Novel genetic contributions in patients with orofacial cleft anomalies

Alharatani, Reham

Awarding institution:
King's College London

Novel Genetic Contributions in Patients with Orofacial Cleft Anomalies

A thesis submitted to King's College London
for the degree of
Doctor of Philosophy

in

Paediatric Dentistry & Genetics

by

Reham Eisa Alharatani

September 2019

Centre of Oral, Clinical and Translational Science, Faculty of Dentistry,
Oral & Craniofacial Sciences
King's College London, London, United Kingdom

Supervisors

Prof. Karen J. Liu

Centre for Craniofacial and Regenerative Biology

Faculty of Dentistry, Oral & Craniofacial Sciences

King's College London

&

Prof. Marie Thérèse Hosey

Centre of Oral, Clinical and Translational Science

Faculty of Dentistry, Oral & Craniofacial Sciences

King's College London

Examined by

Prof. Peter A. Mossey & Prof. David Rice

ABSTRACT

Nowadays there is an emphasis on detecting *de novo* gene changes. The aim of this project was to identify novel *de novo* gene mutations in children with orofacial clefts. To achieve this, I targeted cleft children with unknown aetiology at the South Thames regional cleft centre, London, UK. Using a trio-based design, I recruited 90 child probands and 159 relatives and used parental tooth anomalies as a subclinical marker to categorise probands into those with potentially inherited or *de novo* genetic risks. I identified children with 'cleft-only', 'cleft-tooth anomaly' and 'cleft-medical condition +/- tooth anomaly'. Of those dentally examined, (55/127) relatives in 45 trios were found to have dental anomalies, especially hypodontia outside their child's cleft side, suggesting Mendelian inheritance. In Chapter 3, I reported findings from the Clinical Study and described the cohort with non-syndromic clefts. I showed that 16 probands had a 'cleft-only' condition and only six of them had parents with no dental anomalies or family history of clefting. This chapter highlighted that the clinical diagnosis of 'isolated' clefts needs to be more precise. The Clinical Study further identified (28/90) probands who also had other congenital anomalies or medical co-morbidities associated with their orofacial clefts. Family trios from the 'cleft-medical condition +/- tooth anomaly' group were explored further and became the focus of this thesis. Whole exome sequencing was carried out on several trios from this group. In Chapter 4, a novel *de novo* mutation in the catenin delta-1 (*CTNND1*) gene was identified. This gene encodes the p120-catenin protein known for its role in cell-cell adhesion and the regulation of epithelial-to-mesenchymal transition. I expanded on the developmental roles for p120-catenin demonstrated through the phenotypes I described in the human patients and in mouse and *Xenopus* models. I used the Deciphering Developmental Disorders (DDD) dataset to search and recruit further subjects with *CTNND1* mutations and identified 12 more individuals whom I found to have characteristic craniofacial and dental features as well as heart, limb and neurodevelopmental anomalies. Using loss-of-function genetic approaches in mouse and *Xenopus*, I demonstrated novel roles for *CTNND1* in the vocal cords and the velopharynx, craniofacial cartilages and the heart. I suggest that *CTNND1* is a candidate neurocristopathy gene, highlighting both epithelial and mesenchymal roles for p120-catenin. My work expands upon the spectrum of abnormalities associated with C*TNND1* variants beyond those previously described in non-syndromic cleft lip/palate (CLP) and blepharocheilodontic syndrome (BCD1) and that variations in this gene may be expanded to a broader velocardiofacial-like syndrome. In Chapter 5, I conducted exploratory research into phenotypically different

monozygotic twins who had also been recruited through my Clinical Study. Whole exome sequencing identified a novel *de novo* copy number variation in the *AGAP6* gene. I confirmed that *AGAP6* transcripts are strongly expressed during human embryonic development in craniofacial structures. In order to validate the pathogenicity of *AGAP6*, future work involves its implication in other unrelated individuals with rare craniofacial anomalies. Finally, in Chapter 6, I demonstrated the utility of a publicly available dataset (DECIPHER, www.decipher.org, DDD) in gene discovery. I developed a protocol to interrogate and analyse the dataset that included individuals with neurocristopathic anomalies using Human Phenotype Ontology (HPO) search terms honed from the medical conditions that I found in my Clinical Study. I assembled a list of novel putative genetic variants in *DIP2C*, *ABCA2* and *CELSR1*. The protocol I developed could be used for future studies. The segregation of cleft subjects based on 'associated anomalies', be it dental or medical, and on whether family members were affected with subclinical anomalies, emphasizes the *genetic status* underlying their conditions. In conclusion, I found novel *de novo* gene mutations in patients with orofacial clefts particularly in *CTNND1* and suggest a list of other potential candidate genes for future study.

ACKNOWLEDGMENTS

This project began from a meeting of minds between both my supervisors Professor Karen J. Liu and Professor Marie Thérèse Hosey. I believe that we have formed a terrific 'trio' and am excited for what has yet to come! My life in the past couple of years has been far from smooth sailing and so I am very grateful for their trust and continual support that made this PhD happen. Prof. Liu has given me a chance from day one whilst I was receiving my chemotherapy in my first year and believed that I could work post-recovery. Prof. Hosey and Mr Edward Hosey have been my family in London. Working with Prof. Hosey on many diverse projects has been an eye opener showing me that a 'paediatric dentist' should not necessarily be bound to one aspect of research. Prof. Liu and Prof. Hosey have been a true inspiration, with their proficiency, ambition and creative minds, and I am proud and privileged to have worked with such a dynamic female duo in academia.

I would like to thank my KCL examination committee, Prof. Martyn Cobourne and Dr Isabelle Miletich, for their guidance, scientific discussions and continual support throughout the course of my PhD. I am grateful for the support and encouragement I have received from my postgraduate coordinator, Dr Koula Asimakopoulou; with her expertise in psychology, having chats with her has always put me at ease.

I would also like to thank the outstanding team that I have worked with. Many thanks go to: Dr Nabina Bhujel and Dr Alex Cash at the South Thames Cleft Service for their provision of everything necessary to help with the smooth running of the clinical recruitment and their enthusiasm about the research; the Yale team, Prof. Mustafa Khokha and Dr Saquib Lakhani for carrying out the exome sequencing in their facilities and for their excellent and insightful scientific discussions; Dr Weizhen Ji the 'wizard' bioinformatician for her bioinformatic analysis of the candidate genes and for her responsiveness to all my naive questions; Dr Ana Beleza for being the clinical geneticist on our team, her excitement about new findings and love for research has gotten me equally excited. Thank you, Prof. Karen Liu, for introducing me to the Yale team whose input has been an integral part of my project; the whole team has showed nothing but care and genuine trust in my efforts. It has been a tremendous pleasure to work with them from afar and I am positive that there will be many more collaborations to come.

Many thanks go out to the families who have made this project possible and all the clinical geneticist for their collaboration, help and guidance. A special thanks goes to the Ministry of Education and Umm al Qura University, Kingdom of Saudi Arabia, for the funding and scholarship opportunity to gain experience and to undertake my postgraduate degrees in the United Kingdom.

I am grateful to Dr William Barrell, Dr Hadeel Al-Lami and Dr John Griffin for training me, for their support in my early years in the lab and throughout and for their endless patience when answering all my questions. Being in the Liu Lab has been one of the highlights of my PhD due to the fantastic members and the amazing friendships that I have developed. In addition to John, Wills and Hadeel, thank you for the good times goes to Daniel, Annie, Sandra, Anna, Lisa, Kshem, Jia Shang and all past and present members. Thank you to everyone in the Centre for Craniofacial and Regenerative Biology for their help particularly Ms Fernanda Suzano, Mr Alasdair Edgar, Ms Susmitha Rao and Ms Ewa Kolano-Merlin for their technical support. Especially, a big thanks goes to Ms Angela Gates and Dr Christopher Healy for their genuine and boundless care and kindness. Many thanks to the friends who always made sure I was doing alright during my PhD; thanks to Dima, Rana, Michelle, Ana C., Lucia, Beccy, Hannah, P.Q., Ros, Catriona, Arshiya, Maria and Raha. Also, a personal and special thanks goes to Ms Kimberly Riegman, my dear friend and writing buddy. I have been very lucky to have written my thesis alongside such an enthusiastic, hard-working woman in science.

I've also enjoyed having wonderful students who have shown professionalism and love for learning, I have equally learned a lot from them. Thanks to Ms Asma Alshahrani, Ms Sumayyah Hannan, Ms Natsuko Neo, Ms Annie Liu, Mr Reece Clay and Ms Dominika Antoniszczak.

The best part of my PhD was gauging the progress in my academic years with my daughter's growth! We have enjoyed our trips to the office and lab over the weekends and doing our 'homework' alongside each other and her best part was having her chai lattes as a treat! I hope I can be an inspiring example to her. This thesis is dedicated to my husband Raed, my daughter Maria, my mum Hanan, my dad Eisa, my sister Rund, my brothers Ramadan and Mousa and my in-laws Mansour and Manal. You have all been my rock. Without your endless love, patience and encouragement to keep going, especially during the hardest of times, none of this would have been possible.

CONTRIBUTORS

All studies described within this thesis were conducted by me, with the following exceptions:

Whole exome sequencing and bioinformatic analyses were performed by Dr Weizhen Ji & Dr Saquib Lakhani, Paediatric Genomics Discovery Program, Department of Paediatrics, Yale University School of Medicine, New Haven, CT 06520, USA.

CRISPR/Cas9 design & injections (Chapter 4) in *X. tropicalis* embryos were performed by Dr Emily Mis in the Khokha Lab, Paediatric Genomics Discovery Program, Departments of Genetics and Paediatrics, Yale University School of Medicine, New Haven, CT 06520, USA., *or* the European *Xenopus* Resource Centre, Portsmouth, UK.

Genetic Counselling and the feedback of sequencing results to families recruited from the South Thames Cleft Service were carried out by Dr Ana Beleza-Meireles.

The application for **the CAP180 dataset** (Chapter 6) was initiated by Dr Ana Beleza-Meireles, Dr Melita Irving & Prof. Karen J. Liu.

Participants located via **the Deciphering Developmental Disorders** database gave their consent to their referring clinicians for their images to be presented.

The **Human Developmental Biology Resource** (HDBR, http://www.hdbr.org) provided the embryonic material, funded by the Joint MRC/Wellcome Trust (grant #099175/Z/12/Z) to HDBR.

PUBLICATIONS & AWARDS

Recipient of the British Society of Paediatric Dentistry-Faculty of Dental Surgery (BSPD-FDS) (RCSEng) Small Grant Award, July 2017 (Appendix 1).

Winner of the British Society of Paediatric Dentistry Research Prize, The BSPD Annual Conference, Birmingham, UK, September 2019.

Alharatani, R., Griffin, J. N., & Liu, K. J. (2019). Expression of the guanine nucleotide exchange factor, *RAPGEF5*, during mouse and human embryogenesis. *Gene Expression Patterns*, 119057.

Alharatani, R., Ververi, A., Beleza-Meireles, A., Ji, W., Mis, E., Patterson, Q. T., *et a*l. Liu, K. J., (2019). Novel truncating mutations in *CTNND1* cause a dominant craniofacial and cardiac syndrome. ***bioRxiv***, 711184. **This manuscript has now been published in Human Molecular Genetics, https://doi.org/10.1093/hmg/ddaa050, *Alharatani, R. et al., (2020), Human Molecular Genetics.***

Alharatani, R., Alshahrani, A., Bhujel, N., Antoniszczak, D., Cash, A., Good, S., Liu, K. J., Hosey, M.T. (2020). Refining genetic risk models for isolated orofacial clefts – a cohort study of family trios. **This manuscript is currently in submission at a peer-reviewed journal.**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

[MIM: #]	Mendelian Inheritance in Man Number
> or /	Denotes nucleotide substitution
aCGH	Microarray Comparative Genomic Hybridization
AGAP6	ArfGAP with GTPase Domain, Ankyrin Repeat and PH Domain 6
ASD	Autism Spectrum Disorder
Asd	Atrial Septal Defect
Bp	Base pairs
CADD	Combined Annotation-Dependent Depletion
CAP	Complementary Research Proposal
CHD	Congenital Heart Disease
CLP	Cleft Lip and/or Palate
CNC	Cranial Neural Crest
CNV	Copy Number Variant
CoA	Coarctation of the Aorta
CPO	Cleft Palate Only
CTNND1	Catenin Delta-1
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeats/ CRISPR associated protein 9
DDD	Deciphering Developmental Disorders Study
DNA	Deoxyribonucleic Acid
E-cadherin	Epithelial Cadherin
E.	Embryonic Day
EMT	Epithelial-Mesenchymal-Transition
ERN	European Reference Network
EUROCAT	The Surveillance of Congenital Anomalies in Europe
ExAC	Exome Aggregation Consortium
FGF	Fibroblast Growth Factor
FGFs	Fibroblast Growth Factors
g. or chr.	Genomic reference sequence
GDIS	Gene Damage Index Score

GnomAD	Genome Aggregation Database
GSTT	Guy's and St. Thomas' Foundation Trust
GWAS	Genome Wide Association Studies
HH	Hedgehog
HPO	Human Phenotype Ontology
ICBDMS	The International Clearinghouse for Birth Defect Monitoring System
Indels	Insertions-deletions
Kb	Kilobases
LOF	Loss of Function
MAF	Minor Allele Frequency
Mb	Megabases
MCA	Multiple Congenital Anomalies
MDT	Multidisciplinary Team
MEE	Medial Edge Epithelium
MES	Medial Epithelial Seam
MFT	Multi-Factorial Threshold
MLPA	Multiplex Ligation-Dependent Probe Amplification
MMI	Multi-Malformed Infants
MZ	Monozygotic
NCC	Neural Crest Cells
NHS	National Health Services
NS-	Non-Syndromic
P	Postnatal Day
p.	Protein reference sequence
PDA	Patent Ductus Arteriosus
pLI	Loss Intolerance Score
PPH	PolyPhen2 (Polymorphism Phenotyping Versions 2)
PRS	Pierre Robin Sequence
RNA	Ribonucleic Acid
RVIS	Residual Variation Intolerance Score
S-	Syndromic
SIFT	Sorts Intolerant from Tolerant
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
STCS	South Thames Cleft Service

TGF-β	Transforming Growth Factor-β
TOF	Tetralogy of Fallot
UTR	Untranslated Region
VPI	Velopharyngeal Insufficiency
VSD	Ventricular Septal Defect
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WHO	World Health Organisation

Chapter 1 GENERAL INTRODUCTION

1.1 CRANIOFACIAL AND PALATAL DEVELOPMENT AND DISEASE

Fusion of the facial prominences and palate, or roof of the mouth, are crucial events in prenatal development. Failure in one or more of these processes can lead to cleft lip and/or palate (CLP), and is associated with infant mortality and with difficulties in feeding, breathing, speech and other oral problems. Because CLP is the most common congenital birth defect manifesting in live births, ranging from 1:500 to 1:2000 (Mossey & Modell, 2012), many studies in animal models and in humans have focused on the causative events. These could roughly be divided into anatomical and genetic effects and will be addressed in the following sections of this chapter.

The process of palatogenesis depends on highly coordinated, anatomically specific and precisely timed molecular signals for normal development (Lan *et al.,* 2015; Li *et al.,* 2017; Dixon *et al.,* 2011). Among them, cell migration, proliferation, fusion, apoptotic and differentiation events contribute to the complexity of craniofacial organization. In addition, multiple signalling pathways including Hedgehog (HH) (Cobourne and Green, 2012), transforming growth factor-β (TGF-β) (Nakajima *et al.,* 2018) and fibroblast growth factor (FGF) signalling (Weng *et al.,* 2018) complement each other, and aberration from any of this programming is likely to lead to pathogenesis of the palate (Han *et al.*, 2009; Lan *et al.,* 2015; Nawshad *et al.*, 2004; Riley *et al.*, 2007). The development of the palate is also closely associated with tooth formation, which occurs concurrently, sharing many of the same molecular pathways and tissue-tissue interactions (described in Sections 1.1.3.1 and 1.2.3.3). Therefore, tooth anomalies should also be taken into account when examining palatal vault changes.

The palatal shelves are composed of pharyngeal ectoderm and cranial neural crest-derived mesenchyme. In general, CLPs are caused by failure in any of the processes that develop the primary palate from the facial compartments or secondary palate from the palatal compartments of the first pharyngeal arch (Panetta *et al.*, 2008). Mouse models have been key in forming our understanding of the mechanistic processes involved in orofacial clefting anomalies and of other human developmental disorders (Cox *et al.*, 2019; Liu, 2016). Genetic knockouts have highlighted the stages at which pathology takes place, be it during downward growth of the palatal shelves, during elevation of the shelves or during the final process of fusion and adhesion. Indeed, hundreds of reported mouse strains display orofacial clefts as a component of their phenotypes and their pathogenic variants have often been extensively

characterised well before the identification of the corresponding human disorder. While, in contrast to humans, the vast majority of clefts in mice involve defects of the secondary palate and just a few exhibit cleft lip defects, mouse cleft palate loci have been vital in informing the list of candidate genes for human orofacial clefts in general (Cox *et al.*, 2019). The following section provides a brief overview of cleft palate in mice, looks at some of the hypotheses underlying changes in palatal morphology and takes into consideration *anatomic modifiers*.

1.1.1 Facial Development and Contribution of the Neural Crest

Development of the pharyngeal arches

Development of the face involves an evolutionarily conserved series of morphological events and patterning of facial primordia (Cox *et al.,* 2004; Szabo-Rogers *et al.,* 2010). In mammals, early facial primordia have a central core of mesenchymal cells, derived from mesoderm and neural crest cells (NCCs), covered by epithelia derived externally from ectoderm and internally from endoderm (Sperber *et al.,* 2001). Development of the face begins with the formation of five facial prominences (pharyngeal arches) surrounding the primitive mouth, a single median frontonasal process and two maxillary and mandibular processes, positioned on the ventral surface of the embryo. The first pharyngeal arch constitutes the pair of the maxillary and mandibular processes. The outgrowth of the facial prominences is largely determined by the significant contribution from the neural crest cells (Cox *et al.,* 2004; Sperber *et al.,* 2001; Szabo-Rogers *et al.,* 2010)**.** In **(Figure 1-1)**, I utilised a neural crest specific lineage tracing technique to label neural crest tissues. As shown in **(Figure 1-1, A-B),** *Wnt1-cre* positive cells populate the frontonasal, maxillary and mandibular processes, which is in accordance with previous reports (Chen *et al.,* 2017; Yoshida *et al.,* 2008). As development proceeds, fusion of the lateral and medial nasal processes forms the nasal openings whereas fusion of the medial nasal process with the maxillary prominences forms the upper lip and primary palate. Two outgrowths from the medial nasal and maxillary processes called the palatal shelves form the secondary palate (Cox *et al.,* 2004; Szabo-Rogers *et al.,* 2010).

Tissue requirements for the neural crest

Neural crest cells (NCCs) arise from the neural plate border early during development. The cells go through a process of induction, de-lamination, migration into their destined locations and finally differentiation (Szabo-Rogers *et al.*, 2010). Neural

crest cells are also multipotent, being able to differentiate into many cell types (Simoes-Costa & Bronner, 2015). Development of neural crest cells within the pharyngeal arches depends on the action of numerous signaling molecules and transcription factors that subsequently guide migrating neural crest cells and later play a role in lineage determination, expansion and differentiation of neural crest derivatives (Clouthier *et al.*, 1998; Mayor & Theveneau, 2013; Theveneau & Mayor, 2012).

At the induction phase, NCCs begin as tightly adjoining epithelial cells with distinct apical-basal polarity, where *Snail/Slug* transcription factors are among the earliest known markers for neural crest formation (Trainor, 2005). Following their induction at the lateral edges of the neural plate, and prior to neural tube closure in mammals, a program of gene regulatory networks is initiated that leads to NCC delamination closely followed by the disintegration of cell-cell adhesion complexes including desmosomes, adherens and tight junctions (Cox, 2004; Theveneau and Mayor 2012). This reprogramming transforms the epithelial neural crest into cells with more mesenchymal-like properties. This enables their migration from the neural epithelium dorsally to populate the ventral aspects of the embryo. This process is aided by the down-regulation of cell adhesion molecules such as N-cadherin and the upregulation of others such as cadherin-7 and cadherin-11, along with a breakdown of the cytoskeletal factors in the basement membrane surrounding the neural tube (Mayor and Theveneau, 2013; Trainor, 2005).

In order for neural crest cells to be able to populate the facial prominences, they need to migrate in a highly patterned and directed manner away from their source of origin (Bolande, 1997; Szabo-Rogers *et al.*, 2010; Twigg & Wilkie, 2015). Like other cells during embryogenesis, directional cell migration of the neural crest cells is equally important to achieve normal development. Cell migration requires cell polarization and the formation of protrusions at the leading edge of the cell (Carmona-Fontaine *et al.,* 2008). As shown in **(Figure 1-1, B-C)**, we can see that as the *Wnt1-cre* positive cells begin to migrate, they form cellular protrusions. Reorganization of the actin cytoskeleton is also a pre-requisite for the formation of these cellular protrusions at the leading edge that are either in the form of broad sheet-like protrusions called lamellipodia or spiky extensions called filopodia. In addition to the orchestrated signalling machinery that is required for NCCs to initiate migration and confer full migratory potential, certain mechanisms have been proposed to regulate the directional migration of NCCs such as planar cell polarity (PCP), contact inhibition of locomotion (CIL), co-attraction and collective chemotaxis (Scarpa *et al.*, 2015; Theveneau & Mayor,

2011). Finally, differentiation of the ectomesenchyme generates all of the neural crest derivatives such as the hard tissues of the head including bone, cartilage and teeth (Chai and Maxson, 2006).

Failure of the neural crest to migrate to its destination, such as the palatal mesenchyme, leads to various pathologies termed neurocristopathies. Neurocristopathies are developmental malformations arising from failure of the neural crest. Induction, proliferation, or survival problems of the neural crest usually give rise to dysplasia (i.e. abnormal development of an organ or part of the body, including congenital absence), while cranial neural crest (CNC) migration defects lead to malformations which include cleft lip and cleft palate defects, and in diseases like CHARGE [MIM: # 214800], DiGeorge [MIM: # 188400], and Goldenhar [MIM: # 164210] syndromes (Theveneau & Mayor, 2011). Aberrant neural crest cell migration takes place if any of these processes fail and can lead to orofacial clefting. As shown in **(Figure 1-1, A)**, overall, the neural crest cell contribution expands the anterior-posterior axis of the developing embryo including the cranial, vagal and trunk neural crest, hence the diversity of phenotypes seen in neurocristopathic patients and the multiple organs affected in the syndromic patients.

Despite orofacial clefts being classified as neurocristopathies, and although much of the vertebrate craniofacial skeleton and connective tissue is derived from cranial NCCs, studies investigating the molecular mechanisms required for palatogenesis and animal models of cleft palate have predominately focused on the epithelial reorganisation of the palate and its cross-talk with the underlying mesenchyme. Fewer molecular studies explored the contribution of NCC to orofacial clefts (Brewer *et al.,* 2004; Dudas *et al.,* 2004; Hill *et al.,* 2015; Wang *et al.,* 2020). Most recently, Wang *et al.* (2020) showed that *Wnt1-Cre* mediated *Meis2* inactivation – which encodes a family of proteins that directly regulate HOX protein activity – in cranial neural crest cells led to cleft of the secondary palate in mice (Wang *et al.,* 2020). Likewise, conditional knockout of *Meis2* by *AP2α-Cre* in the developing neural crest cells resulted in defective craniofacial skeleton and abnormal palate (Machon *et al.,* 2015) as well as the *Wnt1-Cre* mediated deletion of *AP2α* (Brewer *et al.,* 2004).

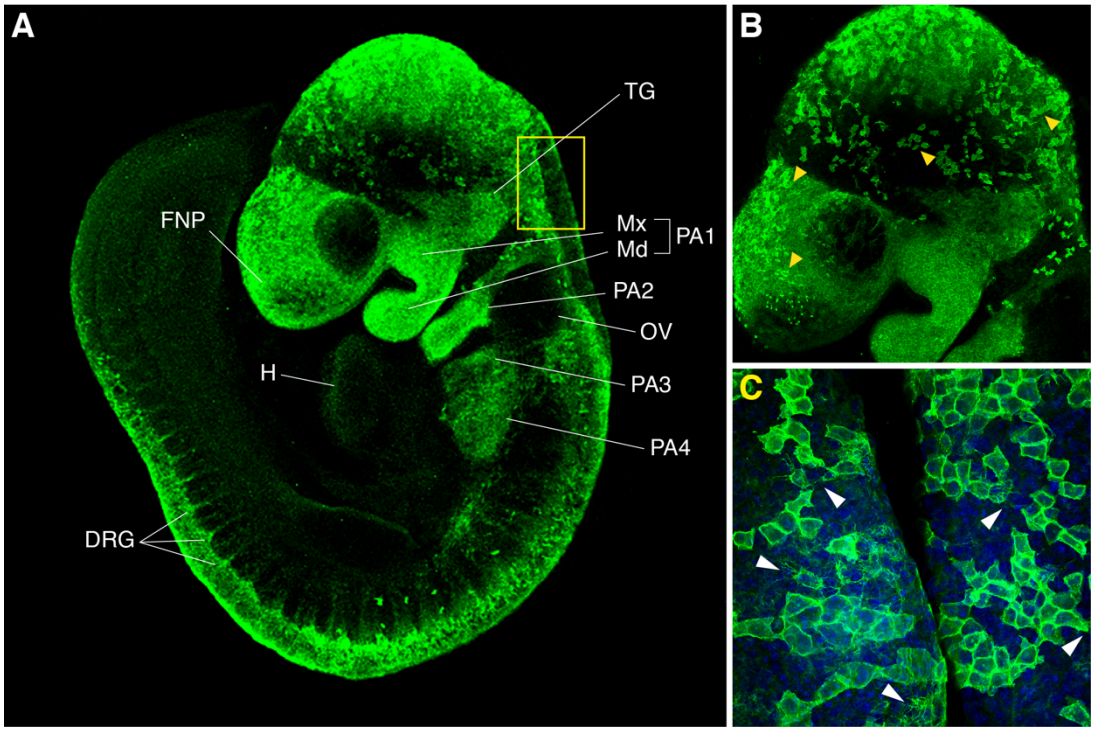


Figure 1-1 Lineage labelling of neural crest cells during embryonic development

[A-C] Embryonic day (E) 9.5 *Wnt1-cre; Rosa26R*^{mTmG} mouse embryo carrying the *ROSA26R-Cre* reporter and Cre recombinase driven by the *Wnt1* promoter. *Wnt1-cre*-mediated recombination labels neural crest-derived cells and tissues with GFP fluorescence (green). **[A] The population of pharyngeal arches with neural crest cells.** Side view of the whole embryo shows trunk and cranial crest contributions. Distinct streams of neural crest cells are shown rostrally. Cranial neural crest-derived cells populate the pharyngeal arches (PA 1,2,3 and 4). The cells condense to form the trigeminal ganglia (TG), the frontonasal process (FNP) and the maxillary and mandibular processes (Mx and Md, respectively). **[B] Migratory neural crest cells.** Lateral view of the craniofacial region. Migration of cranial NC (yellow arrowheads) from the hindbrain region as they leave the neural tube towards the ventral portions of the face. **[C] Cellular protrusions.** Close up of boxed area in [A] seen from the dorsal view. Cell migration requires the formation of protrusions at one end of the neural crest cell (white arrowheads) (note, direction of migration cannot be determined from this static image). Abbreviations: FNP, frontonasal process; H, heart; TG, trigeminal ganglia; Mx, maxillary process; Md, mandibular process; PA1, first pharyngeal arch; PA2, second pharyngeal arch; PA3, third pharyngeal arch; PA4, fourth pharyngeal arch; OV, otic vesicle; DRG, dorsal root ganglia.

1.1.2 **Palatogenesis and Cleft Palate**

In the mouse, at embryonic day (E)11.5, the secondary palate begins to appear as an outgrowth of the maxillary prominences. At E13.5 the palatal shelves grow vertically along the sides of the tongue and then at E14 as the tongue begins to drop in the oral cavity due to forward and downward growth of the mandible, the palatal shelves elevate above the tongue (Lough *et al.,* 2017). After they elevate, the bilateral palatal shelves approximate and establish contact, first in the region of the second ruga (middle third of the palate) from which point fusion spreads in posterior and anterior directions (Ferguson, 1988). So, by E14.5, the palatal shelves are now in contact with one another at the midline and form a seam, often referred to as the medial epithelial seam (MES). Finally, by around E15.5-E16.6 the shelves fully fuse **(Figure 1-2)**. In humans, palatogenesis is initiated in the 6th week of gestation, lip and primary palate fusion is completed between the 4th and 9th week, and secondary palate fusion is completed by 12 weeks (Bush and Jiang, 2012; Cox *et al.,* 2004).

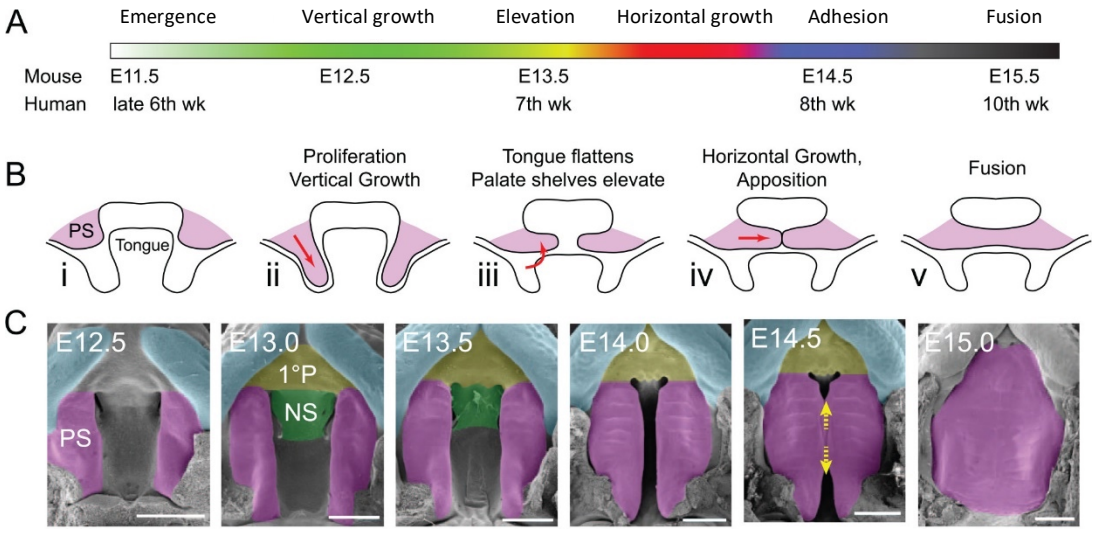


Figure 1-2 Comparison between human and mouse secondary palatal development

(A) Timeline of morphogenetic processes that occur during palate growth and closure in mice and humans. Human data is based on the timing of hard palate closure, with soft palate fusion occurring later. (B) Schematics, in the coronal plane, of the position of the secondary palatal shelves (PS, purple) relative to the tongue during representative stages of palatogenesis. PS initiate outgrowth from the maxillary prominence at ~E11.5 to E12 (i), depending on the mouse strain (Walker 1956), and initially grow downward (ii) before elevating above the tongue at ~E13.5 to E14.0 (iii). Horizontal growth follows until opposing medial edge epithelia (MEE) meet at the midline (iv). PS fusion occurs between E14.5 and E15.5 and proceeds anteriorly and posteriorly over the course of ~6 h (Walker 1956) (v). (C) Scanning electron microscopy images of the roof of the mouth at indicated ages. Taken from (Lough *et al.*, 2017).

Failure in any of the mechanisms outlined above can lead to palatal clefts. Many signalling pathways have been implicated in the aetiology of such defects with some examples elucidated below. Palatal shelf defects could be divided into the following categories:

Failure of palatal shelf formation and growth

Growth of the palatal shelves is the first key step towards successful palatal development. The vertical outgrowth of the shelves bilaterally is controlled by reciprocal epithelial-mesenchymal interactions and involves distinct molecular mechanisms and morphological heterogeneity along the anterior-posterior and medio-lateral axes of the developing palate (Bush and Jiang, 2012; Ferguson and Honig, 1984; Tyler and Koch, 1977). There are key signalling pathways that drive early palatal shelf growth.

Fibroblast growth factor ligands (FGFs) and their receptors are implicated in early palatal development by mediating a variety of cellular responses (Tabler *et al.*, 2016; Weng *et al.*, 2018; Wu *et al.*, 2015c). Mutations in *Fgf10* and *FgfR2b* have been shown to affect the initial development of the palatal shelves in mice (Rice *et al.*, 2004). In a normal state, *Fgf10* in the mesenchyme signals via its receptor *Fgfr2b* in the palatal shelf epithelium supporting epithelial cell proliferation and inducing epithelial *Shh*. In return, *Shh* signals back to the mesenchyme and stimulates cell proliferation (Rice *et al.*, 2004). Lack of SHH signalling in the epithelium results in decreased levels of proliferation in palatal mesenchymal cells necessary for shelf growth (Rice *et al.,* 2004; Stanier and Paws, 2012). Crosstalk between SHH and BMP signalling has also been shown to regulate the growth of the anterior region of the palate (Zhang *et al.,* 2002). A feedback loop between *Bmp4* in the mesenchyme maintains *Shh* expression in the MEE and *Shh* in turn induces *Bmp2* expression in the mesenchyme, with the latter inducing mesenchymal cell proliferation thus leading to palatal growth (Zhang *et al.,* 2002).

A number of transcription factors have also been shown to be modulated by *Fgfs,* particularly *Fgf8.* For instance, T-box transcription factor (*Tbx1)* knock out mice display primary defects in palate elongation and aberrant palatal proliferation that may be explained by decreased *Fgf8* and increased *Fgf10* expression (Goudy *et al.*, 2010). Interestingly, dysregulation of FGF function has been attributed to craniofacial defects in ciliopathic mutant mice and that genetic reduction of *Fgf8* rescues maxillary hyperplasia and high-arched palates seen in ciliopathic *Fuz* mutant mice (Tabler *et al.*, 2013).

Failure of palatal shelf elevation

The process of palatal shelf elevation is very rapid and is triggered by intrinsic forces and signalling within the palatal shelves, in addition to extrinsic factors from other craniofacial structures that ensue temporally and spatially in concert with shelf elevation (Bush and Jiang, 2012). A number of factors have been proposed to drive palatal shelf elevation, these include the forward displacement of the tongue, the downward and forward growth of the mandible, mesenchymal cell remodelling and/or the accumulation of glycosoaminoglycans (Ferguson, 1988; Jin *et al.,* 2010). Therefore, malformation of these structures could also lead to failure in the elevation process. Further, it has been observed that the two palatal shelves do not elevate in unison, but rather sequentially in a very short time frame (Jin *et al.,* 2010; Peters *et al.,* 1998).

A critical criterion to successful elevation is to overcome the physical obstruction of the tongue. This has been shown in the *Gli3* deficient mouse model (Huang *et al.*, 2008). Severe delays in palatal shelf elevation was found in *Gli3*-mutants accompanied by defective tongue development, where the size of the palatal shelves was comparable with wild-type littermates and no differences in proliferation in the epithelium or mesenchymal compartments of the palatal shelves at any stage of palatogenesis, nor abnormal apoptosis were found (Huang *et al.*, 2008). Jin *et al.* (2010) proposed a model that explains how palatal shelves overcome the physical obstruction of the tongue during elevation and re-orientation by showing that the horizontally positioned medial edge epithelium is in effect an outgrowth from the side of the vertical shelf rather than its conventionally assumed position (the distal end of the vertical shelf). Therefore, the authors proposed that the movement of the distal end of the palatal shelf above the dorsal level of the tongue was a rather inconsequential step towards palatal elevation (Jin *et al.,* 2010).

Murray *et al.,* (2007) proposed that the forward growth of the mandible, driven by the extension of the neural-crest-derived Meckel's cartilage, provides the mechanism to lower the tongue, thus permitting palatal shelf elevation. They showed that Meckel's cartilage was dramatically shorter in *Snai1/2*-double knockout embryos compared to controls (Murray *et al.,* 2007). A similar mechanism responsible for cleft palate is seen in mouse models for Pierre Robin Sequence supported by analyses of the A/WySn mouse strain, a strain that shows a highly spontaneous, genetically determined clefting rate of up to 20% to 44% (Schubert *et al.,* 2005), and mice heterozygous for *Dmm* (Ricks *et al.,* 2002).

The localised accumulation and hydration of glycosaminoglycans, predominantly hyaluronic acid, in the palatal mesenchyme has also been proposed as the force driving the elevation of the palatal shelves (Ferguson, 1988; Pelton *et al.,* 1990). The hypothesis being that the regional build-up of hyaluronic acid results in the swelling of the extracellular matrix and a decrease in mesenchyme cell density (Ferguson, 1988). *In vitro* experiments with cultured palatal mesenchymal cells have shown that TGF- β1 stimulates the production of hyaluronic acid, hence playing a critical role in regulating elevation (Dixon and Ferguson, 1992; Ferguson *et al.,* 1988).

The intrinsic function of some genes has also been shown to be involved in palatal elevation. *Pax9*-deficient mice have been shown to have unelevated palatal shelves owing to mechanical hindrance (Peters *et al.,* 1998). However, since unilateral palatal shelf elevation was observed in some cases, Peters and co-workers (1998) proposed that *Pax9* was not necessary for the capability of the shelves to elevate but is required to regulate their shape around embryonic day (E) 13.5, a critical stage of secondary palate formation (Peters *et al.,* 1998). In *Fgf10* null embryos, anomalous adhesions between the epithelium of the palatal shelves and the tongue anteriorly and the mandible in the middle and posteriorly, prevented the elevation of the shelves, which might be a cause for the cleft palate phenotypes seen in these mice (Alappat *et al.,* 2005; Rice *et al.,* 2004). In the Wnt/β-catenin signalling pathway, glycogen synthase kinase-3b (Gsk3b) functions in facilitating the degradation of β-catenin. He *et al.* (2010) demonstrated that *Gsk3b* is required in the epithelium, primarily functioning in the medial edge epithelium (MEE) for normal palatal development. Mice lacking *Gsk3b* also showed impaired elevation of the palatal shelves leading to cleft palate. The regulation of palatal shelf elevation via *Gsk3b* was demonstrated to be independent of alterations in both *Axin2* (another direct target of the Wnt/β-catenin signalling pathway) and *Shh* signaling (He *et al.*, 2010), suggesting that *Gsk3b* is an intrinsic regulator of palatal shelf elevation.

Failure of palatal shelves fusion

Failure of fusion of the palatal shelves is the most common type of defect leading to cleft palate in animal models and is the most extensively studied type both *in vivo* and *ex vivo*. Persistence of the medial edge epithelium (MEE) is one of the major reasons why this occurs. The MEE comprises of two layers of epithelial cells; a flat layer at the leading edge of the palatal shelf called the periderm, and an underlying layer of basal cuboidal cells on a basement membrane (Smiley and Dixon, 1968; DeAngelis and

Nalbandian, 1968). Theories such as epithelial-mesenchymal transition (EMT), programmed cell death and apoptosis and migration of the periderm cells from the MEE to the oral and nasal epithelial triangles have all been proposed to study palatal shelf fusion and consequently confluence of the mesenchyme (Cuervo & Covarrubias, 2004; Cuervo *et al.*, 2002; Dudas *et al.*, 2007; He *et al.*, 2008; Martinez-Alvarez *et al.*, 2000; Nawshad, 2008) .

Using heterologous palatal explant cultures from mouse, chick and alligator embryos, Ferguson *et al.,* (1984) demonstrated that palatal closure did not occur between combinations of shelves from the different species when their MEE was cultured in contact, implying that medial edge epithelium adherence is specific (Ferguson *et al.,* 1984). Yet, when homologous pairs of palatal explants were cultured, the shelves fused as seen *in vivo* in all species (apart from the chick that has a physiologic cleft), suggesting that the palates have intrinsic species-specific epithelial mesenchymal crosstalk. To further resolve this interaction, in a series of sophisticated experiments, Ferguson and Honig (1984) carried out epithelial-mesenchymal recombination explants between various epithelial/mesenchymal tissues from the species mentioned above (Ferguson and Honig, 1984). They demonstrated that regional palatal (nasal, medial edge and oral) epithelial differentiation is specified by the mesenchyme, probably via a complex interaction of extracellular matrix molecules and growth factors (Sharpe and Ferguson, 1988), which were not yet specified at the time, and that signalling of medial edge epithelial differentiation goes across different vertebrates in a species-specific fashion (Ferguson and Honig, 1984). Following contact of the shelves and epithelial seam formation in mouse explants, Ferguson *et al.,* (1984), showed that thinning of the seam was achieved through an expansion in palatal height (oronasally), epithelial cell migration onto the oral and nasal aspects of the palate and apoptotic cell death (Ferguson *et al.,* 1984; Ferguson *et al.,* 1988).

Transforming growth factor beta-3 (*Tgfb3)* has been a strong contender in the study of palatal fusion, irrespective of the mechanism in question. *Tgfb3* has been shown to both mediate medial edge epithelium fusion through the flipodia (long cellular protrusions on the surface of the MEE) and to induce apoptosis of the medial epithelial seam (MES). Consequently, *Tgfb3* mutant mice exhibit cleft palate due to failure of complete fusion between the palatal shelves. The effect of *Tgfb3* on the MES regression is mediated by the *Tgfb* type II and type I receptor (Alk5)/Smad pathway (Cui *et al.*, 2005; Xu *et al.*, 2006). Interestingly, using the EMT model, Cui *et al*. (2005) showed that overexpression of *Smad2 in vivo* in the MEE of *Tgfb3* mutant palates rescued the fusion

defect in some, but not in all cases. The authors proposed that this was attributed to *Smad2* being able to bypass an initial requirement for Tgfb3 ligand binding to the Tgfb receptors to initiate Smad2 phosphorylation in the MEE. However, the fact that the palate was rescued only partially could either be due to lack of *Tgfb* signalling required for the induction of other mesenchymal factors necessary for epithelial remodelling in the *Tgfb3* mutant mice (Gritli-Linde, 2007), or due to other unexamined factors necessary for completion of the epithelial-to-mesenchymal transition process. He *et al.* (2011) on the other hand used a β-catenin mutant model and found a downregulation of *Tgfb3* expression in the MEE and suppression of apoptosis in the MEE cells. Indeed, they showed that palatal cultures of E13.5 mutant embryos exhibited rescue of the fusion defect following the addition of exogenous Tgfb3 protein. He *et al.* (2011) were the first to look at upstream signalling cascades regulating *Tgfb3* in the context of palatal development and confirmed that Wnt/β-catenin was essential for the activation of *Tgfb3* in the palatal epithelium thus directing palatal fusion. Others studied the downstream signalling cascades of *Tgfb3* in palatal development demonstrating that loss of Epithelial (E)-cadherin in palatal MEE was essential for palatal fusion to occur and that Lef1 mediated by the Smad complex (Smad2-P and Smad4) acted as a transcriptional repressor of E-cadherin in response to Tgfb3 signalling during epithelial-mesenchymal transition of the MEE cells (Nawshad *et al.*, 2007). This is an interesting finding, as Lef1 has commonly been known to be activated by the Wnt/β-catenin pathway. Other modulators of *Tgfb3* at the MEE during epithelial-mesenchymal transition have been shown to include *Irf6,* a known cleft gene. Knockdown of *Irf6* has been shown to delay TGFβ3 mediated palatal fusion (Ke *et al.*, 2015). Indeed, *Irf6* mutant mice exhibited abnormal periderm development which subsequently altered the palatal shelves' ability to complete the fusion process (Richardson *et al.*, 2009). It has been demonstrated that TGFβ3-mediated down-regulation of the transcription factor p63, a key regulator of ectodermal and orofacial development, in the MEE allows periderm migration out of the medial epithelial seam and that p63 and TGFβ3 function in concert with IRF6 to drive medial edge epithelium cell fate (Hammond *et al.*, 2017). The various mechanisms underlying the disappearance and fate of the MES and the key studies supporting these mechanisms are described further in Section 1.1.3.1 below.

1.1.3 Tissue Contributions and Anatomical Variations in Craniofacial Development and Disease

1.1.3.1 The palate, tooth and heart as models for tissue-to-tissue crosstalk during development

Tissue-tissue interactions such as in neural crest cells and their migratory abilities and epithelial-mesenchymal interactions are fundamental processes driving development of many craniofacial and other structures.

Mechanisms that drive the disappearance of the medial epithelial seam

The interaction between the ectoderm and the underlying mesenchyme is a fundamental mechanism during the development of many organs. The tooth and palate are excellent models depicting this process (Zhang *et al.,* 2002). In the palate, this interaction is unique whereby a key step in palatal fusion requires the dissolution of the medial epithelial seam (MES) following contact of the medial-edge epithelia (MEE) of the opposing palatal shelves post elevation (Ferguson and Honig, 1984; Tyler and Koch, 1977; Zhang *et al.,* 2002) **(Figure 1-3)**. Epithelium in the developing palate is generally divided into oral, nasal and medial edge epithelium. While the oral and nasal epithelia are later differentiated into pseudo-stratified and squamous epithelia, the medial edge epithelium must be removed from the fusion line between the opposing palatal shelves for normal palatal development to take place (Ferguson and Honig, 1984; Tyler and Koch, 1977; Zhang *et al.,* 2002). If this process fails, confluence of the palatal mesenchyme fails to occur leading to palatal clefting.

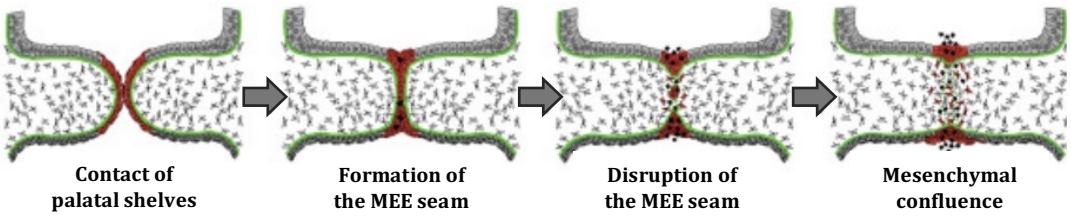


Figure 1-3 Summary diagram illustrating the process of MEE cell disappearance in the palate

Palatal shelf contact and midline epithelial seam formation are prerequisites for medial-edge epithelium (MEE) cell differentiation as well as palatal fusion. MEE cells can disappear from the medial edge by means of apoptosis, epithelial-mesenchymal transformation and epithelial cell migration toward the oral and nasal epithelia. Red cells represent MEE cells and MEE-derived cells; black cells represent dead cells; green lines and dots indicate the basement membrane and its fragments, respectively. Adapted from (Takigawa & Shiota, 2004).

Epithelial-to-mesenchymal transition is a programme that underlies a variety of tissue remodelling events during development and is essential for orchestrating many biological processes including neural crest and heart valve development and has been proposed as a mechanism for secondary palate fusion (Yang & Weinberg, 2008). During heart valve development and secondary palate fusion, epithelial-to-mesenchymal transition occurs in relatively well-differentiated epithelial cells that are destined to become defined mesenchymal cell types (Yang & Weinberg, 2008).

Establishment and dissolution of the medial edge epithelia is tightly controlled and involves an integration of factors such as extracellular signalling and epithelial turnover. In the palate, blocking fusion of the palatal shelves has been shown to prevent basal epithelial cells from undergoing EMT, suggesting that signals from the midline epithelial seam are critical to triggering the EMT process (Griffith & Hay, 1992). Signals mediated through the TGF-β superfamily have been implicated as major induction signals of EMT in the medial edge epithelia (Cui *et al.*, 2005; Yang & Weinberg, 2008). Although not fully determined in the palate, other developmental studies suggest that the TGF-β pathway collaborates with Wnt, Notch and tyrosine kinase receptors to facilitate EMT in various morphogenetic processes (Yang & Weinberg, 2008). As I will point to further in Chapter 4, the process of EMT also requires the removal of the E-cadherin complex from the medial epithelial seam in the palate. Research has shown that E-cadherin is downregulated by the initiation of epithelial-to-mesenchymal transition (Sun *et al.*, 1998). Surprisingly, little is known about this turnover in the palatal seam. One possibility is control from the catenin-cadherin complex, which is known to stabilize cell junctions and perhaps is important in the assembly and disassembly of the medial edge epithelia. Mutations that disrupt this interaction could potentially result in failure of palatal fusion.

The concept of EMT in the palate was first proposed by Fitchett and Hay (1989) using transmission electron microscopy demonstrating that cell death occurred only in the periderm cells where they begin to slough after the shelves assume a horizontal position. They showed that this allowed the opposing basal MEE cells to form junctions and transform to mesenchymal cells which then was followed by the expression of vimentin and degradation of the basal lamina (Fitchett & Hay 1989).

While EMT is one of several mechanisms that lead to palatal shelf fusion, other modes have investigated the fate of the medial edge epithelia, all signifying the need for the medial epithelial seam to disappear in order for mesenchymal confluence of the opposing palatal shelves to take place. Sani-Vaziri *et al.* (2005) ruled out the occurrence

of EMT during palatal confluence by using irreversible cell marking to trace the fate of MES cells *in vivo* during palatal shelf contact and fusion. By using the *Rosa-loxP-stop-lacZ* reporter mice (*R26R*) – that when crossed with epithelially-restricted cre expressing transgenic strains *ShhGFPCre* or *K14-Cre* , the STOP sequence is removed and lacZ is expressed in cells/tissues where cre is expressed (Soriano, 1999) – they hypothesised that if EMT was the expected mechanism, lacZ-positive mesenchymal cells would be visible in the fused palate (Sani-Vaziri *et al*., 2005). Their results, however, demonstrated the disappearance of the MES cells from the confluent palate and the total lack of β-galactosidase activity in the mesenchymal cells in any of the developmental stages examined (E15-E18.5) (Sani-Vaziri *et al*., 2005). Further, they confirmed that the lacZ-positive MES regressed by undergoing apoptosis demonstrated by their immunoreactivity for activated caspase-3 (Sani-Vaziri *et al*., 2005). Around the same time, another group (Xu *et al.,* 2006) used a similar approach where they carried out lineage tracing using the *K14-Cre; R26R* mouse and expanded the results demonstrated by Sani-Vaziri *et al.* (2005) by showing that K14-Cre;R26R embryos from E12.5 to newborn did not reveal any β-galactosidase-positive cells in the palatal mesenchyme when examining serial sections through the entirety of their palates (Xu *et al.,* 2006).

The results above lend further support to the presence of various other mechanisms regulating the degeneration of the MES. A widely accepted candidate is programmed cell death. Using TUNEL staining for the detection of nuclear DNA fragmentation, a hallmark of programmed cell death, Mori and co-workers (1994) observed cytochemical evidence for death in the cells of the disappearing epithelial seam, corroborating earlier reports that used morphological criteria to suggest apoptosis during fusion of the secondary palatal shelves (Ferguson, 1988; Mori *et al.,* 1994; Shapiro and Sweney, 1969) and paving the way for newer techniques to confirm this mechanism in other palatal processes too (Alappat *et al.,* 2005; Cuervo and Covarrubias, 2004; Lee *et al.,* 2008; Rice *et al.,* 2004).

Using siRNA knockdown of *Wnt11* on murine palatal explants, Lee *et al.* (2008) showed a disruption in the process of palatal fusion due to the staggering of apoptotic cells at the MES when *Wnt11* was knockdown and the significantly higher mRNA levels of *Fgfr1b* (Lee *et al.,* 2008). The authors attributed the importance of *Wnt11* in the induction of apoptosis in the palatal epithelium and mesenchyme through inhibition of *Fgfr1b* (Lee *et al.,* 2008). FGF signalling has been shown to play significant roles in palatogenesis through its interaction with other pathways, particularly highlighting the role of apoptosis (Alappat *et al.,* 2005; Rice *et al.,* 2004). Indeed, *Fgf10* and *Fgfr* knockout

mice all exhibit cleft palate because of increased apoptosis in the palatal shelves, albeit at earlier stages than palatal fusion, among other mechanisms such as decreased cell proliferation in the palatal epithelium and/or anomalous fusion of the palate with the tongue or mandible (Alappat *et al.,* 2005; Rice *et al.,* 2004).

A third argument for the fate of the medial edge epithelial cells is their migration to other compartments of the palate. Carette and Ferguson (1992) reported the migration of the cells orally and nasally during seam disruption and their incorporation into the oral and nasal epithelia to form the oral and nasal epithelial triangles (Carette and Ferguson, 1992). This mechanism has formed the basis for a number of elegant studies looking into the function of the periderm during palatal fusion, lending support to this hypothesis. It was shown that it is the periderm cells, in particular, that migrate to form the epithelial triangles (Cuervo and Covarrubias, 2004).

The periderm is a transient epithelial layer that comprises the outer most surface of the MEE. It is crucial during embryogenesis as it acts as a protective barrier by preventing pathologic or premature adhesions between exposed basal cells within intimately opposed epithelial layers (Hammond *et al.,* 2017; Richardson *et al.,* 2014). Nevertheless, the periderm must be removed at a precise time to initiate the fusion process between the palatal shelves and for successful completion of palatal confluency (Hammond *et al.,* 2017; Richardson *et al.,* 2014). One such method is the migration of the periderm cells out of the midline seam (Cuervo and Covarrubias, 2004; Richardson *et al.,* 2017). Cuervo and Covarrubias (2004) inhibited cell migration with cytochalasin D, which blocks actin polymerization, and showed that periderm cells did not migrate, epithelial triangles did not form, and complete adhesion of palatal shelves did not occur as cell death failed to be triggered (Cuervo and Covarrubias, 2004).

Newer theories are evolving, and older hypotheses are being refined. For example, Kim and co-workers (2015) expanded on a less well explored model for fusion of the palate which is the process of convergent extension (Kim *et al.,* 2015). Cuervo and Covarrubias (2004) have alluded to the possibility of the basal MEE cells of each shelf intercalating via convergent extension resulting in a single epithelial layer (Cuervo and Covarrubias, 2004). Kim *et al.* (2015) found that cellular protrusions from the opposing shelves establish contacts and cellular bridges, leading to the formation of a transient multicellular epithelial seam that converges towards the midline. Driven by an actomyosin contractility pathway, the convergence occurs along with the displacement of the epithelium and apoptotic cell extrusions that squeeze epithelial cells out from between the palatal shelves, mediating mesenchymal continuity of the palate (Kim *et al.,*

2015). Taken together, the variability in the outcome seen between the myriad of studies carried out on the topic of palatal fusion is perhaps a reflection of the differences in the techniques used, the stages studied, and sections of the palate examined. Whether it is EMT, apoptosis, or migration, perhaps multiple mechanisms take place in the right place at the right time spatio-temporally to achieve palatal confluency.

Epithelial-mesenchymal interactions during tooth development

During tooth formation, sequential and reciprocal interactions occur between the epithelium and the neural crest-derived mesenchyme, with signalling molecules mediating this communication. Three epithelial signalling centres exist in the tooth to regulate these steps. The placodes which are thickenings present in the dental lamina, the primary enamel knot, and the secondary enamel knots (Cobourne & Sharpe, 2003; Miletich & Sharpe, 2004).

Prior to any histological sign of tooth development, the jaws begin a pattern of restricted expression of certain homeobox genes that later contribute to the patterning of teeth in precise positions in the jaws (Miletich and Sharpe, 2004). Formation of molar teeth in proximal areas of the jaws are directed by homeobox genes like *Dlx1*, *Dlx2* and *Barx1* whereas other homeobox genes such as *Msx1, Msx2, Alx3* and *Alx4*, expressed in the distal parts of the ectomesenchymes of the jaws, result in the formation of incisor teeth (Cobourne and Sharpe 2003; Miletich and Sharpe, 2004). Determination of tooth forming sites in the jaws occurs around embryonic day (E) 10.5 and by E11.5 in mice and six weeks (wk) of gestation in humans, thickenings of the dental epithelium (known as the lamina stage of tooth development) provide the first morphological evidence of tooth formation (Zhang *et al.,* 2005). This is followed by the early bud (E12.5, 7wk), late bud (E13.5, 9-10wk) and cap (E14.5, 11-12wk) stages of tooth development (Zhang *et al.,* 2005).

During the initiation stage, four spots of *Shh* expression in the epithelium mark the sites of the developing tooth germs where *Shh* regulates epithelial cell proliferation to produce a tooth bud (Tucker and Sharpe, 2004). The first sign of interaction between epithelium and mesenchyme in tooth development is observed when signals from *Fgf8* and *Bmp4* in the oral epithelium influence the location of mesenchymal expression of *Pax9*, the earliest mesenchymal marker for the site of future tooth buds. *Pax9* is positively regulated by *Fgf8* and negatively regulated by *Bmp4* (Tucker and Sharpe, 2004). In mouse knock-out models, mice that are $Pax9^{-/-}$ lack all teeth and exhibit cleft palate (Peters *et al.,* 1998). Furthermore, *Fgf8* and *Bmp4* are capable of inducing gene

expression in the dental mesenchyme via *Msx1* which reciprocates this interaction thus playing an important role in mediating epithelial-mesenchymal interactions during tooth and palate development (Bei and Mass, 1998). The interaction between *Bmp4* and *Msx1* ensures that expression of *Msx1* remains localized to the developing ectomesenchyme of the tooth bud (Cobourne and Sharpe, 2003).

Other signalling molecules and transcription factors that are implicated in palatogenesis and/or CLP have also been shown to play important roles in tooth development (Menezes and Viera, 2008; Letra *et al.,* 2009; Letra *et al.,* 2012). WNT is a fundamental pathway in craniofacial development. Axin2, Lef1 and Wnt10a are few of the proteins involved in WNT signalling that have been reported to play significant roles in palate and tooth development. The function of *Lef1* is critical around the bud stages of tooth formation by linking the WNT and FGF signalling pathways (Kratochwil *et al.,* 1996; Kratochwil *et al.,* 2002; Vadlamudi *et al.,* 2005). In their model, Shimizu *et al.* (2013) proposed that *Wnta10a* acted upstream of *Lef1*, and that *Fgf3* and *Fgf4* were downstream targets for *Lef1* that mediate epithelial-mesenchymal interactions in early tooth development (Shimizu *et al.,* 2013). Chen *et al.* (1996) proposed a regulatory hierarchy between *Msx1*, *Lef1* and *Bmp4* in early mouse molar tooth development. They showed a 10-fold decrease in the levels of *Bmp4* and *Lef1* transcripts in the *Msx1* mutant E14.5 dental mesenchyme, indicating that *Msx1* is required for the normal expression of *Bmp4* and *Lef1* in the dental mesenchyme (Chen *et al.,* 1996). Despite its importance in tooth development, the *LEF1* gene has not been implicated in cases of human hypodontia thus far. This perhaps is due to the fact that its function could well be overcome by other proteins involved in tooth development. Finally*, Wnt10a* has been shown to function at multiple stages of tooth development. *WNT10A* encodes components of the Wnt/β -catenin signalling pathway and functions as a ligand for the frizzled family of receptors (Liu and Millar, 2010). It is particularly important at the earlier stages where it facilitates in activating the dental mesenchyme to form the tooth (Mostowska *et al.,* 2013). It also plays a role in the induction and maintenance of the enamel knot (Mostowska *et al.,* 2013).

The role of the neural crest during development of the heart

The heart is the first organ to form in vertebrates (Buckingham *et al.,* 2005). Cardiogenesis begins with the migration of myocardial progenitor cells from the primitive streak to the anterior segment of the embryo at about embryonic day (E) 6.5 in mice. At E7.5, the progenitor cells align under the head folds and form the cardiac

crescent where differentiated myocardial cells are now observed. By E8, the cardiac crescent fuses with the midline to form the early cardiac tube and by E8.5 the tube undergoes rightward looping (Buckingham *et al.,* 2005). By E10.5 the heart has acquired well-defined chambers but still resembles a tube. The looping process and expansion of the myocardium lead to the formation of recognisable cardiac chambers. By E14.5, the chambers are now separated as a result of septation and are connected to the pulmonary trunk and aorta (Buckingham *et al.,* 2005). Additionally, molecular data established that there are two heart fields with progenitor cells that make distinct regional contributions to the heart (Buckingham *et al.,* 2005). Heart morphogenesis involves the complex cycling between epithelial-to-mesenchymal transition (EMT) and mesenchymal to epithelial transition (MET) that is initiated during gastrulation (Lim and Thiery, 2012). In the heart, induction of EMT in the atrioventricular endocardial cells is modulated by signalling pathways such as TGF-β1, 2 ,3 and BMP, canonical Wnts, Notch1 and transcription factors such as Snail and Slug, with the process completed by establishing the presumptive cardiac septa and valves (Person *et al.*, 2005).

As outlined in Section 1.1.1, several neurocristopathies comprise craniofacial abnormalities, including CLP, and congenital heart disease (CHD) as the two major anomalies, and as discussed in Section 1.2.3.2, many individuals with orofacial clefts have CHD as an associated anomaly. The co-occurrence of these two congenital conditions could either be due to shared genetic contributions (Homsy *et al.*, 2015), due to neural crest migration and/or differentiation defects, or both. Human neurocristopathic cardiac anomalies include persistent truncus arteriosus (no outflow tract septation), double outlet right ventricle or Tetralogy of Fallot (abnormal septation), or abnormal patterning of the aortic arch arteries leading to interruption of the aortic arch and aberrant vascular structures (Abu-Issa *et al.,* 2002).

Neural crest cells (NCCs) are essential for cardiovascular patterning (Abu-Issa *et al.,* 2002). Indeed, mesenchymal cells that contribute to the septation of the heart arise from migrating neural crest cells (Buckingham *et al.,* 2005). The cardiac neural crest cells arise from the region of the cranial neural fold at the level of the otic placode and the caudal end of the third somite. This neural crest cell-population migrates into the cardiac outflow tract and contributes to the aorticopulmonary septum, conotruncal cushions, and differentiate into smooth muscle cells populating the derivatives of the third, fourth and sixth pharyngeal arch arteries (Jiang, 2000; Kirby and Waldo, 1990; Kockilas *et al.,* 2002; Trainor, 2005). Ablation of cardiac NCCs results in characteristic

cardiovascular abnormalities such as failure of outflow tract septation and aortic arch artery defects (Abu-Issa *et al.,* 2002; Kirby and Waldo, 1990).

Genetic mutations have also been shown to cause both craniofacial and cardiac anomalies through aberrant neural crest function. The transcription factor *Hand2* (heart and neural crest derivatives expressed transcript 2, dHAND) is expressed in the pharyngeal arches and has been implicated in NCC development. *Hand2* null mouse embryos are embryonically lethal at E10.5 due to heart failure. These mouse mutants also display hypoplasia of the pharyngeal arches as well as cardiovascular patterning defects (Srivastava *et al.,* 1997). Interestingly, *Hand2* and *Hand1* (heart and neural crest derivatives expressed transcript 1, eHAND) represent the earliest cardiac chamber-specific transcription factors owing to their regionally-restricted expression patterns (Srivastava *et al.,* 1997). *Hand2* is expressed throughout the straight heart tube until it becomes restricted to the conotruncus and future right ventricle as the heart tube forms a loop. By E10, its expression is downregulated in the cardiac mesoderm but is maintained in the neural crest-derived aortic arch arteries (Srivastava *et al.,* 1997). Conversely, *Hand1* expression is restricted to the rostral and caudal regions of the straight heart tube which are fated to form the conotruncus and left ventricle, respectively (Srivastava *et al.,* 1997).

FGF signalling is necessary for the proliferation of cells in the second heart field (Buckingham *et al.,* 2005). *Fgf8*$^{neo/-}$ mutant mouse embryos show perfectly migratory NCCs, however, the NCCs undergo cell death as soon as they reach areas that are both adjacent and distal to where *Fgf8* is normally expressed (Abu-Issa *et al.,* 2002). Although *Fgf8* is not expressed in the neural crest – but is normally expressed in the developing pharyngeal arch ectoderm and endoderm – neural crest cells migrate from the mesenchyme of the pharyngeal arches to the outflow region of the heart, where they are consequently compromised by abnormal FGF signalling in the mutant (Abu-Issa *et al.,* 2002). Indeed, the *Fgf8*$^{neo/-}$ mice have malformations of the aorta and pulmonary trunk and the aberrant effect of *Fgf8* on the NCCs is a reflection of theses defects since NCCs are required for the septation and normal development of these structures (Abu-Issa *et al.,* 2002; Buckingham *et al.,* 2005).

1.1.3.2 Timing and variability of events and physical constraints in craniofacial development

Anatomical influences in CLP are understudied. Evidence is mainly based on hypothetical reasoning prior to the gene discovery era (Fraser *et al.*, 1957) and more

recent evidence is based on phenotypes observed in mouse models. One factor to consider for instance, is the *timing of the pathogenic event*. Could there be a critical response period for which a cleft could be prevented? In the context of palatogenesis, variability in the timing of palatal closure is likely unique to each individual; indeed, variability in embryogenesis has been demonstrated in mouse foetuses (Thiel *et al.*, 1993; Wahlsten & Wainwright, 1977). Most of the hypotheses on timing come from conditional mouse lines. Genetically modified mice, engineered for the time of the insult (by deleting a gene crucial in development or introducing a teratogen), showed various responses to developing cleft anomalies (Harbison & Becker, 1969; Liu *et al.*, 2007). Equally, small changes in overall development of the craniofacial complex, such as onset of ossification could constrain growth of the palate. *Physical or indirect effects* from anatomically adjacent craniofacial and/or oral structures are other factors to consider.

In some cases, differences in palatal morphology could be due to a direct or indirect physical effect that could alter growth and development of the palate. For instance, the tongue is constantly in contact placing pressure on the palate in severe micrognathia and retrognathia, a consequence exemplified by the narrow, high-arched palate seen in Rubinstein–Taybi syndrome [MIM: # 180849] in which 100% of mice were also shown to have a narrow palate (Oike *et al.*, 1999). Infants diagnosed with Pierre Robin Sequence (PRS) [MIM: # 261800] are often born with a cleft palate likely due to severe micrognathia and glossoptosis (Breugem *et al.*, 2016). Developmentally, it is thought that Meckel's cartilage and mandibular outgrowth defects and/or palatal shelf obstruction by failure of the tongue to descend due to muscular defects underly the pathogenesis of PRS (Tan *et al.*, 2013). This is now coupled with recent evidence for genetic contributions to PRS (Benko *et al.*, 2009; Jakobsen *et al.*, 2006). Based on these examples, one would question whether macroglossia (enlarged tongue) could also influence how the palate shapes? Indeed, patients with Down syndrome [MIM: # 190685] are known to have macroglossia and narrow palates, a correlation that has not been tested. Nonetheless, mouse models of Down syndrome were shown to have narrower palates rostro-caudally compared to their littermate controls (Hill *et al.*, 2007). In an overgrowth syndrome characterized by macroglossia, macrosomia, and abdominal wall defects, Beckwith-Wiedemann syndrome [MIM: # 130650] (BWS) patients have also been shown to have intraoral features of macroglossia and cleft palates (Romanelli *et al.*, 2010). Contrary to this observation, a question arises of whether patients with neuromuscular defects have shallower palates. Nonetheless, mouse models of neuromuscular defects with mutations in the neurotransmitter gamma-aminobutyric acid (GABA) or *Gad1*, a gene that encodes a component required

for GABA neurotransmission including GABA synthesis, showed that the mice had cleft palates (Condie *et al.*, 1997) and that in *Gad1*$^{-/-}$ mice, cleft palates occurred due to the abnormal position of the tongue between the palatal shelves (Iseki *et al.*, 2007; Oh *et al.*, 2010). Regardless, whether it is because of a small mandible or an abnormally enlarged tongue, the tongue is an organ that could cause anatomical obstruction of timely palatal closure.

Differences in cranial base angles and restrictions during growth in some syndromes could also pose indirect effects on the palatal shelves. Craniofacial malformations such as craniosynostosis, mid-face hypoplasia, or fusion of the frontal bones have been thought to indirectly affect the palate because of the effects they have on growth and development of the maxilla. Factors as such could be summed up in what is known as the Functional Matrix Hypothesis (Moss, 1968; 1997). Lastly, some variability in the palatal phenotypes seen in some children could, in fact, be iatrogenic. For instance, palatal grooves could be acquired from continuous intubation for very long periods of time during infancy (Hennekam *et al.*, 2010). All in all, these factors could perhaps be contributing to the differing palatal anomalies seen, but do not fully explain why the palate appears to have a hyperplastic, high-arched appearance in some cases or why it has a 'true' cleft in others.

1.2 PHENOTYPIC VARIABILITIES IN CLEFT RESEARCH

1.2.1 Cleft Lip/Palate

Clefts of the primary and/or secondary palate (CLP) are one of the most common structural human birth defects and are recognised as the most common craniofacial congenital abnormalities worldwide. Birth prevalence rates vary based on the ethnic groups reported in epidemiological surveys, with Asians (1:500) and Caucasians (1:700) being the most affected and Africans being the least affected populations (1:2500) (Tanaka *et al.,* 2012; Vanderas, 1987). Orofacial clefts are described based on the structures involved (lip, alveolus, hard palate, soft palate), the laterality of the defect (unilateral left, unilateral right, median, bilateral) and the severity of the clefts (complete, incomplete) (Dixon *et al.,* 2011). In the last few decades, orofacial clefts have been more genetically investigated since epidemiological studies demonstrated strong familial recurrence. Twin studies indicated a phenotypic concordance rate of 40–60% in monozygotic twins compared with a 3-5% concordance rate in dizygotic twins (Jugessur *et al.,* 2009a; Leslie and Marazita, 2013; WHO, 2002). Variations between the

types of clefts have also been reported among the population in question such as the low prevalence rates seen for non-syndromic cleft palate only (NS-CPO) compared to non-syndromic cleft lip/palate (NS-CLP) (worldwide prevalence for NS-CLP is 1:1000 vs. 1:2400 for NS-CPO) (Leslie and Marazita, 2013; Sivertsen *et al.,* 2008). However, studies have also shown a 56-fold increase in risk of recurrence for first degree relatives of NS-CPO whereas the overall risk of recurrence of NS-CLP was 10 to 32-fold compared to the reference populations (Jugessur *et al.,* 2009a; Sivertsen *et al.,* 2008).

Asymmetry in the distribution of orofacial clefts is well documented whereby unilateral clefts are twice as common as bilateral clefts in cases of cleft lip and palate, contrastingly, they are 10 times more prevalent in cases with cleft lip only (Shapira *et al.,* 1999). For reasons not well understood, unilateral left-sided clefts of the lip (with or without cleft palate) are twice as common as right-sided clefts. Current assumptions for the affinity towards expressing the dysmorphology in the left side of the face have been attributed to processes such as directional asymmetry (Gallagher *et al.,* 2017; Weinberg *et al.,* 2006) and attempts have been made to model the variation in laterality statistically by applying some modifications to the multifactorial threshold model (Hallgrímsson *et al.,* 2005). However, demonstrations in familial cases suggest random lateralization; where some families show the same side of cleft affecting multiple individuals within a family, others show random distributions of the cleft defect among relatives within the same family (Gallagher *et al.,* 2017). The sex of the individual is another variable that influences the laterality of the cleft defect (Mossey and Castilla, 2003, WHO). The distribution of cleft lip and cleft lip & alveolus is equal between males and females; however, cleft of the lip and palate is two times more frequent in males than in females and the left side predominance is also observed more frequently in males (Shapira *et al.,* 1999). Contrarily, isolated cleft palate occurs more often in females than in males (3:1 ratio) (Mossey *et al.,* 2009; Shapira *et al.,* 1999). Nevertheless, the risk of clefts among children of affected mothers and affected fathers has been shown to be similar (Sivertsen *et al.,* 2008). Taken together, CLP is a complex multifactorial condition reflected by its heterogenous presentations.

1.2.2 **The Spectrum of Palatal Defects**

Many studies in animal models and in humans have focused on the causative events in 'true' (complete) cleft palate (Han *et al.*, 2009; Hoebel *et al.*, 2017; Huang *et al.*, 2008; Jugessur *et al.*, 2009a). However, these studies do not account for the diversity of palatal morphologies observed in the clinic. Unusual palatal shapes such as "pseudo-cleft" are

seen frequently and there are a range of relatively underappreciated palatal abnormalities, including high-arched palates, palatal crowding and swelling of the palatal rugae or gingival swelling and broad or narrow palates. Some of these are accompanied by tooth anomalies, and most have previously been assumed to be a result of sub-phenotype clefting. Many syndromes known to affect craniofacial structures can influence the morphology of the palatal vault. However, to my knowledge, none of the studies in syndromic patients have specifically examined the palatal phenotypes while considering where these anomalies lie in the spectrum of malformations. This could be due to the small sample size in studies of patients with rare syndromes and the variable expressivity of a mutation in different patients with the same syndrome.

Recent research into phenotype-genotype relationships has been focused on the need to investigate subclinical features in patients and their family members. Having a high-arched palate in a family member for instance, could provide valuable clues to the likelihood of inheritance of a genetic mutation. However, this 'subclinical' phenotype has often been overlooked.

A brief survey on the range of palatal phenotypes seen in syndromic patients as documented in the clinical literature is summarised in **(Table 1-1)**; the hypothesis being that many syndromes are accompanied by characteristic anomalies in palatal shape and that similar syndromes can be grouped together suggesting common aetiology. By doing so, this could provide some insight into diagnostic and prognostic indicators.

To illustrate this point, our group has used this approach to identify common molecular features of Orofacial Digital Syndrome Type 1 (OFD1) [MIM: # 311200] and craniosynostosis syndromes such as Apert [MIM: # 101200], two seemingly unrelated disorders (Tabler *et al.*, 2013). In this example, our group had noted that patients with either syndrome frequently showed high-arched palate accompanied by palatal swellings, severe dental crowding and hypodontia. In mouse models, our group and others then demonstrated that the palatal features of both syndromes are likely caused by upregulation of fibroblast growth factor (FGF) signaling affecting the neural crest lineage, prior to formation of the palatal shelves (Tabler *et al.*, 2013). Perhaps more importantly, our findings on the pathophysiology of high-arched palates suggest that the mechanism of its occurrence is entirely independent of that of the 'true' form of palatal clefting. Our group's successful use of this approach suggested that we could expand this to a survey of other syndromes. Based on this, one could postulate that palatal phenotypes, particularly the high-arched palate, are part of a cleft phenotypic

continuum and that cleft palate comprises one point within a broader spectrum of anomalous palatal morphologies.

Table 1-1 Syndromic disorders manifesting high-arched palate or cleft lip/palate

Condition with high-arched palate	Palatal Anomaly	References	Possible Mechanisms	References
Apert syndrome	Cleft palate or high arched-narrow palate +/- enlarged palatine ridges	(Wilkie *et al.*, 1995), (Letra *et al.*, 2007a)	*Fgfr2* gain of function affecting neural crest cells	(Holmes & Basilico, 2012)
Ciliopathies: Oral-facial-digital syndrome (OFD types I & IV)	CL/P, CPO, high-arched palate	(Prattichizzo *et al.*, 2008) (Beales *et al.*, 1999)	Excessive migration of neural crest to maxillary processes	(Tabler *et al.*, 2013)
Bardet Biedl syndrome (BBS)	High-arched palate			
Treacher Collins syndrome	Cleft palate or high-arched palate	(Trainor *et al.*, 2009)	Diminished number of migrating neural crest cells Neuroepithelial cells death	(Conley *et al.*, 2016)
Mucopolysaccharidoses: Maroteaux-Lamy syndrome Hurler syndrome	High-arched palate & hyperplasia of palatine ridges and gingival tissues	(Trowbridge and Gallo, 2002) (Alpöz *et al.*, 2006) (Pan *et al.*, 2005)	Inborn error of metabolism and glycosaminoglycans (GAG) accumulation in palatal tissues. Accumulation of GAGs perturb FGF and BMP4 signalling	(Alpöz *et al.*, 2006) (Trowbridge and Gallo, 2002) (Campos & Monaga, 2012)
Kabuki syndrome	Cleft palate or high-arched palate	(Banka *et al.*, 2012) (Adam *et al.*, 2019)	More 'true' clefts produced from *KMT2D* mutations; more high palates from *KDM6A* mutations	(Adam *et al.*, 2019)
Down syndrome	Narrowness and steepness of the palate, +/- CL/P	(Abeleira *et al.*, 2015; Källén *et al.*, 1996)	Morphology of the skull and mandible are affected by dosage imbalance in the genes	(Richtsmeier *et al.*, 2002) (Hill *et al.*, 2007)
Ectodermal dysplasia	Shallow palates, or CL/P	(Gunduz Arslan *et al.*, 2007; Paschos *et al.*, 2003)	Causes for reported shallow palates are unknown. True clefts occur from mutations in *PVRL1*	(Suzuki *et al.*, 2000)

1.2.3 **Cleft-Associated Anomalies**

1.2.3.1 Classifications used to define cleft lip/palate conditions

For novel gene discovery to take place, it is essential to differentiate between descriptions of the classifications used in cleft research. Accurate comparison between studies has been hampered by the inconsistencies in defining the cleft condition (Benirschke *et al.*, 1979; Rittler *et al.*, 2011; Spranger *et al.*, 1982). While the WHO report in 2001 highlighted the lack of consensus on the terms used to describe craniofacial or other birth malformations, nearly 20 years later, there is still some confusion on the terminology used to define the various orofacial cleft categories (Mossey and Castilla, 2003). Clinically and surgically, cleft classifications have focused on defining the cleft defect in terms of completeness and laterality (Carroll & Mossey, 2012; McBride *et al.*, 2016). Binary groupings in relation to wider medical and craniofacial changes have also been employed. *Aetiologic descriptions* include syndromic vs. non-syndromic clefts. Syndromic clefts are, "*those that include cognitive deficits or structural abnormalities outside of the region of the cleft*" (Murray, 1995). However, in research conducted in congenital disorders, the term *syndromic* is often used to describe clinical perceptions rather than clinical diagnoses. As such, syndromic cases in cleft research are defined as patients with CLP presenting with a congenital anomaly, developmental disorder or a distinct facial gestalt, and not necessarily patients with an established genetic diagnosis or syndrome. These cases are also often referred to as 'Mendelian' conditions in the literature. *Anatomic descriptions* involve 'isolated' vs. 'associated' clefts, which refer to the *number* of defects associated with clefts regardless of the cause.

The ICBDMS defines probands with anomalies associated with their cleft phenotype as multi-malformed infants (MMI) or multi-malformed cases (Mossey and Castilla, 2003). The definitions of associated anomalies vary among researchers, as there are variabilities in the data-collection method and the length of follow-up time (Mossey and Castilla, 2003). For the classification of a congenital anomaly, the WHO recommends differentiating between isolated and multiple congenital anomalies (MCA), separating subclasses within these two categories and separately evaluating the known aetiologic MCA patterns (MCA syndromes vs. MCA associations) (Mossey and Castilla, 2003). MCA syndromes are defined as those that have recognised patterns of component congenital anomalies that presumably have the same aetiology, these include chromosomal aberrations, teratogenic factors, or X-linked, dominant or recessive genetic mutations. On the other hand, MCA associations are those that have recognised patterns of non-random associations of two or more different component

congenital anomalies that do not have a known shared aetiology and do not currently constitute MCA syndromes (Mossey and Castilla, 2003).

To fully be able to distinguish between isolated and associated cases into those that are non-syndromic, clues for the aetiology of the defect is necessary (Mossey and Castilla, 2003). Isolated clefts are defined as cases devoid of other anomalies, whereas associated clefts are those that comprise two or more unrelated anomalies (Calzolari *et al.*, 2007). The term 'isolated' CLP is often, mistakenly, used interchangeably with non-syndromic CLP (Rittler *et al.*, 2011; Tan *et al.*, 2009). Non-syndromic clefts are not truly reflective of 'isolated' clefts since it is likely that other structures are affected, whether overtly or sub-clinically. Therefore, 'isolated' is a *clinical* diagnosis that is interim until further information is available, and 'non-syndromic' is an expression used to describe underlying *genetic* diagnoses whenever possible.

1.2.3.2 Multiple congenital anomalies in cleft patients

It is not unusual for clefts to be associated with other congenital anomalies, indeed more than 400 syndromes comprise of orofacial clefts (Dixon *et al.*, 2011). Many comprehensive population-based clinical surveys have been conducted delineating malformations associated with CLP with extensive sample sizes (Calzolari *et al.*, 2007; Milerad *et al.*, 1997; Rittler *et al.*, 2011; Stoll *et al.*, 2000). Various congenital anomalies have been found to commonly occur with cleft, with the prevalence varied based on the population studied, such that the most common recurring malformations across the studies were congenital anomalies of the cardiovascular system and skeletal/vertebral malformations (Calzolari *et al.*, 2007; Milerad *et al.*, 1997; Rittler *et al.*, 2011; Stoll *et al.*, 2000). Patients with cleft birth defects are known to suffer from cardiac anomalies, where 1.3-27% of orofacial cleft patients are reported to have a form of congenital heart disease, and more than 60 syndromes comprise of cleft and a cardiovascular anomaly (Setó-Salvia and Stanier, 2014). One survey on cleft-associated anomalies showed that malformations of the cardiovascular system were found in 29% (280/970) of their subjects, of which ventricular (60 cases) and atrial septal defects (106 cases) were the most common, followed by Tetralogy of Fallot (22 cases) (Calzolari *et al.*, 2007). The association between congenital heart disease and CLP could either be due to mutations in a single gene or chromosomal malformations, in addition to environmental factors, with neural crest cells a likely shared denominator imperative in the development of both tissues (Setó-Salvia and Stanier, 2014).

Many children with clefts and 'associated' anomalies have uncharacterized syndromes of which the underlying genetic cause is unknown. Studies that surveyed cohorts with multiple congenital anomalies categorized subjects into those with known syndromes or chromosomal defects and those that were unexplained but associated with clefts. In such survey, 15% (94/616) of children were found to have a cleft and an unexplained medical condition (Milerad *et al.*, 1997). Similarly, as part of the EUROCAT registries, Calzolari *et al.* (2007) studied 5,449 cases of cleft and found 1,589 (29.1%) cases with 'associated' medical anomalies. Among these cases, 455 had chromosomal defects and 164 had recognized syndromes and sequences. Interestingly, nine hundred and seventy cases with cleft (17.8%) had multiple congenital anomalies of unknown aetiology (Calzolari *et al.*, 2007). Finally, Rittler and co-workers examined 710 infants with oral clefts and found that those with associated clefts (26.1%) had initial diagnoses of either chromosomal abnormalities (38), non-chromosomal syndromes (33), malformation sequences (16) or multiple anomalies of unknown aetiology 98 (13.8%) (Rittler *et al.*, 2011).

The associated anomalies coexisted with various cleft subtypes. While Milerad *et al.* (1997) showed that the anomalies were more common in CLP than cleft palate only or isolated cleft lip, Stoll *et al.* (2000) showed that in their cohort the anomalies were more frequent in cleft palate (46.7%) than in CLP (36.8%) or isolated cleft lip (13.6%). As for the type of anomalies, consensus for the high prevalence for skeletal and limb defects and congenital heart disease was shown across all the surveys. Other anomalies have been seen, these have included developmental delay, gastrointestinal, central nervous system, eye, urogenital, anal atresia, hydronephrosis and endocrine defects (Calzolari *et al.*, 2007; Milerad *et al.*, 1997; Rittler *et al.*, 2011; Stoll *et al.*, 2000). Interestingly, some of these cleft-associated anomalies occurred in isolation or in combination with other systemic malformations. For instance, Milerad *et al.* (1997) reported that although limb and vertebral column anomalies were the defects that most often coexisted with clefts, they occurred in infants who were diagnosed with more than two anomalies, while cardiovascular anomalies prevailed when found as the only associated malformation. Likewise, Calzolari *et al.* (2007) reported two unrelated anomalies in 351 cases, three in 242, and four or more in 377 cases.

Most of the studies described above surveyed the prevalence of associated anomalies in infants. While most of the anomalies are diagnosed during the first month of life (Rittler *et al.*, 2011), the occurrence of diagnosed syndromes and other associated anomalies increases with age. Indeed, Rittler *et al.* found that 7% of infants that were

initially diagnosed as 'isolated' were later re-classified as 'associated' cases during their 1-year follow-up. Inclusion of phenotypic details as such, will help inform clinical genetic testing, diagnosis and management.

1.2.3.3 Dental anomalies in cleft patients

Dental anomalies have not traditionally been included under the 'multiple congenital anomalies' grouping. None of the surveys mentioned above have looked at dental anomalies in their cohorts. Separate attempts, however, have been adopted to sub-classify clefts based on dental subclinical phenotypes, whereby hypodontia has been regarded as part of the cleft spectrum and has been incorporated into cleft classifications (Klein *et al.*, 2013; Thesleff, 2014; Weinberg *et al.*, 2006; Zhang *et al.*, 2005). The inclusion of dental anomalies, found in the cleft individual or an unaffected family member, has been carried out with the aim of allowing more biologically relevant groupings (Leslie & Marazita, 2013; Vieira *et al.*, 2008).

Mouse studies have demonstrated the close temporal and spatial relationship between tooth and palate development. In early embryonic development, the face and cranium develop from migrating neural crest cells, whereby different cell compartments differentiate and populate different areas of the jaws. The contribution of neural crest cells to tooth development has been well documented (Miletich & Sharpe, 2004). Interestingly, several of the transcription factors involved in the patterning of teeth in their positions in the jaws, such as the homeobox genes, have also been implicated in orofacial clefting in humans and mice, confirming their importance in craniofacial patterning. In mice, both Msx1 and Dlx5 have been shown to operate in parallel in regulating downstream target gene expression during palatogenesis and that indirect inhibition of Shh signalling by *Dlx5* affected the oral-nasal patterning of palate and rescued cleft palate in *Msx1*-null mice (Han *et al.*, 2009). *Dlx5* homozygous null mice also have a cleft of the secondary palate (Depew *et al.*, 1999), so do *Dlx1/Dlx2* double homozygous null mice (Jeong *et al.*, 2012). In humans, a *DLX5* mutation has been found in a patient with PRS (Wolf *et al.*, 2014) and *DLX4* in a patient with bilateral CLP (Wu *et al.*, 2015a). Numerous genetic mutations in *MSX1* have been shown to cause hypodontia, CLP or both (Abid *et al.,* 2017; Peters *et al.* 1998; Satokata and Maas 1994).

Patients with CLP are reported to have higher prevalence rates for dental anomalies. This can range from a single malformed tooth within the cleft region to multiple defects outside the cleft region. The most common type of dental anomalies is hypodontia. The most commonly missing tooth is the maxillary lateral incisor in the

region of the cleft, but almost a third of patients with isolated CLP have teeth missing from outside the cleft region (Slayton *et al.*, 2003), suggesting that the anomaly is not due to physical disruption of the dental lamina alone but is part of the cleft condition. Premolars are also commonly affected regardless of type or side of clefting (Letra *et al.*, 2007b). Lammi *et al.* (2004) provided the first evidence for the involvement of WNT signalling in hypodontia, particularly the more severe forms. Additionally, Callahan *et al.* (2009) showed that mutations in *AXIN2* were involved in single incisor agenesis, however, none of the families reported histories of clefts. Mutations in *WNT10A* cause isolated and syndromic tooth agenesis with the second premolar being the most commonly missing tooth followed by the maxillary lateral incisor (Mostowska *et al.,* 2013; van den Boogaard *et al.* 2012). Additionally, *Wnt10a* has been shown to regulate the proliferation and apoptosis of mouse embryonic palatal mesenchymal cells through the WNT signalling pathway (Feng *et al.,* 2013). Other dental anomalies found in patients with CLP include supernumeraries, mineralisation defects, delayed dental development and microdontia to name a few (Weinberg *et al.,* 2006). The severity of dental anomalies has often been shown to directly correlate with the severity of the cleft defect (Eerens *et al.*, 2001; Slayton *et al.*, 2003; van den Boogaard *et al.*, 2000).

A common finding in cleft groups is the occurrence of single incisor agenesis on the contralateral side of the cleft region such that hypodontia on the right side is more frequent with unilateral left clefts, and vice versa. This phenotype led to the hypothesis that unilateral clefts with missing lateral incisors on the opposite sides could be 'unsuccessful' bilateral clefts (Letra *et al.*, 2007b). This is also in accordance with a study by Menezes & Vieira (2008), whereby 12.5% of their cleft patients presented with anomalies of the maxillary lateral incisors on the non-cleft side. The above further reiterates the potential genetic contribution to such combination of phenotypes in an individual with CLP. Vieira *et al.* (2008) conducted a study aimed at testing the hypothesis that dental anomalies were part of the cleft spectrum. They retrospectively revisited genotyping data of cleft patients and found highly significant associations between cleft and dental anomalies in the *ANKS6* region involved in the SMAD pathway and in the *ERBB2* region that also includes the *RARA* gene (retinoic acid receptor), a gene previously reported as candidate for non-syndromic CLP (Chenevix-Trench *et al.,* 1992). Finally, syndromic forms of CLP that include dental anomalies suggest a potential for genetic aetiology; examples include the *IRF6* gene in van der Woude syndrome and the *FGFR1* gene in Type 1 Pfeiffer syndrome (Dode *et al.*, 2003; Letra *et al.*, 2007b; Muenke *et al.*, 1994b).

1.2.3.4 Subclinical phenotypes in parents and siblings

Often, cleft defects are thought of as binary traits, in other words, an individual is either 'affected' or 'unaffected' (Aspinall *et al.*, 2014). This is a crude classification that hinders finding candidate genes, since detailed phenotyping and the reporting of subclinical anomalies in the cleft individual or an 'unaffected' family member is missed. Subclinical features have been defined as "soft tissue microforms (a minimal manifestation or subclinical sign) that is detected in non-cleft subjects, indicating a greater propensity to clefting in their offspring" (Huston *et al.*, 1984). Evidence suggests that the associated traits that the 'unaffected' parents and relatives present with, may represent cleft microforms or may be due to more generalized developmental disturbances.

Subclinical features in family members have recently been the turning point in cleft research. Information on inheritance is essential for accurate interpretation of genetic findings. Through carefully phenotyping parents and other family members, cases of incomplete penetrance and variable expressivity could be recognized. Phenotypic expressivity varies and can range from a known syndrome to a more subtle subphenotype such as velopharyngeal insufficiency (VPI), nasal speech or a tooth anomaly. Approaches to analyzing subclinical phenotypes in family members included: velopharyngeal variation (Huston *et al.,* 1984) and lateral cephalograms of parents (McIntyre & Mossey, 2002; Mossey *et al.*, 2010), three-dimensional 3D captures of faces of unaffected parents (Weinberg *et al.*, 2009), subepithelial defects in the orbicularis oris muscle (Neiswanger *et al.*, 2007) and most recently, dental anomalies of parents of children with cleft (Aspinall *et al.*, 2014; Howe *et al.*, 2015).

Studies on sub-epithelial defects in the orbicularis oris muscle hypothesize that these defects represent the mildest form of cleft lip. The orbicularis oris is the muscle that encircles the mouth, it functions in closing the mouth and contracting the lips. The Pittsburgh Oral-Facial Cleft Study, a large on-going project designed to identify candidate genes contributing to CLP, confirmed that the orbicularis oris muscle is affected in families of children with cleft (Neiswanger *et al.*, 2007). Morphological analysis of 3D images of faces of unaffected parents have also been examined in depth and significant differences in face shapes were found compared to controls (Weinberg *et al.*, 2009). 'Unaffected' parents showed flattening of the facial profile due to a combination between mid-face retrusion and mandibular protrusion, decreased upper face height, increased lower face height, increased inter-orbital width and altered dimensions of the nasolabial complex (Weinberg *et al.*, 2009). While these studies provided persuasive evidence to validate the presence of subclinical phenotypes in

family members, the questionable practicality and feasibility of these approaches limits their application in clinical practice. The availability of ultrasound and trained personnel is required in detecting orbicularis oris deformities.

Evidence from family surveys also suggests that a dental anomaly might be a microform of the underlying gene change (Aspinall *et al.*, 2014; Howe *et al.*, 2015). Multiple studies have considered the dentition of non-cleft relatives and dental anomalies could become firm contenders in the study of subclinical phenotypes in family members. Yet, the evidence to their value in studying cleft associations is conflicting (Anderson & Moss, 1996; Aspinall *et al.*, 2014; Howe *et al.*, 2015) and very few studies have verified their findings genetically. A survey conducted on children with cleft (n=54), their siblings without cleft (n=63), and controls without cleft (n=250) all with an age range of around 4-15 years, showed the prevalence of hypodontia (of one or more teeth outside the cleft region) to be significantly higher in both the cleft and sibling group. The most significantly missing tooth was the second premolars in all groups (Eerens *et al.*, 2001). More recently, Aspinall *et al.* (2014) carried out a comprehensive dental assessment of unaffected parents. Out of 101 parents, 50% of them had at least one dental anomaly. The most common were enamel defects (27%) but the type of defect observed was not specified. Other dental anomalies included hypodontia (12%), with the most commonly missing tooth being the upper right lateral incisor, followed by microdontia and supernumeraries (5% and 4%, respectively). Their results reinforce the hypothesis that dental anomalies are part of the 'extended cleft phenotype' and that they can be used as a useful tool to study heritability of CLP. However, their microform or subclinical clefts are hard to detect. Therefore, phenotypic data on family members may not often be as detailed or accurate (Shi *et al.*, 2009). Thus, careful phenotyping of patients with orofacial clefts, and their relatives, in the search for new genes and diseases is essential.

1.3 GENETIC ASPECTS IN CLEFT RESEARCH

1.3.1 Background on Cleft Genetics

The seminal paper by P. Fogh-Andersen in 1942 revolutionised our concepts of the mechanism and aetiology behind cleft birth defects by proposing that a large fraction of CLP had a strong genetic component (Fogh-Andersen, 1946; 1967). He noted from the population he studied that CLP and cleft palate only (CPO) were genetically distinct types due to the difference in their incidence. Further, he noted the hereditary pattern

of CLP cases and that dominant and recessive patterns were present, whereas CPO was distributed as a simple dominant event with incomplete manifestations (Eiberg *et al.*, 1987). Before the 1940s, genetic influences were not primarily included as significant components of cleft origin in families (Murray, 1995). In that same decade, it was hypothesised by Warkany *et al.* (1943) that craniofacial abnormalities might also be caused by environmental exposures or deficiencies (Murray, 1995). The following section first looks at human gene discovery approaches in orofacial clefts and then looks at hypotheses underlying changes in palatal morphology and attributes these to *genetic modifiers*.

Generally, genetic disorders are categorised into either multifactorial or monogenic disorders. Multifactorial disorders are usually complex with both multiple genes and lifestyle or environmental factors contributing to the disease (Rabbani *et al.*, 2012). This category best describes non-syndromic clefts. Monogenic disorders, on the other hand, include simple and rare disorders caused by single gene defects; a category that best describes syndromic clefts. Rare genetic disorders, unlike the more common multifactorial disorders, have low prevalence rates estimated in the region of 6.5 out of every 10,000 individuals according to the World Health Organisation (Lander, 2011; Rabbani *et al.*, 2012). Syndromic clefts have been the first types of orofacial clefts to be genetically studied simply because they have been amenable to early gene-discovery techniques. Initial pioneering studies in the 80's and 90's identified cytogenetic locations and, in a few cases, specific genes or environmental factors that appeared causal to syndromic clefts. In his review, Murray (1995) highlighted some of the early examples of ground-breaking discoveries for loci involved in now well-studied cleft syndromes. These include the location of X-linked cleft palate (Moore *et al.*, 1987), van der Woude syndrome (1q32) (Murray *et al.*, 1990), Treacher Collins syndrome (5q) (Dixon *et al.*, 1992), holoprosencephaly (Muenke *et al.*, 1994a), velocardiofacial syndrome (22q) (Driscoll *et al.*, 1992), and the genes for Waardenburg syndrome (PAX3) (Tassabehji *et al.*, 1992), Marfan syndrome (Fibrillin) (Hollister *et al.*, 1990), Stickler syndrome (COL2A1) (Francomano *et al.*, 1987), and diastrophic dysplasia (DTD) (Hästbacka *et al.*, 1994).

From the earliest stages of gene discovery in orofacial clefting, it has been recognised that the study of non-syndromic clefts in humans has been complicated by cleft inheritance patterns. Even early population-based studies on orofacial clefts have provided evidence that segregation analysis of non-syndromic clefts suggests a mixed model with elements of Mendelian (both autosomal recessive and dominant)

inheritance with variable contributions of reduced penetrance, sex differences, and environmental overlays (Murray, 1995). However, offspring of consanguineous parents were shown to have an almost two-fold increase in the risk for non-syndromic clefting (Sabbagh *et al.*, 2014), suggesting a strong genetic input into the phenotype.

