *EIF4A3* expression may disturb functions in cells with high *EIF4A3* demand during embryonic development. In addition, ribosomal protein composition is heterogeneous, endowing ribosomes with selectivity for translating specific mRNAs (Shi et al. 2017), which may reflect tissue requirements and explain tissue-specific defects.

Our analyses showed alteration in expression of genes associated with regionalization and specification, as well as transcription factors related to neural development in patient-derived NC cells. *EIF4A3* deficient NCCs revealed upregulation of canonical neurogenesis regulators such as *PAX5* and *SOX1*, which play a role in neural cell fate determination and differentiation, attenuating Notch signaling and Wnt pathways (Tafforeau et al. 2013; Badertscher et al. 2015; Kan et al. 2004). Mouse mutant data corroborate this finding and revealed a similar proportion of genes related to neural development altered (22.45 % mDEGs compared to 23.14 % of hDEGs). It suggests a potential imbalance of important signaling pathways for cell fate decisions and differentiation, which should be confirmed in future analysis.

It has been shown that haploinsufficiency for each of the core EJC components in the developing brain leads to aberrant neurogenesis and causes p53-mediated microcephaly in mice (Mao et al. 2016). We noted an enrichment of p-53 signaling pathway elements in the overlapping DEGs between the published data from *EIF4A3* deficient NSCs (*Emx1-Cre*) (Mao et

al. 2016) and our data from *EIF4A3* deficient mouse NCCs (*Wnt1-Cre*). Since at E9.5 mouse embryos, *Wnt1* can also be detected in the telencephalon beyond cranial NCCs (Harrison-Uy and Pleasure 2012; Lewis et al. 2013) and no significant difference in apoptosis were detected between RCPS patients and control-derived NCCs (Miller et al. 2017), it may reflect a little contamination of mNCC samples with neural stem cells. This finding suggests that p53-associated apoptosis is not a mechanism underlying defective NCC development.

Altogether, this study reveals novel mechanisms to help explain RCPS pathogenesis. We demonstrated altered expression of cell-ECM interaction and focal adhesion components, suggesting a cell adhesion dynamic impairment as a major contributor to reduced migratory capacity in *EIF4A3* deficient NCCs. Furthermore, we showed a dysregulation in expression of genes with central roles in pattern specification and determination of cell fate during embryonic development. Finally, this study associates aberrant splicing of riboproteins and defective NCC development underlying RCPS.

This work focused on neural crest cells and craniofacial development although RCPS is also characterized by limb anomalies. The limb bud has distinct origins than craniofacial structures (Zuniga 2015) and *Wnt1-Cre* is not active in the developing limb, thus not useful in the investigation of limb development. It will be of interest to employ other Cre drivers to investigate additional, relevant RCPS phenotypes in future experiments.

MATERIALS AND METHODS

Ethics statement

The experimental procedures using human samples were approved by the Ethics Committee of the Instituto de Biociências at Universidade de São Paulo (accession number 1.463.852). All subjects donated biological samples after providing signed informed consent. The animal experiments were approved by Duke University IACUC and were performed in agreement with the guidelines of the Division of Laboratory Animal Resources at Duke University School of Medicine.

Human neural crest cell samples

Induced pluripotent stem cells (iPSCs) were generated by episomal (Okita et al. 2013) or retroviral (Takahashi et al. 2007) reprogramming of dermal fibroblast cultures derived from skin punch biopsies or erythroblast cultures derived from peripheral blood collections as previously described (Miller et al. 2017). Three RCPS patient lines (F6099-1 cl1, F8741-1 cl3 and F8417-2 cl3) and three control lines (F7405-1 cl1 and cl2, F8799-1 cl1 and F9048-1 cl2) were genotyped and characterized elsewhere (Miller et al. 2017; Favaro et al. 2014) (Table S1). For this study, an additional iPSC line was generated from a mild RCPS patient (F11167-1 cl1) harboring a 14-repeat expansion in homozygosity (Bertola et al. 2018), characterized and differentiated to NCCs (Fig. S1). iPSC colonies were cultured on Matrigel-coated dishes (BD-Biosciences) using Essential 8™ Medium (Life Technologies) supplemented with 100 µg/ml Normocin (Invivogen).

We generated NCCs from iPSCs following methodology based on Wnt signaling activation and a Smad pathway blockage (Miller et al. 2017; Menendez et al. 2013; Fukuta et al. 2014). Detailed procedures for hNCC derivation were previously described (Miller et al. 2017). hNCCs were cultivated on Matrigel-coated dishes using Essential 6™ Medium (Life Technologies) supplemented with 8 ng/ml bFGF (Life Technologies), 20 mM SB431542 (Sigma-Aldrich), 1 mM CHIR99021 (Sigma-Aldrich), and 100 µg/ml Normocin. Medium was replenished daily.

During the expansion process, both iPSCs and NCCs were split before reaching confluence using Accutase (Life Technologies) at 37 °C for up to 5 min. Cell suspensions were then centrifuged at 200 g for 4 min and re-seeded into new Matrigel-coated dishes using fresh medium supplemented with 5 µM ROCK inhibitor (Sigma-Aldrich).

Mouse husbandry and genetics

Plug dates were defined as embryonic day (E) 0.5 on the morning the plug was identified. The Wnt1-Cre2 (Lewis et al. 2013) and Ai14 transgenic mice used in this study were acquired from JAX (129S4.Cg-*E2f1*^{Tg(Wnt1-cre)2Sor}/J and B6;129S6-*Gt(ROSA)26Sor*^{tm14(CAG-tdTomato)Hze}/J, respectively), and the conditional *Eif4a3* allele was generated as previously described (Mao et al. 2016; Lewis et al. 2013). We obtained mouse embryos with *Eif4a3* haploinsufficient and tdTomato-expressing neural crest cells (Wnt1-Cre; *Eif4a3*^{lox/+}; Ai14^{lox/+}) by crossing Wnt1-Cre2

males with *Eif4a3*^{lox/lox}; Ai14^{lox/+} females. Control embryos (Wnt1-Cre; *Eif4a3*^{+/+}; Ai14^{lox/+}) were generated using Wnt1-Cre2 males crossed with *Eif4a3*^{+/+}; Ai14^{lox/lox} females. Genotyping primers and PCR conditions are listed in Table S2.

Mouse neural crest cell samples

E9.5 mouse embryos were isolated from extra-embryonic tissues in ice-cold DEPC PBS under a dissection microscope. Tails were trimmed and used for genotyping (Table S2). The segments containing mouse cranial neural crest cells (mNCCs) were dissected by placing the embryos laterally, inserting the scissors at the level of the otic vesicle and making one transverse cut down to the second pharyngeal arch (Fig. 4A). The tissue samples were washed in ice-cold DEPC PBS, transferred to a collagenase-dispase solution by using sterile transfer pipets and incubated in a 37 °C water bath for 7 min. After digestion, they were transferred to ice-cold HBSS with 10 % fetal bovine serum (FBS), 5 µg/ml DAPI, and 2 units/ml DNase I (New England Biolabs) then triturated 6-10 times with glass pipettes. Single-cell suspensions were obtained by passing samples through a 30 µm strainer into sorting tubes. Finally, mNCCs were isolated by FACS using the lineage-labeled expression of tdTomato (Fig. 4A).

RNA-sequencing and bioinformatics

Human RNA samples were extracted from NCC lysates with NucleoSpin TriPrep kit (Macherey-Nagel) according to the manufacturer's protocol (n=3 biological replicates for each group). The libraries were generated with TruSeq Stranded mRNA kit. Coding poly-A RNAs were clustered with TruSeq Rapid PE Cluster kit and then sequenced using TruSeq Rapid SBS kit on the HiSeq 2500 system (paired-end, 2x100 cycles) (all provided by Illumina). RNA-seq data were mapped using the GSNAP alignment (Wu and Nacu 2010) and gene counts were performed using the HTSeq tool (<http://htseq.readthedocs.io/>). Subsequent differential gene expression analysis was carried out using DESeq2 (Love et al. 2014) and alternative splicing analysis performed using rMATS (Shen et al. 2014).

Mouse RNA samples were extracted from sorted NC cells using the RNeasy Plus Micro kit (Qiagen) following the manufacturer's instructions (n=3 biological replicates each group). The

libraries were generated and sequenced with the Ultra Low Input RNA kit for sequencing (Clontech Laboratories) on the HiSeq 4000 (Illumina) (single-end, 1x50). RNA-seq data were processed using the TrimGalore tool (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore), which employs Cutadapt (Martin 2011) to trim low-quality bases and Illumina sequencing adapters. Only reads that were 20 nt or longer after trimming were kept for further analysis. Reads were mapped to the GRCh38v73 version of the mouse genome and transcriptome (Kersey et al. 2012) using the STAR RNA-seq alignment tool (Dobin et al. 2013). Reads were kept for subsequent analysis if they mapped to a single genomic location. Gene counts were compiled using the HTSeq tool (<http://htseq.readthedocs.io/>) and differential expression analysis was carried out using the DESeq2 Bioconductor package (Love et al. 2014; Huber et al. 2015).

Only genes that had at least 10 reads in any library were used in the analyses. Heatmaps were generated using z-score transformed normalized expression for genes with an FDR, $q < 0.1$ using the program Expander. RT-qPCR assays were carried out to validate RNA-seq results (primers are listed in Table S3), along with inspection of reads using integrative genomics viewer (IGV) software. Gene set enrichment analysis (GSEA) was carried out using the Database for Annotation, Visualization and Integrated Discovery (DAVID 6.7) (<http://david.ncifcrf.gov>), to identify Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways associated with altered gene expression.

Mass spectrometry and data processing

For extraction of proteins, hNCC pellets were obtained by rinsing cells with PBS followed by dissociation with Accutase for up to 5 min at room temperature (RT) and centrifugation at 200 g for 4 min (n=3 biological replicates each group). Cell samples were lysed using a buffer containing 100 mM Tris-HCl pH 7.5, 5 mM EDTA, 1 mM PMSF, 8 M Urea, 2 M Thiourea, and protease inhibitor cocktail. Homogenates were sonicated for 20 sec and centrifuged for 10,000 g for 10 min at 4 °C. The supernatants were collected and protein content was determined by Bradford assay (Bradford 1976). Protein samples were digested using trypsin for 16 h at 37 °C and then desalted using Stage Tips (Rappsilber et al. 2007).

Raw mass spectrometry data were processed with MaxQuant software (<http://www.maxquant.org/>). The label-free quantification (LFQ) values were imported into Perseus for subsequent analyses according to the LNBio/CNPEM tutorial (http://lnbio.cnpem.br/wp-content/uploads/2012/11/Tutorial_MaxQuant-protein-identification_release-v1_28052015.pdf). The differentially expressed proteins (DEPs) were further subjected to GSEA by DAVID 6.7 (<http://david.ncifcrf.gov>) and protein-protein interaction analysis using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (www.string-db.org). Statistical significance was tested by two-sample t-test using Perseus software.

Real-time quantitative PCR (RT-qPCR)

Total RNA was obtained from hNCCs using the NucleoSpin RNA II kit (Macherey-Nagel) and 1 µg of RNA was converted into cDNA using Superscript IV (Life Technologies) with oligo-dT primers according to the manufacturer's recommendations. RT-qPCR reactions were performed with Fast SYBR Green PCR Master Mix (Life Technologies) in the QuantStudio 5 System (Life Technologies) under standard cycling protocol. Total RNA from mNCCs was extracted using RNeasy Plus Micro kit (Qiagen) and 500 ng of RNA was converted into cDNA using iScript cDNA Synthesis kit (Bio-Rad) following the manufacturer's recommendations. RT-qPCR reactions were performed with iTaq Universal SYBR Green Supermix (Bio-Rad) in the Step One Plus Real-Time PCR System (Applied Biosystems) under standard cycling protocol.

Primer pairs were designed with Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>) and are listed in Table S3. 100–400 nM of each primer was used for RT-qPCR reactions and their amplification efficiencies (%) were determined by log₁₀ serial cDNA dilutions plotted against Ct values. NormFinder (Andersen et al. 2004) was used to determine the most stable endogenous control for human cells (among *HMBS*, *HPRT1* and *TBP*). *ACTB* was used as an endogenous control for mouse gene expression normalization. The relative expression values were determined by dividing $E^{-\Delta C_t}$ of target genes by normalization factors ($E^{-\Delta C_t}$ of endogenous control) (Pfaffl 2001).

Statistical comparisons were performed using the GraphPad Prism software and the level of statistical significance was set at $p < 0.05$. All relative expression values were log-transformed for analysis and graphed in linear scale.

Acknowledgements

We thank members of the Silver and Passos-Bueno labs for helpful discussions and Laís Giraldi of the Winck lab for technical support in protein samples preparation. This study received funding from FAPESP/CEPID (2013/08028-1) and CNPq (305405/2011-5) to MRPB, NIH (R01NS0830897) to DLS, and CAPES research fellowships (1470032 and 88881.187400/2018-01) to CMM. We also thank the Company of Biologists for the travelling fellowship provided to CMM.

REFERENCES

- Albers, C.A., Paul, D.S., Schulze, H., et al. 2012. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nature Genetics* 44(4), pp. 435–9, S1.
- Alexandrov, A., Colognori, D. and Steitz, J.A. 2011. Human eIF4AIII interacts with an eIF4G-like partner, NOM1, revealing an evolutionarily conserved function outside the exon junction complex. *Genes & Development* 25(10), pp. 1078–1090.
- Andersen, C.L., Jensen, J.L. and Ørntoft, T.F. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* 64(15), pp. 5245–5250.
- Armistead, J., Khatkar, S., Meyer, B., et al. 2009. Mutation of a gene essential for ribosome biogenesis, EMG1, causes Bowen-Conradi syndrome. *American Journal of Human Genetics* 84(6), pp. 728–739.
- Badertscher, L., Wild, T., Montellese, C., et al. 2015. Genome-wide RNAi Screening Identifies Protein Modules Required for 40S Subunit Synthesis in Human Cells. *Cell reports* 13(12), pp. 2879–2891.
- Bertola, D.R., Hsia, G., Alvizi, L., et al. 2018. Richieri-Costa-Pereira syndrome: Expanding its phenotypic and genotypic spectrum. *Clinical Genetics* 93(4), pp. 800–811.
- Bolande, R. 1974. The neurocristopathies: A unifying concept of disease arising in neural crest maldevelopment. *Human Pathology* 5(4), pp. 409–429.
- Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72, pp. 248–254.
- Brunetti-Pierri, N., Berg, J.S., Scaglia, F., et al. 2008. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nature Genetics* 40(12), pp. 1466–1471.
- Choesmel, V., Bacqueville, D., Rouquette, J., et al. 2007. Impaired ribosome biogenesis in Diamond-Blackfan anemia. *Blood* 109(3), pp. 1275–1283.
- 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. *American Journal of Medical Genetics. Part A* 155A(2), pp. 270–279.
- Dobin, A., Davis, C.A., Schlesinger, F., et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1), pp. 15–21.
- Favaro, F.P., Alvizi, L., Zechi-Ceide, R.M., et al. 2014. A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. *American Journal of Human Genetics* 94(1), pp. 120–128.
- Favaro, F.P., Zechi-Ceide, R.M., Alvarez, C.W., et al. 2011. Richieri-Costa-Pereira syndrome: a unique acrofacial dysostosis type. An overview of the Brazilian cases. *American Journal of Medical Genetics. Part A* 155A(2), pp. 322–331.
- Fukuta, M., Nakai, Y., Kirino, K., et al. 2014. Derivation of mesenchymal stromal cells from pluripotent stem cells through a neural crest lineage using small molecule compounds with defined media. *Plos One* 9(12), p. e112291.
- Goetz, J.G., Joshi, B., Lajoie, P., et al. 2008. Concerted regulation of focal adhesion dynamics by galectin-3 and tyrosine-phosphorylated caveolin-1. *The Journal of Cell Biology* 180(6), pp. 1261–1275.
- Grande-García, A., Echarri, A., de Rooij, J., et al. 2007. Caveolin-1 regulates cell polarization and directional migration through Src kinase and Rho GTPases. *The Journal of Cell Biology* 177(4), pp. 683–694.
- Harrison-Uy, S.J. and Pleasure, S.J. 2012. Wnt signaling and forebrain development. *Cold Spring Harbor Perspectives in Biology* 4(7), p. a008094.
- Hayashi, R., Handler, D., Ish-Horowicz, D. and Brennecke, J. 2014. The exon junction complex is required

for definition and excision of neighboring introns in *Drosophila*. *Genes & Development* 28(16), pp. 1772–1785.

Hsia, G.S.P., Musso, C.M., Alvizi, L., et al. 2018. Complexity of the 5' untranslated region of EIF4A3, a critical factor for craniofacial and neural development. *Frontiers in genetics* 9, p. 149.

Hsu, C.-K., Lin, H.-H., Harn, H.I., et al. 2018. Caveolin-1 Controls Hyperresponsiveness to Mechanical Stimuli and Fibrogenesis-Associated RUNX2 Activation in Keloid Fibroblasts. *The Journal of Investigative Dermatology* 138(1), pp. 208–218.

Huang, M., Miller, M.L., McHenry, L.K., et al. 2016. Generating trunk neural crest from human pluripotent stem cells. *Scientific reports* 6, p. 19727.

Huber, W., Carey, V.J., Gentleman, R., et al. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* 12(2), pp. 115–121.

Kan, L., Israsena, N., Zhang, Z., et al. 2004. Sox1 acts through multiple independent pathways to promote neurogenesis. *Developmental Biology* 269(2), pp. 580–594.

Kersey, P.J., Staines, D.M., Lawson, D., et al. 2012. Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species. *Nucleic Acids Research* 40(Database issue), pp. D91-7.

Kim, D.-H. and Wirtz, D. 2013. Focal adhesion size uniquely predicts cell migration. *The FASEB Journal* 27(4), pp. 1351–1361.

Kurosaka, S. and Kashina, A. 2008. Cell biology of embryonic migration. *Birth Defects Research. Part C, Embryo Today: Reviews* 84(2), pp. 102–122.

Le Hir, H., Saulière, J. and Wang, Z. 2016. The exon junction complex as a node of post-transcriptional networks. *Nature Reviews. Molecular Cell Biology* 17(1), pp. 41–54.

Lewis, A.E., Vasudevan, H.N., O'Neill, A.K., Soriano, P. and Bush, J.O. 2013. The widely used Wnt1-Cre transgene causes developmental phenotypes by ectopic activation of Wnt signaling. *Developmental Biology* 379(2), pp. 229–234.

Lin, H.-H., Lin, H.-K., Lin, I.-H., et al. 2015. Mechanical phenotype of cancer cells: cell softening and loss of stiffness sensing. *Oncotarget* 6(25), pp. 20946–20958.

Love, M.I., Huber, W. and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15(12), p. 550.

Mao, H., McMahan, J.J., Tsai, Y.-H., Wang, Z. and Silver, D.L. 2016. Haploinsufficiency for Core Exon Junction Complex Components Disrupts Embryonic Neurogenesis and Causes p53-Mediated Microcephaly. *PLoS Genetics* 12(9), p. e1006282.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1), p. 10.

Mefford, H.C., Sharp, A.J., Baker, C., et al. 2008. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *The New England Journal of Medicine* 359(16), pp. 1685–1699.

Menendez, L., Kulik, M.J., Page, A.T., et al. 2013. Directed differentiation of human pluripotent cells to neural crest stem cells. *Nature Protocols* 8(1), pp. 203–212.

Miller, E.E., Kobayashi, G.S., Musso, C.M., et al. 2017. EIF4A3 deficient human iPSCs and mouse models demonstrate neural crest defects that underlie Richieri-Costa-Pereira syndrome. *Human Molecular Genetics* 26(12), pp. 2177–2191.

Mills, E.W. and Green, R. 2017. Ribosomopathies: There's strength in numbers. *Science* 358(6363).

Nguyen, L.S., Kim, H.-G., Rosenfeld, J.A., et al. 2013. Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders. *Human Molecular Genetics* 22(9), pp. 1816–1825.

Okita, K., Yamakawa, T., Matsumura, Y., et al. 2013. An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells* 31(3), pp. 458–466.

Palacios, I.M., Gatfield, D., St Johnston, D. and Izaurralde, E. 2004. An eIF4AIII-containing complex required

for mRNA localization and nonsense-mediated mRNA decay. *Nature* 427(6976), pp. 753–757.

Perris, R. and Perissinotto, D. 2000. Role of the extracellular matrix during neural crest cell migration. *Mechanisms of Development* 95(1–2), pp. 3–21.

Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* 29(9), p. e45.

Rappsilber, J., Mann, M. and Ishihama, Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* 2(8), pp. 1896–1906.

Ridley, A.J., Schwartz, M.A., Burridge, K., et al. 2003. Cell migration: integrating signals from front to back. *Science* 302(5651), pp. 1704–1709.

Sakai, D. and Trainor, P.A. 2009. Treacher Collins syndrome: unmasking the role of Tcof1/treacle. *The International Journal of Biochemistry & Cell Biology* 41(6), pp. 1229–1232.

Shen, S., Park, J.W., Lu, Z., et al. 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America* 111(51), pp. E5593–601.

Shi, Z., Fujii, K., Kovary, K.M., et al. 2017. Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Molecular Cell* 67(1), p. 71–83.e7.

Singh, G., Kucukural, A., Cenik, C., et al. 2012. The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* 151(4), pp. 750–764.

Singh, G., Pratt, G., Yeo, G.W. and Moore, M.J. 2015. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annual Review of Biochemistry* 84, pp. 325–354.

Suga, M., Hayashi, Y. and Furue, M.K. 2017. In vitro models of cranial neural crest development toward toxicity tests: frog, mouse, and human. *Oral Diseases* 23(5), pp. 559–565.

Tafforeau, L., Zorbas, C., Langhendries, J.-L., et al. 2013. The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. *Molecular Cell* 51(4), pp. 539–551.

Takahashi, K., Tanabe, K., Ohnuki, M., et al. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131(5), pp. 861–872.

Trainor, P.A. 2013. *Neural Crest Cells: Evolution, Development and Disease*. ed. 1.

Vega-Lopez, G.A., Cerrizuela, S., Tribulo, C. and Aybar, M.J. 2018. Neurocristopathies: New insights 150 years after the neural crest discovery. *Developmental Biology* 444 Suppl 1, pp. S110–S143.

Wang, Z., Murigneux, V. and Le Hir, H. 2014. Transcriptome-wide modulation of splicing by the exon junction complex. *Genome Biology* 15(12), p. 551.

Wu, T.D. and Nacu, S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26(7), pp. 873–881.

Zhang, W., Razani, B., Altschuler, Y., et al. 2000. Caveolin-1 inhibits epidermal growth factor-stimulated lamellipod extension and cell migration in metastatic mammary adenocarcinoma cells (MTLn3). Transformation suppressor effects of adenovirus-mediated gene delivery of caveolin-1. *The Journal of Biological Chemistry* 275(27), pp. 20717–20725.

Zuniga, A. 2015. Next generation limb development and evolution: old questions, new perspectives. *Development* 142(22), pp. 3810–3820.

