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:

_____ _____

Kimberly Kamille Diaz Perez Date

Uncovering the contribution of rare genetic variants in orofacial clefts

By

Kimberly Kamille Diaz Perez

Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences

Genetics and Molecular Biology

_____

Elizabeth J. Leslie, Ph.D.
Advisor

_____

Karen N. Conneely, Ph.D.
Committee Member

_____

David J. Cutler, Ph.D.
Committee Member

_____

Michael P. Epstein, Ph.D.
Committee Member

_____

Michael J. Gambello, MD, Ph.D.
Committee Member

Accepted:

_____

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

_____

Date

Uncovering the contribution of rare genetic variants in orofacial clefts

By

Kimberly Kamille Diaz Perez
B.S., Georgia State University, 2018

Advisor: Elizabeth J. Leslie, Ph.D.

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 Graduate Division of Biological and Biomedical Sciences
Genetics and Molecular Biology
2023

Abstract

Uncovering the contribution of rare genetic variants in orofacial clefts
By Kimberly Kamille Diaz Perez

Orofacial clefts (OFCs) are common craniofacial congenital anomalies that occur when the palate and/or the lip fail to form properly during development. Approximately 70% of OFCs are considered non-syndromic, occurring as isolated events without additional structural or cognitive anomalies. They are heterogeneous etiologies involving environmental and genetic risk factors that include Mendelian and non-Mendelian causes. However, much of the focus has been on common variants and genome-wide association studies, which only account for ~25% of the heritable risk for OFCs. The unaccounted heritability may be attributed to other sources of variation, including rare variants that are examined using sequencing approaches (e.g., whole-exome sequencing (WES) and whole-genome sequencing (WGS)). We hypothesize that rare variants contribute to the OFC risk through different genetic mechanisms. We first investigated rare variants using WES in 31 multiplex families, each with multiple affected individuals that were consistent with Mendelian inheritance. Nineteen of these families were deeply phenotyped to assess subclinical phenotypes, which are subtle phenotypes hypothesized to exist within the same etiological spectrum as OFCs. We identified likely causal variants in 25% of multiplex families. Several variants were in genes that are mutated in Mendelian syndromes that include OFCs as a feature. Although some variants were found in individuals with subclinical phenotypes, we cumulatively failed to find clear evidence supporting the subclinical phenotype hypothesis. To explore all possible types of inheritance, we next analyzed 50 consanguineous families from Colombia and Turkey. The identified variants of interest included dominant-acting heterozygous variants as well as recessive homozygous variants. Finally, we estimated the diagnostic yield of 841 cases and 294 controls with WGS using 418 genes previously implicated in OFCs. We found 9.04% of cases had a pathogenic variant and nine genes alone accounted for 4.64% of the yield. Taken together, these findings provide evidence of the role of rare variants in OFCs and underscore the etiologic heterogeneity of OFCs. Further elucidation of the genetic architectures of all OFC types will aid in improving recurrence risk estimations and developing tailored management plans for individuals with OFCs.

Uncovering the contribution of rare genetic variants in orofacial clefts

By

Kimberly Kamille Diaz Perez
B.S., Georgia State University, 2018

Advisor: Elizabeth J. Leslie, Ph.D.

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 Graduate Division of Biological and Biomedical Sciences
Genetics and Molecular Biology
2023

Acknowledgements

I would like to express my gratitude to my parents for helping me get here and to my brothers for their support. I would also like to thank the Koukoutsis-Pilote family for helping us transition from Puerto Rico to the US exactly 10 years ago. Lastly, I would like to thank my colleagues in the lab and my advisor for providing assistance and guiding me through my graduate training.

Table of Contents

Table of Tables

Table of Supplementary Tables

Table of Figures

Table of Supplementary Figures

CHAPTER I. Introduction

What are OFCs?

Orofacial clefts (OFCs) are common structural birth defects that arise due to the incomplete fusion of primordial facial tissues in the lip and/or the palate during embryonic development. OFCs affect, on average, 1 in 1,000 live births worldwide (Mossey, 2007; Mossey *et al*., 2009) but the prevalence of OFCs varies by ethnicity and geographical region (Marazita, 2012). For example, individuals from the Asian and Native American populations have the highest prevalence (1 in 500), individuals of European descent have an intermediate prevalence (1 in 1,000), and individuals from African populations have the lowest prevalence of OFCs (1 in 2,500) (Cooper *et al*., 2006; Mossey, 2007; Mossey & Modell, 2012).

In addition to the public health burden, OFCs present a significant personal burden to individuals with OFCs as well as their families. Although OFCs are surgically repaired during the first years of life, individuals with OFCs experience difficulties in different areas of life, including speech, hearing, feeding, and dentition (Marazita, 2012; Wehby & Cassell, 2010), and can require significant clinical interventions. Consequently, OFCs pose a significant financial burden as individuals with OFCs, on average, undergo six corrective surgeries, five years of orthodontic and dental care, and ongoing speech therapy (Berk & Marazita, 2002; Wehby & Cassell, 2010). The mean medical treatment costs are eight times greater for infants with OFCs compared to infants without OFCs (Boulet *et al*., 2009). Moreover, individuals with OFCs experience higher risks of comorbidities, including breast, brain, lung, and colon cancer, as well as higher mortality rates, especially in less developed countries (Bille *et al*., 2005; Christensen *et al*., 2004; Dietz *et al*., 2012; Wehby *et al*., 2006). Lastly, individuals with OFCs can experience significant psychosocial effects, such as experiencing social discrimination, and behavioral and

mental health problems (Christensen & Mortensen, 2002; Hunt *et al*., 2006). Thus, given the pronounced burden of OFCs across the world, studies of OFCs are needed to explore the epidemiology and their causes.

Overview of Craniofacial Development

The development of the face involves several complex coordinated interactions based on the growth, migration, and fusion of cells and tissues at different time points of embryogenesis (Dixon *et al*., 2011; Hammond & Dixon, 2022; Mossey *et al*., 2009). Craniofacial development starts with the migration of cranial neural crest cells from the neural fold. Together with ectodermal and mesodermal cells, these cells form the branchial arches and facial processes that will ultimately give rise to the face (Mossey *et al*., 2009; O'Rahilly & Müller, 2007; Rinkoff & Adlard, 2022; Roth *et al*., 2021; Schutte & Murray, 1999). Development of the lip begins in the 4th week of embryonic development, with the formation of the five facial primordia that surround the oral cavity. These processes are the frontonasal prominence, two paired mandibular processes, and two paired maxillary processes (Ferguson *et al*., 2000) (Figure 1-1A). Soon after, the frontonasal prominence is divided into the paired medial and lateral nasal processes by the formation of the nasal pits (Figure 1-1B). During the 6th week of embryogenesis, the medial nasal processes fuse with the maxillary and lateral nasal processes to form the upper lip and medial nasal processes fuse at the midline, ultimately creating the philtrum and the primary palate (alveolus) (Mossey *et al*., 2009) (Figure 1-1C). Failure of the growth, migration, and fusion of these processes in the 4th-6th weeks of development can lead to a cleft lip.

During the 6th week of development, the secondary palate starts to form through the growth of maxillary processes and palatal shelves, the latter of which grow vertically along the

sides of the tongue (Figure 1-1D). The palatal shelves then elevate to a horizontal position and grow before fusing at the midline epithelial seam. The palatal mesenchyme differentiates into bone and muscle to create the hard and soft palate (Figure 1-1E). The nasal septum and the primary palate also fuse to distinguish the oral and nasal cavities (Roth *et al*., 2021; Sperber, 2002) (Figure 1-1F). The formation of the palate is completed by the 10th week of development. Failure to complete any of these processes can lead to a cleft palate.

Broadening the Phenotypic Spectrum of OFCs

OFCs can affect different facial structures with varying degrees of severity, resulting in the heterogeneity of the phenotype. The most common way to classify OFCs is by the affected structure(s) into those affecting the upper lip (with or without the involvement of the primary palate/alveolus) (i.e., cleft lip, CL), the secondary palate (i.e., cleft palate, CP), or both (i.e., cleft lip and palate, CLP) (Figure 1-2). Most commonly, CL and CLP are grouped together into cleft lip with or without cleft palate (CL/P) as some have hypothesized these to be the same defect but with a different severity (Fogh-Andersen, 1942; Fraser, 1955; Marazita, 2012). This is because CL and CLP share the same anatomical defect in the lip, which develops prior to the palate. In addition, there are characteristic sex biases that distinguish CL/P and CP, where CP has a 2:1 female-to-male ratio and CL/P has a 2:1 male-to-female bias (Marazita, 2012). These epidemiological findings suggest that CP may be etiologically distinct from CL or CLP (Fraser, 1955); however, it is also worth considering other groupings as the prevalence of CLP (~0.45 per 1,000 live births) is higher than the prevalence for CL (~0.3 per 1,000 live births) or CP (~0.33 per 1,000 live births) (Salari *et al*., 2022). In addition, the types of OFCs within a family can support other groupings. There are many families with both CL and CLP, but others with only

one type (i.e., all CL). Some families have both CL/P and CP; such families are considered "mixed clefting" families (Rahimov *et al*., 2012), and have been more commonly observed with Mendelian OFC syndromes (e.g., families with Van der Woude syndrome) (Kondo *et al*., 2002; Rahimov *et al*., 2012; Rutledge *et al*., 2010; Schutte *et al*., 1993).

Phenotypic descriptions of OFCs can also include descriptions of cleft lip laterality or severity of a cleft palate. For example, CL/P can be further divided into unilateral and bilateral clefts (Gundlach & Maus, 2006; Marazita, 2012). Bilateral clefts occur at a lower frequency (24%) compared to unilateral clefts (24% – 52%) (Gundlach & Maus, 2006) and the estimated sibling recurrence risk for individuals with bilateral clefts is 1.5-2 times higher than the sibling recurrence risk for individuals with unilateral clefts, indicating that bilateral clefts occur more often in families than unilateral clefts (Grosen *et al*., 2010; Mitchell & Risch, 1993). These recurrence risk findings also suggest that unilateral and bilateral clefts could have distinct risk factors. Alternatively, unilateral clefts can also be divided into left-sided and right-sided unilateral clefts, where unilateral left CL/P occurs almost twice as frequently (52%) as unilateral right CL/P (24%) (Carlson, Taub, *et al*., 2017; Gundlach & Maus, 2006). The sibling recurrence risks of specific subtypes also vary by the laterality, where the sibling recurrence risk for unilateral CL is 1.4 compared to 2.5 for unilateral CLP; the same can be said for the sibling recurrence risk of bilateral CL (2.0) compared to bilateral CLP (4.6) (Grosen *et al*., 2010). Cleft palate can also affect different parts of the palate (hard and/or soft palate) with varying severities (overt or submucosal) (Butali *et al*., 2018; He *et al*., 2020). However, few genetic studies have considered these cleft palate subtypes (Ludwig *et al*., 2016; Tsuda *et al*., 2010). Ludwig *et al*. (2016) found a marker in *GREM1* associated with a two-fold risk increase in individuals with CL and a cleft of the soft palate compared to individuals with CL and cleft of the soft and hard

palates. These findings were supported by mouse expression of Grem1 in the soft palate (but not in the hard palate) and the identification of rare variants in *GREM1* in cases with a cleft soft palate (Ludwig *et al*., 2016), indicating that genetic risk factors can also modify the severity of clefts of the palate.

The phenotypic spectrum of OFCs can be further expanded to include very mild forms of OFCs, including microform cleft lip (Castilla & Martínez-Frías, 1995; Fujiwara *et al*., 2021; Yuzuriha & Mulliken, 2008), which resembles a scar, and submucous cleft palate, which occurs when the mucous membrane is closed but the muscles in the soft palate are clefted (Reiter *et al*., 2015; Reiter *et al*., 2012). Submucous cleft palate often includes a missing or bifid uvula (Sales *et al*., 2018; Shprintzen *et al*., 1985), which alone occurs at a frequency of 0.2% to 10% in the general population (Chosack & Eidelman, 1978; Feka *et al*., 2019). There is also evidence that the OFC spectrum may include subclinical phenotypes that are not visible to the eye (Marazita, 2007; Marazita, 2012; Martin *et al*., 1993; Neiswanger *et al*., 2007; Weinberg *et al*., 2008; Weinberg *et al*., 2006; Young & Spinner, 2022). These subclinical phenotypes include orbicularis oris muscle (OOM) discontinuities and velopharyngeal insufficiency (VPI). OOM discontinuities are a disorganization of the muscle fibers surrounding the mouth and are best identified through high-resolution ultrasound of the upper lip (Zhang *et al*., 2015). OOM defects were initially found as phenotypes in the lip from a histological study of cadaver dissections (Briedis & Jackson, 1981). These defects have been found at a higher frequency in 1st-degree relatives of CL cases and unaffected relatives of non-syndromic OFC cases compared to the general population (Martin *et al*., 2000; Neiswanger *et al*., 2007). VPI is caused by the incomplete closure of the velopharyngeal sphincter resulting in an unusual connection between the nasal and oral cavity that leads to hypernasal or hyponasal speech (Fisher & Sommerlad,

2011), and is assessed through speech evaluations by a speech pathologist or by videofluoroscopy (Rudnick & Sie, 2008; Smith & Kuehn, 2007). Although CP is a common cause of VPI, VPI can also occur in the absence of an overt cleft (Andres *et al*., 1981; Vantrappen *et al*., 2002; Young & Spinner, 2022). VPI has been observed at significantly higher rates in unaffected relatives from multiplex OFC families compared to controls (Weinberg *et al*., 2006). Because of the overlap between these phenotypes and normal phenotypic variation, some studies have hypothesized that the phenotypic spectrum of OFCs could be expanded to consider both OOM discontinuities and/or VPI as the "mildest" forms of OFCs (Marazita, 2007; Marazita, 2012; Martin *et al*., 1993; Neiswanger *et al*., 2007; Weinberg *et al*., 2008; Weinberg *et al*., 2006).

Syndromic and Non-Syndromic Forms of OFCs

Another way of classifying and grouping OFCs is by the presence or absence of additional major or minor structural anomalies, or developmental or cognitive disabilities. Individuals with overt clefts (i.e., CL, CLP, and CP) and additional anomalies or disabilities are commonly referred to as "syndromic". To date, more than 400 syndromes that include OFCs as a clinical feature have been documented (Medicine). Some of the more common OFC syndromes are listed in Table 1-1. When additional phenotypic features are recognized, individuals with suspected syndromic diagnoses often receive referrals for genetic counseling and clinical genetic testing to identify their causes and help establish proper clinical management. The causal mechanisms for syndromic OFCs are varied, although most are caused by chromosomal abnormalities (Davies *et al*., 1995; Shenoy *et al*., 2018; Zellweger *et al*., 1975), copy number variants (Conte *et al*., 2016; Lansdon *et al*., 2023; Simioni *et al*., 2015), or point mutations that

act in a Mendelian manner (Braybrook *et al*., 2001; Celli *et al*., 1999; Clifton-Bligh, Roderick J. *et al*., 1998; Hoornaert *et al*., 2010; Kalay *et al*., 2012; Kondo *et al*., 2002; Milunsky *et al*., 2008). Although there are hundreds of OFC syndromes, each syndrome individually has a low population prevalence. Van der Woude syndrome (VWS) is the most common OFC syndrome, occurring at a rate of 1 in 35,000 live births (Cervenka *et al*., 1967; Yow *et al*., 2021). However, in aggregate, ~ 30% of CL/P and 50% of CP cases are considered syndromic (Jones, 1988; Marazita, 2012; Tolarová & Cervenka, 1998).

Variable expressivity and incomplete penetrance are common within OFC families (Jugessur *et al*., 2009). These factors represent a major challenge for clinicians and research teams to accurately identify individuals with syndromes, interpret genetic variation, or predict disease in subsequent generations. For example, VWS accounts for ~2% of OFC cases (de Lima *et al*., 2009; Dronamraju, 1971), and is primarily caused by point mutations in *IRF6*. Although VWS is often diagnosed based on the presence of CL/P with lower lip pits (Schutte *et al*., 1993), it can also be diagnosed if the proband has an OFC and a 1st-degree relative with lip pits, or vice versa. Approximately ~15% of individuals with VWS lack lip pits (Cervenka *et al*., 1967; Schutte *et al*., 1993), making these individuals indistinguishable from those with isolated OFCs. Unsurprisingly, *IRF6* mutations have been identified in individuals with reportedly non-syndromic OFCs (Bishop *et al*., 2020; Leslie, E. J. *et al*., 2016; Rutledge *et al*., 2010).

What causes OFCs?

Environment in OFCs

OFCs are causally influenced by both genetics and the environment, indicating that the etiology is heterogeneous (Beames & Lipinski, 2020). Environmental exposures have been

associated with an increased OFC risk at various degrees, and these factors include maternal smoking, alcohol consumption, maternal intake of various medications, and exposure to teratogens, such as pesticides (Garland, Michael A. *et al*., 2020; Murray, 2002). Some environmental risk factors alone can be substantial causes for OFCs (e.g., antiseizure medications). However, the others are considered risk factors that may be one of many factors influencing the occurrence of OFCs, although the supporting evidence is often inconsistent. Meta-analyses of maternal smoking identified relative risks of 1.34 -1.37 for CL/P and 1.22-1.24 for CP (Little *et al*., 2004; Xuan *et al*., 2016), whereas exposure to second-hand smoke had relative risks of 1.14-1.87 for OFCs (Kummet *et al*., 2016; Oldereid *et al*., 2018; Zheng *et al*., 2019). Although less consistently associated, the data suggests that consuming substantial amounts of alcohol or "binge drinking" is associated with an increased risk for CL (OR:1.48), CP (OR:1.12), and OFCs overall (OR: 1.1-2.6) (Boyles *et al*., 2010; DeRoo *et al*., 2016). Maternal exposure to pesticides has also been associated with a significant increase in OFC risk (OR: 1.37-5.97) (Hao *et al*., 2015; Romitti *et al*., 2007; Yang *et al*., 2014). Other environmental factors associated with OFC risk include the intake of maternal medications during early pregnancy, such as antimicrobial (e.g., amoxicillin, ampicillin) (Cooper *et al*., 2009; Crider *et al*., 2009; Czeizel *et al*., 2001; Lin *et al*., 2012; Puhó *et al*., 2007) and antiepileptic medications (Etemad *et al*., 2012; Hernández-Díaz *et al*., 2000; Kerr *et al*., 2020; Rezaallah *et al*., 2019), folate antagonists (Hernández-Díaz *et al*., 2000), as well as medications for pregnancy complications, such as diazepam and phenobarbital (Aarskog, 1975; Puhó *et al*., 2007; Safra & Oakley, 1975). In contrast, pregnancy nutrition may be an important protective factor (Krapels *et al*., 2004). Although inconsistent, folic acid supplementation and multivitamin intake reduce the risk for OFCs with an odds ratio of 0.66-0.75 for folic acid (Bille *et al*., 2007; Jahanbin *et al*.,

2018; Wilcox *et al*., 2007) and 0.65-0.77 for multivitamin intake (Jahanbin *et al*., 2018; Mitchell *et al*., 2003; Wilcox *et al*., 2007), especially when administered early during pregnancy.

Genetics and Environment Contributions in OFCs

Additional association studies have been performed to evaluate the interaction of environmental and genetic risk factors for OFCs. Candidate gene by environment interaction (GxE) studies have identified significant interactions between maternal smoking or environmental smoking exposure and multivitamin intake and CL/P in several genes, including *BMP4*, *RUNX2*, and *TGFA* (Beaty *et al*., 2002; Hwang *et al*., 1995; Lin *et al*., 2010; Sull *et al*., 2009; Wu *et al*., 2012). These approaches have also been extended genome-wide and identified multiple significant or near-significant GxE associations for alcohol consumption, smoking, and multivitamin intake both for CP (Beaty *et al*., 2011; Haaland Ø *et al*., 2017; Wu *et al*., 2014) and for CL/P (Carlson *et al*., 2022; Haaland Ø *et al*., 2018; Haaland Ø *et al*., 2019). These findings collectively provided insight into additional mechanisms that influence risk for OFCs.

Epigenetics Contributions in OFCs

Another mechanism by which environmental factors could lead to OFC risk is via epigenetic modifications (Garland, M. A. *et al*., 2020; Seelan *et al*., 2019), leading to changes in gene expression without changes in the DNA sequence. As one example, maternal smoking was previously found to be associated with atypical DNA methylation epigenome-wide, altering the gene expression of genes involved in critical growth and development pathways (Suter *et al*., 2011). Joubert *et al*. (2016) also investigated the effect of maternal smoking on DNA methylation and found differentially methylated regions that mapped to OFC risk genes,

including *BMP4* and *TP63*. Epigenetic-wide association studies have also been performed in OFC cohorts in multiple populations, including the United Kingdom and Brazil. Alvizi *et al*. (2017) compared methylation differences in CLP cases compared to controls and identified 578 significant regions associated with CLP risk. Sharp *et al*. (2017) compared individuals with OFCs of different cleft types (i.e., CL vs. CLP, CP vs. CLP, and CL vs. CP) and found four, 17, and 294 differentially methylated regions for CL vs. CLP, CP vs. CLP, and CL vs. CP, respectively. Some of these methylated regions were mapped to known OFC-associated genes, such as *COL11A2* and *TBX1*, and pathways associated with craniofacial development, including the WNT Beta-Catenin signaling pathway (Alvizi *et al*., 2017; Sharp *et al*., 2017). Epigenetic studies of discordant monozygotic CLP twin pairs have conflicting results with some identifying differentially methylated regions that were located near known OFC genes (Young *et al*., 2021), whereas others (Kimani *et al*., 2007) have not found significant regions. These findings suggest a role of epigenetics in the risk for OFCs. However, there are some technical considerations involving epigenetic studies, including whether there is a relevant tissue at an appropriate time point to sample, given that methylation patterns may change throughout development and embryonic tissues relevant to OFCs or craniofacial development are inaccessible.

Genetic Contributions in OFCs

There is significant and compelling evidence pointing to the role of genetics in OFC risk (Christensen & Mitchell, 1996; Chung *et al*., 1986; Fraser, 1955; Marazita, 2002; Mitchell & Risch, 1993). One of the best early pieces of evidence for a genetic contribution to any phenotype is strong familial aggregation and the identification of families with multiple affected individuals (multiplex families) (Fogh-Andersen, 1942; Marazita *et al*., 1984; Rischbieth, 1910;

Sproule, 1863; Trew, 1757). In the early 20th century, OFCs were initially hypothesized to be a dominant disease that was inherited (Bateson, 1909; Melnick, 1997). As nicely summarized by Fogh-Andersen, early studies described the pattern of OFCs in collections of multiplex families (Fogh-Andersen, 1942). Some displayed a completely penetrant phenotype while other families were incompletely penetrant or displayed variability in the OFC subtype. In aggregate, the incomplete penetrance and variable expressivity did not support the conclusion that all OFCs were caused by a dominant "gene," but rather suggested that OFCs are heterogeneous in cause. In other words, some families supported a role for individual genes of large effect whereas others may be caused by some combination of genetic or environmental risk factors (Marazita, 2012). Formal segregation analyses agreed with this view of OFCs being a mixed model that had both a single gene and a polygenic component (Chung *et al*., 1986; Chung *et al*., 1974; Clementi *et al*., 1995; Hecht *et al*., 1991; Marazita *et al*., 1986; Marazita *et al*., 1984; Ray *et al*., 1993).

Epidemiological studies later helped dissect the potential differences in genetic architectures between OFC subtypes. In a population-based study in Norway, the relative risk for 1st-degree relatives was 56 times higher for CP and 32 times higher for CL/P compared to the general population risk, which indicates a strong genetic component (Sivertsen *et al*., 2008). Risk estimates are reduced as genetic distance increases, in a manner consistent with multifactorial etiology (Sivertsen *et al*., 2008). Estimating the "cross-over" risk for different OFC subtypes lent some support to the historical grouping of CL and CLP into CL/P. Although the recurrence risk was highest for the same type of OFC (e.g., CL after CL), the crossover risk for CL after CLP and vice versa was still higher (0.7 and 1%, respectively) than CP after CL/CLP or vice versa (0.1% and 0.2%) (Grosen *et al*., 2010). However, they also support the possibility of some shared genetic risk factors for subtypes of OFCs that had been categorized as having distinct

etiologies on the basis of developmental timing and epidemiology (Grosen *et al*., 2010). Similarly, Sivertsen *et al*. (2008) found that the recurrence risk for CP after the occurrence of CL and vice versa increased three-fold, which also helps explain the occurrence of "mixed clefting" within families. The familial component of OFCs has also been further supported by twin studies that have identified concordance rates of ~50-60% for monozygotic or identical twins compared to 8-10% concordance rates for dizygotic or fraternal twins (Christensen & Fogh-Andersen, 1993; Grosen *et al*., 2011), suggesting that OFCs are highly heritable (heritability ~90%) (Grosen *et al*., 2011), but are not exclusively caused by genetics.

Syndromic Gene Mapping

Various approaches have been used to map genes for the hundreds of OFC syndromes. These include linkage analyses, identification of chromosomal anomalies, and sequencing. The first Mendelian OFC syndrome locus was mapped to the region on Xq21 for CP and ankyloglossia (Moore *et al*., 1987). Next-generation sequencing studies have allowed rapid and direct mapping of syndromic OFC loci, including very rare OFC syndromes that preclude linkage analyses. Examples of syndromes mapped by whole-exome sequencing include Miller syndrome (Ng, Buckingham, *et al*., 2010), Batsocas-Papas syndrome (Mitchell *et al*., 2012), Van der Woude syndrome (Peyrard-Janvid *et al*., 2014), and Kabuki syndrome (Ng, Bigham, *et al*., 2010), among others (Leslie, 2022).

Genetics in Non-Syndromic OFCs

Many approaches have been used to identify genetic risk factors for OFCs. Because of the availability of large, multiplex families, genome-wide linkage studies were attempted.

However, none identified genome-wide significant results and there was little replication across studies (Mangold *et al*., 2009; Marazita *et al*., 2002; Marazita, M. L. *et al*., 2004; Prescott *et al*., 2000; Radhakrishna *et al*., 2006; Riley *et al*., 2007; Wang *et al*., 2010; Wyszynski *et al*., 2003; Zeiger *et al*., 2003). The inability to consistently replicate findings could have been due to limitations in the sample size in either the number of families or the sizes of any given family studied (Marazita, 2012). A meta-analysis in 2004 by Marazita, Mary L. *et al*. (2004) combining 13 linkage studies from non-syndromic CL/P finally yielded genome-wide significant results at 16 loci, including 1q32, 2q32-35, and 9q21. Fine-mapping of these regions identified the causal genes associated with non-syndromic OFCs at several of these regions, including *IRF6* at 1q32 (Marazita *et al*., 2009) and *FOXE1* at 9q21 (Marazita *et al*., 2009; Moreno *et al*., 2009). In addition to linkage analyses based on CL/P, genome-wide scans for CLP were performed (Riley *et al*., 2007), but only a candidate gene linkage study was done for CP (Koillinen *et al*., 2003), the latter of which was unsuccessful. Linkage studies (both successful and unsuccessful) supported a genetic heterogeneity of OFCs and hinted at the potential differences in genetic etiologies by cleft type.

Candidate gene studies are an approach in which a priori selected genes and loci are tested for association with disease. These include genes/loci implicated by previous genetic studies (e.g., linkage or association) analyses or through animal models. One significant benefit of candidate gene analyses is that they are not limited by the requirement of multiplex families. Candidate gene studies were used to study both common variants (Chiquet *et al*., 2007; Jezewski *et al*., 2003; Marazita *et al*., 2009; Zucchero *et al*., 2004) and rare variants (Jezewski *et al*., 2003; Lidral *et al*., 1998). Many studies reported significant associations, including *MSX1* (Beaty *et al*., 2002; Jezewski *et al*., 2003; Lidral *et al*., 1998), *MTHFR* (Blanton *et al*., 2000; Botto & Yang,

2000; Mills *et al*., 1999), and *TGFB3* (Beaty *et al*., 2002; Lidral *et al*., 1998), for CL/P and CP only. Although many candidate gene studies were performed, just as in linkage, there was little replication between studies (Marazita, 2012). Only *IRF6* was consistently associated with OFCs across many studies and populations (Ghassibé *et al*., 2005; Park *et al*., 2007; Zucchero *et al*., 2004).

Genome-wide association studies (GWASs) are an unbiased approach to simultaneously test the association of thousands of single nucleotide polymorphisms (SNPs) and a disease or trait. Most GWASs for OFCs have studied CL/P as large cohorts of CL/P cases and/or families had already been assembled for previous genetic studies. Most GWASs used case-control study designs of OFC cases and controls and identified significant regions at 1q32 and 8q24, among others (Birnbaum *et al*., 2009; Butali *et al*., 2018; Grant *et al*., 2009; He *et al*., 2020; Huang *et al*., 2019; Leslie, Elizabeth J., Carlson, Jenna C., *et al*., 2016; Leslie, Elizabeth J., Liu, Huan, *et al*., 2016; Li *et al*., 2022; Mangold *et al*., 2010; Sun *et al*., 2015; van Rooij *et al*., 2019; Yu *et al*., 2017). Other GWAS have family-based designs in case-parent trios (Beaty *et al*., 2010; Beaty *et al*., 2011; Leslie, Elizabeth J., Liu, Huan, *et al*., 2016) or extended pedigrees (Mukhopadhyay *et al*., 2021; Mukhopadhyay *et al*., 2022). In total, there have been 16 independent GWASs of OFCs and 4 meta-analyses (Leslie *et al*., 2017; Ludwig *et al*., 2016; Ludwig *et al*., 2012; van Rooij *et al*., 2019). Cumulatively OFC GWASs have: (1) identified over 50 loci associated with OFCs, (2) identified several loci that overlapped genes and loci identified by linkage and candidate gene studies (e.g., *IRF6*, *TFAP2A*, *TP63*), and (3) identified novel susceptibility loci associated with OFC risk.

The vast majority of GWASs have been on CL/P and other cleft types have been less frequently studied. Only 5 GWASs have reported results for CP (Beaty *et al*., 2011; Butali *et al*.,

2018; Carlson *et al*., 2019; Huang *et al*., 2019; Leslie, Elizabeth J., Liu, Huan, *et al*., 2016). Recent studies have focused on more specific OFC subtypes. These include CL (Carlson *et al*., 2019; Huang *et al*., 2019) and CLP (Carlson *et al*., 2019; Yu *et al*., 2017), which identified novel loci not previously seen in other studies of CL/P (e.g., *DLK1* and *GRM5* for CL, and, and *KRT18* and *WNT9B* for CLP) (Huang *et al*., 2019; Yu *et al*., 2017). However, it is difficult to disentangle true subtype differences from differences in statistical power or population. Several studies have found risk loci associated specifically with each cleft subtype, and other studies have found variants with different magnitudes of effect for CL and CLP (Carlson, Standley, *et al*., 2017; Rahimov *et al*., 2008). Given the evidence for shared and unique loci for OFC subtypes, a study by Carlson *et al*. (2019) investigated the combination of shared and subtype-specific loci by creating a "cleft map" of associations across and specific to the three subtypes. Carlson *et al*. (2019) reported specific subtype associations (e.g., *GRHL3* for CP, *MSX2* for CLP). Similar to Rahimov *et al*. (2008), a cluster of SNPs containing rs642961 (the SNP found in Rahimov *et al*. (2008)) on *IRF6* supported this region to be strongly associated with CL compared to CLP. These studies have found the contribution of shared and subtype-specific genetic factors to OFC risk.

Other GWASs have performed case vs. case analyses to identify modifiers of CL vs. CLP and the side and laterality of a cleft lip. Evidence of a modifier of CL vs. CLP could be found in Carlson, Standley, *et al*. (2017), where SNPs at 16q21 were found to be strongly associated with CL over CLP with a higher frequency of the allele in cases from families with only CL. In addition, Carlson, Taub, *et al*. (2017) found associations in laterality in CL/P, where SNPs in IRF6 were more strongly associated with unilateral CL/P than bilateral CL/P. The influence of genetic modifiers of laterality can also be found in CL only. A study by Curtis, Chang, Lee, *et al*.

(2021) found a significant modifier near *PAX1*, which was associated with an increased risk for bilateral CL compared to unilateral CL. Lastly, Curtis, Chang, Sun, *et al*. (2021) found a candidate region near *FAT4* that was associated with differences in sidedness (increased risk for left CL vs. right CL) of cleft lip. Although GWASs are widely used and have been relatively successful for OFCs compared to GWASs for other structural birth defects (Lupo *et al*., 2019), there are a number of areas of genetic research that need more work. GWASs and other related methods require large sample sizes to robustly identify significant signals when the effect sizes are modest. The requirement for large sample sizes makes studies of less prevalent cleft types or more specific cleft type definitions impractical with current datasets. GWASs also merely identify regions of the genome associated with disease. Causal variants may or may not be genotyped and there is no requirement that they be within the relevant gene. The majority of SNPs associated with OFCs occur in non-coding parts of the genome and the interpretation and linking of these signals to genes is difficult. Finally, GWASs collectively account for only 20-30% of the estimated heritability of OFCs (Beaty *et al*., 2016). Other sources of variation are currently being explored in the field of OFCs, including low-frequency or rare "private" variants, presumably with larger effect sizes.

Rare variant studies typically require sequencing technologies. Whole-exome sequencing (WES) and whole-genome sequencing (WGS) have been widely applied to study Mendelian diseases and are being increasingly applied to common diseases as technologies have improved and costs have decreased. Multiple groups have studied non-syndromic multiplex families with exome sequencing and found several *de novo* and inherited mutations that segregate in these families in a variety of genes (Aylward *et al*., 2016; Bureau *et al*., 2014; Cox *et al*., 2018; Hoebel *et al*., 2017; Liu *et al*., 2017; Pengelly *et al*., 2016). Collectively, clinically-relevant rare coding

variants may account for 9%-14% of multiplex families (Basha *et al*., 2018; Bureau *et al*., 2014; Cox *et al*., 2018). The sample sizes of these studies have been generally small, and cohorts have been selected in different ways, so the true proportion of multiplex families with causal rare variants is not yet known. Whole-genome sequencing is beginning to be used to explore rare and common single nucleotide variants, small insertions/deletions, and larger structural alterations (Awotoye, Mossey, Hetmanski, Gowans, Eshete, Adeyemo, Alade, Zeng, Adamson, James, *et al*., 2022; Mukhopadhyay *et al*., 2020). Because the available WGS cohorts consist of case-parent trios, initial studies have started with *de novo* mutations, narrowing the number of variants under investigation (Awotoye, Mossey, Hetmanski, Gowans, Eshete, Adeyemo, Alade, Zeng, Adamson, Naicker, *et al*., 2022; Bishop *et al*., 2020). Protein-altering *de novo* mutations were found to be enriched in 756 OFC probands (Bishop *et al*., 2020), and non-coding *de novo* mutations are enriched in established GWAS loci and candidate transcription factors (Zieger *et al*., 2023). It is likely that whole-genome sequencing will continue to uncover genes and variants important for OFC risk.

Overlap in Genetics Between Syndromic and Non-Syndromic OFCs

One of the major findings from genetic research is that the same genes in which mutations cause syndromic forms of OFCs are also associated with non-syndromic OFCs. For some genes, such as *IRF6* and *FOXE1*, rare coding variants cause Van der Woude syndrome (Kondo *et al*., 2002) and Bamforth-Lazarus syndrome (Clifton-Bligh, R. J. *et al*., 1998), respectively, whereas common variants in regulatory elements are associated with non-syndromic OFCs (Lidral *et al*., 2015; Rahimov *et al*., 2008). For *TP63*, missense mutations cause several allelic OFC syndromes (Celli *et al*., 1999; McGrath *et al*., 2001; van Bokhoven *et al*.,

2001) and truncating mutations are associated with incompletely penetrant non-syndromic OFCs (Khandelwal *et al*., 2019). In addition, common variants in a large regulatory element in the first intron of *TP63* were associated with non-syndromic CL/P (Leslie *et al*., 2017). Another gene found to be associated with syndromic and non-syndromic OFCs is *GRHL3*. Protein-altering mutations in *GRHL3* cause VWS (Peyrard-Janvid *et al*., 2014), whereas a common missense variant (Leslie, Elizabeth J., Liu, Huan, *et al*., 2016; Mangold *et al*., 2016) and protein-altering, rare (Eshete *et al*., 2018; Hoebel *et al*., 2017; Mangold *et al*., 2016) variants have been associated with non-syndromic CP. Interestingly, Moreno Uribe *et al*. (2017) found that common variants found through GWASs of non-syndromic OFCs are associated in the same direction and magnitude of effect in a population of individuals with non-isolated OFCs, strengthening the support for the etiological overlap between these two phenotypic designations (Moreno Uribe *et al*., 2017).

Rare variants have also been found in non-syndromic families in multiple OFC-associated genes, including *TP63* (Basha *et al*., 2018; Khandelwal *et al*., 2019), *IRF6* (Leslie, E. J. *et al*., 2016), *CTNND1* (Cox *et al*., 2018), *GRHL3* (Eshete *et al*., 2018; Hoebel *et al*., 2017; Mangold *et al*., 2016), and *COL2A1* (Lace *et al*., 2022) (Table 1-1). In aggregate, Bishop *et al*. (2020) found 6% of 756 OFC cases had a *de novo* mutation in genes previously associated with OFCs. There was a particularly strong enrichment of *de novo* mutations among genes associated with autosomal dominant forms of OFCs. Taken together, these findings are evidence of the contribution of rare variants in a proportion of isolated OFC cases in genes previously associated solely with Mendelian OFC syndromes.

A Phenotypic Spectrum from Syndromic to Non-Syndromic OFCs

It is logical to categorize OFC cases into syndromic and non-syndromic based on their phenotypic presentation and family history. However, clearly grouping cases in this way can be difficult to accomplish because of the variable expressivity and incomplete penetrance of clinical features (Jugessur *et al*., 2009). The fact that rare, coding variants in certain genes can be found in either phenotypically syndromic or non-syndromic cases suggests that it is not possible to assume a genetic model for OFC causality from phenotype alone. This has implications for researchers and clinicians. In research, not recognizing possible etiologic heterogeneity in a sample could lead to either false negative or false positive results in genetic studies. Clinically, individuals that are not suspected to have a Mendelian syndrome are rarely offered genetic testing.

There are several phenotypes that are described as part of OFC syndromes but also occur as part of the normal range of phenotypic variation. Many OFC genes are known to regulate tooth development in addition to lip and palate development (Phan *et al*., 2016). Dental anomalies, such as tooth agenesis, microdontia, or oligodontia, are associated with several OFC syndromes, including VWS (Schinzel & Kläusler, 1986; Schutte *et al*., 1993). They are also common in individuals with OFCs in general (Haque & Alam, 2015; Nicholls, 2016; Vieira *et al*., 2008b). Multiple studies have found that dental anomalies, such as microdontia, teeth malposition or transposition, developmental enamel defects, hypodontia, tooth rotation, and supernumerary teeth, were present more frequently in OFC cases than in controls (Eerens *et al*., 2001; Letra *et al*., 2007; Marzouk *et al*., 2021; Schroeder & Green, 1975; Shapira *et al*., 1999). In support of an overlap between the etiologies of dental anomalies and OFCs, linkage and association studies have identified several loci that contribute to both OFCs and dental anomalies

(Vieira *et al*., 2008a, 2008b). Variants in and around *AXIN2*, *CDH1*, *MSX1*, and *PAX9* are associated with both OFCs and tooth agenesis (Das *et al*., 2003; Letra *et al*., 2009; Phan *et al*., 2016; Seo *et al*., 2013; Slayton *et al*., 2003; van den Boogaard *et al*., 2000).

Subclinical phenotypes may also be useful in classifying OFCs for genetic studies. OOM discontinuities have been observed in a significant proportion of non-syndromic CP cases (Weinberg *et al*., 2008). If an OOM is a form of a cleft lip (as hypothesized), these individuals may be better classified as CLP instead of CP. In addition to OOM defects, the cleft lip spectrum also includes microform CL. Missense and nonsense mutations in *BMP4* were found at a significantly higher frequency in individuals with overt or microform CL or OOM defects than in controls (Suzuki *et al*., 2009). Although no genome-wide studies have been published yet on OOM discontinuities, these data provide modest evidence that OOM defects may share an etiologic spectrum with OFCs. The genetic relationship between OFCs and VPI is less clear. Although common genetic variants associated with VPI occur near genes involved in craniofacial development, they are not the variants or loci identified to date from genome-wide studies of OFCs (Chernus *et al*., 2018).

It is also clear that the same genes that are associated with OFCs also contribute to normal facial shape. One of the first GWAS of facial shape by Liu *et al*. (2012) found significant associations between candidate SNPs in OFC genes, such as *PAX3*, *PRDM16*, and *TP63*, and horizontal facial width. A GWAS by White *et al.* 2021 found an enrichment of lead SNPs surrounding genes associated with craniofacial development and orofacial clefting. A subset of these candidate genes was found to be positively associated with different facial features, such as *NOG* and philtrum shape, and *MSX1* with nose shape (Indencleef *et al*., 2018). More recently, Indencleef *et al*. (2021) found 29 genome-wide significant loci by comparing unaffected parents

of OFC cases and controls, where 22 loci were associated with normal face variation, 16 loci were in/near genes strongly associated with OFCs, and 12 loci had weaker associations with OFCs. These findings suggest that unaffected family members could harbor a subset of genetic risk factors as their affected offspring, and thus, might be at a higher risk of expressing differences in facial shape patterns compared to the general population (Weinberg, 2022). Taken together, these data provide support for the diverse role of OFC genes and shared genetic architecture of facial development and OFCs, which suggest that refining phenotypes to include facial shape variation could be meaningful for genotype-phenotype correlations.

Conclusion

Like many human diseases, OFCs are phenotypically and etiologically heterogeneous. Genetic studies indicate that an etiologic distinction between syndromic and non-syndromic is less clear than once hypothesized. One way of conceptualizing this is to consider OFC risk on a liability scale in which some number of risk factors (whether genetic or environmental) contribute to the risk. The liability threshold makes no requirements as to the type or frequency of risk factor(s); all affected individuals have surpassed the threshold to develop an OFC because of some unknown combination of variants. In the case of Mendelian forms, that is typically a single heterozygous mutation and disease expression could be modified by other, as yet unknown, variants. In the case of individuals with non-syndromic OFCs, GWASs and rare variant studies tell us that some fraction of affected individuals are likely to have a polygenic burden of common variants while others may have a collection of higher effect rare variants. Classifying an individual as syndromic or non-syndromic is, therefore, more akin to a differential

diagnosis that helps to prioritize possible genetic etiologies. This can be helpful when designing genetic studies that are limited by time and resources.

It is clear that much of the genetic heterogeneity of OFCs is unknown and this is especially true when considering more specific OFC types (e.g., laterality, severity) and the variable expressivity that is characteristic of most OFC syndromes. It is therefore important that research study designs are able to capture detailed and precise phenotype information at the individual and family levels. Although there is power in collecting large case-control cohorts, previous OFC research makes it clear that family history and sampling from multiple affected individuals is important to identify and prioritize rare likely causal variants. However, most family-based studies have used relatively small sample sizes. Thus, conducting deep phenotyping and considering additional family history information using larger study cohorts will be useful approaches for sequencing-based studies to fully understand the role of rare or novel genetic variants in the etiology of OFCs. In this thesis, we considered the value of deep phenotyping and using various family structures and histories in studying rare variants.

