

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Kelsey Rebecca Robinson Wallace

Date

Identifying Novel Genetic Causes of Cleft Palate

By

Kelsey Rebecca Robinson Wallace
Doctor of Philosophy

Graduate Division of Biomedical and Biological Sciences
Genetics and Molecular Biology

Elizabeth Leslie
Advisor

David Cutler
Committee Member

Michael Epstein
Committee Member

Judith Fridovich-Keil
Committee Member

Michael Gambello
Committee Member

Accepted:

Kimberly Jacob Arriolla, Ph.D, MPH
Dean of the James T. Laney School of Graduate Studies

Date

Identifying Novel Genetic Causes of Cleft Palate

By

Kelsey Rebecca Robinson Wallace
BS, Clemson University, 2013
DVM, University of Georgia College of Veterinary Medicine, 2016
MSc, University of Georgia College of Veterinary Medicine, 2020

Advisor: Elizabeth J. Leslie, PhD

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate Division of Biomedical and Biological Sciences
Genetics and Molecular Biology
2024

Abstract

Identifying Novel Genetic Causes of Cleft Palate

By: Kelsey Rebecca Robinson Wallace

Cleft palate (CP) occurs in approximately 1 in 1700 live births per year and is one of the most common craniofacial birth defects. Although it is highly heritable, there are few known associated genetic risks. Therefore, we performed a comprehensive investigation in a trio-based cohort to evaluate both common and rare variant risks for CP using whole genome sequencing. We hypothesized there would be risk factors associated with any type of CP, as well as risks specific to CP subtypes (*i.e.*, clefts affecting the hard and/or soft palate), proband sex, and based on the presence or absence of additional clinical features. We performed genome-wide association studies (GWAS) for 435 trios with any type of CP and stratified by subtype, identifying a genome-wide significant locus at 9q33.3 (rs7035976, $p=4.24 \times 10^{-8}$) associated with any cleft of the hard palate. When stratified by proband sex, we found distinct GWAS signals, prompting a genome-wide gene-by-sex interaction analysis. We identified 13 loci with significant interactions. Our top finding was within an intron of *LTBP1* in the 2p33.3 locus (relative risk 3.37 (95% CI 2.04 - 5.56), $p=1.93 \times 10^{-8}$), for which we found a female-specific association between imputed genetically regulated gene expression for *LTBP1* and CP in cultured fibroblasts. We next evaluated *de novo* variants (DNs) and found global enrichment of protein-altering DNs and gene-specific enrichment for several genes, including three not previously associated with CP: *PRKCI*, *POLRIF*, and *SCL25A41*. We found subtype-specific enrichments including *SATB2* and *TGFBR2* in cleft hard and soft palate, and *PRKCI* for cleft soft palate. Moreover, *PRKCI* was specifically enriched in syndromic CP probands. We found identical DN (N383S) in *PRKCI* in two individuals with identical phenotypes—Van der Woude syndrome (VWS) with lip pits and cleft soft palate—and a third individual with a Y136C DN with cleft soft palate and abnormal lip morphology. Functional testing of this variant using zebrafish embryos confirmed loss-of-function, implicating *PRKCI* as a novel gene for VWS. These findings expand our understanding of the genetic basis of CP and related syndromes, emphasizing the importance of considering sex-specific effects and performing subtype-specific analyses in genetic studies of CP.

Identifying Novel Genetic Causes of Cleft Palate

By

Kelsey Rebecca Robinson Wallace

BS, Clemson University, 2013

DVM, University of Georgia College of Veterinary Medicine, 2016

MSc, University of Georgia College of Veterinary Medicine, 2020

Advisor: Elizabeth J. Leslie, PhD

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate Division of Biomedical and Biological Sciences
Genetics and Molecular Biology
2024

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincerest gratitude to my advisor Elizabeth Leslie for her guidance and wisdom throughout this endeavor. With her mentorship, I have developed skills I never even knew existed when starting this journey and have been endlessly supported in my personal and professional goals as a scientist.

I am also grateful for my always supportive committee members, Dave Cutler, Mike Epstein, Michael Gambello, and Judy Fridovich-Keil. However, I am especially thankful for Judy, whose cookie day emails were always a highlight of the week.

To my labmates—particularly Kim and Lizzie—thank you for your friendship, for always dealing with my nonsense, for always listening and being there for me, and for always brightening my days. Pancake Wednesdays (with special shout out to Morgan) resulted in so many great ideas and problem solving!

I would also like to thank my family, my parents Ken and Becky, my sister Kyra, and my wonderful partner Luke, for their unwavering support and encouragement through what I have assured them will be my last degree.

Lastly, but certainly not least, I would not have made it through without all the wonderful canine companionship. Thank you to Tilly, who was the best girl, and to Margot Moonpie, who is also the best girl, for all the serotonin and smiles.

Table of Contents

Chapter I: Introduction	1
Introduction.....	1
Background.....	2
Epidemiology.....	2
Palatogenesis.....	3
Classification.....	5
Animal models.....	6
Etiology.....	8
Environmental.....	8
Epigenetics.....	10
Genetics.....	12
Cleft palate syndromes.....	13
Common variants.....	15
Rare variants.....	16
Emerging areas of investigation.....	18
Conclusion.....	19
Figures.....	21
Tables.....	24
Chapter II: Trio-based GWAS identifies novel associations and subtype-specific risks for cleft palate	29
Abstract.....	30
Introduction.....	31
Materials and methods.....	32
Study population and phenotyping.....	32
Sample preparation and whole genome sequencing.....	33
Quality control.....	34
Principal component analysis.....	34
Statistical analysis.....	34
DECIPHER variants.....	35
Animal studies and gene expression assays.....	36
Replication of previous published SNPs.....	37
Results.....	38
GWAS of any CP type.....	38
Subtype-specific GWAS.....	38
Expression during mouse palatogenesis and limb development.....	40
Attempted replication of previous published SNPs.....	42
Discussion.....	42
Description of supplemental information.....	47
Acknowledgements.....	47
Declaration of interests.....	48
Data availability.....	48
Web resources.....	48
Figures.....	50
Tables.....	54

Supplemental material	58
Supplemental figures	58
Supplemental tables	61
Chapter III: Genome-wide study of gene-by-sex interactions identifies risks for cleft palate	72
Abstract	73
Introduction	73
Methods	75
Study population and phenotyping	75
Sample preparation and whole genome sequencing	76
Quality control	77
Statistical analysis	77
Genetically regulated gene expression imputation and association testing	78
Rare variants	79
Results	80
Discussion	81
Figures	86
Tables	88
Supplemental data	93
Chapter IV: Cleft palate probands are significantly enriched for protein-altering <i>de novo</i> variants	98
Abstract	99
Introduction	100
Methods	102
Study cohort	102
Whole genome sequencing	103
Identification of <i>de novo</i> variants	103
Variant annotation	104
<i>De novo</i> enrichment	104
Enrichment analysis and creation of gene sets	105
Results	106
Exome-wide analyses	106
Gene-specific analyses	108
Gene set enrichments	110
Discussion	112
Figures	117
Tables	123
Supplemental data	125
Chapter V: Rare variants in PRKCI cause Van der Woude syndrome and other features of peridermopathy	185
Abstract	186
Introduction	187
Methods	189

Study cohort	189
Sample preparation and whole genome sequencing	190
Variant filtering and annotation	191
DN enrichment.....	191
Functional validation in zebrafish.....	191
Results.....	192
Discussion.....	195
Figures.....	199
Tables	202
Chapter VI: Discussion.....	205
Summary of Results	205
Interpretation and implications	209
Study limitations	211
Future directions	213
Conclusion	214
Chapter VII: References	215

List of Figures

Figure 1.1: Human palatal development	21
Figure 1.2: Correlation between human and mouse palates	22
Figure 1.3: Phenotypic heterogeneity in orofacial clefts	23
Figure 2.1: Genome-wide significant locus at 9q33.3 spans craniofacially-expressed gene <i>ANGPTL2</i> and a craniofacial super enhancer	50
Figure 2.2: Regional association plots illustrate differences between groups	51
Figure 2.3: Comparison of odds ratios for any suggestive loci demonstrates subtype-specific effects	52
Figure 2.4. <i>Angptl2</i> expression during mouse palate and limb development	53
Figure 3.1: Miami plot highlighting suggestive SNPs in females and males	86
Figure 3.2: Relative risks for males and females demonstrating GxS effect SNPs	86
Figure 3.3: Comparing relative risks with 95% confidence intervals	87
Figure 4.1: Exome-wide enrichment for CP	117
Figure 4.2: Gene-specific enrichment for CP	118
Figure 4.3: Gene-specific enrichment for protein-altering variants by CP subtype	119
Figure 4.4: Gene set enrichment analysis with ToppFun	120
Figure 4.5: Enrichment for CP in a specific set of OFC-associated genes as determined by denovolyzeR	121
Figure 4.6: Enrichment for CP in a set of marker genes from the secondary mouse palate at E15.5	122
Figure 5.1: Expression of <i>PRKCI</i> in embryogenesis	199
Figure 5.2: Structure and distribution of <i>PRKCI</i> variants	200
Figure 5.3: Representative zebrafish embryo assays for functional and non-functional <i>PRKCI</i> variants	201

List of Tables

Table 1.1: Common syndromes featuring cleft palate	24
Table 1.2: Genes associated with cleft palate	26
Table 2.1: Suggestive and significant loci from any CP type and subtype-specific GWAS	54
Table 2.2: Credible sets from fine-mapping and probability for inclusion of listed SNPs.....	55
Table 2.3: Previous published OFC-associated SNPs with evidence of replication in the current study.....	56
Table 3.1: Loci of suggestive significance from sex stratified TDTs.....	88
Table 3.2: Loci of suggestive significance in the LRT 2df.....	89
Table 3.3: Tissues and genes with significant association between imputed gene expression and phenotype	92
Table 4.1: Subtype enrichment for protein-altering DNPs in palatal osteogenesis marker genes	123
Table 4.2: Overlap in CPSeq genes with protein-altering DNPs and DDD genes with P/LP DNPs	124
Table 5.1: DNPs and rare variants in PRKCI	202

List of Supplemental Figures

Figure S2.1: Principal Components for Genetic Ancestry.....	58
Figure S2.2: Asian-Ancestry Specific Analysis.....	59
Figure S2.3: Any Cleft Palate Manhattan & QQplots.	60
Figure S2.4: Soft Cleft Palate Manhattan & QQplots.....	60
Figure S2.5: Hard Cleft Palate Manhattan & QQplots.	60
Figure S3.1: Manhattan and QQplots for male probands	93
Figure S3.2: Manhattan and QQplots for female probands	93
Figure S3.3: Manhattan and QQplots for all probands	93
Figure S3.4: Manhattan and QQplots for LRT2df probands	94
Figure S3.5: Manhattan and QQplots for GxS probands	94
Figure S3.6: <i>Post hoc</i> power calculations.....	95
Figure S4.1: Distribution and frequency of de novo variants	125
Figure S4.2: Enrichment of genes associated with the OFC-gene panel	125
Figure S4.3: Enrichment of genes associated with the secondary palate at E15.5	126

List of Supplemental Tables

Table S2.1: CPSeq Cohort Description	61
Table S2.2: <i>Angptl2</i> gene-specific primers	61
Table S2.3: RT-qPCR primers	61
Table S2.4: All SNPs tested for replication	62
Table S3.1: Rare variants in <i>LTBP1</i>	96
Table S4.1: Breakdown of CP subtypes by proband sex and self-reported race	127
Table S4.2: Features of syndromic CP probands	127
Table S4.3: All <i>de novo</i> variants in CP probands	129
Table S4.4: Expected versus observed DNs in males versus females with respect to prevalence differences	149
Table S4.5: Global DN enrichment for all categories	150
Table S4.6: Protein altering gene-specific DN enrichment in probands with and without syndromes and/or PRS	152
Table S4.7: Protein altering gene-specific DN enrichment in probands by subtype	165
Table S4.8: Gene set enrichment terms	173
Table S4.9: Enrichment in OFC-gene panel	178
Table S4.10: Enrichment in snRNAseq marker genes	180

CHAPTER I – Introduction

INTRODUCTION

Orofacial clefts (OFCs) are a common birth defect, affecting approximately 1 in 1000 live births (1). OFCs are often divided into those that affect the upper lip (cleft lip with or without cleft palate, CL/P) and cleft palate only (CP). CP accounts for about a third of OFC cases, occurring in approximately 1 in 1700 live births. Without intervention, it can be fatal early in life, although prognosis is much more favorable with surgical correction. However, individuals born with CP often go on to face speech or hearing problems, require advanced orthodontic care, and can experience additional comorbidities as they age (2, 3). As such, CP creates both individual and public health burdens. Improving our understanding of its origin can lead to enhanced prevention, treatment, and prognosis for affected individuals.

The etiology of CP is heterogenous. Environmental factors such as maternal exposure to tobacco and alcohol can increase the risk for CP, while maternal vitamin supplementation is suggested to reduce risk. However, CP is also highly heritable. Approximately 20% of affected individuals have a positive family history of clefting, and concordance among twin pairs suggests heritability up to 90%. Despite this known genetic component, the established risk factors for CP do not fully account for this heritability. Genome-wide association studies (GWAS) have discovered several associated loci, but collectively account for little of the heritability as the associated alleles are relatively rare and/or population specific. Thus, there is “missing heritability” for which the genetic causes of CP are unknown.

This research aims to further the understanding of genetic causes of CP through exploration of both common and rare variants in a primarily trio-based cohort. This chapter will provide

background and context for this research, along with main questions and approaches with which to answer to them, including their significance and limitations.

BACKGROUND

Epidemiology

Cleft palate occurs in approximately 5.9 per 10,000 births in the United States but these estimates vary by state and range from 3.1 in Maryland to 13.8 in North Dakota (4). Globally, the overall reported rate is approximately 5.0 per 10,000 births, but is similarly highly variable between countries, with the lowest rates reported in Nigeria (0.32 in 10,000 births) and Cuba (1.35 in 10,000 births) and the highest documented in British Columbia (25.31 in 10,000 births) and Finland (14.31 in 10,000 births) (5). Some of these differences may be due to under diagnosis in certain regions, as CP is not always recognized immediately at birth especially when compared to an outwardly recognizable CL/P. A Netherlands-based study reports almost 25% of CP cases received a diagnosis later than 12 months of age (6). In the case of the Netherlands, all OFC cases are registered prior to corrective surgery; however, organizational structures for birth defect registries are highly variable. Late-diagnosed cases may not always be registered in some localities, and it is possible that CP-related infant mortality in areas with minimal access to healthcare are never detected. Furthermore, most studies report only live births, and do not include other pregnancy outcomes such as termination or stillbirth.

Within the United States, there are also variable rates across different racial and ethnic groups. One study compiling data from 29 states reports the highest prevalence (per 10,000 births) in non-Hispanic American Indian/Alaska Native populations at 6.5, followed by non-Hispanic White populations at 6.4, non-Hispanic Asian or Pacific Islanders and Hispanic populations both

at 5.6, and non-Hispanic Black populations at 4.4 (4). However, a California-specific investigation reported the highest rates among non-Hispanic White populations at 7.60, followed by non-Hispanic Asian or Pacific Islanders at 4.90, Hispanic populations at 4.79, non-Hispanic Black populations at 4.12, and non-Hispanic American Indian/Alaska Native populations at 2.11 (7). The discrepancy in non-Hispanic American Indian/Alaska Native populations is likely due to limited sample sizes resulting in larger variance across studies, though other factors such as environmental exposure rather than genetic ancestry may also contribute. Excluding this group, the overall observed trend in the US is that prevalence of CP is highest in non-Hispanic White Individuals, and lowest in non-Hispanic Black individuals, which largely mirrors the global observations.

Lastly, CP also consistently exhibits a sex-bias. Females are affected more frequently than male, with the reported male:female ratios ranging from 0.64-0.93:1 (4, 5, 8). One reason this may occur is due to a known delay of palatal closure where this process occurs approximately one week later in female embryos compared to male embryos (9). This may contribute to a lower genetic liability in which females are overall more susceptible to CP than males.

Overall, the descriptive epidemiology for CP is complex with variability among both geographical and ancestral backgrounds.

Palatogenesis

The palate is made up of the primary palate, or the alveolus, at the anterior portion of the mouth (just behind the upper teeth) and the secondary palate, which makes up most of the roof of the mouth (**Figure 1.1**). The secondary palate is the structure disrupted in CP and is itself made up of two parts: an anterior bony hard palate and a posterior muscular soft palate (**Figure 1.2A, 1.2C**).

General facial development begins around week 4 in embryogenesis, with cellular contributions from both the facial ectoderm and neural crest mesenchyme (10). The nose, upper lip/philtrum, and primary palate form from the fusion of three facial prominences—the medial nasal process, lateral nasal process, and maxillary process—at the lambdoidal junction. As the forebrain grows, the facial ectoderm expands into the frontonasal process followed shortly thereafter by expanse of the first branchial arch into two maxillary processes and two mandibular processes during week 5. Both maxillary and mandibular processes quickly grow towards midline and merge into single structures that will become the upper and lower jaw, respectively. At the same time, paired ectodermal thickenings lateral to the frontonasal process emerge as nasal placodes. As superficial epithelial cells and the deeper mesenchymal cells proliferate, the lateral and nasal medial processes arise in the early part of the 5th week. The medial nasal processes then fuse with each other at midline and with the maxillary process to form the philtrum and upper lip, while the lateral nasal processes fuse with the maxillary process to complete the nose (11). Simultaneously, the medial nasal process contributes to the anterior primary palate as well as what will become the nasal septum.

As the upper lip and primary palate formation nears completion, the secondary palate begins to develop. It begins to form as paired outgrowths from the maxillary process called the palatine shelves. These outgrowths first grow vertically alongside the tongue, then as the tongue moves downward into place, they elevate and grow horizontally towards each other during week 8. At this point in development, many structures meet and merge: the palatine shelves fuse to each other at midline, to the primary palate anteriorly in a Y-shaped pattern, and to the nasal septum dorsally.

Throughout palatogenesis, the periderm plays a key role in preventing adhesions between growing structures until the proper time of fusion (12). The periderm consists of a single, superficial layer of epithelial cells that are of ectodermal origin, unlike the mesenchymal cells originating from the neural crest. As fusion begins, the cells of the medial edge epithelium (MEE) come together and form the medial epithelial seam (MES). Epithelial cells extend filopodia and adhere to one another through glycoproteins and desmosomes (13). These cells will eventually disappear and form a solid mesenchymal layer, but the process by which this occurs is still unclear. Multiple theories have been postulated, and it appears that the elimination of the MES is due to a combination of apoptosis and epithelial to mesenchymal transition (EMT) (14, 15). Complete fusion of these various structures begins around weeks 9-10 and is complete by week 12 (16, 17).

The process of palatogenesis is complex and well-orchestrated. Disruption between weeks 4-7 can result in CL/P, while disruption from weeks 7-12 can result in CP.

Classification

There are three broad categories of OFCs: cleft lip (CL, occurring in 1 in 2,800 live births), cleft lip with cleft palate (CLP, occurring in 1 in 1,600 live births), and cleft palate (CP, occurring in 1 in 1,700 live births) (**Figure 1.3**) (4). Those involving the upper lip are often combined into cleft lip with or without cleft palate (CL/P), which are thought to be distinct from CP, and account for ~2/3 of all OFCs (1). However, within these categories there is extensive phenotypic heterogeneity such as differences in laterality (i.e., bilateral versus unilateral CL and left versus right) and degree of severity. CP is similarly phenotypically heterogenous.

The secondary palate—the structure disrupted in CP—is divided into the anterior, bony portion and the posterior, muscular portion. Individuals with CP can have a cleft of either or both

structures. Cleft of only the soft palate is the most frequent manifestation, followed by a cleft of both the hard and soft palate, where cleft of only the hard palate is rare (18, 19). Additionally, individuals may have a submucous cleft palate (SMCP), in which the mucous membranes of the palate are continuous and fused, but the underlying structures are not formed appropriately (20). This can result in velopharyngeal insufficiency (VPI) and is often accompanied by a bifid uvula. However, a bifid uvula can also occur as a single anatomical variation and is reported in 0.42% of the population (21). Despite association with additional phenotypes (such as VPI or hearing concerns), both SMCP and bifid uvula are often diagnosed later in life and may be considered subclinical due to lack of overt clinical signs associated with the physical anomaly.

Animal models

Because palatogenesis occurs so early in development, access to human tissue at the time of palatal growth is incredibly limited. Much of what we know about the development of OFCs and specifically CP is due to animal models. Mice have been instrumental in understanding CP pathogenesis: their anatomy and palatogenesis are remarkably like those of humans (**Figure 1.2B, 2D**) (22). The stages of palatal shelf growth and fusion are the same, with a key difference being muscle anatomy. Mice do not have a uvula, so they have 4 paired muscles in the soft palate, as opposed to 5 pairs in humans, and are missing the *musculus uvulus*.

In general, there are several mouse lines in which spontaneous OFCs occur (23, 24). These genetic backgrounds in themselves are insightful, as deeper investigation of this susceptibility shows developmental differences (often related to facial geometry) compared to mice that do not develop spontaneous OFCs (23). In more specific investigations, knockout mouse models have improved our understanding of the mechanisms by which CP occurs. A review on the molecular

basis of CP (25) in mice presents six categories of defects resulting in CP: failure of palatal shelf formation, abnormal fusion of palatal shelves and the tongue or mandible, failure of or delayed palatal shelf elevation, failure of palatal shelf growth after elevation, persistence of medial edge epithelium, and CP due to a secondary defect (such as other structural anomalies impeding normal development). Many of these knockouts are of genes known to cause CP in humans, further highlighting the similarities between species and how we can better understand these mechanisms in humans by utilizing mouse models.

In addition to similar developmental processes and genetic risks, mice have also proven to be representative models for environmental risk for CP. For example, early CP research carried out in mice demonstrated increased risk with maternal administration of corticosteroids, finding that it led to a delay of shelf movement and thus, increased incidence of CP (26). This mimics what we know about systemic corticosteroid use during pregnancy in women, which has been an established CP risk factor for many years (27). This is also true for anti-epileptic drugs (28) and exposure to retinoic acid (29) during pregnancy.

Zebrafish are also a commonly used model for investigating CP. Although their anatomy is quite divergent from humans, they have parallel structures that can be invaluable in understanding palatogenesis. In zebrafish, the ethmoid plate is the roof of the oropharynx and develops similarly to the human palate. Zebrafish also have an enveloping layer, which is the most superficial layer of the epidermis and the embryonic origin of the periderm (30). As with mice, disruption of several orthologous genes that lead to CP in humans also lead to disruption of the periderm or ethmoid plate (31, 32). The relative ease of genetic manipulation, as well as quick reproduction with numerous offspring are also benefits in using zebrafish as a model for CP.

ETIOLOGY

Environmental

There are several environmental factors that affect CP risk. The influence of both maternal and paternal ages has been investigated, with fathers over the age of 40 having a 58% higher risk of having offspring with CP compared to those from ages 20-29; the same comparison showed a 28% higher risk for mothers over 40 (30). Unsurprisingly, most studied factors focus on maternal behaviors throughout pregnancy.

One of the most consistent risk factors is that of active and passive maternal smoking, with reported odds ratios for CP ranging from 1.09-2.11 (31-34). Studies on alcohol use during pregnancy have been less consistent. While some studies suggest an increased risk ranging from 2.28-2.89 (35, 36), others report odds ratios from 0.94-1.3, with confidence intervals inclusive of 1 and thus inconclusive (37, 38). The risk for CP from maternal multivitamin or folate supplementation is also inconsistent. Some studies find no clear preventive or increased risk effects (34, 39, 40), however when broken down into timing of supplementation, one study found that pre-conceptional use was associated with lower risk of CP (41). The true association or effect of multivitamin/folate use may be difficult to ascertain due to multiple factors—multivitamin ingredients and dosage can vary by brand, some people may be inconsistent in taking them, and the duration of supplementation can all affect folate levels. Maternal exposure to drinking water with high levels of water disinfection by-products (42) and nitrates (43) have also been associated with higher risk for CP.

Maternal medical conditions are another possible environmental influence on CP risk. Maternal diabetes increases the risk for any birth defect, including CP when stratified by specific defects (44, 45). Both pregestational and gestational diabetes mellitus (PDM and GDM,

respectively) have been associated with an increased risk for CP. One study reports PDM increased risk of CP with other defects (OR 10.73), and GDM association with isolated CP (OR 1.54) (46). Similarly, another study found an RR of 2.35 for PDM and 1.40 for GDM (47), while a third study reported odds ratios of 2.2, 0.4, and 0.3 for Type 1 DM, Type 2 DM, and GDM, respectively; however, the confidence interval ranges include 1, precluding conclusive associations with risk for CP in this sample (48). A study in a Hungarian population found other medical conditions associated with CP included anemia complicating pregnancy (OR 1.84), excessive vomiting during pregnancy (OR 3.19), and threatened abortion (*i.e.*, pregnancy complications with the potential for spontaneous abortion, OR 4.94) (49). Maternal illness leading to acute inflammation or hyperthermia is also implicated in CP risk. A study out of Hungary (50) found that mothers of children with CP were more commonly affected with influenza (OR 1.6, 95% CI 1.0-2.6), the common cold (OR 1.5, 95% CI 1.0-2.2), and cystitis (OR 2.1, 95% CI 1.2-3.5) in the third and fourth month of gestation compared to unaffected population controls. Compared to mothers with children affected by congenital anomalies other than CP, mothers of children with CP were more frequently affected by influenza (OR 1.6, 95% CI 1.1-2.3), sinusitis (OR 3.5, 95% CI 1.4-8.8), and bronchitis (OR 2.2, 95% CI 1.0-4.7).

There are several teratogenic medications that increase risk for CP. A well-established teratogen is vitamin A, or retinoic acid (RA), and its derivatives. Although epidemiologic data for exposure is difficult to obtain, one study of 154 pregnancies exposed to isotretinoin (51) found a relative risk of 25.6 (95% CI 11.4-57.5) for major congenital anomalies, including cleft palate. Another study suggests a particular sensitivity of the cranial neural crest (CNC) to retinoids, with prevalence of CNC birth defects increasing with vitamin A supplementation (52). However, as a vital embryogenic regulator (53), both too much and too little vitamin A are associated with birth

defects. Specific to CP, a study in Norway identified a lower risk in mothers who were in the top quartile for vitamin A intake versus the lowest quartile (OR 0.46, 95% CI 0.24-0.94) (54). Another category of potentially risk inducing medications is anti-epileptic drugs (AEDs). Many AEDs have been associated with CL/P (55), though data on CP alone is less common. Valproic acid as monotherapy has been reported to increase the odds for CP from 5.2-11.3-fold compared to the general population (56-58), as has carbamazepine (58, 59). Lastly, systemic corticosteroid use has been a longstanding risk factor for CP (27).

There are a number of additional medications for which epidemiological data are conflicting. These include inhaled B2-agonists for asthma (60), inhaled corticosteroids (60), the anti-emetic drug ondansetron (61, 62), analgesic use of aspirin (63), opioids (64, 65), and multiple classes of antibiotics (59, 66, 67). Lastly, the broad category of nitrosatable drugs, particularly amides, have been associated with CP (68)—these are drugs which produce N-nitroso compounds in acidic environments such as the stomach. Unsurprisingly, the substances considered nitrosatable amides include drugs which individually have been associated with CP, including the AED carbamazepine, antibiotics (amoxicillin, tetracycline), and opioids (codeine, morphine, oxycodone, etc.) (69).

Epigenetics

Many maternal environmental exposures exert their effects through epigenetic mechanisms (70), including DNA methylation, histone modifications, and non-coding RNAs (71), ultimately resulting changes in gene expression without changes to the DNA sequence. However, studying epigenetic changes in human cohorts with OFCs is difficult. A key problem is limited access to relevant tissue at a relevant time. Because palatogenesis occurs so early in development, it is nearly

impossible to obtain and compare normal and abnormal tissues. There have been efforts to study markers in whole blood as well as discarded palatal tissue, but the correlation of epigenetic markers in these tissues to the those at time of development is unknown. Regardless, there have been multiple efforts to better characterize epigenetic changes in OFCs.

In an epigenome-wide association study (EWAS), blood samples for individuals with CP, CL, and CLP were compared to each other (72). After filtering out age-related CpGs, there were 121 sites associated with CP versus CL with the top site in the gene body of *NFIX*. There were also 17 differentially methylated regions (DMRs) between CP and CLP and 294 between CP and CL. Of these, 9 have been previously reported as associated with OFCs, such as *TBX1* and *CRISPLD2*. A second EWAS using newborn blood spots found CpGs in *CLASRP* were significantly associated with pooled samples of CP and CLP compared to controls (73). When evaluating DMRs, there were 6 significant differences between controls and CP, 10 between CL and CP, and 13 between CLP and CP. There was one DMR in *FOXK2* that was significantly differentially methylated in all three comparisons, suggesting a possible role in CP.

Related to histone modifications, one group evaluated Ezh2-dependent methylation of H3K27 in a mouse model (74). Using a targeted method, they performed conditional knockout of Ezh2 in the surface-ectoderm derived oral epithelium. They concluded that Ezh2 methylation of H3K27 represses *Cdkn1a* expression and promotes to proliferation of epithelium in the palatal shelves; thus, disruption of this regulation can interfere with palatogenesis.

Lastly, microRNAs (miRNAs) are another avenue of epigenetic regulation that may affect the risk for CP, although our understanding of their impact on palatogenesis is incomplete as much of the work in this area uses mouse or zebrafish models. In a study focused on retinoic acid-induced CP in mice, *all-trans* RA administration led to upregulation of mir-124-3p (75). Consequently,

there was significant reduction in cell proliferation in the palatal shelves resulting in CP, which could be partially rescued by specific mir-124-3p inhibition. Another study using human and mouse palatal cell lines found that overexpression of mir-140-5p resulted in decreased cell proliferation and reduced expression of known CP-associated genes: *BMP2* and *FGF9* in human cells, and *Fgf9* and *Pdgfra* in mouse cells (76). A systematic effort combining genes associated with CP in mice with bioinformatic methods found 18 miRNA candidates for CP based on their roles in regulation of multiple genes (77). There have been efforts to summarize the known roles of miRNA in embryogenesis globally as well as during each stage of palatogenesis (78, 79), and even efforts to use miRNAs in small vesicles as potential early biomarkers for CP (80).

Epigenetic associations with CP and, more broadly, OFCs remain largely unexplored. This is in part due to limited access to relevant tissue and the cost associated with current technology. However, as methods improve with time, further investigation into this expansive field will be needed to further elucidate the mechanisms of CP.

Genetics

OFCs have long been recognized to have a heritable component, with the earliest reports of familial cases published in the mid 1700s (81, 82). In 1942, Fogh-Andersen postulated an epidemiological difference between CL/P and CP on the basis that given a proband with a cleft, affected relatives were more likely to have the same type (*i.e.*, if the proband had a cleft palate, relatives with a cleft were most likely to also have a cleft palate) (83). More recent large-scale twin studies have demonstrated heritability for both CL/P and CP around 80-90% (84), and additional investigation of crossover risks in Denmark and Norway between subtypes has provided additional evidence to support an etiological distinction between CL/P and CP (85, 86).

Despite its known heritability, CP has remained understudied, in part due to insufficient sample sizes. Much of the early work in understanding causes of CP was carried out in syndromic cases. Overall, CP occurs as part of a syndrome (*i.e.*, in conjunction with other congenital anomalies) about 50% of the time (87). There are hundreds of syndromes listed with cleft palate as a feature in OMIM (<https://omim.org/>), some of which are very well-recognized and relatively frequent syndromes.

Cleft palate syndromes

The most common syndromic form of clefting is van der Woude syndrome (VWS, OMIM:119300), which occurs in 1-3 in 100,000 live births and accounts for ~2% of OFC cases (88, 89). VWS is an autosomal dominant syndrome most often presenting as bilateral lower lip pits, with or without an OFC (90, 91). It is also a prominent example of what is often referred to as “mixed clefting”, in which CL/P and CP can be seen within the same family (91, 92). Around 70% of VWS cases are caused by missense or loss-of-function variants in *IRF6* (93-95), and around 5% are due to missense or loss-of-function variants in *GRHL3* (96). Interestingly, there is a degree of genotype-phenotype correlation between cases with *IRF6* or *GRHL3* variants: individuals with *GRHL3* variants are more likely to have CP and less likely to have CL/P compared to those with *IRF6* variants (96).

It has been shown that both genes work within the same pathway in the periderm (96). Popliteal pterygium syndrome (PPS, OMIM:119500) is an allelic disorder also caused by missense variants in *IRF6*, and clinical features can include lower lip pits, OFCs, sygnathia, pterygia (webbing) of the lower limbs, syndactyly, and genital abnormalities (97, 98). While there is overlap in phenotypes and genetic variants of VWS and PPS, it is postulated that dominant

negative mutations results in the more severe PPS, whereas loss-of-function mutations lead to VWS (88), though other studies have found the overall mechanism is likely more complicated (98). Both syndromes result from disruption of the periderm, a single layer of epithelial cells that functions to prevent pathological adhesions during embryogenesis (i.e., functions as “Teflon” to prevent structures from fusing together until the correct time) (12). Other syndromes also arise due to periderm disruption, including Bartsocas-Papas syndrome (BPS, OMIM: 263650) and cocoon syndrome (OMIM: 613610) from variants in *RIPK4* (99) and *CHUK (IKKa)* (100), respectively. In some cases, the phenotype of BPS is indistinguishable from PPS, but overall BPS and cocoon syndrome are described by very severe congenital anomalies that are rarely compatible with life.

Beyond the periderm, there are several other syndromes in which CP is a common feature. An incomplete list of some of the most recognized syndromes can be found in **Table 1.1**. Broadly, the most common additional clinical features observed with CP are congenital heart defects and anomalies of the brain, urinary tract, and digits (5), although the phenotypic spectrum is incredibly diverse.

Historically, genetic causes of syndromic CP were thought to be distinct from isolated CP but as sequencing has become more common, shared causes have become apparent with identification of individuals with similar genetic variants manifesting across a phenotypic spectrum (101). This is evident in the *IRF6*-related disorders, where we see association with both isolated and syndromic CP (91, 102-104). Additionally, one study using whole genome sequencing of 30 isolated CP cases found that 5 individuals harbored pathogenic or likely pathogenic variants in 5 different OFC syndrome-associated genes (105). Altogether, these findings suggest that the

distinction between isolated or syndromic is likely not easily delineated due to the poorly understood, but common, phenomena of variable expressivity and reduced penetrance.

Common variants

Although there is growing evidence of overlap in genetic etiology of isolated and syndromic OFCs, most studies on CP claim to be restricted to isolated cases. This in itself poses several challenges. First, many cases are recruited as infants or very young children, making ruling out some common syndromes difficult since it may be too early in life to detect intellectual disabilities or myopia (near-sightedness). Alternatively, cases with two common structural anomalies may be coincidental rather than due to the same genetic variant (such as CP and a benign heart murmur). Family history can also be complicated. As discussed earlier, some phenotypes such as SMCP or bifid uvula may be subclinical and undiagnosed, thereby making a familial case of CP appear “sporadic”, as there is no known family history of clefting.

There have been varying degrees of success in identifying genetic associations with CP in presumably isolated cases (**Table 1.2**). A targeted study of 40 candidate genes found SNPs or haplotypes in *IRF6*, *COL2A1*, *COL11A2*, *WNT3*, *FGFR1*, and *CLPTM1* to be associated with isolated CP, all of which are also associated with syndromic OFCs (106). On a larger scale, GWAS has found fewer loci for CP compared to CL/P. In addition to fewer studies of CP, there are also fewer significant associations per study. To date there have been eight genome-wide association studies (GWAS) for CP (102, 107-113), of which several have either failed to find any associations or identified significant variants occurring either at low frequency or in specific populations.

For example, the earliest GWAS for CP was conducted as a typical association study along with gene-by-environment interactions with maternal smoking, alcohol use, and supplementation

with multivitamins (107). For genetic association alone, no SNPs reached genome-wide significance. When considering environmental interactions, however, there were suggestive associations, including increased risk of CP for SNPs in *MLLT3* and *SMC2* with alcohol use and *TBK1* and *ZNF236* with smoking, and decreased risk for SNPs in *BAALC* with multivitamin supplementation. In 2017, Ludwig, et. al revisited and expanded this dataset with imputation of variants based on 1000 Genomes reference panels; however, there remained no genome-wide significant loci (114). The first successful GWAS for CP found only a single locus with genome-wide significance in *GRHL3* (108), which we know is also involved in VWS. In this study, the associated SNP has an MAF of only ~3% in Europeans and even lower in other populations, highlighting one of the challenges in finding common variants associations. Similarly, the discovery of the 2p12 locus was carried out in an African cohort, for which the lead SNP has MAF <2% in that population (110), and a SNP in the *IRF6* enhancer region is found nearly exclusively in Finnish Europeans (112). Other studies have found variants that are more common (102, 111), but these results have yet to be replicated outside of the discovery population (Han Chinese). Lastly, some GWAS have failed to find significant associations in their respective cohorts when divided into CL/P and CP, but combined meta-analysis shows shared risks for any OFC that are otherwise not detectable (109, 113). One such gene is *FOXE1*, which is also associated with Bamforth-Lazarus syndrome, a rare syndrome which includes CP as a feature (115).

Rare variants

Although common variant associations are an important part of the genetic architecture, the relative dearth of signals may indicate a role for rare variants in CP pathogenesis. There appears to be etiologic overlap in common and rare associations for CP, with both types of variants having

been identified for *IRF6* (92) and *GRHL3* (116, 117). However, the more common theme for rare variants is syndromic-associated genes also harboring rare variants in isolated CP cases. For example, in one study using WGS of 30 CP cases, there were pathogenic/likely-pathogenic (P/LP) variants identified in 5 syndrome-associated genes. In this study, however, it is worth noting that phenotypic description of four out of five of these cases indicate additional features that may represent mild forms of the respective syndromes, such as speech and/or developmental delays (105). A similar study using whole exome sequencing (WES) from 37 confirmed isolated CP cases identified a burden of ultra-rare variants (those never reported in gnomAD) in *COL2A1* and *GLI3*, and P/LP variants in four genes associated with CP syndromes (118).

Another useful approach to identification of rare variants causing CP has been with multiplex families. Using WES, one study found multiple rare, predicted damaging variants in three genes not previously association with CP (*ACACB*, *PTPRS*, and *MIB1*), two genes from candidate gene studies (*ARID5B* and *MNI*), and two genes associated with syndromic CP (*CREBBP* and *GRHL3*) (119). An additional familial case report found a missense variant in *ARHGAP29* segregating with CP; notably, *ARHGAP29* was previously only associated with CL/P (120, 121).

Other studies have used a more targeted sequencing approach. A study in a Polish population found eight novel and four reported variants in 38 CP cases when using a panel of 423 genes previously associated with OFCs or facial development (122). Another study used a panel of 418 OFC-associated genes found that 17.6% of CP cases harbored a P/LP variant (123). Interestingly, that same study found only 9.1% of cases with CL/P and 2.8% with CL with PL/P variants—again highlighting the difference in genetic architecture between subtypes of OFCs, as well as pointing for a larger role for rare variants in CP than for CL/P.

The largest study to date analyzed 603 syndromic OFC cases (of which 443 had CP) to better understand the molecular pathways underlying syndromic OFCs (124). They found 36.5% of cases had P/LP variants with 20 genes harboring 3 or more of these. On pathway analysis, comparison to traditionally “non-syndromic” genes found shared enrichment for EMT, ossification, and other signaling pathways; however, proteins of ATP-dependent chromatin remodelers were associated with syndromic genes, posing a possible explanation for phenotypic variability.

Lastly, an area that remains largely unexplored for rare variants in CP is that of *de novo* variants (DNs). A previous investigation using WGS for case-parent trios identified exome-wide enrichment of DNs for CL/P (125). This study also found enrichment for CP cases, although the sample size was small, leading to wide confidence intervals for enrichment estimates. Despite this limitation, this suggests a role for DNs that needs to be further explored.

Emerging areas of investigation

As technology has advanced, there are emerging methods that allow for both better understanding of normal palatogenesis pathways and improved gene prioritization based on biological relevance. Rich datasets of gene expression in bulk and specific cell tissues are now accessible for relevant tissues in both human and mouse.

Wilderman, *et al.* generated a 25-state epigenome atlas based on histone markers in human embryonic craniofacial tissue across multiple time points (126). The same group has also performed single nucleus RNA sequencing (snRNAseq) in Carnegie Stage (CS) 20 human embryonic craniofacial tissue (127), which is particularly relevant as this is around the time the secondary palate beings developing. Similarly detailed datasets on more specific structures exist

for mice. One group generated scRNAseq data for the lambdoidal junction in mice at E11.5 (~6 weeks human gestational age), generating five distinct cell clusters (128). Another group generated three sets of data for the developing soft palate at timepoints E13.5, E14.5, and E15.5, and from this data were able determine the role of *Runx2* in soft palate myogenesis with a mouse model (129). A third group focused specifically on secondary palate osteogenesis at E15.5, identifying key drivers of this process and reporting genes not previously known to play a role (130). This same study also employed spatial transcriptomics, allowing observation of changes in gene expression in place as they occur. Additional applications of these types of data to gene discovery can be fruitful—for example, enrichment in specific epithelial and mesenchymal cell types was found in a set of DNs from a cohort of isolated OFCs (125), providing insight into more specific mechanisms of clefting.

Beyond snRNAseq, other uses of expression data have been by integration of GWAS data with expression quantitative trait loci (eQTL). Although this can improve understanding of non-coding variant effects on gene expression, the access to expression data in relative tissue is a limitation for application to OFCs in general, but particularly for CP as there are no large-scale expression databases for human palatal tissue. To date, these studies have not been performed in CP, but a few have been done in CLP with several reported associations (131-133).

CONCLUSION

Cleft palate has historically been understudied, especially compared to CL/P; however, as sample sizes have grown and data becomes more available, studies focused on CP alone have increased. Despite these emerging studies, there remains a gap in discovered genetic variants associated with CP. Based on our current knowledge, it is clear that genetic heterogeneity is a factor in CP

pathogenesis, yet many of the potential sources of this heterogeneity have yet to be explored. This includes common variant associations across specific subtypes of CP and between sexes, as well as a large-scale analysis of rare variants, either inherited or *de novo*. The work presented here adds to the current knowledge using approaches that are both traditional and novel as applied to a CP cohort.

FIGURES

Figure 1.1: Human palatal development. A) Week six of human palatal development, with the secondary palate shown vertically on each side of the tongue and a gap between the secondary palate, nasal septum and primary palate. B) After descent of the tongue, the secondary palatal shelves elevate and orient horizontally, allowing them to come in contact and begin fusing. C) Fusion of the primary and secondary palate and the nasal septum separating the oropharynx from the nasopharynx. Figure used with copyright permission from *Taylor & Francis*: Benjamin Levi, Samantha Brugman, Victor W. Wong, Monica Grova, Michael T. Longaker & Derrick C. Wan (2011) Palatogenesis, *Organogenesis*, 7:4, 242-254, DOI: [10.4161/org.7.4.17926](https://doi.org/10.4161/org.7.4.17926)

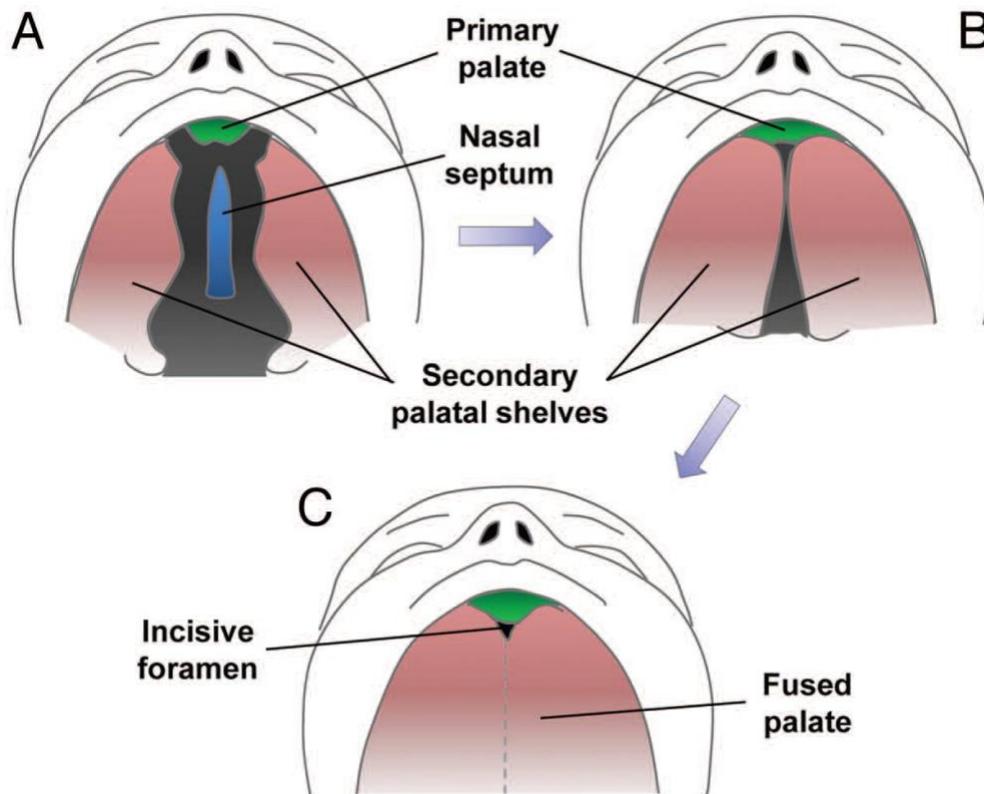


Figure 1.2: Correlation between human and mouse palates. A) Normal human upper lip, hard palate and soft palate. B) Normal mouse upper lip, hard palate and soft palate. C) Cleft of the secondary palate in human patient. D) Clefting of secondary palate in transgenic mouse.

Figure used with copyright permission from *Taylor & Francis*:

Benjamin Levi, Samantha Brugman, Victor W. Wong, Monica Grova, Michael T. Longaker & Derrick C. Wan (2011) Palatogenesis, *Organogenesis*, 7:4, 242-254, DOI: [10.4161/org.7.4.17926](https://doi.org/10.4161/org.7.4.17926)

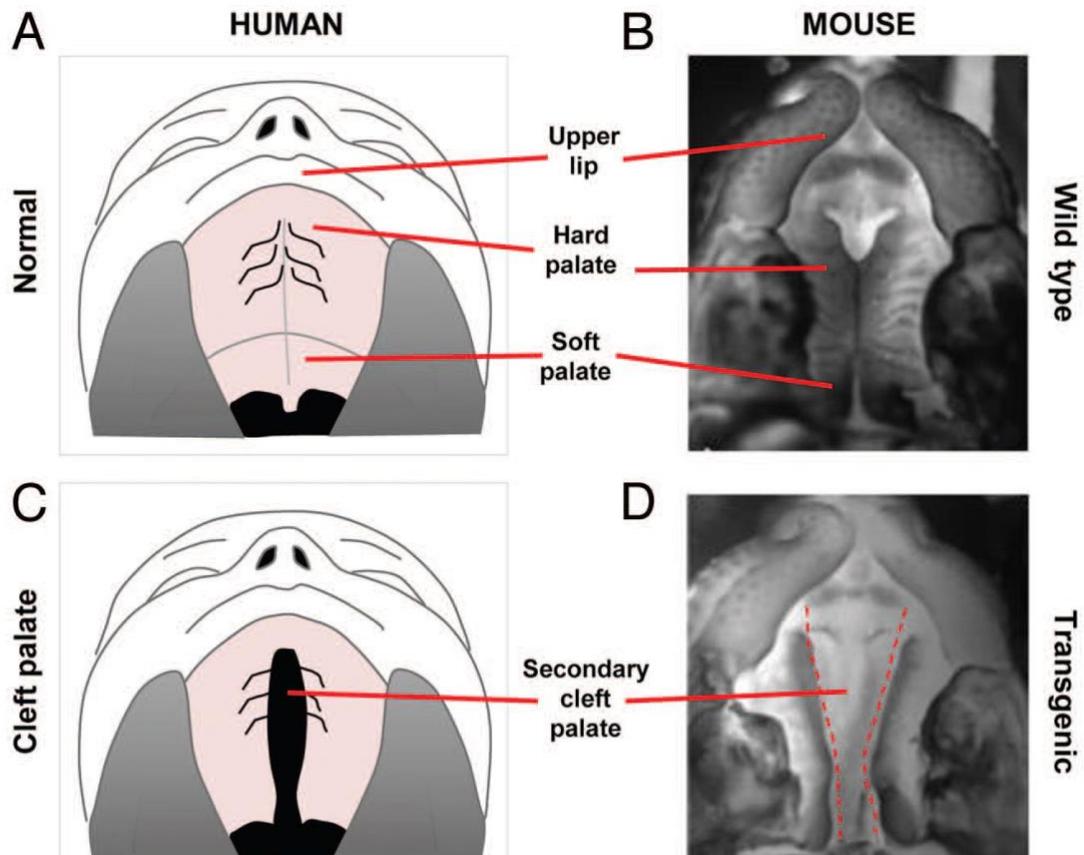
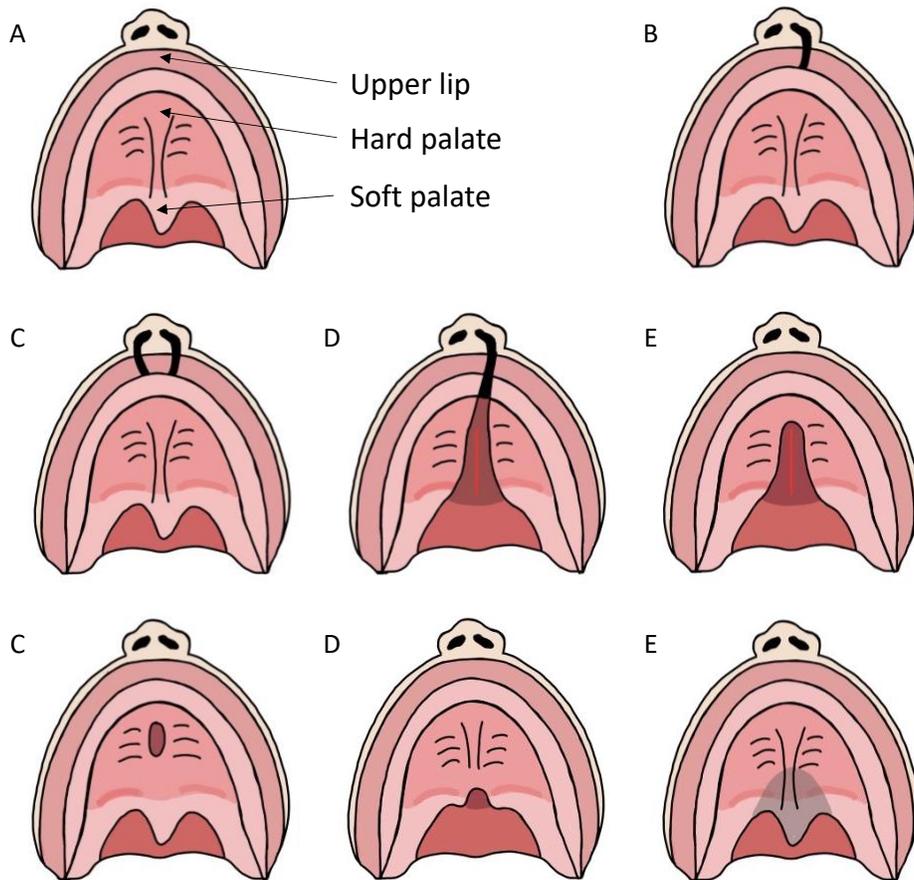


Figure 1.3: Phenotypic heterogeneity in orofacial clefts. A) Normal upper lip and palate. B) Unilateral cleft lip—may occur on the left or right side. C) Bilateral cleft lip. D) Unilateral cleft lip with cleft palate. E) Cleft hard and soft palate. F) Cleft hard palate. G) Cleft soft palate. H) Submucous cleft palate.



TABLES

Table 1.1: Common syndromes featuring cleft palate				
Syndrome	Prevalence	Gene(s)	Types of Variants	Clinical Features
22q11.2 deletion (137, 138)	1 in 4000	<i>TBX1</i> , others	Structural variant	Congenital heart defects, cleft soft palate or velopharyngeal dysfunction, endocrine abnormalities immunodeficiency or autoimmune disease, gastroenterological issues, skeletal abnormalities, neuropsychiatric and cognitive deficits
Van der Woude (97, 99)	1 in 35,000	<i>IRF6</i> <i>GRHL3</i>	Heterozygous missense, nonsense	Bilateral lower lip pits, orofacial clefts
Popliteal Pterygium (97, 101, 139)	1 in 300,000	<i>IRF6</i>	Heterozygous missense, nonsense, splicing	Bilateral lower lip pits, orofacial clefts, popliteal webbing, syndactyly, genitourinary anomalies, nail anomalies, syngnathia, ankyloblepharon, talipes, digital reduction defects
		<i>RIPK4</i>	Homozygous missense	
Stickler (140)	1 in 7500 to 1 in 9000	<i>COL2A1</i> <i>COL11A1</i>	Heterozygous missense, nonsense, splicing	Congenital vitreous anomaly, retinal detachment, congenital megalophthalmos, deafness, arthropathy, cleft palate
Treacher-Collins (141)	1 in 50,000	TCOF1	Heterozygous nonsense, missense, structural variants	External and middle ear abnormalities, conductive hearing loss, bilateral and symmetric downslanting palpebral fissures, coloboma, cleft palate, heart defects, choanal atresia
		<i>POLRID</i>	Heterozygous missense, nonsense, splicing	
		<i>POLRIC</i>	Homozygous missense, compound heterozygous	
		<i>POLRIB</i>	Heterozygous missense	

Apert (142)	1 in 80,000 to 1 in 160,000	<i>FGFR2</i>	Heterozygous missense	Craniosynostosis, midface retrusion, syndactyly, cleft palate, nonprogressive ventriculomegaly, structural cardiac abnormalities, gastrointestinal malformations, and anomalies of the genitourinary tract
SATB2-associated (143, 144)	Unknown	<i>SATB2</i>	Heterozygous missense, nonsense, splicing; structural variants	Developmental delay, intellectual disability, absent or limited speech, cleft palate, dental anomalies, behavioral concerns

Gene/Locus	Study Type	Variant Type	Function	Reference
1p12 - <i>NOTCH2</i>	Rare – targeted NGS	Coding, missense	Transmembrane protein	(125)
1p22.1 - <i>ARHGAP29</i>	Rare – targeted NGS	Coding, missense	GTPase coactivating protein	(124, 125)
1p36.11 - <i>GRHL3</i>	Common - GWAS	Coding, missense	TF	(111, 112)
	Rare	Coding, missense	TF	(122)
1p32.2 – <i>FGGY</i>	Common – GWAS (combined CP and CL/P)	Non-coding, intronic	Carbohydrate kinase	(116)
1q32.2 - <i>IRF6</i>	Common - GWAS	Coding, missense	TF	(105, 114)
	Common - GWAS	Non-coding, enhancer region	TF	(115)
	Rare – targeted NGS	Coding, missense	TF	(121, 125)
2p12 - <i>CTNNA2</i>	Common - GWAS	Non-coding	Unknown	(113)
3p14.1 – <i>ADAMTS9</i>	Common – GWAS (combined CP and CL/P)	Non-coding, intronic	Metalloprotease	(116)
3p14.3 - <i>FLNB</i>	Rare – targeted NGS	Coding, missense		(125)
3p14.3 – <i>ERC2</i>	Common – GWAS (combined CP and CL/P)	Non-coding, intronic	Membrane traffic protein	(116)
3p21.31 – gene dense region	Common – GWAS (combined CP and CL/P)	Non-coding, intronic	Unknown	(116)
3p22.1 <i>POMGNT2</i>	Common - GWAS	Non-coding	TF	(105)
3q28 - <i>TP63</i>	Rare – targeted NGS	Coding, missense	TF	(125)
3q29 - <i>DLG1</i>	Rare – targeted NGS	Splicing	Adaptor protein	(125)
4p16.3 - <i>WHSC1</i>	Common - GWAS	Non-coding	TF	(105)
6p21.32 - <i>COL11A2</i>	Common – haplotype association	N/A	ECM structural protein	(109)
6q14.3 - <i>TBX18</i>	Rare – targeted NGS	Coding, missense	TF	(125)
7p14.1 - <i>GLI3</i>	Rare - WES	Coding, missense	TF	(121)

8p11.23 - <i>FGFR1</i>	Common – haplotype association	N/A	Transmembrane signal receptor	(109)
9q22.33 - <i>FOXE1</i>	Common – GWAS (combined CP and CL/P)	Non-coding	TF	(112)
10q21.3 - <i>ARID5B</i>	Rare - WES	Coding, missense	Transcription cofactor	(122)
10q25.1 - <i>COL17A1</i>	Rare – targeted NGS	Coding, missense	ECM structural protein	(125)
10q25.3 - <i>VAX1</i>	Rare – targeted NGS	Coding, missense	TF	(125)
11q24.2 - <i>FCDON</i>	Rare - WES	Coding, missense	Cell adhesion molecule	(121)
12p13.2 - <i>LRP6</i>	Rare – targeted NGS	Coding, missense	Transmembrane signal receptor	(125)
12p13.33 - <i>WNT5B</i>	Rare – targeted NGS	Coding, missense	Intercellular signal molecule	(125)
12q13.11 - <i>COL2A1</i>	Rare – WGS	Coding, splicing Non-coding, intronic	ECM structural protein	(104, 109, 121)
12q24.11 - <i>ACACB</i>	Rare - WES	Coding, missense	Enzyme	(122)
13q22.1 – <i>KLF12</i>	Common – GWAS (combined CP and CL/P)	Non-coding, intronic	TF	(116)
13q32.3 - <i>DOCK9</i>	Common - GWAS	Non-coding	TF	(105)
14.q13.3 - <i>PAX9</i>	Common - GWAS	Non-coding	TF	(105)
14q32.2 - <i>DLK1</i>	Common - GWAS	Non-coding	TF	(105)
15q24.3	Common - GWAS	Non-coding	Unknown	(114)
16p13.3 – <i>CREBBP</i>	Rare - WES	Coding, missense	Histone modifier	(122)
16q24.2 - <i>FOXC2- FOXL1</i>	Common - GWAS	Non-coding	TF	(105)
17q21.31 - <i>WNT3</i>	Common – haplotype association	N/A	Intercellular signal molecule	(109)
18q11.1 - <i>MIB1</i>	Rare - WES	Coding, missense	Ubiquitin protein ligase	(122)
19p13.11 - <i>MAU2</i>	Common - GWAS	Non-coding	TF	(105)

19p13.3 - <i>PTPRS</i>	Rare - WES	Coding, missense	Protein phosphatase	(122)
19q13.32 - <i>CLPTM1</i>	Common – haplotype association	N/A	Transmembrane protein	(109)
20p13 - <i>SNRPB</i>	Rare - WES	Coding, missense	RNA splicing factor	(121)
22q12.1 - <i>MNI</i>	Rare - WES	Coding, missense	Transcription cofactor	(122)
Xp22.2 - <i>NHS</i>	Rare – targeted NGS	Coding, missense		(125)

CHAPTER II:
Trio-based GWAS identifies novel associations and subtype-specific risk factors for cleft palate

Adapted from an original article published in Human Genetics and Genomics Advances:

Kelsey Robinson, Trenell J. Mosley, Kenneth S. Rivera-González, Christopher R. Jabbarpour, Sarah W. Curtis, Wasii Lanre Adeyemo, Terri H. Beaty, Azeez Butali, Carmen J. Buxó, David J. Cutler, Michael P. Epstein, Lord JJ Gowans, Jacqueline T. Hecht, Jeffrey C. Murray, Gary M. Shaw, Lina Moreno, Seth M. Weinberg, Harrison Brand, Mary L. Marazita, Robert J. Lipinski, Elizabeth J. Leslie. (2023). Trio-based GWAS identifies novel associations and subtype-specific risk factors for cleft palate. *Human Genetics and Genomics Advances* 4(4):100234.

ABSTRACT

Cleft palate (CP) is one of the most common craniofacial birth defects; however, there are relatively few established genetic risk factors associated with its occurrence despite high heritability. Historically, CP has been studied as a single phenotype although it manifests across a spectrum of defects involving the hard and/or soft palate. We performed GWAS using transmission disequilibrium tests of 435 case-parent trios to evaluate broad risks for any cleft palate (ACP, n=435), and subtype-specific risks for any cleft soft palate (CSP, n=259) and any cleft hard palate (CHP, n=125). We identified a single genome-wide significant locus at 9q33.3 (lead SNP rs7035976, $p=4.24 \times 10^{-8}$) associated with CHP. One gene at this locus, angiotensin-like 2 (*ANGPTL2*), plays a role in osteoblast differentiation. It is expressed both in craniofacial tissue of human embryos and developing mouse palatal shelves. We found 19 additional loci reaching suggestive significance ($p < 5 \times 10^{-6}$), of which only one overlapped between groups (chromosome 17q24.2, ACP and CSP). Odds ratios (ORs) for the 20 loci were most similar across all three groups for SNPs associated with the ACP group, but more distinct when comparing SNPs associated with either subtype. We also found nominal evidence of replication ($p < 0.05$) for 22 SNPs previously associated with orofacial clefts. Ours is the first study to evaluate CP risks in the context of its subtypes and we provide newly reported associations affecting the broad risk for CP as well as evidence of subtype-specific risks.

INTRODUCTION

Orofacial clefts (OFCs) are the most common craniofacial congenital anomalies in humans, occurring at a birth prevalence of ~1 in 1000 live births (1, 4). Prognosis is favorable with surgical intervention, though individuals with OFCs face many healthcare challenges such as multiple corrective surgeries, abnormal dentition, hearing and speech problems, and increased morbidity and mortality throughout life (1, 87). As such, these anomalies are associated with increased healthcare costs and long-term psychosocial burdens for individuals and their families (134).

OFCs are typically classified into groups including cleft lip (CL), cleft lip with cleft palate (CLP), and cleft palate (CP). Etiologically, cleft lip with or without a cleft palate (CL/P) is considered anatomically and epidemiologically distinct from CP, although both are highly heritable (85, 86) with estimates up to 90% for both CL/P and CP (84). However, compared to dozens of known risk loci for CL/P (135, 136), common risk variants for CP remain largely undiscovered. Large-scale studies of CP, including seven genome-wide association studies (GWAS) (102, 108-112, 137), have mostly identified variants occurring either at low frequency or in specific populations. For example, a missense variant in *GRHL3* occurs at ~3% only in Europeans (108), a variant near *CTNNA2* occurs at ~1.5% in an African population (110), and a variant in an enhancer region of *IRF6* occurs almost exclusively in Finnish and Estonian Europeans (112).

Historically, CP has been evaluated as a single group, but it encompasses a phenotypic spectrum including overt clefts of the hard and/or soft palate and submucous cleft palate. There is evidence for specific gene expression patterns in the hard versus soft palate (138), so we hypothesized there may be underlying genetic differences among CP subtypes. To investigate this further, we performed GWAS of three groups of case-parent trios via transmission disequilibrium

tests (TDT). These groups include all cases of CP regardless of subtype, cases involving the soft palate (i.e., hard and soft palate, soft palate only, and submucous cleft palate), and cases involving the hard palate (i.e., hard and soft palate plus hard palate only).

MATERIALS AND METHODS

Study population and phenotyping

The study population comes from multiple domestic and international sites where recruitment and phenotypic assessment occurred following institutional review board (IRB) approval for each local recruitment site and the coordinating center (University of Iowa, University of Pittsburgh, and Emory University). We assembled a collection of 435 case-parent trios ascertained on proband affection status (e.g., cleft palate) (Table S2.1). The majority of trios (96%) consist of affected CP probands with unaffected parents. Although probands/trios were not excluded based on additional clinical features consistent with a syndromic diagnosis, only 45 trios were classified as possibly or probably syndromic based on a reported presence of additional major or minor clinical features. Trios represent all major ancestry groups affected by CP including those with European ancestry (recruited from Spain, Turkey, Hungary, United States), as well as understudied populations from Latin America (Puerto Rico, Argentina), Asia (China, Singapore, Taiwan, the Philippines), and Africa (Nigeria, Ghana). All probands and parents were assessed for the presence of a CP with ~2/3 of the assembled samples undergoing additional phenotyping to assess the location and severity of the CP. Here we designate these probands as having a cleft of the hard and soft palate (n=82), cleft of the hard palate only (n=43), cleft of the soft palate only (n=152), and submucous cleft palate (n=25). For the purposes of analysis, submucous cleft palate was grouped with cleft soft palate.

Sample preparation and whole genome sequencing

Whole genome sequencing was performed at the Center for Inherited Disease Research (CIDR) at Johns Hopkins University (Baltimore, MD). Prior to sequencing, samples were tested for adequate quantity and quality of genomic DNA using a Fragment Analyzer system and were processed with an Illumina InfiniumQCArray-24v1-0 array to confirm gender, relatedness, and known duplicates. For each sample, 500-750ng of genomic DNA was sheared to 400-600bp fragments, then processed with the Kapa Hyper Prep kit for End-Repair, A-Tailing, and Ligation of IDT (Integrated DNA Technologies) unique dual-indexed adapters according to the Kapa protocol to create a final PCR-free library.

A NovaSeq 6000 platform using 150bp paired-end runs was used for sequencing followed by base calling through the Illumina Real Time Analysis software (version 3.4.4). Files were demultiplexed from binary format (BCL) to individual fastq files with Illumina Isas bcl2fastq (version 1.37.1) and aligned to the human hg38 reference sequence (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39). The DRAGEN Germline v3.7.5 pipeline on the Illumina BaseSpace Sequence Hub platform was used for alignment, variant calling, and quality control, which produced single sample VCF files. The DRAGEN contamination detection tool was used to check for any cross-human sample contamination. Genotype concordance with existing array-based genotypes was performed using CIDRSeqSuite (version 7.5.0), and genotype concordance checks amongst replicate samples was performed in Picard GenotypeConcordance (Picard 2019). Following data quality steps and confirmation of adequate coverage (at least 80% of the genome at 20X or autosomal coverage at 30X), joint variant calling was performed, generating a multi-sample VCF file.

Quality control

Variants aligning outside of standard chromosomes (1-22, X, Y), those with a filter flag and variants with a minor allele count (MAC) of <1 were removed. Genotypes with a quality score of <20 or a read depth of <10 were set to missing, and sites with missingness values of >10% were subsequently filtered out. Sample-level quality control metrics included transition/transversion (Ts/Tv) ratio, silent/replacement rate, and heterozygous/homozygous ratio; outlier samples with values outside of 3 standard deviations from the cohort mean were discarded. Samples with high missing data (>5% missing) were removed.

Principal component analysis

A principal component (PC) analysis was performed for probands only using PLINK (v2.0) to determine genetic ancestry within our cohort. Variants were excluded if the minor allele frequency (MAF) <15% or if any site was missing a genotype (missingness >0%). We performed linkage disequilibrium (LD) pruning for $R^2 > 0.1$ prior to analysis. Ultimately, there were 67,584 variants for which we generated 15 PCs. After visualization of PC plots, PCs 1-3 were used to group by ancestry (Figure S2.1, S2.2A).

Statistical analysis

We performed transmission disequilibrium tests (TDT) to statistically analyze common variant associations with CP. The TDT, originally described by Spielman et. al. (139) tests for the rate of transmission of the minor allele to an affected proband using a McNemar's test (i.e., a modified Chi-squared test for paired data). Because it tests transmission rather than allele frequency, it is robust to population stratification; however, it is only informative for sites at which parents are

heterozygous for a variant. As differences in heterozygosity between populations may mask signals, we also performed a Chi-square test of homogeneity on our three CP groups to verify there were no significant differences in population makeup. Following filtering steps described above, the multisample VCF was imported into PLINK (version 1.90b53). Trios for which all individuals had a genotype missingness rate <5% and a Mendelian error rate <2% were included. Variants were included if they met the following criteria: MAF of $\geq 3\%$, Mendelian error rate <0.1%, Hardy-Weinberg exact test (HWE) p-value of $> 1 \times 10^{-6}$, and missingness rate of <5%. Results were considered genome-wide significant at $p < 5 \times 10^{-8}$, and suggestive of significance at $p < 5 \times 10^{-6}$. We report our odds ratio (OR) in reference to the alternate allele, and list the effect allele as that which increases risk for CP. Following TDT output, we applied FINEMAP (140) to our genome-wide significant locus. Briefly, FINEMAP uses a stochastic shotgun search to calculate posterior probability of SNP association with disease based on effect size (for which we used the natural log of the OR), MAF, and an LD matrix (generated in PLINK). We ran FINEMAP on SNPs within 1Mb in either direction of the lead SNP using default settings.

DECIPHER variants

We queried the DECIPHER database (141) for copy number variants (CNVs) affecting *ANGPTL2* for individuals with phenotypes related to both palate and limb abnormalities. Terms included for palatal phenotypes were: cleft palate, high palate, narrow palate, narrow mouth, and micrognathia. Terms included for limb phenotypes were: 2-3 toe syndactyly, arachnodactyly, camptodactyly, long toe, abnormality of finger, tapered finger, upper limb undergrowth, and short foot. We then compared the rate of these phenotypes in individuals with CNVs to the general population based on the EUROCAT prevalence data using a two tailed Fisher's exact test.

Animal studies and gene expression assays

Animal studies were conducted in strict accordance with recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the University of Wisconsin School of Veterinary Medicine Institutional Animal Care and Use Committee (protocol number 13-081.0). C57BL/6J mice were purchased from The Jackson Laboratory and housed in rooms maintained at 22 ± 2 °C and 30–70% humidity on a 12 hour dark cycle. Mice were fed Irradiated Soy Protein-Free Extruded Rodent Diet (Catalog No. 2920x; Envigo Teklad Global) until day of plug, when dams received Irradiated Teklad Global 19% Protein Extruded Rodent Diet (Catalog No. 2919; Envigo Teklad Global). For timed matings, one or two nulliparous female mice were placed with a single male mouse for 1-2 hours and then examined for copulation plugs. The beginning of the mating period was designated as gestational day (GD)0, and pregnancy was confirmed by assessing weight gain between GD7 and GD10, as previously described (142). Dams were euthanized by carbon dioxide inhalation followed by cervical dislocation between GD10-14.5 \pm 1 hour for embryo collection. One cohort of embryos collected for *in situ hybridization* assays were dissected in PBS and fixed in 4% paraformaldehyde for 18 h. Embryos subsequently underwent graded dehydration (1:3, 1:1, 3:1 v/v) into 100% methanol and were stored at -20°C indefinitely. Riboprobes were synthesized with gene-specific primers (Table S2.2), and *in situ hybridization* was performed as previously described (143, 144). Embryos were subsequently embedded in 4% agarose gel and cut in sections (130 μ M for head, 60 μ M for limb) using a vibrating microtome. Images were captured using a MicroPublisher 5.0 camera (QImaging) mounted on an Olympus SZX-10 stereomicroscope. Another cohort of embryos was generated for quantitative gene expression analysis. Embryos were collected and microdissected in PBS, and enzymatic separation and isolation of the mesenchyme from maxillary

process (GD10-12) or palatal shelf tissue (GD13-14) was performed as previously described (145, 146). RNA was isolated using the Qiagen RNeasy Mini Kit with on-column DNase I digestion according to the manufacturer recommendations. cDNA was synthesized from 100 ng of total RNA using the GoScript reverse transcription reaction kits (Promega). Singleplex quantitative real-time polymerase chain reaction (RT-qPCR) was performed using SSoFast EvaGreen Supermix (Bio-Rad) on a Bio-Rad CFX96 real-time PCR detection system (Bio-Rad Laboratories). RT-qPCR primers were designed using PrimerQuest (IDT), and sequences are listed in Table S2.3. Target gene specificity was confirmed using National Center for Biotechnology Information Primer Basic Local Alignment Search Tool (NCBI Primer-BLAST). *Gapdh* was used as the housekeeping gene, and analyses were conducted with the $2^{-\Delta\Delta C_t}$ method.