In recent years, children with clefts, particularly those with associated anomalies, have been undergoing clinical baseline testing that includes microarray comparative genomic hybridization (array CGH), multiplex ligation-dependent probe amplification (MLPA) studies for subtelomeric rearrangements and sites of common microdeletions, as well as fluorescence *in situ* hybridization (FISH) (Hills *et al.*, 2010). Array CGH has been introduced in place of routine karyotyping in clinical practice for patients with a suspected genomic imbalance. Additionally, candidate gene tests for common associations based on the child's symptoms are also carried out. For instance, a child with cleft – especially those associated with a Pierre Robin sequence – and eye anomalies is tested for Stickler syndrome. Similarly, a child with cleft and congenital heart disease is tested for variants in the 22q11 region or CHARGE syndrome. Often normal or 'negative' results are returned. Although array CGH has many uses, particularly in a clinical and diagnostic setting, such as excluding the above syndromes and other major chromosomal abnormalities, its use is fairly limited from a research perspective. With the advent of next generation sequencing technologies like whole exome sequencing (WES) and whole genome sequencing (WGS), many of those that had inconclusive results could now be given a chance to receive a genetic diagnosis which will inform future medical needs, genetic counselling and familial planning **(Figure 1-4)**.

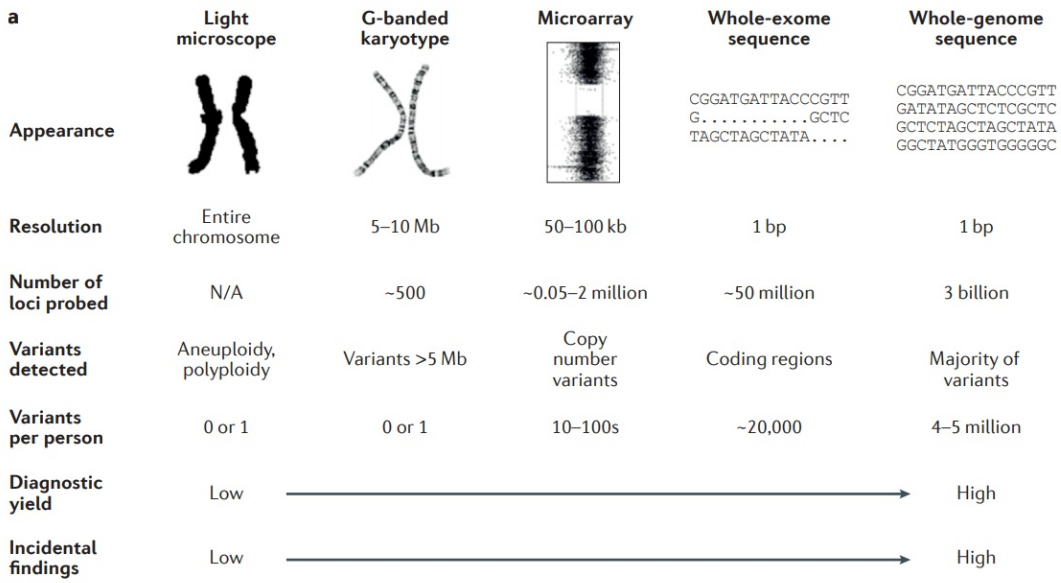


Figure 1-4 Clinical gene testing strategies

Genome-wide assays used in clinical genetics have developed from traditional methods for visualizing chromosomes using a light microscope to assaying copy number variation across the genome and to sequencing the entire genome. As the resolution of the test increases, the number of detectable variants also increases. Although this increase in the number of variants leads to an increase in diagnostic yield across a range of conditions, it also substantially increases the likelihood of detecting incidental findings and variants of uncertain clinical significance. Taken from (Wright *et al.*, 2018a).

1.3.2 **Approaches to Gene Mutation Discovery in Cleft Research**

Familial forms of orofacial clefting have provided most of the genetic information on non-syndromic clefts to date because of their suitability to early research designs and techniques, including linkage studies and candidate gene approaches. Indeed, many of the genes and genomic regions associated with clefts were identified using these early genetic techniques. Genetic linkage study designs have been successful in identifying genomic regions pointing to genes for inherited diseases or Mendelian subsets of complex diseases. However, a key requirement to the use of linkage analysis is the need for large families with multiple affected individuals (also known as 'multiplex families') segregating diseases according to Mendelian principles. The candidate gene approach is a method that has been widely applied and it compares the prevalence of genetic variants in known biological genes or pathways between cases with disease and controls (Wu *et al.*, 2015b). Although this approach has proven successful in terms of cost and time effectiveness, the design does not typically discover novel genetic loci beyond the candidate genes as it relies on testing hypothesised associations (Wu *et al.*, 2015b). Definitions of the approaches used for gene discovery in cleft research, their advantages and disadvantages are summarized in **(Table 1-2).**

Despite their limited power to detect single gene changes, numerous genes and regions of the genome associated with CLP and CPO have been identified through linkage and candidate gene approaches such as the interferon regulatory factor 6 (*IRF6*), Msh homeobox 1 (*MSX1*), transforming growth factor alpha (*TGF-a)* and many more (Dixon *et al.*, 2011; Marazita, 2012). These discoveries were later supported using population-based genome-wide association studies (GWAS), thus contributing to our understanding of the aetiology of orofacial clefts. A summary of some of the cleft candidate genes discovered using these approaches is found in **(Table 1-3).**

Table 1-2 The various types of gene discovery approaches in humans, and the advantages and disadvantages of each

Gene Discovery Approach	Description	Advantages	Disadvantages
Genome-Wide Linkage Analysis	Method traditionally used to identify disease genes. Family pedigrees used in study design. Linkage is a property of loci (not alleles, i.e. the allele is only used as a marker of location) The idea is if the two loci of the marker and the susceptibility gene were close enough, the stronger their co-segregation as a unit within the family under study.	Could serve as a useful tool in syndromic cases with a Mendelian inheritance pattern.	Does not localize to a particular location on the chromosome, instead it guides to a region on the chromosome that appears to co-segregate with the susceptibility gene. This region could turn out to be quite large with a fair amount of candidate genes that could be influencing the trait of interest. Has little power to detect variants that have small to moderate effects on a given disease.
Genetic Association Studies (GAS)	Have often been used as follow-ups to linkage analysis with the aim of narrowing down the region identified via linkage. The idea is to assess correlations between genetic variants and disease phenotype using the genetic markers. In order for association to be positive with a disease phenotype, the genetic marker needs to be in high linkage disequilibrium (LD) with the causal disease variant. Genetic association studies could be broadly divided into genome-wide association studies (GWAS) and candidate-gene studies (discussed below).	Much smaller regions are identified than in linkage analysis. The study design is easier, as the typical case-control design is mostly used & does not require related individuals. More powerful at detecting susceptibility genes that have smaller effects. Faster to conduct. Cheaper.	One of the major difficulties with case-control GAS is the interpretation of findings. If a positive association was found, it is not clear whether the associated allele is the disease-causing one (the functional variant) or instead, is in strong LD with the disease-causing allele. GAS have particularly proven difficult to replicate. Population stratification has long been a significant issue. The analysis is complicated by admixed populations. To solve this problem, family-based designs have been introduced.
Genome-Wide Association Studies (GWAS)	Are forms of comprehensive association studies that survey the genome for risk variants using an unbiased approach. As with GAS, GWAS follow the same concept of marker-association.	The HapMap project will enhance efficiency in GWAS as it focuses only on common SNPs, those with a minimum allele frequency (MAF) of > 1% in a population.	The vast number of SNPs that exist in the human genome. The dbSNP database has so far identified more than 15 million SNPs. Difficulty lies within choosing between the millions of SNPs and genotyping a comprehensive set of variants for a large patient sample.

Candidate-Gene Studies	These are hypothesis-based approaches that predict the identity of the correct genes. The basic idea is to test whether polymorphisms in a candidate gene, whose function is believed to be involved in the disease of interest, is associated with the disease. Coding and /or non-coding regions could be selected for re-sequencing the candidate gene in cases and controls.	Can aid in the gene discovery process of non-syndromic cases, as candidate genes can be selected based on their role in syndromes. The rationale is the less deleterious variants of the same gene can contribute to a less severe phenotype.	Largely relies on previous knowledge about the biological process involved.
Whole-Exome Sequencing (WES)	Approximately 85% of disease-causing mutations will be identified within an exome, the protein-coding part of the genome, representing about 1.5-3% of the total genome. Focuses on capturing rare and very rare (MAF < 0.1%) mutations and detects many of the classes of sequence variations including: single nucleotide variants (SNVs), insertions/deletions (indels) & mosaicism.	Advantage over GWAS is lower cost and fewer data to analyse and store. WES generates 1/15 of the data generated by Whole Genome Sequencing (WGS) and the price is 1/5 of the cost. It provides effective 'depth of coverage', which refers to the number of times that a base (i.e. one letter) has been sequenced and is read by each fragment in the total sequencing process, meaning fewer false positive results and inaccuracies will occur.	Variability between different laboratories in bioinformatics. Current capture kits can only target exons that have been identified to date. Variant filtering. Duplicated sequences not represented in the reference genome are not removed and thus spuriously produce false positive results in the form of new variants in the sequence.

References: (Altmuller *et al.*, 2001); (Thomas & Witte, 2002); (Wacholder *et al.*, 2002); (Hirschhorn & Daly, 2005); (Ng *et al.*, 2010b). (Bamshad *et al.*, 2011); (Biesecker *et al.*, 2011); (Sims *et al.*, 2014).

1.3.2.1 Summary on genome wide association studies (GWAS)

For some time, GWAS has been the predominant gene discovery approach in non-syndromic CLP, an approach that has shown at least 39 genomic loci to confer to the risk of non-syndromic CLP (Beaty *et al.,* 2010; Birnbaum *et al.,* 2009; Böhmer *et al.*, 2018; Leslie *et al.,* 2016a; Ludwig *et al.,* 2012; Mangold *et al.,* 2010).This strategy is not readily applicable in cases of syndromic CLP, as large populations with 'similar' disease phenotypes are required for testing. Moreover, extreme sporadic syndromic cases are often not inherited and are of *de novo* origin; since mutations as such are considered the most extreme form of rare genetic variation (when compared to inherited variation), as they have been subjected to less stringent evolutionary selection (Veltman & Brunner, 2012). The GWAS design, therefore, is based on the hypothesis that common diseases could be explained by common variants (Hirschhorn & Daly, 2005). GWAS studies combine the case-control design with high-throughput genotyping and have relied on preselected common single nucleotide polymorphisms (SNPs) as biological markers from across the genome to agnostically characterise genetic susceptibility to complex disease (Wu *et al.*, 2015b). SNPs represent the most abundant form of genetic variation among people and occur almost once in every 1,000 nucleotides on average, which means there are roughly 4 to 5 million SNPs in a human genome.

Birnbaum and co-workers were the first to analyse a non-syndromic CLP sample using GWAS. In their case-control study, 146 SNPs spanning the 8q24.1 region were chosen in a sample of 462 non-syndromic CLP cases and 954 controls. Genotyping was successful for 125 of those markers; ten were highly significant. Since then, evidence suggesting the involvement of a locus on chromosome eight with CLP has gained great recognition (Birnbaum *et al.*, 2009). Although the evidence is strong, the genes in this locus that contribute to the aetiology of CLP have not been described. Another robust GWAS with a case-parent trio design showed significant linkage and association for multiple SNPs on chromosome 8q24 as well as *IRF6* (Beaty *et al.*, 2010). They discovered SNPs in two genes not previously associated with non-syndromic CLP that achieved genome-wide significance; namely *ABCA4* (on chromosome 1p22.1) and *MAFB* (on 20q12). *Abca4* showed no expression in the palate of mouse embryos. However, *Mafb* was expressed between E13.5-14.5 in the epithelium of the palatal shelves and in the medial edge epithelium (Beaty *et al.*, 2010). Other loci identified through GWAS are summarised in **(Table 1-3).**

Table 1-3 Genetic risk loci and variants identified using various approaches to gene discovery in orofacial clefting

Gene discovery approach	Number of studies done	Loci; or candidate gene in region	Reference
Genome-Wide Linkage Analysis	13	*9q21 (PTCH, FOXE1, ROR2, TGFBR1); 2p13 (TGFA); 10q26 (FGFR2); 6q25.1 (ESR1); 14q22-24 (PAX9, TGFB3, BMP4); 16q24 (FOXC2, CRISPLD2); CYP1B1; FAM82A; SUMO1; FGF10; TFAP2A; TBX1*	(Osoegawa *et al.*, 2008); (Shi *et al.*, 2009); (Marazita *et al.*, 2004); (Marazita *et al.*, 2009)
Candidate-Gene Studies	Multiple	*IRF6; TGFA; F13A; FOXE1; MSX1; GLI2; JAG2; LHX8; MSX2; SATB2; SPRY2; TBX10; PTCH; PVRL1*	(Ardinger *et al.*, 1989): first candidate gene association study; (Eiberg *et al.*, 1987): first candidate gene linkage study; (Jezewski *et al.*, 2003); (Zucchero *et al.*, 2004); (Vieira *et al.*, 2005); (Mansilla *et al.*, 2006); (Avila *et al.*, 2006); (Moreno *et al.*, 2009); (Ludwig *et al.*, 2014)
Genome-Wide Association Studies (GWAS)	9 for NS-CLP	*8q24; 2p24.2; 1q32.2; 17q23; 19q13; 16p13.3 loci. 10q25 (VAX1); 17q22 (NOG); 1q23 (IRF6); 2p21 (THADA); 3p11 (EPHA3); 13q31 (SPRY2); 15q22 (TPM1); 17p13 (NTN1); 1p22 (ARHGAP29); 1p36 (PAX7); ABCA4; 20q12 (MAFB); ADAMTS20; GRHL3*	(Birnbaum *et al.*, 2009); (Grant *et al.*, 2009); (Mangold *et al.*, 2010); (Beaty *et al.*, 2010);(Ludwig *et al.*, 2012); (Wolf *et al.*, 2015); (Sun *et al.*, 2015); (Leslie *et al.*, 2016a); (Leslie *et al.*, 2016c)
Whole-Exome Sequencing (WES)	-*0 for NS-CLP -Multiple for S-CLP	-	-

*No studies on WES for non-syndromic cleft lip and palate were carried out when the current project first began. Since then, 8 studies have been conducted, see (Table 1-4). NS-CLP, non-syndromic cleft lip/palate; S-CLP, syndromic cleft lip/palate.

A downside of GWAS is that it does not take into consideration inheritance models like linkage studies or the more recent exome sequencing methods. In an attempt to test the applicability of inheritance models in GWAS, a study by Böhmer and co-workers was carried out addressing the impact of recessive or dominant effects to non-syndromic CLP in non-related populations. The authors highlighted that in all GWAS of non-syndromic CLP conducted to date, statistical analyses were performed by methods fitting with a multiplicative mode of inheritance (Böhmer *et al.*, 2018), that is, a model that indicates that the risk of disease is increased by a constant factor with each

additional copy of the risk allele (Clarke *et al.*, 2011). This implies that if at some loci the underlying genetic model is recessive or dominant (i.e. non-multiplicative), this misspecification can result in a substantial loss of statistical power (Böhmer *et al.*, 2018). Therefore, the authors re-analysed their GWAS data by methods that were more sensitive to dominant and recessive models. Interestingly, when this segmentation was applied, none of the promising common variants from genome-wide data were replicated, neither were novel genetic findings for non-syndromic CLP identified. The authors concluded that their data did not support the dominant or recessive models to confer risk for non-syndromic CLP in outbred populations. This is in line with another study that showed that dominance variation at common SNPs explains only a small fraction of phenotypic variation in complex traits (Zhu *et al.*, 2015). These results advocate the likelihood of private/rare mutations within families that cannot be found on a population level, that some variants are unique to isolated populations, that a number of cases occur sporadically, or that a fraction of GWAS variants that were not tested in their dataset might still display recessive or dominant inheritance patterns.

Although GWAS has been successful in detecting variation in population-based diseases such as diabetes, auto-immune diseases and schizophrenia (Visscher *et al.*, 2017), we now know through evidence from GWAS in CLP, and other congenital birth defects (Stranger *et al.*, 2011), that only a modest degree of phenotypic variation has been explained by common SNPs. This has led to the concept of pursuing the "missing heritability" (Manolio *et al.*, 2009) and has put an emphasis on other genetic variations that contribute to phenotypic variance such as *de novo* single and copy number variants, including duplications and deletions. Single nucleotide variants (SNVs) and copy number variants (CNVs) can influence gene expression by disrupting gene coding sequences, perturbing long-range gene regulation, or altering gene dosage. Variants as such can be detected via next generation sequencing technologies.

1.3.2.2 Next generation sequencing (NGS)

Next generation sequencing technologies include approaches such as whole exome sequencing (WES) and whole genome sequencing (WGS). Approximately 85% of disease-causing mutations will be identified within an exome, the protein-coding part of the genome representing about 1.5-3% of the total genome. Exome sequencing focuses on capturing rare and very rare (minor allele frequency (MAF) < 0.1%) mutations and detects many of the classes of sequence variations including single nucleotide variants (SNVs), insertions/deletions (indels) and mosaicism (Bamshad *et*

al., 2011; Sims *et al.*, 2014). WES generates 1/15 of the data generated by WGS and so provides reduction in computational complexity and cost (Biesecker *et al.*, 2011). It also provides effective 'depth of coverage', which refers to the number of times that a base has been sequenced and is read by each fragment in the total sequencing process. This is an important aspect of WES as it ensures that fewer false positive results and inaccuracies will occur when the coverage is high. Variants in genes identified as underlying syndromic phenotypes have been shown to represent almost a third of positive findings from clinical whole exome sequencing (Yang *et al.*, 2014). Learning from cases as such will not only enhance our understanding of the genetic and allelic architecture of the syndromic conditions the genes are involved in, but also will facilitate the diagnosis of non-syndromic diseases or complex disorders.

Roughly, four main strategies have been adopted for designing studies using WES **(Figure 1-5).** These include sequencing multiple unrelated affected individuals (i.e. a proband-only approach), sequencing multiple affected individuals within a pedigree (i.e. multiplex families), sequencing parent-child trios and lastly, sampling and sequencing based on phenotype.

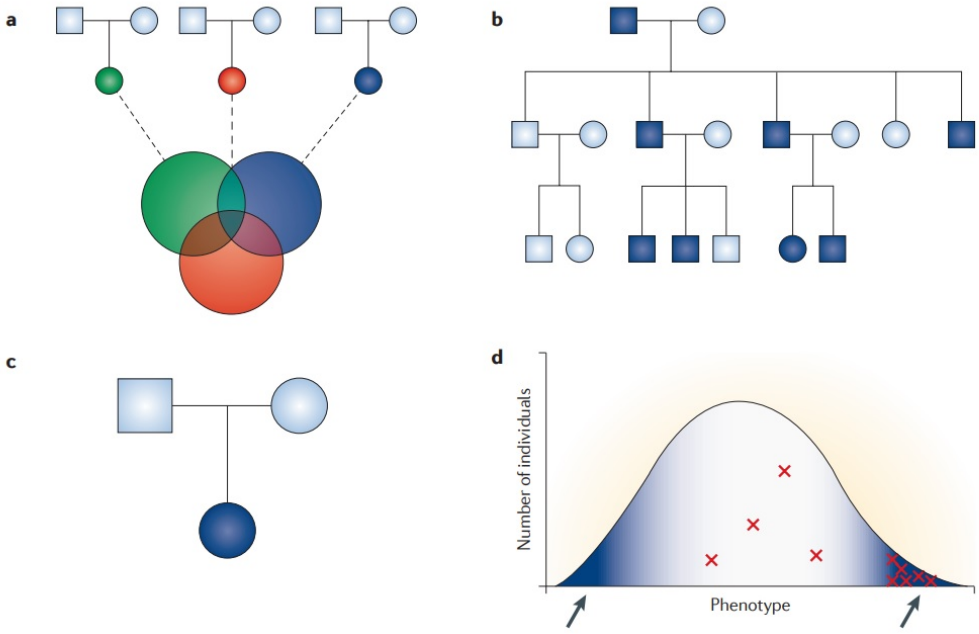


Figure 1-5 Strategies for finding disease-causing rare variants using exome sequencing

Four main strategies are illustrated. (a) Sequencing and filtering across multiple unrelated, affected individuals (indicated by the three coloured circles). This approach is used to identify novel variants in the same gene (or genes), as indicated by the shaded region that is shared by the three individuals in this example. (b) Sequencing and filtering among multiple affected individuals from within a pedigree (shaded circles and squares) to identify a gene (or genes) with a novel variant in a shared region of the genome. (c) Sequencing parent–child trios for identifying *de novo* mutations. (d) Sampling and comparing the extremes of the distribution (arrows) for a quantitative phenotype. As shown in panel d, individuals with rare variants in the same gene (red crosses) are concentrated in one extreme of the distribution. Taken from (Bamshad *et al.*, 2011).

Exome sequencing in syndromes associated with clefts

The contribution of WES to the analysis of Mendelian diseases has been established as the majority of allelic variants known to underlie Mendelian disorders disrupt protein coding sequences, however, its application to complex diseases has only in recent years been applied. Indeed, Ng and co-workers published the first successful study that applied WES to discover the causal gene in a monogenic disorder, Miller's syndrome [MIM: # 263750], or postaxial acrofacial dysostosis, a rare autosomal recessive disorder characterized clinically by severe micrognathia, cleft lip and/or palate, hypoplasia or aplasia of the postaxial elements of the limbs, coloboma of the eyelids, and supernumerary nipples (Ng *et al.*, 2010b). They found mutations in the *DHODH* gene in three affected pedigrees with the disease. Following this success, many other clinically ascertained syndromes with clear phenotypic descriptions but unknown molecular bases now have genetic diagnoses. Indeed, from 2010-2012, more than 100 causative genes in various Mendelian disorders have been identified by means of exome sequencing; 56.5% followed autosomal recessive, 37% followed autosomal dominant and 1.85% followed X-linked inheritance patterns (Rabbani *et al.*, 2012). Since the review by Rabbani and co-workers, growing numbers of reports have been published. A Medline search on Mendelian disorders using WES revealed more than 400 diseases. Most notably, some of the first publications on Mendelian disease genes identified by WES, that include cleft lip and/or palate as part of their phenotype, included *MASP1* in 3MC syndrome (Sirmaci *et al.*, 2010) , *MLL2* in Kabuki syndrome (Ng *et al.*, 2010a), *FLNA* in terminal osseous dysplasia (Sun *et al.*, 2010), *NOTCH2* in Hajdu-Cheney syndrome (with high-arched palates) (Isidor *et al.*, 2011), *SMAD4* in Myhre syndrome (Le Goff *et al.*, 2011), *ASXL1* in Bohring-Opitz syndrome (Hoischen *et al.*, 2011) and *SF3B4* in Nager syndrome (Bernier *et al.*, 2012).

More exclusively, exome sequencing studies were carried out on specific pedigrees with syndromic orofacial clefts. Pengelly and co-workers carried out exome sequencing in seven individuals with syndromic clefts, Nager syndrome, incontinentia pigmenti and Pierre Robin sequence (PRS), from 3 families by targeting 800 known cleft/palatal genes (Pengelly *et al.*, 2015). Known gene mutations related to these syndromes were identified, however, a novel variant in *IRF6* (p.Gly604Ala) was found in the proband with PRS and three other distant relatives with CPO or CLP. The role of *IRF6* in PRS has not been previously described (Pengelly *et al.*, 2015). More recently, Cox and co-workers linked a missense variant in *GDF11,* a gene previously unreported in association with any human monogenic disorder, in a multiplex family that segregated

with CLP or submucous cleft palate and both rib and vertebral hypersegmentation (Cox *et al.*, 2019), mirroring phenotypes seen in *Gdf11* knockout mice (McPherron *et al.*, 1999). The *GDF11* gene encodes the growth differentiation factor 11 protein that is a member of the bone morphogenetic protein family and the TGF-β superfamily (Cox *et al.*, 2019).

Despite 75% of syndromic clefts having a known genetic cause (Leslie & Marazita, 2013), the pathogenesis of such cases is often multifaceted and the heterogenous genetic aetiology is inexhaustible. This is reflected by the hundreds of studies that defined causation in these syndromes either by mutation of a single genetic locus, chromosomal abnormalities, or teratogens (Leslie & Marazita, 2013; Mossey *et al.*, 2009). Despite these efforts, more syndromic orofacial clefts are being discovered.

Exome sequencing in non-syndromic clefts

The first study to utilize WES in a non-syndromic CLP population was carried out by Bureau *et al.* (2014). Sequencing was done among multiple affected individuals from within a pedigree in 55 multiplex cleft families. Since then, only seven more WES studies have been carried out in non-syndromic CLP cohorts using a similar study design to that of Bureau's and co-workers. These were mostly carried out on European populations, with some in Honduran and Chinese populations. WES studies in non-syndromic clefts, their study designs, whether sub-clinical features were included, and the outcomes are summarised in **(Table 1-4)**. As demonstrated from **(Table 1-4)**, the multiplex strategy for gene discovery was used in almost all studies. Another aspect highlighted was the lack of phenotypic information on subclinical features in parents and 'unaffected' individuals. Other WES design strategies are yet to be undertaken in the field of non-syndromic CLP research.

Gene identification for non-syndromic clefts using WES has yielded mostly known but some novel findings. This is because most panels used in the sequencing process were specific to known cleft genes from human or mouse studies. While focusing on 'established' candidate genes maximises the chance that any novel, damaging variant found in the tested subjects would actually be causal and would lower the threshold for statistical significance (Bureau *et al.*, 2014), this approach minimises the chances of reporting variants in novel candidate genes. Although the first study to use WES on families originally recruited for linkage studies used a targeted sequencing approach as such (Bureau *et al.*, 2014), most of the subsequent studies used a non-biased approach (Aylward *et al.*, 2016; Fu *et al.*, 2017a; Hoebel *et al.*, 2017).

Reported cleft genes confirmed in additional non-syndromic cleft families via WES included novel single nucleotide variants in *CDH1* [Epithelial (E)-Cadherin] (Bureau *et al.*, 2014), *GRHL3* [Grainyhead Like Transcription Factor 3], *CREBBP* [CREB binding protein] (Hoebel *et al.*, 2017) and *ARHGAP29* [Rho GTPase Activating Protein 29] (Liu *et al.*, 2017).

On the other hand, novel genes not formerly reported with orofacial cleft anomalies discovered through WES included genes such as *ACACB* [Acetyl-CoA Carboxylase Beta] with 4 missense variants identified in 4 of 148 non-syndromic cleft-palate only individuals (Hoebel *et al.*, 2017) and *ACSS2* [Acyl-CoA Synthetase Short Chain Family Member 2] whereby a missense variant was identified in three different families with non-syndromic CLP (Aylward *et al.*, 2016); this gene has been shown to express in mouse facial tissues during development (Loikkanen *et al.*, 2002). Aylward *et al.* (2016) have also identified three different variants in *PHYH* [Phytanoyl-CoA 2-Hydroxylase], a gene that has been associated with rhizomelic chondrodysplasia punctata, which can include craniofacial anomalies such as micrognathia and high-arched palate (Barr *et al.*, 1993) and has been shown to interact with *PEX7,* a gene possibly linked to clefting (Aylward *et al.*, 2016; Jugessur *et al.*, 2009b). In addition to the novel single nucleotide gene variants reported above, Cai *et al.* (2017) identified small gene copy number variants using WES on multiplex families. A duplication event of 7.7 kb in the *ADH7* gene [Alcohol Dehydrogenase 7] in two affected brothers and their unaffected mother was found. Another family was found to have a deletion of 13.3-23.6 kb in the *AHR* gene [Aryl Hydrocarbon Receptor] in two affected brothers, their unaffected mother, and unaffected grandfather (Cai *et al.*, 2017). With regards to their significance in biological pathways, *ADH7* may participate in the synthesis of retinoic acid, and retinoic acid plays an important role in cellular differentiation and is a well-established cause of cleft palate (Abbott & Birnbaum, 1990); *AHR* encodes the arylhydrocarbon receptor (AHR), which is expressed in the developing mouse palate and is upregulated early in palatogenesis (Abbott *et al.*, 1999); a receptor that mediates the toxicities of aromatic hydrocarbons (Cai *et al.*, 2017). The aromatic hydrocarbon dioxin (2,3,7,8-tetrachlorodibenzo-p-dioxin) is a ligand of AHR and has been shown to induce cleft palate in pregnant mice (Abbott *et al.*, 1989; Cai *et al.*, 2017; Pratt *et al.*, 1984).

Table 1-4 WES studies in non-syndromic cleft lip and/or cleft palate populations

Cleft Exome Study	Cohort & Study Design	Sequencing Strategy Targeted vs. Unbiased	Subclinical Features Included Yes vs. No	Summary of Findings
(Bureau *et al.*, 2014)	NS-CLP 2-3 affected relatives from 55 multiplex families	Targeted-by filtering through 348 candidate genes/loci for oral clefts	Not determined. but Sanger sequenced unaffected relatives	None of the variants were found in more than one multiplex family. Found 4 missense & 1 nonsense novel SNVs that were shared by the affected distant relatives. One damaging SNV in *CDH1*, shared by three affected second cousins from a single family, attained statistical significance
(Liu *et al.*, 2015)	NS-CLP 8 foetuses from Chinese families	Unbiased	Not determined. Trios not sequenced	Found 16 new missense variants with unidentified pathogenicity, 5/16 were in one individual, and 13 reported missense variants all in one individual
(Aylward *et al.*, 2016)	NS-CLP 2-4 affected relatives from 27 multiplex Honduran families	Unbiased	Not determined. Sanger sequenced unaffected relatives if variant found in affected members	Four genes with candidate variants in ≥3 families. Candidate variants in two genes, *ACSS2* and *PHYH*, consistently segregate with NSCLP as a dominant variant with incomplete penetrance
(Fu *et al.*, 2017a)	NS-CLP 2-3 affected relatives from 115 multiplex cleft families from different ethnicities	Unbiased. Searched for CNV	Not determined	53 rare hemizygous deletions, 45 occurring in only one family member. Members of the same family shared a rare deletion in only eight regions
(Liu *et al.*, 2017)	NS-CPO 5 individuals from a multiplex family	Unbiased	Not determined in other family members	Novel missense variation in *ARHGAP29*

(Hoebel *et al.*, 2017)	NS-CPO 2 affected first-degree relatives from each family	Unbiased. 16 with NS-CPO had WES. Candidate genes were re-sequenced in other 132 NS-CPO cases.	Not determined. Parents included in study	2 or more missense variants in each of *ACACB, PTPRS, MIB1* in individuals from independent families. A novel variant in *GRHL3* in 1 patient and a variant in *CREBBP* in 2 siblings
(Cai *et al.*, 2017)	NS-CLP two to four members of 27 multiplex Honduran families. 52 affected individuals and 139 relatives	Unbiased. Searched for copy number changes (CNC)	Not determined.	3 CNCs corresponding to *ADH7, AHR*, and *CRYZ* segregating with NS-CLP
(Basha *et al.*, 2018)	NS-CLP 84 individuals from 46 multiplex families	Unbiased	Yes. Included subclinical features of probands & participants	Implicated syndromic genes in NS-CLP

Highlights and challenges of next generation sequencing

Exome and genome sequencing are the most phenotypically agnostic assays that can be used to diagnose a wide range of disorders (Wright *et al.,* 2018a). In the context of disease types, it is thought that genetic heterogeneity increases as phenotypic specificity decreases; that is, the less specific the phenotype associated with a disease is, the more likely it is to be caused by variants in a large number of individual genes; neuropsychiatric disorders are good examples of the latter. The converse of this is also true, in the sense that genetic heterogeneity decreases as phenotypic specificity increases, in other words, if the phenotype is defined the genotype becomes more specific. Therefore, if the disease is phenotypically and/or genetically very homogeneous, testing a single gene or small number of genes is preferable (Wright *et al.,* 2018a). Hence the success and diagnostic potential of WES in identifying causation in single gene/syndromic disorders. Studies have shown that in cases of congenital anomalies with 'associated' features, the diagnostic rate of exome sequencing is increased significantly, from 10% in isolated-sporadic cases, to 26-30% in 'associated' *syndromic* cases (Blue *et al.*, 2017; Powis *et al.*, 2017). Moreover, the detection of copy number variants is also significantly increased, from 3-10% in isolated cases, to up to 25% in syndromic cases. These CNVs could be detected using chromosomal microarrays or exome sequencing. Thus, demonstrating the utility and power of WES in genetic studies of *syndromic* forms of congenital birth conditions.

Sequencing family trios (parents and child) has helped overcome many of the rigid requirements for research design including the need for large case-control samples or the need for large families with multiple affected individuals. Sequencing of parent-child trios rather than individual probands, for families in which neither parent is affected by the same disorder, offers around a ten-times reduction in the number of candidate variants, thus substantially increasing the speed and likelihood of reaching an accurate diagnosis (Wright *et al.*, 2015). Thus, WES of parent-child trios, where only the child is affected, is a highly effective strategy in identifying *de novo* mutations (Bamshad *et al.*, 2011). Indeed, 2000 cases of children referred for clinical WES provided a molecular diagnosis rate of 25% with 58% of the diagnostic mutations not previously reported. Of the patients diagnosed with an autosomal dominant disease pattern (n=280) and with parental samples available, around 87% were shown to result from *de novo* mutations (Yang *et al.*, 2014). Some WES studies have adopted a proband-based approach; this is also the case in trio-based studies where either or both of the parents were not obtainable. Although moving from a family trio-based approach to a proband-only approach has practical and financial advantages, it has been shown to substantially reduce diagnostic yield from 40% to ~28% in studies of intellectual disability (Wright *et al.,* 2018a). This is because *de novo* status cannot be directly assigned to observed genomic variants to determine if they are on the same or different chromosomes.

While exome and genome sequencing have helped discover many new genes and have given genetic answers for a multitude of clinically ascertained syndromes, they are bioinformatically demanding. A major challenge in the field of high-throughput sequencing is the need to distinguish disease-causing sequence variants from the many potentially functional variants present in any human genome, thus avoiding detrimental false assignments of pathogenicity (MacArthur *et al.*, 2014). This can be a complex, multidimensional task and is particularly true for single nucleotide changes that produce missense variants as we often overlook or fail to consider the pathogenicity of these variants compared with the more explicit truncating mutations. Variant prioritization tools such as SIFT (Sorts Intolerant From Tolerant) (Kumar *et al.*, 2009), PolyPhen2 (polymorphism phenotyping version 2) (Adzhubei *et al.*, 2013) and Combined Annotation-Dependent Depletion (CADD) (Kircher *et al.*, 2014) have aided in computationally predicting the impact of missense and other variants (Eilbeck *et al.*, 2017). Yet, often, a large category of variants are assigned the term 'genes of uncertain significance'. This is a category that represents variants in genes that have not been

previously associated with a disease or have limited evidence for association and are often assigned research status according to the American College of Medical Genetics (ACMG) guidelines (Richards *et al.*, 2015).

Some of the guidelines proposed to overcome false reporting of gene variant causality include taking advantage of information available on public datasets of genomic variation and animal-model phenotypes, validating new genes by assessing that variants in the same gene and similar clinical presentations have been confidently implicated in multiple unrelated individuals, experimentally validating the predicted damaging impact of candidate variants using assays of patient-derived tissue or well-established cell or animal models of gene function and avoiding the assumption that implicated variants are fully penetrant, or completely explanatory in any specific disease case (MacArthur *et al.*, 2014). Put simply, to determine pathogenicity of a certain gene variant, the following factors are usually considered: a) the *de novo* status of a mutation increases the likelihood of its implication in disease, b) the variant is rare (i.e. rarer than disease frequency; the frequency of the variant to occur in a population must be rarer than the disease itself), c) the presence of other patients with the same variant and similar phenotypes, d) the genetic region is constrained for missense or loss of function (LOF) mutations and e) the variant is located in a functional domain on the protein.

Large-scale datasets have become instrumental in the diagnosis of rare disorders, as they can establish with greater confidence whether an observed mutation is likely causal for the phenotype (Bragin *et al.*, 2013). Despite advances in technologies used to sift through the millions of variants and the ability to narrow them down to just a handful, the tasks undertaken to differentiate whether an identified mutation is benign or pathogenic are challenging partly due to the rarity and novelty of the disorders we come across (Bragin *et al.*, 2013). The DECIPHER database (https://decipher.sanger.ac.uk) is an accessible online repository that is designed for the analysis and identification of potential candidate genes implicated in disease (Swaminathan *et al.*, 2012). The Deciphering Developmental Disorders (DDD) Study is a UK-wide collaborative project that links genomic sequencing technologies to health care provision within the National Health Service (NHS); it feeds genotypic and phenotypic data from exome sequenced family-trios of children with severe, undiagnosed developmental disorders into DECIPHER. The DDD study was established with the dual aim of *''assisting the translation of new high-throughput genomic technologies into clinical practice and elucidating the underlying genetic architecture of*

developmental disorders" (Wright *et al.,* 2015). Another advantage of having such datasets is the added value of revisiting them over time, thus aiding in the identification of new correlations and increasing diagnostic rates given the continual and rapid advancements in the area of gene discovery (Wright *et al.*, 2018b). With the ever growing era of next generation sequencing, large generated datasets are enriched with multiple novel and rare alleles that are now appreciated as important contributors to complex human diseases (McClellan & King, 2010). Utilizing these large-scale genomic datasets is pivotal in the diagnosis and management of rare disorders.

1.3.3 **Genetic Modifiers**

Perhaps the difficulty in gene discovery in orofacial clefting relates to the fact that varied inheritance modalities are always observed and varied segregation of phenotypes with a gene mutation in affected families is often seen. Indeed, Chong *et al*. (2015) showed that when multiple modes of inheritance are consistent with the segregation pattern observed in a pedigree or there is otherwise uncertainty about the correct mode of inheritance for a phenotype, the rate of gene discovery is considerably lower than when the mode of inheritance is known or easily predicted. Therefore, other *genetic modifiers* that deviate from the three classic Mendelian laws of inheritance (dominant, recessive and X-linked) include concepts such as 'variable expressivity' and 'incomplete penetrance' **(Figure 1-6)**. Deviations as such have been shown to reflect on the variation in the 'expression' of the palatal phenotype seen or the number of members affected in a family; 'penetrance'. This is true in the context of orofacial clefting and has been demonstrated in multiplex families (Bureau *et al.*, 2014; Fu *et al.*, 2017a). Recent advances in genomic technology attributes some of these differences to epigenetic changes in an individual's genome.

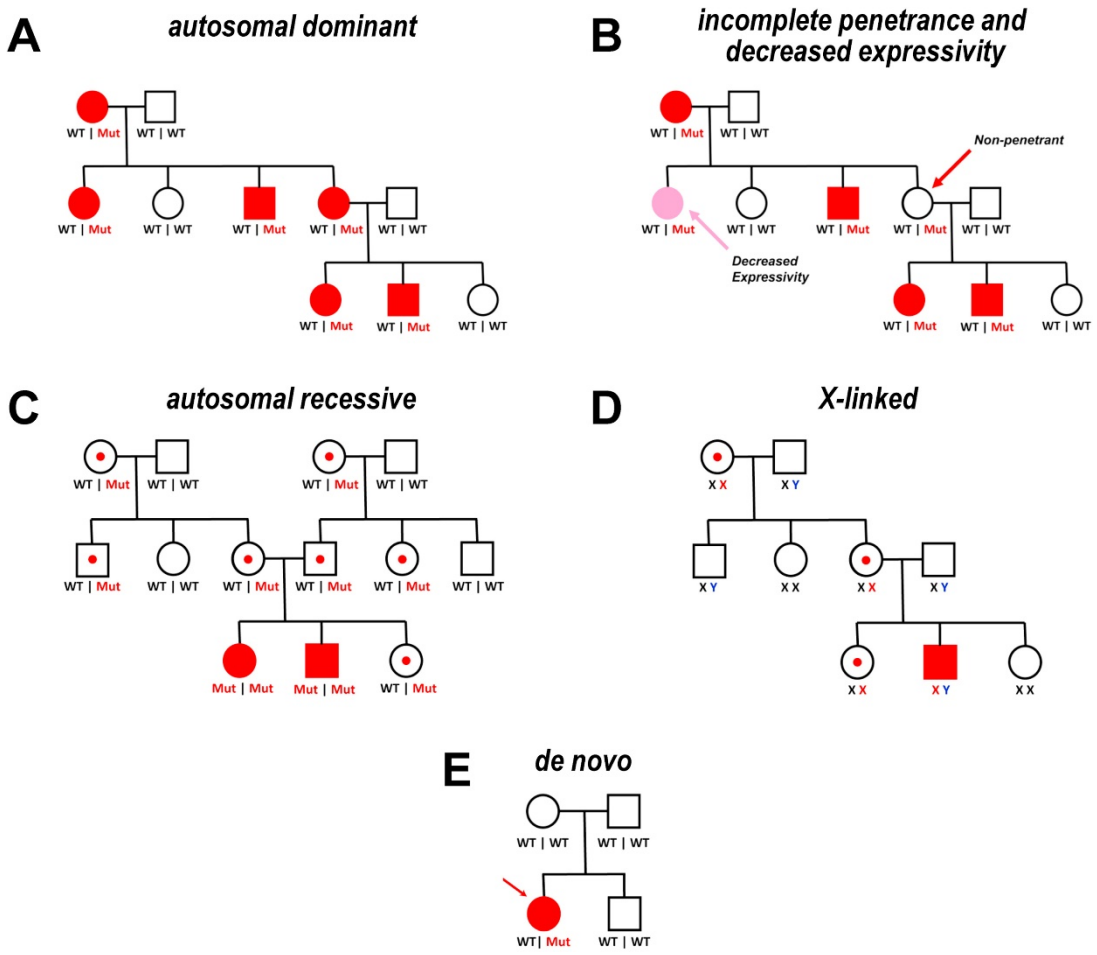


Figure 1-6 Example pedigrees illustrating mendelian inheritance and co-segregation

[A] A pedigree illustrating autosomal dominant transmission of a trait. Red symbols represent family members with the trait ("affected"). Genotypes are given beneath each pedigree symbol to indicate presence of wildtype (WT) or mutant (Mut) alleles. **[B]** Same pedigree as in (A) modified to illustrate incomplete penetrance (e.g., presence of mutant allele in a phenotypically unaffected person) and decreased expressivity (e.g., presence of mutation but with less severe disease). **[C]** Pedigree illustrating autosomal recessive inheritance. Open symbols with a central red dot represent unaffected heterozygous mutation carriers. **[D]** Pedigree illustrating X-linked inheritance. Genotypes are given beneath each pedigree symbol to indicate presence of wildtype (black X, blue Y) or mutant (red X) sex chromosomes. **[E]** Pedigree illustrating occurrence of a *de novo* mutation. Adapted from (McNally & George, 2015).

The distinction between complex and Mendelian forms of disease is becoming increasingly blurred and eventually might be viewed as a continuum apart from single gene disorders (Blue *et al.*, 2017). The concept of additional genetic burden in specific genes has become evident in a number of sporadic and syndromic forms of diseases such as autism and congenital heart disease, suggesting a combination of *de novo* and inherited rare variants in disease causation (Blue *et al.*, 2017). Indeed, in families where a presumed single causal variant for congenital heart disease was found, additional genetic variation comprising rare and low frequency variants (i.e. MAF <0.05) were discovered, however, the relative contribution of the additional variants to the development of the heart defect is still unclear (Blue *et al.*, 2014). An example of this concept in the context of congenital heart disease is shown in **(Figure 1-7)**. On the same note, it has been proposed that in cleft cases in which a copy number variation (deletion) is confined to one gene, that this hypothetically may present a cleft-only anomaly, whereas if the deletion encompassed multiple genes, the cleft becomes associated with other developmental and physical anomalies (Shi *et al.*, 2009). When family-based samples are used, these deletions/duplications can either be found by dosage differences in the probands compared to controls, or by evidence of apparent non-Mendelian transmissions from parents to a hemizygous child (Shi *et al.*, 2009). In a study of 2000 clinically exome-sequenced cases referred for suspected genetic conditions, 4.6% of the cases with multiple phenotypes (n=504) that had positive results harboured two molecular diagnoses within their genomes, highlighting the oligogenic models of disease aetiology and demonstrating that simple Mendelian gene effects can compound to yield complex genetic profiles (Yang *et al.*, 2014). More research is required to understand the concept of 'burden of genetic variation' in disease as this could potentially explain the reduced penetrance and variable expressivity that often accompany familial clefts.

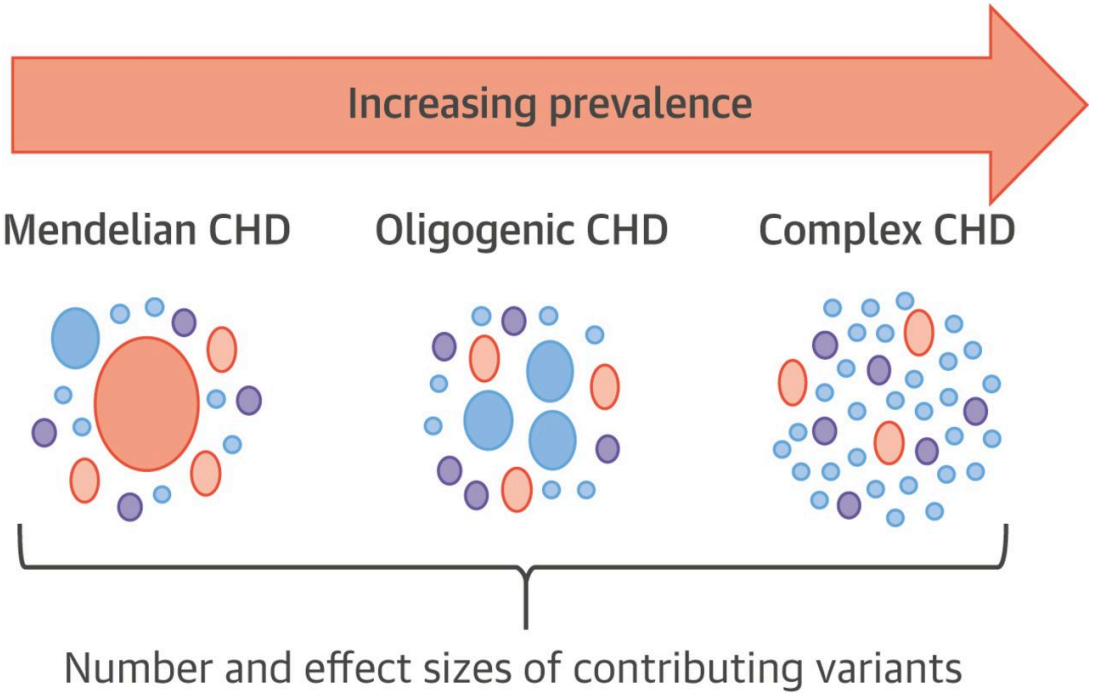


Figure 1-7 The number and effect sizes of the contributing genetic variants for the different inheritance modes observed in congenital heart disease (CHD) in conjunction with disease prevalence

Each **spot** represents a variant contributing to the phenotype and the **size** of the circle is representative of the effect of the variant on the phenotype. Single-gene disorders are caused by rare variants with large effect sizes. In addition to the main causal variant, which typically exhibits a Mendelian pattern of inheritance, several other non-Mendelian variants contribute to expression of the phenotype. On the opposite end of the spectrum are the common complex traits, which are caused, partly, by the cumulative effects of a large number of sequence variants, each imparting a modest effect size. In oligogenetic phenotypes, several alleles with moderate size effects and a large number of alleles with small effect sizes contribute to the phenotype. Taken from (Marian, 2012) and (Blue *et al.*, 2017).

With great advancements in the field of genetics in orofacial clefting, we now appreciate that variation in clinical phenotypes could partly be explained by the *nature* of the mutation and the subsequent effects it has on protein domains or motifs. For instance, mutations in the interferon regulatory factor 6 (*IRF6*) are associated with non-syndromic CLP (Beaty *et al.*, 2010; Ludwig *et al.*, 2012) and isolated hypodontia (Vieira *et al.*, 2007) and with both van der Woude syndrome [MIM: # 119300], an autosomal dominant disorder in which the lower lip pits and tooth agenesis are the only features distinguishing it from isolated CLP (Kondo *et al.*, 2002), and popliteal pterygium syndrome [MIM: # 119500], which shares some clinical features of VWS with the addition of webbed skin and fibrous bands in the mouth (Peyrard-Janvid *et al.*, 2005).

Locus and/or mutation class differences for *IRF6*-related disorders have been proposed for non-syndromic CLP. A statistically significant association between *IRF6* and bilateral CLP at the haplotype level, but not at a SNP-based level, has been reported (Kerameddin *et al.*, 2015) and *IRF6* has been suggested to serve as a potential marker of severity for non-syndromic CLP (Leslie *et al.*, 2016b). The same concept could be applied to the well-known associations between msh homeobox 1 (*MSX1)* and CLP. *MSX1* has been associated with syndromic [MIM: # 106600] (van den Boogaard *et al.*, 2000), and non-syndromic cases of clefting and is linked to isolated hypodontia. Recent associations showed that *MSX1* truncations cause more severe phenotypes than in-frame variants and that mutations in the homeodomain of the protein always cause tooth agenesis with or without other phenotypes while mutations outside the homeodomain are mostly associated with non-syndromic orofacial clefts (Liang *et al.*, 2016). Alterations in protein structures have also been shown to reflect the phenotypic variations produced. Take for example nonsense mediated decay (NMD), a biological process that acts as a quality control measure that degrades mRNA harbouring a premature termination codon to prevent the synthesis of truncated proteins. There are genetic conditions in which NMD can modulate phenotypes. Even when disease results from NMD-induced protein deficiency, the disease phenotype may, arguably, be milder than, and different from, that caused by an expressed truncated protein (Hwang & Maquat, 2011; Maquat, 2004).

1.3.4 **Establishing Genotype-Phenotype Relationships**

The recognition of craniofacial phenotypic features in combination with genetic testing, i.e. establishing phenotype-genotype relationships, greatly aids in the genetic diagnosis of many undiagnosed syndromes and rare developmental diseases.(Bragin *et al.*, 2013; Twigg & Wilkie, 2015). The 'Expansion of Phenotypes' is a revised concept in genetics that could be applied in the findings of many Mendelian disorders and sporadic conditions (Chong *et al.*, 2015). The prevailing method by which new Mendelian conditions are genetically analysed relies on assembling persons with highly similar phenotypes and subsequent gene discovery within the assembled group. This approach successfully enhanced gene discovery in clinically ascertained syndromes particularly when hindered by limitation of the previous standard gene testing strategies. However, undefined or uncharacterised conditions or mutations in the same gene that result in new or different sets of phenotypic features are difficult to discover using this strategy. Interestingly, an analysis of all genes associated with a Mendelian phenotype on OMIM has shown that 24% of the genes are responsible for at least two 'clinically discrete'

phenotypes/syndromes (Chong *et al.*, 2015). Moreover, having relatives affected by one or more of cleft subtypes in a family allows one to consider whether subforms of a trait are different degrees of the same process or individually transmitted developmental processes (Fraser, 1980). Intrafamilial variability in phenotypic expression is seen in autosomal dominant conditions in successive generations, an example of this concept is often seen in Treacher Collins syndrome where there is often considerable variance in the extent of phenotypic expression of the disorder, usually tending to be more severe in subsequent generations (Argenta and Iacobucci, 1989).

Phenotypic characterization of families affected by Mendelian conditions has been key in understanding genetic contribution to disease. A world-wide multi-centre project led by the Centres for Mendelian Genomics (CMGs) – an initiative for undertaking large-scale WES for the discovery of genetic variants for Mendelian phenotypes – showed that to date, more than 2,937 genes underlying 4,163 Mendelian phenotypes have been discovered. Yet, genes underlying ~50% (i.e., 3,152) of all known Mendelian phenotypes are still unknown, suggesting many more Mendelian (or syndromic) conditions have yet to be recognized (Chong *et al.*, 2015). Indeed, the CMGs collaboration, alone, has identified 956 genes that underlie a Mendelian phenotype, including 375 novel genes not previously associated with human health. Most notably, this collaborative effort led to the expansion of clinical features for 219 known syndromic conditions.

We have seen from the evidence displayed in this chapter that genetic discovery in syndromic clefts has often preceded non-syndromic gene discovery regardless of the technology used. This is true in the case of linkage studies in cleft and is also true for more recent WES approaches. In his hallmark paper, Eiberg *et al.* (1987) commented, "*while many kinds of monogenic syndromic orofacial cleft have been recognized as Mendelian traits, and with all of the three major modes of inheritance represented, the results of genetic analyses concerning non-syndromic kinds of cleft have been less definite*", (Eiberg *et al.*, 1987). This holds true to date.

In the context of orofacial clefting, evidence on the contribution of syndromic cleft genes in non-syndromic clefts is abundant. Poliovirus Receptor Like-1 (*PVRL1*) a gene implicated in cleft lip and palate/ectodermal dysplasia 1 syndrome [MIM: # 225060] encodes a protein that acts in the initiation and maintenance of epithelial adherens junctions. Its contribution to cases of non-syndromic CLP in multiple populations has also been confirmed (Avila *et al.*, 2006). Likewise, mutations in *PTCH* are implicated in nevoid basal cell carcinoma syndrome, or Gorlin syndrome [MIM: #

109400] which includes craniofacial anomalies in which cleft palate has been reported in 4% of the cases. Variants near *PTCH* may act as modifiers of CLP and missense mutations in *PTCH* may be considered as rare causes for non-syndromic CLP (Mansilla *et al.*, 2006). Interestingly, syndromic/non-syndromic gene implications are seen in other multifactorial diseases. For instance, the first congenital heart disease (CHD)-targeted gene panel identified the cause for the heart defects in a third of the cohort comprising non-syndromic CHD. Interestingly, in around 50% of those diagnosed, the causal gene was associated with a specific syndrome (Blue *et al.*, 2014). Furthermore, syndromic cleft genes such as *FGFR1* and *IRF6* have been implicated in isolated hypodontia, isolated clefting, or clefting with hypodontia. *FGFR1* has been shown to cause Kallmann syndrome 2 [MIM: # 147950] which is characterized by an impaired sense of smell and incomplete or delayed puberty; CLP is interestingly associated with 30% of the cases and hypodontia in 5-10 % of the cases (Dode *et al.*, 2003). *FGFR1* also causes Type 1 Pfeiffer syndrome [MIM: # 101600] characterized by fusion of the bones in the skull, hands and feet (Muenke *et al.*, 1994b; White *et al.*, 2005). *FGFR1* has also been implicated in non-syndromic CLP (Riley *et al.*, 2007) and in cases of non-syndromic tooth agenesis (Vieira *et al.*, 2007). What is interesting is that these different presentations can occur within the same family with the same mutation. Despite the vast number of the aforementioned associations, full gene variant-phenotype relationships have not yet been elucidated for many of these conditions.

Cleft subtypes have also been the subject of phenotype-genotype studies. It has long been affirmed that sub-grouping clefts by means of their anatomical types into those that include the palate only and those that involve the primary palate including clefts of the lip, is a categorization that reflects the biologic, embryonic and familial bases of their occurrences. This has been postulated on the basis of embryological observations, where the primary and secondary palate form independently and on observed familial patterns where it has been thought that it is unusual to find an occurrence of cleft palate only (CPO) in a family if the index case had CLP or vice versa (Fraser, 1970; Murray, 1995). This, in fact, holds true on account of evidence from GWAS studies. In the only GWAS of non-syndromic cleft palate only to date, single-marker association analyses failed to identify any genome wide significant association, and none of the regions previously associated with non-syndromic CLP have yet been shown to confer an effect in non-syndromic CPO (Ludwig *et al.*, 2014; Mangold *et al.*, 2010).

With more recent evidence on single-gene variants from sequencing studies, it is now believed that it is not unusual to find mixed subtypes of clefts within the same

family and that a single gene can be implicated in both subtypes. Perhaps the first to shed light on this was the study by (Ludwig *et al.*, 2014). The authors showed that *FOXE1* was the first gene that can be considered a common risk factor for both non-syndromic CLP and CPO, whereby two markers or SNPs (rs3758249, rs4460498) were strongly associated with both subtypes, indicating that *FOXE1* plays a role in two phenotypes thought to have been genetically distinct. Notably, mutations in *FOXE1* cause Bamforth-Lazarus syndrome, a syndromic form of orofacial clefting in which thyroid dysgenesis is an additional symptom (Castanet *et al.*, 2002; Clifton-Bligh *et al.*, 1998). Likewise, it is often thought that it is unusual for a single syndrome or genetic disorder to include various types of clefting such as CLP and CPO. Nonetheless, this type of mixed clefting has been shown to occur with *MSX1*, *IRF6* and *FGFR1*-related disorders (Bjork *et al.*, 2003; van den Boogaard *et al.*, 2000).

Taking the anatomic and genetic modifiers outlined in this chapter into consideration, and combining those with genetic variants and environmental influences, provides probable explanations for the phenotypic heterogeneity seen in orofacial clefting. This heterogeneity could form a phenotypic continuum model, with each cleft phenotype harbouring its own underlying interacting mechanisms. In view of everything we have learned from linkage analyses, GWAS and more recently, exome sequencing, we have a much clearer picture of cleft causality than ever before. Although we cannot yet pinpoint the genetic variant(s) causing the cleft defect in every patient, we have better understandings of the types of variants involved and the functional roles they play, and the opportunity available to uncover many more genes and gene variants.

1.4 AIMS OF THIS THESIS

The overarching aim of this thesis was to identify novel *de novo* gene mutations in patients with orofacial clefts and to analyse the novel candidate genes for their pathogenicity.

The aim was to find child and family trios with either 'cleft-only', 'cleft-tooth anomaly' or 'cleft-medical & congenital conditions +/- tooth anomaly', and to categorize them into those with a potentially inherited genetic aetiology and those potentially arising from *de novo* gene variations.

Hypothesis: probands from family trios in the 'cleft-medical' group are likely to have underlying novel *de novo* gene variants.

Chapter 2 MATERIALS & METHODS

SUMMARY

This chapter will describe the methods used in the studies. **Sections 2.1-2.20** include general clinical and/or laboratory methods commonly used throughout the project. Specific methods or protocols used for each results chapter will be described in their relevant sections below (See **Sections 2.21, 2.22, 2.23 & 2.24** for **Chapters 3,4, 5 & 6,** respectively).

2.1 PATIENT ETHICS

2.1.1 Clinical Study Ethics

Ethical approval was granted by the Office for Research Ethics Committees, Northern Ireland (Research Ethic Committee Reference Number, REC REF: 16/NI/0026) and by the Research and Development Department at Guy's and St. Thomas' NHS Foundation Trust (IRAS Project ID: 185686). Ethical Approval correspondence can be found in (Appendix 2). Patient Information Sheets and Consent Forms were designed by the author and approved by the Ethics Committee; these can be found in (Appendix 3).

2.1.2 Patients from the Deciphering Developmental Disorders Study

The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003], DDD (10/H0305/83, Cambridge South REC, and GEN/284/12, Republic of Ireland REC). This study makes use of DECIPHER (http://decipher.sanger.ac.uk), which is funded by Wellcome.

2.2 PATIENT CONSENT

Informed consent from all participants included in this thesis was obtained for publication of data and photographs in the medical literature. All families tested were offered genetic counselling.

2.2.1 The Clinical Study

Informed consent was sought from each participating parent(s) and assent was sought from participating children. Participants were given the option to withdraw from the study at any timepoint. Participants were also given the option to consent to store their samples for future cleft-related studies or to discard them at the closure of the study. All information disclosed in the study was kept confidential; each participant was

anonymised. Samples collected had the participant's ID as the only identifier. Research data were stored using an encrypted USB, a password protected computer and a secure locked cabinet.

2.2.2 The Deciphering Developmental Disorders Study

Patients located via the DDD study also gave their consent to their respective clinical geneticists for their data to be used and shared through the DDD study and database.

2.3 PARTICIPANTS' INCLUSION AND EXCLUSION CRITERIA

A prospective study design of family 'trios' of children with non-syndromic or 'associated' clefts without a genetic diagnosis was carried out from May 2016 to July 2018. The setting was the South Thames Cleft Service (STCS), at Guy's and St Thomas' NHS Foundation Trust, which is one of the regional cleft centres in the United Kingdom. Families with children who had been diagnosed with a cleft-related syndrome, a chromosomal anomaly or a verified genetic diagnosis were excluded as well as children accompanied by someone other than their biologic parents and parents that required an interpreter to provide written informed consent.

2.4 PATIENT RECRUITMENT

Parent(s) and family members that verbally displayed their agreement to take part were consented and their child assented. At least two close family members (e.g. parent, sibling or grandparent) were interviewed by the author and underwent a head, neck and oral exam in a dental setting and provided medical and dental histories and history of cleft conditions in the extended family. The child's detailed cleft condition was obtained from the Trust's medical and dental records. Participants' phenotypes were collected in a clinical proforma (the Family Booklet, see Appendix 4). Findings were confirmed by a second examiner (Dr Nabina Bhujel (NB), Consultant Cleft Paediatric Dentist) when needed, particularly for dental anomalies in parents such as hypodontia and molar-incisor-hypomineralisation (MIH). The inter-rater reliability score was 0.83 (i.e. almost perfect agreement). Fifty-nine of the children had dental radiographs available and these were viewed by the author and a third researcher (Ms. Asma Alshahrani (AA), as part of her MSc in Paediatric Dentistry). The radiographic images used for the probands were requested as part of the patient's routine clinical management and used in this project to report dental anomalies. The radiographs of the

participants were viewed on Planmeca Romexis dental imaging. The dental anomalies were recorded, together with the relationship to the proband's cleft site, according to accepted definitions (See **Table 2-21).** If an anomaly was present in the child or family member, it was sub-classified as 'within' the cleft site or 'outside' the cleft site and further classified according to location in the maxilla or in the mandible. Isolated enamel defects were separated out from the definition of a 'dental anomaly' because some of these might have an environmental aetiology (e.g. fluorosis, caries or trauma).

Since family trios were included in the study, often the third family member was not present. Hence, an envelope was prepared with the study documents and a saliva kit and sent with the parent/family member accompanying the child. In this circumstance, if the person could not attend, a photograph of their dentition was emailed to the author following the consent of the participant or the participant only took part in the genetics study (sample collection). The children were grouped into one of the following categories: 'cleft-only' (medically healthy/no dental anomalies), 'cleft-tooth' (tooth anomaly present but no medical condition) and 'cleft-medical condition' (associated congenital anomaly or medical co-morbidity +/- dental anomaly). Sixty-three families also provided saliva for gene testing. These were taken for studies involving whole exome sequencing. The data were recorded, anonymised and tabulated and entered on an SPSS spreadsheet and descriptive data were tabulated and analysed using IBM® SPSS® Statistics, Version 25. A flowchart of the recruitment process is shown in **(Figure 2-1)**

Participants recruited from the Deciphering Developmental Disorders Study were located either via the DECIPHER website (https://decipher.sanger.ac.uk) or through the DDD Complementary Research Proposal (CAP)180 dataset described in Chapter 6, Section 6.2.1.

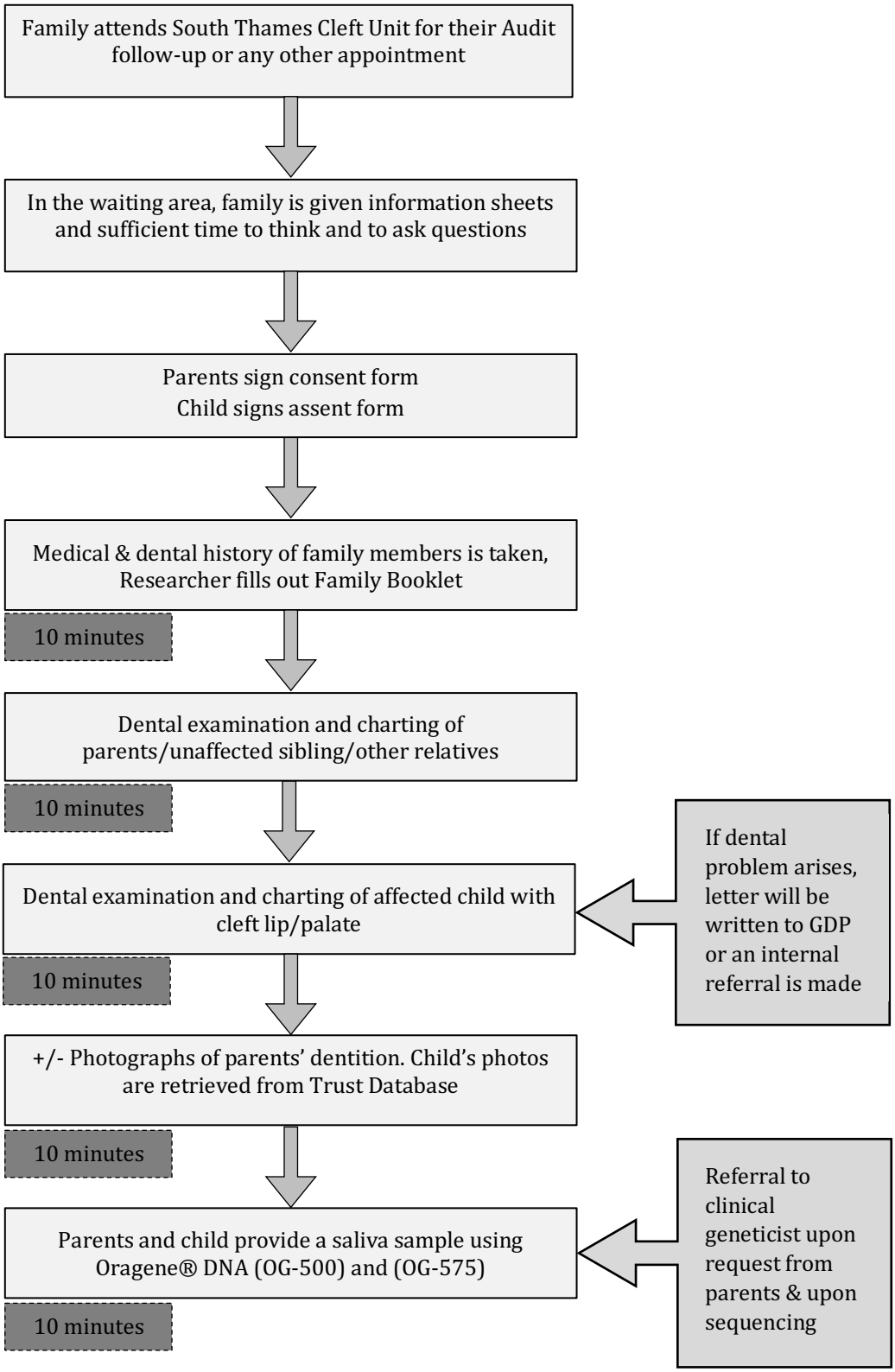


Figure 2-1 Flow diagram demonstrating families' pathway on the day of their recruitment

2.5 HUMAN SPECIMENS

Human embryonic and foetal material was provided by the Joint MRC/Wellcome Trust (Grant # 099175/Z/12/Z) Human Developmental Biology Resource (HDBR, http://www.hdbr.org) as whole embryos (Carnegie stage 13 (C13, day 28-32)) or sectioned embryos (Carnegie stage 21 (C21, day 50-52)).

2.6 MODEL ORGANISMS ETHICS

Animal work was performed in accordance with the UK Home Office Project License P8D5E2773 at King's College London (KJL), University of Texas Southwestern Medical Centre Institutional Animal Care and Use Committee Protocols (for mice provided by Dr Denise Marciano for Chapter 4), the European *Xenopus* Resource Centre, Portsmouth UK, or the Yale University Institutional Animal Care and Use Committee protocols (for frogs provided by Dr Mustafa Khokha for Chapter 4).

2.7 REAGENTS

Table 2-1 General reagents

Reagent	***Supplier***
Ethanol	Fisher Chemical, 1730528
Isopropanol	Acros Organics, 389710025
Methanol	Honey Well, 179957
Phosphate Buffered Saline (PBS)	Fisher, BP-665-1
Triton®X-100	Sigma, X100
Tween-20	Sigma, P7949
Ethylenediaminetetraacetic acid (EDTA)	VWR, 20303.293
Bovine Serum Albumin (BSA)	Sigma, A9647
Goat Serum	Sigma, G6767
Proteinase K	20µg/ml Sigma, P2308
Sodium Chloride (NaCl)	Fisher, S/3160/60
Sodium Acetate (NaOAc)	VWR, 27653-260
Trizma® base (Tris base)	Sigma, T1503
Sodium Hydroxide (NaOH) Pellets	Sigma, 1310-73-2
Hydrochloric Acid (HCL)	Sigma, H1758
Phenol-Chloroform	Qiagen 79306
Chloroform	Sigma, C2432

Table 2-2 Fixatives

Fixative	***Supplier***
Formaldehyde	Sigma, F8775
Glutaraldehyde	Sigma, G5882
Paraformaldehyde (PFA)	Sigma, P6148

Table 2-3 Polymerase chain reactions for sanger sequencing and gel electrophoresis

Reagent	***Supplier***
GoTaq® G2 Flexi DNA Polymerase	Promega, M780A
5X Colourless GoTaq® Flexi Buffer	Promega, M890A
5X Green GoTaq® Flexi Buffer	Promega, M891A
MgCl2	Promega, A351B
UltraPure™ Agarose	Fisher, 16500500
Ethidium Bromide	Fisher Chemical, 1239-45
DNA Gel Loading Dye (6X)	Fisher, R0611

Table 2-4 Cloning

Reagent	***Supplier***
Agar Bacteriological	Oxoid, LP0011
Tryptone	Oxoid, LP0042
One Shot® TOPO 10 competent E. coli	Fisher, C404003

Table 2-5 Wax sections and histology

Reagent	***Supplier***
Ultraplast Polyisobutylene Histological Wax	Solmedia, WAX060
Ehrlich's Haematoxylin	Solmedia, HST003
Eosin, Aqueous Solution	Riedel-de Haen, 32618
Histoclear	National Diagnostics, H3-204
Xylene	Sigma, 534056
DPX New	Merck, 100579
Neo-Mount	VWR, Cat. No. 1.09016.0500

Table 2-6 Probe synthesis, buffers and restriction enzymes

Reagent	***Supplier***
10x Buffer	Promega
BamH1 Restriction Enzyme	Promega
KpnI Restriction Enzyme	Promega
RNA Loading Dye	Promega
RNA Polymerase Enzyme (T7)	Promega
T7 high yield RNA synthesis kit	NEB, E2040S
Rnase Inhibitor	Promega, N251A
DIG (NTPs) RNA Labelling Mix (10X)	Roche, 1127707
DL-Dithiothreitol (DTT)	Promega
5x Transcription Buffer	Promega
DNaseI	Promega
Glycogen	Invitrogen, AM9510
Lithium Chloride (LiCl)	Sigma, L7026

Table 2-7 In situ hybridization reagents and powders

Reagent	***Supplier***
Sarcosyl (N-lacrolysercosine sodium salt solution)	Sigma, L5777
Proteinase K	20µg/ml Sigma, P2308
RNaseA	Invitrogen, 12091-0391
Anti-Digoxigenin-AP Fab Fragments	Roche, 110932274910
BM Purple	Roche, 11442074001
Tetramisole hydrochloride (Levamisole)	Sigma, L9756
Dextran Sulphate	Chemicon, 0702051849
Formamide	Merck, K36952408
Trieholamine (TEA)	Sigma, T58300
Acetic Anhydride	DBH, 100022M
Yeast tRNA (10 mg/mL)	Fisher, AM7119
Sodium Citrate Dihydrate	Sigma, 6132-04-3
Citric Acid	Sigma, 251275
Magnesium Chloride (MgCl2)	Fisher, BP214-500
WHOLE MOUNT	
Glycine	Alfa Aesar, A13816
Maleic Acid	Sigma, 110-16-7
Glutaraldehyde	Alfa Aesar, A17876
Sodium dodecyl sulfate (SDS)	Sigma, 151-21-3

Table 2-8 Immunofluorescence

Reagent	***Supplier***
Goat Serum	Sigma, G6767
Fluoroshield Mounting Medium with DAPI	Abcam, ab104139
Citi Fluor	EMS, E17970-100
Glycine powder	Alfa Aesar, A13816

2.8 COMMERCIAL KITS

Table 2-9 List of commercial kits

Commercial Kit	***Cat. Number***	***Usage***
TOPO® TA Cloning Kit -Dual Promoter pCR®II-TOPO® Vector	Invitrogen, 45-0640	Synthesizing plasmids
Oragene® DNA (OG-500)	DNA Genotek	DNA saliva collection tubes for adults and older children
Oragene® DNA (OG-575)	DNA Genotek	DNA saliva collection tubes for toddlers
Oragene® prepIT•L2P	DNA Genotek	DNA extraction and purification kit
QIAquick PCR Purification Kit Print	Qiagen, 28106	For the purification of PCR products
QIAquick Gel Extraction Kit	Qiagen, 28706	Extracting bands from agarose gels

2.9 ONLINE RESOURCES

Table 2-10 Online resources used throughout the study

Website	URL	Use
NCBI	https://www.ncbi.nlm.nih.gov	Gene & protein transcripts and annotations
NCBI BLAST	https://blast.ncbi.nlm.nih.gov	Search regions of similarity between biological sequences
ENSEMBL	https://www.ensembl.org	Gene transcripts and nucleotide sequence mark-ups
UCSC	https://genome.ucsc.edu	Genome browser
UNIPROT	https://www.uniprot.org	Protein annotation & domains
MUSCLE	https://www.ebi.ac.uk/Tools/msa/muscle	Multiple sequence alignment tool
WatCut	http://watcut.uwaterloo.ca	Online tool for analysing DNA sequences with restriction enzymes
EXPASY	https://web.expasy.org/translate	DNA to protein translator
Primer3Plus	http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi	Forward & Reverse primer designing software
ExAC	http://exac.broadinstitute.org	Exome Aggregation Consortium
gnomAD	https://gnomad.broadinstitute.org	Genome Aggregation Database
DECIPHER	https://decipher.sanger.ac.uk	Patient & variant database for the Deciphering Developmental Disorders study
Genic Intolerance	http://genic-intolerance.org	A database for RVIS (Residual Variation Intolerance Score)
GDI Server	http://pec630.rockefeller.edu:8080/GDI	The Gene Damage Index (GDI) Server
SnapGene® Viewer	Computer Software https://www.snapgene.com	To analyse cloning vectors, sequences and chromatograms
SWISS-MODEL	https://swissmodel.expasy.org/	A structural bioinformatics webserver dedicated to homology modelling of 3D protein structures
Mutation Taster	http://www.mutationtaster.org/	*In silico* software to predict cDNA and protein position from genomic position and vice versa, based on transcript IDs
Mutalyzer	https://mutalyzer.nl/position-converter	A software that converts variant coordinates. Provides a series of tools to check variant HGVS nomenclature and convert between different reference sequence systems

Variant Validator	https://variantvalidator.org	A software that converts variant coordinates. Enables accurate validation, mapping and formatting of sequence variants using HGVS nomenclature
OMIM	https://omim.org/	Online Mendelian Inheritance in Man®. An Online Catalogue of Human Genes and Genetic Disorders

2.10 **DNA EXTRACTION FROM SALIVA**

Saliva samples were collected from family trios using the Oragene® (OG-500) and (OG-575) collection tubes. Parents were asked to collect around 2mL of saliva, children under 5 were assisted by the author to collect around 0.75mL of saliva using a collection sponge that comes with the Oragene® (OG-575) kit. DNA extraction was carried out using the reagent provided with the Oragene® prepIT•L2P kit as per manufacturer's protocol.

2.11 **EXOME SEQUENCING**

Specifics for each patient are described in the methods sections for each chapter (See **Sections 2.22.4** & **2.23.1**, for Chapters 4 & 5, respectively).

The flow diagram **(Figure 2-2)** describes general methods for a whole exome sequencing process and highlights steps carried out by the bioinformatician and those carried out by the author. The flowchart in **(Figure 2-3)** describes the steps used in the project when gene variants were returned following whole exome sequencing. For the current project, exome sequencing was carried out on six families of children with cleft & associated congenital and medical anomalies. Eleven children within the study were referred for counselling with the team clinical geneticist (Dr Ana Beleza).

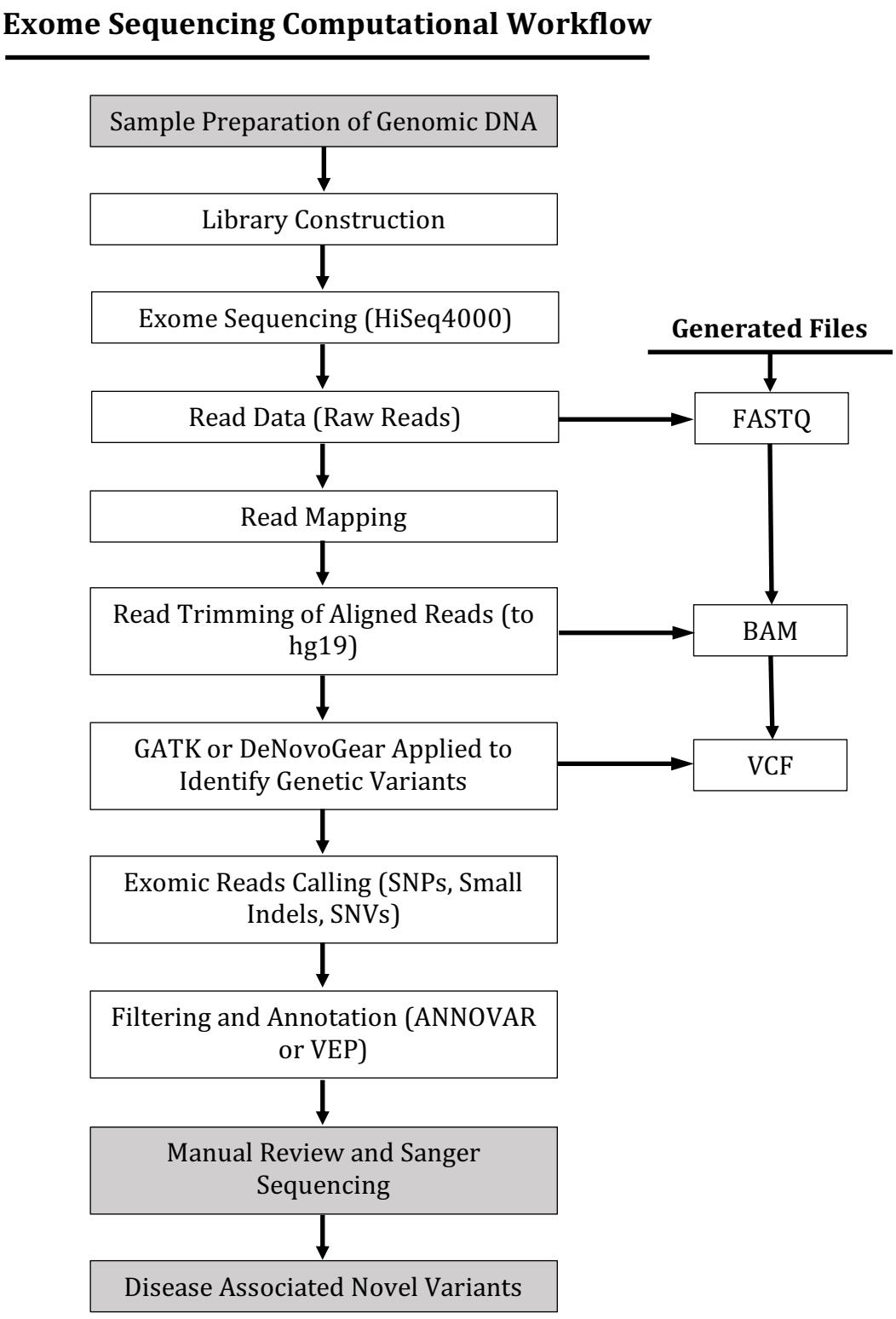


Figure 2-2 Overview of whole exome sequencing workflow

The flowchart shows an example workflow used generally for WES. White boxes are computational methods carried out by Dr. Weizhen Ji, the team bioinformatician. Grey boxes are parts where the author performed the method of interest.

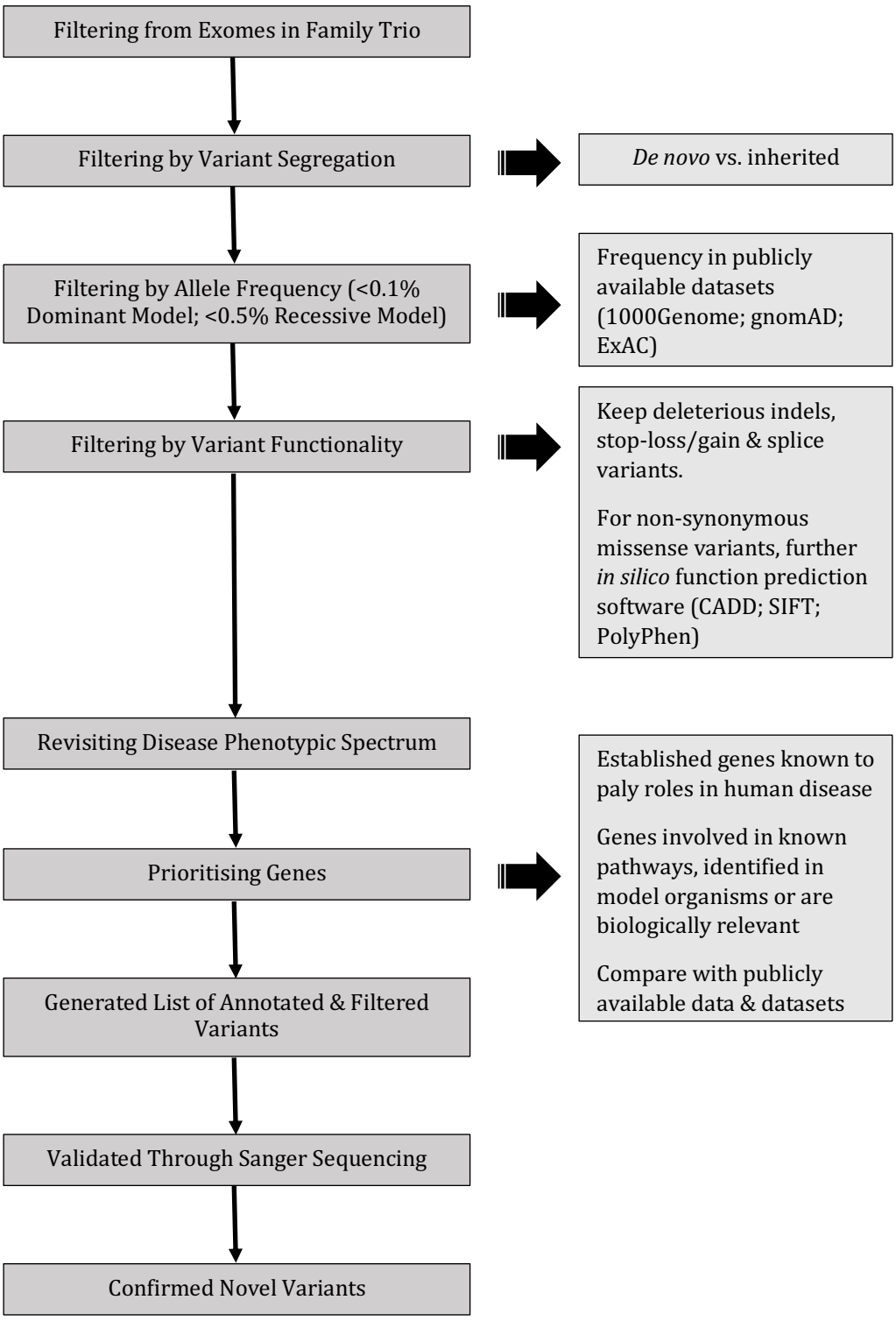


Figure 2-3 Filtering pipeline for the prioritization of variants from whole exome sequencing

The flowchart describes the steps used in the project when gene variants were returned following WES.

2.12 GEL EXTRACTION AND SAMPLE PREPARATION FOR SANGER SEQUENCING

First, genomic DNA extracted from saliva for the probands and their family members included in the study was used in the polymerase chain reaction (PCR) using primer pairs designed for each gene **(see Table 2-11).** The Taq polymerase GoTaq® G2Flexi DNA polymerase was used and 0.5-2µl of DNA was added to the PCR reaction with the following mix (n=1, 12µl): H20 6.2µl, 5X Buffer 2.5µl, MgCl2 0.75µl, dNTPs 0.1µl, primers (diluted 10-fold: 10µm forward and 10µm reverse primers) 0.75µl, Taq polymerase 0.1µl. Generally, the following thermocycling conditions were set-up: 95°C 4-5 minutes; 95°C 30 seconds; varied annealing temperatures (see **Table 2-11** for optimised temperatures) 30 seconds; 72°C 30 seconds; for 30-35 cycles and 72°C 2-5 minutes. A gel of 2% Agarose and 0.05% Ethidium Bromide in 1x TAE (tris-acetate) buffer was used to perform electrophoresis.

Following the initial confirmation of the presence of a band at the expected amplicon size, the same PCR reaction was repeated for each target gene, only this time a total of 50µl of PCR product was used for the gel extraction procedure and the 5X Colourless GoTaq® Flexi Buffer was used instead. Following gel electrophoresis using an 0.8% agarose gel in 1X TAE buffer and 0.05% Ethidium Bromide, the gel was visualized under the Syngene™ UV Transilluminator (2020LM) and photographed using the Syngene™ gel documentation system. Bands were extracted at the predicted amplicon size, melted and purified using the QIAquick Gel Extraction Kit. The samples were then Sanger sequenced as per SourcebioScience sample preparation instructions.

Table 2-11 Primers for polymerase chain reaction

Gene Name/Mutation	Sequence	Optimised Annealing Temp	Expected band size	Project used for
CTNND1, Position g.57569629	5'-AAAAGGAAGTGCACCTTGGA 3'-GAGAGCACATGCTCCAATCA	58	481	Chapter 4
CTNND1, Position g.57569629	5'-CGCTTGGATGCTGTCAAGT 3'-AGATGGATTTGACCCACAGC	52	420	Chapter 4
CTNND1 het Position g.57564451	5'- GCATTGAGGAGCGGTATAGG 3'- GCCAAGTCAGAAAAGGGAAA	52	419	Chapter 4
DMXL2 Position g.51773257	5'- GGCGCTCATAGGAACCAATA 3'- TGACTGGAGTCAGCCAATAGT	52	~500	Chapter 4

DMXL2 Position g.51791234	5'- TCCTGGATTTGAAACAGCTC 3'- GAGGCTGCACATGTACTTTCC	52	~450	Chapter 4
TENM3	5'-GCTGTCAGCCTTCTGGTCA 3'- GGTAGAAGGCGTTGTTGAGC	54	~420	Chapter 5
CDK16	5'-TGGTTGTCATGACGATGAGTG 3'-CATGGATGGGGATCTTTGTC	52	393	Chapter 5
AGAP6-a	5'-CCAAAGCTGTGAGCAGAGG 3'-TGTGCTTGACTCTGGATTGG	54	683	Chapter 5
AGAP6-b	5'-TGTGTCTCTCAGCGCTTGTT 3'-GTGGGATGCCTCAAAGGGAA	56	592	Chapter 5
AGAP6-c	5'-CCAAAGCTGTGAGCAGAGG 3'-GTGGCAACCTCCATTCTGAT	55	985 Gel extract top band	Chapter 5
TP63	5'-TTTCCCTTATCTCGCCAATG 3'-CAAGCTAAGGAAGATTGATTGC	50	385	Chapter 5
DROSHA	5'-AACCCAAGTGCTTTCCTCTG 3'-AAAAGTGTCCTGGGAATTGG	52	544	Chapter 5
ANKRD2	5'-CTGTGAAAGCCTTCAGGACA 3'-TCGCAATTAGCAAAAACAGC	52	383	Chapter 5
AP3B2	5'-ACTGTCCACAGAAGCCTGGT 3'-GACGTTCTGCCTGAAAGTGTC	54	399	Chapter 5

2.13 **SANGER SEQUENCING**

Sanger sequencing was carried out to validate candidate gene variants found in the study. SourcebioScience Sanger sequencing services were used.

Primers used for sequencing are outlined in **(Table 2-12).** For DNA plasmids cloned using the TOPO® kit, the M13-F or M13-R primers that are in the TOPO® vector were used and are also outlined in **(Table 2-12)**.

Table 2-12 Primers used for sanger sequencing

Gene Name	Sequence	Project used for
CTNND1, Position g.57569629	5'-TGGACGTGACCAGGATAACA	Chapter 4
CTNND1, Position chr.57564451	5'- CGCTTTCATCCAGAGCCTTA	Chapter 4
DMXL2 Position g.51773257	5'- GGCGCTCATAGGAACCAATA	Chapter 4
DMXL2 Position g.51791234	5'- TCCTGGATTTGAAACAGCTC	Chapter 4
TENM3	5'-ATCTCGAGTGCTCCCTCCAC	Chapter 5

CDK16	5'-CACACAATCCCACCCAGTG	Chapter 5
AGAP6-a	5'-AGCGGGAAGACCATCTCTG	Chapter 5
AGAP6-b	5'-TGTGTCTCTCAGCGCTTGTT 3'-GTGGGATGCCTCAAAGGGAA	Chapter 5
AGAP6-c	5'-CCAAAGCTGTGAGCAGAGG 3'-GTGGCAACCTCCATTCTGAT	Chapter 5
TP63	5'-CACCAGTAATCTCCAGACCTCA	Chapter 5
DROSHA	5'-AACCCAAGTGCTTTCCTCTG	Chapter 5
ANKRD2	5'-TGGTGGTGTCAATGAGGGTA	Chapter 5
AP3B2	5'-ACCCTCAGGTGAGATGATGC	Chapter 5
M13-F	TGTAAAACGACGGCCAGT	Chapter 5
M13-R	CAGGAAACAGCTATGACC	Chapter 5

2.14 CLONING, TRANSFORMATION AND PLASMID PREPARATION

Table 2-13 Cloning solutions

Solution	***Composition and concentrations***
Luria-Bertani (LB) agar	1% Tryptone, 1% NaCl, 0.5% Yeast and 1.5% Agar
Luria-Bertani (LB) broth	1% Tryptone, 1% NaCl, 0.5% Yeast
S.O.C Outgrowth Medium *New England Biolab, B9020S*	1X SOC Outgrowth Medium: 2% Vegetable Peptone 0.5% Yeast Extract 10 mM NaCl 2.5 mM KCl 10 mM MgCl2 10 mM MgSO4 20 mM Glucose
Ampicillin	100µg/mL
Spectinomycin	50µg/mL

2.14.1 Purifying Plasmid DNA from Bacterial Cultures

To do this, I used the kit-free alkaline lysis plasmid miniprep protocol by addgene (https://www.addgene.org/protocols/purify-plasmid-dna/). The solutions I made for this protocol are described in **(Table 2-14).** Briefly, after discarding the supernatant, the bacterial pellet was resuspended in 250µL of Solution 1 to resuspend the bacteria, vortexed for two minutes and transferred to a 2mL Eppendorf at this stage. 400µL of Solution II was then added and incubated on ice for 5 minutes. Then, 300µL of Solution III was added to form a white precipitate that contains the bacterial proteins and genomic DN. This was incubated again on ice for 5 minutes and centrifuged for 5 minutes at 12,000g. Around 750µL of the supernatant containing the plasmid DNA was

added to a new tube; 0.5µL of RNase A (20mg/mL) was added to the supernatant and incubated for 5 minutes at 37°C. Phenol chloroform extraction was performed, followed by ethanol precipitation of DNA as per manufacturer's instructions (addgene). The DNA was resuspended in 40µL of 0.1x TE buffer. A Nanodrop 2000 (Thermofisher) spectrophotometer was used to measure DNA concentration.

Table 2-14 Kit-free alkaline lysis plasmid miniprep solutions

Solution	***Preparation***
Solution I – Resuspension Buffer	25mM Tris-HCL (pH8) 50mM Glucose 10mM EDTA Stored at 4°C
Solution II- Denaturing Solution	0.2N NaOH 1.0% SDS Stored at room temperature
Solution III- Renaturing Solution	120mL 5M Potassium Acetate 23mL Glacial Acetic Acid 57mL ddH2O Stored at 4°C

2.15 ISOFORM DESCRIPTION AND SEQUENCE ALIGNMENT

Data on isoforms and amino acid sequences was obtained from Ensembl (ensembl.org) and the National Centre for Biotechnology Information (ncbi.nlm.nig.gov). Sequence alignment was conducted using the online Multiple Sequence Comparison by Log-Expectation tool (MUSCLE).

2.16 PLASMID LINEARIZATION AND MRNA PROBE SYNTHESIS

Table 2-15 Solutions for probe synthesis

Solution	***Preparation***
Tris-EDTA (TE) Buffer	1mM TRIS-HCL, pH 7.5 0.1 mM EDTA Made in ddH_2O
Plasmid Linearization/Digestion Reaction Mix	20µg DNA (plasmid) 2µL restriction endonuclease 1µL BSA 10µL 10x buffer Made in ddH_2O up to 100µL
Probe Synthesis Reaction Mix	2µL Rnase inhibitor 5µL DIG NTPs mix 1µg/µL DNA (linearized plasmid) 10µL 5x transcription buffer 3µL RNA polymerase enzyme (T7) 5µL 0.1M DTT (when using T7) Up to 50µL in ddH_2O

Each plasmid used in this study is described in its relevant section (Section 2.22.5 and 2.23.3), however, a brief description of each plasmid and its vector is shown in (**Table 2-16)** below.

Table 2-16 Synthesised DNA plasmids

Plasmid	***Vector***	***Polymerase Enzyme***	***Source***
Human P120	pENTR223 cloning vector	T7	Human ORFeome Collaboration (Rual *et al.*, 2004) clone # HsCD00513511
Human AGAP6	pCR2.1 TOPO cloning vector	T7	Alharatani, Thesis 2019

Following sequencing of the plasmids to confirm their specificity for the DNA fragment of interest, probes were synthesized in order to apply them for downstream mRNA expression experiments. Firstly, the plasmids were linearized using specific restriction enzymes; 20µg of plasmid DNA was added into a reaction mix together with 2µL of the restriction enzyme of choice, 1µL BSA and 10µL 10x reaction buffer up to 100µL nuclease free H_2O and incubated at 37°C overnight.

To clean and precipitate the linearized DNA, the PCR purification kit and columns were used. The DNA was resuspended in 30µL H_2O. A few µL of the cut and uncut DNA (.5-4µL) was then resolved on a 1% Agarose gel with 0.05% Ethidium Bromide; the DNA was mixed with 2µL of the DNA Gel Loading Dye (6X) and 10µL H_2O.

To make the digoxigenin labelled RNA probe, I added 1µg of the linearized and purified DNA to a reaction volume of 50µL containing 2µL of RNase inhibitor (40µg/µL), 5µL Digoxigenin labelled NTPs, 10µL 5x reaction buffer, 5µL 0.1M DTT and 3µL of the T7 polymerase enzyme and the rest with H_2O. The reaction was incubated at 37°C for 2 hours. After that, 1µL of 1mg/ml DNAse1 was added to the reaction and incubated at 37°C for 20 minutes in order to digest any DNA template still present. To precipitate RNA, 26.6µL of lithium chloride and 1µL of glycogen were added and incubated at -20°C overnight. The following day, the reaction was spun for 10 minutes at 13,000rpm, the supernatant was removed, and the pellet was washed with 70% EtOH. The RNA was resuspended in 55µL nuclease free H_2O and measured, then ran on a 1% Agarose gel with 0.05% Ethidium Bromide, whereby 1µL of the probe was mixed with 2.5µL of RNA Loading Dye and 2µL H_2O. Before loading the gel, this mix was placed on a 70°C heating block for 10 minutes and then immediately on ice. To store the probe, it was mixed with

hybridization buffer to make a 5x or 10x stock (10µg/mL), and labelled probe -in-hyb, which was stored at -20°C. For downstream in situ hybridization experiments, a 1x dilution in hybridisation buffer should contain at least 1µg/mL of probe.

2.17 TISSUE PROCESSING FOR HISTOLOGICAL SECTIONS

After fixation in PFA, mouse embryos or pups used in this study were washed multiple times in 1x PBS. For postnatal stage (P1) and P2.5 mice, the pups were de-skinned, and placed to incubate in 10% EDTA/PBS pH7.4 to decalcify for 3-5 days. They were then washed multiple times in PBS followed by a series of ethanol (EtOH) washed to dehydrate the samples going from 25%, 50%, 2x 70%, for 2 hours each then in 70% EtOH overnight. The following day they were dehydrated further in 80%, 90% and 2x 100% 2 hours each, then in 100% EtOH overnight. The samples were then processed and embedded in wax as per standard protocol. The length of time used for the xylene solvent and solvent/wax incubations was 2 hours. Each sample was then oriented either coronally or sagittally and mounted in wax. The Microtome (Leica RM2145) machine was used to section the wax blocks. Wax strips were produced at a thickness of 7µm. Wax strips were mounted sequentially over 3 glass slides (SuperfrostPlus®, VWRTM) using 30% EtOH or ddH_2O. The slides were dried on a hotplate at 42°C, then overnight at 37°C. For haematoxylin and eosin (H&E) staining, slides were fixed, sectioned and stained according to standard protocols. Slides were then cover slipped with Neo-Mount.