Tables

Table 1-1: Known Genes Associated with Syndromic and Non-Syndromic OFCs

Genes	Syndromes	Prevalence	Phenotypes	Common Variants in Non-Syndromic OFCs	Rare Variants in Non-Syndromic OFCs
COL2A1	Stickler Syndrome (Francomano C *et al.* (1987); Ahmad N *et al.* (1990))	1:7,500 - 1:9,000	Myopia, retinal detachment, CP, cataracts, hearing loss, hyperflexible or enlarged joints, mid-facial flatness	Nikopensius T *et al.* (2010)	Bishop M *et al.* (2020); Lace B *et al.* (2022)
CTNND1	Blepharocheilodontic Syndrome (Ghoumid J *et al.* (2017))	1:1,000,000	Eyelid anomalies, CL/P, ectodermal dysplasia	-	Bishop M *et al.* (2020); Cox L *et al.* (2018)
FOXE1	Bamforth-Lazarus Syndrome (Clifton-Bligh R *et al.* (1998))	1:1,000,000	Congenital hypothryoidism, CP, spiky hair, choanal atresia	Marazita M *et al.* (2009); Moreno L *et al.* (2009); Lennon C *et al.* (2012); Ludwig K *et al.* (2014) Lidral A *et al*. (2015); Leslie E *et al*. (2017); Carlson J *et al*. (2019); Xiao *W et al*. (2020)	-
IRF6	Van der Woude Syndrome (VWS) (Kondo S *et al.* (2002); Popliteal Pterygium Syndrome (PPS) (Kondo S *et al.* (2002))	1:35,000 (VWS); 1:300,000 (PPS)	CL/P, lip pits, hypodontia (VWS); Lower lip pits, CL/P, sygnathia, skin and genital abnormalities (webbing of limbs, hypoplasia of labia majora) (PPS)	Zucchero T *et al.* (2004); Scapoli L *et al.* (2005); Blanton S *et al.* (2005); Park J *et al.* (2007); Rahimov F *et al.* (2008); Marazita M *et al.* (2009); Beaty T *et al.* (2010); Nikopensius T *et al.* (2010); Leslie *et al.* (2016); Carlson J *et al*. (2019);	Rutledge K *et al.* (2010); Leslie *et al*. (2016); Zhao H *et al.* (2018); Bishop M *et al.* (2020); Wang Y *et al.* (2021)

				Mukhopadhyay N *et al*. (2022)	
GRHL3	Van der Woude Syndrome (Koillinen H *et al.* (2001); Peyrard-Janvid M *et al.* (2014))	1:35,000	CL/P, lip pits, hypodontia	Leslie E *et al.* (2016); Wang *et al.* (2016); Mangold E *et al.* (2016); Carlson J *et al*. (2019)	Mangold E *et al.* (2016); Hoebel A *et al.* (2017); Eshete M *et al.* (2018); Basha M *et al.* (2018): Bishop M *et al.* (2020); Huang W *et al.* (2022)
TP63	Ankyloblepharon-Ectodermal Defects-Cleft Lip/Palate (AEC) Syndrome (McGrath J *et al.* (2001)); Ectrodactyly, Ectodermal Dysplasia, and Cleft Lip/Palate (EEC) Syndrome (Celli J *et al.* (1999); van Bokhoven H *et al*. (1999); van Bokhoven H *et al.* (2001))	1:1,000,000 (AEC); 1:90,000 (EEC)	Ankyloblepharon, ectodermal dysplasia, CL/P, split-hand/foot malformation, syndactyly, hypodontia, hypopigmentation (AEC); CL/P, ectrodactyly of hands and feet, ectodermal dysplasia, hypopigmentation (EEC)	Leslie E *et al.* (2017)	Leoyklang P *et al.* (2006); Basha M *et al.* (2018); Khandelwal K *et al.* (2019); Xu *et al.* (2021); Awotoye *et al.* (2022)

Figures

Figure 1-1: Schematic Drawing of the Development of the Face. (A) Facial development involves a series of growth, migration, and fusion of cells that start with the migration of neural crest cells to the face. These cells along with mesoderm cells form five facial prominences (i.e., frontonasal prominence, paired mandibular processes, and paired maxillary processes in week 4th of development. (B) The frontonasal prominence divides into the medial and lateral nasal processes. (C and D) The medial and lateral nasal processes fuse with the maxillary processes to form the upper lip and primary palate. (D) In the 6th week of embryonic development, the palatal shelves will grow vertically along the side of the tongue with the maxillary processes to form the secondary palate. (E) The palatal shelves then elevate and fuse at the midline epithelial seam to establish the soft and hard palates. (F) Finally, the nasal septum and primary palate fuse to distinguish the nose and oral cavity. Adapted with copyright permission from *Springer Nature* (License Number: 5475500419170): Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: synthesizing genetic and environmental influences. Nature Reviews Genetics, 2011.

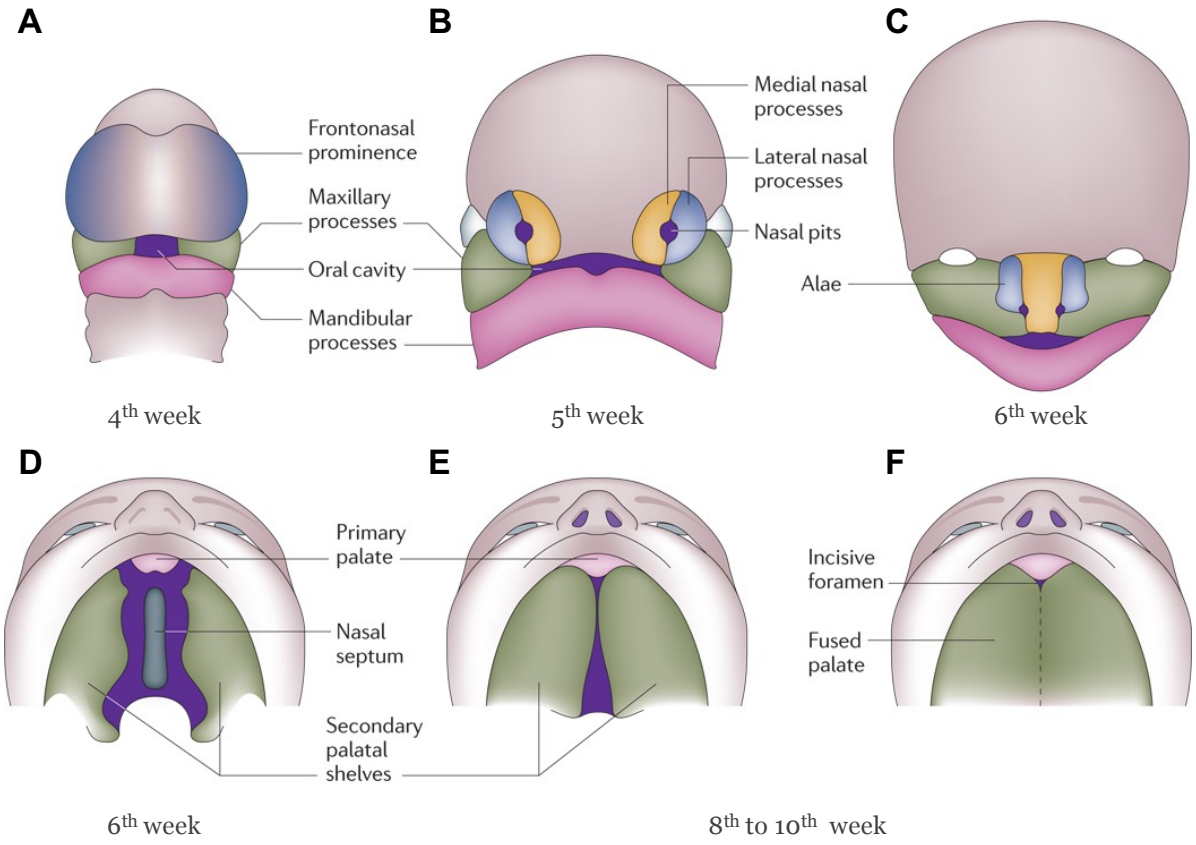
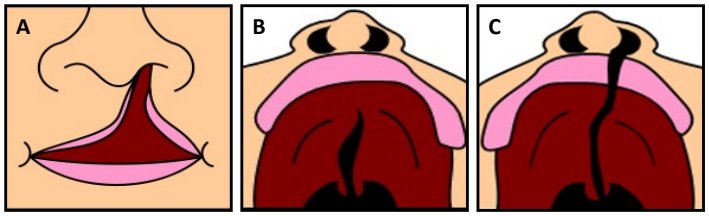


Figure 1-2: A photomontage of the most common orofacial cleft subtypes. These subtypes include (A) Cleft Lip (CL) only, (B) Cleft Palate (CP) only, and (C) Cleft Lip with Cleft Palate (CLP).



References

Aarskog, D. (1975). Letter: Association between maternal intake of diazepam and oral clefts. Lancet, 2(7941), 921. doi:10.1016/s0140-6736(75)92153-4

Alvizi, L., Ke, X., Brito, L. A., Seselgyte, R., Moore, G. E., Stanier, P., & Passos-Bueno, M. R. (2017). Differential methylation is associated with non-syndromic cleft lip and palate and contributes to penetrance effects. Scientific Reports, 7(1), 2441. doi:10.1038/s41598-017-02721-0

Andres, R., Bixler, D., Shanks, J. C., & Smith, W. L. (1981). Dominant inheritance of velopharyngeal incompetence. Clinical Genetics, 19(6), 443-447. doi:10.1111/j.1399-0004.1981.tb02062.x

Awotoye, W., Mossey, P. A., Hetmanski, J. B., Gowans, L. J. J., Eshete, M. A., Adeyemo, W. L., . . . Butali, A. (2022). Damaging Mutations in AFDN Contribute to Risk of Nonsyndromic Cleft Lip With or Without Cleft Palate. Cleft Palate Craniofac J, 10556656221135926. doi:10.1177/10556656221135926

Awotoye, W., Mossey, P. A., Hetmanski, J. B., Gowans, L. J. J., Eshete, M. A., Adeyemo, W. L., . . . Butali, A. (2022). Whole-genome sequencing reveals de-novo mutations associated with nonsyndromic cleft lip/palate. Scientific Reports, 12(1), 11743. doi:10.1038/s41598-022-15885-1

Aylward, A., Cai, Y., Lee, A., Blue, E., Rabinowitz, D., & Haddad, J., Jr. (2016). Using Whole Exome Sequencing to Identify Candidate Genes With Rare Variants In Nonsyndromic Cleft Lip and Palate. Genetic Epidemiology, 40(5), 432-441. doi:10.1002/gepi.21972

Basha, M., Demeer, B., Revencu, N., Helaers, R., Theys, S., Bou Saba, S., . . . Vikkula, M. (2018). Whole exome sequencing identifies mutations in 10% of patients with familial non-syndromic cleft lip and/or palate in genes mutated in well-known syndromes. J Med Genet, 55(7), 449-458. doi:10.1136/jmedgenet-2017-105110

Bateson, W. (1909). Mendel's Principles of Heredity. Cambridge Univ. Press.

Beames, T. G., & Lipinski, R. J. (2020). Gene-environment interactions: aligning birth defects research with complex etiology. Development, 147(21). doi:10.1242/dev.191064

Beaty, T., Marazita, M., & Leslie, E. (2016). Genetic factors influencing risk to orofacial clefts: today?s challenges and tomorrow?s opportunities [version 1; peer review: 2 approved]. F1000Research, 5(2800). doi:10.12688/f1000research.9503.1

Beaty, T. H., Hetmanski, J. B., Zeiger, J. S., Fan, Y. T., Liang, K. Y., VanderKolk, C. A., & McIntosh, I. (2002). Testing candidate genes for non-syndromic oral clefts using a case-parent trio design. Genetic Epidemiology, 22(1), 1-11. doi:10.1002/gepi.1039

Beaty, T. H., Murray, J. C., Marazita, M. L., Munger, R. G., Ruczinski, I., Hetmanski, J. B., . . . Scott, A. F. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nature Genetics, 42(6), 525-529. doi:10.1038/ng.580

Beaty, T. H., Ruczinski, I., Murray, J. C., Marazita, M. L., Munger, R. G., Hetmanski, J. B., . . . Scott, A. F. (2011). Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. Genetic Epidemiology, 35(6), 469-478. doi:10.1002/gepi.20595

Berk, N., & Marazita, M. L. (2002). Costs of Cleft Lip and Palate: Personal and Societal Implications. In O. U. Press (Ed.), Cleft Lip & Palate: From Origin to Treatment (1st ed., pp. 458-467). New York.

Bille, C., Olsen, J., Vach, W., Knudsen, V. K., Olsen, S. F., Rasmussen, K., . . . Christensen, K. (2007). Oral clefts and life style factors--a case-cohort study based on prospective Danish data. European Journal of Epidemiology, 22(3), 173-181. doi:10.1007/s10654-006-9099-5

Bille, C., Winther, J. F., Bautz, A., Murray, J. C., Olsen, J., & Christensen, K. (2005). Cancer risk in persons with oral cleft--a population-based study of 8,093 cases. American Journal of Epidemiology, 161(11), 1047-1055. doi:10.1093/aje/kwi132

Birnbaum, S., Ludwig, K. U., Reutter, H., Herms, S., Steffens, M., Rubini, M., . . . Mangold, E. (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nature Genetics, 41(4), 473-477. doi:10.1038/ng.333

Bishop, M. R., Diaz Perez, K. K., Sun, M., Ho, S., Chopra, P., Mukhopadhyay, N., . . . Leslie, E. J. (2020). Genome-wide Enrichment of De Novo Coding Mutations in Orofacial Cleft Trios. The American Journal of Human Genetics. doi:10.1016/j.ajhg.2020.05.018

Blanton, S. H., Kolle, B. S., Hecht, J. T., Mulliken, J. B., & Martin, E. R. (2000). No evidence supporting MTHFR as a risk factor in the development of familial NSCLP. American Journal of Medical Genetics, 92(5), 370-371. doi:10.1002/1096-8628(20000619)92:5<370::aid-ajmg17>3.0.co;2-4

Botto, L. D., & Yang, Q. (2000). 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. American Journal of Epidemiology, 151(9), 862-877. doi:10.1093/oxfordjournals.aje.a010290

Boulet, S. L., Grosse, S. D., Honein, M. A., & Correa-Villaseñor, A. (2009). Children with orofacial clefts: health-care use and costs among a privately insured population. Public Health Reports, 124(3), 447-453. doi:10.1177/003335490912400315

Boyles, A. L., DeRoo, L. A., Lie, R. T., Taylor, J. A., Jugessur, A., Murray, J. C., & Wilcox, A. J. (2010). Maternal Alcohol Consumption, Alcohol Metabolism Genes, and the Risk of Oral Clefts: A Population-based Case-Control Study in Norway, 1996–2001. American Journal of Epidemiology, 172(8), 924-931. doi:10.1093/aje/kwq226

Braybrook, C., Doudney, K., Marçano, A. C., Arnason, A., Bjornsson, A., Patton, M. A., . . . Stanier, P. (2001). The T-box transcription factor gene TBX22 is mutated in X-linked cleft palate and ankyloglossia. Nature Genetics, 29(2), 179-183. doi:10.1038/ng730

Briedis, J., & Jackson, I. T. (1981). The anatomy of the philtrum: observations made on dissections in the normal lip. British Journal of Plastic Surgery, 34(2), 128-132. doi:10.1016/s0007-1226(81)80077-x

Bureau, A., Parker, M. M., Ruczinski, I., Taub, M. A., Marazita, M. L., Murray, J. C., . . . Beaty, T. H. (2014). Whole exome sequencing of distant relatives in multiplex families implicates rare variants in candidate genes for oral clefts. Genetics, 197(3), 1039-1044. doi:10.1534/genetics.114.165225

Butali, A., Mossey, P. A., Adeyemo, W. L., Eshete, M. A., Gowans, L. J. J., Busch, T. D., . . . Adeyemo, A. A. (2018). Genomic analyses in African populations identify novel risk loci for cleft palate. Human Molecular Genetics, 28(6), 1038-1051. doi:10.1093/hmg/ddy402

Carlson, J. C., Anand, D., Butali, A., Buxo, C. J., Christensen, K., Deleyiannis, F., . . . Leslie, E. J. (2019). A systematic genetic analysis and visualization of phenotypic heterogeneity among orofacial cleft GWAS signals. Genetic Epidemiology, 43(6), 704-716. doi:10.1002/gepi.22214

Carlson, J. C., Shaffer, J. R., Deleyiannis, F., Hecht, J. T., Wehby, G. L., Christensen, K., . . . Leslie, E. J. (2022). Genome-wide Interaction Study Implicates VGLL2 and Alcohol

Exposure and PRL and Smoking in Orofacial Cleft Risk. Front Cell Dev Biol, 10, 621261. doi:10.3389/fcell.2022.621261

Carlson, J. C., Standley, J., Petrin, A., Shaffer, J. R., Butali, A., Buxó, C. J., . . . Leslie, E. J. (2017). Identification of 16q21 as a modifier of nonsyndromic orofacial cleft phenotypes. Genetic Epidemiology, 41(8), 887-897. doi:10.1002/gepi.22090

Carlson, J. C., Taub, M. A., Feingold, E., Beaty, T. H., Murray, J. C., Marazita, M. L., & Leslie, E. J. (2017). Identifying Genetic Sources of Phenotypic Heterogeneity in Orofacial Clefts by Targeted Sequencing. Birth defects research, 109(13), 1030-1038. doi:10.1002/bdr2.23605

Castilla, E. E., & Martínez-Frías, M. L. (1995). Congenital healed cleft lip. American Journal of Medical Genetics, 58(2), 106-112. doi:10.1002/ajmg.1320580203

Celli, J., Duijf, P., Hamel, B. C., Bamshad, M., Kramer, B., Smits, A. P., . . . van Bokhoven, H. (1999). Heterozygous germline mutations in the p53 homolog p63 are the cause of EEC syndrome. Cell, 99(2), 143-153. doi:10.1016/s0092-8674(00)81646-3

Cervenka, J., Gorlin, R. J., & Anderson, V. E. (1967). The syndrome of pits of the lower lip and cleft lip and/or palate. Genetic considerations. Am J Hum Genet, 19(3 Pt 2), 416-432.

Chernus, J., Roosenboom, J., Ford, M., Lee, M. K., Emanuele, B., Anderton, J., . . . Marazita, M. L. (2018). GWAS reveals loci associated with velopharyngeal dysfunction. Scientific Reports, 8(1), 8470. doi:10.1038/s41598-018-26880-w

Chiquet, B. T., Lidral, A. C., Stal, S., Mulliken, J. B., Moreno, L. M., Arcos-Burgos, M., . . . Hecht, J. T. (2007). CRISPLD2: a novel NSCLP candidate gene. Hum Mol Genet, 16(18), 2241-2248. doi:10.1093/hmg/ddm176

Chosack, A., & Eidelman, E. (1978). Cleft uvula: prevalence and genetics. Cleft Palate Journal, 15(1), 63-67.

Christensen, K., & Fogh-Andersen, P. (1993). Cleft lip (+/- cleft palate) in Danish twins, 1970-1990. American Journal of Medical Genetics, 47(6), 910-916. doi:10.1002/ajmg.1320470620

Christensen, K., Juel, K., Herskind, A. M., & Murray, J. C. (2004). Long term follow up study of survival associated with cleft lip and palate at birth. BMJ, 328(7453), 1405. doi:10.1136/bmj.38106.559120.7C

Christensen, K., & Mitchell, L. E. (1996). Familial recurrence-pattern analysis of nonsyndromic isolated cleft palate--a Danish Registry study. Am J Hum Genet, 58(1), 182-190.

Christensen, K., & Mortensen, P. B. (2002). Facial Clefting and Psychiatric Diseases: A Follow-Up of the Danish 1936–1987 Facial Cleft Cohort. The Cleft Palate Craniofacial Journal, 39(4), 392-396. doi:10.1597/1545-1569_2002_039_0392_fcapda_2.0.co_2

Chung, C. S., Bixler, D., Watanabe, T., Koguchi, H., & Fogh-Andersen, P. (1986). Segregation analysis of cleft lip with or without cleft palate: a comparison of Danish and Japanese data. Am J Hum Genet, 39(5), 603-611. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1684048/pdf/ajhg00147-0047.pdf

Chung, C. S., Ching, G. H., & Morton, N. E. (1974). A genetic study of cleft lip and palate in Hawaii. II. Complex segregation analysis and genetic risks. Am J Hum Genet, 26(2), 177-188.

Clementi, M., Tenconi, R., Collins, A., Calzolari, E., & Milan, M. (1995). Complex segregation analysis in a sample of consecutive newborns with cleft lip with or without cleft palate in Italy. Human Heredity, 45(3), 157-164. doi:10.1159/000154277

Clifton-Bligh, R. J., Wentworth, J. M., Heinz, P., Crisp, M. S., John, R., Lazarus, J. H., . . . Chatterjee, V. K. (1998). Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. Nature Genetics, 19(4), 399-401. doi:10.1038/1294
Clifton-Bligh, R. J., Wentworth, J. M., Heinz, P., Crisp, M. S., John, R., Lazarus, J. H., . . . Chatterjee, V. K. (1998). Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. Nature Genetics, 19(4), 399-401. doi:10.1038/1294
Conte, F., Oti, M., Dixon, J., Carels, C. E., Rubini, M., & Zhou, H. (2016). Systematic analysis of copy number variants of a large cohort of orofacial cleft patients identifies candidate genes for orofacial clefts. Hum Genet, 135(1), 41-59. doi:10.1007/s00439-015-1606-x

Cooper, M. E., Ratay, J. S., & Marazita, M. L. (2006). Asian oral-facial cleft birth prevalence. Cleft Palate Craniofac J, 43(5), 580-589. doi:10.1597/05-167

Cooper, W. O., Hernandez-Diaz, S., Arbogast, P. G., Dudley, J. A., Dyer, S. M., Gideon, P. S., . . . Ray, W. A. (2009). Antibiotics potentially used in response to bioterrorism and the risk of major congenital malformations. Paediatric and Perinatal Epidemiology, 23(1), 18-28. doi:10.1111/j.1365-3016.2008.00978.x

Cox, L. L., Cox, T. C., Moreno Uribe, L. M., Zhu, Y., Richter, C. T., Nidey, N., . . . Roscioli, T. (2018). Mutations in the Epithelial Cadherin-p120-Catenin Complex Cause Mendelian Non-Syndromic Cleft Lip with or without Cleft Palate. Am J Hum Genet, 102(6), 1143-1157. doi:10.1016/j.ajhg.2018.04.009

Crider, K. S., Cleves, M. A., Reefhuis, J., Berry, R. J., Hobbs, C. A., & Hu, D. J. (2009). Antibacterial medication use during pregnancy and risk of birth defects: National Birth Defects Prevention Study. Arch Pediatr Adolesc Med, 163(11), 978-985. doi:10.1001/archpediatrics.2009.188

Curtis, S. W., Chang, D., Lee, M. K., Shaffer, J. R., Indencleef, K., Epstein, M. P., . . . Leslie, E. J. (2021). The PAX1 locus at 20p11 is a potential genetic modifier for bilateral cleft lip. HGG Adv, 2(2). doi:10.1016/j.xhgg.2021.100025

Curtis, S. W., Chang, D., Sun, M. R., Epstein, M. P., Murray, J. C., Feingold, E., . . . Leslie, E. J. (2021). FAT4 identified as a potential modifier of orofacial cleft laterality. Genetic Epidemiology, 45(7), 721-735. doi:10.1002/gepi.22420

Czeizel, A. E., Rockenbauer, M., Sørensen, H. T., & Olsen, J. (2001). A population-based case-control teratologic study of ampicillin treatment during pregnancy. American Journal of Obstetrics and Gynecology, 185(1), 140-147. doi:10.1067/mob.2001.113907

Das, P., Hai, M., Elcock, C., Leal, S. M., Brown, D. T., Brook, A. H., & Patel, P. I. (2003). Novel missense mutations and a 288-bp exonic insertion in PAX9 in families with autosomal dominant hypodontia. Am J Med Genet A, 118a(1), 35-42. doi:10.1002/ajmg.a.10011

Davies, A. F., Stephens, R. J., Olavesen, M. G., Heather, L., Dixon, M. J., Magee, A., . . . Ragoussis, J. (1995). Evidence of a locus for orofacial clefting on human chromosome 6p24 and STS content map of the region. Hum Mol Genet, 4(1), 121-128. doi:10.1093/hmg/4.1.121

de Lima, R. L., Hoper, S. A., Ghassibe, M., Cooper, M. E., Rorick, N. K., Kondo, S., . . . Schutte, B. C. (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(4), 241-247. doi:10.1097/GIM.0b013e318197a49a

DeRoo, L. A., Wilcox, A. J., Lie, R. T., Romitti, P. A., Pedersen, D. A., Munger, R. G., . . . 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. European Journal of Epidemiology, 31(10), 1021-1034. doi:10.1007/s10654-016-0171-5

Dietz, A., Pedersen, D. A., Jacobsen, R., Wehby, G. L., Murray, J. C., & Christensen, K. (2012). Risk of breast cancer in families with cleft lip and palate. Annals of Epidemiology, 22(1), 37-42. doi:10.1016/j.annepidem.2011.09.003

Dixon, M. J., Marazita, M. L., Beaty, T. H., & Murray, J. C. (2011). Cleft lip and palate: understanding genetic and environmental influences. Nature Reviews: Genetics, 12(3), 167-178. doi:10.1038/nrg2933

Dronamraju, K. R. (1971). Genetic studies of a cleft palate clinic population. Birth Defects Original Article Series, 7(7), 54-57.

Eerens, K., Vlietinck, R., Heidbüchel, K., Van Olmen, A., Derom, C., Willems, G., & Carels, C. (2001). Hypodontia and tooth formation in groups of children with cleft, siblings without cleft, and nonrelated controls. Cleft Palate Craniofac J, 38(4), 374-378. doi:10.1597/1545-1569_2001_038_0374_hatfig_2.0.co_2

Eshete, M. A., Liu, H., Li, M., Adeyemo, W. L., Gowans, L. J. J., Mossey, P. A., . . . Butali, A. (2018). Loss-of-Function GRHL3 Variants Detected in African Patients with Isolated Cleft Palate. J Dent Res, 97(1), 41-48. doi:10.1177/0022034517729819

Etemad, L., Moshiri, M., & Moallem, S. A. (2012). Epilepsy drugs and effects on fetal development: Potential mechanisms. Journal of Research in Medical Sciences, 17(9), 876-881.

Feka, P., Banon, J., Leuchter, I., & La Scala, G. C. (2019). Prevalence of bifid uvula in primary school children. International Journal of Pediatric Otorhinolaryngology, 116, 88-91. doi:10.1016/j.ijporl.2018.10.026

Ferguson, C. A., Tucker, A. S., & Sharpe, P. T. (2000). Temporospatial cell interactions regulating mandibular and maxillary arch patterning. Development, 127(2), 403-412. doi:10.1242/dev.127.2.403

Fisher, D. M., & Sommerlad, B. C. (2011). Cleft lip, cleft palate, and velopharyngeal insufficiency. Plastic and Reconstructive Surgery, 128(4), 342e-360e. doi:10.1097/PRS.0b013e3182268e1b

Fogh-Andersen, P. (1942). Inheritance of Harelip and Cleft Palate: Contribution to the Elucidation of the Etiology of the Congenital Clefts of the Face. Journal of the American Medical Association, 133(4), 276. doi:10.1001/jama.1947.02880040062031

Fraser, F. C. (1955). Thoughts on the etiology of clefts of the palate and lip. Acta Genetica et Statistica Medica, 5(4), 358-369. doi:10.1159/000150783

Fujiwara, K., Yoshida, M., Nakamichi, N., Saitoh, S., Takaichi, M., Ishizaka, R., . . . Noguchi, M. (2021). Mini-microform cleft lip with complete cleft alveolus and palate: A case report. Congenit Anom (Kyoto), 61(4), 133-137. doi:10.1111/cga.12415

Garland, M. A., Reynolds, K., & Zhou, C. J. (2020). Environmental mechanisms of orofacial clefts. Birth defects research, 112(19), 1660-1698. doi:10.1002/bdr2.1830

Garland, M. A., Sun, B., Zhang, S., Reynolds, K., Ji, Y., & Zhou, C. J. (2020). Role of epigenetics and miRNAs in orofacial clefts. Birth defects research, 112(19), 1635-1659. doi:10.1002/bdr2.1802

Ghassibé, M., Bayet, B., Revencu, N., Verellen-Dumoulin, C., Gillerot, Y., Vanwijck, R., & Vikkula, M. (2005). Interferon regulatory factor-6: a gene predisposing to isolated cleft lip with or without cleft palate in the Belgian population. Eur J Hum Genet, 13(11), 1239-1242. doi:10.1038/sj.ejhg.5201486

Grant, S. F., Wang, K., Zhang, H., Glaberson, W., Annaiah, K., Kim, C. E., . . . Hakonarson, H. (2009). A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. Journal of Pediatrics, 155(6), 909-913. doi:10.1016/j.jpeds.2009.06.020

Grosen, D., Bille, C., Petersen, I., Skytthe, A., Hjelmborg, J. v. B., Pedersen, J. K., . . . Christensen, K. (2011). Risk of oral clefts in twins. Epidemiology (Cambridge, Mass.), 22(3), 313-319. doi:10.1097/EDE.0b013e3182125f9c

Grosen, D., Chevrier, C., Skytthe, A., Bille, C., Mølsted, K., Sivertsen, A., . . . 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(3), 162-168. doi:10.1136/jmg.2009.069385

Gundlach, K. K., & Maus, C. (2006). Epidemiological studies on the frequency of clefts in Europe and world-wide. Journal of Cranio-Maxillo-Facial Surgery, 34 Suppl 2, 1-2. doi:10.1016/s1010-5182(06)60001-2

Haaland Ø, A., Jugessur, A., Gjerdevik, M., Romanowska, J., Shi, M., Beaty, T. H., . . . Gjessing, H. K. (2017). Genome-wide analysis of parent-of-origin interaction effects with environmental exposure (PoOxE): An application to European and Asian cleft palate trios. PloS One, 12(9), e0184358. doi:10.1371/journal.pone.0184358

Haaland Ø, A., Lie, R. T., Romanowska, J., Gjerdevik, M., Gjessing, H. K., & Jugessur, A. (2018). A Genome-Wide Search for Gene-Environment Effects in Isolated Cleft Lip with or without Cleft Palate Triads Points to an Interaction between Maternal Periconceptional Vitamin Use and Variants in ESRRG. Front Genet, 9, 60. doi:10.3389/fgene.2018.00060

Haaland Ø, A., Romanowska, J., Gjerdevik, M., Lie, R. T., Gjessing, H. K., & Jugessur, A. (2019). A genome-wide scan of cleft lip triads identifies parent-of-origin interaction effects between ANK3 and maternal smoking, and between ARHGEF10 and alcohol consumption. F1000Res, 8, 960. doi:10.12688/f1000research.19571.2

Hammond, N. L., & Dixon, M. J. (2022). Revisiting the embryogenesis of lip and palate development. Oral Diseases, 28(5), 1306-1326. doi:https://doi.org/10.1111/odi.14174

Hao, Y., Tian, S., Jiao, X., Mi, N., Zhang, B., Song, T., . . . Zhuang, D. (2015). Association of Parental Environmental Exposures and Supplementation Intake with Risk of Nonsyndromic Orofacial Clefts: A Case-Control Study in Heilongjiang Province, China. Nutrients, 7(9), 7172-7184. doi:10.3390/nu7095328

Haque, S., & Alam, M. K. (2015). Common dental anomalies in cleft lip and palate patients. Malaysian Journal of Medical Sciences, 22(2), 55-60.

He, M., Zuo, X., Liu, H., Wang, W., Zhang, Y., Fu, Y., . . . Bian, Z. (2020). Genome-wide Analyses Identify a Novel Risk Locus for Nonsyndromic Cleft Palate. J Dent Res, 99(13), 1461-1468. doi:10.1177/0022034520943867

Hecht, J. T., Yang, P., Michels, V. V., & Buetow, K. H. (1991). Complex segregation analysis of nonsyndromic cleft lip and palate. Am J Hum Genet, 49(3), 674-681.

Hernández-Díaz, S., Werler, M. M., Walker, A. M., & Mitchell, A. A. (2000). Folic acid antagonists during pregnancy and the risk of birth defects. New England Journal of Medicine, 343(22), 1608-1614. doi:10.1056/nejm200011303432204

Hoebel, A. K., Drichel, D., van de Vorst, M., Böhmer, A. C., Sivalingam, S., Ishorst, N., . . . Ludwig, K. U. (2017). Candidate Genes for Nonsyndromic Cleft Palate Detected by Exome Sequencing. J Dent Res, 96(11), 1314-1321. doi:10.1177/0022034517722761

Hoornaert, K. P., Vereecke, I., Dewinter, C., Rosenberg, T., Beemer, F. A., Leroy, J. G., . . . Mortier, G. R. (2010). Stickler syndrome caused by COL2A1 mutations: genotype-phenotype correlation in a series of 100 patients. European journal of human genetics : EJHG, 18(8), 872-880. doi:10.1038/ejhg.2010.23

Huang, L., Jia, Z., Shi, Y., Du, Q., Shi, J., Wang, Z., . . . Yang, Z. (2019). Genetic factors define CPO and CLO subtypes of nonsyndromicorofacial cleft. PLoS Genetics, 15(10), e1008357. doi:10.1371/journal.pgen.1008357

Hunt, O., Burden, D., Hepper, P., Stevenson, M., & Johnston, C. (2006). Self-reports of psychosocial functioning among children and young adults with cleft lip and palate. Cleft Palate Craniofac J, 43(5), 598-605. doi:10.1597/05-080

Hwang, S. J., Beaty, T. H., Panny, S. R., Street, N. A., Joseph, J. M., Gordon, S., . . . Francomano, C. A. (1995). Association study of transforming growth factor alpha (TGF alpha) TaqI polymorphism and oral clefts: indication of gene-environment interaction in a population-based sample of infants with birth defects. American Journal of Epidemiology, 141(7), 629-636. doi:10.1093/oxfordjournals.aje.a117478

Indencleef, K., Hoskens, H., Lee, M. K., White, J. D., Liu, C., Eller, R. J., . . . Claes, P. (2021). The Intersection of the Genetic Architectures of Orofacial Clefts and Normal Facial Variation. Front Genet, 12, 626403. doi:10.3389/fgene.2021.626403

Indencleef, K., Roosenboom, J., Hoskens, H., White, J. D., Shriver, M. D., Richmond, S., . . . Claes, P. (2018). Six NSCL/P Loci Show Associations With Normal-Range Craniofacial Variation. Front Genet, 9, 502. doi:10.3389/fgene.2018.00502

Jahanbin, A., Shadkam, E., Miri, H. H., Shirazi, A. S., & Abtahi, M. (2018). Maternal Folic Acid Supplementation and the Risk of Oral Clefts in Offspring. Journal of Craniofacial Surgery, 29(6), e534-e541. doi:10.1097/scs.0000000000004488

Jezewski, P. A., Vieira, A. R., Nishimura, C., Ludwig, B., Johnson, M., O'Brien, S. E., . . . Murray, J. C. (2003). Complete sequencing shows a role for MSX1 in non-syndromic cleft lip and palate. J Med Genet, 40(6), 399-407. doi:10.1136/jmg.40.6.399

Jones, M. C. (1988). Etiology of facial clefts: prospective evaluation of 428 patients. Cleft Palate Journal, 25(1), 16-20.

Joubert, B. R., Felix, J. F., Yousefi, P., Bakulski, K. M., Just, A. C., Breton, C., . . . London, S. J. (2016). DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. American journal of human genetics, 98(4), 680-696. doi:10.1016/j.ajhg.2016.02.019

Jugessur, A., Farlie, P. G., & Kilpatrick, N. (2009). The genetics of isolated orofacial clefts: from genotypes to subphenotypes. Oral Dis, 15(7), 437-453. doi:10.1111/j.1601-0825.2009.01577.x

Kalay, E., Sezgin, O., Chellappa, V., Mutlu, M., Morsy, H., Kayserili, H., . . . Akarsu, N. A. (2012). Mutations in RIPK4 cause the autosomal-recessive form of popliteal pterygium syndrome. Am J Hum Genet, 90(1), 76-85. doi:10.1016/j.ajhg.2011.11.014

Kerr, S. M., Parker, S. E., Mitchell, A. A., Tinker, S. C., & Werler, M. M. (2020). Folic acid antagonist use before and during pregnancy and risk for selected birth defects. Birth defects research, 112(18), 1526-1540. doi:https://doi.org/10.1002/bdr2.1789

Khandelwal, K. D., van den Boogaard, M. H., Mehrem, S. L., Gebel, J., Fagerberg, C., van Beusekom, E., . . . van Bokhoven, H. (2019). Deletions and loss-of-function variants in TP63 associated with orofacial clefting. Eur J Hum Genet, 27(7), 1101-1112. doi:10.1038/s41431-019-0370-0

Kimani, J. W., Shi, M., Daack-Hirsch, S., Christensen, K., Moretti-Ferreira, D., Marazita, M. L., . . . Murray, J. C. (2007). X-chromosome inactivation patterns in monozygotic twins and sib pairs discordant for nonsyndromic cleft lip and/or palate. Am J Med Genet A, 143a(24), 3267-3272. doi:10.1002/ajmg.a.32098

Koillinen, H., Ollikainen, V., Rautio, J., Hukki, J., & Kere, J. (2003). Linkage and linkage disequilibrium searched for between non-syndromic cleft palate and four candidate loci. J Med Genet, 40(6), 464-468. doi:10.1136/jmg.40.6.464

Kondo, S., Schutte, B. C., Richardson, R. J., Bjork, B. C., Knight, A. S., Watanabe, Y., . . . Murray, J. C. (2002). Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nature Genetics, 32(2), 285-289. doi:10.1038/ng985

Krapels, I. P. C., van Rooij, I. A. L. M., Ocké, M. C., West, C. E., van der Horst, C. M. A. M., & Steegers-Theunissen, R. g. P. M. (2004). Maternal Nutritional Status and the Risk for Orofacial Cleft Offspring in Humans. The Journal of Nutrition, 134(11), 3106-3113. doi:10.1093/jn/134.11.3106

Kummet, C. M., Moreno, L. M., Wilcox, A. J., Romitti, P. A., DeRoo, L. A., Munger, R. G., . . . Wehby, G. L. (2016). Passive Smoke Exposure as a Risk Factor for Oral Clefts-A Large International Population-Based Study. American Journal of Epidemiology, 183(9), 834-841. doi:10.1093/aje/kwv279

Lace, B., Pajusalu, S., Livcane, D., Grinfelde, I., Akota, I., Mauliņa, I., . . . Inashkina, I. (2022). Monogenic Versus Multifactorial Inheritance in the Development of Isolated Cleft Palate: A Whole Genome Sequencing Study. Front Genet, 13, 828534. doi:10.3389/fgene.2022.828534

Lansdon, L. A., Dickinson, A., Arlis, S., Liu, H., Hlas, A., Hahn, A., . . . Manak, J. R. (2023). Genome-wide analysis of copy-number variation in humans with cleft lip and/or cleft palate identifies COBLL1, RIC1, and ARHGEF38 as clefting genes. Am J Hum Genet, 110(1), 71-91. doi:10.1016/j.ajhg.2022.11.012

Leslie, E. J. (2022). Genetic models and approaches to study orofacial clefts. Oral Dis, 28(5), 1327-1338. doi:10.1111/odi.14109

Leslie, E. J., Carlson, J. C., Shaffer, J. R., Butali, A., Buxo, C. J., Castilla, E. E., . . . Marazita, M. L. (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(3), 275-286. doi:10.1007/s00439-016-1754-7

Leslie, E. J., Carlson, J. C., Shaffer, J. R., Feingold, E., Wehby, G., Laurie, C. A., . . . Marazita, M. L. (2016). A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. Human Molecular Genetics, 25(13), 2862-2872. doi:10.1093/hmg/ddw104

Leslie, E. J., Koboldt, D. C., Kang, C. J., Ma, L., Hecht, J. T., Wehby, G. L., . . . Marazita, M. L. (2016). IRF6 mutation screening in non-syndromic orofacial clefting: analysis of 1521 families. Clinical Genetics, 90(1), 28-34. doi:10.1111/cge.12675

Leslie, E. J., Liu, H., Carlson, J. C., Shaffer, J. R., Feingold, E., Wehby, G., . . . Marazita, M. L. (2016). A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an

Etiologic Missense Variant in GRHL3. American journal of human genetics, 98(4), 744-754. doi:10.1016/j.ajhg.2016.02.014

Letra, A., Menezes, R., Granjeiro, J. M., & Vieira, A. R. (2007). Defining subphenotypes for oral clefts based on dental development. J Dent Res, 86(10), 986-991. doi:10.1177/154405910708601013

Letra, A., Menezes, R., Granjeiro, J. M., & Vieira, A. R. (2009). AXIN2 and CDH1 polymorphisms, tooth agenesis, and oral clefts. Birth Defects Research. Part A: Clinical and Molecular Teratology, 85(2), 169-173. doi:10.1002/bdra.20489

Li, B., Yong, L., Yu, Y., Yu, Y., Zhen, Q., Ge, H., . . . Sun, L. (2022). Genome-wide analyses of nonsyndromic cleft lip with or without palate identify 20 new risk loci in the Chinese Han population. J Genet Genomics, 49(9), 903-905. doi:10.1016/j.jgg.2022.02.004

Lidral, A. C., Liu, H., Bullard, S. A., Bonde, G., Machida, J., Visel, A., . . . Cornell, R. A. (2015). A single nucleotide polymorphism associated with isolated cleft lip and palate, thyroid cancer and hypothyroidism alters the activity of an oral epithelium and thyroid enhancer near FOXE1. Hum Mol Genet, 24(14), 3895-3907. doi:10.1093/hmg/ddv047

Lidral, A. C., Romitti, P. A., Basart, A. M., Doetschman, T., Leysens, N. J., Daack-Hirsch, S., . . . Murray, J. C. (1998). Association of MSX1 and TGFB3 with nonsyndromic clefting in humans. Am J Hum Genet, 63(2), 557-568. doi:10.1086/301956

Lin, J. Y., Luan, R. S., Guo, Z. Q., Lin, X. Q., Tang, H. Y., & Chen, Y. P. (2010). [The correlation analysis between environmental factors, bone morphogenetic protein-4 and transforming growth factor beta-3 polymorphisms in nonsyndromic cleft lip with or without cleft palate]. Zhonghua Kou Qiang Yi Xue Za Zhi, 45(10), 596-600.

Lin, K. J., Mitchell, A. A., Yau, W. P., Louik, C., & Hernández-Díaz, S. (2012). Maternal exposure to amoxicillin and the risk of oral clefts. Epidemiology, 23(5), 699-705. doi:10.1097/EDE.0b013e318258cb05

Little, J., Cardy, A., & Munger, R. G. (2004). Tobacco smoking and oral clefts: a meta-analysis. Bulletin of the World Health Organization, 82(3), 213-218.

Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G., . . . Kayser, M. (2012). A genome-wide association study identifies five loci influencing facial morphology in Europeans. PLoS Genetics, 8(9), e1002932. doi:10.1371/journal.pgen.1002932

Liu, H., Busch, T., Eliason, S., Anand, D., Bullard, S., Gowans, L. J. J., . . . Butali, A. (2017). Exome sequencing provides additional evidence for the involvement of ARHGAP29 in Mendelian orofacial clefting and extends the phenotypic spectrum to isolated cleft palate. Birth defects research, 109(1), 27-37. doi:10.1002/bdra.23596

Ludwig, K. U., Ahmed, S. T., Böhmer, A. C., Sangani, N. B., Varghese, S., Klamt, J., . . . Peters, H. (2016). Meta-analysis Reveals Genome-Wide Significance at 15q13 for Nonsyndromic Clefting of Both the Lip and the Palate, and Functional Analyses Implicate GREM1 As a Plausible Causative Gene. PLoS Genetics, 12(3), e1005914. doi:10.1371/journal.pgen.1005914

Ludwig, K. U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A., . . . Nöthen, M. M. (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nature Genetics, 44(9), 968-971. doi:10.1038/ng.2360

Lupo, P. J., Mitchell, L. E., & Jenkins, M. M. (2019). Genome-wide association studies of structural birth defects: A review and commentary. Birth defects research, 111(18), 1329-1342. doi:10.1002/bdr2.1606

Mangold, E., Böhmer, A. C., Ishorst, N., Hoebel, A. K., Gültepe, P., Schuenke, H., . . . Ludwig, K. U. (2016). Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate. Am J Hum Genet, 98(4), 755-762. doi:10.1016/j.ajhg.2016.02.013

Mangold, E., Ludwig, K. U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., . . . Nothen, M. M. (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nature Genetics, 42(1), 24-26. doi:10.1038/ng.506

Mangold, E., Reutter, H., Birnbaum, S., Walier, M., Mattheisen, M., Henschke, H., . . . Diaz Lacava, A. (2009). Genome-wide linkage scan of nonsyndromic orofacial clefting in 91 families of central European origin. Am J Med Genet A, 149a(12), 2680-2694. doi:10.1002/ajmg.a.33136

Marazita, M. L. (2002). Segregation analyses [Press release]

Marazita, M. L. (2007). Subclinical features in non-syndromic cleft lip with or without cleft palate (CL/P): review of the evidence that subepithelial orbicularis oris muscle defects are part of an expanded phenotype for CL/P. Orthodontics & Craniofacial Research, 10(2), 82-87. doi:10.1111/j.1601-6343.2007.00386.x

Marazita, M. L. (2012). The evolution of human genetic studies of cleft lip and cleft palate. Annual review of genomics and human genetics, 13, 263-283. doi:10.1146/annurev-genom-090711-163729

Marazita, M. L., Field, L. L., Cooper, M. E., Tobias, R., Maher, B. S., Peanchitlertkajorn, S., & Liu, Y. E. (2002). Genome scan for loci involved in cleft lip with or without cleft palate, in Chinese multiplex families. Am J Hum Genet, 71(2), 349-364. doi:10.1086/341944

Marazita, M. L., Field, L. L., Tunçbilek, G., Cooper, M. E., Goldstein, T., & Gürsu, K. G. (2004). Genome-scan for loci involved in cleft lip with or without cleft palate in consanguineous families from Turkey. Am J Med Genet A, 126a(2), 111-122. doi:10.1002/ajmg.a.20564

Marazita, M. L., Goldstein, A. M., Smalley, S. L., & Spence, M. A. (1986). Cleft lip with or without cleft palate: reanalysis of a three-generation family study from England. Genetic Epidemiology, 3(5), 335-342. doi:10.1002/gepi.1370030506

Marazita, M. L., Lidral, A. C., Murray, J. C., Field, L. L., Maher, B. S., Goldstein McHenry, T., . . . Arcos-Burgos, M. (2009). Genome scan, fine-mapping, and candidate gene analysis of non-syndromic cleft lip with or without cleft palate reveals phenotype-specific differences in linkage and association results. Human Heredity, 68(3), 151-170. doi:10.1159/000224636

Marazita, M. L., Murray, J. C., Lidral, A. C., Arcos-Burgos, M., Cooper, M. E., Goldstein, T., . . . Roddick, L. G. (2004). Meta-Analysis of 13 Genome Scans Reveals Multiple Cleft Lip/Palate Genes with Novel Loci on 9q21 and 2q32-35. The American Journal of Human Genetics, 75(2), 161-173. doi:https://doi.org/10.1086/422475

Marazita, M. L., Spence, M. A., & Melnick, M. (1984). Genetic analysis of cleft lip with or without cleft palate in Danish kindreds. American Journal of Medical Genetics, 19(1), 9-18. doi:10.1002/ajmg.1320190104

Martin, R. A., Hunter, V., Neufeld-Kaiser, W., Flodman, P., Spence, M. A., Furnas, D., & Martin, K. A. (2000). Ultrasonographic detection of orbicularis oris defects in first degree relatives of isolated cleft lip patients. American Journal of Medical Genetics, 90(2), 155-161. doi:10.1002/(sici)1096-8628(20000117)90:2<155::aid-ajmg13>3.0.co;2-v

Martin, R. A., Jones, K. L., & Benirschke, K. (1993). Extension of the cleft lip phenotype: the subepithelial cleft. American Journal of Medical Genetics, 47(5), 744-747. doi:10.1002/ajmg.1320470529

Marzouk, T., Alves, I. L., Wong, C. L., DeLucia, L., McKinney, C. M., Pendleton, C., . . . Shope, E. T. (2021). Association between Dental Anomalies and Orofacial Clefts: A Meta-analysis. JDR Clin Trans Res, 6(4), 368-381. doi:10.1177/2380084420964795

McGrath, J. A., Duijf, P. H., Doetsch, V., Irvine, A. D., de Waal, R., Vanmolkot, K. R., . . . van Bokhoven, H. (2001). Hay-Wells syndrome is caused by heterozygous missense mutations in the SAM domain of p63. Hum Mol Genet, 10(3), 221-229. doi:10.1093/hmg/10.3.221

Medicine, M.-N. I. o. G. Online Mendelian Inheritance in Man, OMIM. Retrieved from https://omim.org/

Melnick, M. (1997). Cleft lip and palate etiology and its meaning in early 20th century England: Galton/Pearson vs. Bateson; polygenically poor protoplasm vs. Mendelism. J Craniofac Genet Dev Biol, 17(2), 65-79.

Mills, J. L., Kirke, P. N., Molloy, A. M., Burke, H., Conley, M. R., Lee, Y. J., . . . Scott, J. M. (1999). Methylenetetrahydrofolate reductase thermolabile variant and oral clefts. American Journal of Medical Genetics, 86(1), 71-74.

Milunsky, J. M., Maher, T. A., Zhao, G., Roberts, A. E., Stalker, H. J., Zori, R. T., . . . Lin, A. E. (2008). TFAP2A mutations result in branchio-oculo-facial syndrome. American journal of human genetics, 82(5), 1171-1177. doi:10.1016/j.ajhg.2008.03.005

Mitchell, K., O'Sullivan, J., Missero, C., Blair, E., Richardson, R., Anderson, B., . . . Dixon, M. J. (2012). Exome sequence identifies RIPK4 as the Bartsocas-Papas syndrome locus. Am J Hum Genet, 90(1), 69-75. doi:10.1016/j.ajhg.2011.11.013

Mitchell, L. E., Murray, J. C., O'Brien, S., & Christensen, K. (2003). Retinoic acid receptor alpha gene variants, multivitamin use, and liver intake as risk factors for oral clefts: a population-based case-control study in Denmark, 1991-1994. American Journal of Epidemiology, 158(1), 69-76. doi:10.1093/aje/kwg102

Mitchell, L. E., & Risch, N. (1993). Correlates of genetic risk for non-syndromic cleft lip with or without cleft palate. Clinical Genetics, 43(5), 255-260. doi:10.1111/j.1399-0004.1993.tb03813.x

Moore, G. E., Ivens, A., Chambers, J., Farrall, M., Williamson, R., Page, D. C., . . . Jensson, O. (1987). Linkage of an X-chromosome cleft palate gene. Nature, 326(6108), 91-92. doi:10.1038/326091a0

Moreno, L. M., Mansilla, M. A., Bullard, S. A., Cooper, M. E., Busch, T. D., Machida, J., . . . Lidral, A. C. (2009). FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. Hum Mol Genet, 18(24), 4879-4896. doi:10.1093/hmg/ddp444

Moreno Uribe, L. M., Fomina, T., Munger, R. G., Romitti, P. A., Jenkins, M. M., Gjessing, H. K., . . . Wehby, G. L. (2017). A Population-Based Study of Effects of Genetic Loci on Orofacial Clefts. J Dent Res, 96(11), 1322-1329. doi:10.1177/0022034517716914

Mossey, P. (2007). Epidemiology underpinning research in the aetiology of orofacial clefts. Orthodontics & Craniofacial Research, 10(3), 114-120. doi:10.1111/j.1601-6343.2007.00398.x

Mossey, P. A., Little, J., Munger, R. G., Dixon, M. J., & Shaw, W. C. (2009). Cleft lip and palate. Lancet, 374(9703), 1773-1785. doi:10.1016/s0140-6736(09)60695-4

Mossey, P. A., & Modell, B. (2012). Epidemiology of oral clefts 2012: an international perspective. Frontiers of Oral Biology, 16, 1-18. doi:10.1159/000337464

Mukhopadhyay, N., Bishop, M., Mortillo, M., Chopra, P., Hetmanski, J. B., Taub, M. A., . . . Marazita, M. L. (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(2), 215-226. doi:10.1007/s00439-019-02099-1

Mukhopadhyay, N., Feingold, E., Moreno-Uribe, L., Wehby, G., Valencia-Ramirez, L. C., Muñeton, C. P. R., . . . Marazita, M. L. (2021). Genome-Wide Association Study of Non-syndromic Orofacial Clefts in a Multiethnic Sample of Families and Controls Identifies Novel Regions. Front Cell Dev Biol, 9, 621482. doi:10.3389/fcell.2021.621482

Mukhopadhyay, N., Feingold, E., Moreno-Uribe, L., Wehby, G., Valencia-Ramirez, L. C., Restrepo Muñeton, C. P., . . . Marazita, M. L. (2022). Genome-wide association study of multiethnic nonsyndromic orofacial cleft families identifies novel loci specific to family and phenotypic subtypes. Genetic Epidemiology. doi:10.1002/gepi.22447

Murray, J. (2002). Gene/environment causes of cleft lip and/or palate. Clinical Genetics, 61(4), 248-256. doi:https://doi.org/10.1034/j.1399-0004.2002.610402.x

Neiswanger, K., Weinberg, S. M., Rogers, C. R., Brandon, C. A., Cooper, M. E., Bardi, K. M., . . . Marazita, M. L. (2007). Orbicularis oris muscle defects as an expanded phenotypic feature in nonsyndromic cleft lip with or without cleft palate. Am J Med Genet A, 143a(11), 1143-1149. doi:10.1002/ajmg.a.31760

Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., . . . Shendure, J. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature Genetics, 42(9), 790-793. doi:10.1038/ng.646

Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., . . . Bamshad, M. J. (2010). Exome sequencing identifies the cause of a mendelian disorder. Nature Genetics, 42(1), 30-35. doi:10.1038/ng.499

Nicholls, W. (2016). Dental anomalies in children with cleft lip and palate in Western Australia. Eur J Dent, 10(2), 254-258. doi:10.4103/1305-7456.178317

O'Rahilly, R., & Müller, F. (2007). The development of the neural crest in the human. Journal of Anatomy, 211(3), 335-351. doi:10.1111/j.1469-7580.2007.00773.x

Oldereid, N. B., Wennerholm, U. B., Pinborg, A., Loft, A., Laivuori, H., Petzold, M., . . . Bergh, C. (2018). The effect of paternal factors on perinatal and paediatric outcomes: a systematic review and meta-analysis. Human Reproduction Update, 24(3), 320-389. doi:10.1093/humupd/dmy005

Park, J. W., McIntosh, I., Hetmanski, J. B., Jabs, E. W., Vander Kolk, C. A., Wu-Chou, Y. H., . . . Beaty, T. H. (2007). Association between IRF6 and nonsyndromic cleft lip with or without cleft palate in four populations. Genet Med, 9(4), 219-227. doi:10.1097/gim.0b013e3180423cca

Pengelly, R. J., Arias, L., Martínez, J., Upstill-Goddard, R., Seaby, E. G., Gibson, J., . . . Briceño, I. (2016). Deleterious coding variants in multi-case families with non-syndromic cleft lip and/or palate phenotypes. Scientific Reports, 6, 30457. doi:10.1038/srep30457

Peyrard-Janvid, M., Leslie, E. J., Kousa, Y. A., Smith, T. L., Dunnwald, M., Magnusson, M., . . . Schutte, B. C. (2014). Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. Am J Hum Genet, 94(1), 23-32. doi:10.1016/j.ajhg.2013.11.009

Phan, M., Conte, F., Khandelwal, K. D., Ockeloen, C. W., Bartzela, T., Kleefstra, T., . . . Carels, C. E. (2016). Tooth agenesis and orofacial clefting: genetic brothers in arms? Hum Genet, 135(12), 1299-1327. doi:10.1007/s00439-016-1733-z

Prescott, N. J., Lees, M. M., Winter, R. M., & Malcolm, S. (2000). Identification of susceptibility loci for nonsyndromic cleft lip with or without cleft palate in a two stage genome scan of affected sib-pairs. Hum Genet, 106(3), 345-350. doi:10.1007/s004390051048

Puhó, E. H., Szunyogh, M., Métneki, J., & Czeizel, A. E. (2007). Drug Treatment during Pregnancy and Isolated Orofacial Clefts in Hungary. The Cleft Palate Craniofacial Journal, 44(2), 194-202. doi:10.1597/05-208.1

Radhakrishna, U., Ratnamala, U., Gaines, M., Beiraghi, S., Hutchings, D., Golla, J., . . . Nath, S. K. (2006). Genomewide scan for nonsyndromic cleft lip and palate in multigenerational Indian families reveals significant evidence of linkage at 13q33.1-34. Am J Hum Genet, 79(3), 580-585. doi:10.1086/507487

Rahimov, F., Jugessur, A., & Murray, J. C. (2012). Genetics of nonsyndromic orofacial clefts. Cleft Palate Craniofac J, 49(1), 73-91. doi:10.1597/10-178

Rahimov, F., Marazita, M. L., Visel, A., Cooper, M. E., Hitchler, M. J., Rubini, M., . . . Murray, J. C. (2008). Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. Nature Genetics, 40(11), 1341-1347. doi:10.1038/ng.242

Ray, A. K., Field, L. L., & Marazita, M. L. (1993). Nonsyndromic cleft lip with or without cleft palate in west Bengal, India: evidence for an autosomal major locus. Am J Hum Genet, 52(5), 1006-1011.

Reiter, R., Brosch, S., Goebel, I., Ludwig, K. U., Pickhard, A., Högel, J., . . . Borck, G. (2015). A post GWAS association study of SNPs associated with cleft lip with or without cleft palate in submucous cleft palate. American Journal of Medical Genetics Part A, 167(3), 670-673. doi:https://doi.org/10.1002/ajmg.a.36891

Reiter, R., Brosch, S., Lüdeke, M., Fischbein, E., Haase, S., Pickhard, A., . . . Maier, C. (2012). Genetic and environmental risk factors for submucous cleft palate. European Journal of Oral Sciences, 120(2), 97-103. doi:10.1111/j.1600-0722.2012.00948.x

Rezaallah, B., Lewis, D. J., Zeilhofer, H. F., & Berg, B. I. (2019). Risk of Cleft Lip and/or Palate Associated With Antiepileptic Drugs: Postmarketing Safety Signal Detection and Evaluation of Information Presented to Prescribers and Patients. Ther Innov Regul Sci, 53(1), 110-119. doi:10.1177/2168479018761638

Riley, B. M., Schultz, R. E., Cooper, M. E., Goldstein-McHenry, T., Daack-Hirsch, S., Lee, K. T., . . . Murray, J. C. (2007). A genome-wide linkage scan for cleft lip and cleft palate identifies a novel locus on 8p11-23. Am J Med Genet A, 143a(8), 846-852. doi:10.1002/ajmg.a.31673

Rinkoff, S., & Adlard, R. E. (2022). Embryology, Craniofacial Growth, And Development. In StatPearls. Treasure Island (FL): StatPearls Publishing

Rischbieth, H. (1910). Hare-lip and cleft palate. In K. Pearson (Ed.), Treasury of Human Inheritance, Part IV. (pp. 79–123). London.

Romitti, P. A., Herring, A. M., Dennis, L. K., & Wong-Gibbons, D. L. (2007). Meta-analysis: pesticides and orofacial clefts. Cleft Palate Craniofac J, 44(4), 358-365. doi:10.1597/06-100.1

Roth, D. M., Bayona, F., Baddam, P., & Graf, D. (2021). Craniofacial Development: Neural Crest in Molecular Embryology. Head and Neck Pathology, 15(1), 1-15. doi:10.1007/s12105-021-01301-z

Rudnick, E. F., & Sie, K. C. (2008). Velopharyngeal insufficiency: current concepts in diagnosis and management. Curr Opin Otolaryngol Head Neck Surg, 16(6), 530-535. doi:10.1097/MOO.0b013e328316bd68

Rutledge, K. D., Barger, C., Grant, J. H., & Robin, N. H. (2010). IRF6 mutations in mixed isolated familial clefting. American Journal of Medical Genetics Part A, 152A(12), 3107-3109. doi:https://doi.org/10.1002/ajmg.a.33053

Safra, M. J., & Oakley, G. P., Jr. (1975). Association between cleft lip with or without cleft palate and prenatal exposure to diazepam. Lancet, 2(7933), 478-480. doi:10.1016/s0140-6736(75)90548-6

Salari, N., Darvishi, N., Heydari, M., Bokaee, S., Darvishi, F., & Mohammadi, M. (2022). Global prevalence of cleft palate, cleft lip and cleft palate and lip: A comprehensive systematic review and meta-analysis. J Stomatol Oral Maxillofac Surg, 123(2), 110-120. doi:10.1016/j.jormas.2021.05.008

Sales, S. A. G., Santos, M. L., Machado, R. A., Dias, V. O., Nascimento, J. E., Swerts, M. S. O., . . . Martelli, D. R. B. (2018). Incidence of bifid uvula and its relationship to submucous cleft palate and a family history of oral cleft in the Brazilian population. Brazilian Journal of Otorhinolaryngology, 84(6), 687-690. doi:10.1016/j.bjorl.2017.08.004

Schinzel, A., & Kläusler, M. (1986). The Van der Woude syndrome (dominantly inherited lip pits and clefts). J Med Genet, 23(4), 291-294. doi:10.1136/jmg.23.4.291

Schroeder, D. C., & Green, L. J. (1975). Frequency of dental trait anomalies in cleft, sibling, and noncleft groups. J Dent Res, 54(4), 802-807. doi:10.1177/00220345750540041801

Schutte, B. C., & Murray, J. C. (1999). The many faces and factors of orofacial clefts. Hum Mol Genet, 8(10), 1853-1859. doi:10.1093/hmg/8.10.1853

Schutte, B. C., Saal, H. M., Goudy, S., & Leslie, E. J. (1993). IRF6-Related Disorders. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. W. Gripp, G. M. Mirzaa, & A. Amemiya (Eds.), GeneReviews(®). Seattle (WA): University of Washington, Seattle

Seelan, R. S., Pisano, M., & Greene, R. M. (2019). Nucleic acid methylation and orofacial morphogenesis. Birth defects research, 111(20), 1593-1610. doi:10.1002/bdr2.1564

Seo, Y. J., Park, J. W., Kim, Y. H., & Baek, S. H. (2013). Associations between the risk of tooth agenesis and single-nucleotide polymorphisms of MSX1 and PAX9 genes in nonsyndromic cleft patients. Angle Orthodontist, 83(6), 1036-1042. doi:10.2319/020513-104.1

Shapira, Y., Lubit, E., & Kuftinec, M. M. (1999). Congenitally missing second premolars in cleft lip and cleft palate children. American Journal of Orthodontics and Dentofacial Orthopedics, 115(4), 396-400. doi:10.1016/s0889-5406(99)70258-1

Sharp, G. C., Ho, K., Davies, A., Stergiakouli, E., Humphries, K., McArdle, W., . . . Relton, C. L. (2017). Distinct DNA methylation profiles in subtypes of orofacial cleft. Clinical Epigenetics, 9, 63. doi:10.1186/s13148-017-0362-2

Shenoy, R. D., Shenoy, V., & Shetty, V. (2018). Chromosomal Abnormalities in Syndromic Orofacial Clefts: Report of Three Children. Case Rep Genet, 2018, 1928918. doi:10.1155/2018/1928918

Shprintzen, R. J., Schwartz, R. H., Daniller, A., & Hoch, L. (1985). Morphologic significance of bifid uvula. Pediatrics, 75(3), 553-561.

Simioni, M., Araujo, T. K., Monlleo, I. L., Maurer-Morelli, C. V., & Gil-da-Silva-Lopes, V. L. (2015). Investigation of genetic factors underlying typical orofacial clefts: mutational screening and copy number variation. Journal of Human Genetics, 60(1), 17-25. doi:10.1038/jhg.2014.96

Sivertsen, A., Wilcox, A. J., Skjaerven, R., Vindenes, H. A., Abyholm, F., Harville, E., & Lie, R. T. (2008). Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives. BMJ (Clinical research ed.), 336(7641), 432-434. doi:10.1136/bmj.39458.563611.AE

Slayton, R. L., Williams, L., Murray, J. C., Wheeler, J. J., Lidral, A. C., & Nishimura, C. J. (2003). Genetic association studies of cleft lip and/or palate with hypodontia outside the cleft region. Cleft Palate Craniofac J, 40(3), 274-279. doi:10.1597/1545-1569_2003_040_0274_gasocl_2.0.co_2

Smith, B. E., & Kuehn, D. P. (2007). Speech evaluation of velopharyngeal dysfunction. Journal of Craniofacial Surgery, 18(2), 251-261; quiz 266-257. doi:10.1097/SCS.0b013e31803ecf3b

Sperber, G. H. (2002). Formation of the Primary Palate. In Cleft Lip and Palate: From Origin to Treatment, 5–24.

Sproule, J. (1863). Hereditary Nature of Hare-Lip. British Medical Journal, 1(120), 412.

Sull, J. W., Liang, K. Y., Hetmanski, J. B., Wu, T., Fallin, M. D., Ingersoll, R. G., . . . Beaty, T. H. (2009). Evidence that TGFA influences risk to cleft lip with/without cleft palate through unconventional genetic mechanisms. Hum Genet, 126(3), 385-394. doi:10.1007/s00439-009-0680-3

Sun, Y., Huang, Y., Yin, A., Pan, Y., Wang, Y., Wang, C., . . . Yang, Y. (2015). Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun, 6, 6414. doi:10.1038/ncomms7414

Suter, M., Ma, J., Harris, A., Patterson, L., Brown, K. A., Shope, C., . . . Aagaard-Tillery, K. M. (2011). Maternal tobacco use modestly alters correlated epigenome-wide placental DNA methylation and gene expression. Epigenetics, 6(11), 1284-1294. doi:10.4161/epi.6.11.17819

Suzuki, S., Marazita, M. L., Cooper, M. E., Miwa, N., Hing, A., Jugessur, A., . . . Murray, J. C. (2009). Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. Am J Hum Genet, 84(3), 406-411. doi:10.1016/j.ajhg.2009.02.002

Tolarová, M. M., & Cervenka, J. (1998). Classification and birth prevalence of orofacial clefts. American Journal of Medical Genetics, 75(2), 126-137.

Trew, C. (1757). Sistens plura exempla palati deficientis. Nova Acta Phys Med Acad Caesarae Leopoldion-Carolinae, 1, 445-447.

Tsuda, M., Yamada, T., Mikoya, T., Sogabe, I., Nakashima, M., Minakami, H., . . . Yoshiura, K. (2010). A type of familial cleft of the soft palate maps to 2p24.2-p24.1 or 2p21-p12. Journal of Human Genetics, 55(2), 124-126. doi:10.1038/jhg.2009.131

van Bokhoven, H., Hamel, B. C., Bamshad, M., Sangiorgi, E., Gurrieri, F., Duijf, P. H., . . . Brunner, H. G. (2001). p63 Gene mutations in eec syndrome, limb-mammary syndrome,

and isolated split hand-split foot malformation suggest a genotype-phenotype correlation. Am J Hum Genet, 69(3), 481-492. doi:10.1086/323123

van den Boogaard, M. J., Dorland, M., Beemer, F. A., & van Amstel, H. K. (2000). MSX1 mutation is associated with orofacial clefting and tooth agenesis in humans. Nature Genetics, 24(4), 342-343. doi:10.1038/74155

van Rooij, I. A., Ludwig, K. U., Welzenbach, J., Ishorst, N., Thonissen, M., Galesloot, T. E., . . . Mangold, E. (2019). Non-Syndromic Cleft Lip with or without Cleft Palate: Genome-Wide Association Study in Europeans Identifies a Suggestive Risk Locus at 16p12.1 and Supports SH3PXD2A as a Clefting Susceptibility Gene. Genes (Basel), 10(12). doi:10.3390/genes10121023

Vantrappen, G., Rommel, N., Wellens, W., Cremers, C. W., Fryns, J. P., & Devriendt, K. (2002). Autosomal dominant isolated velopharyngeal insufficiency. Clinical Genetics, 61(1), 74-76. doi:10.1034/j.1399-0004.2002.610115.x

Vieira, A. R., McHenry, T. G., Daack-Hirsch, S., Murray, J. C., & Marazita, M. L. (2008a). Candidate gene/loci studies in cleft lip/palate and dental anomalies finds novel susceptibility genes for clefts. Genet Med, 10(9), 668-674. doi:10.1097/gim.0b013e3181833793

Vieira, A. R., McHenry, T. G., Daack-Hirsch, S., Murray, J. C., & Marazita, M. L. (2008b). A genome wide linkage scan for cleft lip and palate and dental anomalies. Am J Med Genet A, 146a(11), 1406-1413. doi:10.1002/ajmg.a.32295

Wang, Y., Li, X., Zhu, W. L., Guo, J. Z., Song, X. M., Li, S. Q., & Li, Y. (2010). Genome-wide and interaction linkage scan for nonsyndromic cleft lip with or without cleft palate in two multiplex families in Shenyang, China. Biomedical and Environmental Sciences, 23(5), 363-370. doi:10.1016/s0895-3988(10)60077-3

Wehby, G. L., & Cassell, C. H. (2010). The impact of orofacial clefts on quality of life and healthcare use and costs. Oral diseases, 16(1), 3-10. doi:10.1111/j.1601-0825.2009.01588.x

Wehby, G. L., Castilla, E. E., Goco, N., Rittler, M., Cosentino, V., Javois, L., . . . Murray, J. C. (2006). Description of the methodology used in an ongoing pediatric care interventional study of children born with cleft lip and palate in South America [NCT00097149]. BMC Pediatrics, 6(1), 9. doi:10.1186/1471-2431-6-9

Weinberg, S. M. (2022). What's Shape Got to Do With It? Examining the Relationship Between Facial Shape and Orofacial Clefting. Front Genet, 13, 891502. doi:10.3389/fgene.2022.891502

Weinberg, S. M., Brandon, C. A., McHenry, T. H., Neiswanger, K., Deleyiannis, F. W. B., de Salamanca, J. E., . . . Marazita, M. L. (2008). Rethinking isolated cleft palate: Evidence of occult lip defects in a subset of cases. American Journal of Medical Genetics Part A, 146A(13), 1670-1675. doi:https://doi.org/10.1002/ajmg.a.32291

Weinberg, S. M., Neiswanger, K., Martin, R. A., Mooney, M. P., Kane, A. A., Wenger, S. L., . . . Marazita, M. L. (2006). The Pittsburgh Oral-Facial Cleft study: expanding the cleft phenotype. Background and justification. Cleft Palate Craniofac J, 43(1), 7-20. doi:10.1597/04-122r1.1

White, J. D., Indencleef, K., Naqvi, S., Eller, R. J., Hoskens, H., Roosenboom, J., . . . Claes, P. (2021). Insights into the genetic architecture of the human face. Nature Genetics, 53(1), 45-53. doi:10.1038/s41588-020-00741-7

Wilcox, A. J., Lie, R. T., Solvoll, K., Taylor, J., McConnaughey, D. R., Abyholm, F., . . . Drevon, C. A. (2007). Folic acid supplements and risk of facial clefts: national population based case-control study. BMJ, 334(7591), 464. doi:10.1136/bmj.39079.618287.0B

Wu, T., Fallin, M. D., Shi, M., Ruczinski, I., Liang, K. Y., Hetmanski, J. B., . . . Beaty, T. H. (2012). Evidence of gene-environment interaction for the RUNX2 gene and environmental tobacco smoke in controlling the risk of cleft lip with/without cleft palate. Birth Defects Research. Part A: Clinical and Molecular Teratology, 94(2), 76-83. doi:10.1002/bdra.22885

Wu, T., Schwender, H., Ruczinski, I., Murray, J. C., Marazita, M. L., Munger, R. G., . . . Beaty, T. H. (2014). Evidence of gene-environment interaction for two genes on chromosome 4 and environmental tobacco smoke in controlling the risk of nonsyndromic cleft palate. PloS One, 9(2), e88088. doi:10.1371/journal.pone.0088088

Wyszynski, D. F., Albacha-Hejazi, H., Aldirani, M., Hammod, M., Shkair, H., Karam, A., . . . Bailey-Wilson, J. E. (2003). A genome-wide scan for loci predisposing to non-syndromic cleft lip with or without cleft palate in two large Syrian families. Am J Med Genet A, 123a(2), 140-147. doi:10.1002/ajmg.a.20283

Xuan, Z., Zhongpeng, Y., Yanjun, G., Jiaqi, D., Yuchi, Z., Bing, S., & Chenghao, L. (2016). Maternal active smoking and risk of oral clefts: a meta-analysis. Oral Surg Oral Med Oral Pathol Oral Radiol, 122(6), 680-690. doi:10.1016/j.oooo.2016.08.007

Yang, W., Carmichael, S. L., Roberts, E. M., Kegley, S. E., Padula, A. M., English, P. B., & Shaw, G. M. (2014). Residential agricultural pesticide exposures and risk of neural tube defects and orofacial clefts among offspring in the San Joaquin Valley of California. American Journal of Epidemiology, 179(6), 740-748. doi:10.1093/aje/kwt324

Young, A., & Spinner, A. (2022). Velopharyngeal Insufficiency. In StatPearls. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK563149/

Young, J. I., Slifer, S., Hecht, J. T., & Blanton, S. H. (2021). DNA Methylation Variation Is Identified in Monozygotic Twins Discordant for Non-syndromic Cleft Lip and Palate. Front Cell Dev Biol, 9, 656865. doi:10.3389/fcell.2021.656865

Yow, M., Jin, A., & Yeo, G. S. H. (2021). Epidemiologic trends of infants with orofacial clefts in a multiethnic country: a retrospective population-based study. Scientific Reports, 11(1), 7556. doi:10.1038/s41598-021-87229-4

Yu, Y., Zuo, X., He, M., Gao, J., Fu, Y., Qin, C., . . . Bian, Z. (2017). Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. Nature Communications, 8, 14364-14364. doi:10.1038/ncomms14364

Yuzuriha, S., & Mulliken, J. B. (2008). Minor-form, microform, and mini-microform cleft lip: anatomical features, operative techniques, and revisions. Plastic and Reconstructive Surgery, 122(5), 1485-1493. doi:10.1097/PRS.0b013e31818820bc

Zeiger, J. S., Hetmanski, J. B., Beaty, T. H., VanderKolk, C. A., Wyszynski, D. F., Bailey-Wilson, J. E., . . . McIntosh, I. (2003). Evidence for linkage of nonsyndromic cleft lip with or without cleft palate to a region on chromosome 2. Eur J Hum Genet, 11(11), 835-839. doi:10.1038/sj.ejhg.5201052

Zellweger, H., Bardach, J., Bordwell, J., & Williams, K. (1975). The short arm deletion syndrome of chromosome 4 (4p- syndrome). Archives of Otolaryngology, 101(1), 29-32. doi:10.1001/archotol.1975.00780300033008

Zhang, W. H., Chen, Y. Y., Liu, J. J., Liao, X. H., Du, Y. C., & Gao, Y. (2015). Application of ultrasound imaging of upper lip orbicularis oris muscle. International Journal of Clinical and Experimental Medicine, 8(3), 3391-3400.

Zheng, Z., Xie, G., Yang, T., & Qin, J. (2019). Congenital malformations are associated with secondhand smoke among nonsmoking women: A meta-analysis. Birth, 46(2), 222-233. doi:10.1111/birt.12401

Zieger, H. K., Weinhold, L., Schmidt, A., Holtgrewe, M., Juranek, S. A., Siewert, A., . . . Ludwig, K. U. (2023). Prioritization of non-coding elements involved in non-syndromic cleft lip with/without cleft palate through genome-wide analysis of de novo mutations. HGG Adv, 4(1), 100166. doi:10.1016/j.xhgg.2022.100166

Zucchero, T. M., Cooper, M. E., Maher, B. S., Daack-Hirsch, S., Nepomuceno, B., Ribeiro, L., . . . Murray, J. C. (2004). Interferon Regulatory Factor 6 (IRF6) Gene Variants and the Risk of Isolated Cleft Lip or Palate. New England Journal of Medicine, 351(8), 769-780. doi:10.1056/NEJMoa032909

CHAPTER II. Rare variants found in multiplex families with orofacial clefts: Does expanding the phenotype make a difference?

Kimberly K. Diaz Perez, Sydney Chung, S. Taylor Head, Michael P. Epstein, Jacqueline T. Hecht, George L. Wehby, Jeffrey C. Murray, Mary L. Marazita, Elizabeth J. Leslie

Adapted from an original publication submitted to American Journal of Medical Genetics:
Diaz Perez, K., Chung, S., Head, S. T., Epstein, M., Hecht, J., Wehby, G., Murray, J. C., Marazita, M. L., Leslie, E. J. (2023). Rare variants found in multiplex families with orofacial clefts: Does expanding the phenotype make a difference? medRxiv, 2023.2002.2001.23285340. doi:10.1101/2023.02.01.23285340

Introduction

Orofacial clefts (OFCs) represent a human disorder where rare and common variant studies have been successful (Leslie, 2022). OFCs are common birth defects (affecting 1/1000 live births worldwide) that occur on an etiological spectrum that includes Mendelian genetic causes as well as environmental causes, such as exposure to teratogens during pregnancy (Garland *et al*., 2020). However, most OFCs are thought to occur as complex disorders resulting from the interaction of multiple genetic risk factors and environmental influences (Beaty *et al*., 2016). Mendelian forms of OFCs are often syndromes that can include non-cleft phenotypes in some affected as opposed to isolated, non-syndromic cases with accompanying additional clinical features (Dixon *et al*., 2011). It is now clear from multiple studies that non-syndromic and syndromic forms of OFCs have overlapping etiological spectrums (Basha *et al*., 2018; Leslie, 2022). One hypothesis arising from sequencing studies suggests that pathogenic variants causing syndromic OFCs tend to be deleterious exonic variants in genes involved in craniofacial development (Kondo *et al*., 2002; Peyrard-Janvid *et al*., 2014) whereas variants associated with non-syndromic OFCs may have less severe effects on protein function or occur in regulatory variants of the same genes (Leslie *et al*., 2016; Rahimov *et al*., 2008; Zucchero *et al*., 2004). However, the genetic mechanisms for risk in non-syndromic OFCs are varied and include complex/oligogenic/multigenic mechanisms (Alade *et al*., 2022; Stanier & Moore, 2004), Mendelian variants (Cox *et al*., 2018; Liu *et al*., 2017), and *de novo* mutations (Awotoye *et al*., 2022; Bishop *et al*., 2020); but much of the risk for OFCs is still unknown.

Genetic studies of non-syndromic OFCs have recently favored genome-wide association studies (GWAS) and over 15 GWAS or meta-analyses have cumulatively identified over 50 associated genes or loci (Alade *et al*., 2022; Birnbaum *et al*., 2009; Leslie, 2022; Leslie *et al*.,

2016; Mangold *et al*., 2010; Yu *et al*., 2017). These loci are estimated to account for only ~20-25% of the known heritable risk of OFCs, leaving a substantial portion of risk variants unaccounted for (Alade *et al*., 2022; Leslie, 2022). Decreases in the cost of sequencing that allow for far larger sample sizes to be studied have facilitated a shift toward the analysis of rare genetic variation as a possible source of OFC risk, as they are hypothesized to have larger effect sizes compared to common variants (Kryukov *et al*., 2007).

One approach to identify rare variants is to focus on family-based study designs as rare variants with large effects might segregate with OFCs in multiplex families. In support of this, Bureau *et al*. (2014) and Cox *et al*. (2018) identified rare, "likely pathogenic" variants shared by affected relatives that segregated in a dominant manner within ostensibly non-syndromic OFC families. Basha *et al*. (2018) estimated that rare "likely pathogenic" variants in genes associated with OFC syndromes could be identified in ~10% of multiplex non-syndromic OFC families.

Approximately 15% of families with non-syndromic OFCs are multiplex, but the pattern of affected relatives does not always follow classic Mendelian patterns. Imposing a Mendelian structure on these families would require high levels of incomplete penetrance as there can be multiple unaffected individuals linking the affected individuals (Kingdom & Wright, 2022). We have previously hypothesized that this "incomplete penetrance" could be explained by the phenotypic misclassification of individuals who lack overt OFCs but have subclinical phenotypes (Marazita, 2012; Weinberg *et al*., 2006). Under this hypothesis, individuals manifesting these subclinical cleft features could represent "genetic carriers" who, because the phenotype is so subtle, are mischaracterized as unaffected. This expanded phenotypic spectrum of OFCs includes subclinical phenotypes such as discontinuities in the orbicularis oris muscle (OOM), velopharyngeal insufficiency (VPI), or mild phenotypes, such as bifid uvula (Weinberg

et al., 2006). OOM discontinuities are subepithelial defects of the muscle surrounding the upper lip and are only detected through ultrasonography. Similarly, VPI is not readily observable and occurs when the muscular valve between the oral and nasal cavity fails to close, resulting in hypernasal speech and phonation challenges (Weinberg *et al*., 2006). These subclinical phenotypes are hypothesized to be mild forms of OFCs in part because they have been observed at higher frequencies in apparently unaffected individuals from OFC families compared to controls (Neiswanger *et al*., 2007; Weinberg *et al*., 2006). Here, we hypothesize that including these subclinical phenotypes could clarify the inheritance patterns in multiplex OFC families and help the identification of genetic risk factors segregating in these families.

Therefore, we aimed to investigate rare coding variants in multiplex OFC families with whole-exome sequencing by testing two complementary hypotheses. First, we hypothesized that multiplex families with inheritance patterns consistent with a Mendelian mechanism would segregate private, rare variants among affected individuals. Second, we hypothesized that subclinical OFC phenotypes would increase support for specific inheritance patterns and that likely causal variants would segregate among individuals with either overt phenotypes or subclinical phenotypes.

Methods

Cohort Information

This study cohort consists of 31 families from national and international recruitment sites in the United States (Colorado, Iowa, Pennsylvania, Texas) (N=13), Europe (Hungary) (N=2), Asia (China, India, Philippines) (N=13), and Central America (Guatemala) (N=3) originally recruited for the Pittsburgh Orofacial Cleft Study at the University of Pittsburgh. All participants

provided informed consent; the study was approved by the IRB at the University of Pittsburgh and local recruiting sites. We selected apparently non-syndromic OFC families for sequencing if they met the criteria for one of the following groups: (I) OFC multiplex families: characterized by the presence of at least one set of 2nd-degree or closer relative pairs where each member had an OFC (CL, CLP, or CP) and lack sequenced individuals with subclinical phenotypes (N=12); (II) multiplex families with subclinical phenotypes: contains multiple sequenced affected individuals as well as relatives with at least one subclinical phenotype (N=19). Most families had demographic and medical histories as well as photographs of the study participants. A total of 150 individuals (75 males, 75 females) with sufficient DNA quantities were sequenced (Table S2-1).

Sequencing

Whole-exome sequencing was performed using the Agilent SureSelectXT HumanAllExon V6 + UTR S07604624 exome capture kit at the Center for Inherited Disease Research. A low-input library prep protocol developed at CIDR was performed (Marosy *et al*., 2017). Libraries were prepared from 50ng of genomic DNA, sheared for 80s using the Covaris E220 instrument (Covaris). The KAPA Hyper prep kit was used to process the sheared DNA into amplified dual indexed adapter-ligated fragments. 750ng of the amplified library was used in an enrichment reaction following Agilent protocols. Libraries were sequenced on the NovaSeq 6000 platform with onboard clustering using 125 base pairs paired-end runs and sequencing chemistry kit NovaSeq 6000 S4 Reagent Kit v1.

Variant Calling and Quality Control

Fastq files were aligned with BWA-MEM version 0.7.15 to the 1000 genomes phase 2 (GRCh37) human genome reference (Li, 2013). Duplicate molecules were flagged with Picard version 2.17.0. Base call quality score recalibration and binning (2,10,20,30) were performed using the Genome Analysis Toolkit (GATK) version v4.0.1.1 (McKenna *et al*., 2010). Cram files were generated using SAMTools version 1.5. GATK's reference confidence model workflow was used to perform joint sample genotyping using GATK version 3.7. Briefly, this workflow entails: 1) Producing a gVCF (genomic Variant Call Format (VCF)) for each sample individually using Haplotype Caller (--emitRefConfidence GVCF) and –max_alternate_alleles was set to 3 to all bait intervals to generate likelihoods that the sites are homozygote reference or not, and 2) Joint genotyping the single sample gVCFs together with GenotypeGVCFs to produce a multi-sample VCF file. Variant filtering was done using the Variant Quality Score Recalibration (VQSR) method (DePristo *et al*., 2011). For single-nucleotide variants (SNVs), the annotations of MQRankSum, QD, FS, ReadPosRankSum, MQ, and SOR were used in the adaptive error model. HapMap3.3, Omni2.5, and 1000G phase high confidence SNP calls were used as training sets with HapMap3.3 and Omni2.5 used as the truth set. SNVs were filtered to obtain all variants up to the 99.5th percentile of truth sites (0.5% false negative rate). For indels, the annotations of FS, ReadPosRankSum, MQRankSum, QD, and SOR were used in the adaptive error model (4 maximum Gaussians allowed). A set of curated indels obtained from the GATK resource bundle (Mills_and_1000G_gold_standard.indels.b37.vcf) were used as training and truth sites. Indels were filtered to obtain all variants up to the 99th percentile of truth sites (1% false negative rate). Prior to the analysis, additional filters on genotype calls were applied based on a read depth ≥ 15 and genotype quality ≥ 20 via VCFtools (version 0.1.13).

Variant Filtering

All variants within each family were annotated using Bystro Genomics (Kotlar *et al*., 2018), an in-house variant annotation and filtering tool, and VarSeq v2.2.5 (Golden Helix, Inc., Bozeman, MT). We retained and analyzed variants that met the following criteria: 1) exonic, 2) missense, nonsense, frameshift, and canonical splice variants, and 3) a global minor allele frequency (MAF) ≤ 0.5% in the Genome Aggregation Database (gnomAD) exomes and genomes v.2 (Karczewski *et al*., 2020). We also considered predictors of missense pathogenicity using various *in silico* tools, such as CADD scores (Rentzsch *et al*., 2018) and gene tolerance to variation metrics from gnomAD (Karczewski *et al*., 2020). Gene tolerance measures included Z-scores for missense variants, the probability of being loss-of-function intolerant (pLI) (Lek *et al*., 2016), and loss-of-function observed/expected upper bound fraction (LOEUF) for loss-of-function variants from gnomAD (Karczewski *et al*., 2020).

Single-Family Segregation Analyses

For single-family analyses, we defined individuals with either an overt cleft or a subclinical phenotype as "affected". In families with an apparent dominant inheritance pattern, we analyzed heterozygous variants shared among all affected individuals. In families with an apparent recessive mode of inheritance, we analyzed compound heterozygous variants or homozygous variants in the affected individuals and both parents had to be carriers for the variant. For multiplex families with only overt clefts, we allowed variants to be present in their unaffected relatives to allow for an incompletely penetrant model. After filtering for variants using the criteria noted above, we performed literature searches using databases, such as ClinVar (Landrum *et al*., 2018) and Online Mendelian Inheritance in Man (OMIM) (Hamosh *et al*.,

2005), to support the plausibility of the variant and the gene to cause OFCs or a craniofacial phenotype.

Mixed Model Linear Regression

We conducted linear mixed-effect models to compare the number of variants in individuals with OFCs and subclinical phenotypes. We utilized the "lme4" (version 1.1-29) package (Bates *et al*., 2015) along with the "afex" package (version 1.2-0) (Henrik *et al*., 2022) in R (version 3.6.3) (Team, 2021). We computed the number of heterozygous rare, protein-altering variants (MAF ≤ 0.5%) for each sample in protein-coding genes and OFC genes. We considered affection status (presence of an OFC) and the presence of subclinical phenotypes as indicator variables. We added a family-specific random intercept to account for relatedness within families. For models considering variants in OFC genes, we utilized a gene list comprised of 418 genes previously associated with craniofacial development and OFCs. These were compiled from four sources, including OMIM (Amberger *et al*., 2015), the PreventionGenetics Cleft Lip/Cleft Palate Panel (PreventionGenetics, Marshfield, WI), the Genomics England PanelApp Clefting Panel (v2.2, March 2020; Martin *et al*., 2019), and an additional set of literature-curated genes (more information on this gene list can be found in Diaz Perez *et al*. (2022). We considered a significance level for both the "all genes" and "OFC genes" at $P < 0.05$.

Results

We filtered for rare variants in protein-coding regions (MAF ≤ 0.5%) in 31 total families and identified an average of 195 variants in affected individuals (range 12 to 655 variants) per family.

Variants in Families with Overt Clefts

In the 12 families that only had individuals with overt clefts, we identified likely causal variants in three families (3/12, 25%) (Figure 2-1).

Family 1: Family 1 was a Guatemalan family comprised of four siblings with CLP, four unaffected siblings, and unaffected parents. All four siblings shared a novel 1 bp deletion in *TP63* leading to a frameshift (NM_003722.5: c.1606delC; NP_003713.3: p.His536Thrfs*18). The variant was inherited from the unaffected mother and was also found in their unaffected sister. *TP63* is highly intolerant to loss-of-function variation (pLI = 1, LOEUF = 0.27) and the variant was not present in gnomAD. Heterozygous missense mutations in *TP63* cause allelic syndromes impacting the face and/or limbs including ectrodactyly, ectodermal dysplasia and CL/P (EEC) syndrome, and ankyloblepharon-ectodermal defects-CL/P (AEC) syndrome. Deletions and frameshift variants in *TP63* have recently been identified in non-syndromic OFC families (Khandelwal *et al*., 2019). Interestingly, like Family 1, most of the published variants were inherited from unaffected parents, suggesting an incompletely penetrant effect for truncating mutations in *TP63*. To support the initial diagnosis of non-syndromic OFC, we examined photographs of the family which did not reveal any evidence of ectodermal dysplasia or limb defects associated with *TP63*-associated syndromes. Because all affected individuals were male and there was no male-to-male transmission that would rule out an X-linked inheritance model, we also examined genes on the X chromosome. There was one variant in *SEPTIN6* (NP_665798.1: p.Ser408Cys) that was heterozygous in the mother and was transmitted to the affected male offspring but not the unaffected male or unaffected sister. However, it is found as a hemizygous variant in 23 males in the Latino/Admixed American population in

gnomAD (0.3% allele frequency) and this makes it a less compelling candidate than the novel, truncation variant in *TP63*.