Replication of previously published SNPs

We searched for previously published SNPs associated with OFCs using the NHGRI-EBI GWAS catalog (147), which reports any SNPs with p-values less than 1×10^{-5} . We initially identified 202 SNPs from GWAS data; however, after filtering for duplicates (i.e., SNPs reported in multiple studies) there were 166 SNPs of interest. When reporting the p-value for duplicated SNPs, we chose the most significant value. We then evaluated these variants for association with CP or CP subtypes in our current dataset and found 139 SNPs with data in at least one analysis. When comparing the two datasets, we reported the effect allele as the allele with increased OR as found in our current study.

RESULTS

GWAS of Any CP Type

We performed TDT using all 435 CP case-parent trios (Table S2.1) for 6,946,419 variants (Figure S2.3). In the combined analysis of all trios with CP (hereafter referred to as any CP, or ACP), no loci reached genome-wide significance ($p < 5 \times 10^{-8}$), although there were 10 loci with at least one SNP surpassing a suggestive threshold of $p < 5 \times 10^{-6}$ (Table 2.1). Several GWAS of CP have been published to date (102, 108-112, 137), however we did not find any of these previous loci beyond nominal significance in this study, as further discussed below (Table S2.4).

Subtype-Specific GWAS

CP is phenotypically heterogeneous, where clefts can occur in the hard palate (the bony, anterior portion) and/or the soft palate (the muscular, posterior portion). Given that such phenotypic distinctions arise from different developmental origins, we hypothesized some of the relative lack of associated SNP signals could be attributed to this underlying heterogeneity. To determine if CP subtypes were associated with unique loci, we performed TDTs on two subgroups: “any cleft of the soft palate” which included 259 trios with clefts of the hard and soft palate, soft palate only, or submucous cleft (CSP) (Figure S2.4) and “any cleft of the hard palate” which included 125 trios with hard and soft palate or hard palate only (CHP) (Figure S2.5). Those with clefts of both the hard and soft palate were included in both analyses because that group may be etiologically heterogeneous and may have hard palate and/or soft palate risk factors.

The CSP analysis included 7,286,217 variants. There were no loci reaching genome-wide significance, although there were 8 loci of suggestive significance (Table 2.1). Our CHP analysis included 7,337,001 variants, and we identified a genome-wide locus on chromosome 9q33.3

(Figure 2.1A, 2.1B) spanning the genes *RALGPS1* and *ANGPTL2*, as well as one locus of suggestive significance (Table 2.1). We used FINEMAP (140) to perform statistical fine-mapping of the 9q33.3 locus, which identified three SNPs within the credible set for which there was 100% confidence at least one is association with disease: rs2417050, rs777676, rs12350252 (Figure 2.1C).

We next compared the CSP and CHP analyses to each other and to the ACP group to determine to what extent these loci were associated with a specific CP subtype. The only overlap in suggestive loci was 17q24.2 (near *ARSG*), shared by the ACP (lead SNP rs75850252, 2.89×10^{-6}) and CSP analyses (lead SNP rs3785607, $p=9.44 \times 10^{-7}$). The 17q24.2 signal was driven by the CSP group but also showed nominal evidence of association in CHP ($p=0.006$). Due to the overlap in samples, however, we cannot distinguish between an association of this locus with any type of CP or with a cleft of both the hard and soft palate. There were additional regions, such as the 1q41 locus shown in Figure 2.2A, in which the association patterns were similar between ACP, CSP, and CHP, but did not reach the suggestive threshold in the subtype analyses. In contrast, we observed loci with stronger association signals in one subtype versus the other or in ACP. For example, the top locus in CHP, 9q33.3, reached genome-wide significance but this signal was less significant in CSP ($p=2.28 \times 10^{-4}$) and ACP ($p=1.79 \times 10^{-3}$) (Figure 2.2B), suggesting this signal was driven by clefts of the hard palate (either with or without cleft soft palate). Similarly, the 9q22.31 locus identified in CSP is less significant in ACP ($p=5.41 \times 10^{-4}$) and was essentially absent when looking at CHP ($p=0.53$) (Figure 2.2C), suggesting it is driven exclusively by cleft soft palate.

We then compared the estimated odds ratios (ORs) for the lead SNPs in each region showing suggestive significance for the three subgroups (Figure 2.3) to identify subtype-specific risks suggested by comparisons of p-values. As predicted, the ORs were very similar across all

groups for loci identified in the ACP analysis. However, for loci identified in either CSP or CHP, the differences in ORs were more pronounced. Although most confidence intervals overlap between CSP and CHP, there was some evidence for subtype-specific effects. For the SNPs identified from the CHP analysis, there was no overlap in the range of effects for the 5q15 locus in CHP and CSP (which also contained 1), indicating this locus is specific to clefts involving the hard palate. It is less clear for the SNPs identified in the CSP analysis. For 4 of the 10 SNPs, the confidence interval for CHP contained 1, which allows the possibility of no effect in that group. Additionally, all of the confidence intervals overlapped, indicating these loci may not be subtype-specific. However, taken altogether, these findings suggest there are subtype-specific genetic risks for CP.

Lastly, because our cohort represents diverse populations, we evaluated population-specific signals. Although our test of homogeneity did not show any significant differences between each analysis ($p=0.137$), differences in site heterozygosity between ancestries can mask signals. Using principal component analysis, we subdivided groups into Asian, European, and African ancestry groups (Figure S2.1); however, only the Asian subgroup contained enough trios ($N=262$) to perform TDT for 6,491,466 variants (Figure S2.1). The results from this analysis did not reveal any additional loci not already present in the full cohort, but did demonstrate that some signals in the full cohort, such as 5q14.1 are likely driven by this population. This is also supported by the population frequencies in Table 2.1.

Expression during mouse palatogenesis and limb development

We further investigated the 9q33.3 locus, as the only genome-wide significant locus, to locate a candidate gene for cleft palate. Based on available expression and chromatin segmentation data

from human embryonic craniofacial tissue (126), *ANGPTL2* is actively transcribed in neural crest cells and human craniofacial tissue during embryogenesis at Carnegie Stages 13, 14, 15, and 17, whereas this pattern is not apparent for *RALGPS1* (Figure 2.1B). We therefore hypothesized *ANGPTL2* was more likely than *RALGPS1* to be involved in palatal development (Figure 2.1B). We performed *in situ* hybridization on mouse embryos at two key stages of palatogenesis: initial vertical outgrowth of the palatal shelves, and subsequent horizontal outgrowth just prior to their fusion (148). *Angptl2* staining was observed in the palatal shelves and the lateral aspects of the upper lip at both stages of development (Figure 2.4A, 2.4B). Tissue sectioning revealed *Angptl2* staining in the mesenchymal compartment of the palatal shelves (Figure 2.4A', 2.4B'). Subsequent quantitative assessment demonstrated that mesenchymal *Angptl2* expression increases during palatogenesis in a manner consistent with that of *Runx2*, an established marker of osteogenic differentiation (Figure 2.4C); this is consistent with our observation that the association was driven by clefts of the hard palate.

We also observed strong *Angptl2* staining during limb development in a domain restricted to the mesenchyme adjacent to the apical ectodermal ridge (Figure 2.4D, E). The apical ectodermal ridge secretes signals that maintain the adjacent mesenchyme (i.e., progress zone) in a highly proliferative state, driving proximodistal outgrowth of the limb and digits (149). We therefore investigated overlapping phenotypes using the DECIPHER database (141). We found 38 reported individuals with copy number variants (CNVs) affecting *ANGPTL2* and adjacent genes in this region. Of these, 13/38 (34.2%) had craniofacial phenotypes (e.g., cleft palate, narrow palate, high palate) and 15/38 (39.5%) had limb abnormalities (e.g., camptodactyly of finger, arachnodactyly) with an overlap of 10/38 (26.3%) with clinical features of both craniofacial and limb abnormalities. Using prevalence data from EUROCAT, we compared the rate of OFCs (14.95 per 10,000 births)

and limb anomalies (38.18 per 10,000 births) in the general population to that of patients with CNVs of *ANGPTL2* and found a significant difference between these two groups ($p < 2.2 \times 10^{-16}$, Fisher's exact two-tailed test) for both OFCs and limb anomalies.

Attempted replication of previous published SNPs

We investigated SNPs from previously published studies of OFCs for evidence of replication in our dataset. Using a list from the GWAS Catalog (147), we tested 139 unique SNPs for association with ACP, CSP, or CHP. There were 22 SNPs within 18 loci achieving nominal significance ($p < 0.05$) in at least one group from our analyses (Table 2.2, Table S2.4). When we applied Bonferroni multiple-test correction ($p < 3.59 \times 10^{-4}$), a single SNP (rs1838105, $p = 2.95 \times 10^{-4}$) remained significant for replication in the ACP group.

DISCUSSION

CP has historically been evaluated as a single phenotype, but here we have identified CP subtype-specific risk factors, including one genome-wide significant locus and 19 regions of suggestive significance. Our genome-wide significant locus—found in the CHP group—spans both *RALGPS1* and *ANGPTL2* genes on 9q33.3. Although fine-mapping analysis did not implicate any single gene as all three of the SNPs in the credible set fall within intronic regions of *RALGPS1*; two SNPs (rs2417050 and rs12350252), as well as the lead SNP at this locus, also fall within a region considered a craniofacial super enhancer upstream of *ANGPTL2*. We then showed using *in situ* hybridization that *Angptl2* was expressed in the developing mouse palate.

Specifically, we found that *Angptl2* is expressed during initial vertical and subsequent horizontal outgrowth of the palatal shelves (Figure 2.4A, 2.4B) prior to the approximation and

fusion at the midline that forms the secondary palate. Expression appeared restricted to the mesenchymal compartment, which is primarily comprised of cranial neural crest cells that rapidly proliferate to drive palatal shelf outgrowth and differentiate into osteoblasts that form the bones of the hard palate. As palatogenesis proceeded, mesenchymal *Angptl2* expression increased along with osteoblast marker *Runx2* (Figure 2.4C), consistent with previous evidence suggesting that *ANGPTL2* positively regulates osteoblast differentiation (150). These data are also consistent with bulk RNAseq from human craniofacial tissue at CS 13, 14, 15, and 17 showing a similar increase with time, and single-cell RNAseq of the mouse palate showing *Angptl2* is expressed in cells consistent with osteoprogenitors at GD15.5 (129). These findings, in combination with significantly higher rates of OFCs with CNVs of this locus, support a role for *ANGPTL2* in palatal development, particularly as related to the hard palate.

The most strongly associated locus in ACP was located in an intergenic region at 12q21.1, closest to the gene *TRHDE* (~330kb upstream). The second most significant locus was on chromosome 4p14. Although also in an intergenic region, the nearest genes are *UBEK2* and *PDS5A* (~20kb upstream and downstream, respectively), both of which are strongly expressed in craniofacial tissue during embryogenesis (126). However, *PDS5A*, which plays a role in sister chromatid cohesion during mitosis, is of particular interest: both null and heterozygous loss in mice leads to a cleft palate phenotype with variable expressivity and penetrance (151).

In the CSP group, the top non-overlapping locus was at 9q22.31 spanning *FAM120A*, and RNA-binding protein (152). Although its function has been primarily studied in the context of gastric carcinoma, *FAM120A* plays a role in protecting against oxidative stress-induced apoptosis. Additionally, it has been shown to directly bind insulin-like growth factor II (*IGF-II*) mRNA (153), which is spatially and temporally expressed in developing murine secondary palate (154), and can

result in cleft palate—among other clinical features—when dysregulated (155). *FAM120A* is expressed in both mouse soft palate tissue (129) and in human craniofacial tissue during embryological development (126).

Another locus of interest from this analysis is at 12q13.11, in which the lead SNP is approximately 650kb away from *COL2A1*. Variants within this gene are well established as causal for Stickler syndrome (OMIM: 108300), in which CP is a common feature (156); however, our associated signal does not appear to be within the same topological associated domain (TAD) as *COL2A1*, at least in embryonic stem cells, so additional evaluation of this finding is needed.

Because differences in sample sizes (and statistical power) prevent direct comparisons of p-values, we compared odds ratios between analysis groups. We found that for all loci, the risks between the ACP, CHP, and CSP groups were either in the same direction or null (i.e., no loci conveyed opposite effects for different groups). For loci belonging to the ACP group, the ORs and confidence intervals were similar for CHP and CSP; however, for loci identified in the CSP or CHP analyses, differences in ORs were more pronounced. For example, the 5q14.1 region was associated solely with CHP. Differences between estimated ORs of the CSP loci were less apparent, although 4 of the 10 SNPs may have no risk in CHP. Because so many individuals in the dataset had a cleft hard and soft palate and were included in both the CHP and CSP analyses, the overlapping confidence intervals were expected. We cannot rule out influence of CSP SNPs on CHP risk because cleft hard palate only is not as common as other forms of CP and much larger sample sizes will be required to contrast cleft hard palate only and cleft soft palate only. Overall, our results suggest some variants could contribute more to risk to a specific type of CP.

Although the allelic TDT is not confounded by population stratification, examination of allele frequencies across populations suggests some of our findings may be, in part, driven by

certain populations. The most pronounced of these occurs for rs1468036 (effect allele G) at 17q25.3, which is found at a frequency of 36.6% in African populations, and less than 6% in other populations studied here. Another example is the 5q14.1 locus near *HOMER1*, which reached suggestive significance in the ACP group ($p=2.32 \times 10^{-6}$) but is more significant in the Asian ancestry-stratified study ($p=8.35 \times 10^{-7}$) and occurred with a minor allele frequency of 12.2% in East Asian populations compared to 4% in South Asians and <0.5% in the remaining study populations. Presently, ancestry-specific analysis was only possible for the Asian population, but future studies on both additional ancestral risks as well as combined stratification by ancestry and subtype would be of interest; however, limited sample sizes in this study preclude these evaluations here.

All of the risk loci identified in this study were of novel association with CP, although three of these of these loci have been previously reported in studies associated with CL/P. Both 9q22.31 and 5p14.1 have been reported as suggestive in a consanguineous GWAS of 40 families (157), and the 12q21.1 region was reported in a Chinese Han population (137); however, the lead SNPs for each of these loci are approximately 1Mb away from each of our lead SNPs. Additionally, none of the identified SNPs within the same region are in linkage disequilibrium, and therefore may or may not be tagging the same causal variant(s).

We were able to show nominal evidence of replication for 22 previously published SNPs associated with OFCs. Interestingly, there were only 2 SNPs from previous studies that replicated in all three of our studies (i.e., $p < 0.05$ in ACP, CSP, and CHP), both of which were originally published as associated with CP. There were 9 previously published SNPs replicating at nominal significance in any two of our groups, only one of which was shared between only CSP and CHP. Unfortunately, deeper phenotype data is not available to classify subtypes from previous studies

for more detailed comparisons, but this general lack of overlap may further support subtype-specific differences. An additional striking finding was that for all 7 previously published SNPs with reported ORs associated with CL/P, the risk allele conveyed opposite effects for CP in our dataset, whereas this was only true for 2 of 9 in SNPs previously associated with CP. Although some of these findings could be a result of unclear effect allele reporting, there is evidence of opposite effects for the same allele in CL versus CP has been previously demonstrated near *IRF6* (112) (rs72741048, in Table 2.3), indicating additional investigation of these SNPs in the context of CL/P versus CP is warranted.

Our results support the hypothesis there are subtype-specific risks for CP, although this study has limitations. First, due to sample size we chose to evaluate clefts affecting both the hard and soft palate in both subtype groups, rather than as three separate groups. Given both structures are affected in these cases, it is likely they share risks for hard or soft palate clefts as suggested by our distinct findings between analyses. A lack of genome-wide significant signal in the ACP group could be due to our sample being underpowered to identify common variants of modest effect, or may result from SNPs of opposite effects negating signal when all phenotypes are combined. Alternatively, this may support a more prominent role for rare as opposed to common variants in the pathogenesis of CP in general, or there may be environmental effects not captured by our study. We also failed to replicate some well-established risk loci, such as *GRHL3* or *CTNNA2*, although this is likely explained by our study population. These risk variants occur in ~3% of Europeans and ~2% of Africans, respectively. In individuals of Asian ancestry, the MAF for rs41268753 in *GRHL3* is <1% (SAS) and <0.02% (EAS), and the MAF for both reported variants near *CTNNA2* (rs113691307 and rs80004662) is ~4% (SAS) and <0.04% (EAS); therefore, our cohort is unlikely to harbor these variants at a rate detectable above our filter for common variants at MAF >3%.

Despite these limitations, this is the first study to evaluate CP risks in the context of its subtypes and our findings show there are broad factors affecting the risk for cleft palate in general, as well as variants influencing the risk of specific CP subtypes.

DESCRIPTION OF SUPPLEMENTAL INFORMATION

Supplemental Information contains five figures and four tables.

ACKNOWLEDGEMENTS

We are incredibly grateful to participating families and colleagues who have made this research possible. Sequencing services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268201700006I. Patient recruitment, assembly of phenotypic information, sequencing services, and data analysis were supported by National Institutes of Health (NIH) grants: X01-HG010835 (EL), R01-DE016148 (MM, SW), R01-DE030342 (EL), R01-DE011931 (JH), R01-DE028300 (AB), R01-DE014581 (TB), R37-DE008559 (JM), R00-DE024571 (CB), S21-MD001830 (CB) U54-GM133807 (CB) , T32-GM008490 (KR), F31-DE032588 (KR), R56-DE030917 (RL), F32-DE032260 (SWC). Some of this work was supported through cooperative agreements under PA 96043 from the Centers for Disease Control and Prevention to the Centers for Birth Defects Research and Prevention participating in the NBDPS. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the California Department of Public Health. This article was prepared while Trenell J. Mosley was employed at Emory University. The opinions expressed in this article are the

author's own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government. This study makes use of data generated by the DECIPHER community. A full list of centers who contributed to the generation of the data is available from <https://deciphergenomics.org/about/stats> and via email from contact@deciphergenomics.org. Funding for the DECIPHER project was provided by Wellcome [grant number WT223718/Z/21/Z].

DECLARATION OF INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY

Sequence and phenotype data is available from the Database of Genotypes and Phenotypes (dbGaP) under study accession phs002220.v1.p1.

WEB RESOURCES

1. Hg38 human genome reference:
https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39
2. Plink 2.0: <https://www.cog-genomics.org/plink/2.0/>
3. Plink 1.9: <https://www.cog-genomics.org/plink/1.9/>
4. FINEMAP: <http://www.christianbenner.com>
5. DECIPHER database: <https://www.deciphergenomics.org>
6. EUROCAT database: https://eu-rd-platform.jrc.ec.europa.eu/eurocat/eurocat-data/prevalence_en

7. NHGRI-EBI GWAS catalog: <https://www.ebi.ac.uk/gwas/>
8. UCSC genome browser: <https://genome.ucsc.edu/cgi-bin/hgGateway>
9. OMIM database: <https://www.omim.org>

FIGURES

Figure 2.1: Genome-wide significant locus at 9q33.3 spans craniofacially-expressed gene *ANGPTL2* and a craniofacial superenhancer.

A) Regional association plot for 9q33.3. The labelled SNPs were identified by FINEMAP with 100% confidence of belonging to the credible set of SNPs associated with disease. B) UCSC genome browser tracks for craniofacial-specific gene expression and regulatory regions. C) Zoomed in view of the region with high density of SNPs in LD with SNPs labelled in 1A. Point color corresponds to linkage disequilibrium (r^2) with rs2417050 across all populations. For browser tracks, yellow indicates an enhancer region (darker shades represent stronger elements), green indicates active transcription, and red indicates a transcription start sites. CNCC=cranial neural crest cells, CS=Carnegie stage.

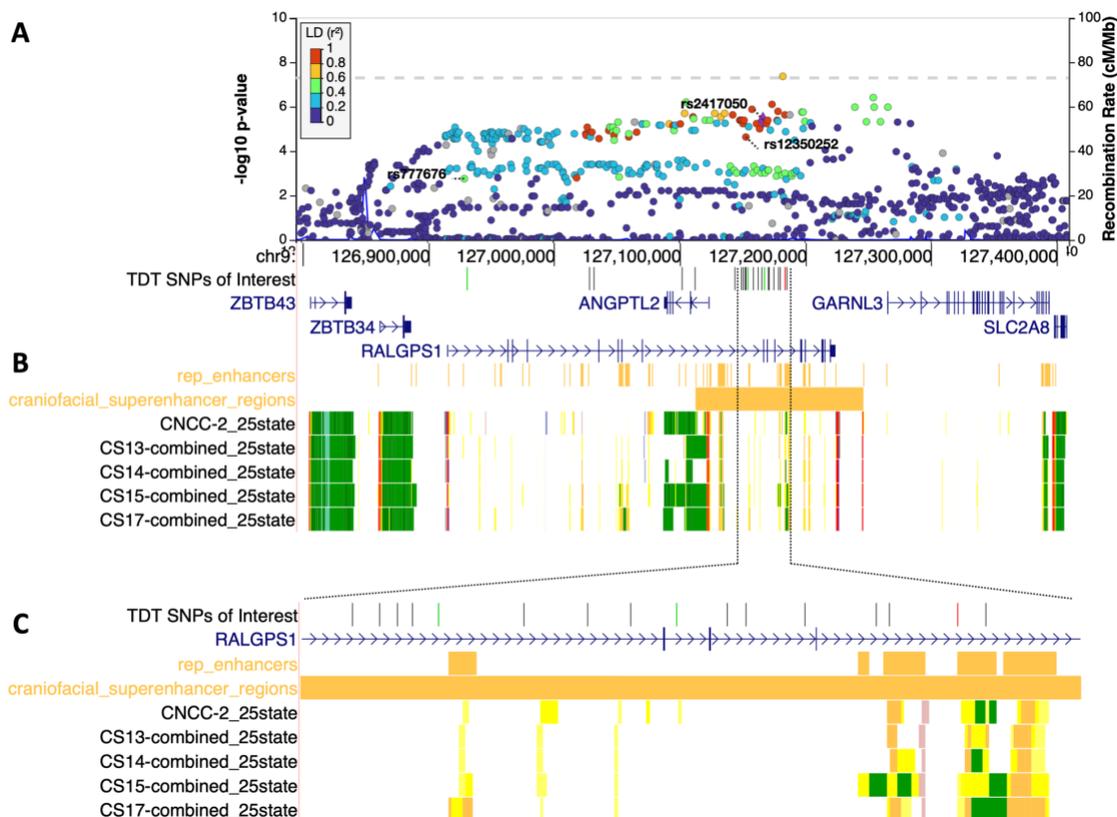


Figure 2.2: Regional association plots illustrate differences between groups at A) the 1q41 locus spanning *LYPAL1* with index SNP rs10779347 B) the 9p33.3 locus spanning *RALGPS1* with index SNP rs7035976, and C) the 9q22.31 locus spanning *FAM120A* with index SNP 4127438 demonstrating similar association patterns (A) and subtype specific associations (B, C). Point color corresponds to linkage disequilibrium (r^2) and the blue lines represent linkage block boundaries.

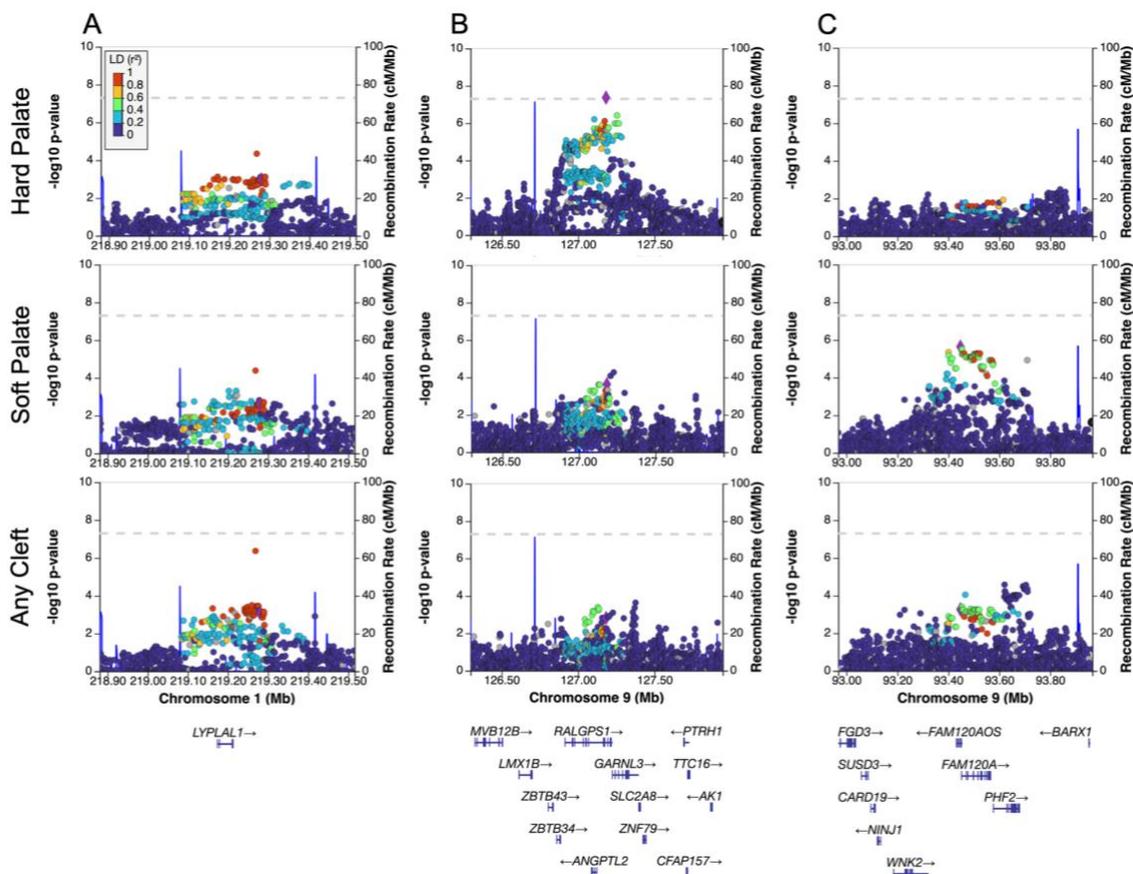


Figure 2.3: Comparison of odds ratios for any suggestive loci demonstrates subtype-specific effects. Loci associated with any cleft palate convey similar ORs for both subtypes (left). Loci associated with a specific subtype (right) carry less extreme ORs and/or are insignificant in the opposing group. Loci marked with an asterisk are featured in Figure 2. ACP=any cleft palate. CSP=cleft soft palate. CHP=cleft hard palate.

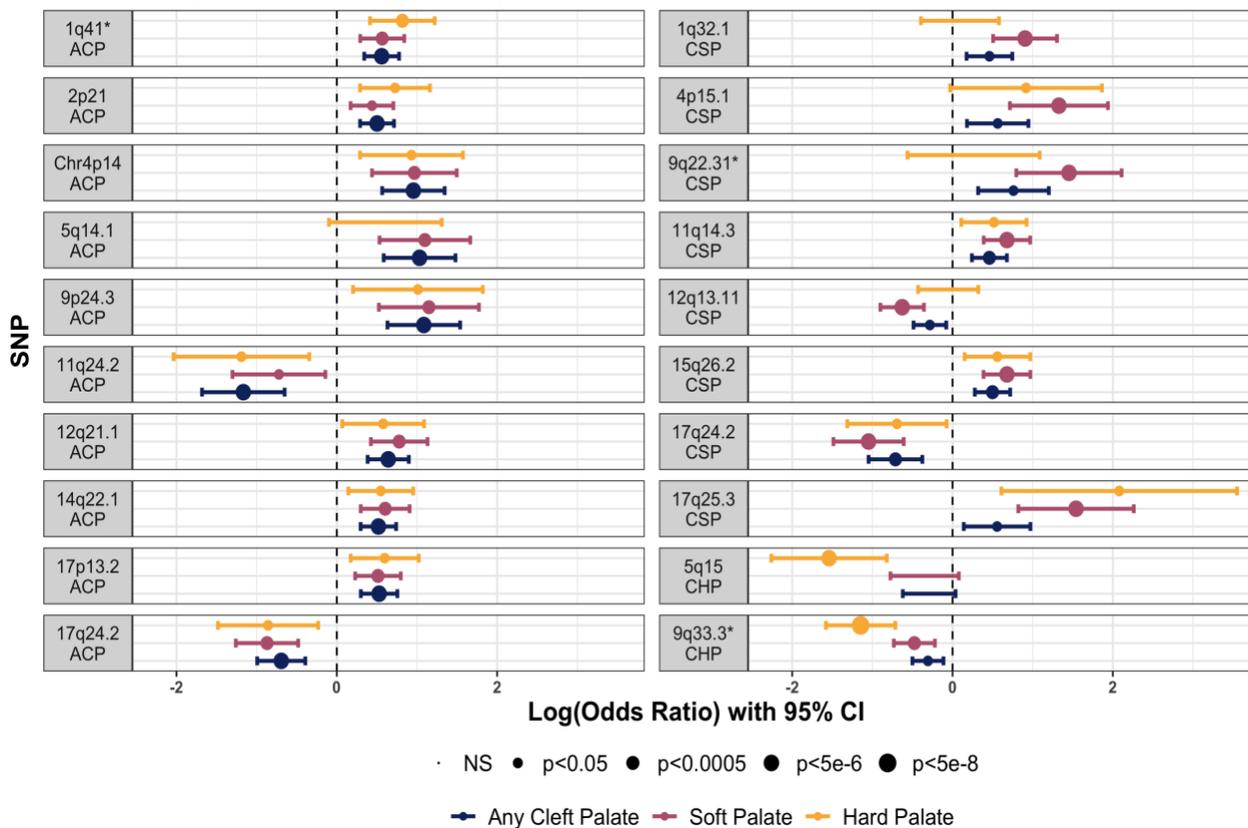


Figure 2.4. *Angptl2* expression during mouse palate and limb development. Mouse embryos at gestational day (GD)13 or 14.5 were stained by *in situ* hybridization to visualize *Angptl2* expression. Whole mount tissues were imaged to view the developing palatal shelves and upper lip (A,B). Subsequent coronal sections illustrate prominent staining in mesenchymal tissue of the palatal shelves (A',B'). qPCR was conducted on mesenchyme isolated from maxillary process/palatal shelf tissue from mice at indicated time points (C). Each value represents the mean \pm SEM of n=3 samples isolated from individual embryos. Forelimbs from GD13 and 14.5 mouse embryos were also stained by *in situ* hybridization to visualize *Angptl2* expression (D,E). A section through the first digit illustrates staining restricted to the mesenchyme adjacent to apical ectodermal ridge. Similar domains of expression were observed in hindlimbs. Scale bars = 0.50 mm.

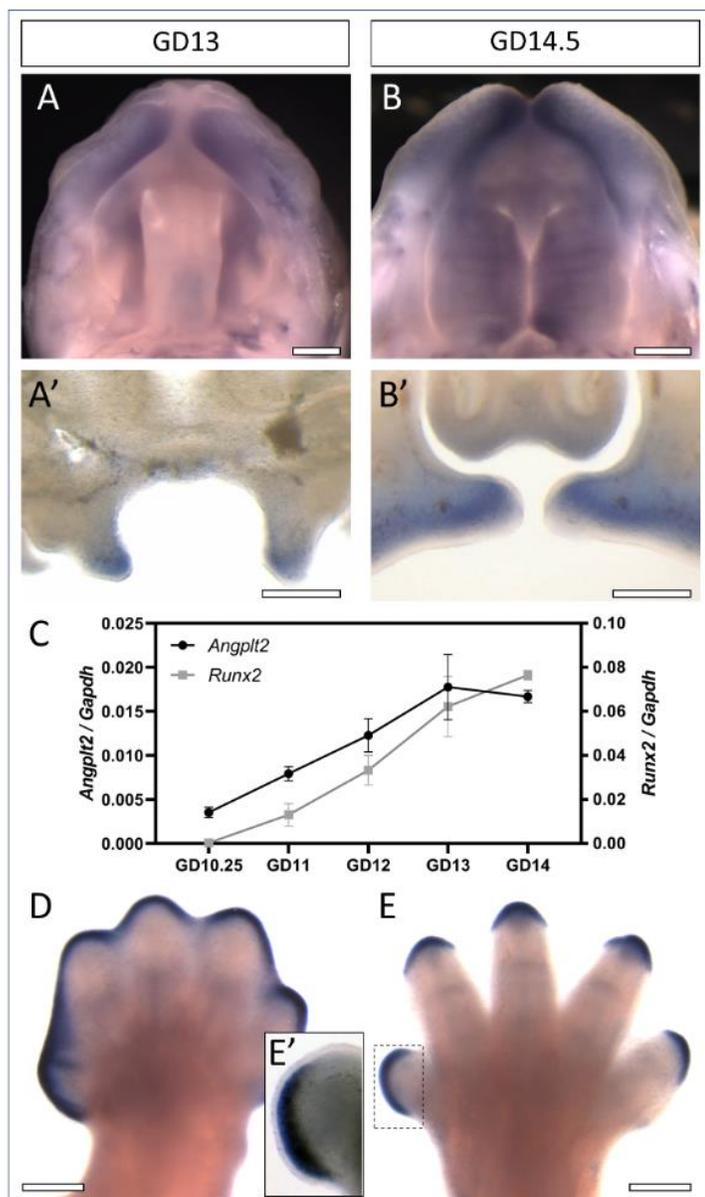


Table 2.1. Suggestive and significant loci from any CP type and subtype-specific GWAS								
Locus (nearest gene)	Lead SNP	Ref/Alt*	OR (95% CI)	p-value	gnomAD v3.1.2 minor allele frequencies			
					EAS (%)	SAS (%)	AFR (%)	EUR (%)
ACP								
1q41 (<i>LYPLAL1</i>)	rs59611530	G/GAAT	1.75 (1.41-2.17)	3.11x10 ⁻⁷	68.2	61.4	39.3	49.5
2p21 (<i>MTA3</i>)	rs57081889	C/T	1.65 (1.34-2.04)	2.55x10 ⁻⁶	22.6	37.3	51.6	36.7
4p14 (<i>UBEK2, PDS5A</i>)	rs10000967	T/C	2.60 (1.76-3.84)	6.07x10 ⁻⁷	7.7	14.6	9.6	11.3
5q14.1 (<i>HOMER1</i>)	rs79156100	T/C	2.81 (1.80-4.39)	2.32x10 ⁻⁶	12.2	4.0	0.1	0.5
9p24.3 (<i>DMRT2</i>)	rs12002920	G/T	2.96 (1.88-4.66)	8.45x10 ⁻⁷	2.7	3.2	18.7	5.4
11q24.2 (<i>OR8B8</i>)	rs375612889	C/T	0.31 (0.19-0.52)	2.66x10 ⁻⁶	1.7	13.0	3.9	9.0
12q21.1	rs7955287	G/A	1.90 (1.47-2.46)	6.34x10 ⁻⁷	88.8	59.6	69.8	62.2
14q22.1 (<i>FRMD6-AS2</i>)	rs10431684	C/A	1.68 (1.35-2.02)	3.42x10 ⁻⁶	70.8	67.2	59.1	52.9
17p13.2 (<i>ANKFY1</i>)	rs58695167	G/A	1.70 (1.35-2.13)	4.26x10 ⁻⁶	76.4	64.5	23.6	61.5
17q24.2 (<i>ARSG</i>)	rs75850252	C/T	0.50 (0.37-0.68)	3.86x10 ⁻⁶	14.3	2.4	25.2	1.5
CSP								
1q32.1 (<i>LINC00862</i>)	rs61824892	C/T	2.47 (1.66-3.68)	4.17x10 ⁻⁶	17.4	8.8	2.8	11.6
4p15.1 (<i>LINC02497</i>)	rs61795400	T/C	3.77 (2.05-6.95)	4.83x10 ⁻⁶	11.7	6.3	10.3	3.9
9q22.31 (<i>FAM120A</i>)	rs41274384	G/A	4.27 (2.22-8.24)	2.28x10 ⁻⁶	7.1	4.1	0.6	3.3
11q14.3 (<i>DISC1FP1</i>)	rs11019136	C/T	1.97 (1.48-2.63)	2.88x10 ⁻⁶	30.0	23.1	25.8	38.3
12q13.11 (<i>LOC105369747</i>)	rs855134	T/C	0.53 (0.41-0.70)	3.92x10 ⁻⁶	41.0	37.4	20.5	35.5
15q26.2 (<i>SPATA8-AS1</i>)	rs36062094	T/C	1.97 (1.47-2.64)	3.42x10 ⁻⁶	32.0	17.5	20.5	22.5
17q24.2 (<i>ARSG, WIP11</i>)	rs3785607	T/C	0.35 (0.23-0.54)	9.44x10 ⁻⁷	17.0	0.9	0.2	0.1
17q25.3 (<i>RPTOR</i>)	rs1468036	C/G	4.67 (2.27-9.59)	3.82x10 ⁻⁶	0.2	5.8	36.6	2.5
CHP								
5q15 (<i>LINC02234</i>)	rs72781553	C/T	0.21 (0.10-0.44)	3.82x10 ⁻⁶	5.3	14.2	16.9	6.7
9q33.3 (<i>RALGPS1, ANGPTL2</i>)	rs7035976	T/C	0.32 (0.21-0.49)	4.24x10 ⁻⁸	38.8	35.9	71.5	48.7
EAS=East Asian; SAS=South Asian; AFR = African; EUR=European								
*Odds ratio reported for the alternate allele								
**Genome-wide significance (p<5x10 ⁻⁸)								

Table 2.2. Credible sets from fine-mapping and probability for inclusion of listed SNPs										
Set (log ₁₀ (BF))	1 (393)*		2 (392)*		3 (333)*		4 (260)		5 (105)	
Chr:BP rsID:Ref/Alt	9:127166841 rs2417050:A/G	1	9:127153352 rs12350252:C/A	1	9:126929634 rs777676:T/A	1	9:126617239 rs10819189:G/A	0.25	9:126994978 rs144622817:A/G	0.33
	NA		NA		NA		9:126623752 rs35622144:ACCT/A	0.25	9:126953456 rs7854363:C/T	0.33
	NA		NA		NA		9:126615117 rs2235057:A/G	0.25	9:126927357 rs78702307:G/A	0.33
	NA		NA		NA		9:126608676 rs10819188:G/A	0.25	NA	
*SNPs are labelled in Figure 1A										

Previous Study							Current Study					
Variant	Locus (hg38)	Alleles	Effect Allele*	OR (95% CI)	Trait	PubMed ID	CP OR (95% CI)	CP p-value	CSP OR (95% CI)	CSP p-value	CHP OR (95% CI)	CHP p-value
rs12065278	1p36.11	A/G	G	2.43 (1.66-3.56)	CP	27018472 ¹⁰	1.18 (0.92-1.52)	2.00E-01	1.22 (0.89-1.67)	2.25E-01	1.63 (1.04-2.57)	3.25E-02
rs481931	1p22.1	G/T	G	1.25	CLP	28232668 ¹²	1.25 (1.01-1.54)	3.96E-02	1.22 (0.93-1.61)	1.61E-01	1.08 (0.70-1.61)	7.51E-01
rs4839542	1p13.3	C/T	C	0.83	CLP	28232668 ¹²	1.23 (0.95-1.59)	1.14E-01	1.22 (0.88-1.72)	2.33E-01	1.96 (1.20-3.12)	5.27E-03
rs2235371	1q32.2	C/T	T	1.49 (1.37-1.61)	CL/P	25775280 ³⁰	NA	NA	1.38 (1.0-1.89)	4.60E-02	1.53 (1.0-2.33)	4.64E-02
rs6540559	1q32.2	G/A	A	1.67 (1.47-1.85)	CL/P	28054174 ¹¹	1.24 (1.01-1.52)	3.57E-02	1.30 (1.0-1.70)	5.11E-02	1.48 (0.99-2.2)	5.62E-02
rs75477785	1q32.2	T/G	G	1.75 (1.54-2.04)	CL/P	28054174 ¹¹	NA	NA	1.38 (1.0-1.9)	5.16E-02	1.61 (1.04-2.48)	3.10E-02
rs72741048	1q32.2	A/T	T	1.31	CP	31609978 ¹³	1.35 (1.09-1.66)	5.58E-03	1.28 (0.97-1.67)	7.55E-02	1.67 (1.15-2.44)	6.85E-03
rs11119394	1q32.2	A/G	G	1.30	CP	32758111 ¹⁵	NA	NA	1.48 (1.09-2.01)	1.16E-02	1.46 (0.98-2.18)	5.87E-02
rs3815854	2q35	C/T	C	0.80 (0.74-0.87)	CL/P	25775280 ³⁰	1.25 (1.03-1.52)	2.45E-02	1.35 (1.06-1.72)	1.38E-02	1.19 (0.83-1.72)	3.53E-01
rs9347594	6q26	T/C	T	2.22 (1.60-3.08)	CP	27018472 ¹⁰	NA	NA	1.11 (0.83-1.47)	4.77E-01	1.49 (1.0-2.22)	4.77E-02
rs11774066	8q22.2	C/T	C	0.86	CLP	28232668 ¹²	1.28 (1.03-1.61)	2.44E-02	1.16 (0.87-1.54)	3.07E-01	1.09 (0.741.59)	6.92E-01
rs1487022	8q22.2	G/T	G	1.17	CLP	28232668 ¹²	1.45 (1.16-1.85)	1.36E-03	1.47 (1.08-2.0)	1.58E-02	1.14 (0.76-1.72)	5.27E-01

rs4246129	8q24.3	C/G	G	2 (1.47-2.63)	CP	28054174 ¹¹	1.28 (1.03-1.61)	2.75E-02	1.47 (1.1-1.97)	8.65E-03	1.16 (0.77-1.74)	4.73E-01
rs7928246	11q22.3	A/G	<u>A</u>	0.41	CP	32758111 ¹⁵	1.77 (1.05-2.99)	2.95E-02	2.18 (1.07-4.45)	2.80E-02	NA	NA
rs730643	14q13.3	G/A	A	1.35	CP	31609978 ¹³	1.34 (1.09-1.66)	5.85E-03	1.38 (1.05-1.82)	2.18E-02	1.60 (1.05-2.44)	2.77E-02
rs4901118	14q22.1	G/A	G**	NR	CL/P	28087736 ³¹	1.41 (1.16-1.72)	4.57E-04	1.45 (1.12-1.89)	4.04E-03	1.19 (0.84-1.69)	3.29E-01
rs152745	16p12.2	G/A	<u>G</u>	0.46 (0.33-0.64)	CP	27018472 ¹⁰	NA	NA	1.19 (0.88-1.59)	2.57E-01	1.61 (1.04-2.50)	2.91E-02
rs57933945	16p12.1	C/T	C	2.86 (1.85-4.35)	CP	28054174 ¹¹	1.92 (1.32-2.86)	6.25E-04	2.5 (1.45-4.35)	6.70E-04	2.22 (1.04-4.55)	3.39E-02
rs9911652	17p13.1	C/T	<u>C</u>	0.64 (0.55-0.74)	CL/P	28054174 ¹¹	1.33 (1.01-1.75)	3.94E-02	1.3 (0.92-1.85)	1.35E-01	1.37 (0.85-2.17)	1.92E-01
rs3785888	17q21.32	C/T	<u>T</u>	0.88 (0.58-0.97)	CL/P	28054174 ¹¹	1.28 (1.06-1.55)	9.13E-03	1.19 (0.93-1.51)	1.59E-01	1.08 (0.76-1.53)	6.60E-01
rs1838105	17q21.32	A/G	<u>G</u>	0.82	CLP	28232668 ¹²	1.44 (1.18-1.75)	2.40E-04	1.34 (1.05-1.72)	2.00E-02	1.32 (0.9-1.93)	1.51E-01
rs227731	17q22	T/G	<u>G</u>	0.81 (0.72-0.89)	CL/P	28054174 ¹¹	1.30 (1.06-1.59)	1.07E-02	1.27 (0.97-1.64)	8.24E-02	1.59 (1.08-2.33)	1.77E-02

*Effect allele is reported for CPSeq data: underlines indicate opposite effect reported in original publication.

**Effect allele for original publication is not reported.

Bold indicates $p < 0.05$

NA indicates absence of data at this SNP location in the CPSeq data.

NR indicates absence of data from original publication.

SUPPLEMENTAL MATERIAL

Figure S2.1: Principal Components for Genetic Ancestry. Principal components 1 vs 2 (top) and 2 vs 3 (bottom) demonstrating group separation by self-reported race and by country of origin.

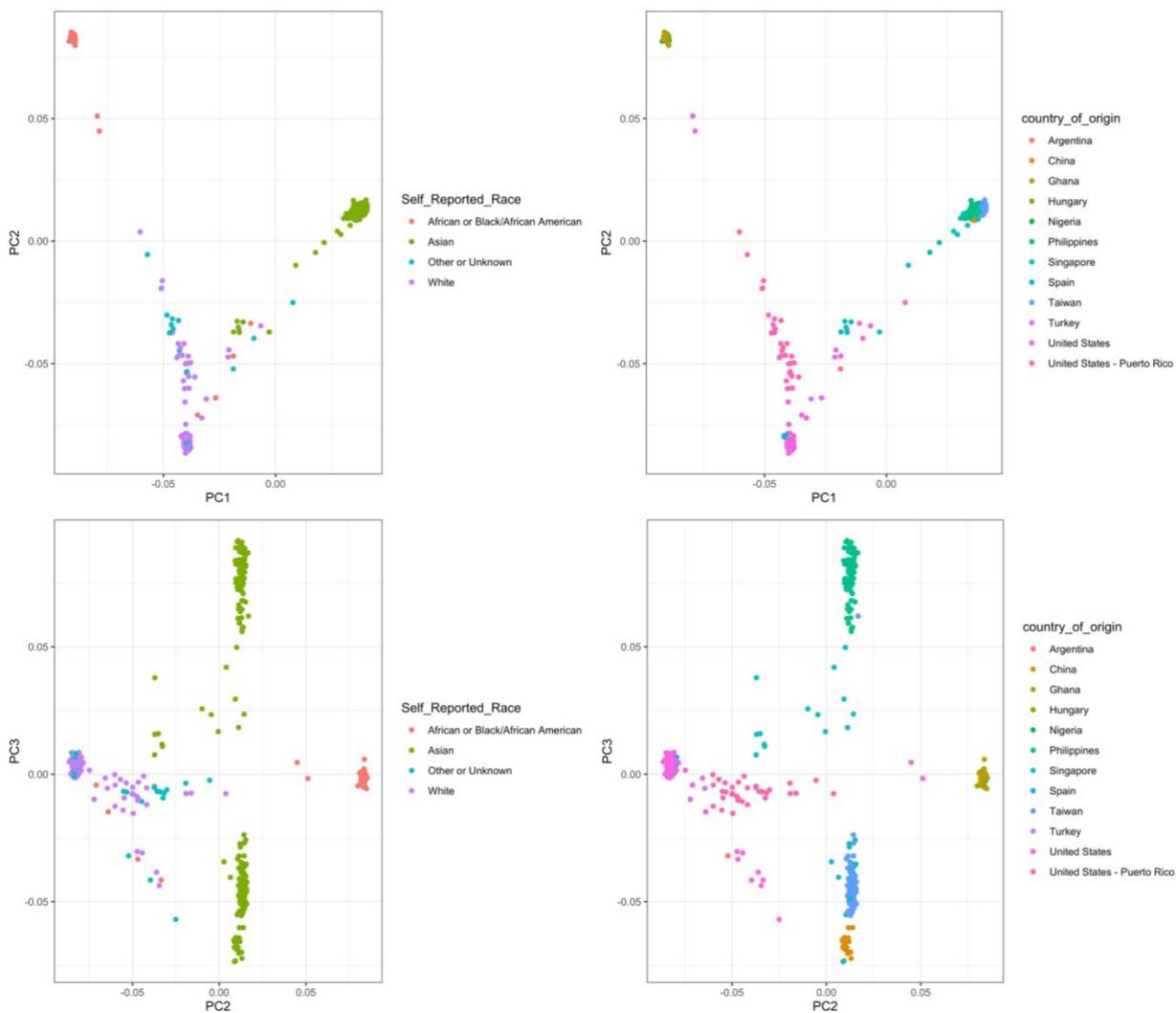


Figure S2.2: Asian-Ancestry Specific Analysis. A) Asian ancestry as determined by PC grouping – pink dots represent probands of Asian descent (n=262). B) qqplot for Asian ancestry-specific TDT. C) Manhattan plot for proband analysis. The blue line represents the suggestive threshold (5×10^{-6}) and the red line represents genome-wide significance (5×10^{-8}).

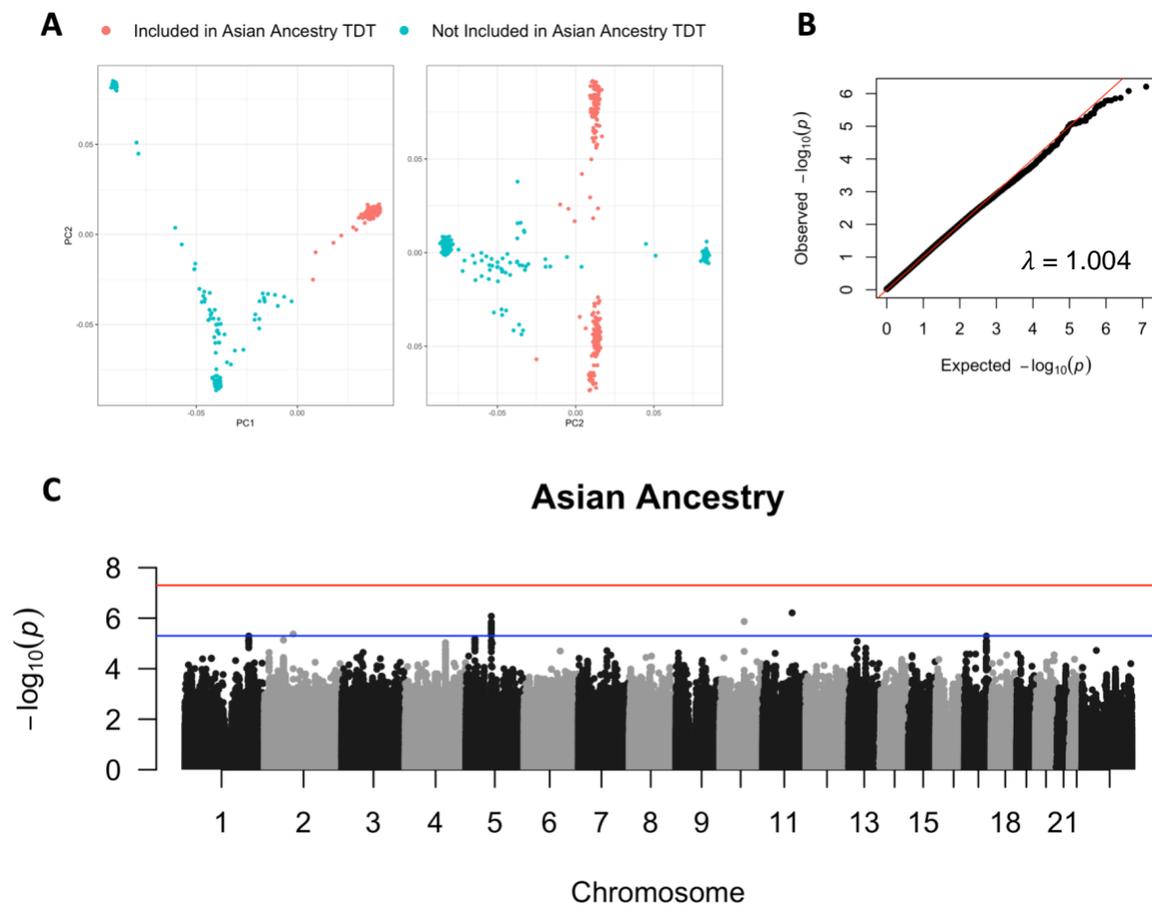


Figure S2.3: Any Cleft Palate Manhattan & QQplots. Manhattan plot (left) and qqplot (right) for the ACP analysis. The blue line represents the suggestive threshold (5×10^{-6}) and the red line represents genome-wide significance (5×10^{-6}).

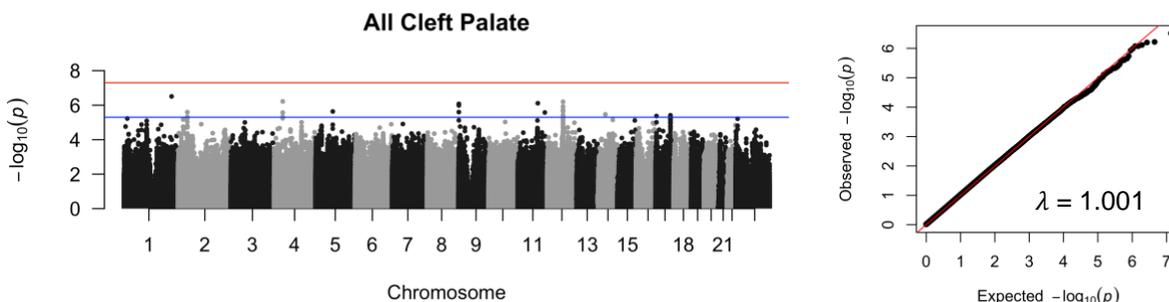


Figure S2.4: Soft Cleft Palate Manhattan & QQplots. Manhattan plot (left) and qqplot (right) for the CSP analysis. The blue line represents the suggestive threshold (5×10^{-6}) and the red line represents genome-wide significance (5×10^{-6}).

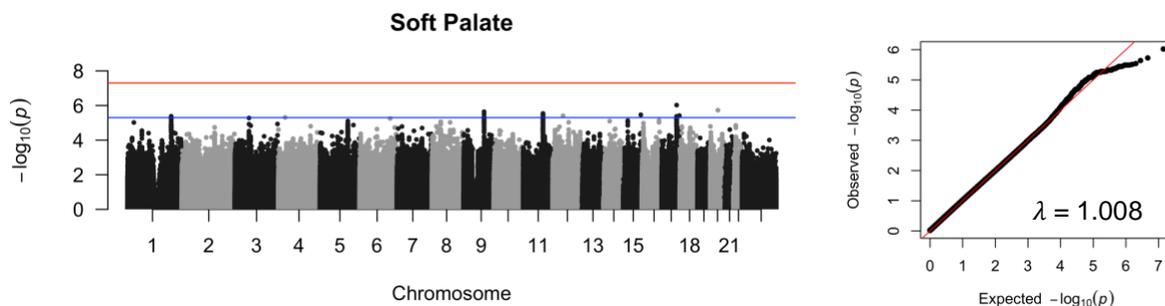


Figure S2.5: Hard Cleft Palate Manhattan & QQplots. Manhattan plot (left) and qqplot (right) for the CHP analysis. The blue line represents the suggestive threshold (5×10^{-6}) and the red line represents genome-wide significance (5×10^{-6}).

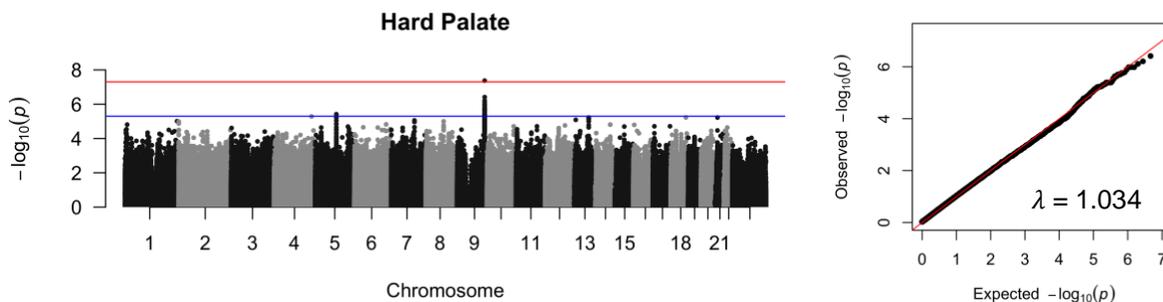


Table S2.1: CPSeq Cohort Description							
Cleft Type	Male						
	Asian	Black/African American	White	Other	Unknow n	Total Male	
Both	20	6	8	1	3	38	
CP - Type Unknown	20	8	10	0	9	47	
Hard Palate	16	0	3	0	0	19	
SMCP	2	6	6	0	0	14	
Soft Palate	37	3	13	0	2	55	
Total Male	95	23	40	1	14	173	
Cleft Type	Female						
	Asian	Black/African American	White	Other	Unknow n	Total Female	
Both	31	2	8	0	3	44	
CP - Type Unknown	38	24	15	0	9	86	
Hard Palate	21	0	3	0	0	24	
SMCP	3	7	1	0	0	11	
Soft Palate	71	5	14	1	6	97	
Total Female	164	38	41	1	18	262	
Total Cases	259	61	81	2	32	435	

Table S2.2: <i>Angptl2</i> gene-specific primers	
Gene/Direction	Sequence
<i>Angptl2</i> -fwd	CTGCAGAGTCTTCCAATCAG
<i>Angptl2</i> -rev + T7 leader	CGATGTTAATACGACTCACTATAGGGATTCTGTGGAGGCTAAAGTATG

Table S2.3: RT-qPCR primers	
Gene/Direction	Sequence
<i>Gapdh</i> -fwd	AACACTGAGCATCTCCCTCACA
<i>Gapdh</i> -rev	GGTGGGTGCAGCGAACTTTATT
<i>Angptl2</i> -fwd	TCTCTGTCCCTACTGCGTTCTT
<i>Angptl2</i> -rev	GGCGATATTGGCAGTTGTGTCT
<i>Runx2</i> -fwd	GGACCCACACAGCCATGTAA
<i>Runx2</i> -rev	GGCAACGTTCCCTAACCTGAAA

Table S2.4: All SNPs tested for replication

Current version is modified – original table available for viewing at

<https://www.sciencedirect.com/science/article/pii/S2666247723000660>

Variant	Locus	CPSeq Ref/Alt	Rep. Effect Allele	P value	Reported OR (95% CI)	Trait	Author (PMID)	CP OR	CP pvalue	CSP OR	CSP pvalue	CHP_OR	CHP pvalue
1:11913831 rs7554547	1p36.22	A/G	A	5.00E-07	1.16 (NR)	CLP	Yu Y (28232668)	1.09 (0.9-1.32)	0.3983	0.97 (0.75-1.26)	0.8435	0.83 (0.57-1.21)	0.3359
1:18646282 rs9439713	1p36.13	G/A	A	6.00E-13	1.48 (1.27-1.73)	CL/P	Leslie EJ (28054174)	NA	NA	0.94 (0.64-1.39)	0.7653	0.96 (0.56-1.67)	0.8886
1:18649995 rs9439714	1p36.13	T/C	C	8.00E-10	1.47 (1.31-1.65)	CL/P	Leslie EJ (28054174)	NA	NA	0.93 (0.62-1.42)	0.7505	0.96 (0.55-1.68)	0.8864
1:18653380 rs742071	1p36.13	G/T	T	4.00E-07	1.4 (NR)	CLP	Yu Y (28232668)	NA	NA	0.95 (0.66-1.37)	0.7778	1 (0.59-1.69)	1
1:19691584 rs11582254	1p36.13	T/C	C	9.00E-07	1.16 (NR)	CLP	Yu Y (28232668)	0.94 (0.76-1.16)	0.5514	1.08 (0.83-1.41)	0.5467	1 (0.69-1.45)	1
1:24550967 rs12065278	1p36.11	A/G	G	6.00E-06	2.43 (1.66-3.56)	CP	Leslie EJ (27018472)	1.18 (0.92-1.52)	0.2004	1.22 (0.89-1.67)	0.2253	1.63 (1.04-2.57)	0.03254
1:94092554 rs66515264	1p22.1	G/T	T	4.00E-17	1.51 (1.31-1.73)	CL/P	Leslie EJ (28054174)	0.94 (0.72-1.24)	0.6774	0.94 (0.66-1.33)	0.7237	0.91 (0.56-1.48)	0.7098
1:94104460 rs481931	1p22.1	G/T	G	1.00E-12	1.25 (NR)	CLP	Yu Y (28232668)	0.8 (0.65-0.99)	0.03956	0.82 (0.62-1.08)	0.1614	0.93 (0.62-1.42)	0.7505
1:110612973 rs4839542	1p13.3	C/T	T	7.00E-06	1.2 (NR)	CLP	Yu Y (28232668)	0.81 (0.63-1.05)	0.1135	0.82 (0.58-1.14)	0.2334	0.51 (0.32-0.83)	0.005272
1:209790735 rs2235371	1q32.2	C/T	C	9.00E-22	1.49 (1.37-1.61)	CL/P	Sun Y (25775280)	NA	NA	1.38 (1-1.89)	0.04602	1.53 (1-2.33)	0.0464

1:209808680 rs6540559	1q32.2	G/A	G	6.00E-16	1.67 (1.47-1.85)	CL/P	Leslie EJ (28054174)	1.24 (1.01-1.52)	0.0356 9	1.3 (1-1.7)	0.0510 9	1.48 (0.99-2.2)	0.0561 9
1:209811125 rs75477785	1q32.2	T/G	T	4.00E-29	1.75 (1.54-2.04)	CL/P	Leslie EJ (28054174)	NA	NA	1.38 (1-1.9)	0.0515 8	1.61 (1.04-2.48)	0.0310 3
1:209815747 rs72741048	1q32.2	A/T	T	3.00E-15	1.31 (NR)	CP	Huang L (31609978)	1.35 (1.09-1.66)	0.0055 78	1.28 (0.97-1.67)	0.0755 2	1.67 (1.15-2.44)	0.0068 46
1:209875474 rs2064163	1q32.2	G/T	G	9.00E-19	1.3 (NR)	CLP	Yu Y (28232668)	1.18 (0.97-1.45)	0.1016	1.18 (0.9-1.53)	0.2291	1.12 (0.77-1.63)	0.5637
1:210021340 rs11119394	1q32.2	A/G	G	3.00E-13	1.3 (NR)	CP	He M (32758111)	NA	NA	1.48 (1.09-2.01)	0.0116 2	1.46 (0.98-2.18)	0.0586 8
2:9832313 rs287982	2p25.1	C/T	T	6.00E-09	1.22 (NR)	CLP	Yu Y (28232668)	1.04 (0.84-1.29)	0.7055	0.95 (0.72-1.25)	0.7256	0.96 (0.65-1.42)	0.8415
2:16548089 rs7566780	2p24.2	A/G	G	4.00E-09	1.2 (1.08-1.34)	CL/P	Leslie EJ (28054174)	1.1 (0.89-1.36)	0.3869	1.05 (0.8-1.39)	0.7243	1.16 (0.77-1.75)	0.4679
2:16552660 rs7552	2p24.2	A/G	G	6.00E-22	1.37 (NR)	CLP	Yu Y (28232668)	1.13 (0.91-1.41)	0.2512	1.12 (0.85-1.48)	0.4355	1.07 (0.71-1.62)	0.7505
2:41954539 rs6740960	2p21	A/T	T	6.00E-13	NR	CL/P	Ludwig KU (28087736)	NA	NA	NA	NA	0.59 (0.3-1.17)	0.1282
2:43312986 rs7590268	2p21	T/G	NR	2.00E-06	NR	CL/P	Howe LJ (30067744)	1.04 (0.72-1.5)	0.8514	1.04 (0.6-1.8)	0.8886	1.78 (0.79-4.02)	0.1615
2:147748280 rs74819164	2q22.3	A/G	NR	6.00E-07	NR	CP	Butali A (30452639)	0.8 (0.6-1.06)	0.1226	0.75 (0.52-1.09)	0.134	0.61 (0.35-1.06)	0.0741 5
2:219789274 rs6734545	2q35	C/A	A	8.00E-07	1.2 (NR)	CLP	Yu Y (28232668)	0.89 (0.66-1.2)	0.451	0.73 (0.5-1.08)	0.1167	0.89 (0.56-1.43)	0.6326
2:219804017 rs3815854	2q35	C/T	T	3.00E-07	1.25 (1.15-1.35)	CL/P	Sun Y (25775280)	0.8 (0.66-0.97)	0.0244 5	0.74 (0.58-0.94)	0.0138 2	0.84 (0.58-1.21)	0.3532
2:232687954 rs77866552	2q37.1	G/A	G	7.00E-07	1.56 (NR)	CLP	Yu Y (28232668)	NA	NA	0.63 (0.31-1.3)	0.2087	1.29 (0.48-3.45)	0.6171
3:43215252 rs6791526	3p22.1	C/T	T	2.00E-10	1.46 (NR)	CP	Huang L (31609978)	1.06 (0.79-1.43)	0.7038	0.92 (0.61-1.38)	0.6799	1 (0.53-1.89)	1
3:99784884 rs1038294	3q12.1	A/G	G	4.00E-07	1.23 (NR)	CLP	Yu Y (28232668)	0.91 (0.71-1.18)	0.4768	0.87 (0.62-1.21)	0.4014	1.09 (0.68-1.73)	0.7218

3:99929006 rs1688766	3q12.1	G/A	G	5.00E-07	1.23 (NR)	CLP	Yu Y (28232668)	1.16 (0.9-1.49)	0.2472	1.25 (0.9-1.75)	0.1794	0.97 (0.61-1.55)	0.9055
3:99972678 rs68092024	3q12.1	C/T	T	5.00E-08	1.35 (1.19-1.52)	CL/P	Leslie EJ (28054174)	0.85 (0.68-1.05)	0.1366	0.88 (0.67-1.16)	0.3639	0.93 (0.64-1.35)	0.7029
3:99982290 rs7616988	3q12.1	G/A	A	3.00E-06	1.56 (1.32-1.85)	CL/P	Leslie EJ (28054174)	0.87 (0.68-1.11)	0.2549	0.83 (0.6-1.14)	0.2498	1.06 (0.67-1.67)	0.8162
3:189835583 rs76479869	3q28	C/T	T	1.00E-08	1.79 (1.41-2.29)	CL/P	Leslie EJ (28054174)	1.29 (0.76-2.2)	0.3452	1.31 (0.68-2.52)	0.4111	2 (0.68-5.85)	0.1967
3:197204950 rs7649443	3q29	T/C	C	9.00E-06	1.16 (NR)	CLP	Yu Y (28232668)	NA	NA	0.84 (0.65-1.1)	0.2185	0.75 (0.52-1.08)	0.116
4:3503721 rs3468	4p16.3	G/A	G	5.00E-11	1.26 (NR)	CP	Huang L (31609978)	1.1 (0.9-1.33)	0.3451	1.14 (0.89-1.45)	0.3135	0.88 (0.62-1.25)	0.4725
4:4817198 rs1907989	4p16.2	A/G	G	2.00E-08	1.18 (NR)	CLP	Yu Y (28232668)	1.03 (0.85-1.25)	0.763	0.96 (0.74-1.24)	0.7453	1 (0.68-1.47)	1
4:76628406 rs6838801	4q21.1	G/T	T	6.00E-08	1.24 (1.11-1.38)	CL/P	Leslie EJ (28054174)	0.86 (0.7-1.06)	0.1599	0.83 (0.64-1.07)	0.1469	0.87 (0.59-1.28)	0.4904
4:116802931 rs10009620	4q26	C/T	T	4.00E-06	2.34 (1.63-3.35)	CP	Leslie EJ (27018472)	0.99 (0.8-1.22)	0.9136	1.05 (0.8-1.38)	0.7256	0.98 (0.65-1.47)	0.9174
4:123789429 rs7692299	4q28.1	C/T	T	9.00E-07	1.32 (NR)	CLP	Yu Y (28232668)	0.87 (0.62-1.21)	0.4014	0.92 (0.59-1.45)	0.7324	1.1 (0.59-2.06)	0.7518
4:123985102 rs908822	4q28.1	C/T	T	4.00E-08	1.31 (NR)	CLP	Yu Y (28232668)	1.09 (0.68-1.73)	0.7218	1.19 (0.67-2.13)	0.5553	1.67 (0.61-4.59)	0.3173
4:156387152 rs72688980	4q32.1	A/G	G	2.00E-08	1.46 (NR)	CP	Huang L (31609978)	0.99 (0.77-1.28)	0.948	0.89 (0.64-1.24)	0.499	1 (0.6-1.66)	1
5:36048294 rs1287275	5p13.2	T/C	T	8.00E-06	2 (1.28-3.12)	CP	Leslie EJ (28054174)	0.97 (0.67-1.39)	0.8514	0.92 (0.57-1.47)	0.718	1.06 (0.54-2.1)	0.8618
5:44068744 rs10462065	5p12	C/A	A	1.00E-08	1.22 (NR)	CLP	Yu Y (28232668)	0.9 (0.7-1.16)	0.4024	0.96 (0.7-1.32)	0.8059	0.83 (0.56-1.24)	0.3657
5:68187904 rs6449957	5q13.1	T/C	T	2.00E-07	1.18 (1.05-1.32)	CL/P	Leslie EJ (28054174)	1.05 (0.85-1.29)	0.6698	0.93 (0.71-1.22)	0.5791	1.06 (0.71-1.58)	0.7607

5:156329289 rs11744154	5q33.3	A/G	A	5.00E-06	2.14 (1.53-2.98)	CP	Leslie EJ (27018472)	0.99 (0.8-1.22)	0.9131	1.03 (0.77-1.38)	0.8254	0.8 (0.54-1.19)	0.2737
5:156372456 rs7732608	5q33.3	G/A	G	4.00E-06	1.75 (1.38-2.22)	CP	Leslie EJ (27018472)	0.95 (0.76-1.19)	0.6506	0.93 (0.7-1.23)	0.6144	0.9 (0.6-1.35)	0.6041
5:158560211 rs2963482	5q33.3	A/G	NR	8.00E-06	NR	CL/P	Ludwig KU (28087736)	0.97 (0.79-1.19)	0.7551	0.93 (0.72-1.21)	0.5913	1.08 (0.76-1.53)	0.6573
5:173986033 rs7727165	5q35.2	C/A	A	6.00E-06	1.58 (NR)	CP	Huang L (31609978)	0.92 (0.72-1.17)	0.4943	1.1 (0.81-1.49)	0.5371	0.79 (0.49-1.27)	0.332
6:9469005 rs9381107	6p24.3	G/A	G	3.00E-09	1.2 (NR)	CLP	Yu Y (28232668)	0.98 (0.79-1.21)	0.8273	1.02 (0.78-1.34)	0.8902	1.13 (0.76-1.68)	0.5445
6:31172964 rs1265158	6p21.33	C/G	NR	3.00E-06	NR	CL/P	Ludwig KU (28087736)	0.91 (0.74-1.13)	0.3897	1 (0.76-1.32)	1	1.06 (0.72-1.56)	0.7675
6:162166860 rs9347594	6q26	T/C	T	2.00E-06	2.22 (1.6-3.08)	CP	Leslie EJ (27018472)	NA	NA	0.9 (0.68-1.2)	0.4773	0.67 (0.45-1)	0.0476 7
7:127615231 rs698406	7q32.1	C/G	C	6.00E-09	1.23 (NR)	CP	Huang L (31609978)	NA	NA	NA	NA	1.16 (0.77-1.75)	0.4679
8:13601285 rs17227506	8p22	C/T	C	3.00E-06	1.15 (NR)	CLP	Yu Y (28232668)	1.15 (0.95-1.4)	0.1496	1.08 (0.85-1.39)	0.5287	1.18 (0.83-1.67)	0.3692
8:17923748 rs12718385	8p22	T/C	T	9.00E-06	2.27 (1.581-3.261)	CP	Leslie EJ (27018472)	1.01 (0.83-1.22)	0.9221	0.93 (0.73-1.19)	0.5745	0.84 (0.6-1.19)	0.3328
8:24297141 rs11992342	8p21.2	C/T	T	4.00E-06	3.77 (2.14-6.62)	CP	Leslie EJ (27018472)	0.71 (0.49-1.02)	0.0656	0.72 (0.45-1.17)	0.1854	0.5 (0.21-1.17)	0.1025
8:32476054 rs1878918	8p12	G/C	C	3.00E-07	1.29 (1.15-1.45)	CL/P	Leslie EJ (28054174)	0.88 (0.71-1.08)	0.2156	0.88 (0.67-1.16)	0.3708	0.88 (0.59-1.31)	0.5445
8:38411996 rs13317	8p11.23	T/C	T	4.00E-08	1.18 (NR)	CLP	Yu Y (28232668)	1.02 (0.83-1.27)	0.8283	1.06 (0.81-1.39)	0.6759	1.12 (0.76-1.66)	0.5525
8:76593073 rs10808812	8q21.13	T/C	NR	1.00E-06	NR	CL/P	Ludwig KU (28087736)	1.03 (0.83-1.27)	0.7853	0.98 (0.75-1.29)	0.8886	0.98 (0.63-1.5)	0.9126