2.18 TISSUE PROCESSING FOR MRNA *IN SITU* HYBRIDIZATION

Table 2-17 Solutions for mRNA in situ hybridization

Solution	***Preparation***
20x SSC pH 4.5 (1 L)	175.3g NaCl (3M) 88.2g Sodium citrate dihydrate (0.3M) Adjusted pH to 4.5 using citric acid Dissolved in 800mL ddH_2O up to 1L
5M NaCL (1 L)	Added 292.2g NaCl Dissolved in 800mL ddH_2O up to 1L
1M Tris-HCL pH 9.5 and pH 8 (1 L)	Added 121.1g Tris-base Adjusted pH with concentrated HCL Dissolved in 800mL ddH_2O up to 1L
Trietholamine (TEA)	To make 0.1M: Added 9.282g of TEA powder (MW=185.65) up to 500mL ddH_2O Adjusted pH to 7.5
50x Denhardt's (Rnase-free)	1% (w/v) Ficoll 400, 1% (w/v) Polyvinylpyrrolidone, 1% (w/v) BSA, 50%

	Dextran sulphate 100mg/mL in H_2O, stored at -20°C
Hybridization Buffer	For 200mL: 100mL Formamide 40mL 50% Dextran Sulphate 4mL 50x Denhardt's Rnase-free 5mL Yeast tRNA (10mg/mL) 12mL 5M NaCl 4mL 1M Tris-HCL pH8 2mL 0.5M EDTA 2mL 1M $NaPO_4$ 10mL 20% Sarcosyl 23mL DEPC-H_2O Aliquots were stored in -20°C
Acetylation solution	125µL of acetic anhydride in 50mL 0.1M TEA.
50%Formamide-50% 2x SSC	For 50mL (to soak tissue for 1-2 humid chambers): 25mL Formamide 5mL 20x SSC (pH 4.5) 20mL ddH_2O
5x standard saline citrate (SSC)	20x SSC was diluted 4 times in ddH_2O
2x standard saline citrate (SSC)	20x SSC was diluted 10 times in ddH2O
0.1x standard saline citrate (SSC)	20x SSC was diluted 200 times in ddH2O
High stringency wash	25mL Formamide and 25mL 2x SSC, left after preparation at 65°C
RNase Buffer	100mL 5M NaCl 10mL 1M Tris-HCL, pH 7.5 10mL 0.5M EDTA Made up to 1L ddH_2O
PBSTw	1mL Tween20 in 1L PBS to make a 0.1% solution
Blocking Solution	To make a 10% goat-serum blocking solution, for 50mL add: 5mL heat-inactivated goat serum 50µL Tween-20 45mL PBS
Anti-Dig AP in 1% Goat Serum	Prepare while the slides are incubating in the blocking solution and place at 4°C until use. For 50mL: 500µL heat-inactivated goat serum 10µL Anti-Dig AP (1:5000) in 49.49mL 1x PBS
Sodium (Na) chloride-Tris-magnesium-Tween20 buffer (NTMT)	2mL 5M NaCl 10mL 1M Tris-HCL, pH9.5 5mL 1M $MgCl_2$ 0.1mL Tween-20 Made up to 100mL with ddH_2O immediately prior to use and added 1mL Levamisole (of a stock made of 0.5g in 20mL H_2O)
Other Solutions for WHOLE MOUNT	
PBSTw	1mL Tween20 in 1L PBS to make a 0.1% solution
Glycine	2mg/mL in PBSTw
Fixative	0.2% Glutaraldehyde/4%PFA in PBT: 20mL 4%PFA, 20 µL Tween 20, 160µL 25% Glutaraldehyde
Proteinase K	100mg of proteinase K dissolved in 5mL of ddH_2O

5x NaCl	5mL 5M NaCl 25mL 1M Tris ddH_2O up to 100mL
10% SDS	10g SDS in 100mL ddH_2O
Solution I	20mL Formamide 8mL 20xSSC (pH 4.5) 4mL 10% SDS Up to 40mL ddH_2O
Solution II	10mL 4M NaCl 0.8mL 1M Tris pH 7.5 80µL 10% Tween 20 Up to 80mL ddH_2O
Solution III	20mL Formamide 4mL 20xSSC (pH 4.5) Up to 40mL ddH_2O
BB:BA	2 parts benzyl benzoic; 1 part benzyl alcohol
MAB pH7.5	For 1L of 1x MAB: 100mM Maleic Acid (added 11.6g) 150mM NaCl (added 8.8g) Adjusted pH to 7.5 using NaOH Dissolved in 500mL ddH_2O then up to 1L
Pre-block solution	2% BBR (or BMB) + 1x MAB. To do this, add: 5mL of the 10% BBR up to 25mL 1x MAB
TBST	35mL 4M NaCl 2.7mL 1M KCL 25mL 1M Tris HCL 1mL Tween 20 Up to 1L ddH_2O
AP Buffer	20mL 5x NaCl-Tris 5mL $MgCl_2$ 500µL Tween-20 ddH_2O up to 100mL

2.18.1 Method for Sectional mRNA *In Situ* Hybridisation

In situ hybridization of mRNA on paraffin wax embedded tissue sections was carried out as per standard protocols (Wilkinson *et al.*, 1987). Specific information on the synthesised probes used in this study are found in Section 2.22.5 for h*CTNND1* and Section 2.23.4 for h*AGAP6*. Reagents and solutions used for *in situ* on tissue sections are described in **(Tables 2-7 & 2-17)**, respectively. Sections were rehydrated in a coplin jar through a series of xylene and ethanol (EtOH) washes as follows: 3x 3 minutes washes in xylene, 2x 2 minutes washes in 100% EtOH, then 95% and 70% EtOH washes, followed by a rinse twice in water. This was followed by brief fixation in 4% PFA for 10 minutes at room temperature, then rinsed in PBS for 5 minutes. In order to permeabilise the membranes, proteinase K treatment (50µL of 20µg/mL in 50mL 1x PBS) was carried out for 8 minutes at room temperature, followed by washes in 1x PBS, another brief fixation in 4% PFA and more 1x PBS washes. Meanwhile, the acetylation solution was

prepared immediately before the samples were incubated in this solution for 10 minutes, followed by 3x five-minute washes in 1x PBS.

The probes of interest were then heated on a heat-block at 80°C in order to uncoil the probe, while the slides were dehydrated in 70% then 95% EtOH prior to the application of probes. One microgram of the probe in the hybridisation buffer is required for the incubation. Therefore, the volume required was made accordingly from the 5x or 10x stock probe-in-hyb solution described in Section 2.16. Two hundred to 300µL of the 1x probe of interest was added to the slides, which were cover-slipped with parafilm and incubated horizontally in a pre-heated humid chamber in a 65°C water bath overnight. The humid chamber (tip boxes) was prepared in advance to contain tissue soaked with 50% formamide- 50% 2x SSC.

The following day, incubation of the slides was carried out in glass coplin jars throughout. The slides were incubated briefly in pre-warmed (65°C) 5x SSC in order for the parafilm coverslips to float off the slides without damaging the tissues. The slides were then incubated in in the high stringency wash at 65°C for 30 minutes, followed by 3x-10-minute washes in RNAse buffer at 37°C. Right before the last RNAse buffer washes, 50µL of RNAse A (20mg/mL) was added to 50mL RNAse buffer and the slides were incubated in this solution at 37°C for 30 minutes in order to digest any unbound RNA; this was followed by a final wash in RNAse buffer for 15 minutes at the same temperature. Two more incubations in the pre-warmed high stringency wash were carried out at 65°C, 20 minutes each. The slides were then washed in pre-warmed (37°C) 2x SSC and 0.1x SSC at 37°C, 15 minutes each, followed by a 15-minute wash at room temperature in PBSTw. To prevent non-specific antibody binding, the sections were incubated in a blocking solution for 1-hour at room temperature. After that, the slides were incubated in a solution containing the alkaline phosphatase-coupled anti-digoxigenin (Anti-Dig AP) antibody in 1% goat serum overnight.

On the third and final day, the slides were washed with PBSTw at room temperature 4x, 15 minutes each to ensure that the unbound antibody was is removed. These were then washed in freshly prepared NTMT buffer 2x, ten minutes each at room temperature. To begin the colour reaction, the coplin jars were covered with foil to carry out the incubation in the dark, and the slides were emerged in the BM purple AP substrate. Prior to this, the BM purple substrate was spun at 4°C for 10 minutes at 4000rpm and 1% levamisole was added to each tube. The slides were checked regularly for the developing colour reaction and once the sections were stained, the reaction was stopped with PBS washes, followed by dehydration with EtOH washes. To coverslip the

slides, they were washed in xylene solvent three time for three minutes each and mounted with DPX.

2.18.2 Method for Whole Mount *in Situ* Hybridisation

The HDBR Carnegie stage 13 embryo was fixed upon shipment with PFA. The embryo was rehydrated in a series of graded methanol (MeOH) washes, beginning with 50% MeOH for 5 minutes, 30% MeOH for 5 minutes and the embryo was taken into PBSTw. The sample was incubated in 10µg/mL proteinase K for 10 minutes and washed with freshly prepared glycine for 5 minutes followed by more PBSTw washes. The sample was then re-fixed using the 0.2% Glutaraldehyde/4%PFA fixative for 20 minutes and washed further with PBSTw. For prehybridization, the sample was then incubated shaking in hybridisation buffer for 1 hour at 70°C. The sample was incubated with 1µg/mL of the probe of interest-in-hyb at 70°C overnight.

The following day, any unbound probe was removed in a series of formamide/SSC washes and maleic acid buffer (MAB) washes. Briefly, washes were done with Solution I 2x, 30 minutes at 70°C; Solution I + Solution II for 10 minutes at 70°C, and Solution II 3x, 5 minutes at room temperature. In order to eliminate RNA transcripts, the samples were then incubated in Solution II containing 100µg/mL RNAse A for 30 minutes at 37°C. This was followed by washes in Solution II for 5 minutes and Solution III 2x, 30 minutes at 65°C. Finally, more washes in TBST 3x, 5 minutes and MAB 2x, 10 minutes at room temperature were carried out before incubating the embryo in the pre-block solution for three hours at room temperature. After that, to detect bound riboprobe, Anti-Dig AP antibody was added to fresh pre-block solution at a dilution of 1:3000 and the embryo was incubated at 4°C overnight. On the third day, the sample was washed multiple times in TBST over eight hours and incubated again at 4°C overnight. The following day, the embryo was washed in TBST followed by 2x washes with AP buffer, 10 minutes at room temperature. Signal was detected using the BM purple AP substrate (with the preparation similar to that described in Section 2.18.1). After a signal appeared, the embryo was washed several times in PBSTw and fixed in 4% PFA for 10 minutes, followed by 2 further washes with PBS. The sample was stored in 50% glycerol.

2.19 TISSUE PROCESSING FOR IMMUNOFLUORESCENT STAINING

Table 2-18 Solutions for immunofluorescent staining on wax sections

Solution	*Preparation*
Heat-inactivated goat serum	Goat serum thawed at room temperature and heat-inactivated at 56°C for 1 hour, made at 1% and 10% in PBT, stored in aliquots at -20°C
PBST (0.1% Triton)	1mL Triton-100X was added to 1L of PBS prior to use to make 0.1%
4% PFA	Paraformaldehyde was dissolved in PBS with stirring and heating at 60°C. Aliquots stored at -20°C
Tris-EDTA pH9	1M TRIS-HCL, pH 9 (5mL) 0.5M EDTA, pH 8 (1mL) Made in 500mL ddH2O
15% Glycine	3g Glycine powder in 20mL H_2O
Slide blocking solution	150µL/slide: 15µL 10% heat-inactivated goat serum 7.5µL 10% Triton 15µL 15% glycine 15µL BSA (20mg/mL) 97.5µL 1x PBS
Antibody blocking solution	100µL/slide: 1µL 1% heat-inactivated goat serum 5µL 10% Triton 10µL 15% glycine 10µL BSA (20mg/mL) x µL primary antibody 1 x µL primary antibody 2 Up to 100µL in 1x PBS

Table 2-19 Solutions for immunofluorescent whole mount staining for Xenopus

Solution	*Preparation*
MEMFA Fixative	10X stock solution for MEMFA: 1M MOPS 20mM EGTA 10mM MgSO4 38% Formaldehyde
Series of methanol (MeOH) washes	Prepared: 75%, 50%, 25%
PBT	1X PBS 2mg/mL BSA 0.1% Triton X-100

Table 2-20 Primary and secondary antibodies used

Antigen	*Host*	*Supplier*	*Dilution*
pY p120-catenin, 2B12	Mouse	Biolegend 828301	1:150
CTNND1/CAS (pS268) [EPR2380]	Rabbit	Abcam ab79545	1:150
E-Cadherin [M168]	Mouse	Abcam ab76055	1:150
E-cadherin (24E10)	Rabbit	Cell Signalling 3195	1:250

Pax-2	Rabbit	ThermoFisher 71-6000	1:100
Anti-collagen, Type II, 6B3	Mouse	Merck MAB887	1:50
Secondary Antibodies	**Host**	**Product Details**	**Dilution**
Goat anti-Rabbit IgG, Alexa Fluor 488	Goat	Invitrogen, A-11008	1:400
F(ab')2-Rabbit anti-Mouse IgG Alexa Fluor 488	Rabbit	Invitrogen, A-21204	1:400
Rabbit anti-Mouse IgG Alexa Fluor 546	Rabbit	Invitrogen, A-11060	1:400
Goat anti-Rabbit IgG Alexa Fluor 568	Goat	Invitrogen, A-11011	1:400
Donkey anti-Rabbit IgG Alexa Fluor 594	Donkey	Invitrogen, A-21207	1:400
Goat anti-Mouse IgG Alexa Fluor 647	Goat	Invitrogen, A-21235	1:400

2.19.1 Method for Immunofluorescent Staining on Wax Sections

For immunostaining, mouse embryos at the indicated stages were fixed and processed according to standard protocols. Wax sections were dewaxed and rehydrated using a series of graded ethanol washes as per standard protocol and washed in PBS. Permeabilization of tissues was carried out using 0.1% triton 2x, for 10 minutes at room temperature. Antigen retrieval was carried out by placing the slides in pre-heated Tris-EDTA (pH 9) in a 90°C water-bath for 30 minutes. After the solution had cooled, the slides were washed with PBS. Slides were then incubated with the slide blocking solution placed horizontally in a humid chamber at room temperature and cover-slipped with parafilm. An hour later, the slides were incubated with the primary antibodies of choice **(Table 2-20)** made in the antibody blocking solution. The slides were covered with parafilm in the humid chamber and incubated at 4°C overnight.

The following day, the slides were washed with PBST several times shaking and incubate with the secondary antibody of choice, listed in **(Table 2-20).** All secondary antibodies were diluted to 1:400 in in the same blocking buffer used for the primary antibodies. This was carried out for one hour at room temperature in the humid chamber. After that, the slides were washed several times in PBST followed by PBS. Slides were mounted in Fluoroshield Mounting Medium with DAPI and cover slipped.

2.19.2 Method for Immunofluorescent Staining for *Xenopus*

Immunos carried out on *Xenopus* whole mount embryos and tadpoles that were fixed in MEMFA and dehydrated and stored in 100% MeOH were carried out in 1.5mL Eppendorfs and began with a series of rehydration steps in MeOH 5 minutes each at room temperature followed by 2x washed in PBT for 5 minutes. Blocking was carried out in PBT beginning with 2x 5-minute incubations, followed by an hour at room temperature. Embryos and tadpoles were incubated in the primary antibody in PBT at 4°C overnight. The following day, samples were washed multiple times in PBT at room temperature for 5x, 20-minutes each and incubated in with secondary antibody for 4 hours at room temperature or at 4°C overnight. All secondary antibodies were diluted to 1:400 in PBT. Following this incubation, tadpoles and embryos were incubated with Hoechst (1:5000 of 20mg/ml, diluted in PBT) for 10 minutes. Finally, multiple washes in PBT were carried out.

2.20 IMAGE ACQUISITION

2.20.1 Imaging for Slides and Whole Mount Embryos

Images for sectional *in situ* hybridization experiments and for haematoxylin and eosin (H&E) slides were captured using a brightfield microscope (Nikon ECLIPSE Ci-L), with an attached camera (Nikon digital sight DS-Fi1) or with a NanoZoomer 2.ORS Digital Slide Scanner (Hamamatsu). NDP.view2 Viewing Software (U12388-01) was used to analyse the scanned images. Whole mount images of mouse pups and embryos, *Xenopus* and human embryos were captured using a Nikon SMZ1500 stereomicroscope with a Nikon digital sight DS-Fi1 (112031) camera. Fluorescent images of mouse palates and *Xenopus* epithelial cells were either acquired on a Leica SP5 confocal or Nikon A1R point scanning confocal; z-stacks of whole mount *Xenopus* tadpoles were captured by mounting the tadpoles on a Cellview Cell Glass Bottom Culture Dish (PS, 35/10 mm, CELLview™, Cat. No. 627860) in PBS. Image sequences were processed using the FIJI (Image J) analysis software.

2.20.2 Micro-Computed Tomography (µCT)

For soft tissue scanning, mouse embryos were stained with a near isotonic 1% I2, 2% potassium iodine solution for 3 days and scanned to produce 6um voxel size volumes, using X-ray settings of 90kVp, 66uA and a 0.5 mm aluminium filter to attenuate harder X-rays. Camera binning was used to improve signal to noise ratios. For hard tissue

staining, perinatal mice were scanned to produce 7.4um voxel size volumes using X-ray settings of 70kVp, 114uA and a 0.5 mm aluminium filter to attenuate harder X-rays. Specimens were scanned using a Scanco μCT50 microcomputed tomographic scanner (Scanco, Brüttisellen, Switzerland). The specimens were immobilised in appropriately sized scanning tubes using cotton gauze. The above technique was carried out by Dr Christopher Healy. The author processed and analysed the soft and hard tissue specimens using Parallax Microview software package (Parallax Innovations Inc., Ilderton, ON Canada).

2.21 SPECIFIC METHODS FOR CHAPTER 3

2.21.1 Definitions and Criteria for the Recording of Tooth Anomalies in the Clinical Study

Table 2-21 Definitions and criteria for recording tooth anomalies in the Clinical Study

Anomaly	*Definition*	*Selection criteria*
Agenesis	Congenitally missing tooth/germ.	Based on child's age and assessed against 'The London Atlas of Tooth development' (Al Qahtani et al 2010).
Ectopic eruption	A tooth erupting in an abnormal location (Toutountzakis and Kastaris 1990).	A tooth not erupting along its path of eruption.
Transposition	Two teeth exchange positions (Jamal *et al.* 2010).	-
Impaction	Following clinical and radiographic assessment the tooth is not expected to erupt to its functional location (Thilander and Jakobsson, 1968).	Teeth remaining in the jaw 2 years following their mean age of eruption (Wedl *et al*. 2005; AlQahtani et al 2010).
Peg lateral	A hereditary condition that results in one or more smaller lateral incisor (Kocabalkan and Özyemisci, 2005).	The incisal mesiodistal width of the lateral incisor's crown was narrower than the cervical.
Microdontia	A hereditary condition that results in one or more smaller teeth (Kocabalkan and Özyemisci, 2005).	When a tooth is much smaller than its homolog (Backman and Wahlin 2001).
Supernumerary	An additional tooth, thought to result from a disturbance to the dental development at the initiation and proliferation stages (AAPD 2015; Regezi*, et al.* 2016).	Based on child's age and assessed against 'The London Atlas of Tooth development' (Al Qahtani et al 2010).
Rotation	The rotation of a tooth along its long axis.	A tooth that is not in line with its normal long axis is recorded as rotated (Tortora *et al.* 2008).
Retained primary teeth	A tooth that is retained beyond the time of its exfoliation (Robinson and Chan 2009).	The permanent tooth has not erupted more than a year later than expected or has not erupted within 6 months of the eruption of the contralateral tooth and the deciduous tooth is retained (Robinson and Chan 2009).
Taurodontism	A tooth with an apically displaced pulp chamber and	Based on the radiographic appearance of mandibular first

	furcation area, shortened roots and absence of the usual constriction of the pulp at the cemento-enamel junction (Darwazeh *et al*. 1998; Jamal *et al.* 2010).	permanent molars and measured as follows: crown –from the deepest point in the occlusal surface to the cement-enamel junction, body – from cement-enamel junction to the furcation of the roots, root- from the furcation area to the root apices. Tooth with crown body-root ratio (CB:R) equal or greater than 1:1.10 was considered as taurodont. (Kim and Lai 1989).
Pulp stone	A calcified collection in the pulp tissue of a tooth that is healthy, diseased or unerupted (Hamasha and Darwazeh, 1998)	A radiopaque mass within the pulp chambers of the premolars and molars (Ranjitkar *et al.* 2002).
Dilaceration	An angulation or a curve anywhere between the crown and the root of a developed tooth (Hamasha *et al*. 2002).	It is when the root is bent mesially, distally, bucally or lingually. Buccal or lingual bends are detected by examining the appearance of the apical portion of the root. When a round opaque area with a dark shadow in its central part cast by the apical foramen and root canal presented a "bull's-eye'' appearance, it was recorded as a dilaceration (White and Pharoah, 2014).
Short/blunt roots	Developmentally very short roots with blunted root tip (Apajalahti *et al.* 2002; Puranik *et al.* 2015), roots are the same length or shorter than the incisor crown.	The root, crown ratio of the teeth on the participant's right side measured and compared to the teeth on left side, if the dimensions are the same, the right side will be measured only; if not, both sides were measured. Lind's method was used to determine the root, crown ratio on radiographs. Two points were drawn on the tooth. One point was placed on the mesial aspect of the tooth at the intersection of the crown and root (y). Another point was placed on distal aspect (x). A horizontal line was drawn connecting the two points and the midpoint of this line was noted as (M). Apex of the root was marked as (A), in teeth with multiple roots the point was selected on apex of the longest buccal root. The incisal edge or cusp tip midpoint was noted as (I), for teeth with several cusps, a line was drawn to connect the buccal cusps and a midpoint was defined. Lines were drawn connecting A-M and M-I. Crown height was determined as follows: IM length. Root length was determined as follows: AM length (Lind 1972). Measurements were rounded to the

		nearest ½ or whole mm (Hölttä *et al*. 2004). Normal R:C ratio was based a Finnish population study, since it was the closest population ethnically to our sample of patients.
Dens evaginatus	A supernumerary tubercle (talon cusp) protruding from the occlusal aspect of posterior teeth, and lingual aspect of anterior teeth, extending from cement-enamel junction to at least half the distance to the incisal edge (Meon 1991; McCulloch *et al.* 1997, Levitan and Himel, 2006).	Clinical inspection or a well-defined triangular radio-opacity of enamel on the occlusal surface of posterior teeth or lingual surface of anterior teeth.
Dens invaginatus	An invagination, before calcification, of the enamel and dentine beginning from foramen coecum or the cusp tip and could extend into the root (Hülsmann 1997; Hamasha *et al.* 2002).	Seen radiographically as an infolding of enamel and dentine.
Enamel defects	**Hypoplasia:** an inherited or acquired condition where the dental enamel has either a surface loss or a break in the continuity (Lai and Seow, 1989). **Hypomineralisation:** an inherited or acquired condition with a defect in the mineralization of the matrix formed by the ameloblasts resulting in reduced mineralisation in a localised or generalised manner. The tooth will have white/creamy/yellowish well-defined opacities within the enamel (Mast *et al.* 2013).	A tooth with any part of the crown having ill-defined radiolucency was recorded as a tooth with hypoplasia/hypomineralisation (Jamal *et al.* 2010).
MIH	'Demarcated, qualitative defects of enamel of systemic origin affecting one or more first permanent molars with or without incisor involvement' (Weerheijm, 2001).	Clinically demarcated opacities of different colour, that is opacities with a clear and distinct border to the adjacent enamel (Weerheijm, 2003). Hypomineralised second primary molars were also recorded (Elfrink *et al.,* 2015). Varied severities were recorded.

2.22 SPECIFIC METHODS FOR CHAPTER 4

2.22.1 Recruitment & Consent

In addition to the Clinical Study and DDD participants, other subjects presented in **Chapter 4** were recruited from the University of Calgary, Alberta Children's Hospital, Canada; or from the Children's Hospital of Philadelphia, USA. All individual study protocols were approved by local Institutional Review Boards. Medical and dental histories were taken, as well as detailed phenotyping and assessment by clinical geneticists with expertise in dysmorphology. All patients also underwent high-resolution analysis for copy number abnormalities using array-based comparative genomic hybridization.

2.22.2 Mouse Models

Mice were genotyped according to standard procedures. Gestational ages for mice were determined by the observation of vaginal plugs, which was considered embryonic day 0.5 (E0.5) and further staging of animals according to Kaufman (Kaufman & Kaufman, 1992). Mice strains used are summarized in (**Table 2-22)**. For each mouse experiment, a minimum of n=3 was examined unless otherwise noted.

Table 2-22 Mouse strains used in the study

Mouse Line	*Source*	*Reference*
Ctnnd1$^{fl/fl}$	MGI ID, 3640772	(Elia *et al.*, 2006)
β-actin::cre	JAX strain, 019099	(Lewandoski *et al.*, 1997)
Wnt1::cre	JAX strain, 022501	(Lewis *et al.*, 2013)

2.22.3 *Xenopus Tropicalis*

X. tropicalis embryos were produced by *in vitro* fertilization and raised to appropriate stages in 1/9MR + gentamycin as per standard protocols (Khokha *et al.*, 2002). For *Xenopus* experiments, experimental numbers are stated in figures, with a minimum of n=30 in all experimental conditions.

2.22.3.1 CRISPR/Cas9 Knockouts in Xenopus Tropicalis

Two non-overlapping sgRNAs were designed to target *Xenopus tropicalis ctnnd1*:

sgRNA1: CTAGCtaatacgactcactataGGAACGGGTGTGGGAGCCATgttttagagctagaa;

sgRNA2: CTAGCtaatacgactcactataGGGGTGGTATCCCACGCAAGgttttagagctagaa.

sgRNA1 targets exon 3 and is thus predicted to disrupt isoform 1 only, while sgRNA2 targets exon 7 and is thus predicted to disrupt all four isoforms. Embryos were injected at the one or two cell stage and raised until indicated stages. CRISPR injections were carried out by Dr Emily Mis. For CRISPR/Cas9 experiments. The rest of the frog experiments in this chapter were carried out by the author. Statistical significance was defined as $P<0.05$ and analysed by chi-squared test or Fishers exact test. Data processing was carried out in Microsoft Excel and Graphpad Prism 7.

2.22.4 **Whole Exome Sequencing Computational Workflow and Variant Screening**

Whole exome sequencing (WES) from trios was performed to identify gene variants. In this chapter, two different WES algorithms were applied depending on the recruitment site of the participant. For patients recruited from DDD (Study *et al.*, 2017), genomic DNA samples from trios were analysed at the Wellcome Trust Sanger Institute. WES was performed using a custom Agilent SureSelect Exome bait design (Agilent Human All Exon V3 Plus with custom ELID # C0338371), 8-plex sample multiplexing and an Illumina HiSeq with 4 samples per lane and a mean depth of 50X. The exome analysis targeted 58.62 Mb of which 51.64 Mb consisted of exonic targets (39 Mb) and their flanking regions and 6.9 Mb consisted of regulatory regions. Alignment was performed using BWA1. Putative *de novo* variants were identified from trio BAM files using DeNovoGear5. Variants were annotated with the most severe consequence predicted by Ensembl Variant Effect Predictor (VEP version 2.6), and minor allele frequencies from a combination of the 1000 Genomes project (www.1000genomes.org), UK10K (www.uk10k.org), the NHLBI Exome Sequencing Project (esp.gs.washington.edu), Scottish Family Health Study (www.generationscotland.org), UK Blood Service and unaffected DDD parents. All flagged variants were automatically annotated with pathogenicity scores from two variant prioritisation algorithms (SIFT and PolyPhen) and compared against the public Human Gene Mutation Database (HGMD) and the Leiden Open Variation Database (LOVD). This sequencing was carried out by the DDD consortium.

For some probands, WES performed at the Yale Centre for Genomic Analysis used genomic DNA isolated from saliva from the probands and their parents. The exons and their flanking regions of the genome were captured using IDT xGen exome capture

kit followed by Illumina DNA sequencing (HiSeq 4000). Paired end sequence reads were converted to FASTQ format and were aligned to the reference human genome (hg19). GATK best practices were applied to identify genetic variants, and variants were annotated by ANNOVAR. Probands and parents were sequenced to a mean depth of 93-123 independent reads per targeted base across all the samples. In an average of 94.0% of targeted bases in all of the samples, the coverage was greater than 20X independent reads. Trio WES analysis on variants with allele frequency of less than 1% was carried out to identify *de novo* variants that are absent from the parents. Putative disease-causing variants were validated in-house using whole genome amplified DNA, PCR and capillary sequencing. The sequencing above was carried out by Dr Weizhen Ji in the Lakhani Lab, Yale School of Medicine.

2.22.5 Generation and Description of the *CTNND1* Probe and mRNA *In Situ* Hybridization

A human *CTNND1* clone was identified from the Human ORFeome Collaboration (Rual *et al.*, 2004) (clone HsCD00513511), encoding *CTNND1* isoform 4, including the entirety of the armadillo repeats and the C-terminal domain. Probes made from this clone should recognize all four *CTNND1* transcripts. Digoxigenin-labeled antisense mRNA probes were produced by linearizing human *CTNND1* clones using BamH1 restriction enzyme, which produces a probe size of 900 base pairs, and *in vitro* transcription with the T7 High Yield RNA Synthesis Kit (E2040S) from New England Biolabs. *In situ* hybridization of mRNA on whole mount and paraffin embedded tissue sections was carried out as per standard protocols (Wilkinson *et al.*, 1987) using an anti-digoxigenin-alkaline phosphatase coupled antibody (described in Section 2.18.1 & 2.18.2).

2.22.6 Site Directed Mutagenesis (For Future Work)

Site directed mutagenesis was carried out on the P120-pENTR223 construct to mutagenize the nucleotide in some of the human variants found in this study (described in **Table 2-23).** Primers were designed as per manufacturer's instructions and the DPN1 kit was used to carry out the reaction as per manufacturer's instructions.

Table 2-23 Variants included for site-directed mutagenesis

Variant	*Primers*	*GC Content (%)*	*Melting Tm (°C)*
*g.57569629, C>T; p.Arg461**	Forward with mutation **CCTGCCCTTGTGCGATTGCTTTGAA AGGCTCGTGATATGGACC** Reverse complement with mutation **GGTCCATATCACGAGCCTTTCAAAG CAATCGCACAAGGGCAGG**	53	71
*g.57571265, TG>T; p. Gly532Alafs**	Forward with mutation **CGGTGCTCACCAACACAGCTGCTGC CTTAGGTAACAGTAG** Reverse complement with mutation **CTACTGTTACCTAAGGCAGCAGCTG TGTTGGTGAGCACCG**	55	71

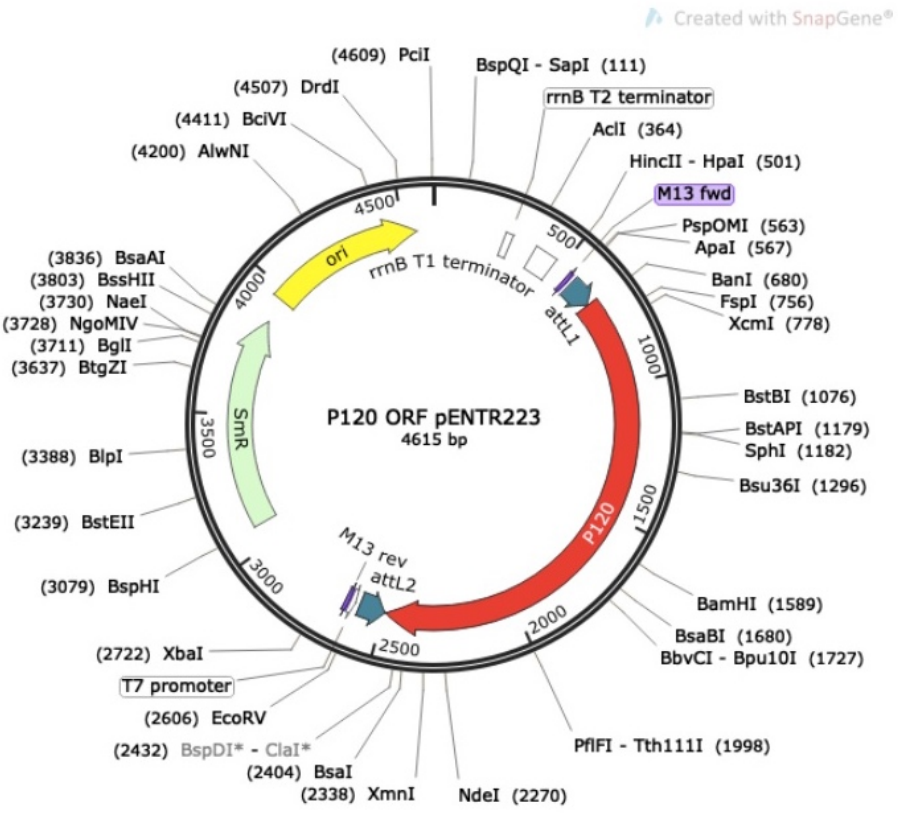


Figure 2-4 Map of the P120 human ORF (clone HsCD00513511) in the pENTR223 Vector

Schematic representation of a vector map generated using SnapGene® containing the human p120 ORF (clone HsCD00513511) represented in red, inserted into the pENTR223 vector. Shown are restriction and promoter sites.

2.23 SPECIFIC METHODS FOR CHAPTER 5

2.23.1 Whole Exome Sequencing

De novo copy number variants (CNVs) were identified using the XHMM algorithm (XHMM (eXome-Hidden Markov Model) (Fromer *et al.*, 2012). The aligned reads from WES data of the family and 320 unrelated European controls were imported into XHMM (eXome-Hidden Markov Model) (Fromer *et al.*, 2012). Potential CNVs were inspected visually and prioritized based on genomic length, GC content of targets, and low sequence complexity followed by Quantitative PCR validation.

2.23.2 *AGAP6* Copy Number Variant Quantitative Analysis qPCR

Quantitative PCR (qPCR) was used to validate the *AGAP6* CNV in the proband (described in Chapter 5) using a method described by (Ma & Chung, 2014). Specifically, isolated and quantified DNA was diluted to a final concentration of 20ng/μL using Rnase free H_2O. PCR primers were diluted to a final concentration of 10μM each (forward and reverse). A total reaction volume of 20μl was used and contained 20ng of genomic DNA, 0.2μl of forward and reverse primers (10μM each), 5μl of SensiMix™ SYBR® Hi-ROX Kit (BIOLINE, QT605-05) with SYBR Green. Note, a hybridization probe was not used in the reaction. Instead, standard curves and a duplicate series of tenfold serial dilutions (from 10x to 0.01x) beginning with a mixture of all the DNA samples, forming the most concentrated standard (100x), were used. Relative gene copy number was estimated using DNA fragments amplified from two different regions of the duplication site. Two primer pairs were used, one primer pair (Primer Pair 1, PP1) was designed to flank the beginning of the duplication site and the other primer pair (Primer Pair 2, PP2) was designed to flank a region within the duplication site. The *CTNND1* gene was used as the control. Primers used to amplify these portions were as follows:

Table 2-24 Primers for the AGAP6 CNV quantitative analysis qPCR

Gene	*Primer*	*Comments*
AGAP6	AGAP6-PP1F: 5'-AGCGGAAGACCATCTCTG-3' AGAP6-PP1R: 5'-TCCCTAGCTCCTGCCTCATA-3'	Location 1 Beginning of duplication site
AGAP6	AGAP6-PP2F: 5'-CCTGCTGAGGTGACTGTTGA-3' AGAP6-PP2R: 5'-AGAGCCAGCTTTTGTTCCTG-3'	Location 2 Middle of duplication site

CTNND1	CTNND1-F: 5'-CGGGCACCTAGTAGACAGGA-3' CTNND1-R:5'-AGCATGGGTGAAAGAGCAAG-3'	Control

The following thermocycling conditions were used

Step	Temperature	Duration	Cycles
Enzyme activation	95	3 minutes	1
Denaturation	95	15 seconds	40
Annealing/Extension/Data acquisition	58	30 seconds	
	72	30 seconds	
Dissociation	95	10 seconds	1

2.23.3 Designing and Cloning of the Human *AGAP6* Plasmid

The *AGAP6* human plasmid was generated by cloning a human cDNA (IPS human CTR-M205 control male 2/clone 5) in a TOPO® vector using the TOPO® TA Cloning Kit as per manufacturer's instructions.

The following primer pair was used:

F- 5'-GTGTCTCTCAGCGCTTGTTG-3'

R- 5'-CGCACAGCAGATACATGGTT-3'

The forward primer is located within the 5'UTR (untranslated region) of the *AGAP6* gene, and the reverse primer spans exons 4 and 5. The amplicon size is 697 bp. Human *GAPDH* was used as the control gene.

Briefly, the PCR reaction was as follows: for an (n=1, 14.65µl): H20 6.2µl, 5X Buffer 2.5µl, MgCl2 0.75µl, dNTPs 0.1µl, primers 0.75µl (diluted 10-fold and mixed: 10µm forward and 10µm reverse primers), Taq polymerase 0.1µl (Taq polymerase GoTaq® G2Flexi DNA polymerase) and 5µl of human cDNA. The following thermocycling conditions were set-up: 95°C 2 minutes; 95°C 30 seconds; 60°C 30 seconds; 72°C 1 minute; for 35 cycles and 72°C 5 minutes. The PCR reaction with these primers produces two bands, one at the expected size and the other at 900bp. Both bands were extracted, melted and purified using the QIAquick Gel Extraction Kit. Both bands were Sanger sequenced and the lower band confirmed sequence alignment to *AGAP6.* After producing and gel-extracting the PCR product, a total volume of 6µl of TOPO® cloning reaction was carried out by mixing 4µl of the gel-extracted PCR product with 1µl of salt and 1µl of the TOPO® vector that contains a T7 anti-sense promoter that

could be used for downstream synthesis of the RNA probe. The reaction was carried out for 30 minutes-1 hour at room temperature.

One Shot® TOPO 10 competent E. coli cells were used for the chemical transformation as per manufacturer's instructions. Briefly, 2µl of the TOPO® cloning reaction was added to 25µl of the One Shot® cells and incubated on ice for 30 minutes. Cells were then heat-shocked for 30 seconds at 42°C and immediately transferred again to ice. Room temperature S.O.C media was added (250µl). The cells were then horizontally incubated at 37°C shaking for 1 hour, 50µl of the transformation were spread on a couple of pre-warmed LB agar plates with 100µg/mL of ampicillin and incubated in a 37°C oven overnight (with the lid facing downwards). All the procedures above were carried out under using aseptic techniques. Several colonies were then picked, inoculated into 5ml of LB media with ampicillin and cultured shaking at 37°C overnight. The following day, pellets were spun down at 4,000g for 5 minutes and used for downstream plasmid synthesis (described in Section 2.14).

2.23.4 *AGAP6* mRNA Probe Synthesis

A digoxigenin-labelled antisense mRNA probe was produced by linearizing the human plasmid (synthesized above) using the KpnI restriction enzyme. *In vitro* transcription was carried out with T7 High Yield RNA Synthesis Kit (E2040S) from New England Biolabs.

2.24 SPECIFIC METHODS FOR CHAPTER 6

2.24.1 Definitions of *In-Silico* Annotation Tools for Variant Prioritisation

This chapter involves the use of an exome sequencing dataset to develop a pipeline for filtering through patient phenotypic data in order to find novel candidate genes. The development of this pipeline and the specific methods for this chapter are described in Chapter 6**.** When gene variants were found, they were assessed based on a number of *in silico* prediction scores. The definitions for these scores and their cut-offs are described in **(Table 2-25)** below.

Table 2-25 Definitions and cut offs for in silico annotation tools

Annotation Tool	*Definition*
Loss Intolerance pLI score:	Computed by the ExAC consortium. Indicates the probability that a gene is intolerant to a loss of function mutation. Genes that are very likely intolerant of loss-of-function **score > 0.9**. The pLI score is the probability that a given gene falls into the haploinsufficient category, therefore is extremely intolerant of loss-of-function variation. Genes with high pLI scores **(pLI ≥ 0.9)** are extremely LoF intolerant, whereby genes with low pLI scores **(pLI ≤ 0.1)** are LoF tolerant.
Residual Variation Intolerance Score RVIS	Taken from the Genic Intolerance Website. Is a gene-based score intended to help in the interpretation of human sequence data. The score is designed to rank genes in terms of whether they have more or less common functional genetic variation relative to genome wide expectation. A gene with a positive score has more common functional variation. A gene with a negative score has less and referred to as 'intolerant'. For example, a gene called *ATP1A3* has an RVIS score of 1.53 and a percentile of 3.37% meaning it is amongst the 3.37% most intolerant human genes.
Polyphen-2 PPH	A tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein. 0.0-0.15 is B=benign, 0.15-1.0 is P= possibly damaging, 0.85-1 is D= confidently damaging.
Sorting Intolerant From Tolerant SIFT	Is a popular web-based tool that uses a sequence homology from multiple sequence alignments (MSAs) to predict if amino acid substitutions would be tolerated or damaging. Scores = 'Tolerated' (T) normalised probability >0.05; i.e. amino acid substitution is predicted to be tolerant; 'Damaging' (D) normalised probability <= 0.05; i.e. mutations are predicted to be deleterious. Note: Polyphen and SIFT scores use the same range 0.0-1 BUT with opposite meanings and are only used for missense mutations.
Combined Annotation Dependent Depletion CADD	Determines pathogenicity. A CADD score that is <20 is less likely pathogenic.
Human Gene Damage Index. GDI	The GDI is the accumulated mutational damage of each human gene in healthy human population, based on the 1000 genomes project. 'Highly Damaged' human genes are likely to be disease causing.

Chapter 3 CLINICAL STUDY OF CHILDREN WITH ISOLATED CLEFTS

3.1 **SUMMARY**

Cleft lip and/or palate (CLP) is the most common congenital malformation and affects 1/700 births in the United Kingdom. It occurs in isolation, or in combination with other dental and medical conditions. Our understanding of the genetics of cleft lip and palate is rapidly expanding; however, many questions particularly those relating to sporadic forms of disease, remain unanswered. Twin and family studies imply genetic causes, but inheritance patterns are not always observed, suggesting that some cases are caused by *de novo* genetic changes, arising in subjects with no associated family histories. One complication in assigning *de novo* causation is that associated phenotypes, such as familial tooth anomalies, are often overlooked. The **aim** of this chapter was to assess parental and child tooth anomalies and define conditions in families with clefts to find those with 'cleft-only', 'cleft-tooth anomaly' and 'cleft-medical condition +/- tooth anomaly'. A further aim was to categorise subjects into those with potential inherited or *de novo* genetic risk. **Method:** To do this, I carried out a clinical survey using a family-trio design examining the dentition of affected children and their parents (or sibling/additional family member); radiographic and medical findings were recorded. **Results:** In this chapter, the 'cleft-only' and 'cleft-tooth anomaly' groups are presented. In the children, I found an overrepresentation of tooth anomalies outside the cleft region for hypodontia, transpositions, taurodontism, dilacerations and dens invaginatus. In the parents, I found that they also had an overrepresentation of tooth anomalies, including hypodontia. Ten parents from nine trios from the combined 'cleft-only' and 'cleft-tooth anomaly' groups had congenitally missing teeth, all outwith their child's cleft region. Most notably, sixteen children had a cleft without a dental anomaly or an associated medical condition but only six of them had dentally unaffected parents and no family history of cleft or dental anomalies. **Conclusion:** The findings in this study suggest that the diagnosis of 'isolated' clefts needs to be more precise to focus on the absence of any other associated medical or dental conditions in order to refine genetic risk models for 'isolated' cleft lip and palate. Affected children with 'isolated' clefts with families who do not have any other dental or medical anomaly may have *de novo* gene changes that should be explored further.

3.2 **INTRODUCTION**

Clefts of the lip and/or palate (CLP) are among the most common craniofacial birth defects ranging from 1:500 to 1:2000 live births worldwide (Mossey & Modell, 2012). CLP is a complex disorder with phenotypic heterogeneity and with many susceptibility loci across the genome. Less than a third of the candidate genes for CLP have been identified although this is increasing with advances in gene sequencing. However, the multi-genetic nature of CLP, the different inheritance patterns and penetrance, environmental influences, and under-diagnosed microforms of CLP make *de novo* gene discovery challenging (Dixon *et al.*, 2011; Mossey *et al.*, 2009). As described in Chapter 1, cleft conditions can be divided into isolated vs. associated or non-syndromic vs. syndromic clefts. Approximately 70% of CLP cases and 50% of cleft palate only (CPO) cases are non-syndromic (Dixon *et al.*, 2011; Jugessur *et al.*, 2009a). However, new classifications based on biologically relevant groupings are emerging because studies have been confounded by incomplete penetrance and failure to diagnose microforms (Carroll & Mossey, 2012; Howe *et al.*, 2015; Leslie & Marazita, 2013; McBride *et al.*, 2013; Veltman & Brunner, 2012; Vieira *et al.*, 2008). The sibling of a child with a cleft has a 50 times higher risk of also having one (Dixon *et al.*, 2011). When this is compared with the tenfold risk of siblings of a child diagnosed with ASD (Ronemus *et al.*, 2014) and the low risk (~ 2.7%) for siblings of children born with a congenital cardiac condition, the genetic profile suggests that there are causative mutations in some cleft families (Sifrim *et al.*, 2016), but, comparatively fewer novel genes have been found in isolated cleft cases, particularly cleft palate-only cases, leading to a multi-factorial aetiology label.

Dental anomalies, especially hypodontia, are commonly associated with cleft. The commonest missing tooth is the lateral incisor and a third of patients with CLP have teeth missing from outside the cleft area, suggesting a genetic aetiology rather than an anatomic or surgical cause. It is well known that relatives of cleft children have more dental anomalies than the normal population (Aspinall *et al.*, 2014; Eerens *et al.*, 2001; Mossey *et al.*, 2010; Slayton *et al.*, 2003; van den Boogaard *et al.*, 2000; Weinberg *et al.*, 2006). Single gene changes that have been found so far have usually had a syndromic presentation such as in *MSX1, PAX9*, the *IRF6* gene (associated with van der Woude syndrome), *ANKS6*, *ERBB2* and *FGFR1* in Type 1 Pfeiffer syndrome and in Kallmann syndrome Type 2 (Anderson & Moss, 1996; Dode *et al.*, 2003; Letra *et al.*, 2009; Muenke *et al.*, 1994b; Peters *et al.*, 1998; Satokata & Maas, 1994; Slayton *et al.*, 2003). Somewhat

surprisingly, very few genes have been identified that cause clefting alone and very few have considered that there may be *de novo* gene changes, despite the many genetic steps involved in the adhesion and fusion of the palatal shelves. The advances in gene research and methodology will enable researchers to have a renewed focus on this group.

The field of genetic studies in non-syndromic CLP has focused in the past few decades on inherited variation. This is reflected by the number of family-based linkage and candidate gene studies in multiplex families (thereby assuming a Mendelian mode of inheritance) (Veltman & Brunner, 2012), and by the shift towards exploring microform traits in parents to explore heritability. However, evidence regarding these perhaps incomplete penetrant phenotypes is mixed, with some finding no differences between the general population and seemingly unaffected parents and others concluding significant differences (Aspinall *et al.*, 2014; Howe *et al.*, 2015).

Medical conditions such as cardiac defects, neurological and skeletal conditions have also been commonly associated with CLP, but few studies have examined the parental medical history, and even fewer have reported both dental and medical histories. Therefore, microforms of CLP, especially those with incomplete penetrance may have been overlooked (Calzolari *et al.*, 2007; Milerad *et al.*, 1997; Rittler *et al.*, 2011; Seto-Salvia & Stanier, 2014; Stoll *et al.*, 2000)

The overall aim of this chapter was to assess parental tooth anomalies in families of children with non-syndromic cleft conditions in order to categorise cleft subjects into those with a potentially inherited genetic aetiology and those potentially arising from *de novo* gene variations. The study was based on the hypothesis that inherited variants might manifest in parents as associated tooth anomalies and that those without a family history of oral or tooth anomalies would be good candidates for the identification of *de novo* cleft-only genes. The assessment of children with syndromic manifestations, and the detailed phenotyping of their families, is reported in Chapter 6. The purpose of the current chapter is to identify cleft children who had no other anomaly and who might therefore hold the key to finding novel putative cleft genes.

3.3 **RESULTS**

3.3.1 **Cohort Description**

I recruited ninety subjects (probands) with a diagnosis of an orofacial cleft. Of these subjects, 51 were female and 39 males, with a mean age of ten years (range 2-23 years, standard deviation, sd=4.6 years). One hundred and fifty-nine of their parents/first- or second-degree cousins gave written consent to take part, of which 130 were the parents, and 127 of these relatives underwent a dental examination. Family members who took part and their ethnicities are shown in **(Figure 3-1)**. The number of children per cleft type were as follows: cleft lip only (n=14), cleft lip and palate (n=37), cleft lip and alveolus (n=13) and cleft palate only (n=26). The laterality of each cleft type and the number of children per group is summarised in **(Table 3-1)**. The number of children who presented with non-syndromic and 'associated' clefts and their families are shown in **(Figure 3-2).** Children with non-syndromic clefts are represented in this chapter.

Table 3-1 Number of children per cleft type

Cleft Type	**Cleft Side/Sub-type**	**Number of Children (N=90)**
Cleft lip only		**14**
	right	7
	left	7
	bilateral	0
Cleft lip & palate		**37**
	right	9
	left	16
	bilateral	12
Cleft lip & alveolus		**13**
	right	3
	left	8
	bilateral	2
Cleft palate only		**26**
	hard palate	22
	submucous and VPI	4

Abbreviations: VPI, velopharyngeal insufficiency.

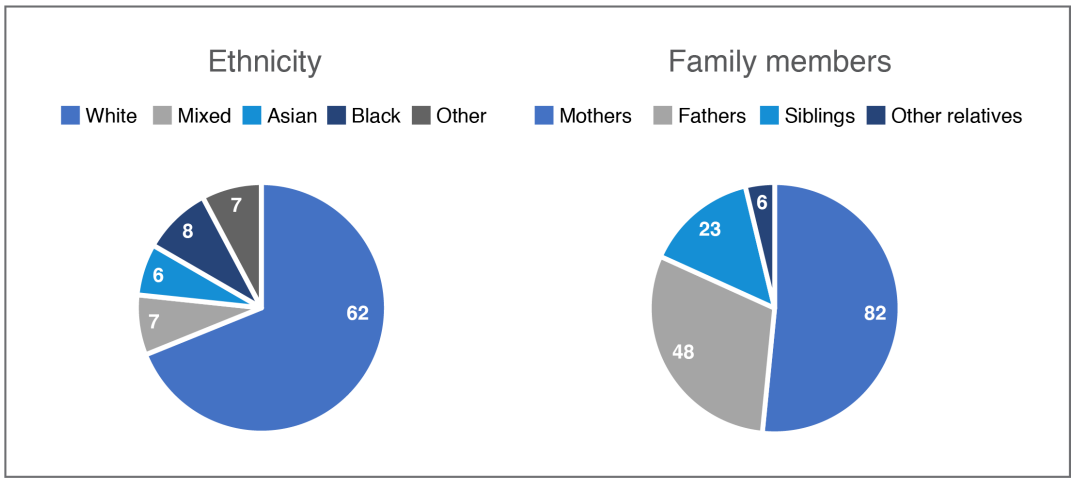


Figure 3-1 Clinical survey demographics

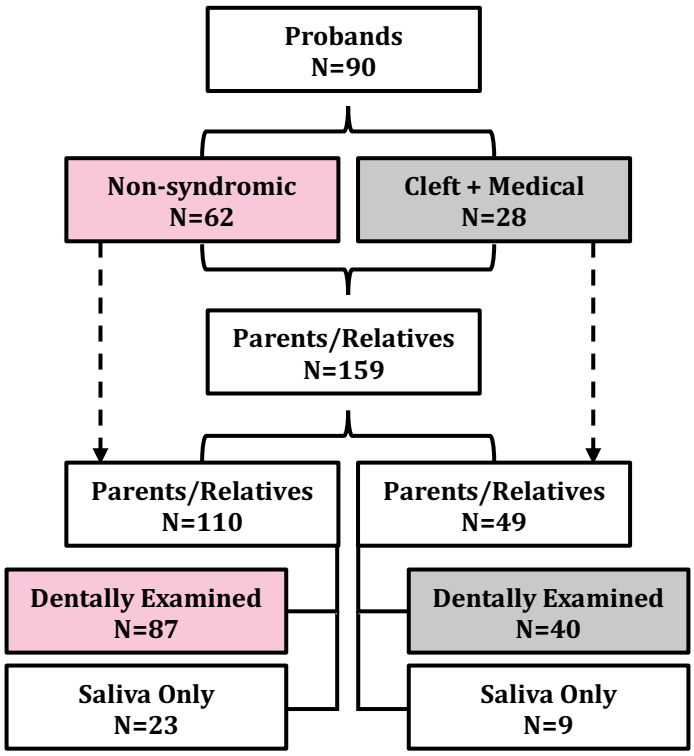


Figure 3-2 Flowchart representing the number of children and their families

The non-syndromic group of probands (N=62) and their families (N=87) are represented in this chapter (pink boxes). The ones in the grey boxes are represented in Chapter 6.

3.3.2 The Dental Profile of Children with Non-Syndromic Clefts

The probands were grouped into 'cleft-only' (medically healthy/no dental anomalies), 'cleft- tooth anomaly' and 'cleft-medical condition +/- tooth anomaly'. Mild dysmorphic features, speech and language or hearing phenotypes were included within the isolated grouping since they were thought, at the time, to be common manifestations in cleft cohorts and might be due to anatomical deformities or surgical consequences. The number of children in each category and the number of parents/relatives with dental anomalies per group are summarised in **(Table 3-2).**

Details of the children's dental anomalies are shown in **(Table 3-3).** Based on the clinical and radiographic assessments, anomalies were recorded based on their location against the cleft site. Anomalies observed outside the cleft region were hypodontia and taurodontism. Notably, 42% (26/62) of the children had hypodontia, in which 65.4% of them had congenitally missing teeth outwith their cleft site. Transpositions, pulp stones and dens invaginatus were only seen in children with non-syndromic clefts (when compared to those with 'cleft-medical condition' group). Anomalies such as peg lateral permanent incisors, impactions, supernumeraries, and tooth rotations were predominately observed within the cleft site.

Sixteen children had a cleft without a dental anomaly or medical condition but only seven of these children also had an unaffected parent/sibling, full details are shown in **(Table 3-4)**. When considering whether any familial clefts were present **(Table 3-5)**, 1/7 of the probands was excluded from this group due to history of familial clefting. The other 9/16 had a parent or sibling with a dental anomaly in the form of hypodontia, ectopic canines or transpositions.

Taking a closer look at the specific cleft profile of the six remaining probands, I found that 4 presented with a cleft palate only (one of which was a submucous cleft palate). The data above suggests that according to the criteria set out in the study, I identified six children and families who may have a *de novo* cleft defect from sixty-two child probands who had already been given a clinical diagnosis of an 'isolated' cleft.

Table 3-2 Full sample of children with clefts and their families categorised by dental anomalies & medical conditions

Cleft phenotype and definition	**Total number of children**	**Number of *children* with a dental anomaly**	**Number of *family* members with a dental anomaly**
Cleft – only: cleft lip and/or palate only without a tooth anomaly and without a medical condition	**16**	NA	**9/16** (56.3%) **(details in Table 3-6 and breakdown in Table 3-4)**
Cleft – tooth: cleft lip and/or palate only but with an associated tooth anomaly and otherwise fit and healthy	**46**	**46** **(Table 3-3)**	**22/46** (47.8%) **(details in Table 3-6)**
Cleft – medical condition +/- tooth cleft lip and/or palate with a systemic congenital medical anomaly and/or craniofacial dysmorphology	**28**	**25**	**15/28** (53.6%)

Children were grouped into three categories based on the presence or absence of a dental anomaly. Data for cells in pink are shown in their respective tables. The group in the grey cells are discussed in Chapter 6.

Table 3-3 Dental anomalies found in the children with non-syndromic clefts

Type of anomaly	Number of children with the dental anomaly (n=62)	Number of affected teeth	Number and location of affected teeth				Number of children with the anomaly outside the cleft site and its location		
			Inside the cleft site	Outside the cleft site			Maxilla	Mandible	Total
				Maxilla	Mandible	Total	Both		
Hypodontia	26	37	13	17	7	24	11	3	17
							3		
Ectopic eruption	13	17	8	7	2	9	5	-	7
							2		
Transposition	6	8	0	5	3	8	4	2	6
Impaction	14	15	12	2	1	3	1	-	2
							1		
Peg lateral	24	27	23	4	-	4	3	-	3
Microdontia	4	5	2	3	-	3	3	-	3
Supernumerary	19	30	29	1	-	1	1	-	1
Rotation	25	34	29	5	-	5	4	-	4
Retained primary	3	6	5	1	-	1	3	-	3
Taurodontism	3	16	-	8	8	16	-	0	3
							3		
Pulp stone	3	5	-	5	-	5	3	-	3
Dilaceration	11	17	5	8	4	12	3	2	7
							2		
Short/blunt roots	2	Generalized	NA	NA	NA	NA	-	-	1
							1		
Dens evaginatus	0	0	0	-	-	-	-	-	-
Dens invaginatus	6	17	2	15	-	15	6	-	6
MIH	10	77							
Other anomalies	7	20	1	4	15	19	-	5	6
							1		

Abbreviations: NA, not applicable. Note: boxes in dark purple are dental anomalies most commonly observed outside the affected child's cleft region.

Table 3-4 Details of probands with 'cleft-only' with no dental anomalies and their parent's/sibling's dental phenotypes

Proband cleft type N=16	Proband ID	Proband sex	Tooth anomaly present in parents/siblings (Y/N)	Comments on parent/siblings' tooth anomalies
CPO	CLP-8	M	No	-
CPO	CLP-13	F	Yes	Father: ectopic UR3 and mandibular exostoses.
CPO (submucous cleft and bifid uvula)	CLP-24	F	No	-
CPO	CLP-30	F	No	-
CPO	CLP-50	F	Yes	Mother: missing maxillary lateral incisors (UR2, UL2). Father: missing UL2; transposed UR2 & UR3; retained maxillary primary canines (URC&ULC); spacing between UL3-ULC, LL3-LL2, LR3-LR2.
URCLO	CLP-57	M	Yes	Father: missing UL2.
ULCLP	CLP-81	F	No	-
ULCLP	CLP-85	F	Yes	Mother: missing LL5; microdont maxillary second molars (UR7&UL7).
CPO	CLP-86	F	Yes	Mother: ectopic UL3.
CPO	CLP-87	F	No	-
URCLO	CLP-04	M	Yes	Mother: transposed UR2.
URCLP	CLP-26	F	Yes	Mother: microdont maxillary second molars (UR7&UL7).
CPO	CLP-42	F	No	-
BCLP	CLP-44	M	Yes	Mother: transposed LLE & LL5 (mandibular).
ULCLP	CLP-56	M	No	-
ULCLP	CLP-59	F	Yes	Mother: MIH (moderate). Sibling: MIH (mild).

This table represents details for probands with isolated clefts (the cleft – only group). Those that fit the *de novo* criteria are highlighted in orange (7/16). Abbreviations: CPO, cleft palate only; CLO, cleft lip only; CLP, cleft lip and palate; UR, unilateral right; UL, unilateral left; B, bilateral. Note, those in the shaded grey cells are probands that do not have any dental anomalies but that have isolated enamel defects.

Table 3-5 Family medical, dental & cleft history for probands with 'cleft-only' and no dental anomalies

Study ID	Ethnicity	Family History of Cleft	Family History of Dental Anomalies	Other Medical History in the Family
CLP-8	WO	No	No	No
CLP-13	Mixed: White/Indian	No	No	No
CLP-24	Mixed: WB/WO	Mother's maternal uncle and his daughter CP, his granddaughter has cleft; Mother's aunt's son has cleft	No	No
CLP-30	WB	No	No	Proband has a 1-year old paternal stepsister born with a PRS affecting only the mandible but not the palate & bilateral coloboma of the iris
CLP-50	WB	No	Maternal grandmother two missing incisors; mother's nephew missing laterals	No
CLP-57	WB	Proband's father and grandfather have clefts	No	Brother has a tongue tie
CLP-81	WB	No	No	No
CLP-85	WB	Maternal grandfather's cousin has a cleft	No	No
CLP-86	WO	No	No	No
CLP-87	WO	No	No	No
CLP-04	WB	Mother's niece CP	Mother's sister has 6 'extra' teeth	Maternal grandmother's grandmother had 5 miscarriages of which all were males
CLP-26	Indian	No	No	Mother's paternal uncle has VPI and unintelligible speech. Mother's younger sister has epilepsy and severe ID
CLP-42	Indian	No	No	No
CLP-44	WB	No	No	No
CLP-56	Sri Lankan	No	No	No
CLP-59	WB	No	No	No

This table represents the family history of probands with isolated clefts (the cleft-only group). Those that fit the *de novo* criteria are highlighted in orange (and are now 6/16). Abbreviations: WO, White 'Other'; WB, White British; CP, cleft palate; PRS, Pierre Robin Sequence; VPI, velopharyngeal insufficiency; ID, intellectual disability.

3.3.3 The Dental Profile of Parents and Relatives of Children with Non-Syndromic Clefts

Regarding the children's family members, of the overall N=90 cohort, 24 families reported having a previous history of cleft defects within the family and 28 reported having dental anomalies. Three of the mothers and two of the fathers had a cleft themselves and one child had an affected sibling. Regarding the dental anomalies found in the 127 relatives examined overall, 55 from 45 trios were found to have a dental anomaly based on the criteria set out in this study (Chapter 2, Section 2.21.1); 29 were mothers, 11 fathers, 12 siblings and 3 'others' (second-degree relatives).

All the family members' dental anomalies are summarised in **(Table 3-6).** Those parents who had tooth anomalies had multiple ones, with hypodontia (11.5%) and Molar-Incisor-Hypo mineralization (MIH) (15%) being the most common, excluding isolated enamel defects. Interestingly, I saw some evidence of a familial pattern of MIH in my cohort. Further, all those who had hypodontia, were shown to have missing teeth outside their child's cleft region; (4/10) of those were parents of children in the 'cleft-only' group. Furthermore, six had microdont-molars and seven had transpositions outside their child's cleft area. Five individuals had unusual dental spacing found bilaterally in the mandible in the segment between the lateral incisors and canines (in two unrelated individuals), or bilaterally in the maxilla in the segments between the lateral incisors and canines or central and lateral incisors (in three unrelated individuals). The 'other' dental anomalies included: mandibular exostoses (n=2) and palatal tori (n=1); fissured tongue (n=1); fused/geminated teeth (n=2); dens evaginatus/dens invaginatus (n=3).

Table 3-6 Dental anomalies found in the parents & relatives of children with non-syndromic clefts

Type of anomaly	Number of family trios with a relative with the anomaly	Number of relatives with the anomaly **N=87**	Number of affected teeth/areas	Number and location of affected teeth compared with the proband's cleft site				Number of family members with the anomaly outside the proband's cleft site		
				Same as cleft site	Outside the cleft site			Maxilla	Mandible	Total
					Maxilla	Mandible	Total	Both		
Hypodontia	9	10	17	2	6	9	15	3	4	10
								3		
Ectopic canine	5	5	8	4	4	-	4	4	-	4
Transposition	9	9	10	2	5	3	8	4	2	7
								1		
Peg lateral/other microdont incisor	4	4	12	6	2	4	6	0	-	1
								1		
Microdont molar	5	6	9	-	8	1	9	5	1	6
Megadont	0	0	0	0	0	0	0	0	0	0
Supernumerary	3	3	3	2	1	-	1	1	-	1
Retained primary/delayed eruption-impaction	2	2	4	0	4	0	4	2	0	2
Unusual spacing	5	5	10	2	3	5	8	1	1	4
								2		
Enamel defect	21	22	46							
MIH	11	13	62							
Other anomalies	10	11								

Note: boxes in dark blue are dental anomalies that were observed outside the affected child's cleft region.

3.4 **DISCUSSION**

In this study, children and families were segregated to identify a small subset of six truly 'cleft-only' probands that may have *de novo* gene changes and as such may be candidates for whole exome sequencing. Interestingly, four of them had a defect in the palate only and this has been previously explained as having an X-linked or multifactorial aetiology such as nutritional and environmental influences. This is clearly the reason why population-based cleft research such as the Cleft Collective and EuroCleft are collecting data about family medical and social history. However, the newly emerging next-generation sequencing studies, such as whole exome sequencing, do not use population studies and rather use individual cases to find *de novo* gene changes and rely upon careful clinical identification and phenotyping. The issue of the mismatch between the surgical and clinical classifications has already been raised by genetic researchers (Aylsworth *et al.*, 2015; Farina *et al.*, 2002; Schutte & Murray, 1999; Weinberg *et al.*, 2008) and this study adds further evidence to demonstrate that clinical categorization alone is insufficient for genetic research to explain cleft aetiology. My findings also suggest that for some families with sporadic cases previously thought to have an environmental cause that there may in fact be an underlying *de novo* aetiology. Further genetic sequencing of these families might avoid mothers who thought that they took every care during pregnancy from being unjustly labelled (Honein *et al.*, 2007; Källén, 1997; Khoury *et al.*, 1987), especially since gene changes have been reported in the cleft palate only group by others using exome sequencing (Hoebel *et al.*, 2017; Liu *et al.*, 2017).

My findings suggest that a subset of cleft patients are erroneously categorised as having an isolated defect, e.g., no associated dental or medical anomaly. Despite the overall recruitment of my sample being targeted towards those diagnosed clinically as 'isolated' cleft patients, almost a third of them were found to have medical conditions known to be associated with clefting (a cohort described in Chapter 6), and over half had a dental anomaly. Furthermore, systematic oral phenotyping of these patients, and their families, demonstrated that the majority of subjects may have an inherited condition, which may manifest in the parent as a microform cleft or a minor tooth anomaly. Dental anomalies were found outside of the cleft area in both the cleft child and their relatives. Family members had hypodontia, microdontia and transposed teeth at a rate double to that reported in the normal population. Therefore, it is unlikely that these children had truly isolated cleft defects either genetically or biologically.

This study also identified a set of subjects with no family history of associated tooth anomalies. Isolated clefting is frequently thought to be a complex multifactorial disorder arising from nutritional, environmental and other influences. As a consequence, there are very few known cleft-only genes, and we also have a poor understanding of genetic predispositions towards clefting. Previous efforts have focused primarily on genome-wide association studies, in part because exome sequences are difficult to analyse at a population level. Identification of this subset of patients with non-familial clefting provides us with an opportunity to propose that their clefts arise from a *de novo* mutation event. Therefore, this set of subjects may be good candidates for future identification of *de novo* cleft genes.

A number of limitations exist in the current study; the methodology and the sampling were not designed to estimate association rates. Also, as recruitment was based on a convenience sample, other close relatives with tooth anomalies may have been missed. Due to ethical reasons, radiographs of parents and relatives were not taken, although some participants were able to provide information from their primary practice. Therefore, the six with isolated clefts may shrink even further. In addition, I included children and families with a diagnosis of speech/language/hearing conditions; despite the known associations with orofacial clefting. As such, it is possible that those families may not truly be presenting with a *de novo* genetic change but rather already have families with laryngeal/pharyngeal microform presentations; an example of this might even be interpreted as a child who presented with the submucous cleft (CLP-24), who was later excluded since it was revealed that she came from a 'multiplex' family with multiple cleft-affected individuals. Another example of this misclassification as 'isolated', is a child from the 'cleft-only' group whose maternal uncle has velopharyngeal insufficiency and unintelligible speech (ascertained by the mother who is a speech and language therapist).

On the other hand, a strength of this study is the rigor of the dental exclusion criteria such as the exclusion of families who had tooth transposition from the 'isolated group', a recognised associated dental anomaly (Aspinal et al, 2014), and also those with an enamel defect when it was the only dental anomaly found. Excluding those with an enamel anomaly only might be considered to be too severe especially since the research into the link between enamel defects and genetic inheritance has been scarcely explored but there is too great a likelihood of environmental causes (Aine *et al.*, 2000; Alaluusua *et al.*, 1996; Darchuk *et al.*, 2012; Jan & Vrbič, 2000). However, this study showed that over a third of child probands had relatives with an enamel defect, adding weight to

other researches that found an association with cleft defects and a potential shared genetic inheritance, especially with molar-incisor hypomineralisation (Hubbard *et al.*, 2017; Vieira & Kup, 2016), a common but relatively unexplored anomaly in cleft research. Previous studies have not defined what constituted an 'enamel defect', especially when examining the parents. Aspinal *et al.* (2014) did not use specific criteria to define them and Howe *et al.* (2015) restricted their data to hypoplasia only because it could be recorded reproducibly from clinical photos since they did not examine everyone in person.

This study adds weight to the need to recategorize clinical diagnoses in line with the genetic/biological aetiology in cleft lip/palate research since it is clear that many more families have an inheritable disorder than the clinical/surgical label of 'isolated cleft' implies. Most research to date has focused on either common variants in population-based studies or inherited modalities. Thus, defining and grouping isolated cleft probands and their families based on the absence of dental or medical anomalies might provide valuable insights into the aetiology. This study also highlights the importance of a thorough dental and medical screening of affected probands and their relatives for accurate phenotyping and the potential role of paediatric dentists and the wider primary care dental team; after all, McBride and co-workers suggested that comparing families with similar cleft conditions, who have not had other causative explanations, might help to find the common aetiology (McBride *et al.*, 2016) .This is especially true when reported alongside the wider medical history and familial dental findings, but as yet, few studies have reported both together. Thus, segregating non-syndromic probands (and their families) into those with 'true' isolated clefts and those with clefts and tooth anomalies appreciates the genetic heterogeneity underlying non-syndromic CLP and would be of great aid in the design of future studies and in the early genetic diagnosis and counselling of affected families.

3.5 OUTSTANDING QUESTIONS

Since I have identified the subset of six probands who only had the cleft with no dental anomalies and with no family history of clefts or dental anomalies, the questions that require further exploration are: do these patients have genetic variants and if so, what is the nature of the genetic variation? Is whole exome sequencing an appropriate method to uncover new candidate genes in this group of patients that fit the *de novo* model? Future clinical studies could be carried out with a particular focus on this subgroup.

3.6 CONCLUSIONS

This study highlights the need to rethink genetic risk models in cleft lip/palate research since most of the research focus on non-syndromic clefts has been on common variants in population-based studies or on inherited modalities. Thus, the small group of children with sporadic clefts fitting the *de novo* model will provide valuable insights into the aetiology of isolated clefts. This highlights the role of the dentist for the accurate assessment of factors conferring genetic susceptibility/predisposition to orofacial-related congenital anomalies particularly in those with cleft palate only.

Chapter 4 NOVEL TRUNCATING MUTATIONS IN *CTNND1* CAUSE A DOMINANT CRANIOFACIAL AND CARDIAC SYNDROME

4.1 SUMMARY

The Clinical Study identified 28 children that had cleft-associated congenital anomalies in undefined syndromes, one of whom I identified to have a novel *de novo* mutation in the catenin delta-1 gene (*CTNND1*). In this chapter I will describe a series of patients with *CTNND1* variants, all of whom presented with multisystem involvement that demonstrates a broad spectrum craniofacial and cardiac syndrome. *CTNND1* encodes the p120-catenin (p120) protein, which has a wide range of functions, including the maintenance of cell-cell junctions, regulation of the epithelial-mesenchymal transition and transcriptional signaling. Due to advances in next generation sequencing, *CTNND1* has been implicated in human diseases including cleft palate and blepharocheilodontic syndrome (BCD) albeit only recently. It was therefore an **aim** central to this project to expand on the *CTNND1* human variants and phenotypes to better reflect and appreciate the abundant data known so far from model organisms and *in vitro* experiments. **Methods:** Whole exome sequencing was first carried out in a female proband who presented through the Clinical Study with an undiagnosed craniofacial syndrome. Other individuals were then found through multiple collaborations. **Results:** In this chapter, I identified eight novel protein-truncating variants, six *de novo,* in thirteen participants presenting with craniofacial dysmorphisms including orofacial clefts and hypodontia, as well as congenital cardiac anomalies, limb dysmorphologies and neurodevelopmental disorders. Using conditional deletions in mice as well as CRISPR/Cas9 approaches to target *CTNND1* in *Xenopus*, I identified a subset of phenotypes that can be linked to p120-catenin in epithelial integrity and turnover, and additional phenotypes that suggest mesenchymal roles of *CTNND1*. **Conclusions:** From the findings in this chapter, I propose that *CTNND1* variants have a wider developmental role than previously described, and that variations in this gene underlie not only CLP and BCD but may be expanded to a broader velocardiofacial-like syndrome.

4.2 INTRODUCTION

This project arose from a participant recruited through my Clinical Study (Chapters 2 & 3). The female patient presented with an undiagnosed syndromic form of cleft. She was 11 years of age and presented with a submucous cleft palate, diagnosed when she was two years of age; velopharyngeal insufficiency; mild facial dysmorphologies including asymmetric ears and mid-facial hypoplasia; a single transverse palmar crease on the right hand; autism spectrum disorder (ASD) and behavioural problems. She had 12 missing permanent teeth. Her older brother who was 15 years of age had three missing permanent adult teeth, namely the left maxillary and mandibular premolars, and the right maxillary lateral incisor and had a peg left maxillary lateral incisor. He was reported by his mother to have Asperger syndrome and attention deficit hyperactivity disorder. The mother and father had normal dentitions, the father and his sister were diagnosed with Chron's disease and there was no family history of craniofacial conditions or cleft lip/palate. The proband had multiple genetic testing throughout her lifetime, including testing for 22q11 deletion and Down syndrome, in which all returned 'normal' findings. Previously, her chromosomal analysis identified an additional small 2p12 duplication that she shares with her mother, a region that contains *CTNNA2;* however, this had been deemed insignificant.

The family took part in the exome sequencing study of my project and WES revealed that the proband had a double heterozygous mutation in the *CTNND1* gene, one of which was a novel, *de novo* truncating mutation (p.Arg461*). The other variant in *CTNND1* was a paternally inherited rare variant (p.Arg315Cys). The father, however, does not share any of the phenotypes with his daughter.

CTNND1 was not associated with congenital human conditions at the time the pathogenicity of this genotype was to be determined. Therefore, a goal was to obtain other patients who had an underlying diagnosis of a *CTNND1* variant, ideally an identical mutation to my patient's, and examine their phenotypes to determine if there were overlaps. The utilisation of the Deciphering Developmental Disorders dataset to discover and recruit similar patients and the validation of the *de novo* gene findings by testing expression in human embryos, then investigating the human manifestations in moue and *Xenopus* models in order to establish a link between *CTNND1* with congenital human anomalies will be described in detail in this chapter.

4.2.1 **Catenin Delta 1 (*CTNND1*) Gene and Protein**

Genetic variation in *CTNND1,* which encodes for the armadillo-repeat protein p120-catenin (p120), is associated with human birth defects, most notably non-syndromic cleft palate and blepharocheilodontic (BCD) syndrome, which involves eyelid, lip and tooth anomalies [MIM: 617681] (Cox *et al.*, 2018; Ghoumid *et al.*, 2017; Kievit *et al.*, 2018). In contrast, *CTNND1* has broader developmental roles in animal models. For example, conditional deletions in mice demonstrate the importance of *CTNND1* for development not only for skin and teeth, but also for kidneys and other structures (Bartlett *et al.*, 2010; Davis & Reynolds, 2006; Elia *et al.*, 2006; Hendley *et al.*, 2015; Marciano *et al.*, 2011; Oas *et al.*, 2010; Perez-Moreno *et al.*, 2006), and complete deletion of *CTNND1* leads to prenatal lethality (Davis & Reynolds, 2006; Oas *et al.*, 2010). Similarly, loss-of-function experiments in *Xenopus* implicate *CTNND1* in craniofacial development (Ciesiolka *et al.*, 2004; Geis *et al.*, 1998).

p120-catenin is a member of the catenin superfamily of proteins studied in catenin-cadherin interactions; notably, it binds to and stabilizes epithelial cadherin (E-cadherin (*CDH1*)) at junctional complexes in epithelia (Davis *et al.*, 2003; Fukumoto *et al.*, 2008; Ireton *et al.*, 2002; Ishiyama *et al.*, 2010; Reynolds *et al.*, 1994). This binding is via the p120-catenin armadillo repeat domain, and displacement of p120-catenin from E-cadherin is a key regulatory event at the adherens junction, that results in endocytosis of E-cadherin and loss of the junction. The protein has a second function as a scaffolding protein for the GTPase RhoA and associated Rho regulatory proteins (Anastasiadis *et al.*, 2000; Wildenberg *et al.*, 2006). In addition, it can also directly interact with the zinc finger transcriptional repressor Kaiso (ZBTB33), facilitating Wnt signal transduction (del Valle-Pérez *et al.*, 2011; Park *et al.*, 2005). Thus, p120-catenin appears to be a multi-functional protein, promoting epithelial stability when in complex with E-cadherin, and regulating RhoA and transcriptional activities. p120-catenin is also able to associate with mesenchymal cadherins such as N-cadherin and cadherin-11 (Reynolds *et al.*, 1994; Yanagisawa & Anastasiadis, 2006). In mesenchymal cells, p120-catenin associates with non-epithelial cadherins, regulating motility and invasion via cytoskeletal events and transcription. Given its functions in both epithelia and mesenchyme, it is unsurprising that both loss and gain of p120-catenin have been associated with oncogenesis (Reynolds & Roczniak-Ferguson, 2004; Schackmann *et al.*, 2013; Stairs *et al.*, 2011).

In humans, the *CTNND1* gene is located at 11q11 and consists of 21 exons, of which exons 11, 18 and 20 are alternatively spliced. Inclusion of exon 11, which is

predominantly neural, disrupts a nuclear localization signal (NLS), while exon 20 contains a nuclear export signal (NES) (Reynolds *et al.*, 1996). In addition, splicing in the N-terminus region of p120 gives rise to four alternative start codons, therefore, there are four isoforms of the p120 protein which vary in their transcriptional start sites. Of the four major isoforms, isoforms 1 and 3 are observed most frequently and retain both the phosphorylation domain and the ARM domain (Ireton *et al.,* 2002). Isoform-1 is abundant in mesenchymal cells, while isoform-3 appears preferentially expressed in epithelial cells (Aho *et al.*, 2002; Hong *et al.*, 2016; Keirsebilck *et al.*, 1998; Montonen *et al.*, 2001). Isoforms 2 and 4 are less well characterized.

4.2.2 **The Junctional Complex and the Cadherin-Catenin Relationship**

Cell-cell adhesion is a crucial mechanism in many organisms for the maintenance of structural integrity. In animal epithelia, this is maintained by the many classes of junctions that cross-bound one cell to the other such as tight junctions, adherens junctions, desmosomes and gap junctions (Davis *et al.*, 2003; Harris, 2012). The cadherin-catenin complex is what makes the adherens junctions unique, constituted by classical cadherins and catenins (Pieters *et al.*, 2012). Across the cell membrane, these complexes contain an extracellular component that mediates intercellular binding outside the cell and maintains intracellular integrity by associating with the actin cytoskeleton inside the cell (Ishiyama et al., 2010) **(Figure 4-1).**

The maintenance of the cadherin-catenin complex is essential for normal embryogenesis and development, likewise, its dissociation is a key biologic machinery particularly for events such as cell migration, epithelial folding and epithelial to mesenchymal transition (EMT). The loss of binding between cadherins and catenins at the cell membrane can occur normally via transcriptional regulation, through mutations in either of the proteins or by aberrant cadherin internalization (Mosesson *et al.*, 2008). This fine balance between prevention and promotion of E-cadherin internalisation and turnover is what dictates development and disease in planar cell polarity, gastrulation and EMT. In the latter, E-cadherin expression levels are reduced in the cells through the process of cadherin internalization (Miller & McClay, 1997).

The domain by which E-cadherin binds to p120-catenin is known as the juxtamembrane domain (JMD) **(Figure 4-1)**. After the Ca^{+2} binding domain in E-cadherin, three other highly conserved domains are found, the transmembrane domain, the JMD and the catenin binding domain (CBD). The latter three domains constitute the residues that make up the cytoplasmic region of E-cadherin (Ishiyama *et al.*, 2010). The

JMD particularly consists of a 50 amino acid sequence motif that specifically provides a binding site for the p120 subfamily of proteins, including, p120 catenin, ARVCF, p0071 and δ-catenin2. Ishiyama and co-workers showed that p120 residues Lys401, Lys444 and Asn478 are crucial for the interaction with the JMD (Ishiyama *et al.*, 2010).

The regulation of cadherins turnover is induced by a number of mechanisms one of which is their modulation by catenins. It is thought that alpha, beta and p120-catenin contribute to the regulation of adherens junctions (Harris, 2012). Although clear evidence is lacking, p120-catenin was particularly shown to act as a 'set-point' for cadherin expression and as a key inhibitor of rapid cadherin turnover (Davis *et al.*, 2003; Ireton *et al.*, 2002; Xia *et al.*, 2003b). The mechanism by which this occurs has been described by Chiasson and co-workers in the context of vascular endothelial (VE)-cadherin, which has analogous roles to E-cadherin except that it is specific to endothelial cells (Chiasson *et al.*, 2009). When p120-catenin binds to cadherins an endocytic adaptor binding site is masked. Upon dissociation of p120 from cadherins, the binding site is exposed, allowing the endocytic adaptor to bind to cadherins prompting cadherin endocytosis (Chiasson *et al.*, 2009). In light of this, p120 was proposed as a master regulator of cadherins levels in cells (Harris, 2012).

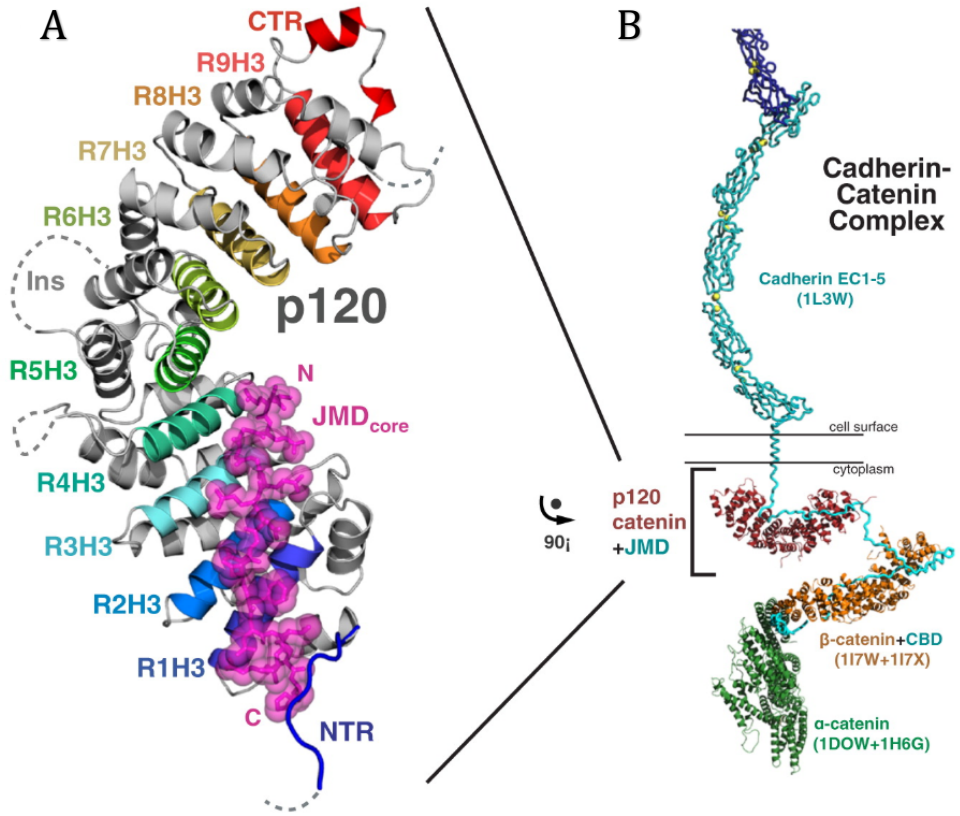


Figure 4-1 The Cadherin-Catenin Complex

[A] Crystal Structure of p120 (isoform 4A) in complex with the JDM core of E-cadherin. R1-9 represents the armadillo arm repeats of p120, H3 are the three helices formed at each arm. Ins is the region of p120 isoform 4 that has been modified. [B] Schematic representation of the cadherin-catenin complex. NTR: N-terminal region; CTR: C-terminal region. Adapted from (Ishiyama et al., 2010).

4.2.3 **The p120-Catenin Superfamily**

The p120 superfamily includes p120-catenin itself (δ-catenin1), δ-catenin2 (CTNND2) and ARVCF (armadillo repeat gene deleted in velocardiofacial syndrome) all of which can compete for E-cadherin binding. Although it is unclear whether they substitute for one another in other cellular functions (Hatzfeld, 2005; Mariner *et al.*, 2000), evidence from animal studies suggests some compensatory roles. For instance, δ-catenin2 (CTNND2) knockdown phenotypes can be rescued with p120-catenin, and the combined depletion of δ-catenin2 and p120 generates more pronounced effects. However, levels of p120 are not altered by reducing δ-catenin protein levels (Gu *et al.*, 2009). In humans, *CTNND2* variants have been associated with autism spectrum disorders and other neurodevelopmental conditions (Belcaro *et al.*, 2015; Hofmeister *et al.*, 2015; Lu *et al.*, 2016; Medina *et al.*, 2000; Nivard *et al.*, 2014; Turner *et al.*, 2015). Interestingly, the other p120 family member, *ARVCF,* lies in 22q11. While loss of *TBX1* in 22q11 is thought to cause the key malformations associated with velocardiofacial (VCF) syndrome [MIM: 192430], evidence from animal models suggests that *ARVCF* may also play a role in craniofacial development (Butts, 2009; Cho *et al.*, 2011; Shprintzen *et al.*, 1978; Sirotkin *et al.*, 1997).

4.2.4 **Current *CTNND1* Study Synopsis**

Although both p120-catenin and its binding partner E-cadherin have been proposed as causative genes in non-syndromic palatal clefting and BCD syndrome (Cox *et al.*, 2018; Ghoumid *et al.*, 2017; Kievit *et al.*, 2018), the patients that I describe in the current project present with a multisystem condition broader than the previously described p120-associated BCD cases. While the patients in this study consistently possess palatal phenotypes (cleft or high-arched palate) as well as tooth anomalies, they also display additional features including severe hypodontia, cardiac, limb and neurodevelopmental anomalies. I hypothesize that these novel truncating variants in *CTNND1* affect both E-cadherin-dependent and -independent functions of p120-catenin, and, given the range of phenotypes seen in this study cohort, should be considered more broadly to cause a velocardiofacial-like syndrome.

4.3 RESULTS

4.3.1 Identification of *CTNND1* Variants

Here, I identified 13 individuals with protein-truncating variants in *CTNND1*. Previously, all patients had undergone an array-based comparative genome hybridization analysis with normal results. A subset of patients had also been referred for other diagnostic tests, including 22q11 deletion, Down syndrome, CHARGE syndrome (*CHD7* sequencing), Noonan syndrome (*PTPN11* sequencing) and other conditions, but with no definitive diagnoses. Whole exome sequencing of the patients revealed eight novel variants in *CTNND1*, including six confirmed *de novo* variants (in 7 patients). Two individuals inherited their variant from affected parents while two other participants inherited a variant from a parent with a mild phenotype **(Figure 4-2, A)**. These truncating mutations included nonsense, splicing and frameshift variants **(Table 4-1)**.

CTNND1 variants identified could be grouped according to the overall structure of the protein **(Figure 4-3)**. One variant falling within the N-terminal regulatory region was identified in Patient 1. Patient 1 has a *de novo CTNND1* c.443_444delTG (p.Val148Aspfs*24) mutation in exon 6. Targeted sequencing for this variant was carried out on the affected daughter (Patient 2), which segregates with the phenotypes in the family.

Four variants fell within the armadillo repeats, which are predicted to be crucial for interactions with E-cadherin. Two unrelated individuals (Patients 3 and 4) both had a *de novo* mutation in *CTNND1*: c.1381C>T (p.Arg461*) **(Figure 4-2 & 4-3)**. This variant results in a nonsense substitution and creates a stop codon in exon 7. In addition, Patient 3 had a rare variant in *CTNND1*, inherited paternally c.943C>T (p.Arg315Cys), which is present at a frequency of $2x10^{-4}$ in reference populations (Karczewski *et al.*, 2019). As the parent shares none of the phenotypes with the patient, this second variant is unlikely to be causative. Patient 5 had a *CTNND1* variant c.2389C>T (p.Arg797*) on exon 15. A *CTNND1* frameshift variant c.1481_1485del (p.Leu494Argfs*5) in exon 8 was identified in a mother and child; both are affected (Patients 6 and 7, respectively). In the same exon, Patient 8 had a *CTNND1* variant c.1594del (p.Gly532Alafs*6).

Three variants affecting the C-terminal domain were found, these were present in five patients in three families. The variant c.2598_2601dupTGAT (p.Ser868*) was paternally inherited in a family with two affected siblings (Patients 9 and 10). The father

is fit and healthy; however, his palate is narrow and high, and his nose is prominent. Patient 11 has a *de novo CTNND1* variant at the splice acceptor site of exon 19 designated as c.2702-5A>G, which is predicted to create a cryptic splice site, leading to a premature termination codon at the start of exon 19. Finally, Patients 12 and 13 are monozygotic twins carrying a *de novo* frameshift variant in *CTNND1*: c.2737dupC (p.His913Profs*3).

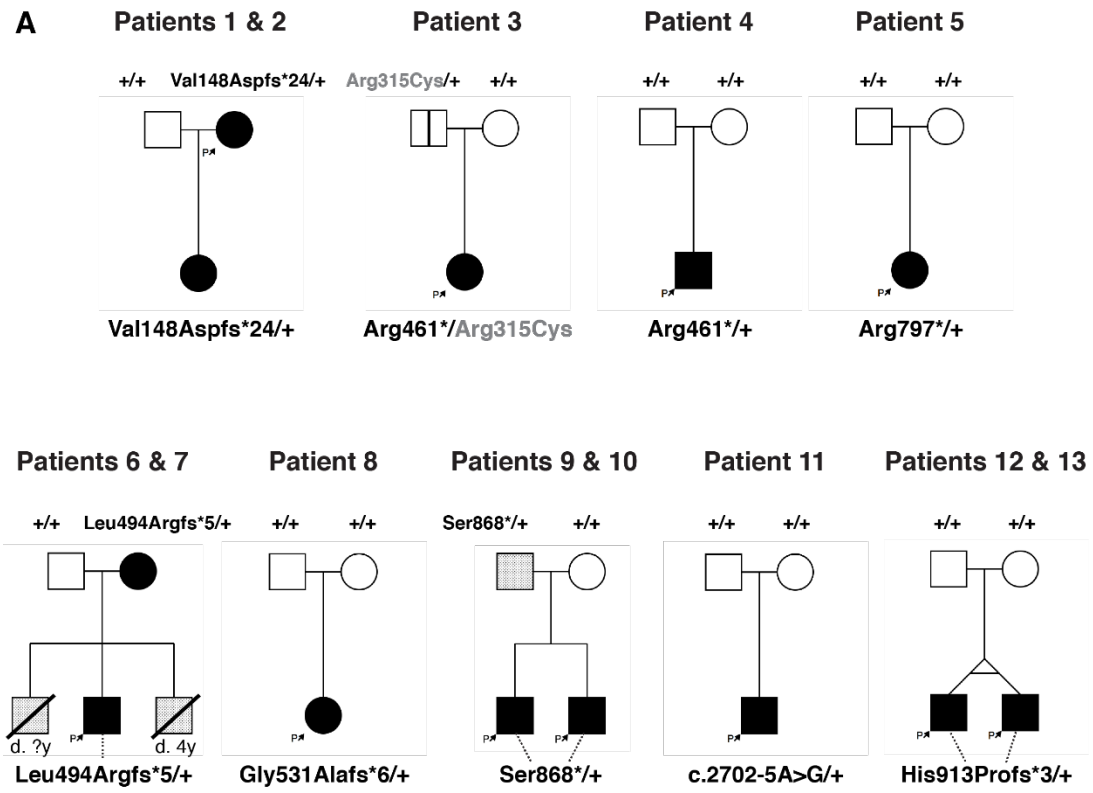


Figure 4-2 Pedigrees of index patients and their CTNND1 variants

[A] Pedigrees of individuals with identified variants. Filled boxes indicate affected individuals demonstrating collective phenotypes described in our cohort. A blank box with a vertical black line indicates an asymptomatic carrier (clinically unaffected). A box with an oblique line indicates a deceased individual. Lightly shaded boxes indicate individuals affected with one or more of the conditions described.

Table 4-1 CTNND1 variants in index patients

Patient ID	Mutation	Protein	Variant type	Exon	gnomAD
Patient 1	c.443_444delTG	p.Val148Aspfs*24	frameshift	6	novel
Patient 2	c.443_444delTG	p.Val148Aspfs*24	frameshift	6	novel
Patient 3	c.943C>T	p.Arg315Cys	missense	6	2.44e-4 8 FE, 39 NFE, 4 A
Patient 3	c.1381C>T	p.Arg461*	nonsense	7	novel
Patient 4	c.1381C>T	p.Arg461*	nonsense	7	novel
Patient 6	c.1481_1485del	p.Leu494Argfs*5	frameshift	8	novel
Patient 7	c.1481_1485del	p.Leu494Argfs*5	frameshift	8	novel
Patient 8	c.1595del	p.Gly532Alafs*6	frameshift	8	novel
Patient 5	c.2389C>T	p.Arg797*	nonsense	15	novel
Patient 9	c.2598_2601dupTGAT	p.Ser868*	nonsense	17	novel
Patient 10	c.2598_2601dupTGAT	p.Ser868*	nonsense	17	novel
Patient 11	c.2702-5A>G	p.?	splice site	18-19	Novel
Patient 12	c.2737dupC	p.His913Profs*3	frameshift	19	novel
Patient 13	c.2737dupC	p.His913Profs*3	frameshift	19	novel

The Human GRCh37 (hg19) Assembly was used to identify transcript positions. The annotations are all based on the NM_001085458 transcript. Confirmations using https//:variantvalidator.org. *Denotes termination codon; FE, Finnish European; NFE, Non-Finnish European; A, African.