SUPPLEMENTARY MATERIAL

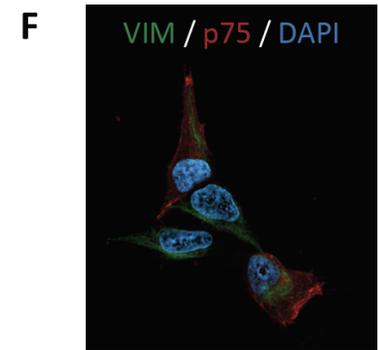
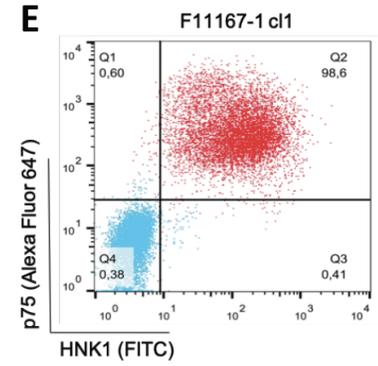
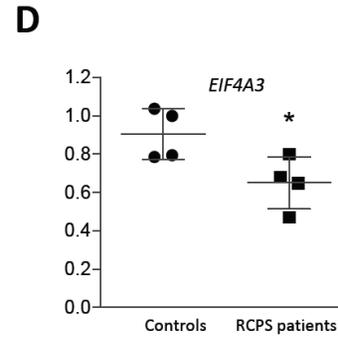
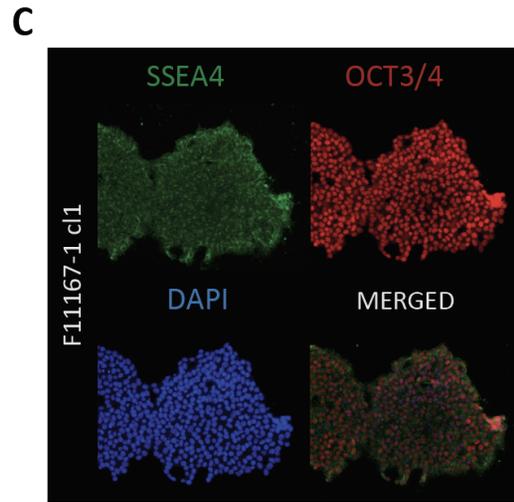
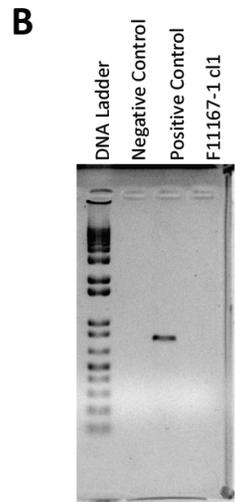
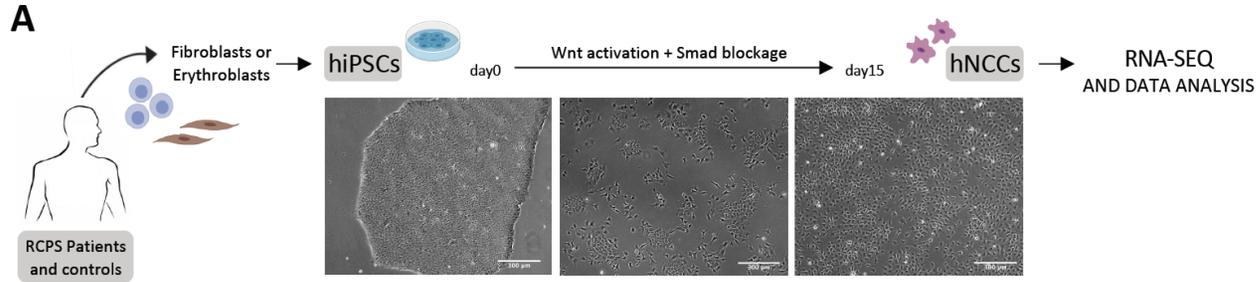


Figure S1. (A) Diagrammatic overview of RNA sequencing of human NCCs. Patient- and control-derived fibroblasts or erythroblasts were reprogrammed into iPSCs, which were subsequently differentiated toward NCCs. **(B-D)** Characterization of the novel iPSC lineage F11167-1 cl1. **(B)** End-point PCR for vector backbone gene *OriP*; DNA from an iPSC line in which genomic integration had been detected was used as positive control. **(C)** Immunofluorescence staining of pluripotency markers SSEA4 (green) and OCT3/4 (red); DAPI nuclear staining is in blue (10X magnification). **(D)** Graph depicting expression of *EIF4A3* in the NCCs from control and RCPS patients. RNAseq data are represented as mean \pm SEM; Student's *t* test, * $p < 0.05$. **(E-F)** Characterization of the novel NCC lineage F11167-1 cl1. **(E)** Biparametric flow cytometry dot plots for HNK1/FITC and p75/Alexa Fluor 647 expression. Value in upper right quadrant represent p75+/HNK1+ events. **(F)** Immunofluorescence staining of NCC markers Vimentin (green) and p75 (red); DAPI nuclear staining is in blue (10X magnification).

Table S1. Human cell samples

Subject	ID / clone	Cell type	Phenotype	Age*	Gender	<i>EIF4A3</i> genotype	Reprogramming method
F11167-1	F11167-1 cl1	Erythroblasts	RCPS patient	16	Female	14 repeats / 14 repeats	Episomal
F6099-1	F6099-1 cl1	Fibroblasts	RCPS patient	10	Male	14 repeats / c.809A>G	Episomal
F7405-1	F7405-1 cl1	Fibroblasts	Control	27	Male	7 repeats / 6 repeats	Retroviral
	F7405-1 cl2						
F8417-1	F8417-1 cl3	Fibroblasts	RCPS patient	11	Female	16 repeats / 16 repeats	Episomal
F8417-2	F8417-2 cl3	Fibroblasts	RCPS patient	9	Male	16 repeats / 16 repeats	Episomal
F8799-1	F8799-1 cl1	Erythroblasts	Control	20	Female	7 repeats / 6 repeats	Episomal
F9048-1	F9048-1 cl2	Fibroblasts	Control	20	Male	6 repeats / 6 repeats	Episomal

*age upon sample collection

Table S2. Mouse genotyping primers and conditions

Target	Forward primer (5'-3')	Reverse primer (5'-3')	Conditions
Eif4a3 ^{lox}	CTTGCAGTTGTCTTTCTGCGG	CACACATGGCGATCCGCTCG	94 °C x 30s (1x); 94 °C x 30s, 62 °C x 30s, 65 °C x 30s (35x); 65 °C 10min (1x)
Cre	GCATTACCGGTCGATGCAACGAGTGATGAG	GAGTGAACGAACCTGGTCGAAATCAGTGCG	94 °C x 3min (1x); 94 °C x 30s, 61 °C x 30s, 72 °C x 30s (36x); 72 °C 10min (1x)
FABP (internal control)	TGGACAGGACTGGACCTCTGCTTTCCTAGA	TAGAGCTTTGCCACATCACAGGTCATTCAG	94 °C x 3min (1x); 94 °C x 30s, 61 °C x 30s, 72 °C x 30s (36x); 72 °C 10min (1x)
Ai14 WT	AAGGGAGCTGCAGTGGAGTA	CCGAAAATCTGTGGGAAGTC	96 °C x 3min (1x); 96 °C x 30s, 58 °C x 30 s, 72 °C x 1min (35x); 72 °C 7min (1x)
Ai14 ^{lox}	GGCATTAAAGCAGCGTATCC	CTGTTCTGTACGGCATGG	96 °C x 3min (1x); 96 °C x 30s, 58 °C x 30 s, 72 °C x 1min (35x); 72 °C 7min (1x)

Table S3. RT-qPCR primer sequences

Human target	Forward primer (5'–3')	Reverse primer (5'–3')
<i>CAV1</i>	GAGAAGCAAGTGTACGACGC	CCGTCAAAACTGTGTGTCCC
<i>CCDC80</i>	ACTGGGCCTATTCACAGCAG	GTTATGTGGCGCAGACCAAAAT
<i>CELF2</i>	ACGGAGCTTTGGATCACTCA	AACAACAACCTTTACTCTGCGG
<i>COL4A1</i>	GGGCCAAGACGAAGACAT	CAACACCCTTGCCGTTGTCTG
<i>COL4A2</i>	TCCCTAGCAAAAGGTGCCAG	GTCTCCGATGAAGCCCTTGG
<i>EIF4A3</i>	GGAGATCAGGTCGATACGGC	GATCAGCAACGTTTCATCGGC
<i>FANCA</i>	CCATGACCAGCGTGAACATC	CTGCAGACTGGCAGAGAGC
	GGTAAGTTAACGTGAACATCTTCCT	CTGCAGACTGGCAGAGAGC
<i>HMBS</i>	AGCTTGCTCGCATAACAGACG	AGCTCCTTGGTAAACAGGCTT
<i>HPRT1</i>	CCTGGCGTTCGTGATTAGTGAT	AGACGTTTCAGTCTGTCCATAA
<i>LNFG</i>	TCGCTGTCAAGACCACCAAA	CCTCATCTTCCCCGTCAGTG
<i>NANOG</i>	TGGACACTGGCTGAATCCTTC	CGTTGATTAGGCTCCAACCAT
<i>OCT3/4</i>	GTGGTCAGCCAACTCGTCA	CCAAAACCCTGGCACAACCT
<i>OriP</i>	TTCCACGAGGGTAGTGAACC	TCGGGGGTGTTAGAGACAAC
<i>PAX3</i>	AAGCCCAAGCAGGTGACAAC	CTCGGATTTCCAGCTGAAC
<i>POSTN</i>	GACCGTGTGCTTACACAAATTG	AAGTGACCGTCTCTTCCAAGG
<i>RPL9</i>	CTCTGAAAGAGACACTTACGCT	GGTGTTACACTGGTAAGCAGAT
	TAACGGAAGCCCCAGTGTA	TGACAAATGGTGGGGTAACAG
<i>RPS9</i>	GCAGCTGGACAGGGGTC	AGCGCAGAGAGAAGTCGATG
	GTTCTGAAGGGGTCCGCAA	GGAGAGCGCAGAGAGAAGTC
<i>SAT2</i>	CTGAGAGCAGATGGCTTTGGA	AATCCCTTGACCCCAGTAGC
	ATGAGGGATGGGTTCCCCAG	GCTCTTACTGCCGAGATTGGAG
<i>SOX1</i>	TGACACCAGACTTGGGTTTTAG	CAAAAGTGGGCTTCGCCTCT
<i>SOX10</i>	CCTCACAGATCGCCTACACC	CATATAGGAGAAGGCCGAGTAGA
<i>TBP</i>	GTGACCCAGCATCACTGTTTC	GCAAACCAGAAACCCTTGCG
<i>ZIC1</i>	AAGGTCCACGAATCCTCCTC	TTGTGGTCGGGTTGTCTG

SUPPLEMENTAL TABLES (excel files/.xlsx)

Table S4. Transcriptome analysis of *EIF4A3* deficient human iPSC-derived NCCs

Table S5. Gene set enrichment analysis of DEGs between RCPS patient- and control-derived NCCs

S6 Table. Splicing analysis of *EIF4A3* deficient human iPSC-derived NCCs

S7 Table. Gene set enrichment analysis of alternative splicing events between RCPS patient- and control-derived NCCs; KEGG pathways and GO terms

S8 Table. Proteome analysis of *EIF4A3* deficient iPSC-derived NCCs

S9 Table. Gene set enrichment analysis of DEPs and EEPs between RCPS patient- and control-derived NCCs; KEGG pathways and GO terms

S10 Table. Transcriptome analysis of *eIF4A3* deficient mouse NCCs

S11 Table. Gene set enrichment analysis of DEGs between control and mutant NCCs; KEGG pathways and GO terms

CHAPTER V

Recapitulation of Neural Crest Specification and EMT via Induction from Neural Plate Border-Like Cells

Gerson S. Kobayashi^{1†*}, Camila M. Musso^{1†}, Giovanna Pontillo-Guimarães¹, Gabriella S. P. Hsia¹, Danielle P. Moreira¹, Luiz C. de Caires-Júnior¹, Ernesto Goulart¹, Maria Rita Passos-Bueno¹

¹Centro de Estudos do Genoma Humano e Células Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

†Co-first authors

***Correspondent author**

Keywords: iPSCs, neuroepithelial cells, neural crest cells, epithelial-mesenchymal transition, disease modeling, cell fate, early neural crest development

ABSTRACT

Neural crest cells (NCCs) contribute to the morphogenesis of several cell types and tissues during vertebrate embryonic development. NCCs are specified from neuroepithelial cells at the neural plate border (NPB), undergoing a series of tightly regulated cellular events, such as epithelial-mesenchymal transition (EMT), migration, and differentiation. Disturbances in these processes are associated with several human cancers and congenital malformations, highlighting the importance of establishing experimental frameworks that enable recapitulation of human embryonic NCC development. Here, we report that the E6 minimal neural induction method transiently produces PAX6^{low}/SOX1⁻/Musashi1⁺ neural plate border-like cells (NBCs), which can be converted into NCC populations with near 100% efficiency. We also show that NBC-to-NCC induction elicits a shift of cellular states showing sequential molecular and cellular hallmarks of NCC specification and EMT, including downregulation of NPB factors and upregulation of NCC specifiers and EMT inducers coupled with EMT-associated E/N-cadherin modulation, recapitulating observations so far largely restricted to animal models. Recapitulation of these early NCC developmental steps *in vitro* will be useful in future research focusing on human development and disease.

Highlights

- PAX6^{low}/SOX1⁻/Musashi1⁺ NBCs emerge during the E6 neural induction method
- NBCs can be efficiently redirected towards the NCC fate
- NBC-to-NCC induction recapitulates sequential steps involved in NCC specification and EMT

In Brief

In this work, Kobayashi et al. show that neural conversion of hiPSCs in minimal conditions produces PAX6^{low}/SOX1⁻/Musashi1⁺ neural plate border-like cells that can be efficiently converted into NCCs. This enables observation and recapitulation of sequential molecular and cellular events involved in NCC specification and EMT so far mostly explored in animal models, providing an *in vitro* platform for investigating human NCC specification and associated disorders.

INTRODUCTION

Neural crest cells (NCCs) are a multipotent cell population that plays pivotal roles in vertebrate embryonic development (Zhang et al., 2014). In late-blastula stages, intermediate BMP levels specify epiblast cells residing between the prospective neural plate and non-neural ectoderm into neural plate border (NPB) cells (Pla and Monsoro-Burq, 2018). Within the NPB territory, inductive signals that include FGF, BMP and WNT pathway molecules initiate NCC specification and epithelial-mesenchymal transition (EMT), a series of molecular events that orchestrates transitioning from an epithelial to a mesenchymal cell phenotype. These events depend upon tightly controlled molecular programs within a gene regulatory network involving sequential activation and/or repression of NPB and NCC specifiers, cadherins and other adhesion molecules (Theveneau and Mayor, 2012). After EMT, NCCs migrate along stereotypical routes through the head and trunk regions of the embryo, contributing to the morphogenesis of a wide range of cells and tissues, such as craniofacial cartilage and bones, melanocytes, sensory and autonomic neurons, Schwann cells, and smooth muscle (Le Douarin and Dupin, 2003; Trainor et al., 2003).

Several human congenital disorders are caused by neural crest-related disturbances. These disorders (termed 'neurocristopathies') have been attributed to genetic mutations leading to neuroepithelial cell death in the NPB prior to NCC migration (e.g. Treacher Collins syndrome (Dixon et al., 2006; Sakai et al., 2016), and Nager syndrome (Devotta et al., 2016) and reduced migratory behavior (e.g. familial dysautonomia (Lee et al., 2009), while others arise due to defects in genes closely associated with NCC EMT, including *SNAI2* (Waardenburg syndrome type 2D/piebaldism (Cordero et al., 2011), *CHD7* (CHARGE syndrome (Bajpai et al., 2010), and *ZEB2* (Mowat-Wilson syndrome (Van de Putte et al., 2007; Wakamatsu et al., 2001). These observations place NCCs under focus of intense research aimed at understanding and treating these conditions; however, access to human embryonic tissues to study NCC biology is limited, and knowledge on the pathophysiological mechanisms behind neurocristopathies has been largely derived from experimentation and observation in animal embryos, which may not completely recapitulate human embryonic development (de Bakker et al., 2016; Thyagarajan et al., 2003).

Human induced pluripotent stem cells (hiPSCs) can be propagated indefinitely and

generate derivatives of all three embryonic germ layers, including NCCs. hiPSCs are a powerful tool to investigate human embryonic development and pathophysiological mechanisms, and also to develop therapeutic strategies via drug screening or tissue engineering and cell therapy (Avior et al., 2016). The first methods for generating human NCCs majorly relied on BMP pathway blockade to produce neuroepithelial intermediates from which NCCs arose, in accordance with *in vivo* observations (Curchoe et al., 2010; Jiang et al., 2009; Lee et al., 2007, 2010; Pomp et al., 2005). NCC generation was further refined mainly through modulation of BMP signaling activity combined or not with WNT pathway activation (Fukuta et al., 2014; Huang et al., 2016b; Kreitzer et al., 2013; Lee et al., 2009; Leung et al., 2016; Liu et al., 2012; Menendez et al., 2011, 2013; Mica et al., 2013; Miller et al., 2017; Noisa et al., 2014; Tchieu et al., 2017; Thier et al., 2019). Importantly, many of these strategies either lack a NPB stage or are unable to fully convert neuroepithelial cell populations into NCCs, potentially hindering investigation of the various stages of neural crest specification.

Here, we report a simple approach to recapitulate developmental events involved in NCC specification from NPB cells. This process recapitulates gene expression dynamics and hallmarks associated with embryonic NPB cells transitioning from a neuroepithelial to a neural crest mesenchymal state, allowing for time-series assessment of neural crest specification and EMT. This strategy can be used to model human NCC development and associated disorders *in vitro*, particularly those associated with these early events of neural crest development.

RESULTS

E6 neural induction transiently produces PAX6^{low}/SOX1⁻/Musashi1⁺ cells

In order to evaluate early steps of NCC induction, we first converted hiPSCs into neuroepithelial cell cultures, from which NCCs would be induced. This step was performed under with the “E6 method”, which enables generation of PAX6-positive neuroepithelial cells in minimal conditions without BMP inhibitors or other exogenous factors, in 6 days (Lippmann et al., 2014) (Fig. 1A).

Three hiPSC lines derived from healthy donors (F7405-1c1, F8799-1c1 and F9048-1c2) were grown in minimal conditions (E8 medium) then subjected to the E6 neural induction, and expression of selected markers was assessed to follow alterations in cellular differentiation states. After 4 days of differentiation, epiblast/pluripotency markers *OCT3/4* and *NANOG* were transcriptionally downregulated, while neural markers *SOX2* and *PAX6* showed elevated expression at day 4, persisting until day 6 (Fig. 1B). At the end of the neuroepithelial induction (day 6), *PAX6* was expressed in cell nuclei, while Musashi-1, a stem cell marker that regulates cell fate transitions (MacNicol et al., 2015), was expressed in the cytoplasm, and N-Cadherin (*CDH2*) was localized to plasma membranes at cell-cell contact sites (Fig. 1C). Flow cytometry analysis confirmed that all cells attained neural identity at day 6, as evidenced by positive staining for *PAX6* (99.80 ± 0.07 %) and Musashi-1 (99.70 ± 0.13 %). *SOX1*, a neural marker co-expressed with *PAX6* in human neural progenitor cells (Li et al., 2017) was detected in 46.67 ± 8.02 % of cells (Fig. 1D).

Considering the peak expression for *PAX6*, we also assessed lineage commitment at day 4 of differentiation. In this timepoint, cells were negative for *SOX1* expression (1.61 ± 1.36 %), while nearly all cells already expressed Musashi-1 (98.6 ± 0.6 %) and *PAX6* (98.6 ± 0.8 %) (Fig.1D). Moreover, despite staining positive, *PAX6* protein was expressed at lower intensity (*PAX6*^{low}) in cells at day 4 compared to day 6 (*PAX6*^{high}) (Fig. 1D and 1D'). Since activation of *SOX1* is observed only at day 6, together, results suggest that *PAX6*^{low}/*SOX1*⁻/*Musashi1*⁺ cells at day 4 represent a transient stem cell population still not completely committed towards the neural fate.

***PAX6*^{low}/*SOX1*⁻/*Musashi1*⁺ cells show a NPB-like transcriptional profile**

PAX6^{low}/*SOX1*⁻/*Musashi1*⁺ (day 4) neuroepithelial cells showed elevated expression of *PAX6* and *SOX2*, which have also been reported in embryonic NPB cells that contribute to NCC formation (Roellig et al., 2017; Thier et al., 2019). Thus, to further investigate the identity of day-4 cells, we assessed mRNA expression patterns for additional NPB factors during neural induction. We found that key NPB specifiers *DLX5*, *PAX3*, *ZIC1* and *MSX2* were markedly upregulated in comparison to hiPSCs (day 0), rising to their highest expression values at day 4 of differentiation; *TFAP2A* followed the same behavior, but starting at day 2 since it was not detected in hiPSCs (Fig. 1E). Cadherin interplay is important for definition of

the neural crest territory in the NPB (Dady and Duband, 2017), so we also examined the expression profile of cadherins (*CDH1*, *CDH2*, *CDH6*, *CDH7* and *CDH11*) known to modulate NCC development across animal models (Taneyhill and Schiffmacher, 2017). Throughout the 6 days of differentiation, mRNA expression of *CDH1* (E-cadherin) and *CDH2* (N-cadherin) is maintained and upregulated at day 4, respectively. *CDH6* and *CDH7*, which are expressed in the NPB region in chick and rat early embryos (Dady and Duband, 2017; Takahashi and Osumi, 2008), are upregulated and reach peak expression at day 4. Finally, *CDH11*, expressed in chick neural plate/tube (Bell et al., 2004; Darnell et al., 2007), is also upregulated at day 4, and reaches peak expression at day 6 (Fig. 1F). Given the observed expression patterns for NPB genes, these results indicate that day-4 neuroepithelial cells may represent or be enriched with NPB cells, so we renamed this differentiation timepoint to 'Neural plate Border-like Cells' (NBCs), which were selected to undergo NCC induction.