8:82347996 rs7000204	8q21.13	T/C	C	5.00E-06	1.89 (1.43-2.5)	CP	Leslie EJ (28054174)	1.07 (0.88-1.3)	0.4839	1.09 (0.84-1.4)	0.5203	1.11 (0.77-1.59)	0.5807
8:87856112 rs12543318	8q21.3	C/A	C	9.00E-12	1.3 (1.17-1.45)	CL/P	Leslie EJ (28054174)	0.97 (0.8-1.18)	0.7624	1.03 (0.8-1.33)	0.7946	1 (0.71-1.41)	1
8:94389037 rs12681366	8q22.1	T/C	T	2.00E-10	1.2 (NR)	CLP	Yu Y (28232668)	NA	NA	1.01 (0.78-1.3)	0.9478	0.79 (0.54-1.15)	0.2172
8:94529074 rs957448	8q22.1	A/G	A	1.00E-12	1.23 (NR)	CLP	Yu Y (28232668)	0.86 (0.71-1.05)	0.1455	0.92 (0.72-1.19)	0.5203	0.75 (0.51-1.11)	0.147
8:98443083 rs11774066	8q22.2	C/T	T	3.00E-07	1.16 (NR)	CLP	Yu Y (28232668)	0.78 (0.62-0.97)	0.0244	0.86 (0.65-1.15)	0.3072	0.92 (0.63-1.36)	0.6921
8:99517598 rs1487022	8q22.2	G/T	G	8.00E-08	1.17 (NR)	CLP	Yu Y (28232668)	0.69 (0.54-0.86)	0.0013	0.68 (0.5-0.93)	0.0158	0.88 (0.58-1.32)	0.5271
8:128921474 rs72728734	8q24.21	A/G	G	6.00E-26	2.06 (1.81-2.35)	CL/P	Leslie EJ (28054174)	1 (0.63-1.58)	1	0.9 (0.49-1.68)	0.7518	0.79 (0.36-1.73)	0.5485
8:128938598 rs7017252	8q24.21	C/T	T	8.00E-16	1.6 (NR)	CLP	Yu Y (28232668)	1.01 (0.76-1.34)	0.9423	0.88 (0.6-1.28)	0.4986	0.84 (0.47-1.5)	0.5553
8:128963890 rs55658222	8q24.21	G/A	A	8.00E-44	2.13 (1.78-2.54)	CL/P	Leslie EJ (28054174)	1.04 (0.7-1.56)	0.8366	0.93 (0.54-1.59)	0.7815	0.87 (0.41-1.82)	0.7055
8:140502973 rs4246129	8q24.3	C/G	G	5.00E-06	2 (1.47-2.63)	CP	Leslie EJ (28054174)	1.28 (1.03-1.61)	0.0275	1.47 (1.1-1.97)	0.0086	1.16 (0.77-1.74)	0.4726
8:143423963 rs13267937	8q24.3	G/C	G	1.00E-06	1.47 (NR)	CP	Huang L (31609978)	1.02 (0.82-1.26)	0.8721	1.01 (0.77-1.33)	0.9443	1 (0.69-1.44)	1
9:88074295 rs11142081	9q22.1	T/G	T	2.00E-06	NR	CL/P	Carlson JC (30277614)	1.03 (0.78-1.36)	0.8299	1.19 (0.83-1.7)	0.3573	1.22 (0.66-2.28)	0.5271
9:89421512 rs4132699	9q22.2	A/C	NR	9.00E-08	NR	CL/P	Ludwig KU (28087736)	1 (0.82-1.23)	1	1.01 (0.78-1.31)	0.9466	1.13 (0.78-1.62)	0.5175
9:89594672 rs7871395	9q22.2	C/T	A	6.00E-09	1.21 (NR)	CLP	Yu Y (28232668)	0.99 (0.8-1.24)	0.9554	1.02 (0.77-1.35)	0.8875	0.8 (0.52-1.24)	0.3232
9:89595154 rs12375983	9q22.2	G/A	NR	1.00E-06	NR	CL/P	Ludwig KU (28087736)	1.22 (0.76-1.95)	0.4061	0.87 (0.48-1.58)	0.6473	0.57 (0.17-1.95)	0.3657
9:92220837 rs147121504	9q22.31	C/T	C	2.00E-06	1.47 (NR)	CP	Huang L (31609978)	0.97 (0.68-1.38)	0.8551	0.86 (0.53-1.39)	0.5413	1.33 (0.63-2.82)	0.4497

9:95497421 rs10512248	9q22.32	T/G	T	5.00E-10	1.22 (NR)	CLP	Yu Y (28232668)	0.97 (0.79-1.2)	0.7924	0.93 (0.71-1.21)	0.5845	0.88 (0.6-1.28)	0.4945
9:95524196 rs28434654	9q22.32	A/G	NR	5.00E-07	NR	CP	Butali A (30452639)	1.07 (0.86-1.33)	0.543	0.93 (0.71-1.23)	0.6249	0.87 (0.58-1.29)	0.4817
9:95529166 rs3858092	9q22.32	A/C	A	3.00E-07	NR	CP	Butali A (30452639)	1.05 (0.84-1.3)	0.6924	0.89 (0.67-1.18)	0.4261	0.85 (0.57-1.26)	0.4142
9:95531283 rs28689133	9q22.32	A/C	NR	4.00E-07	NR	CP	Butali A (30452639)	1.05 (0.84-1.31)	0.6506	0.89 (0.67-1.18)	0.4261	0.85 (0.57-1.26)	0.4142
9:95532166 rs28637199	9q22.32	T/G	NR	5.00E-07	NR	CP	Butali A (30452639)	1.05 (0.84-1.3)	0.6924	0.89 (0.67-1.18)	0.4261	0.85 (0.57-1.26)	0.4142
9:95532733 rs3847316	9q22.32	T/C	NR	5.00E-07	NR	CP	Butali A (30452639)	1.03 (0.83-1.29)	0.7768	0.87 (0.65-1.16)	0.3443	0.85 (0.57-1.26)	0.4142
9:95549844 rs2405373	9q22.32	G/A	NR	4.00E-07	NR	CP	Butali A (30452639)	1.05 (0.85-1.3)	0.626	0.95 (0.72-1.25)	0.7256	0.83 (0.56-1.21)	0.3314
9:96255802 rs2479824	9q22.32	C/T	C	5.00E-06	1.20 (NR)	CP	He M (32758111)	0.95 (0.74-1.21)	0.6586	1.03 (0.74-1.43)	0.8658	1.18 (0.74-1.88)	0.4795
9:116337594 rs13294988	9q33.1	C/T	C	7.00E-06	1.14 (NR)	CLP	Yu Y (28232668)	1.07 (0.88-1.29)	0.5203	1.09 (0.84-1.4)	0.5203	1.12 (0.78-1.61)	0.5211
10:57804720 rs72804706	10q21.1	C/T	C	4.00E-08	NR	CL/P	Carlson JC (30277614)	0.91 (0.67-1.23)	0.5371	0.98 (0.67-1.44)	0.9215	1.04 (0.59-1.82)	0.8864
10:96354046 rs11597348	10q24.1	A/T	NR	4.00E-07	NR	CP	Butali A (30452639)	1.27 (0.81-2.01)	0.2987	0.95 (0.51-1.78)	0.8728	1.14 (0.41-3.15)	0.7963
10:97935437 rs474558	10q24.2	C/A	A	5.00E-06	1.17 (NR)	CLP	Yu Y (28232668)	1.08 (0.86-1.35)	0.4969	1.01 (0.76-1.35)	0.9414	0.98 (0.65-1.48)	0.9156
10:11706804 9 rs7078160	10q25.3	G/A	A	3.00E-10	1.29 (1.19-1.39)	CL/P	Sun Y (25775280)	1.12 (0.91-1.38)	0.2864	0.95 (0.72-1.25)	0.7256	0.84 (0.57-1.24)	0.3705
10:11708678 3 rs10886040	10q25.3	C/G	G	9.00E-11	1.31 (1.17-1.46)	CL/P	Leslie EJ (28054174)	1.12 (0.91-1.38)	0.2905	0.95 (0.72-1.25)	0.7269	0.82 (0.56-1.21)	0.3221
10:11713372 0 rs6585429	10q25.3	A/G	A	7.00E-13	1.23 (NR)	CLP	Yu Y (28232668)	0.89 (0.73-1.08)	0.2266	1 (0.78-1.29)	1	1.41 (0.97-2.06)	0.0713 3
11:33859470 rs3740617	11p13	T/C	T	6.00E-06	2.76 (1.747-4.361)	CP	Leslie EJ (27018472)	1.04 (0.86-1.26)	0.6602	1.1 (0.86-1.4)	0.4533	0.93 (0.64-1.35)	0.7029

11:10764366 5 rs7928246	11q22.3	A/G	A	2.00E-06	2.44 (NR)	CP	He M (32758111)	1.77 (1.05-2.99)	0.02951	2.18 (1.07-4.45)	0.02799	NA (NA-NA)	NA
12:52952966 rs3741442	12q13.1 3	C/T	C	4.00E-12	1.22 (NR)	CLP	Yu Y (28232668)	NA	NA	1.09 (0.82-1.47)	0.5488	1.16 (0.77-1.74)	0.4726
12:52963551 rs2363632	12q13.1 3	T/G	T	2.00E-07	1.37 (1.19-1.56)	CL/P	Leslie EJ (28054174)	1.18 (0.94-1.47)	0.1563	1.08 (0.81-1.45)	0.6008	1.16 (0.77-1.74)	0.4726
12:71686492 rs2304269	12q21.1	T/C	T	1.00E-12	1.23 (NR)	CLP	Yu Y (28232668)	NA	NA	1.3 (0.97-1.76)	0.0821	0.96 (0.63-1.45)	0.833
12:11097665 7 rs12229654	12q24.1 1	T/G	T	1.00E-08	1.27 (NR)	CLP	Yu Y (28232668)	NA	NA	0.64 (0.38-1.08)	0.09056	NA (NA-NA)	NA
12:11125432 0 rs4766453	12q24.1 1	C/T	A	8.00E-06	1.14 (NR)	CLP	Yu Y (28232668)	1.11 (0.91-1.36)	0.3011	1.06 (0.82-1.38)	0.6422	0.96 (0.67-1.4)	0.8501
12:11208099 9 rs11066150	12q24.1 3	G/A	G	3.00E-13	1.25 (NR)	CLP	Yu Y (28232668)	NA	NA	1.04 (0.75-1.44)	0.8046	1.35 (0.85-2.15)	0.1979
12:11248558 9 rs12229892	12q24.1 3	G/A	G	2.00E-10	1.2 (NR)	CLP	Yu Y (28232668)	NA	NA	NA	NA	1.43 (0.93-2.2)	0.1037
13:26007063 rs7329196	13q12.1 3	T/C	C	2.00E-06	1.55 (NR)	CP	Huang L (31609978)	1.03 (0.84-1.25)	0.8014	1.03 (0.8-1.34)	0.7937	0.9 (0.62-1.3)	0.5673
13:80065270 rs9545308	13q31.1	G/T	T	2.00E-09	1.29 (NR)	CLP	Yu Y (28232668)	0.94 (0.7-1.25)	0.6565	0.82 (0.56-1.19)	0.2965	1.22 (0.7-2.11)	0.4838
13:80105167 rs11841646	13q31.1	T/A	A	4.00E-10	1.33 (1.17-1.50)	CL/P	Leslie EJ (28054174)	1.06 (0.85-1.33)	0.6027	1.07 (0.8-1.43)	0.6547	1.47 (0.92-2.33)	0.1036
13:80118676 rs8001641	13q31.1	G/A	NR	4.00E-08	NR	CL/P	Howe LJ (30067744)	1 (0.79-1.27)	1	0.99 (0.72-1.35)	0.9368	1.39 (0.86-2.26)	0.179
13:80197272 rs2183342	13q31.1	C/G	C	8.00E-06	2.61 (1.686-4.031)	CP	Leslie EJ (27018472)	0.97 (0.79-1.2)	0.793	0.93 (0.72-1.21)	0.5913	0.87 (0.61-1.23)	0.4245
13:98743340 rs4646211	13q32.3	T/C	C	5.00E-12	1.28 (NR)	CP	Huang L (31609978)	1.08 (0.86-1.36)	0.484	1 (0.76-1.32)	1	1.16 (0.78-1.72)	0.4772
14:36769942 rs730643	14q13.3	G/A	A	3.00E-16	1.35 (NR)	CP	Huang L (31609978)	1.34 (1.09-1.66)	0.005851	1.38 (1.05-1.82)	0.02181	1.6 (1.05-2.44)	0.02771
14:37068147 rs10133673	14q13.3	T/C	C	1.00E-06	1.17 (NR)	CLP	Yu Y (28232668)	1.12 (0.88-1.43)	0.355	1.3 (0.96-1.77)	0.08963	0.85 (0.56-1.3)	0.4581

14:51372927 rs7148069	14q22.1	C/T	T	2.00E-08	1.22 (NR)	CLP	Yu Y (28232668)	0.86 (0.69-1.08)	0.2025	0.82 (0.61-1.12)	0.2143	1.13 (0.73-1.74)	0.5831
14:51389391 rs4901118	14q22.1	G/A	A	7.00E-10	NR	CL/P	Ludwig KU (28087736)	0.71 (0.58-0.86)	0.0004568	0.69 (0.53-0.89)	0.004041	0.84 (0.59-1.19)	0.329
14:94913162 rs1243572	14q32.13	T/C	C	4.00E-10	1.2 (NR)	CLP	Yu Y (28232668)	NA	NA	1.19 (0.9-1.58)	0.2211	1.16 (0.77-1.75)	0.4679
14:100676553 rs730570	14q32.2	G/A	A	7.00E-10	1.28 (NR)	CP	Huang L (31609978)	0.94 (0.76-1.17)	0.5909	0.83 (0.63-1.09)	0.1866	0.83 (0.57-1.21)	0.3404
15:63020433 rs1873147	15q22.2	G/A	NR	4.00E-07	NR	CL/P	Howe LJ (30067744)	NA	NA	1.03 (0.78-1.37)	0.8282	0.78 (0.48-1.24)	0.2855
15:74452058 rs2289187	15q24.1	C/T	C	4.00E-11	1.21 (NR)	CLP	Yu Y (28232668)	1.04 (0.85-1.28)	0.6733	1.12 (0.86-1.47)	0.4098	0.96 (0.64-1.44)	0.8366
15:74596822 rs11072494	15q24.1	C/T	C	2.00E-08	1.19 (1.05-1.33)	CL/P	Leslie EJ (28054174)	1.04 (0.85-1.28)	0.711	1.12 (0.86-1.47)	0.4076	0.98 (0.65-1.47)	0.9174
15:74607159 rs6495117	15q24.1	T/C	T	6.00E-11	1.2 (NR)	CLP	Yu Y (28232668)	1.02 (0.83-1.25)	0.8738	1.09 (0.83-1.43)	0.5355	0.98 (0.65-1.46)	0.9183
15:76381375 rs3765115	15q24.3	T/C	C	2.00E-10	1.27 (NR)	CP	He M (32758111)	1 (0.82-1.23)	0.9596	1.01 (0.78-1.31)	0.9473	1.05 (0.73-1.52)	0.7778
16:3919885 rs2283487	16p13.3	A/G	A	1.00E-10	1.2 (NR)	CLP	Yu Y (28232668)	1.15 (0.95-1.39)	0.1575	1.05 (0.82-1.34)	0.7066	1.16 (0.81-1.65)	0.4171
16:3930444 rs8049367	16p13.3	T/C	C	9.00E-12	1.35 (1.25-1.47)	CL/P	Sun Y (25775280)	0.92 (0.75-1.11)	0.374	0.99 (0.77-1.28)	0.9484	0.98 (0.69-1.41)	0.927
16:23355101 rs152745	16p12.2	G/A	A	3.00E-06	2.17 (1.57-3.01)	CP	Leslie EJ (27018472)	NA	NA	0.84 (0.63-1.13)	0.2568	0.62 (0.4-0.96)	0.0291
16:25190659 rs57933945	16p12.1	C/T	C	3.00E-06	2.86 (1.85-4.35)	CP	Leslie EJ (28054174)	0.52 (0.35-0.76)	0.0006247	0.4 (0.23-0.69)	0.0006697	0.45 (0.22-0.96)	0.03389
16:28483153 rs34837	16p12.1	A/C	C	3.00E-07	1.32 (NR)	CP	He M (32758111)	1.04 (0.77-1.4)	0.8185	0.96 (0.67-1.4)	0.8501	1.2 (0.71-2.04)	0.5002
16:86482353 rs78669990	16q24.1	C/T	T	2.00E-08	1.3 (NR)	CP	Huang L (31609978)	1.06 (0.81-1.4)	0.6744	0.94 (0.66-1.34)	0.7172	1 (0.61-1.65)	1
16:87763960 rs8061677	16q24.2	C/T	C	9.00E-11	1.29 (NR)	CP	Huang L (31609978)	0.99 (0.79-1.23)	0.9101	0.93 (0.7-1.23)	0.6144	0.79 (0.53-1.19)	0.2591

17:9011376 rs2872615	17p13.1	T/C	T	9.00E-12	1.22 (NR)	CLP	Yu Y (28232668)	1.05 (0.86-1.28)	0.6439	1.04 (0.8-1.36)	0.7389	1.32 (0.91-1.91)	0.1374
17:9028802 rs4791774	17p13.1	A/G	G	5.00E-19	1.56 (1.42-1.72)	CL/P	Sun Y (25775280)	0.87 (0.7-1.08)	0.202	0.88 (0.67-1.17)	0.3889	0.79 (0.53-1.17)	0.2301
17:9039746 rs9911652	17p13.1	C/T	T	3.00E-08	1.56 (1.35-1.81)	CL/P	Leslie EJ (28054174)	0.75 (0.57-0.99)	0.03936	0.77 (0.54-1.09)	0.1345	0.73 (0.46-1.17)	0.1917
17:9044391 rs12944377	17p13.1	T/C	T	8.00E-21	1.52 (1.35-1.72)	CL/P	Leslie EJ (28054174)	NA	NA	NA	NA	1.39 (0.94-2.04)	0.09711
17:46928337 rs3785888	17q21.3 2	C/T	C	4.00E-07	1.14 (1.03-1.27)	CL/P	Leslie EJ (28054174)	1.28 (1.06-1.55)	0.009127	1.19 (0.93-1.51)	0.1593	1.08 (0.76-1.53)	0.6598
17:46931569 rs1838105	17q21.3 2	A/G	A	1.00E-11	1.22 (NR)	CLP	Yu Y (28232668)	1.44 (1.18-1.75)	0.0002399	1.34 (1.05-1.72)	0.02001	1.32 (0.9-1.93)	0.1508
17:56695877 rs227731	17q22	T/G	G	2.00E-09	1.24 (1.12-1.38)	CL/P	Leslie EJ (28054174)	0.77 (0.63-0.94)	0.01072	0.79 (0.61-1.03)	0.08235	0.63 (0.43-0.93)	0.01765
17:62999067 rs1588366	17q23.2	A/G	A	9.00E-09	1.54 (1.35-1.79)	CL/P	Leslie EJ (28054174)	0.96 (0.74-1.24)	0.7378	0.92 (0.67-1.27)	0.6195	0.86 (0.53-1.39)	0.5413
19:2050824 rs3746101	19p13.3	G/T	T	2.00E-08	NR	CL/P	Ludwig KU (28087736)	1.02 (0.78-1.32)	0.8942	1.03 (0.74-1.43)	0.8658	0.71 (0.43-1.16)	0.1724
19:19329619 rs1009136	19p13.1 1	A/G	A	3.00E-09	1.25 (NR)	CP	Huang L (31609978)	0.85 (0.7-1.04)	0.1169	0.92 (0.71-1.2)	0.5537	0.88 (0.61-1.25)	0.4689
20:40632414 rs6072081	20q12	A/G	A	2.00E-12	1.43 (1.28-1.59)	CL/P	Leslie EJ (28054174)	0.96 (0.79-1.16)	0.6609	0.96 (0.76-1.23)	0.7569	1 (0.71-1.41)	1
20:40640434 rs13041247	20q12	T/C	T	2.00E-11	1.32 (1.20-1.41)	CL/P	Sun Y (25775280)	1 (0.82-1.21)	1	0.92 (0.72-1.18)	0.5304	0.95 (0.67-1.35)	0.7884
20:40646963 rs6129653	20q12	T/C	T	9.00E-12	1.23 (NR)	CLP	Yu Y (28232668)	0.92 (0.75-1.12)	0.4068	0.88 (0.68-1.15)	0.3496	1 (0.69-1.45)	1

20:40656135 rs6029258	20q12	G/A	G	4.00E-12	1.54 (1.37-1.72)	CL/P	Leslie EJ (28054174)	0.93 (0.76-1.12)	0.4339	0.94 (0.74-1.21)	0.6547	0.98 (0.69-1.41)	0.9276
20:48172924 rs228218	20q13.1 3	G/A	G	3.00E-06	2.32 (1.632-3.311)	CP	Leslie EJ (27018472)	0.93 (0.76-1.14)	0.4726	1.02 (0.79-1.31)	0.896	1.11 (0.77-1.62)	0.5673
20:51790963 rs6126344	20q13.2	A/C	C	6.00E-08	1.18 (NR)	CLP	Yu Y (28232668)	0.9 (0.74-1.1)	0.3087	1 (0.77-1.29)	1	0.89 (0.62-1.27)	0.5175
22:19811887 rs2073764	22q11.2 1	C/T	C	2.00E-07	1.59 (NR)	CLP	Yu Y (28232668)	NA	NA	1 (0.49-2.05)	1	0.89 (0.34-2.3)	0.8084
22:29990369 rs5763674	22q12.2	C/T	C	1.00E-07	1.26 (NR)	CLP	Yu Y (28232668)	1 (0.73-1.37)	1	0.9 (0.58-1.4)	0.6547	1 (0.54-1.86)	1
22:31602626 rs2006771	22q12.2	G/A	A	3.00E-08	1.21 (NR)	CLP	Yu Y (28232668)	0.9 (0.71-1.13)	0.3474	1.02 (0.75-1.4)	0.8751	0.83 (0.53-1.31)	0.4189
22:36288285 rs5756130	22q12.3	C/T	T	6.00E-06	1.21 (NR)	CLP	Yu Y (28232668)	1.22 (0.87-1.73)	0.2524	1.15 (0.72-1.84)	0.5529	2.11 (0.96-4.67)	0.0587 8

**CHAPTER III:
Genome-wide study of gene-by-sex interactions identifies risks for cleft palate**

This chapter is adapted from a manuscript in preparation for submission for publication.

Kelsey Robinson, Randy Parrish, Wasiu Lanre Adeyemo, Terri H. Beaty, Azeez Butali, Carmen J. Buxó, Lord JJ Gowans, Jacqueline T. Hecht, Jeffrey C. Murray, Gary M. Shaw, Lina Moreno Uribe, Seth M. Weinberg, Harrison Brand, Mary L. Marazita, David J. Cutler, Michael P. Epstein, Jingjing Yang, Elizabeth J. Leslie

ABSTRACT

Structural birth defects (SBDs) affect 3-4% of all live births and, depending on the type, tend to manifest in a sex-biased manner. Orofacial clefts (OFCs) are the most common craniofacial SBD and are often divided into cleft lip with or without a cleft palate (CL/P) and cleft palate (CP). Previous studies have found sex-specific risks for CL/P, but these risks have yet to be evaluated in CP. CL/P is more common in males and CP is more frequently observed in females, so we hypothesized that there would similarly be sex-specific differences for CP. We performed sex-stratified genome-wide association studies (GWAS) based on proband sex followed by a genome-wide gene-by-sex (GxS) interaction testing. There were 13 loci significant for GxS interactions, with the top finding in *LTBP1* (OR 3.37 [2.04 - 5.56], $p=1.93 \times 10^{-6}$). *LTBP1* plays a role in regulating TGF-B bioavailability, and knockdown in both mice and zebrafish lead to craniofacial anomalies. Further, there is evidence for differential expression of *LTBP1* between males and females in both mice and humans; therefore, we performed association testing of imputed genetically regulated gene expression and the CP phenotype, finding significant association for *LTBP1* in cell cultured fibroblasts only in female probands ($p=0.0013$). Taken altogether, we show there are sex-specific risks for CP that are otherwise undetectable in a combined sex cohort, and *LTBP1* is a candidate risk gene, particularly in females.

INTRODUCTION

Structural birth defects (SBDs) occur in 3-4% of all live births (158, 159) and account for 20% of infant mortality (160, 161). There are several factors associated with increased risk for SBDs, including maternal and paternal ages, environmental exposures, and genetic variation. In general, more males are born with SBDs than females (162, 163). Although there are varying degrees of

both phenotypic and genetic heterogeneity among different SBDs, this sex discrepancy suggests inherent sex-specific risks. One such example is that of orofacial clefts (OFCs). For clefts involving the upper lip (cleft lip with or without a cleft palate, CL/P), males are affected at a rate about twice that of females, while for cleft palate (CP), females are affected more frequently (4, 16).

OFCs are a common congenital anomaly, occurring in approximately 1 in 1000 live births (4). The need for surgical intervention at an early age, advanced orthodontic care, and speech and hearing therapies contribute to a substantial individual and public health burden. As such, better understanding of their etiology can lead to prevention and/or improved health outcomes. OFCs are etiologically heterogeneous, with environmental exposures such as maternal multivitamin use (164), and maternal alcohol consumption (165) or smoking (166) contributing to occurrence risk, as well as genetic factors. CP is highly heritable with estimates around 80-90% (85, 86), (135, 136), but CP is still understudied compared to CL/P (167, 168).

Several models exist to explain how sex biases manifest in SBDs and specifically in CP. An X-linked CP with ankyloglossia phenotype has been reported with variants in *TBX22*, primarily affecting males but appearing to follow a semi-dominant X-linked inheritance pattern (169). Although seemingly counterintuitive, aberrant and incomplete X inactivation secondary to p53 loss has been suggested as a mechanism for increased incidence of neural tube defects in females (170), and there are other disorders in which female heterozygotes exhibit a wider spectrum of phenotypes as compared to more severely affected males (171, 172). Another likely contributing factor is the timing of palatal closure in embryogenesis: female embryos undergo this process approximately one week later than males (9). This delay in closure allows more time for disruption of the growth and fusion of palatal shelves compared to male embryos and may ultimately

contribute to a lower genetic liability in females (*i.e.*, higher susceptibility to CP due to genetic variation). Several genome-wide association studies (GWAS) for CP are published; however, all are sex-combined cohorts (102, 107, 108, 110, 111, 173). Stratification by sex in CL/P has led to discovery of novel associated sex-specific factors (167, 168), but no such study has been conducted for CP. Although the observed sex bias is opposite for CP, it is plausible that there are similar sex-specific risk factors. One genome-wide interaction study identified SNPs associated with environmental exposures such as maternal smoking, alcohol intake, or vitamin use (107), but none have specifically examined gene-by-sex interactions.

Taken altogether, we hypothesized that there are sex-specific risks influencing the occurrence of CP. To test this hypothesis, we first performed sex-stratified GWAS, followed by gene-by-sex (GxS) interaction analysis on 435 cleft palate case-parent trios. After identification of 13 loci with significant interactions, we also tested association of predicted gene expression with OFCs in the full cohort and by sex for the nearest genes for the significant loci.

METHODS

Study population and phenotyping

Following institutional review board (IRB) approval for each local recruitment site and coordinating centers (University of Iowa, University of Pittsburgh, and Emory University), subjects underwent phenotypic assessment. The study population was recruited from domestic and international sites in North America (United States, Puerto Rico), Europe (Spain, Turkey, Hungary), South America (Argentina), Asia (China, Singapore, Taiwan, the Philippines), and Africa (Nigeria, Ghana). In total, 435 case-parent trios were ascertained on proband affection status (e.g., cleft palate). We did not exclude subjects based on parent affection status or the

presence of additional clinical features. There were 19 trios (4.4%) with an affected parent and 46 trios (10.6%) with minor or major clinical features in addition to cleft palate. Proband sex was confirmed genetically for all cases during quality control processes.

Sample preparation and whole genome sequencing

Whole genome sequencing was performed at the Center for Inherited Disease Research (CIDR) at Johns Hopkins University (Baltimore, MD). Prior to sequencing, a Fragment Analyzer system was used to test samples for adequate quantity and quality of genomic DNA, which were then processed with an Illumina InfiniumQCArray-24v1-0 array to confirm sex, relatedness, and known duplicates. For each sample, 500-750ng of genomic DNA was sheared to 400-600bp fragments, then processed with the Kapa Hyper Prep kit for End-Repair, A-Tailing, and Ligation of IDT (Integrated DNA Technologies) unique dual- indexed adapters according to the Kapa protocol to create a final PCR-free library.

Sequencing was performed on 150bp paired-end reads on a NovaSeq 6000 platform with base calling through the Illumina Real Time Analysis software (version 3.4.4). Files were demultiplexed from binary format (BCL) to individual fastq files with Illumina Isas bcl2fastq (version 1.37.1) and aligned to the human hg38 reference sequence (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39). Alignment, variant calling, and quality control was done with the DRAGEN Germline v3.7.5 pipeline on the Illumina BaseSpace Sequence Hub platform, which produced single sample VCF files. Contamination for any cross-human sample was checked with the DRAGEN contamination detection tool. We performed genotype concordance with existing array-based genotypes using CIDRSeqSuite (version 7.5.0), and genotype concordance checks amongst replicate samples was performed in Picard

GenotypeConcordance (Picard 2019). After data quality steps, samples with at least 80% of the genome at 20X coverage or autosomal coverage at 30X underwent joint calling to generate a multi-sample VCF.

Quality control

Variants underwent quality control in the multi-sample VCF. Only variants aligning to the standard chromosomes (1-22, X, Y) were included. Variants with any flag other than “PASS” and a minor allele count (MAC) of <2 were removed. Genotypes were set to missing if the quality score was <20 or had a read depth of <10, and then variants with missingness >10% were removed. Sample-level metrics were then evaluated for heterozygous/homozygous ratio, transition/transversion (Ts/Tv) ratio, and silent/replacement rate. Outlier samples with values outside of 3 standard deviations from the cohort mean were discarded. Samples with high missing data (>5% missing) were removed.

Statistical analysis

We performed GWAS using transmission disequilibrium tests (TDT) on the full cohort and stratified by proband sex. The TDT uses the rate of transmission of the minor allele to determine association. This is beneficial as it is not confounded by population stratification, however, it is only informative at sites for which a parent is heterozygous. TDTs were performed using the R package *trio* (174) for trios in which all individuals had a genotype missing rate of <5% and did not exceed a Mendelian error rate of 2%. Variants were removed based on the following filters: minor allele frequency (MAF) <3%, Mendelian error rate >0.1%, Hardy-Weinberg disequilibrium as determined by exact test p-value of 1×10^{-6}, and missingness rate of >5%. Relative risk is

reported in respect to the alternate allele. We considered $p < 1 \times 10^{-5}$ to be of suggestive significance and $p < 5 \times 10^{-8}$ genome-wide significance. We then performed a TDT-based gene-by-sex interaction analysis in *trio*, using the `colGxE()` function, which uses a two-step procedure as described by Gauderman et. al (175). We first evaluated all variants using a likelihood ratio test with two degrees of freedom (LRT 2df), which simultaneously tests the SNP main effect and environmental interaction. For this function, a χ^2 -distribution is used as the asymptotic null distribution for p-values determination.

SNPs reaching suggestive levels of association ($p < 1 \times 10^{-5}$) were then evaluated specifically for gene-by-sex interactions. The GxS relative risk (RR) is reported as the RR in females divided by the RR in males. Because these tests are considered independent of one another, we used the number of suggestive SNPs in the LRT 2df test to calculate the significance threshold for the GxS step based on a Bonferroni correction ($0.05/71$ SNPs; $p < 7.04 \times 10^{-4}$). We performed a post hoc power calculation in Quanto (version 1.2.4) to evaluate our power to detect interactions in the same direction with different magnitude or antagonistic interactions.

Genetically regulated gene expression imputation and association testing

For each gene, we first trained sex-specific gene expression imputation models with cis-SNPs as predictors using GTEx V8 reference data (176) for all 49 tissues in female samples and 48 tissues in male samples. Covariates including age, body mass index, top five genotype principal components, and top probabilistic estimation of expression residual (PEER) factors (177) were regressed out from the gene expression quantitative traits. For each tissue, we used the TIGAR-V2 tool (178) to train both nonparametric Dirichlet process regression (DPR) model and penalized linear regression model with Elastic-net penalty. Second, the trained sex-specific gene

expression imputation models, i.e., coefficients of cis-SNP predictors (referred to as eQTL weights), were used to impute genetically regulated gene expression using the corresponding sex-specific GWAS data. For each gene, the sex-specific association between the genetically regulated gene expression (imputed by the DPR or Elastic-Net model) and CP were tested by the S-PrediXcan Z-score statistic (179) as implemented in the TIGAR-V2 tool. Third, for each gene, the sex-specific test p-values based on DPR and Elastic-Net model were then combined by using the aggregated Cauchy association test (ACAT) method (180). Previous studies have shown that the ACAT method can leverage multiple tests based on different statistical models to improve power. As a result, we obtained sex-specific ACAT p-values for testing the association between the genetically regulated gene expression of test genes and the CP phenotype. A significant ACAT p-value means the genetic effect of the test gene (aggregated over multiple test cis-SNPs with non-zero eQTL weights) on CP is potentially mediated through the corresponding gene expression.

Rare Variants

A bed file for hg38 coordinates for *LTBP1* was generated using the UCSC Genome Browser. All variants within this region were extracted from a multisample VCF with vcfutils (version 0.1.13) followed by annotation with Annovar (version 201910) (181). The commands “dbnsfp42a” and “revel” were included to obtain SIFT, CADD, and REVEL scores for all variants. Additionally, predictions values from AlphaMissense (182) were accessed from the publicly available data. Predicted damaging thresholds were as follows: SIFT ≤ 0.05 , CADD ≥ 20 , REVEL ≥ 0.75 , and AlphaMissense ≥ 0.564 .

RESULTS

Using a dataset of 435 trios, we performed sex-stratified TDTs split into 262 female and 173 male probands based on genetic sex (Figure 3.1, Figure S3.1, 3.2). For all evaluations, we tested 6,774,204 SNPs. There were 16 SNPs in 9 loci of suggestive significance in the female analysis, and 3 SNPs in 3 loci in males (Table 3.1). The significant loci from each test were non-overlapping between sexes, and in all cases were more significant in the respective analysis than for the full cohort (Figure S3.3). To explore these sex-specific differences more formally, we next performed a GxS analysis to find sex-specific interactions.

We first tested SNPs with the LRT 2df model of association considering both gene and GxS effects (Figure S3.4). There were no SNPs at genome-wide significance in this step, though there were 71 SNPs in 32 regions with $p < 1 \times 10^{-5}$ (suggestive significance threshold) (Table 3.2). Of those, 26 SNPs in 13 regions were considered significant for gene-by-sex interactions (Bonferroni correction for 71 SNPs, $p < 7.04 \times 10^{-4}$) (Supp Figure 5). We noted that for all GxS significant SNPs, the direction of effect was opposite in males versus females (Figure 3.2A, B). A post-hoc power analysis shows that given our sample size, we have good power to detect large differences (including opposite effects) when SNPs have a MAF of at least ~15%. However, we have very low power to detect modest interactions, or even large differences at lower MAFs (Figure S3.6). We did not detect significant GxS interactions in the remaining 45 SNPs in 19 regions, suggesting that these associations are primarily driven by the effect of the genotype rather than sex-specific interactions (Figure 3.2C).

For the 13 loci from the sex-stratified TDTs (Table 3.1), two of the three loci from the male-specific analysis, and six of nine loci from the female-specific analysis, we found significant for GxS effects. Using the results from all four tests, we compared the relative risks for the top ten

most significant SNPs in the GxS to those from the sex-stratified and sex-combined TDTs, shown in Figure 3.

The most significant finding for GxS interaction was at 2p22.3, within intron 28/33 of latent TGF-beta binding protein 1 (*LTBP1*). We expanded our analysis of *LTBP1* to look at rare variants (MAF <0.05%) in the full cohort and identified 21 unique protein-altering variants in 25 probands (Table S3.1). 14 variants in 15 probands had at least one damaging prediction as determined by CADD, REVEL, or AlphaMissense scores. We performed a Fisher's exact test differences in the number of predicted damaging rare variants in male versus female probands and found no statistically significant differences (p=0.22).

Lastly, there is evidence for differential expression of *LTBP1* between sexes in both the pituitary of mice (183) and whole blood of humans (183). Therefore, we evaluated the association between the sex-specifically imputed genetically regulated gene expression and the CP phenotype for 12 genes present at the GxS significant loci. We also evaluated the association between the imputed genetically regulated gene expression and the CP phenotype in all samples. There were 11 genes with nominal significance (p<0.05) in at least one of the tested tissues (Table 3.3). With correction for 12 loci, there were only two loci that remained significant: *GTF2AIL* for females in the substantia nigra of the brain (p=0.0010), and *LTBP1* for females in cultured fibroblast cells (p=0.0013). The latter association was also observed to a lesser extent for the full cohort (p=0.0125).

DISCUSSION

CP is more prevalent in females than males, yet little is known about sex-specific risks. Here, we initially found 12 loci of suggestive evidence of association with CP when stratifying our cohort

by proband sex, followed by identification of 13 loci with suggestive evidence of gene-by-sex interactions in CP probands. In all instances, the identified SNPs increased risk of CP for one sex while appearing protective for the other. Given our sample size, this was expected as we have more power to detect these large differences in effects; much larger sample sizes are needed to detect effects that differ in magnitude but are in the same direction in males and females. We also found that most loci were associated with a higher relative risk in females compared to males, which supports the epidemiological data where females are affected more commonly.

In the LRT 2df analysis, we identified 13 loci that were associated with CP but not significant for GxS interactions. This suggests that these associations are primarily gene-driven, where the genotype influences the risk for CP regardless of patient sex. We reported the effects of these gene-driven SNPs in our previously published GWAS results for the full cohort (173), and our results were consistent across both studies, as expected. However, none of the loci with significant GxS interactions were significant in that study, as the effect in male and females likely negated each other. We also compared these loci to sex-specific risks previously reported for OFCs (167, 168), but none were replicated in our cohort. Altogether, these findings illustrate heterogeneity in genetic architecture both between CP and CL/P as well as between males and females, and highlight the importance of investigating sex-specific differences and/or GxS interactions in CP. While our study contributes to the knowledge of the genetic basis of CP and sex-specific risks, it is not without limitation. Our sample size of 435 trios is underpowered to identify gene-by-sex interactions in the same direction but with differences in the magnitude of risk. Further, due to the use of TDTs for case-parent trios, we are restricted to evaluation of sites for which parents are heterozygous.

Our top finding was an intronic SNP in *LTBP1* at 2p22.3. There are several lines of evidence suggesting that *LTBP1* could play a role in palatogenesis. *LTBP1* is an extracellular matrix protein and a component of the large latent complex (184). This complex binds to a second complex of latency associated proteins and mature TGF-B, working to regulate its biological availability in the extracellular matrix. Disruption of the TGF-B signaling pathway is associated with CP (185, 186), so dysregulation caused by variants in *LTBP1* could hypothetically lead to disrupted palatogenesis. We further identified 21 unique, rare variants (MAF <0.5%) in 12 males and 13 females. Of the 21 total, 15 were predicted to be damaging by at least one *in silico* tool. These were found in 9 males and 6 females, although this difference was not significant with a Fisher's exact test.

From a clinical standpoint, bi-allelic truncating variants in *LTBP1* have been reported as causal for autosomal recessive cutis laxa syndrome type IIE (OMIM: 619451). In addition to connective tissue abnormalities, affected individuals present with craniofacial dysmorphisms including high arched palates in 62.5% of cases (187). Further, *LTBP1* directly competes with GARP (glycoprotein A repetitions predominant, encoded by *LRRC32*), a type I membrane protein that plays a role in maturation, processing, and tethering of TGF-B to the cell surface (188). Both *LTBP1* and GARP compete for binding via disulfide link to the same cysteine (Cys-4) on proTGFB1, and direct TGF-B to either the cell surface or deposited within the ECM (189). Recently, homozygous variants in *LRRC32* have been reported to cause a cleft palate syndrome (190), suggesting imbalance in this process can disrupt palatogenesis.

In support of clinical data, morpholino knockdown of *ltbp1* in zebrafish causes severe jaw malformation in which the mandible has a reduced length and width. Similarly, knockout of *Ltbp-1* in mice resulted in a modified facial profile including a shortened maxilla and mandible, as well

as reduced cellular activation and differentiation of myofibroblasts, supporting reduced availability of *Tgfb* when *Ltbp-1* is disrupted (191). In humans, shortened mandibles are a cause of CP in Pierre Robin sequence (192). Our cohort included 18 trios reported to have Pierre Robin sequence. While this was too few to perform a separate analysis, we did carry out a sensitivity analysis by removing these trios and repeating the GxS test. Our results did not meaningfully change, suggesting that—in this study—Pierre Robin cases are not driving the signal in this gene.

Although the mechanism by which variants in or around *LTBPI* mediate their effect on a sex-specific basis is unknown, there is evidence for sex-based differences in expression of this gene. A study in mice found that *Ltbp1* expression in the pituitary was lower in females than males, and that both sexes showed reduced expression in response to estradiol, suggesting a role for sex hormones in this differentiation (183). A similar finding has been shown in human peripheral blood, with *LTBPI* expression belonging to the top 1% of sex-biased genes (based on log fold change) where males had higher levels of *LTBPI* mRNA than females (193). To explore this based on our GWAS data, we performed an association analysis for CP and the imputed genetically regulated gene expression and found a female-specific significant association for *LTBPI* in cultured fibroblast cells. Despite a lack of access to a more relevant tissue (i.e., palatal tissue), the association of *LTBPI* expression with CP in female probands is of interest as fibroblasts have many of the same traits as mesenchymal stem cells, including an ability to differentiate into chondrocyte and osteoblasts (194).

In summary, we have demonstrated that there are gene-by-sex interactions for CP which cannot be detected in combined-sex cohorts as opposite effects can negate each other. We also identified *LTBPI* as a gene of interest for CP risk, particularly for females, and show that it may mediate its effect on the phenotype via gene expression. In conclusion, sex-specific effects

contribute to the genetic heterogeneity of CP and should be further explored as sample sizes continue to grow.

FIGURES

Figure 3.1: Miami plot highlighting suggestive SNPs in females (top) and males (bottom). The dotted line represents suggestive significance at $p=1 \times 10^{-5}$, and the solid line represents $p=5 \times 10^{-8}$.

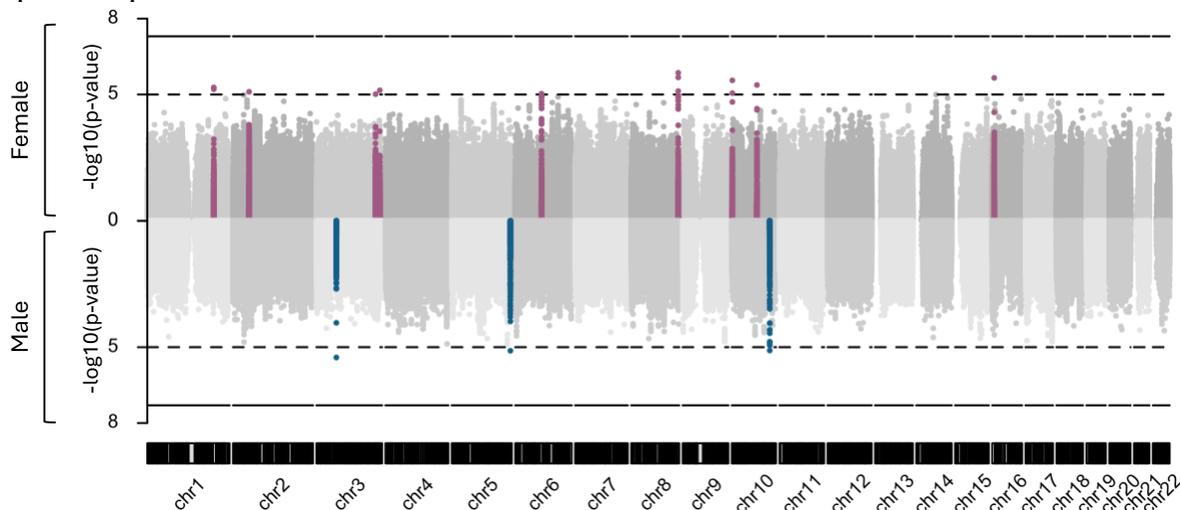


Figure 3.2: Relative risks for males and females demonstrating GxS effect SNPs at (A) 2p22.3 and (B) 15q21.1, and (C) gene-driven SNP at 112q21.1 with 95% confidence intervals. The dotted line at 1 represents null risk. Light blue and dark blue represent heterozygous and homozygous males, respectively while light and dark pink represent heterozygous and homozygous females, respectively.

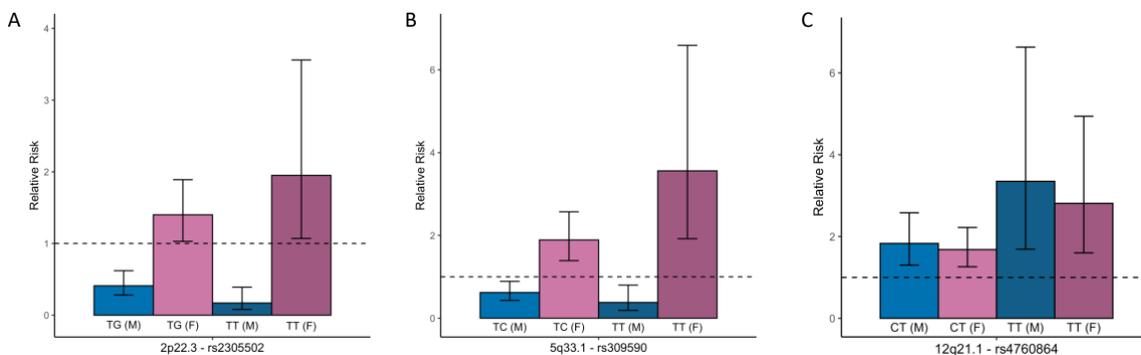


Figure 3.3: Comparing relative risks with 95% confidence intervals across the top 10 SNPs for GxS effects (top) and two gene-driven SNPs (bottom) to relative risk from TDTs in the full cohort and stratified by sex. Risks are shown for GxS (black), female only TDT (pink), male only TDT (blue), and full cohort TDT (green). An asterisk next to the locus name denotes relative risk by heterozygosity shown in Figure 2.

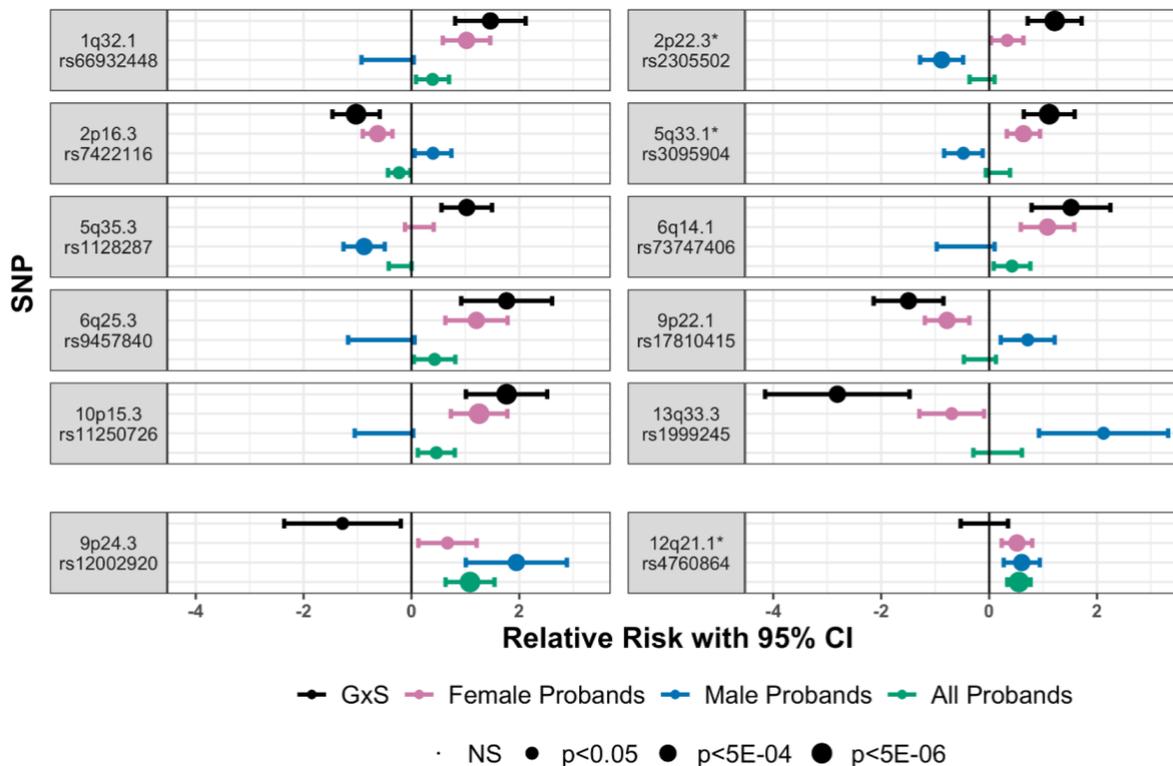


Table 3.1: Loci of suggestive significance from sex stratified TDTs									
Locus (nearest gene)	CHR:BP (rsID)	Ref/Alt	Female RR	Pvalue	Male RR	Pvalue	All RR	Pvalue	Alt AF [^]
Male									
3p14.2* (<i>FHIT</i>)	3:60050040 (rs212048)	A/G	1.04 [0.77 - 1.4]	0.82	0.37 [0.24 - 0.56]	3.92E- 06	0.71 [0.56 - 0.9]	5.11E- 03	0.33
5q35.3* (<i>RAB24</i>)	5:177301457 (rs1128287)	G/T	1.16 [0.89 - 1.51]	0.28	0.42 [0.28 - 0.61]	7.22E- 06	0.81 [0.66 - 1]	0.05	0.8
10q25.3 (<i>ABLIM1, FAM160B1</i>)	10:114803389 (rs7075095)	T/C	1.12 [0.8 - 1.57]	0.5	2.66 [1.73 - 4.07]	7.40E- 06	1.6 [1.23 - 2.07]	3.81E- 04	0.43
Female									
1q32.1* (<i>CACNAIS</i>)	1:201074717 (rs66932448)	ACT/A	2.78 [1.79 - 4.31]	5.31E- 06	0.64 [0.4 - 1.04]	0.07	1.48 [1.09 - 2.01]	0.01	0.72
2p16.3* (<i>GTF2AIL</i>)	2:48656339 (rs7422116)	T/A	0.53 [0.41 - 0.7]	7.83E- 06	1.49 [1.06 - 2.1]	0.02	0.8 [0.65 - 0.98]	0.03	0.43
3q26.32 (<i>KCNMB2</i>)	3:178712151 (rs9290663)	A/T	0.52 [0.39 - 0.69]	9.72E- 06	0.98 [0.68 - 1.41]	0.93	0.66 [0.53 - 0.83]	3.02E- 04	0.21
3q28 (<i>LINCRCR-0002</i>)	3:191499526 (rs9790048)	T/A	0.48 [0.35 - 0.66]	6.82E- 06	0.85 [0.57 - 1.26]	0.42	0.6 [0.47 - 0.76]	4.26E- 05	0.89
6q14.1* (<i>SH3BGRL2</i>)	6:79611392 (rs73476169)	T/C	3.05 [1.86 - 4.99]	9.37E- 06	0.81 [0.48 - 1.36]	0.43	1.7 [1.21 - 2.38]	2.23E- 03	0.21
8q24.3* (<i>MIR4472-1</i>)	8:142171910 (rs55671251)	G/A	2.29 [1.63 - 3.2]	1.39E- 06	0.82 [0.56 - 1.19]	0.29	1.48 [1.16 - 1.88]	1.66E- 03	0.19
10p15.3* (<i>ADARB2</i>)	10:1683542 (rs11250726)	A/G	3.5 [2.07 - 5.91]	2.77E- 06	0.6 [0.35 - 1.03]	0.06	1.58 [1.12 - 2.24]	8.66E- 03	0.17
10q22.3* (<i>LRMDA</i>)	10:76007479 (rs11001574)	T/C	1.93 [1.46 - 2.56]	4.22E- 06	1.25 [0.89 - 1.75]	0.2	1.62 [1.31 - 2.01]	9.81E- 06	0.19
16p13.3 (<i>RBFOX1</i>)	16:6057173 (rs17139255)	T/G	0.46 [0.34 - 0.64]	2.21E- 06	0.94 [0.63 - 1.4]	0.76	0.61 [0.47 - 0.77]	6.00E- 05	0.18
*Locus also identified in GxS analysis ^gnomAD v 3.1.2									

Locus (nearest gene)	CHR:BP* (rsID)	Ref/Alt Allele	GxS RR	GxS p-value	LRT 2df p- value	Alt AF [^]
1q32.1 (<i>CACNA1S</i>)	1:201074717 (rs66932448)	ACT/A	4.32 [2.25 - 8.31]	<u>1.14E-05</u>	1.52E-06	0.72
1q41 (<i>LYPALI</i>)	1:219268643 (rs59611530)	G/GAAT	1.01 [0.65 - 1.56]	0.96	2.39E-06	0.47
2p22.3 (<i>BIRC6</i>)	2:32447229 (rs555243726)	G/A	0.21 [0.08 - 0.54]	1.16E-03	2.02E-06	0.11
2p22.3 (<i>LTBP1</i>)	2:33361709 (rs2305502)	T/G	3.37 [2.04 - 5.56]	<u>1.93E-06</u>	3.21E-06	0.27
2p21 (<i>MTA3</i>)	2:42509602 (rs57081889)	C/T	1.22 [0.79 - 1.88]	0.37	9.30E-06	0.39
2p16.3 (<i>GTF2AIL</i>)	2:48656339 (rs7422116)	T/A	0.36 [0.23 - 0.56]	<u>4.49E-06</u>	1.92E-06	0.41
3p14.2 (<i>FHIT</i>)	3:60050040 (rs212048)	A/G	2.82 [1.68 - 4.75]	<u>9.41E-05</u>	5.82E-06	0.34
4p14	4:39803327 (rs10000967)	T/C	0.48 [0.21 - 1.09]	0.08	6.24E-07	0.10
5q14.1 (<i>HOMER1</i>)	5:79403345 (rs79156100)	T/C	0.38 [0.14 - 1.07]	0.07	1.43E-06	0.01
5q33.1 (<i>NDST1</i>)	5:150536578 (rs309590)	T/C	3.04 [1.9 - 4.87]	<u>3.65E-06</u>	5.36E-06	0.25
5q35.3 (<i>RAB24</i>)	5:177301457 (rs1128287)	G/T	2.79 [1.75 - 4.44]	<u>1.68E-05</u>	8.76E-06	0.78
6q14.1 (<i>SH3BGRL2</i>)	6:79609515 (rs73747406)	T/A	4.56 [2.2 - 9.47]	<u>4.56E-05</u>	6.92E-06	0.20
6q25.3 (<i>SLC22A1</i>)	6:160120316 (rs9457840)	T/C	5.83 [2.51 - 13.55]	<u>4.10E-05</u>	9.18E-06	0.05
8p23.1 (<i>RP1L1</i>)	8:10669884 (rs55782357)	A/T	0.01 [0 - 0.13]	1.48E-03	5.81E-08	0.09
8q24.3 (<i>MIR4472-1</i>)	8:142171910 (rs55671251)	G/A	2.8 [1.69 - 4.64]	<u>6.50E-05</u>	1.82E-06	0.19

9p24.3 (<i>LOC105375951</i>)	9:1705036 (rs12002920)	G/T	0.28 [0.09 - 0.82]	0.02	1.42E-07	0.09
9p22.1 (<i>SAXO1</i>)	9:19032909 (rs17810415)	A/G	0.22 [0.12 - 0.43]	<u>5.93E-06</u>	9.00E-06	0.19
9q21.12 (<i>TRPM3</i>)	9:70937619 (rs67149721)	A/G	0.18 [0.05 - 0.72]	0.02	1.20E-06	0.10
10p15.3 (<i>ADARB2</i>)	10:1683542 (rs11250726)	A/G	5.83 [2.75 - 12.39]	<u>4.43E-06</u>	3.04E-07	0.17
10q21.2 (<i>RTKN2</i>)	10:62241751 (rs112160719)	T/A	0.06 [0.01 - 0.56]	0.01	5.75E-06	0.08
10q22.3 (<i>LRMDA</i>)	10:76007479 (rs11001574)	T/C	1.55 [1 - 2.4]	0.05	6.23E-06	0.19
11q14.3 (<i>FAT3</i>)	11:92264783 (rs76502406)	C/T	0.74 [0.3 - 1.8]	0.5	2.46E-06	0.01
11q24.2	11:124488761 (rs375612889)	C/T	1.25 [0.44 - 3.53]	0.68	8.43E-06	0.07
12q21.1 (<i>LOC105369838</i>)	12:72976347 (rs4760864)	C/T	0.92 [0.59 - 1.42]	0.69	1.88E-06	0.25
13q33.3	13:106844528 (rs78276410)	C/T	0.06 [0.02 - 0.23]	<u>3.85E-05</u>	3.40E-06	0.03
14q22.1 (<i>FRMD6</i>)	14:51469594 (rs10431684)	C/A	1.53 [0.97 - 2.41]	0.07	3.68E-06	0.57
15q26.2	15:96732159 (rs36062094)	T/C	1.37 [0.88 - 2.14]	0.17	7.63E-06	0.23
16p13.3 (<i>RBFox1</i>)	16:6057173 (rs17139255)	T/G	0.49 [0.3 - 0.82]	6.02E-03	5.58E-06	0.19
16q24.2 (<i>C16orf95</i>)	16:87082043 (rs145642800)	CAGA/C	0.28 [0.12 - 0.63]	2.09E-03	8.87E-06	0.06
17p13.2 (<i>ANKFY1</i>)	17:4224286 (rs58695167)	G/A	1.57 [0.98 - 2.5]	0.06	4.65E-06	0.50
17q24.2 (<i>ARSG</i>)	17:68398218 (rs7222535)	A/G	1.85 [1 - 3.43]	0.05	3.43E-06	0.11

17q25.3 (<i>RPTOR</i>)	17:80629981 (rs62068342)	A/G	0.11 [0.03 - 0.43]	1.49E-03	8.77E-06	0.06
<u>Locus</u> is significant with Bonferroni correction for 71 tested SNPs ($p < 7.04 \times 10^{-4}$) ^gnomAD v 3.1.2						

Table 3.3: Tissues and genes with significant association between imputed gene expression and phenotype				
Gene (Locus)	Tissue	All	Male	Female
<i>CACNA1S</i> (1q32.1)	Minor Salivary Gland	NS	NS	0.0385
<i>LTBP1</i> (2p22.3)	Artery - Tibial	NS	NS	0.0193
	Cells Cultured - Fibroblasts	0.0125	NS	0.0013
<i>GTF2AIL</i> (2p16.3)	Artery - Aorta	NS	NS	0.0385
	Brain - Cerebellum	NS	NS	0.0273
	Brain - Substantia nigra	NS	NS	0.0010
	Colon - Sigmoid	NS	NS	0.0051
	Pancreas	NS	NS	0.0311
<i>STON1-GTF2AIL</i> (2p16.3)	Brain - Spinal cord cervical C1	NS	NS	0.0097
	Colon - Sigmoid	NS	NS	0.0237
<i>FHIT</i> (3p14.2)	Esophagus - Mucosa	NS	NS	0.0248
<i>NDST1</i> (5q33.1)	Brain - Substantia nigra	NS	NS	0.0181
	Cells Cultured - Fibroblasts	NS	NS	0.0306
	Minor Salivary Gland	NS	NS	0.0436
<i>RAB24</i> (5q35.3)	Brain - Hypothalamus	NS	NS	0.0343
	Cells - EBV-transformed lymphocytes	NS	0.0052	NS
	Esophagus - Muscularis	NS	0.0405	NS
	Minor Salivary Gland	NS	0.0265	NS
	Whole Blood	NS	0.0151	NS
<i>SH3BGRL2</i> (6q14.1)	Whole Blood	NS	NS	0.0055
<i>SLC22A1</i> (6q25.3)	Brain - Cerebellar Hemisphere	NS	NS	0.0112
	Brain - Hypothalamus	NS	NS	0.0380
	Cells - EBV-transformed lymphocytes	NS	0.0106	NS
<i>SAXO1</i> (9p22.1)	Prostate	NS	0.0195	NA
<i>ADARB2</i> (10p15.3)	Brain - Anterior cingulate cortex BA24	NS	NS	0.0325
	Brain - Amygdala	0.0353	NS	NS
	Brain - Caudate basal ganglia	0.0142	NS	NS
	Brain - Cerebellar Hemisphere	0.0354	NS	NS

NS=not significant, $p>0.05$; NA=not available

SUPPLEMENTAL DATA

Figure S3.1: Manhattan plot for male probands (left) with qqplot of p-values (right). The red line represented genome-wide significance at $p=5E-08$, and the blue line represents suggestive significance at $p=1E-05$.

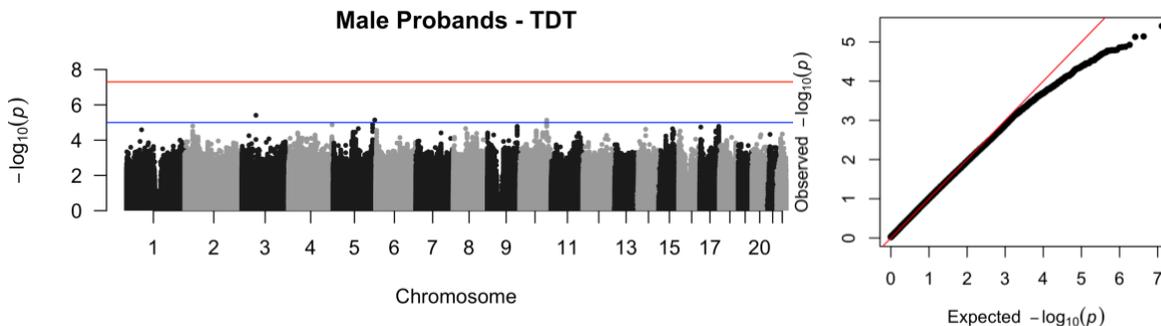


Figure S3.2: Manhattan plot for female probands (left) with qqplot of p-values (right). The red line represented genome-wide significance at $p=5E-08$, and the blue line represents suggestive significance at $p=1E-05$.

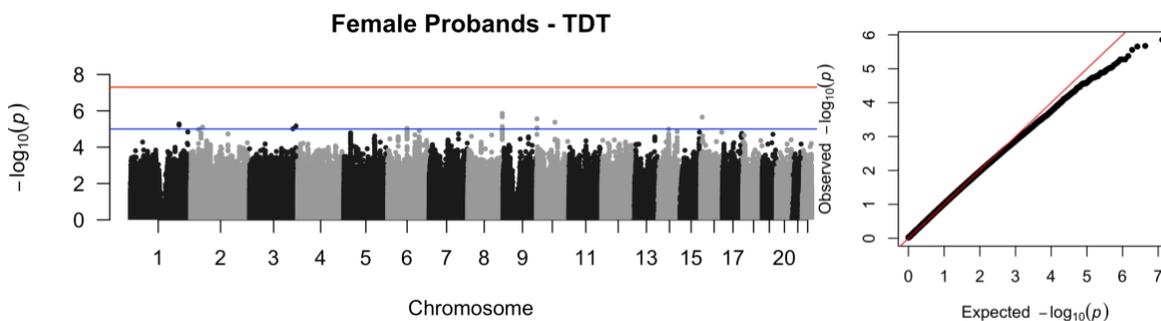


Figure S3.3: Manhattan plot for all probands (left) with qqplot of p-values (right). The red line represented genome-wide significance at $p=5E-08$, and the blue line represents suggestive significance at $p=1E-05$.

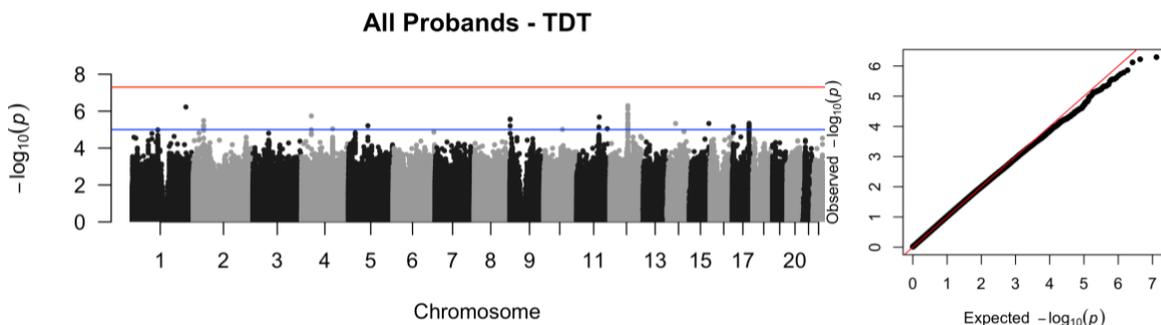


Figure S3.4: Manhattan plot for the LRT2df (left) with qqplot of p-values (right). The red line represented genome-wide significance at $p=5E-08$, and the blue line represents suggestive significance at $p=1E-05$.

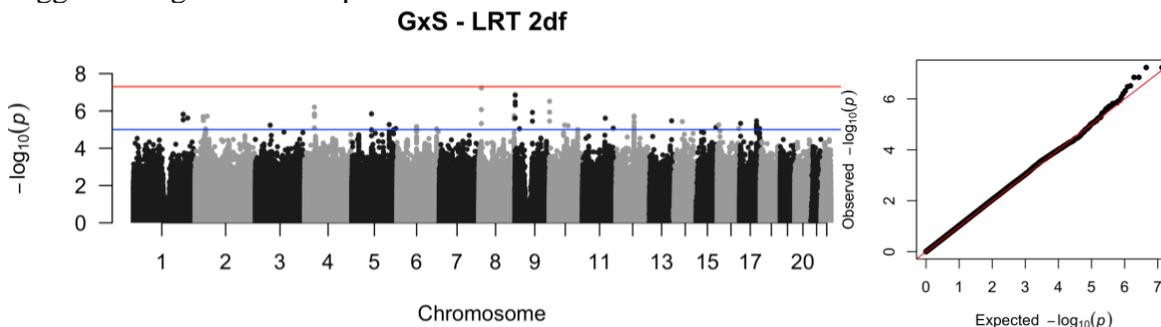


Figure S3.5: Manhattan plot for GxS interactions (left) with qqplot of p-values (right). The red line represented genome-wide significance at $p=5E-08$, and the blue line represents suggestive significance at $p=1E-05$.

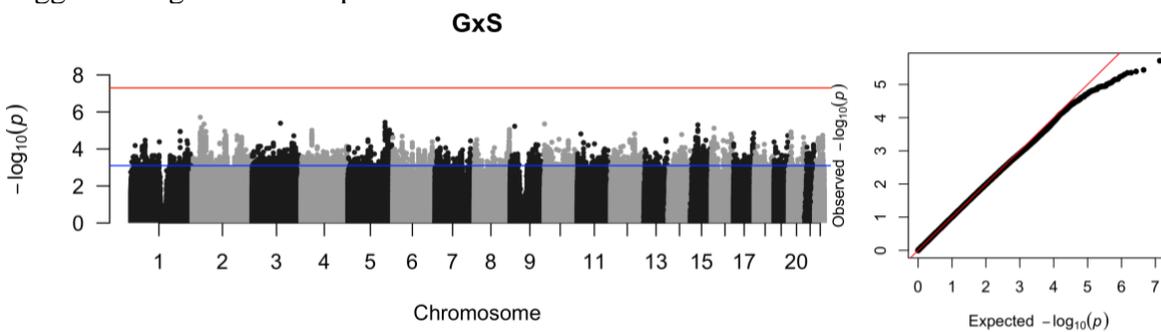


Figure S3.6: Post hoc power calculations for the most significant SNP for GxS effect (left), the SNP just below the significance threshold for GxS effect (middle), and a SNP with no significant effect for GxS (right). In each plot, the dotted vertical line represents the output R_{GE} determined by Trio, and the colored dot represents each SNPs respective MAF at the intersection of its R_{GE} along the line of power of detection. All SNPs were tested using a population risk of 0.00058 and ‘environmental exposure’ of 0.5, with additional SNP-specific corresponding input parameters used for each SNP are listed below the image.

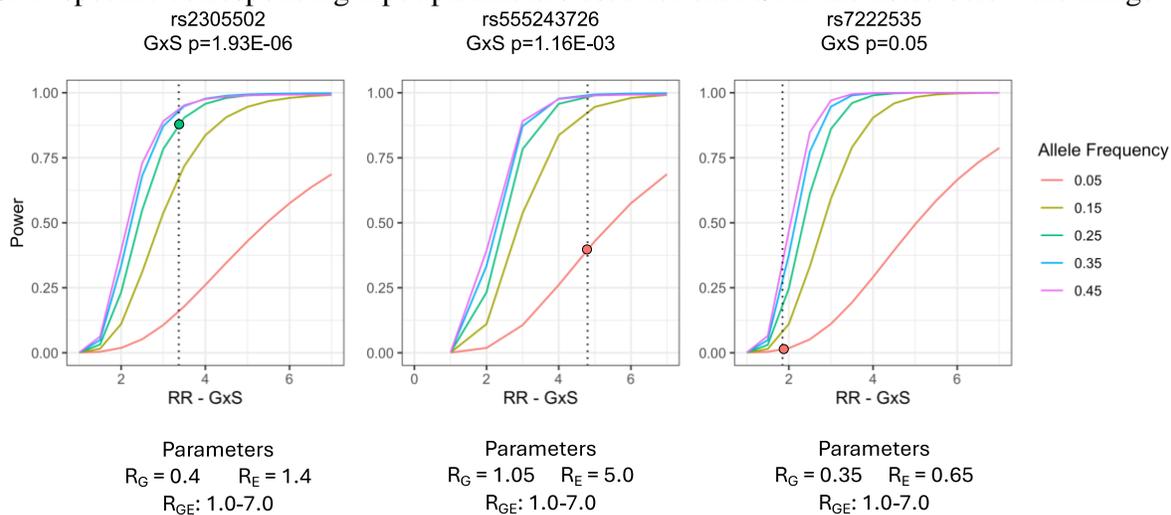


Table S3.1: Rare variants in *LTBP1*

Position	Sex	Population	cDNA	AA change (NM_206943)	CADD	REVEL	Alpha Missense	AF popmax gnomad211
chr2:33020931-GA/G	M	Asian	c.589delA	p.N197Mfs*25	.	.	.	
chr2:33188755- TTCCACATAC/T rs779600504	M	Asian	c.1606_1614del	p.T537_S539del	.	.	.	0
chr2:33361513-A/C rs1293018624	F	Asian	c.A4268C	p.K1423T	22.8	0.555	0.1874	0.001
chr2:33364354- A/G rs139267139	M	Black	c.A4538G	p.N1513S	14.57	0.251	0.0642	0.0000545
chr2:33389287-T/A rs149372970	F	Black	c.T4815A	p.D1605E	12.19	0.215	0.1936	0.0012
chr2:33397262-C/T rs752902860	F	Black	c.C4964T	p.T1655M	16.15	0.219	0.0499	0.0008
chr2:33360602-G/A rs1363963010	F	Unknown	c.G4006A	p.D1336N	16.04	0.066	0.0749	0.0005
chr2:33110640-G/T rs117874429	M	Asian	c.G922T	p.V308L	15.57	0.162	0.1198	0.0000241
chr2:33110640-G/T rs117874429	F	Asian	c.G922T	p.V308L	15.57	0.162	0.1198	0.0046
chr2:33110640-G/T rs117874429	F	Asian	c.G922T	p.V308L	15.57	0.162	0.1198	0.0046
chr2:33110640-G/T rs117874429	F	Asian	c.G922T	p.V308L	15.57	0.162	0.1198	0.0046
chr2:33342945-G/A rs140281612	F	Black	c.G3838A	p.D1280N	26.7	0.612	0.1267	0.0046
chr2:33361488-T/G rs116888774	M	Asian	c.T4243G	p.S1415A	14.42	0.118	0.078	0.0018
chr2:33188802- C/T rs141814525	M	Black	c.C1652T	p.A551V	23.6	0.345	0.1304	0.0013
chr2:33252832-C/G rs1359295062	M	White	c.C2155G	p.L719V	22.2	0.273	0.0873	0.0000483

chr2:33389223-G/A rs377345873	M	White	c.G4751A	p.R1584H	22.8	0.155	0.075	0.0000294
chr2:33342910-G/A rs371235125	F	Asian	c.G3803A	p.R1268H	25.5	0.341	0.1305	0.0006
chr2:33021023-C/T rs377541507	F	Black	c.C680T	p.S227L	4.968	0.139	0.0842	0.0022
chr2:33263300-T/C rs2093072594	F	Asian	c.T2525C	p.I842T	23.1	0.425	0.1038	0.0008
chr2:33342876-C/G rs778553662	F	Black	c.C3769G	p.H1257D	22.6	0.756	0.4901	0
chr2:33360693-A/G rs139331984	F	White	c.A4097G	p.N1366S	24.2	0.689	0.0939	0.0007
chr2:33398499-C/T rs141171373	M	Black	c.C5120T	p.P1707L	23.3	0.531	0.0981	0.0002
chr2:33398499-C/T rs141171373	M	Unknown	c.C5120T	p.P1707L	23.3	0.531	0.0981	0.0047
chr2:33188640-T/C rs537655642	M	Unknown	c.T1490C	p.I497T	24.2	0.771	0.995	0.0047
chr2:33275861-G/C rs79998390	M	Black	c.G2930C	p.G977A	25.9	0.94	0.5278	0.0000176

CHAPTER IV: Cleft palate probands are significantly enriched for protein-altering *de novo* variants

This chapter is adapted from a manuscript in preparation for submission for publication.