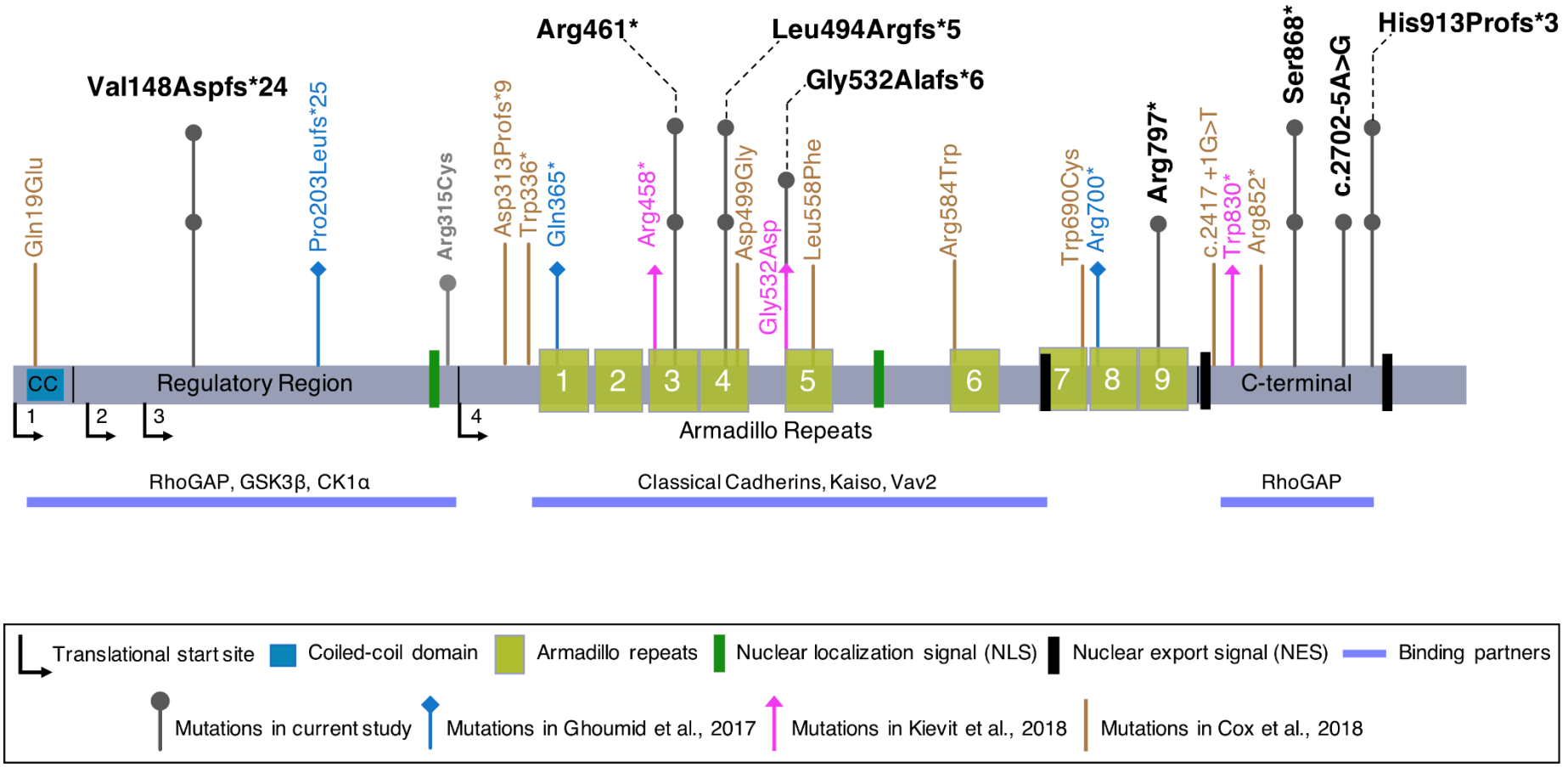

Arg461*
Leu494Argfs*5
His913Profs*3
Val148Aspfs*24
Gly532Alafs*6
Ser868*
Arg797*
c.2702-5A>G
Gln19Glu
Pro203Leufs*25
Arg315Cys
Asp313Profs*9
Trp336*
Gln365*
Arg458*
Asp499Gly
Gly532Asp
Leu558Phe
Arg584Trp
Trp690Cys
Arg700*
c.2417+1G>T
Trp830*
Arg852*
CC
Regulatory Region
1
2
3
4
5
6
7
8
9
C-terminal
Armadillo Repeats
RhoGAP, GSK3β, CK1α
Classical Cadherins, Kaiso, Vav2
RhoGAP
Translational start site
Coiled-coil domain
Armadillo repeats
Nuclear localization signal (NLS)
Nuclear export signal (NES)
Binding partners
Mutations in current study
Mutations in Ghoumid et al., 2017
Mutations in Kievit et al., 2018
Mutations in Cox et al., 2018

Figure 4-3 Schematic of the CTNND1 protein structure & human variant mapping

Schematic representation of the human p120-catenin protein structure and its domains. The variants described in our cohort are shown above the protein with a dark gray arrow. The light gray arrow with the (p.Arg315Cys) variant indicates the other *CTNND1* mutation found in Patient 3 which was inherited from the unaffected father [A]. Arrows in blue, pink and brown represent the variants and their locations reported in (Ghoumid et al., 2017), (Kievit et al., 2018) and (Cox et al., 2018), respectively.

4.3.2 Clinical Presentation of Patients with *CTNND1* Variants

Clinical phenotypes are summarized in **(Table 4-2)**, and further details can be found in (Appendix 5). Photographs from participants show a number of shared craniofacial and oral features **(Figure 4-4 and Figure 4-6, respectively)** as well as other affected structures (eyes, ears and limbs **(Figure 4-5)**). Additional features including heart anomalies and neurodevelopmental conditions are noted in **(Table 4-2** and Appendix 5**).**

Participants shared several distinctive eye features including short, up-slanted palpebral fissures (9/13), hooded eyelids (8/13), telecanthus (7/13), highly arched (8/13) and thin lateral eyebrows (8/13) and other eyelid anomalies such as nasolacrimal obstructions (1/13). These eye anomalies were clear from a young age **(Figure 4-5, A)**. A subset had ectropion (drooping lower eyelids, 4/13) and distichiasis (double eyelashes, 4/13). Many individuals had wide nasal bridges (11/13) with broad nasal tips (7/13), choanal atresia (4/13), either unilateral or bilateral atresia; malar flattening (mid-face hypoplasia) (9/13); mandibular prognathism (5/13); thin upper lips (7/13) and auricular abnormalities (9/13), particularly low-set ears and overfolded helices **(Figure 4-5, B)**.

Phenotypes with high penetrance involved oropharyngeal abnormalities including cleft lip and/or palate (CLP) (8/13), high-arched palate (7/13) or a combination of cleft and high-arched palate **(Figures 4-6, A-D)**. A range of cleft sub-types was seen**,** three had left complete unilateral CLP, two had right complete unilateral CLP, one had bilateral complete CLP, one had a cleft palate and had a submucous cleft of the soft palate. The latter participant also had velopharyngeal insufficiency (VPI) and a bifid uvula. Of interest, three individuals presented with vocalization defects causing stridor and hoarseness or nasal speech.

Upon dental examination, all subjects were found to have intra-oral anomalies **(Figure 4-6 & 4-7)**. In particular, congenital tooth agenesis (hypodontia) was frequently seen, with eight subjects missing between three and twelve adult teeth **(Table 4-3, Figure 4-6 G-H & Figure 4-7, A-D)**. Other anomalies included retained primary teeth and delayed eruption of the permanent teeth (6/13) (Appendix 5). Morphologic tooth anomalies were present, including diminutive permanent teeth/peg-shaped lateral incisors and fissured crowns of the permanent central and lateral incisors **(Figures 4-6, E-F;** Appendix 5**)**.

Beyond the craniofacial structures, the majority of the participants had limb and heart anomalies. Mild limb phenotypes (9/13) were present, including shorter fifth fingers, single transverse palmar crease, mild syndactyly between the 2,3 toes, sandal gaps and camptodactyly of the toes **(Figure 4-5, C)**. Congenital cardiac defects, which have not previously been associated with *CTNND1* variants, consistently occurred in this study cohort. Six subjects had cardiovascular anomalies including tetralogy of Fallot, hypoplastic aortic arch, coarctation of the aorta, ventricular septal defect, atrial septal defect, mitral valve stenosis, patent ductus arteriosus and patent foramen ovale **(Table 4-2)**. Finally, in addition to the craniofacial and cardiac anomalies, individuals presented with other phenotypes that added to the complexity of their conditions. Developmental delay and other neurodevelopmental problems were also observed (8/13). These often appeared from early toddler and school years and included mild learning difficulties, autism spectrum disorder, speech and language delay, and behavioural problems **(Table 4-2)**. One individual was diagnosed with ovarian dysgerminoma stage III in the left ovary at the age of 12 years, which was treated with left oophorectomy followed by chemotherapy. Other infrequent anomalies included urogenital problems, scoliosis and partial agenesis of the corpus callosum **(Table 4-2;** Appendix 5**)**.

Table 4-2 Clinical summary of individuals with CTNND1 variants

	Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	**tot**
	Sex	F	F	F	M	F	F	M	F	M	M	M	M	M	6F/ 7M
Craniofacial	Cleft lip/palate	-	-	+	-	-	+	+	-	+	+	+	+	+	8/ 13
	High-arched palate	+	+	+	-	-	-	+	+	-	ND	-	+	+	7/ 13
	Thin upper lip	+	+	-	-	-	-	+	+	+	-	+	+	-	7/ 13
	Choanal atresia	+	+	-	-	-	-	-	+	-	-	+	-	-	4/ 13
	Ear anomaly	-	+	+	+	-	+	+	+	+	+	+	-	-	9/ 13
	Wide nasal bridge	+	+	+	-	-	+	+	+	+	+	+	+	+	11/ 13
	Broad nasal tip	+	-	+	-	-	-	+	+	+	+	+	-	-	7/ 13
	Mid-facial hypoplas.	+	+	+	-	-	+	+	+	-	-	+	+	+	9/ 13
	Mandib. prognathism	+	-	+	-	-	-	-	+	-	-	+	-	+	5/ 13
	Brachy cephaly	-	+	-	+	-	-	-	-	-	-	+	-	-	3/ 13
Eyes and eyelids															
	Narrow, upslanted palpebral fissures	-	-	+	+	-	-	+	+	+	+	+	+	+	9/ 13
	Hooded eyelids	-	-	+	+	-	-	-	+	+	+	+	+	+	8/ 13
	Telecanthus	-	-	+	+	-	-	-	-	+	+	+	+	+	7/ 13
	High arched eyebrows	+	+	-	-	-	+	+	+	-	-	+	+	+	8/ 13
	Thin lateral eyebrows	+	-	-	-	+	+	+	+	+	+	+	-	-	8/ 13
	Mild ectropion	+	-	-	+	+	+	-	-	-	-	-	-	-	4/ 13
	Distichiasis	+	+	-	-	+	-	+	-	-	-	-	-	-	4/ 13
	Ankyloble pahron	-	+	-	-	-	-	+	-	+	-	-	-	-	3/ 13
Dental anomalies															
	Hypodontia	+	+	+	+	+	-	-	+	ND	ND	+	-	+	8/ 13
	Delayed dentition	+	+	-	+	+	-	-	ND	ND	ND	+	-	+	6/ 13
	Abnormal crown form	+	+	+	-	+	-	+	+	+	ND	+	+	-	9/ 13

Cardiac disease	VSd	+	+	-	+	-	-	-	-	-	-	+	+	-		
	TOF	-	-	-	-	-	-	-	+	-	-	-	-	-		
	Asd or PFO	+	+	-	+	-	-	-	-	-	-	-	-	-		
	MVS	+	-	-	-	-	-	-	-	-	-	-	-	-	tot 6/ 13	
	PS or COA	-	-	-	-	-	-	-	-	-	-	+	-	-		
	PDA	-	+	-	-	-	-	-	-	-	-	-	-	-		
	Hypoplastic aortic arch	+	-	-	-	-	-	-	-	-	-	+	-	-		
Neurodevelopmental	ASD	-	UI	+	+	-	-	-	-	UI	-	+	-	-		
	ADHD	-	+	+	-	-	-	-	-	-	-	+	-	-		
	DD/LD	-	+	+	-	-	-	-	-	-	+	+	+	+	tot 8/ 13	
	Speech & language delay	-	-	+	-	-	-	-	-	-	+	+	-	-		
	Aggressive behaviour	-	+	+	-	-	-	-	-	+	+	-	-	-		
Limb anomalies															tot 9/ 13	
	Hands	-	-	+	-	-	+	+	+	-	-	+	+	+	7/ 13	
	Feet	-	+	+	+	-	+	+	-	-	-	-	+	+	7/ 13	
Voice anom.		-	+	-	-	ND	+	-	-	-	-	+	-	-	3/ 13	
Skeletal		+	-	+	+	-	+	-	-	-	-	+	-	-		
	Scoliosis	+	-	-	-	-	+	-	-	-	-	-	-	-	tot 5/ 13	
	Short stature	-	-	-	-	-	+	-	-	-	-	+	-	-		
Cancer		-	-	-	-	-	-	-	+	-	-	-	-	-	1/ 13	
Other	Patient 1: restrictive lung disease Patient 2: partial agenesis of corpus callosum Patient 3: VPI, early onset puberty, bowel problems Patient 4: joint laxity Patient 6: hypothyroid Patient 8: macroglossia Patient 11: cryptorchidism Patient 12: coronal hypospadias															

Abbreviations: tot, total; UI, under investigation, ND; not determined because of non-availability; VSd, ventricular septal defect; Asd, atrial septal defect; TOF, tetralogy of Fallot; CoA, coarctation of the aorta; MVS, mitral valve stenosis; PDA, patent ductus arteriosus; PFO, patent foramen ovale; ASD, autism spectrum disorder; ADHD, attention deficit hyperactivity disorder; DD, developmental delay; LD, learning difficulty; VPI, velo-pharyngeal insufficiency.

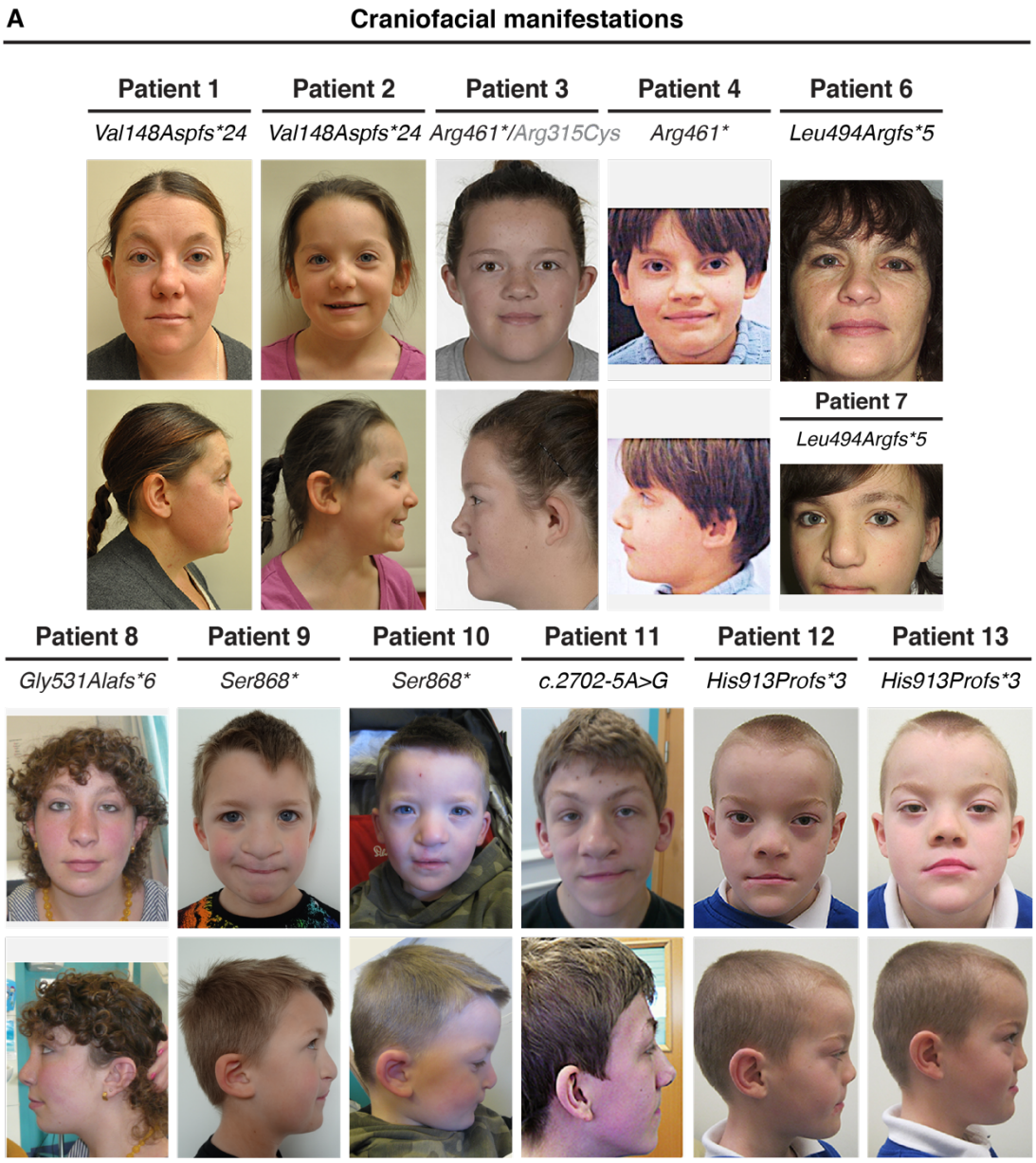


Figure 4-4 Clinical presentation of individuals with a CTNND1 mutation

Facial photos (frontal and profile) show craniofacial features of patients. Note the narrow up-slanting palpebral fissures in Patients 3,4, 7-13; the hooded eyelids in patients 3, 4, 8-13; telecanthus in Patients 3,4,9-13; the high arched eyebrows in patients 1, 2, 6-8, 11-13 and the thin lateral eyebrows in Patients 1,5-11. Patients 1 and 4 had missing eyelashes medially from the inner canthus; Patients 1,2, 5 and 7 have distichiasis (double row of lashes), and mild ectropion of the lower eyelids were seen in Patients 1,5 and 6. As evident, no patient shows signs of hair sparsity. Most patients had wide nasal bridges with broad nasal tips while Patients 1,2, 8 and 11 were also diagnosed with congenital choanal atresia. Patients 1,2,7-9, 11 and 12 showed thin upper lips and while mid-face hypoplasia was observed, Patients 1,3,8,11 and 13 also had mandibular prognathism. Scars from cleft lip operations are seen in Patients 7, 9-13. Patient 3 was born with a submucous cleft palate, a bifid uvula and velopharyngeal insufficiency.

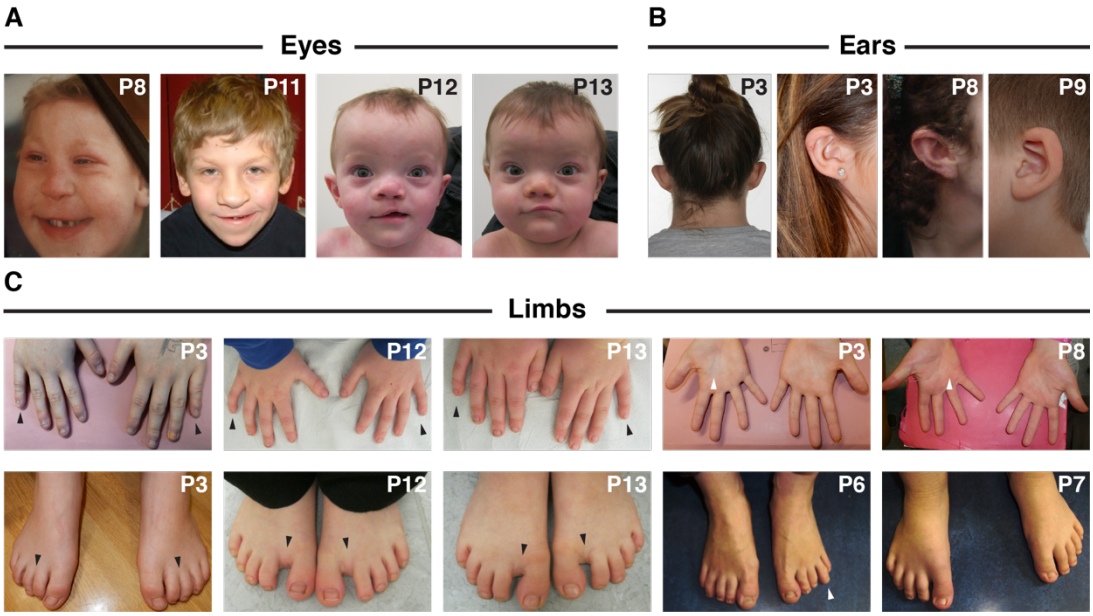


Figure 4-5 Eye, ear and limb anomalies in individuals with a CTNND1 mutation

[A] The eye phenotypes of the narrow palpebral fissures, the hooded eyelids and highly-arched, thin lateral eyebrows were evident from a young age. [B] Ear anomalies included: low-set ears, sometimes asymmetric and/or small; overfolded helices of the external ears; a pre-auricular pit was also seen in one of the patients (data not shown). [C] Upper limb anomalies included: slightly shorter 5th fingers as seen in Patients 3, 12 and 13; and a single transverse palmar crease on the right hand seen in both Patients 3 and 8. Lower limb anomalies included: 2,3-cutaneous syndactyly of the feet; sandal gaps and camptodactyly of the 2nd toe as seen in Patients 12 and 13; a longer 4th toe in Patient 6 and short toes in Patient 7.

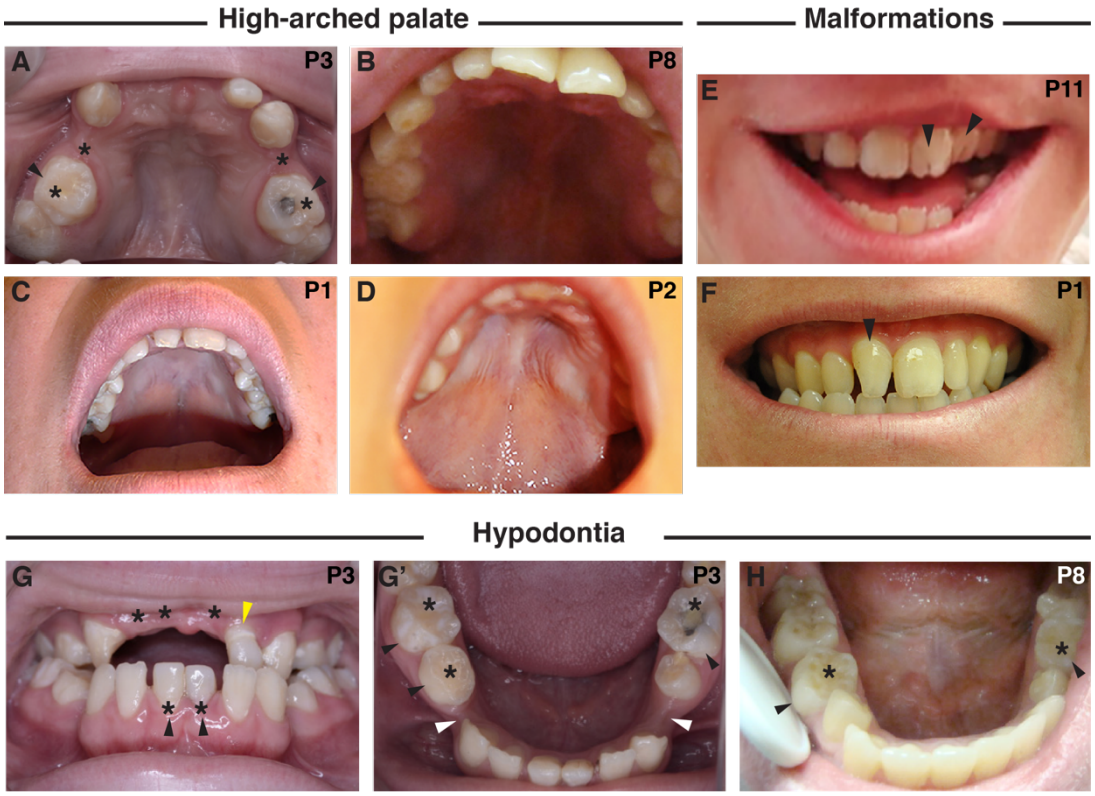


Figure 4-6 Dental manifestations and intra-oral phenotypes of patients with a CTNND1 mutation

[A-D] A high-arched palate was seen, shown are palates of Patients 1, 2, 3 and 8. [E-F] Abnormalities in the morphology of the dentition included: fissured incisors in Patient 11 [E, black arrowheads] and rotation of the incisors from the normal alignment shown in the non-cleft Patient 1 [F, black arrowhead]. [G-H] Hypodontia (tooth agenesis) was a common phenotype, indicated by the black asterisk. Black arrowheads indicate retained primary teeth. Patient 3 also has a diminutive upper left lateral incisor [G, yellow arrowhead] and wide inter-dental spacing [G', white arrowheads].

Table 4-3 Reported congenitally missing teeth

Patient ID	Missing teeth
Patient 1	*16*, 15, **23**, 25, *26*, *36*, 35, 45, *46*
Patient 2	54, 84 and **23**, *36*, 44
Patient 3	15, 14, 12, 11, 21, 24, 25, 35, 31, 41, 44, 45
Patient 5	**23**,25, 45
Patient 8	*16*, 15, **23**, 25, *26*, *36*, 35, 45
Patient 11	15, 35, 45
Patient 13	22, 35, 45

Missing permanent canines are in **bold** and missing permanent first molars are in *italics.*

Dental Orthopantograms

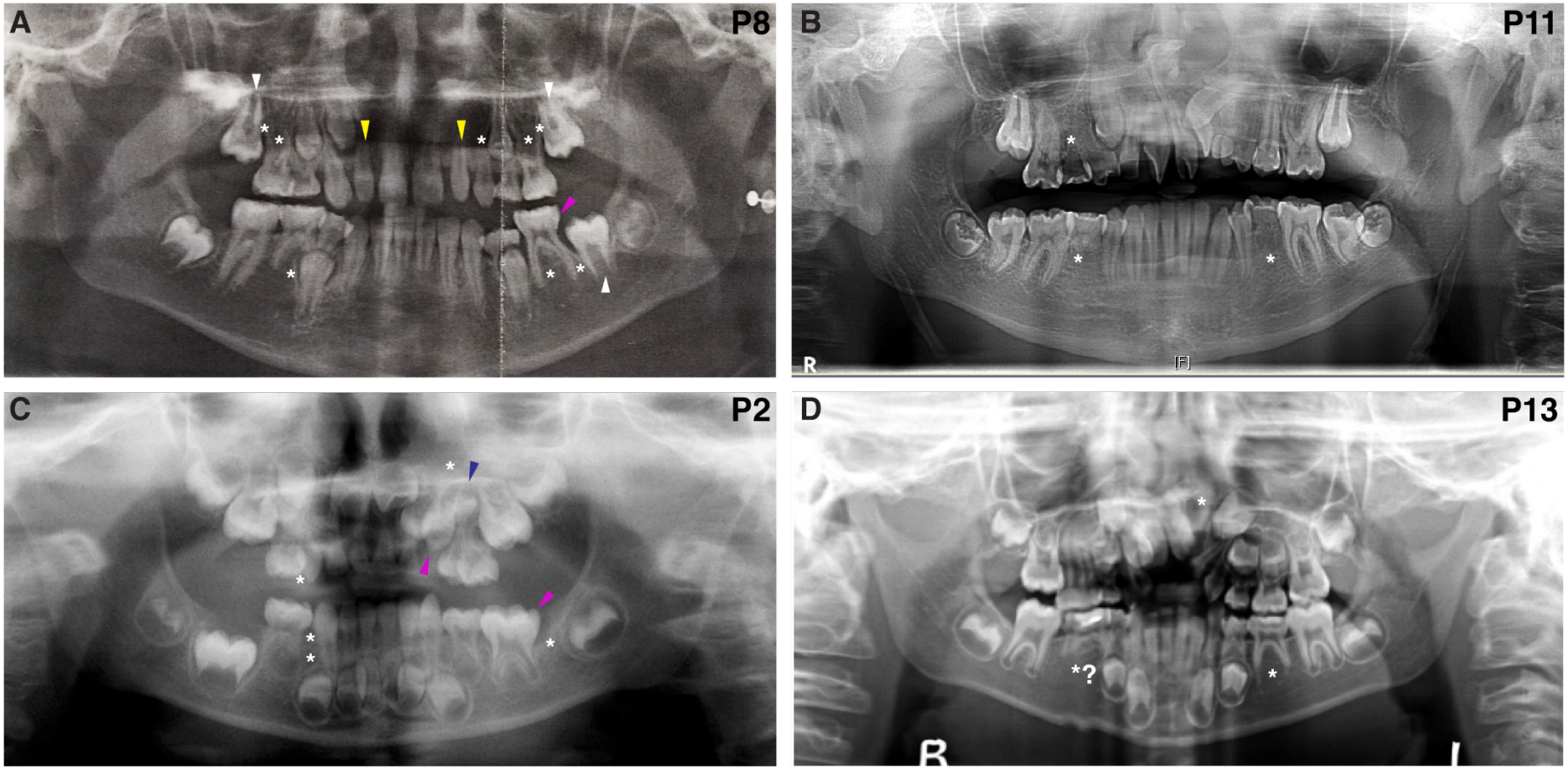


Figure 4-7 Dental radiographs of subjects with CTNND1 variants

[A-D] Dental orthopantograms (OPGs); missing teeth are indicated by white asterisks; diminutive teeth by yellow, macrodont teeth by magenta and supernumerary teeth by blue arrowheads, respectively. [A] OPG of Patient 8 at age 11, shows 8 missing permanent teeth (white asterisks) and shows the eruption of the second permanent molars (white arrowheads) in place of the missing first permanent molars. Also shown are diminutive upper right and left lateral incisors (peg-shaped) (yellow arrowheads), and a macrodont lower left second primary molar (magenta arrowhead). [B] OPG of Patient 11, at the age of 14, shows 3 missing permanent teeth (white asterisks), an ectopic maxillary left permanent canine and rotated maxillary centrals and left lateral incisors and dilacerated roots of the lower second permanent molars. [C] OPG of Patient 2, taken at 4 years, shows missing teeth including a missing lower left first permanent molar (white asterisks); a reported macrodont upper left primary canine (magenta arrowhead) with an underlying missing successor (white asterisk); a macrodont lower left second primary molar (magenta arrowhead) and a supernumerary tooth (blue arrowhead). [D] OPG for Patient 13, taken at 7.5 years, confirms absence of the upper left permanent lateral incisor and possibly the lower second permanent premolars.

4.3.3 **P120 is Expressed During Human Embryonic Development**

Although *P120* mRNA expression patterns have recently been documented during human and mouse palate development (Cox *et al.*, 2018), less is known about expression in the pharyngeal arch stages, which are likely to be important given the range of patient phenotypes. Therefore, I carried out mRNA *in situ* hybridization on human embryos using a probe that binds to all four C*TNND1* mRNA transcripts.

At Carnegie stage 13 (CS13), I found expression at multiple sites within the developing head, including the frontonasal processes, the forebrain, midbrain and rhombomeres **(Figure 4-8, B-C)**. Robust expression was also detected in the maxillary and mandibular processes of the first pharyngeal arch (PA1), the second and third pharyngeal arches (PA2 and PA3, respectively) as well as in the proximal domains of the upper and lower limb buds **(Figure 4-8, A-B)**. Signal was also weakly detected in the somites; however, strong expression was seen in the developing heart, trigeminal ganglion and the 10th cranial nerve **(Figure 4-8, A-B)**.

By Carnegie stage 21, *CTNND1* mRNA was expressed in the brain (data not shown), tooth bud **(Figure 4-8, E)**, the epithelial lining of the tongue and oral cavity and in the tongue mesenchyme **(Figure 4-8, D)**. Expression was particularly strong in the intrinsic muscles of the tongue: the superior longitudinal and transversal muscles and in the extrinsic genioglossus muscle **(Figure 4-8, D)**. Moreover, expression was evident in the dorsal epithelial lining of the developing palatal shelves **(Figure 4-8, F)**. In the heart, *P120* expression was found in cardiomyocytes of the ventricular wall and interventricular septum, in addition to strong expression in the endocardium **(Figure 4-8, G)**. Expression was also found in the intrinsic epithelial lining of the stomach wall; both in the pyloric part of the stomach and in the inner walls of the stomach body, the pancreatic islets, the germinal centre of the spleen, the epithelial lining of the bladder, hindgut and in the spinal cord and vertebral body **(Figure 4-9)**.

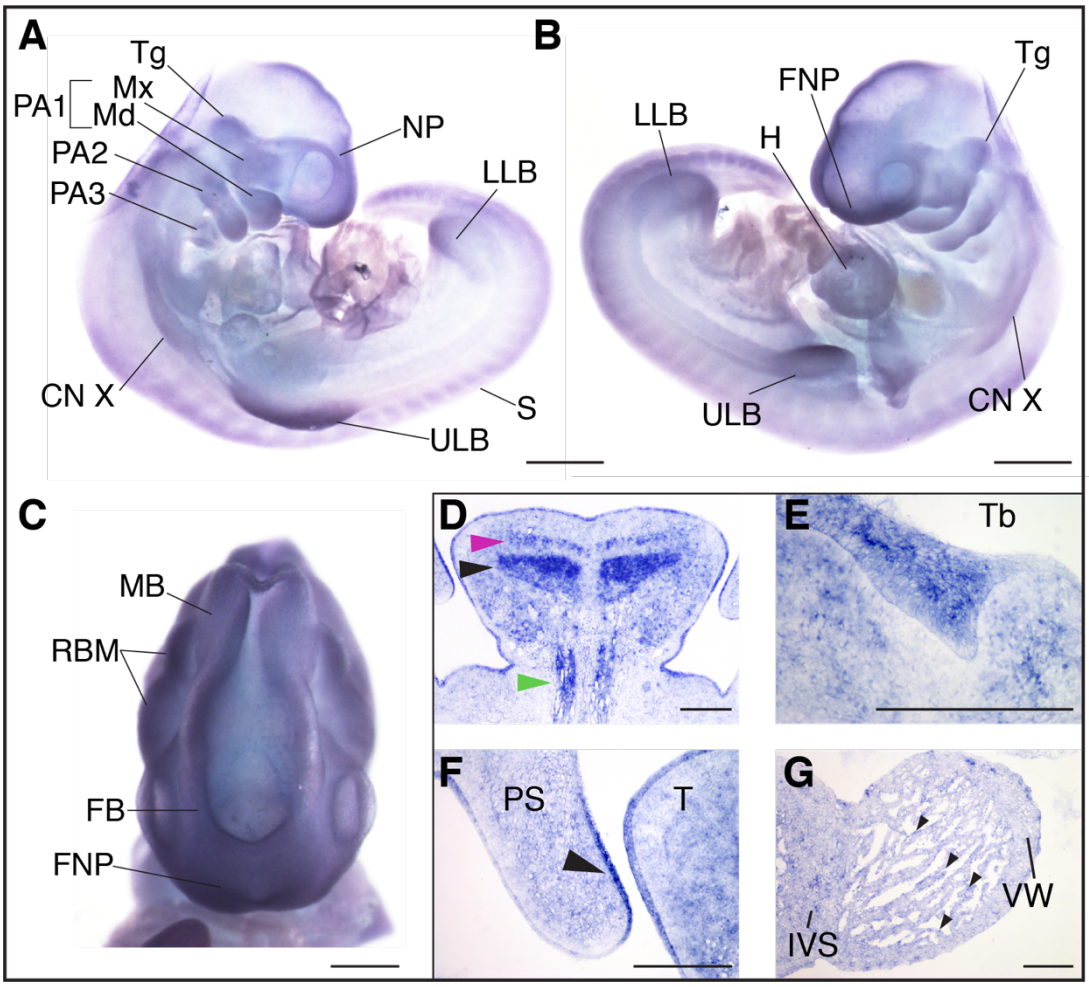


Figure 4-8 CTNND1 is expressed in the pharyngeal arches and craniofacial structures during early human embryonic development

CTNND1 mRNA *in situ* hybridization at human Carnegie stages 13 (CS13) [A-C] and 21 [D-G]. [A] Right lateral view of a CS13 human embryo, *CTNND1* mRNA is strongly expressed in the head in all three pharyngeal arches (PA1, PA2, PA3) and limb buds. Expression is specifically strong around the nasal placode and the maxillary and mandibular prominences. [B] Left lateral view, P120 is strongly expressed in the developing heart, frontonasal process, the trigeminal ganglion and the tenth cranial nerve. [C] P120 is ubiquitously expressed in the developing brain region in the rhombomeres, the forebrain and midbrain. [D-G] Coronal section through the head of a CS21 human embryo through a mid-palatal plane. [D] Strong expression is seen in the intrinsic muscles of the tongue: the superior longitudinal (magenta arrowhead), the transversal muscles of the tongue (black arrowhead) and the extrinsic genioglossus muscle (blue arrowhead). [E] *CTNND1* mRNA is strongly expressed in the epithelium of the developing tooth bud. [F] *CTNND1* is expressed on the dorsal epithelium of the palatal shelf (arrowhead) and in the epithelium of the tongue. [G] Expression is seen in the cardiomyocytes of the ventricular wall and the interventricular septum and in the cells of the endocardium (arrowhead). Scale bars = 100μm. Abbreviations: PA1, first pharyngeal arch; PA2, second pharyngeal arch; PA3, third pharyngeal arch; Tg, trigeminal ganglion; Mx, maxillary process; Md, mandibular process; CN X, tenth cranial nerve; ULB, upper limb bud; S, somites; LLB, lower limb bud; NP, nasal placode; H, heart, FNP, frontonasal process; Tb, mandibular tooth bud; PS, palatal shelf; T, tongue; IVS, interventricular septum; VW, ventricular wall.

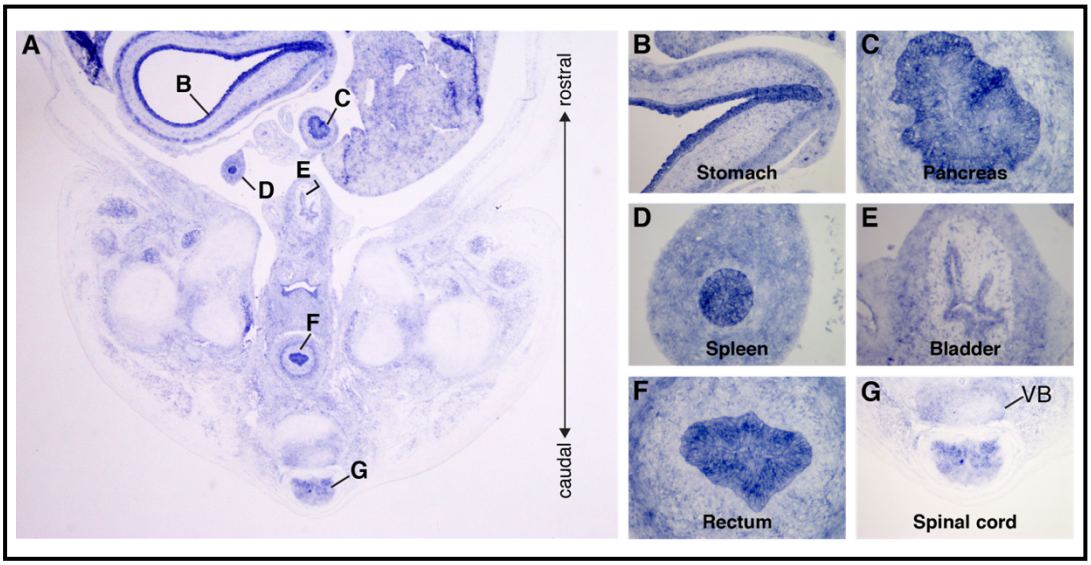


Figure 4-9 CTNND1 is expressed in other organs during human embryonic development

[A] Coronal cross-section through the torso at CS21. [B] *CTNND1* is expressed in the columnar epithelial lining of the stomach wall and continues through the pyloric part of the stomach. [C] Expression is seen in the islet of Langerhans in the pancreas. [D] Expression in the germinal center of the spleen. [E-G] Progressing caudally through the pelvis, *CTNND1* is expressed in the epithelial lining of the bladder [E], the rectum/hindgut [F], the spinal cord and vertebral body (VB) [G].

4.3.4 Expression of Phosphorylated P120-Catenin Predicts Fusion of the Palatal Seam

Because all of the study participants had either cleft palate or associated palatal anomalies, I also assessed p120-catenin expression during palatal fusion in the mouse, which occurs from embryonic day 12.5 (E12.5) to E15.5 **(Figure 4-10, A-D)**. To examine this, I used two antibodies recognizing phosphorylated forms of p120-catenin: a tyrosine-phosphorylated form, or phosphorylation at serine 268 (pS-268), which is proposed to trigger disruption of epithelial cadherin-catenin complexes (Vinyoles *et al.*, 2014; Xia *et al.*, 2003b). Neither of these forms of p120-catenin had been previously analysed in the palate. In palatal cross-sections at E14.5, the medial epithelial seam (MES) is evident **(Figure 4-10, B)**, followed a few hours later with dissolution of the seam at E14.75 **(Figure 4-10, C)**. While E-cadherin is expressed as expected in the MES (Sun *et al.*, 1998) **(Figure 4-10, F & J)**, the two forms of p120-catenin show very distinctive distributions. As the seam undergoes EMT, at E14.5, pS-268 is strongly expressed as predicted in cell-cell interfaces of the periderm layer along the medial seam, clearly co-localising with E-cadherin **(Figure 4-10, E-F)**. As the seam degrades, E-cadherin expression is lost while p120-catenin expression remains **(Figure 4-10, G-H, white arrowheads)**. To my surprise, I found phospho-tyrosine p120 staining in both the mesenchymal and the epithelial cells, with a clear enrichment marking the border between the epithelial and mesenchymal populations **(Figure 4-10, I-J, pink arrowheads)**. This distribution appears unique to this stage of palate formation consistent with reports that p120-catenin is tyrosine phosphorylate in an EGFR-dependent manner (Mariner *et al.*, 2004), and continues during degradation of the seam while E-cadherin expression decreases **(Figure 4-10, K-L, pink arrowheads)**. As a control, in earlier stages (E11-12.5), the phospho-tyrosine expression is much lower and nearly identical to the pS-268 staining (data not shown).

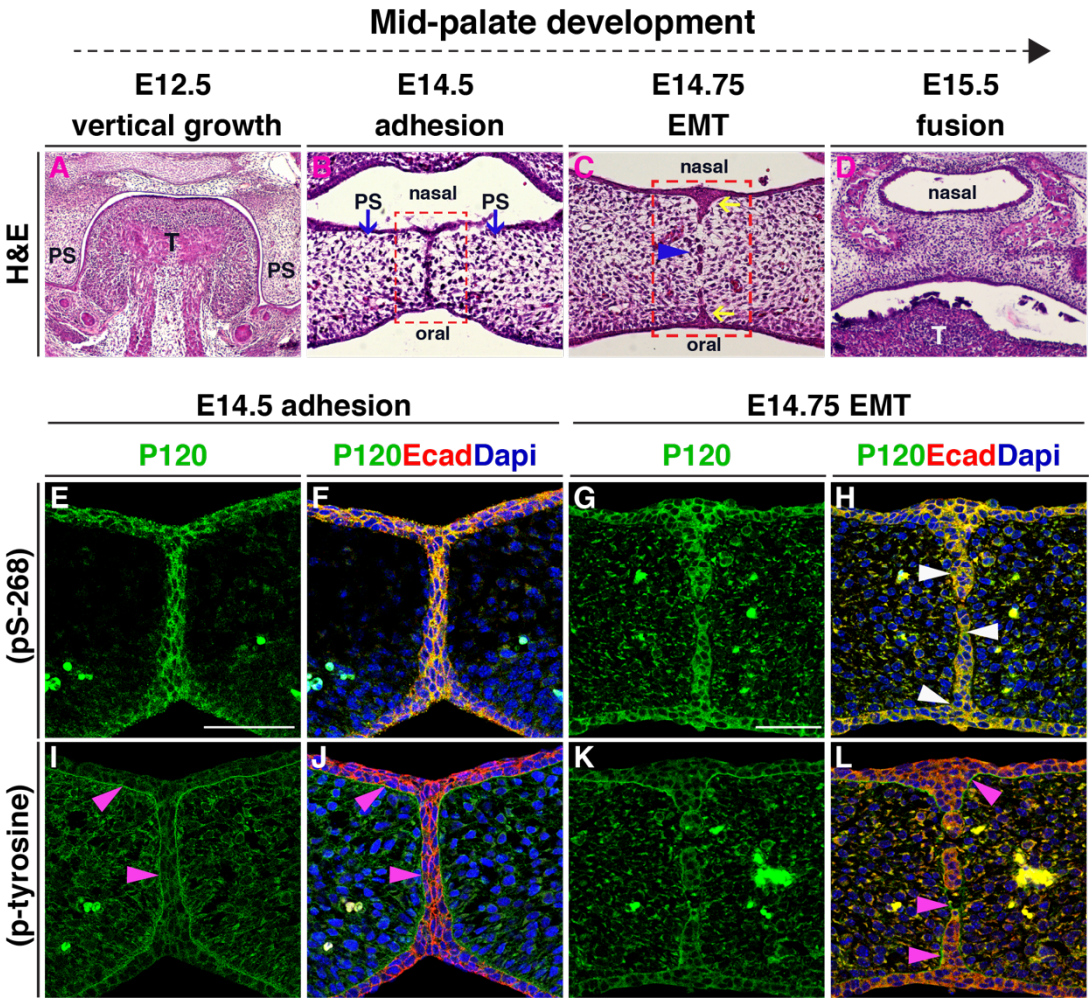


Figure 4-10 Expression of phosphorylated p120-catenin predicts fusion of the palatal seam

[A-L] All images are coronal sections of CD1 wild-type murine embryos at consecutive stages of palatal development. [A-D] Haematoxylin and eosin (H&E) staining illustrates successive stages of palatogenesis from embryonic day (E) 12.5 to E15.5. [B] At E14.5, following horizontal elevation, the opposing palatal shelves (blue arrows) meet and adhere to form the medial epithelial seam (MES). [C] EMT occurs at E14.75 when the MES breaks down, forming epithelial islands (blue arrowhead); the nasal and oral epithelial triangles form (yellow arrows). [D] At E15.5 palatal shelves are fused. Red box in [B] marks the regions shown in [E-F, I-J]. Red box in [C] marks the regions shown in [G-H, K-L]. [E-L] Immunofluorescent staining for either pS-268 or p-tyrosine p120-catenin antibodies (green) shown independently in [E, G, I, K], or in a merge with E-cadherin antibody staining (red) and DNA/DAPI stain (blue) [F, H, J, L]. [E-F, I-J] At E14.5, both forms of p120-catenin are expressed, with pS-268 strongly expressed in the periderm at the midline seam co-localizing with E-cadherin [E-F], while p-tyrosine clearly enriched in the area marking the border between the epithelial and mesenchymal populations [I-J, pink arrowheads]. [G-H, K-L] At E14.75, pS-268 p120-catenin is strongly expressed in the epithelial islands and the oral and nasal epithelial triangles; this is co-localised with E-cadherin during EMT and endocytosis while p120-catenin expression remains in some areas [H, white arrowheads]. In contrast, p-tyrosine p120-catenin expression surrounds E-cadherin positive epithelial islands, while E-cadherin expression has disappeared in the intervening mesenchymal cells (L, pink arrowheads). Scale bars = 50μm. Abbreviations: T, tongue; PS, palatal shelf.

4.3.5 Heterozygous Loss of P120-Catenin Leads to Structural Changes in the Laryngeal Apparatus and Velopharynx

Some of the study participants presented with anomalies associated with dysfunction of their velopharyngeal muscles and voice irregularities **(Table 4-2** and Appendix 5**)**, a phenotype described in patients with velocardiofacial syndrome (Fokstuen *et al.*, 1997; Leopold *et al.*, 2012; Miyamoto *et al.*, 2004). Antibody staining confirmed presence of p120-catenin protein during development of the laryngeal and pharyngeal tissues in the mouse **(Figure 4-11, A-B')**. I then examined the laryngeal structures of mutant mice compared to their littermate controls at E16.5, P1 and P2.5 **(Figure 4-12)**. To do this, a mouse carrying the ubiquitous *β-actin::cre* driver was crossed with *Ctnnd1*$^{fl/fl}$ mice in order to generate heterozygous mutants (Elder *et al.*, 1988; Shawlot *et al.*, 1998) **(Figure 4-12, C, H, M & R)**. Because the Liu Lab previously showed that the vocal ligaments originated from the neural crest (Tabler *et al.*, 2017), tissue-specific *Ctnnd1* heterozygotes using the neural crest specific driver, *Wnt1::cre* (Danielian *et al.*, 1998) was also generated for this study **(Figure 4-12, E, J & O)**. I found identical laryngeal anomalies in the heterozygous mutants in both mouse crosses, confirming the neural crest-specificity of these phenotypes.

Specifically, in control *Ctnnd1*$^{fl/+}$ mice, the palatopharyngeus (PLP) muscle, which elevates the larynx, is well defined and runs uniformly perpendicular to the epiglottis thereby attaching to the superior pharyngeal constrictor muscle (SPC) on either side **(Figure 4-12, A, B & D)**. On the other hand, the PLP and the SPC were both severely disorganized in both sets of heterozygous mice with an apparent increase in the cranio-caudal thickness of the PLP muscle **(Figure 4-12, C & E)**. Second, a striking phenotype known as laryngeal webbing was observed **(compare controls, Figure 4-12, G, I, Q to mutants Figure 4-12, H, J, R)**. Typically, the bilateral vocal cords are parallel and meet at the midline **(Figure 4-12, F-G, with inset schematized and shown in 4-12, P & Q)**. The outer layer of the vocal fold is made of an epithelium that encapsulates the lamina propria comprising the vocal ligaments **(Figure 4-12, P & Q)**. These two layers function as the vibratory components for phonation and oscillation. Instead, in heterozygous mutant mice, the vocal ligaments show only a brief contact point between the opposing epithelia **(Figure 4-12, H, with inset schematized and shown in 4-12, R & S)**. The vocal cords are also thinner, lacking the lamina propria **(Figure 4-12, R)**. Laryngeal webbing was also seen in the *Wnt1::cre* heterozygotes **(Figure 4-12, J)** compared to their littermate controls **(Figure 4-12, I)**.

While the vestibular folds were well demarcated and the ligaments within them clearly defined in controls **(Figure 4-12, G)**, the vestibular folds in the heterozygous mice were ectopically fused and the ligaments sparse and dispersed **(Figure 4-12, H)**. Caudally, where the vestibular folds surrounded the normal corniculate cartilage (COC) **(Figure 4-12, K & L)**; the folds have separated in the *Ctnnd1* heterozygotes, albeit hypoplastic **(Figure 4-12, M)**. Similarly, the COC appeared hypoplastic and devoid of the underlying lamina propria **(Figure 4-12, M)**. Finally, in mutants, the muscles were ectopically fused to the levator veli palatini muscles, which were then fused to the cranial base **(Figure 4-12, M)**. This, in turn, gave the impression of a high-arched epiglottal area; a defect also found in the *Wnt1::cre* heterozygous mutants **(Figure 4-12, O)**.

I also explored other craniofacial phenotypes in the heterozygous mouse model. Compared to their littermate controls **(Figure 4-13, A, a-e)**, mutant mice did not show any cleft lip **(Figure 4-13, A, f)**, face or limb dysmorphologies **(Figure 4-13, A, f-h)** or cleft palate **(Figure 4-13, A, i)** (n=12). This was confirmed by micro-computed tomography (μCT) to check for associated bony defects (n=6) **(Figure 4-13, A, j)**. This was also true for the *Wnt1::cre* heterozygotes **(Figure 4-13, B, m & n).** Further, since the heterozygous patients exhibited congenital heart disease and hypodontia, I also examined the heterozygous mice from both crosses for cardiac and tooth phenotypes using soft tissue and hard tissue μCT scans, respectively, in addition to histological sections. I did not observe any abnormalities within these tissues (data not shown).

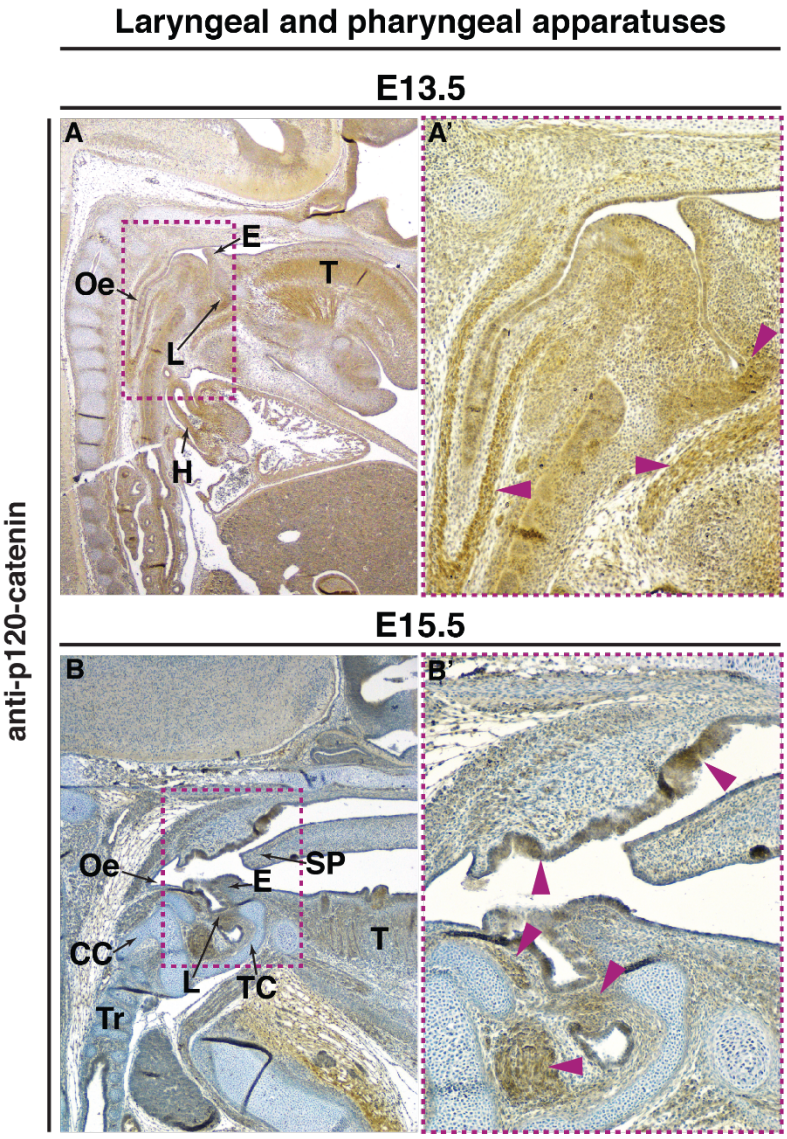
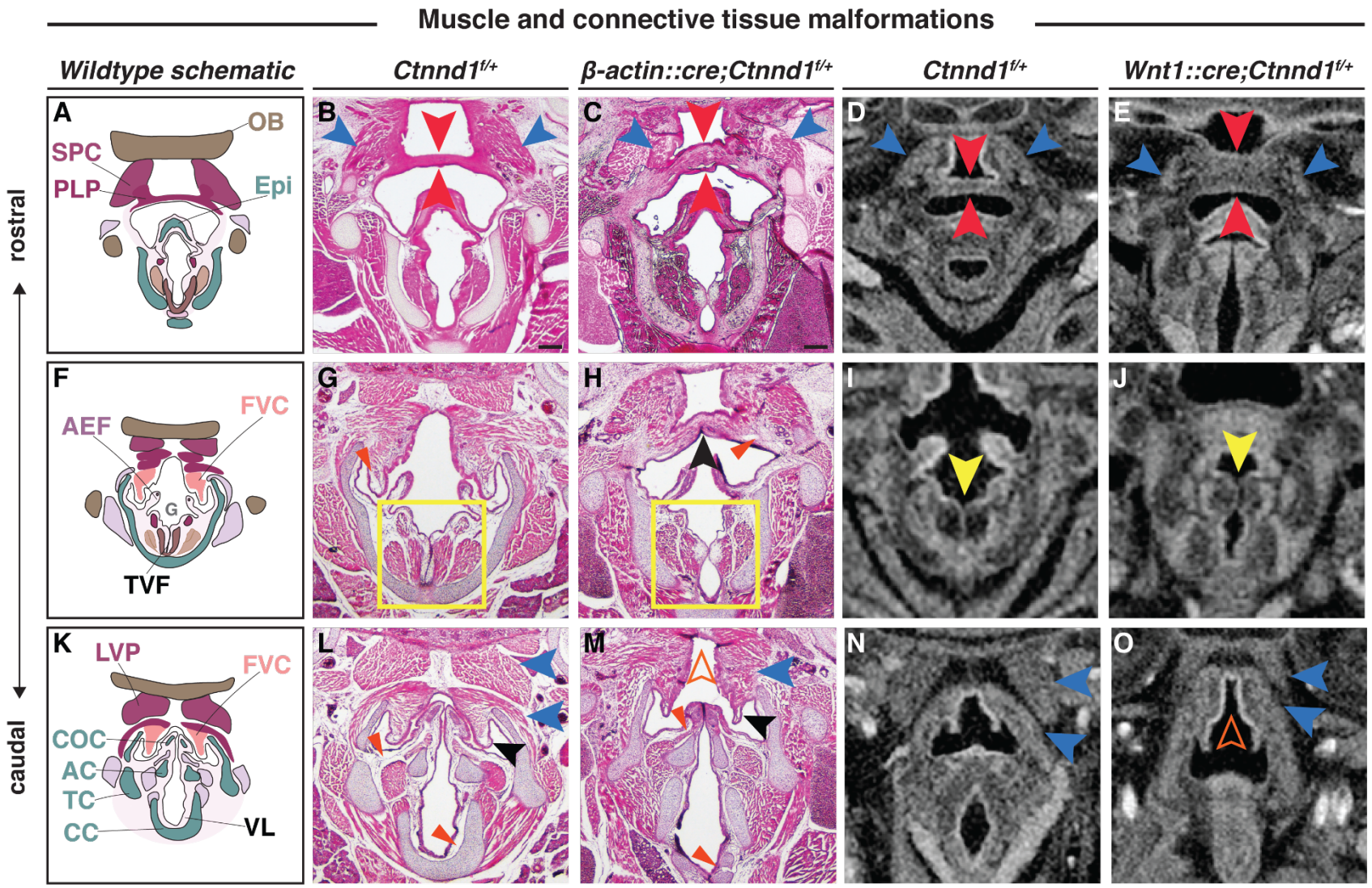


Figure 4-11 Mouse p120-catenin is expressed in the epithelial and mesenchymal compartments of the laryngeal and pharyngeal apparatuses

[A-B'] Immunohistochemistry using the anti-phosphotyrosine p120-catenin antibody on sagittal sections through wild-type mice at E13.5 (A-A') or E15.5 (B-B'). [A, B] Positive staining is seen in the epiglottis, esophagus and the larynx. (A', B') Insets from (A and B, respectively). Muscles that express p120-catenin in the laryngeal and pharyngeal apparatuses are shown (pink arrowheads). Abbreviations: E, epiglottis; Oe, entrance to oesophagus; L, laryngeal auditus; H, heart; SP, soft palate; TC, thyroid cartilage; CC, cricoid cartilage; T, tongue; Tr, trachea


Muscle and connective tissue malformations
Wildtype schematic
Ctnnd1f/+
β-actin::cre;Ctnnd1f/+
Ctnnd1f/+
Wnt1::cre;Ctnnd1f/+
rostral
caudal
A
OB
SPC
PLP
Epi
B
C
D
E
F
AEF
FVC
G
TVF
G
H
I
J
K
LVP
FVC
COC
AC
TC
CC
VL
L
M
N
O

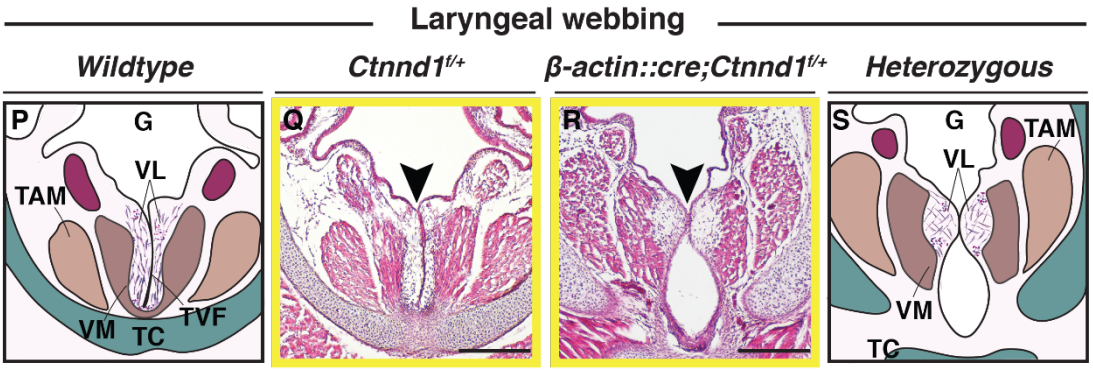


Figure 4-12 Heterozygous loss of p120-catenin leads to structural changes in the laryngeal apparatus and velopharynx

[A-O] (Page 151) ***Progression of the pharyngeal and laryngeal anomalies.*** [A, F, K] Schematics show the organization of the wildtype oropharynx from the more rostral (A) to caudal (K) planes. Haematoxylin and eosin (H&E) staining of coronal sections through control [B, G, L: *Ctnnd1$^{fl/+}$*] and heterozygous mutants [C, H, M: *β-actin::cre/+; Ctnnd1$^{fl/+}$*] littermate at postnatal stage (P1). [B-C] The SPC (blue arrowhead) and PLP (red arrowhead) in mutants are disorganized with an increased thickness in the PLP cranio-caudally [C] as compared to the controls [B]. [G-H] The FVC (vestibular folds) are well-defined in the controls with abundant ligaments [G, red arrowhead]. The FVC are fused in the mutant mice [H, black arrowhead] with ill-defined vestibular ligaments (H, red arrowhead). [L-M] The muscle attachments (blue arrowheads) superior to the FVC (black arrowhead) are well organized bilaterally in the controls surrounding the COC [L]. Caudally, when the FVC separated in the mutants it appeared hypoplastic (black arrowhead) as did the COC. The muscles (blue arrowheads) were ectopically fused to the LVP, producing an appearance of a 'high-arched' epiglottal area [M, orange hollow arrowhead]. **[D-E, I-J, N-O] (Page 151)** ***Neural crest specific mutants showed comparable laryngeal phenotype.*** Microcomputed tomographic (μCT) soft tissue scans of E16.5 control [D, I, N: *Ctnnd1$^{fl/+}$*] or neural-crest specific [E, J, O: *Wnt1::cre/+; Ctnnd1$^{fl/+}$*] heterozygous mutant littermates. [D-E] Compare the PLP in control [D] to the very thick PLP muscle seen in mutant [E, red arrowheads]. Compare the SPC in control [D] to the disorganized and hypoplastic SPC muscles seen in mutants [E, blue arrowheads]. [I-J] Laryngeal webbing was observed in mutant TVF [J, yellow arrowhead] compared to parallel TVF in control littermate [I, yellow arrowhead]. [N-O] Note aberrant muscle attachments (blue arrowheads) in [O] compared to control [N]. Control [N] epiglottal region compared to the high-arched epiglottal area observed in mutant littermate [O, orange hollow arrowhead]. **[P-S] (Page 152)** ***The laryngeal webbing phenotype.*** [P and S] Schematic representations of the wild-type [P] and mutant [S] anatomy at the vocal folds (TVF) from yellow-boxed insets in [G] and [H], respectively. [Q-R] H&E staining of coronal sections through control [Q: Ctnnd1$^{fl/+}$] and heterozygous mutant [R: *β–actin::cre/+;Ctnnd1$^{fl/+}$*] littermate at P1.[Q] In controls, well-defined vocal ligaments (VL) run parallel to the true vocal fold/cords (TVF). Underlying, the vocalis muscle (VM) and the thyroarytenoid muscle (TAM) are clearly attached and well-organised. [R] Laryngeal webbing is seen in the heterozygous mutant mice, where the vocal ligaments (VL) accumulate at a thin contact point (black arrowhead) thus perturbing the correct muscle attachments of the VM and TAM. Scale bars = 100μm. *Abbreviations: SPC, Superior Pharyngeal Constrictor; PLP, Palatopharyngeus Muscle; TAM, Thyroarytenoid Muscle; VM, Vocalis Muscle; LGF; HB, Hyoid Bone; Epi, Epiglottis; OB, Occipital Bone; LVP, Levator Veli Palatini Muscle; AEF, Aryepiglottic Fold; TVF, True Vocal Fold; VL, Vocal Ligament; FVC, False Vocal Cord; CC, Cricoid Cartilage; TC, Thyroid Cartilage; AC, Arytenoid Cartilage; COC, Corniculate Cartilage.*

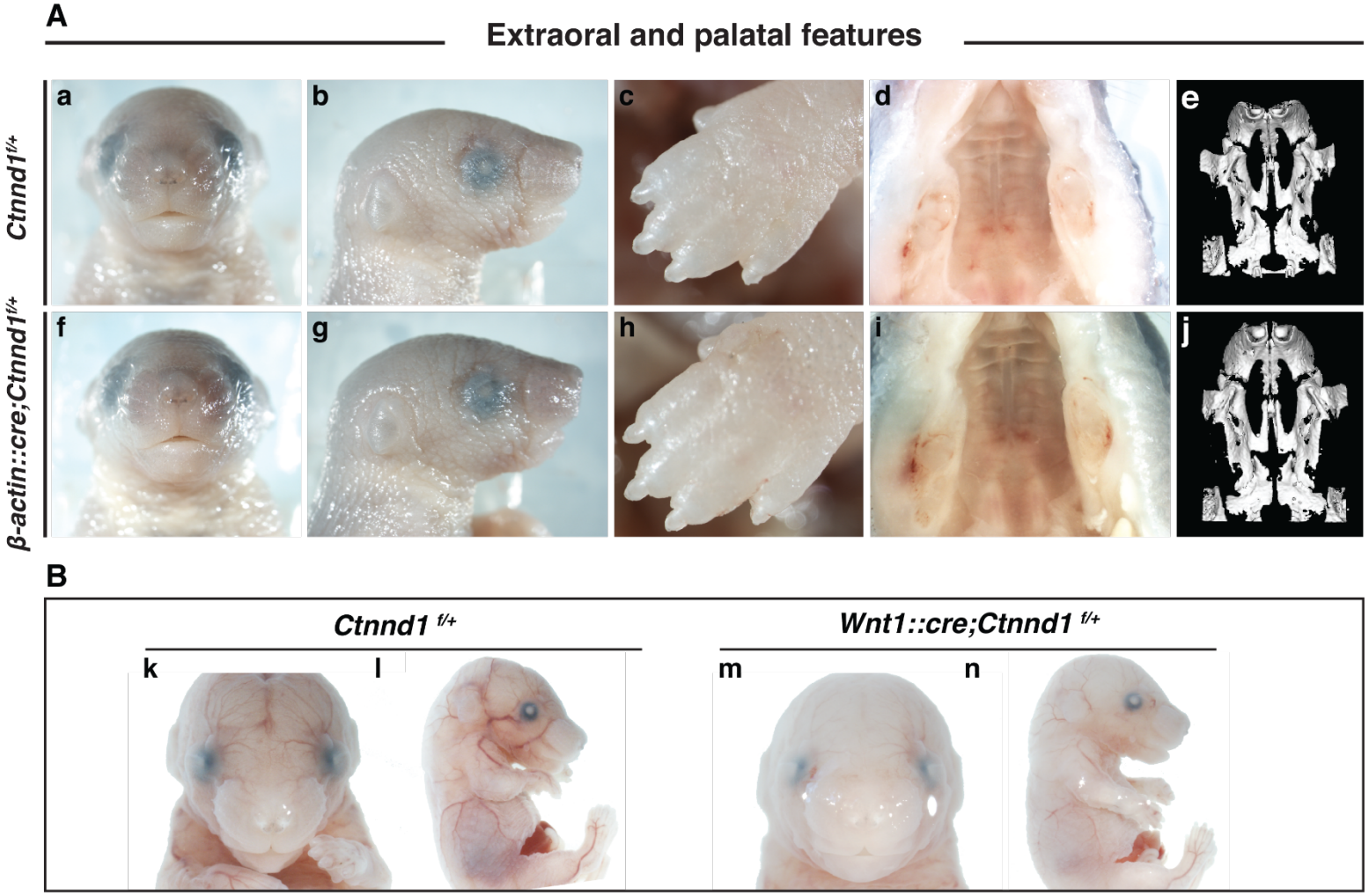

A
Extraoral and palatal features
Ctnnd1f/+
β-actin::cre;Ctnnd1f/+
a
b
c
d
e
f
g
h
i
j
B
Ctnnd1 f/+
Wnt1::cre;Ctnnd1 f/+
k
l
m
n

Figure 4-13 Heterozygosity in p120-catenin leads to normal facial and oral phenotypes

[A] Shown are postnatal P2.5 mice. Heterozygous mutant *β-actin::cre/+;Ctnnd1*$^{fl/+}$ mice do not exhibit facial or lip anomalies (f-g) and are comparable to littermate controls (a-b). (c, h) No limb anomalies are observed. (d, i) Postnatal P1 mice. Intra-oral views of the palate of wild-type (d) and heterozygous mutant littermate (i), cleft palate defects were not observed. (e, j) Microcomputed tomography (μCT) scans showed normal palates in P2.5 control (e) and heterozygous mutant littermate (j). [B] (k-n) Embryonic day (E)16.5 mice. Heterozygous mutant *Wnt1::cre/+;Ctnnd1*$^{fl/+}$mice do not exhibit facial or lip anomalies (m-n) compared with their littermate controls (k-l).

4.3.6 P120-Catenin Isoform 1 Function is Required in Multiple Organ Systems

While genetic mutation of *p120-catenin* in mouse models revealed a role for the neural crest in oropharyngeal development, analysis of multi-system involvement of p120-catenin was difficult due to embryonic lethality of the homozygous null mice (Davis & Reynolds, 2006; Oas *et al.*, 2010). I therefore turned to the frog *Xenopus,* where *in vivo* function of p120-catenin has been well studied (Ciesiolka *et al.*, 2004; Geis *et al.*, 1998; Paulson *et al.*, 1999). Previous analyses of p120-catenin requirements were mainly performed with antisense morpholino oligonucleotide (MO) knockdowns, which transiently prevent protein translation (Ciesiolka *et al.*, 2004). Instead, to create genetic mutants, CRISPR/Cas9 approaches were used for this study, allowing me to specifically delete different p120-catenin isoforms (Bhattacharya *et al.*, 2015). As noted in the introduction, isoform 1 (full length at 968 amino acids (aa)) is most abundant in mesenchymal cells, while isoform 3 (start at aa 102) is preferentially expressed in epithelial cells (Aho *et al.*, 2002; Hong *et al.*, 2016; Keirsebilck *et al.*, 1998; Montonen *et al.*, 2001). Isoforms 2 and 4, which start at 55 aa and 324 aa, respectively, are less well characterized.

Embryos were injected at the one cell stage with single guide RNAs (sgRNAs) targeting either of two coding exons, exon 3 or exon 7 (sgRNA1 and sgRNA2 respectively**, Figure 4-14, A**). Disruptions in exon 3 are predicted to only affect isoform 1, while sgRNA2 targeting exon 7 disrupt all four isoforms.

When embryos were scored at gastrula stages following sgRNA1 injections, disrupted or delayed blastopore closure was evident (n=30/42 vs. 2/30 in the controls) **(Figure 4-14, B)**. Furthermore, I noted severe early lethality **(Figure 4-15, A)**, especially using sgRNA2 which blocked all isoforms **(Figure 4-15, A)**. Notably, by neurula stages the majority of these mutants died due to a loss of integrity in the epithelium (data not shown).

Since the most well-established epithelial role for p120-catenin is in complex with E-cadherin at cell-cell junctions, I first examined E-cadherin localization in the neurectoderm at stage 11, as gastrulation was concluding. Indeed, in uninjected controls, high levels of p120-catenin and E-cadherin were found co-localized at the cell interface **(Figure 4-15, B, a-d)**. E-cadherin is expressed throughout the cell membrane **(Figure 4-15, B, b)**, whereas p120-catenin, though localized to the cell membrane, appears distributed in puncta **(Figure 4-15, B, a)**. Upon p120-catenin deletion, the

expression levels of endogenous E-cadherin in the epithelial cells was diminished particularly at the interface between the cells, leaving only spot-like localization of both proteins at the tricellular junctions of these epithelial cells **(Figure 4-15, B, e-h)**. The residual expression of p120-catenin may be due to maternal loading of the protein, as the CRISPRs should only affect zygotic transcription, or due to mosaicism of the CRISPR deletion.

As the sgRNA2 CRISPR was predicted to disrupt all four isoforms and led to severe lethality by neurula stages, the majority of analyses were performed using the sgRNA1 CRISPR, which is predicted to disrupt the predominantly mesenchymal isoform 1. A proportion of the knockout animals survived past the neurula stages, possibly due to mosaicism, and were examined at stage 46 to determine whether craniofacial and organ development had occurred normally. I observed obvious craniofacial defects in the CRISPR mutants **(Figure 4-16, A)**, including a reduction in the width and height of the head **(Figure 4-16, B, d-f)**, a hypoplastic mouth opening **(Figure 4-16, B, e)**, delayed breakdown of the cement gland **(Figure 4-16, B, d-e)** and heart and gut looping anomalies **(Figure 4-16, B, f)**.

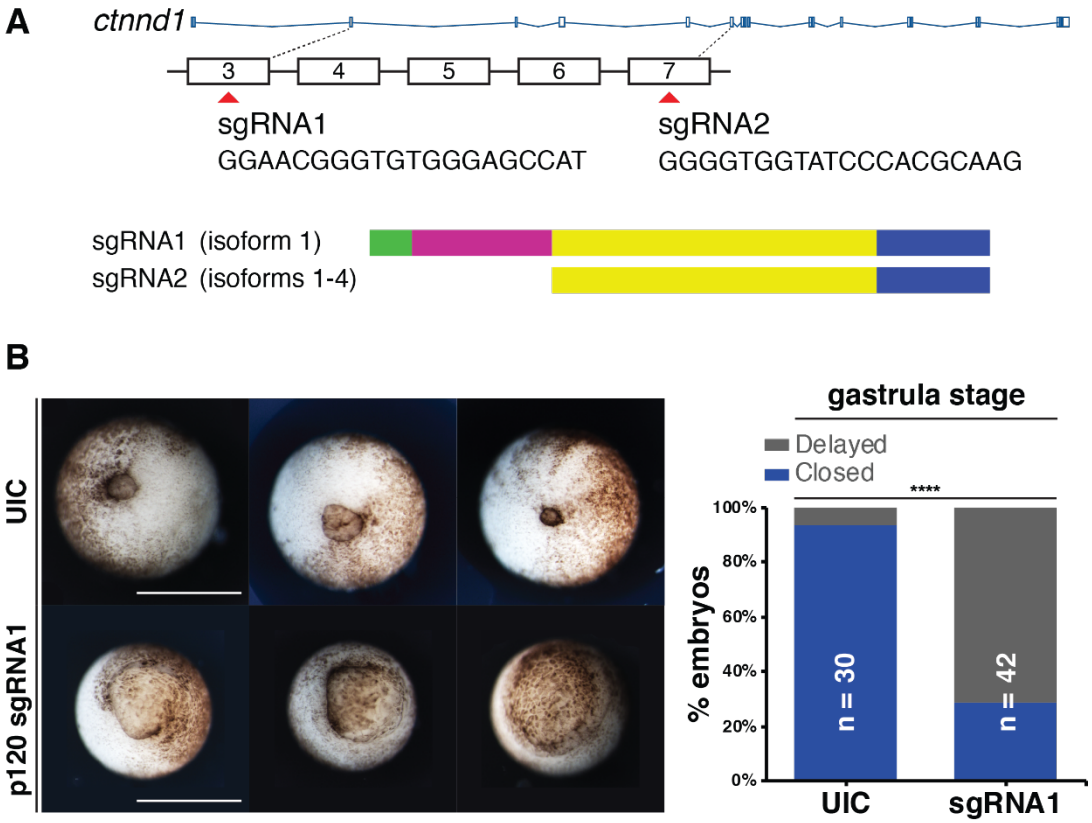


Figure 4-14 Ctnnd1 CRISPR/Cas9 delays blastopore closure

[A] Embryos were injected at the one cell stage with single guide RNAs (sgRNA), sgRNA1 and sgRNA2 targeting exons 3 and 7, respectively. A schematic representation of the p120 protein with its domains is shown (green, coiled-coil domain; pink, N-terminal phosphorylation domain; yellow, armadillo domain; blue, C-terminal region). [B] Ventral view showing blastopores at stage 11. Embryos injected with sgRNA1 had delayed blastopore closure (bottom row) compared to un-injected controls (UIC) (top row). The bar chart shows quantitation. Scale bars = 100μm. sgRNA, single guide RNA; UIC, un-injected control; ****p<0.0001.

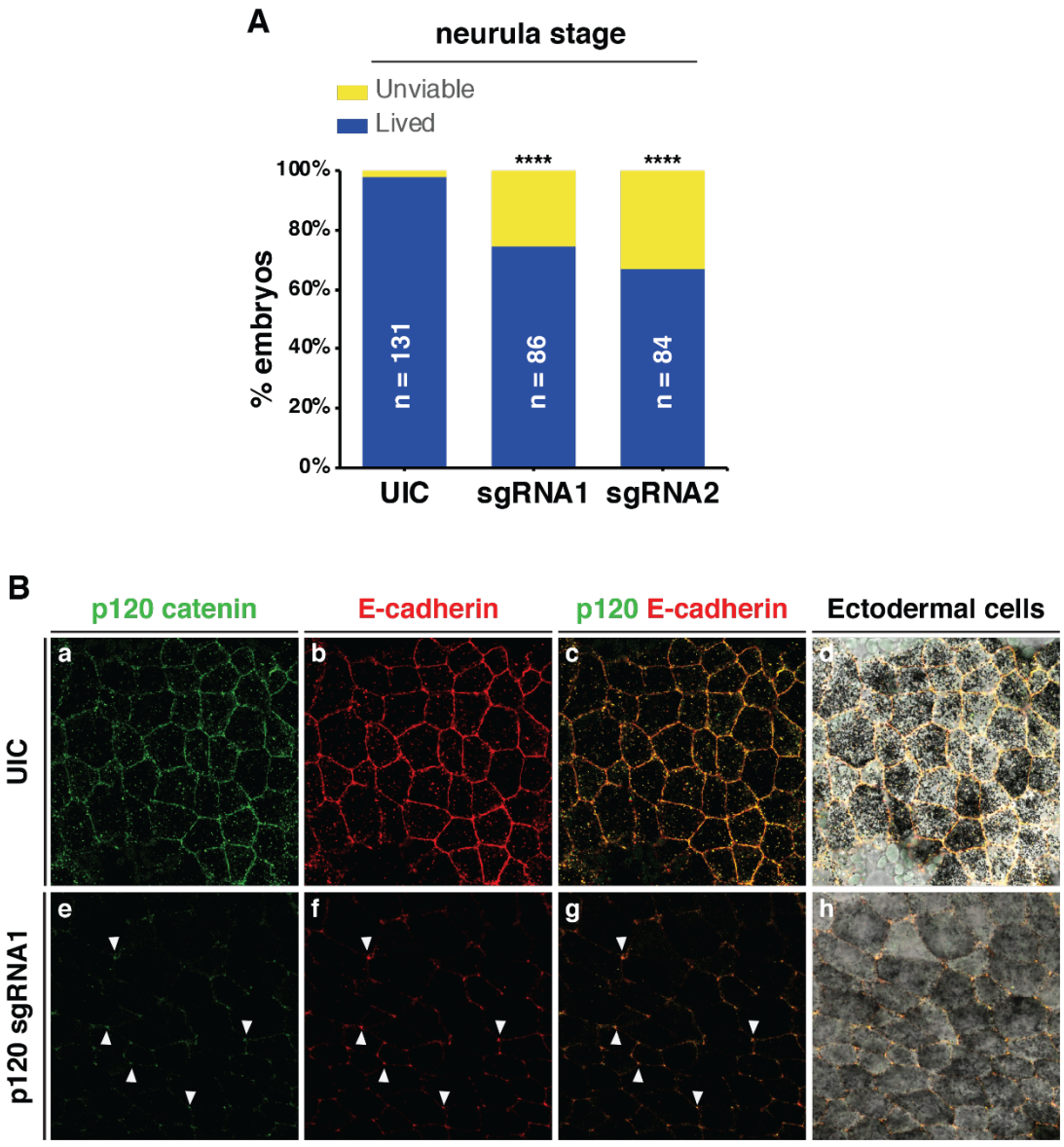


Figure 4-15 Ctnnd1 CRISPR/Cas9 leads to early lethality and loss of catenin-cadherin expression in ectodermal cells

[A] p120-catenin depletion led to lethality in embryos by the neurula stage. [B] Confocal sections through the apical surface of ectodermal cells at stage 11 of embryos injected with sgRNA1 (e-h) and UICs (a-d). (a-d) p120-catenin (a, green) is expressed in puncta at the cell membranes. E-cadherin (b, red) is expressed more evenly through the cell membranes. Both are colocalized at the cell-cell interface (c, d). Endogenous levels of p120-catenin and E-cadherin are diminished at the cell-cell interface in the sgRNA1-injected embryos (e-f). Residual p120-catenin and E-cadherin are seen in a spot-like pattern, only at the tricellular junctions (e-h, white arrowheads). Scale bars = 100µm. sgRNA, single guide RNA; UIC, un-injected control; ****p<0.0001.

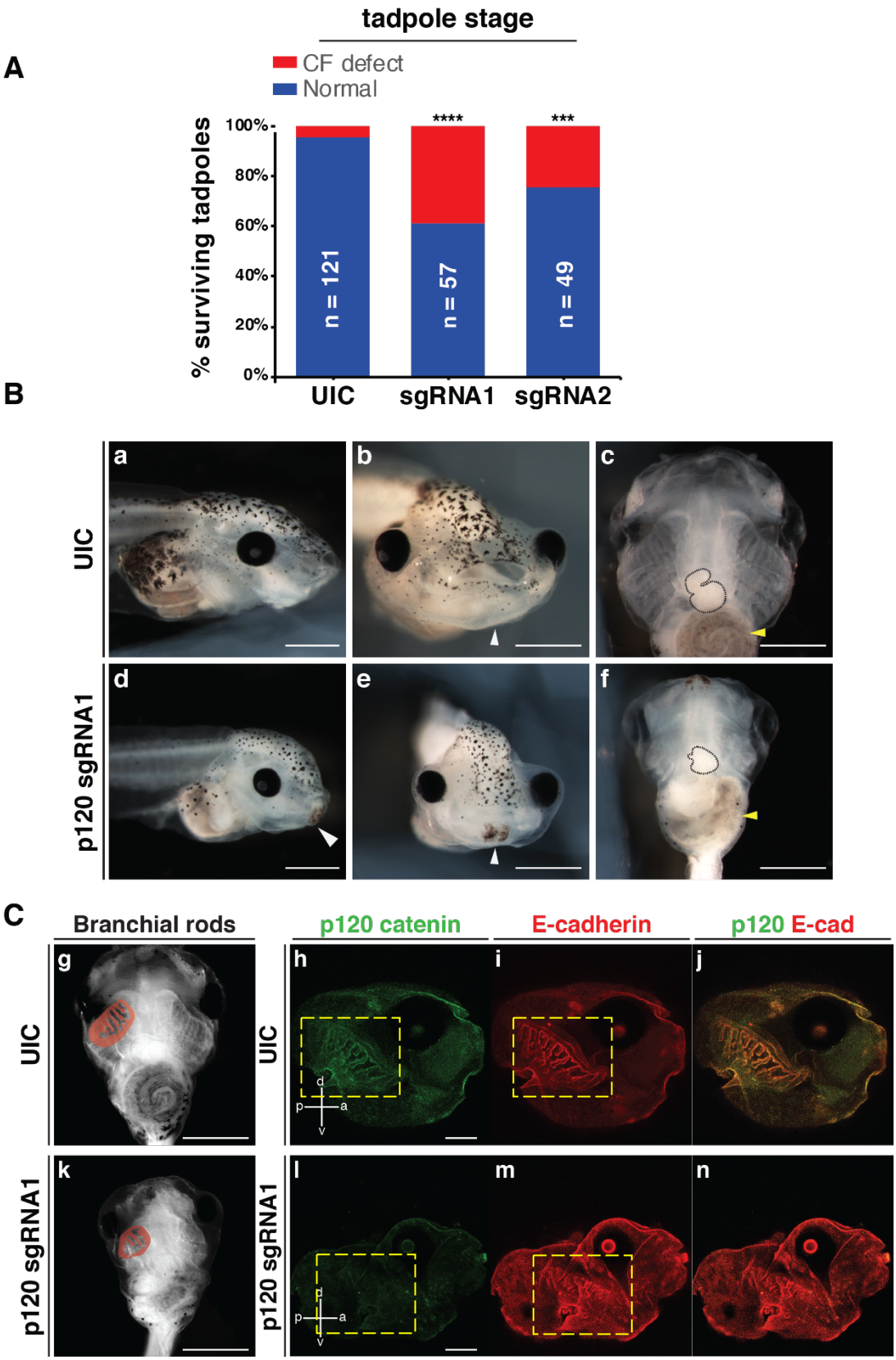

A
tadpole stage
CF defect
Normal
% surviving tadpoles
100%
80%
60%
40%
20%
0%
****
***
n = 121
n = 57
n = 49
UIC
sgRNA1
sgRNA2
B
UIC
p120 sgRNA1
a
b
c
d
e
f
C
Branchial rods
p120 catenin
E-cadherin
p120 E-cad
UIC
p120 sgRNA1
g
h
i
j
k
l
m
n
d
p
a
v

Figure 4-16 Ctnnd1 knockouts in Xenopus give rise to craniofacial and heart defects

[A-B] Stage 46 tadpoles. [A] Quantification of craniofacial defects in UIC and p120 depleted tadpoles. [B] (a, d) Lateral views show a flattened profile in *p120* CRISPR tadpoles (d) compared to UICs (a). (b, e) Frontal views showing a reduction in the size of mouth opening and a persistent cement gland (white arrowhead) in *p120* CRISPR tadpoles (e) compared to UICs (b). (c, f) Ventral views showing a reduction in the size of craniofacial cartilages, altered cardiac looping (black-dashed outline) and altered gut coiling (yellow arrowhead) in *p120* CRISPR tadpoles (f) compared to UICs (c). [C] (g) Normal branchial cartilages (marked over in red). (k) *Ctnnd1* knock-out perturbs the morphology of the branchial cartilages. (h-j) In the UIC, E-cadherin and p120-catenin are expressed in the rods of the branchial cartilages. (l-n) In the CRISPR tadpoles, the morphology of the rods is perturbed, and p120-catenin is no longer expressed in the rods (compare yellow boxes in h, i with l,m). Scale bars = 100μm. sgRNA, single guide RNA; UIC, un-injected control; ****p<0.0001; ***p<0.001.

Following on from the disorganization of the laryngeal muscles seen in the mouse mutants **(Figure 4-12)**, antibody staining against Pax2 was used to label the muscle fibres while anti-collagen 2 (col2) antibody labelled craniofacial cartilages in the mutants **(Figure 4-17, A, a-h)**. In control animals, the muscle fibres were well-organised and straight while in the mutants, the muscle morphology appeared disorganized, particularly the rectus abdominus muscle, with muscle striations being replaced by irregularly shaped fibres **(Figure 4-17, A, f-g)**. Consistent with previous observations **(Figure 4-16)**, craniofacial cartilages were hypomorphic and compacted both in the anterior-posterior and dorsal-ventral axes **(Figure 4-17, A, a & e)**. However, morphology of the chondrocytes appeared normal **(Figure 4-17, A, d & h)**. To my surprise, I also found that both p120 and E-cadherin were strongly expressed in the rods of the branchial cartilages in the control tadpoles **(Figure 4-16, C, h-j)**. The morphology of these rods was consistently perturbed in the CRISPR mutants **(Figure 4-16, C, k-n)**. These extensions were reduced in size, they lose the distinct streams that run parallel to one another, as in the controls **(Figure 4-16, C, g-j),** and are shortened and overlap haphazardly. In addition, most of these rods lose their p120-catenin and E-cadherin expressions **(Figure 4-16, C, l-n)**.

Finally, since the participants (6/13) had a high frequency of congenital heart defects and because I showed p120 to be strongly expressed in the heart of human, mouse and frog embryos, I examined the hearts in the CRISPR-knockout tadpoles. Notably, the strong expression of p120 seen in the different heart chambers in the control tadpoles was lost when p120 was knocked out **(Figure 4-17 B, p)**. The majority of mutant tadpoles had heart anomalies including heart-looping defects **(Figure 4-16, B, f & Figure 4-17, B, n)**. Notably, E-cadherin is not expressed in the normal heart or the muscles **(Figure 4-17, B, l)**, suggesting that the heart and muscle phenotypes may be manifestations of E-cadherin-independent functions of p120-catenin.

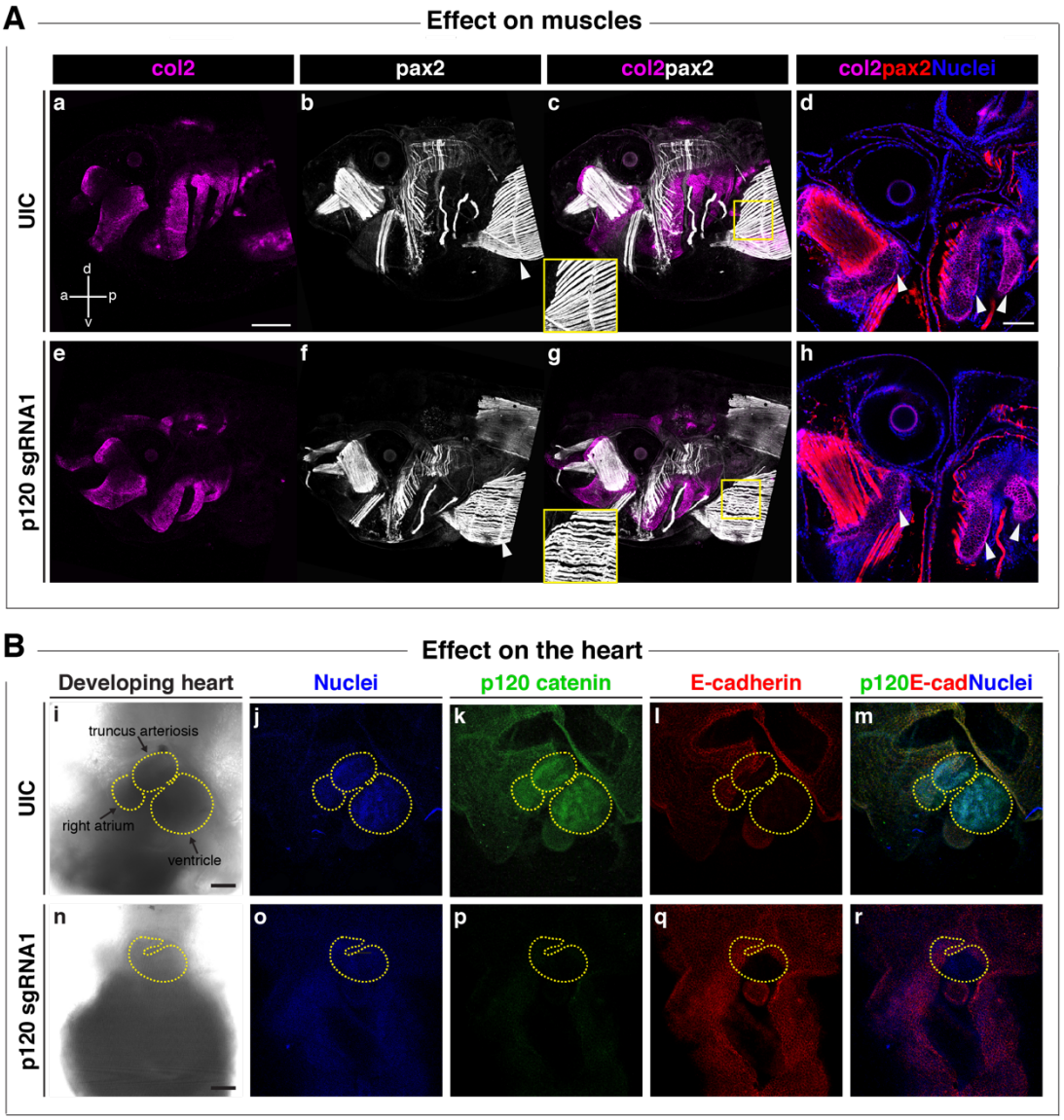


Figure 4-17 Ctnnd1 knockouts in Xenopus give rise to altered morphogenesis of the muscles and heart

[A] Immunofluorescent staining for collagen 2 (col2, magenta), muscle/pax2 (white) and nuclei (DAPI, blue); (a, anterior; p, posterior; d, dorsal; v, ventral). [A] (a, e) A lateral view of col2-positive branchial cartilages in UIC (a) and *p120* CRISPR mutant (e) reveals hypoplasia of mutant cartilages; however, cell morphology appears normal in *p120* CRISPR mutants (h) (d and h, white arrowheads). [A] (b-c, f-g) Pax2-expressing muscles revealed a defect in the fibril organization of the rectus abdominus muscle in the *p120* CRISPR tadpoles (f, white arrowhead) compared to the UIC muscles (b, white arrowhead); note insets in (c, g). [B] Ventral views of hearts of stage 46 tadpoles. Immunofluorescent staining for p120-catenin (green), E-cadherin (red) and DNA (blue). [B] (i-m) Controls; (n-r) *p120* CRISPR mutant tadpoles. Morphologic defects are evident in the size of the heart and directionality of the loops (compare control heart (i) to mutant heart (n), yellow-dashed outlines). [B] (k, p) p120-catenin is strongly expressed in the heart of UIC tadpoles (k) but is lost in *p120* CRISPR tadpoles (p). [B] (l, q) Note the absence of E-cadherin in the control and mutant hearts. Scale bars = 100μm.

4.4 DISCUSSION

4.4.1 The Phenotypic Expansion of *CTNND1*-Associated Anomalies

My work on this project expands upon the spectrum of abnormalities associated with *CTNND1* variants beyond non-syndromic cleft lip/palate (CLP) and blepharocheilodontic (BCD) syndrome (Cox *et al.*, 2018; Ghoumid *et al.*, 2017; Kievit *et al.*, 2018). Most notably, I described in detail characteristic craniofacial features including choanal atresia and unusual patterns of hypodontia as well as heart, limb, laryngeal and neurodevelopmental anomalies. I found expression of *CTNND1* mRNA during development of the pharyngeal arches in human embryos and I defined the profile of two phosphorylated forms of p120 in the mouse palate. Finally, genetic approaches in mouse and *Xenopus* demonstrated novel roles for *CTNND1* in the oropharynx, craniofacial cartilages and in the heart. Thus, the data I presented in this study implicate *CTNND1* variants as causative of a broad-spectrum syndrome that overlaps with DiGeorge velocardiofacial syndrome as well as other disorders of craniofacial development such as CHARGE and Burn McKeown syndromes (Corsten-Janssen *et al.*, 2013; Goos *et al.*, 2017; Vissers *et al.*, 2004; Wong *et al.*, 2015). All of these syndromes could be collectively considered to be neurocristopathies. Notably, the neural crest specific disruption of *CTNND1* in the animal models described here supports this role for *CTNND1* as a candidate neurocristopathy gene and I suggest that these newly identified variants likely highlight both epithelial and mesenchymal roles for p120-catenin.

Prior to this study, the majority of the participants I described did not have a recognizable or a diagnosed condition when they were seen and recruited by their respective clinical geneticists. Here, I demonstrated that they collectively shared consistent characteristic phenotypic features that suggest that mutations in *CTNND1* may lead to a much broader phenotypic spectrum than previously described (Ghoumid *et al.*, 2017; Kievit *et al.*, 2018). For instance, low set ears were reported in one case of BCD by Kievit and colleagues (Kievit *et al.*, 2018); here I found multiple participants with auricular anomalies particularly the low-set ears and over-folded helices **(Figure 4-5, B)**. Similarly, syndactyly was reported in one of the *CTNND1* patients described in (Ghoumid *et al.*, 2017), and clinodactyly (one patient) and camptodactyly (two patients) were reported by (Kievit *et al.*, 2018). Again, I found limb anomalies consistently associated with *CTNND1* variation **(Figure 4-5, C)**. The cardinal features of BCD include ectropion of the lower eyelids, eurybleparon and lagopthalmos (Ababneh *et al.*, 2014;

Lopes *et al.*, 2003); these were not evident. However, five of the patients showed other BCD-eyelid manifestations such as distichiasis and ankyloblepharon **(Figure 4-4; Table 4-2)**, I also found short up-slanting palpebral fissures, hooded eyelids, high arched eyebrows and telecanthus **(Figure 4-4, Figure 4-5, Table 4-2)**. As BCD is associated with both *CTNND1* and *CDH1* (E-cadherin) variants, some of these phenotypes may represent distinctive functions of the E-cadherin-p120 complex; the majority of these functions could be attributed to a role for the cadherin-catenin in epithelia (Hammond *et al.*, 2017).