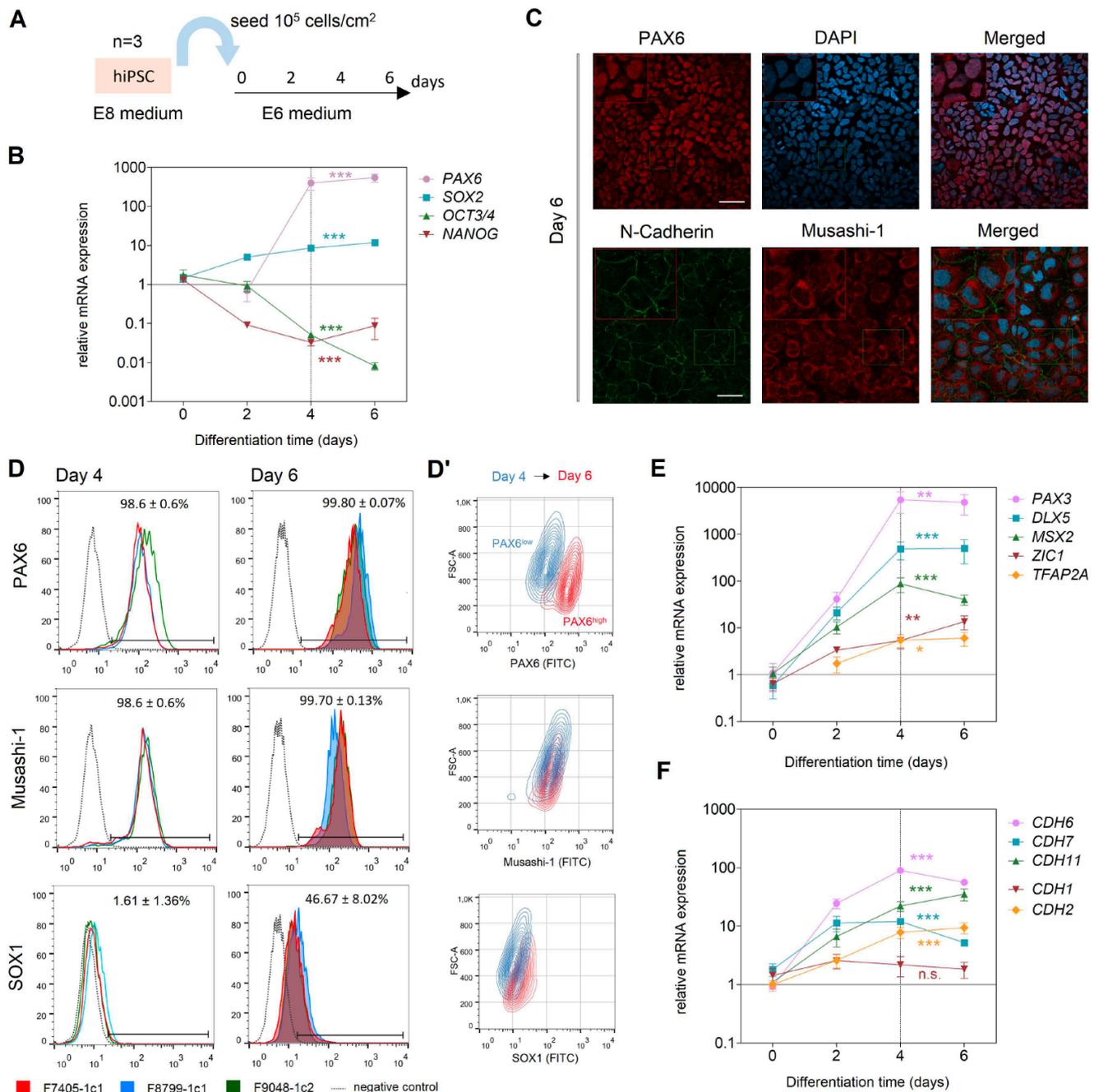


Figure 1. Neural induction with the E6 method elicits a NPB-like (NBC) transcriptional profile to differentiating cells. **(A)** Schematic representation of the E6 differentiation method and time points assessed to investigate cellular identity. **(B)** Time-series assessment of neural conversion via RT-qPCR for neural fate markers *SOX2* and *PAX6*, and epiblast/pluripotency markers *OCT3/4* and *NANOG*. **(C)** Representative immunofluorescence staining for neural markers *PAX6* (nuclear) and *Musashi-1* (cytoplasmic) at day 6 of neural induction. Insets: 2x magnification. Scale bar = 45 μm . **(D)** Flow cytometry analysis for *PAX6*, *Musashi-1* and *SOX1* at day 4 and day 6 of neural induction. Quantification of the fraction of positively stained cells (colored lines) was performed in relation to a negative control (dashed gray line). **(D')** Representative contour plots comparing marker expression between day 4 (blue) and day 6 (red) of neural induction. FSC values were plotted against fluorescence to discriminate populations. **(E)** RT-qPCR for NPB specifiers *PAX3*, *DLX5*, *MSX2*, *ZIC1* and *TFAP2A* during neural induction. **(F)** RT-qPCR for cadherins involved in NCC development *CDH1*,

CDH2, *CDH6*, *CDH7* and *CDH11* during neural induction. n = 3 biological replicates; RT-qPCR data are plotted as mean \pm SEM. ***p<0.001 ** p<0.01, *p<0.05; One-way ANOVA with Dunnett's correction.

NBCs can be efficiently directed towards NBC-derived neural crest cells (NBC-NCCs)

Within the NPB domain, NCCs are specified under still unclear mechanisms involving inductive signals from the mesoderm that include WNT and FGF ligands. Therefore, NCCs were induced from NBCs with a modified NCC induction method that employs E6 medium supplemented with bFGF, an Activin pathway blocker (SB431542), and a WNT pathway activator (CHIR99021), for 15 days (Menendez et al., 2013; Miller et al., 2017). Additionally, this NCC induction strategy was chosen because it efficiently converts hiPSCs into NCCs with use of a defined E6-based medium in monolayer cultures, obviating the need for complex media formulations, suspension cultures, or cell sorting.

During the NBC-to-NCC differentiation, we observed gradual upregulation of NCC specifiers *SOX9* and *SOX10*, in addition to NCC marker *NGFR* (p75), and neural crest/mesenchymal marker *VIM* (Fig. 2A). Flow cytometry analysis of NBC-NCCs at the end of differentiation (day 15) evidenced high conversion efficiency of NBCs into NCCs, showing 94.04 ± 9.06 % of double-positive cells for neural crest markers p75 and HNK-1 (n=5 attempts; Fig. 2B and data not shown).

We further characterized day 15 NBC-NCCs by evaluating their differentiation potential. After random differentiation, we observed cells positive for α SMA (Fig. 2C), and after treatment with Heregulin- β 1, expression of glial/Schwann cell precursor marker S100B was evident (Fig. 2D), indicating lineage commitment towards smooth muscle and glia, respectively. We also converted NBC-NCCs into ectomesenchymal stem cells, as evidenced by positive staining (>94 % of cells) for mesenchymal stem cell markers CD29, CD73, CD166 and CD90, and negative staining (<0.45 %) for endothelial marker CD31 and haematopoietic marker CD45 (Fig 2E). These cells could be further differentiated towards adipocyte, osteoblast and chondrocyte lineages, as evidenced by Oil Red-O, Alizarin Red, and Alcian Blue staining, respectively (Fig. 2F). Taken together, these results show that this strategy generates multipotent NCCs with high efficiency from a NPB-like neuroepithelial stage (NBCs).

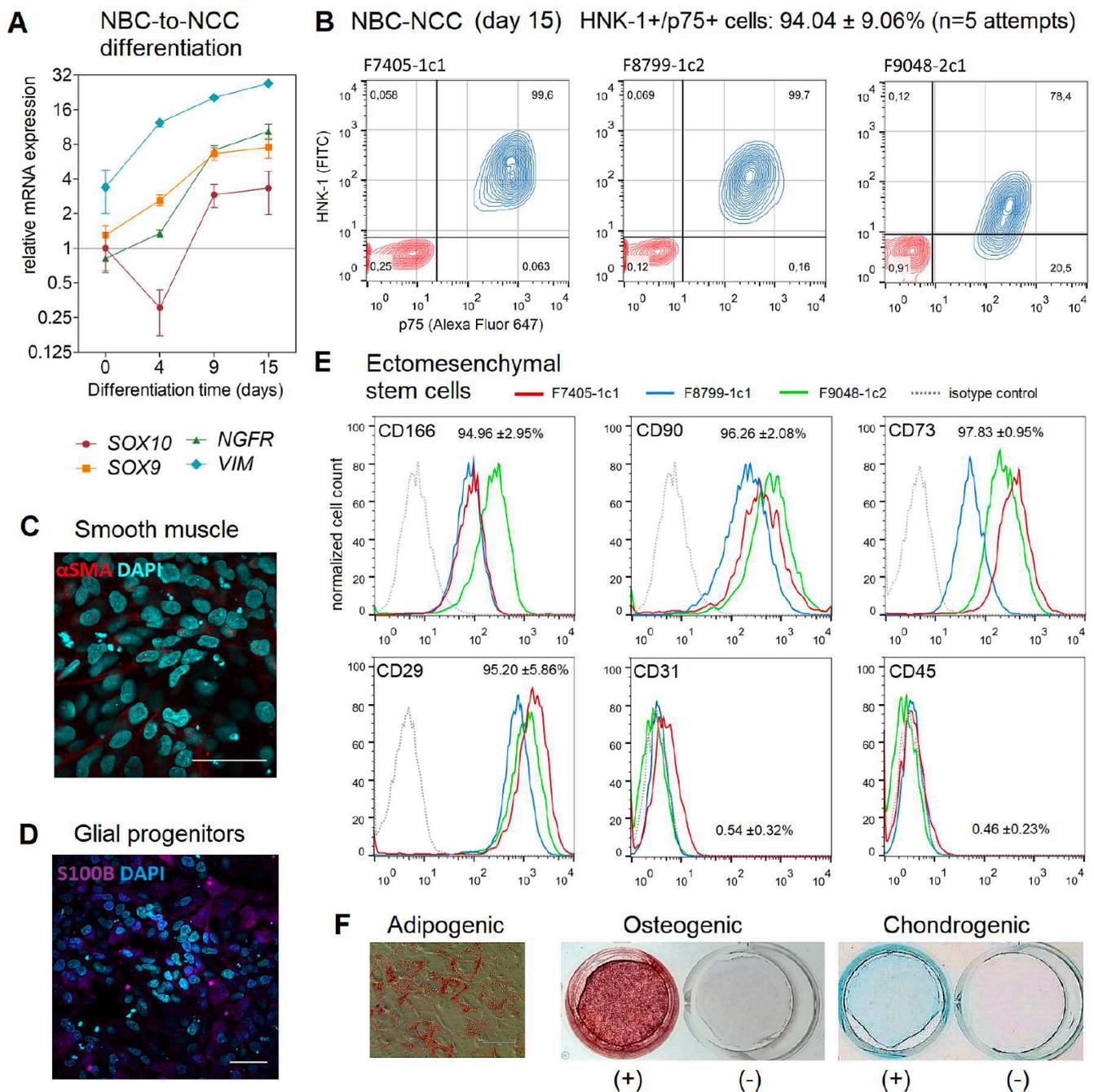


Figure 2. NBCs can be converted to multipotent NCCs with high efficiency. **(A)** RT-qPCR for neural crest specifiers *SOX9* and *SOX10*, neural crest marker *NGFR* (p75), and mesenchymal marker *VIM* (vimentin) during the 15 days of NBC-to-NCC induction. Data are plotted as mean \pm SEM. **(B)** Representative biparametric flow cytometry analysis for neural crest markers HNK-1 and p75 depicting double-positive cells (blue) versus isotype control staining (red). **(C-F)** Differentiation potential of NBC-NCCs. **(C)** Immunofluorescence for α SMA (smooth muscle). **(D)** Immunofluorescence for S100B (glial progenitors). **(E)** Flow cytometry profiles of ectomesenchymal stem cells differentiated from NBC-NCCs for mesenchymal markers CD73, CD90, CD166 and CD29, endothelial marker CD31, and hematopoietic marker CD45. Experimental data (colored lines) are plotted in overlay with data from representative isotype controls. **(F)** Representative Oil Red-O, Alizarin Red and Alcian Blue staining for adipogenic, osteogenic and chondrogenic differentiation, respectively. Scale bars: 45 μ m (immunostaining), 400 μ m (Oil Red-O). (+) differentiated cells; (-) negative control.

NBC-to-NCC differentiation activates the neural crest specification transcriptional program and EMT

The NPB domain is established by activation of NPB specifiers (e.g. *TFAP2A/B*, *DLX5/6*, *PAX3/7*, and *ZIC1*) in epiblast cells. Within the NPB, NCCs are specified via activation of neural crest specifiers (e.g. *SOX8/9/10*, *FOXD3*, *SNAI1/2*, *TWIST1*) that act in concert with already active NPB specifiers (e.g. *TFAP2A/B*, *PAX3/7*, *ZIC1*), while other NPB factors are downregulated (e.g. *SOX2*, *DLX5/6*) (Pla and Monsoro-Burq, 2018; Prasad et al., 2019). In this process, the resultant cellular regulatory state triggers neural crest EMT, transitioning cells from a neuroepithelial to a mesenchymal cell phenotype; this is invariably accompanied by transition from expression of E/N-cadherin, usually associated with epithelial cells, to expression of mesenchymal cadherins (Strobl-Mazzulla and Bronner, 2012).

From day 0 to 15 of NBC-to-NCC differentiation, we observed gradual morphological changes in all cell populations reminiscent of EMT. At day 2 of differentiation, cells were still largely organized as an epithelial monolayer with evident cell-cell adhesions, which were gradually lost as cells detached from each other until fully individualized by the end of 15 days of differentiation (Fig. 3A). To confirm whether these changes result from activation of the neural crest specification program we examined mRNA expression of several specifiers and markers. To better visualize gene expression dynamics associated with early NCC specification events, we included the initial 4 days (hiPSC to NBC) followed by the 15 days (NBC to NCC) of induction (Fig. 3B).

After upregulation from hiPSCs to NBCs, neural/NPB factors *PAX6* and *SOX2* were downregulated during the 15 days of NBC-to-NCC induction, with undetectable *PAX6* expression by the end (at days 12 and 15); NPB factor *DLX5* is also downregulated, remains undetected from days 9-12, and is later expressed at day 15, possibly reflecting later functions in NCC patterning in the pharyngeal arches (Shimizu et al., 2018) (Fig. 3C). After initial upregulation in NBCs (day 0), expression of NPB specifiers also required in NCC specification was either generally maintained high (*TFAP2A* and *MSX2*) or further increased (*PAX3*) until day 15 (Fig. 3D).

From days 0 to 15 of NBC-NCC induction we also observed upregulation of NCC specifiers/EMT inducers *TWIST1*, *SNAI2*, *FOXD3* and *ETS1*. Contrary to *TWIST1*, which is continually upregulated from day 0 onwards, the EMT inducers *FOXD3*, *SNAI2* and *ETS1* are

upregulated only after day 2: *SNAI2*, a NPB factor in *Xenopus*, was initially upregulated at day 0, remains stable at day 2, and further increases expression from days 4-9; *ETS1*, a cranial-specific NCC specifier, initially remains stably expressed and is also upregulated after day 2; *FOXD3*, an NCC specifier that is transiently expressed in the neural plate/border region and later reactivated during NCC specification/EMT (Khudyakov and Bronner-Fraser, 2009) is initially upregulated at day 0, downregulated at day 2, but then recovers expression from day 4 onwards (Fig. 3E).

Finally, a shift in expression was evident for cadherin transcripts throughout the entire process: expression of *CDH1* and *CDH2* was already observed at day 0 and then reduced, with *CDH1* particularly showing pronounced downmodulation from day 7 onwards; in contrast, mesenchymal cadherins *CDH6/7/11*, which exert different functions in NCC specification and migration across organisms (Taneyhill and Schiffmacher, 2017), either retained high expression (*CDH6* and *CDH7*) or were further upregulated (*CDH11*) from day 0 to 15 (Fig. 3F). These observations show that the differentiating NBC-NCC populations recapitulate the transcriptional program for neural crest specification, showing transition of a NPB-like (neuroepithelial) to a NCC (mesenchymal) transcriptional profile.

During EMT, de-epithelization takes place via deconstruction of cell-cell junctions by different mechanisms, including silencing of epithelial adhesion molecules by Snail family genes (*SNAI1/2*) and cleavage of epithelial cadherins (E/N-Cadherin) at the plasma membrane (Clay and Halloran, 2011; Lamouille et al., 2014). Since we observed transcriptional silencing of *CDH1/2* and upregulation of *SNAI2* during NBC-NCC induction, we sought to investigate protein expression and localization of *CDH1* (E-cadherin) and its transcriptional repressor *SNAI2* during the 15 days of NBC-to-NCC induction. Confocal immunofluorescence analysis revealed evident E-Cadherin expression localized to plasma membranes at cell-cell contact sites from day 0 to day 2 in all samples (Fig. 3G and Fig. S1). From day 4 onwards, sharp E-cadherin expression could be detected at cell-cell boundaries until day 6 for F8799-1c1, and until day 4 for F9048-1c2, and day 8 for F7405-1c1 (Fig. S1). On average, E-Cadherin disappearance from these locations takes place between days 2 and 8 of differentiation; in the timepoints following day 8, E-Cadherin was no longer expressed at plasma membranes (Fig. 3G). *SNAI2* protein expression was also confirmed at day 0 (NBCs), localized to DAPI-depleted nuclear foci. During differentiation, this occasionally shifted to a less defined, scattered nuclear *SNAI2* staining pattern (day 2 of F8799-1c1, and

day 8 of F7405-1c1 and F9048-1c2). Nevertheless, nuclear expression of *SNAI2* could be observed in all timepoints, and by the end of the NBC-to-NCC differentiation, *SNAI2* was strongly expressed in nuclear foci (Fig. 3H and Fig. S1).

We also examined protein expression of *CDH2* (N-Cadherin), which is cleaved at cell junctions and translocated to nuclei to promote expression of pro-EMT genes during chick NCC development (Shoval et al., 2007; Strobl-Mazzulla and Bronner, 2012). Like E-Cadherin, N-Cadherin expression is lost at intercellular contact sites as cells move from epithelial to mesenchymal identity. Initially, N-cadherin was clearly observed at those sites from days 0-4 in all samples (Fig. 3G' and Fig. S2). From day 4 onwards, N-Cadherin could still be detected in patches of plasma membrane between neighbouring cells until day 6 in F7405-1c1, and day 8 in F8799-1c1 and in F9048-1c2. In one sample (F7405-1c1), a diffuse N-cadherin expression pattern could still be observed at day 11 in some plasma membranes, although cells were already fully individualized at this day in this and the other cell lines (Fig. S2 and data not shown). On average, strong expression and localization of N-Cadherin between epithelial cells gradually become less evident from days 4 to 8 (Fig. 3G'). Contrasting observations in chick embryos (Shoval et al., 2007; Strobl-Mazzulla and Bronner, 2012), N-Cadherin showed no evidence of nuclear accumulation during NBC-NCC induction; instead, at 15 days of induction, the mesenchymal staining pattern of N-Cadherin was cytoplasmic, while E-Cadherin showed evident nuclear accumulation in all samples (Fig. 3I). Overall, we observed gradual disappearance of E-Cadherin and N-Cadherin at epithelial cell-cell junctions between days 2 and 8, which roughly coincides with the upregulation time frame observed for some NCC specifiers/EMT inducers, including *SNAI2*, *ETS1* and *FOXD3* (Fig.3E). Together, the above findings show that the NCC specification and EMT programs are activated during NBC-to-NCC induction.

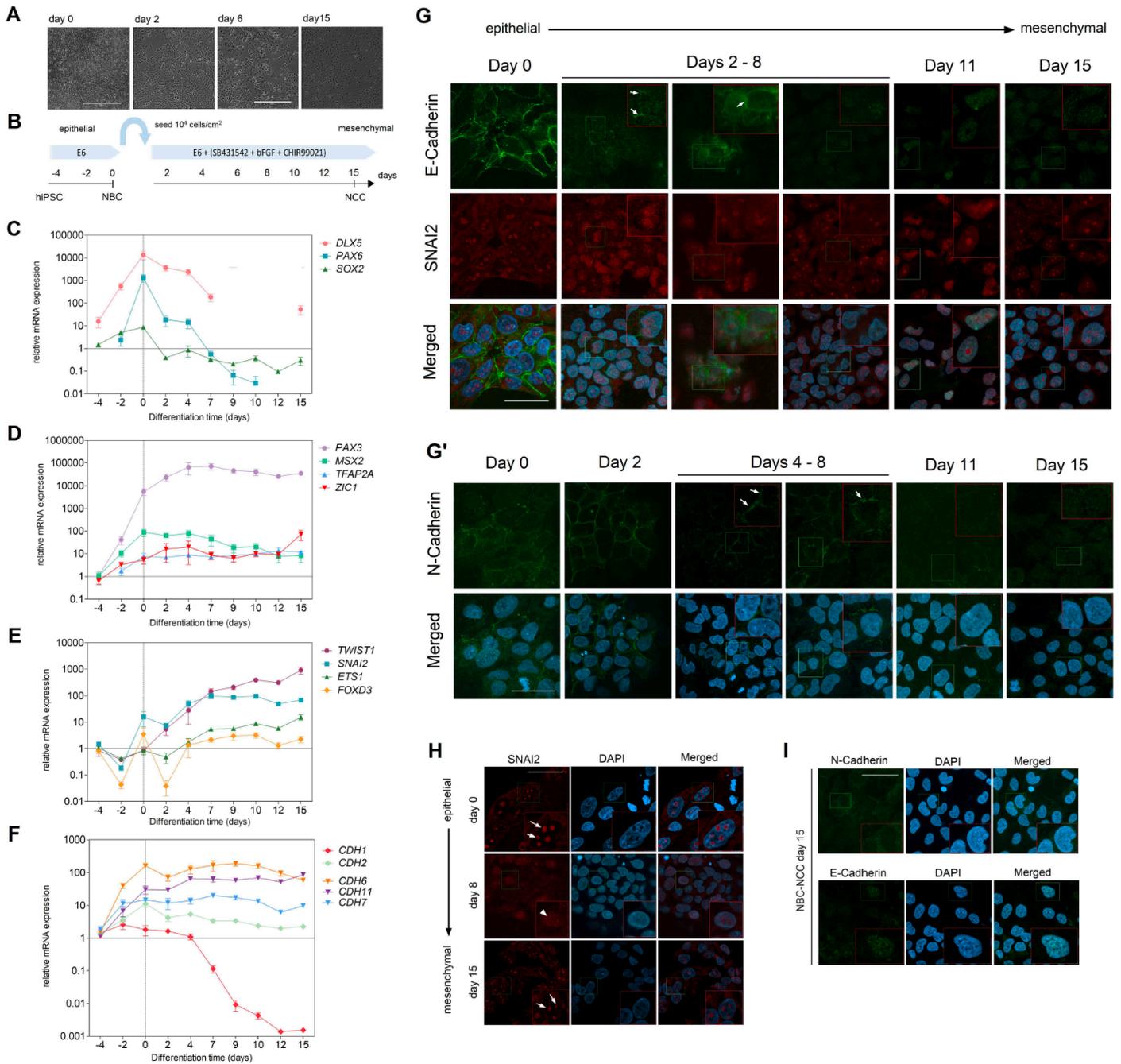


Figure 3. NBC-to-NCC differentiation recapitulates neural crest specification and EMT. **(A)** Representative phase-contrast micrographs depicting epithelial-to-mesenchymal morphological changes during NBC-to-NCC differentiation. Scale bars: 400 μ m. **(B)** Schematic representation of the differentiation setup encompassing hiPSC-to-NBC followed by NBC-to-NCC derivation. **(C-F)** Time-series RT-qPCR analysis of: **(C)** NPB factors *PAX6*, *SOX2* and *DLX5*, **(D)** NPB/NCC specifiers *PAX3*, *MSX2*, *TFAP2A* and *ZIC1*, **(E)** NCC specifiers/EMT inducers *TWIST1*, *SNAI2*, *ETS1* and *FOXD3*, and **(F)** cadherins involved in NCC specification and EMT *CDH1*, *CDH2*, *CDH6*, *CDH7* and *CDH11*. Expression values were plotted relative to day -4 except for *PAX6* and *TFAP2A*. RT-qPCR data are plotted as mean \pm SEM (n=3 biological replicates). **(G,G')** Representative time-series immunofluorescence analysis showing expression patterns for **(G)** E-cadherin (*CDH1*) and *SNAI2*, and for **(G')** N-cadherin. Arrows indicate E/N-cadherin expression at cell-cell junctions. **(H)** Representative immunofluorescence showing *SNAI2* expression patterns observed in: NBCs (day 0), during NBC-NCC

induction (occasionally observed from day 2-8), and in NBC-NCCs (day 15). (I) Representative immunofluorescence showing mesenchymal expression patterns for E-cadherin and N-cadherin in NBC-NCCs at day 15. For G-I: Scale bars = 45 μ m. Insets: 2x magnification. See also figure S1 and S2.