Kelsey Robinson, Sarah Curtis, Terri H. Beaty, Azeez Butali, Carmen J. Buxó, David J. Cutler, Michael P. Epstein, Jacqueline T. Hecht, Gary M. Shaw, Lina M. Moreno, Jeffrey C. Murray, Harrison Brand, Seth M. Weinberg, Mary L. Marazita, Elizabeth J. Leslie

ABSTRACT

De novo variants (DNVs) are sporadically occurring mutations that most commonly arise in the germline and are present in offspring but absent in both parents. As they are not under selective pressure, they may be enriched for disease-causing alleles and have been implicated in multiple rare genetic disorders. Cleft palate (CP) is a common craniofacial birth defect for which few genetic associations are known. Because approximately 50% of CP occurs as a syndrome, rare variants may be a significant contributor to CP in general. Exome and targeted sequencing studies have been performed in family-based and case-control cohorts, but these studies have lacked statistical power to conclusively identify causal genes; therefore, we analyzed a set of 475 case-parent trios to study DNVs in CP. We identified global enrichment of protein-altering DNVs (1.34, $p=2.18 \times 10^{-9}$), and significant ($p < 8.7 \times 10^{-5}$) gene-specific enrichment for *COL2A1*, *IRF6*, *PRKCI*, *SATB2*, *POLR1F*, and *SLC25A41*. We also evaluated subtype-specific risks for clefts affecting the hard and soft palate (CH&SP) and a combined group of cleft soft palate and submucous cleft palate (CSP+SMCP). We found gene-specific enrichment for *SATB2* and *TGFBR2* in CH&SP and *PRKCI* in CSP+SMCP. We also looked at whether any specific cell types harbored excess DNVs in marker genes from a single nucleus RNA sequencing dataset of the secondary palate in mice. We found both subtype groups were enriched for chondrocyte progenitors, but CSP+SMCP was specifically enriched for the epithelial and endothelial clusters. Lastly, we compared our data to pathogenic/likely pathogenic DNVs reported for syndromic CP probands from the Deciphering Developmental Disorders (DDD) study. 11 genes overlapped between the DDD cohort and our CPSeq study, including *COL2A1* and *SATB2*. Notably, some genes (*IRF6*, *GRHL3*) were absent from the DDD cohort. This suggests a comprehensive cohort inclusive of the full spectrum of isolated and syndromic CP is needed to further elucidate phenotypes in genes associated with OFC

syndromes, and to explore subtype-specific genetic risks to better understand palatogenesis and the genetic architecture of CP.

INTRODUCTION

Cleft palate (CP) is one of three major subtypes of orofacial clefts (OFCs), together comprising the most common craniofacial birth defect. CP occurs in approximately 1 in 1700 live births worldwide (4). In the United States, approximately 2,600 babies are born with visible CP each year and up to 80,000 babies are born with the more subtle submucous cleft palate (SMCP) (195). The palate makes up the roof of the mouth, separating the oral and nasal cavities. It is composed of an anterior bony hard palate, which is required for normal feeding, and a posterior muscular soft palate that elevates to close off the pharynx during swallowing and speech. Infants with CP suffer from feeding difficulties; without intervention, CP can be fatal early in life. However, prognosis is much more favorable with surgical correction, though individuals born with CP often go on to face speech or hearing problems, require advanced orthodontic care, and can experience additional comorbidities as they age (2, 3). As such, CP creates both individual and public health burdens. Improving our understanding of its origin can lead to enhanced prevention, treatment, and prognosis for affected individuals.

Compared to other types of OFC, the causes of CP are not well understood. Multiple lines of evidence suggest that the genetic architecture and etiology of CP is different from other types of OFCs. CP is highly heritable (84) and approximately 25% of affected individuals have a positive family history (196, 197). However, the identity of the genetic variants that cause or increase risk for CP are not known. There have been six well-powered genome-wide association studies (GWAS) that have collectively discovered fewer than ten loci associated with CP (102, 108, 110-

112). Two of these studies identified variants that are population-specific: a low frequency missense variant in *GRHL3* that is most frequent in Europeans but rarer in all other populations (108), and a non-coding allele near *IRF6* limited to the Finish and Estonian populations(112). One study in a Chinese population identified 9 associated loci (111) but many loci were not replicated in a subsequent GWAS in an independent Chinese population (102). The known CP loci show little overlap with the dozens of loci identified in GWAS of cleft lip with or without cleft palate (CL/P) and few associations specific to CP are known. This suggests that CP may have a unique genetic etiology and genetic architecture from CL/P and that many of the genetic causes of CP are unknown.

Approximately 50% of CP occurs as a syndrome so rare variants may be a significant contributor to CP in general. Exome and targeted sequencing studies have been carried out in family-based and case-control cohorts, but these studies have lacked statistical power to conclusively identify causal genes (119, 198). Using a set of genes assembled from clinical genetic testing panels, we found that individuals with CP were more likely to have rare pathogenic or likely pathogenic variants than other types of OFCs (123), but this study was limited to 58 CP trios. Similarly, we found a nominally significant enrichment of coding *de novo* variants (DNs) in these trios (enrichment 1.39, $p=9.32 \times 10^{-3}$), but lacked power to identify new genes associated with CP (125).

DNs have been implicated in both structural congenital anomalies such as congenital heart disease(199) or complex diaphragmatic hernia(200), as well as developmental disorders such as autism spectrum disorder(201-203). Exome-wide enrichment has also been shown in a cohort of 698 case-parent trios with cleft lip with or without cleft palate (CL/P) (125), in which three genes (*IRF6*, *TFAP2A*, and *ZFHX4*) also reached exome-wide significance on an individual basis.

Because of prior studies indicating that rare or *de novo* variants are good candidates to explain some of the unknown genetic causes of CP, we set out to analyze DN variants in a set of 475 case-parent trios from the CPSeq whole genome sequencing study. We further explored the heterogeneous nature of CP by stratifying the cohort by proband sex (as CP more commonly affects females) (4, 5, 8), CP subtypes based on the affected parts of the palate, and by the presence or absence of non-cleft phenotypic features.

METHODS

Study cohort

We assembled a collection of 475 case-parent trios ascertained on proband affection status (e.g., cleft palate) from the CPSeq (N=427) and Gabriella Miller Kids First (GMKF) (N=48) whole genome sequencing projects(125) (hereafter referred to as CPSeq). Trios represent all major ancestry groups affected by CP including those with European ancestry (recruited from Spain, Turkey, Hungary, United States), Latin America (Puerto Rico, Argentina), Asia (China, Singapore, Taiwan, the Philippines), and Africa (Nigeria, Ghana). Recruitment and phenotypic assessment occurred at multiple domestic and international sites following institutional review board (IRB) approval for each local recruitment site and coordinating centers (University of Iowa, University of Pittsburgh, and Emory University).

There were 279 female probands and 196 male probands. All probands and parents were assessed for the presence of a CP with ~2/3 of the assembled samples undergoing additional phenotyping to assess the location and severity of the CP. There were 119 probands with cleft hard and soft palate (CH&SP), 149 cleft soft palate (CSP), 3 cleft hard palate (CHP), 24 submucous cleft palate

(SMCP), and 180 with unspecific CP subtypes. Although probands/trios were not excluded based on additional clinical features consistent with a syndromic diagnosis, only 33 trios were classified as possibly or probably syndromic based on a reported presence of additional major or minor clinical features. Additional breakdown based on self-reported race categories is available in Supplemental Table 1.

Whole genome sequencing

The full description of sequencing and variant calling methodology for the CPSeq trios is detailed in Robinson et. al.(173) and for the GMKF trios in Bishop et.al(125). The WGS for CPSeq was performed by the Center for Inherited Disease Research (CIDR) at Johns Hopkins University (Baltimore, MD). The DRAGEN Germline v3.7.5 pipeline on the Illumina BaseSpace Sequence Hub platform was used for alignment, variant calling, and quality control, resulting in a single multisample VCF file. For GMKF, WGS for European samples was carried out by the McDonnell Genome Institute (MGI) the Washington University School of Medicine (St. Louis, MO) followed by alignment to hg38 and variant calling at the GMKF Data Resource Center at the Children's Hospital of Philadelphia. WGS for Colombian and Taiwanese samples was carried out by the Broad Institute, with alignment to hg38 and variant calling by GATK pipelines(204-206).

Identification of *de novo* variants

CPSeq: The DRAGEN 3.7.5 aligner and variant caller was used to generate gVCF files for each trio. Individual trio VCFs with *de novo* variant tags were then generated by using gVCF files combined with pedigree information as input to the DRAGEN 3.7.5 joint caller. Genotypes were set to missing if GQ<20 or read depth <10. To be considered a DN, variants had to have a quality

score of 30, $DQ > 2$, and parental genotypes had to be confirmed homozygous reference (0/0), pass all filtering steps, and have an allele balance (AB) ratio of < 0.05 .

GMKF: Mendelian errors were called in trios using PLINK (version 1.90b53) and bcftools (v1.9). These mendelian errors were then underwent further filtering to yield high quality *de novo* mutation calls including filtering for passing variants, bi-allelic variants only, a minor allele count (MAC) = 1, genotype quality (GQ) ≥ 20 , and depth (DP) ≥ 10 using VCFtools (version 0.1.13). Furthermore, we filtered variants on the basis of allele balance (AB), with *de novo* calls requiring an AB ratio ≥ 0.30 and ≤ 0.70 in the proband and an AB ratio < 0.05 in the parents.

Variant Annotation

Variants were annotated with ANNOVAR (version 201707). Variant with coding consequences were selected based on classification as “exonic” or “splicing”, and only those variants with MAF of $< 0.5\%$ in either gnomAD exomes v 2.1.1 or gnomAD v3.1.2 were retained for analysis.

DN enrichment

We evaluated 475 CP case-parent trios for coding DN enrichment using the R package ‘DenovolyzeR’ package (version 0.2.0). The cohort was tested for an excess of DNs exome-wide and per gene using the functions ‘DenovolyzeByClass’ and ‘DenovolyzeByGene’, respectively. These functions utilize mutation models described by Samocha, et al.(207) to determine if there are more observed DNs in a dataset than would be expected by chance. Using mutational rates, the number of DNs is expected to follow a Poisson distribution under the null model of no association between a variants class and a phenotype(208). Given a fixed sample size and expected mutation

rate for a given genetic sequence, we can determine M (both the mean and variance) as well as the standard deviation, resulting in the “known” constant. Under the alternate model, the number of observed mutations, A, also follows a Poisson distribution, but A may not equal M. We used the Poisson distribution to determine enrichment shown by A/M with 95% confidence intervals.

For the DenovolyzeByGene analysis, we used a multiple test correction for the number of genes with DNAs and considered genes significant at $p < 8.74 \times 10^{-5}$ (0.05/572 genes). We also considered a more conservative threshold for exome-wide significance at $p < 1.3 \times 10^{-6}$.

In analyses where we are interested in the enrichment in specific genes sets, the function ‘includeGenes’ was applied. For analyses using groups of gene sets (for example, analysis of marker genes from single cell sequencing clusters), we corrected for the number of clusters because of the overlap of genes in each set.

We tested for significant differences between DN enrichment in males and females using a Z test for the observed versus expected number of variants in males versus females while considering the prevalence differences between sexes. We assumed the variance of observed variants was equal to the expected variance, based on the Poisson distribution, and determined Z using $(\text{observed} - \text{expected}) / \sqrt{\text{expected}}$ for each class of variant.

Enrichment analyses and creation of gene sets

We evaluated our dataset for enrichment in several different ways. First, we performed a gene set enrichment analysis (GSEA) on all protein-altering DNAs using ToppFun as part of the ToppGene Suite(209). We also created two sets of genes directly relevant to CP: an OFC-specific gene panel(123) and a set of marker genes generated from single nucleotide RNA sequencing (snRNAseq) of the secondary palate in mice at embryonic day 15.5 (E15.5) (130). The OFC gene

panel was curated from four sources, including the National Health Service (NHS) Genomic Medicine Service cleft panel (v2.2), the Prevention Genetics CL/P clinical genetic testing panel, genes from the Online Mendelian Inheritance in Man (OMIM) that include orofacial clefts with a known inheritance and molecular basis, and a manually curated list from recent research studies on OFC genetics—additional details on curation are published in Diaz Perez, et. al (123). Full details on snRNAseq and marker gene generation can be found in Piña, et. al (130). We filtered marker genes for FDR <0.01 prior to enrichment testing for DNs.

RESULTS

Exome-wide analyses

We generated a set of high confidence DNs from 475 CP case-parent trios originating from diverse backgrounds (Table S4.1). These included 33 syndromic or suspected syndromic cases (6.9%) with additional features such as heart murmurs and developmental delays, but none had an official molecular diagnosis for a specific syndrome (Table S4.2). Ultimately, we identified 597 protein-coding variants in 572 genes (Table S4.3), with each trio averaging 1.25 DNs (Figure S4.1A). Variant frequency followed a Poisson distribution with no significant deviation tested by chi-square goodness-of-fit ($p=0.99$).

We classified DNs based on the variant type and predicted function: synonymous variants, missense variants (including single amino acid substitutions and in-frame insertions or deletions), putative loss-of-function variants (LOF, including nonsense, frameshift insertions or deletions, and splice acceptor or donor sites), and a category referred to as protein-altering variants, which includes the combination of all missense and LOF variants. Broken down into these categories, we

had 138 synonymous, 391 missense, 68 LOF, and a combined 459 protein-altering DNAs (Figure S4.1B).

Next, we used denovolyzeR to test for DN enrichment in CP trios on an exome-wide basis. We found that CP probands had significantly more coding DNAs (1.23, $p=1.24 \times 10^{-7}$) than would be expected by chance based on mutational models (Figure 4.1A) (207). We found that there was no enrichment in synonymous variants (1.01, $p=0.45$), as would be expected because these types of variants are typically not causal for disease, but there was significant enrichment of protein-altering variants (1.34, $p=2.18 \times 10^{-9}$), driven by both missense (1.29, $p=6.57 \times 10^{-7}$) and LOF (1.64, $p=1.02 \times 10^{-4}$) classes.

We next stratified the cohort based on characteristics of the probands: male versus female (Figure 4.1B) and by CP subtype (where numbers allowed) (Figure 4.1C, Table S4.5). Although females are affected with CP more commonly than males, females (1.30, $p=2.74 \times 10^{-5}$) and males (1.37, $p=1.53 \times 10^{-5}$) were significantly enriched for protein-altering DNAs with similar signals from missense (female: 1.25, $p=6.64 \times 10^{-4}$; male: 1.34, $p=1.93 \times 10^{-4}$), and LOF DNAs (female: 1.63, $p=2.44 \times 10^{-3}$; male: 1.63, $p=0.01$). We confirmed no statistically significant differences between sexes in any variant class based observed versus expected values while accounting for male and female prevalence differences (Table S4.4).

We then compared enrichment by CP subtype for cleft hard and soft palate (CH&SP), cleft soft palate (CSP), and submucous cleft palate (SMCP) (Figure 4.1C). Although the overall protein-altering enrichment for each subtype was similar (CH&SP 1.67, $p=1.93 \times 10^{-9}$; CSP 1.47, $p=3.38 \times 10^{-6}$; SMCP 1.80, $p=1.75 \times 10^{-3}$), the distribution of variant classes was interesting as the CSP signal was stronger for LOF than missense variants (missense 1.36, $p=4.69 \times 10^{-4}$; LOF 2.23, $p=9.28 \times 10^{-5}$) whereas CH&SP had similar enrichment for both (missense 1.67, $p=7.36 \times 10^{-8}$; LOF

1.72, $p=0.022$), and SMCP was only significant for missense variants (missense 1.79, $p=3.66 \times 10^{-3}$; LOF 1.91, $p=0.16$). We performed pairwise comparisons for the number of individuals with or without missense variants and with or without LOF variants across each of the subtypes using a Fisher's one-tailed test, with no significant difference found for any test.

Although we have a small sample of syndromic individuals ($n=33$) and individuals with Pierre Robin Sequence (PRS, $n=17$), we were interested in knowing whether either of these categories were heavily contributing to our overall enrichment signal as either group could have different etiologies than other types of CP. Thus, we performed sensitivity analyses in which we removed syndromic individuals (Figure 4.1D) or individuals with PRS (Figure 4.1E). However, neither comparison resulted in any substantial change to the overall enrichment in the cohort (Table S4.5).

Gene-specific analyses

We next performed analysis on a per-gene basis to identify individual genes with a significant excess of DNs. Because of limited samples, the rest of our evaluations combined CSP and SMCP. Genes were considered statistically significant if they reached a Bonferroni multiple-testing threshold correcting for 572 genes with at least one DN ($p < 8.7 \times 10^{-5}$) and were considered exome-wide significant at a more conservative threshold ($p < 1.3 \times 10^{-6}$) (Figure 4.2). In the full cohort, we identified two genes reaching exome-wide significance: *COL2A1* for pLOF ($p=1.91 \times 10^{-13}$) and protein-altering ($p=2.39 \times 10^{-9}$) DNs and *IRF6* for protein-altering DNs ($p=6.19 \times 10^{-7}$). There were an additional five genes with significant enrichment: pLOF variants in *MYH3* ($p=1.68 \times 10^{-5}$) and protein-altering variants in *PRKCI* ($p=1.71 \times 10^{-6}$), *SAT2B* ($p=2.81 \times 10^{-6}$), *POLR1F* ($p=5.27 \times 10^{-5}$), and *SLC25A41* ($p=7.75 \times 10^{-5}$) (Table S4.6)

Of these, *COL2A1* (210), *IRF6* (94), *MYH3* (211), and *SATB2* (212, 213) are associated with known syndromes featuring CP, as are *GRHL3* (96) and *TGFBR 2*(214) for which we had nominal significance ($p=1.74 \times 10^{-4}$ for both genes). Given that we did not restrict our cohort by the presence of additional clinical features, we repeated our analysis in 440 presumed isolated CP probands, and removed anyone classified as possibly, probably, or confirmed syndromic. This did not substantially change our results, and in fact, each of the named genes was more significant in this analysis as none of the syndromic probands harbored DNs in these genes (Table S4.6). Similarly, we repeated this analysis without CP probands diagnosed with PRS; however, in this analysis, *IRF6* was no longer significant ($p=1.07 \times 10^{-4}$) as one of our PRS probands harbors a missense DN in this gene (Table S4.6)..

Although our syndromic probands did not have DNs in the top genes, we did look at these probands individually to determine if there were any likely causal variants (Table S4.3). We identified a novel frameshift variant in *RPL5* in a proband with CP, a vascular ring anomaly, myopia, and ADHD. Heterozygous truncating mutations in *RPL5* are associated with Diamond-Blackfan Anemia (OMIM: 612561) (215), which often features cleft palate and congenital heart defects as reported in our proband. We also found a pathogenic/likely pathogenic (P/LP) splicing variant in *CBL*, which is associated with Noonan syndrome-like disorder with or without juvenile myelomonocytic leukemia (OMIM: 613563), characterized by facial dysmorphism, reduced growth, developmental delays, and other structural anomalies (216). This proband features a CP, developmental delay, growth concerns, epilepsy, and an enlarged median ventricle, which all fit within the phenotypic spectrum of this disorder. Lastly, we identified a novel frameshift variant in *KAT6B* in a proband with clinical features resembling Roselli-Gulienetti syndrome (OMIM: 225000). Although a more detailed description of this proband is not available, the overlap in

features caused by heterozygous variants in *KAT6B* (217) (Say-Barber-Biesecker-Young-Simpson syndrome, OMIM: 603736 and genitopatellar syndrome, OMIM:606170) is plausible that this variant is responsible for disease.

Finally, we compared gene-specific enrichment in probands with CH&SP and CSP or SMCP (Table S4.7, Figure 4.3). In the CH&SP group, we found *SATB2* ($p=4.59 \times 10^{-8}$), *TGFBR2* ($p=1.12 \times 10^{-5}$), and *COL2A1* ($p=7.77 \times 10^{-5}$) were significantly enriched for protein-altering DNAs. In the CSP group, we found exome-wide significance for *PRKCI* ($p=5.34 \times 10^{-8}$), as well as nominal significance for *COL2A1* ($p=1.61 \times 10^{-4}$). Notably, however, all 5 *COL2A1* variants are putative LOF, resulting in more significant enrichment if we only consider the LOF category (CH&SP, $p=1.77 \times 10^{-6}$ and CSP $p=3.68 \times 10^{-6}$). Except for *COL2A1*, the DNAs in these genes were found only in individuals with the subtype showing enrichment of DNAs (i.e., *SATB2* variants were only observed in probands with CH&SP). We note however that there were 5 variants in *COL2A1* and 2-3 in the other genes, which limits our ability to firmly establish genotype-phenotype correlations. In addition, we only had three individuals with cleft hard palate, an extremely rare CP subtype, so it is not possible to contrast individuals whose clefts affect separate parts of the palate.

Gene set enrichments

We next looked at enrichment in gene ontology for biological process, cellular component, and molecular function, as well as the Human Phenotype Atlas and Disease categories using ToppFun (209) (Figure 4.4). We identified a significant enrichment in terms related to embryonic development, ATP-dependent activity, cytoskeletal motor activity, actin-based cell projections, and cleft phenotypes in humans. In general, these findings were consistent with would be expected from a dataset of CP probands (Table S4.8).

We then evaluated enrichment of CP DNs within a panel of 418 OFC-associated genes (123), which was unrestricted in regard to syndromic or isolated clefting (Figure 4.5, Table S4.9). Our list of protein-altering DNs was significantly enriched for all genes in this panel (3.49, $p=4.27 \times 10^{-11}$). When divided by inheritance patterns, there were 178 genes associated with autosomal dominant (AD) conditions, for which we had 33 DNs in 22 genes, resulting in significant enrichment (6.27, $p=4.48 \times 10^{-16}$). Comparatively, there was no significant enrichment for the 180 genes associated with autosomal recessive (AR) disease, for which we had 8 DNs in 7 genes (1.58, $p=0.14$). We also compared subtypes (Figure S4.2), which largely followed the same trends as the full cohort, though we did find marginal enrichment for CSP in AR genes ($p=0.04$) whereas CH&SP was not significant ($p=0.72$). To relate our findings to the biological process of palatogenesis, we used a dataset of marker genes from snRNAseq of the secondary palate of mice at the time of palatal shelf fusion. We investigated our DNs for enrichment in each cell type to pinpoint cell-type specificity for risk for CP or its subtypes. Protein-altering DNs, missense DNs, and LOF DNs were significantly enriched among all palatal marker genes (1.60, $p=3.27 \times 10^{-6}$; 1.39, $p=2.78 \times 10^{-3}$; 3.03, $p=1.24 \times 10^{-6}$, respectively) (Figure 4.6A). Like the exome-wide analysis, we noted that marker genes with LOF DNs the most enriched for CSP whereas marker genes with missense DNs was more significant for CH&SP (Figure 4.6A, Table S4.10).

We then evaluated enrichment of DNs within specific cell clusters (Table 4.1, Figure 4.6B). The most significantly enriched cluster was chondrocyte progenitors (4.46, $p=4.47 \times 10^{-5}$), followed by the endothelium (1.92, $p=6.46 \times 10^{-4}$), the epithelium (1.71, $p=5.41 \times 10^{-3}$), muscle progenitors (1.64 $p=0.020$), and early osteoprogenitors (2.32, $p=0.025$). When split by subtype, both CSP and CH&SP were individually enriched for chondrocyte progenitors and the endothelium. CSP was enriched for the epithelium, whereas CH&SP was enriched for late osteoprogenitor and muscle

progenitor cells (Table S4.10). The list of individual genes harboring DNs for each of these subtypes can be found in Table 1.

Lastly, we investigated our data in the context of probands with an CP in the Deciphering Developmental Disorders (DDD) study (124). Probands in DDD were ascertained based on undiagnosed neurodevelopmental disorders (NDDs) and/or congenital anomalies, abnormal growth parameters, dysmorphic features, and unusual behavioral phenotypes (218). In 488 CP trios from DDD, they report 68 P/LP DNs in 102 genes. We compared the overlap of these 102 genes with the list of 439 genes with protein-altering DNs from CPSeq. There were 11 genes in common: *ARID1A*, *BRPF1*, *COL2A1*, *DYNCH1L1*, *EFTUD2*, *KAT6B*, *KMT2D*, *NEDD4L*, *SATB2*, *SF3B4*, and *TGFBR1* (Table 4.2). In CPseq, there were 19 probands with protein-altering DNs in these genes, and only 2 were reported syndromic: one with an *ARID1A* missense variant with unknown additional features and another with a frameshift deletion in *KAT6B* with features similar to Rosselli-Gulienetti Syndrome, described above. Conversely, we considered the top genes in CPSeq within the DDD study, noting an absence of P/LP variants in genes such as *IRF6* or *GRHL3*. We also did not find *POLR1F*, *PRKCI*, or *SLC25A41* in the DDD list. However, given that this set of genes from DDD is not comprehensive and only includes P/LP variants, we would not expect to find these newly discovered genes here, and their impact on individuals with both NDDs and CP is yet to be elucidated.

DISCUSSION

Here we performed a large-scale investigation of coding DNs in CP, providing additional evidence for several genes with known roles in clefting and identifying new genes of interest. We also utilized a variety of resources to prioritize genes with biological relevance and investigate

differences across several groupings, focusing on syndromic versus presumed isolated CP and comparison of CP subtypes.

We found that CP probands are significantly enriched for protein-altering DNs as a whole, and when divided into the various groups. There were no significant differences in enrichment for male versus female probands or across subtypes, though there was a trend toward higher LOF enrichment in CSP cases compared to CH&SP. This finding is interesting, as some have postulated that CSP could be a less severe form of CP compared to CH&SP as it affects fewer palatal structures and thus, one may expect a higher rate of LOF variants in CH&SP. Although the difference in groups is not significant, evidence from this cohort suggests that the severity or subtype of cleft may not be correlated with the class of variant.

Our results were also not substantially different on sensitivity analysis in which we dropped syndromic or PRS probands, suggesting DNs in these probands are not driving the overall enrichment of DNs. We did, however, identify several individual genes commonly associated with syndromic CP, including *COL2A1*, *IRF6*, *SATB2*, and *MYH3*, as well as *GRHL3* and *TGFBR2* at nominal significance. It is worth noting, however, that our isolated cases are presumed to be so—a limitation of this dataset is that some individuals in this cohort were likely recruited as infants and some syndromes or features are not apparent until a child is older; thus, we may have individuals that did not have additional congenital anomalies but who went on to display signs of developmental delay or farsightedness (as is seen with Stickler syndrome) later in life. While variable expressivity (e.g., isolated clefts) has been documented for *IRF6* (92, 94), *GRHL3* (96, 108), and *COL2A1* (105, 106), this is less true for the others. Although further validation is needed to confirm, our findings here suggest an expansion of the phenotypic spectrum for *SATB2*, *MYH3*, and *TGFRB2* to include isolated CP.

Furthermore, when we compared our DN gene list to P/LP variants from the DDD study, we found 19 probands with protein-altering DNs in overlapping genes. Because of the ascertainment criteria for DDD, it is unsurprising that the gene with the most P/LP variants is *SATB2* (accounting for 2.3% of individuals), as we expect the DDD cohort to be enriched for NDDs and intellectual disabilities. Similarly, as CPSeq was ascertained on CP, we are likely depleted for these individuals. While *SATB2* DNs were found in only 0.6% of CPSeq, we were more likely to find genes associated with isolated CP or those with significant variable expressivity such as *IRF6*. Taken altogether, this implies that there is not currently a single list of genes that represents all individuals with cleft palate; thus, a combined study encompassing all forms of syndromic and isolated cases is needed to fully characterize the underlying genetic architecture. Until such a study is performed, care should be taken when creating lists of “cleft palate genes.”

Unsurprisingly, when we restricted our target gene list to a curated OFC-specific panel, we found significant enrichment for our cohort. Further, genes associated with conditions with AD inheritance had over six times the expected number of protein-altering DNs, whereas genes associated with AR conditions were not significantly enriched in this dataset, largely reflecting the known inheritance patterns for CP and the principals of population genetics. Importantly, this gene panel is not restricted to genes associated with isolated OFCs, again highlighting the idea of an expanded spectrum of phenotypes.

We identified three genes (*POLR1F*, *PRKCI*, and *SLC25A41*) not previously associated with CP. *POLR1F* is a component of the RNA polymerase I complex, a key component of ribosome biogenesis. Interestingly, other components of this complex have been implicated in dominant (*POLR1D* (219), *POLR1B* (220)) and recessive (*POLR1C* (219)) Treacher-Collins syndrome and acrofacial dysostosis, Cincinnati-type (*POLR1A* (221)), both of which often feature

CP as part of the syndrome. *SLC25A41* is a poorly characterized calcium-independent mitochondrial carrier important for adenine nucleotide transfer across the inner mitochondrial membrane (222), and thus its role in CP is unclear. *PRKCI*, which also reached exome-wide significance in CSP probands, encodes for atypical protein kinase C iota (PKCi), a serine/threonine kinase. PKCi plays a variety of roles in cell differentiation, migration, and proliferation (223), though most studies of its function have been performed in the context of various cancers (224). However, given that there were three DNPs (including two at the same residue (N383S)) in individuals with cleft soft palate, additional functional testing to validate its role in palatal development is needed.

We found subtype-specific enrichment for *SATB2* and *TGFBR2* with CH&SP, *PRKCI* with CSP, but no subtype-specificity for *COL2A1*, which was similarly enriched in both subtypes. Although we only have a handful of DNPs in each of these genes, this may represent a genotype-phenotype correlation. To explore these subtype differences further, we utilized a dataset of snRNAseq marker genes from the mouse palate at the time of osteogenesis (E15.5). Although mouse and humans undergo remarkably similar palatogenesis (22), there are some differences and thus our human data may not be directly translatable in all instances. Still, due to how early embryogenesis occurs, a similar human dataset does not exist. We found the most significantly enriched cluster for all CP probands was chondrocyte progenitor cells, with DNPs in *COL2A1* driving most of the signal. This enrichment was similar even between CH&SP and CSP as both subtypes harbor DNPs in this gene. Both groups were also enriched for the endothelial cells, albeit with differences in the contributing genes. In contrast, we found CSP probands were specifically enriched in the epithelial cell cluster, whereas CH&SP was nominally enriched for muscle progenitor and late osteoprogenitor cell clusters. This suggests that CP sub-phenotypes may arise

due to variants in specific cell types, and disruption of specific functions or processes could lead to the observed phenotypic differences.

Unfortunately, due to small sample sizes for rarer subtypes, particularly SMCP, PRS, and cleft hard palate, we could not perform statistical analyses on these groups in the context of most of our subtype analyses. While we can estimate the contribution of DNPs in these subtypes with sensitivity analyses, we cannot yet identify subtype-specific effects or new genes exclusively associated with these forms of OFC.

In conclusion, we have shown an enrichment of protein-altering DNPs in CP probands, and identified three new risk genes of interest. We showed differences between CP subtypes in both specific gene enrichment as well as within specific cell types in the developing mouse secondary palate, suggesting distinct genetic risks or modifiers for sub-phenotypes. We also found enrichment across many genes in isolated CP that have largely been associated with CP syndromes, such as *COL2A1* and *SATB2*. Although additional validation of these findings is needed, our results suggest that some instances of isolated CP may be part of a phenotypic spectrum rather than independent of syndromic CP. When considering this in a larger clinical picture, genetic testing for any individual born with a CP regardless of the presence of additional clinical features may be fruitful. However, the extent to which genes such as *COL2A1* or *SATB2* contribute to isolated CP remains to be seen. Future studies focused on the spectrum of phenotypes in genes associated with OFC syndromes, as well as deeper exploration of subtype genetic risks are warranted to better understand palatogenesis and the genetic architecture of CP.

FIGURES:

Figure 4.1: Exome-wide enrichment for CP as determined by denovolyzeR. A) Radar plot for $-\log_{10}(p\text{-value})$ for each category tested. The middle, dotted line represents $p=0.005$ and the outside line represents $p=1 \times 10^{-10}$. B) Comparison of enrichment for males versus females. C) Enrichment for all CP compared to subtypes. D) Enrichment for isolated versus syndromic CP. E) Enrichment for non-PRS vs PRS CP. The horizontal dotted line at 1 represents no enrichment (observed = expected value). Variants classes are represented by the following colors: gray=synonymous, blue=missense, red=LOF, purple=protein-altering. CH&SP=cleft hard & soft palate, SMCP=submucous cleft palate, PRS=Pierre Robin Sequence.

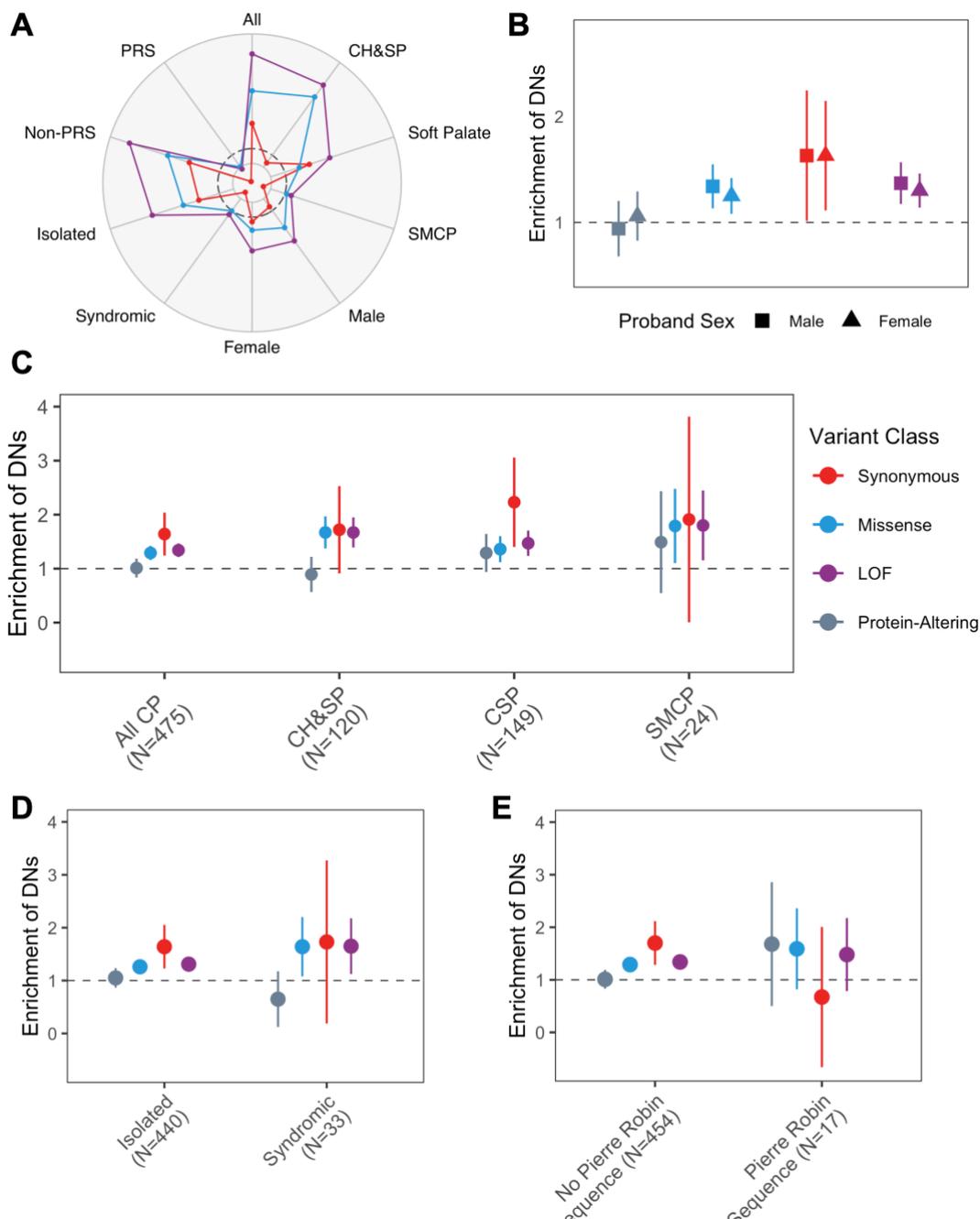


Figure 4.2: Gene-specific enrichment for CP as determined by denovolyzeR. Loss of function variants are shown on the top in red, and all protein-altering variants are shown on the bottom, in purple. The dotted lines represent $p=8.7 \times 10^{-5}$ (Bonferroni correction for 572 genes) and the solid lines represent exome-wide significance at $p=1.3 \times 10^{-6}$. Significant or near significant genes are labelled with the gene name and the number of variants in parenthesis.

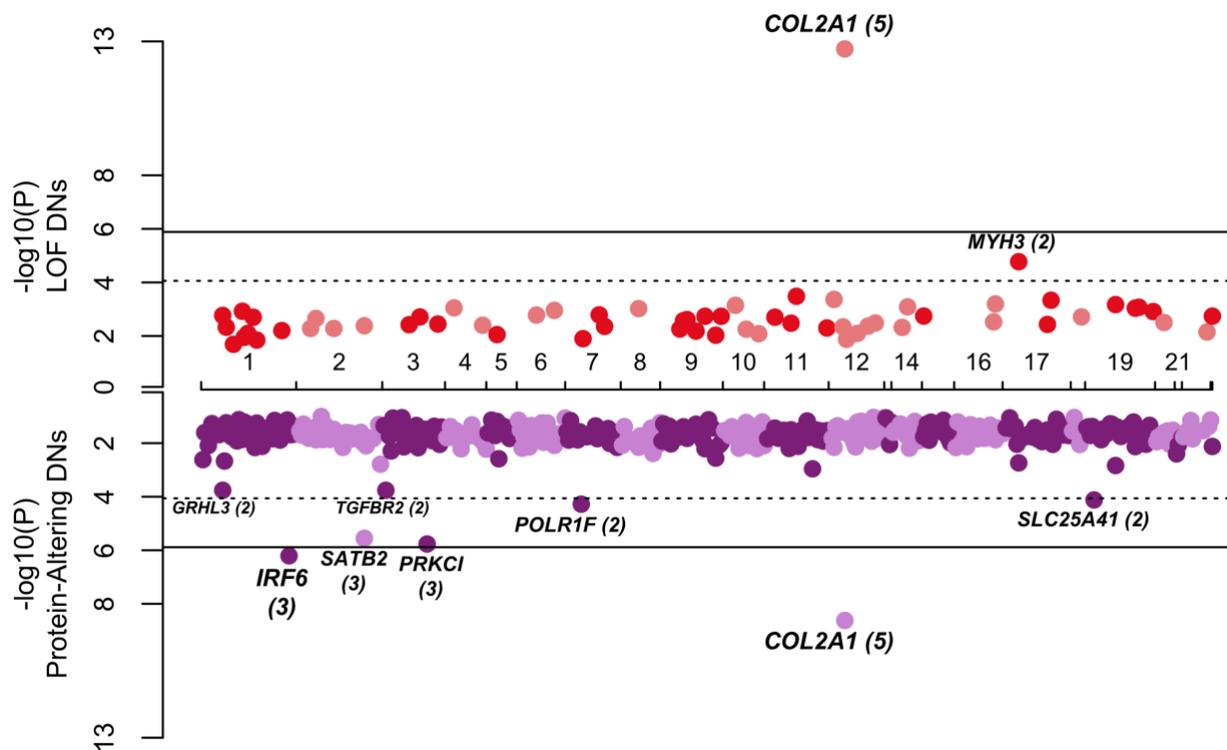


Figure 4.3: Gene-specific enrichment for protein-altering variants in cleft soft palate + submucous cleft palate (top) and cleft hard & soft palate (bottom) as determined by denovolyzeR. The dotted lines represent $p=8.7 \times 10^{-5}$ (Bonferroni correction for 572 genes) and the solid lines represent exome-wide significance at $p=1.3 \times 10^{-6}$. Significant or near significant genes are labelled with the gene name and the number of variants in parenthesis.

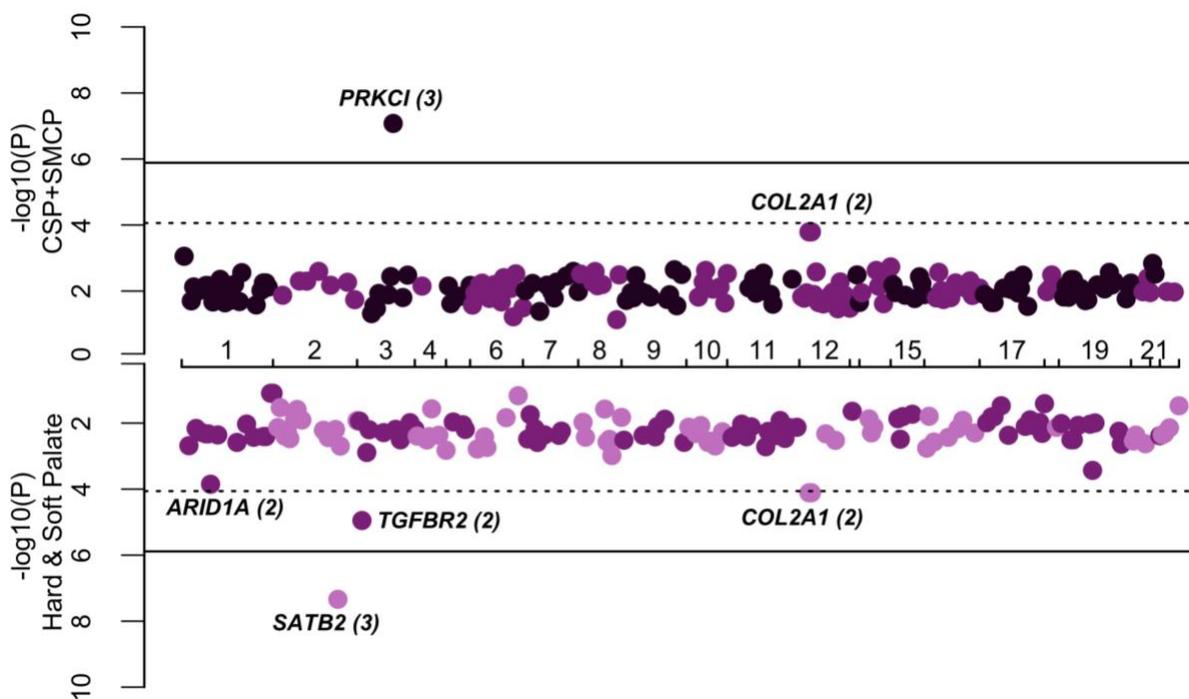


Figure 4.4: Gene set enrichment analysis with ToppFun. Terms on the y-axis are in descending order of significance and grouped by data source on the x-axis. Points are colored by $-\log_{10}(\text{FDR})$ with size determined by the gene ratio (the degree of overlap of genes with DNPs in CPSeq and the number of genes in the respective terms).

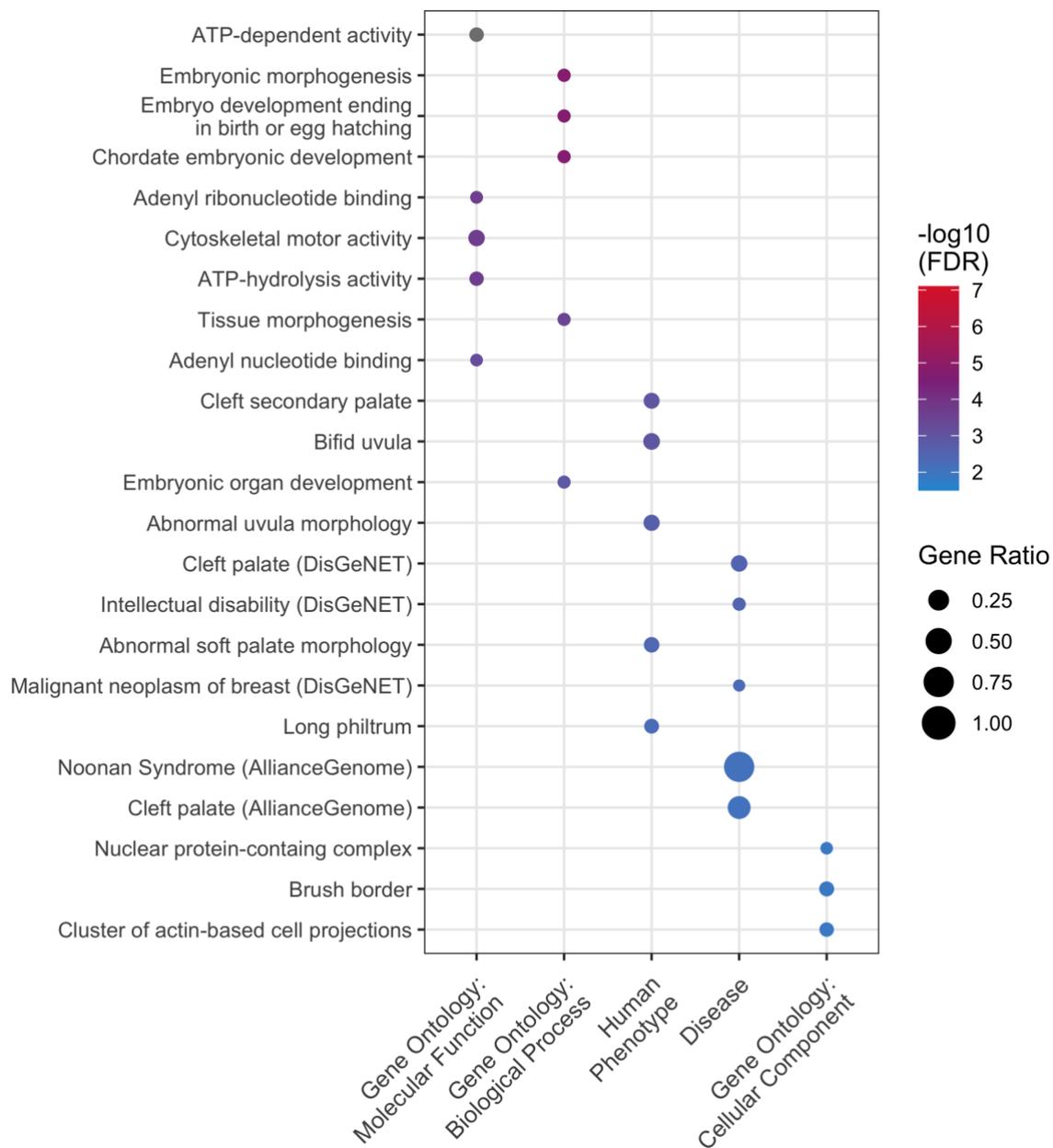


Figure 4.5: Enrichment for CP in a specific set of OFC-associated genes as determined by denovolyzeR. Enrichment for all CP compared to subtypes. The horizontal dotted line at 1 represents no enrichment (observed = expected value). Variants classes are represented by the following colors: gray=synonymous, blue=missense, red=LOF, purple=protein-altering. CH&SP=cleft hard & soft palate, SMCP=submucous cleft palate.

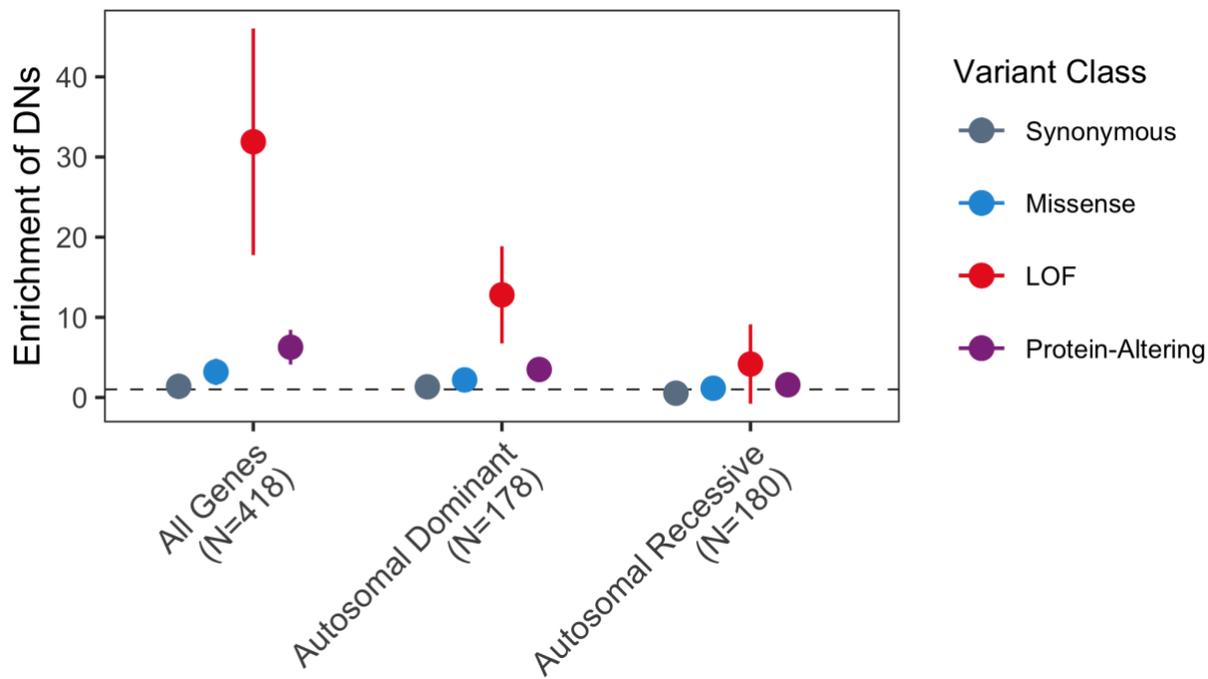
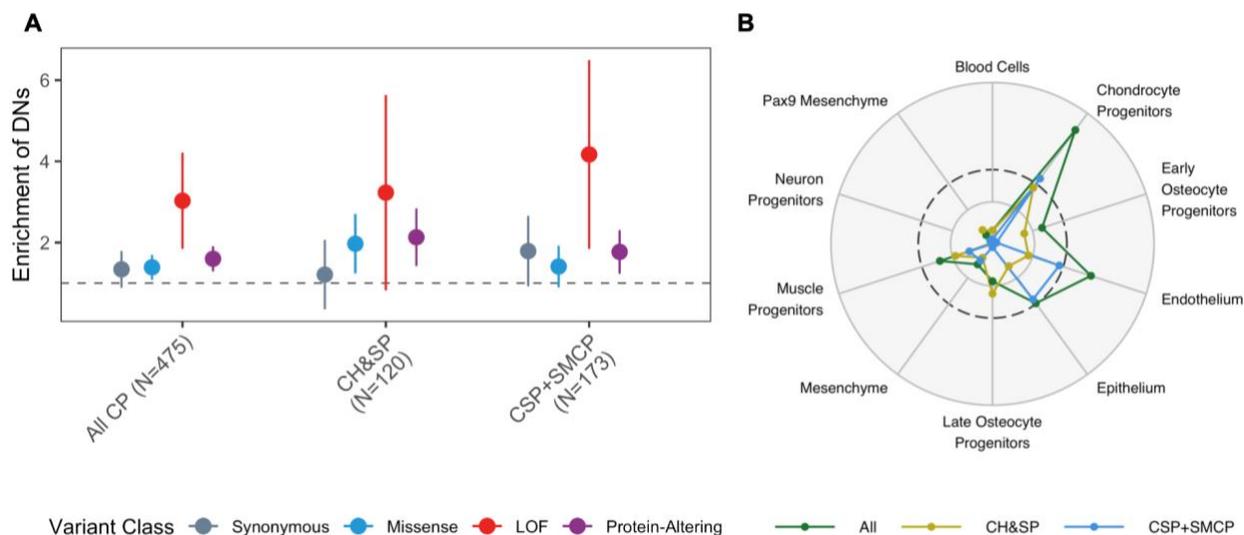


Figure 4.6: Enrichment for CP in a set of marker genes from the secondary mouse palate at E15.5 as determined by denovolyzeR. A) Enrichment for all CP compared to subtypes. The horizontal dotted line at 1 represents no enrichment (observed = expected value). B) Radar plot for $-\log_{10}(\text{p-value})$ for each subtype. The middle, dotted line represents $p=0.005$ and the outside line represents $p=1 \times 10^{-5}$. Variants classes are represented by the following colors: gray=synonymous, blue=missense, red=LOF, purple=protein-altering. CH&SP=cleft hard & soft palate, SMCP=submucous cleft palate. Subtypes are represented by the following colors: green=all CP, yellow=CH&SP, blue=CSP+SMCP.



TABLES:

Table 4.1: Subtype enrichment for protein-altering DNPs in palatal osteogenesis marker genes					
Cluster (# genes)	All CP p-value	CSP		CH&SP	
		Genes	p-value	Genes	p-value
Blood cells (137)	0.386	<i>IKZF1</i>	0.683	<i>SIRPA</i> <i>HDAC9</i>	0.302
Chondrocyte progenitors (81)	4.47E-05	<i>COL2A1</i> (2) <i>COL11A1</i> <i>PEG3</i> <i>FLNB</i>	0.00211	<i>COL2A1</i> (2) <i>SYTL2</i> <i>FRY</i>	0.00354
Early osteoprogenitors (137)	0.0247	<i>NAV3</i>	0.715	<i>SATB2</i> (3)	0.058
Endothelium (661)	0.000646	<i>DYNC1H1</i> (2) <i>C8ORF46</i> <i>ARPC1B</i> <i>ACTB</i> <i>GALNT18</i> <i>TEK</i> <i>FMNL3</i> <i>HEG1</i> <i>COL4A2</i> <i>PTPRB</i> <i>CCDC88C</i> <i>NAV3</i> <i>FLNB</i>	0.00303	<i>TGFBR2</i> (2) <i>TSPAN15</i> <i>MCU</i> <i>RBMS1</i> <i>SIGIRR</i> <i>SWAP70</i> <i>FRY</i> <i>RNF213</i>	0.0237
Epithelium (599)	0.00541	<i>PRKCI</i> (3) <i>MEIS2</i> <i>MSI1</i> <i>PTPRU</i> <i>INADL</i> <i>DNAH11</i> <i>TACR1</i> <i>IRF6</i> <i>EYA1</i> <i>GALNT18</i> <i>HIF3A</i> <i>DSC2</i>	0.00339	<i>ARID1A</i> (2) <i>ARHGDIG</i> <i>CELSR1</i> <i>IRF6</i> <i>KIAA1377</i> <i>FRY</i> <i>ALMS1</i>	0.0592
Late osteoprogenitors (332)	0.0703	<i>COL11A1</i> <i>HEG1</i>	0.746	<i>SATB2</i> (3) <i>HPSE</i> <i>FRY</i> <i>SYNE1</i>	0.0119
Mesenchyme (229)	0.165	<i>RPL12</i> <i>EMILIN2</i>	0.186	<i>PAPPA</i> <i>MAB21L2</i> <i>CDH11</i>	0.203

		<i>EPHB3</i> <i>CACNA1G</i>			
Muscle progenitors (508)	0.0197	<i>HIPK3</i> <i>MYH3</i> <i>TSPAN33</i> <i>PYGM</i> <i>HEG1</i> <i>PEG3</i> <i>IGF2R</i> <i>MDN1</i>	0.123	<i>MYH3</i> <i>MCU</i> <i>DBN1</i> <i>TFRC</i> <i>SMC6</i> <i>SYTL2</i> <i>SYNE1</i>	0.023
Neuron progenitors (451)	0.98	<i>MSI1</i> <i>KIF21A</i>	0.932	<i>GAD2</i> <i>SLC7A14</i>	0.805
Pax9 mesenchyme (124)	0.464	<i>COL11A1</i>	0.721	<i>RBMS1</i> <i>CDH11</i>	0.222

Table 4.2: Overlap in CPSeq genes with protein-altering DNPs and DDD genes with P/LP DNPs

Gene	CPSeq # Variants	CPSeq Percent (out of 475)	DDD # Variants	DDD Percent (out of 488)
ARID1A	2	0.4%	1	0.2%
BRPF1	1	0.2%	1	0.2%
DYNCH1L1	2	0.4%	1	0.2%
COL2A1	5	1.1%	2	0.4%
EFTUD2	1	0.2%	3	0.6%
KAT6B	1	0.2%	3	0.6%
KMT2D	1	0.2%	2	0.4%
NEDD4L	1	0.2%	2	0.4%
SATB2	3	0.6%	11	2.3%
SF3B4	1	0.2%	1	0.2%
TGFBR1	1	0.2%	1	0.2%
IRF6	3	0.6%	0	0.0%
GRHL3	2	0.4%	0	0.0%
TGFBR2	2	0.4%	0	0.0%
MYH3	2	0.4%	0	0.0%

SUPPLEMENTAL DATA

Figures

Figure S4.1: Distribution and frequency of *de novo* variants. A) Histogram showing frequency of DNs in CPSeq trios. B) Distribution of DNs by class.

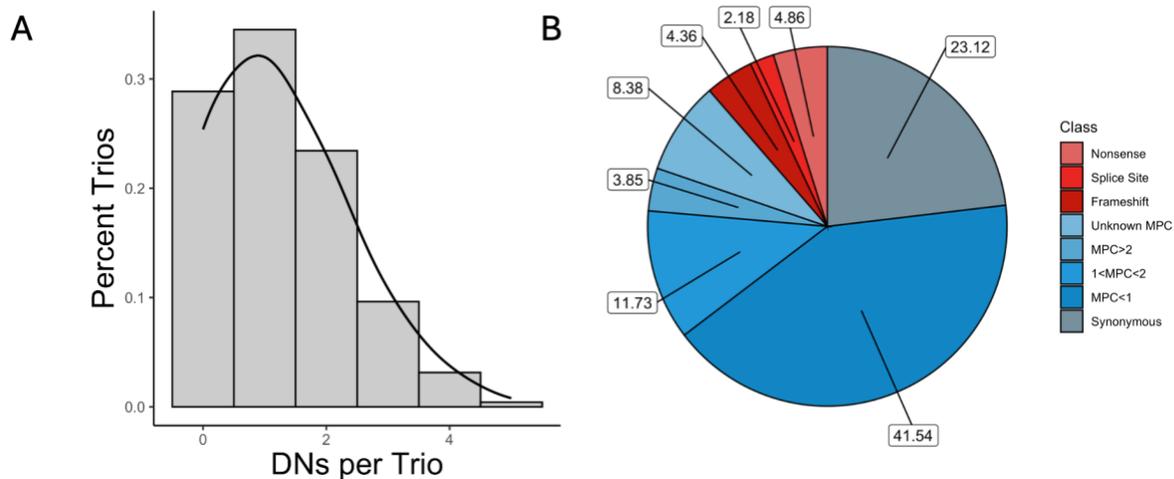


Figure S4.2: Enrichment of genes associated with the OFC-gene panel. Enrichment should be observed for each subtype in all genes (left), autosomal dominant (middle), and autosomal recessive (right) conditions.

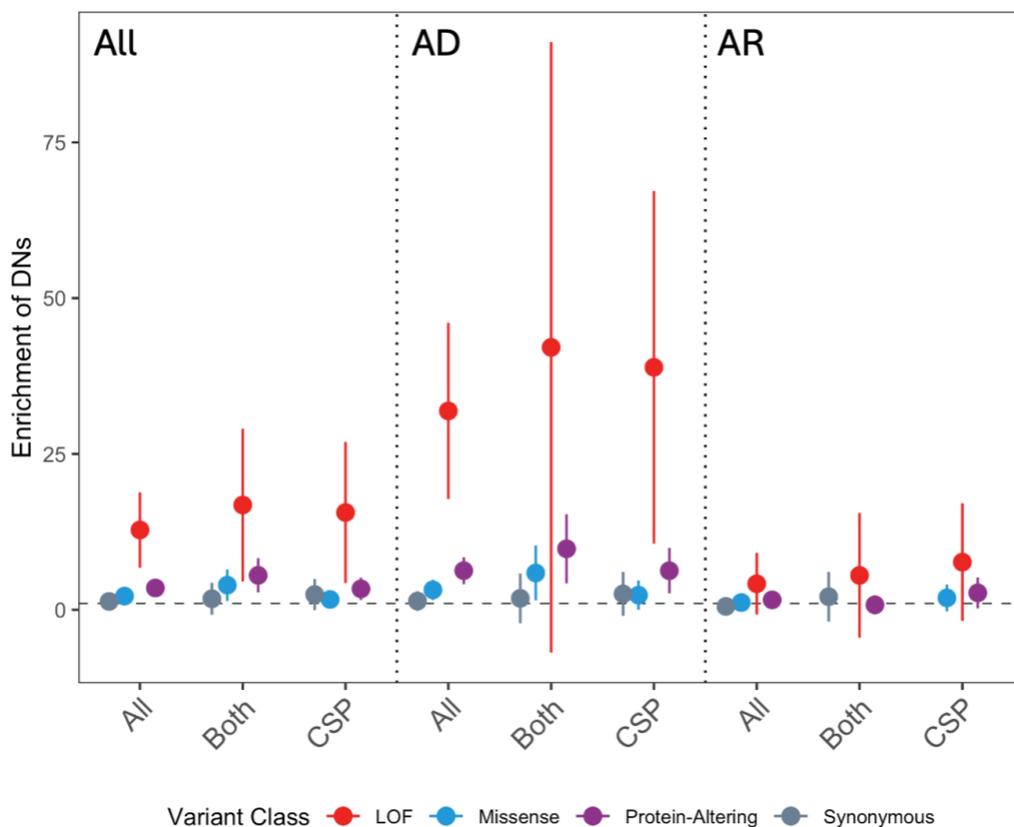
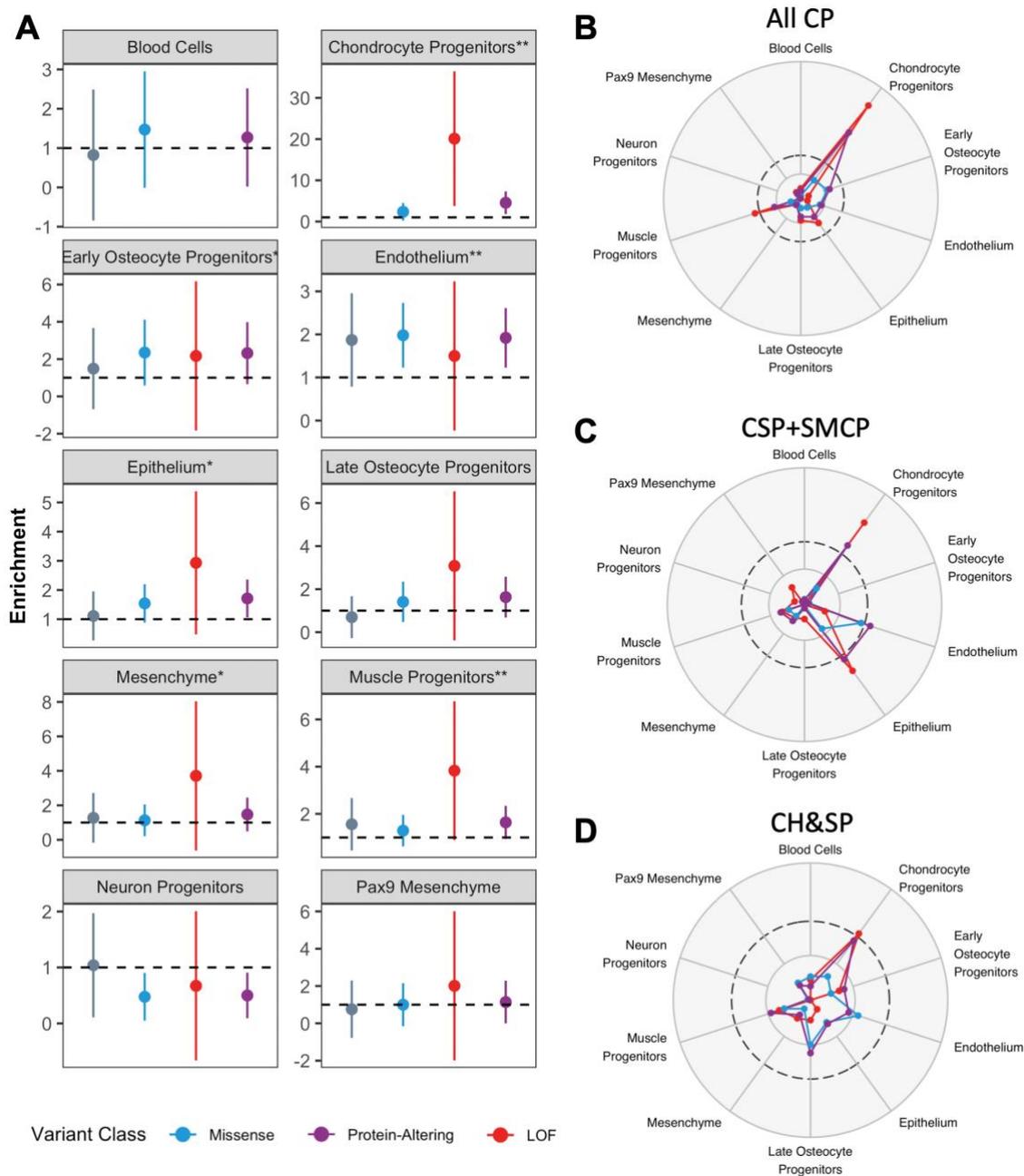


Figure S4.3: Enrichment of genes associated with the secondary palate at E15.5. A) Enrichment for all CP within each cluster. B) Radar plot for $-\log_{10}(\text{p-value})$ for each cluster in all CP, C) CSP+SMCP, and D) CH&SP.