Of note, eight individuals had severe hypodontia, including missing permanent canines and first permanent molars, even in those without cleft lip/palate. While the missing permanent cuspids (canine teeth) in patients with CLP is common, the occurrence of missing cuspids in those who do not have a CLP phenotype is thought to be very rare. The absence of first permanent molars is also a rare occurrence (Abe *et al.*, 2010); therefore, to find this pattern in three individuals in the current cohort in two unrelated families was also surprising **(Figure 4-7, A & C; Table 4-3).** Interestingly, individuals with missing molars also had missing cuspids. Furthermore, one of these patients was reported to have surgical exposure of a macrodont maxillary left primary canine with a missing successor tooth; there was additional history of missing cuspids in their family. Finally, I also noted a patient with missing permanent maxillary and mandibular central incisors. Syndromes with missing lower permanent incisors include velocardiofacial syndrome, Down syndrome and Kallmann syndrome (HeliÖvaara *et al.*, 2011). Thus, missing canines and molars could be classified as a microform cleft anomaly, especially when found in association with high-arched palates.

Beyond the known phenotypes associated with *CTNND1* and *CDH1*, I noted the novel phenotypes seen in the patients, which included the heart anomalies and behavioural disorders. These have not been reported previously in patients with a BCD diagnosis. Nevertheless, my findings suggest that both *CTNND1* and *CDH1* should be tested in patients with congenital orofacial and cardiac anomalies. A key finding was choanal atresia in four individuals; given the rarity of this anomaly, both *CTNND1* and *CDH1* should be considered during genetic profiling of patients with this anomaly, in addition to CHARGE and other syndromes noted above. Indeed, Nishi *et al.* (2016) reported cleft lip, right choanal atresia, a congenital cardiac anomaly (tetralogy of Fallot), agenesis of the corpus callosum, upslanted palpebral fissures and ear anomalies in a patient with a *CDH1* mutation; however, at the time, this was not diagnosed as BCD.

4.4.2 **Novel Variants and Functions in *CTNND1***

While all of the variants found in the present study resulted in truncations of p120-catenin, they fell broadly into three distinct groups: those falling within the N-terminal regulatory region (p.Val148Aspfs*24), those disrupting the armadillo repeat region and presumably subsequent interactions with E-cadherin (e.g., p.Arg461*, p.Leu494Argfs*5, p.GLy532Alafs*6 and p.Arg797*), and those falling in the C-terminal domain (p.Ser868*, the splice variant c.2702-5A>G and p.His913Profs*3). Interestingly, those probands with C-terminal truncations had the most complete cleft lip and palate phenotypes. Interestingly, this complete form of palatal anomaly was consistent with patients with C-terminal variants described in previous reports by Kievit *et al.* (2018) who reported a nonsense mutation (p.Trp830*) and Cox *et al.* (2018) who reported p.Arg852* and a splice site mutation (c.2417+G>T). As these C-terminal truncations would all be predicted to retain E-cadherin binding, but lose crucial RhoGAP interactions (Schackmann *et al.*, 2013), one might hypothesize that a mutation in this region prevents p120-catenin clearing from the epithelial complex, which is necessary for seam dissolution during palate closure. Therefore, future analyses should focus on whether these C-terminal truncations are acting in a dominant-negative manner and preventing clearance of E-cadherin from the seam. Interestingly, clefts of the lip with or without cleft palate and isolated cleft palate have been thought to be developmentally and genetically distinct (Mossey and Castilla, 2003, WHO; Shaw *et al.,* 2004), yet, I have shown that a single-gene disorder can encompass both clefting phenotypes.

With regards to non-epithelial functions of p120, some of the phenotypes that this study, and others, have reported, could be explained by the known interactions of p120-catenin in the Wnt signalling pathway (Park *et al.*, 2005). Epithelial-specific knockouts of p120 (using a *keratin-14* promoter) did not show tooth agenesis (Bartlett *et al.*, 2010), suggesting that the tooth anomalies in the patients described in my study do not arise from the epithelial functions of p120. In support of this, two key genes implicated in tooth development and in tooth agenesis are the Wnt ligand, *Wnt10A* and a Wnt target gene *Axin2* (Callahan *et al.*, 2009; Lammi *et al.*, 2004; Laurikkala *et al.*, 2001; Liu *et al.*, 2008; Lohi *et al.*, 2010; Mostowska *et al.*, 2006; Mostowska *et al.*, 2013; Mues *et al.*, 2014; Song *et al.*, 2014; van den Boogaard *et al.*, 2012; Wang *et al.*, 2014). The Wnt signalling pathway may also explain the laryngeal findings **(Figure 4-12)**, as knockout of the Wnt transducer β-catenin has also recently been shown to lead to similar vocal fold anomalies (Lungova *et al.*, 2018) as those seen in our neural crest specific *p120-catenin* heterozygotes **(Figure 4-12)**. Furthermore, knockout of the mesenchymal form

of p120 (isoform 1) in *Xenopus* **(Figure 4-14 to Figure 4-17)**, confirm prior studies on p120-catenin in the neural crest, where the p120-catenin association with Wnt signalling is well-established (Hatzfeld, 2005; Kim *et al.*, 2004). Thus, I hypothesise that a subset of p120 phenotypes can also be attributed to Wnt perturbation in the neural crest **(Figure 4-18)**. The heart defects seen in the study patients could also be attributed to a failure in neural crest development, which is known to be crucial for development of the septum and valves (Buckingham *et al.*, 2005; Eley *et al.*, 2018; Kochilas *et al.*, 2002; Peterson *et al.*, 2018; Srivastava *et al.*, 1997).

4.4.3 Additional *CTNND1* Phenotypes

In addition to the phenotypes shared commonly across the study cohort, some participants had scoliosis, and one family reported two deceased children, who had bifid uvula, congenital cardiac disease (VSD, PDS), eye anomalies, developmental delay and chronic bowel immotility and gastroesophageal reflux disease; however, no genetic testing had been carried out. One patient presented at a young age with an ovarian dysgerminoma. To my knowledge, this is the first patient with a *CTNND1* variant associated with an early onset cancer, though p120 has been associated with cancer and tumorigenesis (Lehman *et al.*, 2015; Reynolds & Roczniak-Ferguson, 2004; Schackmann *et al.*, 2013; Smalley-Freed *et al.*, 2010; Stairs *et al.*, 2011). Although additional functional analysis will be necessary, this may perhaps be important for counselling. Finally, a number of patients reported in DECIPHER have copy number variants (CNV) affecting *CTNND1* (data not shown). Interestingly, both deletions and duplications have been associated with partially overlapping phenotypes. For instance, two patients with a deletion of less than 4Mb had anomalies including bulbous nose, limb anomalies, delayed speech and language development, intellectual disability, nasal speech, ventricular septal defect, and cleft lip. To my knowledge, copy number variants in the 11q11 region encompassing *CTNND1* have not been reported to date.

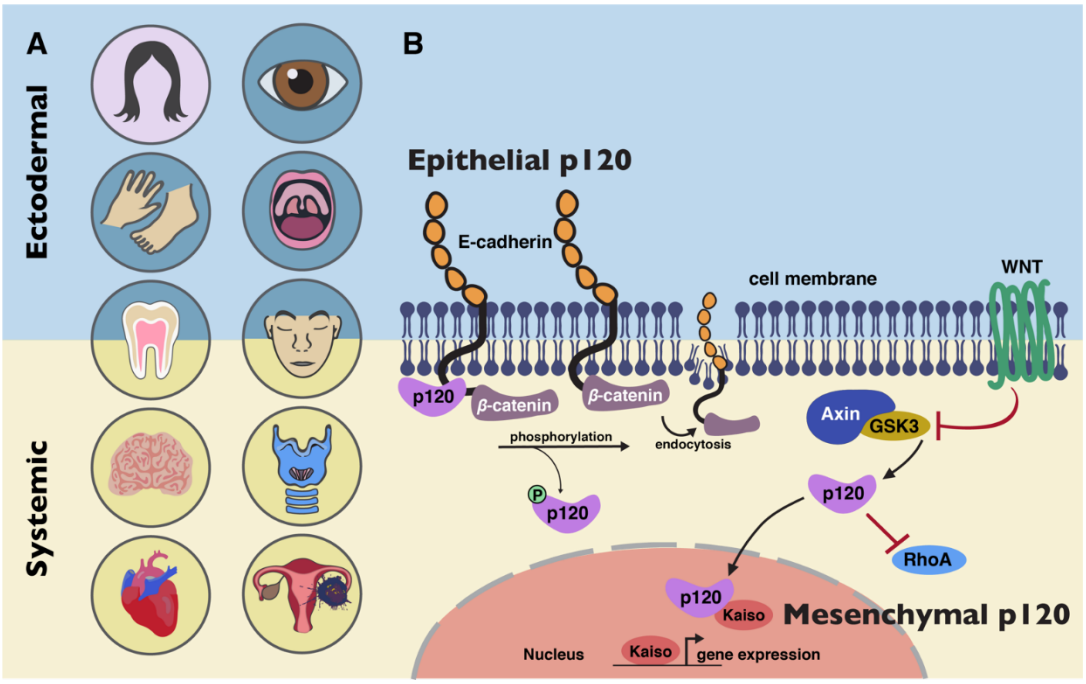


Figure 4-18 Model of CTNND1 function in systemic disease

[A] *CTNND1* mutations are not only implicated in conditions that affect epithelial structures but also systemic conditions that originate from mesenchymal roles of p120-catenin. Structures in pink circles have been described in previous publications on *CTNND1* (Ghoumid *et al.*, 2017; Kievit *et al.*, 2018); structures in blue circles have been implicated previously in *CTNND1*-related disorders (Ghoumid *et al.*, 2017; Kievit *et al.*, 2018) and in this study; structures in yellow circles have been identified in this study. [B] Blepharocheilodontic syndrome (BCD) is primarily due to disturbances in E-cadherin/p120 interactions. The inclusion of other organ systems described here highlights the involvement of other known molecular functions of p120, such as its role in the WNT signaling pathway and its interactions with Rho-GTPases, demonstrating its mesenchymal roles in producing these systemic conditions.

4.5 OUTSTANDING QUESTIONS

4.5.1 Functional Assays with the Human *CTNND1* Variants

Because p120-catenin has numerous binding partners and because the human variants found in this study span the various protein domains, the next step is to ask, how are these mutations affecting p120-catenin function and in what context? To assess the effects of the human variants on p120-catenin function, I am planning to perform *in vivo* and *in vitro* assays.

Since I demonstrated that the CRISPR/Cas9 experiments worked successfully and showed interesting craniofacial and cardiac phenotypes in the injected tadpoles compared to controls, microinjection of mRNA of the human *CTNND1* variants will be used to carry out *in vivo* assays in *Xenopus* to test a) whether they produce phenotypes similar to the ones I reported using the CRISPR/Cas9 knock-outs particularly craniofacial and cardiac anomalies, and b) to test whether injecting wild-type and/or mutant *CTNND1* variants will rescue the phenotypes observed in the p120-CRISPR'd tadpoles.

Moreover, since *in vitro* assays of p120-catenin are well established (Aho *et al.*, 2002; Anastasiadis *et al.*, 2000; Davis *et al.*, 2003; Seidel *et al.*, 2004), p120-catenin constructs that contain various deleted regions of the p120 protein such as Armadillo repeat 3 or the N-terminal domain will be tested. By transfecting these constructs in various cell lines, I am planning to carry out assays using the wild-type and human *CTNND1* variants to explore the nature of these mutations and the mechanism by which they alter p120-catenin function and activity with its binding partners, including E-cadherin and RhoA.

As briefly described in Chapter 2, I carried out site-directed mutagenesis on a wild-type human p120-ORF, thus mutagenizing and producing two constructs, one specific for the (p.Arg461*) variant and the other for the (p.Gly532Alafs*6) variant found in this study. Further optimisation of these constructs and the synthesis of constructs for the other human mutations described in this study are underway.

4.5.2 Variants in the *DMXL2* Gene in Patient 3

Patient 3, for whom the project was based on, had a sibling who also presented with autism spectrum and missing permanent teeth, albeit a less severe form of hypodontia.

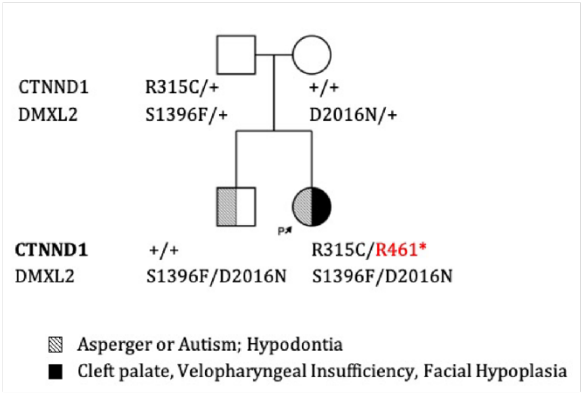
In addition to the *CTNND1* mutations, the proband and her brother were both found to be heterozygous for rare *DMXL2* gene variants **(Figure 4-19, A-B)**.

Recent research has shown that *Dmxl2* is involved in the maturation of neurons in the adult mouse brain and is expressed on synaptic vesicles. Mice with a homozygous deletion of *Dmxl2* were embryonic lethal. Furthermore, conditional heterozygous deletion of *Dmxl2* in adult mice resulted in delayed puberty as well as very low fertility. This functional validation stemmed from a study by Tata *et al.* (2014) where the authors identified a 15-bp (c.5824_5838del) in-frame deletion in exon 24 of *DMXL2* in a consanguineous family in which (3/5) of the siblings had a diagnosis of polyendocrine-polyneuropathy syndrome [MIM: # 616113]. This variant resulted in the removal of five amino acids (p.1942_1946del) including a highly conserved Serine1942 (Tata *et al.*, 2014). How *Dmxl2* alters neuronal and metabolic development in mice is unknown, however, the *DMXL2* gene has been recently described as a novel player in Notch signalling, regulating the acidification of intracellular compartments through the vacuolar protonic pump (V-ATPase) both in Drosophila and in mammalian systems (Faronato *et al.*, 2015; Sethi *et al.*, 2010). The role of Notch signalling in the tooth (Cai *et al.*, 2011; Mitsiadis *et al.*, 2005) and the brain (Ables *et al.*, 2011; Zhang *et al.*, 2018) is well established. An integral step in Notch signalling is γ-secretase mediated Notch cleavage of the Notch intracellular domain (NICD), a step which requires functional V-ATPase. *DMXL2*, through formation of a complex with *DMXL1*, regulates V-ATPase function. Therefore, DXML2 could be considered an upstream regulator of Notch signalling.

To examine whether *DMXL2* was indeed expressed during mammalian embryonic development, I carried out expression assays using the DMXL2 antibody on wildtype mouse embryos at different stages of development. Preliminary immunohistochemistry results suggest that the DMXL2 protein is expressed in the brain and teeth at various stages during development (data not shown). Whether the phenotypes seen in the proband with the *CTNND1* and *DMXL2* mutations are the result of a polygenic effect from the variants in these genes is a question that remains unanswered. Indeed, autism has been shown to be a polygenic disorder (Chaste *et al.*, 2017; Ronemus *et al.*, 2014). Most notably, of all the probands with *CTNND1* mutations presented in this chapter, this proband had the most severe form of hypodontia compared to others. Could the added genetic risk alleles from *DMXL2* augment the severity of the hypodontia phenotype observed? With further advancements on our knowledge of data generated from exome sequencing and how multiple variants could

influence multiple phenotypes is a complex area that is undergoing development, and future research will enhance our understanding on such impacts in disease.

A



B

Position	Ref/ mut	Gene	PLI	RVIS	GDI	Variant	1KG	MAF ExAC_All	CADD	SIFT	PPH	Father	Mother	Brother
15:51773257	C>T	DMXL2	1.0	0.49%	9.5	D2016N	4.0E-4	2.1E-3	16.59	T	D	ref	mut	mut
15:51791234	G>A	DMXL2				S1396F	1.0E-3	1.7E-3	26.9	D	D	mut	ref	mut

C

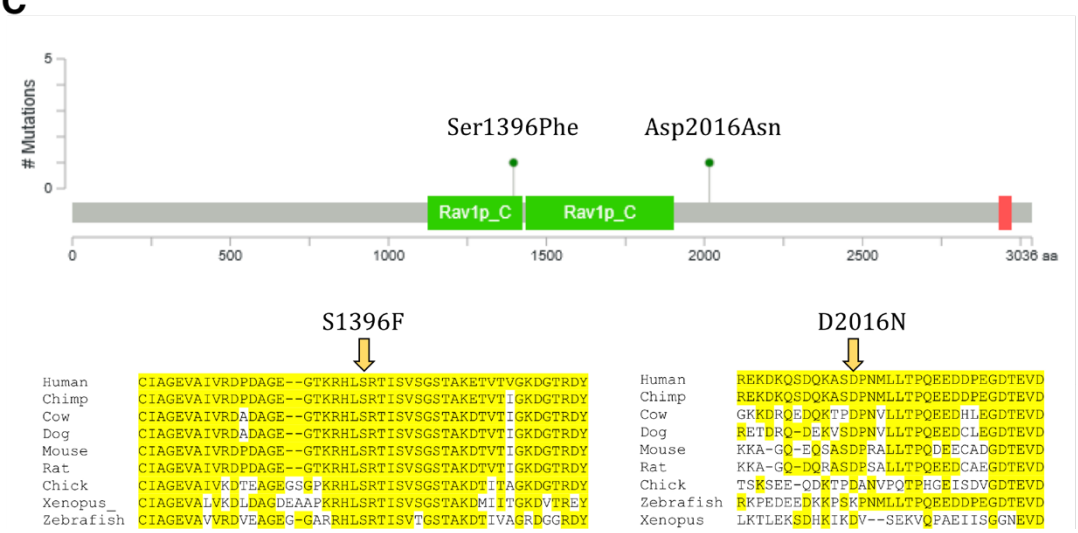


Figure 4-19 The family with the DMXL2 variants

[A] Family pedigree shows that the proband and her sibling have a double heterozygous mutation in *DMXL2,* with each variant inherited from one of the parents. [B] This table summarises the data regarding the two *DMXL2* mutations. The gene variants are c. 4277G>A and c.6136.C>T. These result in the Serine to Phenylalanine and Aspartate to Asparagine substitutions at positions 1396 and 2016, respectively. Abbreviations: MAF, minor allele frequency; CADD, SIFT and PPH are measures of how damaging these mutations are to protein function. A CADD score of over 15 is considered damaging. D, Deleterious, T, Tolerant. Both parents are heterozygous for one of the *DMXL2* variants. [C] Schematic representation of the location of the variants on the DMXL2 protein and the level of conservation of the two amino acids substitutions across various species.

4.6 **CONCLUSIONS**

In this chapter, I demonstrated for the first time that p120-catenin is not only implicated in human conditions involving epithelial integrity, most likely caused by aberrant E-cadherin/p120 interactions, but also in other important intracellular functions **(Figure 4-18)**. I concluded that *CTNND1*-related disorders span a spectrum of phenotypes ranging from multi-system involvement, to non-syndromic clefting. While further studies will be necessary to definitively understand the phenotype-genotype relationships, *CTNND1*, and perhaps *CDH1,* should be considered when patients present with characteristic craniofacial anomalies, congenital cardiac defects and neurodevelopmental disorders.

Chapter 5 IDENTIFICATION OF A COPY NUMBER VARIATION THAT LINKS THE *AGAP6* GENE TO SYNDROMIC CLEFTS

5.1 SUMMARY

In this chapter I will illustrate the implication of a copy number variant from whole exome sequencing (WES) in the identification of phenotypic discordance in a twin pair from another proband that I recruited from my Clinical Study with cleft and associated congenital anomalies. Copy number variations (CNVs) are strong contributors to differing clinical pictures in twins and to human diseases in general. Next generation sequencing (NGS) techniques made it possible to detect small CNVs that were previously undetectable. Sequencing of phenotypically discordant monozygotic (MZ) twins provides valuable insights into the molecular events underpinning their dissimilarities and allows the identification of CNVs. The **aim** of the current study was to identify the genetic aetiology behind the phenotypic discordance in a pair of twin siblings. **Method:** Whole exome sequencing was carried out in a male proband who presented through the Clinical Study with an undiagnosed craniofacial syndrome, WES was also carried out for the mother and the unaffected twin. **Results:** Exome sequencing results confirmed monozygosity and revealed a small *de novo* copy number variation in the proband in a previously undesignated gene, *AGAP6.* This finding was otherwise absent from the unaffected twin. The proband's craniofacial phenotypes included a complete bilateral cleft lip and palate, hearing loss and a hypoplastic corpus callosum. He also had extra-facial anomalies such as congenital kidney abnormalities and mild developmental delay. All conditions were absent from his monozygotic twin. I identified the expression pattern of *AGAP6* mRNA in the right place at the right time during human embryonic development in the various structures affected in the patient and I described three other subjects with single nucleotide variants (SNVs) in *AGAP6* who also had similar craniofacial anomalies. **Conclusion:** I propose that *AGAP6* may be a causal gene for the craniofacial and kidney anomalies seen in the proband. This study demonstrates the value of finding discordant MZ twins to discover candidate genes and suggests that *AGAP6* variation may underlie congenital birth defects, particularly those associated with craniofacial disorders. Future research is needed to confirm the copy number variation in *AGAP6* and to direct future studies to investigate this gene further in cleft children particularly in those with syndromic forms of orofacial clefting.

5.2 INTRODUCTION

5.2.1 Copy Number Variation in Twin Studies

Copy number variants (CNVs) are wide spread in the human genome and account for human disease and population diversity; more than 2000 CNVs have been described affecting 12% of the genome (Daar *et al.*, 2006; Girirajan *et al.*, 2011). DECIPHER (https://decipher.sanger.ac.uk) defines CNVs as variants that exceed 50 base pairs (bp), while variants less than 50bp are considered to be sequence variants. Copy number variants can also underlie monogenic diseases (Veltman & Brunner, 2012). For instance, a *de novo* CNV at chromosome 8q12 led to the discovery of the gene implicated in CHARGE syndrome, *CHD7* (Veltman & Brunner, 2012; Vissers *et al.*, 2004). Traditionally, methods such as fluorescent *in situ* hybridization, array comparative genomic hybridization (aCGH), and SNP arrays have been employed to discover CNVs of sizes ranging from one kilobase (kb) to several megabases (Tan *et al.*, 2014). Small variants as short as 50bp have since become possible to detect, by utilizing various algorithms on next generation sequencing (NGS) data (Alkan *et al.*, 2011; Tan *et al.*, 2014). While the majority of benign CNVs in the human genome are <100kb in size and many copy number variants are usually prioritised the larger their size, small CNVs could still be pathogenic and many have been implicated in disease (Bucan *et al.*, 2009; Hitz *et al.*, 2012; Silversides *et al.*, 2012).

Studying genetic contributions in discordant monozygotic (MZ) twins is of great value to understanding disease causality, as single phenotypically discordant MZ twin pair share gestational, environmental and genetic sequences, therefore differences between them might have been caused by non-shared genetic variation (Zwijnenburg *et al.*, 2010). As dizygotic twin pairs and siblings only share 50% of their genome with the added confounding effects of polymorphisms, correct determination of zygosity is of paramount importance for the interpretation of research findings, for genetic counselling and for the implementation of preventative medical strategies to affected families (Castillo-Fernandez *et al.*, 2014; Cutler *et al.*, 2015; Zwijnenburg *et al.*, 2010). Indeed, the *IRF6* gene was first identified and linked to van der Woude (VWS) from a monozygotic twin pair who were discordant for the syndrome; with an additional 45 unrelated families with VWS later confirmed to have mutations in *IRF6* (Kondo *et al.,* 2002). The authors provided proof of principle that discordant monozygotic twin pairs

allow the search for modifiers or mutations, especially in regard to complex traits (Kondo *et al.,* 2002).

The study of copy number variation in phenotypically dissimilar monozygotic twins may provide a powerful tool for identifying disease-predisposition loci. This has been demonstrated by several studies. A group that studied MZ twins with either concordant or discordant phenotypes found that CNVs existed within the pairs in both groups and that the estimated frequency with which *de novo* CNVs occur could be as high as 10% per twinning event (Bruder *et al.*, 2008). Differences in copy number variation have also been explored in MZ twins discordant for congenital heart disease; three copy number differences were found in 1 out of 6 MZ twin pairs (Breckpot *et al.*, 2012). In pairs of MZ twins discordant for schizophrenia, 10% of the CNVs were *de novo* and 80% represented gains (Maiti *et al.*, 2011). Conversely, other studies that explored CNVs in discordant MZ twins could not reproduce the presence of intertwin CNV differences (Lasa *et al.*, 2010; Ono *et al.*, 2010).

5.2.2 What is ArfGAP with GTPase Domain, Ankyrin Repeat and PH Domain 6 [*AGAP6*]?

Very little is known about AGAP6 [ArfGAP with GTPase domain, ankyrin repeat and PH domain 6]. It is located on chromosome 10. There are two annotated transcripts for *AGAP6* [transcript variant 1: NM_001077665.2] which is 2772 bp long. This transcript has eight exons, of which all are coding exons and encodes isoform 1 [NP_001071133.2] which is 686 amino acids (aa) long. Transcript variant 2 [NM_001365867.1] is 3,123 bp long and the translated protein is made of 485 residues [NP_001352796.1 (isoform 2)].

The *AGAP6* gene is unique to primates and has no known mouse or frog orthologs. However, there are many known human homologous *AGAPs*. Eleven human genes are predicted to encode AGAP-type proteins; *AGAP6* is one of the numerous copies of *AGAPs* clustered at human chr10q11 arising from chromosomal amplifications.

AGAPs are one of the numerous subtypes of a family of proteins called the ADP-ribosylation factor GTPase activating proteins (Arf GAPs). ADP-ribosylation factors (Arfs) are GTP-binding proteins that lack detectable intrinsic GTPase activities. Therefore, hydrolysis of GTP bound to Arf is mediated by GTPase-activating proteins (GAPs) (D'Souza-Schorey & Chavrier, 2006; Donaldson & Honda, 2005; Inoue & Randazzo, 2007; Randazzo *et al.*, 2007; Zhu *et al.*, 2009). Arf GAPs regulate actin dynamics, lipid modification and membrane trafficking (Randazzo & Hirsch, 2004; Zhu *et al.*, 2009). They are a large family of proteins all of which have a conserved Arf GAP

domain consisting of a zinc finger motif (Randazzo & Hirsch, 2004). Arf GAP-subfamily proteins contain, with differing quantities, a catalytic core of pleckstrin homology (PH), Arf GAP, and ankyrin repeat domains which many are known to function as protein-protein interaction domains. The protein interactions between all these domains may directly modify Arf GAP activity. Notably, of all the AGAPs, AGAP1 and AGAP2 have been the most extensively characterized members of this group. They function in the endocytic system; AGAP1 working with AP-3 and AGAP2 with AP-1 (Nie *et al.*, 2005), whereby the PH domain of AGAP1 was found to bind to the clathrin coat protein AP-3 (Nie *et al.*, 2005), the PH domain of AGAP2 was found to directly interact with AP-1 which functions at the trans-Golgi network and endosomes and also recognises sorting motifs in cargo molecules (Nie *et al.*, 2005). AGAP2 was also shown to form a complex with focal adhesion kinase (FAK) by increasing FAK's activity, and provoking the focal adhesion disassembly during cell migration (Zhu *et al.*, 2009).

5.3 RESULTS

5.3.1 Case Report and Clinical Findings

A 15-year old male proband, who is the second of twins participated in the Clinical Study. He presented with a history of the conditions described below. He has two other un-affected siblings.

The proband was prematurely born (32/40 weeks) with complete bilateral cleft lip and palate and hydrocephalus, with a ventriculoperitoneal (VP) shunt later inserted. He was diagnosed with congenital renal malformations comprising antenatal hydronephrosis (right sided hydronephrosis) and postnatal right ureteropelvic junction (UPJ) obstruction which was relieved by surgical reconstruction involving right pyeloplasty. At two years of age he had a bilateral inguinal hernia repair and at five years of age an onset of seizures and was previously diagnosed with hypothyroidism.

He also presented with a number of distinctive facial features including a widows' peak, a broad forehead, brachycephaly, repair of the complete bilateral cleft lip and palate, hypertelorism, downslanted palpebral fissure, broad nasal root and bridge in addition to a nose tip rhinoplasty, and history of eye surgery for a squint. Additionally, he has myopia and hypermetropia and has left sided-hearing loss (only 40% hearing in his left ear), a hypoplastic corpus callosum with reported mild developmental delay and learning difficulties. Skeletally, he has a chest deformity and asymmetric shoulders, broad thumbs and long fingers. He was reported to have a normal echocardiogram.

Intra-orally, the proband presented with malaligned, hypoplastic anterior teeth typical of cleft lip/palate dentitions, absence of the maxillary left permanent lateral incisor and a crowded dentition.

There is no family history of clefting and no history of any syndromes. The proband's twin brother was reported to have none of the above conditions or anomalies described, apart from history of hypothyroidism.

5.3.2 **Genetics**

5.3.2.1 Previous molecular genetics findings

Previous genetic tests were carried out on the proband to investigate genetic contributions that may explain the conditions seen. All tests showed normal outcomes which included a karyotype, chromosome 22q11 deletion, a multiplex ligation-dependent probe amplification (MLPA) sub-telomere test at 6 years of age, and a recent array comparative genomic hybridization (CGH). Although the MLPA test showed a small imbalance in the long arm of chromosome 4 at the time it was carried out, this small duplication is now considered a common benign populational polymorphism and unlikely to be the cause of the proband's conditions.

5.3.2.2 Whole exome sequencing findings

Whole exome sequencing (WES) results revealed monozygosity of the twins with 95% identity of their genomes. At the first time of sequencing, DNA from the unaffected twin brother and the proband's mother was available for WES. Based on a recessive model of inheritance, where the MAF was set at <0.5%, the twins were found to have two novel homozygous deleterious mutations in *TENM3* (teneurin transmembrane protein 3) and *CDK16* (cyclin-dependent kinase 16). Details of the variants are shown in **(Table 5-1)**. However, since the twin sibling is clinically unaffected, these gene variants were most likely not implicated in the proband's condition.

A number of other very rare heterozygous mutations (MAF=<1E-04) were found in the twins that were absent from the mother. Examples included variants in *TP63* (tumor protein p63), *DROSHA* (drosha, ribonuclease type III), *ANKRD2* (ankyrin repeat domain 2), and *AP3B2* (adaptor-related protein complex 3, beta 2 subunit). Again, these were excluded and were deemed unlikely candidates. Sanger sequencing of *TP63*, *DROSHA*, *ANKRD2*, *AP3B2* and the novel homozygous variants, *TENM3* and *CDK16,* was carried out and confirmed sequence similarity in the twins.

Whole exome sequencing identified a *de novo* copy number variant (CNV) on chromosome 10 which is a 724-bp duplication in the *AGAP6* gene (chr10:51,748,453-51,749,177) **(Figure 5-1, A-C).**

Table 5-1 Homozygous gene variants found in the twin brothers

Genomic position	Ref/ mut.	Gene	Protein variant	1KG	MAF ExAC	CADD	Meta SVM	SIFT	PPH	M	T
4:183720759	G>C	TENM3	Gly2452Ala	Novel	Novel	24.8	D	D	D	ht	hm
X:47082561	C>T	CDK16	¥Arg23*	Novel	Novel	28.4	stop-gain	-	-	ht	hm

¥This mutation was present in only one CDK16 isoform. CADD, Meta SVM, SIFT and PPH are all *in silico* prediction scores. Abbreviations: M, mother; T, twin; ht, heterozygous; hm, homozygous; D, damaging.

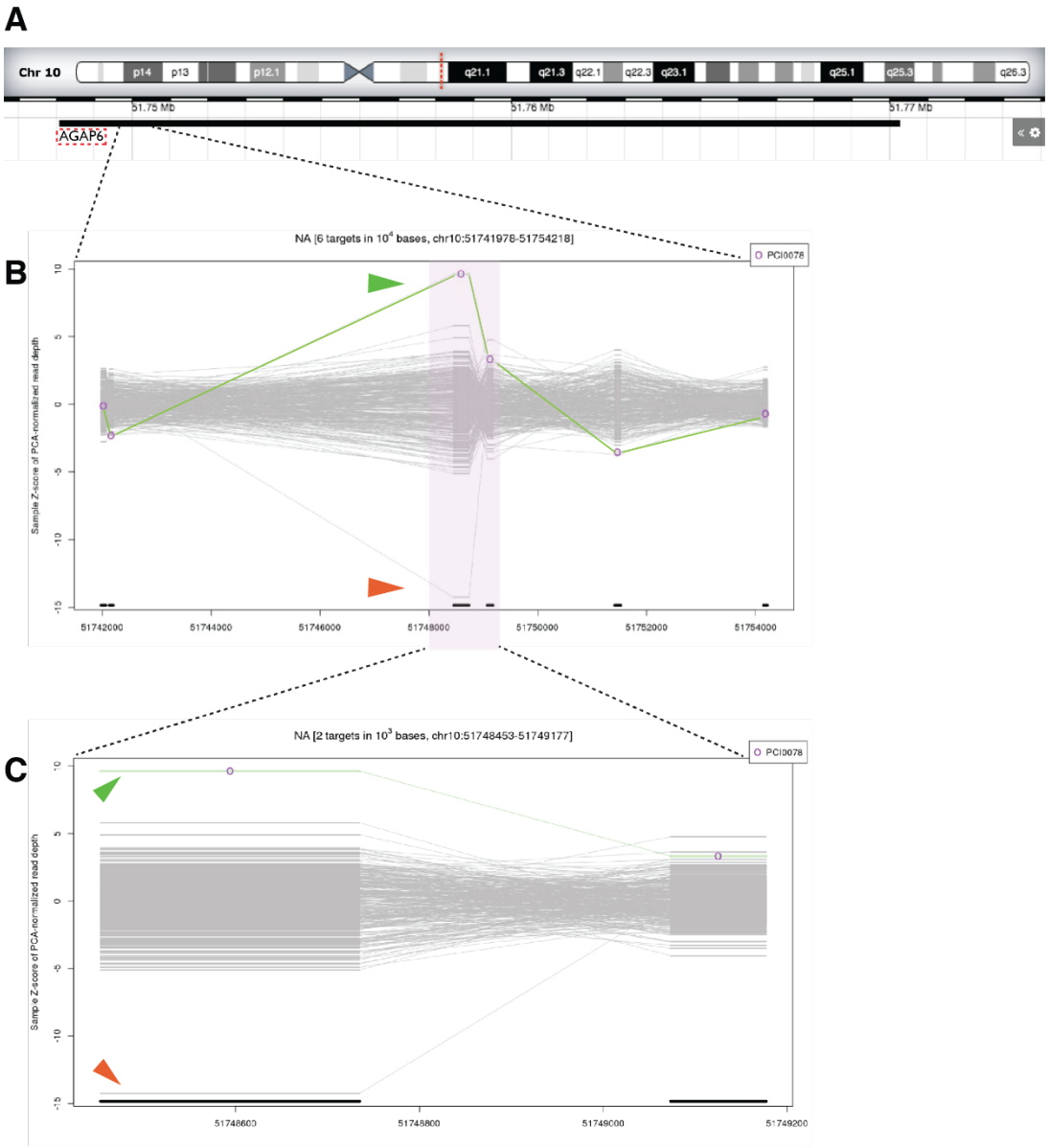


Figure 5-1 AGAP6 copy number variant (CNV) in the proband

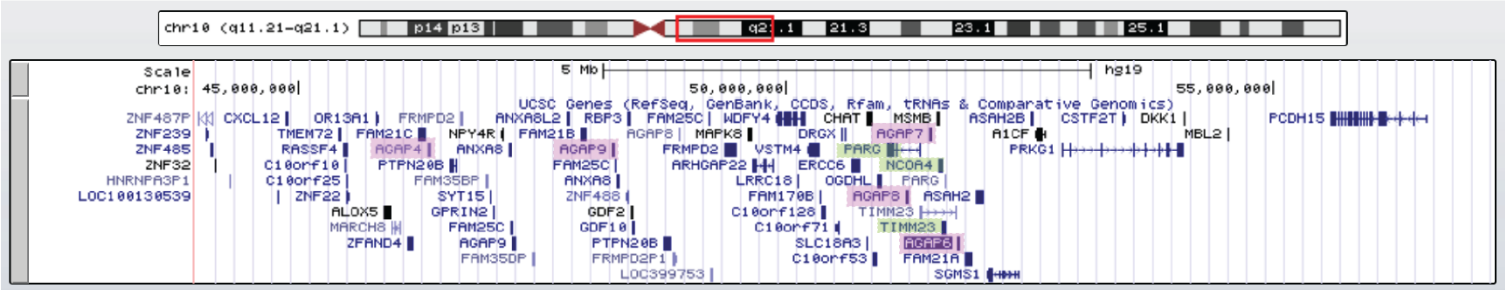
[A] Genome browser image of human chromosome 10 showing the region containing the *AGAP6* gene (red-dotted line). [B-C] A single *de novo* duplication (<1kb) was identified in a male proband. In this XHMM analysis, the sequence reads of the proband, twin brother and 350 European controls were run together. No CNVs were identified in the twin sibling. Sequence Exon Capture shows a duplication in *AGAP6* in the proband (green lines and green arrowhead), covering the region between chr10:51,748,453-51,749,177 (a duplication of 724bp). Black horizontal line within the lower end of the pink-shaded region is expanded in [C] and indicates exonic sequence at 51,748,000. [B-C] The depth coverage was normalized so that the majority of the samples were close to zero, only the outliers were shown up. (Z-score: standard deviation from the mean coverage of the population tested). Each of the grey lines represents normalized mean coverage in the region from an individual, showing 350 lines in the central around Z=0. Two outliers were shown in this region, one is the proband (purple circle) which is gain of copy number, and the other from the control group (non-craniofacial phenotypes), which is loss of copy number (orange arrowhead).

5.3.3 **AGAP6 and Sequence Homology with Other AGAPs**

Several AGAPs are encoded by one of the numerous copies of centaurin gamma-like genes clustered in the q11 region of chromosome 10. These are AGAP 4 (also known as AGAP8), AGAP5, AGAP6, AGAP7, and AGAP9 **(Figure 5-2, A, pink boxes)**. Amino acid sequence alignment revealed close sequence identity between AGAP6 and AGAPs 4, 5 and 9. However, there is less sequence homology with AGAP1 and AGAP2 **(Figure 5-2, B)**. The AGAP6 protein structure and domains are shown in **(Figure 5-2, C).**

There is no mouse ortholog for *AGAP6* When examining the mouse syntenic region on chromosome 14, only neighbouring genes were found such as *Ncoa4*, *Timm23*, *Parg*, but not *Agap6* **(Figure 5-2, A, green boxes; human chromosome).** Moreover, when examining mouse chromosome 1 where *Agap1* is located, its neighbouring genes *Sh3bp4*, *Iqca* and *Gbx2* are all the same as the neighbouring genes of human *AGAP1*. Therefore, mouse *Agap1* is not a true ortholog of human *AGAP6*.

A



B

AGAP homologs	Chromosomal location	Transcript ID	Length (aa)	Overall identity (%)
AGAP6	10q11.23	NM_001077665.2	686	100
AGAP4 (aka AGAP 8)	10q11.22	NM_001263272.2	686	98
AGAP5	10q22.2	NM_001144000.1	686	96.2
AGAP9	10q11.22	NM_001190810.1	658	94
AGAP1	2q37.2	NM_001037131.3	857	51.4
AGAP2	12q14.1	NM_001122772.2	1192	26

C

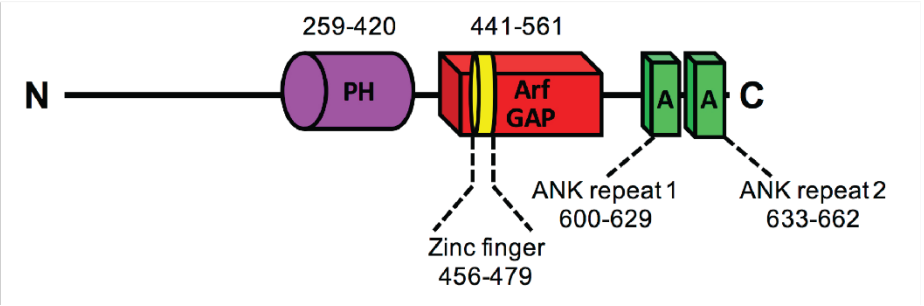


Figure 5-2 AGAP6 homologs and protein domains

[A] Diagrammatic representation of human chromosome 10 expanding on the 10q11 region. A number of AGAPs are clustered in this region (pink boxes). [B] Table of amino acid sequence homology between other AGAPs with AGAP6; transcript variant 1 (isoform 1) from each AGAP was chosen as the representative transcript. [C] Schematic representation of the AGAP6 protein containing a pleckstrin homology domain (PH) at the N-terminal region (N), an Arf GAP domain and two ankyrin repeats (A) at the C-terminal region (C). Numbers indicate amino acid position and sequence length.

5.3.4 Subjects with Sequence Variants in *AGAP6* also have Craniofacial Anomalies

In order to explore whether other pathogenic variants in *AGAP6* existed and whether individuals with these variants phenocopied the proband, I identified three subjects that had validated sequence variants in *AGAP6* identified through WES. However, the pathogenicity of their *AGAP6* variants to their clinical conditions had not been determined thus far. Details of these *AGAP6* variants are described in **(Table 5-2)** and the subject's phenotypes are described in **(Table 5-3).**

Although the c.152C>T (p.Pro51Leu) variant was novel (ExAC=0), *in silico* prediction showed that this variant was a polymorphism. The variant in the other patients, c.55dup (p.Gln19Profs*10), was predicted to be disease causing. Interestingly, this single nucleotide variant is located within the upstream region of the *AGAP6* duplication site. Furthermore, the third patient (ID 301334) was found to have a clinical picture reminiscent of the proband's.

Table 5-2 DECIPHER probands with sequence variants in AGAP6

DDD ID	Sex	Genomic location	Mutation cDNA	Protein variant	Type	Inheritance	ExAC count
266765	F	10: 51,748,627, C/T	c.152C>T	p.(Pro51Leu)	missense	Maternally inherited	0
290806	M	10:51,748,528, A/AC	c.55dup	p.(Gln19Profs*10)	frame shift	¥Unknown Biallelic	24
301334	M	10:51,748,528, A/AC	c.55dup	p.(Gln19Profs*10)	frame shift	¥Unknown Biallelic	24 MAF (dbSNP): C=0.0002/24 (ExAC) C=0.1589/4659 (GnomAD)

¥Unknown because parents were not sequenced. Note, the last two patients have identical mutations. Annotation used: hg19; NM_001077665.2; NP_001071133.2.

Table 5-3 Clinical description of the DECIPHER probands with sequence variants in AGAP6

DDD ID	266765	290806	301334
Sex	F	M	M
		Phenotypes	
Face	Triangular face	-	Long face
Teeth	-	-	Macrodontia of permanent maxillary central incisor; Misalignment of teeth; Yellow-brown discoloration of the teeth
Other craniofacial	Abnormality of the nasal tip; Smooth philtrum; Thin vermilion border	Prominent nasal bridge; Broad neck; Macrotia; Mandibular prognathia; Microcephaly	High palate; Long palpebral fissure; Downslanted palpebral fissures; Abnormal location of ears; Broad columella; Mandibular prognathia; Gingival overgrowth
Neurodevelopmental/ Brain	Mild intellectual disability	Cognitive impairment	Abnormal emotion/affect behaviour; Delayed speech and language development; Mild intellectual disability; Specific learning disability
Other	Umbilical hernia	Short stature	Long fingers

5.3.5 Expression of *AGAP6* mRNA During Human Development

In order to determine whether *AGAP6* is expressed during human development, I designed and synthesised a plasmid that spans the 5'-UTR of *AGAP6* up to exon 5. I assessed *AGAP6* mRNA expression in Carnegie stage 21 human embryos (week 8 of gestation). I found *AGAP6* robustly expressed in craniofacial structures and various other organ systems **(Figure 5-3 to Figure 5-5)**. In the developing head, *AGAP6* expression was seen in the epithelial lining of the tongue and in the tongue muscles **(Figure 5-3, A-A'; 5-3, B)** as well as the developing palatal shelves. Interestingly, expression appeared relatively restricted within in the mesenchymal region of the palatal shelf **(Figure 5-3, B', black-dashed line).** Transcripts were strongly detected in the developing laryngeal apparatus **(Figure 5-3, C-D').** Expression was robust in the epithelium surrounding the laryngeal inlet, particularly in the developing epiglottis and aryepiglottic fold bilaterally **(Figure 5-3, C').** Further posteriorly, *AGAP6* expression was still seen in the ventral laryngeal epithelium and in the epithelium surrounding the laryngeotracheal groove **(Figure 5-3, D').** In the developing central nervous system, transcripts were robustly detected in various regions within the brain **(Figure 5-4).** In the eye, signal was detected in the neuroblastic (sensory) layer of the retina. Strong signal was particularly seen in the outer neuroblastic layer, whereby neuroblasts in this region form rods and cones **(Figure 5-4, B #3, red-dotted line).** Signal was weakly detected within the inner neuroblastic layer, in which the neuroblasts here form bipolar neurons **(Figure 5-4, B #2, red-dotted lines).** Interestingly, expression was spared from the innermost layer, the ganglion cell layer **(Figure 5-4, B #1, red-dotted lines).** Other areas expressing *AGAP6* in the developing eye included the lens epithelium and the hyaloid plexus **(Figure 5-4, B)**. Finally, expression was also strongly detected in the trigeminal ganglia **(Figure 5-4, I).**

As for the other organs, I found *AGAP6* mRNA robustly expressed in the cardiomyocytes of the atria, ventricles and interventricular septum, in addition to its strong expression in cells of the endocardium **(Figure 5-5, A-B & E-F).** In the lower half of the body, transcripts were strongly detected in the developing organs of the gastrointestinal tract (GIT). Particularly, signal was strong in the epithelial linings of the stomach wall **(Figure 5-5, C-D & G)**, and parts of the small intestine (shown are sections through the duodenum and jejunum) **(Figure 5-5, C #3&4).** Expression was also strong in the pancreas **(Figure 5-5, C #2).** Posterior to this, expression was detected in other neuronal ganglia such as the aortic abdominal sympathetic ganglion and the celiac ganglia **(Figure 5-5, D).** Interestingly, robust expression of *AGAP6* was detected in the

developing kidney capsules, specifically in the metanephric mass of mesoderm (or mesenchymal cell cluster) sparing the collecting tubules **(Figure 5-5, H)**. Finally, *AGAP6* expression was faintly observed in the spinal cord and the dorsal root ganglia **(Figure 5-5, I).**

Anterior
Posterior
Oral cavity
A
A'
B
B'
C
C'
D
D'
M
E
Af
Li
Gf
Tc
Le
Li
Lt

Figure 5-3 Craniofacial expression of AGAP6 at Carnegie stage 21 of human development

AGAP6 mRNA *in situ* hybridization. [A-D] Coronal sections through the head with structures shown anteriorly to posteriorly. [A'-D'] Higher magnification images of structures boxed in [A-D]. [A-B, A'-B'] Expression in the tongue epithelium and intrinsic muscles. [B'] Expression in the right palatal shelf is stronger in the mesenchyme (dashed-black line). [C-D, C'-D'] *AGAP6* is strongly expressed in the laryngeal apparatus. [C'] Expression is seen in the epithelium surrounding the laryngeal inlet (Li), specifically in the epiglottis (E), the aryepiglottic fold of mucous membrane (Af) and the glossoepiglottic fold (Gf). [D'] Transcripts are strongly expressed in the ventral laryngeal epithelium (Le, green arrowhead) and in the epithelium of the laryngeotracheal groove (Lt, red arrowhead). Note the expression of *AGAP6* in the epithelium of the soft palate (black arrowheads). Abbreviations: Tc, Thyroid cartilage.

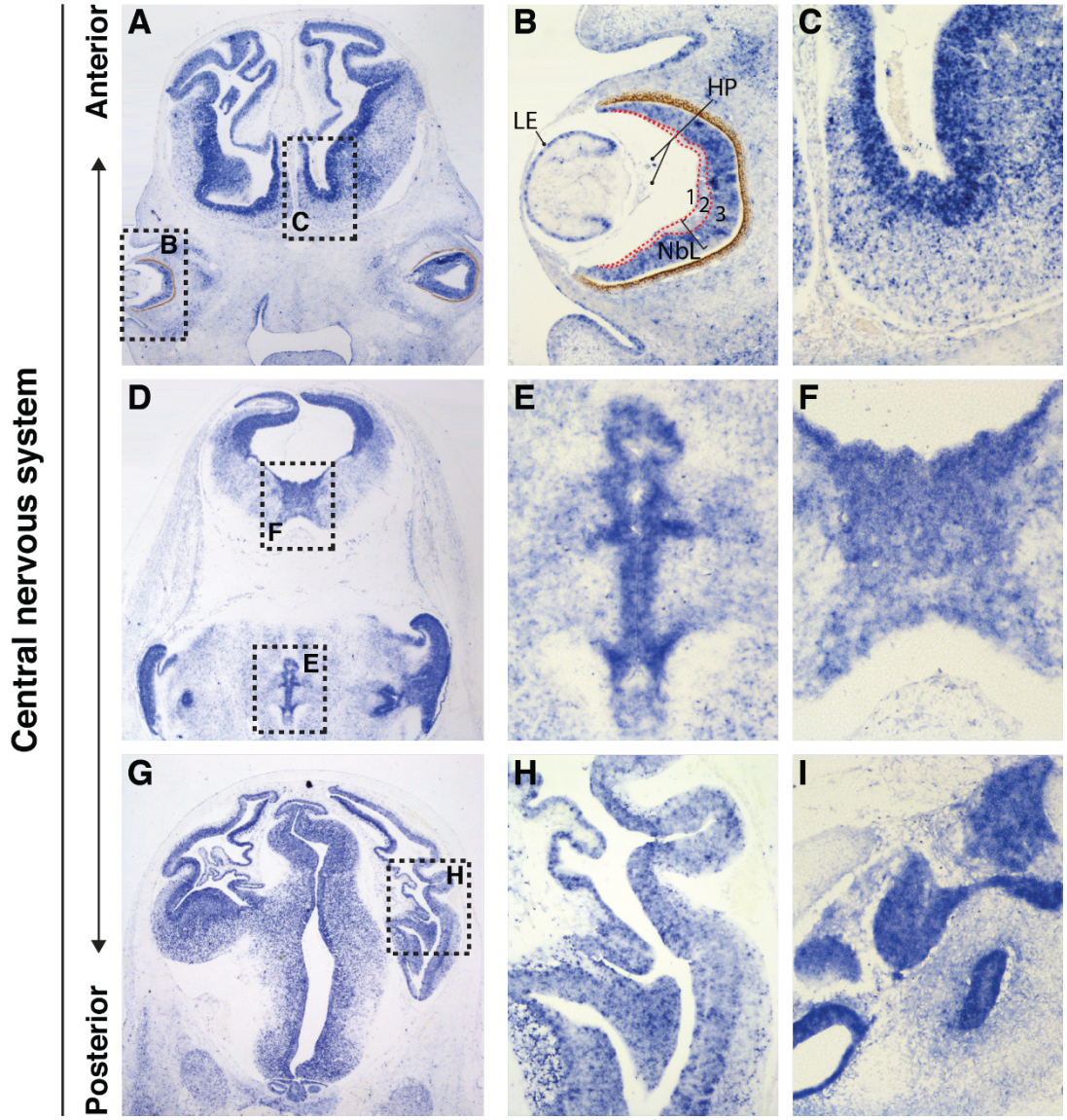


Figure 5-4 Expression of AGAP6 at Carnegie stage 21 in the central nervous system

[A-I] *AGAP6* expression in structures of the central nervous system. [A, D, G] Coronal structures through the brain anteriorly to posteriorly. [B-C, E-F, H-I] Magnified views of *AGAP6* expressing regions in [A, D, G]. [B] Expression in the developing eye is specific to the lens epithelium (LE), hyaloid plexus (HP) and the neuroblastic (sensory) layer of the retina (NbL). Specifically, it is strongly expressed in the outer neuroblastic layer (3), slightly in the inner neuroblastic layer (2), but is not expressed in the ganglion cell layer (1). [C, E-F, H] Transcripts are strongly expressed in different regions of brain cells. [I] Boxed area in (Figure 5-3, C) shows clear expression in the trigeminal ganglia.

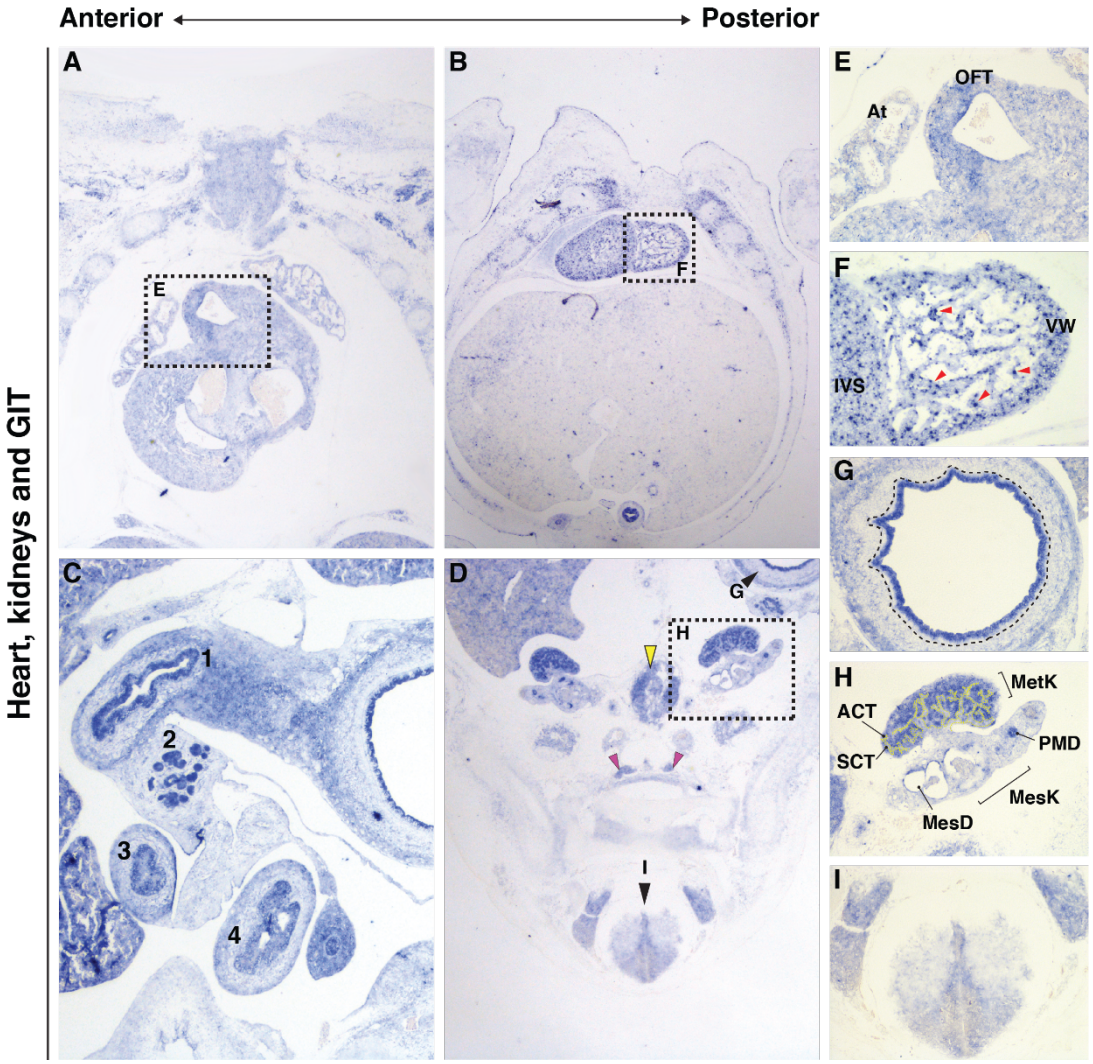


Figure 5-5 Expression of AGAP6 in other organ systems during Carnegie stage 21 of human development

AGAP6 mRNA *in situ* hybridization. [A-D] Coronal sections through the heart [A-B], gastrointestinal (GIT) organs [C] and kidneys [D]. [E-F] Higher magnification images of structures boxed in [A-B, respectively]. [A-B, E-F] Expression is observed in the developing heart in the atrium (At), outflow tract (OFT) [E], myocardiocytes of the ventricular walls (VW) and interventricular septum (IVS) and cells of the endocardium (red arrowheads) [F]. [C] Transcripts are strongly expressed in various organs of the GIT system particularly in the inner epithelium of stomach pylorus (1), pancreas (2), duodenum (3) and jejunum (4). [D] Coronal section posterior to [C]. Expression is observed in the aortic abdominal sympathetic ganglion (yellow arrowhead) and caudal to that, in the celiac ganglia (pink arrowheads). [G-I] Magnified views of *AGAP6* expressing regions in [D]. [G] *AGAP6* is expressed in the inner lining epithelium of the stomach wall sparing the basement membrane (emphasized by the dashed-black line). [H] Left kidney shown. Expression is strong in the metanephric part of the kidney and only surrounding the duct epithelium in the mesonephric part. Note lack of expression in the collecting tubules (yellow-dotted outlines). [I] Faint expression is seen in the spinal cord and the dorsal root ganglia bilaterally. Abbreviations: STC, straight collecting tubule; ACT, arched collecting tubule; MesD, mesonephric duct; MesK, mesonephric kidney; MetK, metanephric kidney; PMD, paramesonephric duct.

5.4 DISCUSSION

5.4.1 *AGAP6* Variants and Phenotypes

In this study, I described the phenotypic discordance of a monozygotic twin pair in which one was found to have a *de novo* copy number variant (CNV) in the *AGAP6* gene following exome sequencing. Thus, for the first time, I reported on the involvement of *AGAP6* with congenital human disease. I also showed the expression of *AGAP6* mRNA during development of human embryos particularly in various craniofacial, neuronal, kidney and gastrointestinal organs.

While *AGAP6* has not been linked to human disorders, it was suggested as a potential candidate gene from recent sequencing data that investigated undiagnosed cases of Rett syndrome (RTT); a neurodevelopmental disorder which is often classically associated with mutations in *MECP2*, and with *CDKL5* and *FOXG1* in atypical cases. The authors found previously undetermined novel pathogenic gene variants one of which was an *AGAP6* mutation designated as chr10:51,748,528 [c.53insC, (p.Asp18Ala_fs10*)] (Lucariello *et al.*, 2016). In this same region, another study reported the same variant, which they designated as [c.53_54insC, (p.Asp18Ala_fs11*)], that was excluded as the causative mutation, in a patient with intractable epilepsy, involuntary movements, microcephaly, and developmental and growth retardation; which the authors eventually linked to a homozygous mutation in *SV2a* (Serajee & Huq, 2015). Interestingly, these variants are only one amino acid upstream of the single nucleotide variant (SNV) I described here (p.Gln19Profs*10); suggesting that the 5'-UTR region of *AGAP6* is enriched with pathogenic gene variants. Functional elements within 5'-UTRs fine tune protein expression and genetic variations within 5'-UTRs have been involved in a number of human diseases (Chatterjee *et al.*, 2001). However, the (p.Gln19Profs*10) variant found in the additional subjects is common in the overall population examined on ExAC (occurring in 1.9% and is especially common in the European Finnish population (5%)). The presence of this common frameshift variant suggests that perhaps *AGAP6* is neutral/tolerant for loss of function and is, on the other hand, intolerant to duplication/gain of function mutations.

Despite the difficulty in identifying CNVs from WES data when compared to whole genome sequencing (WGS), the identification of CNVs is still possible. It relies mainly on read depths and requires different statistical models and algorithms from the ones used in WGS (Tan *et al.*, 2014). In this study, I demonstrated the possibility of detecting small CNVs via WES. Despite the advantages of using WES, a number of

limitations could be outlined. The phenotypes seen in the proband could be caused by a chromosomal aberrance; although this has been excluded since his array CGH analysis was consistent with a normal male complement. This is not surprising as chromosomal array CGH resolution is often limited to 10-20 Mb. Therefore, anything smaller than that will not be detected and since the proband's duplication is <1kb in size, it would not have been picked-up by array CGH.

5.4.2 **Monozygosity and Discordance in Twins**

Factors contributing to the phenotypic discordance in affected and unaffected twin pairs include a difference in DNA methylation or histone acetylation (Castillo-Fernandez *et al.*, 2014). In this study, the possibility of epigenetic differences between the twins has not been addressed. This could be tested by carrying out genome-wide methylation sequencing. Other gestational factors or genetic mechanisms that could account for discordance in MZ twins are postzygotic twinning mutations that could result in somatic mosaicism, environmental differences, differential telomere length reduction, X-chromosome inactivation or the presence of copy number variation (Cutler *et al.*, 2015; Leslie *et al.*, 2017; McNamara *et al.*, 2016; Zwijnenburg *et al.*, 2010), as shown in the present study.

The mother and father were both fit and well and were dentally examined, since it is not unusual to find subclinical phenotypes in 'unaffected' individuals particularly in families with a child with cleft lip and palate (Eerens *et al.*, 2001; Leslie *et al.*, 2017; Mossey *et al.*, 2010; Neiswanger *et al.*, 2007). The parents were not found to have any dental or orofacial anomalies. Despite WES revealing underlying novel homozygous mutations in *TENM3* and *CDK16* in the twins, these were ruled out as likely candidates since the twin sibling was phenotypically unaffected. *TENM3* is associated with Microphthalmia, Isolated, with Coloboma 9 [MIM: 615145] (Chassaing *et al.*, 2016) with the phenotypes recently expanding to include motor developmental delay and intellectual disability (Singh *et al.*, 2019; Stephen *et al.*, 2018). Recently it has also been implicated in developmental hip dysplasia (Feldman *et al.*, 2019). As for *CDK16*, it has not been implicated in congenital human disorders yet.

5.4.3 **The Distinctiveness of *AGAP6***

ArfGaps and AGAPs, which are both subtypes of the larger Arf GAP family of proteins, could often be confused. Although ARF6 has been extensively studied (Bourmoum *et al.*, 2018; D'Souza-Schorey & Chavrier, 2006; Hu *et al.*, 2019; Lin *et al.*, 2017; Mukhamedova

et al., 2016), AGAP6 has not. Despite them being members of the larger Arf GAP family of proteins, they do not share amino acid sequence homology. *ARF6* [NM_001663] is located on chromosome 14q21.3. ARF6 has been shown to have multiple functions including its effect on endocytosis, phagocytosis and receptor recycling, in addition to the regulation of actin and membrane remodelling through the formation of actin rich protrusions and actin rich membrane ruffles (D'Souza-Schorey & Chavrier, 2006). Yet, like *AGAP6*, it has not been linked thus far to any congenital human disorders.

Arf GAP subfamilies defer such that each has its own unique domain(s) at the N-terminal region. For instance, AGAP1 and AGAP2 contain an N-terminal Ras homology domain, called G-domain (aka GLD) which has 28% identity to Ras family members (Nie *et al.*, 2002; Xia *et al.*, 2003a; Zhu *et al.*, 2009). Although AGAPs have been named as such for the presence of a GTP-binding protein-like domain (GLD) in their N-termini (Randazzo & Hirsch, 2004), AGAP6 does not contain this domain.

5.5 OUTSTANDING QUESTIONS

The results presented in this chapter are preliminary. Further validation of the *AGAP6* duplication needs to be carried out. Further, more individuals with novel and rare pathogenic variants in *AGAP6* need to be found to ascertain the implication of *AGAP6* in congenital birth disorders.

5.5.1 Copy Number Variant Analysis in *AGAP6*

Preliminarily, in order to validate the novel *de novo* copy number variant in *AGAP6* found through exome sequencing, I performed quantitative analysis of genomic copy number using the family's DNA which I isolated from saliva samples. To do this, I designed two different primer pairs, the first flanking the beginning of the duplication and the second, flanking a region within the duplication. Indeed, the relative gene copy number analysis showed that the proband had an increased quantity of genomic *AGAP6* when compared to all other family members **(Figure 5-7, B-C).**

Next, I set out to investigate whether the nucleotides duplicated in the proband were inserted within or close to their original genomic location. I carried out a polymerase chain reaction (PCR) (n=3) covering the chr10:51,748,453-51,749,177 region containing the duplication, and the samples were Sanger sequenced. Sequencing results showed that the proband had the same sequence alignment, in that region, as all the other family members (data not shown). In other words, the patient had no extra nucleotide copies in this region, suggesting that the duplication in the proband is located either upstream or downstream of this region on chromosome 10 or that it was inserted into another chromosome.

Before drawing any firm conclusions, the results of the quantitative CNV analysis should be validated with more replicates. The first primer pair I used **(grey regions in Figure 5-7, A)** was perhaps not an accurate set to test copy number variation. Based on my hypothesis that the duplication is inserted within another gene or chromosomal region, this indicates that any nucleotides outwith the duplication sequence should not be amplified. Since the first 70bp amplified from the PCR generated from this primer set should not theoretically anneal, as they lie 'outside' the duplicated region, the results should be interpreted with caution despite the proband showing higher copy numbers of *AGAP6* in all three runs. Therefore, perhaps primer pair two **(yellow regions in Figure 5-7, A)** that I designed within the duplication site was, in hindsight, a more accurate location to test copy number variation. Although one

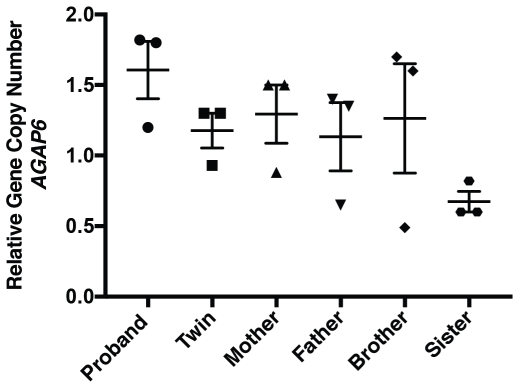
experimental run was carried out, results show higher copy numbers in *AGAP6* for the proband compared to the rest of the family. This experiment, however, needs to be repeated in order to confirm the aforementioned results.

A

>hg19_dna range=chr10:51747953-51749677 5'pad=500 3'pad=500 strand=+ repeat Masking=N

GCAGCAGCATCATAAGCCACAGGGTGGGGCAGCCAAGGCAGGGGCATTCT
GAGCTGTTGGGGAGGGGTGGCAGGCAGGGTGGGGCACTGTGAGCTGTCGG
GGAGGGCATTGTGAAGTGTGGGGTGGGGCATTGTGTGCCACATGCCTGGG
CTCCCACCTGGGGCCAGTGGGCTTCAGTCTGTAGGTGACTACAGAAGGAG
GAGGAACTCCGTCTGTTCTCTCTTCAGGCAGTTGTTGTGTCTCTCAGCGC
TTGTTGGTTTCACAACCTATTAAATAAGCCGGCTGGTCTTCACCCTCCCA
GACAAGTCAACTCAGGGGAGGCAGCAGGGTGCGGGCCTTGGCCCGCAGCC
CTAGCCGGGGCCGGGGCCAGGGCTGGTGCCCGGGGCCTCGCTGTGAGGTG
GGCAGGCGAGGAGCGGGAAGACCATCTCTGCAAGTGCAGCATAGCCTCGG
CCTAGGACAGCGGGAGTGCGTGGCCAAAGCTGTGAGCAGAGGCACAGGTG
GTGGCAGACAGTAGAGGCGCCCCATGGGGAACATACTGACCTGTCGTGTG
CACCCTAGCGTCAGCCTCGAGTTTGACCAGCAGCAGGGGTCGGTGTGTCC
CTCTGAATCTGAGACCTATGAGGCAGGAGCTAGGGACAGGATGGCAGGAG
CGCCCATGGCTGCTGCTGTACAGCCTGCTGAGGTGACTGTTGAAGTTGGT
GAGGACCTCCACATGCACCACGTTCGTGACCGGGAGATGCCTGAAGGTGA
GGAGGTGATAGGTGCCATCTACCCTCGGTTTGCCTCTGGCTGCTGCTGTC
CCCATGGTTCCCTTTGAGGCATCCCACACTTCGAGCTCCTTTCTGCTTGT
AGCCAGCTTTCCCGGGGGCTGGCCAGGAACAAAAGCTGGCTCTGCCTTGA
ATTCCCACCCCTTAGTCTTTCCCCACCGAGTCCAGTCAGTTTCTTTTCGC
CTCCCCTCCCAATCGCCCAGTTCTTGCTCTCTCATCTCATTCTCCCAGGC
TGGCATGGGACCATTTATTTATGGCTCTTGTCGAATAAGCAGCAGTTGAA
TAAATAAGTTGATAAATTTTTATAAATGATTACATCCTTTTTCTTTTCTC
CCTCTATACATATAGCTTTGGAGTTTAACCTTTCTGCCAATCCAGAGTCA
AGCACAATATTCCAGAGGAACTCTCAAACAGAAGGTGAGACAACAGTGTC
TGTAGCTCTATTTATTATCCTGTGGTACTTTGTTTAGGCTTCTTTGAGCT
ATTCTCTTCCTTTTCTCAGTAAAAACTCAAATATCCCAACTTTTCAGTAC
CCATCTTATTTTTTCTTTGTACCTATCCAGATGGTACCTAAGTGAAGGAA
CCAGGTAAGTGCCTAATTGTTTCCTTTGTTAAAGTAGCCAAATCTCAGGA
CAGTTCCTATTCAAATATTTGGGGATTTCTTATTTAAAATCAGAATGGAG
GTTGCCACGGGAGAGGCTATATGGTATTCTTAATGGGCTGCTTTAAGTCA
CCTTGATAGAAGCTGCTTAGTTTCTTCTAACTGTAATTTGAACACAAAAG
GAAAAAGAAAAAAGGAGAGTGCTTAAAGTAATTGTGAAAGGTGTAAAATG
TCACAGCCGGGGCTGCAGAAAAATGGTTGTGTGTGTGTGTTTGGGGTTTC
TCAAAGGAGTTTACCTATGAGGCTCTGATTACTTTAAAATTCTTACTTTA
ACAGAAAATGTGTCTCCAGATTTAT

B



C

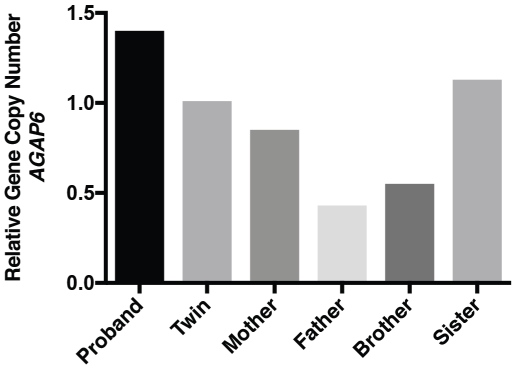


Figure 5-6 AGAP6 relative gene copy number analysis

[A] Nucleotide sequence that is thought to be duplicated in the A*GAP6* gene in the proband (in bold) with 500 base pairs shown before and after this portion. Primer pairs used to test copy number variation are highlighted in grey and are shown to cover the first few base pairs of the duplication. Another primer pair was used (highlighted in yellow) and were designed within the duplication site. [B] Genomic DNA was isolated from the saliva of all individual family members and analysed by qPCR using the primers in grey corresponding to the duplicated region. [C] Genomic DNA was analysed by qPCR using the primers in yellow. Data were normalised to the level of a gene that was normal (had no variation) in all family members. Data are presented as mean and standard errors. Note, the proband shows the highest copy number of *AGAP6* in [B] and [C].

5.6 **CONCLUSIONS**

In conclusion, I reported the discordance for congenital malformations in a monozygotic twin pair where one twin presented with congenital craniofacial and kidney anomalies. Most notably, the affected twin presented with a bilateral cleft of the lip and palate, congenital hydrocephalus, a hypoplastic corpus callosum, eye and ear anomalies with hearing loss and mild developmental delay. I provided evidence from whole exome sequencing that a small duplication in *AGAP6* may have accounted for the phenotypic discordance seen between the twins. Future work needs to be done to replicate these findings.

Chapter 6 NOVEL GENE VARIANTS FOR SYNDROMIC CONGENITAL CRANIOFACIAL AND CARDIAC ANOMALIES

6.1 **SUMMARY**

Congenital birth disorders caused by aberrant neural crest function – be it induction, migration or differentiation – are numerous, yet a large subset remains undefined. The Clinical Study of child and family trios presented in Chapter 3, identified a subset of children with clinically and genetically undiagnosed syndromes involving clefts but who had other phenotypic features including congenital heart disease (CHD). Craniofacial anomalies and congenital heart disease often co-exist possibly due to their shared embryonic origins. The **aim** of this chapter was to identify novel candidate genes in patients with cranio-cardiac phenotypes and to develop new genome-phenome relationships. **Method:** A Complementary Research Proposal (CAP180) was retrieved from the Deciphering Developmental Disorders Study (DDD), a proposal aimed at analysing craniofacial developmental disorders caused by defects in neural crest migration and differentiation. Within a subset of the CAP180, I developed a cranio-cardiac dataset to explore the relationship between craniofacial dysmorphisms and CHD, filtered through the probands' phenotypes and utilized exome sequencing data to compare their genotypes for genes intolerant to variation. **Results:** Genes, gene variants, and copy number variants (CNVs) that are likely candidates for cranio-cardiac neurocristopathic malformations were found. These candidate genes were *EFTUD2*, *FBOX11*, *CELSR1*, *ABCA2* and *DIP2C*. While *EFTUD2* and *FBXO11* have already been explored in craniofacial disorders, I highlighted novel genome-phenome relationships in *CELSR1, ABCA2 and DIP2C*. **Conclusion:** Further research into *CELSR1, ABCA2* and *DIP2C* is needed to define novel syndromic craniofacial disorders, particularly exploring more subjects in other datasets and carrying out molecular assays investigating the pathogenicity produced by the novel mutations.

6.2 **INTRODUCTION**

Anomalies of the cardiovascular system and skeletal or vertebral malformations commonly occur with orofacial clefts (Calzolari *et al.*, 2007; Milerad *et al.*, 1997; Rittler *et al.*, 2011; Stoll *et al.*, 2000). Indeed, more than 60 syndromes comprise of cleft and congenital heart disease (Seto-Salvia & Stanier, 2014). These could either be due to shared genetic contributions (Homsy *et al.*, 2015), due to neural crest migration defects, or both. The Clinical Study of 90 family trios of children with cleft lip/palate (discussed in Chapter 3) has identified (N=28) children with an undiagnosed cleft mutation and a medical diagnosis of other multiple congenital anomalies. Indeed, syndromic phenotypes caused in part by *de novo* mutations have been shown to be are a rich source of novel discoveries (Chong *et al.*, 2015). To do this, often several affected persons with *de novo* mutations in the same gene can be identified nowadays either through recruitment from clinical studies or available exome sequencing datasets.

A significant proportion of craniofacial structures are formed of mesenchymal tissues that are derived embryonically from mesoderm and neural crest (Szabo-Rogers *et al.*, 2010). Neurocristopathies are a group of disorders that share an underlying pathology of the neural crest (Bolande, 1974; Bolande, 1997). It is now appreciated that conditions of the head, neck and heart are pathogenetically related as neurocristopathies (Bolande, 1997). Treacher Collins, CHARGE, 3MC, BOR and tricho-dento-osseous syndrome are craniofacial neurocristopathies commonly affecting the eyes, ears, mandible and palate (Jones *et al.*, 2008; Theveneau & Mayor, 2012; Vega-Lopez *et al.*, 2018). The neural crest is also important in cardiac septation (Kirby & Waldo, 1990; Schievink *et al.*, 1996) and is the precursor of the ascending aorta, intracranial arterial tree and the cardiac ventricular outflow tracts (Sattur *et al.*, 2016; Schievink *et al.*, 1996). Therefore, some forms of congenital heart disease originate anatomically from aberrant neural crest cells such as, persistent truncus arteriosus, bicuspid aortic valve, coarctation of the aorta, pulmonic valve stenosis and pulmonary artery stenosis (Kirby & Waldo, 1990; Schievink *et al.*, 1996).

With the ever growing era of exome and genome sequencing, large datasets of patient gene variants and phenotypes are now generated and are enriched with multiple novel and rare alleles (McClellan & King, 2010). The Deciphering Developmental Disorders Study (DDD) database contains genetic details and phenotypical features of clinically ascertained, genetically undiagnosed patients recruited to DDD by their individual genetics services; a family trio design is undertaken

and family trios are included where possible (Wright *et al.*, 2015). At present, diagnoses for 35-40% of DDD patients have been made and around 90% of the candidate diagnostic variants from DDD have been communicated back to the referring NHS Clinical Genetics Services (www.ddduk.org). Furthermore, the DDD Study team has discovered more than 30 genes not previously associated with developmental disorders and a total of around 157 peer-reviewed publications have been produced from data utilized from the DDD Study (www.ddduk.org). Given the continual rapid advancement in the area of genomics and gene discovery, often revisiting large datasets as such also aids in the identification of new correlations and increases diagnostic rates (Wright *et al.*, 2018b). The aim of this project was to draw novel phenome-genome links by utilizing clinical and exome sequencing information from a large-sequencing dataset (DDD) and clustering patients with rare craniofacial developmental disorders by phenotype and genotype, particularly reinforcing the contribution of previously uncategorized gene variants to novel craniofacial and cardiac-associated disorders.

6.3 **PROTOCOL**

The phenotypic features that were found in a group of children in the Clinical Study who had medical anomalies known to be associated with clefts were used to inform the protocol for interrogating the DDD dataset.

6.3.1 **Protocol Development**

6.3.1.1 Initiation of the CAP180 dataset for neurocristopathies

A research proposal was submitted to the Deciphering Developmental Disorders (DDD) Consortia through a Complementary Research Proposal (CAP)- named CAP180. To develop the CAP180 dataset, subjects with one of the following phenotypes, based on their clinician's annotation, were added (these are based on consensus HPO terms)(Köhler *et al.*, 2018):

- Mandibulofacial dysostosis; HP:0005321
- Abnormal facial shape; HP:0001999
- Abnormality of facial skeleton; HP: 0011821
- Abnormality of mandible condylar process; HP:3000077
- Abnormality of mandible coronoid process; HP:3000078
- Abnormality of mandibular ramus; HP:3000003
- Abnormality of malar bones; HP:0012369
- Hemifacial hypoplasia; HP:0011332
- Hemifacial atrophy; HP:0011331

6.3.1.2 Variant calling parameters applied in the CAP180

Variants in CAP180 are all quality control (QC)-passed, and include SNVs (single nucleotide variants), indels (insertions/deletions) and CNVs (copy number variants). Variants were called using GATK, SAMtools and Dindel for SNV and indels, and in-house CoNVex for CNVs. The reference human genome used was build GRCh37 (hg19). High quality candidate variants in probands were defined in the CAP180 as follows:

- MAF <0.1% in trios and <0.01% in singletons for dominant variants and ExAC count<5.
- MAF <0.1% in trios and singletons and ExAC hemizygous count=0 for hemizygous variants; MAF <1% in trios and singletons for recessive variants.

- VEP annotation predicts the variant with the most severe consequence will be loss-of function or protein altering; inherited missense variants predicted to be benign by PolPhen2 are excluded; CNVs (deletions or duplications) >1MB.
- Genotype is consistent with a dominant (*de novo* or dominantly inherited from affected parent), recessive (homozygous or compound heterozygous) or X-linked (hemizygous) mechanism.
- Inheritance (where known) is consistent with a dominant (*de novo* or dominantly inherited from affected parent), recessive (homozygous or compound heterozygous) or X-linked (hemizygous) mode of inheritance.

Filtering through the MAF (minor allele frequency) and the variant consequence (type) reduces the number of variants down to a limited number of potential candidate variants. Assuming a dominant mode of inheritance, the MAF was set to <0.01. Initially the following variant types were included: splice donor variant, splice acceptor variant, stop gained, frameshift variant, stop lost, initiator codon variant, inframe insertion, inframe deletion, missense variant, transcript amplification, coding sequence variant, in addition to missense variants within regulatory regions or untranslated (UTR) regions and missense variants with deleterious SIFT and Polyphen scores.

6.3.1.3 Identifying subjects with cranio-cardiac malformations

Because a number of patients in the Clinical Study presented with congenital heart diseases, I collated and established a 'cranio-cardiac' sub-dataset for those who had palate and heart anomalies from the CAP180. The datasets described in this chapter are shown in **(Figure 6-1)**.

First, palatal phenotypes were searched for by using Human Phenotype Ontology (HPO) terms under the category 'Palate'(Köhler *et al.*, 2018). Phenotypes for those with a palatal condition were scrutinized to identify those that had a cardiovascular deformity **(Figure 6-2).** To facilitate sifting through the candidate gene variants for the CAP180 probands, patients on DECIPHER with undiagnosed syndromes were also identified (Firth *et al.*, 2009), by searching specifically for those that had 'Abnormality of the head or neck' (HP:0000152) and 'Abnormality of the cardiovascular system' (HP:0001626) **(Figure 6-3).**

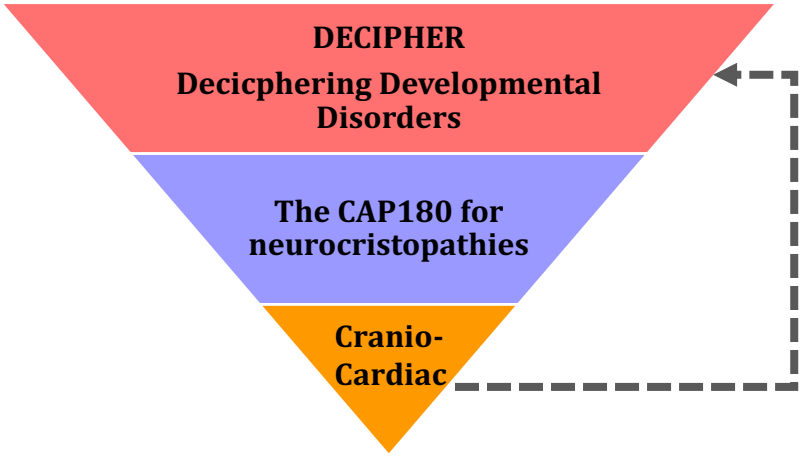


Figure 6-1 The datasets referred to in Chapter 6

The Deciphering Developmental Disorders Study which feeds into the larger DECIPHER online database was used to generate the CAP180 dataset. The CAP180 includes patients with anomalies thought to represent neurocristopathies. From the CAP180, I generated a smaller dataset for patients with palatal and cardiac conditions. The arrow indicates that after subjects and genes were identified through CAP180 and the cranio-cardiac dataset, further subjects with same gene changes were looked for in the larger DECIPHER dataset to find any overlapping or missed subjects.

6.3.1.4 Identifying candidate gene variants in subjects with cranio-cardiac malformations

Once cranio-cardiac patients from both datasets, CAP180 and DECIPHER, were identified, variants were compared in order to identify novel genes implicated in cranio-cardiac disease **(Figure 6-3).** Particularly, genes in cranio-cardiac DECIPHER subjects with known *de novo* mutations were prioritized and cross-checked with variants found in the CAP180 subjects. The RVIS (Residual Variation Intolerance Scores), GDIS (Gene Damage Index Score) and gene constraint (missense Z-scores and probability of being loss-of-function intolerant (pLI) scores) were checked in order to distinguish whether the genes that were found are more intolerant to variation/mutation (Appendix 6-A & Appendix 6-B). The pathogenicity of the variants, when given, was classified according to American College of Medical Genetics (ACMG) standards and guidelines as likely pathogenic and pathogenic scores of ($0.9 < p \leq 0.99$) and ($p \geq 0.99$), respectively (Richards *et al.*, 2015).

6.3.1.5 Identifying and categorizing additional subjects with variants in the putative candidate genes

Following this initial surveillance, all cases in CAP180 and the broader DECIPHER dataset with variants in the candidate genes that were found were surveyed regardless of their initial presentation and were tabulated to examine phenotype commonality.

Once all the data was compiled, variants of these subjects along with the variants of the cranio-cardiac patients were checked, tolerant gene variants and genetic changes that were commonly found in the general population were excluded. For each gene, **tables of variants** and **tables of phenotypes** were compiled (all can be found in Appendix 6).

6.4 **RESULTS**

6.4.1 **The Medical and Dental Profile of Children from the Clinical Study with Associated Medical Conditions and Their Families**

From 90 children recruited to the Clinical Study, 28 had a medical condition associated with their cleft **(Table 6-1)**. Of the 90 children, 26 of them had previously undergone some form of genetic testing (such as an array comparative genomic hybridization (CGH) or candidate gene tests) but their results had been inconclusive. Of the 26 children, five were otherwise medically healthy (reported in Chapter 3) and 21 had other systemic medical diagnoses. During the course of this study, four were subsequently diagnosed to have a cleft-associated syndrome, namely, Kabuki, van der Woude, holoprosencephaly or 3MC syndrome.

Full details of those with 'associated' clefts are shown in **(Table 6-1)**. Some of the conditions such as neurodevelopmental, skeletal and cardiovascular anomalies were equally present in both cleft lip/palate and cleft palate only categories. However, conditions affecting the kidneys, brain and metabolic disorders were more prevalent in the cleft lip/palate group. Almost all children with a medical co-morbidity had various dysmorphic craniofacial features.

In the present study, more than half of the children with clefts had more dental anomalies outwith the cleft area than the general population, these were mainly hypodontia, microdontia, dens invaginatus and taurodontism **(Table 6-2).** Interestingly, 11/25 of those that had hypodontia and/or taurodontism outside their cleft site were from the group with 'clefts-medical/congenital anomalies'. Moreover, parents/siblings of children in this group appeared to have more dental anomalies (53.6%) compared to parents/siblings of children with non-syndromic clefts (42.3%) (**Table 6-2** and **Table 3-2** in Chapter 3).

Since a total of 9/28 subjects were diagnosed with congenital heart disease associated with their cleft conditions, the genotypic findings for cranio-cardiac malformations was the focus of the exome-sequencing dataset interrogation.

Table 6-1 The children's medical co-morbidities

Cleft type Number of children with clefts N=90	Classification of co-morbidity Number of children with 'associated' clefts N=28	Type of medical condition in children with 'associated' clefts	Number of children with co-morbidity
Cleft lip only (14) right (7) left (7) bilateral (0) **Cleft lip & palate (37)** right (9) left (16) bilateral (12) **Cleft lip & alveolus (13)** right (3) left (8) bilateral (2)	N=17		
	Neurodevelopmental	Speech delay, ASD/ADHD, developmental delay	**9**
	Brain	Missing/hypoplastic corpus callosum; epilepsy; missing septum pellucidum, hydrocephalus; polymicrogyria; subependymal/periventricular cysts	**8**
	Cardiac	Congenital cardiac disease	**4**
	Metabolic	Auto immune hypothyroidism; premature puberty; hypopituitarism; diabetes; ketotic hypoglycaemia	**7**
	Skeletal	Skeletal abnormalities; stature anomalies and rheumatology; scoliosis; joint hypermobility; pectus excavatum; low spine termination	**11**
	Kidneys	Congenital Kidney problem Genito-urinary condition	**5** **3**
	Sensory-neural	hearing deficit; vision problems; 8th cranial nerve (vestibule cochlear) aplasia or severe hypoplasia	**7; 2; 2**
	Limbs	Polydactyly, clinodactyly	**2**
	Others	Asthma/allergies; GORD; alloimmune thrombocytopaenia	**6; 1; 1**
Cleft palate only (26) hard palate (22) submucous and VPI (4)	N=11		
	Neurodevelopmental	Speech delay, ASD, global developmental delay	**5**
	Brain	microcephaly, bilateral deep lesions of white matter	**1**
	Cardiac	Congenital cardiac disease	**5**
	Skeletal	Skeletal abnormalities, stature anomalies and rheumatology; scoliosis; joint hypermobility; spinal abnormalities; hip dysplasia	**6**
	Sensory-neural	hearing deficit; vision problems/retinopathy; 6th & 7th cranial nerve palsy	**3; 5; 1**
	Limbs	femoral hypoplasia; missing digit; clinodactyly, foot clubbing	**4**
	Vocalisation	Voice problems/quality of the voice; Laryngomalacia	**3**
	Others	GORD; abdominal hernia; congenital urinary problem; prolonged bleeding; vit D deficiency	**1; 1; 3; 1; 1**

Table 6-2 Comparison between dental anomalies in probands and parents/siblings in non-syndromic patients (Chapter 3) vs. those with 'cleft-medical condition'

Dental Anomaly	**Children with non-syndromic clefts N=62**	**Children with 'cleft-medical condition' N=28**
Hypodontia	26 (42%)	15 **(53.6%)**
Peg laterals	24 **(38.7%)**	8 (28.6%)
Microdontia	4 (6.5%)	4 **(14.3%)**
MIH	10 (16%)	6 **(21.4%)**
	Parents/family members N=87	**Parents/family members N=40**
Hypodontia	10 (11.5%)	6 **(15%)**
Peg laterals	4 (4.6%)	4 **(10%)**
Microdont molars	6 (5.7%)	2 (5%)
MIH	13 **(15%)**	4 (10%)
Megadonts	0	3 **(7.5%)**

6.4.2 The Cranio-Cardiac Population Derived from the CAP180 Dataset for Neurocristopathies

Following the assembly of CAP180 dataset, a large number **(~32,769)** of quality control (QC)-passed variants from **1547** probands were available. I first identified patients from the CAP180 that were diagnosed by their referring geneticists to have both cardiac malformations and cleft lip and palate phenotypes **(Figure 6-2)**. Thirty-four terms were identified and summarized, and probands with diagnoses with one of these categories were identified **(Table 6-3).** Each proband would have several high-quality candidate variants that have been identified by DDD following exome sequencing.

Forty-three genes replicated in both DECIPHER and CAP180 patients with cardiac and cleft/head & neck malformations. Of the 43 genes assessed, the following five genes met all the criteria: *EFTUD2, FBXO11, CELSR1, ABCA2* and *DIP2C*. Their ExAC constraint results are shown in (Appendix 6-B). Additional subjects with variants in these genes were identified in order to establish a wider genome-phenome correlation.

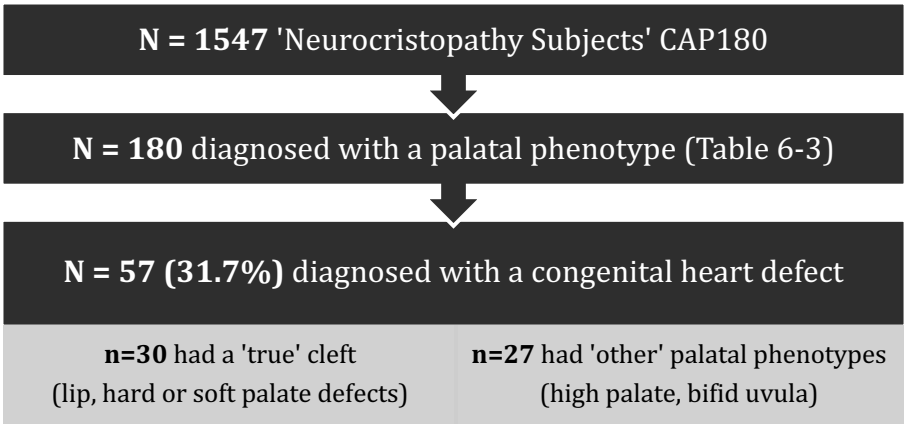


Figure 6-2 Number of probands in the CAP180 dataset that were diagnosed with a congenital palatal phenotype <u>and</u> congenital heart disease

Table 6-3 'Palate' anomalies in the CAP180

Term ID	HPO Disorder	Number affected in CAP180
HP:0000174	Abnormality of the palate	3
HP:0000175	Cleft palate	59
HP:0000176	Submucous cleft hard palate	9
HP:0011819	Submucous cleft soft	1
HP:0000185	Cleft soft palate	15
HP:0000202	Oral cleft	4
HP:0009099	Median cleft palate	14
HP:0002744	Bilateral cleft lip and palate	3
HP:0008501	Median cleft lip and palate	2
HP:0100334	Unilateral cleft palate	4
HP:0000189	Narrow palate	4
HP:0000218	High palate	49
HP:0002705	High, narrow palate	5
HP:0010290	Short hard palate	1
HP:0010650	Hypoplasia of the premaxilla	1
HP:0010759	Prominence of the premaxilla	2
Total		**180**

'Palate' HPO disorders as identified by the Human Phenotype Ontology (https://hpo.jax.org) and the total number of palatal phenotypes in CAP180 probands is shown.

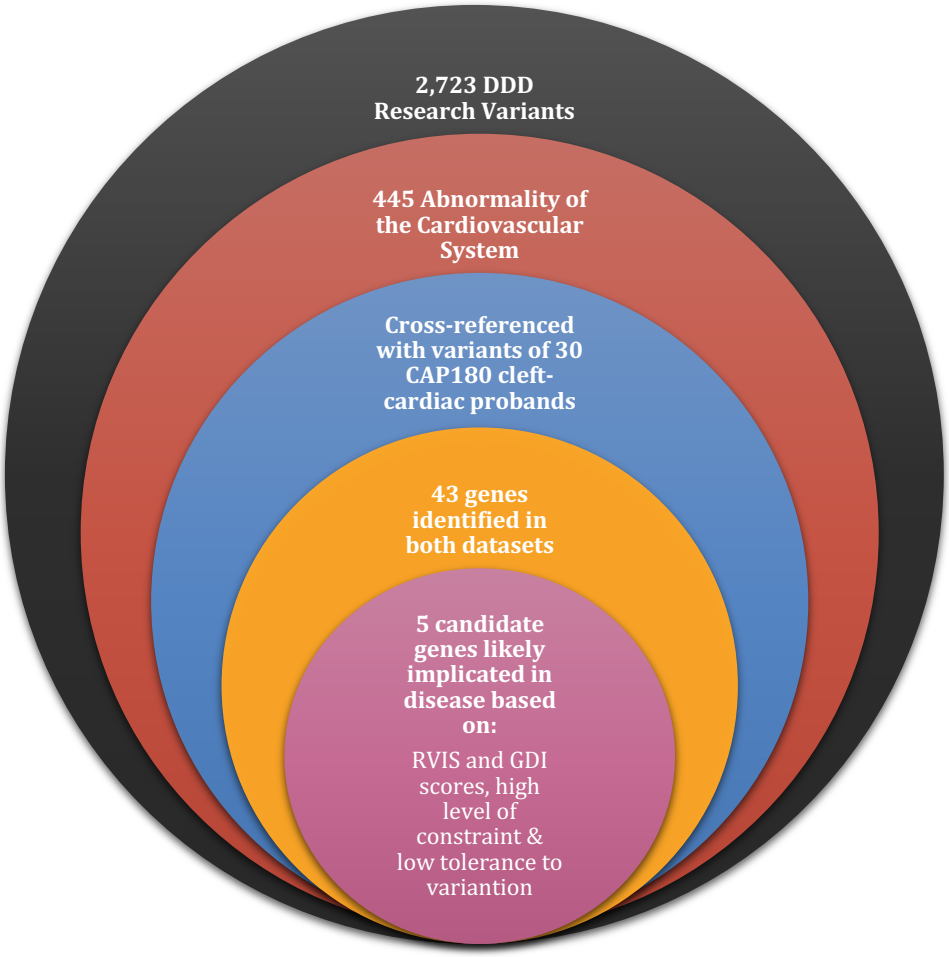


Figure 6-3 Flowchart showing patients identified in DECIPHER and then explored in the CAP180 cardiac-cleft subjects

Forty-three genes replicated in both DECIPHER and CAP180 patients with cardiac and cleft/head & neck malformations. Following filtering and prioritisation of candidate genes, only five met my inclusion criteria. These were: *EFTUD2, FBXO11, CELSR1, ABCA2* and *DIP2C.*

6.4.3 **Putative Genes and Phenotypes Identified from the Protocol**

The following genes were identified using the pipeline developed above: ***EFTUD2*** (elongation factor Tu GTP binding domain containing 2), ***FBXO11*** (F-box protein 11), ***CELSR1*** (cadherin EGF LAG seven-pass G-type receptor 1), ***DIP2C*** (disco interacting protein 2 homolog C) and ***ABCA2*** (ATP binding cassette subfamily A member 2).

Twelve patients that fit the criteria outlined were found to have mutations in *EFTUD2*, of which ten were *de novo* mutations. Ten had pathogenic or likely pathogenic single nucleotide variants (SNV), three frameshift, three missense, three splice donor, and one stop-gained mutation. Two other subjects had copy number variants (CNV) (<12kb deletions) encompassing *EFTUD2* (Appendix 6-C). Of the phenotypes observed, two with an SNV and one with a CNV had cleft of the soft or hard palate. Six with an SNV were diagnosed with laryngeal/oesophageal phenotypes which included laryngeal clefts, oesophageal atresia and tracheoesophageal fistula. Microcephaly was seen in (11/12) subjects, sensorineural or conductive hearing impairment in (7/12), anatomical auricular malformations in (7/12), micrognathia (4/12), choanal atresia (4/12), facial asymmetry (5/12) and skeletal/limb anomalies (8/12). Interestingly, congenital heart disease was common among the *EFTUD2* cohort described here, with (10/12) patients affected; of which total and partial anomalous pulmonary venous return were seen in two patients with SNV and one patient with CNV, respectively. Other clinical details are tabulated (Appendix 6-D).