DISCUSSION

Neural crest development occurs through sequential activation of a series of regulatory circuits comprising NPB and neural crest factors that carry out NCC specification/EMT, transitioning cells from a neuroepithelial towards a mesenchymal state. Owing to the relevance of the neural crest in vertebrate body formation, understanding the mechanisms orchestrating these steps in human cells is important to shed light on human development and neural crest-related disorders. In view of the obvious difficulties to study the human embryo, development of pluripotent stem cell-based strategies to dissect human NCC development has been a promising alternative approach. Here, we show that neuroepithelial induction of hiPSCs without altering BMP signaling (E6 method) confers a NPB-like transcriptional state to differentiating cells, which can be directed towards the neural crest fate with near 100% efficiency, recapitulating for the first time the sequential steps that orchestrate NCC specification and EMT.

We observed that neural induction with the E6 method confers a NPB-like (NBC) transcriptional state to differentiating hiPSCs. This method was originally designed to produce PAX6⁺/N-cadherin⁺ neuroepithelial cells, in 6 days without modulating BMP signaling (Lippmann et al., 2014). Although we confirmed production of PAX6⁺ neuroepithelial cells under these conditions, we noted that at day 4 nearly all cells (>98 %) already expressed Musashi-1 and PAX6 at lower levels (PAX6^{low}) compared to cells at day 6 (PAX6^{high}). Notably, SOX1, which is co-expressed with PAX6 in neural progenitor cells (Li et al., 2017), was not expressed at day 4 and is only partially expressed at day 6 here and in the original report describing the E6 method (Lippmann et al., 2014). Considering the marked upregulation of key NPB specifiers at this differentiation day, we hypothesize that neural induction without BMP inhibition in fact enables hiPSCs to attain a transient, PAX6^{low}/SOX1⁻/Musashi1⁺ NPB identity at day 4, which is further directed towards neural commitment at day 6 under PAX6^{high} expression by maintaining E6 induction. *In vivo* experimentation shows that the NPB domain is specified under intermediate BMP activity orchestrated by signals

from the prospective neural (low BMP activity) and non-neural (high BMP activity) ectoderm, allowing for transient co-expression of both neural and non-neural factors that act as NPB specifiers (Maharana and Schlosser, 2018; Roellig et al., 2017). This includes activation of AP2-alpha (*TFAP2A*), followed by stabilization of *PAX3*, *ZIC1*, and *MSX1/2* (Betancur et al., 2010; de Croz e et al., 2011; Reichert et al., 2013; Schille and Schambony, 2017; Sim oes-Costa and Bronner, 2015), in addition to Cadherin-6B, which modulates BMP signaling and segregation of the neural crest territory from the NPB and promotes neural crest EMT in chick embryos (Dady and Duband, 2017; Park and Gumbiner, 2010). In line with those findings, 4 days of E6 induction enabled elevated expression of several NPB transcripts, including *TFAP2A* and downstream specifiers *DLX5*, *PAX3*, *ZIC1*, and *MSX2*, in addition to peak expression of *CDH6*. Thus, lack of BMP blockade could be recapitulating early inductive signals and producing NPB cells under native (intermediate) BMP pathway activity, which translates into the highly effective derivation of NCCs observed here. Precise adjustment of BMP activity via combination of exogenous inhibitors and activators has been recently proposed to produce human NCCs in vitro either directly or through a PAX6⁺ hindbrain NPB-like intermediate; however, these strategies require complex media formulations or cell sorting to obtain pure populations (Hackland et al., 2017; Thier et al., 2019), whereas 94 % efficiency was attained under the simple NCC induction setup reported here. Still, although homogeneous expression of PAX6^{low} in NBCs was evident, PAX6 is not an exclusive NPB marker and more studies will be necessary to further characterize NBC populations and evaluate the proportion of bona-fide NPB cells within NBCs.

Our results show that redirecting NBCs towards the neural crest fate generates NCCs with high efficiency while recapitulating molecular events of NCC specification. At the end of the 15 days of differentiation, NBC-NCC populations showed over 94 % of p75/HNK1 double-positive cells, upregulation of NCC markers such as *SOX9/10*, *NGFR* (p75) and *VIM*, and they could be further directed towards neural crest derivatives. These results attest the neural crest identity of NBC-NCCs and demonstrate that they possess equivalent characteristics to NCCs directly derived from pluripotent stem cells (Fukuta et al., 2014; Menendez et al., 2013; Miller et al., 2017). Notably, the gene expression profile of differentiating NBC-NCCs recapitulates the NCC developmental program ascertained in animal models (Prasad et al., 2019): from day 0 to 15 of differentiation, we observed downregulation of NPB factors not involved in NCC specification (*DLX5*, *PAX6*, *SOX2*),

increment or maintenance in expression of factors involved both in NPB and in NCC specification (*PAX3*, *MSX2*, *TFAP2A*, *ZIC1*), and upregulation of NCC specification factors (*SNAI2*, *TWIST1*, *ETS1*, and *FOXD3*). Accordingly, as expected during transition from an epithelial to a mesenchymal cellular state, these changes were accompanied by a shift in cadherin expression, where *CDH1/2* expression is downregulated in favor of mesenchymal cadherins *CDH6/7/11*, a hallmark of neural crest EMT (Strobl-Mazzulla and Bronner, 2012).

Another important finding is that NBC-NCC differentiation allows for stepwise observation of events associated with EMT. In vivo, NCC EMT is initiated upon activation of a gene regulatory network that drives acquisition of a mesenchymal, migratory phenotype (Milet and Monsoro-Burq, 2012; Nieto, 2009; Theveneau and Mayor, 2012) that depends on proper regulation of multiple cadherins and adhesion molecules by EMT inducers (Scarpa et al., 2015; Taneyhill, 2008). This was evidenced by gradual morphological and molecular hallmarks of EMT in the course of 15 days, including upregulation of NCC EMT inducers *SNAI2*, *ETS1* and *FOXD3* from days 2 to 7, coupled to downregulation of E/N-Cadherin mRNA, and reduction of E/N-Cadherin protein localization to cell-cell contact sites from days 2 to 8. Therefore, despite the variation in timing for these events across samples, EMT-related changes in gene expression and cadherin localization take place between days 2 and 8. Remarkably, after the shift from epithelial to mesenchymal cell states, E-Cadherin and N-Cadherin protein expression is not completely ablated; rather, we observed nuclear accumulation of E-Cadherin and cytoplasmic expression of N-Cadherin in NBC-NCCs. These observations are in line with the novel roles ascribed to these cadherins in NCC EMT and migration in animal models, in which N-Cadherin fragments are translocated to nuclei instead of E-Cadherin (Bahm et al., 2017; Cousin, 2017; Huang et al., 2016a; Piloto and Schilling, 2010; Powell et al., 2015; Strobl-Mazzulla and Bronner, 2012). In humans, E-cadherin fragments are translocated to nuclei via p120 catenin activity to activate gene expression in some cell lines (Ferber et al., 2008), and nuclear accumulation of E-cadherin is also found in several cancers (Rodriguez et al., 2012). Our results support the notion that E/N-cadherin can exert functions beyond cell adhesion, and indicate that E-cadherin may play novel roles in human NCC development. Further investigation will be required to shed light on the roles played by cadherins and other factors involved in human neural crest specification, as well as to elucidate the similarities and dissimilarities between human and animal models.

The present differentiation setup provides time-series information on human neural crest development while attaining high NBC-to-NCC conversion rates. The ability to assess progression of EMT and other key molecular and cellular processes participating in NCC specification will be important to investigate both normal human development and disorders suspected to arise from alterations in these events of NCC development. These disorders include, besides congenital neurocristopathies, neural crest-derived tumors such as neuroblastoma, melanoma, ganglioneuroma and others, in which the embryonic neural crest developmental program is recapitulated, contributing to malignancy and metastasis (Kerosuo and Bronner-Fraser, 2012; Maguire et al., 2015; Powell et al., 2013). In addition, the simple and efficient method for generating NBC-NCCs reported here could be further developed towards drug/genetic screening and cytotoxicity tests for future therapeutic applications. Finally, it provides a human in vitro model to investigate the functional effects of candidate genetic variants detected by next-generation sequencing or other methods, which is a current challenge in Medical Genetics research.

MATERIAL AND METHODS

Generation and maintenance of hiPSCs

Human induced pluripotent stem cells (hiPSCs) were reprogrammed from two dermal fibroblast (F9048-1c2 and F7405-1c1) and one erythroblast culture (F8799-1c1) according to previous protocols (Miller et al., 2017; Okita et al., 2013). hiPSCs were maintained in defined medium consisted of Essential 8™ Medium (E8 medium; Life Technologies) supplemented with 100 µg/mL of Normocin (Invivogen), in feeder-free condition. Cells were cultured in Matrigel (BD Biosciences)-coated vessels (60-mm petri dishes or 6-well plates), and routinely passaged with Accutase (Life Technologies). The hiPSC lines used in this work had already been generated and characterized, as reported elsewhere (Ishiy et al., 2015; Miller et al., 2017). Importantly, hiPSCs must be conditioned to E8 minimal medium before neuroepithelial induction (Lippmann et al., 2014).

Differentiation of hiPSCs to NPB-like neuroepithelial cells (NBCs)

hiPSC-derive NBCs were generated based on previously published methodology ('E6 method') (Lippmann et al., 2014). In brief, hiPSCs were seeded onto 60-mm Matrigel-coated dishes (1×10^5 cells/cm²) in E8 medium with 5 μ M ROCK inhibitor (Sigma-Aldrich), and one day post-seeding, medium was changed to Essential 6™ Medium (E6 medium; Life Technologies) supplemented with 100 μ g/mL of Normocin. Medium was replenished daily, and differentiation was monitored every 2 days via RT-qPCR analysis, for 6 days.

Differentiation of NBCs to neural crest cells (NBC-NCCs)

NCCs were differentiated from NBCs under WNT pathway activation and TGF- β /Activin pathway blockade (Fukuta et al., 2014; Menendez et al., 2013; Miller et al., 2017). At day 4 of neuroepithelial differentiation, cells were washed once with PBS, detached with Accutase, and collected by centrifugation. Then, 3×10^4 cells/cm² were seeded on Matrigel-coated dishes in E6 medium supplemented with 5 μ M ROCK inhibitor. In the following day, medium was changed to NCC differentiation medium, composed of E6 medium supplemented with 100 μ g/mL of Normocin, 8 ng/mL bFGF (Life Technologies), 20 μ M SB431542 (Tocris), 1 μ M CHIR99021 (Sigma-Aldrich). Cells were split before reaching confluence with Accutase at room temperature (RT) whenever necessary. Since cells continue to replicate as differentiation takes place, cell collection was timed with subculturing to comprise the 15 days of differentiation in 8 timepoints for RT-qPCR and 8 timepoints for confocal immunofluorescence analyses. For flow cytometry characterization, a total of 5 differentiation experiments were carried out (once for F7405-1c1 and twice for F8799-1c1 and F9048-1c2).

Differentiation of NBC-NCCs towards neural crest derivatives

To produce smooth muscle cells, NBC-NCC cultures were spontaneously differentiated in NCC medium for 7 days, according to previous work (Kreitzer et al., 2013). For peripheral glia commitment, NCCs were treated with NCC medium supplemented with 20 ng/mL Heregulin- β 1 (PeproTech) for 10 days (Liu et al., 2012). NCC-derived ectomesenchymal stem cell populations were obtained by culturing NCCs with DMEM/F12 supplemented with 10 % fetal bovine serum, 2 mM GlutaMAX, 0.1 mM non-essential

aminoacids (all provided by Life Technologies), and 100 µg/mL of Normocin. NCCs were seeded at 2×10^4 cells/cm² onto non-coated 60-mm dishes and were differentiated for 6 days, and passaged with TrypLE™ Express (Life Technologies) whenever needed. Cell cultures were expanded for up to 6 passages with medium changes every 3 days, before further differentiation into osteoblasts, chondrocytes and adipocytes.

Alizarin Red, Alcian Blue, and Oil Red staining

Osteogenic, chondrogenic and adipogenic differentiation were induced with the use of commercial kits (StemPro Osteogenesis, Chondrogenesis and Adipogenesis kit, Life Technologies), following the manufacturer's recommendations. Staining procedures were performed on cells treated with differentiation medium and on non-treated cells (negative controls), as previously reported (Miller et al., 2017).

Confocal immunofluorescence analysis

For immunofluorescence, cells were fixed in 4 % paraformaldehyde for 20 min at RT. Permeabilization was carried out with PBS 0.2 % Triton X-100 for 30 min followed by blocking with PBS 5 % BSA for 1 h at RT. Then, cells were incubated with primary antibodies overnight at 4 °C, washed three times with PBS and incubated with secondary antibodies at 4 °C for 1 h, protected from light. After washing with PBS two times, cells were counterstaining with DAPI solution (Life Technologies) for 2 min at RT. For NBCs, primary anti-PAX6 (42-6600; Thermo Scientific), anti-N Cadherin (ab19348; Abcam) and anti-Musashi 1 (ab52865; Abcam) were used. For neural crest derivatives, we used anti-alpha smooth muscle Actin (α SMA; ab5694; Abcam) and anti-S100B (SAB1402349; Sigma Aldrich). For staining of cells under NBC-to-NCC differentiation, anti-E Cadherin (ab1416; Abcam), anti-N Cadherin (ab19348; Abcam) and anti-Slug (SNAI2) (#9585; Cell Signaling Technology) were used. Secondary Goat anti-Mouse IgG (H+L) antibody - Alexa Fluor 488 conjugate (A-11001; Life Technologies) and Goat anti-Rabbit IgG (H+L) antibody – Cyanine3 conjugate (A10520; Invitrogen) were used. Antibody concentrations followed the manufacturer's recommendations. Slides were analyzed under a confocal microscope (Zeiss LSM 800) and captured Z-stacks were processed on ImageJ.

Flow cytometry

For flow cytometry analyses, NBCs and NBC-NCCs were dissociated with Accutase, and washed twice with PBS supplemented with 4 % BSA. Cells were incubated with conjugated antibodies, in the dark for 1 h at 4 °C, followed by two PBS washes and fixed in PBS 1 % paraformaldehyde. FIX & PERM (Life Technologies) was used for fixation and permeabilization. Primary anti-SOX1 (Cell Signaling, #4194), anti-Musashi (Abcam, ab52865) and anti-PAX6 (Thermo Fisher; 42-6600) antibodies were used with secondary Goat anti-Mouse IgG (H+L) antibody - Alexa Fluor 488 conjugate to assess NBC identity. To assess NCC immunophenotype, IgM K FITC 23 Mouse Anti-Human CD57 (anti-HNK1; BD Pharmingen, 561906), IgG1 K Alexa Fluor 647 Mouse Anti-Human CD271 (anti-p75NTR; BD Pharmingen, 560877), FITC Mouse IgM K isotype control (BD Pharmingen 555583), and Alexa Fluor 647 Mouse IgG1 K isotype control (BD Pharmingen 557714) were used. Mesenchymal stem cells were detached with TrypLE Express and washed twice with PBS 4 % BSA. Incubation procedures were done as aforementioned, using FITC Mouse Anti-Human CD31 (BD Pharmingen 555445), APC Mouse Anti-Human CD73 (BD Pharmingen 560847), PE Mouse Anti-Human CD90 (BD Pharmingen 555596), PE Mouse Anti-Human CD166 (BD Pharmingen 559263), and FITC, PE and APC Mouse IgG1 K Isotype Controls (BD Pharmingen 555748, 554681 and 555749, respectively). All antibodies were used at concentrations recommended by the manufacturers. At least 5,000 events were acquired in a FACS Aria II equipment and analyzed on FlowJo (X 10.0.7r2). For best visual comparison, histograms were normalized to mode of cell count (y-axis).

Real-time quantitative PCR (RT-qPCR)

NucleoSpin RNA II extraction kit (Macherey-Nagel) was used to obtain total RNA from cells, and 1 µg of RNA was converted into cDNA with Superscript IV (Life Technologies) and oligo-dT primers according to the manufacturer's recommendations. RT-qPCR reactions were performed with 2X Fast SYBR Green PCR Master Mix (Life Technologies) and 50–400 nM of each primer. Fluorescence was detected in a 7500 Fast Real-Time PCR System or QuantStudio 5 (Life Technologies), under standard cycling protocol. Primer sequences are described in Table S1 and were supplied by Exxtend. Primer pairs were either designed with Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) or retrieved from Primer

Bank (<http://pga.mgh.harvard.edu/primerbank/>) and their amplification efficiencies (E) were determined by serial cDNA dilutions or directly calculated from amplification plots with LinRegPCR (Ramakers et al., 2003). NormFinder (Andersen et al., 2004) was used to determine the most stable endogenous control (among *ACTB*, *TBP*, *HMBS*, *GAPDH*, and *HPRT1*). Relative mRNA quantity was calculated as previously described (Pfaffl, 2001), relative to a calibrator sample. Results were log-transformed before statistical analyses. Primer sequences are shown in table S1.

Statistical analyses

Gene expression values were log-transformed before statistical analyses. One-way ANOVA was conducted to compare the effect of differentiation (days) on gene expression across samples. When means were significantly different ($p < 0.05$), Dunnett's multiple comparison post-hoc test ($\alpha = 0.05$; 95 % CI) was performed, and significance was indicated in comparison to day 0 of differentiation, unless otherwise stated. Analyses were carried out on GraphPad Prism v5.03 (GraphPad Software Inc.).

Ethics statement

The Ethics Committee of Instituto de Biociências at Universidade de São Paulo, Brazil (accession number 1.463.852) approved the experimental procedures involving human samples. Subjects donated biological samples after providing signed informed consent.

Acknowledgments

The authors thank members of the Passos-Bueno lab for helpful discussions. This work was supported by FAPESP/CEPID 2013/08028-1, FAPESP 2016/24188-7 (GSK), CNPq 305405/2011-5 (MRPB), CAPES 1470032 (CMM) research fellowships.

Author disclosure statement

The authors declare that they have no conflict of interest.

REFERENCES

- Andersen, C.L., Jensen, J.L., and Ørntoft, T.F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* 64, 5245–5250.
- Avior, Y., Sagi, I., and Benvenisty, N. (2016). Pluripotent stem cells in disease modelling and drug discovery. *Nat Rev Mol Cell Biol* 17, 170–182.
- Bahm, I., Barriga, E.H., Frolov, A., Theveneau, E., Frankel, P., and Mayor, R. (2017). PDGF controls contact inhibition of locomotion by regulating N-cadherin during neural crest migration. *Development* 144, 2456–2468.
- 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.
- De Bakker, B.S., de Jong, K.H., Hagoort, J., de Bree, K., Besselink, C.T., de Kanter, F.E.C., Veldhuis, T., Bais, B., Schildmeijer, R., Ruijter, J.M., et al. (2016). An interactive three-dimensional digital atlas and quantitative database of human development. *Science* 354.
- Bell, G.W., Yatskievych, T.A., and Antin, P.B. (2004). GEISHA, a whole-mount in situ hybridization gene expression screen in chicken embryos. *Dev Dyn* 229, 677–687.
- Betancur, P., Bronner-Fraser, M., and Sauka-Spengler, T. (2010). Assembling neural crest regulatory circuits into a gene regulatory network. *Annu Rev Cell Dev Biol* 26, 581–603.
- Clay, M.R., and Halloran, M.C. (2011). Regulation of cell adhesions and motility during initiation of neural crest migration. *Curr Opin Neurobiol* 21, 17–22.
- 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.
- Cousin, H. (2017). Cadherins function during the collective cell migration of *Xenopus* Cranial Neural Crest cells: revisiting the role of E-cadherin. *Mech Dev* 148, 79–88.
- De Crozé, N., Maczkowiak, F., and Monsoro-Burq, 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.
- Curchoe, C.L., Maurer, J., McKeown, S.J., Cattarossi, G., Cimadamore, F., Nilbratt, M., Snyder, E.Y., Bronner-Fraser, M., and Terskikh, A.V. (2010). Early acquisition of neural crest competence during hESCs neuralization. *PLoS ONE* 5, e13890.
- Dady, A., and Duband, J.-L. (2017). Cadherin interplay during neural crest segregation from the non-neural ectoderm and neural tube in the early chick embryo. *Dev Dyn* 246, 550–565.
- Darnell, D.K., Kaur, S., Stanislaw, S., Davey, S., Konieczka, J.H., Yatskievych, T.A., and Antin, P.B. (2007). GEISHA: an in situ hybridization gene expression resource for the chicken embryo. *Cytogenet Genome Res* 117, 30–35.
- Devotta, A., Juraver-Geslin, H., Gonzalez, J.A., Hong, C.-S., and Saint-Jeannet, J.-P. (2016). Sf3b4-depleted *Xenopus* embryos: A model to study the pathogenesis of craniofacial defects in Nager syndrome. *Dev Biol* 415, 371–382.
- Dixon, J., Jones, N.C., Sandell, L.L., Jayasinghe, S.M., Crane, J., Rey, J.-P., Dixon, M.J., and Trainor, P.A. (2006). Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proc Natl Acad Sci U S A* 103, 13403–13408.
- Le Douarin, N.M., and Dupin, E. (2003). Multipotentiality of the neural crest. *Curr Opin Genet Dev* 13, 529–536.
- Ferber, E.C., Kajita, M., Wadlow, A., Tobiansky, L., Niessen, C., Ariga, H., Daniel, J., and Fujita, Y. (2008). A role for the cleaved cytoplasmic domain of E-cadherin in the nucleus. *J Biol Chem* 283, 12691–12700.
- Fukuta, M., Nakai, Y., Kirino, K., Nakagawa, M., Sekiguchi, K., Nagata, S., Matsumoto, Y., Yamamoto, T., Umeda, K., Heike, T., et al. (2014). Derivation of mesenchymal stromal cells from pluripotent stem cells through a neural crest lineage using small molecule compounds with defined media. *PLoS ONE* 9, e112291.

Hackland, J.O.S., Frith, T.J.R., Thompson, O., Marin Navarro, A., Garcia-Castro, M.I., Unger, C., and Andrews, P.W. (2017). Top-Down Inhibition of BMP Signaling Enables Robust Induction of hPSCs Into Neural Crest in Fully Defined, Xeno-free Conditions. *Stem Cell Reports* 9, 1043–1052.

Huang, C., Kratzer, M.-C., Wedlich, D., and Kashef, J. (2016a). E-cadherin is required for cranial neural crest migration in *Xenopus laevis*. *Dev Biol* 411, 159–171.

Huang, M., Miller, M.L., McHenry, L.K., Zheng, T., Zhen, Q., Ilkhanizadeh, S., Conklin, B.R., Bronner, M.E., and Weiss, W.A. (2016b). Generating trunk neural crest from human pluripotent stem cells. *Sci Rep* 6, 19727.

Ishiy, F.A.A., Fanganiello, R.D., Griesi-Oliveira, K., Suzuki, A.M., Kobayashi, G.S., Morales, A.G., Capelo, L.P., and Passos-Bueno, M.R. (2015). Improvement of In Vitro Osteogenic Potential through Differentiation of Induced Pluripotent Stem Cells from Human Exfoliated Dental Tissue towards Mesenchymal-Like Stem Cells. *Stem Cells Int* 2015, 249098.