Tables

		Synd- romic	CH&SP	CSP	SMCP	CHP	Unknown	Total
Asian	Male	3	34	37	2	1	18	92
	Female	6	51	69	3	0	39	162
African or Black/ Afr. American	Male	1	6	3	6	0	8	23
	Female	3	2	5	7	0	24	38
Other/ Unknown	Male	1	4	2	0	0	9	15
	Female	5	3	7	0	0	9	19
European	Male	7	10	13	5	1	37	66
	Female	7	9	13	1	1	36	60
		33	119	149	24	3	180	475

CP subtype category	Sex	Clinical Features
CP - Type Unknown	F	CP - Pierre Robin, heart murmur, double inguinal hernias
CP - Type Unknown	F	CP, vascular ring heart defect, myopia, ADHD
CP - Type Unknown	M	Stickler Syndrome; CP - Pierre Robin, strabismus, small mandible, far sighted
CP - Type Unknown	F	CP, intellectual disability, obese, colon atresia, renal disease, talipes, hearing loss
CP - Type Unknown	F	Unknown; no details provided
CP - Type Unknown	F	Unknown; no details provided
Soft Palate	M	Possible Robinow (Tetralogy of Fallot, macrocephaly, hearing loss, hypermobility, learning disability, dry skin, maxillary hypoplasia, hyperopia, dysnomia, hypertelorism, hammer toes)
CP - Type Unknown	M	CP, methemoglobinemia, small ears
CP - Type Unknown	F	CP, lp mounds but no VWS diagnosis
CP - Type Unknown	F	Stickler (no other details)
SMCP	M	Unknown; no details provided
CH&SP	F	Unknown; no details provided
CH&SP	M	CP, bifid uvula, developmental delay, thyroid problem, small jaw, and open valve that closed with medication
CP - Type Unknown	F	Unknown; no details provided
CH&SP	F	CP, developmental delay, growth concerns, epilepsy, and enlarged medial ventricle
Soft Palate	F	VWS; soft cleft palate, lip pits, peg lateral incisors

CP - Type Unknown	F	Moebius syndrome
CP - Type Unknown	F	Possible Rosselli-Gulienetti Syndrome
CH&SP	F	CP, antiphospholipid syndrome, heart murmur, ADHD, oppositional defiant disorder
CH&SP	F	CP, persistent ductus arteriosus
Soft Palate	F	CP, pyloric stenosis, low set ears, prominent brow ridge, curly hair, speech delay, eczema
CP - Type Unknown	M	CP, plagiocephaly, pupillary defect, blepharophimosis, hypospadias, hydronephrosis, vesicoureteral reflux, auricular anomaly, epicanthal folds, overlapping toes
CH&SP	F	Questionable Sticker (CP, micrognathia, flat nasal bridge, bilateral epicanthal folds)
CP - Type Unknown	M	Unknown; no details provided
CP - Type Unknown	M	CP - Pierre Robin, Acral Peeling Skin Syndrome
CH&SP	F	22q11DS
CH&SP	M	Questionable 22qDS (CP + heart murmur)
CH&SP	M	Otopalatodigital Syndrome
CH&SP	F	Pierre Robin Sequence, possible Stickler (CP, micrognathia, very flat nasal bridge, microstomia)
CH&SP	M	VWS; cleft soft palate, lip pits
Soft Palate	M	VWS; cleft soft palate, lip pits
CP - Type Unknown	F	VWS; cleft soft palate, lip pits
CH&SP	F	CP, omphalocele

CHR	POS	Gene	Class	subtype	sex	MPC	HGVS Consequence
1:976624	C/T	Clorf170	mis	CSP	M	NA	NM_001291367 c.G1808A:p.G603E
1:1298378	T/G	ACAP3	mis	TU	F	0.859	NM_030649 c.A907C:p.K303Q
1:3625767	C/T	TPRG1L	syn	CSP	M	NA	NM_182752 c.C348T:p.S116S
1:6628633	G/A	THAP3	mis	CH&SP	F	1.219	NM_138350 c.G209A:p.R70H
1:10404164	C/T	PGD	mis	TU	M	1.085	NM_002631 c.C334T:p.R112W
1:11517478	G/A	PTCHD2	mis	CSP	F	NA	NM_020780 c.G1765A:p.D589N
1:12123329	C/G	TNFRSF8	mis	CSP	M	0.207	NM_001243 c.C992G:p.T331S
1:15101559	G/A	KAZN	mis	CH&SP	M	0.780	NM_201628 c.G1564A:p.E522K
1:15725356	C/T	PLEKHM2	mis	CSP	M	0.234	NM_015164 c.C752T:p.T251I
1:17240669	G/C	PADI1	mis	CH&SP	F	0.223	NM_013358 c.G1667C:p.R556P
1:20722897	C/T	SH2D5	syn	CSP	M	NA	NM_001103160 c.G675A:p.L225L
1:24334680	TGG/T	GRHL3	FS del	CSP	M	NA	NM_021180 c.256_257del:p.G87Qfs*3
1:24347552	C/T	GRHL3	mis	CH&SP	M	1.151	NM_021180 c.C1643T:p.A548V
1:26731597	C/A	ARID1A	mis	CH&SP	F	0.479	NM_139135 c.C1796A:p.P599Q
1:26763175	T/G	ARID1A	mis	CH&SP	F	0.597	NM_139135 c.T2622G:p.N874K
1:29304850	G/C	PTPRU	splicing	CSP	F	NA	NM_001195001 c.2743+1G>C:
1:35392310	C/G	ZMYM4	mis	CSP	M	1.029	NM_005095 c.C2686G:p.P896A
1:36358827	G/A	STK40	syn	TU	F	NA	NM_001282546 c.C123T:p.A41A
1:37540377	T/G	SNIP1	mis	CH&SP	M	1.488	NM_024700 c.A706C:p.T236P
1:39322634	C/G	MACF1	non	CHP	M	NA	NM_012090 c.C4071G:p.Y1357X
1:40239229	A/C	RLF	mis	TU	M	0.452	NM_012421 c.A4527C:p.R1509S
1:42682643	G/T	YBX1	mis	CSP	M	2.330	NM_004559 c.G78T:p.K26N
1:43591380	C/T	PTPRF	mis	TU	M	0.691	NM_130440 c.C1358T:p.P453L
1:45342542	C/T	TOE1	syn	CSP	M	NA	NM_025077 c.C651T:p.D217D
1:54782479	T/C	TTC22	splicing	TU	M	NA	NM_001114108 c.1021-2A>G:
1:55072389	G/C	USP24	non	CSP	F	NA	NM_015306 c.C7617G:p.Y2539X
1:56949598	C/T	C8B	mis	CSP	F	0.041	NM_001278544 c.G635A:p.R212Q
1:61797341	C/T	INADL	non	CSP	F	NA	NM_176877 c.C1315T:p.R439X

1:74724741	A/G	CRYZ	syn	CSP	M	NA	NM_001889 c.T81C:p.D27D
1:84937840	A/C	MCOLN2	mis	CSP	F	0.485	NM_153259 c.T1250G:p.V417G
1:92833429	GAT/G	RPL5	FS del	CH&SP	F	NA	NM_000969 c.45_46del:p.Y16Pfs*5
1:93180226	C/G	TMED5	mis	TU	F	0.427	NM_001167830 c.G17C:p.W6S
1:103008517	CT/C	COL11A1	splicing	CSP	F	NA	NM_001190709 c.1513-2A>-:
1:109013888	C/T	WDR47	mis	TU	M	0.622	NM_001142550 c.G280A:p.A94T
1:112710254	C/T	PPM1J	mis	TU	F	0.402	NM_005167 c.G1427A:p.R476H
1:114779228	A/G	SIKE1	mis	SMCP	M	1.311	NM_001102396 c.T334C:p.Y112H
1:149926709	C/A	SF3B4	mis	TU	F	2.582	NM_005850 c.G373T:p.V125F
1:153944249	C/T	DENND4B	syn	CSP	F	NA	NM_001367466 c.G159A:p.R53R
1:155207869	G/A	THBS3	mis	CH&SP	F	0.369	NM_001252607 c.C8T:p.T3M
1:156844261	C/A	INSRR	syn	CSP	F	NA	NM_014215 c.G2757T:p.G919G
1:158638194	G/A	SPTA1	syn	CSP	F	NA	NM_003126 c.C5028T:p.S1676S
1:167021712	G/A	MAEL	mis	CH&SP	F	0.218	NM_032858 c.G1162A:p.V388I
1:171537347	G/A	PRRC2C	mis	SMCP	M	0.216	NM_015172 c.G2372A:p.R791Q
1:180945402	C/T	KIAA1614	syn	CSP	F	NA	NM_020950 c.C3387T:p.D1129D
1:203225783	G/C	CHIT1	mis	TU	M	0.398	NM_001256125 c.C143G:p.P48R
1:203347564	C/T	FMOD	mis	TU	F	0.283	NM_002023 c.G707A:p.R236Q
1:203840332	C/T	ZC3H11A	non	CSP	F	NA	NM_001319239 c.C1000T:p.R334X
1:204469734	C/T	PIK3C2B	syn	TU	F	NA	NM_002646 c.G69A:p.E23E
1:205271922	T/C	TMCC2	mis	TU	F	1.933	NM_001297613 c.T1208C:p.L403P
1:208084424	C/T	PLXNA2	mis	TU	M	0.274	NM_025179 c.G2254A:p.A752T
1:209788625	C/T	IRF6	mis	TU	F	1.767	NM_006147 c.G1199A:p.R400Q
1:209789719	C/A	IRF6	mis	CSP	F	1.840	NM_001206696 c.G842T:p.G281V
1:209796476	C/T	IRF6	mis	CH&SP	M	2.622	NM_006147 c.G251A:p.R84H
1:223112230	C/T	TLR5	mis	CSP	M	0.356	NM_003268 c.G802A:p.G268R
1:228274388	G/A	OBSCN	mis	CSP	F	0.558	NM_001098623 c.G5627A:p.G1876E
1:228277026	A/T	OBSCN	mis	CH&SP	F	0.523	NM_001098623 c.A6584T:p.D2195V
1:228378717	G/A	OBSCN	syn	CSP	M	NA	NM_001098623 c.G23829A:p.R7943R
1:229602832	G/A	TAF5L	mis	CHP	F	0.273	NM_014409 c.C335T:p.P112L
1:236559825	G/A	HEATR1	syn	TU	F	NA	NM_018072 c.C4659T:p.T1553T

2:17655256	G/A	VSNL1	syn	CSP	F	NA	NM_001366806 c.G438A:p.T146T
2:17721207	T/C	SMC6	mis	CH&SP	M	0.347	NM_024624 c.A781G:p.I261V
2:27508010	A/G	GCKR	mis	TU	M	0.060	NM_001486 c.A1274G:p.Q425R
2:27577963	C/T	C2orf16	mis	TU	M	0.353	NM_032266 c.C1391T:p.T464I
2:27579635	G/C	C2orf16	syn	TU	F	NA	NM_032266 c.G3063C:p.G1021G
2:32415569	A/G	BIRC6	mis	CH&SP	F	0.149	NM_016252 c.A2278G:p.I760V
2:37020712	C/T	HEATR5B	syn	TU	F	NA	NM_019024 c.G3978A:p.A1326A
2:37951514	A/T	RMDN2	mis	TU	F	0.018	NM_144713 c.A299T:p.D100V
2:38995285	T/A	SOS1	non	CSP	F	NA	NM_005633 c.A3184T:p.K1062X
2:46759335	C/T	SOCS5	mis	CH&SP	F	0.447	NM_144949 c.C805T:p.R269W
2:48474766	T/G	PPP1R21	mis	CH&SP	F	0.038	NM_001135629 c.T1172G:p.M391R
2:58084880	G/A	VRK2	splicing	CH&SP	F	NA	NM_001288837 c.187-1G>A:
2:69509340	T/G	AAK1	syn	CSP	F	NA	NM_014911 c.A1897C:p.R633R
2:71423074	G/A	ZNF638	mis	CH&SP	F	0.031	NM_001014972 c.G3560A:p.S1187N
2:73455221	G/A	ALMS1	mis	CH&SP	M	0.104	NM_015120 c.G7600A:p.E2534K
2:73847657	A/T	STAMPB	mis	SMCP	F	0.297	NM_001353976 c.A241T:p.I81L
2:74047251	C/T	TET3	mis	CH&SP	F	0.598	NM_001366022 c.C1055T:p.T352M
2:74506790	G/A	PCGF1	syn	CSP	F	NA	NM_032673 c.C294T:p.L98L
2:74880402	G/A	HK2	mis	TU	M	0.516	NM_000189 c.G1403A:p.R468H
2:75049561	G/C	TACR1	mis	CSP	F	0.245	NM_001058 c.C1095G:p.H365Q
2:97179917	A/C	ANKRD36	syn	CH&SP	F	NA	NM_001354587 c.A1719C:p.G573G
2:98538957	C/T	INPP4A	non	TU	F	NA	NM_001134225 c.C646T:p.R216X
2:101402652	G/T	RFX8	syn	TU	F	NA	NM_001367510 c.C516A:p.V172V
2:108497028	C/T	GCC2	syn	CH&SP	M	NA	NM_181453 c.C4701T:p.S1567S
2:110658489	T/C	BUB1	syn	CSP	F	NA	NM_001278617 c.A1437G:p.T479T
2:112559333	T/G	POLR1B	syn	SMCP	F	NA	NM_001282779 c.T954G:p.A318A
2:119436946	C/T	TMEM37	mis	SMCP	F	1.092	NM_183240 c.C79T:p.R27W
2:120090450	G/A	EPB41L5	mis	TU	M	0.502	NM_001184938 c.G977A:p.R326Q
2:127505167	G/A	IWS1	mis	TU	M	0.369	NM_017969 c.C736T:p.R246C
2:127623363	C/T	MYO7B	syn	CSP	M	NA	NM_001080527 c.C3729T:p.A1243A
2:132318335	G/C	ZNF806	mis	CSP	F	NA	NM_001304449 c.G1369C:p.A457P

2:134412896	G/C	MGAT5	mis	CH&SP	F	1.860	NM_002410 c.G1558C:p.E520Q
2:159747937	G/T	MARCHF7	mis	CH&SP	M	NA	NM_001282806 c.G479T:p.R160M
2:160318177	T/A	RBMS1	mis	CH&SP	F	1.702	NM_016836 c.A302T:p.K101I
2:171728835	C/T	DYNC1I2	mis	CSP	F	0.561	NM_001271786 c.C1352T:p.A451V
2:173913081	C/T	SP3	mis	TU	M	3.016	NM_003111 c.G2018A:p.R673K
2:178733412	C/T	TTN	non	TU	F	NA	NM_001267550 c.G15881A:p.W5294X
2:179047392	T/C	CCDC141	syn	CSP	F	NA	NM_173648 c.A117G:p.V39V
2:182756417	C/T	DNAJC10	mis	CH&SP	M	0.397	NM_001271581 c.C1619T:p.P540L
2:189754267	T/C	OSGEPL1	mis	TU	F	0.234	NM_001354347 c.A721G:p.K241E
2:199348708	C/T	SATB2	mis	CH&SP	F	2.464	NM_001172517 c.G1166A:p.R389H
2:199348709	G/A	SATB2	mis	CH&SP	F	2.561	NM_001172517 c.C1165T:p.R389C
2:199348775	G/A	SATB2	non	CH&SP	M	NA	NM_001172517 c.C1099T:p.Q367X
2:199911637	C/A	C2orf69	mis	CH&SP	F	1.127	NM_153689 c.C199A:p.R67S
2:201209096	C/T	CASP10	mis	TU	F	0.440	NM_032974 c.C949T:p.L317F
2:203461358	C/T	RAPH1	syn	CSP	F	NA	NM_203365 c.G1017A:p.V339V
2:207977003	G/A	PLEKHM3	syn	CH&SP	M	NA	NM_001080475 c.C1194T:p.D398D
2:210192859	C/T	ACADL	mis	CSP	F	0.325	NM_001608 c.G1144A:p.D382N
2:219482798	C/T	SPEG	syn	SMCP	M	NA	NM_005876 c.C5580T:p.G1860G
2:226797617	G/T	IRS1	syn	CH&SP	F	NA	NM_005544 c.C1122A:p.S374S
2:228018794	T/G	SPHKAP	mis	CSP	F	0.050	NM_001142644 c.A2060C:p.N687T
2:232548691	T/C	TIGD1	mis	TU	M	NA	NM_145702 c.A1192G:p.T398A
2:232796138	A/C	GIGYF2	mis	CH&SP	F	1.018	NM_001103147 c.A1619C:p.H540P
3:9739737	C/G	BRPF1	mis	CH&SP	M	1.962	NM_004634 c.C1338G:p.I446M
3:30691459	G/A	TGFBR2	mis	CH&SP	F	1.458	NM_001024847 c.G1639A:p.D547N
3:30691478	G/A	TGFBR2	mis	CH&SP	F	1.649	NM_001024847 c.G1658A:p.R553H
3:40209865	T/C	MYRIP	syn	CSP	M	NA	NM_001284425 c.T1410C:p.D470D
3:42658515	C/T	ZBTB47	mis	TU	M	1.469	NM_145166 c.C160T:p.R54C
3:44754283	C/G	KIAA1143	mis	CH&SP	F	1.238	NM_020696 c.G194C:p.G65A
3:47007275	A/C	NBEAL2	mis	TU	F	1.726	NM_001365116 c.A7157C:p.H2386P
3:48632015	G/A	SLC26A6	mis	CH&SP	F	NA	NM_001281732 c.C511T:p.R171C
3:49662269	G/A	BSN	mis	CSP	F	0.318	NM_003458 c.G10424A:p.R3475Q

3:52438011	G/A	SEMA3G	syn	TU	M	NA	NM_020163 c.C1698T;p.H566H
3:53723989	C/T	CACNA1D	mis	CSP	F	1.237	NM_001128840 c.C2090T;p.T697I
3:58170664	G/A	FLNB	mis	CSP	F	0.807	NM_001164318 c.G7678A;p.V2560I
3:98498122	G/A	OR5K2	mis	TU	F	0.008	NM_001004737 c.G442A;p.A148T
3:101671273	T/G	ZBTB11	mis	SMCP	F	0.718	NM_014415 c.A635C;p.Q212P
3:111118681	A/G	PVRL3	syn	CSP	M	NA	NM_015480 c.A528G;p.K176K
3:113784237	T/TG	ATP6V1A	FS ins	CH&SP	M	NA	NM_001690 c.226dupG;p.D77Rfs*8
3:122879353	A/G	DIRC2	mis	TU	F	NA	NM_032839 c.A1412G;p.Y471C
3:122926604	G/A	SEMA5B	syn	CSP	M	NA	NM_001256347 c.C1086T;p.H362H
3:124997805	C/T	HEG1	mis	CSP	F	0.202	NM_020733 c.G3536A;p.R1179Q
3:129467063	A/G	IFT122	mis	TU	F	0.168	NM_052985 c.A890G;p.N297S
3:133587907	T/C	CDV3	mis	TU	F	0.765	NM_001282763 c.T332C;p.M111T
3:133805867	CGCCGGGTGG CAGATGGCGG CGGT/C	SRPRB	FS del	SMCP	F	NA	NM_021203 c.20_42del;p.V9Gfs*58
3:141292404	G/T	ACPL2	syn	TU	M	NA	NM_001282728 c.G528T;p.G176G
3:157159430	T/G	CCNL1	mis	TU	M	2.167	NM_001308185 c.A353C;p.K118T
3:169121139	C/A	MECOM	mis	TU	F	1.212	NM_001105078 c.G485T;p.C162F
3:170267957	A/G	PRKCI	mis	CSP	M	1.975	NM_002740 c.A407G;p.Y136C
3:170284541	A/G	PRKCI	mis	CSP	M	0.689	NM_002740 c.A1148G;p.N383S
3:170284541	A/G	PRKCI	mis	CSP	F	0.689	NM_002740 c.A1148G;p.N383S
3:170480577	C/T	SLC7A14	mis	CH&SP	M	0.587	NM_020949 c.G1705A;p.V569M
3:180957892	T/C	FXR1	syn	CSP	F	NA	NM_005087 c.T954C;p.I318I
3:184341784	G/A	FAM131A	mis	CH&SP	M	0.529	NM_001366133 c.G37A;p.E13K
3:184572783	G/A	EPHB3	mis	CSP	M	0.759	NM_004443 c.G463A;p.V155M
3:186785917	T/C	EIF4A2	mis	TU	F	1.732	NM_001967 c.T383C;p.M128T
3:188609689	C/T	LPP	non	CH&SP	M	NA	NM_005578 c.C958T;p.Q320X
3:189238788	C/G	TPRG1	mis	CSP	F	0.232	NM_198485 c.C358G;p.L120V
3:193454176	G/A	ATP13A4	mis	CH&SP	F	0.565	NM_032279 c.C1952T;p.T651M
3:196067608	T/C	TFRC	mis	CH&SP	F	0.597	NM_003234 c.A950G;p.N317S
3:197990668	C/T	LMLN	mis	CH&SP	M	1.417	NM_001136049 c.C1039T;p.R347C

4:9908291	T/C	SLC2A9	mis	CH&SP	F	0.120	NM_020041 c.A1057G:p.I353V
4:10444583	T/A	ZNF518B	syn	CSP	F	NA	NM_053042 c.A1746T:p.A582A
4:22388906	C/T	GPR125	mis	TU	M	NA	NM_145290 c.G2765A:p.G922E
4:37839887	A/G	PGM2	mis	TU	M	0.196	NM_018290 c.A481G:p.M161V
4:38829053	C/A	TLR6	non	CSP	F	NA	NM_006068 c.G421T:p.E141X
4:40432691	G/A	RBM47	mis	TU	M	0.434	NM_001371113 c.C1502T:p.A501V
4:56523792	C/T	ARL9	syn	SMCP	M	NA	NM_206919 c.C288T:p.Y96Y
4:67838195	A/T	TMPRSS11D	mis	CH&SP	F	0.616	NM_004262 c.T452A:p.I151K
4:70975266	C/T	MOB1B	mis	TU	F	0.155	NM_001244766 c.C404T:p.T135M
4:80362834	A/G	C4orf22	syn	TU	M	NA	NM_001206997 c.A192G:p.A64A
4:83295489	T/C	HPSE	mis	CH&SP	F	0.433	NM_001166498 c.A1265G:p.N422S
4:84772891	G/A	WDFY3	mis	CH&SP	M	0.462	NM_014991 c.C4793T:p.A1598V
4:95219132	G/T	UNC5C	syn	CH&SP	F	NA	NM_003728 c.C1482A:p.L494L
4:112515179	G/T	NEUROG2	syn	TU	F	NA	NM_024019 c.C297A:p.G99G
4:112587755	T/C	C4orf21	mis	TU	M	NA	NM_001350397 c.A3128G:p.N1043S
4:128871355	G/T	JADE1	mis	TU	F	1.658	NM_001287437 c.G1586T:p.G529V
4:146909237	G/A	TTC29	syn	CH&SP	F	NA	NM_001317806 c.C189T:p.S63S
4:150583359	G/A	MAB21L2	mis	CH&SP	F	1.757	NM_006439 c.G330A:p.M110I
4:151629764	C/T	FAM160A1	syn	CH&SP	F	NA	NM_001109977 c.C1041T:p.S347S
4:166039413	C/T	TLL1	syn	CH&SP	F	NA	NM_012464 c.C1233T:p.D411D
4:168181741	TGTAA/T	ANXA10	splicing	TU	M	NA	NM_007193 c.783+6_783+9del:
4:171814706	C/T	GALNTL6	syn	TU	F	NA	NM_001034845 c.C126T:p.P42P
4:185399750	T/G	ANKRD37	mis	CH&SP	F	0.151	NM_181726 c.T453G:p.N151K
5:1218906	G/A	SLC6A19	mis	TU	M	0.536	NM_001003841 c.G1177A:p.V393M
5:13810153	C/A	DNAH5	mis	TU	F	0.143	NM_001369 c.G7515T:p.E2505D
5:37064557	T/A	NIPBL	mis	TU	M	1.411	NM_133433 c.T8080A:p.S2694T
5:41869720	G/A	OXCT1	syn	TU	F	NA	NM_001364300 c.C24T:p.P8P
5:66170713	G/A	SREK1	mis	CSP	F	0.328	NM_001323533 c.G1250A:p.R417Q
5:71467461	AT/A	BDP1	FS del	CSP	M	NA	NM_018429 c.894delT:p.Y299Tfs*15
5:77077362	C/T	ZBED3	mis	TU	M	1.653	NM_001329564 c.G517A:p.E173K
5:95495008	C/T	TTC37	mis	CH&SP	M	0.063	NM_014639 c.G3646A:p.V1216I

5:124701035	C/T	ZNF608	mis	SMCP	F	2.205	NM_020747 c.G1141A:p.V381M
5:140788116	A/G	PCDHA1	mis	CSP	F	1.034	NM_031410 c.A1826G:p.Y609C
5:140808545	C/G	PCDHA4	mis	SMCP	F	NA	NM_031500 c.C1358G:p.P453R
5:140927341	A/G	PCDHAC1	mis	TU	F	0.958	NM_018898 c.A449G:p.D150G
5:154002376	A/C	FAM114A2	mis	TU	M	0.023	NM_001317994 c.T1131G:p.F377L
5:170918839	T/C	RANBP17	mis	CH&SP	M	0.028	NM_022897 c.T1081C:p.F361L
5:177467590	T/C	DBN1	mis	CH&SP	M	1.589	NM_004395 c.A368G:p.D123G
5:177600219	C/G	B4GALT7	syn	CH&SP	M	NA	NM_007255 c.C9G:p.P3P
5:181200014	T/C	TRIM7	mis	CSP	F	0.078	NM_203297 c.A140G:p.E47G
6:12121807	C/T	HIVEP1	mis	CSP	F	0.478	NM_002114 c.C2012T:p.T671M
6:16327959	C/T	ATXN1	mis	SMCP	M	0.584	NM_001128164 c.G352A:p.V118M
6:27837965	G/T	HIST1H2AK	mis	CH&SP	F	0.134	NM_003510 c.C375A:p.H125Q
6:28920065	G/C	TRIM27	mis	CSP	F	1.963	NM_006510 c.C694G:p.L232V
6:30075336	C/T	RNF39	mis	CSP	F	1.566	NM_170769 c.G454A:p.E152K
6:30171994	G/A	TRIM15	mis	CH&SP	M	0.499	NM_033229 c.G1043A:p.R348H
6:31667959	A/G	CSNK2B	mis	CH&SP	F	2.504	NM_001320 c.A164G:p.D55G
6:31882765	C/T	EHMT2	mis	CSP	F	2.225	NM_001289413 c.G3200A:p.R1067Q
6:33279600	G/A	WDR46	mis	CSP	F	0.997	NM_001164267 c.C1469T:p.P490L
6:33313800	G/T	TAPBP	mis	CSP	F	0.386	NM_172209 c.C102A:p.S34R
6:33321352	CT/C	DAXX	FS del	TU	F	NA	NM_001141969 c.422delA:p.K141Rfs*3
6:37316781	T/A	TBC1D22B	mis	CSP	F	2.017	NM_017772 c.T1244A:p.M415K
6:43007271	G/A	PPP2R5D	mis	CSP	M	2.150	NM_180977 c.G280A:p.E94K
6:43041011	T/C	CUL7	mis	CSP	M	0.211	NM_001168370 c.A3962G:p.Q1321R
6:43508873	G/A	LRRC73	mis	CSP	F	0.827	NM_001012974 c.C320T:p.P107L
6:44153792	G/A	TMEM63B	mis	TU	F	2.250	NM_001318792 c.G2059A:p.A687T
6:46688676	A/G	TDRD6	mis	CH&SP	M	0.149	NM_001010870 c.A548G:p.E183G
6:56670691	A/G	DST	syn	CSP	M	NA	NM_001144770 c.T651C:p.T217T
6:75950705	T/C	IMPG1	mis	CSP	M	0.016	NM_001563 c.A1681G:p.T561A
6:89747403	C/T	MDN1	mis	CSP	F	0.702	NM_014611 c.G3830A:p.R1277Q
6:105158487	G/A	POPDC3	non	CSP	F	NA	NM_022361 c.C859T:p.R287X
6:112175374	G/A	LAMA4	syn	TU	F	NA	NM_001105207 c.C1275T:p.S425S

6:134997436	G/A	HBS1L	mis	TU	F	0.695	NM_001363686 c.C268T:p.R90W
6:144440430	A/G	UTRN	mis	TU	M	0.121	NM_007124 c.A1471G:p.S491G
6:152455487	C/T	SYNE1	mis	CH&SP	F	0.153	NM_033071 c.G2852A:p.R951Q
6:154792909	C/T	SCAF8	syn	CH&SP	F	NA	NM_001286199 c.C408T:p.A136A
6:160073940	G/A	IGF2R	mis	CSP	F	1.059	NM_000876 c.G5131A:p.A1711T
7:763936	C/T	HEATR2	mis	CSP	M	NA	NM_017802 c.C1745T:p.P582L
7:1057822	G/A	GPR146	mis	CH&SP	M	0.374	NM_001303473 c.G307A:p.V103M
7:1471134	T/G	INTS1	mis	CH&SP	F	NA	NM_001080453 c.A6346C:p.S2116R
7:2263257	C/T	SNX8	syn	TU	M	NA	NM_013321 c.G888A:p.A296A
7:4781584	G/T	AP5Z1	mis	TU	F	NA	NM_014855 c.G196T:p.V66L
7:5528500	C/T	ACTB	mis	CSP	F	2.110	NM_001101 c.G583A:p.E195K
7:13900796	T/C	ETV1	mis	TU	F	0.438	NM_001163150 c.A1034G:p.N345S
7:18954181	C/G	HDAC9	mis	CH&SP	M	0.302	NM_178423 c.C2964G:p.S988R
7:19700129	C/T	TWISTNB	mis	TU	F	0.346	NM_001002926 c.G548A:p.R183H
7:19708855	T/G	TWISTNB	mis	CH&SP	F	0.199	NM_001002926 c.A162C:p.Q54H
7:21687098	G/C	DNAH11	splicing	CSP	F	NA	NM_001277115 c.5622-1G>C:
7:27245124	G/C	EVX1	syn	CH&SP	F	NA	NM_001989 c.G504C:p.G168G
7:44079602	T/C	POLM	mis	CH&SP	F	0.258	NM_001362683 c.A611G:p.H204R
7:48274925	T/C	ABCA13	syn	TU	M	NA	NM_152701 c.T5259C:p.A1753A
7:50368009	C/T	IKZF1	mis	CSP	M	NA	NM_001291846 c.C164T:p.T55I
7:70790816	C/T	AUTS2	syn	CSP	F	NA	NM_015570 c.C3600T:p.T1200T
7:73434777	C/A	FZD9	mis	SMCP	M	NA	NM_003508 c.C770A:p.P257H
7:73449623	G/A	BAZ1B	mis	SMCP	M	1.807	NM_001370402 c.C3647T:p.P1216L
7:99391224	T/C	ARPC1B	mis	CSP	F	0.468	NM_005720 c.T754C:p.F252L
7:100160641	G/A	GAL3ST4	non	CH&SP	M	NA	NM_024637 c.C748T:p.R250X
7:101039779	C/G	MUC17	mis	TU	M	NA	NM_001040105 c.C8363G:p.A2788G
7:102468083	C/T	LRWD1	mis	CH&SP	M	0.065	NM_152892 c.C700T:p.R234W
7:106015195	G/GC	CDHR3	FS ins	TU	F	NA	NM_152750 c.1310dupC:p.Y440Lfs*3
7:121136089	T/C	CPED1	syn	CH&SP	F	NA	NM_024913 c.T1698C:p.D566D
7:127807539	G/A	SND1	mis	TU	F	1.732	NM_014390 c.G1208A:p.R403Q
7:129145056	CTCT/C	TSPAN33	nonFS del	SMCP	M	NA	NM_178562 c.77_79del:p.F29Nfs*255

7:129671515	C/T	NRF1	mis	TU	F	2.801	NM_005011 c.C310T:p.R104W
7:130504957	C/T	MEST	syn	CSP	F	NA	NM_001253901 c.C780T:p.S260S
7:139716723	G/A	HIPK2	syn	CH&SP	F	NA	NM_001113239 c.C312T:p.T104T
7:143286015	G/A	TMEM139	mis	CSP	F	0.327	NM_153345 c.G58A:p.A20T
7:143478771	T/C	TAS2R41	mis	CSP	M	0.397	NM_176883 c.T899C:p.L300S
7:149104123	G/A	ZNF425	mis	CSP	F	0.301	NM_001001661 c.C1748T:p.A583V
8:9011624	G/C	ERI1	mis	CSP	F	0.007	NM_001354635 c.G136C:p.D46H
8:9709968	C/T	TNKS	mis	CH&SP	M	0.389	NM_003747 c.C1592T:p.A531V
8:11331484	C/T	SLC35G5	syn	TU	F	NA	NM_054028 c.C378T:p.P126P
8:20178492	T/C	SLC18A1	mis	CH&SP	F	0.003	NM_001135691 c.A490G:p.I164V
8:26044806	C/T	EBF2	syn	TU	F	NA	NM_022659 c.G54A:p.S18S
8:54459401	C/T	SOX17	syn	CH&SP	F	NA	NM_022454 c.C651T:p.Y217Y
8:54629597	C/T	RP1	syn	TU	F	NA	NM_006269 c.C5715T:p.D1905D
8:63175359	A/G	YTHDF3	syn	SMCP	M	NA	NM_001277813 c.A87G:p.Q29Q
8:66468294	G/A	ADHFE1	mis	CSP	F	0.088	NM_144650 c.G1346A:p.C449Y
8:66505526	CCGGTA/C	C8orf46	FS del	CSP	F	NA	NM_152765 c.279_280del:p.V94Gfs*4
8:71215387	C/T	EYA1	mis	CSP	M	0.901	NM_172058 c.G1597A:p.G533R
8:94250382	G/A	GEM	syn	CH&SP	F	NA	NM_181702 c.C819T:p.I273I
8:100141560	T/A	FBXO43	mis	CSP	M	0.154	NM_001029860 c.A694T:p.T232S
8:109491884	T/C	PKHD1L1	mis	CH&SP	F	0.157	NM_177531 c.T10126C:p.W3376R
8:123231367	C/A	C8orf76	mis	CH&SP	F	0.057	NM_032847 c.G748T:p.A250S
8:123513316	C/T	FBXO32	mis	CH&SP	M	1.032	NM_148177 c.G98A:p.C33Y
8:130076351	T/C	ASAP1	mis	TU	F	0.403	NM_001362924 c.A2707G:p.K903E
8:132104113	G/A	HHLA1	mis	CH&SP	M	NA	NM_001145095 c.C134T:p.T45I
8:143379349	T/C	RHPN1	syn	CH&SP	F	NA	NM_052924 c.T786C:p.H262H
8:143918902	TAGG/T	PLEC	nonFS del	CSP	F	NA	NM_201382 c.10916_10918del:p.S3639Yfs*909
8:144289154	C/T	BOP1	mis	CSP	M	0.846	NM_015201 c.G250A:p.G84R
8:144394929	C/T	CPSF1	mis	CH&SP	M	0.565	NM_013291 c.G3367A:p.A1123T
9:893968	A/G	DMRT1	mis	CH&SP	M	0.530	NM_001363767 c.A121G:p.N41D
9:5968074	G/C	KIAA2026	mis	CSP	F	NA	NM_001017969 c.C2157G:p.I719M
9:6328733	A/C	TPD52L3	syn	CSP	M	NA	NM_001001875 c.A138C:p.L46L

9:6720967	C/T	KDM4C	mis	CSP	F	NA	NM_001146696 c.C19T;p.P7S
9:19360319	C/T	DENND4C	mis	CSP	M	0.272	NM_001330640 c.C5236T;p.L1746F
9:19378429	G/C	RPS6	mis	CSP	M	NA	NM_001010 c.C435G;p.F145L
9:27173294	T/C	TEK	mis	SMCP	F	0.481	NM_000459 c.T833C;p.V278A
9:32440550	G/C	ACO1	mis	CSP	F	0.785	NM_001362840 c.G2333C;p.S778T
9:35092589	C/T	PIGO	mis	TU	F	0.196	NM_001201484 c.G1298A;p.R433Q
9:35105967	C/T	FAM214B	mis	CH&SP	F	0.233	NM_025182 c.G1261A;p.V421M
9:70315564	G/T	SMC5	non	SMCP	M	NA	NM_015110 c.G1792T;p.E598X
9:76710150	G/A	PRUNE2	syn	SMCP	M	NA	NM_001308048 c.C2124T;p.L708L
9:81992892	CAT/C	SPATA31D1	FS del	CSP	F	NA	NM_001001670 c.2423_2424del;p.M809Dfs*7
9:92500989	A/G	ECM2	mis	CH&SP	F	0.128	NM_001197295 c.T1603C;p.S535P
9:99149246	G/GAAAA	TGFBR1	FS ins	CH&SP	M	NA	NM_004612 c.1453_1454insAAAA;p.A485Efs*6
9:100292953	G/A	INVS	mis	CH&SP	M	0.866	NM_001318381 c.G2408A;p.R803Q
9:109141530	G/A	FRRS1L	syn	CSP	F	NA	NM_014334 c.C675T;p.H225H
9:110514096	C/T	SVEP1	syn	CSP	M	NA	NM_153366 c.G975A;p.S325S
9:113591375	C/T	RGS3	mis	TU	F	0.209	NM_001276260 c.C1021T;p.R341W
9:116211737	C/T	PAPPA	non	CH&SP	F	NA	NM_002581 c.C1723T;p.R575X
9:121006939	G/C	C5	mis	CSP	M	0.437	NM_001735 c.C2387G;p.T796S
9:123457396	C/G	DENND1A	mis	CSP	M	1.111	NM_020946 c.G1138C;p.G380R
9:124898627	T/C	GOLGA1	syn	CSP	F	NA	NM_002077 c.A1329G;p.E443E
9:125160898	G/T	PPP6C	mis	TU	F	1.905	NM_001123355 c.C291A;p.D97E
9:127448337	CAG/C	RPL12	FS del	CSP	F	NA	NM_000976 c.377_378del;p.S126Wfs*4
9:127944856	G/A	FAM102A	syn	TU	F	NA	NM_203305 c.C534T;p.D178D
9:131459238	G/A	PRRC2B	mis	CSP	F	0.148	NM_013318 c.G1286A;p.R429Q
9:131630303	C/T	RAPGEF1	mis	TU	F	0.956	NM_198679 c.G676A;p.E226K
9:133734157	C/T	SARDH	mis	TU	F	0.237	NM_001134707 c.G17A;p.R6Q
9:135119959	C/T	OLFM1	syn	CSP	F	NA	NM_014279 c.C1185T;p.Y395Y
9:136463724	G/A	SEC16A	mis	TU	F	0.130	NM_001276418 c.C4463T;p.T1488M
9:136476452	TGA/T	SEC16A	FS del	TU	M	NA	NM_014866 c.1162_1163del;p.S388Ifs*2
9:136672030	C/T	EGFL7	syn	TU	F	NA	NM_016215 c.C741T;p.L247L
9:137220824	G/A	RNF208	mis	CSP	F	0.210	NM_031297 c.C389T;p.S130L

9:137253286	G/A	C9orf173	non	CH&SP	F	NA	NM_001256699 c.G984A:p.W328X
9:137383339	G/A	EXD3	syn	SMCP	F	NA	NM_001286823 c.C94T:p.L32L
10:3165312	G/T	PITRM1	mis	CH&SP	M	0.065	NM_001347730 c.C532A:p.H178N
10:4966995	A/G	AKR1C1	syn	CSP	M	NA	NM_001353 c.A321G:p.Q107Q
10:24473623	A/G	KIAA1217	syn	CSP	M	NA	NM_001098500 c.A1002G:p.P334P
10:26286418	A/G	GAD2	mis	CH&SP	M	1.358	NM_000818 c.A1310G:p.Y437C
10:27040013	T/C	ANKRD26	mis	CSP	M	0.058	NM_014915 c.A2327G:p.Q776R
10:32021109	T/C	KIF5B	mis	TU	M	0.447	NM_004521 c.A2117G:p.Q706R
10:35640302	G/C	FZD8	non	CH&SP	F	NA	NM_031866 c.C1128G:p.Y376X
10:62376272	C/T	ZNF365	mis	CSP	M	0.963	NM_014951 c.C79T:p.R27C
10:63610208	C/T	REEP3	mis	SMCP	M	1.117	NM_001001330 c.C439T:p.R147C
10:69506154	C/T	TSPAN15	mis	CH&SP	F	1.164	NM_001351263 c.C388T:p.R130W
10:72860431	C/T	MCU	mis	CH&SP	F	1.879	NM_001270680 c.C253T:p.R85C
10:73817065	G/A	CAMK2G	mis	TU	F	1.762	NM_001367521 c.C1345T:p.L449F
10:75029026	ACT/A	KAT6B	FS del	TU	F	NA	NM_001370133 c.2514_2515del:p.S839Cfs*5
10:75399279	G/T	ZNF503	mis	CSP	M	0.737	NM_032772 c.C1411A:p.P471T
10:79557336	T/A	SFTPA2	mis	CH&SP	F	2.259	NM_001320814 c.A650T:p.N217I
10:86944127	C/T	MMRN2	syn	CSP	M	NA	NM_024756 c.G657A:p.E219E
10:98209526	C/T	R3HCC1L	mis	CSP	M	0.026	NM_001351011 c.C1412T:p.S471F
10:101023505	C/T	PDZD7	mis	CH&SP	F	0.859	NM_001351044 c.G473A:p.R158H
10:104123028	G/A	SFR1	mis	TU	F	0.017	NM_001002759 c.G77A:p.S26N
10:110577491	CAGTA/C	SMC3	FS del	TU	F	NA	NM_005445 c.270_270del:p.I91Sfs*21
10:132899631	G/A	TTC40	mis	TU	F	NA	NM_001200049 c.C2960T:p.T987I
10:133218895	T/C	KNDC1	mis	CSP	M	1.002	NM_152643 c.T4742C:p.F1581S
10:133230480	C/A	UTF1	mis	CSP	M	1.607	NM_003577 c.C192A:p.S64R
11:247387	C/T	PSMD13	syn	CSP	F	NA	NM_175932 c.C513T:p.H171H
11:407123	G/A	SIGIRR	mis	CH&SP	M	0.998	NM_021805 c.C667T:p.R223C
11:821747	C/T	PNPLA2	mis	CH&SP	M	0.838	NM_020376 c.C307T:p.P103S
11:1450667	G/A	BRSK2	syn	CH&SP	F	NA	NM_001282218 c.G1188A:p.P396P
11:1761354	G/A	CTSD	syn	CSP	M	NA	NM_001909 c.C183T:p.A61A
11:2412824	TG/T	TRPM5	FS del	CH&SP	F	NA	NM_014555 c.2284delC:p.Q762Rfs*35

11:6616729	G/A	TPP1	mis	CH&SP	M	0.750	NM_000391 c.C818T:p.A273V
11:9729377	T/C	SWAP70	mis	CH&SP	M	1.582	NM_015055 c.T824C:p.F275S
11:11379189	C/T	GALNT18	mis	CSP	M	0.416	NM_198516 c.G671A:p.R224H
11:17296190	A/G	NUCB2	syn	TU	F	NA	NM_001352672 c.A231G:p.K77K
11:17522862	C/T	USH1C	mis	CH&SP	F	0.399	NM_153676 c.G941A:p.R314Q
11:19229918	G/T	E2F8	mis	TU	M	0.158	NM_024680 c.C1429A:p.P477T
11:27340722	T/C	CCDC34	mis	SMCP	M	0.538	NM_030771 c.A881G:p.Y294C
11:27499384	T/C	LIN7C	mis	TU	M	1.392	NM_018362 c.A413G:p.D138G
11:33328606	C/CT	HIPK3	FS ins	CSP	F	NA	NM_001048200 c.1195dupT:p.Y399Lfs*8
11:44244271	A/G	EXT2	mis	CSP	F	0.262	NM_000401 c.A2240G:p.N747S
11:47420357	G/A	PSMC3	mis	CSP	F	2.031	NM_002804 c.C1034T:p.S345L
11:48264269	C/T	OR4X1	non	CSP	F	NA	NM_001004726 c.C409T:p.R137X
11:56276029	G/A	OR5T1	mis	CH&SP	M	0.040	NM_001004745 c.G391A:p.A131T
11:61342771	G/A	DAK	mis	CH&SP	F	NA	NM_001351977 c.G792A:p.M264I
11:62839532	G/C	WDR74	syn	TU	F	NA	NM_001369451 c.C39G:p.V13V
11:64753974	G/C	PYGM	mis	CSP	F	0.822	NM_005609 c.C1144G:p.P382A
11:64807581	C/T	MEN1	mis	TU	M	2.185	NM_000244 c.G769A:p.D257N
11:64896605	G/A	ATG2A	mis	CSP	M	0.996	NM_015104 c.C5284T:p.R1762W
11:65850454	A/C	SNX32	mis	CH&SP	M	0.452	NM_152760 c.A398C:p.K133T
11:66422139	C/G	NPAS4	syn	CH&SP	M	NA	NM_178864 c.C195G:p.P65P
11:66558057	C/G	ACTN3	mis	TU	M	NA	NM_001258371 c.C1288G:p.Q430E
11:66849325	C/T	PC	mis	CH&SP	F	0.414	NM_001040716 c.G3193A:p.V1065M
11:66868867	G/A	PC	mis	TU	F	2.060	NM_001040716 c.C1001T:p.T334M
11:82934036	T/A	C11orf82	mis	TU	F	NA	NM_001363481 c.T2698A:p.L900I
11:85725843	G/A	SYTL2	mis	CH&SP	M	0.013	NM_206927 c.C3515T:p.P1172L
11:94179808	G/A	PANX1	mis	CH&SP	F	0.523	NM_015368 c.G752A:p.G251E
11:95789464	G/A	FAM76B	syn	CSP	F	NA	NM_001330357 c.C15T:p.A5A
11:101962762	A/C	KIAA1377	mis	CH&SP	M	NA	NM_001363543 c.A1130C:p.K377T
11:104951972	G/A	CASP4	mis	CSP	F	0.065	NM_033306 c.C128T:p.S43L
11:105100565	T/C	CARD17	syn	CSP	F	NA	NM_001007232 c.A222G:p.T74T
11:119278508	A/G	CBL	splicing	CH&SP	F	NA	NM_005188 c.1228-2A>G:

11:125038506	A/G	CCDC15	syn	SMCP	M	NA	NM_025004 c.A2487G;p.P829P
12:328988	T/C	KDM5A	syn	TU	F	NA	NM_001042603 c.A1815G;p.L605L
12:7096817	C/A	C1RL	mis	TU	M	0.264	NM_001297640 c.G912T;p.Q304H
12:11022374	C/T	TAS2R19	non	TU	M	NA	NM_176888 c.G198A;p.W66X
12:31984091	G/A	KIAA1551	mis	CSP	F	NA	NM_018169 c.G3136A;p.A1046T
12:39341563	TTCC/T	KIF21A	nonFS del	CSP	F	NA	NM_001173463 c.1821_1823del;p.E608Dfs*1030
12:42466271	TTCATGATATA CC/T	PRICKLE1	nonFS del	TU	F	NA	NM_001144882 c.686_697del;p.R229Kfs*600
12:45924796	G/A	SCAF11	mis	TU	F	0.282	NM_004719 c.C3838T;p.P1280S
12:47750427	T/TG	RAPGEF3	splicing	CSP	F	NA	NM_001098531 c.672-2->C:
12:47980029	G/A	COL2A1	non	CH&SP	F	NA	NM_033150 c.C2452T;p.R818X
12:47980553	C/A	COL2A1	splicing	TU	F	NA	NM_001844 c.2625+1G>T:
12:47985732	G/GTAAGTAAG TAGCAGCTCTG CTACTTA	COL2A1	FS ins	SMCP	M	NA	NM_033150 c.1468_1469insTAAGTAGCAGAGCTGCTACT TACTTA;p.A490Vfs*79
12:47989799	G/A	COL2A1	non	CSP	F	NA	NM_033150 c.C823T;p.R275X
12:47997644	CA/C	COL2A1	FS del	CH&SP	F	NA	NM_033150 c.285delT;p.G96Vfs*34
12:49022793	G/A	KMT2D	non	TU	M	NA	NM_003482 c.C16135T;p.Q5379X
12:49646997	G/A	FMNL3	mis	SMCP	F	1.237	NM_001367835 c.C2884T;p.R962W
12:52553070	C/A	KRT71	mis	TU	M	0.529	NM_033448 c.G8T;p.R3L
12:53279845	C/T	ESPL1	syn	TU	F	NA	NM_012291 c.C2478T;p.L826L
12:54054978	G/T	HOXC4	mis	TU	M	1.529	NM_014620 c.G568T;p.A190S
12:55365141	ATTC/A	OR6C75	nonFS del	CSP	M	NA	NM_001005497 c.32_34del;p.L13Gfs*300
12:70538996	G/A	PTPRB	non	CSP	M	NA	NM_001206972 c.C4873T;p.R1625X
12:76346575	C/T	BBS10	syn	CH&SP	F	NA	NM_024685 c.G1410A;p.Q470Q
12:78006521	C/T	NAV3	mis	CSP	F	0.247	NM_014903 c.C983T;p.A328V
12:99246740	C/T	ANKS1B	syn	TU	M	NA	NM_001352187 c.G1881A;p.G627G
12:101195135	A/G	SLC5A8	mis	CH&SP	F	0.254	NM_145913 c.T497C;p.V166A
12:107696535	T/C	PWP1	syn	SMCP	F	NA	NM_001317962 c.T378C;p.P126P
12:108792507	G/A	SSH1	non	CSP	M	NA	NM_001161330 c.C1672T;p.Q558X
12:112043851	T/C	NAA25	mis	CSP	M	1.275	NM_024953 c.A2024G;p.H675R
12:118024952	G/A	RFC5	mis	CH&SP	F	0.487	NM_007370 c.G523A;p.G175S

12:120162021	G/A	GCN1L1	mis	SMCP	F	NA	NM_006836 c.C2201T;p.S734L
12:120356953	G/A	MSI1	non	CSP	F	NA	NM_002442 c.C601T;p.R201X
12:120554742	C/T	RNF10	syn	CSP	M	NA	NM_014868 c.C579T;p.D193D
12:129074831	G/A	TMEM132D	mis	CSP	F	0.890	NM_133448 c.C2344T;p.R782W
12:131845381	C/T	MMP17	mis	TU	F	0.542	NM_016155 c.C1136T;p.P379L
12:131916480	C/T	ULK1	mis	CSP	F	0.253	NM_003565 c.C1961T;p.T654M
12:132679522	C/T	POLE	mis	CSP	F	0.194	NM_006231 c.G553A;p.D185N
12:132680017	G/A	POLE	syn	CSP	F	NA	NM_006231 c.C360T;p.L120L
13:32224379	T/C	FRY	mis	CH&SP	F	1.340	NM_023037 c.T4910C;p.L1637P
13:41226552	G/A	MTRF1	syn	CH&SP	F	NA	NM_004294 c.C1005T;p.C335C
13:48707668	A/G	CYSLTR2	mis	SMCP	F	0.222	NM_001308471 c.A851G;p.H284R
13:110484931	G/A	COL4A2	mis	CSP	M	1.083	NM_001846 c.G2929A;p.G977R
14:21092458	C/A	ZNF219	mis	TU	M	2.085	NM_001101672 c.G839T;p.G280V
14:21368274	C/T	SUPT16H	mis	CSP	F	0.926	NM_007192 c.G950A;p.R317K
14:22766722	C/T	OXA1L	syn	CH&SP	F	NA	NM_005015 c.C21T;p.C7C
14:22843834	C/T	MMP14	syn	CSP	M	NA	NM_004995 c.C975T;p.T325T
14:22845898	G/A	MMP14	syn	TU	F	NA	NM_004995 c.G1608A;p.A536A
14:35750685	G/A	RALGAPA1	mis	CH&SP	F	1.748	NM_001346245 c.C808T;p.P270S
14:50161519	G/A	SOS2	non	TU	F	NA	NM_006939 c.C1159T;p.R387X
14:54775099	C/G	SAMD4A	syn	TU	F	NA	NM_001161577 c.C654G;p.T218T
14:56603914	A/C	TMEM260	mis	CH&SP	M	0.094	NM_017799 c.A444C;p.L148F
14:64542089	ACC/A	HSPA2	FS del	CH&SP	M	NA	NM_021979 c.1241_1242del;p.P415Tfs*42
14:65084256	C/T	MAX	mis	SMCP	F	NA	NM_145114 c.G215A;p.R72H
14:72940333	A/G	DCAF4	mis	CSP	F	0.221	NM_015604 c.A307G;p.S103G
14:74522854	G/A	LTBP2	syn	CH&SP	F	NA	NM_000428 c.C2595T;p.P865P
14:91273275	G/C	CCDC88C	mis	CSP	F	0.603	NM_001080414 c.C5437G;p.L1813V
14:94383150	C/A	SERPINA1	mis	CSP	F	0.088	NM_001002235 c.G88T;p.D30Y
14:94487437	T/G	SERPINA12	mis	CSP	F	0.158	NM_001304461 c.A1111C;p.T371P
14:102034062	G/A	DYNC1H1	mis	CSP	M	2.459	NM_001376 c.G10500A;p.M3500I
14:102036501	C/G	DYNC1H1	mis	CSP	M	2.726	NM_001376 c.C10767G;p.I3589M
14:104943985	A/G	AHNAK2	syn	TU	M	NA	NM_001350929 c.T11166C;p.I3722I

15:37095615	C/T	MEIS2	splicing	CSP	F	NA	NM_001220482 c.388-1G>A:
15:40459551	G/A	BAHD1	mis	SMCP	F	0.297	NM_001301132 c.G1087A:p.D363N
15:42451163	C/T	ZNF106	mis	CH&SP	M	0.183	NM_022473 c.G1040A:p.R347H
15:42682538	C/G	STARD9	mis	CH&SP	M	NA	NM_020759 c.C2500G:p.Q834E
15:44583963	G/A	SPG11	mis	CH&SP	M	0.037	NM_025137 c.C5717T:p.P1906L
15:52218558	G/A	MYO5C	syn	TU	M	NA	NM_018728 c.C3915T:p.N1305N
15:55628995	G/C	PRTG	mis	TU	F	0.428	NM_173814 c.C2633G:p.T878S
15:59174156	G/A	MYO1E	mis	CSP	M	0.535	NM_004998 c.C2134T:p.R712W
15:62797231	C/T	TLN2	syn	CH&SP	F	NA	NM_015059 c.C6063T:p.L2021L
15:68054344	A/G	PIAS1	syn	CSP	F	NA	NM_016166 c.A18G:p.E6E
15:69038955	C/A	NOX5	mis	TU	F	0.878	NM_001184779 c.C1386A:p.H462Q
15:71898665	G/C	MYO9A	mis	CH&SP	F	0.518	NM_006901 c.C3838G:p.L1280V
15:73322674	C/T	HCN4	mis	CSP	F	0.592	NM_005477 c.G3419A:p.R1140K
15:74410801	G/A	SEMA7A	syn	CSP	M	NA	NM_001146029 c.C1782T:p.Y594Y
15:74842364	C/G	ULK3	mis	TU	M	0.661	NM_001099436 c.G159C:p.K53N
15:74850545	A/G	SCAMP2	mis	CSP	M	1.578	NM_005697 c.T601C:p.C201R
15:75349074	C/A	NEIL1	mis	CSP	F	0.706	NM_001256552 c.C427A:p.R143S
15:89648439	T/C	KIF7	mis	CSP	M	0.019	NM_198525 c.A1259G:p.Y420C
16:282519	G/A	ARHGDIG	mis	CH&SP	M	0.066	NM_001176 c.G467A:p.R156Q
16:655335	C/T	WDR90	mis	CH&SP	F	0.401	NM_145294 c.C1585T:p.R529W
16:787141	G/A	RPUSD1	syn	CSP	M	NA	NM_058192 c.C345T:p.S115S
16:1787700	G/C	NUBP2	mis	CH&SP	F	0.585	NM_012225 c.G358C:p.D120H
16:16177563	C/T	ABCC6	mis	CSP	F	0.066	NM_001171 c.G2479A:p.A827T
16:20625517	C/T	ACSM1	mis	TU	M	0.455	NM_001318890 c.G1433A:p.R478H
16:21652568	G/A	IGSF6	mis	CSP	M	0.372	NM_005849 c.C31T:p.R11C
16:21687609	G/A	OTOA	mis	CSP	F	0.214	NM_144672 c.G596A:p.R199Q
16:22165885	C/T	SDR42E2	syn	CSP	F	NA	NM_001365288 c.C303T:p.S101S
16:24571364	G/A	RBBP6	mis	CSP	F	0.174	NM_018703 c.G4196A:p.R1399H
16:29844849	TCAC/T	MVP	nonFS del	CSP	F	NA	NM_001293204 c.1992_1994del:p.T666Nfs*168
16:30027094	C/T	C16orf92	syn	TU	F	NA	NM_001353379 c.C318T:p.V106V
16:31110243	G/A	BCKDK	syn	TU	F	NA	NM_001122957 c.G462A:p.L154L

16:31507812	T/C	C16orf58	mis	CH&SP	F	0.115	NM_022744 c.A367G:p.N123D
16:48141320	C/A	ABCC12	mis	CSP	F	0.158	NM_033226 c.G309T:p.R103S
16:48277407	A/T	LONP2	syn	CSP	M	NA	NM_031490 c.A1311T:p.P437P
16:58516917	G/C	SETD6	mis	CSP	F	0.644	NM_024860 c.G709C:p.E237Q
16:58716112	C/A	GOT2	mis	CHP	M	0.739	NM_001286220 c.G792T:p.Q264H
16:64948024	C/A	CDH11	mis	CH&SP	F	0.833	NM_001330576 c.G1592T:p.R531L
16:66885829	G/A	PDP2	syn	CSP	M	NA	NM_001329932 c.G1545A:p.V515V
16:67646028	C/T	RLTPR	mis	CH&SP	M	NA	NM_001317026 c.C197T:p.T66M
16:67659589	C/T	ACD	mis	CSP	F	0.940	NM_001082486 c.G361A:p.D121N
16:69118976	C/T	DERPC	mis	TU	F	NA	NM_001366604 c.G1453A:p.A485T
16:75715039	T/G	CPHXL	syn	CH&SP	F	NA	NM_001355613 c.A403C:p.R135R
16:84866983	C/T	CRISPLD2	non	CSP	M	NA	NM_031476 c.C796T:p.Q266X
16:88433069	G/GC	ZNF469	FS ins	CSP	F	NA	NM_001367624 c.5600dupC:p.R1869Pfs*22
16:88577142	C/G	ZC3H18	mis	CSP	M	0.018	NM_001294340 c.C19G:p.P7A
16:88712352	G/A	CTU2	mis	CH&SP	F	NA	NM_001012759 c.G422A:p.G141E
16:89550572	T/C	SPG7	mis	CSP	F	0.840	NM_003119 c.T1742C:p.V581A
16:89595520	G/A	CPNE7	mis	CSP	M	0.329	NM_014427 c.G1681A:p.D561N
17:1732382	C/T	WDR81	syn	TU	M	NA	NM_001163673 c.C606T:p.I202I
17:2376800	G/C	SGSM2	mis	TU	F	0.583	NM_001346700 c.G2542C:p.V848L
17:3691758	G/A	P2RX5	syn	CH&SP	M	NA	NM_002561 c.C174T:p.V58V
17:4142681	A/G	ZZEF1	mis	TU	F	1.812	NM_015113 c.T215C:p.L72P
17:4809998	C/T	PLD2	mis	CSP	M	0.367	NM_001243108 c.C829T:p.R277W
17:5536929	G/A	NLRP1	mis	CH&SP	M	0.230	NM_033004 c.C2882T:p.T961I
17:8146123	C/T	PER1	mis	CSP	M	0.132	NM_002616 c.G2053A:p.A685T
17:8290665	C/A	SLC25A35	mis	TU	F	NA	NM_001320877 c.G677T:p.R226L
17:10635734	C/T	MYH3	splicing	CSP	F	NA	NM_002470 c.3975+1G>A:
17:10647414	G/A	MYH3	non	CH&SP	M	NA	NM_002470 c.C748T:p.R250X
17:16552927	A/G	ZNF287	syn	CH&SP	F	NA	NM_001346168 c.T1215C:p.F405F
17:17794568	C/T	RAI1	syn	CH&SP	F	NA	NM_030665 c.C1620T:p.A540A
17:17994338	G/A	LRRC48	mis	TU	F	NA	NM_001130090 c.G631A:p.E211K
17:18119324	C/G	MYO15A	mis	CH&SP	M	NA	NM_016239 c.C524G:p.P175R

17:28734773	C/T	NEK8	syn	TU	F	NA	NM_178170 c.C255T:p.G85G
17:39187483	T/C	CACNB1	mis	SMCP	M	0.764	NM_199247 c.A410G:p.K137R
17:40983872	C/T	KRT40	syn	CSP	F	NA	NM_182497 c.G402A:p.P134P
17:41489540	C/G	KRT36	mis	CH&SP	M	0.694	NM_003771 c.G325C:p.E109Q
17:41912529	T/C	ACLY	mis	TU	F	1.112	NM_198830 c.A173G:p.K58R
17:42037816	G/T	ZNF385C	mis	CHP	M	NA	NM_001242704 c.C83A:p.P28H
17:42105127	G/T	DHX58	mis	CSP	M	0.298	NM_024119 c.C1292A:p.T431N
17:43022193	A/AGGC	VAT1	nonFS ins	TU	M	NA	NM_006373 c.129_130insGCC:p.A43_S44insA
17:43193582	A/G	NBR1	mis	TU	F	0.130	NM_005899 c.A1468G:p.I490V
17:44074746	C/T	G6PC3	mis	TU	F	0.329	NM_138387 c.C392T:p.S131L
17:44197643	A/C	ATXN7L3	mis	CSP	F	1.881	NM_020218 c.T139G:p.Y47D
17:44261587	C/G	SLC4A1	syn	CH&SP	F	NA	NM_000342 c.G156C:p.P52P
17:44859128	GT/G	EFTUD2	FS del	CSP	F	NA	NM_001258353 c.1913delA:p.D638Afs*3
17:47531368	CCCT/C	NPEPPS	nonFS del	CSP	F	NA	NM_006310 c.69_71del:p.L27Vfs*893
17:47847456	C/CG	SP6	FS ins	TU	M	NA	NM_199262 c.973dupC:p.R325Pfs*113
17:48936863	A/G	SNF8	mis	CSP	F	NA	NM_001317194 c.T29C:p.M10T
17:50525022	C/T	MYCBPAP	mis	CH&SP	M	0.151	NM_032133 c.C1781T:p.P594L
17:50596588	G/C	CACNA1G	mis	CSP	F	0.101	NM_001256328 c.G3006C:p.E1002D
17:59679449	G/A	CLTC	mis	CH&SP	F	2.629	NM_004859 c.G2849A:p.R950Q
17:64177356	T/C	TEX2	mis	CH&SP	F	0.302	NM_001288732 c.A2540G:p.K847R
17:64571890	C/T	SMURF2	syn	SMCP	M	NA	NM_022739 c.G924A:p.T308T
17:67738162	G/A	NOL11	mis	TU	M	0.289	NM_015462 c.G1570A:p.E524K
17:75506320	C/T	CASKIN2	syn	CSP	M	NA	NM_001142643 c.G465A:p.V155V
17:75568992	C/T	LLGL2	mis	CH&SP	F	0.845	NM_004524 c.C1337T:p.T446M
17:75876738	G/A	TRIM47	mis	CH&SP	F	0.594	NM_033452 c.C751T:p.R251C
17:80346035	G/A	RNF213	mis	CH&SP	M	1.161	NM_001256071 c.G7700A:p.R2567Q
18:2891553	C/T	EMILIN2	mis	CSP	M	0.475	NM_032048 c.C1426T:p.R476W
18:7032152	C/T	LAMA1	mis	TU	F	0.536	NM_005559 c.G2188A:p.V730M
18:27993601	C/T	CDH2	mis	TU	F	0.998	NM_001792 c.G1057A:p.D353N
18:31092245	G/C	DSC2	mis	CSP	F	0.358	NM_024422 c.C210G:p.D70E
18:35245679	G/T	ZNF397	mis	CSP	M	0.503	NM_001135178 c.G974T:p.R325M

18:51067189	T/TC	SMAD4	splicing	CSP	M	NA	NM_005359 c.1308+2->C:
18:58329068	C/G	NEDD4L	mis	CH&SP	M	0.539	NM_001144967 c.C754G;p.R252G
18:63655690	G/A	SERPINB3	syn	CSP	M	NA	NM_006919 c.C1140T;p.S380S
19:1108242	G/A	SBNO2	mis	CH&SP	M	1.250	NM_001100122 c.C3908T;p.P1303L
19:1828545	G/A	REXO1	mis	SMCP	M	0.377	NM_020695 c.C244T;p.R82C
19:3964319	T/C	DAPK3	mis	SMCP	F	0.823	NM_001348 c.A478G;p.I160V
19:5700922	G/A	LONP1	mis	CSP	M	0.533	NM_001276480 c.C785T;p.T262I
19:6427477	G/A	SLC25A41	mis	CSP	F	0.056	NM_001321298 c.C235T;p.R79C
19:6430108	C/A	SLC25A41	mis	CH&SP	F	0.049	NM_001321298 c.G3T;p.M1I
19:6495747	C/T	TUBB4A	mis	TU	F	1.925	NM_006087 c.G752A;p.R251H
19:7527893	T/G	MCOLN1	mis	TU	F	1.548	NM_020533 c.T710G;p.L237R
19:8961650	G/A	MUC16	syn	CH&SP	M	NA	NM_024690 c.C15120T;p.P5040P
19:12074059	T/C	ZNF844	mis	TU	F	0.050	NM_001136501 c.T32C;p.V11A
19:13968595	C/T	RFX1	mis	CH&SP	F	0.883	NM_002918 c.G1702A;p.A568T
19:14441323	C/T	PKN1	mis	TU	F	1.229	NM_002741 c.C202T;p.R68C
19:15401473	G/T	AKAP8L	mis	CSP	F	0.833	NM_014371 c.C493A;p.R165S
19:15652592	C/T	CYP4F3	syn	CH&SP	F	NA	NM_001199209 c.C942T;p.S314S
19:16157272	C/T	HSH2D	syn	TU	M	NA	NM_032855 c.C537T;p.N179N
19:17611801	G/A	UNC13A	mis	SMCP	M	1.711	NM_001080421 c.C4613T;p.T1538I
19:19544439	G/A	CILP2	mis	CSP	F	1.144	NM_153221 c.G1894A;p.G632S
19:31279075	G/A	TSHZ3	mis	CH&SP	M	1.252	NM_020856 c.C718T;p.R240C
19:32813735	G/GT	TDRD12	FS ins	CH&SP	M	NA	NM_001366102 c.3101dupT;p.I1035Dfs*11
19:35553757	G/A	ATP4A	mis	CH&SP	M	1.072	NM_000704 c.C2554T;p.R852C
19:35944946	C/A	LRFN3	mis	TU	M	0.884	NM_024509 c.C1814A;p.A605E
19:40367793	G/A	PLD3	mis	CSP	M	0.366	NM_001031696 c.G343A;p.G115S
19:41387568	G/A	EXOSC5	syn	CSP	M	NA	NM_020158 c.C561T;p.S187S
19:42754313	G/T	PSG8	syn	CH&SP	F	NA	NM_182707 c.C1263A;p.S421S
19:45641694	G/A	EML2	syn	TU	F	NA	NM_001352053 c.C405T;p.D135D
19:46312216	G/A	HIF3A	mis	CSP	M	0.186	NM_152795 c.G826A;p.E276K
19:47372758	C/T	DHX34	syn	TU	F	NA	NM_014681 c.C1797T;p.F599F
19:47377145	T/A	DHX34	mis	CHP	M	0.956	NM_014681 c.T2645A;p.V882E

19:48836361	G/A	HSD17B14	syn	CSP	M	NA	NM_016246 c.C51T:p.G17G
19:49601639	ACCGCCGCTG/A	PRR12	nonFS del	TU	M	NA	NM_020719 c.4495_4503del:p.L1501Pfs*534
19:50512667	G/GC	ASPDH	FS ins	SMCP	M	NA	NM_001114598 c.425dupG:p.L143Pfs*18
19:51768895	A/G	FPR2	syn	TU	M	NA	NM_001462 c.A237G:p.P79P
19:52766077	CCA/C	ZNF600	FS del	CSP	M	NA	NM_001321866 c.1884_1885del:p.C628Wfs*15
19:53408304	A/G	ZNF765	mis	SMCP	M	0.084	NM_001040185 c.A749G:p.K250R
19:54596442	C/T	LILRA1	syn	CSP	M	NA	NM_006863 c.C1212T:p.N404N
19:54632607	G/C	LILRB1	mis	CH&SP	F	0.070	NM_001081639 c.G805C:p.A269P
19:55404301	C/T	UBE2S	mis	CH&SP	F	1.330	NM_014501 c.G329A:p.R110Q
19:55651322	A/G	CCDC106	mis	CH&SP	M	0.383	NM_001370471 c.A248G:p.E83G
19:56821703	T/A	PEG3	mis	SMCP	M	0.052	NM_001369722 c.A242T:p.E81V
19:57927330	C/T	ZNF418	mis	CSP	F	0.197	NM_133460 c.G851A:p.R284K
19:57943991	G/A	ZNF256	non	CSP	M	NA	NM_005773 c.C103T:p.Q35X
19:58350401	C/T	A1BG	syn	TU	F	NA	NM_130786 c.G1161A:p.A387A
20:1636488	C/T	SIRPG	mis	CH&SP	M	0.023	NM_001039508 c.G448A:p.V150M
20:1915305	C/A	SIRPA	mis	CH&SP	F	0.603	NM_001330728 c.C286A:p.L96I
20:2655641	C/T	NOP56	syn	SMCP	M	NA	NM_006392 c.C804T:p.S268S
20:31722042	C/T	BCL2L1	syn	CSP	F	NA	NM_001317920 c.G177A:p.L59L
20:35003568	G/A	TRPC4AP	non	TU	F	NA	NM_199368 c.C2074T:p.R692X
20:38239694	T/C	KIAA1755	syn	TU	M	NA	NM_001029864 c.A1581G:p.P527P
20:49236465	G/A	DDX27	mis	CSP	M	0.608	NM_001348187 c.G1828A:p.V610I
20:49984288	G/A	SNAI1	mis	CH&SP	M	0.908	NM_005985 c.G547A:p.G183R
20:56386461	C/T	AURKA	mis	CSP	F	0.401	NM_001323303 c.G115A:p.V39I
20:62064933	G/T	TAF4	mis	CSP	F	1.025	NM_003185 c.C878A:p.A293D
20:64048641	T/C	SOX18	mis	TU	F	NA	NM_018419 c.A680G:p.E227G
21:30425845	C/T	KRTAP13-3	mis	CSP	M	0.013	NM_181622 c.G68A:p.G23D
21:32321896	G/A	URB1	mis	SMCP	F	NA	NM_014825 c.C5389T:p.R1797C
21:38299657	C/T	KCNJ15	syn	CSP	F	NA	NM_001276438 c.C396T:p.V132V
21:45267662	G/A	POFUT2	mis	CH&SP	F	1.036	NM_015227 c.C1064T:p.T355M
22:19134260	G/A	DGCR14	mis	CH&SP	M	NA	NM_022719 c.C1367T:p.P456L

22:19898052	C/T	TXNRD2	mis	CH&SP	M	0.881	NM_001282512 c.G761A:p.R254H
22:20086213	C/T	DGCR8	mis	CSP	F	0.656	NM_001190326 c.C250T:p.R84C
22:20994210	G/A	LZTR1	mis	TU	F	0.061	NM_006767 c.G1556A:p.R519Q
22:24060129	C/A	CABIN1	mis	TU	M	0.320	NM_001199281 c.C1605A:p.N535K
22:24513430	C/T	UPB1	mis	TU	F	0.122	NM_016327 c.C566T:p.A189V
22:26499178	C/T	TFIP11	mis	CH&SP	M	0.434	NM_001346859 c.G1255A:p.D419N
22:27799989	G/A	MN1	syn	CSP	M	NA	NM_002430 c.C555T:p.A185A
22:29140361	A/G	KREMEN1	syn	TU	F	NA	NM_001039570 c.A1203G:p.T401T
22:31345318	G/A	PATZ1	syn	TU	F	NA	NM_032050 c.C285T:p.G95G
22:31815166	G/A	DEPDC5	syn	TU	F	NA	NM_001369901 c.G1536A:p.L512L
22:33337693	G/A	LARGE	mis	SMCP	M	NA	NM_001362951 c.C1240T:p.R414W
22:46326445	C/T	GTSE1	syn	CSP	F	NA	NM_016426 c.C1515T:p.V505V
22:46364573	CG/C	CELSR1	FS del	CH&SP	M	NA	NM_014246 c.8717delC:p.P2906Rfs*24
22:49884387	C/T	ZBED4	mis	TU	M	0.373	NM_014838 c.C725T:p.S242L
22:50287681	G/A	PLXNB2	mis	TU	F	0.335	NM_012401 c.C1594T:p.R532W
X:155919881	G/C	VAMP7	splicing	TU	M	NA	NM_005638 c.501+1G>C:

Table S4.4 Expected versus observed DNs in males versus females with respect to prevalence differences					
Variant Class	All CP (Trios=475)		Expected for 196 Males	Expected for 279 Females	Pvalue M v F Difference
	Male (n=196)	Female (n=279)			
Loss of Function	28	40	17.2	24.5	0.975
Missense and Non-frameshifting indels	166	221	124.1	176.4	0.632
Synonymous	52	83	55.2	78.6	0.857
Protein-altering	194	261	141.3	200.8	0.659
Total	246	344	196.5	279.4	0.897
Total Prevalence	0.0005882				
Male:Female Ratio	0.83				
Male Prevalence	0.0005336				
Female Prevalence	0.0006429				
Male Threshold	3.2722				
Female Threshold	3.2191				
Male Count	196				
Female Count	279				

Table S4.5 Global DN enrichment for all categories									
class	observed	expected	enrichment	pValue	group	z	SE	lowbound	upbound
syn	135	133.2	1.01	0.449	all	0.09	0.17	0.84	1.18
mis	387	299.2	1.29	6.57E-07	all	0.07	0.13	1.16	1.42
lof	68	41.5	1.64	0.000102	all	0.20	0.40	1.24	2.04
prot	455	340.7	1.34	2.18E-09	all	0.06	0.13	1.21	1.47
all	590	473.9	1.24	1.52E-07	all	0.05	0.10	1.14	1.34
syn	30	33.6	0.892	0.758	Both	0.16	0.33	0.57	1.22
mis	126	75.6	1.67	7.36E-08	Both	0.15	0.30	1.37	1.97
lof	18	10.5	1.72	0.0217	Both	0.40	0.81	0.91	2.53
prot	144	86.1	1.67	7.61E-09	Both	0.14	0.28	1.39	1.95
all	174	119.7	1.45	1.95E-06	Both	0.11	0.22	1.23	1.67
syn	54	41.8	1.29	0.0391	CSP	0.18	0.35	0.94	1.64
mis	128	93.8	1.36	0.000469	CSP	0.12	0.24	1.12	1.60
lof	29	13	2.23	9.28E-05	CSP	0.41	0.83	1.40	3.06
prot	157	106.9	1.47	3.38E-06	CSP	0.12	0.23	1.24	1.70
all	211	148.7	1.42	8.63E-07	CSP	0.10	0.20	1.22	1.62
syn	10	6.7	1.49	0.143	SMCP	0.47	0.94	0.55	2.43
mis	27	15.1	1.79	0.00366	SMCP	0.34	0.69	1.10	2.48
lof	4	2.1	1.91	0.161	SMCP	0.95	1.90	0.01	3.81
prot	31	17.2	1.8	0.00175	SMCP	0.32	0.65	1.15	2.45
all	41	23.9	1.71	0.000946	SMCP	0.27	0.54	1.17	2.25
syn	52	55.2	0.941	0.687	male	0.13	0.26	0.68	1.20
mis	166	124.1	1.34	0.000193	male	0.10	0.21	1.13	1.55
lof	28	17.2	1.63	0.0104	male	0.31	0.62	1.01	2.25
prot	194	141.3	1.37	1.53E-05	male	0.10	0.20	1.17	1.57
all	246	196.5	1.25	0.000374	male	0.08	0.16	1.09	1.41
syn	83	78.5	1.06	0.321	female	0.12	0.23	0.83	1.29
mis	221	176.4	1.25	0.000664	female	0.08	0.17	1.08	1.42
lof	40	24.5	1.63	0.00244	female	0.26	0.52	1.11	2.15
prot	261	200.8	1.3	2.74E-05	female	0.08	0.16	1.14	1.46
all	344	279.4	1.23	0.000102	female	0.07	0.13	1.10	1.36

syn	6	9.3	0.648	0.899	Syndromic	0.26	0.53	0.12	1.17
mis	34	20.8	1.64	0.00476	Syndromic	0.28	0.56	1.08	2.20
lof	5	2.9	1.73	0.166	Syndromic	0.77	1.54	0.19	3.27
prot	39	23.7	1.65	0.00237	Syndromic	0.26	0.53	1.12	2.18
all	45	32.9	1.37	0.0261	Syndromic	0.20	0.41	0.96	1.78
syn	129	123.4	1.05	0.318	isoalted	0.09	0.18	0.87	1.23
mis	350	277.1	1.26	1.42E-05	Isolated	0.07	0.14	1.12	1.40
lof	63	38.5	1.64	0.000176	Isolated	0.21	0.41	1.23	2.05
prot	413	315.6	1.31	9.23E-08	Isolated	0.06	0.13	1.18	1.44
all	542	439	1.23	1.14E-06	Isolated	0.05	0.11	1.12	1.34
syn	127	126.2	1.01	0.483	No PR	0.09	0.18	0.83	1.19
mis	367	283.4	1.29	1.14E-06	No PR	0.07	0.14	1.15	1.43
lof	67	39.3	1.7	3.75E-05	No PR	0.21	0.42	1.28	2.12
prot	434	322.8	1.34	2.29E-09	No PR	0.06	0.13	1.21	1.47
all	561	449	1.25	1.96E-07	No PR	0.05	0.11	1.14	1.36
syn	8	4.8	1.68	0.11	PR	0.59	1.18	0.50	2.86
mis	17	10.7	1.59	0.0459	PR	0.39	0.77	0.82	2.36
lof	1	1.5	0.673	0.774	PR	0.67	1.33	-0.66	2.01
prot	18	12.2	1.48	0.0707	PR	0.35	0.70	0.78	2.18
all	26	17	1.53	0.0246	PR	0.30	0.60	0.93	2.13