Family 2: We identified a heterozygous missense mutation in *SHROOM3* (NM_020859.4: c.1088A>G; NP_065910.3: p.Gln363Arg) in a family from the United States that was shared between a set of monozygotic twins, one with CL and the other with CLP, and their brother with CL, and was transmitted from their unaffected mother. *SHROOM3* is associated with the cytoskeleton, and it is important for neural tube morphogenesis (Das *et al*., 2014; Hildebrand & Soriano, 1999). *SHROOM3* has been previously associated with OFCs through genome-wide association studies and rare, *de novo* mutations in OFC trios (Bishop *et al*., 2020; Copp & Greene, 2013; Leslie *et al*., 2017; Ray *et al*., 2021). Moreover, mouse mutants of Shroom3 have been shown to exhibit highly penetrant craniofacial malformations, including exencephaly and facial clefting (Hildebrand & Soriano, 1999).

Family 3: We found a missense substitution in *KLF4* (NM_004235.6: c.203C>G; NP_004226.3: p.Ala68Gly) in a three-generation Chinese family that segregated among all three affected individuals with CL or CLP and was absent from the sequenced unaffected individual. *KLF4* is a transcription factor involved in the differentiation of the epidermis (Segre *et al*., 1999). The expression of *KLF4* is directly regulated by IRF6, a key OFC-associated gene, in the oral epithelium during periderm differentiation in zebrafish (Liu *et al*., 2016). Functional zebrafish studies of rare missense mutations in *KLF4* in non-syndromic OFC cases found alterations in the differentiation of the periderm, indicating that rare variants in *KLF4* may increase the risk for OFCs (Liu *et al*., 2020).

Variants in Families with Overt Clefts and Subclinical Phenotypes

We evaluated 19 multiplex OFC families with at least one sequenced individual with a subclinical phenotype. We found likely causal variants in four families (4/19, 21%) (Figure 2-1).

Family 4: We identified a novel missense mutation in *IRF6* (NM_006147.4: c.65T>C; NP_006138.1: p.Leu22Pro) in a three-generation pedigree from Hungary. This substitution is located in the DNA-binding domain of the IRF6 protein and has been previously reported in Van der Woude syndrome (VWS) (Ghassibé *et al*., 2004). The variant was transmitted from the paternal grandfather, who had a bifid uvula, missing teeth, and syndactyly of the hands and feet. The proband's father, who had CLP and missing teeth, also had the missense variant. Lip pits, one of the diagnostic criteria for VWS, were not reported. Ink lip prints (Neiswanger *et al*., 2009), but not photographs, were collected; however, it is not possible to conclusively confirm the presence or absence of lip pits from these prints.

Family 5: In Family 5 from the Philippines, we found a rare in-frame deletion (NM_005996.4: c.1991_2005delTGGCAGTGGACTCGG; NP_005987.3: p.Val664_Ser668del) in *TBX3*. The deletion was transmitted from the unaffected mother and was present in two affected individuals and a sibling with the OOM phenotype, but not their unaffected siblings. Heterozygous truncation mutations in *TBX3* mutations cause Ulnar-mammary syndrome, characterized by mammary gland hypoplasia and upper limb defects. The proband is a short (5' 2"), but not obese (<100 lbs), female with a missing lateral right incisor and unilateral CL. At the time of enrollment, no developmental delays or other structural anomalies were reported. The mother reported a history of miscarriage but did not report any major medical conditions or structural anomalies; a limited craniofacial physical exam by research staff reported buccal frenula and a high-arched palate. Although OFC rarely occurs in Ulnar-mammary syndrome,

inactivation of *TBX3* in the neural crest in mice leads to postnatal death and a highly penetrant cleft palate (López *et al*., 2018).

Family 6: We identified a novel 32-base pair deletion in *SMC3* (NM_005445.4: c.2019_2050del; NP_005436.1: p.Leu676Argfs*5) in this family from the United States. The variant was shared between the three affected individuals with CP but was not present in the sibling with VPI. The deletion was paternally inherited and the father's unsequenced aunt had CLP. Mutations in *SMC3* cause Cornelia de Lange (CdL) syndrome; however, this family did not have any additional structural anomalies, intellectual disability, or craniofacial features (e.g., microcephaly, arched eyebrows) that are characteristic of CdL (Gil-Rodriguez *et al*., 2015; Kline *et al*., 2007). This family illustrates that the inclusion of subclinical phenotypes could lead to false negatives should the causal variant for OFCs not be also causal for the subclinical phenotype.

Family 7: We identified a nonsense mutation in *COL11A2* (NM_080680.3: c.3181C>T; NP_542411.2: p.Arg1061*) in a family from the United States that was transmitted to the proband with CLP from his unaffected father but was not present in his sibling with an OOM defect. *COL11A2* is associated with autosomal dominant and recessive forms of Fibrochondrogenesis and Otospondylomegaepiphyseal Dysplasia (also known as non-ocular Stickler syndrome), the latter of which sometimes includes cleft palate (van Steensel *et al*., 1997; Vikkula *et al*., 1995). *COL11A2* has also been associated with non-syndromic OFCs through common variants (Nikopensius *et al*., 2010).

Quantitative Variant Analysis

In most families, we were not able to identify a single causal variant, but we did observe many compelling missense variants in genes associated with craniofacial development (Table S2-2). We hypothesized that individuals with overt clefts might have a higher number of such variants compared to their relatives with subclinical phenotypes. Using a curated list of 418 genes, we first calculated the number of rare (MAF ≤ 0.5%), protein-altering variants in individuals with overt clefts or subclinical phenotypes. We found fewer variants in individuals with OFCs (an average of 14.3 variants per person) than in individuals with subclinical phenotypes (an average of 15 variants per person) (Figure 2-2A). After adjusting for affection status and relatedness, there was no difference in the number of variants in all protein-coding genes (p=0.46) or OFC genes (p=0.64). The same was true when restricting to rare variants with a CADD score ≥ 20 (Figure 2-2B; p=0.27 for protein-coding genes and p=0.44 for OFC genes).

Discussion

In this study, we aimed to investigate the contribution of rare variants in the genetic etiology of OFCs by sequencing 31 multiplex families with overt OFCs with or without subclinical phenotypes. Our "hit" rate was ~21-25% for both families with individuals with subclinical phenotypes and families with overt OFCs only, which is higher than the 10% reported by Basha *et al*. (2018), but is not statistically different (p=0.21, Fisher's exact test). Our higher rate may be explained by the smaller sample size but there were also differences in the selection of families and the analysis pipeline. One of our families had an *IRF6* mutation, but this family (and others like them) would have been excluded from the Basha *et al*. study, which were drawn from a database prescreened for *IRF6* mutations. Basha *et al*. (2018) also focused their analysis on a subset of 500 genes plausibly involved in OFCs.

Rare variants in *BMP4* were previously reported to be associated with overt clefts and OOM defects; however, *BMP4* variants were not found among the candidate variants in this study (Suzuki *et al*., 2009). In fact, we did not detect strong evidence to suggest that the inclusion of subclinical phenotypes facilitates gene discovery. Given the small sample sizes in this study, our evidence supporting a common etiology for subclinical phenotypes and overt OFCs is only anecdotal. Additional genetic studies need to be conducted in larger and more phenotypically homogeneous samples to determine the utility of subclinical phenotypes for gene discovery.

Four variants were transmitted from unaffected parents. One explanation for incomplete penetrance of a variant is mosaicism in the transmitting parent (Kingdom & Wright, 2022). We have limited ability to detect mosaicism with a single tissue source and standard exome sequencing, but nonetheless did not find evidence of mosaicism in the parental samples based on the allele balance (43.2-52.4% alternate alleles). It is also possible the effect of the variant is modified by as-yet unknown environmental exposures or additional genetic risk factors, which could influence the expression of OFCs (Beames & Lipinski, 2020; Carlson *et al*., 2017). Similar explanations (e.g., mosaicism, modifiers, or stochastic events) may explain the variable expressivity of overt and more mild forms of OFCs observed in these families. More work is needed to test the hypothesis that OFCs and subclinical phenotypes share an etiology and to determine the impact of rare genetic variation in the etiology of OFCs.

Overall, our results provide further evidence of the Mendelian transmission of rare coding variants in non-syndromic multiplex OFC families. Similar to the findings of Basha *et al*. (2018), Bishop *et al*. (2020), and others, this work provides evidence that individuals and families with apparently non-syndromic OFCs may have rare coding variants in genes associated

with syndromic OFCs. These results can provide support for the recommendation to offer diagnostic genetic testing to families with apparently non-syndromic OFCs and a positive family history. We note, however, that the number of affected family members and the family structure should be carefully considered. Many of our families were relatively small and not all affected or informative individuals had DNA available or were successfully sequenced, limiting our ability to narrow the list of candidate variants. In this study, we found most likely causal variants in families with at least three affected individuals. Specific recommendations for diagnostic testing will continue to evolve as more data on the contribution of rare variants to both isolated and familial clefting accrues. Recent data supporting a role for rare copy number variants (Lansdon *et al*., 2023) and how to incorporate other genomic variants, including those in non-coding regions (Zieger *et al*., 2023), will require additional data and validation through analytic trials. But as some individuals with a positive family history will have questions about risks consideration should be given to sequencing studies to identify variants that might suggest higher than what epidemiologic recurrence risks alone would support.

Figures

Figure 2-1: Likely Causal Variants in Multiplex OFC Families. We found seven likely causal variants in *TP63*, *SHROOM3, KLF4*, *IRF6*, *TBX3*, *SMC3*, and *COL11A2.* Sex symbols with solid black indicate the phenotype of the individual: CL (cleft lip), CP (cleft palate), and CLP (cleft lip and palate). The symbol with a green circle represents the individuals with discontinuities in the orbicularis oris muscle (OOM), the blue circle represents individuals with velopharyngeal insufficiency (VPI), and the black solid square inside the symbol indicates the sample had a bifid uvula. The purple solid lines indicate individuals with whole-exome data while purple solid stars indicate variant carriers.

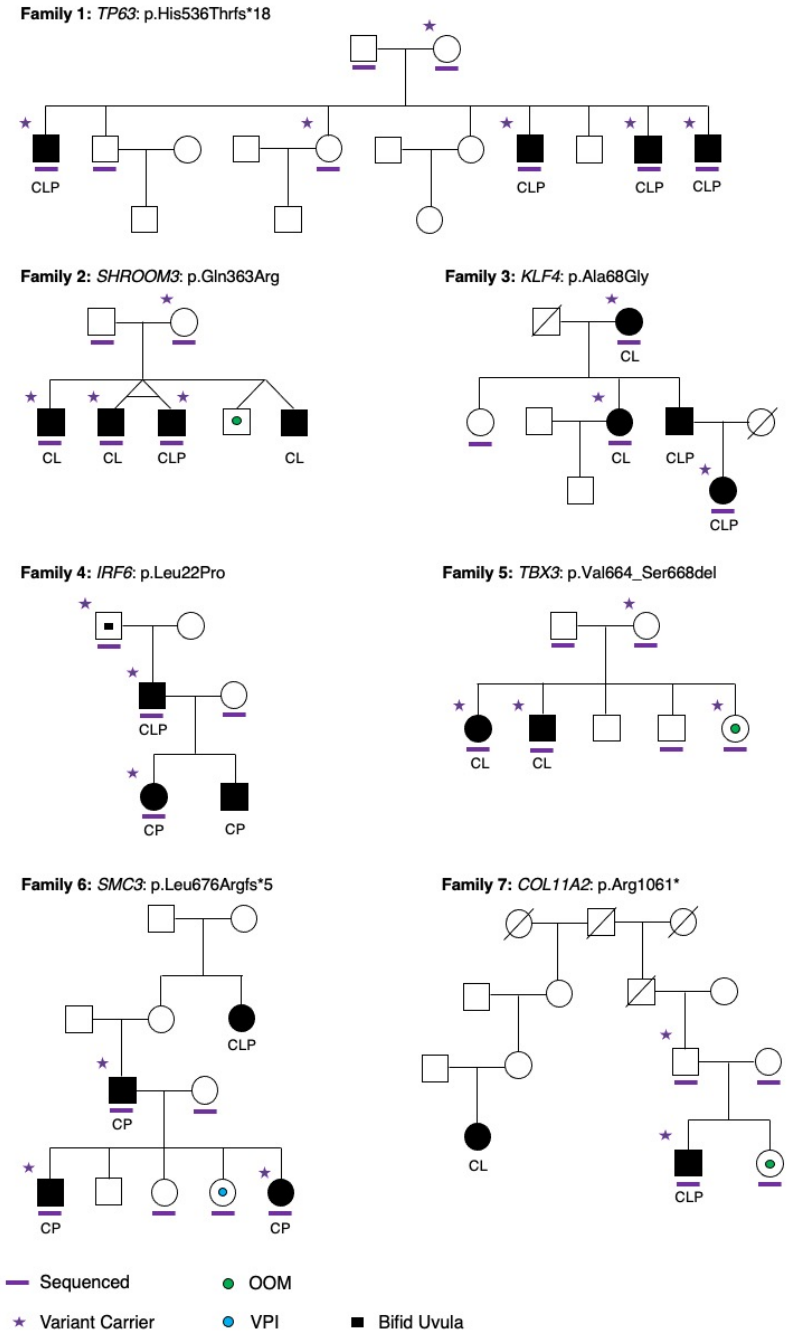
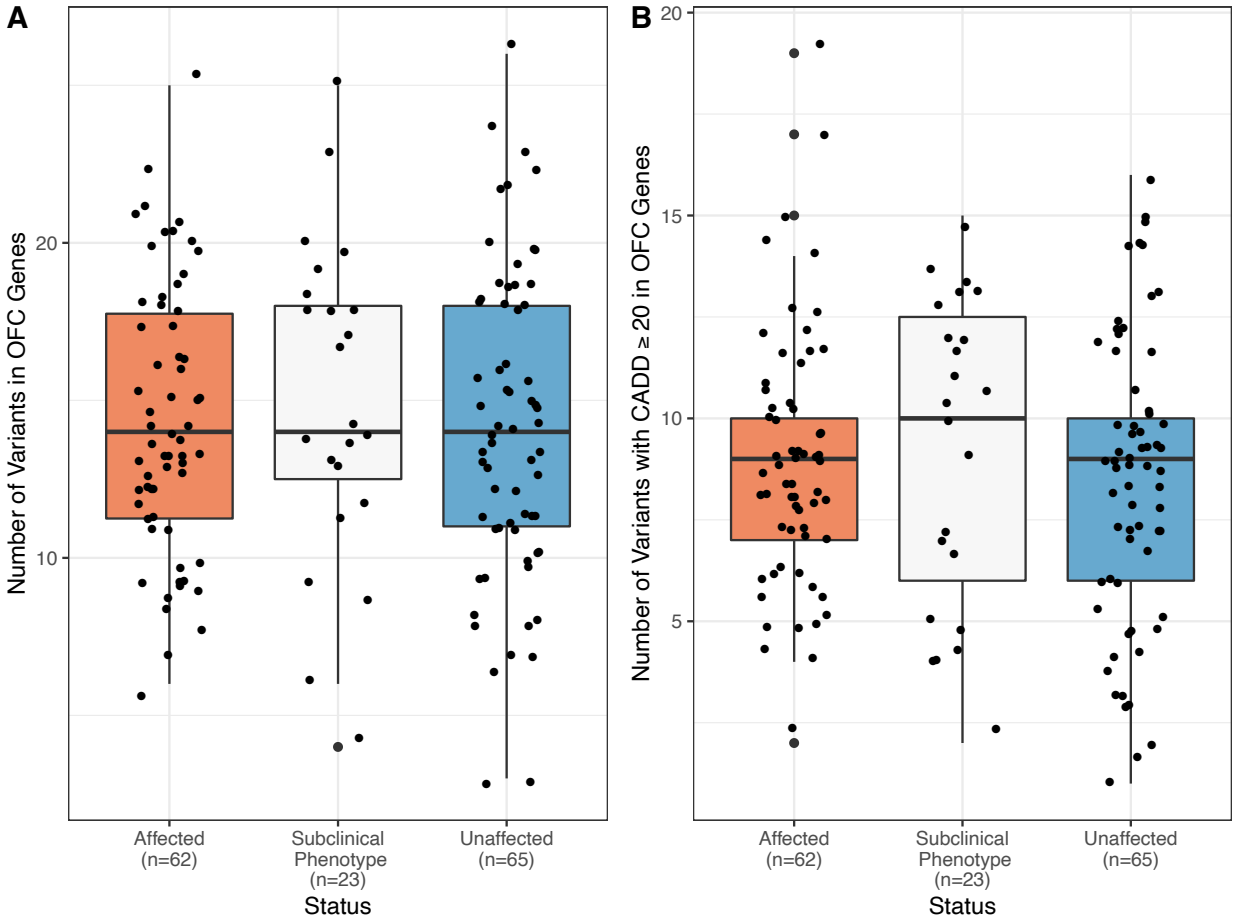


Figure 2-2: Number of Variants in OFC Genes per Family Within Groups. The number of variants in genes associated with OFCs per person (A) overall and (B) variants with a CADD ≥ 20 across affection status, including affected individuals (n=62, orange), individuals with subclinical phenotypes (n=23, gray) and unaffected (n=65, blue) individuals.



Supplementary Tables

Table S2-1: Demographics of Study Cohort

	Number of Samples		
Type of Sample	**Male**	**Female**	**Total**
All Samples	75	75	150
Affected Individuals	37	25	62
Individuals with Subclinical Phenotypes	11	12	23
Individuals with OOM	7	10	17
Individuals with VPI	3	2	5
Individuals with Bifid Uvula	1	0	1
Unaffected Individuals (Including Subclinical Phenotypes)	38	50	88
Unaffected Individuals (No Subclinical Phenotypes)	27	38	65

Table S2-2: Variants of Interest in Multiplex OFC Families

					Segregates Among:	
Family Number	**Gene**	**Variant**	**Allele Frequency**	**CADD**	**OFCs**	**OFCs + Subclinical Phenotypes**
1	*TP63*	NM_003722.5: c.1606delC; NP_003713.3: p.His536Thrfs*18	NA	NA	X	
2	*SHROOM3*	NM_020859.4: c.1088A>G; NP_065910.3: p.Gln363Arg	1.36E-04	22.2	X	
3	*KLF4*	NM_004235.6: c.203C>G; NP_004226.3: p.Ala68Gly	5.84E-04	23.5	X	
4	*IRF6*	NM_006147.4: c.65T>C; NP_006138.1: p.Leu22Pro	NA	19.2		X
5	*TBX3*	NM_005996.4: c.1991_2005delTGGCAGTGGACTCGG; NP_005987.3: p.Val664_Ser668del	9.32E-06	NA		X
6	*SMC3*	NM_005445.4: c.2019_2050del; NP_005436.1: p.Leu676Argfs*5	NA	NA	X	
7	*COL11A2*	NM_080680.3: c.3181C>T; NP_542411.2: p.Arg1061*	7.98E-06	37	X	
8	*HOXA2*	NM_006735.4: c.193C>G; NP_006726.1: p.Pro65Ala	2.52E-04	23.2	X	
9	*AFDN*	NM_001386888.1: c.3545A>G; NP_001373817.1: p.Asn1182Ser	NA	21.7		X
10	*WDR11*	NM_018117.12: c.7C>T; NP_060587.8: p.Pro3Ser	3.73E-05	21.3	X	
	PDGFRB	NM_002609.4: c.1664T>C; NP_002600.1: p.Leu555Pro	7.96E-06	28.9	X	

11	*COL11A2*	NM_080680.3: c.353G>C; NP_542411.2: p.Arg118Pro	2.08E-03	25.1		X
12	*PDGFRA*	NM_006206.6: c.1848C>G; NP_006197.1: p.Ser616Arg	3.98E-06	23.3	X	
13	*COL11A2*	NM_080680.3: c.4408A>G; NP_542411.2: p.Lys1470Glu	NA	24.4		X
	COL11A1	NM_001854.4: c.5401C>G; NP_001845.3: p.Pro1801Ala	1.59E-05	18.2		X
	SHROOM3	NM_020859.4: c.697C>A; NP_065910.3: p.Pro233Thr	1.39E-04	23		X
14	*SMARCE1*	NM_003079.5: c.506C>T; NP_003070.3: p.Pro169Leu	2.01E-05	26.9		X
	IL11RA	NM_001142784.3: c.296G>A; NP_001136256.1: p.Gly99Asp	3.98E-06	26.2		X
15	*GRHL2*	NM_024915.4: c.38C>T; NP_079191.2: p.Ala13Val	NA	17.4		X
16	*ACACB*	NM_001093.4: c.656C>T; NP_001084.3: p.Pro219Leu	4.86E-04	31	X	
	COL11A2	NM_080680.3: c.3173C>T; NP_542411.2: p.Pro1058Leu	3.59E-05	26.4	X	
	CRISPLD2	NM_031476.4: c.142C>T; NP_113664.1: p.Arg48Cys	3.98E-06	29.5		X
	SHROOM3	NM_020859.4: c.3508C>T; NP_065910.3: p.Arg1170Cys	1.72E-04	33		X
17	*RREB1*	NM_001003699.4: c.2411A>G;	1.41E-05	22.7		X

		NP_001003699.1: p.Asn804Ser				
18	*NOTCH2*	NM_024408.4: c.4457C>T; NP_077719.2: p.Thr1486Met	1.19E-05	28.4	X	
	DLG1	NM_001366207.1: c.175A>G; NP_001353136.1: p.Thr59Ala	NA	17.1		X
19	*PTCH1*	NM_000264.5: c.3947A>G; NP_000255.2: p.Tyr1316Cys	5.48E-04	23	X	
	TFAP2C	NM_003222.4: c.58G>T; NP_003213.1: p.Asp20Tyr	5.78E-06	24.6	X	
20	*PVR*	NM_006505.5: c.512G>A; NP_006496.4:p.Gly171Asp	1.61E-05	24.7		X
	PAX3	NM_181457.4: c.1263_1265dupTAC; NP_852122.1: p.Thr424dup	NA	NA		X
21	*SOX9*	NM_000346.4: c.1000G>C; NP_000337.1: p.Val334Leu	5.33E-06	23.4		X
	JAG2	NM_002226.5: c.1886C>T; NP_002217.3: p.Thr629Ile	4.31E-06	25		X
22	*GLI1*	NM_005269.3: c.1165C>G; NP_005260.1: p.Pro389Ala	NA	25.3		X
	RAD21	NM_006265.3: c.1576G>C; NP_006256.1: p.Glu526Gln	1.59E-05	22.4		X
	USF3	NM_001009899.4: c.2144A>G; NP_001009899.3: p.Asn715Ser	NA	20.1	X	
23	*AFDN*	NM_001386888.1: c.1222+1G>T	9.78E-06	25.9	X	
	MYH9	NM_002473.6: c.4745A>G;	NA	28.1	X	

		NP_002464.1: p.Glu1582Gly				
	VWA8	NM_015058.2: c.2003G>A; NP_055873.1: p.Arg668Gln	2.74E-03	34	X	
24	*TULP4*	NM_020245.5: c.2465C>T; NP_064630.2: p.Ala822Val	9.37E-05	23.5	X	
	PLCB4	NM_000933.4: c.1757G>A; NP_000924.3: p.Arg586His	5.63E-05	25.9	X	
25	*MACF1*	NM_012090.5: c.59_76del; NP_036222.3: p.Cys20_Ser25del	1.59E-05	NA	X	
26	*ERBB4*	NM_005235.3: c.259G>C; NP_005226.1: p.Val87Leu	NA	27.4	X	
27	*PVR*	NM_006505.5: c.29C>T; NP_006496.4: p.Pro10Leu	3.45E-05	23	X	
28	*PHF8*	NM_015107.3: c.1726G>A; NP_055922.1: p.Gly576Ser	1.02E-04	28.2	X	
29	*ANK3*	NM_020987.5: c.1448G>A; NP_066267.2: p.Arg483Gln	NA	35	X	
	BOC	NM_001378074.1: c.1186G>C; NP_001365003.1: p.Glu396Gln	4.41E-04	23.9	X	
	NIPBL	NM_133433.4: c.838T>G; NP_597677.2: p.Ser280Ala	2.39E-05	23.6	X	
30	-	-	-	-	-	-
31	-	-	-	-	-	-

References

Alade, A., Awotoye, W., & Butali, A. (2022). Genetic and epigenetic studies in non-syndromic oral clefts. Oral Dis. doi:10.1111/odi.14146

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., & Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res, 43(Database issue), D789-798. doi:10.1093/nar/gku1205

Awotoye, W., Mossey, P. A., Hetmanski, J. B., Gowans, L. J. J., Eshete, M. A., Adeyemo, W. L., . . . Butali, A. (2022). Whole-genome sequencing reveals de-novo mutations associated with nonsyndromic cleft lip/palate. Sci Rep, 12(1), 11743. doi:10.1038/s41598-022-15885-1

Basha, M., Demeer, B., Revencu, N., Helaers, R., Theys, S., Bou Saba, S., . . . Vikkula, M. (2018). Whole exome sequencing identifies mutations in 10% of patients with familial non-syndromic cleft lip and/or palate in genes mutated in well-known syndromes. J Med Genet, 55(7), 449-458. doi:10.1136/jmedgenet-2017-105110

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1 - 48. doi:10.18637/jss.v067.i01

Beames, T. G., & Lipinski, R. J. (2020). Gene-environment interactions: aligning birth defects research with complex etiology. Development, 147(21). doi:10.1242/dev.191064

Beaty, T., Marazita, M., & Leslie, E. (2016). Genetic factors influencing risk to orofacial clefts: today?s challenges and tomorrow?s opportunities [version 1; peer review: 2 approved]. F1000Research, 5(2800). doi:10.12688/f1000research.9503.1

Birnbaum, S., Ludwig, K. U., Reutter, H., Herms, S., Steffens, M., Rubini, M., . . . Mangold, E. (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet, 41(4), 473-477. doi:10.1038/ng.333

Bishop, M. R., Diaz Perez, K. K., Sun, M., Ho, S., Chopra, P., Mukhopadhyay, N., . . . Leslie, E. J. (2020). Genome-wide Enrichment of De Novo Coding Mutations in Orofacial Cleft Trios. The American Journal of Human Genetics. doi:10.1016/j.ajhg.2020.05.018

Bureau, A., Parker, M. M., Ruczinski, I., Taub, M. A., Marazita, M. L., Murray, J. C., . . . Beaty, T. H. (2014). Whole exome sequencing of distant relatives in multiplex families implicates rare variants in candidate genes for oral clefts. Genetics, 197(3), 1039-1044. doi:10.1534/genetics.114.165225

Carlson, J. C., Standley, J., Petrin, A., Shaffer, J. R., Butali, A., Buxó, C. J., . . . Leslie, E. J. (2017). Identification of 16q21 as a modifier of nonsyndromic orofacial cleft phenotypes. Genet Epidemiol, 41(8), 887-897. doi:10.1002/gepi.22090

Copp, A. J., & Greene, N. D. (2013). Neural tube defects--disorders of neurulation and related embryonic processes. Wiley Interdiscip Rev Dev Biol, 2(2), 213-227. doi:10.1002/wdev.71

Cox, L. L., Cox, T. C., Moreno Uribe, L. M., Zhu, Y., Richter, C. T., Nidey, N., . . . Roscioli, T. (2018). Mutations in the Epithelial Cadherin-p120-Catenin Complex Cause Mendelian Non-Syndromic Cleft Lip with or without Cleft Palate. Am J Hum Genet, 102(6), 1143-1157. doi:10.1016/j.ajhg.2018.04.009

Das, D., Zalewski, J. K., Mohan, S., Plageman, T. F., VanDemark, A. P., & Hildebrand, J. D. (2014). The interaction between Shroom3 and Rho-kinase is required for neural tube morphogenesis in mice. Biol Open, 3(9), 850-860. doi:10.1242/bio.20147450

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., . . . Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet, 43(5), 491-498. doi:10.1038/ng.806

Diaz Perez, K. K., Curtis, S. W., Sanchis-Juan, A., Zhao, X., Head, S. T., Ho, S., . . . Leslie, E. J. (2022). Rare variants found in clinical gene panels illuminate the genetic and allelic architecture of orofacial clefting. medRxiv, 2022.2008.2015.22278795. doi:10.1101/2022.08.15.22278795

Dixon, M. J., Marazita, M. L., Beaty, T. H., & Murray, J. C. (2011). Cleft lip and palate: understanding genetic and environmental influences. Nat Rev Genet, 12(3), 167-178. doi:10.1038/nrg2933

Garland, M. A., Reynolds, K., & Zhou, C. J. (2020). Environmental mechanisms of orofacial clefts. Birth defects research, 112(19), 1660-1698. doi:10.1002/bdr2.1830

Ghassibé, M., Revencu, N., Bayet, B., Gillerot, Y., Vanwijck, R., Verellen-Dumoulin, C., & Vikkula, M. (2004). Six families with van der Woude and/or popliteal pterygium syndrome: all with a mutation in the IRF6 gene. Journal of Medical Genetics, 41(2), e15-e15. doi:10.1136/jmg.2003.009274

Gil-Rodriguez, M. C., Deardorff, M. A., Ansari, M., Tan, C. A., Parenti, I., Baquero-Montoya, C., . . . Pie, J. (2015). De novo heterozygous mutations in SMC3 cause a range of Cornelia de Lange syndrome-overlapping phenotypes. Hum Mutat, 36(4), 454-462. doi:10.1002/humu.22761

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., & McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res, 33(Database issue), D514-D517. doi:10.1093/nar/gki033

Henrik, S., Ben, B., Jake, W., Frederik, A., & Ben-Shachar, M. S. (2022). afex: Analysis of Factorial Experiments. Retrieved from https://CRAN.R-project.org/package=afex

Hildebrand, J. D., & Soriano, P. (1999). Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice. Cell, 99(5), 485-497. doi:10.1016/s0092-8674(00)81537-8

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., . . . MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. bioRxiv, 531210. doi:10.1101/531210

Khandelwal, K. D., van den Boogaard, M. H., Mehrem, S. L., Gebel, J., Fagerberg, C., van Beusekom, E., . . . van Bokhoven, H. (2019). Deletions and loss-of-function variants in TP63 associated with orofacial clefting. Eur J Hum Genet, 27(7), 1101-1112. doi:10.1038/s41431-019-0370-0

Kingdom, R., & Wright, C. F. (2022). Incomplete Penetrance and Variable Expressivity: From Clinical Studies to Population Cohorts. Frontiers in Genetics, 13. doi:10.3389/fgene.2022.920390

Kline, A. D., Krantz, I. D., Sommer, A., Kliewer, M., Jackson, L. G., FitzPatrick, D. R., . . . Selicorni, A. (2007). Cornelia de Lange syndrome: clinical review, diagnostic and scoring systems, and anticipatory guidance. Am J Med Genet A, 143a(12), 1287-1296. doi:10.1002/ajmg.a.31757

Kondo, S., Schutte, B. C., Richardson, R. J., Bjork, B. C., Knight, A. S., Watanabe, Y., . . . Murray, J. C. (2002). Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet, 32(2), 285-289. doi:10.1038/ng985

Kotlar, A. V., Trevino, C. E., Zwick, M. E., Cutler, D. J., & Wingo, T. S. (2018). Bystro: rapid online variant annotation and natural-language filtering at whole-genome scale. Genome Biol, 19(1), 14. doi:10.1186/s13059-018-1387-3

Kryukov, G. V., Pennacchio, L. A., & Sunyaev, S. R. (2007). Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am J Hum Genet, 80(4), 727-739. doi:10.1086/513473

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., . . . Maglott, D. R. (2018). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res, 46(D1), D1062-d1067. doi:10.1093/nar/gkx1153

Lansdon, L. A., Dickinson, A., Arlis, S., Liu, H., Hlas, A., Hahn, A., . . . Manak, J. R. (2023). Genome-wide analysis of copy-number variation in humans with cleft lip and/or cleft palate identifies COBLL1, RIC1, and ARHGEF38 as clefting genes. Am J Hum Genet, 110(1), 71-91. doi:10.1016/j.ajhg.2022.11.012

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . MacArthur, D. G. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature, 536(7616), 285-291. doi:10.1038/nature19057

Leslie, E. J. (2022). Genetic models and approaches to study orofacial clefts. Oral Dis, 28(5), 1327-1338. doi:10.1111/odi.14109

Leslie, E. J., Carlson, J. C., Shaffer, J. R., Butali, A., Buxo, C. J., Castilla, E. E., . . . Marazita, M. L. (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(3), 275-286. doi:10.1007/s00439-016-1754-7

Leslie, E. J., Liu, H., Carlson, J. C., Shaffer, J. R., Feingold, E., Wehby, G., . . . Marazita, M. L. (2016). A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. Am J Hum Genet, 98(4), 744-754. doi:10.1016/j.ajhg.2016.02.014

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv, 1303.3997v1(q-Bio.GN).

Liu, H., Busch, T., Eliason, S., Anand, D., Bullard, S., Gowans, L. J. J., . . . Butali, A. (2017). Exome sequencing provides additional evidence for the involvement of ARHGAP29 in Mendelian orofacial clefting and extends the phenotypic spectrum to isolated cleft palate. Birth defects research, 109(1), 27-37. doi:10.1002/bdra.23596

Liu, H., Duncan, K., Helverson, A., Kumari, P., Mumm, C., Xiao, Y., . . . Cornell, R. A. (2020). Analysis of zebrafish periderm enhancers facilitates identification of a regulatory variant near human KRT8/18. eLife, 9, e51325. doi:10.7554/eLife.51325

Liu, H., Leslie, E. J., Jia, Z., Smith, T., Eshete, M., Butali, A., . . . Cornell, R. A. (2016). Irf6 directly regulates Klf17 in zebrafish periderm and Klf4 in murine oral epithelium, and dominant-negative KLF4 variants are present in patients with cleft lip and palate. Hum Mol Genet, 25(4), 766-776. doi:10.1093/hmg/ddv614

López, S. H., Avetisyan, M., Wright, C. M., Mesbah, K., Kelly, R. G., Moon, A. M., & Heuckeroth, R. O. (2018). Loss of Tbx3 in murine neural crest reduces enteric glia and causes cleft palate, but does not influence heart development or bowel transit. Developmental Biology, 444 Suppl 1(Suppl 1), S337-s351. doi:10.1016/j.ydbio.2018.09.017

Mangold, E., Ludwig, K. U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., . . . Nothen, M. M. (2010). Genome-wide association study identifies two susceptibility loci for

nonsyndromic cleft lip with or without cleft palate. Nat Genet, 42(1), 24-26. doi:10.1038/ng.506

Marazita, M. L. (2012). The evolution of human genetic studies of cleft lip and cleft palate. Annu Rev Genomics Hum Genet, 13, 263-283. doi:10.1146/annurev-genom-090711-163729

Marosy, B. A., Craig, B. D., Hetrick, K. N., Witmer, P. D., Ling, H., Griffith, S. M., . . . Doheny, K. F. (2017). Generating Exome Enriched Sequencing Libraries from Formalin-Fixed, Paraffin-Embedded Tissue DNA for Next-Generation Sequencing. Curr Protoc Hum Genet, 92, 18.10.11-18.10.25. doi:10.1002/cphg.27

Martin, A. R., Williams, E., Foulger, R. E., Leigh, S., Daugherty, L. C., Niblock, O., . . . McDonagh, E. M. (2019). PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet, 51(11), 1560-1565. doi:10.1038/s41588-019-0528-2

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res, 20(9), 1297-1303. doi:10.1101/gr.107524.110

Neiswanger, K., Walker, K., Klotz, C. M., Cooper, M. E., Bardi, K. M., Brandon, C. A., . . . Marazita, M. L. (2009). Whorl patterns on the lower lip are associated with nonsyndromic cleft lip with or without cleft palate. Am J Med Genet A, 149A(12), 2673-2679. doi:10.1002/ajmg.a.33089

Neiswanger, K., Weinberg, S. M., Rogers, C. R., Brandon, C. A., Cooper, M. E., Bardi, K. M., . . . Marazita, M. L. (2007). Orbicularis oris muscle defects as an expanded phenotypic feature in nonsyndromic cleft lip with or without cleft palate. Am J Med Genet A, 143a(11), 1143-1149. doi:10.1002/ajmg.a.31760

Nikopensius, T., Jagomägi, T., Krjutskov, K., Tammekivi, V., Saag, M., Prane, I., . . . Metspalu, A. (2010). Genetic variants in COL2A1, COL11A2, and IRF6 contribute risk to nonsyndromic cleft palate. Birth Defects Res A Clin Mol Teratol, 88(9), 748-756. doi:10.1002/bdra.20700

Peyrard-Janvid, M., Leslie, E. J., Kousa, Y. A., Smith, T. L., Dunnwald, M., Magnusson, M., . . . Schutte, B. C. (2014). Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. Am J Hum Genet, 94(1), 23-32. doi:10.1016/j.ajhg.2013.11.009

Rahimov, F., Marazita, M. L., Visel, A., Cooper, M. E., Hitchler, M. J., Rubini, M., . . . Murray, J. C. (2008). Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. Nat Genet, 40(11), 1341-1347. doi:10.1038/ng.242

Ray, D., Venkataraghavan, S., Zhang, W., Leslie, E. J., Hetmanski, J. B., Weinberg, S. M., . . . Beaty, T. H. (2021). Pleiotropy method reveals genetic overlap between orofacial clefts at multiple novel loci from GWAS of multi-ethnic trios. PLoS Genet, 17(7), e1009584. doi:10.1371/journal.pgen.1009584

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2018). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res, 47(D1), D886-D894. doi:10.1093/nar/gky1016

Segre, J. A., Bauer, C., & Fuchs, E. (1999). Klf4 is a transcription factor required for establishing the barrier function of the skin. Nat Genet, 22(4), 356-360. doi:10.1038/11926

Stanier, P., & Moore, G. E. (2004). Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. Hum Mol Genet, 13(suppl_1), R73-R81. doi:10.1093/hmg/ddh052

Suzuki, S., Marazita, M. L., Cooper, M. E., Miwa, N., Hing, A., Jugessur, A., . . . Murray, J. C. (2009). Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. Am J Hum Genet, 84(3), 406-411. doi:10.1016/j.ajhg.2009.02.002

Team, R. C. (2021). R: A Language and Environment for Statistical Computing: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

van Steensel, M. A., Buma, P., de Waal Malefijt, M. C., van den Hoogen, F. H., & Brunner, H. G. (1997). Oto- spondylo-megaepiphyseal dysplasia (OSMED): clinical description of three patients homozygous for a missense mutation in the COL11A2 gene. Am J Med Genet, 70(3), 315-323. doi:10.1002/(sici)1096-8628(19970613)70:3<315::aid-ajmg19>3.3.co;2-y

Vikkula, M., Mariman, E. C., Lui, V. C., Zhidkova, N. I., Tiller, G. E., Goldring, M. B., . . . *et al*. (1995). Autosomal dominant and recessive osteochondrodysplasias associated with the COL11A2 locus. Cell, 80(3), 431-437. doi:10.1016/0092-8674(95)90493-x

Weinberg, S. M., Neiswanger, K., Martin, R. A., Mooney, M. P., Kane, A. A., Wenger, S. L., . . . Marazita, M. L. (2006). The Pittsburgh Oral-Facial Cleft study: expanding the cleft phenotype. Background and justification. Cleft Palate Craniofac J, 43(1), 7-20. doi:10.1597/04-122r1.1

Yu, Y., Zuo, X., He, M., Gao, J., Fu, Y., Qin, C., . . . Bian, Z. (2017). Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. Nat Commun, 8, 14364-14364. doi:10.1038/ncomms14364

Zieger, H. K., Weinhold, L., Schmidt, A., Holtgrewe, M., Juranek, S. A., Siewert, A., . . . Ludwig, K. U. (2023). Prioritization of non-coding elements involved in non-syndromic cleft lip with/without cleft palate through genome-wide analysis of de novo mutations. HGG Adv, 4(1), 100166. doi:10.1016/j.xhgg.2022.100166

Zucchero, T. M., Cooper, M. E., Maher, B. S., Daack-Hirsch, S., Nepomuceno, B., Ribeiro, L., . . . Murray, J. C. (2004). Interferon Regulatory Factor 6 (IRF6) Gene Variants and the Risk of Isolated Cleft Lip or Palate. New England Journal of Medicine, 351(8), 769-780. doi:10.1056/NEJMoa032909

CHAPTER III. Identifying rare recessive mutations in consanguineous families with orofacial clefts

Kimberly K. Diaz Perez, Lina M. Moreno Uribe, Luz Consuelo Valencia-Ramirez, Claudia Restrepo, Eleanor Feingold, Mary L. Marazita, Elizabeth J. Leslie

Adapted from manuscript currently in preparation.

Introduction

Consanguineous unions are a form of non-random mating that occurs between individuals who share a recent common ancestor. Consanguinity is still preferentially practiced worldwide, and the prevalence varies by ethnicity, religion, and geography (Fareed & Afzal, 2017; Hamamy *et al*., 2011). Consanguinity is considered a founder effect due to the presence of a few different common ancestors, which results in an unequal distribution of founder mutations that leads to high levels of homozygosity (Woods *et al*., 2006). In populations with a high prevalence of consanguinity, there is a greater likelihood for these unions to share a substantial portion of the genome and their offspring are at a greater risk of being affected by recessive genetic diseases (Alkuraya, 2012; Hamamy *et al*., 2011). Consanguinity has also been associated with a high prevalence of congenital anomalies (Hamamy, 2012; Stoll *et al*., 1999; Tayebi *et al*., 2010).

Consanguinity has been investigated as a risk factor in orofacial clefts (OFCs), a common craniofacial congenital defect that affects 1 in 1,000 live births worldwide. Previous epidemiological studies suggest that consanguinity is a risk factor for OFCs (Elahi *et al*., 2004; Jamilian *et al*., 2007; Rittler *et al*., 2001; Stoll *et al*., 1991). For example, a meta-analysis of case-control studies using consanguineous families found that consanguinity results in a two-fold increase in the risk for OFCs (Sabbagh *et al*., 2014). In non-syndromic OFCs, which are considered to have a complex etiology comprised of an interplay of genetic and environmental factors, consanguineous families have been examined considering different genetic approaches. For instance, Camargo *et al*. (2012) conducted a genome-wide association study of 40 consanguineous families with extended pedigrees to identify recessive loci associated with OFC risk and identified various novel significant recessive susceptibility regions, such as *CSMD1*

(Camargo *et al*., 2012). Other family-based analyses have found individual rare recessive mutations in OFC cases from consanguineous unions, highlighting the utility of consanguineous pedigrees to discover additional risk loci for OFCs (Al Mahdi *et al*., 2020; Holzinger *et al*., 2017).

In this study, we aim to identify rare variants in a cohort of 50 consanguineous OFC case-parent trios recruited from Colombia and Turkey, where consanguinity is prevalent. The consanguinity prevalence rates for these populations are estimated to range from 1.3% to 18.5% (De Castro & Restrepo, 2015; Kaplan *et al*., 2016).

Methods

Study Cohort Information

The study cohort is derived from two populations of case-parent trios with OFCs sequenced by the Gabriella Miller Kids First (GMKF) Pediatric Research Consortium. We used kinship inference (described in detail below) to identify 50 probands whose parents are 2nd, 3rd, or 4th-degree relatives from Turkey (N = 23) or Colombia (N = 27). In two trios from Turkey, one parent was affected, each with cleft lip (CL). A review of recorded pedigrees showed that most of the identified pedigrees were known to be consanguineous.

Whole-Genome Sequencing and Variant Calling

The sequencing and variant calling procedures for the GMKF cohort are detailed in Bishop *et al*. (2020) and Diaz Perez *et al*. (2022). Briefly, we applied variant call filters in VCFtools (version 0.1.13) and retained genotype calls for variants with a genotype quality ≥ 20,

Quality Normalized by Depth (QD) score > 4, and read depth ≥ 10. Sites with missingness > 5% were excluded.

Kinship Inference of GMKF Trios

We applied a sample allele frequency filter > 0.2% and performed pairwise linkage disequilibrium pruning in PLINK v1.90b5.3 with a 50 kb window size, a five variant window shift, and a pairwise r2 threshold of 0.5 (Purcell *et al*., 2007). We then calculated identity-by-descent sharing using the Kinship-based Inference for GWAS (KING) software (version 2.2.5) (Manichaikul *et al*., 2010) and identified 2nd, 3rd, and 4th-degree relationships between parent pairs.