Jiang, X., Gwyne, Y., McKeown, S.J., Bronner-Fraser, M., Lutzko, C., and Lawlor, E.R. (2009). Isolation and characterization of neural crest stem cells derived from in vitro-differentiated human embryonic stem cells. *Stem Cells Dev* 18, 1059–1070.

Kerosuo, L., and Bronner-Fraser, M. (2012). What is bad in cancer is good in the embryo: importance of EMT in neural crest development. *Semin Cell Dev Biol* 23, 320–332.

Khudyakov, J., and Bronner-Fraser, M. (2009). Comprehensive spatiotemporal analysis of early chick neural crest network genes. *Dev Dyn* 238, 716–723.

Kreitzer, F.R., Salomonis, N., Sheehan, A., Huang, M., Park, J.S., Spindler, M.J., Lizarraga, P., Weiss, W.A., So, P.-L., and Conklin, B.R. (2013). A robust method to derive functional neural crest cells from human pluripotent stem cells. *Am J Stem Cells* 2, 119–131.

Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* 15, 178–196.

Lee, G., Kim, H., Elkabetz, Y., Al Shamy, G., Panagiotakos, G., Barberi, T., Tabar, V., and Studer, L. (2007). Isolation and directed differentiation of neural crest stem cells derived from human embryonic stem cells. *Nat Biotechnol* 25, 1468–1475.

Lee, G., Papapetrou, E.P., Kim, H., Chambers, S.M., Tomishima, M.J., Fasano, C.A., Ganat, Y.M., Menon, J., Shimizu, F., Viale, A., et al. (2009). Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* 461, 402–406.

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.

Leung, A.W., Murdoch, B., Salem, A.F., Prasad, M.S., Gomez, G.A., and García-Castro, M.I. (2016). WNT/ β -catenin signaling mediates human neural crest induction via a pre-neural border intermediate. *Development* 143, 398–410.

Li, Y., Wang, R., Qiao, N., Peng, G., Zhang, K., Tang, K., Han, J.-D.J., and Jing, N. (2017). Transcriptome analysis reveals determinant stages controlling human embryonic stem cell commitment to neuronal cells. *J Biol Chem* 292, 19590–19604.

Lippmann, E.S., Estevez-Silva, M.C., and Ashton, R.S. (2014). Defined human pluripotent stem cell culture enables highly efficient neuroepithelium derivation without small molecule inhibitors. *Stem Cells* 32, 1032–1042.

Liu, Q., Spusta, S.C., Mi, R., Lassiter, R.N.T., Stark, M.R., Höke, A., Rao, M.S., and Zeng, X. (2012). Human neural crest stem cells derived from human ESCs and induced pluripotent stem cells: induction, maintenance, and differentiation into functional schwann cells. *Stem Cells Translational Medicine* 1, 266–278.

MacNicol, A.M., Hardy, L.L., Spencer, H.J., and MacNicol, M.C. (2015). Neural stem and progenitor cell fate transition requires regulation of Musashi1 function. *BMC Dev Biol* 15, 15.

Maguire, L.H., Thomas, A.R., and Goldstein, A.M. (2015). Tumors of the neural crest: Common themes in development and cancer. *Dev Dyn* 244, 311–322.

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

Menendez, L., Yatskievych, T.A., Antin, P.B., and Dalton, S. (2011). Wnt signaling and a Smad pathway blockade direct the differentiation of human pluripotent stem cells to multipotent neural crest cells. *Proc Natl Acad Sci U S A* 108, 19240–19245.

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.

Mica, Y., Lee, G., Chambers, S.M., Tomishima, M.J., and Studer, L. (2013). Modeling neural crest induction, melanocyte specification, and disease-related pigmentation defects in hESCs and patient-specific iPSCs. *Cell Rep* 3, 1140–1152.

Milet, C., and Monsoro-Burq, A.H. (2012). Neural crest induction at the neural plate border in vertebrates. *Dev Biol* 366, 22–33.

Miller, E.E., Kobayashi, G.S., Musso, C.M., Allen, M., Ishiy, F.A.A., de Caires, L.C., Goulart, E., Griesi-Oliveira, K., Zechi-Ceide, R.M., Richieri-Costa, A., et al. (2017). EIF4A3 deficient human iPSCs and mouse models demonstrate neural crest defects that underlie Richieri-Costa-Pereira syndrome. *Hum Mol Genet* 26, 2177–2191.

Nieto, M.A. (2009). Epithelial-Mesenchymal Transitions in development and disease: old views and new perspectives. *Int J Dev Biol* 53, 1541–1547.

Noisa, P., Lund, C., Kanduri, K., Lund, R., Lähdesmäki, H., Lahesmaa, R., Lundin, K., Chokechuwattanalert, H., Otonkoski, T., Tuuri, T., et al. (2014). Notch signaling regulates the differentiation of neural crest from human pluripotent stem cells. *J Cell Sci* 127, 2083–2094.

Okita, K., Yamakawa, T., Matsumura, Y., Sato, Y., Amano, N., Watanabe, A., Goshima, N., and Yamanaka, S. (2013). An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells* 31, 458–466.

Park, K.-S., and Gumbiner, B.M. (2010). Cadherin 6B induces BMP signaling and de-epithelialization during the epithelial mesenchymal transition of the neural crest. *Development* 137, 2691–2701.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29, e45.

Piloto, S., and Schilling, T.F. (2010). *Ovo1* links Wnt signaling with N-cadherin localization during neural crest migration. *Development* 137, 1981–1990.

Pla, P., and Monsoro-Burq, A.H. (2018). The neural border: Induction, specification and maturation of the territory that generates neural crest cells. *Dev Biol*.

Pomp, O., Brokhman, I., Ben-Dor, I., Reubinoff, B., and Goldstein, R.S. (2005). Generation of peripheral sensory and sympathetic neurons and neural crest cells from human embryonic stem cells. *Stem Cells* 23, 923–930.

Powell, D.R., Blasky, A.J., Britt, S.G., and Artinger, K.B. (2013). Riding the crest of the wave: parallels between the neural crest and cancer in epithelial-to-mesenchymal transition and migration. *Wiley Interdiscip Rev Syst Biol Med* 5, 511–522.

Powell, D.R., Williams, J.S., Hernandez-Lagunas, L., Salcedo, E., O’Brien, J.H., and Artinger, K.B. (2015). *Cdon* promotes neural crest migration by regulating N-cadherin localization. *Dev Biol* 407, 289–299.

Prasad, M.S., Charney, R.M., and García-Castro, M.I. (2019). Specification and formation of the neural crest: Perspectives on lineage segregation. *Genesis* 57, e23276.

Van de Putte, T., Francis, A., Nelles, L., van Grunsven, L.A., and Huylebroeck, D. (2007). Neural crest-specific removal of *Zfhx1b* in mouse leads to a wide range of neurocristopathies reminiscent of Mowat-Wilson syndrome. *Hum Mol Genet* 16, 1423–1436.

Ramakers, C., Ruijter, J.M., Deprez, R.H.L., and Moorman, A.F.M. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 339, 62–66.

Reichert, S., Randall, R.A., and Hill, C.S. (2013). A BMP regulatory network controls ectodermal cell fate decisions at the neural plate border. *Development* 140, 4435–4444.

Rodriguez, F.J., Lewis-Tuffin, L.J., and Anastasiadis, P.Z. (2012). E-cadherin’s dark side: possible role in tumor progression. *Biochim Biophys Acta* 1826, 23–31.

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

Sakai, D., Dixon, J., Achilleos, A., Dixon, M., and Trainor, P.A. (2016). Prevention of Treacher Collins syndrome craniofacial anomalies in mouse models via maternal antioxidant supplementation. *Nat Commun* 7, 10328.

Scarpa, E., Szabó, A., Bibonne, A., Theveneau, E., Parsons, M., and Mayor, R. (2015). Cadherin Switch during EMT in Neural Crest Cells Leads to Contact Inhibition of Locomotion via Repolarization of Forces. *Dev Cell* 34, 421–434.

Schille, C., and Schambony, A. (2017). Signaling pathways and tissue interactions in neural plate border formation. *Neurogenesis (Austin, Tex.)* 4, e1292783.

Shimizu, M., Narboux-Nême, N., Gitton, Y., de Lombares, C., Fontaine, A., Alfama, G., Kitazawa, T., Kawamura, Y., Heude, E., Marshall, L., et al. (2018). Probing the origin of matching functional jaws: roles of Dlx5/6 in cranial neural crest cells. *Sci Rep* 8, 14975.

Shoval, I., Ludwig, A., and Kalcheim, C. (2007). Antagonistic roles of full-length N-cadherin and its soluble BMP cleavage product in neural crest delamination. *Development* 134, 491–501.

Simões-Costa, M., and Bronner, M.E. (2015). Establishing neural crest identity: a gene regulatory recipe. *Development* 142, 242–257.

Strobl-Mazzulla, P.H., and Bronner, M.E. (2012). Epithelial to mesenchymal transition: new and old insights from the classical neural crest model. *Semin Cancer Biol* 22, 411–416.

Takahashi, M., and Osumi, N. (2008). Expression study of cadherin7 and cadherin20 in the embryonic and adult rat central nervous system. *BMC Dev Biol* 8, 87.

Taneyhill, L.A. (2008). To adhere or not to adhere: the role of Cadherins in neural crest development. *Cell Adh Migr* 2, 223–230.

Taneyhill, L.A., and Schiffmacher, A.T. (2017). Should I stay or should I go? Cadherin function and regulation in the neural crest. *Genesis* 55.

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.e7.

Theveneau, E., and Mayor, R. (2012). Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. *Dev Biol* 366, 34–54.

Thier, M.C., Hommerding, O., Panten, J., Pinna, R., García-González, D., Berger, T., Wörsdörfer, P., Assenov, Y., Scognamiglio, R., Przybylla, A., et al. (2019). Identification of Embryonic Neural Plate Border Stem Cells and Their Generation by Direct Reprogramming from Adult Human Blood Cells. *Cell Stem Cell* 24, 166–182.e13.

Thyagarajan, T., Totey, S., Danton, M.J.S., and Kulkarni, A.B. (2003). Genetically altered mouse models: the good, the bad, and the ugly. *Crit Rev Oral Biol Med* 14, 154–174.

Trainor, P.A., Melton, K.R., and Manzanares, M. (2003). Origins and plasticity of neural crest cells and their roles in jaw and craniofacial evolution. *Int J Dev Biol* 47, 541–553.

Wakamatsu, N., Yamada, Y., Yamada, K., Ono, T., Nomura, N., Taniguchi, H., Kitoh, H., Mutoh, N., Yamanaka, T., Mushiake, K., et al. (2001). Mutations in SIP1, encoding Smad interacting protein-1, cause a form of Hirschsprung disease. *Nat Genet* 27, 369–370.

Zhang, D., Ighaniyan, S., Stathopoulos, L., Rollo, B., Landman, K., Hutson, J., and Newgreen, D. (2014). The neural crest: a versatile organ system. *Birth Defects Res C Embryo Today* 102, 275–298.

SUPPLEMENTARY MATERIAL

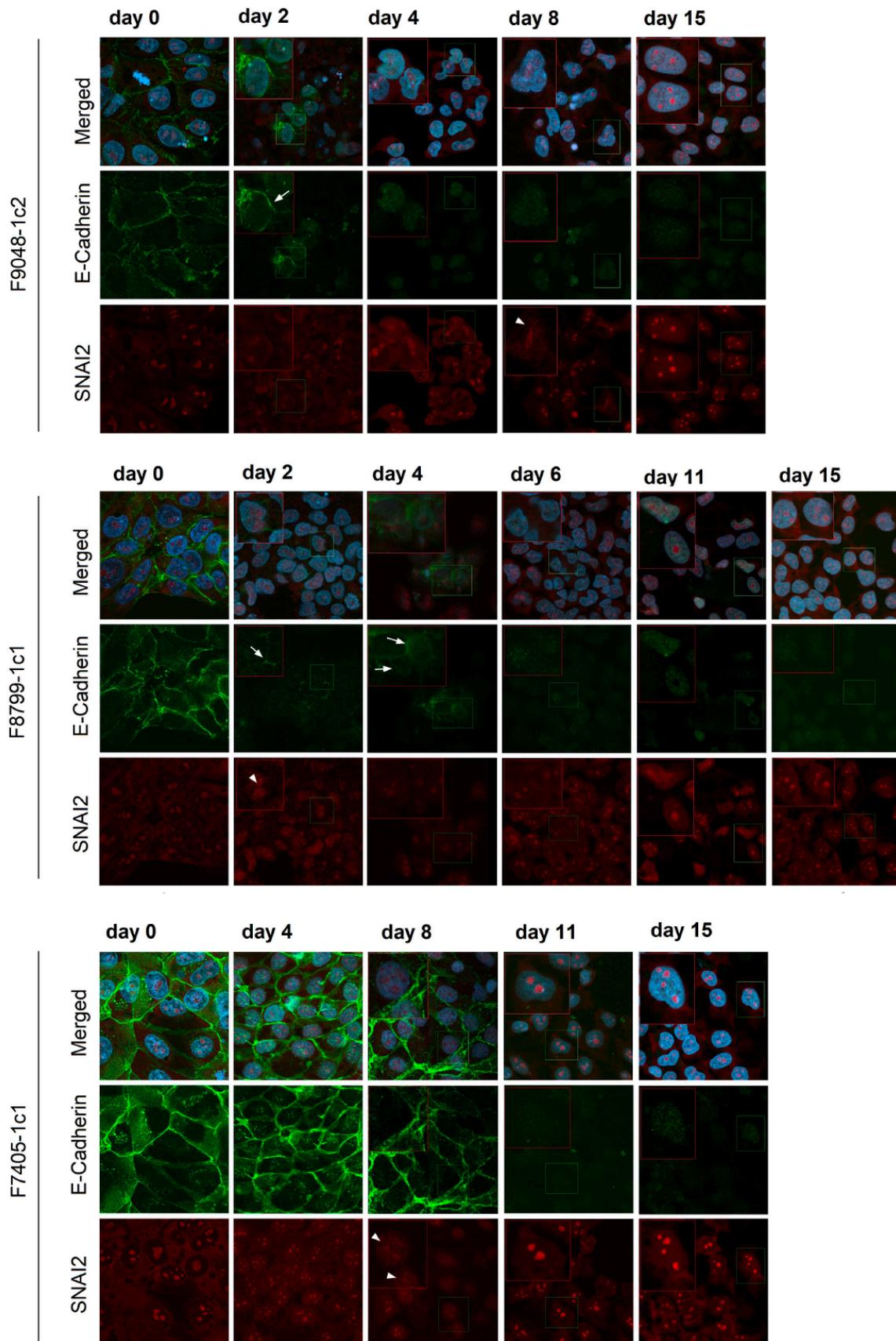


Figure S1. Time-series immunofluorescence analysis showing expression patterns for E-cadherin (*CDH1*) and SNAI2 during NBC-to-NCC differentiation in all samples. Arrows indicate E-cadherin expression at cell-cell junctions. Arrowheads indicate scattered nuclear SNAI2 expression. Scale bars = 45 μ m. Insets: 2x magnification. Related to figure 3.

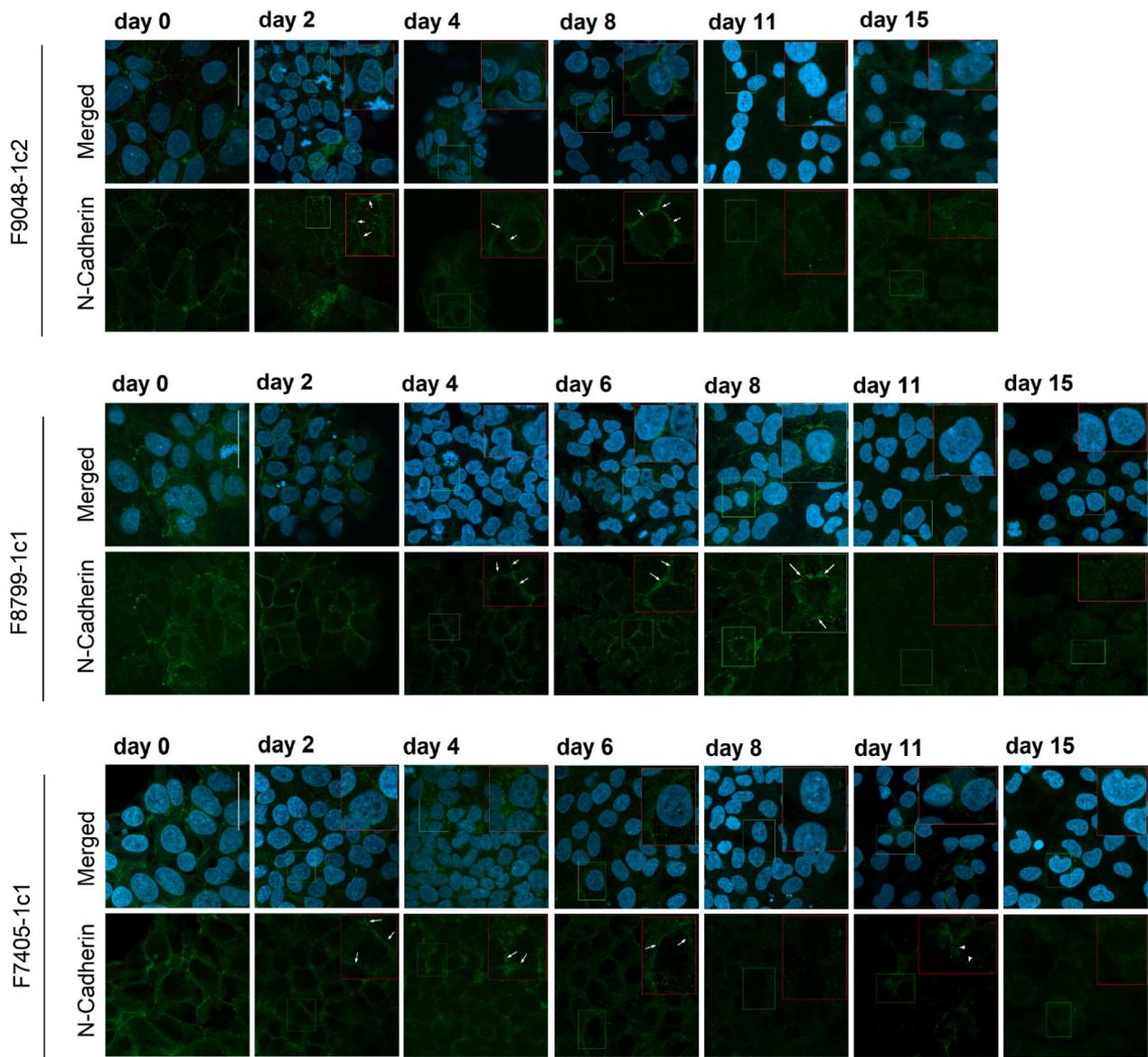


Figure S2. Time-series immunofluorescence analysis showing expression patterns for N-cadherin (*CDH2*) during NBC-to-NCC differentiation in all samples. Arrows indicate N-cadherin expression at cell-cell junctions. Arrowheads indicate membrane patches with diffuse N-cadherin expression. Scale bars = 45 μm . Insets: 2x magnification. Related to figure 3.

Table S1. Primer sequences

Target	Forward primer (5'-3')	Reverse Primer (5'-3')
<i>ACTB</i>	TGAAGTGTGACGTGGACATC	GGAGGAGCAATGATCTTGAT
<i>CDH1</i>	CCATTCACTACAACGCCCAACCC	CACAGTCACACACGCTGACCTC
<i>CDH11</i>	AGAGGTCCAATGTGGGAACG	GGTTGTCCTTCGAGGATACTGT
<i>CDH2</i>	TGCGGTACAGTGTAAGTGGG	GAAACCGGGCTATCTGCTCG
<i>CDH6</i>	AGCTGCAGTTTCAGCCGCGA	AGGGTATCTCTGCTCGCCTTCC
<i>CDH7</i>	TCAAATACATCTTGTGAGCGAA	TGGCATGAATATCCCCAGTGT
<i>DLX5</i>	ACCAGCCAGAGAAAGAAGTGAC	CCTTCTCTGTAATGCGGCCA
<i>ETS1</i>	TCAAGGACTATGTGCGGGAC	TTGGTCCACTGCCTGTGTAG
<i>FOXD3</i>	ATTTCTTTTCCCCTGAGCCC	TTCGGTTTTTCGGTTTTACCTG
<i>GAPDH</i>	ATCACCATCTTCCAGGAGCG	GGGCAGAGATGATGACCCTTT
<i>HMBS</i>	AGCTTGCTCGCATACAGACG	AGCTCCTTGGTAAACAGGCTT
<i>HPRT1</i>	CCTGGCGTCGTGATTAGTGAT	AGACGTTCAAGTCTGTCCATAA
<i>MSX2</i>	CACCCTGAGGAAACACAAGAC	TGCACGCTCTGCAATGGAG
<i>NANOG</i>	TGGACACTGGCTGAATCCTTC	CGTTGATTAGGCTCCAACCAT
<i>NGFR (P75)</i>	ACATAGCCTTCAAGAGGTGGA	TGTCGCTGTGGAGTTTTTCT
<i>OCT3/4</i>	GTGGTCAGCCAACTCGTCA	CCAAAACCCTGGCACAAACT
<i>PAX3</i>	AAGCCCAAGCAGGTGACAAC	CTCGGATTTCCAGCTGAAC
<i>PAX6</i>	TGTCCAACGGATGTGTGAG	TTTCCAAGCAAAGATGGAC
<i>SNAI2</i>	TCTGCGGCAAGGCGTTTTCCAG	GCAAATGCTCTGTTGCAGTGAGGG
<i>SOX2</i>	ATGTCCCAGCACTACCAGAGC	TTACTCTCCTCTTTTGCACCCC
<i>SOX9</i>	AGCGAACGCACATCAAGAC	CTGTAGGCGATCTGTTGGGG
<i>SO10</i>	GCCTTACCACTCCTATGACTCC	TCAAAGCTACTCTCAGCCCC
<i>TBP</i>	GTGACCCAGCATCACTGTTTC	GCAAACCAGAAACCCTTGCG
<i>TFAP2A</i>	CTCCGCCATCCCTATTAACAAG	GACCCGGAAGTGAACAGAAGA
<i>TWIST1</i>	CAATGACATCTAGGTCTCCGGGCC	TACGCCTTCTCGGTCTGGAGGATG
<i>VIM</i>	GACAACCTGGCCGAGGACATCATG	AGACGTGCCAGAGACGCATTGTC
<i>ZIC1</i>	AAGGTCCACGAATCCTCCTC	TTGTGGTCCGGTTGTCTG

CHAPTER VI

GENERAL DISCUSSION

Richieri-Costa-Pereira syndrome (RCPS) is a rare autosomal-recessive disorder that affects craniofacial and limb development and is reported almost entirely in Brazilian patients (Favaro et al. 2011). RCPS is caused by an increased number of repeat motifs within the *EIF4A3* 5'UTR (Favaro et al. 2014). Specifically, the noncoding 5' region presents a complex pattern of three distinct motifs varying in both size and organization among individuals. In chapter II, variability in the *EIF4A3* 5'UTR was characterized in a large cohort in order to investigate the mechanism that originated the RCPS-associated allelic pattern. This approach was important not only to clarify the origins of the pathogenic alleles but also to provide clues on the chance of RCPS arising in other populations. Despite advances in genome sequencing, repetitive regions remain difficult to sequence and noncoding regions lack detailed analysis in routine NGS pipelines. Indeed, the 5'UTR of *EIF4A3* is not covered in GnomAD database. Notably, RCPS is described in only two non-Brazilian patients to date, which may potentially be due to missing diagnoses of this condition leading to an underestimation of affected individuals, reinforcing the importance of characterizing this region and publishing results in an international journal. We have found 43 different variations amongst unaffected individuals, which shows the polymorphic nature and structural complexity of this region (Hsia et al. 2018). Based on these allele patterns and haplotype analyses, we suggest that RCPS-associated alleles have arisen from independent unequal crossing-over events between ancient alleles. The fact that the number of repeats in the *EIF4A3* 5'UTR seems to be stable across generations (Favaro et al. 2014) supports this hypothesis. RCPS etiology is comparable to synpolydactyly, in which polyalanine expansion may have resulted from unequal crossing over of *HOXD13* (Warren 1997), contradicting from several neurological disorders caused by unstable dynamic expansion mutations that have arisen only once and increase in size during meiotic divisions (Cummings and Zoghbi 2000; Gatchel and

Zoghbi 2005; Mirkin 2007; La Spada and Taylor 2010; McMurray 2010; Haeusler et al. 2016; DeJesus-Hernandez et al. 2011; Renton et al. 2011; Usdin et al. 2015; Virtaneva et al. 1997).