Gene	All		Non-syndromic		Syndromic		PRS Removed		Only PRS	
	Obs PA Vars	PA p-value	Obs PA Vars	PA p-value	Obs PA Vars	PA p-value	Obs PA Vars	PA p-value	Obs PA Vars	PA p-value
ABCC12	1	4.29E-02	1	3.98E-02	-	-	1	4.07E-02	-	-
ABCC6	1	4.35E-02	1	4.04E-02	-	-	1	4.13E-02	-	-
ACADL	1	1.48E-02	1	1.37E-02	-	-	1	1.40E-02	-	-
ACAP3	1	2.48E-02	1	2.30E-02	-	-	1	2.35E-02	-	-
ACD	1	2.18E-02	1	2.02E-02	-	-	1	2.07E-02	-	-
ACLY	1	3.83E-02	1	3.55E-02	-	-	1	3.63E-02	-	-
ACO1	1	2.80E-02	1	2.59E-02	-	-	1	2.65E-02	-	-
ACSM1	1	1.79E-02	-	-	1	1.25E-03	-	-	1	6.45E-04
ACTB	1	1.61E-02	1	1.49E-02	-	-	1	1.52E-02	-	-
ADHFE1	1	1.47E-02	1	1.37E-02	-	-	1	1.40E-02	-	-
AKAP8L	1	2.23E-02	1	2.07E-02	-	-	1	2.12E-02	-	-
ALMS1	1	9.88E-02	1	9.19E-02	-	-	1	9.39E-02	-	-
ANKRD26	1	4.11E-02	1	3.81E-02	-	-	1	3.90E-02	-	-
ANKRD37	1	5.94E-03	1	5.50E-03	-	-	1	5.63E-03	-	-
ANXA10	1	1.14E-02	1	1.06E-02	-	-	1	1.08E-02	-	-
AP5Z1	1	2.48E-02	1	2.30E-02	-	-	1	2.35E-02	-	-
ARHGDIG	1	6.87E-03	1	6.37E-03	-	-	1	6.51E-03	-	-
ARID1A	2	2.14E-03	1	6.01E-02	1	4.63E-03	2	1.92E-03	-	-
ARPC1B	1	1.47E-02	1	1.36E-02	-	-	1	1.39E-02	-	-
ASAP1	1	3.61E-02	1	3.35E-02	-	-	-	-	1	1.31E-03
ASPDH	1	7.42E-03	1	6.87E-03	-	-	1	7.03E-03	-	-
ATG2A	1	6.72E-02	1	6.24E-02	-	-	1	6.38E-02	-	-
ATP13A4	1	4.01E-02	1	3.72E-02	-	-	1	3.80E-02	-	-
ATP4A	1	3.95E-02	1	3.66E-02	-	-	1	3.74E-02	-	-
ATP6V1A	1	2.02E-02	1	1.88E-02	-	-	1	1.92E-02	-	-
ATXN1	1	3.09E-02	1	2.86E-02	-	-	1	2.93E-02	-	-
ATXN7L3	1	1.30E-02	1	1.20E-02	-	-	1	1.23E-02	-	-
AURKA	1	1.09E-02	1	1.01E-02	-	-	1	1.04E-02	-	-
BAHD1	1	3.02E-02	1	2.81E-02	-	-	1	2.87E-02	-	-

BAZ1B	1	4.38E-02	1	4.06E-02	-	-	1	4.15E-02	-	-
BDP1	1	6.48E-02	1	6.02E-02	-	-	1	6.15E-02	-	-
BIRC6	1	1.11E-01	1	1.03E-01	-	-	1	1.06E-01	-	-
BOP1	1	8.84E-03	1	8.19E-03	-	-	1	8.38E-03	-	-
BRPF1	1	4.48E-02	1	4.16E-02	-	-	1	4.25E-02	-	-
BSN	1	1.27E-01	1	1.18E-01	-	-	1	1.20E-01	-	-
C11ORF82	1	2.17E-02	1	2.02E-02	-	-	1	2.06E-02	-	-
C16ORF58	1	1.46E-02	1	1.36E-02	-	-	1	1.39E-02	-	-
C1ORF170	1	2.42E-03	1	2.25E-03	-	-	1	2.30E-03	-	-
C1RL	1	1.52E-02	-	-	-	-	-	-	-	-
C2ORF16	1	5.14E-02	1	4.77E-02	-	-	1	4.87E-02	-	-
C2ORF69	1	7.95E-03	1	7.37E-03	-	-	1	7.53E-03	-	-
C4ORF21	1	4.80E-02	-	-	1	3.41E-03	-	-	1	1.76E-03
C5	1	4.83E-02	1	4.48E-02	-	-	1	4.58E-02	-	-
C8B	1	1.81E-02	1	1.67E-02	-	-	1	1.71E-02	-	-
C8ORF46	1	7.00E-03	1	6.49E-03	-	-	1	6.64E-03	-	-
C8ORF76	1	1.06E-02	1	9.84E-03	-	-	1	1.01E-02	-	-
C9ORF173	1	1.03E-02	1	9.54E-03	-	-	1	9.75E-03	-	-
CABIN1	1	7.00E-02	1	6.50E-02	-	-	1	6.64E-02	-	-
CACNA1D	1	7.47E-02	1	6.94E-02	-	-	1	7.09E-02	-	-
CACNA1G	1	7.78E-02	1	7.23E-02	-	-	1	7.39E-02	-	-
CACNB1	1	2.23E-02	-	-	1	1.56E-03	1	2.11E-02	-	-
CAMK2G	1	2.04E-02	1	1.89E-02	-	-	1	1.93E-02	-	-
CASP10	1	1.59E-02	1	1.48E-02	-	-	1	1.51E-02	-	-
CASP4	1	1.19E-02	1	1.10E-02	-	-	1	1.12E-02	-	-
CBL	1	2.92E-02	-	-	1	2.06E-03	1	2.77E-02	-	-
CCDC106	1	1.25E-02	1	1.16E-02	-	-	1	1.18E-02	-	-
CCDC34	1	1.14E-02	-	-	1	7.94E-04	1	1.08E-02	-	-
CCDC88C	1	6.48E-02	1	6.02E-02	-	-	1	6.15E-02	-	-
CCNL1	1	1.86E-02	1	1.72E-02	-	-	1	1.76E-02	-	-
CDH11	1	2.66E-02	1	2.47E-02	-	-	1	2.52E-02	-	-
CDH2	1	2.60E-02	1	2.41E-02	-	-	1	2.46E-02	-	-
CDHR3	1	2.46E-02	1	2.28E-02	-	-	1	2.34E-02	-	-
CDV3	1	6.76E-03	1	6.26E-03	-	-	1	6.40E-03	-	-

CELSR1	1	1.26E-01	1	1.17E-01	-	-	1	1.20E-01	-	-
CHIT1	1	1.73E-02	1	1.60E-02	-	-	1	1.64E-02	-	-
CILP2	1	5.03E-02	1	4.66E-02	-	-	1	4.77E-02	-	-
CLTC	1	4.87E-02	1	4.52E-02	-	-	1	4.62E-02	-	-
COL11A1	1	5.61E-02	1	5.21E-02	-	-	1	5.33E-02	-	-
COL2A1	5	2.39E-09	5	1.63E-09	-	-	5	1.83E-09	-	-
COL4A2	1	5.94E-02	1	5.52E-02	-	-	1	5.64E-02	-	-
CPNE7	1	2.22E-02	1	2.06E-02	-	-	1	2.10E-02	-	-
CPSF1	1	5.77E-02	1	5.36E-02	-	-	1	5.48E-02	-	-
CRISPLD2	1	1.94E-02	1	1.80E-02	-	-	1	1.84E-02	-	-
CSNK2B	1	7.24E-03	1	6.71E-03	-	-	1	6.86E-03	-	-
CTU2	1	1.97E-02	1	1.82E-02	-	-	1	1.87E-02	-	-
CUL7	1	5.73E-02	1	5.32E-02	-	-	1	5.44E-02	-	-
CYSLTR2	1	8.87E-03	1	8.22E-03	-	-	1	8.41E-03	-	-
DAK	1	2.20E-02	1	2.04E-02	-	-	1	2.09E-02	-	-
DAPK3	1	1.94E-02	1	1.80E-02	-	-	1	1.84E-02	-	-
DAXX	1	2.44E-02	-	-	1	1.72E-03	1	2.31E-02	-	-
DBN1	1	2.52E-02	1	2.33E-02	-	-	1	2.39E-02	-	-
DCAF4	1	1.95E-02	1	1.80E-02	-	-	1	1.84E-02	-	-
DDX27	1	2.85E-02	1	2.64E-02	-	-	1	2.70E-02	-	-
DENND1A	1	3.51E-02	1	3.25E-02	-	-	-	-	1	1.28E-03
DENND4C	1	4.31E-02	1	4.00E-02	-	-	1	4.08E-02	-	-
DGCR14	1	1.75E-02	1	1.62E-02	-	-	1	1.65E-02	-	-
DGCR8	1	2.82E-02	1	2.62E-02	-	-	1	2.68E-02	-	-
DHX34	1	4.44E-02	1	4.12E-02	-	-	1	4.22E-02	-	-
DHX58	1	2.42E-02	1	2.25E-02	-	-	1	2.30E-02	-	-
DIRC2	1	1.52E-02	1	1.41E-02	-	-	1	1.44E-02	-	-
DMRT1	1	1.20E-02	1	1.11E-02	-	-	1	1.14E-02	-	-
DNAH11	1	1.10E-01	1	1.02E-01	-	-	1	1.04E-01	-	-
DNAH5	1	1.28E-01	1	1.19E-01	-	-	1	1.22E-01	-	-
DNAJC10	1	2.52E-02	1	2.34E-02	-	-	1	2.39E-02	-	-
DSC2	1	2.37E-02	1	2.20E-02	-	-	1	2.25E-02	-	-
DYNC1H1	2	1.30E-02	2	1.12E-02	-	-	2	1.17E-02	-	-
DYNC1I2	1	1.82E-02	1	1.69E-02	-	-	1	1.73E-02	-	-

E2F8	1	2.48E-02	1	2.30E-02	-	-	1	2.35E-02	-	-
ECM2	1	1.89E-02	1	1.76E-02	-	-	1	1.79E-02	-	-
EFTUD2	1	3.11E-02	1	2.88E-02	-	-	1	2.94E-02	-	-
EHMT2	1	4.52E-02	1	4.20E-02	-	-	1	4.29E-02	-	-
EIF4A2	1	1.45E-02	1	1.34E-02	-	-	1	1.37E-02	-	-
EMILIN2	1	2.84E-02	1	2.64E-02	-	-	1	2.70E-02	-	-
EPB41L5	1	2.43E-02	1	2.25E-02	-	-	1	2.31E-02	-	-
EPHB3	1	4.26E-02	1	3.95E-02	-	-	1	4.04E-02	-	-
ERI1	1	8.38E-03	1	7.77E-03	-	-	1	7.95E-03	-	-
ETV1	1	1.32E-02	1	1.23E-02	-	-	1	1.25E-02	-	-
EXT2	1	2.44E-02	1	2.26E-02	-	-	1	2.31E-02	-	-
EYA1	1	1.88E-02	1	1.74E-02	-	-	1	1.78E-02	-	-
FAM114A2	1	1.49E-02	1	1.38E-02	-	-	1	1.41E-02	-	-
FAM131A	1	1.23E-02	1	1.14E-02	-	-	1	1.16E-02	-	-
FAM214B	1	1.68E-02	1	1.56E-02	-	-	1	1.59E-02	-	-
FBXO32	1	1.26E-02	1	1.17E-02	-	-	1	1.20E-02	-	-
FBXO43	1	1.73E-02	1	1.61E-02	-	-	1	1.64E-02	-	-
FLNB	1	8.94E-02	1	8.31E-02	-	-	1	8.49E-02	-	-
FMNL3	1	3.46E-02	1	3.21E-02	-	-	1	3.28E-02	-	-
FMOD	1	1.30E-02	1	1.20E-02	-	-	1	1.23E-02	-	-
FRY	1	8.74E-02	-	-	1	6.33E-03	1	8.30E-02	-	-
FZD8	1	3.30E-02	1	3.06E-02	-	-	1	3.13E-02	-	-
FZD9	1	2.88E-02	1	2.67E-02	-	-	1	2.73E-02	-	-
G6PC3	1	1.14E-02	1	1.06E-02	-	-	1	1.08E-02	-	-
GAD2	1	1.91E-02	-	-	1	1.34E-03	1	1.81E-02	-	-
GAL3ST4	1	1.68E-02	1	1.55E-02	-	-	1	1.59E-02	-	-
GALNT18	1	2.17E-02	1	2.01E-02	-	-	-	-	1	7.84E-04
GCKR	1	1.98E-02	1	1.83E-02	-	-	-	-	1	7.14E-04
GCN1L1	1	9.17E-02	1	8.52E-02	-	-	1	8.71E-02	-	-
GIGYF2	1	4.63E-02	-	-	1	3.29E-03	1	4.40E-02	-	-
GOT2	1	1.51E-02	1	1.40E-02	-	-	1	1.43E-02	-	-
GPR125	1	4.14E-02	1	3.84E-02	-	-	-	-	1	1.51E-03
GPR146	1	1.30E-02	1	1.21E-02	-	-	1	1.23E-02	-	-
GRHL3	2	1.74E-04	2	1.49E-04	-	-	2	1.56E-04	-	-

HBS1L	1	2.20E-02	1	2.04E-02	-	-	1	2.09E-02	-	-
HCN4	1	4.54E-02	1	4.22E-02	-	-	1	4.31E-02	-	-
HDAC9	1	2.66E-02	-	-	1	1.87E-03	1	2.52E-02	-	-
HEATR2	1	2.65E-02	1	2.46E-02	-	-	1	2.52E-02	-	-
HEG1	1	3.54E-02	1	3.28E-02	-	-	1	3.35E-02	-	-
HHLA1	1	4.11E-03	1	3.81E-03	-	-	1	3.89E-03	-	-
HIF3A	1	2.35E-02	1	2.18E-02	-	-	1	2.23E-02	-	-
HIPK3	1	3.16E-02	1	2.93E-02	-	-	1	2.99E-02	-	-
HIST1H2AK	1	6.47E-03	1	6.00E-03	-	-	1	6.13E-03	-	-
HIVEP1	1	7.14E-02	1	6.63E-02	-	-	1	6.78E-02	-	-
HK2	1	3.37E-02	1	3.12E-02	-	-	-	-	1	1.23E-03
HOXC4	1	1.10E-02	1	1.02E-02	-	-	1	1.04E-02	-	-
HPSE	1	1.53E-02	1	1.42E-02	-	-	1	1.45E-02	-	-
HSPA2	1	3.01E-02	1	2.79E-02	-	-	1	2.85E-02	-	-
IFT122	1	4.57E-02	1	4.24E-02	-	-	1	4.34E-02	-	-
IGF2R	1	8.58E-02	-	-	1	6.22E-03	1	8.15E-02	-	-
IGSF6	1	7.40E-03	1	6.86E-03	-	-	1	7.01E-03	-	-
IKZF1	1	1.81E-02	1	1.68E-02	-	-	1	1.72E-02	-	-
IMPG1	1	2.26E-02	1	2.10E-02	-	-	1	2.14E-02	-	-
INADL	1	5.20E-02	1	4.83E-02	-	-	1	4.93E-02	-	-
INPP4A	1	3.26E-02	1	3.02E-02	-	-	1	3.09E-02	-	-
INTS1	1	6.95E-02	-	-	1	4.99E-03	1	6.60E-02	-	-
INVS	1	3.10E-02	1	2.87E-02	-	-	1	2.94E-02	-	-
IRF6	3	6.19E-07	3	4.93E-07	-	-	2	1.07E-04	1	5.56E-04
IWS1	1	2.51E-02	-	-	1	1.76E-03	1	2.38E-02	-	-
JADE1	1	2.66E-02	1	2.46E-02	-	-	1	2.52E-02	-	-
KAT6B	1	5.93E-02	-	-	1	4.24E-03	1	5.62E-02	-	-
KAZN	1	2.69E-02	1	2.50E-02	-	-	1	2.55E-02	-	-
KDM4C	1	3.01E-02	1	2.79E-02	-	-	1	2.85E-02	-	-
KIAA1143	1	5.11E-03	1	4.73E-03	-	-	1	4.84E-03	-	-
KIAA1377	1	2.84E-02	1	2.63E-02	-	-	1	2.69E-02	-	-
KIAA1551	1	4.05E-02	1	3.76E-02	-	-	1	3.84E-02	-	-
KIAA2026	1	5.18E-02	1	4.81E-02	-	-	1	4.91E-02	-	-
KIF21A	1	4.44E-02	1	4.12E-02	-	-	1	4.21E-02	-	-

KIF5B	1	2.76E-02	1	2.56E-02	-	-	1	2.62E-02	-	-
KIF7	1	3.46E-02	1	3.21E-02	-	-	1	3.28E-02	-	-
KMT2D	1	1.53E-01	1	1.43E-01	-	-	1	1.46E-01	-	-
KNDC1	1	6.17E-02	1	5.73E-02	-	-	1	5.86E-02	-	-
KRT36	1	1.68E-02	1	1.56E-02	-	-	1	1.59E-02	-	-
KRT71	1	1.89E-02	1	1.75E-02	-	-	1	1.79E-02	-	-
KRTAP13-3	1	3.98E-03	1	3.69E-03	-	-	1	3.77E-03	-	-
LAMA1	1	9.11E-02	1	8.47E-02	-	-	1	8.65E-02	-	-
LARGE	1	2.92E-02	1	2.71E-02	-	-	1	2.77E-02	-	-
LILRB1	1	2.22E-02	1	2.06E-02	-	-	1	2.10E-02	-	-
LIN7C	1	6.33E-03	1	5.86E-03	-	-	1	5.99E-03	-	-
LLGL2	1	4.07E-02	1	3.78E-02	-	-	1	3.86E-02	-	-
LMLN	1	2.56E-02	1	2.37E-02	-	-	1	2.43E-02	-	-
LONP1	1	4.18E-02	1	3.88E-02	-	-	1	3.96E-02	-	-
LPP	1	2.10E-02	1	1.95E-02	-	-	1	1.99E-02	-	-
LRFN3	1	2.90E-02	1	2.69E-02	-	-	1	2.75E-02	-	-
LRRC48	1	1.66E-02	1	1.54E-02	-	-	1	1.57E-02	-	-
LRRC73	1	1.07E-02	1	9.96E-03	-	-	1	1.02E-02	-	-
LRWD1	1	2.24E-02	1	2.08E-02	-	-	1	2.13E-02	-	-
LZTR1	1	3.19E-02	1	2.95E-02	-	-	1	3.02E-02	-	-
MAB21L2	1	1.69E-02	1	1.57E-02	-	-	1	1.60E-02	-	-
MACF1	1	1.49E-01	1	1.39E-01	-	-	1	1.42E-01	-	-
MAEL	1	1.46E-02	1	1.36E-02	-	-	1	1.39E-02	-	-
7-Mar	1	2.00E-02	1	1.85E-02	-	-	1	1.89E-02	-	-
MAX	1	6.50E-03	1	6.02E-03	-	-	1	6.16E-03	-	-
MCOLN1	1	2.38E-02	1	2.21E-02	-	-	1	2.25E-02	-	-
MCOLN2	1	1.72E-02	1	1.59E-02	-	-	1	1.63E-02	-	-
MCU	1	1.12E-02	1	1.04E-02	-	-	1	1.06E-02	-	-
MDN1	1	1.57E-01	1	1.46E-01	-	-	1	1.49E-01	-	-
MECOM	1	3.30E-02	1	3.07E-02	-	-	1	3.13E-02	-	-
MEIS2	1	1.76E-02	1	1.63E-02	-	-	1	1.67E-02	-	-
MEN1	1	2.45E-02	1	2.27E-02	-	-	1	2.33E-02	-	-
MGAT5	1	2.39E-02	1	2.22E-02	-	-	1	2.27E-02	-	-
MMP17	1	2.46E-02	1	2.28E-02	-	-	1	2.33E-02	-	-

MOB1B	1	6.11E-03	1	5.66E-03	-	-	1	5.79E-03	-	-
MSI1	1	1.45E-02	1	1.34E-02	-	-	1	1.37E-02	-	-
MUC17	1	1.11E-01	1	1.03E-01	-	-	1	1.05E-01	-	-
MVP	1	3.20E-02	1	2.97E-02	-	-	1	3.03E-02	-	-
MYCBPAP	1	3.11E-02	1	2.88E-02	-	-	1	2.95E-02	-	-
MYH3	2	1.83E-03	2	1.58E-03	-	-	2	1.65E-03	-	-
MYO15A	1	1.27E-01	-	-	1	9.36E-03	1	1.20E-01	-	-
MYO1E	1	3.63E-02	1	3.37E-02	-	-	1	3.45E-02	-	-
MYO9A	1	7.18E-02	1	6.66E-02	-	-	1	6.81E-02	-	-
NAA25	1	2.92E-02	1	2.71E-02	-	-	1	2.77E-02	-	-
NAV3	1	6.59E-02	1	6.12E-02	-	-	1	6.25E-02	-	-
NBEAL2	1	8.64E-02	1	8.03E-02	-	-	1	8.21E-02	-	-
NBR1	1	2.16E-02	-	-	-	-	-	-	-	-
NEDD4L	1	2.91E-02	1	2.70E-02	-	-	1	2.76E-02	-	-
NEIL1	1	1.61E-02	1	1.49E-02	-	-	1	1.52E-02	-	-
NIPBL	1	7.26E-02	1	6.74E-02	-	-	1	6.89E-02	-	-
NLRP1	1	4.01E-02	1	3.72E-02	-	-	1	3.81E-02	-	-
NOL11	1	2.01E-02	1	1.86E-02	-	-	1	1.90E-02	-	-
NOX5	1	2.62E-02	-	-	1	1.84E-03	1	2.48E-02	-	-
NPEPPS	1	2.27E-02	1	2.10E-02	-	-	1	2.15E-02	-	-
NRF1	1	1.86E-02	1	1.73E-02	-	-	1	1.77E-02	-	-
NUBP2	1	1.05E-02	1	9.76E-03	-	-	1	9.98E-03	-	-
OBSCN	2	4.41E-02	2	3.84E-02	-	-	2	4.00E-02	-	-
OR4X1	1	7.84E-03	1	7.26E-03	-	-	1	7.43E-03	-	-
OR5K2	1	7.25E-03	1	6.71E-03	-	-	1	6.87E-03	-	-
OR5T1	1	7.58E-03	1	7.02E-03	-	-	1	7.18E-03	-	-
OR6C75	1	7.23E-03	1	6.70E-03	-	-	1	6.85E-03	-	-
OSGEPL1	1	1.08E-02	1	1.00E-02	-	-	1	1.02E-02	-	-
OTOA	1	2.62E-02	1	2.43E-02	-	-	1	2.48E-02	-	-
PADI1	1	2.07E-02	1	1.92E-02	-	-	1	1.97E-02	-	-
PANX1	1	1.35E-02	1	1.25E-02	-	-	1	1.28E-02	-	-
PAPPA	1	5.10E-02	1	4.73E-02	-	-	1	4.84E-02	-	-
PC	2	1.10E-03	2	9.44E-04	-	-	2	9.87E-04	-	-
PCDHA1	1	3.92E-02	1	3.63E-02	-	-	1	3.72E-02	-	-

PCDHA4	1	3.89E-02	1	3.61E-02	-	-	1	3.69E-02	-	-
PCDHAC1	1	3.34E-02	1	3.10E-02	-	-	1	3.17E-02	-	-
PDZD7	1	2.14E-02	1	1.99E-02	-	-	1	2.03E-02	-	-
PEG3	1	4.62E-02	1	4.29E-02	-	-	1	4.38E-02	-	-
PER1	1	4.28E-02	1	3.98E-02	-	-	1	4.06E-02	-	-
PGD	1	1.56E-02	-	-	1	1.09E-03	1	1.48E-02	-	-
PGM2	1	1.66E-02	1	1.54E-02	-	-	1	1.58E-02	-	-
PIGO	1	3.17E-02	1	2.94E-02	-	-	1	3.00E-02	-	-
PITRM1	1	2.94E-02	1	2.72E-02	-	-	1	2.78E-02	-	-
PKHD1L1	1	1.01E-01	1	9.41E-02	-	-	1	9.62E-02	-	-
PKN1	1	3.43E-02	1	3.18E-02	-	-	1	3.25E-02	-	-
PLD2	1	3.44E-02	1	3.19E-02	-	-	1	3.27E-02	-	-
PLD3	1	1.77E-02	1	1.64E-02	-	-	1	1.68E-02	-	-
PLEC	1	1.89E-01	1	1.76E-01	-	-	1	1.80E-01	-	-
PLEKHM2	1	2.68E-02	1	2.49E-02	-	-	1	2.54E-02	-	-
PLXNA2	1	7.33E-02	1	6.81E-02	-	-	1	6.96E-02	-	-
PLXNB2	1	7.28E-02	-	-	1	5.24E-03	1	6.91E-02	-	-
PNPLA2	1	1.62E-02	1	1.50E-02	-	-	1	1.54E-02	-	-
POFUT2	1	1.63E-02	-	-	1	1.14E-03	1	1.55E-02	-	-
POLE	1	8.79E-02	1	8.17E-02	-	-	1	8.35E-02	-	-
POLM	1	1.73E-02	1	1.61E-02	-	-	1	1.64E-02	-	-
POPDC3	1	8.23E-03	1	7.63E-03	-	-	1	7.80E-03	-	-
PPM1J	1	1.79E-02	1	1.66E-02	-	-	1	1.70E-02	-	-
PPP1R21	1	2.16E-02	1	2.01E-02	-	-	1	2.05E-02	-	-
PPP2R5D	1	2.06E-02	1	1.91E-02	-	-	-	-	-	-
PPP6C	1	1.19E-02	1	1.11E-02	-	-	1	1.13E-02	-	-
PRICKLE1	1	2.51E-02	1	2.33E-02	-	-	1	2.38E-02	-	-
PRKCI	3	1.71E-06	1	2.00E-02	2	1.15E-06	3	1.45E-06	-	-
PRR12	1	6.14E-02	1	5.70E-02	-	-	1	5.83E-02	-	-
PRRC2B	1	7.46E-02	-	-	1	5.37E-03	1	7.08E-02	-	-
PRRC2C	1	7.05E-02	1	6.55E-02	-	-	1	6.70E-02	-	-
PRTG	1	3.27E-02	1	3.03E-02	-	-	1	3.10E-02	-	-
PSMC3	1	1.43E-02	1	1.33E-02	-	-	1	1.36E-02	-	-
PTCHD2	1	5.35E-02	1	4.97E-02	-	-	1	5.08E-02	-	-

PTPRB	1	6.03E-02	1	5.60E-02	-	-	1	5.72E-02	-	-
PTPRF	1	7.53E-02	1	7.00E-02	-	-	1	7.15E-02	-	-
PTPRU	1	5.76E-02	1	5.35E-02	-	-	1	5.47E-02	-	-
PYGM	1	3.41E-02	1	3.16E-02	-	-	1	3.23E-02	-	-
R3HCC1L	1	2.05E-02	1	1.90E-02	-	-	1	1.94E-02	-	-
RALGAPA1	1	5.19E-02	1	4.82E-02	-	-	-	-	1	1.91E-03
RANBP17	1	3.48E-02	1	3.23E-02	-	-	1	3.30E-02	-	-
RAPGEF1	1	3.85E-02	1	3.57E-02	-	-	1	3.65E-02	-	-
RAPGEF3	1	3.07E-02	1	2.85E-02	-	-	1	2.91E-02	-	-
RBBP6	1	4.88E-02	1	4.53E-02	-	-	1	4.63E-02	-	-
RBM47	1	2.94E-02	-	-	1	2.07E-03	1	2.79E-02	-	-
RBMS1	1	1.44E-02	1	1.33E-02	-	-	1	1.36E-02	-	-
REEP3	1	6.46E-03	1	5.99E-03	-	-	1	6.13E-03	-	-
REXO1	1	4.09E-02	1	3.80E-02	-	-	1	3.88E-02	-	-
RFC5	1	1.18E-02	1	1.10E-02	-	-	1	1.12E-02	-	-
RFX1	1	3.34E-02	-	-	1	2.35E-03	-	-	1	1.21E-03
RGS3	1	4.21E-02	1	3.90E-02	-	-	1	3.99E-02	-	-
RLF	1	4.74E-02	1	4.40E-02	-	-	1	4.49E-02	-	-
RLTPR	1	4.69E-02	1	4.35E-02	-	-	-	-	1	1.72E-03
RMDN2	1	1.52E-02	1	1.41E-02	-	-	1	1.44E-02	-	-
RNF208	1	8.61E-03	1	7.97E-03	-	-	1	8.15E-03	-	-
RNF213	1	1.45E-01	1	1.35E-01	-	-	1	1.38E-01	-	-
RNF39	1	1.57E-02	1	1.46E-02	-	-	1	1.49E-02	-	-
RPL12	1	6.20E-03	1	5.75E-03	-	-	1	5.88E-03	-	-
RPL5	1	1.03E-02	-	-	1	7.17E-04	1	9.73E-03	-	-
RPS6	1	9.37E-03	1	8.69E-03	-	-	1	8.88E-03	-	-
SARDH	1	3.67E-02	1	3.41E-02	-	-	1	3.48E-02	-	-
SATB2	3	2.81E-06	3	2.24E-06	-	-	3	2.39E-06	-	-
SBNO2	1	3.79E-02	1	3.52E-02	-	-	1	3.60E-02	-	-
SCAF11	1	3.63E-02	-	-	1	2.56E-03	1	3.44E-02	-	-
SCAMP2	1	1.01E-02	1	9.35E-03	-	-	1	9.56E-03	-	-
SEC16A	2	2.76E-03	2	2.38E-03	-	-	2	2.49E-03	-	-
SERPINA1	1	1.21E-02	1	1.12E-02	-	-	1	1.15E-02	-	-
SERPINA12	1	1.24E-02	1	1.15E-02	-	-	1	1.17E-02	-	-

SETD6	1	1.55E-02	1	1.43E-02	-	-	1	1.47E-02	-	-
SF3B4	1	1.29E-02	1	1.20E-02	-	-	1	1.22E-02	-	-
SFR1	1	6.24E-03	1	5.78E-03	-	-	1	5.91E-03	-	-
SFTPA2	1	8.00E-03	-	-	1	5.58E-04	-	-	1	2.87E-04
SGSM2	1	3.73E-02	1	3.46E-02	-	-	1	3.54E-02	-	-
SIGIRR	1	1.46E-02	-	-	1	1.02E-03	1	1.38E-02	-	-
SIKE1	1	7.58E-03	1	7.02E-03	-	-	1	7.18E-03	-	-
SIRPA	1	1.70E-02	1	1.58E-02	-	-	1	1.61E-02	-	-
SIRPG	1	1.13E-02	1	1.04E-02	-	-	1	1.07E-02	-	-
SLC18A1	1	1.47E-02	1	1.37E-02	-	-	1	1.40E-02	-	-
SLC25A35	1	9.12E-03	1	8.45E-03	-	-	1	8.64E-03	-	-
SLC25A41	2	7.57E-05	2	6.50E-05	-	-	2	6.79E-05	-	-
SLC26A6	1	2.47E-02	1	2.29E-02	-	-	1	2.34E-02	-	-
SLC2A9	1	1.59E-02	1	1.47E-02	-	-	1	1.51E-02	-	-
SLC5A8	1	1.86E-02	1	1.73E-02	-	-	1	1.77E-02	-	-
SLC6A19	1	2.42E-02	1	2.25E-02	-	-	1	2.30E-02	-	-
SLC7A14	1	2.53E-02	1	2.35E-02	-	-	1	2.40E-02	-	-
SMAD4	1	1.65E-02	1	1.53E-02	-	-	1	1.56E-02	-	-
SMC3	1	3.65E-02	1	3.39E-02	-	-	1	3.46E-02	-	-
SMC5	1	3.21E-02	1	2.98E-02	-	-	1	3.05E-02	-	-
SMC6	1	2.87E-02	1	2.66E-02	-	-	1	2.72E-02	-	-
SNAI1	1	9.33E-03	1	8.65E-03	-	-	1	8.85E-03	-	-
SND1	1	3.51E-02	1	3.25E-02	-	-	1	3.32E-02	-	-
SNF8	1	9.02E-03	1	8.36E-03	-	-	1	8.55E-03	-	-
SNIP1	1	1.72E-02	1	1.60E-02	-	-	1	1.63E-02	-	-
SNX32	1	1.63E-02	1	1.51E-02	-	-	1	1.55E-02	-	-
SOCS5	1	1.57E-02	1	1.46E-02	-	-	1	1.49E-02	-	-
SOS1	1	3.63E-02	1	3.37E-02	-	-	1	3.44E-02	-	-
SOS2	1	3.66E-02	1	3.39E-02	-	-	1	3.47E-02	-	-
SOX18	1	1.58E-02	-	-	-	-	1	1.50E-02	-	-
SP3	1	2.19E-02	1	2.03E-02	-	-	1	2.08E-02	-	-
SP6	1	1.44E-02	1	1.33E-02	-	-	1	1.37E-02	-	-
SPATA31D1	1	4.08E-02	1	3.79E-02	-	-	1	3.87E-02	-	-
SPG11	1	6.07E-02	1	5.63E-02	-	-	1	5.76E-02	-	-

SPG7	1	3.13E-02	-	-	1	2.21E-03	1	2.97E-02	-	-
SPHKAP	1	4.93E-02	1	4.57E-02	-	-	1	4.68E-02	-	-
SREK1	1	1.92E-02	1	1.78E-02	-	-	1	1.82E-02	-	-
SRPRB	1	9.94E-03	1	9.21E-03	-	-	1	9.42E-03	-	-
SSH1	1	3.56E-02	1	3.30E-02	-	-	1	3.37E-02	-	-
STAMBP	1	1.36E-02	1	1.26E-02	-	-	1	1.28E-02	-	-
STARD9	1	1.28E-02	-	-	1	8.95E-04	1	1.21E-02	-	-
SUPT16H	1	3.06E-02	1	2.84E-02	-	-	1	2.90E-02	-	-
SWAP70	1	1.52E-02	1	1.41E-02	-	-	1	1.44E-02	-	-
SYNE1	1	2.42E-01	-	-	1	1.91E-02	1	2.31E-01	-	-
SYTL2	1	3.13E-02	1	2.90E-02	-	-	1	2.97E-02	-	-
TACR1	1	1.37E-02	1	1.27E-02	-	-	1	1.30E-02	-	-
TAF4	1	3.03E-02	1	2.81E-02	-	-	1	2.87E-02	-	-
TAF5L	1	2.17E-02	1	2.01E-02	-	-	-	-	1	7.86E-04
TAPBP	1	1.75E-02	1	1.63E-02	-	-	1	1.66E-02	-	-
TAS2R19	1	6.97E-03	-	-	1	4.86E-04	1	6.61E-03	-	-
TAS2R41	1	8.23E-03	1	7.62E-03	-	-	1	7.80E-03	-	-
TBC1D22B	1	1.69E-02	1	1.57E-02	-	-	1	1.60E-02	-	-
TDRD12	1	1.46E-03	1	1.36E-03	-	-	1	1.39E-03	-	-
TDRD6	1	5.65E-02	1	5.24E-02	-	-	1	5.36E-02	-	-
TEK	1	3.25E-02	1	3.01E-02	-	-	1	3.08E-02	-	-
TET3	1	4.82E-02	1	4.47E-02	-	-	1	4.57E-02	-	-
TEX2	1	3.34E-02	1	3.09E-02	-	-	1	3.16E-02	-	-
TFIP11	1	2.81E-02	1	2.61E-02	-	-	1	2.67E-02	-	-
TFRC	1	2.58E-02	1	2.40E-02	-	-	1	2.45E-02	-	-
TGFBR1	1	1.51E-02	1	1.40E-02	-	-	1	1.43E-02	-	-
TGFBR2	2	1.74E-04	2	1.50E-04	-	-	2	1.56E-04	-	-
THAP3	1	8.17E-03	1	7.57E-03	-	-	1	7.74E-03	-	-
THBS3	1	3.64E-02	1	3.38E-02	-	-	1	3.45E-02	-	-
TIGD1	1	1.63E-03	-	-	1	1.14E-04	-	-	1	5.85E-05
TLR5	1	2.07E-02	1	1.92E-02	-	-	1	1.97E-02	-	-
TLR6	1	1.93E-02	1	1.79E-02	-	-	1	1.83E-02	-	-
TMCC2	1	2.93E-02	-	-	1	2.06E-03	1	2.78E-02	-	-
TMED5	1	6.86E-03	1	6.36E-03	-	-	1	6.50E-03	-	-

TMEM132D	1	3.67E-02	1	3.40E-02	-	-	1	3.48E-02	-	-
TMEM139	1	6.95E-03	1	6.44E-03	-	-	1	6.59E-03	-	-
TMEM260	1	1.99E-02	1	1.84E-02	-	-	1	1.88E-02	-	-
TMEM37	1	6.93E-03	1	6.42E-03	-	-	1	6.57E-03	-	-
TMEM63B	1	3.13E-02	1	2.91E-02	-	-	1	2.97E-02	-	-
TMPRSS11D	1	1.21E-02	1	1.12E-02	-	-	1	1.15E-02	-	-
TNFRSF8	1	2.04E-02	1	1.89E-02	-	-	1	1.94E-02	-	-
TNKS	1	4.13E-02	-	-	1	2.93E-03	1	3.92E-02	-	-
TPP1	1	1.80E-02	1	1.67E-02	-	-	1	1.71E-02	-	-
TPRG1	1	9.01E-03	1	8.35E-03	-	-	1	8.54E-03	-	-
TRIM15	1	1.51E-02	1	1.40E-02	-	-	1	1.43E-02	-	-
TRIM27	1	1.73E-02	1	1.61E-02	-	-	1	1.64E-02	-	-
TRIM47	1	1.92E-02	1	1.78E-02	-	-	1	1.82E-02	-	-
TRIM7	1	1.84E-02	1	1.70E-02	-	-	1	1.74E-02	-	-
TRPC4AP	1	2.48E-02	1	2.30E-02	-	-	1	2.36E-02	-	-
TRPM5	1	3.63E-02	1	3.37E-02	-	-	1	3.44E-02	-	-
TSHZ3	1	3.62E-02	1	3.35E-02	-	-	1	3.43E-02	-	-
TSPAN15	1	1.03E-02	1	9.50E-03	-	-	1	9.72E-03	-	-
TSPAN33	1	9.67E-03	1	8.96E-03	-	-	1	9.16E-03	-	-
TTC22	1	1.38E-02	1	1.28E-02	-	-	-	-	1	4.96E-04
TTC37	1	4.22E-02	1	3.91E-02	-	-	1	4.00E-02	-	-
TTC40	1	5.15E-02	1	4.78E-02	-	-	1	4.89E-02	-	-
TTN	1	6.40E-01	1	6.12E-01	-	-	1	6.20E-01	-	-
TUBB4A	1	2.03E-02	-	-	1	1.43E-03	1	1.93E-02	-	-
TWISTNB	2	5.27E-05	2	4.53E-05	-	-	2	4.73E-05	-	-
TXNRD2	1	1.91E-02	-	-	1	1.34E-03	1	1.81E-02	-	-
UBE2S	1	9.07E-03	1	8.41E-03	-	-	1	8.60E-03	-	-
ULK1	1	4.46E-02	1	4.14E-02	-	-	-	-	1	1.63E-03
ULK3	1	1.42E-02	1	1.31E-02	-	-	1	1.34E-02	-	-
UNC13A	1	5.28E-02	1	4.90E-02	-	-	1	5.01E-02	-	-
UPB1	1	1.44E-02	1	1.34E-02	-	-	1	1.37E-02	-	-
URB1	1	8.39E-03	1	7.78E-03	-	-	1	7.95E-03	-	-
USH1C	1	3.02E-02	-	-	1	2.13E-03	1	2.86E-02	-	-
USP24	1	6.11E-02	1	5.67E-02	-	-	1	5.79E-02	-	-

UTF1	1	8.09E-03	1	7.49E-03	-	-	1	7.66E-03	-	-
UTRN	1	1.03E-01	1	9.62E-02	-	-	1	9.83E-02	-	-
VAMP7	1	7.52E-03	1	6.97E-03	-	-	1	7.13E-03	-	-
VAT1	1	1.38E-02	1	1.28E-02	-	-	1	1.30E-02	-	-
VRK2	1	1.33E-02	1	1.23E-02	-	-	1	1.26E-02	-	-
WDFY3	1	1.03E-01	1	9.58E-02	-	-	1	9.79E-02	-	-
WDR46	1	2.27E-02	1	2.11E-02	-	-	1	2.15E-02	-	-
WDR47	1	2.59E-02	1	2.40E-02	-	-	1	2.45E-02	-	-
WDR90	1	6.21E-02	1	5.76E-02	-	-	1	5.89E-02	-	-
YBX1	1	1.21E-02	1	1.12E-02	-	-	1	1.15E-02	-	-
ZBED3	1	2.61E-03	1	2.42E-03	-	-	1	2.47E-03	-	-
ZBED4	1	4.23E-02	1	3.93E-02	-	-	1	4.01E-02	-	-
ZBTB11	1	3.22E-02	1	2.98E-02	-	-	1	3.05E-02	-	-
ZBTB47	1	1.84E-02	1	1.70E-02	-	-	-	-	1	6.63E-04
ZC3H11A	1	2.62E-02	1	2.43E-02	-	-	1	2.48E-02	-	-
ZC3H18	1	3.70E-02	1	3.43E-02	-	-	1	3.51E-02	-	-
ZMYM4	1	4.13E-02	1	3.83E-02	-	-	1	3.92E-02	-	-
ZNF106	1	5.27E-02	1	4.89E-02	-	-	1	5.00E-02	-	-
ZNF219	1	2.55E-02	1	2.37E-02	-	-	1	2.42E-02	-	-
ZNF256	1	1.57E-02	1	1.45E-02	-	-	1	1.49E-02	-	-
ZNF365	1	1.43E-02	1	1.33E-02	-	-	1	1.36E-02	-	-
ZNF397	1	8.91E-03	1	8.26E-03	-	-	1	8.44E-03	-	-
ZNF418	1	1.88E-02	1	1.74E-02	-	-	1	1.78E-02	-	-
ZNF425	1	2.84E-02	1	2.63E-02	-	-	1	2.69E-02	-	-
ZNF469	1	1.34E-02	1	1.24E-02	-	-	1	1.27E-02	-	-
ZNF503	1	2.38E-02	1	2.20E-02	-	-	1	2.25E-02	-	-
ZNF600	1	1.78E-02	1	1.65E-02	-	-	1	1.69E-02	-	-
ZNF608	1	4.40E-02	1	4.08E-02	-	-	1	4.17E-02	-	-
ZNF638	1	5.13E-02	1	4.76E-02	-	-	1	4.87E-02	-	-
ZNF765	1	1.40E-02	1	1.30E-02	-	-	1	1.33E-02	-	-
ZNF844	1	1.34E-02	1	1.24E-02	-	-	1	1.27E-02	-	-
ZZEF1	1	8.78E-02	1	8.16E-02	-	-	1	8.34E-02	-	-

Table S4.7: Protein altering gene-specific DN enrichment in probands by subtype				
Gene	CH&SP		CSP+SMCP	
	Obs PA Vars	PA p-value	Obs PA Vars	PA p-value
ABCC12	-	-	1	0.0158
ABCC6	-	-	1	0.0161
ACADL	-	-	1	0.00542
ACD	-	-	1	0.008
ACO1	-	-	1	0.0103
ACTB	-	-	1	0.00588
ADHFE1	-	-	1	0.0054
AKAP8L	-	-	1	0.0082
ALMS1	1	0.0259	-	-
ANKRD26	-	-	1	0.0152
ANKRD37	1	0.0015	-	-
ARHGDI3	1	0.00174	-	-
ARID1A	2	0.000141	-	-
ARPC1B	-	-	1	0.00536
ASPDH	-	-	1	0.00271
ATG2A	-	-	1	0.025
ATP13A4	1	0.0103	-	-
ATP4A	1	0.0101	-	-
ATP6V1A	1	0.00515	-	-
ATXN1	-	-	1	0.0113
ATXN7L3	-	-	1	0.00474
AURKA	-	-	1	0.00399
BAHD1	-	-	1	0.0111
BAZ1B	-	-	1	0.0162
BDP1	-	-	1	0.0241
BIRC6	1	0.0293	-	-
BOP1	-	-	1	0.00323
BRPF1	1	0.0115	-	-
BSN	-	-	1	0.0481
C16ORF58	1	0.00371	-	-
C1ORF170	-	-	1	0.000884
C2ORF69	1	0.00201	-	-
C5	-	-	1	0.0179
C8B	-	-	1	0.00662
C8ORF46	-	-	1	0.00256
C8ORF76	1	0.00269	-	-
C9ORF173	1	0.00261	-	-
CACNA1D	-	-	1	0.0279
CACNA1G	-	-	1	0.0291
CACNB1	-	-	1	0.00816
CASP4	-	-	1	0.00434

CBL	1	0.00747	-	-
CCDC106	1	0.00317	-	-
CCDC34	-	-	1	0.00416
CCDC88C	-	-	1	0.0241
CDH11	1	0.00679	-	-
CELSR1	1	0.0335	-	-
CILP2	-	-	1	0.0186
CLTC	1	0.0125	-	-
COL11A1	-	-	1	0.0208
COL2A1	2	7.77E-05	2	0.000161
COL4A2	-	-	1	0.0221
CPNE7	-	-	1	0.00814
CPSF1	1	0.0149	-	-
CRISPLD2	-	-	1	0.00712
CSNK2B	1	0.00183	-	-
CTU2	1	0.00501	-	-
CUL7	-	-	1	0.0213
CYSLTR2	-	-	1	0.00324
DAK	1	0.00561	-	-
DAPK3	-	-	1	0.0071
DBN1	1	0.00642	-	-
DCAF4	-	-	1	0.00713
DDX27	-	-	1	0.0105
DENND1A	-	-	1	0.0129
DENND4C	-	-	1	0.0159
DGCR14	1	0.00444	-	-
DGCR8	-	-	1	0.0104
DHX58	-	-	1	0.00889
DMRT1	1	0.00304	-	-
DNAH11	-	-	1	0.0414
DNAJC10	1	0.00643	-	-
DSC2	-	-	1	0.0087
DYNC1H1	-	-	2	0.00185
DYNC1I2	-	-	1	0.00668
ECM2	1	0.00482	-	-
EFTUD2	-	-	1	0.0114
EHMT2	-	-	1	0.0167
EMILIN2	-	-	1	0.0105
EPHB3	-	-	1	0.0157
ERI1	-	-	1	0.00306
EXT2	-	-	1	0.00895
EYA1	-	-	1	0.0069
FAM131A	1	0.00311	-	-
FAM214B	1	0.00427	-	-
FBXO32	1	0.00321	-	-

FBXO43	-	-	1	0.00634
FLNB	-	-	1	0.0335
FMNL3	-	-	1	0.0127
FRY	1	0.0228	-	-
FZD8	1	0.00844	-	-
FZD9	-	-	1	0.0106
GAD2	1	0.00485	-	-
GAL3ST4	1	0.00426	-	-
GALNT18	-	-	1	0.00795
GCN1L1	-	-	1	0.0344
GIGYF2	1	0.0119	-	-
GPR146	1	0.0033	-	-
GRHL3	1	0.00473	1	0.00681
HCN4	-	-	1	0.0168
HDAC9	1	0.00678	-	-
HEATR2	-	-	1	0.00975
HEG1	-	-	1	0.013
HHLA1	1	0.00104	-	-
HIF3A	-	-	1	0.00861
HIPK3	-	-	1	0.0116
HIST1H2AK	1	0.00164	-	-
HIVEP1	-	-	1	0.0266
HPSE	1	0.0039	-	-
HSPA2	1	0.00768	-	-
IGF2R	-	-	1	0.0322
IGSF6	-	-	1	0.0027
IKZF1	-	-	1	0.00663
IMPG1	-	-	1	0.0083
INADL	-	-	1	0.0193
INTS1	1	0.018	-	-
INVS	1	0.00792	-	-
IRF6	1	0.00392	1	0.00565
KAZN	1	0.00688	-	-
KDM4C	-	-	1	0.0111
KIAA1143	1	0.00129	-	-
KIAA1377	1	0.00726	-	-
KIAA1551	-	-	1	0.0149
KIAA2026	-	-	1	0.0192
KIF21A	-	-	1	0.0164
KIF7	-	-	1	0.0128
KNDC1	-	-	1	0.0229
KRT36	1	0.00428	-	-
KRTAP13-3	-	-	1	0.00145
LARGE	-	-	1	0.0107
LILRB1	1	0.00566	-	-

LLGL2	1	0.0105	-	-
LMLN	1	0.00653	-	-
LONP1	-	-	1	0.0154
LPP	1	0.00535	-	-
LRRC73	-	-	1	0.00393
LRWD1	1	0.00571	-	-
MAB21L2	1	0.0043	-	-
MAEL	1	0.00372	-	-
7-Mar	1	0.00508	-	-
MAX	-	-	1	0.00237
MCOLN2	-	-	1	0.00628
MCU	1	0.00285	-	-
MDN1	-	-	1	0.0601
MEIS2	-	-	1	0.00645
MGAT5	1	0.0061	-	-
MSI1	-	-	1	0.00529
MVP	-	-	1	0.0118
MYCBPAP	1	0.00795	-	-
MYH3	1	0.0155	1	0.0223
MYO15A	1	0.0336	-	-
MYO1E	-	-	1	0.0134
MYO9A	1	0.0186	-	-
NAA25	-	-	1	0.0107
NAV3	-	-	1	0.0245
NEDD4L	1	0.00744	-	-
NEIL1	-	-	1	0.00588
NLRP1	1	0.0103	-	-
NPEPPS	-	-	1	0.00832
NUBP2	1	0.00267	-	-
OBSCN	1	0.0802	1	0.114
OR4X1	-	-	1	0.00286
OR5T1	1	0.00192	-	-
OR6C75	-	-	1	0.00264
OTOA	-	-	1	0.00962
PADI1	1	0.00528	-	-
PANX1	1	0.00343	-	-
PAPPA	1	0.0131	-	-
PC	1	0.012	-	-
PCDHA1	-	-	1	0.0145
PCDHA4	-	-	1	0.0143
PDZD7	1	0.00545	-	-
PEG3	-	-	1	0.0171
PER1	-	-	1	0.0158
PITRM1	1	0.0075	-	-
PKHD1L1	1	0.0266	-	-

PLD2	-	-	1	0.0127
PLD3	-	-	1	0.0065
PLEC	-	-	1	0.0733
PLEKHM2	-	-	1	0.00986
PNPLA2	1	0.00412	-	-
POFUT2	1	0.00415	-	-
POLE	-	-	1	0.033
POLM	1	0.00441	-	-
POPDC3	-	-	1	0.00301
PPP1R21	1	0.00551	-	-
PPP2R5D	-	-	1	0.00756
PRKCI	-	-	3	8.35E-08
PRRC2B	-	-	1	0.0278
PRRC2C	-	-	1	0.0263
PSMC3	-	-	1	0.00524
PTCHD2	-	-	1	0.0198
PTPRB	-	-	1	0.0224
PTPRU	-	-	1	0.0214
PYGM	-	-	1	0.0126
R3HCC1L	-	-	1	0.00751
RALGAPA1	1	0.0134	-	-
RANBP17	1	0.0089	-	-
RAPGEF3	-	-	1	0.0113
RBBP6	-	-	1	0.018
RBMS1	1	0.00365	-	-
REEP3	-	-	1	0.00236
REXO1	-	-	1	0.0151
RFC5	1	0.003	-	-
RFX1	1	0.00854	-	-
RLTPR	1	0.0121	-	-
RNF208	-	-	1	0.00314
RNF213	1	0.0388	-	-
RNF39	-	-	1	0.00576
RPL12	-	-	1	0.00226
RPL5	1	0.0026	-	-
RPS6	-	-	1	0.00342
SATB2	3	4.59E-08	-	-
SBNO2	1	0.00972	-	-
SCAMP2	-	-	1	0.00369
SERPINA1	-	-	1	0.00442
SERPINA12	-	-	1	0.00452
SETD6	-	-	1	0.00566
SFTPA2	1	0.00203	-	-
SIGIRR	1	0.0037	-	-
SIKE1	-	-	1	0.00277

SIRPA	1	0.00433	-	-
SIRPG	1	0.00286	-	-
SLC18A1	1	0.00374	-	-
SLC25A41	1	0.00312	1	0.00449
SLC26A6	1	0.00629	-	-
SLC2A9	1	0.00404	-	-
SLC5A8	1	0.00474	-	-
SLC7A14	1	0.00646	-	-
SMAD4	-	-	1	0.00603
SMC5	-	-	1	0.0118
SMC6	1	0.00733	-	-
SNAI1	1	0.00237	-	-
SNF8	-	-	1	0.0033
SNIP1	1	0.00438	-	-
SNX32	1	0.00415	-	-
SOCS5	1	0.004	-	-
SOS1	-	-	1	0.0134
SPATA31D1	-	-	1	0.0151
SPG11	1	0.0157	-	-
SPG7	-	-	1	0.0115
SPHKAP	-	-	1	0.0182
SREK1	-	-	1	0.00702
SRPRB	-	-	1	0.00363
SSH1	-	-	1	0.0131
STAMPB	-	-	1	0.00496
STARD9	1	0.00325	-	-
SUPT16H	-	-	1	0.0113
SWAP70	1	0.00386	-	-
SYNE1	1	0.0676	-	-
SYTL2	1	0.008	-	-
TACR1	-	-	1	0.00501
TAF4	-	-	1	0.0111
TAPBP	-	-	1	0.00642
TAS2R41	-	-	1	0.003
TBC1D22B	-	-	1	0.0062
TDRD12	1	0.00037	-	-
TDRD6	1	0.0146	-	-
TEK	-	-	1	0.012
TET3	1	0.0124	-	-
TEX2	1	0.00853	-	-
TFIP11	1	0.00719	-	-
TFRC	1	0.00659	-	-
TGFBR1	1	0.00384	-	-
TGFBR2	2	1.12E-05	-	-
THAP3	1	0.00207	-	-

THBS3	1	0.00933	-	-
TLR5	-	-	1	0.0076
TLR6	-	-	1	0.00706
TMEM132D	-	-	1	0.0135
TMEM139	-	-	1	0.00254
TMEM260	1	0.00505	-	-
TMEM37	-	-	1	0.00253
TMPRSS11D	1	0.00308	-	-
TNFRSF8	-	-	1	0.00749
TNKS	1	0.0106	-	-
TPP1	1	0.00459	-	-
TPRG1	-	-	1	0.00329
TRIM15	1	0.00383	-	-
TRIM27	-	-	1	0.00634
TRIM47	1	0.00489	-	-
TRIM7	-	-	1	0.00674
TRPM5	1	0.0093	-	-
TSHZ3	1	0.00926	-	-
TSPAN15	1	0.0026	-	-
TSPAN33	-	-	1	0.00353
TTC37	1	0.0108	-	-
TWISTNB	1	0.0026	-	-
TXNRD2	1	0.00486	-	-
UBE2S	1	0.0023	-	-
ULK1	-	-	1	0.0165
UNC13A	-	-	1	0.0195
URB1	-	-	1	0.00306
USH1C	1	0.00772	-	-
USP24	-	-	1	0.0227
UTF1	-	-	1	0.00295
VRK2	1	0.00338	-	-
WDFY3	1	0.0271	-	-
WDR46	-	-	1	0.00833
WDR90	1	0.0161	-	-
YBX1	-	-	1	0.00442
ZBTB11	-	-	1	0.0118
ZC3H11A	-	-	1	0.00962
ZC3H18	-	-	1	0.0136
ZMYM4	-	-	1	0.0152
ZNF106	1	0.0136	-	-
ZNF256	-	-	1	0.00574
ZNF365	-	-	1	0.00523
ZNF397	-	-	1	0.00325
ZNF418	-	-	1	0.00687
ZNF425	-	-	1	0.0104

ZNF469	-	-	1	0.00491
ZNF503	-	-	1	0.00872
ZNF600	-	-	1	0.00652
ZNF608	-	-	1	0.0163
ZNF638	1	0.0132	-	-
ZNF765	-	-	1	0.00513

Category	ID	Name	q-value FDR B&H	# in List	# in Gen ome	Hit in Query List
GO: Molecular Function	GO:0140657	ATP-dependent activity	7.86E-08	42	619	ZGRF1,ARID1A,PGD,LONP1,STARD9,HSPA2,NIPBL,ATP13A4,DHX58,KIF21A,RNF213,NLRP1,DDX27,ABCC6,DHX34,ATP4A,DNAH11,TDRD12,ATP6V1A,MYH3,SPG7,MYO1E,MYO9A,ACSM1,PSMC3,SMC5,MDN1,NAV3,KIF5B,DNAH5,TAPBP,DYNC1H1,DYNC1I2,SMC6,RFC5,NUBP2,KIF7,SMC3,EIF4A2,MACF1,ABCC12,MYO15A
GO: Molecular Function	GO:0016887	ATP hydrolysis activity	2.09E-04	27	412	LONP1,HSPA2,ATP13A4,DHX58,KIF21A,RNF213,NLRP1,DDX27,ABCC6,DHX34,ATP4A,TDRD12,MYH3,SPG7,MYO1E,PSMC3,SMC5,MDN1,NAV3,KIF5B,SMC6,RFC5,KIF7,SMC3,EIF4A2,MACF1,ABCC12
GO: Molecular Function	GO:0003774	cytoskeletal motor activity	2.09E-04	14	127	PGD,STARD9,KIF21A,DNAH11,MYH3,MYO1E,MYO9A,KIF5B,DNAH5,DYNC1H1,DYNC1I2,KIF7,SMC3,MYO15A
GO: Molecular Function	GO:0032559	adenyl ribonucleotide binding	2.09E-04	65	1605	EPHB3,HK2,ACADL,ACLY,ACTB,OBSCN,TTN,PGD,LONP1,RAPGEF3,STARD9,ULK1,HSPA2,ATP13A4,VRK2,DHX58,KIF21A,RNF213,NLRP1,DDX27,ABCC6,ULK3,TKFC,PRKCI,PKN1,DHX34,ATP4A,DNAH11,TDRD12,ATP6V1A,MYH3,SPG7,MYO1E,SWAP70,MYO9A,ACSM1,PSMC3,DAPK3,SMC5,AURKA,MDN1,NAV3,UBE2S,PYGM,POPDC3,KIF5B,DNAH5,DYNC1H1,HCN4,CAMK2G,SMC6,BAZ1B,GAL3ST4,RFC5,TEK,NUBP2,KIF7,HIPK3,TGFBR1,TGFBR2,SMC3,EIF4A2,ABCC12,MYO15A,PC
GO: Molecular Function	GO:0030554	adenyl nucleotide binding	5.78E-04	67	1735	EPHB3,HK2,ACADL,ACLY,ACTB,OBSCN,TTN,PGD,LONP1,RAPGEF3,STARD9,ULK1,HSPA2,ASPDH,ATP13A4,VRK2,DHX58,KIF21A,RNF213,NLRP1,DDX27,ABCC6,ULK3,TKFC,PRKCI,PKN1,DHX34,ATP4A,DNAH11,

						TDRD12,ATP6V1A,MYH3,SPG7,MYO1E,SWAP70,MYO9A,NOX5,ACSM1,PSMC3,DAPK3,SMC5,AURKA,MDN1,NAV3,UBE2S,PYGM,POPDC3,KIF5B,DNAH5,DYNC1H1,HCN4,CAMK2G,SMC6,BAZ1B,GAL3ST4,RFC5,TEK,NUBP2,KIF7,HIPK3,TGFBR1,TGFBR2,SMC3,EIF4A2,ABCC12,MYO15A,PC
GO: Molecular Function	GO:0003777	microtubule motor activity	1.21E-03	10	78	PGD,STARD9,KIF21A,DNAH11,KIF5B,DNAH5,DYNC1H1,DYNC1I2,KIF7,SMC3
GO: Molecular Function	GO:0005524	ATP binding	1.73E-03	59	1527	EPHB3,HK2,ACLY,ACTB,OBSCN,TTN,PGD,LONP1,STARD9,ULK1,HSPA2,ATP13A4,VRK2,DHX58,KIF21A,RNF213,NLRP1,DDX27,ABCC6,ULK3,TKFC,PRKCI,PKN1,DHX34,ATP4A,DNAH11,TDRD12,ATP6V1A,MYH3,SPG7,MYO1E,SWAP70,MYO9A,ACSM1,PSMC3,DAPK3,SMC5,AURKA,MDN1,NAV3,UBE2S,KIF5B,DNAH5,DYNC1H1,CAMK2G,SMC6,BAZ1B,RFC5,TEK,NUBP2,KIF7,HIPK3,TGFBR1,TGFBR2,SMC3,EIF4A2,ABCC12,MYO15A,PC
GO: Molecular Function	GO:0032555	purine ribonucleotide binding	3.14E-03	71	2000	EPHB3,HK2,ACADL,ACLY,ACTB,OBSCN,TTN,PGD,SRPRB,EFTUD2,TUBB4A,LONP1,RAPGEF3,STARD9,ULK1,HSPA2,ATP13A4,VRK2,DHX58,KIF21A,RNF213,NLRP1,DDX27,ABCC6,RANBP17,ULK3,ARHGDIG,TKFC,PRKCI,PKN1,DHX34,ATP4A,DNAH11,TDRD12,ATP6V1A,MYH3,HBS1L,SPG7,MYO1E,SWAP70,MYO9A,ACSM1,PSMC3,DAPK3,SMC5,AURKA,MDN1,NAV3,UBE2S,PYGM,POPDC3,KIF5B,DNAH5,DYNC1H1,HCN4,CAMK2G,SMC6,BAZ1B,GAL3ST4,RFC5,TEK,NUBP2,KIF7,HIPK3,TGFBR1,TGFBR2,SMC3,EIF4A2,ABCC12,MYO15A,PC
GO: Molecular Function	GO:0019904	protein domain specific binding	3.48E-03	39	893	HIVEP1,LIN7C,IKZF1,SIRPG,FZD8,LONP1,HOXC4,MUC17,RAPGEF3,CLTC,HSPA2,NIPBL,UTF1,VRK2,KIF21A,EPB41L5,NLRP1,DENND1A,STAMPB,CSNK2B,UNC13A,SLC26A6,SOS1,DAPK3,GCKR,SIRPA,LILRB1,ASA

						P1,EHMT2,CACNA1D,CACNB1,CASP4,RAPGEF1,CASP10,CBL,CCDC88C,ZNF106,LLGL2,CABIN1
GO: Biological Process	GO:0043009	chordate embryonic development	1.77E-05	59	1255	SOX18,MECOM,EYA1,ARID1A,PLXNB2,BIRC6,TTN,MEIS2,MEN1,SERPINA1,HEG1,HOXC4,NIPBL,PLXNA2,COL2A1,TSHZ3,COL11A1,EPB41L5,MAB21L2,ABCC6,ARHGDIG,CELSR1,MCU,PRKCI,IFT122,SNAI1,GRHL3,DHX34,MYH3,SP3,UPB1,MYO1E,KDM4C,INTS1,PSMC3,IRF6,DBN1,BRPF1,C5,PRRC2B,CRISPLD2,SATB2,HCN4,PRICKLE1,YBX1,RBBP6,CASP4,RAPGEF1,PTPRU,E2F8,TGFBR1,CCDC88C,TGFBR2,KMT2D,LLGL2,PAPPA,CDH2,RPL5,SMAD4
GO: Biological Process	GO:0009792	embryo development ending in birth or egg hatching	1.77E-05	59	1278	SOX18,MECOM,EYA1,ARID1A,PLXNB2,BIRC6,TTN,MEIS2,MEN1,SERPINA1,HEG1,HOXC4,NIPBL,PLXNA2,COL2A1,TSHZ3,COL11A1,EPB41L5,MAB21L2,ABCC6,ARHGDIG,CELSR1,MCU,PRKCI,IFT122,SNAI1,GRHL3,DHX34,MYH3,SP3,UPB1,MYO1E,KDM4C,INTS1,PSMC3,IRF6,DBN1,BRPF1,C5,PRRC2B,CRISPLD2,SATB2,HCN4,PRICKLE1,YBX1,RBBP6,CASP4,RAPGEF1,PTPRU,E2F8,TGFBR1,CCDC88C,TGFBR2,KMT2D,LLGL2,PAPPA,CDH2,RPL5,SMAD4
GO: Biological Process	GO:0048598	embryonic morphogenesis	1.77E-05	56	1189	RPS6,SOX18,MECOM,PDZD7,EXT2,EYA1,ARID1A,PLXNB2,MEIS2,MEN1,FZD8,HOXC4,NIPBL,PLXNA2,COL2A1,COL4A2,COL11A1,EPB41L5,MAB21L2,CELSR1,MCU,PRKCI,IFT122,SNAI1,GRHL3,INVS,SOS1,MYH3,SP3,ACD,SWAP70,PSMC3,IRF6,BRPF1,POFUT2,ADGRA3,CRISPLD2,TRIM15,SATB2,DSC2,PRICKLE1,YBX1,CAMK2G,CASP4,USH1C,ZNF503,TGFBR1,CCDC88C,TGFBR2,KMT2D,LLGL2,LPP,MACF1,MYO15A,CDH2,SMAD4
GO: Biological Process	GO:0048729	tissue morphogenesis	3.29E-04	51	1141	LIN7C,ACTB,SOX18,EXT2,EYA1,ARID1A,PLXNB2,TTN,MEIS2,HEG1,KRT71,FBXO32,NIPBL,PLXNA2,COL11A1,EPB41L5,MAB21L2,CELSR1,CSNK2B,MCU,PRKCI,IFT122,SNAI1,GRHL3,INVS,SOS1,SWAP70,MYO9A,S

						YTL2,LAMA1,ALMS1,BRPF1,POFUT2,ADGRA3,CRISPLD2,TRIM15,SATB2,CARMIL2,DSC2,PRICKLE1,CAMK2G,TEK,TGFBR1,CCDC88C,TGFBR2,KMT2D,LLGL2,LPP,MACF1,CDH2,SMAD4
GO: Biological Process	GO:0048568	embryonic organ development	1.32E-03	45	1004	EPHB3,ZNF219,SOX18,PDZD7,EYA1,ARID1A,BIRC6,MEIS2,MEN1,FZD8,HOXC4,NIPBL,PLXNA2,COL2A1,COL11A1,POLE,EPB41L5,MAB21L2,CELSR1,PRKCI,IFT122,SNAI1,GRHL3,INVS,HDAC9,SP3,LAMA1,PSMC3,IRF6,CRISPLD2,SATB2,PRICKLE1,RBBP6,CAMK2G,CASP4,USH1C,E2F8,ZNF503,TFRC,TGFBR1,TGFBR2,KMT2D,LLGL2,MYO15A,CDH2
GO: Biological Process	GO:0048880	sensory system development	2.22E-03	42	932	MAX,OBSCN,PER1,IKZF1,PDZD7,EXT2,ARID1A,MEIS2,EFTUD2,FZD8,MGAT5,NIPBL,PLXNA2,EPB41L5,MAB21L2,SLC7A14,PRKCI,IFT122,SNAI1,GRHL3,SOS1,SP3,LAMA1,PSMC3,ALMS1,EHMT2,LMLN,MVP,DYNC1H1,NRF1,PRICKLE1,HIF3A,RFC5,USH1C,MCOLN1,ZNF503,TGFBR1,TGFBR2,LLGL2,CDH2,CDH11,LARGE1
GO: Biological Process	GO:0048562	embryonic organ morphogenesis	3.53E-03	33	667	SOX18,PDZD7,EYA1,ARID1A,MEIS2,MEN1,FZD8,HOXC4,NIPBL,PLXNA2,COL2A1,COL11A1,EPB41L5,MAB21L2,CELSR1,IFT122,GRHL3,INVS,SP3,PSMC3,IRF6,CRISPLD2,SATB2,PRICKLE1,CAMK2G,CASP4,USH1C,ZNF503,TGFBR1,TGFBR2,KMT2D,MYO15A,CDH2
GO: Biological Process	GO:0007369	gastrulation	6.21E-03	24	420	RPS6,EXT2,EYA1,ARID1A,COL4A2,COL11A1,EPB41L5,CELSR1,MCU,SNAI1,INVS,SWAP70,POFUT2,ADGRA3,TRIM15,SATB2,DSC2,PRICKLE1,TGFBR2,KMT2D,LPP,MACF1,CDH2,SMAD4
GO: Cellular Component	GO:0098862	cluster of actin-based cell projections	1.33E-02	16	237	ACTB,SLC6A19,PDZD7,ACTN3,MUC17,RAPGEF3,PLD2,PLEC,FLNB,PRKCI,SLC26A6,MYO1E,USH1C,PKHD1L1,SLC5A8,MYO15A
GO: Cellular Component	GO:0005903	brush border	1.33E-02	13	168	ACTB,SLC6A19,ACTN3,MUC17,RAPGEF3,PLD2,PLEC,FLNB,PRKCI,SLC26A6,MYO1E,USH1C,SLC5A8