Nine patients with *FBXO11* variants have been identified, seven of which were *de novo* mutations. Six of those were excluded as these subjects were reported elsewhere (Fritzen *et al.*, 2018; Gregor *et al.*, 2018). The clinical description of the remaining patients is described in (Appendix 6-E & Appendix 6-F).

Ten patients had variants in *CELSR1* (Appendix 6-G), five of them were compound heterozygotes (i.e. they are probably in strong linkage disequilibrium), of which three (Patients 2-4) had similar compound heterozygous mutations, one of which was novel and the other common. Whether these variants in compound are pathogenic is unknown. Multiple genetic testing was carried out on these individuals prior to their inclusion for exome sequencing. Results, however, have always been inconclusive (a summary of those tests is found in Appendix 6-I). All patients had orofacial phenotypes, of which one had a cleft soft palate and four had congenital cardiovascular anomalies (Appendix 6-H).

In this study I found previously unreported single and copy number variants in *DIP2C*. Eighteen patients were found, four with single nucleotide variants and 14 with copy number variants in the 10p15.3 region containing *DIP2C* (Appendix 6-J). Phenotypic analysis revealed craniofacial dysmorphisms particularly with the ears and eyes and abnormalities in the kidneys and cardiovascular system as well as muscular hypotonia and seizures as recurrent features (Appendix 6-K & Appendix 6-L).

I also found previously unreported gene and copy number variants in *ABCA2*. Twelve subjects were found to have variants in *ABCA2*, nine with SNVs and three with CNVs encompassing *ABCA2* (Appendix 6-M). Palatal phenotypes were seen in (2/9) subjects with SNVs and congenital heart disease in (3/9). Craniofacial anomalies, in general, were identified in the majority of subjects (9/12) and all had some form of neurodevelopmental condition (Appendix 6-N). Skeletal conditions were observed, three with generalized hypotonia, two with short stature and two were scored by their referring clinicians to have skeletal abnormalities under the HPO term 'Abnormality of the skeletal system' [HP:0000924] and one had an additional 'Growth abnormality' [HP:0001507]. Of interest, Subject 6 (p.Arg2265Trp) and Subject 7 (p.Arg1513His) were both thought to be CHARGE-like phenotypically and were therefore tested for *CHD7*, however, the test results were normal.

6.5 DISCUSSION

6.5.1 Phenotypic Aspects from the Clinical Study and the Protocol

In this chapter, I identified a subset of cleft subjects with associated medical anomalies through clinical recruitment who did not have an identified syndrome. Exome sequencing and gene identification for two of the cases has been described in detail in Chapters 4 and 5 of this thesis. Since nine subjects were identified through the Clinical Study to have congenital heart disease associated with their clefts, I retrospectively analysed the CAP180 neurocristopathy dataset for craniofacial and cardiac anomalies. The protocol I developed for filtering through patient phenotypes revealed novel links to genes that had not been previously implicated in syndromic cranio-cardiac diseases. Dissecting published literature on the identified genes recognized two genes shown to associate to known human disorders encompassing craniofacial and cardiac manifestations, giving some proof of principle for the pipeline. I focused on three genes (*CELSR1*, *DIP2C* and *ABCA2*) that are less well characterised and that have not been linked to congenital craniofacial or cardiac malformations and I reported on novel single nucleotide variants in some of those genes that have been shown to encompass chromosomal microdeletions.

Multiple congenital anomalies (MCA) are common in cleft populations (Calzolari *et al.*, 2007; Milerad *et al.*, 1997). Although studies with larger sample sizes were carried out (Calzolari *et al.*, 2007; Milerad *et al.*, 1997; Rittler *et al.*, 2011; Stoll *et al.*, 2000), similar figures and trends in prevalence of associated congenital anomalies were consistent with the current findings, despite the methodological differences that may hinder reliable comparisons. A strength of the current study, however, is that all children recruited with 'associated' clefts were on average more than 10 years of age (with the exception of a 2-year old), indicating that any other major or minor associated anomalies would have been diagnosed, genetic tests (chromosomal test) would have been carried out and ruling out 'true' isolated clefts from syndromic clefts would have taken place.

Regarding the dental anomalies seen, hypodontia, microdontia and molar-incisor-hypomineralisation (MIH) were more prevalent in the 'cleft-medical condition' group with a similar trend for their parents/siblings for hypodontia and peg laterals. Megadonts were only seen in parents/siblings of the 'cleft-medical condition' group whom interestingly had zero 'isolated' enamel defects.

In the present study, although most hypodontia cases involved teeth commonly known to be missing, some individuals had missing mandibular first and/or second permanent molars or permanent maxillary central incisors, providing evidence for the importance of the inclusion of dental anomalies within the overall 'multiple congenital anomalies' screened for in clinical surveys involving cleft subjects. These dental anomalies could provide the first sign indicative of a more severe systemic condition. Indeed, one of the subjects recruited through the clinical survey presented at 15-years of age with a missing central incisor, bilateral cleft lip and palate and intellectual disability. Referral to clinical genetics and sequencing of this patient revealed an undiagnosed case of holoprosencephaly. The finding that the large majority of children with medically 'associated' clefts (21/28) clearly had undiagnosed conditions highlights that previous methods were insufficient for the diagnosis of these cases and emphasizes the need for further genetic studies to identify novel candidate genes. Genetic findings for two of the subjects included in the current study's exome sequencing pipeline were discussed in Chapter 4 (*CTNND1* mutation) and Chapter 5 (*AGAP6* mutation).

Recent research into genotype-phenotype relationships has highlighted the link between subclinical features in patients and family members, particularly in cleft cohorts (Howe *et al.*, 2015; Leslie *et al.*, 2017). However, the opposite relationship, i.e. using the association of phenotypic similarities in probands-only to suggest common aetiologies has increased in popularity with next generation sequencing study designs. By clustering subjects based on their phenotypes and linking those back to their shared genotypes either expands on the phenotypic spectrum for certain genes or helps identify novel links for previously unexplored genes to new syndromes. Reviewing DECIPHER patients in addition to those included in the CAP180 had several advantages. First, the CAP180 is a smaller subset of a larger DDD/DECIPHER cohort, both datasets were, therefore, scrutinised to find any overlapping or missed subjects. Another advantage in investigating DECIPHER 'research' cases is that these patients were exome sequenced, genotyped and their other candidate variants excluded with only a handful of candidate genes remaining that are likely contributory to their conditions. These are either new genes that have not been previously linked to human disorders or genes that have been linked to human conditions, but the patient's phenotypic picture is not fitting with known syndromes caused by the candidate gene. These genes are often assigned the term 'genes of uncertain significance' (Richards *et al.*, 2015). Therefore, examining these gene variants from the DECIPHER website provided further validity for the identified genes through the CAP180.

6.5.2 Known Genes in Developmental Disorders Involving Craniofacial and Cardiac Conditions

The findings of the *EFTUD2* and *FBXO11* genes in this project illustrates that they may be proof of principle genes for the pipeline utilized in the current study. A total of 107 individuals with pathogenic *EFTUD2* variants comprising 76 SNVs and seven microdeletions have been reported (Huang *et al.*, 2016). Mutations in *EFTUD2* have recently been well characterized in individuals with 'Mandibulofacial Dysostosis with Microcephaly' (MFDM) syndrome, also known as Guion-Almeida [MIM: # 610536], which often phenocopies Treacher Collins (Lehalle *et al.*, 2014; Lines *et al.*, 2012; Luquetti *et al.*, 2013; Smigiel *et al.*, 2015; Vincent *et al.*, 2016). Craniofacial malformations with a recognizable dysmorphic appearance, microcephaly and some major sequelae including choanal atresia, sensorineural hearing loss, and cleft palate are some of the manifestations seen (Lines *et al.*, 2012). There are also reports expanding on the spectrum of anomalies which include oesophageal atresia (Gordon *et al.*, 2012; Voigt *et al.*, 2013). Here I report additional cardiac and limb anomalies associated with *EFTUD2* variants. Despite cardiac defects occurring in 31.5% of MFDM cases to date (Yu *et al.*, 2018), only a few studies reported the association of congenital heart disease mostly manifesting as atrial and ventricular septal defects. Interestingly, the total anomalous pulmonary venous return phenotype shown recurrently in my cohort was seen in only (1/36) patients reported to date. Although *EFTUD2* is conserved in mammals, functional characterizations using knock-out models do not exist. Patient manifestations suggest that *EFTUD2* could be classified as a neurocristopathy gene since the clinical picture of *EFTUD2* phenocopies that of Treacher Collins, a known neurocristopathy and because of the association of oesophageal atresia to *EFTUD2* mutations. Oesophageal atresia, which often occurs concurrently with congenital heart disease (Morini *et al.*, 2001), is an anatomical defect of the pharyngeal arches related to abnormalities in cephalic neural crest differentiation (Morini *et al.*, 2001).

Human diseases have only recently been linked to mutations in *FBXO11* (Fritzen *et al.*, 2018; Jansen *et al.*, 2019; Lelieveld *et al.*, 2016; Martinez *et al.*, 2017), particularly in syndromic forms of neurodevelopmental disorders (Intellectual Developmental Disorder with Dysmorphic Facies and Behavioural Abnormalities [MIM: # 618089]) (Fritzen *et al.*, 2018; Gregor *et al.*, 2018; Jansen *et al.*, 2019). Palatal defects, including uvula anomalies and cleft palate have been observed in (5/20) individuals with missense mutations and a high-arched palate and cleft lip and alveolus were seen in other subjects who had a frameshift mutation and a nonsense mutation, respectively

(Gregor *et al.*, 2018). Two patients were previously found to have cardiac anomalies including an atrial septal defect (in which the patient also had a cleft palate) and a mild aortic dilation (Gregor *et al.*, 2018). Despite individuals with *FBXO11* mutations showing dysmorphic features, no consistent facial gestalt has been reported (Jansen *et al.*, 2019). Although *FBXO11* has been linked to syndromic neurodevelopmental disorders, the cleft phenotypes seem quite penetrant in the various cohorts observed. Accordingly, *FBXO11* can be considered in the differential diagnoses of syndromic cases of orofacial clefting associated with cardiac, eye and neurodevelopmental disorders.

6.5.3 Novel Links to Relatively Unexplored Genes in Craniofacial and Cardiac Conditions

6.5.3.1 CELSR1

CELSR1 human mutations have been linked to a spectrum of neural tube defects (NTDs) ranging from craniorachischisis, a condition where the neural tube remains open from the midbrain/hindbrain boundary throughout the spinal region, to much milder caudal agenesis defects (Lei *et al.*, 2014; Qiao *et al.*, 2016; Robinson *et al.*, 2012). Four studies reported a total of 34 pathogenic variants in *CLESR1* absent from their respective control populations. Most of the mutations were missense and only (3/34) were indels and one reported a non-sense mutation (Allache *et al.*, 2012; Lei *et al.*, 2014; Qiao *et al.*, 2016; Robinson *et al.*, 2012). How the missense variants affect neural tube closure has not been determined and the severity shown in the NTDs are likely due to synergistic effects from other mutations (Robinson *et al.*, 2012). A large-scale study conducted on a Chinese population identified novel pathogenic *CLESR1* gene variants in a cohort of patients with NTDs and another subset of patients with congenital heart disease (CHD) and only one novel missense mutation in *CELSR1* (p.P870L) was shared amongst both groups. This variant was found to be a gain-of-function mutation up-regulating the planar cell polarity (PCP) pathway and canonical WNT signalling in cells and also induced both NTDs and CHDs in zebrafish embryos (Qiao *et al.*, 2016). These authors were the first to correlate mutations in *CELSR1* to congenital heart disease.

In mice, *Celsr1,* which is a core component of PCP, was found to be essential for initiation of neural tube closure (Murdoch *et al.*, 2014). As Qiao and co-workers (2016) alluded to, our group previously demonstrated that the Fuz PCP protein showed similar effects on mice producing neural tube defects and heart anomalies (Gray *et al.*, 2009). Moreover, our group recently linked a missense mutation in *FUZ* to human craniosynostosis (in press); a phenotype we observed in our mouse knock-outs (Tabler

et al., 2016), suggesting that PCP proteins such as FUZ and CELSR1 have important roles in the aetiology of craniofacial dysmorphisms in humans (Szabo-Rogers *et al.*, 2010). The study by Qiao *et al.* (2016) highlights the potential expansion of the phenotypes implicated in *CELSR1* mutations. Since my cohort did not have neural tube defects, and since previous reports on *CELSR1* mutations have not reported craniofacial and other anomalies, I have described a novel subset of phenotypes linked to variants in *CELSR1*.

6.5.3.2 DIP2C

Upon literature analysis of genes encompassing the 10p15 region, several studies have described a '10p15.3 deletion syndrome'. Two genes were involved in this region, *DIP2C* and *ZMYND11*; the last exon of *ZMYND11* is around 20kb upstream of the first exon of *DIP2C*. Indeed, DeScipio and co-workers (2012) described a group of patients with subtelomeric deletions in 10p15.3. Collectively, their cohort showed a multitude of anomalies that phenocopied the ones described for the *DIP2C* cohort presented here. These included fully penetrant traits such as intellectual disability and developmental delay, structural abnormalities in the brain, craniofacial dysmorphisms, hypotonia, seizures and limb anomalies. Structural cardiac abnormalities were present in (2/9) of their cohort, one of which was the proband with the microdeletion that only included *ZMYND11* but not *DIP2C*. The features reported in their 'other' category were anomalies often recurring in the present study, these included skeletal abnormalities comprising spinal anomalies and quadriplegia, recurrent respiratory infections, structural gastrointestinal conditions and constipation, genitourinary problems, specific skin pigmentations and hair anomalies. Looking particularly at the proband that had the copy number variant that only included *DIP2C*, the patient was reported to have repetitive behaviours, craniofacial dysmorphism including a high-arched palate, a resolved heart murmur, mild hypotonia and hyper-extendible hands and webbed toes (DeScipio *et al.*, 2012). Likewise, both *ZMYND11* and *DIP2C* were re-sequenced in a study that searched for single nucleotide variants in a neuropsychiatric cohort with identified copy number variants in 10p15.3 microdeletions (Coe *et al.*, 2014). Five truncating variants in *ZMYND11* and none in *DIP2C* were found.

Single nucleotide variants in *DIP2C* have not been widely explored in congenital syndromes affecting craniofacial and cardiac systems. However, they have been reported in two cases of cerebral palsy. No craniofacial or cardiac anomalies were reported in these cases (Zarrei *et al.*, 2018). Iossifov and co-workers (2014) tested the genetic contribution to autism spectrum disorder by examining 2,500 simplex families.

One missense and two frameshift variants in *DIP2C* were reported. Additional phenotypes were not described in the paper. Most recently, Maddirevula *et al.* (2018) reported a *DIP2C* variant in a child with a novel form of skeletal dysplasia, born of consanguineous parents. The proband was short in stature, had a short humerus and bilateral shortening of the femora, a hypoplastic middle phalanx of the middle finger, and absent 1st left phalanx; in addition to other limb anomalies (Maddirevula *et al.*, 2018). The only craniofacial anomaly reported was mid-face hypoplasia. Cardiovascular anomalies were not reported. In summary, I highlighted the phenotypes of individuals with single nucleotide variants in *DIP2C* and Patient 5 with the 10p15.3 microdeletion that does not include *ZMYND11* (Appendix 6-J), revealing considerable clinical overlap to those with the '10p15.3 microdeletion syndrome' described by (DeScipio *et al.*, 2012). Although congenital heart disease was not reported in papers documenting SNVs in *DIP2C* (Iossifov *et al.*, 2014; Maddirevula *et al.*, 2018; Zarrei *et al.*, 2018), I reported two individuals with cardiovascular anomalies. Delineating phenotypes in subject with single nucleotide variants in *DIP2C* vs. those with *ZMYND11* helps predict the pathologic mechanisms behind selective phenotypes.

6.5.3.3 ABCA2

The 9q34.3 chromosomal region encompasses both *NOTCH1* and *ABCA2.* During my analysis of variants in *ABCA2,* I excluded many subjects with copy number variations (<4Mb) in the 9q34.3 region since *NOTCH1* is only 500kb upstream of *ABCA2*. *NOTCH1* has been implicated in Adams Oliver Syndrome [MIM: # 616028] and in Left Ventricular Outflow Tract Obstruction (Aortic valve disease 1) [MIM: # 109730]. (Helle *et al.*, 2019; McBride *et al.*, 2008). I also scrutinised phenotypes in subjects with *ABCA2* mutations to assess whether there was any overlap with those affected by *NOTCH1* mutations. Adams-Oliver syndrome (AOS) is a rare developmental disorder defined by the combination of aplasia cutis congenita of the scalp vertex (a patchy round skin lesion in the centre of the scalp) and terminal transverse limb defects (e.g., amputations, syndactyly, brachydactyly, or oligodactyly). Vascular anomalies such as cutis marmorata telangiectatica congenita, pulmonary hypertension, portal hypertension, and retinal hypervascularization are recurrently seen. Congenital heart defects have been estimated to be present in 20% of the cases. From the analysis of the literature, patients with Adams-Oliver are not typically known to exhibit craniofacial anomalies. Mutations in *NOTCH1* are also implicated in congenital heart disease (CHD). In a large-scale exome sequencing study of 1213 CHD parent-offspring trios, Homsy and co-workers (2015) described three probands with *NOTCH1* mutations. The cohort with

ABCA2 variants described in this thesis had some phenotypic overlap with *NOTCH1* subjects particularly some neurologic anomalies (the cerebellar defects) and limb anomalies. However, the skeletal phenotypes were quite different. To differentiate between the two, the scalp lesions in *NOTCH1*-affected patients could possibly be the cardinal feature distinguishing those with *NOTCH1* mutations from others.

Mutations in *ABCA2* have not been widely linked to human syndromes. So far, a few case reports have described individuals with *ABCA2* variants. A recent study reported *ABCA2* as a novel autosomal recessive disease gene in two families with global developmental delay and intellectual disability (Maddirevula *et al.*, 2019). The authors concluded that their cohort could explain a potentially distinct allelic disorder caused by *ABCA2* than the previously reported clinical presentation of amyotrophic lateral sclerosis (ALS) (Steinberg *et al.*, 2015). Most interestingly, the two probands reported by Maddirevula had phenotypes that overlapped the ones reported here. Notably, diffuse hypotonia was reported in one of the patients who additionally had mild dysmorphic features. Although the second individual had apparent lack of gross dysmorphisms, she was microcephalic, had a small pituitary gland, a phenotype seen in the MRI scan of Patient 6 described here, and had significant internal rotation of the hips. All patients reported in the current study had a form of neurodevelopmental disorder and global developmental delay, suggesting a link between variants in *ABCA2* with syndromic forms of intellectual disability. Variants in *ABCA2* were also reported in Alzheimer's disease (Mace *et al.*, 2005) and recently a homozygous mutation was reported in siblings with ataxia and dysarthria from a consanguineous family (Aslam & Naz, 2019). Interestingly, the siblings in the latter report had an ataxic gait; Patient 7 in the current study also manifested a broad-based gait, typical of an ataxic form. However, an MRI scan was not reported for this child. Finally, analysis of additional DECIPHER patients with copy number variants in the 9q34.3 region that contains both *ABCA2* and *NOTCH1* revealed six phenotypically relevant cases. The limb and cardiac anomalies phenocopied those reported in *NOTCH1* mutations and the craniofacial anomalies phenocopied the ones associated with *ABCA2.*

6.6 **OUTSTANDING QUESTIONS**

What are the genetic contributions in the probands with undiagnosed clefts associated with other congenital anomalies? The exome sequencing carried out in Chapters 4 & 5 partly answered this question. Moreover, two interesting questions arose from the database study reported in this chapter: a) if the same analysis was carried out on the CAP180 group with 'other' palatal phenotypes, such as the high-arched palate, would the same genes identified here replicate as candidate variants or would a new subset of candidate genes be identified? b) are the genes and gene variants identified in this study reproducible in other large-scale datasets? To answer the first question, I am planning to test the same 'pipeline' I developed in this chapter to sift through patient genes and phenotypes for those with other palatal variations, in conjunction with congenital heart disease. To address the latter question, I am planning to look for subjects with mutations in the candidate genes identified here (*CELSR1, DIP2C, ABCA2*) from a large-scale cranio-cardiac exome-sequencing dataset developed by the team at the Paediatric Genetics Department, Yale Medical School. Once these variants are established and verified in other datasets, future research would involve developing assays to investigate molecular functions of these genes and understand the nature of the mutations in various biological contexts.

6.7 **CONCLUSIONS**

The use of exome-sequencing datasets to cluster patients with undiagnosed developmental disorders based on their clinical manifestations and underlying genetic variants could in turn provide a better understanding of the underlying mechanisms by which mutations in a particular gene attribute to the phenotypes collectively observed in the patients. The possibility that the craniofacial anomalies including orofacial clefting and/or cardiovascular phenotypes not being fully penetrant is not unlikely. Therefore, not only is the phenotypic variance important but more so the genotypic uniformity when examining subjects. I highlighted novel phenome-genome links in the context of cranio-cardiac neurocristopathy disorders in the genes *CELSR1, DIP2C* and *ABCA2*. In the future, exploring the underlying molecular mechanisms, signalling pathways and binding partners will enhance our understanding of the pathology behind the phenotypes seen in the patients.

Chapter 7 GENERAL DISCUSSION

The field of human genetics relating to orofacial clefting has evolved in the last decade. Integrated approaches described in this thesis provide promise for future gene discovery using multi-disciplinary approaches. The Clinical Study led to the discovery of novel candidate genes for syndromic orofacial clefting and new phenotype-genotype relationships. Very few gene changes will affect cleft development only and it is more likely that a single gene change will have a wider effect, i.e. on multi- overt and subclinical structures. Indeed, novel gene discovery from whole exome sequencing (WES) in the field of orofacial clefting was instigated from sequencing syndromic clefts, these attempts were followed in non-syndromic cases and have also recently begun to produce revolutionary results. This research, alongside that of many others, will expand our knowledge on the involvement of syndromic genes in hitherto diagnosed isolated cleft patients and vice versa.

7.1 THE UNDERUSE OF PAEDIATRIC DENTISTS IN GENETICS. HOW CAN WE CONTRIBUTE?

Paediatric dentists are well placed to contribute to genetic research since they see children with a range of medical conditions. The discordance between the use of genetics for progress on disease aetiology and employing genetics for clinical management is evident. The use of genetics in practice is currently applied toward ascertaining the probability of a given clinical diagnosis and risk assessment for future offspring. Therefore, communicating with families and other health-care professionals is necessary to optimise the decision-making processes (Guttmacher *et al.*, 2009; Guttmacher *et al.*, 2007). Paediatric dentists have a crucial role in identifying the families who might benefit from a genetics referral. They are some of the first to follow-up paediatric patients and monitor their growth and development; they observe the rapid changes that occur in the craniofacial area in the first 16 years of the child's life and are privy to the wider medical diagnoses since they are trained in providing dental care for medically compromised children and those with developmental disorders. Likewise, family dentists in general have the opportunity to be the first in identifying syndromes where subclinical features or carrier status are manifested by dental phenotypes in family members (Harrison *et al.*, 2018). A report from the Royal Society in 2005 highlighted the lag between genetic education at undergraduate, postgraduate and continuing medical education levels and the scientific and technical advances in the fields of genetics and genomics (Mayor, 2005) but attitudes towards genetics vary, with some marginalizing its applicability and clinical utility (Suther & Goodson, 2003).

In a survey of families conducted by Harvey and co-workers (2007) on 5915 individuals with genetic diseases including Marfan, long QT, celiac, CLP and 20 other conditions, 85% of respondents ranked their own understanding of the genetics of the condition as good or excellent. On the other hand, between 17% and 62% of health care providers had 'good/excellent' knowledge of the underpinning genetic changes. Parents reported, "*It is frustrating when you have to be the one to 'teach' the doctor*" and another said, "*I still wanted them to be understanding and knowledgeable about his condition*". Furthermore, 64% of participants reported not receiving any genetics education material from their primary care provider. This emphasizes our need for a basic understanding of genetics in an era where genetic information is readily available (Harrison *et al.*, 2018). Despite this, there is a considerable shortage in knowledge and utility of genetics in clinical paediatric dental practice.

7.2 THE IMPORTANCE OF INFORMATION SHARING

A recurrent theme in this research was that many of the children included already had genetic investigations carried out but the results were inconclusive. This highlights how sequencing technologies and algorithms, and the availability of public datasets (ExAC, gnomAD and dbSNP) is evolving and can contribute to a genetic diagnosis and that cleft patients who were tested in the past with no result found when an 'older' technique was used should be retested using the 'newer' ones.

The success of this project is due to the collaboration between a broad multi-disciplinary team **(Figure 7-1)**, beginning with phenotypic data assembly, to communicating findings so that gene variants are filtered down, combined with computational functional predictions and laboratory functional assays. These were key steps in discovering novel variants and determining the pathogenicity of candidate genes. The feedback of positive findings to the participants was also an approach we took in our research protocol, whereby all sequenced families were seen by a named clinical geneticist dedicated to this study. Indeed, recruiting for exome sequencing research is feasible and does not deter from a daily clinical workflow and the results could be directed back to affected families. Moreover, it is inclusive whereby patients from all backgrounds and variable clinical traits are seen. This is contrast to GWAS that focuses on large cohorts with similar phenotypes and shared ethnic backgrounds and the inability to directly feedback results to the participants. Additionally, the findings of this research support the inclusion of the identified novel candidate genes in gene panels selected to diagnose syndromic and non-syndromic orofacial clefting and

suggest that the genes' biological partners should also be considered as further candidates in genetic research on orofacial clefting disorders.

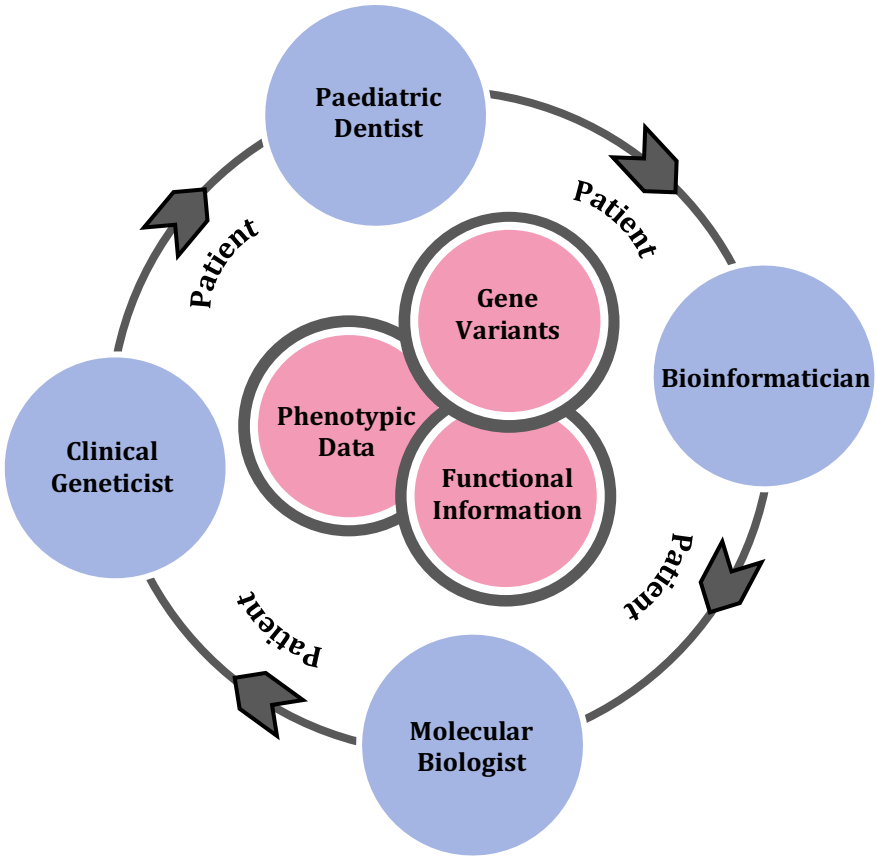


Figure 7-1 Team workflow for gene discovery in the current study

This diagram represents the constant flow in data sharing and crosstalk among all parties involved, with the patient being the crosslinker and centre of focus.

7.3 CHANGING CONCEPTS IN CLEFT GENETIC RESEARCH

Genome-wide association studies in the field of cleft research have yielded replicable associations between certain loci and CLP and it has been an appealing approach to pursue based on the concept that multiple common variants with small effects underlie common disease. This is known as the 'common disease-common variant' hypothesis. With a new emphasis towards associating rare genetic variants in disease, in a concept-change to answer the 'missing heritability' of complex diseases (Manolio *et al.*, 2009), not only have *de novo* mutations been implicated in rare syndromic disorders but also evidence suggests that they are found in more common, yet genetically, heterogeneous diseases such as autism spectrum disorder (ASD) (Shendure & Akey, 2015). However, this concept has been challenged in cleft lip and palate research, for instance, the role of *IRF6* in non-syndromic and syndromic forms of clefting is well known but some researchers have segregated the association of variants in *IRF6*-related disorders and concluded that common variants in *IRF6* are strongly associated with non-syndromic orofacial cases, while rare coding mutations underly syndromic forms such as van der Woude (Leslie *et al.*, 2016b). Leslie *et al.* (2016) have therefore reported that rare coding variants in *IRF6* are unlikely to play a major role in risk for non-syndromic CLP (Leslie *et al.*, 2016b). Finally, while WES studies in non-syndromic clefting were investigating the implication of rare variants, almost all assumed inherited modalities and excluded potential sporadic presentations from their approach.

While reviewing the current role of WES in non-syndromic cleft lip and palate genetics, it became clear that almost all studies were carried out using a 'multiplex family' design (i.e. multiple affected individuals within a family) that does not take sub-clinical phenotypes in 'un-affected' relatives into consideration. A study design such as this assumes Mendelian inheritance and often tests dominant modes of inheritance. Indeed, almost all WES studies carried out on non-syndromic CLP have been based on the probability that rare variants would be shared by the two or three affected relatives (Aylward *et al.*, 2016; Basha *et al.*, 2018; Bureau *et al.*, 2014), a robust approach when multiple family members are available for sequencing. Perhaps, by testing non-syndromic cohorts with *no* parental or family history of clefting, *de novo* or rare gene variants might be found.

Previous researchers have encouraged the segregation of subjects by cleft phenotype (CL/P vs. cleft palate only) into separate cohorts (Grosen *et al.*, 2010) but our results challenge that concept by showing that a variety of clinical appearances of

cleft can spring from the same gene. Thus, research should now look into establishing 'variant-phenotype' relationships rather than 'genotype-phenotype' relationships. In this thesis, segregation was based on the presence or absence of 'associated anomalies', be it dental or medical, and on whether family members were affected with full or subclinical associated anomalies, thereby making the *genetic mode* the focus of our strategy rather the cleft subtype itself. This approach challenges most conventions; however, it is not surprising that we found genes that are involved with non-syndromic CLP also implicated in CPO cases. Other examples of genes include *ARHGAP29* (Leslie *et al.*, 2012; Liu *et al.*, 2017) and *FOXE1* (Ludwig *et al.*, 2014; Moreno *et al.*, 2009).

7.4 INHERITED OR SPORADIC, WHAT IS THE CASE IN CLEFT?

Both inherited and sporadic genetic changes could cause syndromic and non-syndromic orofacial clefting disorders. Sifrim *et al.* (2016) showed that *de novo* mutations play a major role in syndromic congenital heart disease (CHD) and that inherited high-risk variants were more prevalent in non-syndromic CHD. In the context of the current study, this implies that parents with tooth anomalies could signify non-syndromic cases with inherited variants. Other examples that support this are twin and familial studies that showed gene changes in non-syndromic CLP. Nevertheless, a clear Mendelian inheritance pattern is not always observed (Dixon *et al.*, 2011), and so *de novo* changes might also be plausible in non-syndromic CLP. Therefore, probands with 'isolated' clefts who have parents with no dental anomalies might also share the same risk model as those with syndromic-CLP, where cleft events are sporadic and *de novo* risks are high. Yet, *de novo* variation in syndromic-CLP has not often been considered, for instance, Pengelly *et al.* (2015) suggested that syndromic conditions have underlying mutations that follow a Mendelian pattern of inheritance, contradicting the notion stated by Sifrim *et al.* (2016) that most syndromic cases are sporadic and hence mutations occur *de novo.* However, it is more evolutionary probable that sporadic syndromic cleft cases are of *de novo* origin. Indeed, *de novo* germline single nucleotide variants in single genes have been shown through exome sequencing to be the cause of the many syndromic forms of cleft (Veltman & Brunner, 2012). **Figure 7-2** attempts to sum genetic models in cleft research an reflects on how both genetic risk models apply to both types of cleft conditions, syndromic and non-syndromic.

To date, a large-scale exome sequencing study has not been employed for clefts, whether syndromic or non-syndromic. This has already been carried out for other congenital and complex diseases such as autism and congenital heart disease. In cleft

research, the few large-scale sequencing studies that have been carried out have only looked for known genes (Leslie *et al.*, 2016b). Using broader gene panels for exome sequencing to find disease variants, perhaps by developing ones for genes and pathways involved in craniofacial development rather than just cleft alone, and comparing those with others that target candidate genes implicated in other related developmental disorders such as cardiac or autism arrays, would potentially enhance cleft gene discovery.

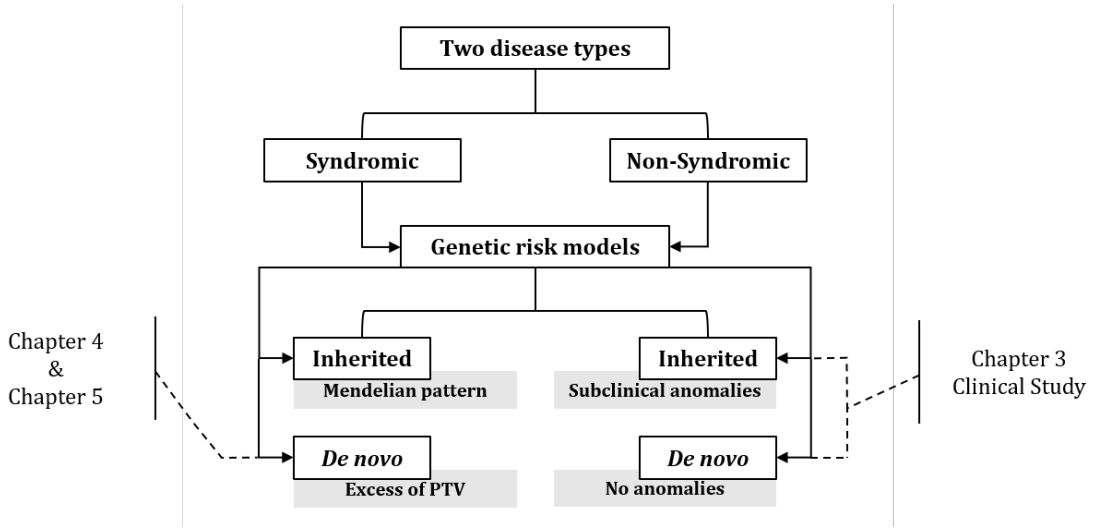


Figure 7-2 Genetic risk models in cleft lip and palate research

Rare gene variants can be found through exploring both Mendelian or *de novo* disease models in syndromic and non-syndromic cleft lip and palate. Detailed clinical phenotyping is crucial to delineate between the subgroups and to draw these genetic models. PTV, protein truncating variants.

7.5 EXPANDING THE PHENOTYPIC SPECTRUM OF PALATE ANOMALIES

Previous findings in mouse models from our research group suggest that there are multiple steps in palate development (Tabler *et al.*, 2013). This thesis shows that patients with *CTNND1* mutations have a range of palatal changes ranging from a high-arched palate to submucous cleft, to complete cleft lip and palate. My study suggests that cleft palate is part of a broader spectrum of anomalous palatal morphologies, and that high-arched palates should be considered to be included within this spectrum. The high-arched palate is underreported and is often missed; for instance, the gene *PHYH* has been associated with rhizomelic chondrodysplasia punctata, which can include craniofacial anomalies such as micrognathia and high-arched palate (Barr *et al.*, 1993; Jansen *et al.*, 1997). More recently, *PHYH* has been described as a candidate gene for non-syndromic CLP following exome sequencing (Aylward *et al.*, 2016). Therefore, a high-arched palate is *not* a mild form of cleft palate that failed to progress further, rather, it is an anomaly within the cleft spectrum and is likely due to a discrete genetic change that is occurring at a different time developmentally to palatal shelf closure. As such, the expression 'pseudo-cleft' that is frequently applied to high-arched palates may be misleading.

7.6 VELOPHARYNGEAL INSUFFICIENCY AS A SUBCLINICAL FINDING

The data I present in this thesis suggests that anomalies of the velopharynx form an integral group of anomalies that can be caused by cleft associated genes. Velopharyngeal dysfunction is a disorder that affects speech production. It is a broad term that can be further divided into velopharyngeal insufficiency (VPI) and velopharyngeal incompetence. VPI is an anatomic defect or structural abnormality, which refers to the inability to achieve complete closure of the velopharyngeal port and commonly occurs in individuals with submucous cleft palate or an overt cleft palate. Velopharyngeal incompetence on the other hand occurs secondary to a congenital or acquired neurological process such as congenital hypotonia. VPI occurs in a number of syndromes, of which 22q11 deletion is the most common particularly in those diagnosed with velocardiofacial syndrome. Indeed, 69% of individuals with a 22q11 deletion have a palatal abnormality in the form of cleft palate (11%), submucous cleft palate (16%), bifid uvula (5%), and cleft lip/cleft lip and palate (2%), with associated VPI in 27% of them (Sweeney *et al.*, 2015). Some children who have their cleft palate

repaired also have velopharyngeal insufficiency, a surgical outcome that is unpredictable (Sell *et al.*, 2015). The first Clinical Standards Advisory Group (CSAG) report on speech concluded that the primary surgery for cleft children in the UK had produced poor speech outcomes (Sell *et al.*, 2001), though this improved following centralization of surgical services (Sell *et al.*, 2015).

The findings in the present study suggest that speech defects in children with clefts should not solely be attributed to poor surgical outcomes. The *Ctnnd1* mutant mice clearly had sub-phenotypic VPI and several patients had laryngeal and pharyngeal anomalies. The anatomic or structural variations seen in VPI result from failure of complete closure of the velopharyngeal port, the region of the velopharynx that comprises of the soft palate anteriorly, posterior pharyngeal wall posteriorly, and the lateral pharyngeal walls laterally. Phonation requires closure of this port initiated by coordinated motor cortical functions and contractions of the velopharyngeal musculature and tightness of this port influences the type of speech produced. The muscles of the soft palate include the tensor veli palatini, palatoglossus, palatopharyngeus, levator veli palatini, and musculus uvulae (Sweeney *et al.*, 2015). The *Ctnnd1* mice did not have clefts or facial dysmorphisms but they did have laryngeal webbing changes affecting the vocal cords as well as defects within their velopharyngeal port. Their palatopharyngeus muscle was severely disorganized with an apparent increase in the cranio-caudal thickness and other muscles within the pharyngeal apparatus were ectopically fused to the levator veli palatini muscles which gave the impression of a high-arched epiglottal area. The abnormal position of the levator veli palatini is critical to the functional deficits observed in cleft palate patients as defects as such are known to cause nasal air emission during speech. In cleft patients, the levator veli palatini becomes discontinuous across the palate and inserts abnormally into the posterior aspect of the hard palate, therefore, contraction of the levator in this position cannot reposition the soft palate against the posterior pharyngeal wall to close the velopharyngeal port. Analogous to the laryngeal webbing found in the mice, pharyngeal webbing has been reported in the past in patients with cleft but is now seldom studied (Walter, 1990).

There can be no doubt that VPI and speech deficits should be within the cleft spectrum. Huston and co-workers (1984) clinically investigated and compared velopharyngeal function in three groups, patients with cleft lip only, unaffected parents/siblings that had positive family history of clefts and unaffected parents with negative family history and found no differences amongst the groups. They also did not

observe VPI in unaffected parents/siblings with positive family history for clefts . This present thesis suggests that more research should be conducted into laryngeal webbing and into patients with residual speech defects, and their families.

The data I present in this thesis, suggests that structural defects affecting the velopharynx such as VPI should be considered within the subclinical manifestations in 'unaffected' parents/siblings and children with 'isolated' clefts. This will have repercussions for classifications since when these types of defects are included within the cleft diagnosis, very few children will have 'true' isolated clefts. Certainly, the presence of velopharyngeal anomalies and speech problems should be considered when designing clinical/genetic studies into clefts in the future.

7.7 WHERE AND HOW COULD WE FIND UNEXPLAINED ANSWERS?

Exome sequencing has proven to be a valuable approach for gene discovery; however, the answer for many unexplained cases might lie in the genome. To date, whole genome sequencing has not been carried out on a large cleft cohort. It is likely that by including non-coding regulatory regions in the analysis of sequencing data that further insights into the genetic architecture of both inherited and sporadic forms of disease will be made. Nowadays, rare *de novo* 'epigenetic' variations are also being discovered in congenital disorders, for instance, *de novo* non-coding epivariations following DNA methylation studies were enriched in cases refractory to exome and genome sequencing in patients with neurodevelopmental conditions associated with congenital malformations (Barbosa *et al.*, 2018).

Another explanation for finding inconclusive results from genetic testing was proposed by Blue *et al.* (2017) whereby some familial cases could in fact be oligogenic or polygenic, in other words, an individual's genetic load for common and rare variation accounts for a substantial proportion of their disease liability. Accurate estimation of genetic load and how rare and common variations interact to confer risks and produce phenotypes are complex areas that are still evolving in cleft research. Therefore, a variant identified by a genetic test may be the full cause of a child's presenting features, a contributory factor or an unrelated finding and so further testing is needed for validation (Wright *et al.*, 2018a). An example of that is reported in this thesis when a benign chromosomal duplication on chromosome 4 was initially identified in the proband that was later revealed to have an *AGAP6* copy number variation.

This study demonstrated the usefulness of the DDD dataset to widen our research sample beyond the initial undiagnosed genetic cohorts of recruits. The CAP180 dataset was designed to specifically include patients with 'neurocristopathies'. Assuming that 1:700 in a general population would have a cleft (0.14%), the CAP180 showed that 118:1547 (8%) of our cohort had 'true clefts' (excluding those with other palate anomalies) and when including all palatal phenotypes, 180:1547 (12%) of them had a 'palatal anomaly'. This suggests that the CAP180 dataset provided a good basis to explore new cleft genes. Neurocristopathies appear to have an overrepresentation of orofacial clefts. A separate CAP that is solely based on clefting/palatal conditions, including or excluding neurocristopathies, could potentially steer towards a targeted approach for discovering cleft candidate genes.

7.8 ENVIRONMENTAL FACTORS

Genetic heterogeneity is to be expected with complex diseases such as cleft lip and palate; therefore, as many have concluded from their exome sequencing studies on non-syndromic CLP, rare variants may only explain part of the "missing heritability" (Chaste *et al.*, 2017; Fu *et al.*, 2017b). The interplay between genetic, epigenetic and environmental factors was not addressed in the current project. Although non-syndromic and syndromic cases of cleft could be explained by a simple genetic change, associations between gene changes and the environment are difficult to replicate (Mossey *et al.*, 2017). The Cleft Collective are distributing questionnaires to families to obtain detailed parental living, pregnancy information and DNA sampling to develop gene banks to address some of the outstanding questions on the effect of environmental co-variates. A large-scale study on gene-environment-joint-effects comprising 1020 families with clefts has shown that the relative risk for mothers carrying a variant in the *MTHFR* gene (involved in folate metabolism) and the lack of folic acid intake was slightly larger when compared to mothers taking vitamin supplements (Mossey *et al.*, 2017).

Mouse studies have also enhanced our understanding on many of the environmental insults and it is known that embryonopathies in mice cause cases of cleft, cardiac and other anomalies (O'Reilly *et al.*, 2014). Birth defects have been shown to be caused by environmental factors such as hypoxia from smoking, altitude, exposure to pollutants or maternal iron deficiency and from teratogens (Moreau *et al.*, 2019; Sparrow *et al.*, 2012). Perhaps it is the general effect of hypoxia stopping translation of proteins? Perhaps it is the interplay between hypoxia and certain receptors that manifest these birth defects? Studies have shown that Fgfr1 is responsive to hypoxia in

mice (Shi *et al.*, 2016) and birth defects caused by maternal iron deficiency can mimic DiGeorge syndrome (Christian & Stewart, 2010; McArdle *et al.*, 2013); perhaps this could also clarify unexplained cases from 22q11 tests or perhaps it is the unexplored novel genes? Diabetic mouse models also provided examples for environmental influences and showed that when glucose levels increased so did the observed penetrance of phenotypes (Hrubec *et al.*, 2006), suggesting that diabetic embryonopathies likely have an effect on neural crest formation, migration or differentiation that might mimic neurocristopathies.

Therefore, genetic causes, especially in the case of rare variants, are useful for family-based research. By contrast, investigating and identifying environmental risk factors has the added effect of possibly reducing risk of birth defects on a public/population level. Hence, a fundamental focus should be on how to prevent or mitigate the occurrence of CLP and other congenital anomalies (Mossey, 2003). This can be achieved by devising a targeted approach to study environmental/peri-natal effects. Studying the metabolome can give insights into genetic or environmental changes. The study of metabonomics in cleft research is sparse and has emerged in recent years. By utilising metabonomic analyses between healthy pregnant mice and those injected with the teratogen dexamethasone, Wu *et al.* (2010) found that the administration of vitamin B12 reverses the CLP induced by dexamethasone by promoting the generation and metabolism of folic acid (Wu *et al.,* 2010). The authors hypothesised that the administration of vitamin B12 allows the accumulation of useful metabolites to supplement the lack of proteins and enzymes caused by the teratogen (Wu *et al.,* 2010). Further, the analysis of nutrient-related maternal biomarkers such as the study of folate-dependent one-carbon metabolism provided evidence for the involvement of this metabolic pathway in the risk of clefts (Munger *et al.,* 2011). To study the interaction between genetic and environmental factors is a complex process and requires extensive collaborative efforts (Blue *et al.*, 2017).

7.9 THE MULTIFACTORIAL THRESHOLD MODEL

Cleft anomalies often are described as disorders arising from *de novo* mutations or Mendelian patterns of inheritance, which are by definition environmentally insensitive (Roff, 1996). However, disease heterogeneity suggests a far more complex origin for genetic liability to clefts, and thus, cleft lip/palate can be considered a multifactorial threshold trait, where multiple events contribute to the expression and extent of the cleft trait (Baxter and Fraser, 1950; Fraser, 1976, Mendell *et al.,* 1980). This is analogous

to other reported multifactorial diseases such as polycystic kidney disease and autism spectrum disorder (McDonald *et al.,* 1990; Chaste *et al.,* 2017). The standard multifactorial threshold model (MFT) assumes susceptibility to a particular trait or disorder by means of three elements: genes affecting expression of a trait, interacting with a set of environmental influences, taking into account the element of chance, enough input from all these elements alters *liability* towards disease (Kurnit *et al.,* 1987; McDonald *et al.,* 1990). Liability determines the probability of an individual succumbing to the disease and is determined by both genes and the environment (Cunrow and Smith 1975). At one end of the distribution is a *threshold* of liability, which must be exceeded in the affected individual for a trait to be fully expressed (McDonald *et al.,* 1990).

The multifactorial threshold model has been used widely in genetic applications and has served to explain orofacial clefts in two important areas: susceptibility towards *disease* and as a model for *inheritance* (Baxter and Fraser 1950; Curnow and Smith, 1975; Fraser 1976; Mendell *et al.,* 1980). The hallmarks for multifactorial inheritance are: (1) most affected children have normal parents; (2) recurrence risk increases with the number of affected children in a family (3) recurrence risk increases with severity of the defect; 4) consanguinity slightly increases the risk for an affected child; (5) the risk decreases most between first and second-degree relatives; and (6) when the two sexes have a different probability of being affected, the least likely sex, if affected, is the most likely sex to produce an affected offspring (Grosen *et al.,* 2010). Most of the historic studies implementing the concept of the multifactorial threshold model have shaped our understanding of the aetiology of orofacial clefts, despite the lack of data on genetic input (Baxter and Fraser 1950; Fraser 1976). Indeed, as Curnow and Smith discussed in 1975, "*In many respects the multifactorial model is a simplistic and "lumping" approach and nature is likely to be much more complex and heterogeneous. Thus, the role of the multifactorial model in familial disease may be as a temporary tool useful during a period of ignorance for estimating risks and for providing indicators about the relations between different diseases and the relation of diseases with measurable continuous characters.*" (Curnow and Smith, 1975).

Therefore, perhaps the MFT model is better suited for interpreting metric data for population risk estimates in diseases such as obesity, cancer and metabolic disorders. As for orofacial clefts and other dysmorphisms, Hallgrímsson *et al.* (2005) noted that although discrete (or ordinal) traits are conventionally modelled as having normally distributed liability, this is not always true, since these traits are due to major gene effects (Hallgrímsson *et al.,* 2005). Moreover, the MFT model has not been

considered for data such as laterality of the cleft defect and manifestation of cleft subtype in offspring.

Thus, my current study may be a better fit for other statistical models. The key findings of my research suggest a much greater weight for genetic effects in cleft aetiology than previously observed. I propose that in the majority of cases, a causative genetic variant (or variants) is present. This changes the threshold of liability causing the phenotype. Furthermore, the additive effects of environmental factors may modify the liability towards disease by either increasing the likelihood of the occurrence of disease in the genetically susceptible individual, increasing the severity of the outcome, or altering disease expressivity.

7.10 MEDICAL DISCOVERY & GENERALISABILITY

Exome studies conducted on cohorts of cardiac patients have already incorporated extra-cardiac phenotypes to their analyses, adopting an approach where generalizability of the pathogenic variants can be extended to the wider disease. This is a step ahead of gene discovery in cleft research. Indeed, two large-scale studies have identified a significant excess of damaging *de novo* variants in cardiac patients with neurodevelopmental disorders and extra-cardiac congenital defects including facial anomalies (Homsy *et al.*, 2015; Sifrim *et al.*, 2016). This suggests that gene changes in congenital heart disease can be considered in other developmental disorders in general (Blue *et al.*, 2017).

As exemplified from the genes identified in Chapter 6, *ABCA2, CELSR1* and *DIP2C* and from the *CTNND1* gene explored in Chapter 4, the concept of pleiotropy is becoming increasingly recognised the more phenotypic traits and new syndromes are being discovered. Pleiotropy is when a single gene causes multiple phenotypic expressions or disorders (Wright *et al.*, 2018a). Variants in *DIP2C* for instance, have been linked to skeletal dysplasia and *ABCA2* has been linked to global developmental delay (Maddirevula *et al.*, 2018; Maddirevula *et al.*, 2019). In this study we suggest that other variants in *DIP2C* and *ABCA2* may lead to broader syndromic craniofacial-cardiac disorders though further research will be needed to confirm this. Similarly, germline mutations in E-cadherin (*CDH1*) either cause isolated cleft lip and palate, blepharocheilodontic syndrome (BCD) or hereditary diffuse gastric cancer (HDGC) (Benusiglio, 2017). It is not yet known, in the *CDH1* scenario, or indeed for other diseases, whether a further somatic change through loss of heterozygosity alters the

tissue affected by cancer when a patient has a heterozygous variant at germline level (Wu *et al.*, 2015b). Thus, knowing the spectrum of disease phenotypes linked to individual gene changes is of paramount importance to inform clinical investigations, counselling, family planning and medical management.

Whether the phenotypes described in the present study, even when previously reported ones are included, are an expansion of the phenotypic spectrum or examples of pleiotropy is a question for debate. Nevertheless, clinical documentation of phenotype-genotype relationships and inclusion of newly published discoveries in public databases is important to support the management of affected patients (Wright *et al.*, 2018a). Ultimately, the holistic approach to patient care is key.

7.11 THE TRANSLATIONAL IMPACTS OF GENE DISCOVERY

It has been estimated that it takes an average of 17 years for only 14% of new scientific discoveries to enter day-to-day clinical practice (Westfall *et al.,* 2007). Yet, the application of preventative and therapeutic measures for human diseases has greatly advanced in the past few years (Ginn *et al.,* 2018). Gene therapy clinical trials have particularly been successful in targeting cancer, infectious diseases, neurological and blood disorders. Despite this progress, many disease phenotypes, including congenital craniofacial birth conditions are still at bench-side experimental phases and lie beyond the reach of existing translational technology and its application into standard clinical care (Ginn *et al.,* 2018). Nevertheless, many of the gene therapy trials have targeted monogenic diseases, which were the most addressed disease entities (11.1%) following cancer (65%). Surprisingly, gene therapies for monogenic diseases, such as cystic fibrosis, are the ones achieving the greatest success in terms of outcome (Ginn *et al.,* 2012), providing proof-of-concept of their applicability to other single gene disorders such as those causing orofacial clefts.

Nonetheless, numerous efforts have been carried out in model organisms in an attempt to translate those clinically. Establishing therapy is fundamentally dependant on preclinical research and in the case of craniofacial anomalies, a thorough understanding of the cellular and molecular mechanisms that regulate neural crest cell events as well as the genetic aetiology of neurocristopathies and other craniofacial disorders. For example, prevention of craniofacial anomalies characteristic of Treacher Collins syndrome (TCS), a neurocristopathy that includes cleft palate as one of its major features, has been undertaken (Jones *et al.,* 2008; Sakai *et al.,* 2016).

The pathologic events contributing to the aetiology of TCS have been revealed in *Tcof1*$^{+/-}$ mouse models in the form of perturbations in the generation and proliferation of neural crest cells, as well as oxidative stress-induced apoptosis of neuroepithelial cells caused by deficient ribosome biogenesis (Dixon *et al.,* 2006; Sakai *et al.,* 2016; Trainor *et al.,* 2009). Attempts to reverse these events have been carried out first by the genetic and pharmacological inhibition of p53, thereby suppressing neuroepithelial apoptosis in *Tcof1*$^{+/-}$ embryos and preventing the pathogenesis of TCS (Jones *et al.,* 2008). This research group further evolved their therapeutic strategies to overcome adverse side effects of tumorigenesis from p53 inhibition by alternatively administering an antioxidant, N-acetyl-cystein (NAC), during the critical period of neural crest cell formation (Sakai *et al.,* 2016). By doing so, Sakai *et al.* (2016) confirmed cell death suppression in the neuroepithelium, suggesting that antioxidant treatment/supplementation can prevent the pathogenesis of severe craniofacial abnormalities in *Tcof1*$^{+/-}$ embryos (Sakai *et al.,* 2016). Certainly, prenatal supplementation has shown success in the prevention of birth defects as indicated by the WHO, *"the success of folic acid interventions in preventing neural tube defects provides a benchmark against which other preventive strategies for birth defects can be measured,"* (WHO, 2002).

Another model that was utilised to investigate therapeutic potential for craniofacial defects was the *Pax9*$^{-/-}$ mouse. These mice consistently exhibit clefts of the secondary palate and die shortly after birth (Jia *et al.,* 2017). Based on RNAseq data from E13.5 *Pax9* deficient palatal shelves, Jia and co-workers (2017) found significantly higher expression of the Wnt signalling pathway inhibitors dickkopf 1 (*Dkk1*) and *Dkk2*. Interestingly, they used a small molecule Wnt agonist that targets *Dkk1* by inhibiting its effect (Jia *et al.,* 2017). The molecule was administered into the tail veins of pregnant *Pax9*$^{-/+}$ mice throughout the critical periods for embryonic palatogenesis (E10.5-E14.5). This pharmacologic inhibition rescued the cleft palate phenotype in the *Pax9*$^{-/-}$ pups, thereby restoring the Wnt pathway (Jia *et al.,* 2017). In another illustration, a mesenchymally expressed *Bmp4* human transgene in *Msx1*$^{-/-}$ mice was capable of rescuing the cleft palate phenotype and neonatal lethality; 41/66 mice exhibited complete palate closure and histologic analysis showed that the rescued palates were able to fuse, and the midline seam disappeared. (Zhang *et al.,* 2002).

The examples above are indications of how 'precision' medicine can work in practice. As shown in the *Pax9*$^{-/-}$ and *Tcof1*$^{+/-}$ rescue models (Jia *et al.,* 2017), perhaps the area where translational medicine is most likely to succeed is pharmacogenomics,

which is the study of how drugs interact with a patient's underlying genetic makeup to influence biological pathways and processes (Khoury *et al.,* 2007). Precision medicine will also impact health and decisions across one's lifespan (Ginsburg and Phillips, 2018); genetic screening for reproductive counselling and prenatal testing can be used prior to conception to accurately predict the risk of passing on genetic conditions to offspring (Ginsburg and Phillips, 2018). Yet, the translation of genomics and precision medicine into useful and cost-effective clinical care will require years of translational research and has certainly lagged behind the pace of basic science discoveries (Ginsburg and Phillips, 2018; Khoury *et al.,* 2007).

7.12 VALUE OF TRANSLATIONAL RESEARCH AND DIALOGUE: HOW TO EXPAND AND BUILD THE CHANGES/BARRIERS

Involving dentists in genomic projects will enhance literacy in cleft genetics. Dentists that are already involved in genetic research mostly focus on tooth anomalies and cleft lip/palate. In the future, genetic diagnosis and planning will become an integral part of daily practice and so the role of genetics, in not only clefts but craniofacial and medical conditions will have a relevant clinical focus, and not just as a scientific work in progress. However, the implementation of genomic medicine has major challenges especially when every aspect of the analytical pipeline is still evolving, from clinical judgements to bioinformatic algorithms to the expansion of populations databases representing diverse ethnicities (Wright *et al.*, 2018a). While utilizing the DDD dataset was a positive means of exploring phenotypes and novel candidate genes for clefts, phenotypic data on patient dentition was lacking. The present study suggests that dental anomalies should be included in the spectrum of 'associated anomalies' in the wider picture of diseases and that dentists should be an integral part of the team providing phenotypic data on the dentition for participants recruited to genomic programmes.

Cleft Care UK, an initiative to monitor outcomes for children with CLP, reported that while some cleft services have improved since the centralization of cleft centres in the UK in 1998, other outcomes have not improved or have become worse such as speech and language or dental decay (Ness *et al.*, 2015; Smallridge *et al.*, 2017). It is clear from these reports that not all outcomes have been monitored. Perhaps as part of the future recommendations for Cleft Care UK, genetic findings of these children should also become an integral part of the care pathway. Detailed clinical information and long-term follow-up will strengthen the management as well as potentially uncover associations

between subclinical phenotypes, genetic variants and environmental risks. Indeed, the results of the current study have been periodically disseminated to the paediatric dental, dysmorphology and cleft teams at local meetings and national and international conferences, with a particular focus on the usefulness, feasibility and success of integrating genetic testing into practice exemplified by the results of my study and the need to join forces to reflect such efforts into mainstream practice.

Population registries and global surveillance of cleft defects have set excellent examples for how worldwide collaborative efforts could be established. Major networks such as the ICBDMS (the International Clearinghouse for Birth Defect Monitoring System) and EUROCAT (the Surveillance of Congenital Anomalies in Europe) have provided us with in-depth data on the prevalence of orofacial cleft anomalies worldwide (Mossey and Castilla, 2003). Now is the time for global collaborative networks to establish a joint effort for bioinformatic data. Indeed, the WHO report advocates the reduction in the duplication of efforts (Mossey, 2003).

In a bid to improve the outcomes of care for patients with craniofacial conditions who face challenges in obtaining correct diagnoses, the European Reference Network on craniofacial anomalies and ENT disorders (ERN CRANIO) has recently been initiated. The ERN CRANIO is a virtual network that brings together experts on craniofacial malformations including orofacial clefts. It aims to improve patient access to diagnostics and reduce the number of undiagnosed or misdiagnosed patients by also using virtual multidisciplinary clinics (Héon-Klin, 2017). The joint network also aims to support the detection of new causative genes by paving the way for clinical research and increasing the numbers of participants (Héon-Klin, 2017). In the UK, syndromic and non-syndromic clefts conditions have been included in the list of rare diseases recruited to the 100K Genomes Project (Genomics England, 2019). Having utilised the Deciphering Developmental Disorders, UK, dataset to facilitate in establishing diagnoses for a number of my patients who have had previously unresolved clinical and molecular findings, I aspire to aid in driving the gene discovery agenda to support the value of CLP in the European Reference Networks.

7.13 THE VALUE AND SIGNIFICANCE OF THE CRURRENT PROJECT

In this thesis I concentrated my efforts on elucidating the aetiology behind craniofacial and associated anomalies in cleft children with complex traits, as they are likely to have underlying novel *de novo* gene mutations. I used an integrated approach for the

identification of cleft candidate genes, from recruiting family trios to identifying monozygotic twins and unrelated families, to then using genomic datasets and animal models, all to confirm the pathogenicity of the identified mutations. In this thesis I developed a checklist library of overt and microform phenotypes to interrogate when reviewing cleft patients and their families (Dental, Table 2-21; Medical, Table 6-1; Spectrum of Palatal Anomalies, Table 6-3). Using the combined approach of deep phenotyping, next generation sequencing and e-dataset mining will inform future precision medicine and genetic counselling beyond the common empiric risk figures that carry no implications concerning aetiology. I demonstrated the power of family trios in uncovering phenotype-genotype relations. I have shown that my deep phenotyping of parent and proband trios has allowed the specification/design of genetic models which have laid the groundwork for the data mining process for our exome sequencing studies. My research emphasised the value of deep phenotyping in unifying the patients, identifying new variants and expanding syndromes. This deep phenotyping enabled the identification of a previously unrecognised group of multiple congenital anomalies (MCA) associated with CLP or expanded on previously delineated ones. Decreasing the proportion of unidentified MCA entities is an important task that has been recommended by the WHO on craniofacial anomalies (Mossey and Castilla, 2003), therefore the data I present in this thesis adds insight into the field of rare craniofacial anomalies. For example, by mining patient-specific genomic data using national repositories, such as the DDD, and through the analysis of the candidate genes found, I was able to confirm the aetiology of the previously undefined MCA entity associated with an underlying *CTNND1* genotype. I was also able to describe a group of craniofacial and other congenital anomalies associated with *AGAP6, CELSR1, ABCA2 and DIP2C,* that require further functional validations. The recent advances in bioinformatic datasets such as those produced by the DDD, ERN or Genomics England, provide unique and special opportunities to further genetic studies in powerful ways. For example, in addition to the cases I have identified through my clinical recruitment, my research also uncovered answers for cases that have remained unresolved in the DDD/DECIPHER datasets and my findings on the *CTNND1* study have now been published (Alharatani *et al.* Human Molecular Genetics, 2020).

Expanding the *CTNND1* related disorders gave insight into possible genotype (variant) -phenotype relations. For instance, the protein truncating variants in the C-terminal region of *CTNND1* were the ones that caused a complete cleft of the lip and palate. Discerning these correlations improves the diagnostic and prognostic estimates

of disease prenatally, if the parents were carriers for the genotypes. It is also fundamental to understand the molecular pathways and mechanisms the identified candidate genes are involved in if novel therapies are to be considered in the future. Mesenchymal roles for p120-catenin have not been previously shown. My results from patient and model organisms shed light on the molecular control of *CTNND1* in craniofacial development reflected by the facial, dental and oral phenotypes seen and by the laryngeal and velopharyngeal dysfunctions. My experimental study design for deep phenotyping was reinforced when examining mouse models for *Ctnnd1.* This facilitated the discovery of the key findings of microform laryngeal and pharyngeal anomalies in the *Ctnnd1* heterozygous mice. My finding on velopharyngeal anomalies from patient and mouse data suggests that common cleft-related mild abnormalities such as speech delay, hearing difficulties and mild dysmorphologies may in fact be manifestations of a shared underlying genetic aetiology and that cases like these should no longer be classified as isolated clefts, since they may constitute a recognised MCA cleft association (Mossey and Castilla, 2003, WHO). Therefore, my clinical study mitigated the ambiguity behind non-syndromic clefts (WHO, 2002); following deep phenotyping of children and parents the definition of true isolated clefts should be much tighter than previously thought. Moreover, my clinical results on cleft subtypes in relation to the underlying *CTNND1* genotypes are in agreement with the report from the WHO on genotype-phenotype correlations whereby it states that *"there is some emerging evidence that traditional separations between cleft lip, with or without cleft palate, and cleft palate only, may be breaking down, and further work in this area is essential*" (WHO, 2002). Finally, my research advocates the need for a global gene and phenotype platform for CLP because an effort of this kind will greatly accelerate the best outcomes for patient diagnosis and care, and deciphering the molecular pathways of the established genes will advance preventative and therapeutic trials in the field of congenital craniofacial anomalies.

BIBLIOGRAPHY

ABABNEH, F. K., AL-SWAID, A., ELHAG, A., YOUSSEF, T. & ALSAIF, S. (2014). Blepharo-cheilo-dontic (BCD) syndrome: Expanding the phenotype, case report and review of literature. *American Journal of Medical Genetics Part A*, 164(6), 1525-1529.

ABBOTT, B. & BIRNBAUM, L. (1990). Retinoic acid-induced alterations in the expression of growth factors in embryonic mouse palatal shelves. *Teratology*, 42(6), 597-610.

ABBOTT, B., DILIBERTO, J. & BIRNBAUM, L. (1989). 2, 3, 7, 8-Tetrachlorodibenzo-p-dioxin alters embryonic palatal medial epithelial cell differentiation in vitro. *Toxicology and applied pharmacology*, 100(1), 119-131.

ABBOTT, B., HELD, G., WOOD, C., BUCKALEW, A., BROWN, J. & SCHMID, J. (1999). AhR, ARNT, and CYP1A1 mRNA quantitation in cultured human embryonic palates exposed to TCDD and comparison with mouse palate in vivo and in culture. *Toxicological sciences: an official journal of the Society of Toxicology*, 47(1), 62-75.

ABE, R., ENDO, T. & SHIMOOKA, S. (2010). Maxillary first molar agenesis and other dental anomalies. *The Angle Orthodontist*, 80(6), 1002-1009.

ABELEIRA, M. T., OUTUMURO, M., DINIZ, M., LIMERES, J., RAMOS, I. & DIZ, P. (2015). Morphometry of the hard palate in Down's syndrome through CBCT-image analysis. *Orthod Craniofac Res*, 18(4), 212-20.

ABID, M. F., SIMPSON, M. A., PETRIDIS, C., COBOURNE, M. T., & SHARPE, P. T. (2017). Non-syndromic severe hypodontia caused by a novel frameshift insertion mutation in the homeobox of the MSX1 gene. *Archives of oral biology*, *75*, 8-13.

ABLES, J. L., BREUNIG, J. J., EISCH, A. J. & RAKIC, P. (2011). Not(ch) just development: Notch signalling in the adult brain. *Nat Rev Neurosci*, 12(5), 269-83.

ABU-ISSA, R., SMYTH, G., SMOAK, I., YAMAMURA, K. I., & MEYERS, E. N. (2002). Fgf8 is required for pharyngeal arch and cardiovascular development in the mouse. *Development*, 129(19), 4613-4625.

ADAM, M. P., BANKA, S., BJORNSSON, H. T., BODAMER, O., CHUDLEY, A. E., HARRIS, J., et al. MERLA, G. (2019). Kabuki syndrome: international consensus diagnostic criteria. *Journal of medical genetics*, 56(2), 89-95.

ADZHUBEI, I., JORDAN, D. M. & SUNYAEV, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, 76(1), 7.20. 1-7.20. 41.

AHO, S., LEVÄNSUO, L., MONTONEN, O., KARI, C., RODECK, U. & UITTO, J. (2002). Specific sequences in p120ctn determine subcellular distribution of its multiple isoforms involved in cellular adhesion of normal and malignant epithelial cells. *Journal of cell science*, 115(7), 1391-1402.

AINE, L., BACKSTRÖM, M., MÄKI, R., KUUSELA, A. L., KOIVISTO, A. M., IKONEN, R. S. & MÄKI, M. (2000). Enamel defects in primary and permanent teeth of children born prematurely. *Journal of oral pathology & medicine*, 29(8), 403-409.

ALALUUSUA, S., LUKINMAA, P.-L., VARTIAINEN, T., PARTANEN, M., TORPPA, J. & TUOMISTO, J. (1996). Polychlorinated dibenzo-p-dioxins and dibenzofurans via mother's milk may cause developmental defects in the child's teeth. *Environmental Toxicology and Pharmacology*, 1(3), 193-197.

ALAPPAT, S. R., ZHANG, Z., SUZUKI, K., ZHANG, X., LIU, H., JIANG, R., et al. CHEN, Y. (2005). The cellular and molecular etiology of the cleft secondary palate in Fgf10 mutant mice. *Developmental biology*, 277(1), 102-113.

ALKAN, C., COE, B. P. & EICHLER, E. E. (2011). Genome structural variation discovery and genotyping. *Nat Rev Genet*, 12(5), 363-76.

ALLACHE, R., DE MARCO, P., MERELLO, E., CAPRA, V. & KIBAR, Z. (2012). Role of the planar cell polarity gene CELSR1 in neural tube defects and caudal agenesis. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 94(3), 176-181.

ALPÖZ, A. R., ÇOKER, M., ÇELEN, E., ERSIN, N. K., GÖKÇEN, D., VAN DIGGELENC, O. P. & HUIJMANSC, J. G. (2006). The oral manifestations of Maroteaux-Lamy syndrome (mucopolysaccharidosis VI): a case report. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, 101(5), 632-637.

ALTMULLER, J., PALMER, L. J., FISCHER, G., SCHERB, H. & WJST, M. (2001). Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet*, 69(5), 936-50.

ANASTASIADIS, P. Z., MOON, S. Y., THORESON, M. A., MARINER, D. J., CRAWFORD, H. C., ZHENG, Y. & REYNOLDS, A. B. (2000). Inhibition of RhoA by p120 catenin. *Nature cell biology*, 2(9), 637.

ANDERSON, P. J. & MOSS, A. L. (1996). Dental findings in parents of children with cleft lip and palate. *The Cleft palate-craniofacial journal*, 33(5), 436-439.

ARDINGER, H. H., BUETOW, K. H., BELL, G. I., BARDACH, J., VANDEMARK, D. R. & MURRAY, J. C. (1989). Association of genetic variation of the transforming growth factor-alpha gene with cleft lip and palate. *Am J Hum Genet*, 45(3), 348-53.

ARGENTA, L. C., & IACOBUCCI, J. J. (1989). Treacher Collins syndrome: present concepts of the disorder and their surgical correction. *World journal of surgery*, *13*(4), 401-409.

ASLAM, F. & NAZ, S. (2019). Ataxia and dysarthria due to an ABCA2 variant: Extension of the phenotypic spectrum. *Parkinsonism & related disorders*.

ASPINALL, A., RAJ, S., JUGESSUR, A., MARAZITA, M., SAVARIRAYAN, R. & KILPATRICK, N. (2014). Expanding the cleft phenotype: the dental characteristics of unaffected parents of Australian children with non-syndromic cleft lip and palate. *International Journal of Paediatric Dentistry*, 24(4), 286-292.

AVILA, J. R., JEZEWSKI, P. A., VIEIRA, A. R., ORIOLI, I. M., CASTILLA, E. E., CHRISTENSEN, K., et al. MURRAY, J. C. (2006). PVRL1 variants contribute to non-syndromic cleft lip and palate in multiple populations. *Am J Med Genet A*, 140(23), 2562-70.

AYLSWORTH, A. S., ALLORI, A. C., PIMENTA, L. A., MARCUS, J. R., HARMSEN, K. G., WATKINS, S. E., et al. MEYER, R. E. (2015). Issues involved in the phenotypic classification of orofacial clefts ascertained through a state birth defects registry for the north carolina cleft outcomes study. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 103(11), 899-903.

AYLWARD, A., CAI, Y., LEE, A., BLUE, E., RABINOWITZ, D. & HADDAD, J., Jr. (2016). Using Whole Exome Sequencing to Identify Candidate Genes With Rare Variants In Nonsyndromic Cleft Lip and Palate. *Genet Epidemiol*, 40(5), 432-41.

BAMSHAD, M. J., NG, S. B., BIGHAM, A. W., TABOR, H. K., EMOND, M. J., NICKERSON, D. A. & SHENDURE, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*, 12(11), 745-755.

BANKA, S., VEERAMACHANENI, R., REARDON, W., HOWARD, E., BUNSTONE, S., RAGGE, N., et al. DONNAI, D. (2012). How genetically heterogeneous is Kabuki syndrome?: MLL2 testing in 116 patients, review and analyses of mutation and phenotypic spectrum. *Eur J Hum Genet*, 20(4), 381-8.

BARBOSA, M., JOSHI, R. S., GARG, P., MARTIN-TRUJILLO, A., PATEL, N., JADHAV, B., et al. SHARP, A. J. (2018). Identification of rare de novo epigenetic variations in congenital disorders. *Nat Commun*, 9(1), 2064.

BARR, D. G., KIRK, J. M., AL HOWASI, M., WANDERS, R. J. & SCHUTGENS, R. B. (1993). Rhizomelic chondrodysplasia punctata with isolated DHAP-AT deficiency. *Archives of Disease in Childhood*, 68(3), 415-417.

BARTLETT, J. D., DOBECK, J. M., TYE, C. E., PEREZ-MORENO, M., STOKES, N., REYNOLDS, A. B., et al. SKOBE, Z. (2010). Targeted p120-catenin ablation disrupts dental enamel development. *PLoS One*, 5(9), e12703.

BASHA, M., DEMEER, B., REVENCU, N., HELAERS, R., THEYS, S., BOU SABA, S., et al. VIKKULA, M. (2018). Whole exome sequencing identifies mutations in 10% of patients with familial non-syndromic cleft lip and/or palate in genes mutated in well-known syndromes. *J Med Genet*, 55(7), 449-458.

BAXTER, H., & FRASER, F. C. (1950). The production of congenital defects in the offspring of female mice treated with cortisone. A preliminary report. *McGill medical journal*, *19*(4), 245.

BEALES, P. L., ELCIOGLU, N., WOOLF, A. S., PARKER, D. & FLINTER, F. A. (1999). New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. *J Med Genet*, 36(6), 437-46.

BEATY, T. H., MURRAY, J. C., MARAZITA, M. L., MUNGER, R. G., RUCZINSKI, I., HETMANSKI, J. B., et al. SCOTT, A. F. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nature Genetics*, 42, 525.

BEI, M., & MAAS, R. (1998). FGFs and BMP4 induce both Msx1-independent and Msx1-dependent signaling pathways in early tooth development. *Development*, *125*(21), 4325-4333.

BELCARO, C., DIPRESA, S., MORINI, G., PECILE, V., SKABAR, A. & FABRETTO, A. (2015). CTNND2 deletion and intellectual disability. *Gene*, 565(1), 146-149.

BENIRSCHKE, K., LOWRY, R. B., OPITZ, J. M., SCHWARZACHER, H. G. & SPRANGER, J. W. (1979). Developmental terms -- some proposals: first report of an international working group. *Am J Med Genet*, 3(3), 297-302.

BENKO, S., FANTES, J. A., AMIEL, J., KLEINJAN, D.-J., THOMAS, S., RAMSAY, J., et al. LYONNET, S. (2009). Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nature Genetics*, 41, 359.

BENUSIGLIO, P. R. (2017). CDH1 germline mutations: different syndromes, same management? *Genetics in Medicine*, 19(9), 965.

BERNIER, F. P., CALUSERIU, O., NG, S., SCHWARTZENTRUBER, J., BUCKINGHAM, K. J., INNES, A. M., et al. PARBOOSINGH, J. S. (2012). Haploinsufficiency of SF3B4, a component of the pre-mRNA spliceosomal complex, causes Nager syndrome. *Am J Hum Genet*, 90(5), 925-33.

BHATTACHARYA, D., MARFO, C. A., LI, D., LANE, M. & KHOKHA, M. K. (2015). CRISPR/Cas9: an inexpensive, efficient loss of function tool to screen human disease genes in Xenopus. *Developmental biology*, 408(2), 196-204.

BIESECKER, L. G., SHIANNA, K. V. & MULLIKIN, J. C. (2011). Exome sequencing: the expert view. *Genome Biol*, 12(9), 128.

BIRNBAUM, S., LUDWIG, K. U., REUTTER, H., HERMS, S., STEFFENS, M., RUBINI, M., et al. ALBLAS, M. A. (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nature Genetics*, 41(4), 473-477.

BJORK, B. C., KRAHN, K. N., SCHUTTE, B. C. & MURRAY, J. C. (2003). IRF6-related disorders. *Gene reviews. Seattle (WA): University of Washington*.

BLUE, G. M., KIRK, E. P., GIANNOULATOU, E., DUNWOODIE, S. L., HO, J. W., HILTON, D. C., et al. WINLAW, D. S. (2014). Targeted next-generation sequencing identifies pathogenic variants in familial congenital heart disease. *J Am Coll Cardiol*, 64(23), 2498-506.

BLUE, G. M., KIRK, E. P., GIANNOULATOU, E., SHOLLER, G. F., DUNWOODIE, S. L., HARVEY, R. P. & WINLAW, D. S. (2017). Advances in the genetics of congenital heart disease: a clinician's guide. *Journal of the American College of Cardiology*, 69(7), 859-870.

BÖHMER, A. C., GÖLZ, L., KREUSCH, T., KRAMER, F.-J., PÖTZSCH, B., NÖTHEN, M. M., et al. LUDWIG, K. U. (2018). Investigation of dominant and recessive inheritance models in genome-wide association studies data of nonsyndromic cleft lip with or without cleft palate. *Birth Defects Research*, 110(4), 336-341.

BOLANDE, R. P. (1974). The neurocristopathies: A unifying concept of disease arising in neural crest maldevelopment. *Human Pathology*, 5(4), 409-429.

BOLANDE, R. P. (1997). Neurocristopathy: its growth and development in 20 years. *Pediatr Pathol Lab Med*, 17(1), 1-25.

BOURMOUM, M., CHARLES, R. & CLAING, A. (2018). ARF6 protects sister chromatid cohesion to ensure the formation of stable kinetochore-microtubule attachments. *J Cell Sci*, 131(11).

BRAGIN, E., CHATZIMICHALI, E. A., WRIGHT, C. F., HURLES, M. E., FIRTH, H. V., BEVAN, A. P. & SWAMINATHAN, G. J. (2013). DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Research*, 42(D1), D993-D1000.

BRECKPOT, J., THIENPONT, B., GEWILLIG, M., ALLEGAERT, K., VERMEESCH, J. R. & DEVRIENDT, K. (2012). Differences in Copy Number Variation between Discordant Monozygotic Twins as a Model for Exploring Chromosomal Mosaicism in Congenital Heart Defects. *Mol Syndromol*, 2(2), 81-87.

BREUGEM, C. C., EVANS, K. N., POETS, C. F., SURI, S., PICARD, A., FILIP, C., et al. HONG, P. (2016). Best Practices for the Diagnosis and Evaluation of Infants With Robin Sequence: A Clinical Consensus Report. *JAMA Pediatrics*, 170(9), 894-902.

BREWER, S., FENG, W., HUANG, J., SULLIVAN, S., & WILLIAMS, T. (2004). Wnt1-Cre-mediated deletion of AP-2α causes multiple neural crest-related defects. *Developmental biology*, *267*(1), 135-152.

BRUDER, C. E., PIOTROWSKI, A., GIJSBERS, A. A., ANDERSSON, R., ERICKSON, S., DIAZ DE STAHL, T., et al. DUMANSKI, J. P. (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet*, 82(3), 763-71.

BUCAN, M., ABRAHAMS, B. S., WANG, K., GLESSNER, J. T., HERMAN, E. I., SONNENBLICK, L. I., et al. BRADFIELD, J. P. (2009). Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. *PLoS genetics*, 5(6), e1000536.

BUCKINGHAM, M., MEILHAC, S. & ZAFFRAN, S. (2005). Building the mammalian heart from two sources of myocardial cells. *Nature Reviews Genetics*, 6(11), 826.

BUREAU, A., PARKER, M. M., RUCZINSKI, I., TAUB, M. A., MARAZITA, M. L., MURRAY, J. C., et al. BEATY, T. H. (2014). Whole exome sequencing of distant relatives in multiplex families implicates rare variants in candidate genes for oral clefts. *Genetics*, 197(3), 1039-44.

BUSH, J. O., & JIANG, R. (2012). Palatogenesis: morphogenetic and molecular mechanisms of secondary palate development. *Development,* 139(2), 231-243.

BUTTS, S. C. (2009). The facial phenotype of the velo-cardio-facial syndrome. *International journal of pediatric otorhinolaryngology*, 73(3), 343-350.

CAI, X., GONG, P., HUANG, Y. & LIN, Y. (2011). Notch signalling pathway in tooth development and adult dental cells. *Cell Prolif*, 44(6), 495-507.

CAI, Y., PATTERSON, K. E., REINIER, F., KEESECKER, S. E., BLUE, E., BAMSHAD, M. & HADDAD, J., Jr. (2017). Copy Number Changes Identified Using Whole Exome Sequencing in Nonsyndromic Cleft Lip and Palate in a Honduran Population. *Birth Defects Res*, 109(16), 1257-1267.

CALLAHAN, N., MODESTO, A., MEIRA, R., SEYMEN, F., PATIR, A. & VIEIRA, A. (2009). Axis inhibition protein 2 (AXIN2) polymorphisms and tooth agenesis. *Archives of oral biology*, 54(1), 45-49.

CALZOLARI, E., PIERINI, A., ASTOLFI, G., BIANCHI, F., NEVILLE, A. J. & RIVIERI, F. (2007). Associated anomalies in multi-malformed infants with cleft lip and palate: An epidemiologic study of nearly 6 million births in 23 EUROCAT registries. *Am J Med Genet A*, 143a(6), 528-37.

CAMPOS, D. & MONAGA, M. (2012). Mucopolysaccharidosis type I: current knowledge on its pathophysiological mechanisms. *Metabolic brain disease*, 27(2), 121-129.

CARETTE, M. J., & FERGUSON, M. W. (1992). The fate of medial edge epithelial cells during palatal fusion in vitro: an analysis by DiI labelling and confocal microscopy. *Development*, 114(2), 379-388.

CARROLL, K. & MOSSEY, P. A. (2012). Anatomical Variations in Clefts of the Lip with or without Cleft Palate. *Plast Surg Int*, 2012, 542078.

CASTANET, M., PARK, S. M., SMITH, A., BOST, M., LEGER, J., LYONNET, S., et al. POLAK, M. (2002). A novel loss-of-function mutation in TTF-2 is associated with congenital hypothyroidism, thyroid agenesis and cleft palate. *Hum Mol Genet*, 11(17), 2051-9.

CASTILLO-FERNANDEZ, J. E., SPECTOR, T. D. & BELL, J. T. (2014). Epigenetics of discordant monozygotic twins: implications for disease. *Genome medicine*, 6(7), 60.

CHAI, Y., & MAXSON JR, R. E. (2006). Recent advances in craniofacial morphogenesis. *Developmental dynamics: an official publication of the American Association of Anatomists*, *235*(9), 2353-2375.

CHASSAING, N., RAGGE, N., PLAISANCIÉ, J., PATAT, O., GENEVIÈVE, D., RIVIER, F., et al. CALVAS, P. (2016). Confirmation of TENM3 involvement in autosomal recessive colobomatous microphthalmia. *American Journal of Medical Genetics Part A*, 170(7), 1895-1898.

CHASTE, P., ROEDER, K. & DEVLIN, B. (2017). The Yin and Yang of Autism Genetics: How Rare De Novo and Common Variations Affect Liability. *Annu Rev Genomics Hum Genet*, 18, 167-187.

CHATTERJEE, S., RAO, S. J. & PAL, J. K. (2001). Pathological mutations in 5′ untranslated regions of human genes. *eLS*, 1-8.

CHEN, G., ISHAN, M., YANG, J., KISHIGAMI, S., FUKUDA, T., SCOTT, G., et al. MISHINA, Y. (2017). Specific and spatial labeling of P0-Cre versus Wnt1-Cre in cranial neural crest in early mouse embryos. *Genesis*, *55*(6), e23034.

CHEN, Y., BEI, M., WOO, I., SATOKATA, I. & MAAS, R. (1996) Msx1 controls inductive signaling in mammalian tooth morphogenesis. *Development*, 122(10), pp. 3035-3044.

CHENEVIX-TRENCH, G., JONES, K., GREEN, A., DUFFY, D. & MARTIN, N. (1992) Cleft lip with or without cleft palate: associations with transforming growth factor alpha and retinoic acid receptor loci. *American Journal of Human Genetics*, 51(6), pp. 1377.

CHIASSON, C. M., WITTICH, K. B., VINCENT, P. A., FAUNDEZ, V. & KOWALCZYK, A. P. (2009). p120-catenin inhibits VE-cadherin internalization through a Rho-independent mechanism. *Mol Biol Cell*, 20(7), 1970-80.

CHO, K., LEE, M., GU, D., MUNOZ, W. A., JI, H., KLOC, M. & MCCREA, P. D. (2011). Kazrin, and its binding partners ARVCF-and delta-catenin, are required for Xenopus laevis craniofacial development. *Developmental Dynamics*, 240(12), 2601-2612.

CHONG, Jessica X., BUCKINGHAM, Kati J., JHANGIANI, Shalini N., BOEHM, C., SOBREIRA, N., SMITH, Joshua D., et al. BAMSHAD, Michael J. (2015). The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *The American Journal of Human Genetics*, 97(2), 199-215.

CHRISTIAN, P. & STEWART, C. P. (2010). Maternal Micronutrient Deficiency, Fetal Development, and the Risk of Chronic Disease. *The Journal of Nutrition*, 140(3), 437-445.

CIESIOLKA, M., DELVAEYE, M., VAN IMSCHOOT, G., VERSCHUERE, V., MCCREA, P., VAN ROY, F. & VLEMINCKX, K. (2004). p120 catenin is required for morphogenetic movements involved in the formation of the eyes and the craniofacial skeleton in Xenopus. *Journal of cell science*, 117(18), 4325-4339.

CLARKE, G. M., ANDERSON, C. A., PETTERSSON, F. H., CARDON, L. R., MORRIS, A. P. & ZONDERVAN, K. T. (2011). Basic statistical analysis in genetic case-control studies. *Nat Protoc*, 6(2), 121-33.

CLIFTON-BLIGH, R. J., WENTWORTH, J. M., HEINZ, P., CRISP, M. S., JOHN, R., LAZARUS, J. H., et al. CHATTERJEE, V. K. (1998). Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. *Nature genetics*, 19(4), 399.

CLOUTHIER, D. E., HOSODA, K., RICHARDSON, J. A., WILLIAMS, S. C., YANAGISAWA, H., KUWAKI, T., et al. YANAGISAWA, M. (1998). Cranial and cardiac neural crest defects in endothelin-A receptor-deficient mice. *Development*, 125(5), 813-824.

COBOURNE, M. T., & GREEN, J. B. (2012). Hedgehog signalling in development of the secondary palate. In *Cleft Lip and Palate* (Vol. 16, pp. 52-59). Karger Publishers.

COBOURNE, M. T. & SHARPE, P. T. (2003). Tooth and jaw: molecular mechanisms of patterning in the first branchial arch. *Archives of Oral Biology*, 48(1), 1-14.

COE, B. P., WITHERSPOON, K., ROSENFELD, J. A., VAN BON, B. W., VULTO-VAN SILFHOUT, A. T., BOSCO, P., et al. EICHLER, E. E. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet*, 46(10), 1063-71.

CONDIE, B. G., BAIN, G., GOTTLIEB, D. I. & CAPECCHI, M. R. (1997). Cleft palate in mice with a targeted mutation in the γ-aminobutyric acid-producing enzyme glutamic acid decarboxylase 67. *Proceedings of the National Academy of Sciences*, 94(21), 11451-11455.

CONLEY, Z. R., HAGUE, M., KUROSAKA, H., DIXON, J., DIXON, M. J. & TRAINOR, P. A. (2016). A quantitative method for defining high-arched palate using the Tcof1+/− mutant mouse as a model. *Developmental Biology*, 415(2), 296-305.

CORSTEN-JANSSEN, N., SAITTA, S. C., HOEFSLOOT, L. H., MCDONALD-MCGINN, D. M., DRISCOLL, D. A., DERKS, R., et al. VAN RAVENSWAAIJ-ARTS, C. M. (2013). More Clinical Overlap between 22q11.2 Deletion Syndrome and CHARGE Syndrome than Often Anticipated. *Mol Syndromol*, 4(5), 235-45.

COX, L. L., COX, T. C., MORENO URIBE, L. M., ZHU, Y., RICHTER, C. T., NIDEY, N., et al. ROSCIOLI, T. (2018). Mutations in the Epithelial Cadherin-p120-Catenin Complex Cause Mendelian Non-Syndromic Cleft Lip with or without Cleft Palate. *Am J Hum Genet*, 102(6), 1143-1157.

COX, T. C. (2004). Taking it to the max: the genetic and developmental mechanisms coordinating midfacial morphogenesis and dysmorphology. *Clinical genetics*, *65*(3), 163-176.

COX, T. C., LIDRAL, A. C., MCCOY, J. C., LIU, H., COX, L. L., ZHU, Y., et al. ROSCIOLI, T. (2019). Mutations in GDF11 and the extracellular antagonist, Follistatin, as a likely cause of Mendelian forms of orofacial clefting in humans. *Human Mutation*, 0(0).

CUERVO, R. & COVARRUBIAS, L. (2004). Death is the major fate of medial edge epithelial cells and the cause of basal lamina degradation during palatogenesis. *Development*, 131(1), 15-24.

CUERVO, R., VALENCIA, C., CHANDRARATNA, R. A. & COVARRUBIAS, L. (2002). Programmed cell death is required for palate shelf fusion and is regulated by retinoic acid. *Dev Biol*, 245(1), 145-56.

CUI, X.-M., SHIOMI, N., CHEN, J., SAITO, T., YAMAMOTO, T., ITO, Y., et al. SHULER, C. F. (2005). Overexpression of Smad2 in Tgf-β3-null mutant mice rescues cleft palate. *Developmental biology*, 278(1), 193-202.

CURNOW, R. N., & SMITH, C. (1975). Multifactorial models for familial diseases in man. *Journal of the royal statistical society: Series A (General)*, 138(2), 131-156.

CUTLER, T. L., MURPHY, K., HOPPER, J. L., KEOGH, L. A., DAI, Y. & CRAIG, J. M. (2015). Why accurate knowledge of zygosity is important to twins. *Twin Research and Human Genetics*, 18(3), 298-305.

D'SOUZA-SCHOREY, C. & CHAVRIER, P. (2006). ARF proteins: roles in membrane traffic and beyond. *Nat Rev Mol Cell Biol*, 7(5), 347-58.

DAAR, A. S., SCHERER, S. W. & HEGELE, R. A. (2006). Implications of copy-number variation in the human genome: a time for questions. *Nature Reviews Genetics*, 7(6), 414.

DANIELIAN, P. S., MUCCINO, D., ROWITCH, D. H., MICHAEL, S. K. & MCMAHON, A. P. (1998). Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Current biology*, 8(24), 1323-S2.

DARCHUK, L., ZAVERBNA, L., WOROBIEC, A. & VAN GRIEKEN, R. (2012) Molecular Spectroscopy Study of Human Tooth Tissues Affected by High Dose of External Ionizing Radiation (Caused by Nuclear Catastrophe of Chernobyl Plant), *Current Topics in Ionizing Radiation Research*IntechOpen.

DAVIS, M. A., IRETON, R. C. & REYNOLDS, A. B. (2003). A core function for p120-catenin in cadherin turnover. *J Cell Biol*, 163(3), 525-534.

DAVIS, M. A. & REYNOLDS, A. B. (2006). Blocked acinar development, E-cadherin reduction, and intraepithelial neoplasia upon ablation of p120-catenin in the mouse salivary gland. *Developmental cell*, 10(1), 21-31.

DEANGELIS, V., & NALBANDIAN, J. (1968). Ultrastructure of mouse and rat palatal processes prior to and during secondary palate formation. *Archives of oral biology*, 13(6), 601-IN7.

DEL VALLE-PÉREZ, B., CASAGOLDA, D., LUGILDE, E., VALLS, G., CODINA, M., DAVE, N., et al. DUÑACH, M. (2011). Wnt controls the transcriptional activity of Kaiso through CK1ε-dependent phosphorylation of p120-catenin. *J Cell Sci*, 124(13), 2298-2309.

DEPEW, M. J., LIU, J. K., LONG, J. E., PRESLEY, R., MENESES, J. J., PEDERSEN, R. A. & RUBENSTEIN, J. L. (1999). Dlx5 regulates regional development of the branchial arches and sensory capsules. *Development*, 126(17), 3831-3846.

DESCIPIO, C., CONLIN, L., ROSENFELD, J., TEPPERBERG, J., PASION, R., PATEL, A., et al. KRANTZ, I. D. (2012). Subtelomeric deletion of chromosome 10p15.3: clinical findings and molecular cytogenetic characterization. *Am J Med Genet A*, 158a(9), 2152-61.

DIXON, J., JONES, N. C., SANDELL, L. L., JAYASINGHE, S. M., CRANE, J., REY, J. P., et al. TRAINOR, P. A. (2006). Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proceedings of the national academy of sciences*, *103*(36), 13403-13408.

DIXON, M. J., DIXON, J., RASKOVA, D., LE BEAU, M. M., WILLIAMSON, R., KINGER, K. & LANDES, G. M. (1992). Genetic and physical mapping of the Treacher Collins syndrome locus: refinement of the localization to chromosome 5q32-33.2. *Human molecular genetics*, 1(4), 249-253.

DIXON, M. J., & FERGUSON, M. W. J. (1992). The effects of epidermal growth factor, transforming growth factors alpha and beta and platelet-derived growth factor on murine palatal shelves in organ culture. *Archives of oral biology*, *37*(5), 395-410.

DIXON, M. J., MARAZITA, M. L., BEATY, T. H. & MURRAY, J. C. (2011). Cleft lip and palate: understanding genetic and environmental influences. *Nature Reviews Genetics*, 12(3), 167-178.

DODE, C., LEVILLIERS, J., DUPONT, J. M., DE PAEPE, A., LE DU, N., SOUSSI-YANICOSTAS, N., et al. HARDELIN, J. P. (2003). Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. *Nat Genet*, 33(4), 463-5.

DONALDSON, J. G. & HONDA, A. (2005). Localization and function of Arf family GTPases. *Biochem Soc Trans*, 33(Pt 4), 639-42.

DRISCOLL, D. A., SPINNER, N. B., BUDARF, M. L., MCDONALD-MCGINN, D. M., ZACKAI, E. H., GOLDBERG, R. B., et al. JONES, M. C. (1992). Deletions and microdeletions of 22q11. 2 in velo-cardio-facial syndrome. *American journal of medical genetics*, 44(2), 261-268.

DUDAS, M., LI, W. Y., KIM, J., YANG, A. & KAARTINEN, V. (2007). Palatal fusion - where do the midline cells go? A review on cleft palate, a major human birth defect. *Acta Histochem*, 109(1), 1-14.

DUDAS, M., SRIDURONGRIT, S., NAGY, A., OKAZAKI, K., & KAARTINEN, V. (2004). Craniofacial defects in mice lacking BMP type I receptor Alk2 in neural crest cells. *Mechanisms of development*, *121*(2), 173-182.

EERENS, K., VLIETINCK, R., HEIDBÜCHEL, K., VAN OLMEN, A., DEROM, C., WILLEMS, G. & CARELS, C. (2001). Hypodontia and tooth formation in groups of children with cleft, siblings without cleft, and nonrelated controls. *The Cleft palate-craniofacial journal*, 38(4), 374-378.

EIBERG, H., BIXLER, D., NIELSEN, L. S., CONNEALLY, P. M. & MOHR, J. (1987). Suggestion of linkage of a major locus for nonsyndromic orofacial cleft with F13A and tentative assignment to chromosome 6. *Clin Genet*, 32(2), 129-32.

EILBECK, K., QUINLAN, A. & YANDELL, M. (2017). Settling the score: variant prioritization and Mendelian disease. *Nature Reviews Genetics*, 18(10), 599.

ELDER, P. K., FRENCH, C. L., SUBRAMANIAM, M., SCHMIDT, L. J. & GETZ, M. J. (1988). Evidence that the functional beta-actin gene is single copy in most mice and is associated with 5'sequences capable of conferring serum-and cycloheximide-dependent regulation. *Molecular and cellular biology*, 8(1), 480-485.