Furthermore, considering the restricted inclusion of CGCA-20nt motifs in the largest alleles along with the difference in allele pattern between Brazilian patients and the sole UK patient (Hsia et al. 2018), we propose that the pathogenic alleles have originated more than once and that RCPS can potentially occur in any population containing alleles with the disease-associated CGCA-20nt motif.

Additionally, we demonstrated a direct association between the number of motifs and *EIF4A3* expression, revealing a potential cis-acting regulatory mechanism for these repeat motifs (Hsia et al. 2018). The inverse correlation between number of motifs and expression levels suggests that the 5'UTR structure plays a central role in phenotypic modulation. Another interesting finding is that the disease-associated CGCA-20nt motif showed a reduction in expression compared to CACA-20nt and CA-18nt motifs, albeit not statistically significant, suggesting that both the number and sequence of the motif regulate *EIF4A3* expression. Indeed, a broad clinical variability is observed in RCPS patients. Specifically, individuals homozygous for the 16-repeat allele present more severe phenotypes than both the compound heterozygote carrying a 14-repeat allele *in trans* with a nucleotide change (Favaro et al. 2014) and the homozygote carrying the 14-repeat alleles, which was recently reported with a very mild phenotype (Bertola et al. 2018). Therefore, RCPS phenotypic variability may depend on the number and sequence of the motifs in the *EIF4A3* 5'UTR.

Since hypermethylation is not the mechanism by which the motifs repress *EIF4A3* expression in RCPS (Hsia et al. 2018), there could be alternative mechanisms responsible for this decreased expression that may involve creation of binding sites for repressors or post-translational events. Further studies are needed to better understand the mechanisms behind *EIF4A3* downregulation.

Next, the chapters III and IV described the functional evaluation of neural crest cells in order to unravel the pathological mechanisms of RCPS. Foremost, this work reinforced the applicability of induced pluripotent stem cell (iPSC)-based disease modeling in deciphering the etiology and pathogenetic mechanisms underlying human craniofacial disorders. Patient iPSC-

derived cell cultures have proven to be a very powerful and effective approach in contributing to our knowledge of embryonic development in both normal and disease contexts. Likewise, haploinsufficient mouse mutants were valuable for investigating *Eif4a3* functions in craniofacial development. The integrative use of iPSC technology with mouse models establishes a foundation for studying craniofacial malformations such as RCPS.

iPSCs highly expressed pluripotency markers, did not present aneuploidies, and demonstrated the ability to generate teratomas *in vivo*. iPSCs were successfully differentiated into NCCs which in turn exhibited p75(NRT)/HNK1 double-positive staining and high expression of NCC markers. RCPS patient-derived NCCs showed decreased *EIF4A3* mRNA and protein levels compared to controls. All NCC lineages were able to generate mesenchymal stem cells (nMSCs) with the typical immunophenotype, which gave rise to osteoblasts, chondrocytes and adipocytes (Miller et al. 2017). Taken together, these data demonstrated that our strategy efficiently yielded disease-relevant cell types, which allowed us to assess cellular and molecular phenotypes associated with RCPS *in vitro*. The mouse models recapitulated several anomalies seen in RCPS patients including mandibular and craniofacial abnormalities, indicating essential requirements of *EIF4A3* for NCC development. Both iPSC-derived cells and haploinsufficient mouse embryos contributed to clarify the NCC dysfunctions behind RCPS pathogenesis.

The cellular mechanisms by which reduced *EIF4A3* causes RCPS have been elucidated in chapter III (Miller et al. 2017). Patient-derived iPSCs differentiated toward a NCC lineage revealed defective migration and those differentiated into mesenchymal derivatives exhibited premature ossification and dysregulated chondrogenesis. Albeit some neurocristopathies have shown increased apoptosis and/or impaired proliferation involved in their pathogenesis (Devotta et al. 2016; Dixon et al. 2006; Brugmann et al. 2010), neither NCCs, nMSCs nor mouse models displayed commitment of either process. Cell death and cell cycle alteration was detectable only in extremely severe mouse mutants. Therefore, aberrant apoptosis, proliferation or mitosis delay are not major causes of RCPS. Depletion of *Eif4a3* in mice disrupted formation of Meckel's cartilage that is critically important for proper mandible morphogenesis and led to precocious clavicle ossification, which parallels observations made with human nMSCs.

We provided evidence that impaired NCC migratory capacity is a key cellular dysfunction involved in RCPS pathogenesis. Delayed migration at early developmental stages can reduce mesenchymal progenitors populating the pharyngeal arches and subsequently impair cartilage and bone formation leading to abnormal or absence of mandibular fusion, which is amongst the most relevant RCPS phenotypes. Migration defects seem to be unique to RCPS etiology, since they have not been observed in other craniofacial disorders (Lehalle et al. 2015; Mills and Green 2017; Jones et al. 2008; Dixon et al. 2006) despite the clinical overlap. Additionally, chondrogenesis and osteogenesis dysregulation have been demonstrated in both *in vitro* and *in vivo* models and may contribute to the mandibular and clavicular anomalies seen in RCPS patients.

Despite presenting phenotypes comparable to humans, whose *EIF4A3* levels are varied depending on cell type, *Eif4a3* haploinsufficient mice also displayed severe phenotypes not so far described in humans (Miller et al. 2017). These phenotypic differences could be due to Cre efficiency or because humans are genetically diverse whereas mice are inbred, which exposes phenotypes that are normally masked in humans. In addition, the reported high rates of spontaneous abortion in RCPS families (Favaro et al. 2011) suggest that severe phenotypes in humans, particularly brain malformations, may be embryonic lethal.

One interesting fact that deserves our attention is that while *EIF4A3* exerts basic and ubiquitous functions as a core component of exon junction complex (EJC) which is involved in many aspects of mRNA metabolism, mutations in this gene lead to very specific phenotypes. Indeed, this paradox does not happen only in RCPS, but also in other disorders such as Diamond-Blackfan anemia, Hoyeraal-Hreidarsson and Bowen-Conradi syndromes (Boria et al. 2010; Armistead et al. 2009; Heiss et al. 1998). Accordingly, these diverse disorders may potentially be related at a molecular level.

Chapter IV described the study of the molecular mechanisms by which *EIF4A3* depletion impairs NCC functions. To assess these mechanisms, we employed transcriptomic and proteomic analyses of RCPS patient-derived NCCs in order to expose candidate targets and pathways controlled by *EIF4A3* at the onset of craniofacial development. In addition, we used conditional mouse models for validation and further understanding of transcriptome data.

Gene set enrichment analysis (GSEA) of both human and mouse data revealed enrichment of extracellular matrix (ECM)-receptor interaction and focal adhesion pathway elements in response to decreased *EIF4A3* expression. Closer inspection demonstrated enrichment for GO categories of extracellular matrix and plasma membrane components. Optimal cell migration requires dramatic changes in cell shape and interaction with the ECM, which involves actin polymerization and continuous formation and disassembly of adhesions (Gardel et al. 2010). Focal adhesions (FA) are macromolecular assemblies that function as a physical connection between the cell and the ECM through interaction of integrin receptors and the actin cytoskeleton, mediated by numerous FA-associated proteins (Wu 2007). The interaction with ECM and focal adhesion features e.g. size, shape, and turnover dynamics, have long been associated with cell migration (Ridley et al. 2003; Kim and Wirtz 2013; Perris and Perissinotto 2000). Moreover, overexpression of *CAV1*, an identified DEG upregulated in RCPS patients, has been linked to decreased migration in tumor cells (Zhang et al. 2000). These findings suggest that impairment of cell adhesion dynamics is implicated in the cell migration defects underlying RCPS pathogenesis. However, future experiments will be needed for further clarifying the involvement of *EIF4A3* on migration.

Several human genetic conditions that manifest as tissue-specific phenotypes restricted to craniofacial structures are caused by mutations affecting mRNA splicing and ribosome biogenesis, known collectively as spliceopathies and ribosomopathies, respectively (Danilova and Gazda 2015; Yelick and Trainor 2015; Armistead and Triggs-Raine 2014; Mills and Green 2017; Lehalle et al. 2015). Treacher Collins and Bowen-Conradi syndromes are examples of such conditions (Choesmel et al. 2007; Sakai and Trainor 2009; Armistead et al. 2009; Jones et al. 2008), and notably both have phenotype overlap with RCPS. Ribosomes appear to be especially sensitive to deficits in EJC components, since we identified aberrant splicing as well as alterations at the proteomic level of ribosomal components in *EIF4A3* deficient hNCCs. Furthermore, haploinsufficiency for each of the core EJC components (i.e. Magoh, Eif4a3 and Rbm8a) resulted in altered expression of ribosomal components in mouse brains (Mao et al. 2016). These data suggest that ribosome defects contribute to RCPS pathogenesis, reinforcing the idea that the EJC is a strong regulator of protein homeostasis machinery.

Bearing in mind that *EIF4A3* is super-stoichiometric to other EJC members in mammalian cells (Singh et al. 2015; Singh et al. 2012) and ribosome composition is heterogeneous and may reflect tissue requirements to translate specific mRNAs (Shi et al. 2017), we hypothesized that the tissue-specific defects in RCPS may be caused by insufficient *EIF4A3* expression in tissues with high EJC demand, leading to aberrant splicing of specific riboproteins. However, how *EIF4A3* influences ribosomal components and whether this is a direct effect is still unknown and may be an important question for future studies. In fact, NSC-specific *Eif4a3* haploinsufficiency alters expression of ribosomal components and triggers aberrant p53 activation leading to microcephaly in mouse model (Mao et al. 2016). Interestingly, microcephaly has recently been included in the phenotypic spectrum of RCPS (Bertola et al. 2018). Given these connections, it is fascinating to consider that *EIF4A3* regulates ribosome biogenesis and consequently contributes to RCPS pathology.

This work was designed with a focus on neural crest cells and craniofacial development although RCPS is also characterized by limb anomalies. The limb bud has different origins than craniofacial structures (Zuniga 2015) and *Wnt1*-Cre mouse models are not useful in the investigation of limb development since Cre is not active in the developing limb (Lewis et al. 2013). In the future, it will be of interest to employ other Cre drivers to investigate additional, relevant RCPS phenotypes.

Finally, owing to the importance of neural crest cells in embryonic development, in parallel with aforementioned studies we developed a simple method to recapitulate developmental events involved in NCC specification from neural plate border (NPB)-like cells.

NCC development depends upon a tightly controlled spatial and temporal molecular program involving a sequential activation and/or repression of multiple molecules (Knecht and Bronner-Fraser 2002). Anomalies in these developmental steps cause neurocristopathies (NCP), frequently encompassing craniofacial abnormalities (Vega-Lopez et al. 2018). A central process which orchestrates transitioning from an epithelial to a mesenchymal cell phenotype during NCC development is designated as the epithelial to mesenchymal transition (EMT). EMT drives the onset of NCC migration and subsequent tissue morphogenesis and is associated with the

segregation of homogeneous precursors into distinct fates (Kalcheim 2015; Theveneau and Mayor 2012).

In view of the limited access to human embryonic tissues and the fact that animal embryos may not completely recapitulate human embryonic development (Thyagarajan et al. 2003; de Bakker et al. 2016; Barriga et al. 2015), human iPSC-based strategies have been largely used as an alternative approach to study NCC biology in pathological conditions. The available methods for generating NCCs rely on BMP pathway modulation (Jiang et al. 2009; Lee et al. 2010; Pomp et al. 2005; Lee et al. 2007; Curchoe et al. 2010) and may or may not be combined with WNT pathway activation (Menendez et al. 2013; Fukuta et al. 2014; Miller et al. 2017; Menendez et al. 2011; Noisa et al. 2014; Huang et al. 2016; Liu et al. 2012; Lee et al. 2009; Mica et al. 2013; Kreitzer et al. 2013; Leung et al. 2016; Thier et al. 2019; Tchieu et al. 2017). However, many of these methods lack a NPB stage or are unable to fully convert neuroepithelial cell populations into NCCs, potentially hindering investigation of the various stages of neural crest specification. Thus, in chapter V we showed that neuroepithelial induction of hiPSCs without altering BMP signaling (Lippmann et al. 2014) confers a NPB-like transcriptional state to differentiating cells, which can be directed towards the neural crest fate with near 100% efficiency, recapitulating for the first time the sequential steps that orchestrate NCC specification and EMT.

This method provided an efficient *in vitro* human model to investigate the functional effects of candidate genetic variants and a promising approach to better understand human development and neural crest-related disorders.

REFERENCES

- Armistead, J., Khatkar, S., Meyer, B., et al. 2009. Mutation of a gene essential for ribosome biogenesis, *EMG1*, causes Bowen-Conradi syndrome. *American Journal of Human Genetics* 84(6), pp. 728–739.
- Armistead, J. and Triggs-Raine, B. 2014. Diverse diseases from a ubiquitous process: the ribosomopathy paradox. *FEBS Letters* 588(9), pp. 1491–1500.
- de Bakker, B.S., de Jong, K.H., Hagoort, J., et al. 2016. An interactive three-dimensional digital atlas and quantitative database of human development. *Science* 354(6315).
- 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(9), pp. 1555–1560.
- Bertola, D.R., Hsia, G., Alvizi, L., et al. 2018. Richieri-Costa-Pereira syndrome: Expanding its phenotypic and genotypic spectrum. *Clinical Genetics* 93(4), pp. 800–811.
- Boria, I., Garelli, E., Gazda, H.T., et al. 2010. The ribosomal basis of Diamond-Blackfan Anemia: mutation and database update. *Human Mutation* 31(12), pp. 1269–1279.
- Brugmann, S.A., Allen, N.C., James, A.W., Mekonnen, Z., Madan, E. and Helms, J.A. 2010. A primary cilia-dependent etiology for midline facial disorders. *Human Molecular Genetics* 19(8), pp. 1577–1592.
- Choismel, V., Bacqueville, D., Rouquette, J., et al. 2007. Impaired ribosome biogenesis in Diamond-Blackfan anemia. *Blood* 109(3), pp. 1275–1283.
- Cummings, C.J. and Zoghbi, H.Y. 2000. Fourteen and counting: unraveling trinucleotide repeat diseases. *Human Molecular Genetics* 9(6), pp. 909–916.
- Curchoe, C.L., Maurer, J., McKeown, S.J., et al. 2010. Early acquisition of neural crest competence during hESCs neuralization. *Plos One* 5(11), p. e13890.
- Danilova, N. and Gazda, H.T. 2015. Ribosomopathies: how a common root can cause a tree of pathologies. *Disease Models & Mechanisms* 8(9), pp. 1013–1026.
- DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., et al. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72(2), pp. 245–256.
- Devotta, A., Juraver-Geslin, H., Gonzalez, J.A., Hong, C.-S. and Saint-Jeannet, J.-P. 2016. Sf3b4-depleted *Xenopus* embryos: A model to study the pathogenesis of craniofacial defects in Nager syndrome. *Developmental Biology* 415(2), pp. 371–382.
- Dixon, J., Jones, N.C., Sandell, L.L., et al. 2006. *Tcof1/Treacle* is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proceedings of the National Academy of Sciences of the United States of America* 103(36), pp. 13403–13408.
- Favaro, F.P., Alvizi, L., Zechi-Ceide, R.M., et al. 2014. A noncoding expansion in *EIF4A3* causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. *American Journal of Human Genetics* 94(1), pp. 120–128.
- Favaro, F.P., Zechi-Ceide, R.M., Alvarez, C.W., et al. 2011. Richieri-Costa-Pereira syndrome: a unique acrofacial dysostosis type. An overview of the Brazilian cases. *American Journal of Medical Genetics. Part A* 155A(2), pp. 322–331.
- Fukuta, M., Nakai, Y., Kirino, K., et al. 2014. Derivation of mesenchymal stromal cells from pluripotent stem cells through a neural crest lineage using small molecule compounds with defined media. *Plos One* 9(12), p. e112291.
- Gardel, M.L., Schneider, I.C., Aratyn-Schaus, Y. and Waterman, C.M. 2010. Mechanical integration of actin and adhesion dynamics in cell migration. *Annual Review of Cell and Developmental Biology* 26, pp. 315–333.
- Gatchel, J.R. and Zoghbi, H.Y. 2005. Diseases of unstable repeat expansion: mechanisms and common principles. *Nature Reviews. Genetics* 6(10), pp. 743–755.

Haeusler, A.R., Donnelly, C.J. and Rothstein, J.D. 2016. The expanding biology of the C9orf72 nucleotide repeat expansion in neurodegenerative disease. *Nature Reviews. Neuroscience* 17(6), pp. 383–395.

Heiss, N.S., Knight, S.W., Vulliamy, T.J., et al. 1998. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nature Genetics* 19(1), pp. 32–38.

Hsia, G.S.P., Musso, C.M., Alvizi, L., et al. 2018. Complexity of the 5' untranslated region of EIF4A3, a critical factor for craniofacial and neural development. *Frontiers in genetics* 9, p. 149.

Huang, M., Miller, M.L., McHenry, L.K., et al. 2016. Generating trunk neural crest from human pluripotent stem cells. *Scientific reports* 6, p. 19727.

Jiang, X., Gwyne, Y., McKeown, S.J., Bronner-Fraser, M., Lutzko, C. and Lawlor, E.R. 2009. Isolation and characterization of neural crest stem cells derived from in vitro-differentiated human embryonic stem cells. *Stem Cells and Development* 18(7), pp. 1059–1070.

Jones, N.C., Lynn, M.L., Gaudenz, K., et al. 2008. Prevention of the neurocristopathy Treacher Collins syndrome through inhibition of p53 function. *Nature Medicine* 14(2), pp. 125–133.

Kalchauer, C. 2015. Epithelial-Mesenchymal Transitions during Neural Crest and Somite Development. *Journal of clinical medicine* 5(1).

Kan, L., Israsena, N., Zhang, Z., et al. 2004. Sox1 acts through multiple independent pathways to promote neurogenesis. *Developmental Biology* 269(2), pp. 580–594.

Kim, D.-H. and Wirtz, D. 2013. Focal adhesion size uniquely predicts cell migration. *The FASEB Journal* 27(4), pp. 1351–1361.

Knecht, A.K. and Bronner-Fraser, M. 2002. Induction of the neural crest: a multigene process. *Nature Reviews. Genetics* 3(6), pp. 453–461.

de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A. and Pollock, D.D. 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genetics* 7(12), p. e1002384.

Kreitzer, F.R., Salomonis, N., Sheehan, A., et al. 2013. A robust method to derive functional neural crest cells from human pluripotent stem cells. *American journal of stem cells* 2(2), pp. 119–131.

La Spada, A.R. and Taylor, J.P. 2010. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nature Reviews. Genetics* 11(4), pp. 247–258.

Lee, G., Chambers, S.M., Tomishima, M.J. and Studer, L. 2010. Derivation of neural crest cells from human pluripotent stem cells. *Nature Protocols* 5(4), pp. 688–701.

Lee, G., Kim, H., Elkabetz, Y., et al. 2007. Isolation and directed differentiation of neural crest stem cells derived from human embryonic stem cells. *Nature Biotechnology* 25(12), pp. 1468–1475.

Lee, G., Papapetrou, E.P., Kim, H., et al. 2009. Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* 461(7262), pp. 402–406.

Lehalle, D., Wiczorek, D., Zechi-Ceide, R.M., et al. 2015. A review of craniofacial disorders caused by spliceosomal defects. *Clinical Genetics* 88(5), pp. 405–415.

Leung, A.W., Murdoch, B., Salem, A.F., Prasad, M.S., Gomez, G.A. and García-Castro, M.I. 2016. WNT/ β -catenin signaling mediates human neural crest induction via a pre-neural border intermediate. *Development* 143(3), pp. 398–410.

Lewis, A.E., Vasudevan, H.N., O'Neill, A.K., Soriano, P. and Bush, J.O. 2013. The widely used Wnt1-Cre transgene causes developmental phenotypes by ectopic activation of Wnt signaling. *Developmental Biology* 379(2), pp. 229–234.

Lippmann, E.S., Estevez-Silva, M.C. and Ashton, R.S. 2014. Defined human pluripotent stem cell culture enables highly efficient neuroepithelium derivation without small molecule inhibitors. *Stem Cells* 32(4), pp. 1032–1042.

Liu, Q., Spusta, S.C., Mi, R., et al. 2012. Human neural crest stem cells derived from human ESCs and induced pluripotent stem cells: induction, maintenance, and differentiation into functional schwann cells. *Stem cells translational medicine* 1(4), pp. 266–278.

Mao, H., McMahon, J.J., Tsai, Y.-H., Wang, Z. and Silver, D.L. 2016. Haploinsufficiency for Core Exon

Junction Complex Components Disrupts Embryonic Neurogenesis and Causes p53-Mediated Microcephaly. *PLoS Genetics* 12(9), p. e1006282.

McMurray, C.T. 2010. Mechanisms of trinucleotide repeat instability during human development. *Nature Reviews. Genetics* 11(11), pp. 786–799.

Menendez, L., Kulik, M.J., Page, A.T., et al. 2013. Directed differentiation of human pluripotent cells to neural crest stem cells. *Nature Protocols* 8(1), pp. 203–212.

Menendez, L., Yatskievych, T.A., Antin, P.B. and Dalton, S. 2011. Wnt signaling and a Smad pathway blockade direct the differentiation of human pluripotent stem cells to multipotent neural crest cells. *Proceedings of the National Academy of Sciences of the United States of America* 108(48), pp. 19240–19245.

Mica, Y., Lee, G., Chambers, S.M., Tomishima, M.J. and Studer, L. 2013. Modeling neural crest induction, melanocyte specification, and disease-related pigmentation defects in hESCs and patient-specific iPSCs. *Cell reports* 3(4), pp. 1140–1152.

Miller, E.E., Kobayashi, G.S., Musso, C.M., et al. 2017. EIF4A3 deficient human iPSCs and mouse models demonstrate neural crest defects that underlie Richieri-Costa-Pereira syndrome. *Human Molecular Genetics* 26(12), pp. 2177–2191.