GO: Cellular Component	GO:0140513	nuclear protein-containing complex	1.33E-02	51	1386	LRWD1,SF3B4,ESS2,MAX,ACTB,MECOM,ARID1A,PGD,MEN1,EFTUD2,CLTC,CPSF1,DGCR8,RBM47,NIPBL,POLE,NOL11,ERI1,BAHD1,RANBP17,CSNK2B,TNKS,TAF5L,HDAC9,ACD,INTS1,TFIP11,SREK1,CUL7,SFR1,POLR1F,ATXN7L3,ZC3H11A,R3HCC1L,CCNL1,UBE2S,BOP1,TAF4,MVP,SATB2,YBX1,SYNE1,BAZ1B,HIF3A,RFC5,E2F8,SNIP1,KMT2D,SUPT16H,SLC5A8,SMAD4
Human Phenotype	HP:0000193	Bifid uvula	1.17E-03	19	162	EYA1,TTN,COL2A1,COL11A1,SKIC3,GRHL3,IRF6,PIGO,SATB2,KIF7,TGFBR1,TGFBR2,KMT2D,SUPT16H,KAT6B,RPL5,CDH11,SMAD4,LARGE1
Human Phenotype	HP:0410004	Cleft secondary palate	1.17E-03	23	232	ESS2,EYA1,TTN,PRR12,DGCR8,COL2A1,COL11A1,PPP2R5D,SKIC3,GRHL3,IRF6,PIGO,SATB2,KIF7,TGFBR1,TGFBR2,KMT2D,SUPT16H,KAT6B,RPL5,CDH11,SMAD4,LARGE1
Human Phenotype	HP:0000172	Abnormal uvula morphology	2.14E-03	20	192	EYA1,TTN,PRR12,COL2A1,COL11A1,SKIC3,GRHL3,IRF6,PIGO,SATB2,KIF7,TGFBR1,TGFBR2,KMT2D,SUPT16H,KAT6B,RPL5,CDH11,SMAD4,LARGE1
Human Phenotype	HP:0100736	Abnormal soft palate morphology	3.79E-03	23	258	SF3B4,ESS2,EYA1,TTN,PRR12,DGCR8,COL2A1,COL11A1,SKIC3,GRHL3,IRF6,PIGO,SATB2,KIF7,TGFBR1,TGFBR2,KMT2D,SUPT16H,KAT6B,RPL5,CDH11,SMAD4,LARGE1
Human Phenotype	HP:0000343	Long philtrum	5.65E-03	29	390	ACTB,EXT2,ARID1A,MEIS2,CLTC,NIPBL,COL2A1,PPP1R21,COL11A1,NLRP1,MAB21L2,PPP2R5D,SKIC3,ATP6V1A,MYH3,INTS1,CUL7,BRPF1,TET3,SATB2,SYNE1,CAMK2G,BAZ1B,CBL,KIF7,SMC3,KAT6B,CDH2,CDH11
Disease - DisGeNET Curated	C3714756	Intellectual Disability	3.16E-03	21	447	RALGAPA1,MEIS2,PRR12,TPP1,PITRM1,PPP2R5D,ZBTB11,INPP4A,LAMA1,CACNA1G,TET3,DYNC1H1,DYNC1H2,GOT2,SYNE1,KIF7,FRY,SUPT16H,MACF1,CDH2,LARGE1
Disease - DisGeNET Curated	C0008925	Cleft Palate	3.16E-03	9	81	MEIS2,COL2A1,COL11A1,INTS1,IRF6,SATB2,NEDD4L,KIF7,RPL5

Disease - DisGeNET Curated	C0006142	Malignant neoplasm of breast	5.23E-03	35	1074	CTU2,RPS6,OBSCN,PER1,MECOM,ARID1A,NAA25,SNX32,PLD2,NIPBL,RUSF1,FLNB,COL11A1,FAM131A,MSI1,ZNF365,SNAI1,DBN1,HPSE,AURKA,ADHFE1,ZC3H11A,YBX1,SYNE1,SMC6,TEK,MCOLN1,TFRC,THBS3,KMT2D,LLGL2,EIF4A2,MACF1,KAT6B,CDH2
Disease - AllianceGenome	DOID:3490	Noonan syndrome (implicated_via_orthology)	7.57E-03	3	4	LZTR1,CBL,KAT6B
Disease - AllianceGenome	DOID:674	cleft palate (is_implicated_in)	8.69E-03	4	12	MEIS2,COL2A1,FLNB,IRF6

Table S4.9: Enrichment in OFC-gene panel

cat	group	class	obs	exp	enrichment	pValue	z	SE	lowbound	upbound
All genes	all	syn	6	4.5	1.33	0.297	0.544	1.089	0.241	2.419
All genes	all	mis	22	10.1	2.19	0.00075	0.464	0.929	1.261	3.119
All genes	all	lof	18	1.4	12.8	2.02E-14	3.030	6.061	6.739	18.861
All genes	all	prot	40	11.5	3.49	4.27E-11	0.550	1.100	2.390	4.590
All genes	all	all	46	16	2.88	7.15E-10	0.424	0.848	2.032	3.728
All genes	CH&SP	syn	2	1.1	1.76	0.315	1.286	2.571	-0.811	4.331
All genes	CH&SP	mis	10	2.5	3.94	0.00032	1.265	2.530	1.410	6.470
All genes	CH&SP	lof	6	0.4	16.8	2.10E-06	6.124	12.247	4.553	29.047
All genes	CH&SP	prot	16	2.9	5.52	7.83E-08	1.379	2.759	2.761	8.279
All genes	CH&SP	all	18	4	4.46	2.81E-07	1.061	2.121	2.339	6.581
All genes	CSP+SMCP	syn	4	1.6	2.44	0.0843	1.250	2.500	-0.060	4.940
All genes	CSP+SMCP	mis	6	3.7	1.64	0.165	0.662	1.324	0.316	2.964
All genes	CSP+SMCP	lof	8	0.5	15.6	7.65E-08	5.657	11.314	4.286	26.914
All genes	CSP+SMCP	prot	14	4.2	3.35	0.00012	0.891	1.782	1.568	5.132
All genes	CSP+SMCP	all	18	5.8	3.09	3.86E-05	0.731	1.463	1.627	4.553
AD	all	syn	3	2.2	1.39	0.366	0.787	1.575	-0.185	2.965

AD	all	mis	15	4.7	3.19	0.00012	0.824	1.648	1.542	4.838
AD	all	lof	18	0.6	31.9	3.09E-21	7.071	14.142	17.758	46.042
AD	all	prot	33	5.3	6.27	4.48E-16	1.084	2.168	4.102	8.438
AD	all	all	36	7.4	4.85	4.41E-14	0.811	1.622	3.228	6.472
AD	CH&SP	syn	1	0.5	1.83	0.42	2.000	4.000	-2.170	5.830
AD	CH&SP	mis	7	1.2	5.9	0.00024	2.205	4.410	1.490	10.310
AD	CH&SP	lof	6	0.1	42.1	1.03E-08	24.495	48.990	-6.890	91.090
AD	CH&SP	prot	13	1.3	9.78	1.91E-09	2.774	5.547	4.233	15.327
AD	CH&SP	all	14	1.9	7.46	1.34E-08	1.969	3.939	3.521	11.399
AD	CSP+SMCP	syn	2	0.8	2.54	0.186	1.768	3.536	-0.996	6.076
AD	CSP+SMCP	mis	4	1.7	2.34	0.0949	1.176	2.353	-0.013	4.693
AD	CSP+SMCP	lof	8	0.2	38.9	6.59E-11	14.142	28.284	10.616	67.184
AD	CSP+SMCP	prot	12	1.9	6.26	8.86E-07	1.823	3.646	2.614	9.906
AD	CSP+SMCP	all	14	2.7	5.18	1.04E-06	1.386	2.772	2.408	7.952
AR	all	syn	1	1.9	0.529	0.849	0.526	1.053	-0.524	1.582
AR	all	mis	5	4.3	1.15	0.438	0.520	1.040	0.110	2.190
AR	all	lof	3	0.7	4.17	0.0365	2.474	4.949	-0.779	9.119
AR	all	prot	8	5.1	1.58	0.14	0.555	1.109	0.471	2.689
AR	all	all	9	7	1.29	0.265	0.429	0.857	0.433	2.147
AR	CH&SP	syn	1	0.5	2.09	0.38	2.000	4.000	-1.910	6.090
AR	CH&SP	lof	1	0.2	5.51	0.166	5.000	10.000	-4.490	15.510
AR	CH&SP	prot	1	1.3	0.782	0.722	0.769	1.538	-0.756	2.320
AR	CH&SP	all	2	1.8	1.14	0.524	0.786	1.571	-0.431	2.711
AR	CSP+SMCP	mis	3	1.6	1.9	0.212	1.083	2.165	-0.265	4.065
AR	CSP+SMCP	lof	2	0.3	7.64	0.0288	4.714	9.428	-1.788	17.068
AR	CSP+SMCP	prot	5	1.8	2.71	0.0397	1.242	2.485	0.225	5.195
AR	CSP+SMCP	all	5	2.5	1.97	0.113	0.894	1.789	0.181	3.759

Table S4.10: Enrichment in snRNAseq marker genes										
pheno	cells	class	obs	exp	enrichment	pValue	z	SE	lower	upper
AllCP	all	syn	35	26.1	1.34	0.0561	0.227	0.453	0.887	1.793
AllCP	all	mis	81	58.3	1.39	0.00278	0.154	0.309	1.081	1.699
AllCP	all	lof	26	8.6	3.03	1.24E-06	0.593	1.186	1.844	4.216
AllCP	all	prot	107	66.8	1.6	3.72E-06	0.155	0.310	1.290	1.910
AllCP	all	all	142	93	1.53	1.40E-06	0.128	0.256	1.274	1.786
AllCP	Blood_Cells	syn	1	1.2	0.823	0.703	0.833	1.667	-0.844	2.490
AllCP	Blood_Cells	mis	4	2.7	1.47	0.29	0.741	1.481	-0.011	2.951
AllCP	Blood_Cells	prot	4	3.2	1.27	0.386	0.625	1.250	0.020	2.520
AllCP	Blood_Cells	all	5	4.4	1.15	0.442	0.508	1.016	0.134	2.166
AllCP	chondro_progs	mis	5	2.1	2.37	0.0633	1.065	2.130	0.240	4.500
AllCP	chondro_progs	lof	6	0.3	20.1	7.58E-07	8.165	16.330	3.770	36.430
AllCP	chondro_progs	prot	11	2.4	4.56	4.47E-05	1.382	2.764	1.796	7.324
AllCP	chondro_progs	all	11	3.4	3.27	0.000737	0.975	1.951	1.319	5.221
AllCP	earlyOP	syn	2	1.3	1.49	0.39	1.088	2.176	-0.686	3.666
AllCP	earlyOP	mis	7	3	2.35	0.0327	0.882	1.764	0.586	4.114
AllCP	earlyOP	lof	1	0.5	2.17	0.369	2.000	4.000	-1.830	6.170
AllCP	earlyOP	prot	8	3.4	2.32	0.0247	0.832	1.664	0.656	3.984
AllCP	earlyOP	all	10	4.8	2.09	0.0249	0.659	1.318	0.772	3.408
AllCP	Endothelium	syn	12	6.4	1.87	0.0308	0.541	1.083	0.787	2.953
AllCP	Endothelium	mis	28	14.1	1.98	0.000727	0.375	0.751	1.229	2.731
AllCP	Endothelium	lof	3	2	1.5	0.324	0.866	1.732	-0.232	3.232
AllCP	Endothelium	prot	31	16.1	1.92	0.000646	0.346	0.692	1.228	2.612
AllCP	Endothelium	all	43	22.5	1.91	8.08E-05	0.291	0.583	1.327	2.493
AllCP	Epithelium	syn	7	6.3	1.11	0.446	0.420	0.840	0.270	1.950
AllCP	Epithelium	mis	22	14.3	1.54	0.035	0.328	0.656	0.884	2.196
AllCP	Epithelium	lof	6	2	2.93	0.0183	1.225	2.449	0.481	5.379

AllCP	Epithelium	prot	28	16.3	1.71	0.00541	0.325	0.649	1.061	2.359
AllCP	Epithelium	all	35	22.7	1.54	0.00971	0.261	0.521	1.019	2.061
AllCP	lateOP	syn	2	2.9	0.699	0.779	0.488	0.975	-0.276	1.674
AllCP	lateOP	mis	9	6.4	1.41	0.193	0.469	0.938	0.473	2.348
AllCP	lateOP	lof	3	1	3.08	0.0758	1.732	3.464	-0.384	6.544
AllCP	lateOP	prot	12	7.3	1.63	0.0703	0.475	0.949	0.681	2.579
AllCP	lateOP	all	14	10.2	1.37	0.151	0.367	0.734	0.636	2.104
AllCP	Mesenchyme	syn	3	2.4	1.27	0.423	0.722	1.443	-0.173	2.713
AllCP	Mesenchyme	mis	6	5.3	1.13	0.438	0.462	0.924	0.206	2.054
AllCP	Mesenchyme	lof	3	0.8	3.71	0.0486	2.165	4.330	-0.620	8.040
AllCP	Mesenchyme	prot	9	6.1	1.47	0.165	0.492	0.984	0.486	2.454
AllCP	Mesenchyme	all	12	8.5	1.41	0.15	0.408	0.815	0.595	2.225
AllCP	muscle_progs	syn	8	5.1	1.56	0.147	0.555	1.109	0.451	2.669
AllCP	muscle_progs	mis	15	11.6	1.29	0.195	0.334	0.668	0.622	1.958
AllCP	muscle_progs	lof	7	1.8	3.83	0.00277	1.470	2.940	0.890	6.770
AllCP	muscle_progs	prot	22	13.4	1.64	0.0197	0.350	0.700	0.940	2.340
AllCP	muscle_progs	all	30	18.6	1.61	0.00896	0.294	0.589	1.021	2.199
AllCP	neuro_progs	syn	5	4.8	1.04	0.523	0.466	0.932	0.108	1.972
AllCP	neuro_progs	mis	5	10.5	0.476	0.979	0.213	0.426	0.050	0.902
AllCP	neuro_progs	lof	1	1.5	0.672	0.774	0.667	1.333	-0.661	2.005
AllCP	neuro_progs	prot	6	12	0.5	0.98	0.204	0.408	0.092	0.908
AllCP	neuro_progs	all	11	16.8	0.655	0.946	0.197	0.395	0.260	1.050
AllCP	Pax9_Mesenchyme	syn	1	1.3	0.755	0.734	0.769	1.538	-0.783	2.293
AllCP	Pax9_Mesenchyme	mis	3	3	0.998	0.578	0.577	1.155	-0.157	2.153
AllCP	Pax9_Mesenchyme	lof	1	0.5	2.01	0.392	2.000	4.000	-1.990	6.010
AllCP	Pax9_Mesenchyme	prot	4	3.5	1.14	0.464	0.571	1.143	-0.003	2.283
AllCP	Pax9_Mesenchyme	all	5	4.8	1.04	0.529	0.466	0.932	0.108	1.972
CH&SP	Blood_Cells	prot	2	1.1	1.81	0.302	1.286	2.571	-0.761	4.381
CH&SP	Blood_Cells	mis	2	0.7	2.91	0.151	2.020	4.041	-1.131	6.951

CH&SP	Blood_Cells	lof	2	0.8	2.51	0.19	1.768	3.536	-1.026	6.046
CH&SP	chondro_progs	prot	4	0.6	6.57	0.00354	3.333	6.667	-0.097	13.237
CH&SP	chondro_progs	mis	2	0.5	3.75	0.101	2.828	5.657	-1.907	9.407
CH&SP	chondro_progs	lof	2	0.1	26.5	0.0027	14.142	28.284	-1.784	54.784
CH&SP	chondro_progs	all	4	0.8	4.71	0.0111	2.500	5.000	-0.290	9.710
CH&SP	earlyOP	prot	3	0.9	3.45	0.058	1.925	3.849	-0.399	7.299
CH&SP	earlyOP	mis	2	0.8	2.65	0.175	1.768	3.536	-0.886	6.186
CH&SP	earlyOP	lof	1	0.1	8.61	0.11	10.000	20.000	-11.390	28.610
CH&SP	earlyOP	all	3	1.2	2.48	0.123	1.443	2.887	-0.407	5.367
CH&SP	Endothelium	prot	9	4.1	2.21	0.0237	0.732	1.463	0.747	3.673
CH&SP	Endothelium	syn	2	1.6	1.24	0.481	0.884	1.768	-0.528	3.008
CH&SP	Endothelium	mis	9	3.6	2.52	0.0111	0.833	1.667	0.853	4.187
CH&SP	Endothelium	all	11	5.7	1.93	0.0312	0.582	1.164	0.766	3.094
CH&SP	Epithelium	prot	8	4.1	1.94	0.0592	0.690	1.380	0.560	3.320
CH&SP	Epithelium	syn	2	1.6	1.25	0.475	0.884	1.768	-0.518	3.018
CH&SP	Epithelium	mis	7	3.6	1.94	0.0743	0.735	1.470	0.470	3.410
CH&SP	Epithelium	lof	1	0.5	1.93	0.404	2.000	4.000	-2.070	5.930
CH&SP	Epithelium	all	10	5.7	1.75	0.0665	0.555	1.110	0.640	2.860
CH&SP	lateOP	prot	6	1.9	3.24	0.0119	1.289	2.578	0.662	5.818
CH&SP	lateOP	mis	5	1.6	3.11	0.0241	1.398	2.795	0.315	5.905
CH&SP	lateOP	lof	1	0.2	4.06	0.218	5.000	10.000	-5.940	14.060
CH&SP	lateOP	all	6	2.6	2.33	0.0474	0.942	1.884	0.446	4.214
CH&SP	Mesenchyme	prot	3	1.5	1.94	0.203	1.155	2.309	-0.369	4.249
CH&SP	Mesenchyme	syn	1	0.6	1.67	0.451	1.667	3.333	-1.663	5.003
CH&SP	Mesenchyme	mis	2	1.3	1.49	0.388	1.088	2.176	-0.686	3.666
CH&SP	Mesenchyme	lof	1	0.2	4.9	0.185	5.000	10.000	-5.100	14.900
CH&SP	Mesenchyme	all	4	2.1	1.87	0.17	0.952	1.905	-0.035	3.775
CH&SP	muscle_progs	prot	8	3.4	2.35	0.023	0.832	1.664	0.686	4.014
CH&SP	muscle_progs	syn	1	1.3	0.772	0.726	0.769	1.538	-0.766	2.310

CH&SP	muscle_progs	mis	6	2.9	2.04	0.0777	0.845	1.689	0.351	3.729
CH&SP	muscle_progs	lof	2	0.5	4.34	0.0787	2.828	5.657	-1.317	9.997
CH&SP	muscle_progs	all	9	4.7	1.92	0.0499	0.638	1.277	0.643	3.197
CH&SP	neuro_progs	prot	2	3	0.66	0.805	0.471	0.943	-0.283	1.603
CH&SP	neuro_progs	syn	2	1.2	1.65	0.342	1.179	2.357	-0.707	4.007
CH&SP	neuro_progs	mis	2	2.7	0.754	0.743	0.524	1.048	-0.294	1.802
CH&SP	neuro_progs	all	4	4.2	0.943	0.612	0.476	0.952	-0.009	1.895
CH&SP	Pax9_Mesenchyme	prot	2	0.9	2.26	0.222	1.571	3.143	-0.883	5.403
CH&SP	Pax9_Mesenchyme	mis	2	0.8	2.63	0.177	1.768	3.536	-0.906	6.166
CH&SP	Pax9_Mesenchyme	all	2	1.2	1.64	0.345	1.179	2.357	-0.717	3.997
CSP_SMCP	Blood_Cells	mis	1	0.4	2.26	0.358	2.500	5.000	-2.740	7.260
CSP_SMCP	Blood_Cells	lof	1	1	1.01	0.629	1.000	2.000	-0.990	3.010
CSP_SMCP	Blood_Cells	all	2	1.6	1.26	0.472	0.884	1.768	-0.508	3.028
CSP_SMCP	Blood_Cells	prot	1	1.1	0.871	0.683	0.909	1.818	-0.947	2.689
CSP_SMCP	chondro_progs	mis	2	0.8	2.6	0.18	1.768	3.536	-0.936	6.136
CSP_SMCP	chondro_progs	lof	3	0.1	27.6	0.000197	17.321	34.641	-7.041	62.241
CSP_SMCP	chondro_progs	all	5	1.2	4.09	0.00838	1.863	3.727	0.363	7.817
CSP_SMCP	chondro_progs	prot	5	0.9	5.7	0.00211	2.485	4.969	0.731	10.669
CSP_SMCP	earlyOP	syn	1	0.5	2.04	0.388	2.000	4.000	-1.960	6.040
CSP_SMCP	earlyOP	mis	1	1.1	0.92	0.663	0.909	1.818	-0.898	2.738
CSP_SMCP	earlyOP	all	2	1.7	1.15	0.521	0.832	1.664	-0.514	2.814
CSP_SMCP	earlyOP	prot	1	1.3	0.797	0.715	0.769	1.538	-0.741	2.335
CSP_SMCP	Endothelium	syn	6	2.3	2.57	0.0317	1.065	2.130	0.440	4.700
CSP_SMCP	Endothelium	mis	12	5.1	2.33	0.00676	0.679	1.358	0.972	3.688
CSP_SMCP	Endothelium	lof	2	0.7	2.74	0.166	2.020	4.041	-1.301	6.781
CSP_SMCP	Endothelium	all	20	8.2	2.44	0.000348	0.545	1.091	1.349	3.531
CSP_SMCP	Endothelium	prot	14	5.9	2.38	0.00303	0.634	1.268	1.112	3.648
CSP_SMCP	Epithelium	syn	2	2.3	0.868	0.67	0.615	1.230	-0.362	2.098
CSP_SMCP	Epithelium	mis	9	5.2	1.73	0.0825	0.577	1.154	0.576	2.884

CSP_SMCP	Epithelium	lof	5	0.7	6.7	0.00104	3.194	6.389	0.311	13.089
CSP_SMCP	Epithelium	all	16	8.3	1.94	0.0108	0.482	0.964	0.976	2.904
CSP_SMCP	Epithelium	prot	14	6	2.35	0.00339	0.624	1.247	1.103	3.597
CSP_SMCP	lateOP	syn	2	1	1.92	0.28	1.414	2.828	-0.908	4.748
CSP_SMCP	lateOP	mis	1	2.3	0.431	0.902	0.435	0.870	-0.439	1.301
CSP_SMCP	lateOP	lof	1	0.4	2.82	0.299	2.500	5.000	-2.180	7.820
CSP_SMCP	lateOP	all	4	3.7	1.08	0.509	0.541	1.081	-0.001	2.161
CSP_SMCP	lateOP	prot	2	2.7	0.748	0.746	0.524	1.048	-0.300	1.796
CSP_SMCP	Mesenchyme	syn	2	0.9	2.32	0.214	1.571	3.143	-0.823	5.463
CSP_SMCP	Mesenchyme	mis	3	1.9	1.55	0.305	0.912	1.823	-0.273	3.373
CSP_SMCP	Mesenchyme	lof	1	0.3	3.4	0.255	3.333	6.667	-3.267	10.067
CSP_SMCP	Mesenchyme	all	6	3.1	1.94	0.0934	0.790	1.580	0.360	3.520
CSP_SMCP	Mesenchyme	prot	4	2.2	1.8	0.186	0.909	1.818	-0.018	3.618
CSP_SMCP	muscle_progs	syn	5	1.9	2.68	0.0416	1.177	2.354	0.326	5.034
CSP_SMCP	muscle_progs	mis	6	4.2	1.42	0.252	0.583	1.166	0.254	2.586
CSP_SMCP	muscle_progs	lof	2	0.7	3.01	0.144	2.020	4.041	-1.031	7.051
CSP_SMCP	muscle_progs	all	13	6.8	1.92	0.0214	0.530	1.060	0.860	2.980
CSP_SMCP	muscle_progs	prot	8	4.9	1.63	0.123	0.577	1.154	0.476	2.784
CSP_SMCP	neuro_progs	syn	2	1.7	1.15	0.521	0.832	1.664	-0.514	2.814
CSP_SMCP	neuro_progs	mis	1	3.8	0.261	0.978	0.263	0.526	-0.265	0.787
CSP_SMCP	neuro_progs	lof	1	0.5	1.84	0.419	2.000	4.000	-2.160	5.840
CSP_SMCP	neuro_progs	all	4	6.1	0.654	0.859	0.328	0.656	-0.002	1.310
CSP_SMCP	neuro_progs	prot	2	4.4	0.458	0.932	0.321	0.643	-0.185	1.101
CSP_SMCP	Pax9_Mesenchyme	syn	1	0.5	2.07	0.383	2.000	4.000	-1.930	6.070
CSP_SMCP	Pax9_Mesenchyme	lof	1	0.2	5.52	0.166	5.000	10.000	-4.480	15.520
CSP_SMCP	Pax9_Mesenchyme	all	2	1.8	1.14	0.525	0.786	1.571	-0.431	2.711
CSP_SMCP	Pax9_Mesenchyme	prot	1	1.3	0.783	0.721	0.769	1.538	-0.755	2.321

CHAPTER V: Rare variants in PRKCI cause Van der Woude syndrome and other features of peridermopathy

This chapter is adapted from a manuscript in preparation for submission for publication.

Kelsey Robinson, Sunil K. Singh, Wasiu Lanre Adeyemo, Terri H. Beaty, Azeez Butali, Carmen J. Buxó, David J. Cutler, Michael P. Epstein, Brooklynn Gasser, Lord JJ Gowans, Jacqueline T. Hecht, Jeffrey C. Murray, Daryl Scott¹, Gary M. Shaw, Mary Ann Thomas, Lina Moreno Uribe, Seth M. Weinberg, Harrison Brand, Mary L. Marazita, Robert A. Cornell, Elizabeth J. Leslie

ABSTRACT

Van der Woude syndrome (VWS) is an autosomal dominant disorder characterized by lower lip pits and orofacial clefts (OFCs). With a prevalence of approximately 1 in 35,000 live births, it is the most common form of syndromic clefting and accounts for ~2% of all OFCs. The majority of cases are due to genetic variants in *IRF6* (~70%) or *GRHL3* (~5%), leaving up to 25% of VWS cases unsolved. Both *IRF6* and *GRHL3* work within the periderm, a single layer of epithelial cells that prevent pathological adhesions during palatogenesis. Disruption of this layer results in a spectrum of phenotypes ranging from lip pits to severe pterygia, facial clefts, and other congenital anomalies that are incompatible with life; thus, understanding the mechanisms of peridermopathies is vital in improving health outcomes for affected individuals. As there is a large gene regulatory network within the periderm during embryogenesis, other genes upstream or downstream of *IRF6* are excellent candidates to harbor mutations resulting in VWS. Here we identified 5 *de novo* mutations (DNs) and 14 rare variants in *PRKCI*, an atypical protein kinase C, in 16 individuals with clinical features consistent with peridermopathy and other phenotypes. Three variants (P351L, N383S, and L385F) were confirmed loss-of-function using zebrafish embryos, one of which (N383S) is an apparent hotspot mutation found in three individuals. A spectrum of phenotypes was observed for probands with loss of function variants including: isolated cleft lip and palate; VWS with cleft soft palate and lip pits; CP, global hypotonia, developmental delays, and autism; and cleft soft palate, sygnathia, ankyloblepharon, elbow and knee contractures, and an atrial septal defect. These findings highlight *PRKCI* as a novel candidate gene in the periderm gene regulatory network and expand our understanding of the genetic basis of VWS and other peridermopathies.

INTRODUCTION

Van der Woude syndrome (VWS, OMIM:119300) is an autosomal dominant disorder that occurs in approximately 1 in 35,000 live births (88, 89). It is commonly characterized by bilateral lower lip pits and an orofacial cleft (OFC) (90, 91). Approximately 70% of VWS is caused by variants in *IRF6*, including whole gene deletions (225) and structural variants(226), missense variants, and truncating variants (93-95). Approximately 5% of VWS is caused by missense or truncating variants in *GRHL3* (96), leaving up to 25% of cases unsolved. While all families with VWS may present with lip pits and/or a cleft lip with or without a cleft palate (CL/P) or a cleft palate only (CP), there is some evidence for a genotype-phenotype correlation: those with *GRHL3* variants are more likely to have CP as compared to *IRF6* (96). However, the unsolved families span the phenotypic spectrum of OFCs and lip pits. Gene discovery for Mendelian conditions, including VWS, is made easier with large families, regions of significant linkage, and low genetic heterogeneity. In the absence of these, gene discovery can be tedious unless new causal genes can be easily tied to known biological pathways.

IRF6 and *GRHL3* are conserved transcription factors that regulate epidermal differentiation. VWS, fundamentally, is caused by disruption of the periderm, a single epithelial cell layer that lines the oral cavity and other structures during embryonic development, preventing pathologic adhesions in early embryogenesis (12). Disruption of this cell layer can lead to many clinical features ranging from mild to incompatible with life. VWS belongs to a group of conditions, now referred to as peridermopathies, that arise due to mutations in genes essential for periderm development and function. Peridermopathies include popliteal pterygium syndrome (PPS, OMIM:119500), caused by mutations in *IRF6* that result in lower lip pits, OFCs, sygnathia, pterygia (webbing) of the lower limbs, syndactyly, and urogenital abnormalities (97, 98). More

severe manifestations of peridermopathies include Bartsocas-Papas syndrome (BPS; OMIM 263650) and Cocoon syndrome (OMIM 613630) caused by recessive mutations in *RIPK4* (99, 227) and *CHUK* (also called *IKKa*) (100), respectively.

The genes mutated in peridermopathies are part of a complex gene regulatory network (GRN) and signaling cascade that begins with activation via phosphorylation of *RIPK4* by protein kinase C (PKC) (228). *IRF6* is phosphorylated by *RIPK4* (228), which then activates *GRHL3* expression (229). Parallel to this signaling cascade, *RIPK4* also phosphorylates *CHUK* and *IKKb* (230), resulting in activation of NF- κ B signaling which plays a role in regulating programmed cell death at the appropriate time (231). However, up to a quarter of individuals with VWS have no identified mutation in *IRF6*, *GRHL3*, *RIPK4*, or *CHUK* but share phenotypes with the conditions associated with this network of genes. Other genes upstream or downstream of *IRF6* are therefore excellent candidates to harbor mutations resulting in VWS.

Although the extent of the genes in this GRN have yet to be elucidated, we suggest a new member of this network, *PRKCI* for which we report variants as causal for VWS and other clinical features of peridermopathies. *PRKCI* encodes for atypical protein kinase C (aPKC) iota (PKC ι), a serine/threonine protein kinase belonging to the protein kinase C family. We identified exome-wide enrichment of *de novo* variants (DNs) in *PRKCI* associated with syndromic CP, including a hotspot mutation found in three individuals (N383S), as well as several rare, protein-altering, predicted damaging variants. We tested the functional effects of 8 variants using a zebrafish model and found that several of these result in a loss of function, leading to peridermopathy.

METHODS

Study cohort

VWS: Families for this study consisted of 16 families diagnosed with VWS with no known mutation in *IRF6* or *GRHL3* based on bidirectional Sanger sequencing of both genes. Ethical approval and oversight was provided by regional care centers where recruitment, consent, and phenotyping occurred and at the University of Iowa.

OFCs: We also analyzed two large OFC cohorts: one ascertained primarily on CP, which will be referred to as CPseq, and the other ascertained on any OFC as part of the Gabriella Miller Kids First (GMKF) Research Consortium. For CPSeq, a total of 435 trios with CP were recruited from diverse ancestral backgrounds including European (recruited from Spain, Turkey, Hungary, United States), Latin American (Puerto Rico, Argentina), Asian (China, Singapore, Taiwan, the Philippines), and African (Nigeria, Ghana). For GMKF, case-parent trios originate from three ancestral groups: 376 trios of European (recruited from sites in the United States, Argentina, Turkey, Hungary, and Spain), 267 trios from Medellin, Colombia, and 116 trios from Taiwan. In total, there were 1,511 affected individuals from both CPSeq and GMKF broken down into 608 CP, 897 CL/P, and 3 unknown OFC type. Phenotyping, recruitment, and details of local ethical approval were previously described for CPSeq (173) and GMKF (125). All recruitment was approved by local review boards and/or coordinating centers at the University of Iowa, the University of Pittsburgh, Johns Hopkins University, or Emory University.

Sample preparation and whole genome sequencing (WGS)

VWS and CPSeq cohorts: WGS was performed at the Center for Inherited Disease Research (CIDR) at Johns Hopkins University (Baltimore, MD). Alignment, variant calling, and quality control was performed using the DRAGEN Germline v3.7.5 pipeline on the Illumina BaseSpace Sequence Hub platform, producing single sample VCF files. Details on the full pipeline for sequencing have been published in Robinson et al (173).

DNs were called using a separate pipeline than the full cohort multisample VCF. First, the DRAGEN 3.7.5 aligner and variant caller was used to generate gVCF files for each trio. The gVCF files along with a pedigree file for each trio were provided as input to the DRAGEN 3.7.5 joint caller, resulting in a trio specific VCF file including tags for potential de novo variants. Genotypes were set to missing if $GQ < 20$ or read depth < 10 . To be considered a DN, variants had to have a quality score of 30, $DQ > 2$, and parental genotypes had to be confirmed homozygous reference (0/0), pass all filtering steps, and have an allele balance (AB) ratio of < 0.05 . Only *de novo* variants with MAF of $< 0.5\%$ in either gnomAD exomes v 2.1.1 or gnomAD v3.1.2 were retained.

GMKF: WGS was performed from blood samples in the majority of cases, but saliva was used in cases where blood was unobtainable. Sequencing for European samples was carried out by the McDonnell Genome Institute (MGI) the Washington University School of Medicine (St. Louis, MO) followed by alignment to hg38 and variant calling at the GMKF Data Resource Center at the Children's Hospital of Philadelphia. Sequencing for Colombian and Taiwanese samples was carried out by the Broad Institute, with alignment to hg38 and variant calling by GATK pipelines(204-206). Additional details on alignment and workflow used to harmonize these datasets have been published in Mukhopadhyay et al (232).

Variant filtering and annotation

Variants with a passing filter flag were included. Next, we removed variants aligning outside of the standard chromosomes (1-22, X, Y) and with a minor allele count (MAC) of <1. Genotypes with quality scores of <20 or read depths of <10 were set to missing, and sites with missingness values of >10% were subsequently removed. Sample-level quality control metrics included transition/transversion (Ts/Tv) ratio, silent/replacement rate, and heterozygous/homozygous ratio; outlier samples with values outside of 3 standard deviations from the cohort mean were discarded. Samples with high missing data (>5% missing) were removed.

All variants were annotated with Annovar (version 201707). Variants with a MAF of <0.5% in any population in either gnomAD exomes v 2.1.1 or gnomAD v3.1.2 were considered rare for this analysis. We retained protein-altering variants, removing any annotated as synonymous. We estimated predicted pathogenicity with the *in silico* tools SIFT, PolyPhen2, and MutationAssessor.

DN enrichment

To determine enrichment of DNs in *PRKCI* across all 475 CP trios, we used the ‘DenovolyzeByGene’ function of the R package ‘DenovolyzeR’ package (version 0.2.0) (207). Gene were considered significantly enriched at $p < 8.74 \times 10^{-5}$ (multiple test correction for 572 genes), with a more conservative exome-wide significance threshold of $p < 1.3 \times 10^{-6}$.

Functional validation in zebrafish

We performed overexpression assays in transgenic zebrafish in which enveloping layer (EVL) cells were labeled with green fluorescent protein (GFP) using a *KRT14*-GFP tag. Activated

(CAAX) versions of *PRKCI* were cloned in pCS2+ plasmids to generate mRNAs for testing. Zebrafish embryos at the 1-2 cell stage were then injected with 100pg mRNA per embryo of either human reference *PRCKI* or variant alleles. Images of GFP signal in un-injected and injected embryos were acquired at 22 hours post fertilization (hpf).

RESULTS

We performed WGS on 17 families diagnosed with VWS but negative for mutations in *IRF6* or *GRHL3* based on bidirectional Sanger sequencing of both genes. We found two individuals with identical DNs in *PRKCI* (N383S) and matching phenotypes: lip pits and cleft soft palate. Because VWS can present with varied phenotypic features, we expanded our search to a trio-based cohort ascertained on CP and identified a third individual with a DN in *PRKCI* (Y136C). This individual had a cleft soft palate, as well as abnormal lip morphology but had not received a confirmed diagnosis of lip pits or VWS. Using denovolyzerR, we then confirmed that there is a statistically significant enrichment of protein-altering DNs in *PRKCI* in this cohort (n=475 trios, 1.71×10^{-6}), and exome-wide enrichment when restricted to syndromic cases alone (n=33 trios, 1.16×10^{-6}).

Using additional clinical databases such GeneMatcher (233), we identified two more DNs in *PRKCI*, including a third N383S mutation. This individual presented with multiple clinical features including CP (subtype unspecified), global hypotonia, developmental delays, and autism. The other identified DN was only two base pairs away (L385F). The clinical features in this individual were consistent with more severe signs of peridermopathy, including a cleft soft palate, sygnathia (adhesions between the upper and lower jaw), ankyloblepharon (adhesions between the upper and lower eyelids), elbow and knee contractures, and an atrial septal defect.

Based on these findings, we expanded our search to all rare variants (MAF <0.5%) in *PRKCI*, regardless of inheritance pattern. Because VWS and other peridermopathies exhibit variable expressivity and can manifest with either CP or CL/P, we looked for variants in 1,491 affected individuals from a combined cohort ascertained on any OFC (598 CP, 888 CL/P, and 3 unknown OFC type). We found 10 additional protein-altering variants in affected probands from this cohort, and another 4 variants of interest from other clinical databases (**Table 5.1, Figure 5.1A**). There were 12 unique missense variants, with R480H occurring twice, and one nonsense variant in the second to last exon (Q526X). Patients harboring these variants had clinical features across a wide phenotypic spectrum, ranging from isolated OFCs to cases including combinations of intrauterine growth restriction, congenital heart defects, and autism or developmental delays with and without an OFC.

PRKCI encodes atypical protein kinase C (aPKC) iota (PKC_ι), a serine/threonine protein kinase. It is expressed relatively ubiquitously, though some tissues have stronger expression than others. For example, in craniofacial tissue from human embryos at Carnegie Stage 20, the highest expression is in the ectoderm, the origin cells for the periderm (**Figure 5.1A**) (127). In mice, whole embryo *in situ* hybridization illustrates *Prkci* expression throughout the body (234), including in the palate (**Figure 5.1B**), and an snRNAseq dataset specifically from the secondary palate shows *Prkci* as an epithelial cell marker gene at embryonic day 15.5 (130).

PKC_ι has three domains: Phox/Bem 1 (PB1), atypical C1, and the protein kinase catalytic domain (**Figure 5.1C**). The PB1 domain is important for guiding cell polarity (235), the PB1 and C1 domains together are considered the regulatory region (236), and the kinase domain is responsible for catalytic activity. We did not see any clear clustering of variants based on domain, with 4 variants falling outside of any defined region, 1 variant in the PB1 domain, and 11 within

the catalytic domain, which is also the largest. However, both the N383S and L385F DN are at ATP-binding sites with the catalytic region (237, 238) which may directly interfere with normal kinase function (239).

Using a combination of *in silico* pathogenicity predictors and phenotype information, we chose 8 variants to test in zebrafish embryos to determine the effects of genetic variation on PKC ι function in the enveloping layer (EVL), a structure analogous to the periderm (**Figure 5.3A, B**). These zebrafish transgenic with a *krt14:gfp* tag as an EVL marker, allowing visualization of this cell layer. First, we found aPKC activity is necessary to induce the EVL, and that constitutively active aPKC induces ectopic EVL. We injected 1-cell stage zebrafish embryos with mRNAs encoding constitutively active forms of the human PKC ι reference allele and the following variants: all DN (Y136C, N383S, L385F), four rare variants (R130H, P351L, A421G, G581V), and a control variant we predicted would be benign as it was only found in an unaffected parent (V186I).

Using this system, the reference PKC ι (**Figure 5.3C, D**) and control variant (V186I) induced ectopic EVL, indicating these variants retained function, as expected. Although predicted to be deleterious by *in silico* methods, R130H, Y136C, A421G, and G581V also induced ectopic EVL (**Figure 5.3E, F**). In contrast, two DN alleles (N383S and L385F), and one rare variant allele (P351L) did not induce ectopic periderm or otherwise disrupt zebrafish development, indicating these variants are loss-of-function (**Figure 5.3G, H**). These findings suggest that patient variants P351L, N383S, and L385F disrupt the periderm, and are causal for a spectrum of peridermopathy phenotypes.

DISCUSSION

In this study, we initially conducted WGS on 17 families diagnosed with VWS but negative for mutations in *IRF6* or *GRHL3*, known causative genes for VWS. We identified two individuals harboring identical DN (N383S) in *PRKCI*, both with VWS characterized by lip pits and cleft soft palate. Expansion of our investigation of *PRKCI* into additional cohorts revealed a third individual with an N383S DN, two further DN (Y163C and L385F), and 13 rare, protein-altering variants. Functional validation of select *PRKCI* variants in zebrafish embryos demonstrated loss-of-function effects, providing evidence for its role in the pathogenesis of cleft palate and related peridermopathies.

PRKCI is an aPKC, and unlike conventional protein kinase C, aPKCs are Ca^{2+} and diacylglycerol (DAG) independent. Instead, they contain Phox/Bem 1 (PB1) domains that play a role in regulating cell polarity (235). Other functions associated with PKC ι have largely been studied in the context of cancer (224), finding a variety of roles in cell differentiation, migration, and proliferation (223). However, there are several studies on *PRKCI* during embryogenesis. Complete knockout of *Prkci* in mice is embryonic lethal by day E9.5 due to failed cavitation from polarity defects (240). *In situ* hybridization has shown *Prkci* is expressed in many tissues throughout embryogenesis, including regions of the brain, olfactory epithelium, liver, lungs, heart, stomach, and kidneys (234). Although the authors did not specifically investigate the palate, *Prkci* expression in whole embryo images is apparent in the palate (**Figure 5.1B**). In studies specific to the palate, *Prkci* was included in a set of epithelial marker genes in snRNAseq derived from the secondary palate at the time of osteogenesis in mice (130). In a similar dataset of snRNAseq from craniofacial tissue from human embryos at Carnegie Stage 20 (approximately gestational week 7), *PRKCI* expression was highest in the ectoderm (127). These data show that, during typical

development, *PRKCI* is expressed in relevant tissues (*i.e.*, the periderm) at the time of early secondary palatogenesis in humans and during the osteogenic phase in mice.

Given the enrichment of DNs in *PRKCI*, its varied roles, and established expression patterns, we performed functional testing of PKC ζ using zebrafish embryos. We found that PKC ζ is necessary to induce the EVL in zebrafish, as has been reported for *irf6* (241) and *poky/chuk/ikk1* (242). Further, constitutively active PKC ζ induced ectopic periderm similar to what has been reported for *grhl3(116)*. Together, these results suggest that *prkci* functions within this same pathway, though further characterization of its role is needed. We tested the effects of 8 *PRKCI* variants. One variant (R130H) was from a family with an affected proband and mother, yet it was inherited from the unaffected father. Based on this inheritance pattern, we suspected this variant would be functional and the zebrafish studies agreed with this prediction. However, it is worth noting that R130H has been predicted to diminish PKC ζ stability (243) may still be a hypomorphic allele or mediating effects in combination with another variant in an oligogenic manner: additional follow up on the impact of this variant is needed.

Three of the variants tested were found to be loss-of-function: P351L, N383S, and L385F. P351L was found in an isolated case of CLP and, as it was inherited from an unaffected parent, could represent reduced penetrance of disease. N383S was of particular interest as it was found as a DN in three people. Two had identical phenotypes of VWS with lip pits and a cleft soft palate, and the third had an unspecific CP subtype, as well as global hypotonia, developmental delay, and autism. Although appearing as a hotspot mutation, it is unclear what may be driving the occurrence of this variant. This area is not GC-rich nor highly repetitive; thus, the common mechanisms of mutation (244) are unlikely factors at this residue. Only two base pairs away, we also identified an L385F DN in an individual featuring a cleft soft palate, sygnathia, ankyloblepharon, elbow/knee

contractures, and an atrial septal defect. Taken together, these variants are implicated in VWS and other peridermopathy.

The remaining tested variants, Y136C, A421G, and G581V were found to retain function. One reason for this may be that these variants are, in fact, benign in nature and an incidental finding in this cohort. Alternatively, the mechanism by which we are testing functionality may not capture hypomorphic variants, in which there is enough activity to induce ectopic EVL but the reduction in protein may still lead to disruption of the periderm clinically. Further, there may be human-specific effects that we cannot detect in zebrafish, or these patients may have other modifier variants ultimately resulting in our observed phenotype.

Specific to Y136, there is *in vitro* evidence this tyrosine plays an important role in the PKC ζ regulatory module and membrane binding, with Y136E substitutions leading to a reduced level of the PKC ζ kinase domain presumably due to a decreased thermostability of the protein (245). Because our model includes a membrane-guided tag, we may be unable to capture effects secondary to defective membrane binding in this assay. Similarly, G581V is predicted to destabilize and alter the protein dynamics of PKC ζ (243), so further testing of this variant is also warranted. Lastly, although we did not test R480H, this variant was found in our dataset in an individual with an isolated CP, as well as in an individual with severe intrauterine growth restriction and multiple congenital heart defects. It has also been noted as a somatic mutation in cancers. Although R480C is more common as a somatic variant, it results in a loss of substrate specificity (224). It remains to be seen if R480H exerts similar effects. Provided the evidence in the literature and contradictory results in our zebrafish assay, additional investigation under alternative conditions is needed to better characterize the effects of these variants. However, this

may be limited to *in vitro* modeling or in conditional approaches due to the early lethality in murine models.

There are limitations to our study. One is the age at which probands ascertained. Although we have many presumed isolated OFCs, these are often recruited at a very young age—neurodevelopmental or behavior concerns may become apparent at a later timepoint and ultimately represent a syndromic case. In the context of *PRKCI* variants, their role in individuals' risk for both OFCs and developmental disorders remains to be elucidated. Lastly, our functional testing method is sensitive to variants with full loss-of-function but does not allow for measurement of more nuanced effects; thus, further validation of these findings and functional testing of additional variants is warranted to better understand their effects.

Although this study was focused on the role of *PRKCI* in VWS, we identified individuals with clinical features outside of the spectrum of peridermopathies, including intrauterine growth restriction, developmental disorders, autism spectrum disorder, congenital heart defects, global hypotonia, and feeding difficulties. As *PRKCI* is expressed in many tissues, it is plausible that variants within this gene could manifest deleterious effects elsewhere in the body; however, the mechanism by which this may occur remains to be seen.

When taken altogether, we provide evidence that variants in *PRKCI* can be causal for VWS and other features of peridermopathies. The 17 VWS families considered in this study are those left unsolved from previous sequencing studies of *IRF6* and *GRHL3* and are estimated to account for ~15-20% of Van der Woude syndrome families. The N383S variant alone therefore accounts for ~2.5% of VWS (12.5% of the estimated 20% unsolved VWS). Future studies expanding into other genes within the periderm GRN may also identify new candidates for the remaining unsolved families.

FIGURES

Figure 5.1: Expression of *PRKCI* in embryogenesis. A) Expression by snRNAseq cell cluster from human craniofacial tissue at Carnegie Stage 20 (http://cotneyweb.cam.uchc.edu/craniofacial_cs20/). B) Darkfield image of E14.5 ISHs for PKC ι/λ embryonic stage E14.5. PKC ι/λ is widely expressed throughout the embryo (di, diencephalon; in, intestine; mb, midbrain; mgt, midgut; mo, medulla oblongata; sp, spinal cord). The yellow arrows highlight the palate and apparent PKCi expression.

Figure 1B is adapted with copyright permission from *Elsevier* under license 5730860890376: Judit Kovac, Henrik Oster, Michael Leitges. Expression of the atypical protein kinase C (aPKC) isoforms ι/λ and ζ during mouse embryogenesis, *Gene Expression Patterns*. Volume 7. Issues 1–2, 2007, Pages 187-196. ISSN 1567-133X.

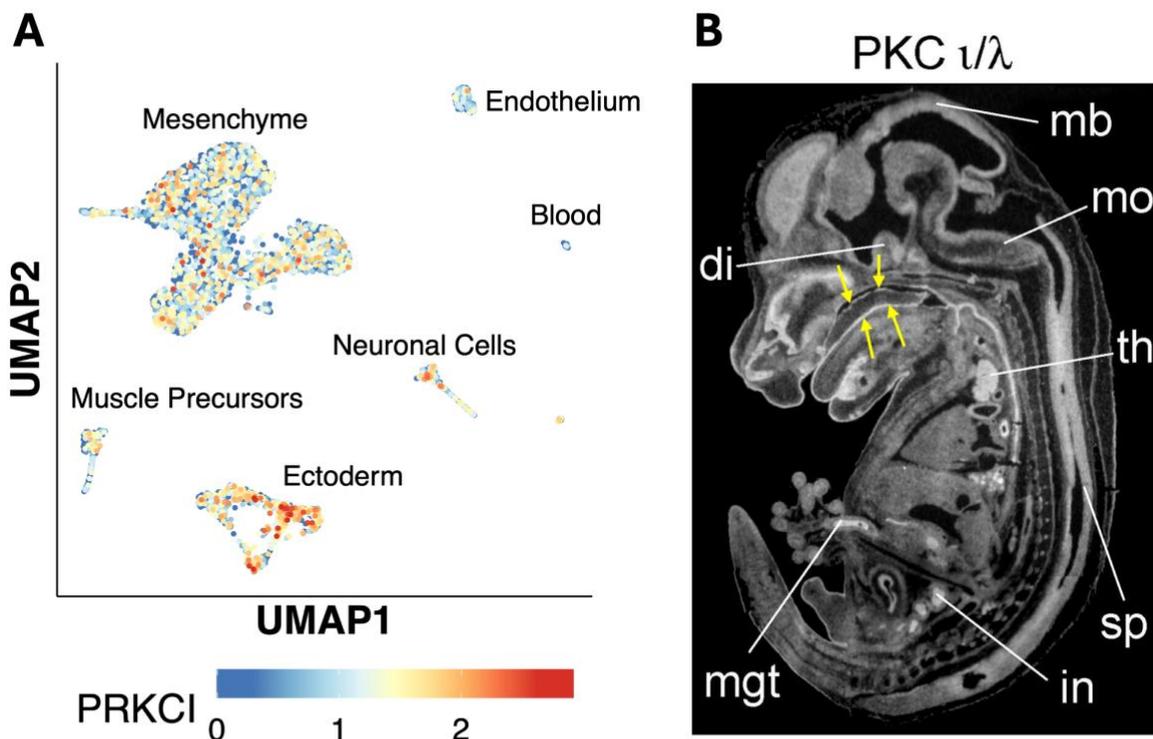


Figure 5.2: Structure and distribution of PRKCI variants. A) Lollipop plot of rare *PRKCI* variants in affected individuals across the linear structure. The number of variants per site is indicated by the number in the circle, where a solid circle indicates a single variant. Lollipops are colored by a combination of predicted pathogenicity and functional testing results: green=variant was functional and predicted functional, yellow with green circle=variant was found functional but predicted deleterious, red=variant was nonfunctional and predicted deleterious, gray=variant not tested. B) Representation of the PKC ζ regulatory module as predicted by AlphaFold Colab (best scoring model). The domains and regions are indicated and colored; from N-C PB1 domain (cyan), pseudosubstrate (PS) region (dark purple), beta-strand linker (BSL) (light purple), C1 domain (blue). C) Representation of the full PKC ζ structure as predicted from I-TASSER. PB1 domain (red), C1 (blue), protein kinase domain (pink), unspecific domain (green).

Figure 2B used with copyright permission from *Elsevier*:

Control of atypical PKC ζ membrane dissociation by tyrosine phosphorylation within a PB1-C1 interdomain interface. Cobbaut, Mathias et al. *Journal of Biological Chemistry*, Volume 299, Issue 7, 104847.

Figure 2C used under the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>):

Shah, H., Khan, K., Khan, N. *et al.* Impact of deleterious missense *PRKCI* variants on structural and functional dynamics of protein. *Sci Rep* **12**, 3781 (2022).

<https://doi.org/10.1038/s41598-022-07526-4>

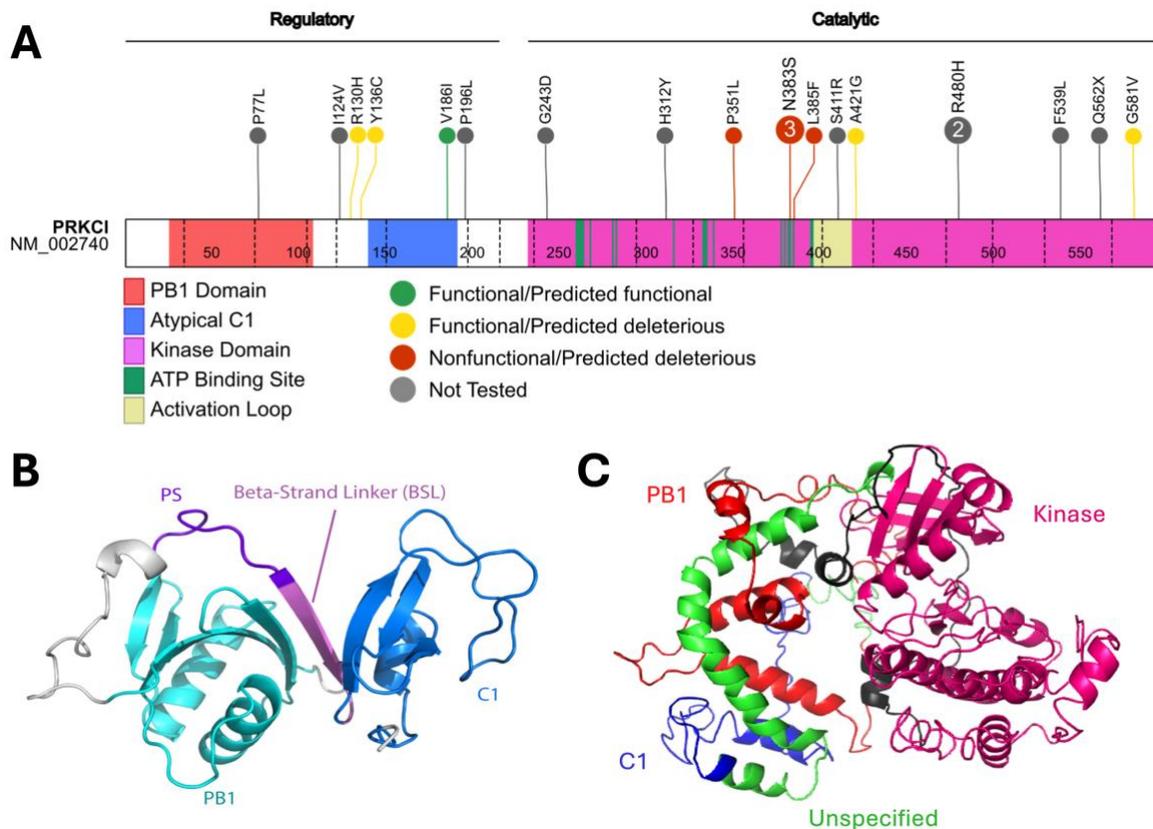
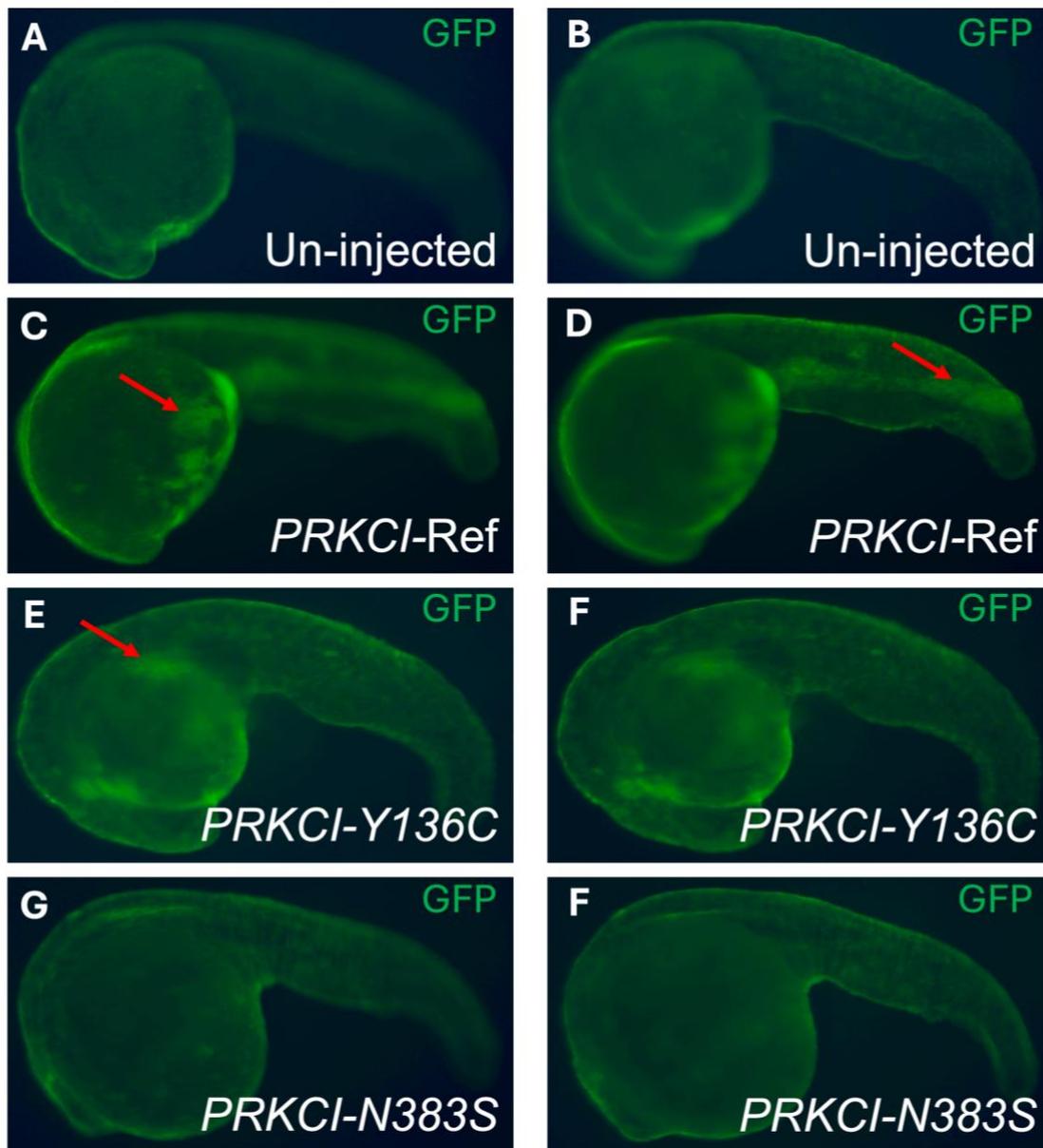


Figure 5.3: Representative zebrafish embryo assays for functional and non-functional PRKCI variants. A) Cranially and B) caudally focused images of an un-injected embryo with typical enveloping layer (EVL) development. C) Cranially and D) caudally focused images of an embryo injected with human reference PRKCI mRNA. E) Cranially and F) caudally focused images of an embryo injected with the PRKCI Y136C mRNA. G) Cranially and H) caudally focused images of an embryo injected with the PRKCI N383S mRNA. EVL is denoted by KRT14-GFP expression. Red arrows indicate ectopic EVL.



TABLES

Table 5.1: DN and rare variants in <i>PRKCI</i>										
Variant (CHR:BP)	CDS*	AA*	Phenotype	MAF**	MAC**	SIFT	Polyphen2	Mutation Assessor	Function	Supporting Literature
Confirmed <i>de novo</i> variants										
3:170267957	A407G	Y136C	Cleft soft palate	-	0	0.002	0.976	3.02	F	Y136 is a conserved Tyr and plays an important role in membrane binding(245)
3:170284541	A1148G	N383S	VWS, cleft soft palate	8.80E-06	1	0	0.999	4.04	LOF	
			VWS, cleft soft palate	8.80E-06	1	0	0.999	4.04	LOF	
			CP, global hypotonia, DD, autism	8.80E-06	1	0	0.999	4.04	LOF	
3:170284548	A1155C	L385F	Cleft soft palate, sygnathia, ankyloblepharon, elbow/knee contractures, atrial septal defect	-	0	0.001	1	2.2	LOF	
Rare variants (inherited from unaffected parent unless noted)										
3:170259975	C230T ¹	P77L	Autism, speech delay, elbow hyperextension	6.56E-05	1	0.086	0.998	2	NT	
3:170267920	A370G	I124V	LCLP	6.04E-05	2	0.282	0.021	1.925	NT	

3:170267939	G389A ²	R130H	BCLP	8.82E-05	6	0.031	0.983	2.84	F	Predicted to be highly deleterious(243)
3:170270557	C587T	P196L	Abnormal ultrasound findings: growth delay, intrauterine growth restriction, OFC, and ventricular septal defect	-	0	0.082	0.005	1.79	NT	
3:170280249	G728A	G243D	BCLP	1.24E-04	14	0.374	0.079	1.825	NT	
3:170281217	C934T ³	H312Y	CP	9.65E-05	4	0.134	0.165	3.3	NT	
3:170281953	C1052T	P351L	LCLP	-	0	0.027	0.999	1.615	LOF	
3:170291881	A1231C ⁴	S411R	BCLP	5.44E-05	1	0.266	0.908	1.095	NT	
3:170291912	C1262G	A421G	CLP	-	0	0.001	0.999	3.7	F	
3:170295932	G1439A	R480H	CP	-	0	0.469	0.04	1.42	NT	The recurrent somatic mutation R480C in cancer affects substrate specificity (LLGL2)(224, 246)
			Severe intrauterine growth restriction, oligohydramnios, premature birth, hypoplastic left heart, aortic valve atresia, mitral valve atresia, hypoplastic aortic arch, and upslanted palpebral fissures	-	0	0.469	0.04	1.42	NT	
3:170299022	T1615C ¹	F539L	Autism, global DD, mostly nonverbal, toe walking, not able to run	1.18E-04	8	0.001	0.001	4.115	NT	

			or jump, chronic constipation, poor suck, feeding difficulties							
3:170299091	C1684T	Q562X	CP	-	0	.	.	.	NT	
3:170303078	G1742T	G581V	LCLP	2.94E-05	2	0	1	3.51	F	Predicted to be highly deleterious(243)
Control rare variant in unaffected parent										
3:170270526	G556A ¹	V186I	None	1.94E-04	1	0.045	0.026	1.57	F	
<p>*All HGVS changes are for NM_002740 **gnomAD exomes v2.1.1 or gnomAD genomes v3.1.2</p> <p>¹Inheritance data not available ²From unaffected father (mother is affected) ³From unaffected mother (father is affected) ⁴From unaffected father (father is homozygous)</p> <p>F=functional, LOF=loss-of-function, NT=not tested</p> <p>BCLP=bilateral cleft lip and palate, LCLP=left cleft lip and palate, RCLP=right cleft lip and palate, CP=cleft palate, DD=developmental delay, OFC=orofacial cleft</p>										

CHAPTER VI: Discussion

Despite being a common and highly heritable congenital anomaly, the genetic etiology of cleft palate (CP) is poorly understood. Although there are dozens of loci associated with the related—but etiologically distinct—cleft lip with or without cleft palate (CL/P), CP has been understudied in comparison. The CPSeq study was created to address this gap in research and resulted in the assembly of whole genomes for over 400 case-parent trios ascertained on CP and originating from diverse backgrounds. Using this cohort, we examined common and rare variant associations for all CP probands and by various categories such as CP subtype, proband sex, and syndromic classification. Taken altogether, the work presented here highlights the vast heterogeneity underlying the genetic architecture of CP.

Summary of results

In our common variant analysis, there were no genome-wide significant loci for the full cohort of CP trios, but when split into subtypes, we found a significant association of the 9q33.3 locus with any cleft hard palate (CHP). We found *Angptl2*, one of the genes at this locus, is expressed in the palatal shelves in mice at embryonic day (E) 13 and E14.5, and that its expression pattern mimics that of *Runx2*, a well-known regulator of osteogenesis (1). Although there were no other genome-wide loci in the analysis stratified by CP subtype, we examined the odds ratios (ORs) for all loci reaching suggestive significance across CP subtypes. In loci identified in the full cohort of all types of CP, ORs were consistent in their effects on risk in each of the subtypes (*i.e.*, the overall OR was similar for all CP, CHP, and any cleft soft palate (CSP)). In contrast, when comparing the ORs of loci identified in either of the subtype analyses there were differences in the magnitude of risk. For example, the 9q22.31 reached suggestive significance in

the CSP group with OR 4.27 (2.22-8.24, $p=2.28 \times 10^{-6}$), whereas this locus was not significant in CHP ($p=0.53$) with OR 1.30 (0.57-2.97). Although overlapping confidence intervals preclude strong conclusions about distinct risks, these findings suggest that certain genotypes may predispose an individual to a specific subtype of CP.

We also investigated sex-specific risks among common variants. Again, there were no genome-wide significant findings, but there were 13 loci significant for gene-by-sex interactions. Notably, all detected interactions conveyed opposite effects in males and females. The most significant finding was an increased risk for female probands for SNPs in *LTBP1* at the 2p22.3 locus (OR 3.37 (2.04 - 5.56), $p=1.93 \times 10^{-6}$). Further, we identified a female-specific significant association for *LTBP1* in cultured fibroblast cells using imputed genetically regulated gene expression. As *LTBP1* is involved in TGF β signaling, a key pathway in palatogenesis (2, 3), these results suggest it is a plausible candidate gene for CP. Although we were underpowered to detect gene-by-sex interactions with different magnitudes of effect for males and females, this study highlights the importance of investigating sex-specific effects to understand the genetic underpinnings of CP more comprehensively.

In summary, our common variant analyses identified several loci not previously reported for CP, including two candidate disease genes: *ANGPTL2* and *LTBP1*. These studies demonstrate the importance of investigating datasets inclusive of all affected individuals, which can reveal shared or smaller effects due to the benefits of larger sample sizes, as well as the benefits of subdividing into specific categories to better understand subtype-specific risks. However, despite having reasonable power to identify genetic associations, the results from these studies do not fully account for the level of genetic influence we know exists for CP from epidemiology and

family studies. Thus, we next investigated the contribution of rare variation, with a focus on protein-coding *de novo* variants (DNs).

Using a cohort of 475 CP trios, we identified 597 protein-coding DNs in 572 genes. The CP probands showed a significant enrichment of protein-altering variants (1.34, $p=2.18 \times 10^{-9}$), driven by both missense (1.29, $p=6.57 \times 10^{-7}$) and putative loss of function (LOF) variants (1.64, $p=1.02 \times 10^{-4}$). Gene-specific analysis found significant enrichment in LOF DNs for *COL2A1* and *MYH3*, and in protein-altering DNs for *IRF6*, *PRKCI*, *SAT2B*, *POLR1F*, and *SLC25A41*.

A key takeaway from this investigation was the overlap of presumed isolated CP in genes commonly associated with CP syndromes. Although this has been documented for *IRF6* and *GRHL3*, there is less, if any, evidence for isolated CP associated with *COL2A1*, *MYH3*, and *SATB2*. However, each of these genes remained significant when syndromic probands were removed from our analysis, and none of the individuals harboring DNs in these genes were reported to have additional clinical features. To explore this further, we compared our dataset to a list of pathogenic/likely pathogenic DNs from syndromic CP probands in the Deciphering Developmental Disorders (DDD) study, in which probands were ascertained on undiagnosed neurodevelopmental disorders and/or congenital anomalies. We identified 11 genes with protein-altering DNs that were common between the two datasets. Interestingly, among the overlapping genes, only two individuals with syndromic features were identified in the CPSeq: one harboring an *ARIDIA* missense variant and another with a frameshift deletion in *KAT6B*, with features resembling Roselli-Gulienetti Syndrome. In the DDD study, *SATB2* was the most prevalent DN, which is consistent with the expected enrichment of neurodevelopmental disorders. Conversely, genes like *IRF6* and *GRHL3*, both prominent in CPSeq, were not found among the P/LP variants in the DDD study. These comparisons highlight differences and similarities in the genetic

landscape between individuals with syndromic and isolated CP, although differences may also be expected between syndromic CP cohorts ascertained on different features or with different exclusion criteria. Further investigation in a large, unselected and well-phenotyped cohort could better characterize the genetic mechanisms underlying these conditions.

We also performed subtype-specific analyses comparing clefts of the soft and hard palate (CH&SP) and cleft soft palate (including submucous cleft palate, CSP+SMCP). We found *SATB2*, *TGFBR2*, and *COL2A1* were significantly enriched in the CH&SP subtype, while *PRKCI* and *COL2A1* were significantly enriched in the CSP+SMCP subtype. Further investigation of these enrichments using a single nucleus RNA sequencing dataset from the secondary palate in mice at E15.5 (the time of palatal osteogenesis) identified cell-specific enrichments also. CSP+SMCP probands were enriched in epithelial (driven by *PRKCI*) and endothelial cell clusters, whereas CH&SP probands showed nominal enrichment for late osteogenic progenitors (driven by *SATB2*) and muscle progenitors. These subtype-specific differences may indicate genotype-phenotype correlations or that there are cell type-specific disruptions leading to one subtype versus another; however, larger sample sizes are needed to investigate this further.

Lastly, as part of our investigation of DNs, we identified variants in *PRKCI* as a causal for Van der Woude syndrome (VWS) and other features of peridermopathy. There was exome-wide enrichment of DNs in *PRKCI* associated with syndromic CP cases ($p=1.16 \times 10^{-6}$), including a hotspot mutation found in three individuals (N383S), as well as several rare, protein-altering, predicted damaging variants. Functional testing in zebrafish embryos confirmed that some of these *PRKCI* variants, including P351L, N383S, and L385F, result in loss-of-function effects, implicating them as likely causal variants for some or all of the phenotypes in their respective

probands. However, the remaining variants retained function, indicating the need for further investigation into their effects and any association with an array of phenotypes, including clinical features of peridermopathies and other congenital anomalies.

In conclusion, investigation of DNs found a significant enrichment of protein-altering variants and identified three genes not previously associated with CP (*PRKCI*, *POLRIF*, and *SLC25A41*). Interestingly, this enrichment also included genes commonly associated with syndromic CP, despite the CPSeq cohort being enriched for probands with isolated CP. Subtype-specific analyses revealed gene and cell type enrichment in different CP subtypes, suggesting potential genotype-phenotype correlations. Lastly, variants in *PRKCI* were identified as causal for VWS peridermopathies, with some variants showing loss-of-function effects. Further investigation is needed to understand the effects of remaining variants and their association with various phenotypes.

Interpretation and implications

When considering this body of work as whole, these results represent a significant advancement in our understanding of the genetic basis of CP. They highlight the heterogeneity of CP etiology, demonstrating the influence of both common and rare genetic variants, sex-specific effects, and subtype-specific correlations.

Throughout our common variant analyses, we identified only a single genome-wide significant locus at 9q33.3 in the CHP subtype. Given previous efforts of well-powered studies to find CP-associated loci (4-11), we did not expect to find many loci and we hypothesized that this dearth of associated loci could be due to studying all types of CP as single phenotype, and that an analysis by subtype would be more rewarding. We were able to find mostly suggestive

associations, in part due to the even smaller sample sizes that result from stratifying the cohort, and the single genome-wide significant 9q33.3 locus; this suggests that the sub-categorization effort did not significantly improve our ability to detect common variant associations. Similarly, we identified 13 loci with significant gene-by-sex interactions but failed to find any of genome-wide significance. Taken together, these results do not account for what we know about the heritability of CP; thus, it is likely that common variation is not a primary driver of the CP phenotype.

Irrespective of variant frequency, an observed pattern throughout these studies is the presence of both shared risks factors that increase the risk for any type of CP, and subtype-specific risk factors that are associated more specifically with a distinct phenotype. As palatogenesis is a complex process, it is not surprising that disruption of specific processes could plausibly result in specific phenotypes. In the context of our findings, for example, *SATB2* was specifically enriched for the CH&SP subtype. Consequently, CH&SP probands were enriched for the late osteogenic progenitor cell cluster, in which *SATB2* was expressed. It is reasonable to speculate that failure of these cells to progress into osteocytes could result in a cleft of the hard palate, which ultimately affects the ability of the soft palate to fuse together, culminating in CH&SP. Similarly, disruption of a gene expressed in many cell types could be more likely to result in any type of CP, as the variant effects may be non-specific. Therefore, continuing to study CP in the context of shared and subtype-specific risks can help identify new pathways of interest and provide insight into palatogenesis.