Variant Filtering and Annotation

We considered an autosomal recessive inheritance model and extracted all homozygous alternate variant calls from OFC cases using VarSeq v2.2.5 (Golden Helix, Inc., Bozeman, MT). We retained exonic variants that met the following filters: 1) missense, canonical splice acceptor or donor, nonsense, or frameshifting variants, and 2) a global minor allele frequency ≤ 0.5% from the gnomAD genomes database (v2 and v3) (Karczewski *et al*., 2020). All homozygous variants were then annotated using additional information from a literature review, Online Mendelian Inheritance in Man (Hamosh *et al*., 2005), ClinVar (Landrum *et al*., 2018), CADD scores (Rentzsch *et al*., 2018), and gene tolerance metrics from gnomAD (Karczewski *et al*., 2020; Lek *et al*., 2016). We also annotated genes for craniofacial expression using single-nuclei RNA-sequencing data from human craniofacial tissues at Carnegie Stage 20 (CS20), which

corresponds to approximately week 11 post-conception (http://cotneyweb.cam.uchc.edu/craniofacial_cs20/) (Yankee *et al*., 2022).

Results

We studied 23 Turkish and 27 Colombian trios in which the parents were 2nd, 3rd, or 4th-degree relatives (Figure 3-1). We identified an average of 13 rare, homozygous variants per trio (range: 3-38 variants, total: 304) in the Turkish trios and an average of 10 variants (range: 0-27 variants, total: 257) per trio in the Colombian trios. Most of these variants (95%) were missense. The number of homozygous variants decreased as the genetic distance between the parents increased (Spearman's rank order correlation: rho=-55.7, p=2.67 x 10-05).

Turkish Consanguineous Trios

After considering variant frequencies, annotations, and bioinformatic predictions, we prioritized "variants of interest." We found such variants in six out of 23 (26.1%) Turkish trios (Table 3-1).

Family 1 (CP): Of the 14 qualifying variants, we considered three of interest: *GRHL3* (NM_198173.3: c.1586C>T; NP_937816.1: p.Ala529Val), *CFAP57* (NM_001378189.1: c.1855C>T; NP_001365118.1: p.Arg619Cys), and *TMEM94* (NM_014738.6: c.1238T>C; NP_055553.3: p.Leu413Pro), the latter of which is absent from gnomAD. All three variants had high CADD scores (34 for *GRHL3*, 35 for *CFAP57*, and 28 for *TMEM94*). Heterozygous mutations in *GRHL3* are associated with Van der Woude syndrome and non-syndromic CP (Eshete *et al*., 2018; Leslie *et al*., 2016; Mangold *et al*., 2016; Peyrard-Janvid *et al*., 2014).

Grhl3-deficient mice exhibit a cleft palate (Peyrard-Janvid *et al*., 2014). *CFAP57* is a target of transcription factor *IRF6* and is expressed in the medial edge and nasal epithelia of the palatal shelves (Rorick *et al*., 2011). Sequencing of individuals with Van der Woude syndrome and no causal variants in *IRF6* (one of the genes known to cause this syndrome) identified a single individual from Brazil with a heterozygous missense variant, p.Asp523Tyr, in *CFAP57* (Rorick *et al*., 2011). There is, however, no precedence for homozygous variants causing human phenotypes in this gene and it shows no evidence of constraint for missense or loss-of-function variants. Recessive variants in *TMEM94* cause intellectual disability disorder with cardiac defects and dysmorphic facial features. However, the causal variants for *TMEM94* have primarily been loss-of-function, not missense as is the case for this variant.

Family 2 (CLP): We found a variant of interest (out of 7 variants total) in *SLC35C1* (NM_018389.5: c.396C>A; NP_060859.4: p.Asn132Lys). Recessive missense and loss-of-function variants in *SLC35C1* cause Congenital Disorder of Glycosylation (CDG), characterized by psychomotor retardation, short stature, and facial dysmorphism, including thickened facial skin and coarse facial features (Lübke *et al*., 2001); however, no individuals with CDG and OFCs have been previously reported. The variant in *SLC35C1* had a CADD score of 32 and was absent from gnomAD.

Family 3 (CL): Out of 10 variants total, we identified one variant of interest in *PLEKHA7* (NM_001329630.2: c.1796C>T; NP_001316559.1: p.Pro599Leu). This amino acid substitution is located in a domain associated with interactions with *CTNND1*, a gene in which dominant mutations cause OFCs (Ghoumid *et al*., 2017). Cox *et al*. (2018) found likely causal compound

heterozygous missense variants (p.Arg513Trp and p.Asp662Gly) in *PLEKHA7* in a single non-syndromic OFC family; both variants were also located in or near the region that interacts with *CTNND1* (Cox *et al*., 2018). In addition, PLEKHA7 is expressed in epithelial structures of the developing human palate, including the oral epithelium, midline epithelial seam, and periderm (Cox *et al*., 2018).

Family 4 (CLP): Out of 7 qualifying variants, we identified two missense variants in *TTC28* (NM_001145418.1: c.774G>T; NP_001138890.1: p.Gln258His) and *DKK1* (NM_012242.4:c.359G>T; NP_036374.1: p.Arg120Leu). The variant in *TTC28* is absent from gnomAD with a CADD score of 24. *TTC28* is highly expressed in the developing face at CS20, with strong expression in the endothelium, ectoderm, and mesenchyme. A previous report described *TTC28* as the likely causal gene among those deleted in Pierre-Robin sequence, which includes CP as a primary clinical feature (Davidson *et al*., 2012). Moreover, multiple copy number variants including *TTC28* have also been observed in CP cases (Conte *et al*., 2016). The variant in *DKK1* has a CADD score of 35 and a frequency of 0.2-0.4% in gnomAD. *DKK1* is a target of PAX9 in palate development, and reducing its activity corrects the palatal defects in Pax9-/- mouse embryos; overexpression of *DKK1* results in cleft palate (Jia *et al*., 2020). *DKK1* is also an inhibitor of the Wnt signaling pathway, where Dkk1-/- mice have craniofacial defects, including anomalies in the frontonasal mass and mandibular processes, supporting its involvement in the regulation of cranial development (Mukhopadhyay *et al*., 2001)

Family 5 (CP): Out of 14 qualifying variants, we found a missense variant in *CSMD1* (NM_033225.6: c.5432G>A; NP_150094.5: p.Gly1811Asp), a gene expressed in mesenchymal

and ectodermal cells of human embryonic facial tissue. This substitution has a CADD score of 27 and it is absent from gnomAD. This variant is present in the CUB 11 domain, and these domains are generally involved in a range of functions in other genes, including developmental patterning and tissue repair (Blanc *et al*., 2007). Common variants in *CSMD1* were associated with non-syndromic OFCs using a recessive model in a family-based association study of 40 extended consanguineous families (Camargo *et al*., 2012).

Family 6 (CP): We found a variant of interest (out of 19 total variants) in *TOPORS* (NM_005802.5: c.1408G>T; NP_005793.2: p.Asp470Tyr). This variant is classified as a "Variant of Uncertain Significance" (VUS) on ClinVar because although it is absent from gnomAD, computational predictions do not agree or cannot model the variant. The variant results in an amino acid change that does not alter the protein properties (change from acidic and polar to neutral and polar). Typically, loss-of-function variants in *TOPORS* cause Retinitis Pigmentosa, but a case study recently reported an individual with Oral-facial-digital syndrome (and CP) with a homozygous missense variant in this gene (Strong *et al*., 2021). TOPORS is a ligase of *SUMO1*, a gene previously associated with OFCs, and the p.Asp470Tyr substitution is located in a region that interacts with *SUMO1* (Alkuraya *et al*., 2006; Czub *et al*., 2016).

Colombian Consanguineous Trios

In the Colombian trios, we identified two out of 27 (7.4%) total trios with interesting homozygous variants (Table 3-1).

Family 7 (CLP): We found a 1 bp deletion (out of 9 qualifying variants) (NM_144966.7: c.318_319delCA; NP_659403.4: p.Arg107Thrfs*6) in *FREM1*, which is absent from gnomAD. Recessive missense and loss-of-function mutations in *FREM1* cause Manitoba Oculotrichoanal syndrome, which is characterized by microphthalmia, bifid or broad nasal tip, and gastrointestinal anomalies, whereas dominant missense mutations cause Trigonocephaly. FREM1 is an extracellular protein involved in the differentiation of the epidermis and its expression has been noted in relevant craniofacial regions in mice, especially in areas of epithelial and mesenchymal transitions (Smyth *et al*., 2004). Common variants near *FREM1* are associated with velopharyngeal dysfunction (Chernus *et al*., 2018). Previously, a case report described an individual with CP with a distant family history of OFCs (from the maternal second cousin) with a *de novo* genomic rearrangement of an inverted duplication of 9p24 to 9p21.3, which captures the complete *FREM1* gene along 44 other genes (Hulick *et al*., 2009).

Family 8 (CLP): We identified a missense substitution on *SUCO* (NM_014283.5: c.1094A>G; NP_055098.1: p.Asp365Gly). This variant is not present in the gnomAD database and has a CADD score of 28. SUCO is necessary for bone modeling during the late stages of embryogenesis, and it is involved in osteoblast proliferation and differentiation (Sha *et al*., 2015; Sohaskey *et al*., 2010).

Discussion

Consanguinity is associated with an increased risk for congenital defects, including orofacial clefts (Hamamy, 2012; Stoll *et al*., 1999; Tayebi *et al*., 2010). Here, we studied 50 consanguineous OFC trios from two different populations to identify rare homozygous variants

possibly contributing to OFC risk based on functional computational annotations that are publicly available. We found several variants of interest in a diverse group of genes, several of which were previously associated with OFCs or craniofacial development. However, only one variant (the 1 bp deletion in *FREM1*) was considered "likely pathogenic" according to the American College of Medical Genetics and Genomics guidelines. A significant challenge of this study was that the remaining variants were classified as "variants of uncertain significance" (VUS). Although a subset of these variants met a few of the criteria for a "pathogenic" classification based on in silico predictions or being absent from major control population databases, these criteria only provide supporting or moderate evidence for the overall variant classification. In addition, all variants classified as VUS were missense, which are typically more challenging to interpret their effect than loss-of-function variants. Additional variant- and gene-specific information is needed to determine whether these variants are likely pathogenic in these individuals.

We found several compelling examples of variants in both genes implicated in OFCs previously, such as *CFAP57* and *CSMD1*, and those that have not been implicated in OFCs, such as *SUCO* and *SLC35C1*. Although a missense variant in *CFAP57* has been identified in an individual with syndromic OFC (Rorick *et al*., 2011), this study is the first instance of a rare homozygous missense variant in this gene being implicated in non-syndromic OFCs, which provides evidence to the etiological spectrum overlap of syndromic/non-syndromic OFC. Similarly, our findings provide evidence for rare homozygous point mutations in *CSMD1*, a candidate gene found in an association study of common variants in consanguineous families from Colombia (Camargo *et al*., 2012). In contrast, *SUCO* has not been implicated in OFCs or craniofacial development previously; however, *SUCO* is highly expressed in mesenchymal,

endothelial, and ectodermal cells in the developing face at CS20, suggesting this gene may be involved in novel pathways involved in craniofacial development. Overall, we identified variants in 11 different genes, which could provide evidence of multiple genes in different families contributing to disease pathogenesis. But we did not identify any genes that had multiple variants of interest or variants of interest that recurred in each population. These findings provide insight into the potential role of rare homozygous variants in these genes as causal mechanisms conferring the OFC risk in Colombian and Turkish cases from this cohort.

We also identified homozygous variants in genes associated with Mendelian syndromes that can include OFCs or craniofacial phenotypes. Some of these variants were missense variants in genes where syndromes are caused by heterozygous loss-of-function variants, such as *GRLH3*. If these variants are causal, we hypothesize these variants might be hypomorphic alleles. With this model, the homozygous variants we identified in dominant genes could produce 50% of the protein levels, and result in the OFC phenotype, while parents who are heterozygous for these variants would produce between 50%-100% of the protein levels and, thus, do not exhibit OFCs (Leslie *et al*., 2015).

In summary, we reported the identification and analysis of consanguineous OFC families from Colombia and Turkey. We found several candidate rare homozygous variants in genes including *CFAP57*, *CSMD1*, and *FREM1*, among others. These findings highlighted the etiological heterogeneity of OFCs and the different mechanisms by which rare homozygous variants can potentially modify OFC risk. Taken together, studying populations with high consanguinity rates proved to be a reasonable mechanism for identifying likely causal variants in candidate genes possibly associated with OFC risk. However, functional studies of these

candidate variants are needed to dissect their molecular function and contribution to the OFC phenotype in these consanguineous families.

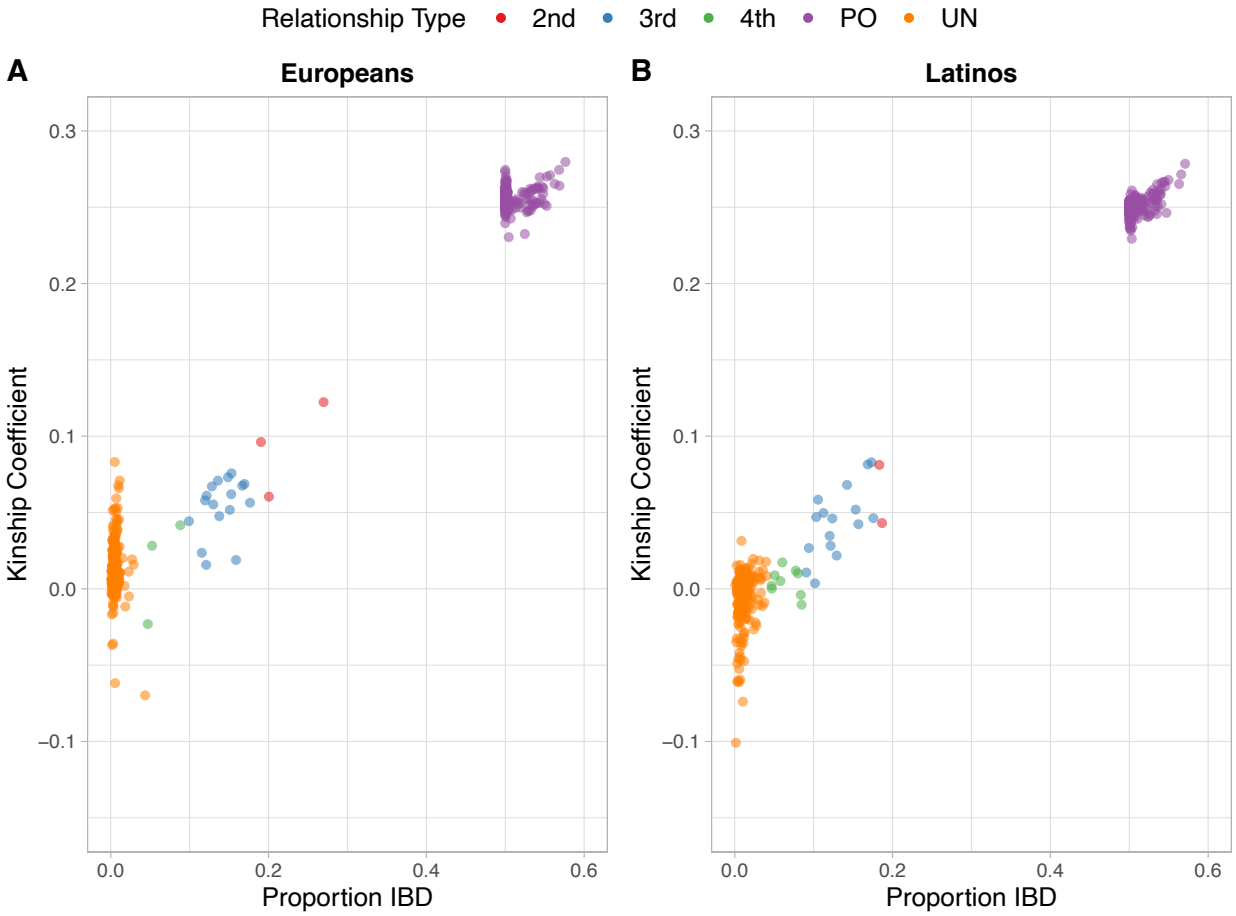
Tables

Table 3-1: Rare Homozygous Variants of Interest in OFC Cases from Turkey and Colombia

Family	Gene	Variant	Allele Frequency	CADD
Turkish Trios				
1	*GRHL3*	NM_198173.3: c.1586C>T; NP_937816.1: p.Ala529Val	5.26E-05	34
	CFAP57	NM_001378189.1: c.1855C>T; NP_001365118.1: p.Arg619Cys	1.54E-03	35
	TMEM94	NM_014738.6: c.1238T>C; NP_055553.3: p.Leu413Pro	-	28
2	*SLC35C1*	NM_018389.5: c.396C>A; NP_060859.4: p.Asn132Lys	-	32
3	*PLEKHA7*	NM_001329630.2: c.1796C>T; NP_001316559.1: p.Pro599Leu	3.09E-04	15
4	*TTC28*	NM_001145418.1: c.774G>T; NP_001138890.1: p.Gln258His	-	24
	DKK1	NM_012242.4: c.359G>T; NP_036374.1: p.Arg120Leu	2.65E-03	35
5	*CSMD1*	NM_033225.6: c.5432G>A; NP_150094.5: p.Gly1811Asp	-	27
6	*TOPORS*	NM_005802.5: c.1408G>T; NP_005793.2: p.Asp470Tyr	-	24
Colombian Trios				
7	*FREM1*	NM_144966.7: c.318_319delCA; NP_659403.4: p.Arg107Thrfs*6	-	-
8	*SUCO*	NM_014283.5: c.1094A>G; NP_055098.1: p.Asp365Gly	-	28

Figures

Figure 3-1: Kinship Estimation of Orofacial Cleft Trios from Turkey and Colombia. This figure shows the kinship coefficient for (A) European and (B) Latino trios along with the proportion shared identity-by-descent (IBD) from the Kinship-based Inference for GWAS (KING) software. The point colors indicate the relationship type of sample pairs, including 2nd-degree (red), 3rd-degree (blue), 4th-degree (green), parent-offspring (purple), and unrelated pairs (orange).



References

Al Mahdi, H. B., Edris, S., Bahieldin, A., Al-Aama, J. Y., Elango, R., Jamalalail, B. A., & Sabbagh, H. J. (2020). Identification of Causative Variants Contributing to Nonsyndromic Orofacial Clefts Using Whole-Exome Sequencing in a Saudi Family. Genetic Testing and Molecular Biomarkers, 24(11), 723-731. doi:10.1089/gtmb.2019.0233

Alkuraya, F. S. (2012). Discovery of rare homozygous mutations from studies of consanguineous pedigrees. Curr Protoc Hum Genet, Chapter 6, Unit6.12. doi:10.1002/0471142905.hg0612s75

Alkuraya, F. S., Saadi, I., Lund, J. J., Turbe-Doan, A., Morton, C. C., & Maas, R. L. (2006). SUMO1 haploinsufficiency leads to cleft lip and palate. Science, 313(5794), 1751. doi:10.1126/science.1128406

Bishop, M. R., Diaz Perez, K. K., Sun, M., Ho, S., Chopra, P., Mukhopadhyay, N., . . . Leslie, E. J. (2020). Genome-wide Enrichment of De Novo Coding Mutations in Orofacial Cleft Trios. The American Journal of Human Genetics. doi:10.1016/j.ajhg.2020.05.018

Blanc, G., Font, B., Eichenberger, D., Moreau, C., Ricard-Blum, S., Hulmes, D. J., & Moali, C. (2007). Insights into how CUB domains can exert specific functions while sharing a common fold: conserved and specific features of the CUB1 domain contribute to the molecular basis of procollagen C-proteinase enhancer-1 activity. Journal of Biological Chemistry, 282(23), 16924-16933. doi:10.1074/jbc.M701610200

Camargo, M., Rivera, D., Moreno, L., Lidral, A. C., Harper, U., Jones, M., . . . Arcos-Burgos, M. (2012). GWAS reveals new recessive loci associated with non-syndromic facial clefting. European Journal of Medical Genetics, 55(10), 510-514. doi:10.1016/j.ejmg.2012.06.005

Chernus, J., Roosenboom, J., Ford, M., Lee, M. K., Emanuele, B., Anderton, J., . . . Marazita, M. L. (2018). GWAS reveals loci associated with velopharyngeal dysfunction. Scientific Reports, 8(1), 8470. doi:10.1038/s41598-018-26880-w

Conte, F., Oti, M., Dixon, J., Carels, C. E., Rubini, M., & Zhou, H. (2016). Systematic analysis of copy number variants of a large cohort of orofacial cleft patients identifies candidate genes for orofacial clefts. Hum Genet, 135(1), 41-59. doi:10.1007/s00439-015-1606-x

Cox, L. L., Cox, T. C., Moreno Uribe, L. M., Zhu, Y., Richter, C. T., Nidey, N., . . . Roscioli, T. (2018). Mutations in the Epithelial Cadherin-p120-Catenin Complex Cause Mendelian Non-Syndromic Cleft Lip with or without Cleft Palate. Am J Hum Genet, 102(6), 1143-1157. doi:10.1016/j.ajhg.2018.04.009

Czub, B., Shah, A. Z., Alfano, G., Kruczek, P. M., Chakarova, C. F., & Bhattacharya, S. S. (2016). TOPORS, a Dual E3 Ubiquitin and Sumo1 Ligase, Interacts with 26 S Protease Regulatory Subunit 4, Encoded by the PSMC1 Gene. PloS One, 11(2), e0148678. doi:10.1371/journal.pone.0148678

Davidson, T. B., Sanchez-Lara, P. A., Randolph, L. M., Krieger, M. D., Wu, S. Q., Panigrahy, A., . . . Erdreich-Epstein, A. (2012). Microdeletion del(22)(q12.2) encompassing the facial development-associated gene, MN1 (meningioma 1) in a child with Pierre-Robin sequence (including cleft palate) and neurofibromatosis 2 (NF2): a case report and review of the literature. BMC Medical Genetics, 13, 19. doi:10.1186/1471-2350-13-19

De Castro, M., & Restrepo, C. M. (2015). Genetics and genomic medicine in Colombia. Molecular genetics & genomic medicine, 3(2), 84-91. doi:10.1002/mgg3.139

Diaz Perez, K. K., Curtis, S. W., Sanchis-Juan, A., Zhao, X., Head, S. T., Ho, S., . . . Leslie, E. J. (2022). Rare variants found in clinical gene panels illuminate the genetic and allelic

architecture of orofacial clefting. medRxiv, 2022.2008.2015.22278795. doi:10.1101/2022.08.15.22278795

Elahi, M. M., Jackson, I. T., Elahi, O., Khan, A. H., Mubarak, F., Tariq, G. B., & Mitra, A. (2004). Epidemiology of cleft lip and cleft palate in Pakistan. Plastic and Reconstructive Surgery, 113(6), 1548-1555. doi:10.1097/01.prs.0000117184.77459.2b

Eshete, M. A., Liu, H., Li, M., Adeyemo, W. L., Gowans, L. J. J., Mossey, P. A., . . . Butali, A. (2018). Loss-of-Function GRHL3 Variants Detected in African Patients with Isolated Cleft Palate. J Dent Res, 97(1), 41-48. doi:10.1177/0022034517729819

Fareed, M., & Afzal, M. (2017). Genetics of consanguinity and inbreeding in health and disease. Annals Of Human Biology, 44(2), 99-107. doi:10.1080/03014460.2016.1265148

Ghoumid, J., Stichelbout, M., Jourdain, A. S., Frenois, F., Lejeune-Dumoulin, S., Alex-Cordier, M. P., . . . Manouvrier-Hanu, S. (2017). Blepharocheilodontic syndrome is a CDH1 pathway-related disorder due to mutations in CDH1 and CTNND1. Genet Med, 19(9), 1013-1021. doi:10.1038/gim.2017.11

Hamamy, H. (2012). Consanguineous marriages : Preconception consultation in primary health care settings. J Community Genet, 3(3), 185-192. doi:10.1007/s12687-011-0072-y

Hamamy, H., Antonarakis, S. E., Cavalli-Sforza, L. L., Temtamy, S., Romeo, G., Kate, L. P. T., . . . Bittles, A. H. (2011). Consanguineous marriages, pearls and perils: Geneva International Consanguinity Workshop Report. Genetics in Medicine, 13(9), 841-847. doi:10.1097/GIM.0b013e318217477f

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., & McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Research, 33(Database issue), D514-D517. doi:10.1093/nar/gki033

Holzinger, E. R., Li, Q., Parker, M. M., Hetmanski, J. B., Marazita, M. L., Mangold, E., . . . Bailey-Wilson, J. E. (2017). Analysis of sequence data to identify potential risk variants for oral clefts in multiplex families. Molecular genetics & genomic medicine, 5(5), 570-579. doi:https://doi.org/10.1002/mgg3.320

Hulick, P. J., Noonan, K. M., Kulkarni, S., Donovan, D. J., Listewnik, M., Ihm, C., . . . Weremowicz, S. (2009). Cytogenetic and array-CGH characterization of a complex de novo rearrangement involving duplication and deletion of 9p and clinical findings in a 4-month-old female. Cytogenet Genome Res, 126(3), 305-312. doi:10.1159/000251966

Jamilian, A., Nayeri, F., & Babayan, A. (2007). Incidence of cleft lip and palate in Tehran. Journal of the Indian Society of Pedodontics and Preventive Dentistry, 25(4), 174-176. doi:10.4103/0970-4388.37013

Jia, S., Zhou, J., & D'Souza, R. N. (2020). Pax9's dual roles in modulating Wnt signaling during murine palatogenesis. Developmental Dynamics, 249(10), 1274-1284. doi:https://doi.org/10.1002/dvdy.189

Kaplan, S., Pinar, G., Kaplan, B., Aslantekin, F., Karabulut, E., Ayar, B., & Dilmen, U. (2016). THE PREVALENCE OF CONSANGUINEOUS MARRIAGES AND AFFECTING FACTORS IN TURKEY: A NATIONAL SURVEY. Journal of Biosocial Science, 48(5), 616-630. doi:10.1017/s0021932016000055

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., . . . MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. bioRxiv, 531210. doi:10.1101/531210

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., . . . Maglott, D. R. (2018). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Research, 46(D1), D1062-d1067. doi:10.1093/nar/gkx1153

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . MacArthur, D. G. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature, 536(7616), 285-291. doi:10.1038/nature19057

Leslie, E. J., Liu, H., Carlson, J. C., Shaffer, J. R., Feingold, E., Wehby, G., . . . Marazita, M. L. (2016). A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. Am J Hum Genet, 98(4), 744-754. doi:10.1016/j.ajhg.2016.02.014

Leslie, E. J., O'Sullivan, J., Cunningham, M. L., Singh, A., Goudy, S. L., Ababneh, F., . . . Dixon, M. J. (2015). Expanding the genetic and phenotypic spectrum of popliteal pterygium disorders. Am J Med Genet A, 167a(3), 545-552. doi:10.1002/ajmg.a.36896

Lübke, T., Marquardt, T., Etzioni, A., Hartmann, E., von Figura, K., & Körner, C. (2001). Complementation cloning identifies CDG-IIc, a new type of congenital disorders of glycosylation, as a GDP-fucose transporter deficiency. Nature Genetics, 28(1), 73-76. doi:10.1038/ng0501-73

Mangold, E., Böhmer, A. C., Ishorst, N., Hoebel, A. K., Gültepe, P., Schuenke, H., . . . Ludwig, K. U. (2016). Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate. Am J Hum Genet, 98(4), 755-762. doi:10.1016/j.ajhg.2016.02.013

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics, 26(22), 2867-2873. doi:10.1093/bioinformatics/btq559

Mukhopadhyay, M., Shtrom, S., Rodriguez-Esteban, C., Chen, L., Tsukui, T., Gomer, L., . . . Westphal, H. (2001). Dickkopf1 is required for embryonic head induction and limb morphogenesis in the mouse. Developmental Cell, 1(3), 423-434. doi:10.1016/s1534-5807(01)00041-7

Peyrard-Janvid, M., Leslie, E. J., Kousa, Y. A., Smith, T. L., Dunnwald, M., Magnusson, M., . . . Schutte, B. C. (2014). Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. Am J Hum Genet, 94(1), 23-32. doi:10.1016/j.ajhg.2013.11.009

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet, 81(3), 559-575. doi:10.1086/519795

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2018). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Research, 47(D1), D886-D894. doi:10.1093/nar/gky1016

Rittler, M., Liascovich, R., López-Camelo, J., & Castilla, E. E. (2001). Parental consanguinity in specific types of congenital anomalies. American Journal of Medical Genetics, 102(1), 36-43. doi:10.1002/1096-8628(20010722)102:1<36::aid-ajmg1394>3.0.co;2-m

Rorick, N. K., Kinoshita, A., Weirather, J. L., Peyrard-Janvid, M., de Lima, R. L., Dunnwald, M., . . . Schutte, B. C. (2011). Genomic strategy identifies a missense mutation in WD-repeat domain 65 (WDR65) in an individual with Van der Woude syndrome. Am J Med Genet A, 155a(6), 1314-1321. doi:10.1002/ajmg.a.33980

Sabbagh, H. J., Hassan, M. H., Innes, N. P., Baik, A. A., & Mossey, P. A. (2014). Parental consanguinity and nonsyndromic orofacial clefts in children: a systematic review and meta-analyses. Cleft Palate Craniofac J, 51(5), 501-513. doi:10.1597/12-209

Sha, Z., Sha, L., Li, W., Dou, W., Shen, Y., Wu, L., & Xu, Q. (2015). Exome sequencing identifies SUCO mutations in mesial temporal lobe epilepsy. Neuroscience Letters, 591, 149-154. doi:https://doi.org/10.1016/j.neulet.2015.02.009

Smyth, I., Du, X., Taylor, M. S., Justice, M. J., Beutler, B., & Jackson, I. J. (2004). The extracellular matrix gene Frem1 is essential for the normal adhesion of the embryonic epidermis. Proceedings of the National Academy of Sciences of the United States of America, 101(37), 13560-13565. doi:10.1073/pnas.0402760101

Sohaskey, M. L., Jiang, Y., Zhao, J. J., Mohr, A., Roemer, F., & Harland, R. M. (2010). Osteopotentia regulates osteoblast maturation, bone formation, and skeletal integrity in mice. Journal of Cell Biology, 189(3), 511-525. doi:10.1083/jcb.201003006

Stoll, C., Alembik, Y., Dott, B., & Roth, M. P. (1991). Epidemiological and genetic study in 207 cases of oral clefts in Alsace, north-eastern France. J Med Genet, 28(5), 325-329. doi:10.1136/jmg.28.5.325

Stoll, C., Alembik, Y., Roth, M. P., & Dott, B. (1999). Parental consanguinity as a cause for increased incidence of births defects in a study of 238,942 consecutive births. Annales de Génétique, 42(3), 133-139.

Strong, A., Simone, L., Krentz, A., Vaccaro, C., Watson, D., Ron, H., . . . Hakonarson, H. (2021). Expanding the genetic landscape of oral-facial-digital syndrome with two novel genes. Am J Med Genet A, 185(8), 2409-2416. doi:10.1002/ajmg.a.62337

Tayebi, N., Yazdani, K., & Naghshin, N. (2010). The prevalence of congenital malformations and its correlation with consanguineous marriages. Oman Medical Journal, 25(1), 37-40. doi:10.5001/omj.2010.9

Woods, C. G., Cox, J., Springell, K., Hampshire, D. J., Mohamed, M. D., McKibbin, M., . . . Inglehearn, C. F. (2006). Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. Am J Hum Genet, 78(5), 889-896. doi:10.1086/503875

Yankee, T. N., Wilderman, A., Winchester, E. W., VanOudenhove, J., & Cotney, J. (2022). Integrative analysis of transcriptomics in human craniofacial development reveals novel candidate disease genes. bioRxiv, 2022.2002.2028.482338. doi:10.1101/2022.02.28.482338

CHAPTER IV. Rare variants found in clinical gene panels illuminate the genetic and allelic architecture of orofacial clefting

Kimberly K. Diaz Perez, Sarah W. Curtis, Alba Sanchis-Juan, Xuefang Zhao, S. Taylor Head, Samantha Ho, Bridget Carter, Toby McHenry, Madison R. Bishop, Luz Consuelo Valencia-Ramirez, Claudia Restrepo, Jacqueline T. Hecht, Lina Moreno Uribe, George Wehby, Seth M. Weinberg, Terri H. Beaty, Jeffrey C. Murray, Eleanor Feingold, Mary L. Marazita, David J. Cutler, Michael P. Epstein, Harrison Brand, Elizabeth J. Leslie

Adapted from manuscript submitted to Genetics in Medicine:
Diaz Perez KK, Curtis S, Sanchis-Juan A, Head T, Ho S, Carter B, McHenry T, Bishop M, Valencia-Ramirez LC, Restrepo C, Moreno L, Wehby G, Beaty T, Murray J, Feingold E, Marazita M, Cutler D, Epstein M, Brand H, Leslie E. 2022. "Rare variants found in clinical gene panels illuminate the genetic and allelic architecture of orofacial clefting." doi: https://doi.org/10.1101/2022.08.15.22278795. medRXiv. Submitted.

Introduction

Orofacial clefts (OFCs) are etiologically heterogeneous structural birth defects (Leslie, 2022), including cleft lip (CL), cleft palate (CP), and cleft lip and palate (CLP). Genetic factors, such as point mutations, copy number variants, and chromosomal abnormalities, contribute to the etiology of OFCs, especially in Mendelian OFC syndromes, which contain other structural anomalies, cognitive anomalies, or intellectual disabilities. Hundreds of rare Mendelian syndromes involving OFCs have been described but most OFC cases occur as apparently isolated birth defects (often termed non-syndromic). These are considered etiologically complex disorders with genetic and environmental risk factors. It is possible, however, that both syndromic and isolated cases have an etiology caused by genetics, environment, or the combined effect of both. The heterogeneity of OFCs is further compounded by phenotypic heterogeneity, incomplete penetrance, and variable expressivity, making clinical diagnostics challenging. Recurrence risk estimates vary with an approximate sibling recurrence of ~4% (Grosen *et al*., 2010), which is much lower than the sibling risk for an autosomal dominant disorder or the empirical risk of being affected with an incompletely penetrant disorder, but is also significantly higher than expected if the risk were driven by *de novo* variants alone. It is therefore important to determine the cause of OFCs as it could inform recurrence risk estimates and the approaches for genetic counseling and clinical management.

Molecular diagnoses from genetic testing is an alternative to diagnosis based on the phenotype alone. However, genetic testing is not common for most OFC cases, which are isolated cases without any family history. Furthermore, existing commercial clinical testing panels are highly variable in their content, leading to potentially missed diagnoses. Multiple studies have explored the use of sequencing to improve diagnostics for OFCs. We previously

investigated the proportion of isolated OFC cases attributable to variants in *IRF6*; however, our estimate of 0.2-0.4% was too low to recommend broad screening of this gene (Leslie, E. J. *et al*., 2016). The use of exome sequencing (WES) and genome sequencing (WGS) in OFCs has recently increased. Basha *et al*. 2018 tested the diagnostic rate from WES in 46 multiplex OFC families, finding 10% of cases carried 'likely pathogenic' (LP) variants, primarily in genes causing autosomal dominant OFC syndromes. However, WES in large cohorts has not been performed so diagnostic yield estimates in OFCs are still uncertain.

We aimed to estimate the diagnostic yield of 418 genes associated with OFCs using WGS in 841 OFC cases and 294 controls. We previously investigated *de novo* variants in 756 OFC trios from this same cohort and found 6% of sequenced trios had a *de novo* variant in genes broadly associated with OFCs (Bishop *et al*., 2020). Two genes (*IRF6* and *TFAP2A*) mutated in OFC syndromes were individually associated with OFCs, raising the question of the clinical impact of *de novo* variants and other types of variants in similar genes. We, therefore, developed the present study to utilize WGS to fully characterize the clinical impact of variants in OFC cases by analyzing de novo and transmitted single nucleotide and structural variants.

Methods

Study Population

The case sample consisted of 841 total OFC cases (765 case-parent trios, 60 parent-child dyads, and 16 singletons) sequenced through the Gabriella Miller Kids First (GMKF) Pediatric Research Program. The case sample was sequenced in three cohorts based on recruitment site/ancestry: 1) "Europeans" from the United States, Argentina, Turkey, Hungary, and Spain; 2) "Latinos" from Colombia; and 3) "Asians" from Taiwan (Table S4-1). Participant recruitment

occurred over many years using different research protocols, but each protocol generally included a physical exam to exclude individuals with major anomalies or known intellectual disability indicative of an OFC syndrome. This cohort is therefore enriched for isolated OFCs and depleted of multiple congenital anomalies and severe manifestations of syndromes. The case sample includes probands with cleft lip only (e.g., CL, cleft lip and cleft alveolus) (107 cases), cleft lip and cleft secondary palate (CLP; 660 cases), and cleft secondary palate only (CP; 74 cases). The cases were primarily male (56% CL, 65% CLP, 53% CP) reflecting the large proportion of the cases having CLP where males are overrepresented. The 756 trios were analyzed previously for *de novo* variants only (Bishop *et al*., 2020).

A total of 621 probands were considered "simplex" as they reported no family history, defined as not reporting any affected relative within the 3rd degree. 220 probands reported having at least one affected relative (1st, 2nd, or 3rd degree) were classified as "multiplex" (Table S4-2); this included 63 probands with at least one affected parent. All probands were confirmed to be unrelated with kinship calculations using KING.

The control sample was comprised of 294 child-parent trios from the 1000 Genomes Project (1KGP) (Auton *et al*., 2015). Because some 1KGP samples are derived from cell lines, these samples have excessive numbers of *de novo* variants acquired through multiple passages and are not comparable to the pattern of variation in the GMKF samples (Ng *et al*., 2021). Therefore, we selected 1KGP trios through a quality control process (described below) to have approximately the same amount of total variation including *de novo* variant rates (all selected 1KGP trios had fewer than 138 *de novo* events per trio) as the case cohorts. This cohort included trios from multiple ancestries: 84 African, 116 US and European, 52 East Asian, and 42 South Asian trios. Although phenotype information is unavailable for 1KGP, we would expect at most

1 OFC in the 294 trios (882 total individuals) based on the prevalence rate of OFCs at 1 in 1000 individuals worldwide. Thus, 1KGP can serve as a comparison group as it is unlikely to have LP variants influencing risk to OFCs.

Sequencing and Quality Control

Sequencing and variant calling of the OFC cohort was described in Bishop *et al*. (2020). Sequencing of the control cohort was described in Byrska-Bishop *et al*. (2021). The same quality control procedures were performed on case and control VCF files. We retained genotype calls with a genotype quality ≥20, read depth ≥10, and biallelic variants passing VSQR with a Quality Normalized by Depth (QD) score >4 using VCFTools (v0.1.13) and BCFtools (v1.9). Variants with >2 Mendelian errors, >5% missingness, or deviations from Hardy-Weinberg equilibrium ($p<10\text{-}7$) in unaffected samples were dropped. For *de novo* variants, we required an allele balance between 0.3 – 0.7 in each proband and < 0.05 in both parents.

Selection of Gene List

We created a comprehensive set of 418 genes (Table S4-3, Figure S4-1A) to prioritize variants possibly associated with OFCs from four sources (all downloaded September 4th, 2020): 1) the National Health Service (NHS) Genomic Medicine Service cleft panel (v2.2), an expert-curated list of genes for familial cleft lip and/or cleft palate (CL/P), isolated and syndromic clefting; 2) the PreventionGenetics CL/P clinical genetic testing panel; 3) Clinical synopses/genes from the Online Mendelian Inheritance in Man (OMIM) that included OFCs with a known inheritance and molecular basis. OMIM clinical synopses search terms included: “cleft lip,” “cleft palate,” “oral cleft,” “orofacial cleft,” and “cleft lip and/or palate;” 4) a manually

curated list from recent OFC genetic studies. The NHS panel included an evidence level indicator corresponding to expert consensus for genes on the panel: “green” for genes of known clinical utility and scientific validity, “amber” for moderate evidence levels, and “red” indicating little evidence (Martin *et al*., 2019). We classified genes based on the mechanism by which variants lead to OFC phenotypes and hereafter refer to these genes as autosomal dominant (AD), autosomal recessive (AR), or X-linked (XL). Genes in which variants have been described as acting in dominant and recessive manners or unspecified modes of inheritance were considered in both AD and AR analyses. Average read depth for each gene was comparable between cases and controls and between all case populations (Figure S4-1B). In our previous work, we analyzed 336 genes associated with OFCs, which included genes from OMIM and those nominated by linkage, candidate gene, and association studies. The current list of 418 genes includes 200 genes not analyzed previously, the majority of which came from NHS and OMIM. There were 116 genes on the Bishop *et al*. (2020) gene list absent from this analysis, most of which were GWAS genes that lacked the necessary support to be included in a clinical gene panel.

SNV and Indel Annotation and Variant Filtering

Variants were annotated using ANNOVAR (version 201707) and Variant Effect Predictor (VEP, release 102, 103, 106). Protein-altering variants were extracted for the 418 genes. Variants were filtered using a maximum allele frequency (AF) threshold of 0.1% for variants in autosomal dominant genes and 0.5% for variants in autosomal recessive genes using gnomAD (v2 and v3) (Karczewski *et al*., 2020) and ExAC (v0.3) (Karczewski *et al*., 2017), and a cohort allele count (AC) ≤ 10. Variant-level annotations used in the prioritization and

interpretation included nine i*n silico* pathogenicity predictions (e.g., SIFT (Ng & Henikoff, 2003), PolyPhen (Adzhubei *et al*., 2010), MutationTaster (Schwarz *et al*., 2014)), CADD (Rentzsch *et al*., 2018) scores, variant pathogenicity classifications from ClinVar (Landrum *et al*., 2018), and constrained regions within genes (Havrilla *et al*., 2019).

Structural Variants (SVs) Identification and Filtering

We detected SVs in the OFC cohort with the GATK-SV discovery pipeline as previously described (Collins *et al*., 2020). GATK-SV (https://github.com/broadinstitute/gatk-sv) relies on an ensemble approach that harmonizes SV detection from multiple tools, including Manta (Chen *et al*., 2016) and Gatk-gCNV (McKenna *et al*., 2010), followed by machine learning to remove likely false positive events and then performs joint genotyping and refined variant resolution. The derived VCF file was annotated by svtk (Werling *et al*., 2018) to predict the functional impact of SVs and compare AF against gnomAD SV (v2.1) (Collins *et al*., 2020). We obtained SVs overlapping the 418 genes and filtered the SVs by AF (OFC cohort AF ≤ 0.03 and gnomAD SV AF ≤ 0.01). SVs overlapping recurrent genomic disorder regions were investigated independently and we reported the gene(s) overlapping our gene list from those regions. Further inheritance-specific genotype and frequency filters were applied to identify *de novo* (gnomAD SV AF ≤ 0.001, cohort AC ≤ 10 and cohort sample count ≤ 5), homozygous (homozygous AC ≤ 10 and absent in unaffected individuals in the cohort), compound heterozygous and X-linked recessive SVs. We also considered SVs inherited from unaffected parents if the cohort AC was ≤ 10, of which ≤ 5 were unaffected individuals. Candidate SVs (Table S4-4) were manually reviewed and visually inspected in normalized read-depth plots using Integrative Genomics Viewer (Robinson *et al*., 2011).

Classification into the Tier System

Rare SNVs and indels located within the 418 genes were classified into a ranked tier system designed to minimize the number of variants undergoing manual American College of Medical Genetics and Genomics (ACMG) review while retaining as many potential LP variants as possible. Each tier was based on gene or variant annotation criteria, including variant type and gene constraint (Figure S4-2A). Tiers were ranked based on qualitative assessments of their likelihood to contain LP variants. Assessments were made by KDP, MRB, and EJL, and the final tiers were formed from a consensus of these assessments.

After sorting variants into tiers, we identified a cutoff point above which variants would be manually reviewed according to ACMG criteria. To determine the cutoff point, we extracted variants in ClinVar from 418 genes classified as either pathogenic ('likely pathogenic' or 'pathogenic') or benign ('likely benign' or 'benign'). We sorted the 526 pathogenic and 274 benign variants into tiers (Figure S4-2B), and identified Tier 1B as a point where 95% of pathogenic variants but only 49% of benign variants would be retained for review.

ACMG Variant Classification

All variants meeting the Tier 1B threshold on this tier system were assessed using ACMG criteria blinded to case-control status (Richards *et al*., 2015). We considered variants with "damaging" pathogenicity predictions from ≥ 5 out of 9 algorithms to meet the PP3 criteria, while variants with ≥ 5 out of 9 "tolerant" predictions met the BP4 criteria (Table S4-5). For criteria based on AF alone (PM2, BS1), we used the maximum AF observed in any population across gnomAD v2 genomes and exomes, gnomAD v3 genomes, and ExAC exomes. Variants

with an AF < 0.001% met criteria PM2, and variants with AF ≥ 0.005% (heterozygous) and AF ≥ 0.2% (homozygous) met BS1. We estimated the maximum credible AF for a variant considering an OFC prevalence of 1 in 1,000, 5% of allelic heterogeneity, 100% of genetic heterogeneity, and 50% penetrance (Whiffin *et al*., 2017). All variants classified as 'pathogenic' or 'likely pathogenic' were counted towards the diagnostic yield calculation and are referred to as 'likely pathogenic' (LP) throughout the manuscript.