Mills, E.W. and Green, R. 2017. Ribosomopathies: There's strength in numbers. *Science* 358(6363).

Mirkin, S.M. 2007. Expandable DNA repeats and human disease. *Nature* 447(7147), pp. 932–940.

Noisa, P., Lund, C., Kanduri, K., et al. 2014. Notch signaling regulates the differentiation of neural crest from human pluripotent stem cells. *Journal of Cell Science* 127(Pt 9), pp. 2083–2094.

Perris, R. and Perissinotto, D. 2000. Role of the extracellular matrix during neural crest cell migration. *Mechanisms of Development* 95(1–2), pp. 3–21.

Pomp, O., Brokhman, I., Ben-Dor, I., Reubinoff, B. and Goldstein, R.S. 2005. Generation of peripheral sensory and sympathetic neurons and neural crest cells from human embryonic stem cells. *Stem Cells* 23(7), pp. 923–930.

Renton, A.E., Majounie, E., Waite, A., et al. 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72(2), pp. 257–268.

Ridley, A.J., Schwartz, M.A., Burridge, K., et al. 2003. Cell migration: integrating signals from front to back. *Science* 302(5651), pp. 1704–1709.

Sakai, D. and Trainor, P.A. 2009. Treacher Collins syndrome: unmasking the role of Tcof1/treacle. *The International Journal of Biochemistry & Cell Biology* 41(6), pp. 1229–1232.

Shi, Z., Fujii, K., Kovary, K.M., et al. 2017. Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Molecular Cell* 67(1), p. 71–83.e7.

Singh, G., Kucukural, A., Cenik, C., et al. 2012. The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* 151(4), pp. 750–764.

Singh, G., Pratt, G., Yeo, G.W. and Moore, M.J. 2015. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annual Review of Biochemistry* 84, pp. 325–354.

Tchieu, J., Zimmer, B., Fattahi, F., et al. 2017. A Modular Platform for Differentiation of Human PSCs into All Major Ectodermal Lineages. *Cell Stem Cell* 21(3), p. 399–410.e7.

Theveneau, E. and Mayor, R. 2012. Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. *Developmental Biology* 366(1), pp. 34–54.

Thier, M.C., Hommerding, O., Panten, J., et al. 2019. Identification of Embryonic Neural Plate Border Stem Cells and Their Generation by Direct Reprogramming from Adult Human Blood Cells. *Cell Stem Cell* 24(1), p. 166–182.e13.

Thyagarajan, T., Totey, S., Danton, M.J.S. and Kulkarni, A.B. 2003. Genetically altered mouse models: the good, the bad, and the ugly. *Critical Reviews in Oral Biology and Medicine* 14(3), pp. 154–174.

Usdin, K., House, N.C.M. and Freudenreich, C.H. 2015. Repeat instability during DNA repair: Insights from model systems. *Critical Reviews in Biochemistry and Molecular Biology* 50(2), pp. 142–167.

- Vega-Lopez, G.A., Cerrizuela, S., Tribulo, C. and Aybar, M.J. 2018. Neurocristopathies: New insights 150 years after the neural crest discovery. *Developmental Biology* 444 Suppl 1, pp. S110–S143.
- Virtaneva, K., D’Amato, E., Miao, J., et al. 1997. Unstable minisatellite expansion causing recessively inherited myoclonus epilepsy, EPM1. *Nature Genetics* 15(4), pp. 393–396.
- Warren, S.T. 1997. Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13. *Science* 275(5298), pp. 408–409.
- Wu, C. 2007. Focal adhesion: a focal point in current cell biology and molecular medicine. *Cell Adhesion & Migration* 1(1), pp. 13–18.
- Yelick, P.C. and Trainor, P.A. 2015. Ribosomopathies: Global process, tissue specific defects. *Rare diseases (Austin, Tex.)* 3(1), p. e1025185.
- Zhang, W., Razani, B., Altschuler, Y., et al. 2000. Caveolin-1 inhibits epidermal growth factor-stimulated lamellipod extension and cell migration in metastatic mammary adenocarcinoma cells (MTLn3). Transformation suppressor effects of adenovirus-mediated gene delivery of caveolin-1. *The Journal of Biological Chemistry* 275(27), pp. 20717–20725.
- Zuniga, A. 2015. Next generation limb development and evolution: old questions, new perspectives. *Development* 142(22), pp. 3810–3820.

CONCLUSIONS

In conclusion, we unraveled novel mechanisms by which *EIF4A3* hypomorphic mutations lead to craniofacial phenotypes observed in RCPS patients.

In summary, we revealed that:

- RCPS-associated alleles have arisen from unequal crossing-over events between ancient alleles at least twice;
- Both the number and sequence of repeat motifs within *EIF4A3* 5'UTR regulate gene expression;
- *EIF4A3* is required for NCC development and its deficiency leads to decreased NCC migratory capacity and compromised osteogenic and chondrogenic differentiation of NCC-mesenchymal derivatives;
- *EIF4A3* deficient NCCs present altered expression of cell-ECM interaction and FA components, which can be implicated in the defective NCC migration;
- Aberrant splicing of ribosomal components contributes to RCPS pathogenesis;

Furthermore, we explored previously published protocols to establish a simple method for studying early NCC development.

- iPSC-derived NCCs were efficiently generated in a manner that recapitulates developmental events involved in NCC specification and EMT from the NPB onward.

CHAPTER VII

ABSTRACT

Richieri-Costa-Pereira syndrome (RCPS) is a rare autosomal-recessive disorder characterized by craniofacial abnormalities, including mandible cleft, microstomia, Robin sequence and microcephaly, as well as limb defects and learning impairments. RCPS is mainly caused by an increased number of repeat motifs within the *EIF4A3* 5'UTR, which in turn leads to decreased expression of the gene product. The DEAD-box RNA helicase eIF4A3 is a core component of the RNA-binding exon junction complex (EJC), which is involved in post-transcriptional events such as alternative splicing, nonsense-mediated mRNA decay (NMD), translation initiation and mRNA localization. The *EIF4A3* 5'UTR varies in both number and organization of three types of motifs between individuals. However, the origin of the RCPS-associated allelic pattern, as well as the functional effects of these motifs on *EIF4A3* expression, remain to be uncovered. Although a relationship between *EIF4A3* hypomorphic biallelic mutations and RCPS has been established, the pathogenetic mechanisms by which decreased levels of *EIF4A3* lead to craniofacial malformation are unknown. To address these gaps, we first characterized the variation in the *EIF4A3* 5'UTR at a populational level. This analysis demonstrated that this noncoding region displays a polymorphic nature and structural complexity presenting multiple patterns. The RCPS-associated allele patterns may have arisen from independent unequal crossing-over events between ancient alleles and can potentially emerge in any population containing alleles with the CGCA-20nt motif. Furthermore, there is a direct association between the number of motifs and *EIF4A3* expression, revealing a potential cis-acting regulatory mechanism for these motifs and suggesting that the 5'UTR structure plays a central role in phenotypic modulation. Next, we unraveled the cellular and molecular mechanisms responsible for RCPS using two complementary models, patient-derived induced pluripotent stem cells (iPSCs) and *EIF4A3* haploinsufficient mouse models. The craniofacial structures compromised in RCPS patients are suggestive of

disturbances in neural crest cells (NCCs), a transient cell population that originates at the neural plate border (NPB) in the developing embryo and gives rise to multiple cell types, generating most of the cranium. Accordingly, we differentiated iPSCs from RCPS patients and control individuals into NCCs and demonstrated that *EIF4A3* deficiency impairs NCC development, leading to defective migration, premature osteogenic and dysregulated chondrogenic differentiation of NCC mesenchymal derivatives. Besides pinpointing that impaired NCC migratory capacity is a key cellular dysfunction underlying RCPS pathogenesis, we provided evidence by transcriptome analyses that impairment to cell adhesion dynamics is implicated in this dysfunction and involves alteration of cell-extracellular matrix (ECM) interaction components. Additionally, we suggest that ribosome defects contribute to RCPS pathogenesis, since aberrant splicing as well as alterations at the proteomic level of ribosomal components were identified in patient-derived NCCs. Elucidating the pathogenetic mechanism underlying RCPS will also aid in clarifying the etiology of other craniofacial syndromes and how mutations in genes with basic and ubiquitous functions such as *EIF4A3* lead to specific phenotypes. Finally, we provided a simple protocol for quickly deriving human NCCs *in vitro* in a manner that recapitulates multiple stages from the early NPB onward. This method represents a promising approach to better understand human craniofacial development.

RESUMO

A síndrome Richieri-Costa-Pereira (SRCP) é uma doença autossômica-recessiva rara caracterizada por anomalias craniofaciais, incluindo fissura de mandíbula, microstomia, seqüência Robin e microcefalia, assim como malformação dos membros e dificuldade de aprendizado. A SRCP é principalmente causada por um aumento no número de motivos repetitivos na região 5' não traduzida (5'UTR) do gene *EIF4A3*, o que por sua vez leva à diminuição da expressão do produto gênico. A DEAD-box RNA helicase eIF4A3 é um dos principais componentes do complexo de junção de éxons (EJC) que se liga ao RNA e está envolvido em eventos pós-transcricionais incluindo *splicing* alternativo, decaimento de RNAm mediado por mutações sem sentido (NMD), iniciação da tradução e localização de RNAm. A 5'UTR do gene *EIF4A3* varia em número e organização de três tipos de motivos entre indivíduos. Entretanto, a origem do padrão alélico associado à SRCP assim como o efeito funcional desses motivos na expressão de *EIF4A3* continua a ser descoberto. Embora a relação entre mutações bialélicas hipomórficas em *EIF4A3* e SRCP foi estabelecida, os mecanismos patogênicos pelos quais a diminuição de *EIF4A3* leva a malformação craniofacial são desconhecidos. Para responder essas questões, nós primeiramente caracterizamos a 5'UTR do *EIF4A3* em nível populacional. Essa análise demonstrou que essa região não codificante apresenta uma natureza polimórfica e estruturalmente complexa, exibindo múltiplos padrões. O padrão alélico associado à SRCP parece ter surgido a partir de eventos independentes de *crossing-over* desigual entre alelos ancestrais e pode potencialmente surgir em qualquer população contendo alelos com o motivo CGCA-20nt. Além disso, existe uma associação direta entre o número de motivos e a expressão de *EIF4A3*, revelando um potencial mecanismo regulatório de ação cis para esses motivos e sugerindo que a estrutura 5'UTR exerce um papel central na modulação fenotípica. Em seguida, nós revelamos os mecanismos celulares e moleculares responsáveis pela SRCP usando dois modelos complementares, células-tronco pluripotentes induzidas (iPSCs) de pacientes e controles e camundongos haploinsuficientes para *EIF4A3*. As estruturas craniofaciais comprometidas nos pacientes acometidos pela SRCP são sugestivas de distúrbios em células de crista neural (NCCs), uma população de células transiente

que se origina a partir da borda da placa neural (NPB) no embrião em desenvolvimento e dá origem a múltiplos tipos celulares, gerando a maior parte do esqueleto craniano. Então, nós diferenciamos iPSCs de pacientes acometidos pela SRCP e indivíduos controles em NCCs e demonstramos que deficiência de *EIF4A3* prejudica o desenvolvimento de NCCs, levando à migração defeituosa, diferenciação osteogênica prematura e condrogênica desregulada dos derivados mesenquimais de NCCs. Além de apontar que a capacidade de migração prejudicada em NCCs é uma disfunção chave na patogênese de RCPS, nós providenciamos evidência por meio de análise de transcriptoma que o comprometimento da dinâmica de adesão celular está implicado nessa disfunção e envolve componentes de interação das células com a matriz extracelular (ECM). Ademais, nós sugerimos que defeitos nos ribossomos contribuem para a patogênese da SRCP, desde que *splicing* aberrante assim como alteração nos níveis proteicos de componentes ribossomais foram identificados em NCCs derivadas de pacientes. Elucidando os mecanismos patogênicos envolvidos na SRCP também auxilia no entendimento da etiologia de outras síndromes craniofaciais e de como mutações em genes de função básica e ubíqua como o *EIF4A3* leva a fenótipos específicos. Finalmente, nós desenvolvemos um protocolo simples para a rápida geração de NCCs *in vitro* de uma maneira que recapitule os múltiplos estágios do desenvolvimento das NCCs desde a NPB. Esse método representa uma abordagem promissora para o melhor entendimento do desenvolvimento craniofacial humano.

APPENDIX

ADDITIONAL PUBLICATIONS

- I- Savastano CP, Brito LA, Faria AC, Setó-Salvia N, Peskett E, **Musso CM**, Alvizi L, Ezquina SAM, James C, GOSgene, Beales P, Lees M, Moore GE, Stanier P, Passos-Bueno MR. Impact of rare variants in *ARHGAP29* to the etiology of oral clefts: role of loss-of-function vs missense variants. **Clinical Genetics** 91:683-689, 2017.
<https://doi.org/10.1111/cge.12823>

ARHGAP29 has emerged as a candidate gene associated to non-syndromic cleft lip with or without cleft palate (NSCL/P), a complex congenital malformation. In this study, we assessed the clinical impact of rare variants in this gene in a cohort of familial NSCL/P cases. A significant mutational burden was found for loss-of-function (LoF) variants but not for missense variants in *ARHGAP29*. This data pinpointed rare variants leading to haploinsufficiency of *ARHGAP29* as an important etiological clefting mechanism.

Original Article

Impact of rare variants in *ARHGAP29* to the etiology of oral clefts: role of loss-of-function vs missense variants

Savastano C.P., Brito L.A., Faria Á.C., Setó-Salvia N., Peskett E., Musso C.M., Alvizi L., Ezquina S.A.M., James C., GOSgene, Beales P., Lees M., Moore G.E., Stanier P., Passos-Bueno M.R.. Impact of rare variants in *ARHGAP29* to the etiology of oral clefts: role of loss-of-function vs missense variants.

Clin Genet 2017; 91: 683–689. © John Wiley & Sons A/S. Published by John Wiley & Sons Ltd, 2016

Non-syndromic cleft lip with or without cleft palate (NSCL/P) is a prevalent, complex congenital malformation. Genome-wide association studies (GWAS) on NSCL/P have consistently identified association for the 1p22 region, in which *ARHGAP29* has emerged as the main candidate gene. *ARHGAP29* re-sequencing studies in NSCL/P patients have identified rare variants; however, their clinical impact is still unclear. In this study we identified 10 rare variants in *ARHGAP29*, including five missense, one in-frame deletion, and four loss-of-function (LoF) variants, in a cohort of 188 familial NSCL/P cases. A significant mutational burden was found for LoF (Sequence Kernel Association Test, $p=0.0005$) but not for missense variants in *ARHGAP29*, suggesting that only LoF variants contribute to the etiology of NSCL/P. Penetrance was estimated as 59%, indicating that heterozygous LoF variants in *ARHGAP29* confer a moderate risk to NSCL/P. The GWAS hits in *IRF6* (rs642961) and 1p22 (rs560426 and rs4147811) do not seem to contribute to the penetrance of the phenotype, based on co-segregation analysis. Our data show that rare variants leading to haploinsufficiency of *ARHGAP29* represent an important etiological clefting mechanism, and genetic testing for this gene might be taken into consideration in genetic counseling of familial cases.

Conflict of interest

The authors have declared no conflicting interests.

C.P. Savastano^a, L.A. Brito^a, Á.C. Faria^a, N. Setó-Salvia^{b,c}, E. Peskett^b, C.M. Musso^a, L. Alvizi^a, S.A.M. Ezquina^a, C. James^b, GOSgene^b, P. Beales^b, M. Lees^d, G.E. Moore^b, P. Stanier^b and M.R. Passos-Bueno^a

^aCentro de Pesquisa sobre o Genoma Humano e Células-Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil, ^bGenetics and Genomic Medicine, UCL Institute of Child Health, London, UK, ^cDepartment of Molecular Neuroscience, UCL Institute of Neurology, London, UK, and ^dDepartment of Clinical Genetics, Great Ormond Street Children's Hospital, London, UK

Key words: cleft lip and palate – GWAS – haploinsufficiency – *IRF6* – nonsense mutations – penetrance – rare variants – 1p22

Corresponding authors: Maria Rita Passos-Bueno, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências da USP, Centro de Pesquisa sobre o Genoma Humano e Células-Tronco, Universidade de São Paulo, Rua do Matão 277, sala 200, Cidade Universitária, São Paulo, SP 05508-090, Brazil.
Tel.: +55 11 3091 7740
e-mail: passos@ib.usp.br
and
Philip Stanier, Genetics and Genomic Medicine, UCL Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK.
Tel.: +44 207 905 2867;
Fax: +44 (0) 207 905 2953
e-mail: p.stanier@ucl.ac.uk

Received 3 May 2016, revised and accepted for publication 16 June 2016

- II- Alves LU, Santos S, **Musso CM**, Ezquina SAM, Opitz JM, Kok F, Otto PA, Mingroni-Netto RC. Santos syndrome is caused by mutation in the WNT7A gene. **Journal of Human Genetics** 62:1073-1078, 2017.
<https://doi.ORG/10.1038/jhg.2017.86>

Santos syndrome (SS) is mainly characterized by fibular agenesis/hypoplasia, hypoplastic femora, clubfeet, oligodactyly, and ungual hypoplasia/anonychia. In this study, we used linkage studies and exome sequencing to describe a novel variant in *WNT7A* as the cause of SS.

ORIGINAL ARTICLE

Santos syndrome is caused by mutation in the *WNT7A* gene

Leandro U Alves¹, Silvana Santos², Camila M Musso¹, Suzana AM Ezquina¹, John M Opitz³, Fernando Kok⁴, Paulo A Otto¹ and Regina C Mingroni-Netto¹

We have recently described a family with a condition (Santos syndrome (SS; MIM 613005)) characterized by fibular agenesis/hypoplasia, hypoplastic femora and grossly malformed/deformed clubfeet with severe oligodactyly, unguis hypoplasia/anonychia, sometimes associated with mild brachydactyly and occasional pre-axial polydactyly. Autosomal dominant inheritance with incomplete penetrance was suggested, but autosomal recessive inheritance could not be ruled out, due to the high frequency of consanguineous matings in the region where the family lived. This report deals with linkage studies and exome sequencing, disclosing a novel variant in *WNT7A*, c.934G>A (p.Gly312Ser), as the cause of this syndrome. This variant was present in homozygous state in five individuals typically affected by the SS syndrome, and in heterozygous state in the son of one affected homozygous individual. The heterozygous boy presented only unilateral complex polysyndactyly and we hypothesize that he either presents a distinct defect or that his phenotype results from a rare, mild clinical manifestation of the variant in heterozygous state. Variants in *WNT7A* are known to cause at least two other limb defect disorders, the syndromes of Fuhrmann and Al-Awadi/Raas-Rothschild. Despite their variable degree of expressivity and overlap, the three related conditions can be differentiated phenotypically in most instances.

Journal of Human Genetics (2017) 62, 1073–1078; doi:10.1038/jhg.2017.86; published online 31 August 2017

INTRODUCTION

In 2008, we described a new hereditary condition, presently known as 'Santos syndrome' (SS; MIM 613005), comprising asymmetric lower limb defects with hypoplastic femora, fibular agenesis/hypoplasia, grossly malformed/deformed clubfeet with severe oligodactyly; upper limbs with acromial dimples, limited extension, flexion, pronation, and supination of forearm and hands, severe unguis hypoplasia/anonychia, sometimes associated with mild brachydactyly, and occasionally pre-axial polydactyly.¹

The condition was compared with two syndromes presenting fibular agenesis associated with other defects/deformities: brachydactyly–ectrodactyly with fibular 'aplasia' or hypoplasia (MIM 113310)^{2–4} and Fuhrmann syndrome (FS; MIM 228930).^{5–8} Its clinical presentation did not perfectly fit the reported phenotypes of either conditions or those of other phenotypically related syndromes.

Pedigree analysis suggested autosomal dominant inheritance with a low penetrance, but the possibility of autosomal recessive inheritance could not be discarded because of the high regional population inbreeding rate, as already stated clearly on the 2008 paper. The frequencies of consanguineous matings in the region range from 9 to 33%, with corresponding average inbreeding coefficients (*F*) varying from 0.3 to 1.0%.^{9,10}

Below we report on the identification of the genetic variant responsible for the syndrome.

MATERIALS AND METHODS

The research was approved by the Ethics Committee of the Institute of Biosciences at the University of São Paulo (CAAE: 37297114.7.0000.5464).

Patients

This study covers 12 individuals from the family with SS, originally described by Santos *et al.*¹ five individuals (IV:1, IV:16; IV:19; IV:20 and IV:31) with the complete SS phenotype, one individual (V:12) with a milder and atypical phenotype (unilateral polysyndactyly without lower limb defects) and six unaffected individuals (IV:17, IV:29, IV:32, V:13, V:14 and V:15). Genomic DNA was extracted from blood samples. Pedigree is shown in Figure 1 and clinical findings are summarized in Table 1.

Array-based comparative genomic hybridization

Investigation of copy number changes was performed by array-based comparative genomic hybridization (array-CGH), using a 180 K chip (OGT, Oxford, UK) with genomic DNA samples from the proband IV:31 and his mildly affected son V:12. Cytosure Genomic DNA Labelling kit (OGT) and Agilent Oligo a-CGH Hybridization kit (Agilent Technologies, Santa Clara, CA, USA) were used, according to the manufacturers' protocol. Gains and losses of genomic sequences were called using the aberration detection statistical algorithm ADM-2, with a sensitivity threshold of 6.0.

¹Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, SP, Brazil; ²Departamento de Biologia, Universidade Estadual da Paraíba, Campina Grande, PB, Brazil; ³Pediatrics (Division of Medical Genetics), Human Genetics, Pathology, Obstetrics and Gynecology, University of Utah Health Sciences Center, Salt Lake City, UT, USA and ⁴Departamento de Neurologia e Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, SP, Brazil
Correspondence: Dr RC Mingroni-Netto, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, Rua do Matao, 277, CEP: 05508-900, São Paulo, SP, Brazil.
E-mail: renetto@ib.usp.br

Received 12 April 2017; revised 26 July 2017; accepted 27 July 2017; published online 31 August 2017

- III- Araujo BHS, Kaid C, De Souza JS, Gomes Da Silva S, Goulart E, Caires LCJ, **Musso CM**, Torres LB, Ferrasa A, Herai R, Zatz M, Okamoto OK, Cavaleiro EA. Down Syndrome iPSC-Derived Astrocytes Impair Neuronal Synaptogenesis and the mTOR Pathway *In Vitro*. **Molecular Neurobiology** 55:5962-5975, 2018.
- <https://doi.org/10.1007/s12035-017-0818-6>

Induced pluripotent stem-cell (iPSC) technology has provided a valuable tool for assessing specific cell types affected by Down syndrome (DS), such as neurons and astrocytes. In this study, we demonstrated that astrocytes may play a key role in the hyperactivation of the Akt/mTOR axis observed in DS brains, using patient-derived iPSCs. Further research in this direction will improve our understanding of DS etiology and facilitate the development of new therapeutic approaches.