Another important finding throughout this work is the association of isolated CP with syndromic CP-related genes. This was apparent in genes enriched for DNs in our cohort, and when we compare CPSeq to the DDD study (12, 13), in which probands were primarily

ascertained on neurodevelopmental disorders, growth delays, and/or other congenital anomalies. From a gene discovery point of view, exploration of genes within shared molecular networks can serve as candidate genes for future studies. We demonstrate this with the discovery of *PRKCI* as causal for peridermopathies, seemingly working within the same pathways as *IRF6*, known to cause ~70% of VWS cases. Importantly, these findings also have potential clinical implications. Genetic testing is rarely offered for isolated CP unless there is a strong family history (14). However, if these individuals only outwardly present with CP, they may fail to receive necessary resources if they were to have subclinical or mild features of a syndrome, such as intellectual disability or other developmental delays that would benefit from early intervention. Alternatively, in individuals with inherited variants, a positive genetic test finding a variant in a known syndromic gene may lead to a change in clinical care or surveillance to identify of features in their relatives who otherwise may not receive a diagnosis. Further exploration of the spectrum of phenotypes between isolated and syndromic CP can lead to gene discovery and potentially improve diagnosis for affected individuals.

In summary, this work highlights the contribution of common and rare variants to CP, demonstrates the presence of broad and subtype-specific risks, and identifies a phenotypic spectrum across which CP can manifest. Although there remains much to detangle about the genetic causes of CP, this research significantly advances current knowledge in the field and lays a foundation for future investigation.

Study limitations

Although this research contributes substantially to our current knowledge of CP, there are some limitations. First, although we have assembled one of the largest trio-based cohorts for CP, we

still lack power in some analyses. For common variants, our capacity to detect signals are dependent on the effect size and minor allele frequencies—we have 80% power for SNPs with relative risk of approximately 2 or greater with minor allele frequencies >15%. Therefore, we may not discover variants with smaller effect sizes or those that are less common in this dataset.

Other limitations of this dataset are largely based around phenotyping and ascertainment information. For many of our participants, the age at which an individual is enrolled is unavailable and an unknown percentage of individuals were likely recruited in infancy or the first few years of life. This can lead to uncertainty around the classification of an isolated or syndromic occurrence of CP, as many of these children are too young to manifest clinical signs consistent with developmental delays or intellectual disabilities. In the absence of additional structural defects, they will be presumed isolated. We are fortunate to have deeper phenotyping, including CP subtypes, for 295 (62%) of our probands; however, that leaves 38% classified as “unknown subtype”. Ultimately the inability to classify these probands reduces our power for subtype-specific analysis. Future research including new recruitment of cohorts should consider this limitation (in the context of the limitations of phenotyping and the challenges of longitudinal studies) when designing study instruments and phenotyping protocols.

Lastly, there are limitations not unique to this cohort including reduced penetrance and variable expressivity of the phenotype. Although CP is largely associated with autosomal dominant inheritance, not all individuals harboring a causal variant will manifest clinical features (14), thus penetrance is not complete. Variable expressivity can result in misclassification of parent affection status if they manifest subclinical phenotypes. For example, submucous CP is often non-symptomatic and difficult to visualize, thus an individual may be considered unaffected unless a close inspection is performed. The presence of a bifid uvula is similar and

may not be documented as an abnormal phenotype as there are no clinical concerns. In the context of this study, misclassification of a parental phenotype would have minimal impact on our analyses. Because CP is likely multifactorial, parent affection status was not an exclusion criterion; however, when considering *de novo* or rare variants, parental affection status could influence interpretation genetic variants.

Future directions

Although the data presented here contribute substantially to the overall understanding of CP and genetic associations, there are many questions that have been generated and or remain unanswered. Firstly, much of our work has been performed using statistical methods associating phenotypes with variants; thus, functional validation of variant effects is needed to confidently establish causal relationships in newly reported genes.

Secondly, within our specific dataset, there are several avenues of research yet to be investigated. We have identified enrichment of protein-altering DNs in this cohort, but the contribution of all rare variants together is not known. In addition to protein-altering variants, the impact of synonymous variants resulting in cryptic splice sites may also be of interest, as this has been documented for *IRF6* in non-syndromic cleft lip and palate (15). Beyond the coding regions, exploration of non-coding DNs is crucial to fully understand the genetic architecture of CP, as variants in regulatory regions like enhancers, promoters, 5' UTRs, and 3' UTRs are increasingly being reported as disease-causing (16), yet little is known about their impact on CP.

Lastly, expanding studies to larger cohorts will enhance statistical power and allow for more robust subtype-specific analyses. With larger samples, a deeper dive into CP subtypes incorporating single cell and spatial transcriptomic data (presumably from mouse models) could

provide a more detailed understanding of the molecular mechanisms of palatogenesis; thus, enabling a better understanding of the development of specific phenotypes. Ultimately, the goal of the current and future work should be focused on application of these genetic discoveries into better prevention, diagnosis, and treatment for individuals with CP and their families.

Conclusion

In conclusion, this body of work identifies both common and rare variant associations with CP, finding sex-specific and subtype-specific effects. These findings highlight the vast genetic heterogeneity of CP etiology, while also establishing many new questions. Future work focused on rare coding and non-coding variation will be instrumental in better understanding the overall genetic architecture of CP. As sample sizes continue to grow, expansion of this understanding to include molecular mechanisms underlying CP subtypes in spatial and temporal context will also be invaluable. Ultimately, the aim of this work and future endeavors is to improve prevention, diagnosis, and treatment strategies for affected individuals and their families.

Chapter VII: REFERENCES

- 1 Mossey, P.A. and Modell, B. (2012) Epidemiology of oral clefts 2012: an international perspective. *Front Oral Biol*, **16**, 1-18.
- 2 Marazita, M.L. (2012) The evolution of human genetic studies of cleft lip and cleft palate. *Annu Rev Genomics Hum Genet*, **13**, 263-283.
- 3 Wehby, G.L. and Cassell, C.H. (2010) The impact of orofacial clefts on quality of life and healthcare use and costs. *Oral Dis*, **16**, 3-10.
- 4 Mai, C.T., Isenburg, J.L., Canfield, M.A., Meyer, R.E., Correa, A., Alverson, C.J., Lupo, P.J., Riehle-Colarusso, T., Cho, S.J., Aggarwal, D. *et al.* (2019) National population-based estimates for major birth defects, 2010-2014. *Birth Defects Res*, **111**, 1420-1435.
- 5 Mossey, P.A. and Catilla, E.E. (2003). World Health Organization, Geneva, in press.
- 6 Hanny, K.H., de Vries, I.A., Haverkamp, S.J., Oomen, K.P., Penris, W.M., Eijkemans, M.J., Kon, M., Mink van der Molen, A.B. and Breugem, C.C. (2016) Late detection of cleft palate. *Eur J Pediatr*, **175**, 71-80.
- 7 Saad, A.N., Parina, R.P., Tokin, C., Chang, D.C. and Gosman, A. (2014) Incidence of Oral Clefts Among Different Ethnicities in the State of California. *Annals of Plastic Surgery*, **72**.
- 8 Yow, M., Jin, A. and Yeo, G.S.H. (2021) Epidemiologic trends of infants with orofacial clefts in a multiethnic country: a retrospective population-based study. *Sci Rep-Uk*, **11**, 7556.
- 9 Burdi, A.R. and Silvey, R.G. (1969) Sexual differences in closure of the human palatal shelves. *Cleft Palate J*, **6**, 1-7.
- 10 Som, P.M. and Naidich, T.P. (2013) Illustrated review of the embryology and development of the facial region, part 1: Early face and lateral nasal cavities. *AJNR Am J Neuroradiol*, **34**, 2233-2240.
- 11 Mossey, P.A., Little, J., Munger, R.G., Dixon, M.J. and Shaw, W.C. (2009) Cleft lip and palate. *Lancet*, **374**, 1773-1785.
- 12 Richardson, R.J., Hammond, N.L., Coulombe, P.A., Saloranta, C., Nousiainen, H.O., Salonen, R., Berry, A., Hanley, N., Headon, D., Karikoski, R. *et al.* (2014) Periderm prevents pathological epithelial adhesions during embryogenesis. *J Clin Invest*, **124**, 3891-3900.
- 13 Jiang, R., Bush, J.O. and Lidral, A.C. (2006) Development of the upper lip: morphogenetic and molecular mechanisms. *Dev Dyn*, **235**, 1152-1166.
- 14 Song, J. (2007) EMT or apoptosis: a decision for TGF- β . *Cell Research*, **17**, 289-290.
- 15 Iseki, S. (2011) Disintegration of the medial epithelial seam: Is cell death important in palatogenesis? *Development, Growth & Differentiation*, **53**, 259-268.
- 16 Nasreddine, G., El Hajj, J. and Ghassibe-Sabbagh, M. (2021) Orofacial clefts embryology, classification, epidemiology, and genetics. *Mutation Research/Reviews in Mutation Research*, **787**, 108373.
- 17 Som, P.M. and Naidich, T.P. (2014) Illustrated review of the embryology and development of the facial region, part 2: Late development of the fetal face and changes in the face from the newborn to adulthood. *AJNR Am J Neuroradiol*, **35**, 10-18.
- 18 Jacobs, F.J. (2008) Variations of the isolated cleft of the hard palate: scientific. *South African Dental Journal*, **63**, 164-168.
- 19 Engelbrecht, H., Bütow, K.-W. and Botha, S.J.P. (2013) Exceptional isolated cleft of the hard palate. *British Journal of Oral and Maxillofacial Surgery*, **51**, e275-e276.

- 20 Reiter, R., Brosch, S., Wefel, H., Schlömer, G. and Haase, S. (2011) The submucous cleft palate: Diagnosis and therapy. *International Journal of Pediatric Otorhinolaryngology*, **75**, 85-88.
- 21 Feka, P., Banon, J., Leuchter, I. and La Scala, G.C. (2019) Prevalence of bifid uvula in primary school children. *Int J Pediatr Otorhinolaryngol*, **116**, 88-91.
- 22 Lan, Y. and Jiang, R. (2022) Mouse models in palate development and orofacial cleft research: Understanding the crucial role and regulation of epithelial integrity in facial and palate morphogenesis. *Curr Top Dev Biol*, **148**, 13-50.
- 23 Millicovsky, G., Ambrose, L.J. and Johnston, M.C. (1982) Developmental alterations associated with spontaneous cleft lip and palate in CL/Fr mice. *Am J Anat*, **164**, 29-44.
- 24 Forbes, D.P., Steffek, A.J. and Klepacki, M. (1989) Reduced epithelial surface activity is related to a higher incidence of facial clefting in A/WySn mice. *J Craniofac Genet Dev Biol*, **9**, 271-283.
- 25 Funato, N., Nakamura, M. and Yanagisawa, H. (2015) Molecular basis of cleft palates in mice. *World J Biol Chem*, **6**, 121-138.
- 26 Walker, B.E. and Fraser, F.C. (1957) The Embryology of Cortisone-induced Cleft Palate. *Development*, **5**, 201-209.
- 27 Carmichael, S.L. and Shaw, G.M. (1999) Maternal corticosteroid use and risk of selected congenital anomalies. *American Journal of Medical Genetics*, **86**, 242-244.
- 28 Paulson, R.B., Paulson, G.W. and Jreissaty, S. (1979) Phenytoin and carbamazepine in production of cleft palates in mice. Comparison of teratogenic effects. *Arch Neurol*, **36**, 832-836.
- 29 Abbott, B.D., Hill, L.G. and Birnbaum, L.S. (1990) Processes involved in retinoic acid production of small embryonic palatal shelves and limb defects. *Teratology*, **41**, 299-310.
- 30 Herkrath, A.P., Herkrath, F.J., Rebelo, M.A. and Vettore, M.V. (2012) Parental age as a risk factor for non-syndromic oral clefts: a meta-analysis. *J Dent*, **40**, 3-14.
- 31 Sabbagh, H.J., Hassan, M.H., Innes, N.P., Elkodary, H.M., Little, J. and Mossey, P.A. (2015) Passive smoking in the etiology of non-syndromic orofacial clefts: a systematic review and meta-analysis. *PLoS One*, **10**, e0116963.
- 32 Little, J., Cardy, A. and Munger, R.G. (2004) Tobacco smoking and oral clefts: a meta-analysis. *Bull World Health Organ*, **82**, 213-218.
- 33 Leite, M., Albiéri, V., Kjaer, S.K. and Jensen, A. (2014) Maternal smoking in pregnancy and risk for congenital malformations: results of a Danish register-based cohort study. *Acta Obstet Gynecol Scand*, **93**, 825-834.
- 34 Butali, A., Little, J., Chevriér, C., Cordier, S., Steegers-Theunissen, R., Jugessur, A., Oladugba, B. and Mossey, P.A. (2013) Folic acid supplementation use and the MTHFR C677T polymorphism in orofacial clefts etiology: An individual participant data pooled-analysis. *Birth Defects Res A Clin Mol Teratol*, **97**, 509-514.
- 35 Lorente, C., Cordier, S., Goujard, J., Aymé, S., Bianchi, F., Calzolari, E., De Walle, H.E. and Knill-Jones, R. (2000) Tobacco and alcohol use during pregnancy and risk of oral clefts. Occupational Exposure and Congenital Malformation Working Group. *Am J Public Health*, **90**, 415-419.
- 36 Leite, I.C. and Koifman, S. (2009) Oral clefts, consanguinity, parental tobacco and alcohol use: a case-control study in Rio de Janeiro, Brazil. *Braz Oral Res*, **23**, 31-37.
- 37 Bell, J.C., Raynes-Greenow, C., Turner, R.M., Bower, C., Nassar, N. and O'Leary, C.M. (2014) Maternal Alcohol Consumption during Pregnancy and the Risk of Orofacial Clefts in

Infants: a Systematic Review and Meta-Analysis. *Paediatric and Perinatal Epidemiology*, **28**, 322-332.

- 38 Romitti, P.A., Sun, L., Honein, M.A., Reefhuis, J., Correa, A., Rasmussen, S.A. and Study, N.B.D.P. (2007) Maternal periconceptional alcohol consumption and risk of orofacial clefts. *Am J Epidemiol*, **166**, 775-785.
- 39 De-Regil, L.M., Peña-Rosas, J.P., Fernández-Gaxiola, A.C. and Rayco-Solon, P. (2015) Effects and safety of periconceptional oral folate supplementation for preventing birth defects. *Cochrane Database Syst Rev*, **2015**, Cd007950.
- 40 Li, S., Chao, A., Li, Z., Moore, C.A., Liu, Y., Zhu, J., Erickson, J.D., Hao, L. and Berry, R.J. (2012) Folic Acid Use and Nonsyndromic Orofacial Clefts in China: A Prospective Cohort Study. *Epidemiology*, **23**, 423-432.
- 41 Johnson, C.Y. and Little, J. (2008) Folate intake, markers of folate status and oral clefts: is the evidence converging? *International journal of epidemiology*, **37**, 1041-1058.
- 42 Kaufman, J.A., Wright, J.M., Evans, A., Rivera-Núñez, Z., Meyer, A. and Narotsky, M.G. (2018) Associations Between Disinfection By-Product Exposures and Craniofacial Birth Defects. *J Occup Environ Med*, **60**, 109-119.
- 43 Brender, J.D., Weyer, P.J., Romitti, P.A., Mohanty, B.P., Shinde, M.U., Vuong, A.M., Sharkey, J.R., Dwivedi, D., Horel, S.A., Kantamneni, J. *et al.* (2013) Prenatal nitrate intake from drinking water and selected birth defects in offspring of participants in the national birth defects prevention study. *Environ Health Perspect*, **121**, 1083-1089.
- 44 Stothard, K.J., Tennant, P.W.G., Bell, R. and Rankin, J. (2009) Maternal Overweight and Obesity and the Risk of Congenital Anomalies: A Systematic Review and Meta-analysis. *JAMA*, **301**, 636-650.
- 45 Block, S.R., Watkins, S.M., Salemi, J.L., Rutkowski, R., Tanner, J.P., Correia, J.A. and Kirby, R.S. (2013) Maternal Pre-Pregnancy Body Mass Index and Risk of Selected Birth Defects: Evidence of a Dose-Response Relationship. *Paediatric and Perinatal Epidemiology*, **27**, 521-531.
- 46 Correa, A., Gilboa, S.M., Besser, L.M., Botto, L.D., Moore, C.A., Hobbs, C.A., Cleves, M.A., Riehle-Colarusso, T.J., Waller, D.K. and Reece, E.A. (2008) Diabetes mellitus and birth defects. *American Journal of Obstetrics and Gynecology*, **199**, 237.e231-237.e239.
- 47 Wu, Y., Liu, B., Sun, Y., Du, Y., Santillan, M.K., Santillan, D.A., Snetselaar, L.G. and Bao, W. (2020) Association of Maternal Prepregnancy Diabetes and Gestational Diabetes Mellitus With Congenital Anomalies of the Newborn. *Diabetes Care*, **43**, 2983-2990.
- 48 Bánhidy, F., Ács, N., Puhó, E.H. and Czeizel, A.E. (2010) Congenital abnormalities in the offspring of pregnant women with type 1, type 2 and gestational diabetes mellitus: A population-based case-control study. *Congenital Anomalies*, **50**, 115-121.
- 49 Ács, L., Bányai, D., Nemes, B., Nagy, K., Ács, N., Bánhidy, F. and Rózsa, N. (2020) Maternal-related factors in the origin of isolated cleft palate-A population-based case-control study. *Orthod Craniofac Res*, **23**, 174-180.
- 50 Métneki, J., Puhó, E. and Czeizel, A.E. (2005) Maternal diseases and isolated orofacial clefts in Hungary. *Birth Defects Res A Clin Mol Teratol*, **73**, 617-623.
- 51 Lammer, E.J., Chen, D.T., Hoar, R.M., Agnish, N.D., Benke, P.J., Braun, J.T., Curry, C.J., Fernhoff, P.M., Grix, A.W., Lott, I.T. *et al.* (1985) Retinoic Acid Embryopathy. *New Engl J Med*, **313**, 837-841.
- 52 Rothman, K.J., Moore, L.L., Singer, M.R., Nguyen, U.-S.D.T., Mannino, S. and Milunsky, A. (1995) Teratogenicity of High Vitamin A Intake. *New Engl J Med*, **333**, 1369-1373.

- 53 Finnell, R.H., Shaw, G.M., Lammer, E.J., Brandl, K.L., Carmichael, S.L. and Rosenquist, T.H. (2004) Gene–nutrient interactions: importance of folates and retinoids during early embryogenesis. *Toxicology and Applied Pharmacology*, **198**, 75-85.
- 54 Johansen, A.M., Lie, R.T., Wilcox, A.J., Andersen, L.F. and Drevon, C.A. (2008) Maternal dietary intake of vitamin A and risk of orofacial clefts: a population-based case-control study in Norway. *Am J Epidemiol*, **167**, 1164-1170.
- 55 Veroniki, A.A., Cogo, E., Rios, P., Straus, S.E., Finkelstein, Y., Kealey, R., Reynen, E., Soobiah, C., Thavorn, K., Hutton, B. *et al.* (2017) Comparative safety of anti-epileptic drugs during pregnancy: a systematic review and network meta-analysis of congenital malformations and prenatal outcomes. *BMC Medicine*, **15**, 95.
- 56 Jackson, A., Bromley, R., Morrow, J., Irwin, B. and Clayton-Smith, J. (2016) In utero exposure to valproate increases the risk of isolated cleft palate. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, **101**, F207-F211.
- 57 Jentink, J., Loane, M.A., Dolk, H., Barisic, I., Garne, E., Morris, J.K. and de Jong-van den Berg, L.T. (2010) Valproic acid monotherapy in pregnancy and major congenital malformations. *N Engl J Med*, **362**, 2185-2193.
- 58 Gilboa, S.M., Broussard, C.S., Devine, O.J., Duwe, K.N., Flak, A.L., Boulet, S.L., Moore, C.A., Werler, M.M. and Honein, M.A. (2011) Influencing clinical practice regarding the use of antiepileptic medications during pregnancy: Modeling the potential impact on the prevalences of spina bifida and cleft palate in the United States. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, **157**, 234-246.
- 59 Puhó, E.H., Szunyogh, M., Métneki, J. and Czeizel, A.E. (2007) Drug treatment during pregnancy and isolated orofacial clefts in Hungary. *Cleft Palate Craniofac J*, **44**, 194-202.
- 60 Garne, E., Hansen, A.V., Morris, J., Zaupper, L., Addor, M.-C., Barisic, I., Gatt, M., Lelong, N., Klungsoyr, K., O'Mahony, M. *et al.* (2015) Use of asthma medication during pregnancy and risk of specific congenital anomalies: A European case-malformed control study. *Journal of Allergy and Clinical Immunology*, **136**, 1496-1502.e1497.
- 61 Danielsson, B., Wikner, B.N. and Källén, B. (2014) Use of ondansetron during pregnancy and congenital malformations in the infant. *Reprod Toxicol*, **50**, 134-137.
- 62 Huybrechts, K.F., Hernández-Díaz, S., Straub, L., Gray, K.J., Zhu, Y., Paterno, E., Desai, R.J., Mogun, H. and Bateman, B.T. (2018) Association of Maternal First-Trimester Ondansetron Use With Cardiac Malformations and Oral Clefts in Offspring. *Jama*, **320**, 2429-2437.
- 63 Hernandez, R.K., Werler, M.M., Romitti, P., Sun, L. and Anderka, M. (2012) Nonsteroidal antiinflammatory drug use among women and the risk of birth defects. *Am J Obstet Gynecol*, **206**, 228.e221-228.
- 64 Broussard, C.S., Rasmussen, S.A., Reefhuis, J., Friedman, J.M., Jann, M.W., Riehle-Colarusso, T. and Honein, M.A. (2011) Maternal treatment with opioid analgesics and risk for birth defects. *American Journal of Obstetrics & Gynecology*, **204**, 314.e311-314.e311.
- 65 Bateman, B.T., Hernandez-Diaz, S., Straub, L., Zhu, Y., Gray, K.J., Desai, R.J., Mogun, H., Gautam, N. and Huybrechts, K.F. (2021) Association of first trimester prescription opioid use with congenital malformations in the offspring: population based cohort study. *Bmj*, **372**, n102.
- 66 Lin, K.J., Mitchell, A.A., Yau, W.P., Louik, C. and Hernández-Díaz, S. (2012) Maternal exposure to amoxicillin and the risk of oral clefts. *Epidemiology*, **23**, 699-705.
- 67 Mølgaard-Nielsen, D. and Hviid, A. (2012) Maternal use of antibiotics and the risk of orofacial clefts: a nationwide cohort study. *Pharmacoepidemiol Drug Saf*, **21**, 246-253.

- 68 Brender, J.D., Werler, M.M., Shinde, M.U., Vuong, A.M., Kelley, K.E., Huber, J.C., Jr., Sharkey, J.R., Griesenbeck, J.S., Romitti, P.A., Malik, S. *et al.* (2012) Nitrosatable drug exposure during the first trimester of pregnancy and selected congenital malformations. *Birth Defects Res A Clin Mol Teratol*, **94**, 701-713.
- 69 Brender, J.D., Kelley, K.E., Werler, M.M., Langlois, P.H., Suarez, L. and Canfield, M.A. (2011) Prevalence and patterns of nitrosatable drug use among U.S. women during early pregnancy. *Birth Defects Res A Clin Mol Teratol*, **91**, 258-264.
- 70 Tammen, S.A., Friso, S. and Choi, S.W. (2013) Epigenetics: the link between nature and nurture. *Mol Aspects Med*, **34**, 753-764.
- 71 Gibney, E.R. and Nolan, C.M. (2010) Epigenetics and gene expression. *Heredity*, **105**, 4-13.
- 72 Sharp, G.C., Ho, K., Davies, A., Stergiakouli, E., Humphries, K., McArdle, W., Sandy, J., Davey Smith, G., Lewis, S.J. and Relton, C.L. (2017) Distinct DNA methylation profiles in subtypes of orofacial cleft. *Clin Epigenetics*, **9**, 63.
- 73 Xu, Z., Lie, R.T., Wilcox, A.J., Saugstad, O.D. and Taylor, J.A. (2019) A comparison of DNA methylation in newborn blood samples from infants with and without orofacial clefts. *Clin Epigenetics*, **11**, 40.
- 74 Sun, B., Reynolds, K., Saha, S.K., Zhang, S., McMahon, M. and Zhou, C.J. (2023) Ezh2-dependent methylation in oral epithelia promotes secondary palatogenesis. *Birth Defects Res*, **115**, 1851-1865.
- 75 Yoshioka, H., Mikami, Y., Ramakrishnan, S.S., Suzuki, A. and Iwata, J. (2021) MicroRNA-124-3p Plays a Crucial Role in Cleft Palate Induced by Retinoic Acid. *Front Cell Dev Biol*, **9**, 621045.
- 76 Li, A., Jia, P., Mallik, S., Fei, R., Yoshioka, H., Suzuki, A., Iwata, J. and Zhao, Z. (2020) Critical microRNAs and regulatory motifs in cleft palate identified by a conserved miRNA-TF-gene network approach in humans and mice. *Brief Bioinform*, **21**, 1465-1478.
- 77 Suzuki, A., Abdallah, N., Gajera, M., Jun, G., Jia, P., Zhao, Z. and Iwata, J. (2018) Genes and microRNAs associated with mouse cleft palate: A systematic review and bioinformatics analysis. *Mech Dev*, **150**, 21-27.
- 78 Schoen, C., Aschrafi, A., Thonissen, M., Poelmans, G., Von den Hoff, J.W. and Carels, C.E.L. (2017) MicroRNAs in Palatogenesis and Cleft Palate. *Front Physiol*, **8**, 165.
- 79 Seelan, R.S., Pisano, M.M. and Greene, R.M. (2022) MicroRNAs as epigenetic regulators of orofacial development. *Differentiation*, **124**, 1-16.
- 80 Zhao, X., Peng, X., Wang, Z., Zheng, X., Wang, X., Wang, Y., Chen, J., Yuan, D., Liu, Y. and Du, J. (2023) MicroRNAs in Small Extracellular Vesicles from Amniotic Fluid and Maternal Plasma Associated with Fetal Palate Development in Mice. *Int J Mol Sci*, **24**.
- 81 Horowitz, S.L., Chabora, A.J. and Horowitz, S.L. (1993) Observations Presenting Examples of Missing Palate. *The Cleft Palate Craniofacial Journal*, **30**, 593-594.
- 82 Trew, C. (1757) Sistens plura exempla palati deficientis. *Nova acta physico-medica academiae caesariae Leopoldion-Carolinae*, **1**, 445-447.
- 83 Fogh-Andersen, P. (1942). Nyt nordisk Forlag Copenhagen, Copenhagen, in press.
- 84 Grosen, D., Bille, C., Petersen, I., Skytthe, A., Hjelmberg, J., Pedersen, J.K., Murray, J.C. and Christensen, K. (2011) Risk of oral clefts in twins. *Epidemiology*, **22**, 313-319.
- 85 Grosen, D., Chevrier, C., Skytthe, A., Bille, C., Mølsted, K., Sivertsen, A., Murray, J.C. and 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. *Journal of medical genetics*, **47**, 162-168.

86 Sivertsen, A., Wilcox, A.J., Skjaerven, R., Vindenes, H.A., Abyholm, F., Harville, E. and Lie, R.T. (2008) Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives. *BMJ (Clinical research ed.)*, **336**, 432-434.

87 Burg, M.L., Chai, Y., Yao, C.A., Magee, W., 3rd and Figueiredo, J.C. (2016) Epidemiology, Etiology, and Treatment of Isolated Cleft Palate. *Frontiers in physiology*, **7**, 67-67.

88 Little, H.J., Rorick, N.K., Su, L.I., Baldock, C., Malhotra, S., Jowitt, T., Gakhar, L., Subramanian, R., Schutte, B.C., Dixon, M.J. *et al.* (2009) Missense mutations that cause Van der Woude syndrome and popliteal pterygium syndrome affect the DNA-binding and transcriptional activation functions of IRF6. *Hum Mol Genet*, **18**, 535-545.

89 Dronamraju, K.R. (1971) Genetic studies of a cleft palate clinic population. *Birth Defects Orig Artic Ser*, **7**, 54-57.

90 Van Der Woude, A. (1954) Fistula labii inferioris congenita and its association with cleft lip and palate. *Am J Hum Genet*, **6**, 244-256.

91 Janku, P., Robinow, M., Kelly, T., Bralley, R., Baynes, A. and Edgerton, M.T. (1980) The van der Woude syndrome in a large kindred: variability, penetrance, genetic risks. *Am J Med Genet*, **5**, 117-123.

92 Leslie, E.J., Koboldt, D.C., Kang, C.J., Ma, L., Hecht, J.T., Wehby, G.L., Christensen, K., Czeizel, A.E., Deleyiannis, F.W., Fulton, R.S. *et al.* (2016) IRF6 mutation screening in non-syndromic orofacial clefting: analysis of 1521 families. *Clin Genet*, **90**, 28-34.

93 Kondo, S., Schutte, B.C., Richardson, R.J., Bjork, B.C., Knight, A.S., Watanabe, Y., Howard, E., de Lima, R.L., Daack-Hirsch, S., Sander, A. *et al.* (2002) Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nat Genet*, **32**, 285-289.

94 de Lima, R.L., Hoper, S.A., Ghassibe, M., Cooper, M.E., Rorick, N.K., Kondo, S., Katz, L., Marazita, M.L., Compton, J., Bale, S. *et al.* (2009) Prevalence and nonrandom distribution of exonic mutations in interferon regulatory factor 6 in 307 families with Van der Woude syndrome and 37 families with popliteal pterygium syndrome. *Genet Med*, **11**, 241-247.

95 Leslie, E.J., Standley, J., Compton, J., Bale, S., Schutte, B.C. and Murray, J.C. (2013) Comparative analysis of IRF6 variants in families with Van der Woude syndrome and popliteal pterygium syndrome using public whole-exome databases. *Genet Med*, **15**, 338-344.

96 Peyrard-Janvid, M., Leslie, E.J., Kousa, Y.A., Smith, T.L., Dunnwald, M., Magnusson, M., Lentz, B.A., Unneberg, P., Fransson, I., Koillinen, H.K. *et al.* (2014) Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. *Am J Hum Genet*, **94**, 23-32.

97 Froster-Iskenius, U.G. (1990) Popliteal pterygium syndrome. *J Med Genet*, **27**, 320-326.

98 Leslie, E.J., O'Sullivan, J., Cunningham, M.L., Singh, A., Goudy, S.L., Ababneh, F., Alsubaie, L., Ch'ng, G.S., van der Laar, I.M., Hoogeboom, A.J. *et al.* (2015) Expanding the genetic and phenotypic spectrum of popliteal pterygium disorders. *Am J Med Genet A*, **167A**, 545-552.

99 Mitchell, K., O'Sullivan, J., Missero, C., Blair, E., Richardson, R., Anderson, B., Antonini, D., Murray, J.C., Shanske, A.L., Schutte, B.C. *et al.* (2012) Exome sequence identifies RIPK4 as the Bartsocas-Papas syndrome locus. *Am J Hum Genet*, **90**, 69-75.

- 100 Lahtela, J., Nousiainen, H.O., Stefanovic, V., Tallila, J., Viskari, H., Karikoski, R., Gentile, M., Saloranta, C., Varilo, T., Salonen, R. *et al.* (2010) Mutant CHUK and Severe Fetal Encasement Malformation. *New Engl J Med*, **363**, 1631-1637.
- 101 Stanier, P. and Moore, G.E. (2004) Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. *Human Molecular Genetics*, **13**, R73-R81.
- 102 Huang, L., Jia, Z., Shi, Y., Du, Q., Shi, J., Wang, Z., Mou, Y., Wang, Q., Zhang, B., Wang, Q. *et al.* (2019) Genetic factors define CPO and CLO subtypes of nonsyndromic orofacial cleft. *PLoS Genet*, **15**, e1008357.
- 103 Khandelwal, K.D., Ishorst, N., Zhou, H., Ludwig, K.U., Venselaar, H., Gilissen, C., Thonissen, M., van Rooij, I.A.L.M., Dreesen, K., Steehouwer, M. *et al.* (2017) Novel IRF6 Mutations Detected in Orofacial Cleft Patients by Targeted Massively Parallel Sequencing. *J Dent Res*, **96**, 179-185.
- 104 Tokat, C., Bilkay, U., Songur, E. and Akin, Y. (2005) Van der Woude Syndrome in Twins. *Journal of Craniofacial Surgery*, **16**, 936-939.
- 105 Lace, B., Pajusalu, S., Livcane, D., Grinfelde, I., Akota, I., Mauliņa, I., Barkāne, B., Stavusis, J. and Inashkina, I. (2022) Monogenic Versus Multifactorial Inheritance in the Development of Isolated Cleft Palate: A Whole Genome Sequencing Study. *Front Genet*, **13**, 828534.
- 106 Nikopensius, T., Jagomägi, T., Krjutškov, K., Tammekivi, V., Saag, M., Prane, I., Piekuse, L., Akota, I., Barkane, B., Krumina, A. *et al.* (2010) Genetic variants in COL2A1, COL11A2, and IRF6 contribute risk to nonsyndromic cleft palate. *Birth Defects Research Part A: Clinical and Molecular Teratology*, **88**, 748-756.
- 107 Beaty, T.H., Ruczinski, I., Murray, J.C., Marazita, M.L., Munger, R.G., Hetmanski, J.B., Murray, T., Redett, R.J., Fallin, M.D., Liang, K.Y. *et al.* (2011) Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. *Genet Epidemiol*, **35**, 469-478.
- 108 Leslie, E.J., Liu, H., Carlson, J.C., Shaffer, J.R., Feingold, E., Wehby, G., Laurie, C.A., Jain, D., Laurie, C.C., Doheny, K.F. *et al.* (2016) A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. *Am J Hum Genet*, **98**, 744-754.
- 109 Leslie, E.J., Carlson, J.C., Shaffer, J.R., Butali, A., Buxo, C.J., Castilla, E.E., Christensen, K., Deleyiannis, F.W., Leigh Field, L., Hecht, J.T. *et al.* (2017) Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Hum Genet*, **136**, 275-286.
- 110 Butali, A., Mossey, P.A., Adeyemo, W.L., Eshete, M.A., Gowans, L.J.J., Busch, T.D., Jain, D., Yu, W., Huan, L., Laurie, C.A. *et al.* (2019) Genomic analyses in African populations identify novel risk loci for cleft palate. *Hum Mol Genet*, **28**, 1038-1051.
- 111 He, M., Zuo, X., Liu, H., Wang, W., Zhang, Y., Fu, Y., Zhen, Q., Yu, Y., Pan, Y., Qin, C. *et al.* (2020) Genome-wide Analyses Identify a Novel Risk Locus for Nonsyndromic Cleft Palate. *J Dent Res*, **99**, 1461-1468.
- 112 Rahimov, F., Nieminen, P., Kumari, P., Juuri, E., Nikopensius, T., Karjalainen, J., Kurki, M., Palotie, A., FinnGen, N. and Heliövaara, A. (2021) High incidence and regional distribution of cleft palate in Finns are associated with a functional variant in an IRF6 enhancer. in press.
- 113 Yu, Y., Zhen, Q., Chen, W., Yu, Y., Li, Z., Wang, Y., Fan, W., Luo, S., Wang, D., Bai, Y. *et al.* (2023) Genome-wide meta-analyses identify five new risk loci for nonsyndromic orofacial clefts in the Chinese Han population. *Mol Genet Genomic Med*, **11**, e2226.

- 114 Ludwig, K.U., Bohmer, A.C., Bowes, J., Nikolic, M., Ishorst, N., Wyatt, N., Hammond, N.L., Golz, L., Thieme, F., Barth, S. *et al.* (2017) Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip +/- cleft palate and cleft palate only. *Hum Mol Genet*, **26**, 829-842.
- 115 Clifton-Bligh, R.J., Wentworth, J.M., Heinz, P., Crisp, M.S., John, R., Lazarus, J.H., Ludgate, M. and Chatterjee, V.K. (1998) Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. *Nature Genetics*, **19**, 399-401.
- 116 Eshete, M.A., Liu, H., Li, M., Adeyemo, W.L., Gowans, L.J.J., Mossey, P.A., Busch, T., Deressa, W., Donkor, P., Olaitan, P.B. *et al.* (2018) Loss-of-Function GRHL3 Variants Detected in African Patients with Isolated Cleft Palate. *J Dent Res*, **97**, 41-48.
- 117 Huang, W., He, Q., Li, M., Ding, Y., Liang, W., Li, W., Lin, J., Zhao, H. and Chen, F. (2023) Two rare variants reveal the significance of Grainyhead-like 3 Arginine 391 underlying non-syndromic cleft palate only. *Oral Dis*, **29**, 1632-1643.
- 118 Iovino, E., Scapoli, L., Palmieri, A., Sgarzani, R., Nouri, N., Pellati, A., Carinci, F., Seri, M., Pippucci, T. and Martinelli, M. (2023) Ultra-Rare Variants Identify Biological Pathways and Candidate Genes in the Pathobiology of Non-Syndromic Cleft Palate Only. *Biomolecules*, **13**.
- 119 Hoebel, A.K., Drichel, D., van de Vorst, M., Böhmer, A.C., Sivalingam, S., Ishorst, N., Klamt, J., Gözl, L., Alblas, M., Maaser, A. *et al.* (2017) Candidate Genes for Nonsyndromic Cleft Palate Detected by Exome Sequencing. *J Dent Res*, **96**, 1314-1321.
- 120 Leslie, E.J., Mansilla, M.A., Biggs, L.C., Schuette, K., Bullard, S., Cooper, M., Dunnwald, M., Lidral, A.C., Marazita, M.L., Beaty, T.H. *et al.* (2012) Expression and mutation analyses implicate ARHGAP29 as the etiologic gene for the cleft lip with or without cleft palate locus identified by genome-wide association on chromosome 1p22. *Birth Defects Res A Clin Mol Teratol*, **94**, 934-942.
- 121 Liu, H., Busch, T., Eliason, S., Anand, D., Bullard, S., Gowans, L.J.J., Nidey, N., Petrin, A., Augustine-Akpan, E.A., Saadi, I. *et al.* (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**, 27-37.
- 122 Dąbrowska, J., Biedziak, B., Bogdanowicz, A. and Mostowska, A. (2023) Identification of Novel Risk Variants of Non-Syndromic Cleft Palate by Targeted Gene Panel Sequencing. *J Clin Med*, **12**.
- 123 Diaz Perez, K.K., Curtis, S.W., Sanchis-Juan, A., Zhao, X., Head, T., Ho, S., Carter, B., McHenry, T., Bishop, M.R., Valencia-Ramirez, L.C. *et al.* (2023) Rare variants found in clinical gene panels illuminate the genetic and allelic architecture of orofacial clefting. *Genetics in Medicine*, **25**.
- 124 Wilson, K., Newbury, D.F. and Kini, U. (2023) Analysis of exome data in a UK cohort of 603 patients with syndromic orofacial clefting identifies causal molecular pathways. *Human Molecular Genetics*, **32**, 1932-1942.
- 125 Bishop, M.R., Diaz Perez, K.K., Sun, M., Ho, S., Chopra, P., Mukhopadhyay, N., Hetmanski, J.B., Taub, M.A., Moreno-Urbe, L.M., Valencia-Ramirez, L.C. *et al.* (2020) Genome-wide Enrichment of De Novo Coding Mutations in Orofacial Cleft Trios. *Am J Hum Genet*, **107**, 124-136.
- 126 Wilderman, A., VanOudenhove, J., Kron, J., Noonan, J.P. and Cotney, J. (2018) High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep*, **23**, 1581-1597.

- 127 Yankee, T.N., Oh, S., Winchester, E.W., Wilderman, A., Robinson, K., Gordon, T., Rosenfeld, J.A., VanOudenhove, J., Scott, D.A., Leslie, E.J. *et al.* (2023) Integrative analysis of transcriptome dynamics during human craniofacial development identifies candidate disease genes. *Nat Commun*, **14**, 4623.
- 128 Li, H., Jones, K.L., Hooper, J.E. and Williams, T. (2019) The molecular anatomy of mammalian upper lip and primary palate fusion at single cell resolution. *Development*, **146**.
- 129 Han, X., Feng, J., Guo, T., Loh, Y.E., Yuan, Y., Ho, T.V., Cho, C.K., Li, J., Jing, J., Janeckova, E. *et al.* (2021) Runx2-Twist1 interaction coordinates cranial neural crest guidance of soft palate myogenesis. *Elife*, **10**.
- 130 Piña, J.O., Raju, R., Roth, D.M., Winchester, E.W., Chattaraj, P., Kidwai, F., Faucz, F.R., Iben, J., Mitra, A., Campbell, K. *et al.* (2023) Multimodal spatiotemporal transcriptomic resolution of embryonic palate osteogenesis. *Nat Commun*, **14**, 5687.
- 131 Masotti, C., Brito, L.A., Nica, A.C., Ludwig, K.U., Nunes, K., Savastano, C.P., Malcher, C., Ferreira, S.G., Kobayashi, G.S., Bueno, D.F. *et al.* (2018) MRPL53, a New Candidate Gene for Orofacial Clefting, Identified Using an eQTL Approach. *J Dent Res*, **97**, 33-40.
- 132 Li, X., Tian, Y., Qiu, L., Lou, S., Zhu, G., Gao, Y., Ma, L. and Pan, Y. (2022) Expression Quantitative Trait Locus Study of Non-Syndromic Cleft Lip with or without Cleft Palate GWAS Variants in Lip Tissues. *Cells*, **11**.
- 133 Yang, J., Yu, X., Zhu, G., Wang, R., Lou, S., Zhu, W., Fu, C., Liu, J., Fan, L., Li, D. *et al.* (2021) Integrating GWAS and eQTL to predict genes and pathways for non-syndromic cleft lip with or without palate. *Oral Dis*, **27**, 1747-1754.
- 134 Berk, N.W. and Marazita, M.L. (2002) *Cleft Lip and Palate: From Origin to Treatment*. Oxford University Press, Oxford.
- 135 Leslie, E.J. and Marazita, M.L. (2013) Genetics of cleft lip and cleft palate. *Am J Med Genet C Semin Med Genet*, **163C**, 246-258.
- 136 Dixon, M.J., Marazita, M.L., Beaty, T.H. and Murray, J.C. (2011) Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet*, **12**, 167-178.
- 137 Yu, Y., Zuo, X., He, M., Gao, J., Fu, Y., Qin, C., Meng, L., Wang, W., Song, Y., Cheng, Y. *et al.* (2017) Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat Commun*, **8**, 14364.
- 138 Smith, T.M., Lozanoff, S., Iyyanar, P.P. and Nazarali, A.J. (2012) Molecular signaling along the anterior-posterior axis of early palate development. *Front Physiol*, **3**, 488.
- 139 Spielman, R.S., McGinnis, R.E. and Ewens, W.J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet*, **52**, 506-516.
- 140 Benner, C., Spencer, C.C.A., Havulinna, A.S., Salomaa, V., Ripatti, S. and Pirinen, M. (2016) FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*, **32**, 1493-1501.
- 141 Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M. and Carter, N.P. (2009) DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet*, **84**, 524-533.
- 142 Heyne, G.W., Plisch, E.H., Melberg, C.G., Sandgren, E.P., Peter, J.A. and Lipinski, R.J. (2015) A Simple and Reliable Method for Early Pregnancy Detection in Inbred Mice. *J Am Assoc Lab Anim Sci*, **54**, 368-371.

- 143 Abler, L.L., Mehta, V., Keil, K.P., Joshi, P.S., Flucus, C.L., Hardin, H.A., Schmitz, C.T. and Vezina, C.M. (2011) A high throughput in situ hybridization method to characterize mRNA expression patterns in the fetal mouse lower urogenital tract. *J Vis Exp*, in press.
- 144 Everson, J.L., Sun, M.R., Fink, D.M., Heyne, G.W., Melberg, C.G., Nelson, K.F., Doroodchi, P., Colopy, L.J., Ulschmid, C.M., Martin, A.A. *et al.* (2019) Developmental Toxicity Assessment of Piperonyl Butoxide Exposure Targeting Sonic Hedgehog Signaling and Forebrain and Face Morphogenesis in the Mouse: An in Vitro and in Vivo Study. *Environ Health Perspect*, **127**, 107006.
- 145 Everson, J.L., Fink, D.M., Yoon, J.W., Leslie, E.J., Kietzman, H.W., Ansen-Wilson, L.J., Chung, H.M., Walterhouse, D.O., Marazita, M.L. and Lipinski, R.J. (2017) Sonic hedgehog regulation of Foxf2 promotes cranial neural crest mesenchyme proliferation and is disrupted in cleft lip morphogenesis. *Development*, **144**, 2082-2091.
- 146 Li, H. and Williams, T. (2013) Separation of mouse embryonic facial ectoderm and mesenchyme. *J Vis Exp*, in press.
- 147 Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E. *et al.* (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*, **47**, D1005-D1012.
- 148 Bush, J.O. and Jiang, R. (2012) Palatogenesis: morphogenetic and molecular mechanisms of secondary palate development. *Development*, **139**, 231-243.
- 149 Rubin, L. and Saunders, J.W., Jr. (1972) Ectodermal-mesodermal interactions in the growth of limb buds in the chick embryo: constancy and temporal limits of the ectodermal induction. *Dev Biol*, **28**, 94-112.
- 150 Takano, A., Fukuda, T., Shinjo, T., Iwashita, M., Matsuzaki, E., Yamamichi, K., Takeshita, M., Sanui, T. and Nishimura, F. (2017) Angiopoietin-like protein 2 is a positive regulator of osteoblast differentiation. *Metabolism - Clinical and Experimental*, **69**, 157-170.
- 151 Zhang, B., Chang, J., Fu, M., Huang, J., Kashyap, R., Salavaggione, E., Jain, S., Kulkarni, S., Deardorff, M.A., Uzielli, M.L. *et al.* (2009) Dosage effects of cohesin regulatory factor PDS5 on mammalian development: implications for cohesinopathies. *PLoS One*, **4**, e5232.
- 152 Kelly, T.J., Suzuki, H.I., Zamudio, J.R., Suzuki, M. and Sharp, P.A. (2019) Sequestration of microRNA-mediated target repression by the Ago2-associated RNA-binding protein FAM120A. *RNA*, **25**, 1291-1297.
- 153 Tanaka, M., Sasaki, K., Kamata, R., Hoshino, Y., Yanagihara, K. and Sakai, R. (2009) A novel RNA-binding protein, Ossa/C9orf10, regulates activity of Src kinases to protect cells from oxidative stress-induced apoptosis. *Mol Cell Biol*, **29**, 402-413.
- 154 Ferguson, M.W., Sharpe, P.M., Thomas, B.L. and Beck, F. (1992) Differential expression of insulin-like growth factors I and II (IGF I and II), mRNA, peptide and binding protein 1 during mouse palate development: comparison with TGF beta peptide distribution. *J Anat*, **181** (Pt 2), 219-238.
- 155 Eggenschwiler, J., Ludwig, T., Fisher, P., Leighton, P.A., Tilghman, S.M. and Efstratiadis, A. (1997) Mouse mutant embryos overexpressing IGF-II exhibit phenotypic features of the Beckwith-Wiedemann and Simpson-Golabi-Behmel syndromes. *Genes Dev*, **11**, 3128-3142.
- 156 Hoornaert, K.P., Vereecke, I., Dewinter, C., Rosenberg, T., Beemer, F.A., Leroy, J.G., Bendix, L., Björck, E., Bonduelle, M., Boute, O. *et al.* (2010) Stickler syndrome caused by COL2A1 mutations: genotype-phenotype correlation in a series of 100 patients. *Eur J Hum Genet*, **18**, 872-880.

- 157 Camargo, M., Rivera, D., Moreno, L., Lidral, A.C., Harper, U., Jones, M., Solomon, B.D., Roessler, E., Velez, J.I., Martinez, A.F. *et al.* (2012) GWAS reveals new recessive loci associated with non-syndromic facial clefting. *Eur J Med Genet*, **55**, 510-514.
- 158 Lary, J.M. and Paulozzi, L.J. (2001) Sex differences in the prevalence of human birth defects: A population-based study. *Teratology*, **64**, 237-251.
- 159 Sokal, R., Tata, L.J. and Fleming, K.M. (2014) Sex prevalence of major congenital anomalies in the United Kingdom: a national population-based study and international comparison meta-analysis. *Birth Defects Res A Clin Mol Teratol*, **100**, 79-91.
- 160 Copeland, G.E. and Kirby, R.S. (2007) Using birth defects registry data to evaluate infant and childhood mortality associated with birth defects: an alternative to traditional mortality assessment using underlying cause of death statistics. *Birth Defects Res A Clin Mol Teratol*, **79**, 792-797.
- 161 Ely, D.M. and Driscoll, A.K. (2019) Infant Mortality in the United States, 2017: Data From the Period Linked Birth/Infant Death File. *Natl Vital Stat Rep*, **68**, 1-20.
- 162 Tennant, P.W.G., Samarasekera, S.D., Pless-Mullooli, T. and Rankin, J. (2011) Sex differences in the prevalence of congenital anomalies: A population-based study. *Birth Defects Research Part A: Clinical and Molecular Teratology*, **91**, 894-901.
- 163 Cui, W., Ma, C.-X., Tang, Y., Chang, V., Rao, P.V., Ariet, M., Resnick, M.B. and Roth, J. (2005) Sex differences in birth defects: A study of opposite-sex twins. *Birth Defects Research Part A: Clinical and Molecular Teratology*, **73**, 876-880.
- 164 Yoshida, S., Takeuchi, M., Kawakami, C., Kawakami, K. and Ito, S. (2020) Maternal multivitamin intake and orofacial clefts in offspring: Japan Environment and Children's Study (JECS) cohort study. *BMJ Open*, **10**, e035817.
- 165 DeRoo, L.A., Wilcox, A.J., Lie, R.T., Romitti, P.A., Pedersen, D.A., Munger, R.G., Moreno Uribe, L.M. and Wehby, G.L. (2016) Maternal alcohol binge-drinking in the first trimester and the risk of orofacial clefts in offspring: a large population-based pooling study. *Eur J Epidemiol*, **31**, 1021-1034.
- 166 Kummet, C.M., Moreno, L.M., Wilcox, A.J., Romitti, P.A., DeRoo, L.A., Munger, R.G., Lie, R.T. and Wehby, G.L. (2016) Passive Smoke Exposure as a Risk Factor for Oral Clefts-A Large International Population-Based Study. *Am J Epidemiol*, **183**, 834-841.
- 167 Carlson, J.C., Nidey, N.L., Butali, A., Buxo, C.J., Christensen, K., Deleyiannis, F.W., Hecht, J.T., Field, L.L., Moreno-Urbe, L.M., Orioli, I.M. *et al.* (2018) Genome-wide interaction studies identify sex-specific risk alleles for nonsyndromic orofacial clefts. *Genet Epidemiol*, **42**, 664-672.
- 168 Awotoye, W., Comnick, C., Pendleton, C., Zeng, E., Alade, A., Mossey, P.A., Gowans, L.J.J., Eshete, M.A., Adeyemo, W.L., Naicker, T. *et al.* (2022) Genome-wide Gene-by-Sex Interaction Studies Identify Novel Nonsyndromic Orofacial Clefts Risk Locus. *J Dent Res*, **101**, 465-472.
- 169 Marçano, A.C., Doudney, K., Braybrook, C., Squires, R., Patton, M.A., Lees, M.M., Richieri-Costa, A., Lidral, A.C., Murray, J.C., Moore, G.E. *et al.* (2004) TBX22 mutations are a frequent cause of cleft palate. *J Med Genet*, **41**, 68-74.
- 170 Delbridge, A.R.D., Kueh, A.J., Ke, F., Zamudio, N.M., El-Saafin, F., Jansz, N., Wang, G.-Y., Iminoff, M., Beck, T., Haupt, S. *et al.* (2019) Loss of p53 Causes Stochastic Aberrant X-Chromosome Inactivation and Female-Specific Neural Tube Defects. *Cell Reports*, **27**, 442-454.e445.

- 171 Pinto, L.L.C., Vieira, T.A., Giugliani, R. and Schwartz, I.V.D. (2010) Expression of the disease on female carriers of X-linked lysosomal disorders: a brief review. *Orphanet Journal of Rare Diseases*, **5**, 14.
- 172 McWilliam, C., Cooke, A., Lobo, D., Warner, J., Taylor, M. and Tolmie, J.L. (2010) Semi-dominant X-chromosome linked learning disability with progressive ataxia, spasticity and dystonia associated with the novel MECP2 variant p.V122A: Akin to the new MECP2 duplication syndrome? *Eur J Paediatr Neuro*, **14**, 267-269.
- 173 Robinson, K., Mosley, T.J., Rivera-González, K.S., Jabbarpour, C.R., Curtis, S.W., Adeyemo, W.L., Beaty, T.H., Butali, A., Buxó, C.J., Cutler, D.J. *et al.* (2023) Trio-based GWAS identifies novel associations and subtype-specific risk factors for cleft palate. *Human Genetics and Genomics Advances*, **4**.
- 174 Schwender H, L.Q., Berger P, Neumann C, Taub M, Ruczinski I. (2023) trio: Testing of SNPs and SNP Interactions in Case-Parent Trio Studies. in press.
- 175 Gauderman, W.J., Thomas, D.C., Murcray, C.E., Conti, D., Li, D. and Lewinger, J.P. (2010) Efficient genome-wide association testing of gene-environment interaction in case-parent trios. *Am J Epidemiol*, **172**, 116-122.
- 176 Consortium, T.G., Aguet, F., Anand, S., Ardlie, K.G., Gabriel, S., Getz, G.A., Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H. *et al.* (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318-1330.
- 177 Stegle, O., Parts, L., Piipari, M., Winn, J. and Durbin, R. (2012) Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nature Protocols*, **7**, 500-507.
- 178 Parrish, R.L., Gibson, G.C., Epstein, M.P. and Yang, J. (2022) TIGAR-V2: Efficient TWAS tool with nonparametric Bayesian eQTL weights of 49 tissue types from GTEx V8. *Human Genetics and Genomics Advances*, **3**, 100068.
- 179 Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L. *et al.* (2018) Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature Communications*, **9**, 1825.
- 180 Liu, Y., Chen, S., Li, Z., Morrison, A.C., Boerwinkle, E. and Lin, X. (2019) ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *The American Journal of Human Genetics*, **104**, 410-421.
- 181 Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, **38**, e164.
- 182 Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T. *et al.* (2023) Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*, **381**, eadg7492.
- 183 Recouvreux, M.V., Lapyckyj, L., Camilletti, M.A., Guida, M.C., Ornstein, A., Rifkin, D.B., Becu-Villalobos, D. and Díaz-Torga, G. (2013) Sex differences in the pituitary transforming growth factor- β 1 system: studies in a model of resistant prolactinomas. *Endocrinology*, **154**, 4192-4205.
- 184 Peeters, S., De Kinderen, P., Meester, J.A.N., Verstraeten, A. and Loeys, B.L. (2022) The fibrillinopathies: New insights with focus on the paradigm of opposing phenotypes for both FBN1 and FBN2. *Hum Mutat*, **43**, 815-831.
- 185 Iwata, J., Parada, C. and Chai, Y. (2011) The mechanism of TGF- β signaling during palate development. *Oral Dis*, **17**, 733-744.

- 186 Taya, Y., O'Kane, S. and Ferguson, M.W. (1999) Pathogenesis of cleft palate in TGF-beta3 knockout mice. *Development*, **126**, 3869-3879.
- 187 Pottie, L., Adamo, C.S., Beyens, A., Lütke, S., Tapanecyaphan, P., De Clercq, A., Salmon, P.L., De Rycke, R., Gezdirici, A., Gulec, E.Y. *et al.* (2021) Bi-allelic premature truncating variants in *LTBP1* cause cutis laxa syndrome. *The American Journal of Human Genetics*, **108**, 1095-1114.
- 188 Tran, D.Q., Andersson, J., Wang, R., Ramsey, H., Unutmaz, D. and Shevach, E.M. (2009) GARP (LRRC32) is essential for the surface expression of latent TGF-beta on platelets and activated FOXP3+ regulatory T cells. *Proc Natl Acad Sci U S A*, **106**, 13445-13450.
- 189 Wang, R., Zhu, J., Dong, X., Shi, M., Lu, C. and Springer, T.A. (2012) GARP regulates the bioavailability and activation of TGFβ. *Mol Biol Cell*, **23**, 1129-1139.
- 190 Hexner-Erlichman, Z., Fichtman, B., Zehavi, Y., Khayat, M., Jabaly-Habib, H., Izhaki-Tavor, L.S., Dessau, M., Elpeleg, O. and Spiegel, R. (2022) A Novel Homozygous Missense Variant in the LRRC32 Gene Is Associated With a New Syndrome of Cleft Palate, Progressive Vitreoretinopathy, Growth Retardation, and Developmental Delay. *Front Pediatr*, **10**, 859034.
- 191 Drews, F., Knöbel, S., Moser, M., Muhlack, K.G., Mohren, S., Stoll, C., Bosio, A., Gressner, A.M. and Weiskirchen, R. (2008) Disruption of the latent transforming growth factor-β binding protein-1 gene causes alteration in facial structure and influences TGF-β bioavailability. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, **1783**, 34-48.
- 192 Resnick, C.M., Estroff, J.A., Kooiman, T.D., Calabrese, C.E., Koudstaal, M.J. and Padwa, B.L. (2018) Pathogenesis of Cleft Palate in Robin Sequence: Observations From Prenatal Magnetic Resonance Imaging. *J Oral Maxil Surg*, **76**, 1058-1064.
- 193 Jansen, R., Batista, S., Brooks, A.I., Tischfield, J.A., Willemsen, G., van Grootheest, G., Hottenga, J.J., Milaneschi, Y., Mbarek, H., Madar, V. *et al.* (2014) Sex differences in the human peripheral blood transcriptome. *BMC Genomics*, **15**, 33.
- 194 Denu, R.A., Nemcek, S., Bloom, D.D., Goodrich, A.D., Kim, J., Mosher, D.F. and Hematti, P. (2016) Fibroblasts and Mesenchymal Stromal/Stem Cells Are Phenotypically Indistinguishable. *Acta Haematol*, **136**, 85-97.
- 195 Stewart, J.M., Ott, J.E. and Lagace, R. (1972) Submucous cleft palate: prevalence in a school population. *Cleft Palate J*, **9**, 246-250.
- 196 Moreira, T., Dias, M., Von Hafe, M., Curval, A.R., Ramalho, C., Maia, A.M., Moura, C.P. and Orofacial Cleft Team of University Hospital Center of São João, E. (2023) Orofacial clefts: Reflections on prenatal diagnosis and family history based on a series of cases of a tertiary children hospital. *Congenital Anomalies*, **63**, 195-199.
- 197 Trezena, S., Machado, R.A., de Almeida Reis, S.R., Scariot, R., Rangel, A., de Oliveira, F.E.S., Borges, A.J., Silva, A.T., Martelli, D.R.B. and Martelli Júnior, H. (2023) Isolated nonsyndromic cleft palate: multicenter epidemiological study in the Brazil. *BMC Oral Health*, **23**, 486.
- 198 Machado, R.A., Martelli-Junior, H., Reis, S.R.A., Kuchler, E.C., Scariot, R., das Neves, L.T. and Coletta, R.D. (2021) Identification of Novel Variants in Cleft Palate-Associated Genes in Brazilian Patients With Non-syndromic Cleft Palate Only. *Front Cell Dev Biol*, **9**, 638522.
- 199 Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J.D., Romano-Adesman, A., Bjornson, R.D., Breitbart, R.E., Brown, K.K. *et al.* (2013) De novo mutations in histone-modifying genes in congenital heart disease. *Nature*, **498**, 220-223.
- 200 Yu, L., Sawle, A.D., Wynn, J., Aspelund, G., Stolar, C.J., Arkovitz, M.S., Potoka, D., Azarow, K.S., Mychaliska, G.B., Shen, Y. *et al.* (2015) Increased burden of de novo predicted

- deleterious variants in complex congenital diaphragmatic hernia. *Hum Mol Genet*, **24**, 4764-4773.
- 201 Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E. *et al.* (2014) The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, **515**, 216-221.
- 202 Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L. *et al.* (2020) Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*, **180**, 568-584.e523.
- 203 Zhou, X., Feliciano, P., Shu, C., Wang, T., Astrovskaia, I., Hall, J.B., Obiajulu, J.U., Wright, J.R., Murali, S.C., Xu, S.X. *et al.* (2022) Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. *Nat Genet*, **54**, 1305-1319.
- 204 Conrad, D.F., Keebler, J.E., DePristo, M.A., Lindsay, S.J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C.L., Torroja, C., Garimella, K.V. *et al.* (2011) Variation in genome-wide mutation rates within and between human families. *Nat Genet*, **43**, 712-714.
- 205 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, **20**, 1297-1303.
- 206 Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J. *et al.* (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, **43**, 11 10 11-11 10 33.
- 207 Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnstrom, K., Mallick, S., Kirby, A. *et al.* (2014) A framework for the interpretation of de novo mutation in human disease. *Nat Genet*, **46**, 944-950.
- 208 Hayat, M.J. and Higgins, M. (2014) Understanding poisson regression. *J Nurs Educ*, **53**, 207-215.
- 209 Chen, J., Bardes, E.E., Aronow, B.J. and Jegga, A.G. (2009) ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*, **37**, W305-311.
- 210 Soh, Z., Richards, A.J., McNinch, A., Alexander, P., Martin, H. and Snead, M.P. (2022) Dominant Stickler Syndrome. *Genes (Basel)*, **13**.
- 211 Carapito, R., Goldenberg, A., Paul, N., Pichot, A., David, A., Hamel, A., Dumant-Forest, C., Leroux, J., Ory, B., Isidor, B. *et al.* (2016) Protein-altering MYH3 variants are associated with a spectrum of phenotypes extending to spondylocarpotarsal synostosis syndrome. *Eur J Hum Genet*, **24**, 1746-1751.
- 212 Zarate, Y.A., Bosanko, K.A., Caffrey, A.R., Bernstein, J.A., Martin, D.M., Williams, M.S., Berry-Kravis, E.M., Mark, P.R., Manning, M.A., Bhambhani, V. *et al.* (2019) Mutation update for the SATB2 gene. *Hum Mutat*, **40**, 1013-1029.
- 213 Zarate, Y.A., Kaylor, J. and Fish, J. (1993) Adam, M.P., Feldman, J., Mirzaa, G.M., Pagon, R.A., Wallace, S.E., Bean, L.J.H., Gripp, K.W. and Amemiya, A. (eds.), In *GeneReviews*(®). University of Washington, Seattle
Copyright © 1993-2024, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved., Seattle (WA), in press.
- 214 Loeys, B.L., Chen, J., Neptune, E.R., Judge, D.P., Podowski, M., Holm, T., Meyers, J., Leitch, C.C., Katsanis, N., Sharifi, N. *et al.* (2005) A syndrome of altered cardiovascular,

- craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. *Nat Genet*, **37**, 275-281.
- 215 Gerrard, G., Valgañón, M., Foong, H.E., Kasperaviciute, D., Iskander, D., Game, L., Müller, M., Aitman, T.J., Roberts, I., de la Fuente, J. *et al.* (2013) Target enrichment and high-throughput sequencing of 80 ribosomal protein genes to identify mutations associated with Diamond-Blackfan anaemia. *Br J Haematol*, **162**, 530-536.
- 216 Martinelli, S., De Luca, A., Stellacci, E., Rossi, C., Checquolo, S., Lepri, F., Caputo, V., Silvano, M., Buscherini, F., Consoli, F. *et al.* (2010) Heterozygous germline mutations in the CBL tumor-suppressor gene cause a Noonan syndrome-like phenotype. *Am J Hum Genet*, **87**, 250-257.
- 217 Clayton-Smith, J., O'Sullivan, J., Daly, S., Bhaskar, S., Day, R., Anderson, B., Voss, A.K., Thomas, T., Biesecker, L.G., Smith, P. *et al.* (2011) Whole-exome-sequencing identifies mutations in histone acetyltransferase gene KAT6B in individuals with the Say-Barber-Biesecker variant of Ohdo syndrome. *Am J Hum Genet*, **89**, 675-681.
- 218 Wright, C.F., Fitzgerald, T.W., Jones, W.D., Clayton, S., McRae, J.F., van Kogelenberg, M., King, D.A., Ambridge, K., Barrett, D.M., Bayzatinova, T. *et al.* (2015) Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet*, **385**, 1305-1314.
- 219 Dauwerse, J.G., Dixon, J., Seland, S., Ruivenkamp, C.A., van Haeringen, A., Hoefsloot, L.H., Peters, D.J., Boers, A.C., Daumer-Haas, C., Maiwald, R. *et al.* (2011) Mutations in genes encoding subunits of RNA polymerases I and III cause Treacher Collins syndrome. *Nat Genet*, **43**, 20-22.
- 220 Sanchez, E., Laplace-Builhé, B., Mau-Them, F.T., Richard, E., Goldenberg, A., Toler, T.L., Guignard, T., Gatinois, V., Vincent, M., Blanchet, C. *et al.* (2020) POLR1B and neural crest cell anomalies in Treacher Collins syndrome type 4. *Genet Med*, **22**, 547-556.
- 221 Smallwood, K., Watt, K.E.N., Ide, S., Baltrunaite, K., Brunswick, C., Inskip, K., Capannari, C., Adam, M.P., Begtrup, A., Bertola, D.R. *et al.* (2023) POLR1A variants underlie phenotypic heterogeneity in craniofacial, neural, and cardiac anomalies. *Am J Hum Genet*, **110**, 809-825.
- 222 Traba, J., Satrustegui, J. and del Arco, A. (2009) Characterization of SCaMC-3-like/slc25a41, a novel calcium-independent mitochondrial ATP-Mg/Pi carrier. *Biochem J*, **418**, 125-133.
- 223 Niessen, M.T., Scott, J., Zielinski, J.G., Vorhagen, S., Sotiropoulou, P.A., Blanpain, C., Leitges, M. and Niessen, C.M. (2013) aPKC λ controls epidermal homeostasis and stem cell fate through regulation of division orientation. *J Cell Biol*, **202**, 887-900.
- 224 Reina-Campos, M., Diaz-Meco, M.T. and Moscat, J. (2019) The Dual Roles of the Atypical Protein Kinase Cs in Cancer. *Cancer Cell*, **36**, 218-235.
- 225 Tan, E.C., Lim, E.C. and Lee, S.T. (2013) De novo 2.3 Mb microdeletion of 1q32.2 involving the Van der Woude Syndrome locus. *Mol Cytogenet*, **6**, 31.
- 226 Al-Kurbi, A.A., Aliyev, E., AlSa'afin, S., Aamer, W., Palaniswamy, S., Al-Maraghi, A., Kilani, H., Akil, A.A., Stotland, M.A. and Fakhro, K.A. (2023) A Complex Intrachromosomal Rearrangement Disrupting IRF6 in a Family with Popliteal Pterygium and Van der Woude Syndromes. *Genes (Basel)*, **14**.
- 227 Kalay, E., Sezgin, O., Chellappa, V., Mutlu, M., Morsy, H., Kayserili, H., Kreiger, E., Cansu, A., Toraman, B., Abdalla, E.M. *et al.* (2012) Mutations in RIPK4 cause the autosomal-recessive form of popliteal pterygium syndrome. *Am J Hum Genet*, **90**, 76-85.

- 228 Kwa, M.Q., Huynh, J., Aw, J., Zhang, L., Nguyen, T., Reynolds, E.C., Sweet, M.J., Hamilton, J.A. and Scholz, G.M. (2014) Receptor-interacting protein kinase 4 and interferon regulatory factor 6 function as a signaling axis to regulate keratinocyte differentiation. *J Biol Chem*, **289**, 31077-31087.
- 229 de la Garza, G., Schleiffarth, J.R., Dunnwald, M., Mankad, A., Weirather, J.L., Bonde, G., Butcher, S., Mansour, T.A., Kousa, Y.A., Fukazawa, C.F. *et al.* (2013) Interferon regulatory factor 6 promotes differentiation of the periderm by activating expression of Grainyhead-like 3. *J Invest Dermatol*, **133**, 68-77.
- 230 Kim, S.W., Schifano, M., Oleksyn, D., Jordan, C.T., Ryan, D., Insel, R., Zhao, J. and Chen, L. (2014) Protein kinase C-associated kinase regulates NF- κ B activation through inducing IKK activation. *Int J Oncol*, **45**, 1707-1714.
- 231 Seitz, C.S., Freiberg, R.A., Hinata, K. and Khavari, P.A. (2000) NF-kappaB determines localization and features of cell death in epidermis. *J Clin Invest*, **105**, 253-260.
- 232 Mukhopadhyay, N., Bishop, M., Mortillo, M., Chopra, P., Hetmanski, J.B., Taub, M.A., Moreno, L.M., Valencia-Ramirez, L.C., Restrepo, C., Wehby, G.L. *et al.* (2020) Whole genome sequencing of orofacial cleft trios from the Gabriella Miller Kids First Pediatric Research Consortium identifies a new locus on chromosome 21. *Hum Genet*, **139**, 215-226.
- 233 Sobreira, N., Schiettecatte, F., Valle, D. and Hamosh, A. (2015) GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat*, **36**, 928-930.
- 234 Kovac, J., Oster, H. and Leitges, M. (2007) Expression of the atypical protein kinase C (aPKC) isoforms ι/λ and ζ during mouse embryogenesis. *Gene Expression Patterns*, **7**, 187-196.
- 235 Moscat, J., Diaz-Meco, M.T., Albert, A. and Campuzano, S. (2006) Cell signaling and function organized by PB1 domain interactions. *Mol Cell*, **23**, 631-640.
- 236 Steinberg, S.F. (2008) Structural basis of protein kinase C isoform function. *Physiol Rev*, **88**, 1341-1378.
- 237 Messerschmidt, A., Macieira, S., Velarde, M., Bädeker, M., Benda, C., Jestel, A., Brandstetter, H., Neufeind, T. and Blaesse, M. (2005) Crystal Structure of the Catalytic Domain of Human Atypical Protein Kinase C- ι Reveals Interaction Mode of Phosphorylation Site in Turn Motif. *Journal of Molecular Biology*, **352**, 918-931.
- 238 Zhou, X., Edmonson, M.N., Wilkinson, M.R., Patel, A., Wu, G., Liu, Y., Li, Y., Zhang, Z., Rusch, M.C., Parker, M. *et al.* (2016) Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nat Genet*, **48**, 4-6.
- 239 Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The Protein Kinase Complement of the Human Genome. *Science*, **298**, 1912-1934.
- 240 Soloff, R.S., Katayama, C., Lin, M.Y., Feramisco, J.R. and Hedrick, S.M. (2004) Targeted deletion of protein kinase C lambda reveals a distribution of functions between the two atypical protein kinase C isoforms. *J Immunol*, **173**, 3250-3260.
- 241 Sabel, J.L., d'Alençon, C., O'Brien, E.K., Van Otterloo, E., Lutz, K., Cuykendall, T.N., Schutte, B.C., Houston, D.W. and Cornell, R.A. (2009) Maternal Interferon Regulatory Factor 6 is required for the differentiation of primary superficial epithelia in Danio and Xenopus embryos. *Developmental Biology*, **325**, 249-262.
- 242 Fukazawa, C., Santiago, C., Park, K.M., Deery, W.J., Gomez de la Torre Canny, S., Holterhoff, C.K. and Wagner, D.S. (2010) poky/chuk/ikk1 is required for differentiation of the zebrafish embryonic epidermis. *Developmental Biology*, **346**, 272-283.

- 243 Shah, H., Khan, K., Khan, N., Badshah, Y., Ashraf, N.M. and Shabbir, M. (2022) Impact of deleterious missense PRKCI variants on structural and functional dynamics of protein. *Sci Rep-Uk*, **12**, 3781.
- 244 Nesta, A.V., Tafur, D. and Beck, C.R. (2021) Hotspots of Human Mutation. *Trends in Genetics*, **37**, 717-729.
- 245 Cobbaut, M., McDonald, N.Q. and Parker, P.J. (2023) Control of atypical PKC ζ membrane dissociation by tyrosine phosphorylation within a PB1-C1 interdomain interface. *J Biol Chem*, **299**, 104847.
- 246 Linch, M., Sanz-Garcia, M., Soriano, E., Zhang, Y., Riou, P., Rosse, C., Cameron, A., Knowles, P., Purkiss, A., Kjaer, S. *et al.* (2013) A Cancer-Associated Mutation in Atypical Protein Kinase C ζ Occurs in a Substrate-Specific Recruitment Motif. *Science Signaling*, **6**, ra82-ra82.