Statistical Analysis

Statistical tests were performed to calculate differences between groups using two-sided chi-square and Fisher exact tests, which were conducted using R (version 3.6.3). We performed 10,000 permutations for the chi-square tests comparing cases and controls (overall, by cleft subtype, population, and sex) to adjust for multiple hypothesis testing under the null hypothesis of no association between the number of individuals with LP variants and case-control status. The significance level was set at $P < 0.05$ for these tests. Odds ratios and 95% confidence intervals were estimated through chi-square tests in R.

We tested gene-based associations in genes with 'variants of uncertain significance' (VUS) using the Optimal Sequence Kernel Association test (SKAT-O), which unites the Sequence Kernel Association test (SKAT) and the burden test to maximize statistical power while allowing for variants of opposite effects (Lee *et al*., 2012). Data were converted to binary PLINK files and imported into the SKAT package (version 2.0.1) (Seunggeun Lee, 2020) in R (version 3.6.3). First, we performed SKAT-O tests for 139 genes with more than one VUS or LP variant. We then excluded "solved" cases and controls with LP variants and conducted SKAT-O

tests for 129 genes with more than one VUS in the remaining samples. We used a Bonferroni correction to adjust for multiple testing.

Results

We identified 2,549 SNVs, small indels, and SVs from 841 OFC cases and 294 controls. After sorting variants into tiers designed to prioritize variants, we narrowed our list to 1,483 variants for manual review under ACMG criteria (Figure S4-2C). On average, we reviewed 1.33 variants per case and 1.24 variants per control ($p=0.07$).

After ACMG review, 79 variants (5.33%) were classified as 'likely pathogenic' (LP) (Table S4-6). The LP variants were dominated by those presumed to be loss-of-function (LoF): 46.8% were stop-gain, frameshifting indels, and canonical splice site variants; 15% were SVs. Overall, 9.04% of cases and 1.02% of controls had LP variants ($p<0.0001$, Figure 4-1A). Stratifying our gene list by the mode of inheritance, we found LP variants were almost exclusively in autosomal dominant genes (8.80% of cases vs. 1.02% of controls; $p<0.0001$). Consistent with previous analysis of an excess of *de novo* variants in clinically relevant genes among OFC cases (Bishop *et al*., 2020), 3.69% of cases (vs. none in controls; $p=0.0008$) had a *de novo* LP variant. Notably, we did not identify any LP homozygous or compound heterozygous variants in autosomal recessive genes. This lack of signal was unexpected because a subset of the trios came from consanguineous families from Turkey and Colombia. Similarly, there was a limited contribution from X-linked genes. Only two individuals (0.24% of cases) had LP variants on the X chromosome: a hemizygous male with a LoF variant in *PHF8* inherited from his unaffected mother and a heterozygous female with a *de novo* in-frame deletion in *FLNA*.

Epidemiology and association studies suggest some differences in the genetic architecture of specific OFC subtypes (Carlson *et al*., 2019; Sivertsen *et al*., 2008). Therefore, we stratified the case cohort to test for differences in diagnostic yield across CL, CLP, and CP subtypes (Figure 4-1B). Among CLP cases, which comprise 78% of the OFC cohort, 9.09% had a LP variant (60 out of 660 total CLP cases) ($p=0.0003$ vs. controls). The diagnostic yield was much higher among CP cases (13 out of 74 total CP cases), where 17.6% had a LP variant ($p<0.0001$ vs. controls and $p=0.035$ vs. CLP). Equally striking was the difference between CL and CLP, which have historically been viewed as a variation in severity of the same disorder and are commonly analyzed together. Only 2.80% of CL cases (3 out of 107 total CL cases) had a LP variant, which was not significantly different than controls ($p=0.353$) and only nominally different from CLP ($p=0.045$). These data suggest the differences in genetic architecture between CL and CLP seen in GWAS studies (Harville *et al*., 2005; Marazita *et al*., 2009) may extend to rare variants.

There are characteristic sex biases in OFCs where CP occurs twice as frequently in females than males, and CL/P occurs twice as frequently in males than females (Marazita, 2012). We considered whether these sex biases were also reflected in the yields. Although the less frequently affected sex had consistently higher diagnostic yields within each subtype, none was statistically significant (Figure 4-1C). These results could be consistent with a "protective effect" model; when there are disease prevalence differences between sexes, affected individuals among the less commonly affected sex have, on average, greater enrichment for disease-causing alleles or alleles of larger effect than members of the more commonly affected sex. This would also be consistent with the observation that sex biases are not as commonly observed in Mendelian OFC syndromes. We also observed small (but non-significant) differences in yield when stratifying by

population by cleft type (Figure S4-3). Although this is loosely correlated with OFC prevalence rates (Mossey, 2007), it is more likely that these differences are due to disparities in the representation of these populations in reference databases that impact the filtering of variants based on AF.

The 76 LP variants in OFC cases were found across 39 genes, constituting 9.33% of the gene list (Figure 4-2). Sixteen genes had multiple variants, and nine of these had at least three LP variants in cases. These nine genes: *CTNND1* (6 cases), *ARHGAP29* (5 cases), *COL2A1* (5 cases), *IRF6* (5 cases), *TFAP2A* (5 cases), *CDH1* (4 cases), *CHD7* (3 cases), *PDGFC* (3 cases), and *TBX1* (3 cases, all 22q11.2 deletions) accounted for 4.64% of OFC cases alone. Eight of these genes (and 35 out of 39 genes with LP variants) were genes in which variants cause disease in an autosomal dominant manner. Of 163 genes associated with autosomal dominant disease, 21.5% had at least one LP variant, demonstrating the genetic heterogeneity of OFCs. Although previous studies have found LP variants in two or more disease loci in the same individual (Posey *et al*., 2017), we did not identify cases with more than one LP variant.

Previous OFC studies report incomplete penetrance for several genes, including *CTNND1* and *TP63* (Basha *et al*., 2018; Cox *et al*., 2018), but few have studied large datasets drawn from both simplex and multiplex families, allowing us to weigh the contribution of *de novo* and transmitted variants and estimate penetrance for autosomal dominant variants. A total of 220 probands were from multiplex families, defined as having at least one other affected relative (up to the 3rd-degree). There was no difference in yield between individuals from multiplex and simplex families (11.8% multiplex vs. 7.73% simplex, p=0.089). However, there were notable differences in the types of variants identified (Figure S4-4). Twenty of the 26 LP variants in individuals from multiplex families were transmitted (Figure S4-5); the rest were *de novo*

(Figure S4-6). In contrast, 52.1% (25/48) of LP variants in simplex families were *de novo*, which were confirmed by visual inspection of aligned reads (Figure S4-6). In three of the six families, one parent was affected, and we cannot exclude the possibility of mosaicism in other tissues. But it is also possible these variants are not the only variants conferring risk for OFCs.

Among transmitted variants in multiplex families, we asked how often the variant was transmitted by the parent with a personal or family history of OFC. We found 82.4% (14/17) of variants were transmitted by the parent with a family history with no differences by maternal vs. paternal history; 64% (9/14) of these variants were transmitted by an affected parent. We can therefore estimate the global penetrance to be 60% among multiplex families. If we count all transmitting parents, including unaffected parents from simplex families, the penetrance of transmitted variants falls to 25% (9/36). Interestingly, most of these variants are predicted to be loss-of-function, and impacted genes included *ARHGAP29*, *CTNND1*, and *TP63*, which are considered haploinsufficient with reduced penetrance (Khandelwal *et al*., 2019; Savastano *et al*., 2017).

The majority (61.8%) of classified variants were variants of uncertain significance (VUS). Overall, we found a significant enrichment of VUS among OFC cases compared to controls (56.6% cases vs. 46.6% controls, p=0.004). This result was consistent across populations but not OFC subtypes (Table S4-7). VUS were not clustered in cases with LP variants or in cases without such variants as removing "solved" cases/controls resulted in a similar enrichment: 56.1% of 765 cases had at least one VUS vs. 46.7% of 291 controls (p=8.03 x 10-3).

One possible hypothesis to explain the excess of VUS among cases is that there are cryptic LP variants among this set of VUS. For both cases and controls, VUS were

overwhelmingly missense variants, which is not surprising given the challenges of interpreting missense variation. We expected VUS among cases would have greater "damaging" prediction scores, but there was no difference in the distribution of pathogenicity predictions aggregated under nine different algorithms (see Table S4-5, Figure S4-7).

We next hypothesized VUS would be more likely to occur in genes with LP variants. Collectively, VUS were similarly enriched among genes with LP variants (OR 1.63, p=0.008) as they were among genes without LP variants (OR 1.36, p=0.033) (Figure 4-3A). To further parse which sets of genes were contributing to the VUS signal, we used the evidence level of genes on the NHS panel, corresponding to three levels of support for the genes' association with OFCs reviewed by an expert panel ("green" for high evidence, "amber" for moderate evidence, and "red" for low evidence). The enrichment of VUS was strongest among 69 "green" genes (OR 2, p=1.36 x 10-4) (Figure 4-3A). We then performed SKAT-O tests for autosomal genes with LP variants and/or VUS to pinpoint individual genes with significant VUS contributions. Although no gene reached formal significance due to an unbalanced sample size favoring cases, *PRICKLE1* (MIM: 608500) was nominally significant with an odds ratio indicating an increased risk for OFC (Figure S4-8). Furthermore, many genes with multiple LP variants had an increase in odds ratio from the VUS (Figure 4-3B).

Lastly, we asked which gene list sources performed the best. The yield for individual lists was between 6.18% and 7.61% but no list had a statistically different yield from the others (Table S4-8). The manually curated, PreventionGenetics, and NHS gene lists had the highest proportion of genes with LP variants, which is not surprising given that 82% of the 418 genes appear on at least one of those three lists. Most autosomal dominant genes with LP variants (27/35) were shared among at least these three sources. In contrast, the OMIM list had the lowest

percentage (10.3%) of LP variants and a yield of 6.18% for all OFCs. The OMIM list performed better than the PreventionGenetics or NHS lists for CP (16.2% vs. 14.8%), but these differences were not statistically different (Table S4-8).

Discussion

Genetic diagnostics are currently performed on OFC cases with an OFC family history consistent with Mendelian inheritance patterns or individuals with syndromic presentations. Consequently, diagnostic testing is conducted in only a small fraction of cases, creating a potential clinical diagnostic gap. Previous OFC studies estimated a diagnostic yield of ~10% using ES in a small set of 46 multiplex families (Basha *et al*., 2018). In this sample of 841 cases from multiple populations and different family structures, we also estimated a yield of ~10%, confirming previous studies (Basha *et al*., 2018; Cox *et al*., 2018). Our study provided confirmation and moves beyond replication in several substantive ways. First, we showed that diagnostic yield varied significantly by OFC subtype. We observed an almost 20% diagnostic yield for CP but a nearly 7-fold lower yield in CL. The CP yield estimate was comparable to a recent report of 30 isolated CP cases, where 17% of cases had LP variants (Lace *et al*., 2022). These findings could be clinically useful as they may suggest a reexamination of whether routine clinical testing of CP cases is warranted as standard of care. It is also scientifically useful as it reinforces the etiologic heterogeneity in OFC subtypes observed from studies of common variants (Carlson *et al*., 2019; Carlson *et al*., 2017; Curtis *et al*., 2021; Mukhopadhyay *et al*., 2022) and extends it to rare variants. It also motivates future CL-specific research which will be necessary to perform genetic testing clinically for OFCs in general. Second, we show that although some risk is clustered in nine genes, OFCs are highly heterogeneous within and

between subtypes. Finally, we found that VUS, which constitute most variants categorized in this study and in clinical tests, are enriched in cases and subsets of these VUS, such as those in high-confidence OFC genes, are likely to contribute to the OFC phenotype.

Our estimated yield (~10%) is similar to those from exome or genome-based studies of other pediatric conditions including congenital heart disease (12.7%) (Mone *et al*., 2021) and autism spectrum disorder (7.5%) (Lowther *et al*., 2020). A major determining factor in these studies is their approach and cohort ascertainment. For example, Lowther and colleagues estimated the yield from sequencing in autism spectrum disorder at 7.5% using a panel of 907 neurodevelopmental genes (Lowther *et al*., 2020). They found a similar yield (12%) for a heterogeneous group of fetal structural anomalies in 2,535 genes but noted cases had been pre-screened by karyotype and chromosomal microarray analyses, lowering their diagnostic yield. Targeted investigations such as ours may favor specificity (but lose sensitivity) since the overall yield will be lower than exome-wide studies or including other first-tier techniques such as karyotyping. Further, diagnostic yields are typically higher in syndromic cases or those with multiple congenital anomalies (Hathaway *et al*., 2021; Scott *et al*., 2022). We note that our study was not population-based so the true diagnostic yield remains to be determined.

We identified LP variants in 39 genes. Although several genes had multiple variants, few individual variants recurred, underscoring the extensive allelic and genetic heterogeneity of OFCs. We observed patterns of variation across OFC subtypes consistent with the literature. For example, we found *COL2A1* variants exclusively among CP cases (Hoornaert *et al*., 2010), *TFAP2A* variants were found exclusively among CLP cases (Reiber *et al*., 2010), and *IRF6* variants were found in either CP or CLP cases (Kondo *et al*., 2002). Interestingly, despite a strong genotype-phenotype correlation between *GRHL3* (MIM: 608317) and CP (Leslie,

Elizabeth J. *et al*., 2016; Mangold *et al*., 2016), the only LP variant identified in *GRHL3* was in a CL case. Due to differences in sample sizes for each OFC subtype, we were unable to quantitatively analyze genotype-phenotype correlations for each gene, so these remain anecdotal observations requiring follow-up in larger datasets.

Interpreting VUS is a considerable challenge. Nearly 62% of variants were classified as VUS, most of which were missense variants. The effect of missense variants is often difficult to interpret without independent functional evidence, especially for genes with high allelic heterogeneity. There were multiple genes (e.g., *PRICKLE1*) for which VUS were identified in cases while zero VUS were identified in controls. *PRICKLE1* was previously evaluated through family-based association studies and showed evidence of association with OFCs (Yang *et al*., 2014). Here, we found a similar, but nominally significant, effect on OFC risk. Therefore, it is likely these datasets are underpowered to detect genes with a burden of rare variants and the top-ranked genes should be considered for further analysis pending functional testing to sort out the effect of identified variants.

It is important to note this study was conducted on a cross-sectional cohort from multiple recruitment protocols intended for research, and it is not representative of all OFC cases that may be referred from craniofacial clinics. Clinical diagnostics and differential diagnoses are aided by detailed phenotyping and collection of family histories, but data availability is limited for specific populations in this cohort. Although those with multiple congenital anomalies and significant developmental delays were likely excluded and such individuals should represent a minority of the dataset, the recruitment timing, varying skills of the clinical and research teams, and different recruitment goals make this a highly heterogeneous cohort with incomplete phenotypic data needed for the clinical setting. However, many OFC syndromes show

incomplete penetrance and variable expressivity, which can complicate a diagnosis based on phenotype alone even when detailed phenotyping is available. In this study, we estimated the penetrance of OFCs for transmitted alleles but could not estimate the extent of variable expressivity of other phenotypic features. Moreover, these estimates represent global penetrance, not gene-level, which requires additional investigation in larger cohorts. Nonetheless, the low penetrance of LP variants was striking, as many variants were predicted to be 'loss of function'. The ideal cohort to fully evaluate penetrance and expressivity would be a prospectively recruited, deeply phenotyped, and sequenced cohort of sequential cases, which are difficult and costly to assemble. Lastly, we only reviewed some variants to limit variant "noise"; however, some variants that did not meet our prioritization criteria could be pathogenic.

Sequencing studies such as this one and those investigating the functional consequences of variants in model systems are necessary to advance research translation to clinical practice. Although panel-based clinical tests will inevitably be replaced by GS, our results offer insight into the breadth of genes that may be found from clinical sequencing. Our results suggest there is value in genetic testing of CP cases but limited utility in using panels for CL cases. It remains to be seen if the yield differences are due to differences in architecture (i.e., fewer highly penetrant variants in CL) or differences in panel content that are insufficient for CL. Besides the potential clinical applications, we highlight the critical need for high-throughput validation to quantitatively distinguish the effects of individual rare variants. Future work in this area should allow for improved variant interpretation in a clinical setting, a greater understanding of the genes influencing craniofacial birth defects, and help explain the variable penetrance observed in this study.

Figures

Figure 4-1: Diagnostic Yield of 418 OFC Genes. (A) The percentage of cases (red) and controls (gray) with at least one 'likely pathogenic' variant by mode of inheritance. (B) The percentage of individuals with 'likely pathogenic' variants in controls (gray) and in cases by OFC subtype: cleft lip (CL, red), cleft lip and palate (CLP, purple), and cleft palate (CP, blue). (C) Percentage of cases with a 'likely pathogenic' variant stratified by cleft type and proband sex (female (pink) and male (blue)). Yields were not significantly different between males and females in any cleft subtype. P-values were calculated using chi-square tests with 10,000 permutations in R.

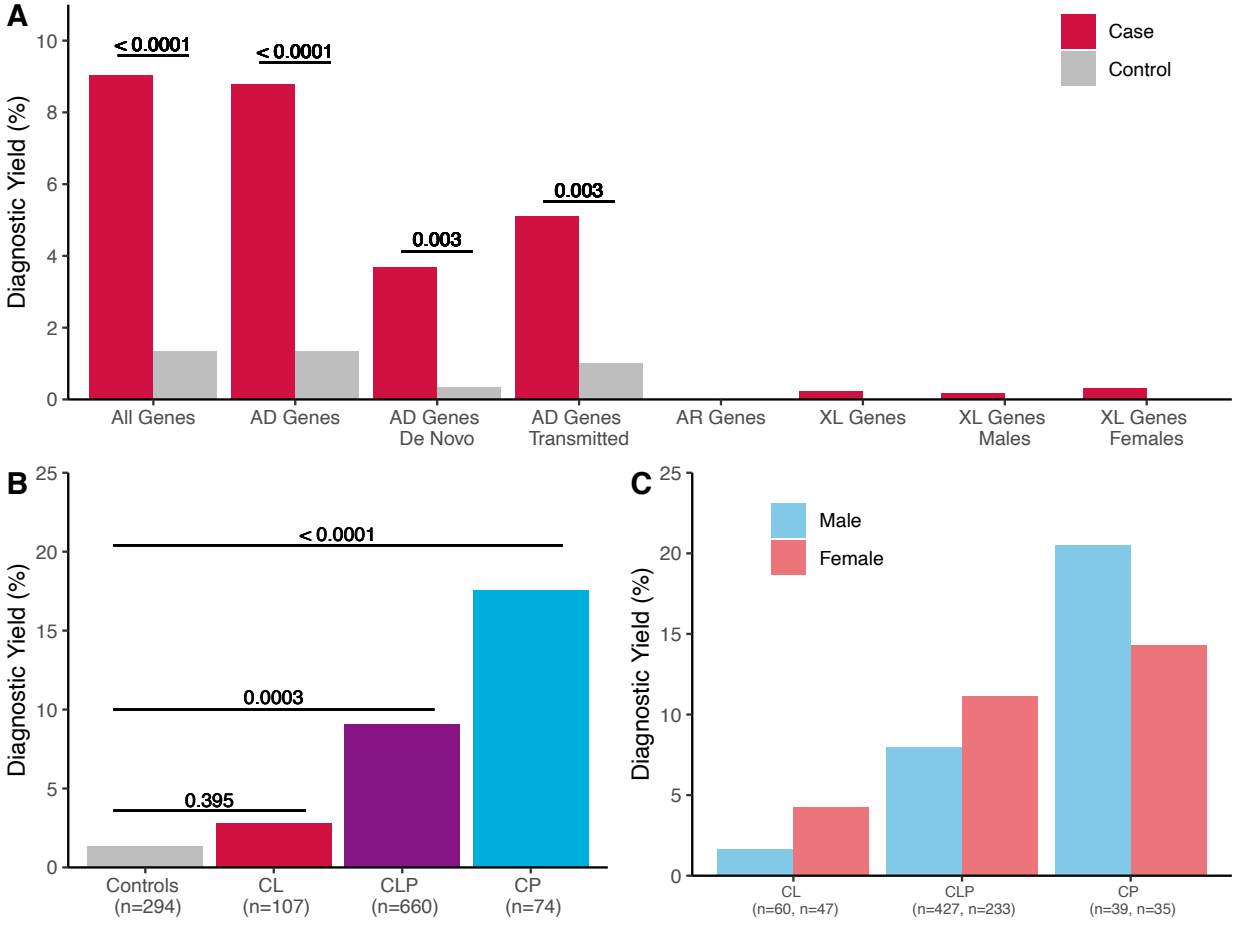


Figure 4-2: Genes with Likely Pathogenic Variants in Cases. The count of 'likely pathogenic' variants in cases in each of the 39 genes with 'likely pathogenic' variants. Variants are in the same order in both panels and are colored in the left panel based on the mode of inheritance: *de novo* (dark purple), transmitted (light purple), or were unknown (gray), and in the right panel based on variant consequence: missense (red) variants, predicted loss-of-function (pLoF, dark red), and structural variants (blue). The center panel shows a filled circle for each source list that each gene was on: GP (curated gene panel), NHS (NHS PanelApp list), OMIM, and PG (PreventionGenetics panel).

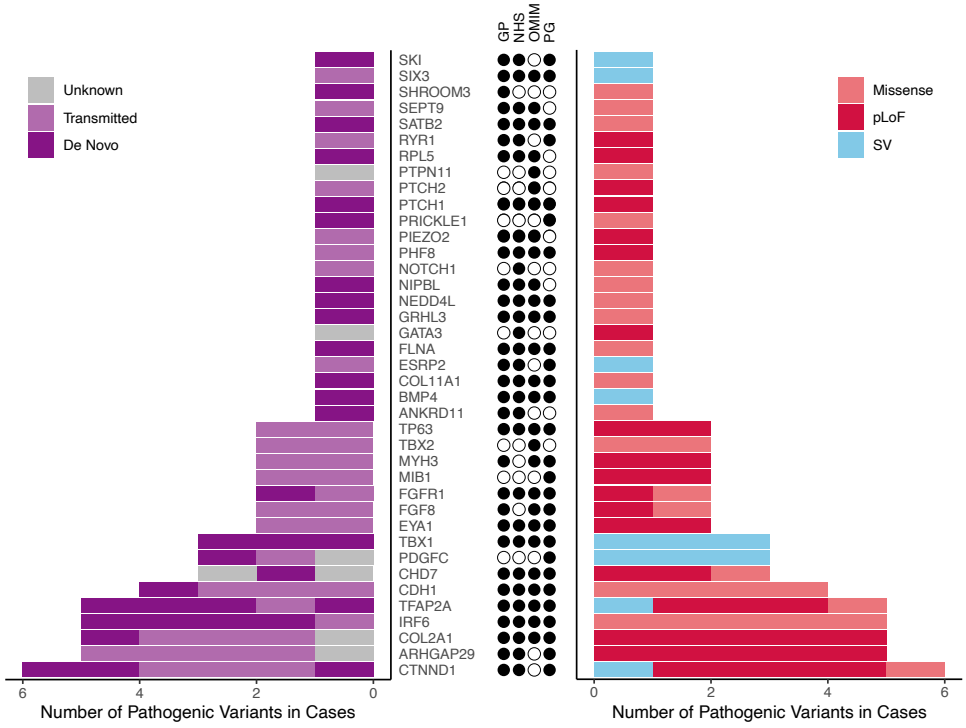
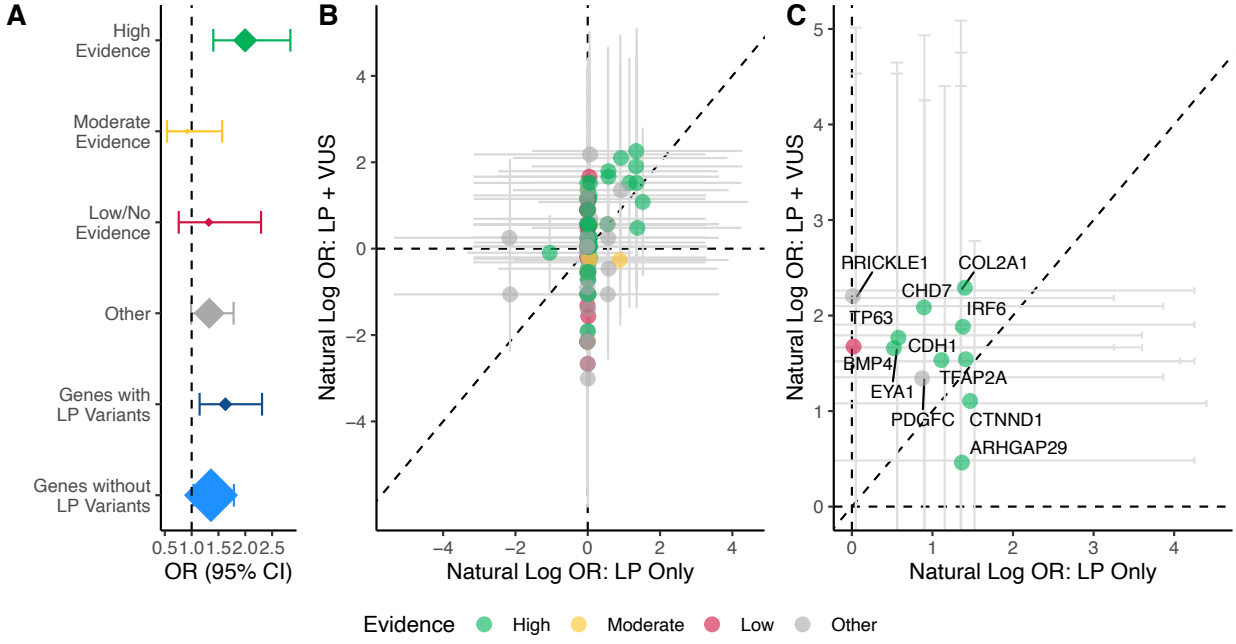


Figure 4-3: Variants of Uncertain Significance (VUS) are enriched in cases vs. controls. (A) Enrichment of VUS in cases for sets of genes. Odds ratios (OR) and 95% confidence intervals (95% CI) are calculated from a chi-square test. Point estimates are scaled by the number of genes in the gene set. (B) The natural log odds ratios for individual genes for 'likely pathogenic' (LP) variants only vs. 'likely pathogenic' and variants of uncertain significance (LP + VUS). (C) Individual genes with natural log odds ratios greater than 1 for 'likely pathogenic' (LP) variants only or likely pathogenic' and variants of uncertain significance (LP + VUS). The dotted lines show OR = 1 for the x and y axes and y=x. In all panels, genes/gene sets are colored by the level of supporting evidence as recorded in the UK NHS Cleft Lip and Palate PanelApp: high evidence (green), moderate evidence (amber), low evidence (red), or not on the PanelApp gene list (gray).



Supplementary Tables

Table S4-1: Demographics of OFC Cohort

	Number of Probands					
	Population				Sex	
Cleft Type	**Asian**	**European**	**Latinos**	**Total**	**Female**	**Male**
Cleft Lip	0	99	8	107	47	60
Cleft Lip & Palate	121	273	266	660	233	427
Cleft Palate	0	74	0	74	35	39
Total	121	446	274	841	315	526

Table S4-2: Family History Information for Study Populations

							Family History			
	Complete Trios	**Dyads**	**Singletons**	**Simplex Trios**	**Multiplex Trios**	**Affected Parent**	**Maternal FamHx**	**Paternal FamHx**	**Both FamHx**	**Unknown FamHx**
Case	765	60	16	621	220	63	86	62	19	53
Control	294	-	-	294	0	-	-	-	-	-

Table S4-3: OFC Gene List

Gene ID	Z-Score	LOEUF	UK PanelApp Evidence	Inheritance Pattern	GP	NHS	OMIM	PGCP
ABCA12	1.63	0.478	NA	Autosomal Recessive	0	0	0	1
ACACB	1.57	0.857	NA	Unspecified	0	0	0	1
ACBD5	0.873	0.529	Red	Autosomal Recessive	1	1	0	1
ACSS2	1.08	0.903	NA	Unspecified	0	0	0	1
ACTB	5.02	0.232	Green	Autosomal Dominant	1	1	1	0
ACTC1	4.52	0.48	NA	Unspecified	0	0	0	1
ACTG1	3.16	0.858	Green	Autosomal Dominant	0	1	1	0
ADH7	-0.613	1.617	NA	Unspecified	0	0	0	1
AHDC1	2.86	0.076	NA	Unspecified	0	0	0	1
ALG9	1.01	0.653	Red	Autosomal Recessive	0	1	0	0
ALX1	0.0961	0.607	Amber	Autosomal Recessive	1	1	1	1
ALX3	0.446	1.032	Amber	Autosomal Recessive	1	1	1	0
AMELX	0.957	0.947	NA	X-Linked	0	0	0	1
AMER1	-0.569	0.371	Green	X-Linked	1	1	1	0
AMMECR1	1.08	0.472	NA	X-Linked	0	0	1	0
ANKRD11	-0.554	0.107	Green	Autosomal Dominant	1	1	0	0
ANOS1	0.89	0.25	NA	X-Linked	0	0	0	1
ARCN1	2.08	0.101	Red	Autosomal Dominant	0	1	1	0
ARHGAP29	1.21	0.249	Green	Autosomal Dominant	1	1	0	1
ARHGAP31	0.797	0.192	Green	Autosomal Dominant	0	1	1	0
ARID5B	2.6	0.11	NA	Unspecified	0	0	0	1
ASXL1	0.636	0.794	Green	Autosomal Dominant	1	1	1	0
ATN1	1.76	0.194	NA	Autosomal Dominant	0	0	1	0
ATR	4.36	0.399	Amber	Autosomal Recessive	0	1	1	0
ATRX	3.1	0.119	Red	X-Linked	0	1	0	0
B3GALT6	1.26	1.425	Amber	Autosomal Recessive	0	1	1	0
B3GAT3	0.692	1.331	Red	Autosomal Recessive	0	1	1	0
B3GLCT	0.39	NA	Green	Autosomal Recessive	1	1	1	0
B4GALT7	-0.311	1.11	Amber	Autosomal Recessive	0	1	1	0
B9D2	-0.13	1.575	NA	Autosomal Recessive	0	0	1	0
BCOR	1.88	0.141	Green	X-Linked	1	1	1	0
BMP2	1.14	0.224	Green	Autosomal Dominant	1	1	0	1
BMP4	1.01	0.334	Red	Autosomal Dominant	1	1	1	1
BMPER	0.0102	0.663	NA	Autosomal Recessive	0	0	1	0
BPNT2	0.83	0.84	Green	Autosomal Recessive	0	1	1	0
BUB1B	0.893	0.58	Amber	Autosomal Recessive	0	1	1	0

C2CD3	0.535	0.644	Green	Autosomal Recessive	1	1	1	0
CADPS	3.1	0.435	NA	Unspecified	0	0	0	1
CASK	4.25	0.073	Red	X-Linked	0	1	0	0
CBFB	1.71	0.33	NA	Unspecified	0	0	0	1
CC2D2A	0.647	0.781	Green	Autosomal Recessive	0	1	1	0
CDC45	0.828	0.598	Amber	Autosomal Recessive	0	1	1	0
CDH1	0.706	0.43	Green	Autosomal Dominant	1	1	1	1
CDH19	-0.708	1.119	NA	Unspecified	0	0	0	1
CDKN1C	1.93	0.529	Green	Autosomal Dominant	1	1	0	1
CDON	0.209	0.779	NA	Autosomal Dominant	1	0	1	1
CENPF	-0.0297	0.733	NA	Autosomal Recessive	0	0	1	0
CEP120	-0.233	0.795	NA	Autosomal Recessive	0	0	1	0
CEP290	0.475	0.985	NA	Autosomal Recessive	0	0	1	0
CFAP57	0.45	NA	NA	Unspecified	0	0	0	1
CFDP1	-0.521	0.754	NA	Unspecified	0	0	0	1
CHD7	3.22	0.076	Green	Autosomal Dominant	1	1	1	1
CHN2	1.13	0.628	NA	Unspecified	0	0	0	1
CHRNA1	0.722	1.079	NA	Autosomal Recessive	1	0	1	0
CHRNA7	1.25	0.768	NA	Unspecified	0	0	0	1
CHRND	0.054	0.978	NA	Autosomal Recessive	0	0	1	0
CHRNG	-0.118	1.316	Green	Autosomal Recessive	1	1	1	0
CHST14	0.683	0.64	Green	Autosomal Recessive	0	1	1	0
CHSY1	0.831	0.405	Red	Autosomal Recessive	0	1	1	0
CILK1	1.24	0.61	Green	Autosomal Recessive	1	1	1	0
CKAP2L	-0.089	0.626	Red	Autosomal Recessive	0	1	0	0
CLPTM1	2.93	0.326	NA	Unspecified	0	0	0	1
CLPTM1L	0.567	0.645	NA	Unspecified	0	0	0	1
CNOT2	3.65	0.189	NA	Autosomal Dominant	0	0	1	0
COG1	-0.181	0.649	NA	Autosomal Recessive	0	0	1	0
COG5	-0.483	0.965	NA	Unspecified	0	0	0	1
COL11A1	1.02	0.216	Green	Autosomal Dominant	1	1	1	1
COL11A2	2.37	0.306	Green	Autosomal Dominant	1	1	1	1
COL2A1	3.29	0.134	Green	Autosomal Dominant	1	1	1	1
COL9A1	0.186	0.814	Green	Autosomal Dominant	1	1	0	1
COL9A2	1.34	0.636	Amber	Autosomal Dominant	1	1	0	1
COL9A3	-0.382	0.746	Red	Autosomal Recessive	1	1	0	1
COLEC10	0.0281	0.965	Green	Autosomal Recessive	1	1	1	0
COLEC11	0.564	0.816	Green	Autosomal Recessive	1	1	1	1

CPLANE1	1.93	0.67	Green	Autosomal Recessive	1	1	1	0
CRYZ	0.381	1.928	NA	Unspecified	0	0	0	1
CTCF	4.44	0.148	Green	Autosomal Dominant	0	1	1	0
CTNND1	2.08	0.185	Green	Autosomal Dominant	1	1	0	1
DAB1	1.52	0.215	NA	Unspecified	0	0	0	1
DDX3X	4.33	0.118	Amber	X-Linked	0	1	1	0
DDX59	0.779	0.876	Amber	Autosomal Recessive	1	1	1	0
DEAF1	1.5	0.698	NA	Unspecified	0	0	0	1
DENND4B	3.23	0.29	NA	Unspecified	0	0	0	1
DHCR24	1.5	0.523	NA	Autosomal Recessive	0	0	1	0
DHCR7	-0.452	1.349	Green	Autosomal Recessive	1	1	1	1
DHODH	-0.337	0.778	Green	Autosomal Recessive	1	1	1	1
DIS3L2	0.944	0.325	Red	Autosomal Recessive	0	1	0	0
DLG1	1.5	0.286	Red	Unspecified	1	1	0	1
DLL4	2.71	0.188	Green	Autosomal Dominant	0	1	0	0
DLX4	-0.0197	1.067	Amber	Autosomal Dominant	1	1	1	1
DNMT3B	1.5	0.375	Red	Autosomal Recessive	0	1	0	0
DOCK6	1.43	0.717	Green	Autosomal Recessive	0	1	0	0
DONSON	0.243	0.907	NA	Autosomal Recessive	0	0	1	0
DUSP22	0.451	0.934	NA	Unspecified	0	0	0	1
DVL1	-1.13	0.733	Green	Autosomal Dominant	0	1	0	0
DVL3	2.32	0.422	Green	Autosomal Dominant	1	1	1	1
DYNC2H1	0.909	0.576	Green	Autosomal Recessive	1	1	1	0
DYNC2I1	0.42	0.67	Amber	Autosomal Recessive	0	1	0	0
DYNC2I2	-0.12	0.99	Amber	Autosomal Recessive	0	1	0	0
DYNC2LI1	-0.326	1.129	Green	Autosomal Recessive	0	1	0	0
EARS2	0.108	1.021	NA	Autosomal Recessive	0	0	1	0
EBP	1.49	0.342	Green	X-Linked	0	1	0	0
ECEL1	0.478	0.831	NA	Autosomal Recessive	0	0	1	0
EDN1	0.169	0.635	Red	Autosomal Recessive	0	1	0	0
EDNRA	2.8	0.254	Green	Autosomal Dominant	0	1	1	0
EFNB1	1.59	0.349	Green	X-Linked	0	1	1	0
EFTUD2	4.03	0.094	Green	Autosomal Dominant	1	1	1	1
EIF2S3	3.61	0.322	Green	X-Linked	0	1	1	0
EIF4A3	4.02	0.119	Green	Autosomal Recessive	1	1	1	0
EMG1	-0.304	1.247	NA	Autosomal Recessive	0	0	1	0
EOGT	-0.0882	0.985	Green	Autosomal Recessive	0	1	0	0
EPG5	1.08	0.456	Green	Autosomal Recessive	1	1	1	0

ERCC5	0.118	0.805	NA	Autosomal Recessive	1	0	1	0
ESCO2	0.408	0.833	Green	Autosomal Recessive	1	1	1	1
ESRP2	0.824	0.715	Amber	Autosomal Dominant	1	1	0	1
EVC	-1.25	1.064	NA	Autosomal Recessive	0	0	1	0
EVC2	-2.14	1.064	NA	Autosomal Recessive	1	0	1	1
EYA1	1.02	0.332	Green	Autosomal Dominant	1	1	1	1
FAM111A	-0.021	1.949	Red	Autosomal Dominant	0	1	0	0
FAM149B1	1.56	1.049	NA	Autosomal Recessive	0	0	1	0
FAM20C	1.02	0.489	Green	Autosomal Recessive	1	1	1	1
FANCL	-1.31	1.614	Red	Autosomal Recessive	0	1	1	0
FBLN1	1.43	0.338	NA	Unspecified	0	0	0	1
FBXO11	4.38	0.097	Amber	Autosomal Dominant	0	1	0	0
FGD1	3.52	0.196	Green	X-Linked	1	1	1	1
FGF10	0.85	0.339	NA	Unspecified	0	0	0	1
FGF17	1.85	0.305	NA	Autosomal Dominant	0	0	1	0
FGF8	1.57	0.51	NA	Autosomal Dominant	1	0	1	1
FGF9	1.54	0.323	NA	Autosomal Dominant	0	0	1	0
FGFR1	2.49	0.215	Green	Autosomal Dominant	1	1	1	1
FGFR2	2.4	0.27	Green	Autosomal Dominant	1	1	1	1
FGFR3	1.26	0.604	NA	Unspecified	0	0	0	1
FLNA	3.78	0.082	Green	X-Linked	1	1	1	1
FLNB	2.14	0.443	Green	Autosomal Dominant	1	1	1	0
FLRT3	1.86	0.52	NA	Autosomal Dominant	0	0	1	0
FOXC2	-0.251	0.788	Green	Autosomal Dominant	1	1	1	0
FOXE1	0.315	0.576	Red	Autosomal Recessive	1	1	0	1
FOXP2	1.9	0.219	Amber	Autosomal Dominant	0	1	0	0
FRAS1	0.078	0.658	Green	Autosomal Recessive	1	1	1	1
FREM2	-0.859	0.475	Red	Autosomal Recessive	1	1	0	1
FTO	0.562	0.85	Amber	Autosomal Recessive	0	1	1	0
FZD2	3.67	0.471	NA	Autosomal Dominant	1	0	1	0
FZD6	1.4	1.079	NA	Unspecified	0	0	0	1
G6PC3	0.696	1.031	NA	Autosomal Recessive	0	0	1	0
GATA3	1.91	0.388	Amber	Autosomal Dominant	0	1	0	0
GATA6	1.28	0.174	Red	Unspecified	0	1	0	0
GDF6	0.934	0.223	NA	Autosomal Dominant	0	0	1	0
GJA1	1.28	0.624	Green	Autosomal Dominant	1	1	1	1
GLI2	0.817	0.309	NA	Autosomal Dominant	1	0	1	1
GLI3	0.522	0.195	Green	Autosomal Dominant	1	1	1	1

GMNN	0.991	0.825	Red	Autosomal Dominant	0	1	1	0
GMPPB	1.02	1.092	NA	Autosomal Recessive	0	0	1	0
GNAI3	1.79	0.611	Red	Autosomal Dominant	1	1	1	0
GNB1	3.83	0.145	Amber	Autosomal Dominant	0	1	1	0
GPC3	1.45	0.162	Green	X-Linked	1	1	1	0
GPC6	0.901	0.511	NA	Autosomal Recessive	0	0	0	1
GREM1	1.55	0.818	NA	Autosomal Dominant	0	0	0	1
GRHL3	1.42	0.277	Green	Autosomal Dominant	1	1	1	1
GRIP1	1.58	0.356	Red	Autosomal Dominant	0	1	0	0
HAAO	-0.177	1.141	NA	Autosomal Recessive	0	0	1	0
HDAC8	2.82	0.297	Green	X-Linked	1	1	1	0
HNRNPK	3.99	0.096	NA	Autosomal Dominant	0	0	1	0
HOXA2	0.414	0.699	Red	Autosomal Dominant	1	1	1	0
HS6ST1	1.92	0.376	NA	Autosomal Dominant	1	0	1	0
HYAL2	1.4	0.535	NA	Unspecified	0	0	0	1
HYLS1	0.177	1.433	Green	Autosomal Recessive	0	1	1	0
IFT140	-0.811	0.818	Green	Autosomal Recessive	1	1	0	0
IFT172	1.19	0.755	Green	Autosomal Recessive	0	1	1	0
IFT52	1.44	0.866	Amber	Autosomal Recessive	0	1	0	0
IFT57	0.302	1.096	NA	Autosomal Recessive	0	0	1	0
IFT80	1.51	0.991	Green	Autosomal Recessive	0	1	0	0
IFT88	0.335	0.708	NA	Unspecified	0	0	0	1
IGF1R	2.73	0.306	NA	Unspecified	0	0	0	1
IGF2	0.931	1.131	NA	Autosomal Dominant	0	0	1	0
IL1B	1.15	0.648	NA	Autosomal Dominant	0	0	1	0
IL1RN	0.00812	0.907	NA	Autosomal Dominant	0	0	1	0
INPPL1	1.6	0.423	NA	Unspecified	0	0	0	1
INTS1	1.41	0.476	Red	Unspecified	0	1	0	0
INTU	0.384	0.74	NA	Autosomal Recessive	1	0	1	0
IRF6	2.74	0.132	Green	Autosomal Dominant	1	1	1	1
ISM1	1.07	0.696	NA	Autosomal Dominant	0	0	0	1
JAG2	2.2	0.143	NA	Unspecified	0	0	0	1
KANSL1	1.23	0.238	Red	Autosomal Dominant	1	1	1	0
KAT6A	2.07	0.069	Green	Autosomal Dominant	0	1	0	0
KAT6B	2.89	0.128	Red	Autosomal Dominant	0	1	1	0
KATNIP	-0.11	0.91	NA	Autosomal Recessive	0	0	1	0
KCNJ2	2.75	0.615	Green	Autosomal Dominant	0	1	1	0
KDM1A	4.68	0.276	Amber	Autosomal Dominant	1	1	1	0