Down Syndrome iPSC-Derived Astrocytes Impair Neuronal Synaptogenesis and the mTOR Pathway In Vitro

Bruno H. S. Araujo^{1,2} · Carolini Kaid³ · Janaina S. De Souza⁴ · Sérgio Gomes da Silva^{5,6} · Ernesto Goulart³ · Luiz C. J. Caires³ · Camila M. Musso³ · Laila B. Torres⁷ · Adriano Ferrasa^{8,9} · Roberto Herai⁹ · Mayana Zatz³ · Oswaldo K. Okamoto³ · Esper A. Cavalheiro¹

Received: 17 June 2017 / Accepted: 2 November 2017 / Published online: 11 November 2017
© Springer Science+Business Media, LLC, part of Springer Nature 2017

Abstract Several methods have been used to study the neuropathogenesis of Down syndrome (DS), such as mouse aneuploidies, post mortem human brains, and in vitro cell culture of neural progenitor cells. More recently, induced pluripotent stem cell (iPSC) technology has offered new approaches in investigation, providing a valuable tool for studying specific cell types affected by DS, especially neurons and astrocytes. Here, we investigated the role of astrocytes in DS developmental disease and the impact of the astrocyte secretome in neuron mTOR signaling and synapse formation using iPSC derived from DS and wild-type (WT) subjects. We demonstrated for the first time that DS neurons derived from hiPSC recapitulate the hyperactivation of the Akt/mTOR axis observed in DS brains and that DS astrocytes may play a key role in this dysfunction. Our results bear out that 21 trisomy in

astrocytes contributes to neuronal abnormalities in addition to cell autonomous dysfunctions caused by 21 trisomy in neurons. Further research in this direction will likely yield additional insights, thereby improving our understanding of DS and potentially facilitating the development of new therapeutic approaches.

Keywords Down syndrome · Induced pluripotent stem cell · mTOR pathway · Astrocyte

Introduction

Down syndrome (DS) is the most common cause of genetic intellectual disability (ID), and its worldwide

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12035-017-0818-6>) contains supplementary material, which is available to authorized users.

✉ Bruno H. S. Araujo
bruno.araujo@lnbio.cnpem.br

¹ Department of Neurosurgery and Neurology, Laboratory of Neuroscience, Universidade Federal de São Paulo (UNIFESP/EPM), São Paulo, São Paulo, Brazil

² Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Rua Giuseppe Máximo Scolfaro, no. 10.000, Campinas, São Paulo 13083-970, Brazil

³ Department of Genetics and Evolutionary Biology, Human Genome and Stem Cell Research Center, Institute of Biosciences, Universidade de São Paulo (USP), São Paulo, São Paulo, Brazil

⁴ Department of Medicine, Laboratory of Endocrinology and Translational Medicine, Universidade Federal de São Paulo (UNIFESP/EPM), São Paulo, São Paulo, Brazil

⁵ Hospital Israelita Albert Einstein (HIAE), São Paulo, São Paulo, Brazil

⁶ Universidade de Mogi das Cruzes, Mogi das Cruzes, São Paulo, Brazil

⁷ São Leopoldo Mandic Institute and Research Center, Campinas, São Paulo, Brazil

⁸ Experimental Multiuser Laboratory (LEM), Graduate Program in Health Sciences (PPGCS), School of Medicine, Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, Paraná 80215-901, Brazil

⁹ Department of Informatics (DEINFO), Universidade Estadual de Ponta Grossa (UEPG), Ponta Grossa, Paraná 84030-900, Brazil

- IV- Caires LCJ, Goulart E, Melo US, Araujo BHS, Alvizi L, Soares-Schanoski A, Oliveira, DF, Kobayashi GS, Griesi-Oliveira K, **Musso CM**, Amara M, Silva LF, Astray RM, (...) Passos-Bueno MR, Zatz M. Discordant congenital Zika syndrome twins show differential in vitro viral susceptibility of neural progenitor cells. **Nature Communications** 9:475, 2018.
<https://doi.ORG/10.1038/s41467-017-02790-9>

Congenital Zika syndrome (CZS) affects early brain development by impairment of neural progenitor cells (NPCs). In this study, we analyzed discordant CZS dizygotic twins by comparing RNA-seq of iPSC-derived NPCs. We demonstrated a different gene expression signature of mTOR and Wnt pathway regulators in NPCs from CZS-affected compared to unaffected babies, prior to Zika virus infection. In addition, cell samples from CZS-affected twins showed higher virus replication and reduced cell growth. A monogenic model of resistance or susceptibility to CZS was excluded by whole-exome analysis.

Discordant congenital Zika syndrome twins show differential in vitro viral susceptibility of neural progenitor cells

Luiz Carlos Caires-Júnior *et al.*[#]

Congenital Zika syndrome (CZS) causes early brain development impairment by affecting neural progenitor cells (NPCs). Here, we analyze NPCs from three pairs of dizygotic twins discordant for CZS. We compare by RNA-Seq the NPCs derived from CZS-affected and CZS-unaffected twins. Prior to Zika virus (ZIKV) infection the NPCs from CZS babies show a significantly different gene expression signature of mTOR and Wnt pathway regulators, key to a neurodevelopmental program. Following ZIKV in vitro infection, cells from affected individuals have significantly higher ZIKV replication and reduced cell growth. Whole-exome analysis in 18 affected CZS babies as compared to 5 unaffected twins and 609 controls excludes a monogenic model to explain resistance or increased susceptibility to CZS development. Overall, our results indicate that CZS is not a stochastic event and depends on NPC intrinsic susceptibility, possibly related to oligogenic and/or epigenetic mechanisms.

Correspondence and requests for materials should be addressed to M.Z. (email: mayazatz@usp.br)

[#]A full list of authors and their affiliations appears at the end of the paper

- V- Oliveira D, Leal GF, Sertié AL, Caires LCJ, Goulart E, **Musso CM**, Oliveira JRM, Krepschi ACV, Vianna-Morgante AM, Zatz M. 10q23.31 microduplication encompassing *PTEN* decreases mTOR signalling activity and is associated with autosomal dominant primary microcephaly. **Journal of Medical Genetics** 0:1-5, 2018.
<http://dx.doi.org/10.1136/jmedgenet-2018-105471>

Hereditary primary microcephaly (MCPH) is an autosomal dominant condition, characterized by decreased occipitofrontal circumference and intellectual disability. In this study, we identified a microduplication at 10q23.31 leading to a downregulation of the mTOR signaling pathway through overexpression of *PTEN*, as the causative variant of MCPH in a Brazilian family.

SHORT REPORT

10q23.31 microduplication encompassing *PTEN* decreases mTOR signalling activity and is associated with autosomal dominant primary microcephaly

Danyllo Oliveira,¹ Gabriela Ferraz Leal,^{2,3} Andréa L Sertié,⁴ Luiz Carlos Caires Jr,¹ Ernesto Goulart,¹ Camila Manso Musso,¹ João Ricardo Mendes de Oliveira,⁵ Ana Cristina Victorino Krepischki,¹ Angela Maria Vianna-Morgante,¹ Mayana Zatz¹

¹Department of Genetics and Evolutionary Biology, Human Genome and Stem-Cell Research Center, Institute of Biosciences, University of São Paulo, São Paulo, Brazil

²Fernando Figueira Integral Medicine Institute, Recife, Brazil

³Amara de Medeiros

Health Center, University of Pernambuco, Recife, Brazil

⁴Hospital Albert Einstein, São Paulo, Brazil

⁵Neuropsychiatry Department, Federal University of Pernambuco, Recife, Brazil

Correspondence to

Professor Mayana Zatz, Department of Genetics and Evolutionary Biology, Human Genome and Stem-Cell Research Center, University of São Paulo, SPO3178-200, Brazil; mayazatz@usp.br

Received 4 May 2018

Revised 7 August 2018

Accepted 29 August 2018



© Author(s) (or their employer(s)) 2018. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Oliveira D, Leal GF, Sertié AL, et al. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmedgenet-2018-105471

ABSTRACT

Background Hereditary primary microcephaly (MCPH) is mainly characterised by decreased occipitofrontal circumference and variable degree of intellectual disability. MCPH with a dominant pattern of inheritance is a rare condition, so far causally linked to pathogenic variants in the *ALFY*, *DPP6*, *KIF11* and *DYRK1A* genes.

Objective This study aimed at identifying the causative variant of the autosomal dominant form of MCPH in a Brazilian family with three affected members.

Methods Following clinical evaluation of two sibs and their mother presenting with autosomal dominant MCPH, array comparative genome hybridisation was performed using genomic DNA from peripheral blood of the family members. Gene and protein expression studies were carried out in cultured skin fibroblasts.

Results A 382 kb microduplication at 10q23.31 was detected, encompassing the entire *PTEN*, *KLLN* and *ATAD1* genes. *PTEN* haploinsufficiency has been causally associated with macrocephaly and autism spectrum disorder and, therefore, was considered the most likely candidate gene to be involved in this autosomal dominant form of MCPH. In the patients' fibroblasts, *PTEN* mRNA and protein were found to be overexpressed, and the phosphorylation patterns of upstream and downstream components of the mammalian target of rapamycin (mTOR) signalling pathway were dysregulated.

Conclusions Taken together, our results demonstrate that the identified submicroscopic 10q23.31 duplication in a family with MCPH leads to markedly increased expression of *PTEN* and reduced activity of the mTOR signalling pathway. These results suggest that the most probable pathomechanism underlying the microcephaly phenotype in this family involves downregulation of the mTOR pathway through overexpression of *PTEN*.

INTRODUCTION

Hereditary primary microcephaly (MCPH), characterised by occipitofrontal circumference (OFC) two or three SD below the population mean,^{1,2} is a genetically heterogeneous disorder whose overall incidence varies across different populations, from 1.3 to 150 per 100 000 newborns. The population prevalence of this condition is directly related to the rates of endogamy, since most cases of MCPH have an autosomal recessive mode of inheritance.^{1,3}

Nevertheless, families have been reported, in which MCPH has either autosomal dominant or X linked inheritance.^{1,4}

Patients with MCPH often display varying degrees of intellectual disability (ID), which is correlated to the level of brain hypoplasia. Despite being structurally smaller, the patients' brains usually exhibit normal cytoarchitecture,² suggesting that the genes implicated in MCPH have an important role in determining human brain size. For instance, the *ASPM* gene, associated with autosomal recessive MCPH, has been reported as undergoing positive selection for cortical size expansion during recent primate evolution.⁵

The main processes through which MCPH-linked genes regulate brain growth involve those associated with neural progenitor cells (NPCs) divisions. During neocortical development, these cells, particularly concentrated at ventricular zone (VZ) and subventricular zone (SVZ) of the developing telencephalon, undergo both symmetric and asymmetric divisions, which maintain the stem cell pool and generate newborn neurons that, through radial migration, originate the cortical plate.⁶ Accurate positioning of centrosomes and spindles is essential for proper NPCs division and cortical development, since vertical cleavage planes perpendicular to the VZ and SVZ usually result in symmetric divisions, while horizontal cleavages lead to asymmetric divisions.⁶ Most of the known pathogenic variants responsible for autosomal recessive MCPH are in genes encoding centrosomal proteins or proteins required for spindle formation and function, such as *ASPM*, *CEP152*, *CEP235*, *CENPJ* and *CDK5RAP2*. In addition, disruption of chromatin dynamics and condensation, caused by pathogenic variants in the *PCH1* and *MCPH1* genes, has also been implicated in the pathogenesis of autosomal recessive MCPH.⁷

Patients with autosomal dominant MCPH⁴ usually show moderate-to-mild ID or no cognitive impairment. The term 'silent microcephaly' has been proposed to refer to this entity without any neurological or dysmorphic manifestations.⁴ Recently, linkage studies and exome sequencing identified a pathogenic variant in the *ALFY* gene, which encodes an autophagy scaffold protein, segregating along with microcephaly through three generations of a family.⁸ Functional studies using animal models showed that the mutated protein

- VI- Tavares VLR, Kague E, **Musso CM**, Alegria TGP, Freitas RS, Bertola DR, Twigg SRF, Passos-Bueno MR. Craniofrontonasal Syndrome Caused by Introduction of a Novel uATG in the 5'UTR of *EFNB1*. **Molecular Syndromology** 10:40-47, 2019.
<https://www.ncbi.nlm.nih.gov/pubmed/30976278>

Craniofrontonasal syndrome (CFNS) is an X-linked disorder caused mainly by mutations in the coding region of *EFNB1*. In this study, we described a CFNS female patient carrying a variant located at c.-411, which creates a novel upstream ATG (uATG) within the *EFNB1* 5'UTR and alters an existing upstream open reading frame (uORF). We demonstrated the regulatory impact of uATG formation on *EFNB1* protein levels, associating uORF disturbance with the etiology of CFNS.

Craniofrontonasal Syndrome Caused by Introduction of a Novel uATG in the 5'UTR of *EFNB1*

Vanessa L. Romanelli Tavares^{a,b} Erika Kague^{a,b,f} Camila M. Musso^{a,b}
Thiago G.P. Alegria^b Renato S. Freitas^{d,e} Debora R. Bertola^{a-c}
Stephen R.F. Twigg^g Maria R. Passos-Bueno^{a,b}

^aCentro de Pesquisa sobre o Genoma Humano e Células-Tronco (CEGH-CEL) and ^bDepartamento de Genética e Biologia Evolutiva, Instituto de Biociências, and ^cInstituto da Criança do Hospital das Clínicas da Faculdade de Medicina, Universidade de São Paulo, São Paulo, ^dCentro de Atendimento Integral ao Fissurado Lábio Palatal (CAIF) and ^eUniversidade Federal do Paraná, Curitiba, Brazil; ^fDepartment of Physiology, Pharmacology and Neuroscience, University of Bristol, Bristol, and ^gClinical Genetics Group, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK

Keywords

Genetic counseling · Molecular diagnosis and testing · Protein translation · Regulatory variant · Upstream ORF and ATG

Abstract

Craniofrontonasal syndrome (CFNS) is an X-linked disorder caused by *EFNB1* mutations in which females are more severely affected than males. Severe male phenotypes are associated with mosaicism, supporting cellular interference for sex bias in this disease. Although many variants have been found in the coding region of *EFNB1*, only 2 pathogenic variants have been identified in the same nucleotide in 5'UTR, disrupting the stop codon of an upstream open reading frame (uORF). uORFs are known to be part of a wide range of post-transcriptional regulation processes, and just recently, their association with human diseases has come to light. In the present study, we analyzed *EFNB1* in a female patient with typical features of CFNS. We identified a variant,

located at c.–411, creating a new upstream ATG (uATG) in the 5'UTR of *EFNB1*, which is predicted to alter an existing uORF. Dual-luciferase reporter assays showed significant reduction in protein translation, but no difference in the mRNA levels. Our study demonstrates, for the first time, the regulatory impact of uATG formation on *EFNB1* levels and suggests that this should be the target region in molecular diagnosis of CFNS cases without pathogenic variants in the coding and splice sites regions of *EFNB1*.

© 2018 S. Karger AG, Basel

Craniofrontonasal syndrome (CFNS; OMIM 304110) is a rare X-linked dominant disorder caused by loss-of-function mutations in *EFNB1* [Twigg et al., 2004; Wieland et al., 2004]. Greater severity is observed in heterozygous female patients compared with hemizygous male individuals. Female patients can present with severe hypertelorism, a central nasal groove, craniofacial asymmetry, and coronal craniosynostosis associated with extracranial

KARGER

© 2018 S. Karger AG, Basel

E-Mail karger@karger.com
www.karger.com/msy

Maria R. Passos-Bueno
Departamento de Genética e Biologia Evolutiva, Instituto de Biociências
Universidade de São Paulo, Rua do Matão, 277, sala 200
São Paulo, SP 05508-090 (Brazil)
E-Mail passos@ib.usp.br

- VII- Guo L, Bertola DR, Takanohashi A, Saito A, Segawa Y, Yokota T, Ishibashi S, Nishida Y, Yamamoto GL, Franco JFS, Honjo RS, Kim CA, **Musso CM**, Timmons M, Pizzino A, Taft RJ, Lajoie B, Knight MA, Fischbeck KH, Singleton AB, Ferreira CR, Wang Z, Yan L, Garbern JY, Simsek-Kiper PO, Ohashi H, Robey PG, Boyde A, Matsumoto N, Miyake N, Spranger J, Schiffmann R, Vanderver A, Nishimura G, Passos-Bueno MR, Simons C, Ishikawa K, Ikegawa S. Bi-allelic *CSF1R* Mutations Cause Skeletal Dysplasia of Dysosteosclerosis-Pyle Disease Spectrum and Degenerative Encephalopathy with Brain Malformation. **The American Journal of Human Genetics** 104:1-11, 2019.

<https://www.sciencedirect.com/science/article/abs/pii/S0002929719301004?via%3Dihub>

Colony stimulating factor 1 receptor (*CSF1R*) plays key roles in monocyte/macrophage lineage development. It is known that a progressive neurodegenerative disorder named hereditary diffuse leukoencephalopathy with spheroids (HDLS) is caused by mono-allelic mutations in *CSF1R*. In this study, we report bi-allelic mutations in five affected individuals who had brain malformations and skeletal dysplasia. We characterized a unique human skeletal phenotype caused by *CSF1R* deficiency.

Bi-allelic *CSF1R* Mutations Cause Skeletal Dysplasia of Dysosteosclerosis-Pyle Disease Spectrum and Degenerative Encephalopathy with Brain Malformation

Long Guo,^{1,26} Débora Romeo Bertola,^{2,3,26,*} Asako Takanohashi,⁴ Asuka Saito,⁵ Yuko Segawa,⁶ Takatori Yokota,⁵ Satoru Ishibashi,⁵ Yoichiro Nishida,⁵ Guilherme Lopes Yamamoto,^{2,3} José Francisco da Silva Franco,² Rachel Sayuri Honjo,² Chong Ae Kim,² Camila Manso Musso,³ Margaret Timmons,⁷ Amy Pizzino,⁴ Ryan J. Taft,⁸ Bryan Lajoie,⁸ Melanie A. Knight,⁹ Kenneth H. Fischbeck,⁹ Andrew B. Singleton,¹⁰ Carlos R. Ferreira,¹¹ Zheng Wang,^{1,12} Li Yan,¹³ James Y. Garbern,^{14,27} Pelin O. Simsek-Kiper,¹⁵ Hirofumi Ohashi,¹⁶ Pamela G. Robey,¹⁷ Alan Boyde,¹⁸ Naomichi Matsumoto,¹⁹ Noriko Miyake,¹⁹ Jürgen Spranger,^{20,21} Raphael Schiffmann,²² Adeline Vanderver,⁴ Gen Nishimura,²³ Maria Rita dos Santos Passos-Bueno,³ Cas Simons,^{24,25} Kinya Ishikawa,⁵ and Shiro Ikegawa^{1,*}

Colony stimulating factor 1 receptor (*CSF1R*) plays key roles in regulating development and function of the monocyte/macrophage lineage, including microglia and osteoclasts. Mono-allelic mutations of *CSF1R* are known to cause hereditary diffuse leukoencephalopathy with spheroids (HDLS), an adult-onset progressive neurodegenerative disorder. Here, we report seven affected individuals from three unrelated families who had bi-allelic *CSF1R* mutations. In addition to early-onset HDLS-like neurological disorders, they had brain malformations and skeletal dysplasia compatible to dysosteosclerosis (DOS) or Pyle disease. We identified five *CSF1R* mutations that were homozygous or compound heterozygous in these affected individuals. Two of them were deep intronic mutations resulting in abnormal inclusion of intron sequences in the mRNA. Compared with *Csf1r*-null mice, the skeletal and neural phenotypes of the affected individuals appeared milder and variable, suggesting that at least one of the mutations in each affected individual is hypomorphic. Our results characterized a unique human skeletal phenotype caused by *CSF1R* deficiency and implied that bi-allelic *CSF1R* mutations cause a spectrum of neurological and skeletal disorders, probably depending on the residual *CSF1R* function.

Colony stimulating factor 1 (*CSF1*) regulates survival, proliferation, differentiation, and phagocytic and chemotactic activity of cells from the monocyte/macrophage lineage, which could be specialized as microglia in brain and as osteoclasts in bone.¹ The effects of *CSF1* are mediated by its receptor, *CSF1R*, which triggers multiple signal transduction pathways.² *CSF1R* is expressed not only in cells of hematopoietic origin, but also in non-hematopoietic cells including Paneth cells,³ renal proximal tubule epithelial cells,⁴ oocyte and female reproductive tract,⁵ as well as decidual cells and trophoblastic cells.⁶ The pleiotropic

effects of *CSF1R* deficiency have been under study extensively.

In mouse, bi-allelic *Csf1r* deficiency is reported to cause sclerosing skeletal dysplasia and gross anatomical abnormalities in developing brains, leading to death within 6 weeks.^{7–9} In contrast, *Csf1r*^{+/-} mice gradually develop cognitive and sensorimotor deficits, depression with anxiety-like behavior, and enlarged ventricles from 6 months of age.¹⁰ In humans, *CSF1R* mono-allelic mutations, which are considered to cause haploinsufficiency or dominant-negative effects, lead to hereditary diffuse

¹Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences, Tokyo 108-8639, Japan; ²Unidade de Genética Clínica, Instituto da Criança do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo 05403-000, Brazil; ³Instituto de Biociências da Universidade de São Paulo, São Paulo 05508-090, Brazil; ⁴Division of Neurology, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA 19104, USA; ⁵Department of Neurology and Neurological Science, Graduate School, Tokyo Medical and Dental University, Tokyo 113-8519, Japan; ⁶Department of Orthopedic Surgery, Graduate School, Tokyo Medical and Dental University, Tokyo 113-8519, Japan; ⁷Developmental and Metabolic Neurology Branch, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD 20892, USA; ⁸Illumina, Inc., 5200 Illumina Way, San Diego, CA 92122, USA; ⁹Neurogenetics Branch, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD 20892, USA; ¹⁰Laboratory of Neurogenetics, National Institute of Aging, NIH, Bethesda, MD 20892, USA; ¹¹Medical Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA, and Division of Genetics and Metabolism, Children's National Health System, Washington, DC 20010, USA; ¹²Department of Medical Genetics, Institute of Basic Medical Sciences, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100005, People's Republic of China; ¹³Department of Neurology, China-Japan Friendship Hospital, Beijing 100029, People's Republic of China; ¹⁴Center of Molecular Medicine and Genetics, Wayne State University, Detroit, MI 48201, USA; ¹⁵Department of Pediatrics, Hacettepe University Medical Faculty, Ankara 06100, Turkey; ¹⁶Division of Medical Genetics, Saitama Children's Medical Center, Saitama 330-8777, Japan; ¹⁷Skeletal Biology Section, National Institute of Dental and Craniofacial Research, NIH, Bethesda, MD 20892, USA; ¹⁸Biophysics, Oral Growth and Development, Dental Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, UK; ¹⁹Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan; ²⁰Central German Competence Center for Rare Diseases (MKSE), Magdeburg 39120, Germany; ²¹Greenwood Genetic Center, Greenwood, SC 29646, USA; ²²Baylor Scott & White Research Institute, Dallas, TX 75204, USA; ²³Intractable Disease Center, Saitama University Hospital, Moro 350-0495, Japan; ²⁴Translational Bioinformatics Group, Murdoch Children's Research Institute, The Royal Children's Hospital, Melbourne, VIC 3052, Australia; ²⁵Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

²⁶These authors contributed equally to this work

²⁷Deceased

*Correspondence: sikegawa@ims.u-tokyo.ac.jp (S.I.), debora.bertola@usp.br (D.R.B.)

<https://doi.org/10.1016/j.ajhg.2019.03.004>

© 2019 American Society of Human Genetics.