ELEY, L., ALQAHTANI, A. M., MACGROGAN, D., RICHARDSON, R. V., MURPHY, L., SALGUERO-JIMENEZ, A., et al. HENDERSON, D. J. (2018). A novel source of arterial valve cells linked to bicuspid aortic valve without raphe in mice. *Elife*, 7.

ELIA, L. P., YAMAMOTO, M., ZANG, K. & REICHARDT, L. F. (2006). p120 catenin regulates dendritic spine and synapse development through Rho-family GTPases and cadherins. *Neuron*, 51(1), 43-56.

FARINA, A., WYSZYNSKI, D. F., PEZZETTI, F., SCAPOLI, L., MARTINELLI, M., CARINCI, F., et al. CARINCI, P. (2002). Classification of oral clefts by affection site and laterality: a genotype–phenotype correlation study. *Orthodontics & Craniofacial Research*, 5(3), 185-191.

FARONATO, M., NGUYEN, V. T., PATTEN, D. K., LOMBARDO, Y., STEEL, J. H., PATEL, N., et al. MAGNANI, L. (2015). DMXL2 drives epithelial to mesenchymal transition in hormonal therapy resistant breast cancer through Notch hyper-activation. *Oncotarget*, 6(26), 22467-79.

FELDMAN, G., KAPPES, D., MOOKERJEE-BASU, J., FREEMAN, T., FERTALA, A. & PARVIZI, J. (2019). Novel mutation in Teneurin 3 found to co-segregate in all affecteds in a multi-generation family with developmental dysplasia of the hip. *Journal of Orthopaedic Research*, 37(1), 171-180.

FERGUSON, M. W. (1988). Palate development. *Development*, 103(Supplement), 41-60.

FERGUSON, M. W., & HONIG, L. S. (1984). Epithelial-mesenchymal interactions during vertebrate palatogenesis. *Current topics in developmental biology*, *19*, 137-164.

FIRTH, H. V., RICHARDS, S. M., BEVAN, A. P., CLAYTON, S., CORPAS, M., RAJAN, D., et al. CARTER, N. P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet*, 84(4), 524-33.

FITCHETT, J. E. & HAY, E. D. 1989. Medial edge epithelium transforms to mesenchyme after embryonic palatal shelves fuse. *Developmental biology*, 131, 455–474.

FOGH-ANDERSEN, P. (1946). Harelip and cleft palate; 1,000 patients submitted to operation. *Acta Chir Scand*, 94(3/4), 213-42.

FOGH-ANDERSEN, P. (1967). Genetic and Non-Genetic Factors in the Etiology of Facial Clefts. *Scandinavian Journal of Plastic and Reconstructive Surgery*, 1(1), 22-29.

FOKSTUEN, S., BOTTANI, A., MEDEIROS, P. F., ANTONARAKIS, S. E., STOLL, C. & SCHINZEL, A. (1997). Laryngeal atresia type III (glottic web) with 22q11. 2 microdeletion: report of three patients. *American journal of medical genetics*, 70(2), 130-133.

FRANCOMANO, C. A., LIBERFARB, R., HIROSE, T., MAUMENEE, I., STREETEN, E., MEYERS, D. & PYERITZ, R. (1987). The Stickler syndrome: evidence for close linkage to the structural gene for type II collagen. *Genomics*, 1(4), 293-296.

FRASER, F. C. (1970). The genetics of cleft lip and cleft palate. *Am J Hum Genet*, 22(3), 336-52.

FRASER, F. C. (1976). The multifactorial/threshold concept—uses and misuses. *Teratology*, 14(3), 267-280.

FRASER, F. C. (1980). The William Allan Memorial Award Address: evolution of a palatable multifactorial threshold model. *Am J Hum Genet*, 32(6), 796-813.

FRASER, F. C., WALKER, B. E. & TRASLER, D. G. (1957). EXPERIMENTAL PRODUCTION OF CONGENITAL CLEFT PALATE: GENETIC AND ENVIRONMENTAL FACTORS. *Pediatrics*, 19(4), 782-787.

FRITZEN, D., KUECHLER, A., GRIMMEL, M., BECKER, J., PETERS, S., STURM, M., et al. STROM, T. M. (2018). De novo FBXO11 mutations are associated with intellectual disability and behavioural anomalies. *Human Genetics*, 137(5), 401-411.

FROMER, M., MORAN, J. L., CHAMBERT, K., BANKS, E., BERGEN, S. E., RUDERFER, D. M., et al. PURCELL, S. M. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet*, 91(4), 597-607.

FU, J., BEATY, T. H., SCOTT, A. F., HETMANSKI, J., PARKER, M. M., WILSON, J. E., et al. SCHARPF, R. B. (2017a). Whole exome association of rare deletions in multiplex oral cleft families. *Genet Epidemiol*, 41(1), 61-69.

FU, W., LIGABUE, A., ROGERS, K. J., AKEY, J. M. & MONNAT, R. J., Jr. (2017b). Human RECQ Helicase Pathogenic Variants, Population Variation and "Missing" Diseases. *Hum Mutat*, 38(2), 193-203.

FUKUMOTO, Y., SHINTANI, Y., REYNOLDS, A. B., JOHNSON, K. R. & WHEELOCK, M. J. (2008). The regulatory or phosphorylation domain of p120 catenin controls E-cadherin dynamics at the plasma membrane. *Experimental cell research*, 314(1), 52-67.

GALLAGHER, E. R., COLLETT, B. R., BARRON, S., ROMITTI, P., ANSLEY, T., & WEHBY, G. L. (2017). Laterality of oral clefts and academic achievement. *Pediatrics*, 139(2), e20162662.

GEIS, K., ABERLE, H., KÜHL, M., KEMLER, R. & WEDLICH, D. (1998). Expression of the Armadillo family member p120 cas 1B in Xenopus embryos affects head differentiation but not axis formation. *Development genes and evolution*, 207(7), 471-481.

GHOUMID, J., STICHELBOUT, M., JOURDAIN, A. S., FRENOIS, F., LEJEUNE-DUMOULIN, S., ALEX-CORDIER, M. P., et al. MANOUVRIER-HANU, S. (2017). Blepharocheilodontic syndrome is a CDH1 pathway-related disorder due to mutations in CDH1 and CTNND1. *Genet Med*, 19(9), 1013-1021.

GINN, S. L., AMAYA, A. K., ALEXANDER, I. E., EDELSTEIN, M., & ABEDI, M. R. (2018). Gene therapy clinical trials worldwide to 2017: An update. *The journal of gene medicine*, 20(5), e3015.

GINSBURG, G. S., & PHILLIPS, K. A. (2018). Precision medicine: from science to value. *Health affairs*, 37(5), 694-701.

GIRIRAJAN, S., CAMPBELL, C. D. & EICHLER, E. E. (2011). Human copy number variation and complex genetic disease. *Annual review of genetics*, 45, 203-226.

GOOS, J. A. C., SWAGEMAKERS, S. M. A., TWIGG, S. R. F., VAN DOOREN, M. F., HOOGEBOOM, A. J. M., BEETZ, C., et al. HURST, J. A. (2017). Identification of causative variants in TXNL4A in Burn-McKeown syndrome and isolated choanal atresia. *Eur J Hum Genet*, 25(10), 1126-1133.

GORDON, C. T., PETIT, F., OUFADEM, M., DECAESTECKER, C., JOURDAIN, A.-S., ANDRIEUX, J., et al. BAUMANN, C. (2012). EFTUD2 haploinsufficiency leads to syndromic oesophageal atresia. *Journal of medical genetics*, 49(12), 737-746.

GOUDY, S., LAW, A., SANCHEZ, G., BALDWIN, H. S. & BROWN, C. (2010). Tbx1 is necessary for palatal elongation and elevation. *Mech Dev*, 127(5-6), 292-300.

GRANT, S. F. A., WANG, K., ZHANG, H., GLABERSON, W., ANNAIAH, K., KIM, C. E., et al. HAKONARSON, H. (2009). A Genome-Wide Association Study Identifies a Locus for Nonsyndromic Cleft Lip with or without Cleft Palate on 8q24. *The Journal of Pediatrics*, 155(6), 909-913.

GRAY, R. S., ABITUA, P. B., WLODARCZYK, B. J., SZABO-ROGERS, H. L., BLANCHARD, O., LEE, I., et al. FINNELL, R. H. (2009). The planar cell polarity effector Fuz is essential for targeted membrane trafficking, ciliogenesis and mouse embryonic development. *Nat Cell Biol*, 11(10), 1225-32.

GREGOR, A., SADLEIR, L. G., ASADOLLAHI, R., AZZARELLO-BURRI, S., BATTAGLIA, A., OUSAGER, L. B., et al. ZWEIER, C. (2018). De Novo Variants in the F-Box Protein FBXO11 in 20 Individuals with a Variable Neurodevelopmental Disorder. *Am J Hum Genet*, 103(2), 305-316.

GRIFFITH, C. M. & HAY, E. D. (1992). Epithelial-mesenchymal transformation during palatal fusion: carboxyfluorescein traces cells at light and electron microscopic levels. *Development*, 116(4), 1087-99.

GRITLI-LINDE, A. (2007). Molecular control of secondary palate development. *Dev Biol*, 301(2), 309-26.

GROSEN, D., CHEVRIER, C., SKYTTHE, A., BILLE, C., MOLSTED, K., SIVERTSEN, A., et al. CHRISTENSEN, K. (2010). A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance. *J Med Genet*, 47(3), 162-8.

GU, D., SATER, A. K., JI, H., CHO, K., CLARK, M., STRATTON, S. A., et al. MCCREA, P. D. (2009). Xenopus δ-catenin is essential in early embryogenesis and is functionally linked to cadherins and small GTPases. *Journal of cell science*, 122(22), 4049-4061.

GUNDUZ ARSLAN, S., DEVECIOGLU KAMA, J., OZER, T. & YAVUZ, I. (2007). Craniofacial and upper airway cephalometrics in hypohidrotic ectodermal dysplasia. *Dentomaxillofacial Radiology*, 36(8), 478-483.

GUTTMACHER, A. E., NABEL, E. G. & COLLINS, F. S. (2009). Why data-sharing policies matter. *Proc Natl Acad Sci U S A*, 106(40), 16894.

GUTTMACHER, A. E., PORTEOUS, M. E. & MCINERNEY, J. D. (2007). Science & society: Educating health-care professionals about genetics and genomics. *Nature Reviews Genetics*, 8(2), 151.

HALLGRÍMSSON, B., DONNABHÁIN, B. Ó., BLOM, D. E., LOZADA, M. C., & WILLMORE, K. T. (2005). Why are rare traits unilaterally expressed?: trait frequency and unilateral expression for cranial nonmetric traits in humans. *American journal of physical anthropology*, 128(1), 14-25.

HAMMOND, N. L., DIXON, J. & DIXON, M. J. (2017) Periderm: Life-cycle and function during orofacial and epidermal development, *Seminars in cell & developmental biology*. Elsevier.

HAN, J., MAYO, J., XU, X., LI, J., BRINGAS, P., MAAS, R. L., et al. CHAI, Y. (2009). Indirect modulation of Shh signaling by Dlx5 affects the oral-nasal patterning of palate and rescues cleft palate in Msx1-null mice. *Development*, 136(24), 4225-4233.

HARBISON, R. D. & BECKER, B. A. (1969). Relation of dosage and time of administration of diphenylhydantoin to its teratogenic effect in mice. *Teratology*, 2(4), 305-311.

HARRIS, T. J. (2012). An introduction to adherens junctions: from molecular mechanisms to tissue development and disease. *Subcell Biochem*, 60, 1-5.

HARRISON, M., BUSHELL, C. J. & IRVING, M. (2018). 32 and you - genetic testing for dental disorders. *Br Dent J*, 224(10), 829-832.

HARVEY, E. K., FOGEL, C. E., PEYROT, M., CHRISTENSEN, K. D., TERRY, S. F. & MCINERNEY, J. D. (2007). Providers' knowledge of genetics: A survey of 5915 individuals and families with genetic conditions. *Genetics In Medicine*, 9, 259.

HÄSTBACKA, J., DE LA CHAPELLE, A., MAHTANI, M. M., CLINES, G., REEVE-DALY, M. P., DALY, M., et al. WEAVER, A. (1994). The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell*, 78(6), 1073-1087.

HATZFELD, M. (2005). The p120 family of cell adhesion molecules. *European journal of cell biology*, 84(2-3), 205-214.

HE, F., POPKIE, A. P., XIONG, W., LI, L., WANG, Y., PHIEL, C. J. & CHEN, Y. (2010). Gsk3β is required in the epithelium for palatal elevation in mice. *Developmental Dynamics*, 239(12), 3235-3246.

HE, F., XIONG, W., WANG, Y., LI, L., LIU, C., YAMAGAMI, T., et al. CHEN, Y. (2011). Epithelial Wnt/β-catenin signaling regulates palatal shelf fusion through regulation of Tgfβ3 expression. *Developmental biology*, *350*(2), 511-519.

HE, F., XIONG, W., YU, X., ESPINOZA-LEWIS, R., LIU, C., GU, S., et al. CHEN, Y. (2008). Wnt5a regulates directional cell migration and cell proliferation via Ror2-mediated noncanonical pathway in mammalian palate development. *Development*, 135(23), 3871-9.

HELIÖVAARA, A., RANTANEN, I. & ARTE, S. (2011). Dental development and tooth agenesis in children with velocardiofacial syndrome. *International journal of paediatric dentistry*, 21(6), 446-450.

HELLE, E., CÓRDOVA-PALOMERA, A., OJALA, T., SAHA, P., POTINY, P., GUSTAFSSON, S., et al. PRIEST, J. R. (2019). Loss of function, missense, and intronic variants in NOTCH1 confer different risks for left ventricular outflow tract obstructive heart defects in two European cohorts. *Genetic Epidemiology*, 43(2), 215-226.

HENDLEY, A. M., PROVOST, E., BAILEY, J. M., WANG, Y. J., CLEVELAND, M. H., BLAKE, D., et al. REYNOLDS, A. B. (2015). p120 Catenin is required for normal tubulogenesis but not epithelial integrity in developing mouse pancreas. *Developmental biology*, 399(1), 41-53.

HENNEKAM, R. C., ALLANSON, J. E. & KRANTZ, I. D. (2010) *Gorlin's Syndromes of the Head and Neck*Oxford University Press.

HÉON-KLIN, V. (2017). European Reference networks for rare diseases: what is the conceptual framework?. *Orphanet journal of rare diseases*, 12(1), 137.

HILL, C., JACOBS, B., KENNEDY, L., ROHDE, S., ZHOU, B., BALDWIN, S., & GOUDY, S. (2015). Cranial neural crest deletion of VEGFa causes cleft palate with aberrant vascular and bone development. *Cell and tissue research*, *361*(3), 711-722**.**

HILL, C. A., REEVES, R. H. & RICHTSMEIER, J. T. (2007). Effects of aneuploidy on skull growth in a mouse model of Down syndrome. *Journal of Anatomy*, 210(4), 394-405.

HILLS, A., AHN, J. W., DONAGHUE, C., THOMAS, H., MANN, K. & OGILVIE, C. M. (2010). MLPA for confirmation of array CGH results and determination of inheritance. *Molecular cytogenetics*, 3(1), 19.

HIRSCHHORN, J. N. & DALY, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, 6(2), 95-108.

HITZ, M.-P., LEMIEUX-PERREAULT, L.-P., MARSHALL, C., FEROZ-ZADA, Y., DAVIES, R., YANG, S. W., et al. CULLUM, R. (2012). Rare copy number variants contribute to congenital left-sided heart disease. *PLoS genetics*, 8(9), e1002903.

HOEBEL, A. K., DRICHEL, D., VAN DE VORST, M., BOHMER, A. C., SIVALINGAM, S., ISHORST, N., et al. LUDWIG, K. U. (2017). Candidate Genes for Nonsyndromic Cleft Palate Detected by Exome Sequencing. *J Dent Res*, 96(11), 1314-1321.

HOFMEISTER, W., NILSSON, D., TOPA, A., ANDERLID, B. M., DARKI, F., MATSSON, H., et al. LINDSTRAND, A. (2015). CTNND2-a candidate gene for reading problems and mild intellectual disability. *J Med Genet*, 52(2), 111-22.

HOISCHEN, A., VAN BON, B. W., RODRIGUEZ-SANTIAGO, B., GILISSEN, C., VISSERS, L. E., DE VRIES, P., et al. DE VRIES, B. B. (2011). De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat Genet*, 43(8), 729-31.

HOLLISTER, D. W., GODFREY, M., SAKAI, L. Y. & PYERITZ, R. E. (1990). Immunohistologic abnormalities of the microfibrillar-fiber system in the Marfan syndrome. *New England Journal of Medicine*, 323(3), 152-159.

HOLMES, G. & BASILICO, C. (2012). Mesodermal expression of Fgfr2S252W is necessary and sufficient to induce craniosynostosis in a mouse model of Apert syndrome. *Dev Biol*, 368(2), 283-93.

HOMSY, J., ZAIDI, S., SHEN, Y., WARE, J. S., SAMOCHA, K. E., KARCZEWSKI, K. J., et al. CHUNG, W. K. (2015). De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science*, 350(6265), 1262-1266.

HONEIN, M. A., RASMUSSEN, S. A., REEFHUIS, J., ROMITTI, P. A., LAMMER, E. J., SUN, L. & CORREA, A. (2007). Maternal Smoking and Environmental Tobacco Smoke Exposure and the Risk of Orofacial Clefts. *Epidemiology*, 18(2), 226-233.

HONG, J. Y., OH, I.-H. & MCCREA, P. D. (2016). Phosphorylation and isoform use in p120-catenin during development and tumorigenesis. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1863(1), 102-114.

HOWE, B. J., COOPER, M. E., VIEIRA, A. R., WEINBERG, S. M., RESICK, J. M., NIDEY, N. L., et al. MORENO URIBE, L. M. (2015). Spectrum of Dental Phenotypes in Nonsyndromic Orofacial Clefting. *Journal of Dental Research*, 94(7), 905-912.

HRUBEC, T. C., PRATER, M. R., TOOPS, K. A. & HOLLADAY, S. D. (2006). Reduction in diabetes-induced craniofacial defects by maternal immune stimulation. *Birth Defects Res B Dev Reprod Toxicol*, 77(1), 1-9.

HU, J., YANG, Q., CHEN, Z., LIANG, W., FENG, J. & DING, G. (2019). Small GTPase Arf6 regulates diabetes-induced cholesterol accumulation in podocytes. *J Cell Physiol*.

HUANG, L., VANSTONE, M. R., HARTLEY, T., OSMOND, M., BARROWMAN, N., ALLANSON, J., et al. LINES, M. A. (2016). Mandibulofacial Dysostosis with Microcephaly: Mutation and Database Update. *Hum Mutat*, 37(2), 148-54.

HUANG, X., GOUDY, S. L., KETOVA, T., LITINGTUNG, Y. & CHIANG, C. (2008). Gli3-deficient mice exhibit cleft palate associated with abnormal tongue development. *Developmental Dynamics*, 237(10), 3079-3087.

HUBBARD, M. J., MANGUM, J. E., PEREZ, V. A., NERVO, G. J. & HALL, R. K. (2017). Molar hypomineralisation: a call to arms for enamel researchers. *Frontiers in physiology*, 8, 546.

HUSTON, J., BIXLER, D. & YU, P. (1984). Velopharyngeal variations in relatives of cleft-affected individuals. *Journal of craniofacial genetics and developmental biology*, 5(2), 167-173.

HWANG, J. & MAQUAT, L. E. (2011). Nonsense-mediated mRNA decay (NMD) in animal embryogenesis: to die or not to die, that is the question. *Curr Opin Genet Dev*, 21(4), 422-30.

ICBDMS (2001) Annual Report 2001 (with data for 1999). Eds. Lowry B, Castilla EE, Mastroiacovo P, Siffel C. ISSN 0743-5703.

INOUE, H. & RANDAZZO, P. A. (2007). Arf GAPs and Their Interacting Proteins. *Traffic*, 8(11), 1465-1475.

IOSSIFOV, I., O'ROAK, B. J., SANDERS, S. J., RONEMUS, M., KRUMM, N., LEVY, D., et al. WIGLER, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, 515(7526), 216-21.

IRETON, R. C., DAVIS, M. A., VAN HENGEL, J., MARINER, D. J., BARNES, K., THORESON, M. A., et al. REYNOLDS, A. B. (2002). A novel role for p120 catenin in E-cadherin function. *J Cell Biol*, 159(3), 465-76.

ISEKI, S., ISHII-SUZUKI, M., TSUNEKAWA, N., YAMADA, Y., ETO, K. & OBATA, K. (2007). Experimental induction of palate shelf elevation in glutamate decarboxylase 67-deficient mice with cleft palate due to vertically oriented palatal shelf. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 79(10), 688-695.

ISHIYAMA, N., LEE, S.-H., LIU, S., LI, G.-Y., SMITH, M. J., REICHARDT, L. F. & IKURA, M. (2010). Dynamic and static interactions between p120 catenin and E-cadherin regulate the stability of cell-cell adhesion. *Cell*, 141(1), 117-128.

ISIDOR, B., LINDENBAUM, P., PICHON, O., BEZIEAU, S., DINA, C., JACQUEMONT, S., et al. LE CAIGNEC, C. (2011). Truncating mutations in the last exon of NOTCH2 cause a rare skeletal disorder with osteoporosis. *Nat Genet*, 43(4), 306-8.

JAKOBSEN, L. P., KNUDSEN, M. A., LESPINASSE, J., GARCÍA AYUSO, C., RAMOS, C., FRYNS, J.-P., et al. TOMMERUP, N. (2006). The Genetic Basis of the Pierre Robin Sequence. *The Cleft Palate-Craniofacial Journal*, 43(2), 155-159.

JAN, J. & VRBIČ, V. (2000). Polychlorinated biphenyls cause developmental enamel defects in children. *Caries research*, 34(6), 469-473.

JANSEN, G. A., MIHALIK, S. J., WATKINS, P. A., MOSER, H. W., JAKOBS, C., HEIJMANS, H. S. & WANDERS, R. J. (1997). Phytanoyl-CoA hydroxylase is not only deficient in classical Refsum disease but also in rhizomelic chondrodysplasia punctata. *J Inherit Metab Dis*, 20(3), 444-6.

JANSEN, S., VAN DER WERF, I. M., INNES, A. M., AFENJAR, A., AGRAWAL, P. B., ANDERSON, I. J., et al. DE VRIES, B. B. A. (2019). De novo variants in FBXO11 cause a syndromic form of intellectual disability with behavioral problems and dysmorphisms. *European Journal of Human Genetics*, 27(5), 738-746.

JEONG, J., CESARIO, J., ZHAO, Y., BURNS, L., WESTPHAL, H. & RUBENSTEIN, J. L. (2012). Cleft palate defect of Dlx1/2-/- mutant mice is caused by lack of vertical outgrowth in the posterior palate. *Dev Dyn*, 241(11), 1757-69.

JEZEWSKI, P. A., VIEIRA, A. R., NISHIMURA, C., LUDWIG, B., JOHNSON, M., O'BRIEN, S. E., et al. MURRAY, J. C. (2003). Complete sequencing shows a role for MSX1 in non-syndromic cleft lip and palate. *J Med Genet*, 40(6), 399-407.

JIA, S., ZHOU, J., FANELLI, C., WEE, Y., BONDS, J., SCHNEIDER, P., et al. D'SOUZA, R. N. (2017). Small-molecule Wnt agonists correct cleft palates in Pax9 mutant mice in utero. *Development*, *144*(20), 3819-3828.

JIN, J. Z., TAN, M., WARNER, D. R., DARLING, D. S., HIGASHI, Y., GRIDLEY, T., & DING, J. (2010). Mesenchymal cell remodeling during mouse secondary palate reorientation. *Developmental Dynamics*, *239*(7), 2110-2117.

JONES, N. C., LYNN, M. L., GAUDENZ, K., SAKAI, D., AOTO, K., REY, J.-P., et al. TRAINOR, P. A. (2008). Prevention of the neurocristopathy Treacher Collins syndrome through inhibition of p53 function. *Nature Medicine*, 14, 125.

JUGESSUR, A., FARLIE, P. & KILPATRICK, N. (2009a). The genetics of isolated orofacial clefts: from genotypes to subphenotypes. *Oral Diseases*, 15(7), 437-453.

JUGESSUR, A., SHI, M., GJESSING, H. K., LIE, R. T., WILCOX, A. J., WEINBERG, C. R., et al. MURRAY, J. C. (2009b). Genetic determinants of facial clefting: analysis of 357 candidate genes using two national cleft studies from Scandinavia. *PLoS One*, 4(4), e5385.

KÄLLÉN, B., MASTROIACOVO, P. & ROBERT, E. (1996). Major congenital malformations in Down syndrome. *American Journal of Medical Genetics*, 65(2), 160-166.

KÄLLÉN, K. (1997). Maternal Smoking and Orofacial Clefts. *The Cleft Palate-Craniofacial Journal*, 34(1), 11-16.

KARCZEWSKI, K. J., FRANCIOLI, L. C., TIAO, G., CUMMINGS, B. B., ALFÖLDI, J., WANG, Q., et al. MACARTHUR, D. G. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210.

KAUFMAN, M. H. & KAUFMAN, M. H. (1992) *The atlas of mouse development*, *428*Academic press London.

KE, C.-Y., XIAO, W.-L., CHEN, C.-M., LO, L.-J. & WONG, F.-H. (2015). IRF6 is the mediator of TGFβ3 during regulation of the epithelial mesenchymal transition and palatal fusion. *Scientific Reports*, 5, 12791.

KEIRSEBILCK, A., BONNÉ, S., STAES, K., VAN HENGEL, J., NOLLET, F., REYNOLDS, A. & VAN ROY, F. (1998). Molecular cloning of the human p120ctnCatenin gene (CTNND1): Expression of multiple alternatively spliced isoforms. *Genomics*, 50(2), 129-146.

KERAMEDDIN, S., NAMIPASHAKI, A., EBRAHIMI, S. & ANSARI-POUR, N. (2015). IRF6 Is a Marker of Severity in Nonsyndromic Cleft Lip/Palate. *J Dent Res*, 94(9 Suppl), 226s-32s.

KHOKHA, M. K., CHUNG, C., BUSTAMANTE, E. L., GAW, L. W., TROTT, K. A., YEH, J., et al. GRAMMER, T. C. (2002). Techniques and probes for the study of Xenopus tropicalis development. *Dev Dyn*, 225(4), 499-510.

KHOURY, M. J., GWINN, M., YOON, P. W., DOWLING, N., MOORE, C. A., & BRADLEY, L. (2007). The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention?. *Genetics in medicine*, 9(10), 665-674.

KHOURY, M. J., WEINSTEIN, A., PANNY, S., HOLTZMAN, N. A., LINDSAY, P. K., FARREL, K. & EISENBERG, M. (1987). Maternal cigarette smoking and oral clefts: a population-based study. *American Journal of Public Health*, 77(5), 623-625.

KIEVIT, A., TESSADORI, F., DOUBEN, H., JORDENS, I., MAURICE, M., HOOGEBOOM, J., et al. VAN HAAFTEN, G. (2018). Variants in members of the cadherin-catenin complex, CDH1 and CTNND1, cause blepharocheilodontic syndrome. *Eur J Hum Genet*, 26(2), 210-219.

KIM, S., LEWIS, A. E., SINGH, V., MA, X., ADELSTEIN, R., & BUSH, J. O. (2015). Convergence and extrusion are required for normal fusion of the mammalian secondary palate. *PLoS biology*, *13*(4).

KIM, S. W., PARK, J.-I., SPRING, C. M., SATER, A. K., JI, H., OTCHERE, A. A., et al. MCCREA, P. D. (2004). Non-canonical Wnt signals are modulated by the Kaiso transcriptional repressor and p120-catenin. *Nature cell biology*, 6(12), 1212.

KIRBY, M. L. & WALDO, K. L. (1990). Role of neural crest in congenital heart disease. *Circulation*, 82(2), 332-340.

KIRCHER, M., WITTEN, D. M., JAIN, P., O'ROAK, B. J., COOPER, G. M. & SHENDURE, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, 46(3), 310-5.

KLEIN, O. D., OBEROI, S., HUYSSEUNE, A., HOVORAKOVA, M., PETERKA, M. & PETERKOVA, R. (2013) Developmental disorders of the dentition: an update, *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*. Wiley Online Library.

KOCHILAS, L., MERSCHER-GOMEZ, S., LU, M. M., POTLURI, V., LIAO, J., KUCHERLAPATI, R., et al. EPSTEIN, J. A. (2002). The role of neural crest during cardiac development in a mouse model of DiGeorge syndrome. *Developmental biology*, 251(1), 157-166.

KÖHLER, S., CARMODY, L., VASILEVSKY, N., JACOBSEN, J. O B., DANIS, D., GOURDINE, J.-P., et al. ROBINSON, P. N. (2018). Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Research*, 47(D1), D1018-D1027.

KONDO, S., SCHUTTE, B. C., RICHARDSON, R. J., BJORK, B. C., KNIGHT, A. S., WATANABE, Y., et al. SANDER, A. (2002). Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nature Genetics*, 32(2), 285-289.

KRATOCHWIL, K., DULL, M., FARINAS, I., GALCERAN, J. AND GROSSCHEDL, R. (1996) Lef1 expression is activated by BMP-4 and regulates inductive tissue interactions in tooth and hair development. *Genes and Development*, 10(11), pp. 1382-1394.

KRATOCHWIL, K., GALCERAN, J., TONTSCH, S., ROTH, W., & GROSSCHEDL, R. (2002). FGF4, a direct target of LEF1 and Wnt signaling, can rescue the arrest of tooth organogenesis in Lef1−/− mice. *Genes & development*, *16*(24), 3173-3185.

KUMAR, P., HENIKOFF, S. & NG, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, 4(7), 1073-81.

KURNIT, D. M., LAYTON, W. M., & MATTHYSSE, S. (1987). Genetics, chance, and morphogenesis. *American journal of human genetics*, *41*(6), 979.

LAMMI, L., ARTE, S., SOMER, M., JÄRVINEN, H., LAHERMO, P., THESLEFF, I., et al. NIEMINEN, P. (2004). Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *The American Journal of Human Genetics*, 74(5), 1043-1050.

LAN, Y., XU, J., & JIANG, R. (2015). Cellular and molecular mechanisms of palatogenesis. In *Current topics in developmental biology* (Vol. 115, pp. 59-84). Academic Press.

LANDER, E. S. (2011). Initial impact of the sequencing of the human genome. *Nature*, 470(7333), 187-97.

LASA, A., Y CAJAL, T. R., LLORT, G., SUELA, J., CIGUDOSA, J., CORNET, M., et al. BAIGET, M. (2010). Copy number variations are not modifiers of phenotypic expression in a pair of identical twins carrying a BRCA1 mutation. *Breast cancer research and treatment*, 123(3), 901-905.

LAURIKKALA, J., MIKKOLA, M., MUSTONEN, T., ÅBERG, T., KOPPINEN, P., PISPA, J., et al. THESLEFF, I. (2001). TNF signaling via the ligand–receptor pair ectodysplasin and edar controls the function of epithelial signaling centers and is regulated by Wnt and activin during tooth organogenesis. *Developmental biology*, 229(2), 443-455.

LEE, J. M., KIM, J. Y., CHO, K. W., LEE, M. J., CHO, S. W., KWAK, S., et al. JUNG, H. S. (2008). Wnt11/Fgfr1b cross-talk modulates the fate of cells in palate development. *Developmental biology*, 314(2), 341-350.

LE GOFF, C., MAHAUT, C., ABHYANKAR, A., LE GOFF, W., SERRE, V., AFENJAR, A., et al. CORMIER-DAIRE, V. (2011). Mutations at a single codon in Mad homology 2 domain of SMAD4 cause Myhre syndrome. *Nat Genet*, 44(1), 85-8.

LEHALLE, D., GORDON, C. T., OUFADEM, M., GOUDEFROYE, G., BOUTAUD, L., ALESSANDRI, J. L., et al. BOUTE-BENEJEAN, O. (2014). Delineation of EFTUD 2 Haploinsufficiency-Related Phenotypes Through a Series of 36 Patients. *Human mutation*, 35(4), 478-485.

LEHMAN, H. L., YANG, X., WELSH, P. A. & STAIRS, D. B. (2015). p120-catenin down-regulation and epidermal growth factor receptor overexpression results in a transformed epithelium that mimics esophageal squamous cell carcinoma. *The American journal of pathology*, 185(1), 240-251.

LEI, Y., ZHU, H., YANG, W., ROSS, M. E., SHAW, G. M. & FINNELL, R. H. (2014). Identification of novel CELSR1 mutations in spina bifida. *PloS one*, 9(3), e92207.

LELIEVELD, S. H., REIJNDERS, M. R., PFUNDT, R., YNTEMA, H. G., KAMSTEEG, E. J., DE VRIES, P., et al. GILISSEN, C. (2016). Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat Neurosci*, 19(9), 1194-6.

LEOPOLD, C., DE BARROS, A., CELLIER, C., DROUIN-GARRAUD, V., DEHESDIN, D. & MARIE, J.-P. (2012). Laryngeal abnormalities are frequent in the 22q11 deletion syndrome. *International journal of pediatric otorhinolaryngology*, 76(1), 36-40.

LESLIE, E. J., CARLSON, J. C., COOPER, M. E., CHRISTENSEN, K., WEINBERG, S. M. & MARAZITA, M. L. (2017). Exploring Subclinical Phenotypic Features in Twin Pairs Discordant for Cleft Lip and Palate. *Cleft Palate Craniofac J*, 54(1), 90-93.

LESLIE, E. J., CARLSON, J. C., SHAFFER, J. R., FEINGOLD, E., WEHBY, G., LAURIE, C. A., et al. MARAZITA, M. L. (2016a). A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. *Human Molecular Genetics*, 25(13), 2862-2872.

LESLIE, E. J., KOBOLDT, D. C., KANG, C. J., MA, L., HECHT, J. T., WEHBY, G. L., et al. MARAZITA, M. L. (2016b). IRF6 mutation screening in non-syndromic orofacial clefting: analysis of 1521 families. *Clinical Genetics*, 90(1), 28-34.

LESLIE, E. J., LIU, H., CARLSON, J. C., SHAFFER, J. R., FEINGOLD, E., WEHBY, G., et al. MARAZITA, M. L. (2016c). A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. *American journal of human genetics*, 98(4), 744-54.

LESLIE, E. J., MANSILLA, M. A., BIGGS, L. C., SCHUETTE, K., BULLARD, S., COOPER, M., et al. MURRAY, J. C. (2012). Expression and mutation analyses implicate ARHGAP29 as the etiologic gene for the cleft lip with or without cleft palate locus identified by genome-wide association on chromosome 1p22. *Birth Defects Res A Clin Mol Teratol*, 94(11), 934-42.

LESLIE, E. J. & MARAZITA, M. L. (2013). Genetics of cleft lip and cleft palate. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 163(4), 246-258.

LETRA, A., BJORK, B., COOPER, M., SZABO-ROGERS, H., DELEYIANNIS, F., FIELD, L., CZEIZEL, A., MA, L., GARLET, G. & POLETTA, F. (2012) Association of AXIN2 with non-syndromic oral clefts in multiple populations. *Journal of dental research*, 91(5), pp. 473-478.

LETRA, A., DE ALMEIDA, A. L., KAIZER, R., ESPER, L. A., SGARBOSA, S. & GRANJEIRO, J. M. (2007a). Intraoral features of Apert's syndrome. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, 103(5), e38-41.

LETRA, A., MENEZES, R., GRANJEIRO, J. & VIEIRA, A. (2007b). Defining subphenotypes for oral clefts based on dental development. *Journal of Dental Research*, 86(10), 986-991.

LETRA, A., MENEZES, R., GRANJEIRO, J. M. & VIEIRA, A. R. (2009). AXIN2 and CDH1 polymorphisms, tooth agenesis, and oral clefts. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 85(2), 169-173.

LEWANDOSKI, M., MEYERS, E. & MARTIN, G. (1997) Analysis of Fgf8 gene function in vertebrate development, *Cold Spring Harbor symposia on quantitative biology*. Cold Spring Harbor Laboratory Press.

LEWIS, A. E., VASUDEVAN, H. N., O'NEILL, A. K., SORIANO, P. & BUSH, J. O. (2013). The widely used Wnt1-Cre transgene causes developmental phenotypes by ectopic activation of Wnt signaling. *Developmental biology*, 379(2), 229-234.

LI, C., LAN, Y., & JIANG, R. (2017). Molecular and cellular mechanisms of palate development. *Journal of dental research*, *96*(11), 1184-1191.

LIANG, J., VON DEN HOFF, J., LANGE, J., REN, Y., BIAN, Z. & CARELS, C. E. L. (2016). MSX1 mutations and associated disease phenotypes: genotype-phenotype relations. *European Journal of Human Genetics*, 24(12), 1663-1670.

LIM, J., & THIERY, J. P. (2012). Epithelial-mesenchymal transitions: insights from development. *Development*, *139*(19), 3471-3486.

LIN, J. S., JEON, J. S., FAN, Q., WONG, H. N., PALMER, M. B. & HOLZMAN, L. B. (2017). ARF6 mediates nephrin tyrosine phosphorylation-induced podocyte cellular dynamics. *PLoS One*, 12(9), e0184575.

LINES, M. A., HUANG, L., SCHWARTZENTRUBER, J., DOUGLAS, S. L., LYNCH, D. C., BEAULIEU, C., et al. GILLESSEN-KAESBACH, G. (2012). Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly. *The American Journal of Human Genetics*, 90(2), 369-377.

LIU, F., CHU, E. Y., WATT, B., ZHANG, Y., GALLANT, N. M., ANDL, T., et al. MILLAR, S. E. (2008). Wnt/beta-catenin signaling directs multiple stages of tooth morphogenesis. *Dev Biol*, 313(1), 210-24.

LIU, F. & MILLAR, S. E. (2010) Wnt/β-catenin signaling in oral tissue development and disease. *Journal of dental research*, 89(4), pp. 318-330.

LIU, H., BUSCH, T., ELIASON, S., ANAND, D., BULLARD, S., GOWANS, L. J. J., et al. BUTALI, A. (2017). Exome sequencing provides additional evidence for the involvement of ARHGAP29 in Mendelian orofacial clefting and extends the phenotypic spectrum to isolated cleft palate. *Birth Defects Res*, 109(1), 27-37.

LIU, K. J. (2016). Animal models of craniofacial anomalies. *Dev Biol*, 415(2), 169-170.

LIU, K. J., ARRON, J. R., STANKUNAS, K., CRABTREE, G. R. & LONGAKER, M. T. (2007). Chemical rescue of cleft palate and midline defects in conditional GSK-3β mice. *Nature*, 446(7131), 79-82.

LIU, W., SUN, X., BRAUT, A., MISHINA, Y., BEHRINGER, R. R., MINA, M., & MARTIN, J. F. (2005). Distinct functions for Bmp signaling in lip and palate fusion in mice. *Development*, *132*(6), 1453-1461.

LIU, Y. P., XU, L. F., WANG, Q., ZHOU, X. L., ZHOU, J. L., PAN, C., et al. XU, L. C. (2015). Identification of susceptibility genes in non-syndromic cleft lip with or without cleft palate using whole-exome sequencing. *Med Oral Patol Oral Cir Bucal*, 20(6), e763-70.

LOHI, M., TUCKER, A. S. & SHARPE, P. T. (2010). Expression of Axin2 indicates a role for canonical Wnt signaling in development of the crown and root during pre-and postnatal tooth development. *Developmental Dynamics*, 239(1), 160-167.

LOIKKANEN, I., HAGHIGHI, S., VAINIO, S. & PAJUNEN, A. (2002). Expression of cytosolic acetyl-CoA synthetase gene is developmentally regulated. *Mechanisms of Development*, 115(1), 139-141.

LOPES, V. L. G. d. S., GUION-ALMEIDA, M. L. & RODINI, E. S. d. O. (2003). Blepharocheilodontic (BCD) syndrome: expanding the phenotype? *American Journal of Medical Genetics Part A*, 121(3), 266-270.

LOUGH, K. J., BYRD, K. M., SPITZER, D. C. & WILLIAMS, S. E. (2017). Closing the Gap: Mouse Models to Study Adhesion in Secondary Palatogenesis. *J Dent Res*, 96(11), 1210-1220.

LU, Q., AGUILAR, B. J., LI, M., JIANG, Y. & CHEN, Y.-H. (2016). Genetic alterations of δ-catenin/NPRAP/Neurojungin (CTNND2): functional implications in complex human diseases. *Human genetics*, 135(10), 1107-1116.

LUCARIELLO, M., VIDAL, E., VIDAL, S., SAEZ, M., ROA, L., HUERTAS, D., et al. JURADO, P. (2016). Whole exome sequencing of Rett syndrome-like patients reveals the mutational diversity of the clinical phenotype. *Human genetics*, 135(12), 1343-1354.

LUDWIG, K. U., BÖHMER, A. C., RUBINI, M., MOSSEY, P. A., HERMS, S., NOWAK, S., et al. MANGOLD, E. (2014). Strong Association of Variants around FOXE1 and Orofacial Clefting. *Journal of Dental Research*, 93(4), 376-381.

LUDWIG, K. U., MANGOLD, E., HERMS, S., NOWAK, S., REUTTER, H., PAUL, A., et al. NOTHEN, M. M. (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nat Genet*, 44(9), 968-71.

LUNGOVA, V., VERHEYDEN, J. M., SUN, X. & THIBEAULT, S. L. (2018). β-Catenin signaling is essential for mammalian larynx recanalization and the establishment of vocal fold progenitor cells. *Development*, 145(4), dev157677.

LUQUETTI, D. V., HING, A. V., RIEDER, M. J., NICKERSON, D. A., TURNER, E. H., SMITH, J., et al. CUNNINGHAM, M. L. (2013). "Mandibulofacial dysostosis with microcephaly" caused by EFTUD2 mutations: expanding the phenotype. *American Journal of Medical Genetics Part A*, 161(1), 108-113.

MA, L. & CHUNG, W. K. (2014). Quantitative analysis of copy number variants based on real-time LightCycler PCR. *Current protocols in human genetics*, 80(1), 7.21. 1-7.21. 8.

MACARTHUR, D. G., MANOLIO, T. A., DIMMOCK, D. P., REHM, H. L., SHENDURE, J., ABECASIS, G. R., et al. GUNTER, C. (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature*, 508, 469.

MACE, S., COUSIN, E., RICARD, S., GENIN, E., SPANAKIS, E., LAFARGUE-SOUBIGOU, C., et al. DELEUZE, J. F. (2005). ABCA2 is a strong genetic risk factor for early-onset Alzheimer's disease. *Neurobiol Dis*, 18(1), 119-25.

MACHON, O., MASEK, J., MACHONOVA, O., KRAUSS, S., & KOZMIK, Z. (2015). Meis2 is essential for cranial and cardiac neural crest development. *BMC developmental biology*, *15*(1), 40.

MADDIREVULA, S., ALSAHLI, S., ALHABEEB, L., PATEL, N., ALZAHRANI, F., SHAMSELDIN, H. E., et al. ALKURAYA, F. S. (2018). Expanding the phenome and variome of skeletal dysplasia. *Genet Med*, 20(12), 1609-1616.

MADDIREVULA, S., ALZAHRANI, F., AL-OWAIN, M., AL MUHAIZEA, M. A., KAYYALI, H. R., ALHASHEM, A., et al. ALKURAYA, F. S. (2019). Autozygome and high throughput confirmation of disease genes candidacy. *Genetics in Medicine*, 21(3), 736-742.

MAITI, S., KUMAR, K. H., CASTELLANI, C. A., O'REILLY, R. & SINGH, S. M. (2011). Ontogenetic de novo copy number variations (CNVs) as a source of genetic individuality: studies on two families with MZD twins for schizophrenia. *PLoS One*, 6(3), e17125.

MANGOLD, E., LUDWIG, K. U., BIRNBAUM, S., BALUARDO, C., FERRIAN, M., HERMS, S., et al. MATTHEISEN, M. (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nature genetics*, 42(1), 24-26.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., et al. VISSCHER, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747-53.

MANSILLA, M. A., COOPER, M. E., GOLDSTEIN, T., CASTILLA, E. E., LOPEZ CAMELO, J. S., MARAZITA, M. L. & MURRAY, J. C. (2006). Contributions of PTCH gene variants to isolated cleft lip and palate. *Cleft Palate Craniofac J*, 43(1), 21-9.

MAQUAT, L. E. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol*, 5(2), 89-99.

MARAZITA, M. L. (2012). The evolution of human genetic studies of cleft lip and cleft palate. *Annual review of genomics and human genetics*, 13.

MARAZITA, M. L., LIDRAL, A. C., MURRAY, J. C., FIELD, L. L., MAHER, B. S., GOLDSTEIN MCHENRY, T., et al. ARCOS-BURGOS, M. (2009). Genome scan, fine-mapping, and candidate gene analysis of non-syndromic cleft lip with or without cleft palate reveals phenotype-specific differences in linkage and association results. *Hum Hered*, 68(3), 151-70.

MARAZITA, M. L., MURRAY, J. C., LIDRAL, A. C., ARCOS-BURGOS, M., COOPER, M. E., GOLDSTEIN, T., et al. RODDICK, L. G. (2004). Meta-Analysis of 13 Genome Scans Reveals Multiple Cleft Lip/Palate Genes with Novel Loci on 9q21 and 2q32-35. *The American Journal of Human Genetics*, 75(2), 161-173.

MARCIANO, D. K., BRAKEMAN, P. R., LEE, C.-Z., SPIVAK, N., EASTBURN, D. J., BRYANT, D. M., et al. REICHARDT, L. F. (2011). p120 catenin is required for normal renal tubulogenesis and glomerulogenesis. *Development*, 138(10), 2099-2109.

MARIAN, A. J. (2012). Molecular genetic studies of complex phenotypes. *Transl Res*, 159(2), 64-79.

MARINER, D. J., DAVIS, M. A. & REYNOLDS, A. B. (2004). EGFR signaling to p120-catenin through phosphorylation at Y228. *Journal of cell science*, 117(8), 1339-1350.

MARINER, D. J., WANG, J. & REYNOLDS, A. B. (2000). ARVCF localizes to the nucleus and adherens junction and is mutually exclusive with p120 (ctn) in E-cadherin complexes. *J Cell Sci*, 113(8), 1481-1490.

MARTINEZ-ALVAREZ, C., TUDELA, C., PEREZ-MIGUELSANZ, J., O'KANE, S., PUERTA, J. & FERGUSON, M. W. (2000). Medial edge epithelial cell fate during palatal fusion. *Dev Biol*, 220(2), 343-57.

MARTINEZ, F., CARO-LLOPIS, A., ROSELLO, M., OLTRA, S., MAYO, S., MONFORT, S. & ORELLANA, C. (2017). High diagnostic yield of syndromic intellectual disability by targeted next-generation sequencing. *J Med Genet*, 54(2), 87-92.

MAYOR, R. & THEVENEAU, E. (2013). The neural crest. *Development*, 140(11), 2247-51.

MAYOR, S. (2005). Royal Society predicts that personalised medicines are still decades away. *BMJ (Clinical research ed.)*, 331(7518), 656-656.

MCARDLE, H. J., GAMBLING, L. & KENNEDY, C. (2013). Iron deficiency during pregnancy: the consequences for placental function and fetal outcome. *Proceedings of the Nutrition Society*, 73(1), 9-15.

MCBRIDE, K. L., RILEY, M. F., ZENDER, G. A., FITZGERALD-BUTT, S. M., TOWBIN, J. A., BELMONT, J. W. & COLE, S. E. (2008). NOTCH1 mutations in individuals with left ventricular outflow tract malformations reduce ligand-induced signaling. *Hum Mol Genet*, 17(18), 2886-93.

MCBRIDE, W. A., MCINTYRE, G. T., CARROLL, K. & MOSSEY, P. A. (2016). Subphenotyping and Classification of Orofacial Clefts: Need for Orofacial Cleft Subphenotyping Calls for Revised Classification. *Cleft Palate Craniofac J*, 53(5), 539-49.

MCBRIDE, W. A., MOSSEY, P. A. & MCINTYRE, G. T. (2013). Reliability, completeness and accuracy of cleft subphenotyping as recorded on the CLEFTSiS (Cleft Service in Scotland) electronic patient record. *Surgeon*, 11(6), 313-8.

MCCLELLAN, J. & KING, M.-C. (2010). Genetic Heterogeneity in Human Disease. *Cell*, 141(2), 210-217.

MCDONALD, A. T., CROCKER, J. F., DIGOUT, S. C., MCCARTHY, S. C., BLECHER, S. R., & COLE, D. E. (1990). Glucocorticoid-induced polycystic kidney disease—A threshold trait. *Kidney international*, *37*(3), 901-908.

MCINTYRE, G. T. & MOSSEY, P. A. (2002). The Craniofacial Morphology of the Parents of Children with Orofacial Clefting: A Systematic Review of Cephalometric Studies. *Journal of Orthodontics*, 29(1), 23-29.

MCNALLY, E. M. & GEORGE, A. L. (2015). New approaches to establish genetic causality. *Trends in Cardiovascular Medicine*, 25(7), 646-652.

MCNAMARA, H. C., KANE, S. C., CRAIG, J. M., SHORT, R. V. & UMSTAD, M. P. (2016). A review of the mechanisms and evidence for typical and atypical twinning. *American journal of obstetrics and gynecology*, 214(2), 172-191.

MCPHERRON, A. C., LAWLER, A. M. & LEE, S. J. (1999). Regulation of anterior/posterior patterning of the axial skeleton by growth/differentiation factor 11. *Nat Genet*, 22(3), 260-4.

MEDINA, M., MARINESCU, R. C., OVERHAUSER, J. & KOSIK, K. S. (2000). Hemizygosity of δ-catenin (CTNND2) is associated with severe mental retardation in cri-du-chat syndrome. *Genomics*, 63(2), 157-164.

MENDELL, N. R., SPENCE, M. A., GLADSTIEN, K., BRUNETTE, J., STEVENS, A., CLIFFORD, E., et al. QUINN, G. W. (1980). Multifactorial/threshold models and their application to cleft lip and cleft palate. *Progress in clinical and biological research*, *46*, 387-406.

MENEZES, R. & VIEIRA, A. R. (2008). Dental anomalies as part of the cleft spectrum. *The Cleft Palate-Craniofacial Journal*, 45(4), 414-419.

MILERAD, J., LARSON, O., HAGBERG, C. & IDEBERG, M. (1997). Associated malformations in infants with cleft lip and palate: a prospective, population-based study. *Pediatrics*, 100(2 Pt 1), 180-6.

MILETICH, I. & SHARPE, P. T. (2004). Neural crest contribution to mammalian tooth formation. *Birth Defects Research Part C: Embryo Today: Reviews*, 72(2), 200-212.

MILLER, J. R. & MCCLAY, D. R. (1997). Characterization of the role of cadherin in regulating cell adhesion during sea urchin development. *Dev Biol*, 192(2), 323-39.

MITSIADIS, T. A., REGAUDIAT, L. & GRIDLEY, T. (2005). Role of the Notch signalling pathway in tooth morphogenesis. *Arch Oral Biol*, 50(2), 137-40.

MIYAMOTO, R. C., COTTON, R. T., ROPE, A. F., HOPKIN, R. J., COHEN, A. P., SHOTT, S. R. & RUTTER, M. J. (2004). Association of anterior glottic webs with velocardiofacial syndrome (chromosome 22q11. 2 deletion). *Otolaryngology—Head and Neck Surgery*, 130(4), 415-417.

MONTONEN, O., AHO, M., UITTO, J. & AHO, S. (2001). Tissue distribution and cell type-specific expression of p120ctn isoforms. *Journal of Histochemistry & Cytochemistry*, 49(12), 1487-1495.

MOORE, G. E., IVENS, A., CHAMBERS, J., FARRALL, M., WILLIAMSON, R., PAGE, D. C., et al. JENSSON, O. (1987). Linkage of an X-chromosome cleft palate gene. *Nature*, 326(6108), 91.

MOREAU, J. L. M., KESTEVEN, S., MARTIN, E., LAU, K. S., YAM, M. X., O'REILLY, V. C., et al. DUNWOODIE, S. L. (2019). Gene-environment interaction impacts on heart development and embryo survival. *Development*, 146(4).

MORENO, L. M., MANSILLA, M. A., BULLARD, S. A., COOPER, M. E., BUSCH, T. D., MACHIDA, J., et al. LIDRAL, A. C. (2009). FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. *Human Molecular Genetics*, 18(24), 4879-4896.

MORI, C., NAKAMURA, N., OKAMOTO, Y., OSAWA, M., & SHIOTA, K. (1994). Cytochemical identification of programmed cell death in the fusing fetal mouse palate by specific labelling of DNA fragmentation. *Anatomy and embryology*, *190*(1), 21-28.

MORINI, F., COZZI, D. A., ILARI, M., CASATI, A. & COZZI, F. (2001). Pattern of Cardiovascular Anomalies Associated with Esophageal Atresia: Support for a Caudal Pharyngeal Arch Neurocristopathy. *Pediatric Research*, 50(5), 565-568.

MOSESSON, Y., MILLS, G. B. & YARDEN, Y. (2008). Derailed endocytosis: an emerging feature of cancer. *Nature Reviews Cancer*, 8, 835.

MOSS, M. L. (1968). A theoretical analysis of the functional matrix. *Acta Biotheoretica*, 18(1), 195-202.

MOSS, M. L. (1997). The functional matrix hypothesis revisited. 1. The role of mechanotransduction. *American journal of orthodontics and dentofacial orthopedics*, 112(1), 8-11.

MOSSEY, P. (2003). Global strategies to reduce the healthcare burden of craniofacial anomalies. *British dental journal*, *195*(10), 613-613.

MOSSEY, P. A., BATRA, P. & MCINTYRE, G. T. (2010). The parental dentocraniofacial phenotype-an orofacial clefting microform. *The Cleft Palate-Craniofacial Journal*, 47(1), 22-34.

MOSSEY, P. A., & CASTILLA, E. E. (2003). Global registry and database on craniofacial anomalies: Report of a WHO Registry Meeting on Craniofacial Anomalies.

MOSSEY, P. A., LITTLE, J., MUNGER, R. G., DIXON, M. J. & SHAW, W. C. (2009). Cleft lip and palate. *The Lancet*, 374(9703), 1773-1785.

MOSSEY, P. A., LITTLE, J., STEEGERS-THEUNISSEN, R., MOLLOY, A., PETERLIN, B., SHAW, W. C., et al. RUBINI, M. (2017). Genetic Interactions in Nonsyndromic Orofacial Clefts in Europe—EUROCRAN Study. *The Cleft Palate-Craniofacial Journal*, 54(6), 623-630.

MOSSEY, P. A. & MODELL, B. (2012). Epidemiology of oral clefts 2012: an international perspective. *Front Oral Biol*, 16, 1-18.

MOSTOWSKA, A., BIEDZIAK, B. & JAGODZINSKI, P. P. (2006). Axis inhibition protein 2 (AXIN2) polymorphisms may be a risk factor for selective tooth agenesis. *Journal of human genetics*, 51(3), 262.

MOSTOWSKA, A., BIEDZIAK, B., ZADURSKA, M., DUNIN-WILCZYNSKA, I., LIANERI, M. & JAGODZINSKI, P. (2013). Nucleotide variants of genes encoding components of the Wnt signalling pathway and the risk of non-syndromic tooth agenesis. *Clinical genetics*, 84(5), 429-440.

MUENKE, M., GURRIERI, F., BAY, C., YI, D. H., COLLINS, A. L., JOHNSON, V. P., et al. LUBINSKY, M. S. (1994a). Linkage of a human brain malformation, familial holoprosencephaly, to chromosome 7 and evidence for genetic heterogeneity. *Proceedings of the National Academy of Sciences*, 91(17), 8102-8106.

MUENKE, M., SCHELL, U., HEHR, A., ROBIN, N. H., LOSKEN, H. W., SCHINZEL, A., et al. ET AL. (1994b). A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome. *Nat Genet*, 8(3), 269-74.

MUES, G., BONDS, J., XIANG, L., VIEIRA, A. R., SEYMEN, F., KLEIN, O. & D'SOUZA, R. N. (2014). The WNT10A gene in ectodermal dysplasias and selective tooth agenesis. *American Journal of Medical Genetics Part A*, 164(10), 2455-2460.

MUKHAMEDOVA, N., HOANG, A., CUI, H. L., CARMICHAEL, I., FU, Y., BUKRINSKY, M. & SVIRIDOV, D. (2016). Small GTPase ARF6 Regulates Endocytic Pathway Leading to Degradation of ATP-Binding Cassette Transporter A1. *Arterioscler Thromb Vasc Biol*, 36(12), 2292-2303.

MUNGER, R. G., TAMURA, T., JOHNSTON, K. E., FELDKAMP, M. L., PFISTER, R., CUTLER, R., et al. CAREY, J. C. (2011). Oral clefts and maternal biomarkers of folate-dependent one-carbon metabolism in Utah. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *91*(3), 153-161.

MURDOCH, J. N., DAMRAU, C., PAUDYAL, A., BOGANI, D., WELLS, S., GREENE, N. D., et al. COPP, A. J. (2014). Genetic interactions between planar cell polarity genes cause diverse neural tube defects in mice. *Dis Model Mech*, 7(10), 1153-63.

MURRAY, J., NISHIMURA, D., BUETOW, K., ARDINGER, H., SPENCE, M., SPARKES, R., et al. HARKNESS, E. (1990). Linkage of an autosomal dominant clefting syndrome (Van der Woude) to loci on chromosome Iq. *American journal of human genetics*, 46(3), 486.

MURRAY, J. C. (1995). Face facts: genes, environment, and clefts. *American journal of human genetics*, 57(2), 227-232.

MURRAY, S. A., ORAM, K. F., & GRIDLEY, T. (2007). Multiple functions of Snail family genes during palate development in mice. *Development*, 134(9), 1789-1797.

NAKAJIMA, A., F SHULER, C., GULKA, A. O., & HANAI, J. I. (2018). TGF-β signaling and the epithelial-mesenchymal transition during palatal fusion. *International journal of molecular sciences*, *19*(11), 3638.

NAWSHAD, A. (2008). Palatal seam disintegration: to die or not to die? that is no longer the question. *Dev Dyn*, 237(10), 2643-56.

NAWSHAD, A., LAGAMBA, D. & HAY, E. (2004). Transforming growth factor β (TGFβ) signalling in palatal growth, apoptosis and epithelial mesenchymal transformation (EMT). *Archives of oral biology*, 49(9), 675-689.

NAWSHAD, A., MEDICI, D., LIU, C.-C. & HAY, E. D. (2007). TGFβ3 inhibits E-cadherin gene expression in palate medial-edge epithelial cells through a Smad2-Smad4-LEF1 transcription complex. *Journal of cell science*, 120(9), 1646-1653.

NEISWANGER, K., WEINBERG, S. M., ROGERS, C. R., BRANDON, C. A., COOPER, M. E., BARDI, K. M., et al. MOONEY, M. P. (2007). Orbicularis oris muscle defects as an expanded phenotypic feature in nonsyndromic cleft lip with or without cleft palate. *American Journal of Medical Genetics Part A*, 143(11), 1143-1149.

NESS, A. R., WILLS, A. K., WAYLEN, A., AL-GHATAM, R., JONES, T. E. M., PRESTON, R., et al. SANDY, J. R. (2015). Centralization of cleft care in the UK. Part 6: a tale of two studies. *Orthodontics & Craniofacial Research*, 18(S2), 56-62.

NG, S. B., BIGHAM, A. W., BUCKINGHAM, K. J., HANNIBAL, M. C., MCMILLIN, M. J., GILDERSLEEVE, H. I., et al. SHENDURE, J. (2010a). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet*, 42(9), 790-3.

NG, S. B., BUCKINGHAM, K. J., LEE, C., BIGHAM, A. W., TABOR, H. K., DENT, K. M., et al. BAMSHAD, M. J. (2010b). Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*, 42(1), 30-5.

NIE, Z., FEI, J., PREMONT, R. T. & RANDAZZO, P. A. (2005). The Arf GAPs AGAP1 and AGAP2 distinguish between the adaptor protein complexes AP-1 and AP-3. *J Cell Sci*, 118(Pt 15), 3555-66.

NIE, Z., STANLEY, K. T., STAUFFER, S., JACQUES, K. M., HIRSCH, D. S., TAKEI, J. & RANDAZZO, P. A. (2002). AGAP1, an endosome-associated, phosphoinositide-dependent ADP-ribosylation factor GTPase-activating protein that affects actin cytoskeleton. *J Biol Chem*, 277(50), 48965-75.

NISHI, E., MASUDA, K., ARAKAWA, M., KAWAME, H., KOSHO, T., KITAHARA, M., et al. IZUMI, K. (2016). Exome sequencing-based identification of mutations in non-syndromic genes among individuals with apparently syndromic features. *Am J Med Genet A*, 170(11), 2889-2894.

NIVARD, M., MBAREK, H., HOTTENGA, J., SMIT, J., JANSEN, R., PENNINX, B., et al. BOOMSMA, D. (2014). Further confirmation of the association between anxiety and CTNND2: replication in humans. *Genes, Brain and Behavior*, 13(2), 195-201.

O'REILLY, V. C., LOPES FLORO, K., SHI, H., CHAPMAN, B. E., PREIS, J. I., JAMES, A. C., et al. DUNWOODIE, S. L. (2014). Gene-environment interaction demonstrates the vulnerability of the embryonic heart. *Dev Biol*, 391(1), 99-110.

OAS, R. G., XIAO, K., SUMMERS, S., WITTICH, K. B., CHIASSON, C. M., MARTIN, W. D., et al. KOWALCZYK, A. P. (2010). p120-Catenin is required for mouse vascular development. *Circulation research*, 106(5), 941.

OH, W.-J., WESTMORELAND, J. J., SUMMERS, R. & CONDIE, B. G. (2010). Cleft palate is caused by CNS dysfunction in Gad1 and Viaat knockout mice. *PLoS One*, 5(3), e9758.

OIKE, Y., HATA, A., MAMIYA, T., KANAME, T., NODA, Y., SUZUKI, M., et al. YAMAMURA, K.-i. (1999). Truncated CBP Protein Leads to Classical Rubinstein—Taybi Syndrome Phenotypes in Mice: Implications for a Dominant-Negative Mechanism. *Human Molecular Genetics*, 8(3), 387-396.

ONO, S., IMAMURA, A., TASAKI, S., KUROTAKI, N., OZAWA, H., YOSHIURA, K. & OKAZAKI, Y. (2010). Failure to confirm CNVs as of aetiological significance in twin pairs discordant for schizophrenia. *Twin Res Hum Genet*, 13(5), 455-60.

OSOEGAWA, K., VESSERE, G. M., UTAMI, K. H., MANSILLA, M. A., JOHNSON, M. K., RILEY, B. M., et al. VAN DER VLIET, W. A. (2008). Identification of novel candidate genes associated with cleft lip and palate using array comparative genomic hybridisation. *Journal of medical genetics*, 45(2), 81-86.

PAN, C., NELSON, M. S., REYES, M., KOODIE, L., BRAZIL, J. J., STEPHENSON, E. J., et al. GUPTA, P. (2005). Functional abnormalities of heparan sulfate in mucopolysaccharidosis-I are associated with defective biologic activity of FGF-2 on human multipotent progenitor cells. *Blood*, 106(6), 1956-64.

PANETTA, N. J., GUPTA, D. M., SLATER, B. J., KWAN, M. D., LIU, K. J. & LONGAKER, M. T. (2008). Tissue engineering in cleft palate and other congenital malformations. *Pediatr Res*, 63(5), 545-51.

PARK, J.-i., JI, H., JUN, S., GU, D., HIKASA, H., LI, L., et al. MCCREA, P. D. (2006). Frodo links Dishevelled to the p120-catenin/Kaiso pathway: distinct catenin subfamilies promote Wnt signals. *Developmental cell*, 11(5), 683-695.

PARK, J.-i., KIM, S. W., LYONS, J. P., JI, H., NGUYEN, T. T., CHO, K., et al. MCCREA, P. D. (2005). Kaiso/p120-catenin and TCF/β-catenin complexes coordinately regulate canonical Wnt gene targets. *Developmental cell*, 8(6), 843-854.

PASCHOS, E., HUTH, K. C. & HICKEL, R. (2003). Clinical management of hypohidrotic ectodermal dysplasia with anodontia: case report. *Journal of Clinical Pediatric Dentistry*, 27(1), 5-8.

PAULSON, A. F., FANG, X., JI, H., REYNOLDS, A. B. & MCCREA, P. D. (1999). Misexpression of the Catenin p120ctn1A PerturbsXenopusGastrulation But Does Not Elicit Wnt-Directed Axis Specification. *Developmental biology*, 207(2), 350-363.

PELTON, R. W., HOGAN, B. L., MILLER, D. A., & MOSES, H. L. (1990). Differential expression of genes encoding TGFs β1, β2, and β3 during murine palate formation. *Developmental biology*, *141*(2), 456-460.

PENGELLY, R. J., UPSTILL-GODDARD, R., ARIAS, L., MARTINEZ, J., GIBSON, J., KNUT, M., et al. BRICENO, I. (2015). Resolving clinical diagnoses for syndromic cleft lip and/or palate phenotypes using whole-exome sequencing. *Clin Genet*, 88(5), 441-9.

PEREZ-MORENO, M., DAVIS, M. A., WONG, E., PASOLLI, H. A., REYNOLDS, A. B. & FUCHS, E. (2006). p120-catenin mediates inflammatory responses in the skin. *Cell*, 124(3), 631-644.

PERSON, A. D., KLEWER, S. E. & RUNYAN, R. B. (2005). Cell biology of cardiac cushion development. *Int Rev Cytol*, 243, 287-335.

PETERS, H., NEUBÜSER, A., KRATOCHWIL, K. & BALLING, R. (1998). Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes & development*, 12(17), 2735-2747.

PETERSON, J. C., CHUGHTAI, M., WISSE, L. J., GITTENBERGER-DE GROOT, A. C., FENG, Q., GOUMANS, M.-J. T., et al. DERUITER, M. C. (2018). Nos3 mutation leads to abnormal neural crest cell and second heart field lineage patterning in bicuspid aortic valve formation. *Disease Models & Mechanisms*, dmm. 034637.

PEYRARD-JANVID, M., PEGELOW, M., KOILLINEN, H., LARSSON, C., FRANSSON, I., RAUTIO, J., et al. KERE, J. (2005). Novel and de novo mutations of the IRF6 gene detected in patients with Van der Woude or popliteal pterygium syndrome. *Eur J Hum Genet*, 13(12), 1261-7.

PIETERS, T., VAN ROY, F. & VAN HENGEL, J. (2012). Functions of p120ctn isoforms in cell-cell adhesion and intracellular signaling. *Front Biosci (Landmark Ed)*, 17, 1669-94.

POWIS, Z., HART, A., CHERNY, S., PETRIK, I., PALMAER, E., TANG, S. & JONES, C. (2017). Clinical diagnostic exome evaluation for an infant with a lethal disorder: genetic diagnosis of TARP syndrome and expansion of the phenotype in a patient with a newly reported RBM10 alteration. *BMC Med Genet*, 18(1), 60.

PRATT, R. M., DENCKER, L. & DIEWERT, V. M. (1984). 2, 3, 7, 8-Tetrachlorodibenzo-p-dioxin-induced cleft palate in the mouse: Evidence for alterations in palatal shelf fusion. *Teratogenesis, carcinogenesis, and mutagenesis*, 4(5), 427-436.

PRATTICHIZZO, C., MACCA, M., NOVELLI, V., GIORGIO, G., BARRA, A. & FRANCO, B. (2008). Mutational spectrum of the oral-facial-digital type I syndrome: a study on a large collection of patients. *Hum Mutat*, 29(10), 1237-46.

QIAO, X., LIU, Y., LI, P., CHEN, Z., LI, H., YANG, X., et al. QIAO, B. (2016). Genetic analysis of rare coding mutations of CELSR1–3 in congenital heart and neural tube defects in Chinese people. *Clinical Science*, 130(24), 2329-2340.

RABBANI, B., MAHDIEH, N., HOSOMICHI, K., NAKAOKA, H. & INOUE, I. (2012). Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet*, 57(10), 621-32.

RANDAZZO, P. A. & HIRSCH, D. S. (2004). Arf GAPs: multifunctional proteins that regulate membrane traffic and actin remodelling. *Cellular signalling*, 16(4), 401-413.

RANDAZZO, P. A., INOUE, H. & BHARTI, S. (2007). Arf GAPs as regulators of the actin cytoskeleton. *Biology of the Cell*, 99(10), 583-600.

REYNOLDS, A. B., DANIEL, J., MCCREA, P. D., WHEELOCK, M. J., WU, J. & ZHANG, Z. (1994). Identification of a new catenin: the tyrosine kinase substrate p120cas associates with E-cadherin complexes. *Molecular and cellular biology*, 14(12), 8333-8342.

REYNOLDS, A. B., JENKINS, N. A., GILBERT, D. J., COPELAND, N. G., SHAPIRO, D. N., WU, J. & DANIEL, J. M. (1996). The gene encoding p120cas, a novel catenin, localizes on human chromosome 11q11 (CTNND) and mouse chromosome 2 (Catns). *Genomics*, 31(1), 127-129.

REYNOLDS, A. B. & ROCZNIAK-FERGUSON, A. (2004). Emerging roles for p120-catenin in cell adhesion and cancer. *Oncogene*, 23(48), 7947.

RICE, R., SPENCER-DENE, B., CONNOR, E. C., GRITLI-LINDE, A., MCMAHON, A. P., DICKSON, C., et al. RICE, D. P. (2004). Disruption of Fgf10/Fgfr2b-coordinated epithelial-mesenchymal interactions causes cleft palate. *Journal of Clinical Investigation*, 113(12), 1692.

RICHARDS, S., AZIZ, N., BALE, S., BICK, D., DAS, S., GASTIER-FOSTER, J., et al. REHM, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics In Medicine*, 17, 405.

RICHARDSON, R. J., DIXON, J., JIANG, R. & DIXON, M. J. (2009). Integration of IRF6 and Jagged2 signalling is essential for controlling palatal adhesion and fusion competence. *Human Molecular Genetics*, 18(14), 2632-2642.

RICHARDSON, R. J., HAMMOND, N. L., COULOMBE, P. A., SALORANTA, C., NOUSIAINEN, H. O., SALONEN, R., et al. DIXON, M. J. (2014). Periderm prevents pathological epithelial adhesions during embryogenesis. *The Journal of clinical investigation*, 124(9), 3891-3900.

RICHTSMEIER, J. T., ZUMWALT, A., CARLSON, E. J., EPSTEIN, C. J. & REEVES, R. H. (2002). Craniofacial phenotypes in segmentally trisomic mouse models for Down syndrome. *Am J Med Genet*, 107(4), 317-24.

RICKS, J. E., RYDER, V. M., BRIDGEWATER, L. C., SCHAALJE, B., & SEEGMILLER, R. E. (2002). Altered mandibular development precedes the time of palate closure in mice homozygous for disproportionate micromelia: an oral clefting model supporting the Pierre-Robin sequence. *Teratology*, 65(3), 116-120.

RILEY, B. M., MANSILLA, M. A., MA, J., DAACK-HIRSCH, S., MAHER, B. S., RAFFENSPERGER, L. M., et al. MOHAMMADI, M. (2007). Impaired FGF signaling contributes to cleft lip and palate. *Proceedings of the National Academy of Sciences*, 104(11), 4512-4517.

RITTLER, M., COSENTINO, V., LOPEZ-CAMELO, J. S., MURRAY, J. C., WEHBY, G. & CASTILLA, E. E. (2011). Associated anomalies among infants with oral clefts at birth and during a 1-year follow-up. *Am J Med Genet A*, 155a(7), 1588-96.

ROBINSON, A., ESCUIN, S., DOUDNEY, K., VEKEMANS, M., STEVENSON, R. E., GREENE, N. D., et al. STANIER, P. (2012). Mutations in the planar cell polarity genes CELSR1 and SCRIB are associated with the severe neural tube defect craniorachischisis. *Human mutation*, 33(2), 440-447.

ROFF, D. A. (1996). The evolution of threshold traits in animals. THE QUARTERLY REVIEW OF BIOLOGY, 71(1), 3-35.

ROMANELLI, V., BELINCHON, A., BENITO-SANZ, S., MARTINEZ-GLEZ, V., GRACIA-BOUTHELIER, R., HEATH, K. E., et al. LAPUNZINA, P. (2010). CDKN1C (p57(Kip2)) analysis in Beckwith-Wiedemann syndrome (BWS) patients: Genotype-phenotype correlations, novel mutations, and polymorphisms. *Am J Med Genet A*, 152a(6), 1390-7.

RONEMUS, M., IOSSIFOV, I., LEVY, D. & WIGLER, M. (2014). The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet*, 15(2), 133-41.

RUAL, J. F., HIROZANE-KISHIKAWA, T., HAO, T., BERTIN, N., LI, S., DRICOT, A., et al. VIDAL, M. (2004). Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res*, 14(10b), 2128-35.

SABBAGH, H. J., HASSAN, M. H., INNES, N. P., BAIK, A. A. & MOSSEY, P. A. (2014). Parental consanguinity and nonsyndromic orofacial clefts in children: a systematic review and meta-analyses. *Cleft Palate Craniofac J*, 51(5), 501-13.

SAKAI, D., DIXON, J., ACHILLEOS, A., DIXON, M., & TRAINOR, P. A. (2016). Prevention of Treacher Collins syndrome craniofacial anomalies in mouse models via maternal antioxidant supplementation. *Nature communications*, *7*(1), 1-13.

SANI VAZIRI, F., HALLBERG, K., HARFE, B. D., MCMAHON, A. P., LINDE, A., & GRITLI-LINDE, A. (2005). Fate-mapping of the epithelial seam during palatal fusion rules out epithelial–mesenchymal transformation. *Developmental biology*, *285*(2), 490-495.

SATOKATA, I. & MAAS, R. (1994). Msx1 deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nature Genetics*, 6(4), 348-356.

SATTUR, M., PINES, A. R. & BENDOK, B. R. (2016). Thinking from the Heart: Neurocristopathy, Aortic Abnormalities, and Intracranial Aneurysms. *World Neurosurgery*, 85, 25-27.

SCARPA, E., SZABO, A., BIBONNE, A., THEVENEAU, E., PARSONS, M. & MAYOR, R. (2015). Cadherin Switch during EMT in Neural Crest Cells Leads to Contact Inhibition of Locomotion via Repolarization of Forces. *Dev Cell*, 34(4), 421-34.

SCHACKMANN, R. C., TENHAGEN, M., VAN DE VEN, R. A. & DERKSEN, P. W. (2013). p120-catenin in cancer–mechanisms, models and opportunities for intervention. *J Cell Sci*, 126(16), 3515-3525.

SCHIEVINK, W. I., MOKRI, B., PIEPGRAS, D. G. & GITTENBERGER-DE GROOT, A. C. (1996). Intracranial Aneurysms and Cervicocephalic Arterial Dissections Associated with Congenital Heart Disease. *Neurosurgery*, 39(4), 685-690.

SCHUBERT, J., JAHN, H., & BERGINSKI, M. (2005). Experimental aspects of the pathogenesis of Robin sequence. *The Cleft palate-craniofacial journal*, *42*(4), 372-376.

SCHUTTE, B. C. & MURRAY, J. C. (1999). The Many Faces and Factors of Orofacial Clefts. *Human Molecular Genetics*, 8(10), 1853-1859.

SEIDEL, B., BRAEG, S., ADLER, G., WEDLICH, D. & MENKE, A. (2004). E-and N-cadherin differ with respect to their associated p120 ctn isoforms and their ability to suppress invasive growth in pancreatic cancer cells. *Oncogene*, 23(32), 5532.

SELL, D., GRUNWELL, P., MILDINHALL, S., MURPHY, T., CORNISH, T. A., BEARN, D., et al. SANDY, J. R. (2001). Cleft lip and palate care in the United Kingdom—the Clinical

Standards Advisory Group (CSAG) Study. Part 3: speech outcomes. *The Cleft palate-craniofacial journal*, 38(1), 30-37.

SELL, D., MILDINHALL, S., ALBERY, L., WILLS, A. K., SANDY, J. R. & NESS, A. R. (2015). The Cleft Care UK study. Part 4: perceptual speech outcomes. *Orthodontics & Craniofacial Research*, 18(S2), 36-46.

SERAJEE, F. J. & HUQ, A. M. (2015). Homozygous mutation in synaptic vesicle glycoprotein 2A gene results in intractable epilepsy, involuntary movements, microcephaly, and developmental and growth retardation. *Pediatric neurology*, 52(6), 642-646. e1.

SETHI, N., YAN, Y., QUEK, D., SCHUPBACH, T. & KANG, Y. (2010). Rabconnectin-3 is a functional regulator of mammalian Notch signaling. *J Biol Chem*, 285(45), 34757-64.

SETO-SALVIA, N. & STANIER, P. (2014). Genetics of cleft lip and/or cleft palate: association with other common anomalies. *Eur J Med Genet*, 57(8), 381-93.

SHAPIRA, Y., LUBIT, E., KUFTINEC, M. M., & BORELL, G. (1999). The distribution of clefts of the primary and secondary palates by sex, type, and location. *The angle orthodontist*, 69(6), 523-528.

SHAPIRO, B. L., & SWENEY, L. (1969). Electron microscopic and histochemical examination of oral epithelial-mesenchymal interaction (programmed cell death). *Journal of dental research*, *48*(5), 652-660.

SHARPE, P. M., & FERGUSON, M. W. J. (1988). Mesenchymal influences on epithelial differentiation in developing systems. *Journal of cell science*, *1988*(Supplement 10), 195-230.

SHAW, W. (2004). Global strategies to reduce the health care burden of craniofacial anomalies: report of WHO meetings on international collaborative research on craniofacial anomalies. *The Cleft palate-craniofacial journal*, *41*(3), 238-243.

SHAWLOT, W., DENG, J. M., FOHN, L. E. & BEHRINGER, R. R. (1998). Restricted β-galactosidase expression of a hygromycin-lacZ gene targeted to the β-actin locus and embryonic lethality of β-actin mutant mice. *Transgenic research*, 7(2), 95-103.

SHENDURE, J. & AKEY, J. M. (2015). The origins, determinants, and consequences of human mutations. *Science*, 349(6255), 1478-1483.

SHI, H., O'REILLY, V. C., MOREAU, J. L. M., BEWES, T. R., YAM, M. X., CHAPMAN, B. E., et al. DUNWOODIE, S. L. (2016). Gestational stress induces the unfolded protein response, resulting in heart defects. *Development*, 143(14), 2561-2572.

SHI, M., MOSTOWSKA, A., JUGESSUR, A., JOHNSON, M. K., MANSILLA, M. A., CHRISTENSEN, K., et al. MURRAY, J. C. (2009). Identification of microdeletions in candidate genes for cleft lip and/or palate. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 85(1), 42-51.

SHIMIZU, T., ICHINOSAWA, T., KIGUCHI, Y., ISHIDA, F., & MAEDA, T. (2013). Lef1 may contribute to agenesis of the third molars in mice. *Open journal of stomatology*, *3*(05), 281.

SHPRINTZEN, R., GOLDBERG, R., LEWIN, M., SIDOTI, E., BERKMAN, M., ARGAMASO, R. & YOUNG, D. (1978). A new syndrome involving cleft palate, cardiac anomalies, typical facies, and learning disabilities: velo-cardio-facial syndrome. *The Cleft palate journal*, 15(1), 56-62.

SIFRIM, A., HITZ, M.-P., WILSDON, A., BRECKPOT, J., TURKI, S. H. A., THIENPONT, B., et al. HURLES, M. E. (2016). Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nature Genetics*, 48, 1060.

SILVERSIDES, C. K., LIONEL, A. C., COSTAIN, G., MERICO, D., MIGITA, O., LIU, B., et al. MARSHALL, C. R. (2012). Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways. *PLoS genetics*, 8(8), e1002843.

SIMOES-COSTA, M. & BRONNER, M. E. (2015). Establishing neural crest identity: a gene regulatory recipe. *Development*, 142(2), 242-57.

SIMS, D., SUDBERY, I., ILOTT, N. E., HEGER, A. & PONTING, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet*, 15(2), 121-32.

SINGH, B., SRIVASTAVA, P. & PHADKE, S. R. (2019). Sequence variations in TENM3 gene causing eye anomalies with intellectual disability: Expanding the phenotypic spectrum. *Eur J Med Genet*, 62(1), 61-64.

SIRMACI, A., WALSH, T., AKAY, H., SPILIOPOULOS, M., SAKALAR, Y. B., HASANEFENDIOGLU-BAYRAK, A., et al. TEKIN, M. (2010). MASP1 mutations in patients with facial, umbilical, coccygeal, and auditory findings of Carnevale, Malpuech, OSA, and Michels syndromes. *Am J Hum Genet*, 87(5), 679-86.

SIROTKIN, H., O'DONNELL, H., DASGUPTA, R., HALFORD, S., ST JORE, B., PUECH, A., et al. KUCHERLAPATI, R. (1997). Identification of a new human catenin gene family member (ARVCF) from the region deleted in velo-cardio-facial syndrome. *Genomics*, 41(1), 75-83.

SIVERTSEN, Å., WILCOX, A. J., SKJÆRVEN, R., VINDENES, H. A., ÅBYHOLM, F., HARVILLE, E., & LIE, R. T. (2008). Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives. *British medical journal*, 336(7641), 432-434.

SLAYTON, R. L., WILLIAMS, L., MURRAY, J. C., WHEELER, J. J., LIDRAL, A. C. & NISHIMURA, C. J. (2003). Genetic association studies of cleft lip and/or palate with hypodontia outside the cleft region. *The Cleft palate-craniofacial journal*, 40(3), 274-279.

SMALLEY-FREED, W. G., EFIMOV, A., BURNETT, P. E., SHORT, S. P., DAVIS, M. A., GUMUCIO, D. L., et al. REYNOLDS, A. B. (2010). p120-catenin is essential for maintenance of barrier function and intestinal homeostasis in mice. *The Journal of clinical investigation*, 120(6), 1824-1835.

SMALLRIDGE, J., WILLS, A. K., MAHMOUD, O., CHONG, A., CLARK, V., COLLARD, M., et al. NESS, A. R. (2017). Centre-level variation in dental treatment and oral health and individual- and area-level predictors of oral health in 5-year-old children with non-syndromic unilateral cleft lip and palate: the Cleft Care UK study. Part 3. *Orthod Craniofac Res*, 20 Suppl 2, 19-26.

SMIGIEL, R., BEZNIAKOW, N., JAKUBIAK, A., BŁOCH, M., PATKOWSKI, D., OBERSZTYN, E. & SASIADEK, M. M. (2015). Phenotype analysis of Polish patients with mandibulofacial dysostosis type Guion-Almeida associated with esophageal atresia and choanal atresia caused by EFTUD2 gene mutations. *Journal of applied genetics*, 56(2), 199-204.

SMILEY, G. R., & DIXON, A. D. (1968). Fine structure of midline epithelium in the developing palate of the mouse. *The anatomical record*, 161(3), 293-310.

SONG, S., ZHAO, R., HE, H., ZHANG, J., FENG, H. & LIN, L. (2014). WNT10A variants are associated with non-syndromic tooth agenesis in the general population. *Human genetics*, 133(1), 117-124.

SORIANO, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature genetics*, 21(1), 70-71.

SPARROW, D. B., CHAPMAN, G., SMITH, A. J., MATTAR, M. Z., MAJOR, J. A., O'REILLY, V. C., et al. DUNWOODIE, S. L. (2012). A mechanism for gene-environment interaction in the etiology of congenital scoliosis. *Cell*, 149(2), 295-306.

SPERBER, G. H., SPERBER, G. H., GUTTMANN, G. D., SPERBER, S. M., & GUTTERMAN, G. D. (2001). *Craniofacial development (book for windows & macintosh)*. PMPH-USA.

SPRANGER, J. a., BENIRSCHKE, K., HALL, J., LENZ, W., LOWRY, R. B., OPITZ, J., et al. SMITH, D. (1982). Errors of morphogenesis: concepts and terms: recommendations of an international working group. *The Journal of pediatrics*, 100(1), 160-165.

SRIVASTAVA, D., THOMAS, T., LIN, Q., KIRBY, M. L., BROWN, D. & OLSON, E. N. (1997). Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nature genetics*, 16(2), 154.

STAIRS, D. B., BAYNE, L. J., RHOADES, B., VEGA, M. E., WALDRON, T. J., KALABIS, J., et al. RUSTGI, A. K. (2011). Deletion of p120-catenin results in a tumor microenvironment with inflammation and cancer that establishes it as a tumor suppressor gene. *Cancer Cell*, 19(4), 470-83.

STEINBERG, K. M., YU, B., KOBOLDT, D. C., MARDIS, E. R. & PAMPHLETT, R. (2015). Exome sequencing of case-unaffected-parents trios reveals recessive and de novo genetic variants in sporadic ALS. *Scientific Reports*, 5, 9124.

STEPHEN, J., NAMPOOTHIRI, S., KUPPA, S., YESODHARAN, D., RADHAKRISHNAN, N., GAHL, W. A. & MALICDAN, M. C. V. (2018). Novel truncating mutation in TENM3 in siblings with motor developmental delay, ocular coloboma, oval cornea, without microphthalmia. *Am J Med Genet A*, 176(12), 2930-2933.

STOLL, C., ALEMBIK, Y., DOTT, B. & ROTH, M. P. (2000). Associated malformations in cases with oral clefts. *Cleft Palate Craniofac J*, 37(1), 41-7.

STRANGER, B. E., STAHL, E. A. & RAJ, T. (2011). Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*, 187(2), 367-83.

STUDY, D. D. D., MCRAE, J. F., CLAYTON, S., FITZGERALD, T. W., KAPLANIS, J., PRIGMORE, E., et al. AKAWI, N. (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature*, 542(7642), 433.

SUN, D., MCALMON, K. R., DAVIES, J. A., BERNFIELD, M. & HAY, E. D. (1998). Simultaneous loss of expression of syndecan-1 and E-cadherin in the embryonic palate during epithelial-mesenchymal transformation. *International Journal of Developmental Biology*, 42(5), 733-736.

SUN, Y., ALMOMANI, R., ATEN, E., CELLI, J., VAN DER HEIJDEN, J., VENSELAAR, H., et al. BREUNING, M. H. (2010). Terminal osseous dysplasia is caused by a single recurrent mutation in the FLNA gene. *Am J Hum Genet*, 87(1), 146-53.

SUN, Y., HUANG, Y., YIN, A., PAN, Y., WANG, Y., WANG, C., et al. YANG, Y. (2015). Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. *Nature Communications*, 6, 6414.

SUTHER, S. & GOODSON, P. (2003). Barriers to the provision of genetic services by primary care physicians: a systematic review of the literature. *Genetics in Medicine*, 5(2), 70.

SUZUKI, K., HU, D., BUSTOS, T., ZLOTOGORA, J., RICHIERI-COSTA, A., HELMS, J. A. & SPRITZ, R. A. (2000). Mutations of PVRL1, encoding a cell-cell adhesion molecule/herpesvirus receptor, in cleft lip/palate-ectodermal dysplasia. *Nature Genetics*, 25(4), 427-430.

SWAMINATHAN, G. J., BRAGIN, E., CHATZIMICHALI, E. A., CORPAS, M., BEVAN, A. P., WRIGHT, C. F., et al. FIRTH, H. V. (2012). DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. *Hum Mol Genet*, 21(R1), R37-44.

SWEENEY, W. M., LANIER, S. T., PURNELL, C. A. & GOSAIN, A. K. (2015). Genetics of cleft palate and velopharyngeal insufficiency. *Journal of pediatric genetics*, 4(01), 009-016.

SZABO-ROGERS, H. L., SMITHERS, L. E., YAKOB, W. & LIU, K. J. (2010). New directions in craniofacial morphogenesis. *Dev Biol*, 341(1), 84-94.

TABLER, J. M., BARRELL, W. B., SZABO-ROGERS, H. L., HEALY, C., YEUNG, Y., PERDIGUERO, E. G., et al. LIU, K. J. (2013). Fuz mutant mice reveal shared mechanisms between ciliopathies and FGF-related syndromes. *Dev Cell*, 25(6), 623-35.

TABLER, J. M., RICE, C. P., LIU, K. J. & WALLINGFORD, J. B. (2016). A novel ciliopathic skull defect arising from excess neural crest. *Dev Biol*, 417(1), 4-10.

TABLER, J. M., RIGNEY, M. M., BERMAN, G. J., GOPALAKRISHNAN, S., HEUDE, E., AL-LAMI, H. A., et al. WALLINGFORD, J. B. (2017). Cilia-mediated Hedgehog signaling controls form and function in the mammalian larynx. *Elife*, 6.

TAKIGAWA, T. & SHIOTA, K. (2004). Terminal differentiation of palatal medial edge epithelial cells in vitro is not necessarily dependent on palatal shelf contact and midline epithelial seam formation. *International Journal of Developmental Biology*, 48(4), 307-317.

TAN, R., WANG, Y., KLEINSTEIN, S. E., LIU, Y., ZHU, X., GUO, H., et al. ZHU, M. (2014). An evaluation of copy number variation detection tools from whole-exome sequencing data. *Human mutation*, 35(7), 899-907.

TAN, T. Y., AMOR, D. J., RILEY, M., HALLIDAY, J., KILPATRICK, N., SIMMS, K. & WHITE, S. M. (2009). Registry-and clinic-based analyses of birth defects and syndromes associated

with cleft lip/palate in Victoria, Australia. *The Cleft Palate-Craniofacial Journal*, 46(6), 583-587.

TAN, T. Y., KILPATRICK, N. & FARLIE, P. G. (2013). Developmental and genetic perspectives on Pierre Robin sequence. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 163(4), 295-305.

TANAKA, S. A., MAHABIR, R. C., JUPITER, D. C., & MENEZES, J. M. (2012). Updating the epidemiology of cleft lip with or without cleft palate. *Plastic and reconstructive surgery*, 129(3), 511e-518e.

TASSABEHJI, M., READ, A. P., NEWTON, V. E., HARRIS, R., BALLING, R., GRUSS, P. & STRACHAN, T. (1992). Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature*, 355(6361), 635.

TATA, B., HUIJBREGTS, L., JACQUIER, S., CSABA, Z., GENIN, E., MEYER, V., et al. DE ROUX, N. (2014). Haploinsufficiency of Dmxl2, encoding a synaptic protein, causes infertility associated with a loss of GnRH neurons in mouse. *PLoS Biol*, 12(9), e1001952.

THE NATIONAL GENOMICS RESEARCH AND HEALTHCARE KNOWLEDGEBASE v5, Genomics England. doi:10.6084/m9.figshare.4530893.v5. 2019.

THESLEFF, I. (2014). Current understanding of the process of tooth formation: transfer from the laboratory to the clinic. *Australian dental journal*, 59, 48-54.

THEVENEAU, E. & MAYOR, R. (2011). Collective cell migration of the cephalic neural crest: The art of integrating information. *genesis*, 49(4), 164-176.

THEVENEAU, E. & MAYOR, R. (2012). Neural crest migration: interplay between chemorepellents, chemoattractants, contact inhibition, epithelial-mesenchymal transition, and collective cell migration. *Wiley Interdiscip Rev Dev Biol*, 1(3), 435-45.

THIEL, R., CHAHOUD, I., JÜRGENS, M. & NEUBERT, D. (1993). Time-dependent differences in the development of somites of four different mouse strains. *Teratogenesis, Carcinogenesis, and Mutagenesis*, 13(6), 247-257.

THOMAS, D. C. & WITTE, J. S. (2002). Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev*, 11(6), 505-12.

TRAINOR, P. A. (2005). Specification and patterning of neural crest cells during craniofacial development. *Brain, behavior and evolution*, *66*(4), 266-280.

TRAINOR, P. A., DIXON, J. & DIXON, M. J. (2009). Treacher Collins syndrome: etiology, pathogenesis and prevention. *European Journal of human genetics*, 17(3), 275.

TUCKER, A. AND SHARPE, P. (2004) The cutting-edge of mammalian development; how the embryo makes teeth. *Nature reviews genetics*, 5(7), pp. 499-508.

TURNER, T. N., SHARMA, K., OH, E. C., LIU, Y. P., COLLINS, R. L., SOSA, M. X., et al. CHAKRAVARTI, A. (2015). Loss of delta-catenin function in severe autism. *Nature*, 520(7545), 51-6.

TWIGG, S. R. & WILKIE, A. O. (2015). New insights into craniofacial malformations. *Hum Mol Genet*, 24(R1), R50-9.

TYLER, M. S., & KOCH, W. E. (1977). In vitro development of palatal tissues from embryonic mice: III. Interactions between palatal epithelium and heterotypic oral mesenchyme. *Development*, *38*(1), 37-48.

VADLAMUDI, U., ESPINOZA, H. M., GANGA, M., MARTIN, D. M., LIU, X., ENGELHARDT, J. F. & AMENDT, B. A. (2005). PITX2, β-catenin and LEF-1 interact to synergistically regulate the LEF-1 promoter. *Journal of cell science*, 118(6), pp. 1129-1137.

VAN DEN BOOGAARD, M.-J., CRÉTON, M., BRONKHORST, Y., VAN DER HOUT, A., HENNEKAM, E., LINDHOUT, D., et al. VAN AMSTEL, H. K. P. (2012). Mutations in WNT10A are present in more than half of isolated hypodontia cases. *Journal of medical genetics*, 49(5), 327-331.

VAN DEN BOOGAARD, M. J., DORLAND, M., BEEMER, F. A. & VAN AMSTEL, H. K. (2000). MSX1 mutation is associated with orofacial clefting and tooth agenesis in humans. *Nat Genet*, 24(4), 342-3.

VANDERAS, A. P. (1987). Incidence of cleft lip, cleft palate, and cleft lip and palate among races: a review. *Cleft palate J*, 24(3), 216-225.

VEGA-LOPEZ, G. A., CERRIZUELA, S., TRIBULO, C. & AYBAR, M. J. (2018). Neurocristopathies: New insights 150 years after the neural crest discovery. *Developmental Biology*, 444, S110-S143.

VELTMAN, J. A. & BRUNNER, H. G. (2012). De novo mutations in human genetic disease. *Nat Rev Genet*, 13(8), 565-75.

VIEIRA, A. R., AVILA, J. R., DAACK-HIRSCH, S., DRAGAN, E., FELIX, T. M., RAHIMOV, F., et al. MURRAY, J. C. (2005). Medical sequencing of candidate genes for nonsyndromic cleft lip and palate. *PLoS Genet*, 1(6), e64.

VIEIRA, A. R. & KUP, E. (2016). On the etiology of molar-incisor hypomineralization. *Caries research*, 50(2), 166-169.

VIEIRA, A. R., MCHENRY, T. G., DAACK-HIRSCH, S., MURRAY, J. C. & MARAZITA, M. L. (2008). A genome wide linkage scan for cleft lip and palate and dental anomalies. *American Journal of Medical Genetics Part A*, 146(11), 1406-1413.

VIEIRA, A. R., MODESTO, A., MEIRA, R., BARBOSA, A. R. S., LIDRAL, A. C. & MURRAY, J. C. (2007). Interferon regulatory factor 6 (IRF6) and fibroblast growth factor receptor 1 (FGFR1) contribute to human tooth agenesis. *American Journal of Medical Genetics Part A*, 143(6), 538-545.

VINCENT, M., GENEVIEVE, D., OSTERTAG, A., MARLIN, S., LACOMBE, D., MARTIN-COIGNARD, D., et al. COLLET, C. (2016). Treacher Collins syndrome: a clinical and molecular study based on a large series of patients. *Genet Med*, 18(1), 49-56.

VINYOLES, M., DEL VALLE-PÉREZ, B., CURTO, J., VIÑAS-CASTELLS, R., ALBA-CASTELLÓN, L., DE HERREROS, A. G. & DUÑACH, M. (2014). Multivesicular GSK3 sequestration upon Wnt signaling is controlled by p120-catenin/cadherin interaction with LRP5/6. *Molecular cell*, 53(3), 444-457.

VISSCHER, P. M., WRAY, N. R., ZHANG, Q., SKLAR, P., MCCARTHY, M. I., BROWN, M. A. & YANG, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*, 101(1), 5-22.

VISSERS, L. E., VAN RAVENSWAAIJ, C. M., ADMIRAAL, R., HURST, J. A., DE VRIES, B. B., JANSSEN, I. M., et al. VAN KESSEL, A. G. (2004). Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet*, 36(9), 955-7.