KDM6A	2.95	0.161	Green	X-Linked	1	1	1	1
KIAA0586	0.371	0.719	Green	Autosomal Recessive	1	1	1	0
KIF22	0.141	0.851	Red	Autosomal Dominant	0	1	0	0
KIF7	-0.689	0.926	Green	Autosomal Recessive	0	1	1	0
KIFBP	1.45	0.51	Green	Autosomal Recessive	0	1	0	0
KLF4	0.1	0.174	NA	Autosomal Dominant	0	0	0	1
KLHL41	0.887	0.795	NA	Autosomal Recessive	0	0	1	0
KMT2D	3.73	0.103	Green	Autosomal Dominant	1	1	1	1
KRAS	2.32	1.24	NA	Autosomal Dominant	0	0	1	0
LHX8	1.61	0.473	NA	Unspecified	0	0	0	1
LMNA	2.37	0.209	Amber	Autosomal Recessive	0	1	1	0
LMX1B	2.02	0.415	Red	Autosomal Dominant	0	1	1	0
LRP8	2.52	0.224	NA	Unspecified	0	0	0	1
MAGEL2	-0.563	0.312	NA	Autosomal Dominant	0	0	1	0
MAP3K7	2.79	0.212	Green	Autosomal Dominant	0	1	1	0
MAPRE2	3.19	0.406	Green	Autosomal Dominant	1	1	1	0
MASP1	-0.16	0.978	Green	Autosomal Recessive	1	1	1	1
MBTPS2	2.09	0.17	Green	X-Linked	1	1	1	0
MED12	6.58	0.071	Red	X-Linked	0	1	1	0
MED13L	3.69	0.064	Amber	Autosomal Dominant	0	1	0	0
MED25	1.25	0.652	Amber	Autosomal Recessive	0	1	1	0
MEGF10	1.05	0.347	NA	Autosomal Recessive	0	0	1	0
MEIS2	2.46	0.184	Green	Autosomal Dominant	1	1	1	1
MEOX1	-0.261	1.268	Amber	Autosomal Recessive	1	1	1	0
METTL23	-0.973	1.77	Red	Autosomal Recessive	0	1	0	0
MIB1	3.21	1.973	NA	Unspecified	0	0	0	1
MID1	2.92	0.304	Green	X-Linked	1	1	1	0
MKS1	0.492	1.042	Green	Autosomal Recessive	0	1	1	0
MN1	2.16	0.087	NA	Autosomal Dominant	1	0	0	1
MSX1	0.278	1.01	Green	Autosomal Dominant	1	1	1	1
MSX2	0.471	0.819	NA	Autosomal Dominant	1	0	1	1
MUSK	1.09	0.698	NA	Autosomal Recessive	0	0	1	0
MYH3	1.74	0.63	NA	Autosomal Dominant	1	0	1	1
MYMK	0.76	NA	Green	Autosomal Recessive	1	1	1	0
NAA10	2.41	0.522	NA	X-Linked	0	0	1	0
NBAS	-0.875	0.777	NA	Autosomal Recessive	1	0	0	1
NBN	0.609	1.01	Red	Autosomal Recessive	0	1	1	0
NEB	-0.0409	0.424	NA	Autosomal Recessive	0	0	1	0

NECTIN1	1.28	0.57	Green	Autosomal Recessive	1	1	1	1
NECTIN2	1.64	0.32	NA	Unspecified	0	0	0	1
NEDD4L	3.73	0.198	Green	Autosomal Dominant	1	1	1	1
NEK1	1.07	0.862	Green	Autosomal Recessive	1	1	1	0
NFASC	2.59	0.215	NA	Autosomal Recessive	0	0	1	0
NIPBL	5.57	0.032	Green	Autosomal Dominant	1	1	1	0
NOTCH1	3.45	0.097	Green	Autosomal Dominant	0	1	0	0
NOTCH2	3.5	0.116	NA	Unspecified	0	0	0	1
NSDHL	0.869	0.289	Red	X-Linked	0	1	1	0
NSMF	1.88	0.443	NA	Autosomal Dominant	0	0	1	0
NUDT6	-0.0242	0.948	NA	Unspecified	0	0	0	1
NUP107	1.89	0.57	NA	Autosomal Recessive	0	0	1	0
NUP188	1.12	0.297	NA	Autosomal Recessive	0	0	1	0
NXN	2.18	0.218	NA	Autosomal Recessive	0	0	1	0
OFD1	0.32	0.322	Green	X-Linked	1	1	1	0
ORC1	0.236	0.881	NA	Autosomal Recessive	0	0	1	0
OTX2	1.05	0.376	NA	Autosomal Dominant	0	0	1	0
PAX3	1.62	0.475	Green	Autosomal Dominant	1	1	1	1
PAX7	0.804	0.558	NA	Autosomal Dominant	0	0	0	1
PAX9	0.619	0.614	NA	Unspecified	0	0	0	1
PDGFC	1.17	0.198	NA	Unspecified	0	0	0	1
PDGFRA	1.94	0.169	NA	Unspecified	0	0	0	1
PEX2	0.0473	1.198	NA	Autosomal Recessive	0	0	1	0
PEX7	0.497	1.416	NA	Autosomal Recessive	0	0	1	0
PGAP2	0.933	0.363	Red	Autosomal Recessive	0	1	1	0
PGAP3	0.685	0.833	NA	Autosomal Recessive	0	0	1	0
PGM1	-0.228	1.319	Red	Autosomal Recessive	1	1	1	1
PHF8	3.98	0.263	Green	X-Linked	1	1	1	1
PHGDH	0.207	0.778	Amber	Autosomal Recessive	1	1	1	0
PHIP	5.14	0.113	NA	Autosomal Dominant	1	0	0	1
PHYH	0.031	1.122	NA	Unspecified	0	0	0	1
PIBF1	-0.456	0.719	NA	Unspecified	0	0	0	1
PIEZO2	3.44	0.371	Green	Autosomal Dominant	1	1	1	0
PIGA	2.17	0.296	Red	X-Linked	1	1	0	1
PIGL	0.0451	1.506	Red	Autosomal Recessive	0	1	1	0
PIGN	0.416	1.025	Green	Autosomal Recessive	0	1	1	0
PIGO	0.957	0.812	NA	Autosomal Recessive	0	0	1	0
PIGV	0.398	1.238	Green	Autosomal Recessive	0	1	1	0

PIK3R2	2.33	0.489	Red	Autosomal Dominant	0	1	1	0
PLCB4	3.57	0.449	Red	Autosomal Dominant	1	1	1	0
PLEKHA5	1.24	0.098	Red	Autosomal Dominant	1	1	0	1
PLEKHA7	0.214	0.531	Amber	Autosomal Dominant	1	1	0	1
POLR1A	2.95	0.256	Amber	Autosomal Dominant	0	1	1	0
POLR1B	2.97	0.246	NA	Autosomal Dominant	0	0	1	0
POLR1C	-1.44	1.401	Green	Autosomal Recessive	1	1	1	1
POLR1D	0.169	1.013	Green	Autosomal Dominant	1	1	1	1
POMT1	0.618	0.934	Red	Autosomal Recessive	0	1	1	0
POMT2	0.0763	0.795	Red	Autosomal Recessive	0	1	1	0
PORCN	2.37	0.135	Green	X-Linked	1	1	1	1
PPP1R21	-1.92	0.686	NA	Unspecified	0	0	0	1
PPP3CA	3.63	0.158	NA	Autosomal Dominant	1	0	1	0
PQBP1	1.92	0.499	Red	X-Linked	0	1	1	0
PREPL	-2.5	0.994	NA	Unspecified	0	0	0	1
PRICKLE1	1.83	0.129	NA	Autosomal Dominant	0	0	0	1
PRRX1	1.02	0.659	NA	Autosomal Dominant	0	0	1	0
PSAT1	0.413	1.273	Red	Autosomal Recessive	0	1	1	0
PTCH1	1.68	0.075	Green	Autosomal Dominant	1	1	1	1
PTCH2	0.785	0.85	NA	Autosomal Dominant	0	0	1	0
PTDSS1	2.38	0.448	Red	Autosomal Dominant	0	1	0	0
PTHLH	1.15	0.48	NA	Unspecified	0	0	0	1
PTPN11	3.13	0.135	NA	Autosomal Dominant	0	0	1	0
PTPRS	3.78	0.254	NA	Unspecified	0	0	0	1
RAD21	2.64	0.256	NA	Autosomal Dominant	0	0	1	0
RAI1	1.12	0.116	Red	Autosomal Dominant	0	1	0	0
RAPSN	0.548	0.741	NA	Autosomal Recessive	0	0	1	0
RARB	2.87	0.12	Amber	Autosomal Recessive	0	1	0	0
RBM10	4.46	0.11	Green	X-Linked	0	1	0	0
RBM8A	2.16	0.564	Red	Autosomal Recessive	0	1	0	0
RBPJ	3.57	0.241	Amber	Autosomal Dominant	0	1	0	0
RIC1	0.07	0.4	NA	Autosomal Recessive	0	0	1	0
RIMS3	0.974	1.06	NA	Unspecified	0	0	0	1
RIPK4	1.89	0.594	NA	Autosomal Recessive	1	0	1	1
ROR2	0.143	0.558	Green	Autosomal Recessive	0	1	0	0
RPGRIP1L	-0.108	0.956	NA	Autosomal Recessive	1	0	1	1
RPL11	1.69	0.296	Red	Autosomal Dominant	0	1	1	0
RPL26	1.64	0.376	NA	Autosomal Dominant	0	0	1	0

RPL5	1.9	0.167	Green	Autosomal Dominant	1	1	1	0
RPS17	NA	1.849	Red	Autosomal Dominant	0	1	0	0
RPS19	1.45	0.368	Red	Autosomal Dominant	0	1	1	0
RPS23	2.36	0.465	NA	Autosomal Dominant	0	0	1	0
RPS26	1.93	0.483	Green	Autosomal Dominant	1	1	1	0
RPS28	1.51	0.796	Amber	Autosomal Dominant	0	1	1	0
RSPO2	0.496	0.762	Red	Autosomal Recessive	1	1	1	0
RUNX2	1.6	0.343	NA	Autosomal Dominant	0	0	1	0
RYK	1.81	0.289	NA	Unspecified	0	0	0	1
RYR1	1.92	0.456	Amber	Autosomal Dominant	1	1	0	1
SALL4	1.08	0.101	Green	Autosomal Dominant	0	1	0	0
SATB2	4.05	0.091	Green	Autosomal Dominant	1	1	1	1
SCARF2	3.02	0.19	Green	Autosomal Recessive	0	1	1	0
SEC23A	1.79	0.741	Amber	Autosomal Recessive	0	1	1	0
SELENOI	1.67	0.44	Red	Autosomal Recessive	0	1	1	0
SEMA3E	0.465	0.679	NA	Autosomal Dominant	1	0	1	1
SEPTIN9	1.62	0.25	Green	Autosomal Dominant	1	1	1	0
SF3B4	3.87	0.208	Green	Autosomal Dominant	1	1	1	1
SHH	2.95	0.242	Green	Autosomal Dominant	1	1	1	1
SHROOM3	1.24	0.352	NA	Autosomal Dominant	1	0	0	0
SIX1	1.19	0.522	Green	Autosomal Dominant	1	1	0	0
SIX3	2.07	0.323	Green	Autosomal Dominant	1	1	1	1
SIX5	-0.563	0.802	Green	Autosomal Dominant	1	1	0	0
SKI	1.51	0.194	Green	Autosomal Dominant	1	1	0	1
SLC10A7	1.34	1	NA	Autosomal Recessive	0	0	1	0
SLC26A2	0.0314	1.106	Green	Autosomal Recessive	1	1	1	0
SLC35D1	0.0255	0.784	NA	Autosomal Recessive	0	0	1	0
SMAD2	3.66	0.259	Red	Autosomal Dominant	1	1	0	1
SMAD3	3.48	0.4	Green	Autosomal Dominant	1	1	1	1
SMAD4	4.13	0.222	Green	Autosomal Dominant	0	1	1	0
SMC1A	6.45	0.062	Green	X-Linked	1	1	0	1
SMC3	6.4	0.037	Green	Autosomal Dominant	0	1	0	0
SMCHD1	3.63	0.153	NA	Autosomal Dominant	1	0	1	0
SMG9	1.6	1.098	Amber	Autosomal Recessive	0	1	1	0
SMO	1.9	0.684	NA	Autosomal Recessive	0	0	1	0
SMOC1	0.743	0.592	Red	Autosomal Recessive	0	1	1	0
SMPD4	0.857	1.022	NA	Autosomal Recessive	0	0	1	0
SMS	2.32	0.305	Green	X-Linked	0	1	1	0

SNAP29	-0.872	0.861	NA	Autosomal Recessive	0	0	0	1
SNRPB	1.87	0.494	Green	Autosomal Dominant	0	1	0	0
SNTG1	-0.329	0.445	NA	Unspecified	0	0	0	1
SON	1.54	0.118	Green	Autosomal Dominant	0	1	1	0
SOX2	2.12	0.569	Red	Autosomal Dominant	0	1	0	0
SOX9	1.63	0.168	Green	Autosomal Dominant	1	1	1	1
SP8	1.69	0.285	NA	Unspecified	0	0	0	1
SPECC1L	1.6	0.332	Green	Autosomal Dominant	1	1	1	1
SPOP	4.14	0.144	NA	Autosomal Dominant	0	0	1	0
SPRY4	0.503	1.374	NA	Unspecified	0	0	0	1
SRCAP	2.13	0.096	NA	Autosomal Dominant	1	0	0	0
STAC3	0.906	0.754	NA	Autosomal Recessive	1	0	1	0
STAG2	4.94	0.09	NA	X-Linked	1	0	1	0
STAMBP	0.886	0.764	Green	Autosomal Recessive	0	1	1	0
STIL	1.13	0.383	Red	Autosomal Recessive	0	1	0	0
STRA6	0.599	0.824	Red	Autosomal Recessive	0	1	0	0
STXBP1	4.26	0.086	Red	Autosomal Dominant	0	1	0	0
SUFU	1.93	0.111	NA	Autosomal Dominant	1	0	1	0
SUMO1	1.51	0.463	Red	Autosomal Dominant	1	1	1	1
TAPT1	1.92	0.564	NA	Autosomal Recessive	0	0	1	0
TBX1	0.735	0.427	Amber	Autosomal Dominant	1	1	1	1
TBX10	-0.0686	1.362	NA	Unspecified	0	0	0	1
TBX15	0.993	0.437	Amber	Autosomal Recessive	0	1	1	0
TBX2	1.54	0.246	NA	Autosomal Dominant	0	0	1	0
TBX22	-0.126	0.297	Green	X-Linked	1	1	0	1
TBX4	0.871	0.445	NA	Autosomal Dominant	0	0	1	0
TCOF1	0.339	0.311	Green	Autosomal Dominant	1	1	1	1
TCTN2	0.237	0.99	NA	Autosomal Recessive	1	0	1	0
TCTN3	0.836	1.04	Green	Autosomal Recessive	1	1	1	0
TELO2	-0.367	0.927	Green	Autosomal Recessive	0	1	0	0
TFAP2A	2.59	0.261	Green	Autosomal Dominant	1	1	1	1
TFAP2B	1.29	0.259	Red	Autosomal Dominant	0	1	0	0
TGDS	0.3	1.148	Green	Autosomal Recessive	1	1	1	1
TGFB1	1.86	0.614	NA	Unspecified	0	0	0	1
TGFB2	2.13	0.148	Red	Autosomal Dominant	1	1	0	0
TGFB3	1.66	0.27	Green	Autosomal Dominant	1	1	1	1
TGFBR1	2.79	0.383	Green	Autosomal Dominant	1	1	1	0
TGFBR2	2.24	0.519	Green	Autosomal Dominant	1	1	1	1

TGIF1	0.178	0.911	NA	Unspecified	0	0	0	1
TMCO1	1.03	1.362	Green	Autosomal Recessive	1	1	1	0
TMEM216	0.491	1.37	NA	Autosomal Recessive	0	0	1	0
TMEM67	0.493	0.944	NA	Autosomal Recessive	0	0	1	0
TP63	2.21	0.267	Green	Autosomal Dominant	1	1	1	1
TRAPPC9	1.48	0.733	Green	Autosomal Recessive	0	1	1	0
TRIM37	0.82	0.642	Green	Autosomal Recessive	0	1	0	0
TRRAP	8.17	0.06	NA	Autosomal Dominant	0	0	1	0
TSR2	0.989	0.561	Red	X-Linked	0	1	0	0
TTC21B	-0.409	0.852	Amber	Autosomal Recessive	0	1	0	0
TUBB	5.63	0.292	Green	Autosomal Dominant	1	1	1	0
TWIST1	1.06	1.056	NA	Autosomal Dominant	0	0	1	0
TWIST2	2.13	1.154	Red	Autosomal Dominant	0	1	0	0
TXNL4A	2.77	1.822	Green	Autosomal Recessive	1	1	1	0
TXNRD2	0.636	0.99	NA	Unspecified	0	0	0	1
UBB	2.73	1.56	Red	Autosomal Dominant	0	1	0	0
UFD1	2.71	0.23	NA	Unspecified	0	0	0	1
UGT1A9	-0.937	1.112	NA	Unspecified	0	0	0	1
UQCC2	0.499	1.039	Red	Autosomal Recessive	0	1	0	0
USP9X	6.41	0.051	Green	X-Linked	0	1	0	0
VAX1	0.869	0.515	Red	Autosomal Recessive	1	1	1	1
WASHC5	1.3	0.74	Red	Autosomal Recessive	1	1	1	0
WDPCP	0.898	0.782	NA	Autosomal Recessive	0	0	1	0
WDR11	1.8	0.532	NA	Unspecified	0	0	0	1
WDR19	1.5	0.561	Amber	Autosomal Recessive	0	1	0	0
WDR35	0.599	0.815	Amber	Autosomal Recessive	1	1	1	1
WNT3	3.18	0.404	Red	Autosomal Recessive	1	1	1	0
WNT4	1.61	0.651	NA	Autosomal Recessive	0	0	1	0
WNT5A	1.98	0.269	Green	Autosomal Dominant	1	1	1	1
WNT7A	1.34	0.71	NA	Unspecified	0	0	0	1
XYLT1	0.586	0.336	Green	Autosomal Recessive	0	1	1	0
YAP1	1.78	0.19	Red	Autosomal Dominant	1	1	1	0
YWHAE	2.83	0.235	NA	Unspecified	0	0	0	1
ZBTB24	1.01	0.664	Amber	Autosomal Recessive	1	1	0	1
ZC4H2	1.51	0.39	NA	X-Linked	0	0	1	0
ZEB2	3.94	0.107	Green	Autosomal Dominant	1	1	1	0
ZIC2	3.21	0.267	Green	Autosomal Dominant	0	1	1	0
ZIC3	2.52	0.361	Green	X-Linked	0	1	0	0

ZMPSTE24	0.372	1.327	Amber	Autosomal Recessive	0	1	1	0
ZSWIM6	4.18	0.067	Green	Autosomal Dominant	1	1	1	0

Table S4-4: Analyzed Structural Variants

Structural Variant ID	Mode of Inheritance	Type	Allele Frequency	Gene
chr1:2251423-2251424	*De Novo*	DEL	9.40E-05	*SKI*
chr1:145564585-146057000	*De Novo*	DEL	2.83E-04	*AC243547.3,ANKRD34A,ANKRD35,CD1 JV,ITGA10,LIX1L,NUDT17,PDZK1,PEX1 LR3C,RNF115,TXNIP*
chr1:224448382-224448383	*De Novo*	CTX	9.40E-05	*CNIH3*
chr11:57721382-57799985	*De Novo*	DEL	.40E-05	*AP001931.1,BTBD18,CTNND1,SELENO 2-CTNND1*
chr14:49952415-55275096	*De Novo*	DEL	9.40E-05	*ABHD12B,ATL1,ATP5S,BMP4,CDKN3,C 1,DDHD1,DLGAP5,ERO1A,FERMT2,FR MFB,GNG2,GNPNAT1,GPR137C,L2HG MAP4K5,MAPK1IP1L,NID2,NIN,PSMC6, ER2,PYGL,RTRAF,SAMD4A,SAV1,SOCS4, MX1,TRIM9,TXNDC16,VCPKMT,WDHD1*
chr14:58504983-58514932	*De Novo*	DUP	9.40E-05	NA
chr2:19931363-19931415	*De Novo*	INS:ME:ALU	2.83E-04	*WDR35*
chr22:18889490-21465674	*De Novo*	DEL	3.77E-04	*AC007731.5,AIFM3,ARVCF,C22orf39,CC 5,CLDN5,CLTCL1,COMT,CRKL,DGCR2, GNB1L,GP1BB,GSC2,HIRA,KLHL22,LZT RPL40,P2RX6,PI4KA,RANBP1,RTL10,SC SEPT5- IND1,SLC25A1,SLC7A4,SNAP29,TANGO 2A,TSSK2,TXNRD2,UFD1,USP41,ZDHH*
chr22:18889490-21465674	*De Novo*	DEL	3.77E-04	*AC007731.5,AIFM3,ARVCF,C22orf39,CC 5,CLDN5,CLTCL1,COMT,CRKL,DGCR2, GNB1L,GP1BB,GSC2,HIRA,KLHL22,LZT RPL40,P2RX6,PI4KA,RANBP1,RTL10,SC SEPT5- IND1,SLC25A1,SLC7A4,SNAP29,TANGO 2A,TSSK2,TXNRD2,UFD1,USP41,ZDHH*
chr22:21345564-22381564	*De Novo*	DEL	1.88E-04	*AC007731.5,AIFM3,ARVCF,C22orf39,CC 188,CDC45,CLDN5,CLTCL1,COMT,CRK CR6L,DGCR8,ESS2,GNB1L,GP1BB,GSC KLHL22,LZTR1,MAPK1,MED15,MRPL40 4,PPIL2,PPM1F,RANBP1,RIMBP3B,RIM SCARF2,SDF2L1,SEPT5,SEPT5- IND1,SLC25A1,SLC7A4,SNAP29,TANGO B,TRMT2A,TSSK2,TXNRD2,UBE2L3,UF REB1,YDJC,YPEL1,ZDHHC8,ZNF74*
chr4:78087238-78087239	Unknown	DEL	9.40E-05	*FRAS1*
chr4:156356665-156780752	*De Novo*	DEL	9.40E-05	*PDGFC*
chr6:10403251-10406453	*De Novo*	DEL	9.40E-05	*TFAP2A*
chr8:8269998-12039000	*De Novo*	DUP	1.88E-04	NA
chr1:93904955-94557144	Inherited	CPX	1.88E-04	*GCLM*
chr1:183652848-213099798	Inherited	INV	1.88E-04	*RGL1,RPS6KC1*

chr10:95687850-95688459	Inherited	INS	3.77E-04	*TCTN3*
chr10:95687850-95688459	Inherited	INS	3.77E-04	*TCTN3*
chr16:1500715-1501728	Inherited	DEL	2.83E-04	*TELO2*
chr16:68134717-68238471	Inherited	DEL	1.88E-04	*ESRP2,NFATC3*
chr18:62152639-62157701	Inherited	DEL	2.83E-04	*PIGN*
chr18:80030730-80059392	Inherited	CPX	1.88E-04	*AC090360.1,RBFA*
chr2:15371695-15469485	Inherited	DEL	1.88E-04	*NBAS*
chr2:44935132-45103884	Inherited	DEL	1.88E-04	*SIX2,SIX3*
chr2:232154308-232176809	Inherited	DEL	2.83E-04	*DIS3L2*
chr2:233639410-233731802	Inherited	DEL	1.88E-04	*UGT1A3,UGT1A4,UGT1A5,UGT1A6,UGT1A9*
chr2:233722899-233729618	Inherited	DEL	1.88E-04	*UGT1A3*
chr4:156827302-156895489	Inherited	DEL	2.83E-04	*PDGFC*
chr4:156827302-156895489	Inherited	DEL	2.83E-04	*PDGFC*
chr6:33698179-33701682	Inherited	DEL	3.77E-04	*UQCC2*
chr7:180179-211785	Inherited	DEL	1.88E-04	*FAM20C*
chr7:29168934-29499005	Inherited	DUP	3.77E-04	NA
chr7:29168934-29499005	Inherited	DUP	3.77E-04	NA
chr7:107294883-107316341	Inherited	DEL	1.88E-04	*COG5*
chrX:77508541-77514542	Inherited	DEL	2.83E-04	*ATRX*

Table S4-5: American College of Medical Genetics & Genomics Classification Modifications

American College of Medical Genetics & Genomics Criteria	**Modifications**
PM2: Absent from controls in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium	Variants with maximum allele frequency in any population < 0.001% among gnomAD (v2 and v3) and ExAC
PP3: Multiple lines of computational evidence supports a deleterious effect on the gene	At least 5 out of 9 "damaging" scores on SIFT, Polyphen, LRT, MutationTaster, FATHMM, PROVEAN, MetaSVM, MetaLR, M-CAP *in silico* predictions
BS1: Allele frequency is greater than expected for disorder	For heterozygous variants: Variants with maximum allele frequency in any population ≥ 0.00005 among gnomAD (v2 and v3) and ExAC
	For homozygous variant: Variants with maximum allele frequency in any population ≥ 0.002 among gnomAD (v2 and v3) and ExAC
BP4: Multiple lines of computational evidence suggest no impact on the gene	At least 5 out of 9 "tolerant" scores on SIFT, Polyphen, LRT, MutationTaster, FATHMM, PROVEAN, MetaSVM, MetaLR, M-CAP i*n silico* predictions

Table S4-6: Likely Pathogenic Variants in Present Study

Gene	Population	Cleft Type	Inheritance	Consequence	Variant
CDH1	Latino	CLP	Inherited	Missense	NM_004360.5: c.2245C>T; NP_004351.1: p.Arg749Trp
CDH1	Latino	CLP	Inherited	Missense	NM_004360.5: c.2245C>T; NP_004351.1: p.Arg749Trp
CTNND1	Latino	CLP	Inherited	Stopgain	NM_001331.3: c.1381C>T; NP_001322.1: p.Arg461*
CTNND1	Latino	CLP	*De Novo*	Stopgain	NM_001331.3: c.2414C>G; NP_001322.1: p.Ser805*
MYH3	Latino	CLP	Inherited	Frameshift	NM_002470.4: c.5511del; NP_002461.2: p.Val1838Leufs*4
SHROOM3	Latino	CLP	*De Novo*	Missense	NM_020859.4: c.5486A>G; NP_065910.3: p.Asn1829Ser
CTNND1	Latino	CLP	*De Novo*	Missense	NM_001331.3: c.2042C>G; NP_001322.1: p.Ala681Gly
GRHL3	Latino	CL	*De Novo*	Missense	NM_006147.4: c.262A>G; NP_006138.1: p.Asn88Asp
EYA1	Latino	CLP	Inherited	Stopgain	NM_000503.6: c.229C>T; NP_000494.2: p.Arg77*
EYA1	Latino	CLP	Inherited	Stopgain	NM_000503.6: c.229C>T; NP_000494.2: p.Arg77*
CHD7	Latino	CLP	*De Novo*	Stopgain	NM_017780.4: c.2572C>T; NP_060250.2: p.Arg858*
IRF6	Latino	CLP	*De Novo*	Missense	NM_021180.4: c.1453C>T; NP_067003.2: p.Pro485Ser
TP63	European	CLP	Inherited	Stopgain	NM_003722.5: c.1177C>T; NP_003713.3: p.Arg393*
ARHGAP29	European	CLP	Inherited	Splice	NM_004815.4: c.1281+2T>G
ANKRD11	European	CLP	*De Novo*	Missense	NM_013275.6: c.5431A>G; NP_037407.4: p.Arg1811Gly
COL11A1	European	CLP	*De Novo*	Missense	NM_080630.4: c.271C>A; NP_542197.3: p.Pro91Thr
FGF8	European	CLP	Inherited	Stopgain	NM_006119.6: c.298C>T; NP_006110.1: p.Arg100*
CDH1	European	CLP	*De Novo*	Missense	NM_004360.5: c.760G>A; NP_004351.1: p.Asp254Asn
FGFR1	European	CLP	Inherited	Stopgain	NM_001174063.2: c.1819C>T; NP_001167534.1: p.Arg607*
GATA3	European	CLP	unknown	Splice	NM_001002295.2: c.242-1G>T
PTPN11	European	CLP	unknown	Missense	NM_001330437.2: c.845T>C; NP_001317366.1: p.Ile282Thr
PTCH2	European	CLP	Inherited	Splice	NM_003738.5: c.617+1G>A
MIB1	European	CLP	Inherited	Stopgain	NM_020774.4: c.2502T>A; NP_065825.1: p.Cys834*
ARHGAP29	Asian	CLP	Inherited	Stopgain	NM_004815.4: c.2392C>T; NP_004806.3: p.Arg798*
FGF8	Asian	CLP	Inherited	Missense	NM_006119.6: c.311C>T; NP_006110.1: p.Thr104Met
TP63	Asian	CLP	Inherited	Stopgain	NM_003722.5: c.1156C>T; NP_003713.3: p.Gln386*

CTNND1	Asian	CLP	Inherited	Frameshift	NM_001331.3: c.1013_1014insT; NP_001322.1: p.Leu339Phefs*5
IRF6	Asian	CLP	*De Novo*	Missense	NM_006147.4: c.251G>A; NP_006138.1: p.Arg84His
ARHGAP29	Asian	CLP	Inherited	Stopgain	NM_004815.4: c.2392C>T; NP_004806.3: p.Arg798*
RYR1	Asian	CLP	Inherited	Splice	NM_000540.3: c.4934+1G>A
SEPT9	Asian	CLP	Inherited	Missense	NM_006640.5: c.1142G>A; NP_006631.2: p.Arg381His
PTCH1	Asian	CLP	*De Novo*	Frameshift	NM_000264.5: c.2011del; NP_000255.2: p.His671Thrfs*22
ARHGAP29	European	CLP	unknown	Stopgain	NM_004815.4: c.217G>T; NP_004806.3: p.Glu73*
MIB1	European	CLP	Inherited	Frameshift	NM_020774.4: c.602_603insT; NP_065825.1: p.Lys202*
COL2A1	European	CP	unknown	Stopgain	NM_001844.5: c.625C>T; NP_001835.3: p.Arg209*
NEDD4L	European	CP	*De Novo*	Missense	NM_001144964.1: c.391C>G; NP_001138436.1: p.Arg131Gly
CDH1	European	CLP	Inherited	Missense	NM_004360.5: c.895G>A; NP_004351.1: p.Ala299Thr
PIEZO2	European	CLP	Inherited	Stopgain	NM_022068.4: c.3412C>T; NP_071351.2: p.Arg1138*
CHD7	European	CLP	unknown	Frameshift	NM_017780.4: c.2500_2504del; NP_060250.2: p.Tyr835Serfs*14
FGFR1	European	CLP	*De Novo*	Missense	NM_001174063.2: c.1992G>C; NP_001167534.1: p.Trp664Cys
ARHGAP29	European	CLP	Inherited	Frameshift	NM_004815.4: c.2430_2431insG; NP_004806.3: p.Pro811Alafs*6
IRF6	European	CLP	*De Novo*	Missense	NM_006147.4: c.250C>T; NP_006138.1: p.Arg84Cys
CTNND1	European	CLP	Inherited	Frameshift	NM_001331.3: c.585_586insC; NP_001322.1: p.Gln196Profs*9
SATB2	European	CLP	*De Novo*	Missense	NM_015265.4: c.1999C>G; NP_056080.1: p.Arg667Gly
TBX2	European	CLP	Inherited	Missense	NM_005994.4: c.1822C>T; NP_005985.3: p.Arg608Trp
NOTCH1	European	CLP	Inherited	Missense	NM_017617.5: c.1981G>A; NP_060087.3: p.Gly661Ser
CHD7	European	CLP	unknown	Missense	NM_017780.4: c.3082A>G; NP_060250.2: p.Ile1028Val
PRICKLE1	European	CL	*De Novo*	Missense	NM_001144881.2: c.715T>C; NP_001138353.1: p.Cys239Arg
TBX2	European	CL	Inherited	Missense	NM_005994.4: c.949G>A; NP_005985.3: p.Gly317Ser
COL2A1	European	CP	Inherited	Splice	NM_001844.5: c.2355+1G>A
COL2A1	European	CP	*De Novo*	Stopgain	NM_001844.5: c.2659C>T; NP_001835.3: p.Arg887*
MYH3	European	CP	Inherited	Stopgain	NM_002470.4: c.4483C>T; NP_002461.2: p.Gln1495*

RPL5	European	CP	*De Novo*	Frameshift	NM_000969.5: c.45_46del; NP_000960.2: p.Tyr16Profs*5
IRF6	European	CP	*De Novo*	Missense	NM_006147.4: c.1127G>T; NP_006138.1: p.Gly376Val
COL2A1	European	CP	Inherited	Frameshift	NM_001844.5: c.2858del; NP_001835.3: p.Pro953Leufs*75
IRF6	European	CP	Inherited	Missense	NM_006147.4: c.17G>A; NP_006138.1: p.Arg6His
COL2A1	European	CP	Inherited	Stopgain	NM_001844.5: c.2818C>T; NP_001835.3: p.Arg940*
TFAP2A	Asian	CLP	*De Novo*	Missense	NM_003220.: c.740C>T; ENSP00000417495.1: p.Ser247Leu
TFAP2A	Asian	CLP	*De Novo*	Frameshift	NM_003220.: c.434delG; ENSP00000417495.1: p.Gly145Glufs*18
TFAP2A	Latino	CLP	*De Novo*	Stopgain	NM_003220.: c.310G>T; ENSP00000417495.1: p.Glu104*
TFAP2A	European	CLP	Inherited	Frameshift	NM_003220.: c.984_985insCGATCCC; p.Asn329Argfs*4
SIX3	Latino	CLP	Inherited	SV	SV
ESRP2	Latino	CLP	Inherited	SV	SV
BMP4	Asian	CLP	*De Novo*	SV	SV
TFAP2A	European	CLP	*De Novo*	SV	SV
TBX1	European	CLP	*De Novo*	SV	SV
TBX1	European	CP	*De Novo*	SV	SV
TBX1	European	CLP	*De Novo*	SV	SV
CTNND1	Latino	CLP	*De Novo*	SV	SV
SKI	Asian	CLP	*De Novo*	SV	SV
PDGFC	Latino	CLP	*De Novo*	SV	SV
PDGFC	European	CLP	Inherited	SV	SV
PDGFC	European	CP	unknown	SV	SV
NIPBL	European	CP	*De Novo*	Missense	NM_133433.4: c.8080T>A; NP_597677.2: p.Ser2694Thr
MAGEL2	Control	Control	Inherited	Frameshift	NM_019066.5: c.93_103del; NP_061939.3: p.Ala32Phefs*15
WDR11	Control	Control	Inherited	Missense	NM_018117.12: c.1575T>G; NP_060587.8: p.Ser525Arg
NOTCH1	Control	Control	Inherited	Missense	NM_017617.5: c.1981G>A; NP_060087.3: p.Gly661Ser
FLNA	Asian	CLP	*De Novo*	In-frame Deletion	NM_001110556.2: c.3266_3268del; NP_001104026.1: p.Thr1089del
PHF8	Latino	CLP	Inherited	Frameshift	NM_015107.3: c.2925_2926insC; NP_055922.1: p.Gly976Argfs*113

Table S4-7: VUS Comparisons by Population and Cleft Type

	Percentage (%) of Individuals with VUS	P-Value (vs. Control)
Cases	56.6	0.004
Controls	46.6	-
Population		
Asian	62.8	0.004
European	55.2	0.027
Latino	56.2	0.027
Cleft Type		
Cleft Lip	56.1	0.117
Cleft Lip & Palate	57.3	0.003
Cleft Palate	51.4	0.548

Table S4-8: Yield by Cleft Type and Source Gene List

	Gene Panel	NHS	OMIM	Prev. Genetics
CL	0.9	0.9	1.9	1.9
CLP	7.7	7.4	5.8	7.7
CP	16.2	14.9	16.2	14.9
All OFCs	7.6	7.3	6.2	7.6

*** P-values for all comparisons > 0.2**

Supplementary Figures

Figure S4-1: OFC Gene List. (A) 418 genes associated with OFCs and craniofacial development obtained from four sources: the Online Mendelian Inheritance in Man (OMIM), the National Health Service (NHS), the Developmental Disorders Genotype to Phenotype (DDG2P), PreventionGenetics Cleft Lip & Palate Panel (PGCP), and a manually curated gene panel (GP). The overlap between sources is shown by the connecting lines on the lower panel while the number of genes in each overlap and individual source is shown by the gray bars. (B) Sequencing coverage of the OFC genes (circles) by study population: Controls (pink), Asian (green), Latino (blue), and European (purple).

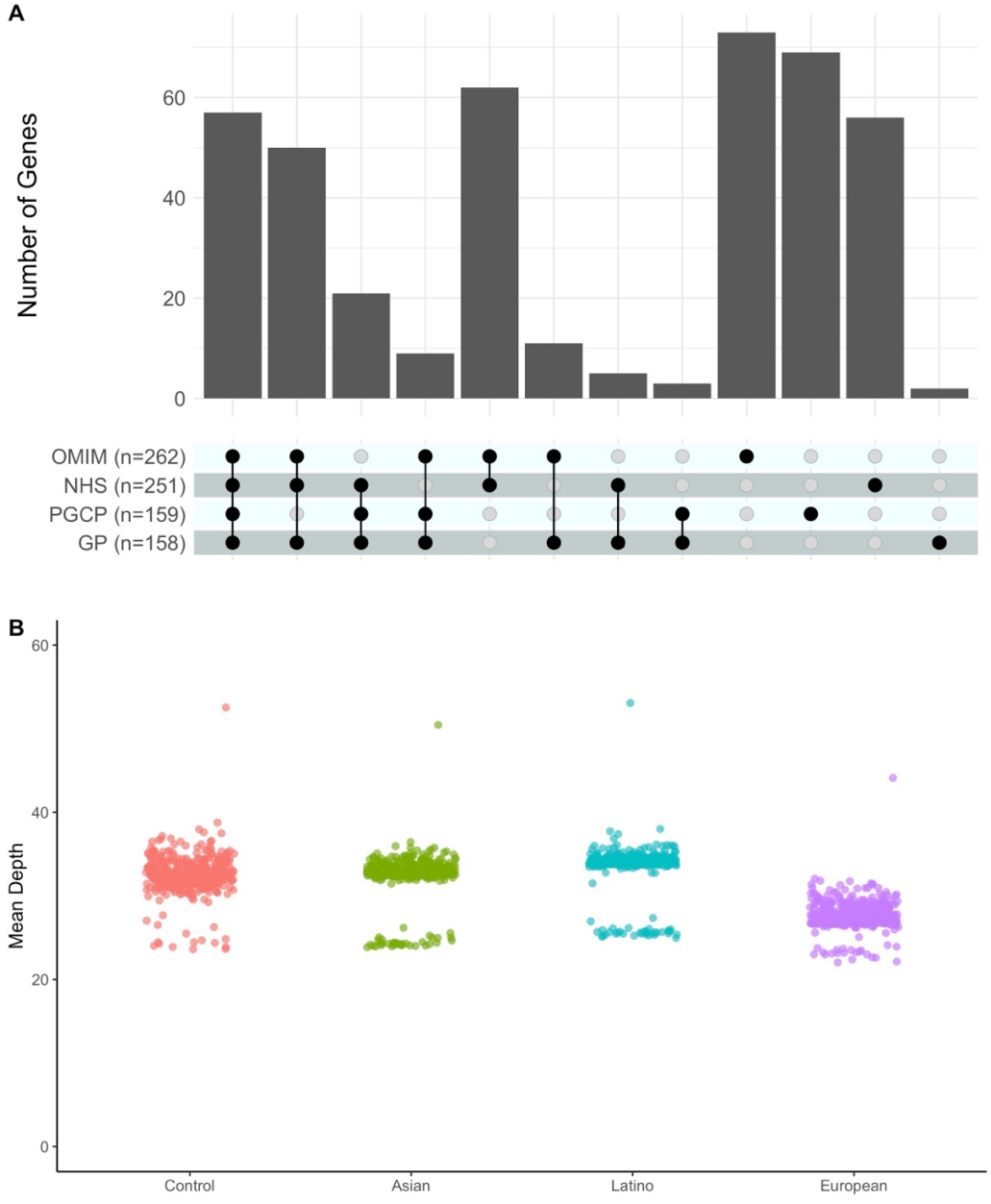


Figure S4-2: Whole-Genome Sequencing and Variant Filtering Pipeline. (A) Pipeline for prioritizing and sorting variants (B) Percentage of pathogenic (black) and benign (gray) ClinVar variants in genes from the OFC gene list at each tier threshold. (C) The number of variants from 841 OFC cases (red) and 294 controls (grey) sorted into tiers.
CADD: Combined Annotation Dependent Depletion; B/LB: 'Benign' or 'Likely Benign' Variants; SIFT: Sorting Intolerant from Tolerant; P/LP: 'Pathogenic' or 'Likely Pathogenic' Variants; pLoF: Predicted Loss-of-Function Variants

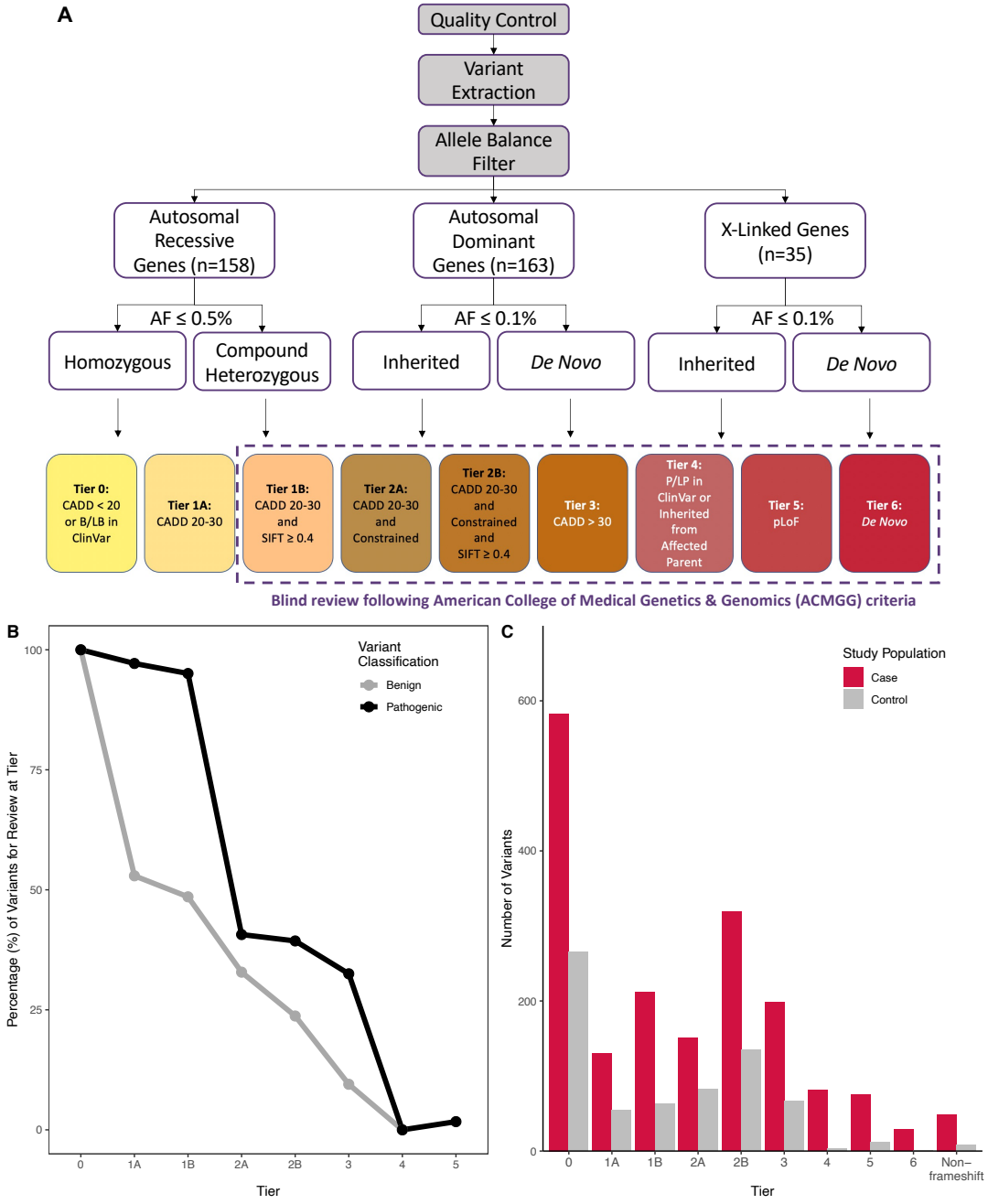


Figure S4-3: Diagnostic Yield by Cleft Type and Population. The sample size for each OFC subtype and population group is denoted below each bar.