VOIGT, C., MÉGARBANÉ, A., NEVELING, K., CZESCHIK, J. C., ALBRECHT, B., CALLEWAERT, B., et al. KÖNIG, R. (2013). Oto-facial syndrome and esophageal atresia, intellectual disability and zygomatic anomalies-expanding the phenotypes associated with EFTUD2 mutations. *Orphanet journal of rare diseases*, 8(1), 110.

WACHOLDER, S., ROTHMAN, N. & CAPORASO, N. (2002). Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. *Cancer Epidemiol Biomarkers Prev*, 11(6), 513-20.

WAHLSTEN, D. & WAINWRIGHT, P. (1977). Application of a morphological time scale to hereditary differences in prenatal mouse development. *Journal of Embryology and Experimental Morphology*, 42(1), 79-92.

WALTER, J. D. (1990). Webbing of the pharyngeal recess in adults with cleft palate. *Cleft Palate J*, 27(4), 411-3; discussion 414.

WANG, B., LI, H., LIU, Y., LIN, X., LIN, Y., WANG, Y., et al. ZHANG, Y. (2014). Expression patterns of WNT/β-CATENIN signaling molecules during human tooth development. *Journal of molecular histology*, 45(5), 487-496.

WANG, L., TANG, Q., XU, J., LI, H., YANG, T., LI, L., et al. CHEN, Y. (2020). The transcriptional regulator MEIS2 sets up the ground state for palatal osteogenesis in mice. *Journal of Biological Chemistry*, *295*(16), 5449-5460.

WEINBERG, S., NAIDOO, S., BARDI, K., BRANDON, C., NEISWANGER, K., RESICK, J., et al. MARAZITA, M. (2009). Face shape of unaffected parents with cleft affected offspring: combining three-dimensional surface imaging and geometric morphometrics. *Orthodontics & craniofacial research*, 12(4), 271-281.

WEINBERG, S. M., BRANDON, C. A., MCHENRY, T. H., NEISWANGER, K., DELEYIANNIS, F. W., DE SALAMANCA, J. E., et al. MARAZITA, M. L. (2008). Rethinking isolated cleft palate: evidence of occult lip defects in a subset of cases. *Am J Med Genet A*, 146a(13), 1670-5.

WEINBERG, S. M., NEISWANGER, K., MARTIN, R. A., MOONEY, M. P., KANE, A. A., WENGER, S. L., et al. DE SALAMANCA, J. E. (2006). The Pittsburgh Oral-Facial Cleft study: expanding the cleft phenotype. Background and justification. *The Cleft palate-craniofacial journal*, 43(1), 7-20.

WENG, M., CHEN, Z., XIAO, Q., LI, R. & CHEN, Z. (2018). A review of FGF signaling in palate development. *Biomed Pharmacother*, 103, 240-247.

WESTFALL, J. M., MOLD, J., & FAGNAN, L. (2007). Practice-based research—"Blue Highways" on the NIH roadmap. *Jama*, 297(4), 403-406.

WHITE, K. E., CABRAL, J. M., DAVIS, S. I., FISHBURN, T., EVANS, W. E., ICHIKAWA, S., et al. ECONS, M. J. (2005). Mutations that cause osteoglophonic dysplasia define novel roles for FGFR1 in bone elongation. *Am J Hum Genet*, 76(2), 361-7.

WHO, R. (2001). Craniofacial anomalies and associated birth defects. *Global registry and database on craniofacial anomalies*.

WILDENBERG, G. A., DOHN, M. R., CARNAHAN, R. H., DAVIS, M. A., LOBDELL, N. A., SETTLEMAN, J. & REYNOLDS, A. B. (2006). p120-catenin and p190RhoGAP regulate cell-cell adhesion by coordinating antagonism between Rac and Rho. *Cell*, 127(5), 1027-1039.

WILKIE, A. O., SLANEY, S. F., OLDRIDGE, M., POOLE, M. D., ASHWORTH, G. J., HOCKLEY, A. D., et al. ET AL. (1995). Apert syndrome results from localized mutations of FGFR2 and is allelic with Crouzon syndrome. *Nat Genet*, 9(2), 165-72.

WILKINSON, D. G., BAILES, J. A. & MCMAHON, A. P. (1987). Expression of the proto-oncogene int-1 is restricted to specific neural cells in the developing mouse embryo. *Cell*, 50(1), 79-88.

WOLF, Z. T., BRAND, H. A., SHAFFER, J. R., LESLIE, E. J., ARZI, B., WILLET, C. E., et al. BANNASCH, D. L. (2015). Genome-wide association studies in dogs and humans identify ADAMTS20 as a risk variant for cleft lip and palate. *PLoS genetics*, 11(3), e1005059.

WOLF, Z. T., LESLIE, E. J., ARZI, B., JAYASHANKAR, K., KARMI, N., JIA, Z., et al. BANNASCH, D. L. (2014). A LINE-1 insertion in DLX6 is responsible for cleft palate and mandibular abnormalities in a canine model of Pierre Robin sequence. *PLoS Genet*, 10(4), e1004257.

WONG, M. T., SCHÖLVINCK, E. H., LAMBECK, A. J. & VAN RAVENSWAAIJ-ARTS, C. M. (2015). CHARGE syndrome: a review of the immunological aspects. *European journal of human genetics*, 23(11), 1451.

WORLD HEALTH ORGANIZATION. (2002). Global strategies to reduce the health-care burden of craniofacial anomalies.

WRIGHT, C. F., FITZGERALD, T. W., JONES, W. D., CLAYTON, S., MCRAE, J. F., VAN KOGELENBERG, M., et al. BAYZETINOVA, T. (2015). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *The Lancet*, 385(9975), 1305-1314.

WRIGHT, C. F., FITZPATRICK, D. R. & FIRTH, H. V. (2018a). Paediatric genomics: diagnosing rare disease in children. *Nature Reviews Genetics*, 19(5), 253.

WRIGHT, C. F., MCRAE, J. F., CLAYTON, S., GALLONE, G., AITKEN, S., FITZGERALD, T. W., et al. LORD, J. (2018b). Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genetics in Medicine*, 20(10), 1216.

WU, D., MANDAL, S., CHOI, A., ANDERSON, A., PROCHAZKOVA, M., PERRY, H., et al. SLAVOTINEK, A. M. (2015a). DLX4 is associated with orofacial clefting and abnormal jaw development. *Human Molecular Genetics*, 24(15), 4340-4352.

WU, L., SCHAID, D. J., SICOTTE, H., WIEBEN, E. D., LI, H. & PETERSEN, G. M. (2015b). Case-only exome sequencing and complex disease susceptibility gene discovery: study design considerations. *J Med Genet*, 52(1), 10-6.

WU, W., GU, S., SUN, C., HE, W., XIE, X., LI, X., et al. LIU, C. (2015c). Altered FGF Signaling Pathways Impair Cell Proliferation and Elevation of Palate Shelves. *PLoS One*, 10(9), e0136951.

WU, X., ZHOU, J., HUANG, J., HUANG, H., XU, B., WANG, X., et al. SHI, B. (2010). Effects of vitamin B 12 on murine embryonic dexamethasone-induced cleft palates using 1 H-NMR-based metabonomic analysis. *Metabolomics*, *6*(4), 467-477.

XIA, C., MA, W., STAFFORD, L. J., LIU, C., GONG, L., MARTIN, J. F. & LIU, M. (2003a). GGAPs, a New Family of Bifunctional GTP-Binding and GTPase-Activating Proteins. *Molecular and Cellular Biology*, 23(7), 2476-2488.

XIA, X., MARINER, D. J. & REYNOLDS, A. B. (2003b). Adhesion-associated and PKC-modulated changes in serine/threonine phosphorylation of p120-catenin. *Biochemistry*, 42(30), 9195-9204.

XU, X., HAN, J., ITO, Y., BRINGAS, P., URATA, M. M. & CHAI, Y. (2006). Cell autonomous requirement for Tgfbr2 in the disappearance of medial edge epithelium during palatal fusion. *Developmental biology*, 297(1), 238-248.

YANAGISAWA, M. & ANASTASIADIS, P. Z. (2006). p120 catenin is essential for mesenchymal cadherin–mediated regulation of cell motility and invasiveness. *The Journal of cell biology*, 174(7), 1087-1096.

YANG, J. & WEINBERG, R. A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell*, 14(6), 818-29.

YANG, Y., MUZNY, D. M., XIA, F., NIU, Z., PERSON, R., DING, Y., et al. ENG, C. M. (2014). Molecular Findings Among Patients Referred for Clinical Whole-Exome SequencingClinical Whole-Exome SequencingClinical Whole-Exome Sequencing. *JAMA*, 312(18), 1870-1879.

YOSHIDA, T., VIVATBUTSIRI, P., MORRISS-KAY, G., SAGA, Y., & ISEKI, S. (2008). Cell lineage in mammalian craniofacial mesenchyme. *Mechanisms of development*, *125*(9-10), 797-808.

YU, K. P. T., LUK, H. M., GORDON, C. T., FUNG, G., OUFADEM, M., GARCIA-BARCELO, M. M., et al. TIONG, Y. T. (2018). Mandibulofacial dysostosis Guion-Almeida type caused by novel EFTUD2 splice site variants in two Asian children. *Clin Dysmorphol*, 27(2), 31-35.

ZARREI, M., FEHLINGS, D. L., MAWJEE, K., SWITZER, L., THIRUVAHINDRAPURAM, B., WALKER, S., et al. SCHERER, S. W. (2018). De novo and rare inherited copy-number variations in the hemiplegic form of cerebral palsy. *Genet Med*, 20(2), 172-180.

ZHANG, R., ENGLER, A. & TAYLOR, V. (2018). Notch: an interactive player in neurogenesis and disease. *Cell Tissue Res*, 371(1), 73-89.

ZHANG, Y. D., CHEN, Z., SONG, Y. Q., LIU, C. & CHEN, Y. P. (2005). Making a tooth: growth factors, transcription factors, and stem cells. *Cell Res*, 15(5), 301-16.

ZHANG, Z., SONG, Y., ZHAO, X., ZHANG, X., FERMIN, C., & CHEN, Y. (2002). Rescue of cleft palate in Msx1-deficient mice by transgenic Bmp4 reveals a network of BMP and Shh signaling in the regulation of mammalian palatogenesis. *Development*, *129*(17), 4135-4146

ZHU, Y., WU, Y., KIM, J. I., WANG, Z., DAAKA, Y. & NIE, Z. (2009). Arf GTPase-activating protein AGAP2 regulates focal adhesion kinase activity and focal adhesion remodeling. *J Biol Chem*, 284(20), 13489-96.

ZHU, Z., BAKSHI, A., VINKHUYZEN, A. A., HEMANI, G., LEE, S. H., NOLTE, I. M., et al. YANG, J. (2015). Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am J Hum Genet*, 96(3), 377-85.

ZUCCHERO, T. M., COOPER, M. E., MAHER, B. S., DAACK-HIRSCH, S., NEPOMUCENO, B., RIBEIRO, L., et al. MACHIDA, J. (2004). Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. *New England Journal of Medicine*, 351(8), 769-780.

ZWIJNENBURG, P. J., MEIJERS-HEIJBOER, H. & BOOMSMA, D. I. (2010). Identical but not the same: the value of discordant monozygotic twins in genetic research. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 153(6), 1134-1149.

APPENDICES

List of included Appendices:

Appendix 1 BSPD-FDS (RCSEng) Small Grant Award

Appendix 2 Research Ethics Favourable Opinion & Substantial Amendment Approval

Appendix 3 Consent Forms, Patient Information Sheets & Study Flyers

Appendix 4 Data Collection Sheet (Family Booklet)

Appendix 5 Detailed Phenotypes of Index Patients with *CTNND1* Variants (Chapter 4)

Appendix 6 Appendices for Chapter 6:

A. RVIS and GDIS scores for all 43 genes examined.
B. ExAC Z-scores and pLI scores for the five chosen candidate genes that met the filtering criteria.
C. *EFTUD2* variants in chosen DECIPHER and CAP180 patients.
D. Clinical description of probands carrying *de novo* mutations in *EFTUD2*.
E. HPO terms and phenotypes in DECIPHER *FBXO11* DDD research variants.
F. Clinical description of CAP180 probands carrying *de novo* mutations in previously un-reported *FBXO11* variants.
G. *CELSR1* Variants in DECIPHER and CAP180.
H. Clinical description of CAP180 probands carrying mutations in *CELSR1*.
I. Previous genetic tests carried out for probands carrying mutations in *CELSR1.*
J. *DIP2C* single and copy number variants in DECIPHER and CAP180 probands.
K. Clinical description of CAP180 and DECIPHER probands carrying mutations or CNVs in *DIP2C.*
L. Detailed clinical description of CAP180 and DECIPHER probands carrying mutations or CNVs in *DIP2C*.
M. *ABCA2* single and copy number variants in DECIPHER and CAP180.
N. Phenotypes in patients with SNV and CNV in the *ABCA2* gene.

Appendix 1

Royal College of Surgeons
FACULTY OF DENTAL SURGERY

Faculty of Dental Surgery
The Royal College of Surgeons of England
35–43 Lincoln's Inn Fields
London WC2A 3PE
www.rcseng.ac.uk/fds

Ms Reham Alharatani
Centre for Craniofacial and Regenerative Biology
Floor 27, Tower Wing
Guy's Hospital
London
SE1 9RT

6th July 2017

Dear Ms Alharatani

Re: Genetic Associations between Oral Clefting and Dental Anomalies in Children and their Families

I am pleased to inform you that your application for a FDS RCSEng- British Society of Paediatric Dentistry grant has been successful.

Your research project has been awarded **£9,025** to cover the following:

Oragene OG-500 (300 kits	£4.950
Oragene OG-575 (50 kits)	£875
Oragene PTL2P-5 PrepIT.L2P (350 kits)	£350
Dig RNA Labelling Mix	£450
Anti-Digoxigenin antibody	£400
BM Purple AP Substrate	£500
Blocking Reagent	£500
Microscopy access	£1,000

The grant will be paid upon receipt of an invoice from your host institution.

You may be aware the Faculty publishes the Faculty Dental Journal (FDJ) quarterly and we are particularly interested in receiving articles relating to dental research. It is a condition of award of the grant that you write a short article for publication in the FDJ upon completion of your research study. This will be a good opportunity for you and an excellent chance for Members and Fellows to see how the research funds are being used and to learn about your accomplishments. The research section editor will provide you with guidance and any necessary help on formatting and designing your article. I will send you a reminder email closer to your completion date.

Congratulations and good luck.

Yours sincerely

Professor Paul Speight
Chair, FDS Research Committee

Dr Sondos Albadri
BSPD, Honorary Editor

Registered Charity No. 212808

Appendix 2

Business Services Organisation

Office for Research Ethics Committees Northern Ireland (ORECNI)

Customer Care & Performance Directorate

Unit 4, Lissue Industrial Estate West
Rathdown Walk
Moira Road
Lisburn
BT28 2RF
Tel: 028 95361400
www.orecni.hscni.net
HSC REC A

01 March 2016

Professor Marie-Therese Hosey
Department of Paediatric Dentistry
136 , 1st Floor, Dental Institute, Denmark Hill, London

Dear Professor Hosey

Study title: Genetic associations between oral clefting and tooth defects in children and their families.
REC reference: 16/NI/0026
Protocol number: NA
IRAS project ID: 185686

Thank you for your letter of 01 March 2016, responding to the Proportionate Review Sub-Committee's request for changes to the documentation for the above study.

The revised documentation has been reviewed and approved by the Chair of the Proportionate Review sub-committee.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager Mrs Katrina Greer, PRS@hscni.net. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a **favourable ethical opinion** for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission must be obtained from each host organisation prior to the start of the study

at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA Approval (England)/ NHS permission for research is available in the Integrated Research Application System, www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where an NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publicly accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" above).

Approved documents

The documents reviewed and approved by the Committee are:

Document	*Version*	*Date*
Copies of advertisement materials for research participants [Child 2-7 years Flyer]	1	25 January 2016
Copies of advertisement materials for research participants [Parent's Flyer]	2	16 February 2016
Copies of advertisement materials for research participants [Child 8-16 years Flyer]	2	16 February 2016
IRAS Checklist XML [Checklist 25012016]		25 January 2016
IRAS Checklist XML [Checklist 16022016]		16 February 2016
IRAS Checklist XML [Checklist 01032016]		01 March 2016
Letter from statistician [Statistician Power Calculation]	1	03 September
Letters of invitation to participant [Child 2-7 years Invitation Letter]	1	25 January 2016
Letters of invitation to participant [Parent's Invitation Letter]	2	16 February 2016
Letters of invitation to participant [Child 8-16 years Invitation Letter]	2	16 February 2016
Non-validated questionnaire [Data Collection Booklet]	1	25 January 2016
Other [Response to outstanding issues]	1	16 February 2016
Other [Response to outstanding issues NUMBER 2]	1	01 March 2016
Participant consent form [Child Assent Form]	1	25 January 2016
Participant consent form [Parent's Consent Form]	3	01 March 2016
Participant consent form [GRANDPARENTS Consent Form]	1	01 March 2016
Participant information sheet (PIS) [Child 8-16 years Information Sheet]	2	16 February 2016
Participant information sheet (PIS) [Child 2-7 years Information Sheet]	2	16 February 2016
Participant information sheet (PIS) [Parent's Info Sheet]	3	01 March 2016
Participant information sheet (PIS) [GRANDPARENTS Information Sheet]	1	01 March 2016
REC Application Form [REC Form 25012016]		25 January 2016
Research protocol or project proposal [Project Protocol]	1	25 January 2016
Summary CV for Chief Investigator (CI) [Chief Investigator's CV]	1	25 January 2016
Summary CV for student [PhD Student's CV]	1	25 January 2016
Summary CV for supervisor (student research) [Academic Supervisor's CV]	1	25 January 2016

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document “After ethical review - guidance for researchers” gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol

- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days - see details at http://www.hra.nhs.uk/hra-training/

16/NI/0026 *Please quote this number on all correspondence*

With the Committee's best wishes for the success of this project.

Yours sincerely

pp Mrs Celia Diver-Hall

Alternate Vice-Chair - Chair of the PRSC Meeting of 01/02/2016

Email: PRS@hscni.net

Enclosures: *"After ethical review - guidance for researchers"*

Copy to: *Mr Keith Brennan, King's College London*

Ms Jennifer Boston, Guy's and St. Thomas' Foundation NHS Trust

HSC Business Services Organisation

Office for Research Ethics Committees Northern Ireland (ORECNI)

Customer Care & Performance Directorate
Unit 4, Lissue Industrial Estate West
Rathdown Walk
Moira Road
Lisburn
BT28 2RF
Tel: 028 95361400
www.orecni.hscni.net
HSC REC A

16 August 2016

Professor Marie-Therese Hosey
Department of Paediatric Dentistry
136, 1st Floor, Dental Institute, Denmark Hill, London

Dear Professor Hosey

Study title:	**Genetic associations between oral clefting and tooth defects in children and their families.**
REC reference:	**16/NI/0026**
Protocol number:	**NA**
Amendment number:	**Substantial Amendment #1**
Amendment date:	**21 July 2016**
IRAS project ID:	**185686**

The above amendment was reviewed at the meeting of the Sub-Committee held on 16 August 2016 in correspondence.

Ethical opinion

The members of the Committee taking part in the review gave a **favourable ethical opinion** of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Notice of Substantial Amendment (non-CTIMP) [Substantial Amendment #1]	Substantial Amendment #1	21 July 2016
Other [Email notice of substantial amendment #1 (Reham Alharatani)]		22 July 2016
Other [Reasons for Substantial Amendment]		17 June 2016
Participant consent form [Parent]	4	17 June 2016
Participant consent form [Grandparents]	2	17 June 2016
Participant information sheet (PIS) [Grandparents/Others]	2	17 June 2016
Participant information sheet (PIS) [Parent]	4	17 June 2016
Research protocol or project proposal	2	17 June 2016

Appendix 3

Prof Marie-Therese Hosey
DDS, MSc, BDS,
FDS RCPS (Glas)
Head of Department

Dept of Paediatric Dentistry
1st Floor Kings College Dental Institute
Bessemer Road
London
SE5 9RS
Tel: 020 3299 4078

Guy's and St Thomas' NHS
NHS Foundation Trust

Participant Study ID

Date:

CONSENT FORM FOR PARENTS AND ADULTS

Title of Project: Genetic associations between tooth defects and oral clefting in children and their families.

1. I confirm that I have read the **information sheet** dated 17.06.2016 (Version 4) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that our participation is voluntary and that we are free to withdraw at any time without giving any reason, without our medical/dental care or legal rights being affected. ☐

3. I understand that relevant sections of any of my child's medical and or dental notes and data collected during the study may be looked at by responsible individuals from the regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my child's records in accordance with the terms of the UK Data Protection Act 1998. ☐

4. I understand that the information collected about us will be used to support other research in the future. I understand that the data will be published as a report but that confidentiality and anonymity will be maintained and it will not be possible to identify me or my child in any publications. ☐

5. I agree to the research team collecting a saliva sample from me and child/children. I agree to my saliva being stored for research, including detailed analysis of my whole genome. ☐

6. I understand that the saliva samples will be analysed in a laboratory at King's College London. ☐

7. I understand that I will not be receiving any feedback unless **a specific genetic cause for cleft** with relevance to my family has been identified with very high confidence. ☐

8. 'I AGREE that our saliva samples surplus to this study and relevant information about me and my child can be used in future research which has been approved by a recognised Research Ethics Committee. I am aware that future research may be in collaboration with a commercial company, but my identity will be kept anonymous at all times'. ☐

9. I am aware that this study does not intend to inform me about the health status of my mouth and/or teeth and does not intend to diagnose any dental disease I might have. ☐

10. We agree to take part in the above study. ☐

Name of Participant Date Signature

Name of Person taking consent Date Signature

**When completed: 1 for participant; 1 for researcher site file; 1 (original) to be kept in medical/dental notes.*

Note: Grandparents and other family member's consent forms are similar to the one above.

Prof Marie-Therese Hosey
DDS, MSc, BDS,
FDS RCPS (Glas)
Head of Department

Dept of Paediatric Dentistry
1st Floor Kings College Dental Institute
Bessemer Road
London
SE5 9RS
Tel: 020 3299 4078

Guy's and St Thomas' NHS
NHS Foundation Trust

Patient study ID:

ASSENT FORM FOR CHILDREN

Research Title:

Genetic associations between tooth defects and oral clefting in children and their families.

Child (or parent/carer on their behalf) to initial.

Please initial boxes on the right if you agree to the following:	**Your Initials**
1) Have you read (or has someone read to you) about this project?	
2) Do you understand what this study is about?	
3) Have you asked all the questions you want?	
4) Have you had your questions answered in a way you understand?	
5) Do you understand it's OK to stop taking part at any time?	
6) Are you happy to take part?	

If any answers are 'no' or you don't want to take part, do not sign your name below!

If you do want to take part, you can write your name below

Your name _____

Signature _____

Date _____

The researcher who explained this project to you needs to sign too:

Name _____

Signature _____

Date _____

THANK YOU FOR YOUR HELP!

**When completed, 1 copy for the patient; 1 copy for the study site file; 1 copy (original) to be kept in dental notes.*

Prof Marie-Therese Hosey
DDS, MSc, BDS,
FDS RCPS (Glas)
Head of Department

Dept of Paediatric Dentistry
1st Floor Kings College Dental Institute
Bessemer Road
London
SE5 9RS
Tel: 020 3299 4078

Guy's and St Thomas' NHS
NHS Foundation Trust

INFORMATION SHEET FOR PARENTS AND ADULTS

Research Title:
Genetic associations between tooth defects and oral clefting in children and their families.

Invitation to take part:
We would like to invite you (and your child) to take part in our research study. We would like to discover what genes could link your teeth to your child's cleft lip and/or palate. Before you decide whether or not to take part, we would like you to understand why the research is being done and what it will involve for you. Please take time to read this information carefully. Reham Alharatani, the researcher, who is a children's dentist, will go through the information sheet with you and answer any questions you have. Talk to others about the study if you wish. Thank you for reading this.

What is the background and purpose of the study?

Cleft of the lip and/or palate is common in babies around the world. Parents of children with cleft often ask why this has happened. They also want to know the chances of it happening again. In most cases, no single cause for the cleft can be found. It is thought that a combination of many different genetic and environmental factors is the cause.

In some children, especially ones with the isolated form of the cleft (where no other medical concerns are involved), a spelling mistake in the DNA (sometimes called a mutation) can result in a cleft. In a few families, there is a strong inheritance tendency. However, many do not know if the cleft in their child has been inherited. Genetic testing is not commonly offered to children with clefts. We think that we can understand inheritance better by searching for clues in parents. We think this can be done by looking at your teeth, whether they were a complete set or missing or extra.

Aim of this study

The main purpose of this study is to discover new genes that may link two very common oral cavity defects: the first, a split of the lip/roof of the mouth (a cleft) and the second, defects in the number or the size of teeth. This study is part of a Doctorate degree.

Why have we been chosen?

We are asking you and your child to take part because you will be having an appointment to see a health care professional at the South Thames Cleft Unit, St. Thomas' Hospital.

It does not matter what treatment is planned or which health care professional you are waiting to see.

Who is expected to take part?

Since this is a genetic association study, we will need to gather information from family 'trios', this means a group of three individuals. This allows us to trace back inheritance patterns of a particular gene. Therefore, **both** sides of your child's family need to be represented. This means we will need to collect 'spit' (saliva) samples from your child who's a patient at the South Thames Cleft Unit, in addition to 'spit' samples from both parents. If both parents can't give a sample, then a sample taken from a brother or sister or grandparent (on the other side of the family) will be just as good.

Do we have to take part?

No. It is up to you and your child to decide whether you wish to take part. If you do decide to take part, you will be asked to sign a consent form. Your child will also sign an assent form. You are free to withdraw at any time and without giving any reason. Whether you wish to take part in this study or not, you and your child will not be disadvantaged in any way and your child's standard of care will remain unaffected.

What will my participation involve?

- If you agree to take part, you will be involved in this study only on the day of your visit.
- When you attend with your child for your appointment at the South Thames Cleft Unit, you will meet the researcher (Reham Alharatani) before or after your appointment with the health care professional.
- Whilst you are waiting, she will give you a copy of the information sheet and another for your child and your signed consent form to keep. A copy of the consent form and child assent form will be kept in the dental records.
- She will then invite you and all your accompanying family members into a dental clinic. She will ask you a series of questions regarding your family history. Then she will examine the teeth of all family members. She will ask to check, count and record the number and shape of all your teeth and may also take photographs of you or child's teeth.
- She may ask you and all accompanying family members to 'spit' into a special tube each, depending on the questions you've answered and the number/shape of your teeth. These will be used to carry out biological research, including analysis of your genome which contains your DNA.
- These 'spit' samples will be analysed in the Craniofacial Laboratory at King's College London. We will only look for special genes that cause clefts and/or tooth defects. The 'spit' tube will be kept anonymised and secure at all times.
- We anticipate that this whole process should not take more than 45 minutes of your time.

- If your child is accompanied by only one of his/her parents on the day of their visit, the saliva kit/consent form for the third family member (parent or other family member) can be provided for mailing into the lab via pre-paid first-class post.

- **What do we need to do?** The third family member will need to take a saliva sample using the kit that will be given/sent to you. Ensure that you follow the instructions outlined in the inside of the kit. You can either give it back to the researcher (Reham

Alharatani) in person when you next come for a visit with your child OR pop it in the post using the freepost *jiffy* envelope she will provide you with. The return address for the Craniofacial Development Laboratory will be provided on the envelope.

- The researcher Reham Alharatani will ask to contact you a few days later to ensure that the saliva sample of the third family member has been collected and sent to our laboratory.

We will follow ethical and legal practices and all information about you will be handled in confidence and will remain anonymous throughout.

What happens to the saliva samples?

- The samples will be delivered by a registered courier to a laboratory based at the Craniofacial Development and Stem Cell Biology Department at King's College London. They will be processed and stored securely and anonymously.
- We will keep the tubes for approximately four years until the study is completed. We also intend to store your samples anonymously for future ethically approved studies.
- You will be asked to indicate on the consent form if you give permission for the storage and use of your samples in future, ethically approved cleft related research.

How will genetic information be stored?

Each family in this study will have a unique number and this number- but no other identifying information such as name- will be on the DNA or any stored samples.

Would I receive any feedback?

Our analysis may, in some cases, uncover **a specific genetic cause for the cleft**. Please note that we will **only** provide feedback **if** a specific genetic cause for cleft with relevance to your family has been identified with very high confidence, and confirmed by an expert team. *If so*, you will be offered a choice for a referral to see a clinical geneticist which our research team will organise for you. However, the overall anonymous findings of the research will be reported in professional publications, meetings and conferences.

What should I NOT expect when taking part in this study?

This study is not intended to inform you about the health status of your mouth and does not intend to diagnose any dental disease you may have. A thorough dental check-up will therefore NOT be carried out. We only plan to count the NUMBER of teeth you have and look at the SHAPE of your teeth. It is recommended you see your dentist every six months to have your gums, teeth and oral mucosa checked.

Are there any benefits involved in participating?

There may be no direct advantage to you or your child in taking part but the information we get out of this study might help families with cleft in the future to better understand some of the genetic causes of cleft.

What are the possible disadvantages and risks of taking part?

Your participation will not affect the care we provide for your child both now and if future treatment is ever required. There are no risks to you or your child in taking part in this study.

Will my taking part in this study be kept confidential?

Yes. All information which is collected about you during the course of the research will be kept strictly confidential. Collected data will be stored anonymously on a secure, password protected database in keeping with the Data Protection Act 1998. This data will not be accessible to anyone other than the immediate study team. Any information about you which leaves the hospital (i.e. the 'spit' tube) will have your name, address and any identifiable information removed so that you cannot be recognised from it.

What will happen if you or your child decide not to take part?

You can stop taking part at any time. This will not affect the care you or your child receive.

If you have concerns

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions. You can contact ***Professor Marie Therese Hosey*** *(Consultant in Paediatric Dentistry) at* ***Telephone: 0203 299 4078 / Fax: 0203 299 4074*** *or at the following email address****: m.t.hosey@kcl.ac.uk****. If you remain unhappy and wish to complain formally, you can do this through the Guy's and St Thomas' Patients Advice and Liaison Service (PALS) on 020 7188 8801,* ***pals@gstt.nhs.uk****. The PALS team are based in the main entrance on the ground floor at St Thomas' Hospital and on the ground floor at Guy's Hospital in the Tower Wing.*

In the event that something does go wrong and you are harmed during the research you may have grounds for legal action for compensation against Guy's and St Thomas' NHS Foundation Trust and/or King's College London but you may have to pay your legal costs. The normal National Health Service complaints mechanisms will still be available to you (if appropriate).

Further Information

If you have any questions or require more information about this study, please contact **Ms. Reham Alharatani** (Paediatric Dentist, PhD student) at: **reham.al-haratani@kcl.ac.uk**

This project has been approved by the Research Ethics Committee.

Thank you for considering taking part in this study, your help is much appreciated.

Note: Grandparents and other family member's PIS are similar to the one above.

Prof Marie-Therese Hosey
DDS, MSc, BDS,
FDS RCPS (Glas)
Head of Department

Dept of Paediatric Dentistry
1st Floor Kings College Dental Institute
Bessemer Road
London
SE5 9RS
Tel: 020 3299 4078

Guy's and St Thomas' NHS
NHS Foundation Trust

INFORMATION SHEET FOR 8-16 YEAR OLDS

Research Title:

Genetic associations between tooth defects and oral clefting in children and their families.

Hello! My name is Reham, I am a children's dentist and I would like to invite you to take part in our research study. Before you decide if you want to take part, take time to read this sheet of information carefully so that you understand why this project is being done and what it will involve. I will go through the information sheet with you and answer any questions you have. Thank you for reading this.

What is this about?

Cleft of the lip and/or palate is common in babies around the world [A cleft is a separation in the lip or the roof of the mouth (palate)]. The cause of a cleft is very complex. For many years, scientists have been trying to discover the causes for cleft. DNA, which is the blueprint for our bodies, has many genes. The genes tell us everything about ourselves. Small changes in the genes can lead to changes in our bodies. For example, your eye colour and hair colour are controlled by different genes. Therefore, lots of genes control whether or not you will get a cleft. In our research project, we think that your cleft might be linked to your parent's teeth (even if they weren't born with a cleft themselves). We think if we explore your parent's set of tooth genes, we might be able to link that to your cleft gene. This is because the roof of the mouth and teeth form around the same time when you are still a foetus.

So, our aim is to discover new genes that may link two very common mouth defects: the first, a split of the lip/roof of the mouth (a cleft) and the second, defects in number/size of teeth.

Why have I been chosen?

If you have been born with a cleft or have a family member born with it, you can help us in this project. We chose you because you or a family member are registered with the South Thames Cleft Unit. It does not matter what treatment is planned or which health care professional you are due to see on the day of your visit.

Who is expected to take part?

Since we are interested in finding out what tooth and cleft genes you might have inherited from your family, we will need to gather information about three members of your family, in scientific terms this is called family 'trios'. Therefore, both sides of your family need to be represented in this project. This means we will need to collect a 'spit' (saliva) sample from you, in addition to 'spit' samples from both your parents. If both parents can't give a sample, then a sample taken from a brother or sister or grandparent will be just as good.

Do I have to take part?

No! It is up to you to decide if you wish to take part. We will give you further explanation and answer any questions you might have about this project. If you and your parents agree to take part, we will get your parent to sign a consent form, and you sign an assent form, showing us that you are happy to participate.

You can stop taking part at any time, without giving any reason. This would NOT affect the care you receive.

What will I have to do if I decide to take part?

1. If you agree to take part, you will be involved in this study only on the day of your visit.
2. You will meet the researcher before or after your appointment with the health care professional. Whilst you are waiting, she will give you a copy of the information sheet and your signed consent/assent form to keep.

3. She will then invite you and all your accompanying family members into a dental clinic.
4. She will ask you some questions about your health and teeth.
5. Then she will examine the mouths and teeth of all family members, including yourself. She will use a mouth mirror and count the number and shape of teeth you have.
6. You will be asked to 'spit' into a special tube.
7. This spit will be sent to a special laboratory, at King's College London, that will explore the DNA in your 'spit' to look for genes that may have caused your cleft.
8. No one will be able to identify if the 'spit' tube belongs to you except for the researcher.

We think that all of this should take about 45 minutes of your time.

What benefit will I get?

We cannot promise the study will help you but the information we get might help other children with cleft and their parents in the future.

What happens to my 'spit' tube?

No one will be able to tell if the 'spit' is yours, only the researcher (Reham) can. Also, any information you or your family gave us will be kept secret, safe and not linked to your name.

What if I want to ask more questions?

You can ask the researcher (Reham) on the day of your visit or contact the research lead: Professor Marie-Therese Hosey on **0203 299 4078**.

Prof Marie-Therese Hosey
DDS, MSc, BDS,
FDS RCPS (Glas)
Head of Department

Dept of Paediatric Dentistry
1st Floor Kings College Dental Institute
Bessemer Road
London
SE5 9RS
Tel: 020 3299 4078

Guy's and St Thomas' NHS
NHS Foundation Trust

INFORMATION SHEET FOR CHILDREN (2-7) YEARS

Hello,
My name is Reham

I am a children's dentist

I want to find out more about what makes your teeth and mouth so special

We are asking you to help us with our quest to find what is so special about your smile

1. We will ask you some questions
2. You will help me count your teeth
3. You will also spit into a little bottle

Reference No.: 16/NI/0026 INFO SHEET (CHILD 2-7) Version 2: Date 16.02.2016

We will count your teeth

You will spit in here

You are a super star !

If you don't want to take part, or change your mind just tell your parents. They will not be cross with you and it will not change your treatment.

We cannot promise that helping us will help you but it might help other children and parents later.

Prof Marie-Therese Hosey
DDS, MSc, BDS,
FDS RCPS (Glas)
Head of Department

Dept of Paediatric Dentistry
1st Floor Kings College Dental Institute
Bessemer Road
London
SE5 9RS
Tel: 020 3299 4078

Guy's and St Thomas' NHS
NHS Foundation Trust

Research Title: **Genetic associations between tooth defects and oral clefting in children and their families.**

Are you or any of your family members attending the Cleft Unit at St. Thomas?

Could the number and shape of your teeth influence your child's cleft lip and/or palate? If so, what are the genes involved in doing so?

We are carrying out a study to learn more about how (if at all) differences in the number, size and shape of teeth in a family member might affect the chances of another family member having a cleft of the lip and/or palate. You can help us by answering some questions and providing us with a sample of your 'spit' (saliva).

We will need to gather information from family 'trios', this means a group of three individuals. Therefore, both sides of your child's family need to be represented. This means we will need to collect 'spit' samples from your child who's a patient at the South Thames Cleft Unit, in addition to 'spit' samples from both parents. If both parents can't give a sample, then a sample taken from a brother or sister or grandparent will be just as good.

So when you next visit the South Thames Cleft Unit at St. Thomas' Hospital for an appointment, we may ask you a few starter questions. We will talk to you about this project and provide you with some more information about the study, if you are interested. Then we can see if you might be willing to help us with this exciting project. Your information/spit samples will stay **anonymous and confidential** at all times.

If you have any questions or would like further information, please contact Prof. MT Hosey on 020 3299 4078.

This project has been approved by the Research Ethics Committee.

REC Reference No.: 16/NI/0026 PARENTS FLYER Version 2: Date 16.02.2016

Prof Marie-Therese Hosey
DDS, MSc, BDS,
FDS RCPS (Glas)
Head of Department

Dept of Paediatric Dentistry
1st Floor Kings College Dental Institute
Bessemer Road
London
SE5 9RS
Tel: 020 3299 4078

Guy's and St Thomas' NHS
NHS Foundation Trust

Research Title: Genetic associations between tooth defects and oral clefting in children and their families

Are you or any of your family members attending the Cleft Unit at St. Thomas?

We are carrying out a study to learn more about how different numbers and shapes of teeth in some families can be linked to a split in the upper lip or roof of the mouth in the same family.

You can help us by answering some questions, counting your teeth and providing us with a sample of your 'spit'. Two family members, other than yourself, will also need to take part. This can be your parents. A sample taken from a brother or sister or a grandparent will be just as good too.

When you next visit the South Thames Cleft Unit at St. Thomas' Hospital for an appointment, we will talk to you about this project. Then we can see if you are happy to help us with this exciting project.

No one else apart from the researcher will know anything about the information you give us about yourself.

If you have any questions or would like further information, please contact Prof. MT Hosey on 020 3299 4078.

This project has been approved by the Research Ethics Committee.

Thanks to Our Volunteers!

REC Reference No.: 16/NI/0026 FLYER (CHILD 8-16) Version 2: Date 16.02.2016

Appendix 4

Prof Marie-Therese Hosey
DDS, MSc, BDS,
FDS RCPS (Glas)
Head of Department

Dept of Paediatric Dentistry
1st Floor Kings College Dental Institute
Bessemer Road
London
SE5 9RS
Tel: 020 3299 4078

Guy's and St Thomas' NHS
NHS Foundation Trust

Research Title: Genetic associations between tooth defects and oral clefting in children and their families.

Data Collection Sheet - 'Family Booklet'

Family Study ID: **Date:**

Inclusion Criteria Checklist

Child accompanied with _____

	YES	NO
Does the child have a cleft of the lip and/or palate? (include, if yes)		
Is the family English literate? (include, if yes)		
Have they had previous genetic testing? (exclude, if yes) **If yes, do they know what their candidate gene was?**		
Are they happy to participate in this study?		
Are they happy to provide a saliva sample?		

- If the family meets the inclusion criteria, the parent should be supplied with:
  - a consent form *(When completed, 1 copy for the patient; 1 copy for the* study site *file; 1 copy (original) to be kept in dental notes).*
  - information sheet.
- The child is given an age appropriate information sheet + assent form.

- Does the parent consent Yes ☐ No ☐
- Does the child assent ☐ Yes ☐ No
- Has the consent form been signed and collected? Yes ☐ No ☐

Child Data Collection Sheet ID:

(note: assign the following sub-codes: child = a; family member1 = b; family member2 = c)

Age: _____

Gender: Female Male

Ethnic origin: _____

Circle one of the following:

- Type of cleft:
  - cleft of the lip only
  - cleft of the lip and alveolus
  - cleft of the lip and palate
  - cleft of the palate only
- Subcategory:
  - Unilateral
  - Bilateral
- Side:
  - Right
  - Left
  - Both

- Syndromic vs. Non-syndromic

If syndromic, what is the syndrome? _____

- Is there any family history of clefting? ☐ Yes ☐ No

If yes, who in the family has a cleft defect? and what type?

_____

_____

Brief medical history of the child with cleft/main participant:

Does the child have or ever had any of the following conditions?

	Yes	No
Congenital heart disease		
Respiratory disease		
Stomach/intestinal disorders		
Kidney/liver diseases		
Urinary tract disease		
Diseases of the joints/muscles		

Other:

_____

Dental Examination:

Dentition: Primary Mixed Permanent

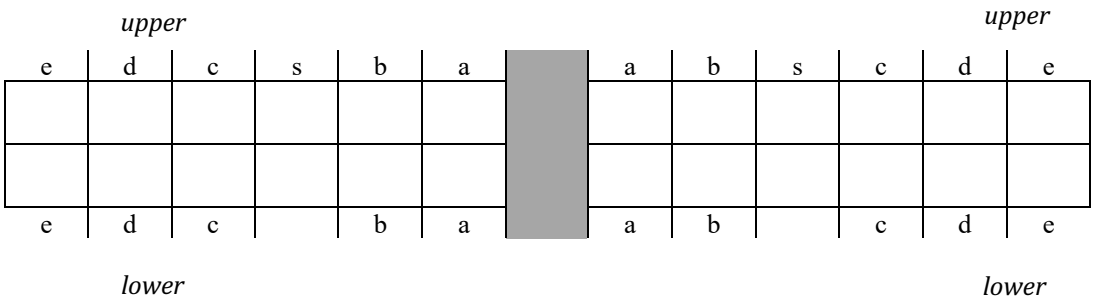
Dental X-ray present? (to check for missing/extra teeth):

_____

_____

Primary dentition

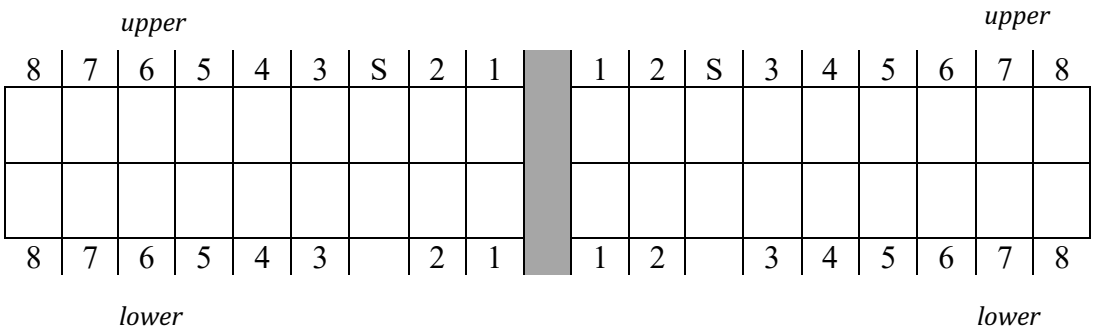
- *Chart status of each tooth in the table using the legend below*



P = tooth present/healthy; **C** = carious; **M** = congenitally missing; **E**= extracted due to caries; **O** = extracted for orthodontics; **EX** = exfoliated; **#** = missing due to trauma or fractured; **F** = filled; **S** = supernumerary; **HM** = hypomineralised; **HP** = hypoplastic; **D** = discoloured**; FG** = fused/geminated; **I** = impacted; **T** = transposed; **mi** = microdont; **ma** = macrodont.

Permanent dentition

- *Chart status of each tooth in the table using the legend below*



P = tooth present/healthy; **C** = carious; **M** = congenitally missing; **E**= extracted due to caries; **O** = extracted for orthodontics; **EX** = exfoliated; **#** = missing due to trauma or fractured; **F** = filled; **S** = supernumerary; **HM** = hypomineralised; **HP** = hypoplastic; **D** = discoloured**; FG** = fused/geminated; **I** = impacted; **T** = transposed; **mi** = microdont; **ma** = macrodont.

Parent1/Family member1 data collection sheet ID:

Family member1 (circle one):

Mother Father Sister Brother Grandfather: paternal – maternal
Grandmother: paternal – maternal Other (please specify):

Does this family member have cleft lip/palate? ☐ Yes ☐ No

Medical history:

Dental history:

Permanent dentition

- *Chart status of each tooth in the table using the legend below*

upper																	*upper*	
8	7	6	5	4	3	S	2	1		1	2	S	3	4	5	6	7	8
8	7	6	5	4	3		2	1		1	2		3	4	5	6	7	8
	lower																*lower*	

P = tooth present/healthy; **C** = carious; **M** = congenitally missing; **E**= extracted due to caries; **O** = extracted for orthodontics; **EX** = exfoliated; **#** = missing due to trauma or fractured; **F** = filled; **S** = supernumerary; **HM** = hypomineralised; **HP** = hypoplastic; **D** = discoloured**; FG** = fused/geminated; **I** = impacted; **T** = transposed; **mi** = microdont; **ma** = macrodont.

- Dental photographs taken: ☐ Yes ☐ No

Parent2/Family member2 data collection sheet ID:

Family member2 (circle one):

Mother Father Sister Brother Grandfather: paternal – maternal
Grandmother: paternal – maternal Other (please specify):

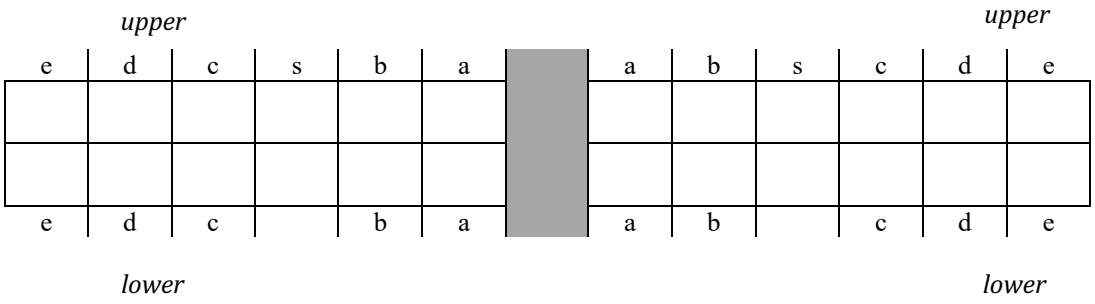
Does this family member have cleft lip/palate? ☐ Yes ☐ No

Medical history:

Dental history:

Primary dentition

- *Chart status of each tooth in the table using the legend below*

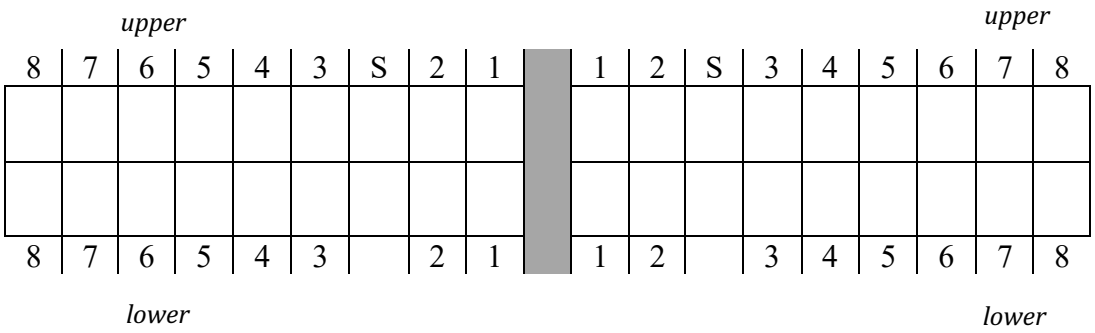


P = tooth present/healthy; **C** = carious; **M** = congenitally missing; **E**= extracted due to caries; **O** = extracted for orthodontics; **EX** = exfoliated; **#** = missing due to trauma or fractured; **F** = filled; **S** = supernumerary; **HM** = hypomineralised; **HP** = hypoplastic; **D** = discoloured**; FG** = fused/geminated; **I** = impacted; **T** = transposed; **mi** = microdont; **ma** = macrodont.

OR:

Permanent dentition

- *Chart status of each tooth in the table using the legend below*



P = tooth present/healthy; **C** = carious; **M** = congenitally missing; **E**= extracted due to caries; **O** = extracted for orthodontics; **EX** = exfoliated; **#** = missing due to trauma or fractured; **F** = filled; **S** = supernumerary; **HM** = hypomineralised; **HP** = hypoplastic; **D** = discoloured**; FG** = fused/geminated; **I** = impacted; **T** = transposed; **mi** = microdont; **ma** = macrodont.

- Dental photographs taken: ☐ Yes ☐ No

Family Saliva Samples

Has a saliva sample been collected from the child? Yes ☐ No ☐

- Has it been labelled with the study ID **ONLY**? Yes ☐ No ☐

Has a saliva sample been collected from family member1? Yes ☐ No ☐

- Has it been labelled with the study ID **ONLY**? Yes ☐ No ☐

Has a saliva sample been collected from family member2? Yes ☐ No ☐

- Has it been labelled with the study ID **ONLY**? ☐ Yes ☐ No

End of booklet

Appendix 5

Participant	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13
DDD ID	NA	NA		294023	278766 NA	-		278000	264418 NA	NA	265667	281120 -	
Variant	V148Dfs*24	V148Dfs*24	R461*	R461*	R468*	L494Rfs*	L494Rfs*	GS 31Afs*6	S868*	S868*	c.2702-5A>G	H913Pfs*3	H913Pfs*3
Sex	Female	Female	Female	Male	Female	Female	Male	Female	Male	Male	Male	Male	Male
Craniofacial													
Cleft lip/palate	no	no	submucous cleft soft palate & bifid uvula	no	no	yes	right unilateral cleft lip and palate	no	bilateral cleft lip and palate	right unilateral cleft lip and palate	left unilateral cleft lip and palate	left unilateral cleft lip and palate	left unilateral cleft lip and palate
High arched palate	yes	yes	yes	no	no	no	yes	yes	no	ND	_	yes	yes
Thin upper lip	yes	yes	no	no	no	no	yes	yes	yes	no	yes	yes	no
Choanal atresia	bilateral	left	no	no	no	no	no	bilateral	no	no	right	no	no
Dysplastic ears	_	low set ears	asymmetric low set ears; overfolded helices.	slightly anteverted	no	small	small	low set ears; overfolded helices.	low set; overfolded helices.	low set; overfolded helices.	preauricular right pit, mild conductive hearing loss bilaterally.	_	_
Wide nasal bridge	yes	yes	yes	no	no	yes	yes	yes	yes	yes	yes	yes	yes
Broad nasal tip	yes	no	yes	no	no	no	yes	yes	yes	yes	yes	no	no
Midface hypoplasia	yes	yes	yes	no	no	yes	yes	yes	no	no	yes	yes	yes
Mandibular prognathism	yes	no	yes	no	no	no	yes	yes	no	no	yes	no	yes
Brachycephaly	no	brachycephaly	no	plagiocephaly & brachycephaly	no	no	no	no	no	no	brachycephaly	no	no
Eyes and eyelids													
	eyelid synechiae bilaterally; distichiasis & mild ectropion.	ankyloblepharon bilaterally; distichiasis & hypertopia.	narrow upslanted palpebral fissures, hooded eyelids and telecanthus; no ophthalmic concerns.	fissures, hooded eyelids and medially deficient eyelashes (everted upper and lower eyelids laterally).	wide palpebral fissues with scleral show, hypertelerism, ectropion, distichiasis.	sleeps with eyes open; ectropion- everted outer third of the lower eyelids.	narrow palpebral fissures, ankyloblepharon, sleeps withteyes open, watery eyes, distichiasis.	narrow upslanted palpebral fissures, hooded eyelids; had a squint. No ophthalmic concerns.	narrow palpebral fissures, hooded eyelids; bilateral eyelid tags at birth, ?ankyloblepharon, telecanthus.	narrow upslanted palpebral fissures, hooded eyelids; telecanthus.	fissures, hooded eyelid and telecanthus. Hooded normal distinction, left side complete, right side partial. No other ophthalmic concerns.	narrow upslanted palpebral fissures, hooded eyelids and telecanthus. No ophthalmic concerns.	narrow upslanted palpebral fissures, hooded eyelids and telecanthus. No ophthalmic concerns.
Highly arched eyebrows	yes	yes	no	no	no	yes & bushy eyebrows	yes & interrupted eyebrows	yes	no	no	yes	yes	yes
Thin lateral eyebrows	yes	no	no	no	yes	yes	yes	yes	yes	yes	yes	no	no
Dental/oral													
Hypodontia	yes, 9 permanent teeth	yes, 5 primary & permanent; other premolars could not be assessed at this age.	yes, 12 permanent teeth	yes (number not determined)	yes, 3 permanent teeth	-	-	yes, 8 permanent teeth	? missing back molars, ND	? missing back molars, ND	yes, 3 permanent teeth	no	yes, 3 permanent teeth
Delayed dentition	yes; supereruption of 63 with apical resorption from 24. Delayed development 34 & 44 with unerupted submerged and resorbed 84.	yes, impacted 63 and delayed eruption of the second primary molars	_	yes	yes	-	-	ND	ND	ND	yes, ectopic 23	no	delayed development of 21
Abnormal crown form	yes; mesoangular eruption of 24, fusion of teeth, long cone-shaped tooth.	yes; macrodontia of 63 & 75; supernumerary tooth in 64-65 area; tooth 51 is dysmorphic and diminutive and tooth 73 is dysmorphic and rotated.	yes; peg 22	no	yes, multiple conical/peg teeth.	-	disorganised teeth	yes ;peg 21 , 22 and history of neonatal tooth.	'pointy teeth'	ND	yes; fissured incisors	yes; diminutive 22	no

Participant	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13	TOTAL
Limbs														
Hands	NAD	NAD	single transverse palmar crease, right hand; thumbs and fifth finger slightly shorter.	NAD	NAD	small hands; short fifth finger.	small hands; short fifth finger.	single transverse palmar crease, right hand.	NAD	NAD	mild syndactyly and camptodactyly	slightly shorter fifth finger	slightly shorter fifth finger	**(7/13)**
Feet	NAD	slight 2, 3-syndactyly of feet bilaterally, long halluces.	slight 2,3-syndactyly on feet, lateral deviated halluces.	pes planus, hallux valgus.	NAD	4th toe longer than others, schilles stretched as a child; 2,3-syndactyly on right foot.	short toes	NAD	NAD	NAD	NAD	sandal gap, camptodactyly 2nd toe.	sandal gap, camptodactyly 2nd toe.	**(7/13)**
Systemic Phenotypes					-									
Cardiac	hypoplastic aortic arch, ASd, VSd, MVS.	VSd, PDA, PFO.	NAD	VSd, secundum ASd.	NAD	NAD	NAD	TOFt - repair done at 4 months, pulmonary regurgitation - valve replacement at 9 years.	NAD	NAD	hypoplastic aortic arch with severe CoA; perimembranous VSd; surgeries at1 month & 2 years.	VSd	NAD	**(6/13)**
Neurologic	NAD	sensory processing disorder. Query autism & ADHD, GDD; aggressive behaviour.	speech delay early life, autism & ADHD, LD, defiant behaviour.	autistic traits, poor coordination.	NAD	NAD	NAD	NAD	incresingly challenging behaviour; speech and language delay; developmental concern re:autistic traits.	increasingly challenging behaviour; speech and language delay; developmental delay.	severe speech and language delay; autism; mild LD; OCD.	mild LD.	mild LD.	**(8/13)**
Voice anomaly	no	nasal speech	no	no	ND	nasal speech	no	no	no	no	stridor; hoarseness	no	no	**(3/13)**
Other skeletal anomalies	severe scoliosis	-	delayed fontanelle closure	joint laxity	-	scoliosis as an adult; short stature	-	-	-	-	short stature; asymetric shoulder; short neck	-	-	**(5/13)**
Cancer	no	no	no	no	no	no	no	ovarian dysgerminoma at 12 years of age stage III in the left ovary; underwent left oophorectomy followed by chemotherapy.	no	no	no	no	no	**(1/13)**
Others	restrictive lung disease	partial agenesis of the corpus callosum	velopharyngeal insufficiency; early onset puberty; bowel problems infant-toddler years	-	-	hypothyroid	abnormal nasal morphology: had rhinoplasty for a collapsed nose	macroglossia (enlarged tongue)	glue ear	glue ear	cryptorchidism on the left side; multiple grommets	coronal hypospadias (repaired); monozygotic twin of Patient 13	monozygotic twin of Patient 12	_

Appendix 5.For pages 303 & 304. *Abbreviation: DDD ID, Deciphering Developmental Disorders patient identification number; NA, not applicable; NAD, no abnormality detected, ND; not determined because of non-availability; 23, upper left permanent canine; 63, upper left primary canine; 24, upper left permanent first premolar; 34, lower left permanent first premolar; 44, lower right permanent first premolar; 84, lower righ primary first molar; 22, upper left permanent lateral incisor; 21, upper left permanent central incisor; 75, lower left primary second molar; 73, lower left primary canine; 64, upper left primary first molar; 65, upper left primary second molar; VSd, ventricular septal defect; ASd, atrial septal defect; TOF, tetralogy of Fallot; CoA, coarctation of the aorta; MVS, mitral valve stenosis; PDA, patent ductus arteriosus; PFO, patent foramen ovale; ADHD, attention deficit hyperactivity disorder; LD, learning difficulty; ID, intellectual disability; OCD, obsessive compulsive disorder; GDD, global developmental delay.*

Appendix 6 - A

GENE	RVIS (ALL_0.1%)	OEratio	%ExAC RVIS_(0.05%popn)	OEratio-percentile[ExAC]	LoF-FDR[ExAC]	ExAC_v2_RVIS_0.05%popn	Edge Case (%OE-ratio)
BAZ2B	-1.42 (4.06%)	NA	5.18%	41.43%	3.16E-15	-1.0079 (13.0890%)	N (35.4715%)
EFTUD2	-1.15 (6.23%)	NA	2.42%	4.77%	4.04E-06	-1.5949 (5.3568%)	N (10.23%)
DMXL1	-0.82 (11.89%)	NA	3.37%	47.02%	3.09E-21	-1.6454 (5.0147%)	N (41.452%)
ABCA3	-1.48 (3.71%)	NA	6.70%	40.91%	0.01145775	0.4362 (69.1887%)	N (48.6829%)
FBXO11	-0.69 (15.12%)	NA	4.09%	2.58%	9.88E-06	-1.3934 (7.0186%)	N (2.23%)
ABCA2	-4.24 (0.12%)	NA	4.13%	7.45%	4.98E-10	-1.4948 (6.0704%)	N (8.2%)
CSMD1	-7.11 (0.02%)	NA	0.17%	NA	NA	-4.4149 (0.3128%)	N
DIP2C	-3.62 (0.29%)	NA	0.16%	4.22%	2.11E-09	-3.5481 (0.6061%)	N
CELSR1	-2.82 (0.63%)	NA	1.54%	35.18%	2.99E-10	-0.3669 (32.8739%)	N
VPS13D	-4.23 (0.14%)	NA	0.08%	43.15%	7.93E-21	-5.7674 (0.1075%)	N
WDR87	4.14 (99.69%)	NA	NA	NA	NA	NA (NA)	NA
WDFY4	4.26 (99.72%)	NA	NA	NA	NA	NA (NA)	NA
GNB2	-0.49 (22.36%)	NA	13.70%	0.52%	0.008431649	-0.6845 (21.4272%)	N 0.7526%
PRR12	NA (NA)	NA	1.32%	12.62%	9.29E-08	-1.5834 (5.4545%)	N 7.5494%
EZR	-1.2 (5.79%)	NA	7.71%	12.48%	0.000108108	-0.9138 (15.2688%)	N 12.1924%
PREP	-1 (8.47%)	NA	6.55%	12.38%	0.000112746	-1.2157 (9.4819%)	N 18.8329%
UPF2	-0.89 (10.46%)	NA	1.64%	10.43%	1.84E-09	-1.8296 (3.8612%)	N 10.0156%
CYP27C1	-0.05 (50.22%)	NA	10.99%	57.40%	0.4823498	-0.6465 (22.5415%)	N 41.5562%
Tolerant Genes							
ATP7B	-0.34 (30.38%)	NA	80.98%	84.45%	0.8485588	0.1038 (54.5161%)	N (85.6829%)
PSG7	NA (NA)	NA	NA	NA	NA	2.2907 (97.9765%)	N
CFAP43	-1.46 (3.76%)	NA	73.65%	65.01%	0.01728628	0.5205 (72.5122%)	N (58.47%)
ZAN	NA (NA)	NA	NA	74.44%	0.8389666	7.0767 (99.8827%)	N (98.24)%
FNDC7	1.38 (94.6%)	NA	92.29%	83.80%	0.1651128	1.3021 (91.4467%)	N (63.0406)%
RBMXL3	4.93 (99.81%)	NA	NA	NA	NA	NA (NA)	NA
INSC	0.89 (89.29%)	NA	95.05%	70.45%	0.6405145	1.7901 (95.8553%)	N 61.88%
TTN	2.17 (98.04%)	NA	99.50%	70.55%	1.60E-123	-2.8532 (1.1926%)	N 60.63%
FCGBP	NA (NA)	NA	99.86%	NA	NA	NA (NA)	NA
OBSCN	NA (NA)	NA	99.98%	78.67%	0.393318	16.3908 (99.9804%)	N 74.57%
SNAPC4	0.32 (72.76%)	NA	98.89%	47.65%	0.3246349	2.0101 (97.0283%)	N 46.37%
KRI1	1.54 (95.59%)	NA	NA	87.56%	0.3333185	0.4861 (71.2121%)	N
ZNF117	0.73 (86.17%)	NA	81.65%	67.11%	0.1255513	0.7351 (79.9120%)	N
SPATA20	-0.59 (18.26%)	NA	56.68%	69.79%	0.07105105	0.2207 (59.9804%)	N
ANKRD30A	2.7 (98.9%)	NA	89.29%	85.80%	0.2100735	1.8391 (96.0704%)	N
MIDN	-1.26 (5.26%)	NA	28.12%	16.12%	0.1334073	0.2236 (60.1173%)	N
TRPM5	-0.84 (11.29%)	NA	89.40%	43.84%	0.7216876	2.9873 (99.0127%)	N
ARHGAP24	-0.62 (17.45%)	NA	81.03%	53.31%	0.002421102	0.4518 (69.9120%)	N 45.8172%
ATP6V1C2	-0.02 (52.09%)	NA	76.31%	53.94%	0.444381	0.3506 (65.6305%)	N 34.2789%
COX7A2L	-0.23 (36.86%)	Y%	58.43%	98.56%	0.7421918	0.1494 (56.7253%)	Y 97.4527%
ZNF646	0.23 (68.55%)	NA	92.26%	46.48%	0.000147622	0.4560 (70.0880%)	N 31.8648%
STON1	1.61 (95.91%)	NA	81.53%	92.38%	0.0473848	0.0659 (52.6588%)	N 88.5718%
ANGPTL7	0.17 (65.76%)	NA	57.48%	58.08%	0.8602305	0.2011 (59.1300%)	N 55.3407%
PHF21B	0.89 (89.24%)	NA	69.86%	60.87%	0.00393352	0.2521 (61.3881%)	N 70.8852%
TSN	-0.08 (47.79%)	Y%	30.85%	2.79%	0.05597885	-0.2994 (35.7771%)	Y 7.62%

Gene	GDI_Score	GDI_Phred	GDI_Damage_Prediction	Selective_pressur(McDonald-Kr	Selective_pressure_prediction		
BZ2B	NA	NA	NA				
EFTUD2	20.07609	0.68608	Medium	0.00423	Moderate_purifying		
DMXL1	2325.86474	8.93384	Medium	0.30531	Moderate_purifying		
ABCA3	188.10316	2.97844	Medium	0.04064	Moderate_purifying		
FBXO11	113.87494	2.32633	Medium	0.37209	Moderate_purifying		
ATP7B	2480.95753	9.28654	Medium	2.95787	Strong_positive		
PSG7	NA	NA	NA	NA	NA		
CFAP43	NA	NA	NA	NA	NA		
ZAN	NA	NA	NA	NA	NA		
FNDC7	1673.11244	7.54892	Medium	0.20292	Moderate_purifying		
RBMXL3	785.37004	5.53655	Medium	0.4223	Moderate_purifying		
INSC	1315.35282	6.81837	Medium	0.25238	Moderate_purifying		
TTN	74772.86558	42.91324	High	0.47058	Moderate_purifying		
FCGBP	18359.5916	28.93384	High	0.42199	Moderate_purifying		
OBSCN	33482.92348	35.92354	High	0.29425	Moderate_purifying		
SNAPC4	4352.03325	13.15433	Medium	0.07841	Moderate_purifying		
ABCA2	470.404	4.488	Medium	0.088	Moderate_purifying		
CSMD1	1048.097	6.215	Medium	0.04	Moderate_purifying		
KRI1	6629.495	17.208	High	7.097	Strong_positive		
ZNF117	5466.629	15.234	High	13.718	Strong_positive		
DIP2C	69.013	1.721	Medium	0.003	Strong_purifying		
SPATA20	4840.851	14.122	High	0.57	Moderate_purifying		
ANKRD30A	3097.854	10.583	Medium	0.303	Moderate_purifying		
MIDN	155.683	2.727	Medium	0.011	Moderate_purifying		
CELSR1	7612.061	18.73	High	0.341	Moderate_purifying		
VPS13D	2635.305	9.634	Medium	0.284	Moderate_purifying		
WDR87	2359.543	9.016	Medium	0.302	Moderate_purifying		
WDFY4	5281.755	15.01	High	0.433	Moderate_purifying		
TRPM5	997.38	6.086	Medium	0.104	Moderate_purifying		
GNB2	28.358	0.908	Medium	0.107	Moderate_purifying		
PRR12	335.021	3.889	Medium	0.017	Moderate_purifying		
EZR	62.542	1.616	Medium	0.12	Moderate_purifying		
PREP	3956.75	12.444	Medium	0.213	Moderate_purifying		
UPF2	286.423	3.625	Medium	0.057	Moderate_purifying		
CYP27C1	93.008	2.072	Medium	0.016	Moderate_purifying		
ARHGAP24	242.679	3.363	Medium	0.023	Moderate_purifying		
ATP6V1C2	2899.989	10.206	Medium	0.678	Moderate_purifying		
COX7A2L	44.469	1.275	Medium	0.41	Moderate_purifying		
ZNF646	5474.894	15.249	High	0.411	Moderate_purifying		
STON1	39.12	1.16	Medium	0.23	Moderate_purifying		
ANGPTL7	419.965	4.281	Medium	0.174	Moderate_purifying		
PHF21B	708.321	5.286	Medium	0.079	Moderate_purifying		
TSN	6.363	0.238	Medium	0.955	Moderate_purifying		

Appendix 6-A. RVIS and GDIS scores for all 43 genes examined.

Appendix 6 - B

Cranio-cardiac genes with high level of constraint and low tolerance to variation

EFTUD2

Constraint from ExAC	Expected no. variants	Observed no. variants	Constraint Metric
Synonymous	155.1	162	z = -0.34
Missense	343.2	178	z = 4.36
LoF	36.9	1	pLI = 1.00
CNV	19.3	0	z = 2.44

FBXO11

Constraint from ExAC	Expected no. variants	Observed no. variants	Constraint Metric
Synonymous	90.7	125	z = -2.23
Missense	250.4	120	z = 4.03
LoF	34.6	1	pLI = 1.00
CNV	8.2	6	z = 0.28

CELSR1

Constraint from ExAC	Expected no. variants	Observed no. variants	Constraint Metric
Synonymous	837.3	753	z = 1.81
Missense	1534.7	1181	z = 4.42
LoF	71.9	9	pLI = 1.00
CNV	7.3	6	z = 0.18

ABCA2

Constraint from ExAC	Expected no. variants	Observed no. variants	Constraint Metric
Synonymous	531.7	615	z = -2.24
Missense	1008.5	660	z = 5.37
LoF	66.6	6	pLI = 1.00
CNV	9.3	5	z = 0.56

DIP2C

Constraint from ExAC	Expected no. variants	Observed no. variants	Constraint Metric
Synonymous	368.3	405	z = -1.19
Missense	688.4	376	z = 5.62
LoF	65.1	5	pLI = 1.00
CNV	16.9	26	z = -0.56

Appendix 6-B. ExAC scores for the five chosen candidate genes that met the filtering criteria. *Missense Z scores (second row in each table) indicate that the higher the Z score, the more intolerant the transcript is to variation (more constrained). pLI scores (probability of being loss-of-function intolerant) (third row in each table) closer to one indicate more intolerance to protein-truncating variation (http://exac.broadinstitute.org & https://gnomad.broadinstitute.org).*

Appendix 6 – C

#	DDD ID	Sex	Genomic location	Mutation	Protein variant	Type	Identified pathogenicity /Contribution	Inheritance	Exac
1	292486	F	17:42960506-42960507, TA>T	c.446del	p.Leu149*	frameshift	Likely pathogenic	De novo Het	0
2	281570 +CAP180	F	17:42956923-42956923, C>T	c.702+1G>A	-	splice donor	Pathogenic	De novo Het	0
3	271922 +CAP180	M	17:42949863-42949864, GC>G	c.944del	p.Ser315Thrfs*71	frameshift	Pathogenic	Maternally inherited Het	0
4	304793 +CAP180	F	17:42937814-42937814, G>A	c.1705C>T	p.Arg569*	stop gained	Likely pathogenic	De novo Het	0
5	295695	F	17:42931646-42931646, A>G	c.2338T>C	p.Cys780Arg	missense	Pathogenic	De novo Het	0
6	263948 +CAP180	M	17:42929088-42929088, C>T	c.2813G>A	p.Arg938His	missense	Likely pathogenic Partial	De novo Het	0
7	266427 +CAP180	M	17:42937272, C/T	c.1860+1G>A	-	splice donor	Pathogenic Full	De novo Het	0
8	280110 CAP180 only	M	17: 42937371, T/TA	c.1763_1764insA	p.Lys589Glnfs*17	frameshift	?	Unknown	0
9	272587 +CAP180	M	17:42957006, C/T	c.620G>A	p.Gly207Glu	missense	Pathogenic Full	De novo Het	0
10	283722 +CAP180	M	17:42930658, GCTCA/G	c.2561+2_2561+5del	-	splice donor	Likely pathogenic Full	De novo Het	0
Copy number variants encompassing EFTUD2									
11	259863	M	17:42920900-42932449	CNV:11.55 kb	-	Deletion	Likely pathogenic	De novo Het	-
12	275934	M	17:42953106-42961328	CNV:8.22 kb	-	Deletion	Pathogenic Full	De novo Het	-

Appendix 6-C. EFTUD2 Variants in chosen DECIPHER and CAP180 Patients.
ENST00000426333, NM_001258353.

Appendix 6 - D

Phenotype	1	2	3	4	5	6	7	8	9	10	11	12
Variant	SVV	SNV	SNV	SNV	SNV	SNV	SNV	SNV	SNV	SNV	CNV	CNV
Sex	F	F	M	F	F	M	M	M	M	M	M	
Cleft	Soft palate	-	-	-	-	Cleft palate; PRS					-	Median cleft palate
Laryngeal/ pharyngeal anomaly	Laryngeal cleft	Esophageal atresia	Tracheoesophageal fistula	Esophageal atresia; Tracheoesophageal fistula	Tracheoesophageal fistula	-	-	Esophageal atresia; Tracheoesophageal fistula	-	-	-	-
Microcephaly	Yes	Yes+ Trigonocephaly	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes+ Progressive
Low set ears	-	Yes	-	-	-	Yes			Yes			
Hearing impairment				conductive	Bilateral sensorineural	Bilateral conductive		Bilateral conductive	sensorineural	sensorineural		conductive
Other ear anomalies	-	Cupped ear; Preauricular skin tag	Abnormality of the pinna; Preauricular skin tag	Preauricular skin tag/	-	Microtia	Microtia	Microtia	-	-	-	Cupped ear; Stenosis of the external auditory canal
Choanal atresia	Yes	-	-	-	Bilateral	-	Bilateral	Yes	-	-	-	
Micrognathia	-	-	-	-	-	Yes	-	Yes	Yes			Yes
Dysmorphic features	Hypertelorism; Low anterior hairline; Low hanging columella	Facial asymmetry	Facial asymmetry; Wide nasal bridge	Facial asymmetry	-	Fine hair	Facial symmetry, carious teeth	Abnormal facial shape Long upper eyelashes, Sparse lower eyelashes	Large beaked nose, Malar flattening, Prominent forehead; Short neck	Hemifacial hypoplasia	-	Smooth philtrum
Global develop. delay	Mild	Mild				Yes		Yes	Moderate	Mild	Yes	Moderate

Other neurodevelop.	-	Cerebral white matter agenesis	-	-	Cognitive impairment	-	Mild ID	-	-			
Congenital heart disease	ASD	Peri-membranous VSD	Total anomalous pulmonary venous return	Right aortic arch (?)	Total anomalous pulmonary venous return	Pulmonic stenosis; VSD	-	Bradycardia	-	Internal carotid artery hypoplasia	ASD; Hypoplasia of right ventricle	Partial anomalous pulmonary venous return; Secundum ASD
Skeletal/Limb	Broad thumb	-	Narrow foot	-	Thoracic scoliosis	Talipes equinovarus	Joint hypermobility	2-3 toe cutaneous syndactyly, Aplasia/Hypoplasia of the thumb, Clinodactyly of the 5th finger, Hypoplastic toenails, Sacral dimple Toe clinodactyly	-	-	Abnormality of the epiphyses of the distal phalanx of finger	Bilateral single transverse palmar creases; Joint laxity
Others	Otitis media; Sleep disturbance	-	Polyhydramnios (?)	-	Generalized-onset seizure	-	-	Premature birth, Recurrent respiratory infections	-	Few cafe-au-lait spots	Strabismus	Cryptorchidism; Micropenis

Appendix 6-D. For pages 308 & 309. Clinical description of probands carrying de novo mutations in EFTUD2.
Abbreviations: PRS, Pierre Robin Sequence; ID, intellectual disability; ASD, atrial septal defect; VSD, ventricular septal defect.

Appendix 6 – E

HPO Phenotype	N of affected patients with FBXO11 mutations
Abnormality of head or neck	8/8
Abnormality of the eye	5/8
Abnormality of the ear	3/8
Abnormality of the cardiovascular system	2/8
Abnormality of the skeletal system	7/8
Abnormality of the musculature	3/8
Growth abnormality	2/8
Abnormality of the nervous system	8/8
Abnormality of abdomen morphology	2/8
Abnormality of the integument	2/8

Appendix 6-E. HPO terms and phenotypes in DECIPHER FBXO11 DDD Research Variants.

Appendix 6 - F

Phenotype	**268635**	**283762**	**303291**
	F	M	M
Cleft	CLEFT palate	-	-
Skull	Plagiocephaly	-	-
Ear dysmorphology	Macrotia	-	-
Dysmorphic features	Abnormal facial shape; Abnormality of the incisor; Short philtrum; Thin vermilion border; Synophrys	Facial asymmetry; Strabismus	Abnormal facial shape
Neurodevelopmental/ Brain	Delayed speech and language development, Global developmental delay	Cognitive impairment; Delayed gross motor development; Poor speech	Global developmental delay
CHD	Ventricular septal defect	-	-
Skeletal/Muscles/Limbs	Pectus carinatum	Muscular hypotonia; Scoliosis; Tall stature; Gait imbalance	-
Other	-	-	-

Appendix 6-F. Clinical description of CAP180 probands carrying de novo mutations in previously un-reported FBXO11 variants.

Appendix 6 – G

#	DDD ID	Sex	Genomic location	Mutation	Protein variant	Type	Inheritance	Exac count
1	DDD RV	?	22:46805021-46805021, C/T	c.5098G>A	p.Val1700Met	missense	De novo	1
2	279464 CAP180	M	22:46832074, C/CGG	?	p.Ala1507Profs*17	frameshift	Unknown Single variant, compound-het	346
	279464	M	22:46832075, A/AGAAGGCCCCACCTGCG	c.4521_4522insGAAGGCCCCACCTGCG	p.Gly1508Glufs*22	frameshift	Unknown Single variant, compound-het	Not on exac although cap180 seems to think freq. is the same as above
3	285856 CAP180	F	22:46832074, C/CGG	?	p.Ala1507Profs*17	frameshift	Unknown Single variant, compound-het	346
	285856	F	22:46832075, A/AGAAGGCCCCACCTGCG	c.4521_4522insGAAGGCCCCACCTGCG	p.Gly1508Glufs*22	frameshift	Unknown Single variant, compound-het	novel
4	290935 CAP180	F	22:46832074, C/CGG	?	p.Ala1507Profs*17	frameshift	Unknown Single variant, compound-het	346
	290935	F	22:46832075, A/AGAAGGCCCCACCTGCG	c.4521_4522insGAAGGCCCCACCTGCG	p.Gly1508Glufs*22	frameshift	Unknown Single variant, compound-het	novel
5	260762 CAP180	F	22:46835249, C/T	c.4243G>A	p.Gly1415Arg	missense	Unknown Monoallelic	1
6	266177 CAP180	M	22:46773155, C/T	c.7387G>A	p.Gly2463Arg	missense	Maternally Inherited Comp Het	15
	266177	M	22: 46931405, C/G	c.1663G>C	p.Val555Leu	missense	Paternally Inherited Comp Het	8
7	274753 CAP180	M	22: 46829361, C/T	c.4540G>A	p.Val1514Met	regulatory region/missense	Unknown Monoallelic	2
8	300750 CAP180	F	22:46777753, G/A	c.7078C>T	p.Arg2360Cys	missense	Paternally inherited Comp Het.	1
	300750	F	22:46932370, G/A	c.698C>T	p.Ala233Val	regulatory region/missense	Maternally Inherited Comp Het.	Novel
9	93958 CAP180	M	22:46777756, G/A	c.7075C>T	Arg2359Cys	missense	Unknown Monoallelic	Novel
10	261208 CAP180	M	22:46860081, A/ACAGCACGGCGGC	?	?	inframe_insertion	Unknown Monoallelic	Novel

Appendix 6-G. Variants in DECIPHER and CAP180 *with mutations in* ***CELSR1****. Note, each colour in the cells of the first column signifies a patient. Note the presence of multiple compound heterozygous patients. NM_014246.1.*

Appendix 6 - H

Patient	DDD RV	279464	285856	290935	260762	266177	274753	300750	93958	261208
Sex		M	F	F	F	M	M	F	M	M
Cleft	?	CLEFT soft palate	-	-	-	-	-	-	-	(mother has a submucous cleft hard palate)
Orofacial phenotypes	Abnormality of head or neck Abnormality of the ear; Abnormality of the eye	Facial asymmetry. Low posterior hairline, Malar flattening, Periauricular skin pits	Abnormal facial shape, Deeply set eye	Abnormal facial shape	Abnormal facial shape	Depressed nasal ridge, Hypertelorism, Low-set ears, Strabismus, Widow's peak	Abnormal facial shape, Widely spaced primary teeth	-	Childhood onset sensorineural hearing impairment, Low hanging columella, Underdeveloped nasal alae. Malar flattening, Microcephaly, Eyes: Glaucoma, Hypoplasia of the iris, Posterior embryotoxon.	Epicanthus, Everted lower lip vermilion, Hypertelorism, Malar flattening, Open mouth
Mandible	-	-	Mandibular prognathia	-	-	Abnormality of the mandible	-	Micrognathia	-	
CHD	Abnormality of the cardiovascular system	Peripheral pulmonary artery stenosis. Primum atrial septal defect		Patent foramen ovale			Mitral regurgitation		-	
Limbs	Abnormality of limbs			Broad hallux		Abnormality of the femur, Abnormality of the humerus. Clinodactyly of the 5th finger. Short phalanx of finger, Short toe	Broad hallux		Tapered finger	

GIT	Abnormality of abdomen morphology	Inguinal hernia			Feeding difficulties in infancy, Gastroesophageal reflux			
Skeletal growth Muscles	Abnormality of the skeletal system; Growth abnormality				Abnormality of the skeletal system	Infantile muscular hypotonia		
Genitourinary	Abnormality of the genitourinary system	Hypospadias, Renal agenesis						
Others	Abnormality of prenatal development or birth, Abnormality of the immune system	Esophageal duplication		Melanocytic nevus		Stridor, NG fed for duration in NICU. Recruited as a baby (6m), not many phentoypes could be observed	Prominent veins on trunk	Lacticaciduria

Appendix 6-H For pages 312 & 313. Clinical description of CAP180 probands carrying mutations in CELSR1.

Appendix 6 - I

Patient	279464	285856	290935	260762	266177	274753	300750	93958
Previous clinical differential diagnoses		Angelman syndrome		Pitt-Hopkins	Aarskog -scott syndrome	Fragile x; Cardiofaciocutaneous;	SMA, Myotonic Dystrophy, PWS	SHORT syndrome
Previous negative genetic tests	CHD7, SOX2, SALL1; 22q11.2, Fanconi anaemia	UBE3A	CREBBP, FRAXA, MLPA P245		karyotype,FGD1 gene	FMR1;HRAS; KRAS; BRAF; MAP2K1/MEK1; MAP2K2/MEK2; SHOC2		PITX2, FOXC1, PAX6

Appendix 6-I. Previous genetic tests carried out for probands carrying mutations in CELSR1

. Appendix 6 – J

N	DDD ID	Sex	Genomic location	Mutation	Protein variant	Type	Identified pathogenicity	Inheritance	Exac count		
Single nucleotide variants in DIP2C											
1	DDD RV1	?	10:403824-403824, C/T	c.2848G>A	p.Ala950Thr	missense	Uncertain	De novo Het.	Novel		
2	DDD RV2	?	10:530768-530768, C/G	c.157+1631G>C	?	missense	Uncertain	De novo Het.	Novel (there's a C/T mut in this region, n=4)		
3	CAP180 264560	F	10:518415, G/A	c.232C>T	p.Arg78Trp	missense	(?) likely pathogenic	Paternally inherited	2		
4	CAP180 263415	F	10:370900, G/C	c.3924+2046C>G	?	stop-gained	?	Unknown Monoallelic	Novel		
Chromosomal copy number variants encompassing DIP2C										DIP2C	ZMYND11
5	249415	F	10:631589-866586	CNV: 235.00 kb	-	deletion	Unknown	De novo	-	+	No
6	272754	M	10:138680-375095	CNV: 236.42 kb	-	deletion	Unknown	Unknown	-	+	+
7	285910	M	10:148006-470372	CNV: 322.37 kb	-	deletion	Likely pathogenic	Unknown	-	+	+
8	248531	Other	10:226083-596534	CNV: 370.45 kb	-	duplication	Unknown	Unknown	-	+	+
9	278831	Other	10:278544-680960	CNV: 402.42 kb	-	Triplication	Uncertain	Maternally inherited	-	+	+
10	270190	M	10:299304-740247	CNV: 440.94 kb	-	deletion	Unknown	De novo	-	+	+
11	1232	F	10:269607-1380732	CNV: 1.11 Mb	-	deletion	Unknown	Unknown	-	+	+
12	2319	Other	10:136361-1758581	CNV: 1.62 Mb	-	deletion	Unknown	Unknown	-	+	+
13	318601	F	10:136391-1818132	CNV: 1.68 Mb	-	deletion	Pathogenic	De novo	-	+	+
14	250462	F	10:125544-3637331	CNV: 3.51 Mb	-	deletion	Unknown	De novo	-	+	+
15	290001	Unknown	10:116829-3822976	CNV: 3.71 Mb	-	deletion	Pathogenic	Unknown	-	+	+
16	257495	F	10:148206-4396320	CNV: 4.25 Mb	-	deletion	Unknown	De novo	-	+	+
17	330942	F	10:208454-7075724	CNV: 6.87 Mb		duplication	Pathogenic Full	Maternally inherited. Het.	-	+	+
18	250441	F	10:136361-10421102	CNV: 10.28 Mb	-	deletion	Unknown	Unknown	-	+	+

Appendix 6-J. DIP2C single and copy number variants in DECIPHER and CAP180 probands. *Note, cells shaded in pink; Patients 1-4 have single nucleotide variants in DIP2C and Patient 5 has a copy number variant encompassing DIP2C but not ZMYND11. NM_014974.2*

Appendix 6 – K

Patient	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Variant	SNV	SNV	SNV	SNV	CNV	CNV	CNV	CNV	CNV	CNV	CNV	CNV	CNV	CNV	CNV	CNV	CNV	CNV
Sex	-	-	F	F	F	M	M	O	O	M	F	O	F	F	U	F	F	F
Cleft			+														+	+
Skull				+ma	+ mi		+ma					+ma				+mr		+mi
Face/H&N	+	+		+		+	+			+					+	+		+
Ears		+	+		+	+						+				+	+	+
Prominent Nose			+	+		+				+						+		+
Eyelids			+		+	+											+	+
Eyes	+	+	+		+												+	+
Hypertelorism			+										+			+		
Lips												+						+
Mouth			+													+		+
Micrognathia			+		+													
CHD		+	+										+				+	
CNS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
Brain structural abn.			+	+											+			
Seizures				+			+				+				+	+		
Muscular hypotonia				+			+							+		+		
Skeletal sys.	+					+		+			+	+				+		+
Short/Growth																		
Limbs	+		+		+								+				+	+
Kidney/Renal					+								+					+
GIT		+	+								+		+	+		+		

Appendix 6-K Clinical description of CAP180 and DECIPHER probands carrying mutations or CNVs in DIP2C. *Note, cells shaded in pink; Patients 1-4 have single nucleotide variants in DIP2C and Patient 5 has a copy number variant encompassing DIP2C but not ZMYND11. Abbreviations: SNV, single number variant; CNV, copy number variant; F, female; M, male; O, other; U, unknow; Mouth, 'Abnormality of the Mouth' based on HPO phenotypes; GIT, 'Abnormality of the Stomach' based on HPO phenotypes; RS, abnormality of the respiratory system; GU, abnormality of the genitourinary system, HC, hypocalcaemia; mi, Microcephaly, ma, Macrocephaly, mr, prominent metopic ridge; AN, absent nipple. Note: only general HPO phenotyping is present for Subject 1,2&9 and the phenotypes are probably not fully documented for many of the patients reported*

Appendix 6 – L

Patient	Detailed Phenotypes
Pt1 **DDD RV1**	Abnormality of head or neck, Abnormality of the eye, Abnormality of the integument, Abnormality of the nervous system, Abnormality of the skeletal system,, Abnormality of limbs.
Pt2 **DDD RV2**	Abnormality of head or neck, Abnormality of the cardiovascular system, Abnormality of the ear, Abnormality of the eye, Abnormality of the nervous system, Abnormality of abdomen morphology, Abnormality of the respiratory system.
Pt3 **CAP180 patient: DDDP110191-264560**	Abnormality of the mouth, Lobulated tongue, Micrognathia, Bilateral microphthalmos, Downslanted palpebral fissures, Hypertelorism, Prominent nasal bridge, Cleft palate, Overfolding of the superior helices, Thickened ears, Uplifted earlobe, Pulmonary hypoplasia, Right aortic arch with mirror image branching Hypoplasia of the corpus callosum, Abnormal cortical gyration, Abnormality of the cerebellar vermis, Ventriculomegaly, Hand clenching, Abnormality of the stomach, Abnormality of the uterus, Polyhydramnios.
Pt4 **CAP180 263415**	Abnormal facial shape, Abnormality of prenatal development or birth, Abnormality of the cerebral ventricles, Aggressive behaviour, Constipation, Frontal bossing, Generalized neonatal hypotonia, Global developmental delay, Hypoplasia of the corpus callosum, Macrocephaly, Recurrent upper respiratory tract infections, Seizures, Sensorineural hearing impairment, Sleep disturbance, Specific learning disability, Wide nasal bridge
Pt5 **249415**	Epicanthus, Low-set ears, Posteriorly rotated ears, Microcephaly, Micrognathia, Hydronephrosis, Multiple renal cysts, Ureteral duplication, Hypocalcaemia, Intellectual disability, Broad thumb.
Pt6 **272754**	Broad neck, Downslanted palpebral fissures, Low-set ears, Prominent nose, Cognitive impairment, Short stature.
Pt7 **285910**	Abnormal facial shape, Macrocephaly, Global developmental delay, Infantile axial hypotonia, Lumbar hypertrichosis, Sacral hypertrichosis, Seizures.
Pt8 **248531**	Autism, Hyperactivity, Hyperextensible skin, Joint laxity.
Pt9 **278831**	Attention deficit hyperactivity disorder, Mild global developmental delay.
Pt10 **270190**	Abnormality of the face, Abnormality of the pinna, Intellectual disability.
Pt11 **1232**	Generalized tonic seizures, Intellectual disability, Feeding difficulties in infancy, Obesity, Proportionate short stature, Small for gestational age.
Pt12 **2319**	Macrocephaly, Thick lower lip vermilion, Thickened ears, Synophrys, Absent nipple, Cafe-au-lait spot, Autism, Delayed speech and language development, Intellectual disability, Stereotypy, Short stature.
Pt13 **318601**	Hypertelorism, Bicuspid aortic valve, Pulmonic stenosis, Ventricular septal defect, Chronic kidney disease, Hydronephrosis, Multicystic kidney dysplasia, Gastrostomy tube feeding in infancy, Moderate global developmental delay, , Severe global developmental delay.
Pt14 **250462**	Intellectual disability, Muscular hypotonia, Truncal obesity.
Pt15 **90001**	Abnormal facial shape, Epileptic spasms, severe Pachygyria, Seizures, Tetraplegia, Delayed speech and language development Global developmental delay, Intellectual disability.
Pt16 **257495**	Depressed nasal ridge, Narrow mouth, Prominent metopic ridge, Round face, Low-set ears, Thickened ears, Generalized tonic-clonic seizures, Global developmental delay, Hypertelorism, Muscular hypotonia, Feeding difficulties in infancy, Intrauterine growth retardation, Postnatal growth retardation.
Pt17 **330942**	Cleft palate, Blepharophimosis, Microphthalmia, Low-set ears, Congestive heart failure, Talipes equinovarus, Camptodactyly.
Pt18 **250441**	Submucous cleft hard palate, Broad forehead, Broad nasal tip, Anteverted nares, Downturned corners of mouth, Thin lower lip vermilion, Thin upper lip vermilion, Epicanthus, Ptosis, Malar flattening, Microcephaly, Low-set ears, Hearing impairment, 2-3 toe syndactyly, Absence seizures, Intellectual disability, Multiple renal cysts, Muscular hypotonia, Recurrent infections, Short stature.

Appendix 6-L Detailed clinical description of CAP180 and DECIPHER probands carrying mutations or CNVs in DIP2C. *Note, cells shaded in pink; Patients 1-4 have single nucleotide variants in DIP2C and Patient 5 has a copy number variant encompassing DIP2C but not ZMYND11. Phenotype Color Code: Palate, heart, ears, seizures, muscular hypotonia, renal, small/short stature/growth, neurodevelopmental, limbs, eyes, eyelid abnormality, GIT. Face: broad nose, hypertelorism, micrognathia, lips thin/thick, skull.*

Appendix 6 – M

N	DDD ID	Sex	Genomic location	Mutation	Protein variant	Type	Identified pathogenicity	Inheritance	+Exac count
1	DDD RV1	-	9:139903856-139903856, G/A	c.6793C>T	p.Arg2265Trp	missense	Uncertain	De novo Het.	2
2	DDD RV2	-	9:139905048-139905048, C/T	c.6174+24G>A ??	p.Trp1416*†	Likely LOF. Stop gained	Uncertain	Biparental Homozygous	allele count 52, but n of homozygous =1
3	DDD OA 359554	M	9:139912529-139912529, T>C	c.1994-2A>G	-	splice acceptor	Uncertain	De novo Het.	Novel
4	CAP180 284294	F	9:139903861, C/T	c.6788G>A	p.Arg2263His	missense	?	1 copy maternally Inherited. Biallelic	6
5	CAP180 294080	M	9:139903402, C/T	c.6930+1G>A	-	splice donor	?	Unknown Monoallelic	Novel
6	CAP180 259046	F	9:139903856, G/A	c.6793C>T	p.Arg2265Trp	missense	?	De novo Monoallelic	2
7	CAP180 300126	M	9:139907925, C/T	c.4538G>A	p.Arg1513His	regulatory region, missense	?	Unknown Monoallelic	1
8	272341	M	9:139904018, G/A	c.6709C>T	p.Arg2237Cys	missense	?	Unknown Monoallelic	Novel
9	267924	M	9:139903856, G/A	c.6793C>T	p.Arg2265Trp	missense	?	Unknown Monoallelic	2
Chromosomal copy number variants encompassing ABCA2									
10	278258	M	9:139665148-141018984	CNV: 1.35 Mb	-	Duplication	Pathogenic	De novo	-
11	282032	M	9:139885077-140103893	CNV: 218.82 kb	-	Duplication	Likely pathogenic	Unknown	-
12	266359	M	9:139863777-140488857	CNV: 625.08 kb	-	Deletion	Unknown	De novo	-

Appendix 6-M. ABCA2 single and copy number variants in DECIPHER and CAP180. *† annotation is for non-canonical transcript. ENST00000341511, NM_001606.5 (used by DECIPHER).*

Appendix 6 – N

Patient	DDD RV1	DDD RV2	DDD OA	284294	294080	259046	300126	272341	267924	278258	282032	266359
Variant	SNV	SNV	SNV	SNV	SNV	SNV	SNV	SNV	SNV	CNV	CNV	CNV
Sex	-	-	M	F	M	F	M	M	M	M	M	M
Cleft					High palate	Submucous cleft hard palate						
Orofacial/ H&N	+	+		Round face	Abnormal facial shape; High anterior hairline	Facial asymmetry; Laryngomalacia; Tracheomalacia	Bilateral choanal atresia/stenosis; Long face;; Malar flattening; Mandibular prognathia; Prominent nasal bridge; Short philtrum	Coarse facial features. Micrognathia, Retrognathia	Abnormal facial shape	Anteverted nares, Hypertelorism; Micrognathia, Narrow mouth; Short neck, Thin upper lip vermilion, Wide nasal bridge		
Eyes	+					Corneal dystrophy	Long upper eyelashes	Iris hypopigmentation, Proptosis		Strabismus		Strabismus
Ears				Cupped ear	Chronic otitis media		Abnormal location of ears; Cupped ear; Low-set ears; Microtia		Chronic otitis media	External ear malformation; Otitis media	Large fleshy ears; Hearing impairment	
Speech				Absent speech	Expressive language delay		Delayed S & L development				Delayed S&L development	
Skull				Congenital microcephaly		Postnatal microcephaly	Brachycephaly			Plagiocephaly		
CHD		+				Patent ductus arteriosus	Pulmonic stenosis					
Skin		+				Hypopigmented streaks; Increased number of skin folds						
Limbs	+					Congenital talipes calcaneovalgus	2-3 toe syndactyly; Broad thumb;; Single transverse palmar crease	Clinodactyly of the 5th finger	Clinodactyly of the 5th finger			

Limbs	+					Congenital talipes calcaneovalgus	2-3 toe syndactyly; Broad thumb;; |Single transverse palmar crease	Clinodactyly of the 5th finger	Clinodactyly of the 5th finger			
Neruodevelopmental	+	+	+	Severe ID; Stereotypic behavior	Specific learning disability	Bimanual synkinesia; GDD	Broad-based gait; Seizures	Delayed speech and language development	Absence seizures, Global developmental delay	GDD; Impaired social interactions; Restrictive behavior	ASD, Specific learning disability	ID
GIT	+	+				Abnorm. of subcutaneous fat tissue; Anteriorly placed anus; Severe GOR						
Skeletal/ Growth	+	+ +			**Generalized hypotonia.** Generalized joint laxity	Joint hypermobility; Thoracic scoliosis|	Kyphosis; Limited elbow extension; Limited pronation/supination of forearm		**Generalized hypotonia.** *Proportionate short stature|Short stature*	**Generalized neonatal hypotonia**		*Short stature,*
Metabolism / Homeostasis		+			Pinealoma							
Others	CT, RS	Bl, GU, IS		Consanguineous parents	Multicystic kidney dysplasia	VPI; Brain MRI: slightly small pituitary gland			Decreased body weight			

Appendix 6-N For pages 320 & 321. Phenotypes in patients with SNV and CNV in the ABCA2 gene. *Abbreviations: CT, Abnormality of connective tissue'; RS, Abnormality of the respiratory system; Bl, Abnormality of blood and blood-forming tissues; GU, Abnormality of the genitourinary system; Abnormality of the immune system; ID, intellectual disability; GOR, Gastroesophageal reflux; GDD, Global developmental delay; VPI, pharyngeal insufficiency; ASD, Autism spectrum disorder; S&L, Speech and Language.*