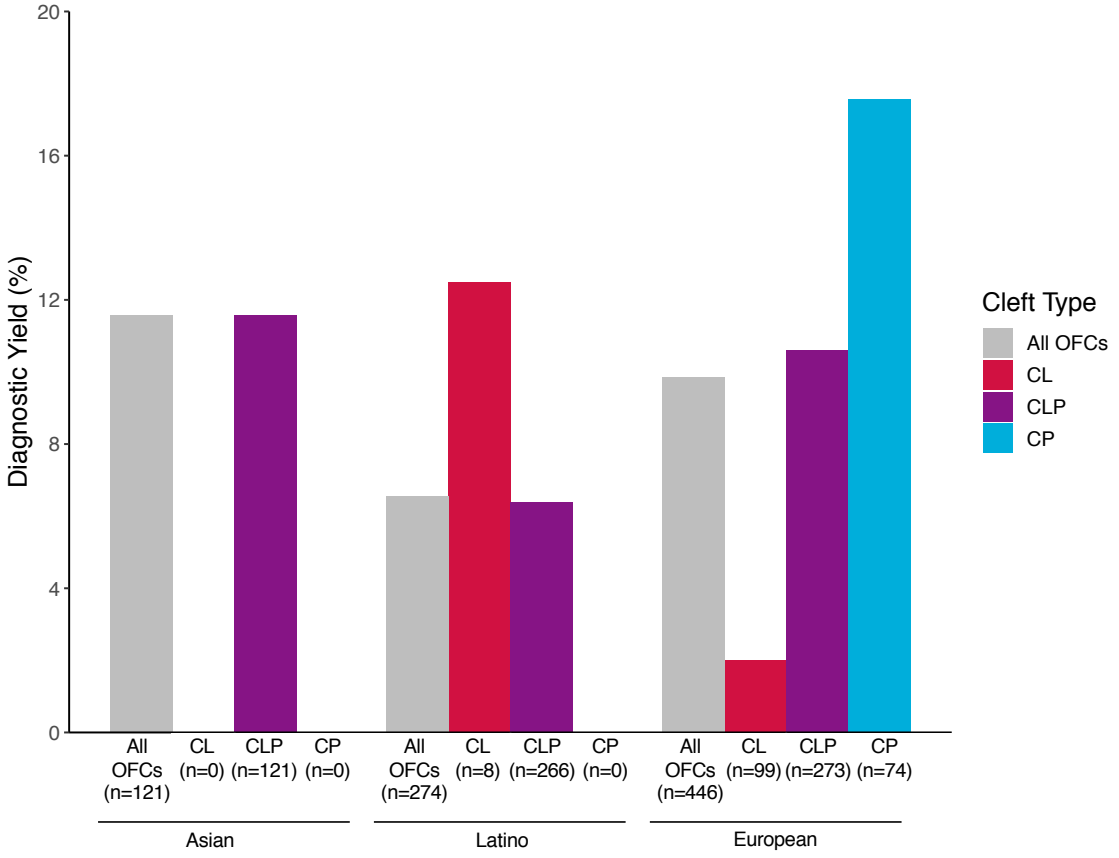


Figure S4-4: Diagnostic Yield by Family Type. The diagnostic yield from multiplex families (N=220) versus simplex families (N=621). 'Likely pathogenic' variants are classified by mode of inheritance: *de novo* (dark purple), transmitted from a parent (light purple), or unknown (gray).

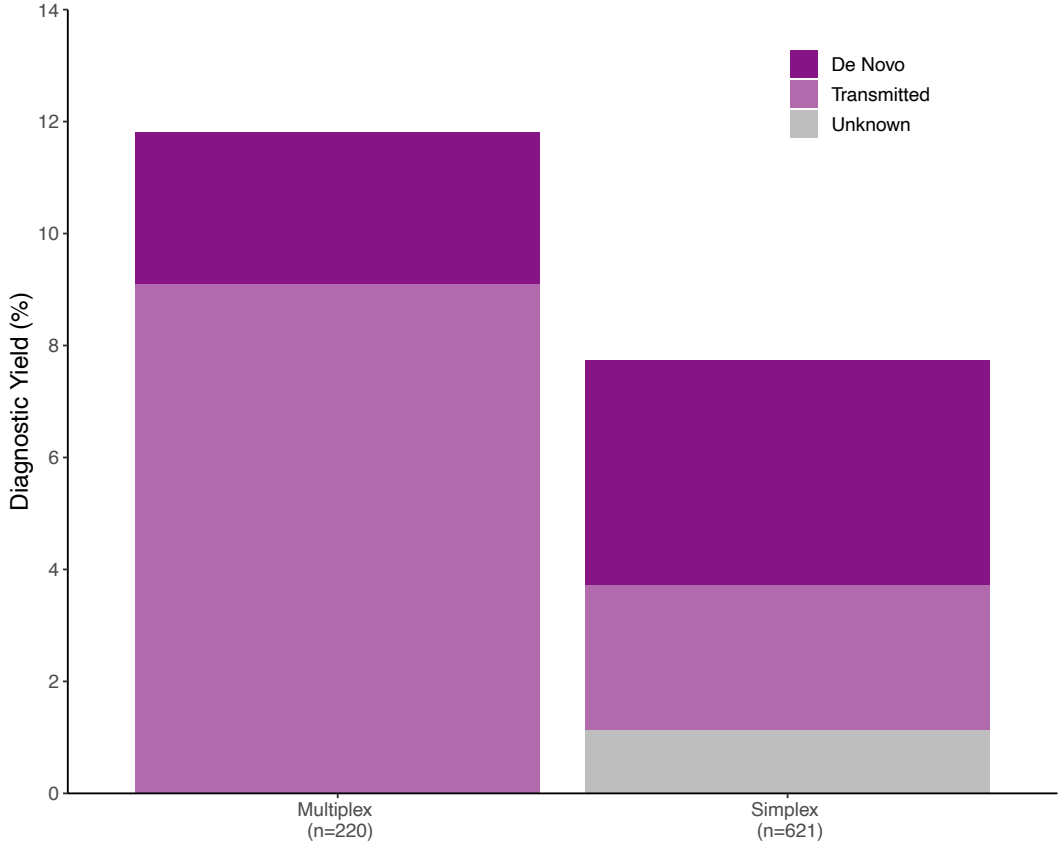


Figure S4-5: Inherited Likely Pathogenic Variants in Multiplex Families. The mutated gene and consequence are noted above each pedigree (refer to Table S4-6 for variant details). Sex symbols with solid black indicate the phenotype of the individual: CL (cleft lip), CP (cleft palate), and CLP (cleft lip and palate). The red solid lines below individuals indicate variant carriers while the blue outline of the sex symbols indicates individuals with WGS data.

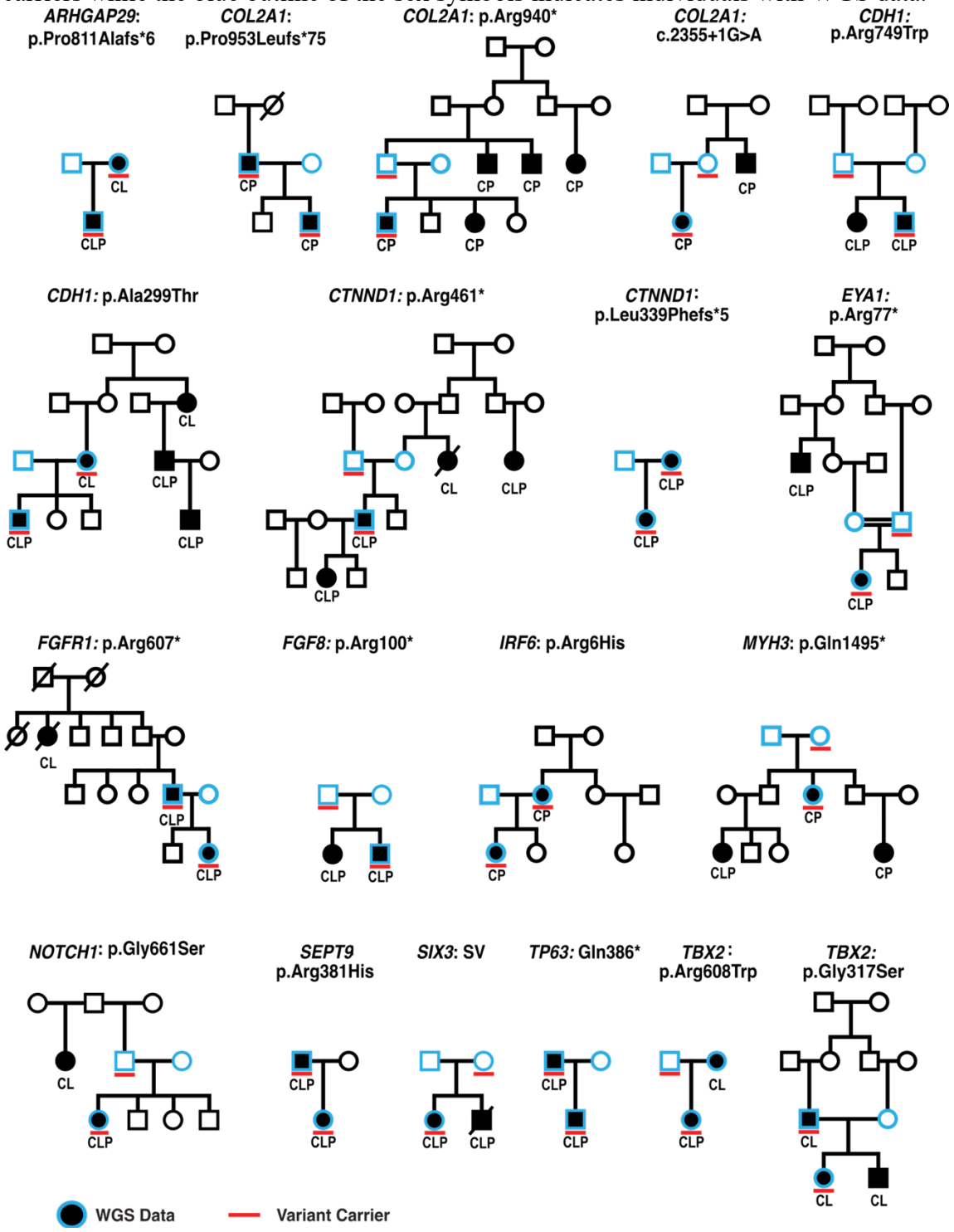


Figure S4-6: *De Novo* Mutations in Multiplex Families. (A) We identified 'likely pathogenic' *de novo* variants in *CDH1*, *COL11A1*, *PRICKLE1*, and *TFAP2A* (refer to Table S4-6 for variant details). We confirmed the *de novo* mutation by visual inspection of the proband (top segment), father (middle segment), and mother (bottom segment) reads using the Integrative Genomics Viewer (IGV) (Robinson *et al*., 2011). (B) We identified a *de novo* deletion encompassing the *TBX1*/22q deletion region. On the right, we the read depth ratio of the proband (red), parents (black), and controls (gray) of the SV region on the right. In each pedigree, the red solid lines below symbols indicate variant carriers while the blue outline corresponds to sequenced individuals.

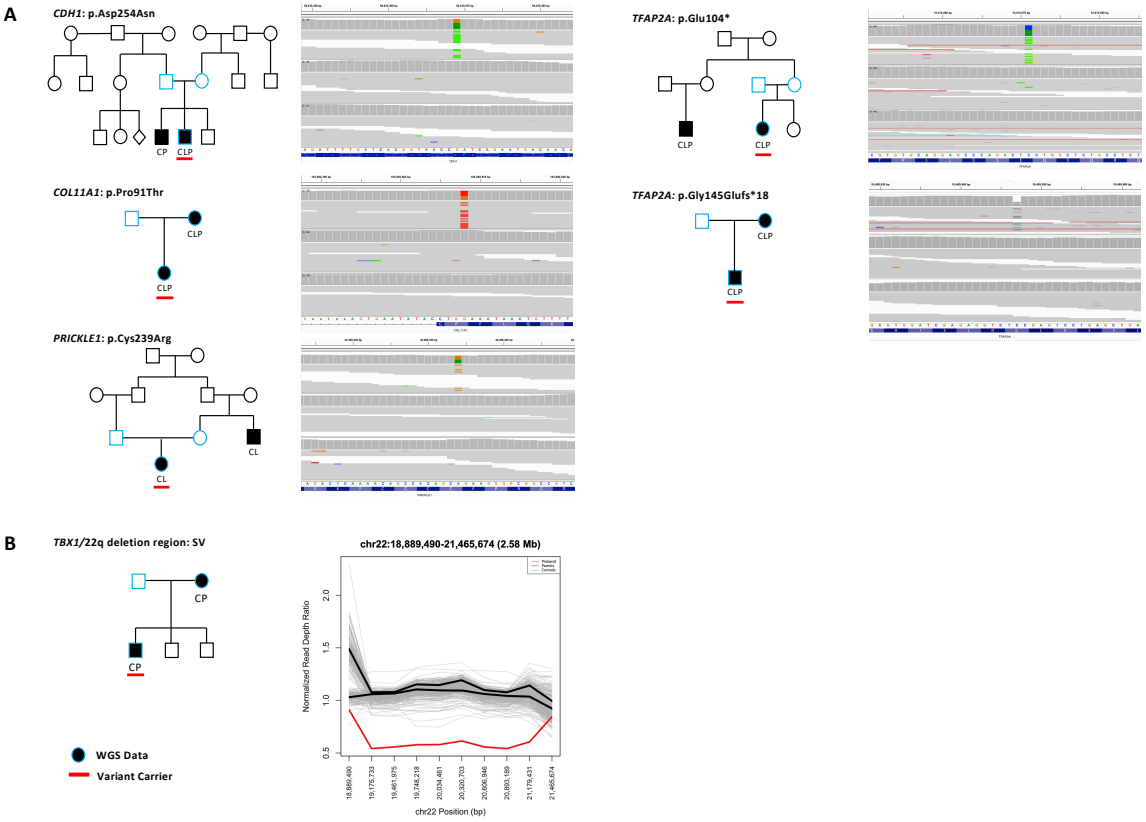


Figure S4-7: Variants of Uncertain Significance in Cases and Controls. (A) Distribution of the number of *in silico* prediction tools from nine different algorithms predicting a missense variant to be likely pathogenic/damaging for 'likely benign', VUS, and 'likely pathogenic' variants. (B) The number of VUS in genes is correlated with transcript length in cases (red, $p=1.28 \times 10^{-05}$) and controls (gray, $p=3.82 \times 10^{-03}$).

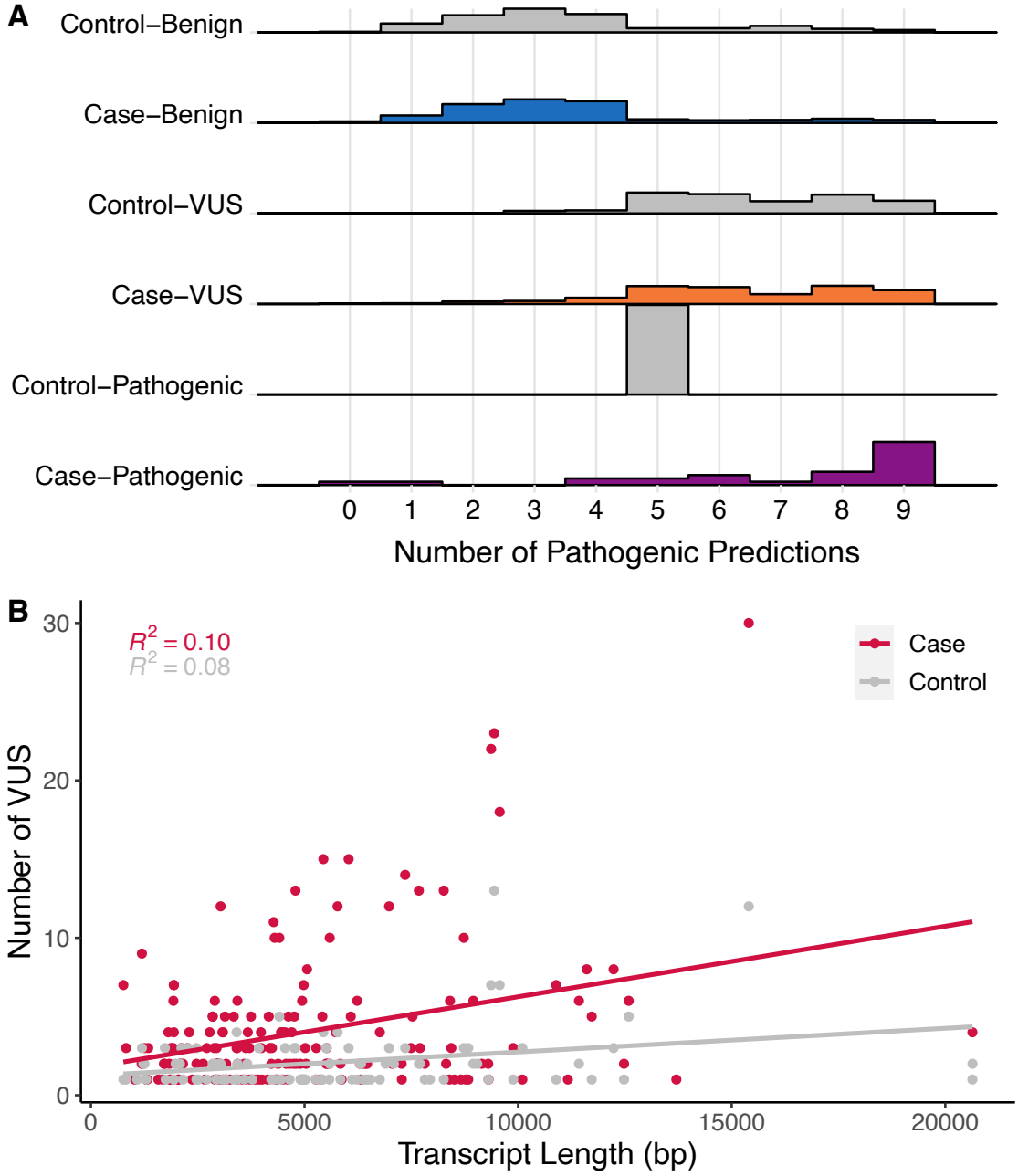
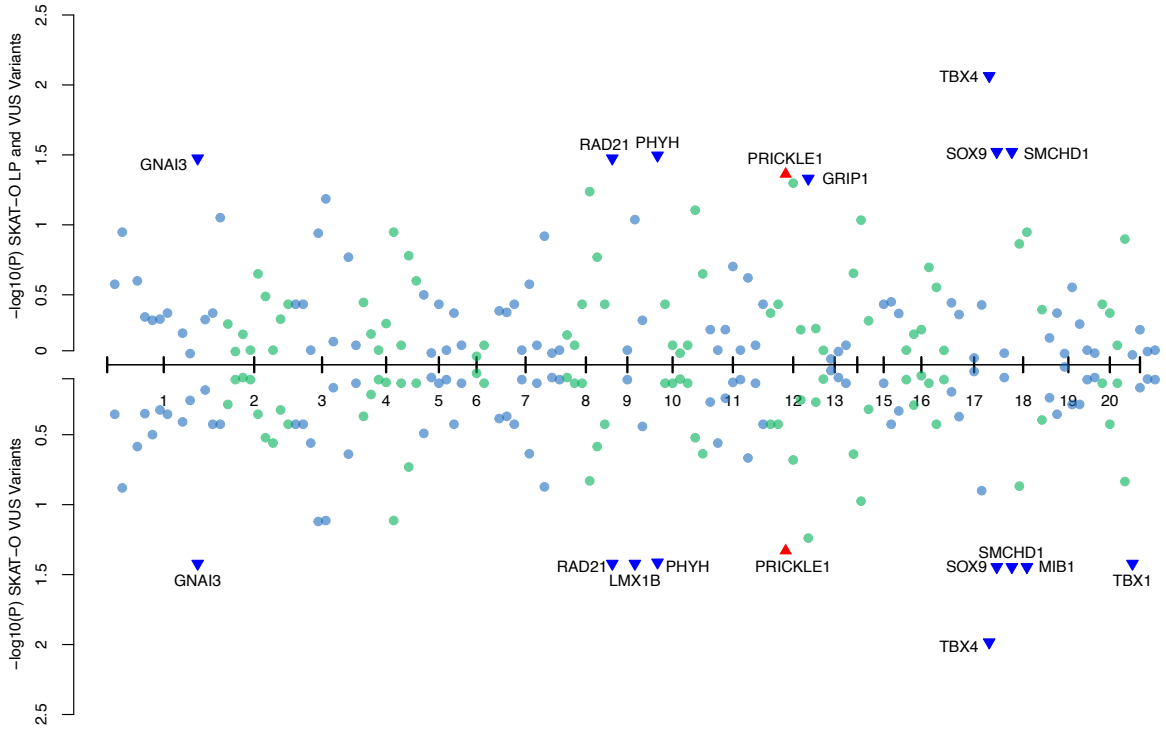


Figure S4-8: Gene-Based Association Tests of VUS. SKAT-O gene-based association tests for 139 genes with VUS and/or 'likely pathogenic' (LP) variants (top) and 129 genes with VUS variants (excluding individuals with 'likely pathogenic' variants) (bottom). Each gene with $p < 0.05$ is labeled according to the direction of effect with a triangle: decreased risk for cases (dark blue) and increased risk for cases (red). No gene reached a formal Bonferroni significance threshold ($p < 3.60 \times 10^{-4}$ (top) and $p < 3.88 \times 10^{-4}$ (bottom)).



References

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., . . . Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. Nature Methods, 7(4), 248-249. doi:10.1038/nmeth0410-248

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., . . . Abecasis, G. R. (2015). A global reference for human genetic variation. Nature, 526(7571), 68-74. doi:10.1038/nature15393

Basha, M., Demeer, B., Revencu, N., Helaers, R., Theys, S., Bou Saba, S., . . . Vikkula, M. (2018). Whole exome sequencing identifies mutations in 10% of patients with familial non-syndromic cleft lip and/or palate in genes mutated in well-known syndromes. J Med Genet, 55(7), 449-458. doi:10.1136/jmedgenet-2017-105110

Bishop, M. R., Diaz Perez, K. K., Sun, M., Ho, S., Chopra, P., Mukhopadhyay, N., . . . Leslie, E. J. (2020). Genome-wide Enrichment of De Novo Coding Mutations in Orofacial Cleft Trios. The American Journal of Human Genetics. doi:10.1016/j.ajhg.2020.05.018

Byrska-Bishop, M., Evani, U. S., Zhao, X., Basile, A. O., Abel, H. J., Regier, A. A., . . . Zody, M. C. (2021). High coverage whole genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. bioRxiv, 2021.2002.2006.430068. doi:10.1101/2021.02.06.430068

Carlson, J. C., Anand, D., Butali, A., Buxo, C. J., Christensen, K., Deleyiannis, F., . . . Leslie, E. J. (2019). A systematic genetic analysis and visualization of phenotypic heterogeneity among orofacial cleft GWAS signals. Genetic Epidemiology, 43(6), 704-716. doi:10.1002/gepi.22214

Carlson, J. C., Standley, J., Petrin, A., Shaffer, J. R., Butali, A., Buxó, C. J., . . . Leslie, E. J. (2017). Identification of 16q21 as a modifier of nonsyndromic orofacial cleft phenotypes. Genetic Epidemiology, 41(8), 887-897. doi:10.1002/gepi.22090

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., . . . Saunders, C. T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics, 32(8), 1220-1222. doi:10.1093/bioinformatics/btv710

Collins, R. L., Brand, H., Karczewski, K. J., Zhao, X., Alföldi, J., Francioli, L. C., . . . Talkowski, M. E. (2020). A structural variation reference for medical and population genetics. Nature, 581(7809), 444-451. doi:10.1038/s41586-020-2287-8

Cox, L. L., Cox, T. C., Moreno Uribe, L. M., Zhu, Y., Richter, C. T., Nidey, N., . . . Roscioli, T. (2018). Mutations in the Epithelial Cadherin-p120-Catenin Complex Cause Mendelian Non-Syndromic Cleft Lip with or without Cleft Palate. Am J Hum Genet, 102(6), 1143-1157. doi:10.1016/j.ajhg.2018.04.009

Curtis, S. W., Chang, D., Lee, M. K., Shaffer, J. R., Indencleef, K., Epstein, M. P., . . . Leslie, E. J. (2021). The PAX1 locus at 20p11 is a potential genetic modifier for bilateral cleft lip. HGG Adv, 2(2). doi:10.1016/j.xhgg.2021.100025

Grosen, D., Chevrier, C., Skytthe, A., Bille, C., Mølsted, K., Sivertsen, A., . . . 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(3), 162-168. doi:10.1136/jmg.2009.069385

Harville, E. W., Wilcox, A. J., Lie, R. T., Vindenes, H., & Abyholm, F. (2005). Cleft lip and palate versus cleft lip only: are they distinct defects? American Journal of Epidemiology, 162(5), 448-453. doi:10.1093/aje/kwi214

Hathaway, J., Heliö, K., Saarinen, I., Tallila, J., Seppälä, E. H., Tuupanen, S., . . . Koskenvuo, J. (2021). Diagnostic yield of genetic testing in a heterogeneous cohort of 1376 HCM patients. BMC Cardiovascular Disorders, 21(1), 126. doi:10.1186/s12872-021-01927-5

Havrilla, J. M., Pedersen, B. S., Layer, R. M., & Quinlan, A. R. (2019). A map of constrained coding regions in the human genome. Nature Genetics, 51(1), 88-95. doi:10.1038/s41588-018-0294-6

Hoornaert, K. P., Vereecke, I., Dewinter, C., Rosenberg, T., Beemer, F. A., Leroy, J. G., . . . Mortier, G. R. (2010). Stickler syndrome caused by COL2A1 mutations: genotype-phenotype correlation in a series of 100 patients. European journal of human genetics : EJHG, 18(8), 872-880. doi:10.1038/ejhg.2010.23

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., . . . MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. bioRxiv, 531210. doi:10.1101/531210

Karczewski, K. J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D. M., Kavanagh, D., . . . MacArthur, D. G. (2017). The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Research, 45(D1), D840-d845. doi:10.1093/nar/gkw971

Khandelwal, K. D., van den Boogaard, M. H., Mehrem, S. L., Gebel, J., Fagerberg, C., van Beusekom, E., . . . van Bokhoven, H. (2019). Deletions and loss-of-function variants in TP63 associated with orofacial clefting. Eur J Hum Genet, 27(7), 1101-1112. doi:10.1038/s41431-019-0370-0

Kondo, S., Schutte, B. C., Richardson, R. J., Bjork, B. C., Knight, A. S., Watanabe, Y., . . . Murray, J. C. (2002). Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nature Genetics, 32(2), 285-289. doi:10.1038/ng985

Lace, B., Pajusalu, S., Livcane, D., Grinfelde, I., Akota, I., Mauliņa, I., . . . Inashkina, I. (2022). Monogenic Versus Multifactorial Inheritance in the Development of Isolated Cleft Palate: A Whole Genome Sequencing Study. Front Genet, 13, 828534. doi:10.3389/fgene.2022.828534

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., . . . Maglott, D. R. (2018). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Research, 46(D1), D1062-d1067. doi:10.1093/nar/gkx1153

Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., . . . Lin, X. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. American journal of human genetics, 91(2), 224-237. doi:10.1016/j.ajhg.2012.06.007

Leslie, E. J. (2022). Genetic models and approaches to study orofacial clefts. Oral Dis, 28(5), 1327-1338. doi:10.1111/odi.14109

Leslie, E. J., Koboldt, D. C., Kang, C. J., Ma, L., Hecht, J. T., Wehby, G. L., . . . Marazita, M. L. (2016). IRF6 mutation screening in non-syndromic orofacial clefting: analysis of 1521 families. Clinical Genetics, 90(1), 28-34. doi:10.1111/cge.12675

Leslie, E. J., Liu, H., Carlson, J. C., Shaffer, J. R., Feingold, E., Wehby, G., . . . Marazita, M. L. (2016). A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. American journal of human genetics, 98(4), 744-754. doi:10.1016/j.ajhg.2016.02.014

Lowther, C., Valkanas, E., Giordano, J. L., Wang, H. Z., Currall, B. B., O'Keefe, K., . . . Talkowski, M. E. (2020). Systematic evaluation of genome sequencing as a first-tier

diagnostic test for prenatal and pediatric disorders. bioRxiv, 2020.2008.2012.248526. doi:10.1101/2020.08.12.248526

Mangold, E., Böhmer, A. C., Ishorst, N., Hoebel, A. K., Gültepe, P., Schuenke, H., . . . Ludwig, K. U. (2016). Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate. Am J Hum Genet, 98(4), 755-762. doi:10.1016/j.ajhg.2016.02.013

Marazita, M. L. (2012). The evolution of human genetic studies of cleft lip and cleft palate. Annual review of genomics and human genetics, 13, 263-283. doi:10.1146/annurev-genom-090711-163729

Marazita, M. L., Lidral, A. C., Murray, J. C., Field, L. L., Maher, B. S., Goldstein McHenry, T., . . . Arcos-Burgos, M. (2009). Genome scan, fine-mapping, and candidate gene analysis of non-syndromic cleft lip with or without cleft palate reveals phenotype-specific differences in linkage and association results. Human Heredity, 68(3), 151-170. doi:10.1159/000224636

Martin, A. R., Williams, E., Foulger, R. E., Leigh, S., Daugherty, L. C., Niblock, O., . . . McDonagh, E. M. (2019). PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nature Genetics, 51(11), 1560-1565. doi:10.1038/s41588-019-0528-2

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297-1303. doi:10.1101/gr.107524.110

Mone, F., Eberhardt, R. Y., Morris, R. K., Hurles, M. E., McMullan, D. J., Maher, E. R., . . . Kilby, M. D. (2021). COngenital heart disease and the Diagnostic yield with Exome sequencing (CODE) study: prospective cohort study and systematic review. Ultrasound in Obstetrics and Gynecology, 57(1), 43-51. doi:10.1002/uog.22072

Mossey, P. (2007). Epidemiology underpinning research in the aetiology of orofacial clefts. Orthodontics & Craniofacial Research, 10(3), 114-120. doi:10.1111/j.1601-6343.2007.00398.x

Mukhopadhyay, N., Feingold, E., Moreno-Uribe, L., Wehby, G., Valencia-Ramirez, L. C., Restrepo Muñeton, C. P., . . . Marazita, M. L. (2022). Genome-wide association study of multiethnic nonsyndromic orofacial cleft families identifies novel loci specific to family and phenotypic subtypes. Genetic Epidemiology. doi:10.1002/gepi.22447

Ng, J. K., Vats, P., Fritz-Waters, E., Sarkar, S., Sams, E. I., Padhi, E. M., . . . Turner, T. N. (2021). de novo variant calling identifies cancer mutation profiles in the 1000 Genomes Project. bioRxiv, 2021.2005.2027.445979. doi:10.1101/2021.05.27.445979

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Research, 31(13), 3812-3814. doi:10.1093/nar/gkg509

Posey, J. E., Harel, T., Liu, P., Rosenfeld, J. A., James, R. A., Coban Akdemir, Z. H., . . . Lupski, J. R. (2017). Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. The New England journal of medicine, 376(1), 21-31. doi:10.1056/NEJMoa1516767

Reiber, J., Sznajer, Y., Posteguillo, E. G., Müller, D., Lyonnet, S., Baumann, C., & Just, W. (2010). Additional clinical and molecular analyses of TFAP2A in patients with the branchio-oculo-facial syndrome. Am J Med Genet A, 152a(4), 994-999. doi:10.1002/ajmg.a.33331

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2018). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Research, 47(D1), D886-D894. doi:10.1093/nar/gky1016

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med, 17(5), 405-424. doi:10.1038/gim.2015.30

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. Nature Biotechnology, 29(1), 24-26. doi:10.1038/nbt.1754

Savastano, C. P., Brito, L. A., Faria Á, C., Setó-Salvia, N., Peskett, E., Musso, C. M., . . . Passos-Bueno, M. R. (2017). Impact of rare variants in ARHGAP29 to the etiology of oral clefts: role of loss-of-function vs missense variants. Clinical Genetics, 91(5), 683-689. doi:10.1111/cge.12823

Schwarz, J. M., Cooper, D. N., Schuelke, M., & Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. Nature Methods, 11(4), 361-362. doi:10.1038/nmeth.2890

Scott, T. M., Campbell, I. M., Hernandez-Garcia, A., Lalani, S. R., Liu, P., Shaw, C. A., . . . Scott, D. A. (2022). Clinical exome sequencing data reveal high diagnostic yields for congenital diaphragmatic hernia plus (CDH+) and new phenotypic expansions involving CDH. J Med Genet, 59(3), 270-278. doi:10.1136/jmedgenet-2020-107317

Seunggeun Lee, Z. Z., with contributions from Larisa Miropolsky and Michael Wu. (2020). SKAT: SNP-Set (Sequence) Kernel Association Test (Version R package version 2.0.1).

Sivertsen, A., Wilcox, A. J., Skjaerven, R., Vindenes, H. A., Abyholm, F., Harville, E., & Lie, R. T. (2008). Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives. BMJ (Clinical research ed.), 336(7641), 432-434. doi:10.1136/bmj.39458.563611.AE

Werling, D. M., Brand, H., An, J. Y., Stone, M. R., Zhu, L., Glessner, J. T., . . . Sanders, S. J. (2018). An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. Nature Genetics, 50(5), 727-736. doi:10.1038/s41588-018-0107-y

Whiffin, N., Minikel, E., Walsh, R., O'Donnell-Luria, A. H., Karczewski, K., Ing, A. Y., . . . Ware, J. S. (2017). Using high-resolution variant frequencies to empower clinical genome interpretation. Genet Med, 19(10), 1151-1158. doi:10.1038/gim.2017.26

Yang, T., Jia, Z., Bryant-Pike, W., Chandrasekhar, A., Murray, J. C., Fritzsch, B., & Bassuk, A. G. (2014). Analysis of PRICKLE1 in human cleft palate and mouse development demonstrates rare and common variants involved in human malformations. Molecular genetics & genomic medicine, 2(2), 138-151. doi:10.1002/mgg3.53

CHAPTER V. Discussion

Summary

Structural birth defects are collectively the leading cause of infant mortality, but the causes of many birth defects remain unaccounted for (Khokha *et al*., 2017). Genetic risk factors in orofacial clefts (OFCs), a common congenital craniofacial defect, have been previously investigated through linkage, candidate gene studies, and genome-wide association studies (GWASs) (Beaty *et al*., 2016; Leslie, 2022). However, these studies have not accounted for all of the heritable risks for OFCs and the genetic architecture of OFCs remains incompletely understood. In this work, we investigated the contribution of rare genetic variants to OFC risk using whole-exome and whole-genome sequencing. This work also addresses the etiologic and phenotypic heterogeneity that exists in OFCs. In Chapter Two, we analyzed single families with multiple affected individuals ("multiplex" families) with whole-exome sequencing. We identified likely causal rare variants in 25% of multiplex families sequenced. In Chapter Three, we analyzed consanguineous OFC families to explore the contribution of homozygous rare variants and found that 16% of probands had homozygous variants of interest in genes associated with autosomal recessive disorders involving facial dysmorphism or craniofacial development. Lastly, in Chapter Four, we expanded on the work described in Chapters Two and Three by estimating the proportion of OFC cases carrying likely pathogenic variants in a large cohort of 841 cases and 294 controls. We found 9.04% of OFC cases carried a likely pathogenic variant in one of 418 OFC-associated genes but that yield varied significantly by cleft type. More than half of the diagnostic yield was accounted for by variants in nine genes. Taken together, our findings suggest that rare variants are significant contributors to the etiology of non-syndromic OFCs, elucidating part of the "missing" heritability in OFCs. This work motivates more widespread use

of clinical genetic testing in OFCs, which could improve genetic counseling, family planning, and clinical management.

Diagnostic Yield Comparisons Across Studies

Previous studies have evaluated the utility of mutation screenings of single genes or sets of genes for potential clinical diagnostics applications in non-syndromic OFCs. For example, Leslie, E. J. *et al*. (2016) conducted a screening and meta-analysis to determine the prevalence of IRF6 mutations in non-syndromic OFC cases. However, only ~0.3% of individuals carried mutations, suggesting that testing this one gene would not account for a large enough fraction of non-syndromic cases to call for changing testing recommendations. Exome-sequencing studies in multiplex families focused on selected gene sets (ranging from 348-500 genes) found ~9-10% of non-syndromic OFC cases had a likely causal mutation (Basha *et al*., 2018; Bureau *et al*., 2014). In this thesis, we interrogated the diagnostic yield of three cohorts of non-syndromic cases, each selected and analyzed using different approaches. One cohort was comprised of multiplex families, one consisted of case-parent trios (with or without family histories of OFCs), and one was a set of consanguineous families. We identified likely causal variants in 25% of multiplex families using an untargeted exome-sequencing approach versus 16% in consanguineous families, and 9% of all cases using a targeted set of 418 OFC-associated genes in Chapter Four. The yield from the multiplex families in Chapter Two was double the yield from multiplex families analyzed in Chapter Four (12%). This discrepancy may be because although the families in Chapter Four were multiplex, we only analyzed case-parent trios and did not benefit from the variant filtering advantages that come from sequencing multiple affected individuals within individual families, as we did in Chapter Two.

There was also a discrepancy in the yield from consanguineous families in Chapter Three, where we identified variants of interest in 16% of families but failed to identify any pathogenic recessive variants in Chapter Four. Because the datasets in these chapters completely overlap, the most plausible explanation is that these variants were "variants of uncertain significance" (VUS) and did not meet the stringent criteria required by the American College of Medical Genetics and Genomics to call a variant pathogenic. Our results from Chapter Four are the most comparable to estimates of ~9-10% from previous studies (Basha *et al*., 2018; Bureau *et al*., 2014). Our large cohort with whole-genome sequencing replicated the previous findings by Bureau *et al*. (2014), which was expected given the similar filtering strategies. Our untargeted exome-sequencing analysis more than doubled these yields, which could be partly due to having a smaller study and highly selected study cohort. The yield for OFCs is still an unsettled question until comprehensive and consistent analyses of all qualifying exome (or, in the future, genome) variants are possible in large, unselected cohorts.

Clinical genetic testing typically involves some combination of tests, including karyotype, chromosomal microarray, gene or panel sequencing, and whole-exome or -genome sequencing. The type of genetic testing offered is dependent on factors such as an individual's clinical presentation, such as the severity of the cleft or the presence of additional structural or developmental anomalies, family history, and insurance coverage. Recommendations for implementing genetic medicine more broadly for OFCs will require assessing the yield from first-tier diagnostic tests in a variety of settings as there is no single standard of care for OFCs and diagnostic approaches vary by clinicians and institutions. It can be difficult to assess or compare the diagnostic yield for research-based cohorts because of differences in recruitment. Research participants may be more likely to be individuals or families not offered genetic testing

or those who received negative or inconclusive results for genetic tests (Carlock *et al*., 2022). In other birth defects, the diagnostic yield from exome sequencing has been estimated by sequentially testing cohorts with karyotype, microarray, and sequencing (Fu *et al*., 2018; Mone *et al*., 2021; Petrovski *et al*., 2019). Others have estimated the possible yield if whole-genome sequencing was the only diagnostic test by inferring that whole-genome sequencing can detect the same rearrangements detectable by other technologies (Lowther *et al*., 2020). Collectively, these studies find that ~10-24% of patients receive a diagnosis by sequencing after a negative karyotype and/or microarray. Although there are few equivalent studies published for OFCs, the yields from this thesis suggest that there may be a similar added value of exome or panel sequencing for OFCs. Future studies with larger and/or prospective cohorts will be needed to replicate and refine these yield estimations to develop guidelines for genetic testing in OFCs.

Genetic Heterogeneity of OFCs

There are hundreds of syndromes in which OFCs are a feature, which is evidence of substantial genetic heterogeneity. However, most of these syndromes are rare and there was no quantitative measure of the extent of the genetic heterogeneity for apparently non-syndromic OFCs. However, our study provides support for the role of many genes contributing to the risk of OFCs. In comparing the genes with dominant-acting likely causal variants in the independent cohorts from Chapter Two and Chapter Four, there were three genes (*IRF6*, *TP63*, *SHROOM3*) with likely causal rare variants in both analyses. Although there is overwhelmingly strong evidence that variants in and around *IRF6* and *TP63* increase the risk for OFCs, less is known about *SHROOM3*. Although all of the genes analyzed in Chapter Four are known to be involved in OFC risk in some way, our analysis gives some insight into the relative importance of some

genes versus others. Our results suggest that larger gene panels (and eventually exome), rather than single gene tests, are going to be most appropriate for genetic testing of OFCs. This work provides further evidence of the etiological overlap between syndromic and non-syndromic OFCs. Previous reports identified causal rare variants in ostensibly non-syndromic OFC cases within genes known to cause Mendelian OFC syndromes, including *TP63*, *CDH1*, and *CTNND1* (Basha *et al*., 2018; Cox *et al*., 2018; Khandelwal *et al*., 2019). We identified variants in these same genes as well as 36 others. Our findings support that the Mendelian transmission of variants in many genes can be found in cohorts enriched for non-syndromic OFCs. However, such variants still account for less than a quarter of the cohort. Taken together, these results suggest that a multifactorial etiology is one of several genetic mechanisms to explain the occurrence of non-syndromic OFCs.

Genetic Architecture of Cleft Subtypes and Subclinical Phenotypes

We explored differences in genetic etiology between specific cleft types. Previous genetic studies of OFCs primarily focused on cleft lip with or without cleft palate (CL/P) (Beaty *et al*., 2010; Birnbaum *et al*., 2009; Dixon *et al*., 2011; Ludwig *et al*., 2012). Although it is clear that some genetic risk factors are associated with both CL and CLP, combining these two subtypes could mask risk factors that are specific to only one subtype. In addition, fewer studies have studied cleft palate (CP) only (Beaty *et al*., 2011; Butali *et al*., 2018; Carlson *et al*., 2019; Leslie, Elizabeth J. *et al*., 2016) or CL and CLP as separate phenotypes using common and rare variants (Carlson *et al*., 2019; Huang *et al*., 2019; Marazita *et al*., 2009; Rahimov *et al*., 2008; Yu *et al*., 2017). Our study builds on previous work by exploring differences in cleft subtypes when considering rare and *de novo* variants. First, we found the highest percentage (18%) of

individuals with pathogenic variants in CP. These findings are similar to a recent finding by Lace *et al*. (2022) estimating a yield of 17% in non-syndromic CP based on a list of 198 genes. These findings along with Lace *et al*. (2022) support genetic testing for non-syndromic CP cases; genetic testing for all CP cases is the standard practice at some institutions (Eric Liao, personal communication). Studies such as these continue to build the evidence base needed to make formal recommendations. Second, our analyses revealed a significant difference in diagnostic yields for CL (3%) and CLP (9%), which mirrors evidence from GWASs that there are differences between CL and CLP (Carlson *et al*., 2019; Curtis, Chang, Lee, *et al*., 2021; Curtis, Chang, Sun, *et al*., 2021; Rahimov *et al*., 2008). Collectively, rare and common variants studies support differences between CL and CLP, but it remains to be seen if the genetic architecture of CL is different from CLP with respect to the relative contribution of *de novo* and rare coding variants versus common variants or if the architecture is similar but the genes responsible are simply different. Further studies in larger CL cohorts are needed to address this question and determine the utility of clinical diagnostics for CL.

Chapter Two was one of only a few studies in which OOM defects or VPI have been included with OFCs in studies of rare variants. Suzuki *et al*. (2009) found that rare variants in BMP4 were associated with a broad definition of "affected" that included cleft lip, microform cleft lip, or OOM defects. Some variants were transmitted from a parent with an OOM defect, providing part of the rationale for our study. Although we identified four families with subclinical phenotypes that had likely causal variants, only two families had variants that segregated cleanly among both individuals with overt clefts and those with subclinical phenotypes, and we did not observe any variants in *BMP4*. Our results provided only modest

evidence supporting a shared etiology between OFCs and subclinical phenotypes and more research is needed to fully assess the shared etiology hypothesis.

Future Directions

We assessed the potential pathogenicity of variants using purely computational predictions, and many variants were considered VUS. These were overwhelmingly missense variants and in-frame insertions or deletions, which are prevalent in the genome and more challenging to interpret computationally compared to predicted loss-of-function variants. The performance of *in silico* predictions varies across computational tools depending on their base model and purpose (evolutionary conservation, statistical, or machine-learning algorithm models), which can result in false positives (Luo *et al*., 2019; Thusberg *et al*., 2011). Computational predictions can also be limited by the availability of public functional, biological, and structural data (Katsonis *et al*., 2022). Because of these limitations, functional validation is considered a stronger piece of pathogenicity evidence than *in silico* predictions when performing variant classification using the American College of Medical Genetics and Genomics guidelines (Richards *et al*., 2015). Our analysis shows that these VUS are enriched in cases compared to controls and it is likely that a subset of these has a functional effect on the protein that would benefit from validation in model systems. A study by Li *et al*. (2017) found that *in silico* predictions (e.g., SIFT and PolyPhen-2) for variants in *IRF6* did not consistently predict the biological function of missense variants. Therefore, it will be important to functionally validate rare variants to understand their contribution to OFCs.

Our study focused solely on the protein-coding regions of the genome, which only account for 1-2% of the human genome. Other sources of variation not covered in this study

likely harbor critical risk variants. For instance, variants in 5'UTRs can create upstream start codon sequences, resulting in protein translation repression (Whiffin *et al*., 2020), and variants within 3'UTRs can impact miRNA binding sites and mRNA stability (Griesemer *et al*., 2021; Hughes, 2006). Variants in non-coding regions are particularly challenging to interpret because we lack the tools and knowledge necessary to comprehensively annotate their location within genomic features (e.g., transcription factor binding sites, enhancers) or regulatory units (e.g., topologically associated domains). Low-frequency, rare, and *de novo* non-coding variants in craniofacial enhancers have all been associated with the risk for OFCs (Morris *et al*., 2020; Shaffer *et al*., 2019; Zieger *et al*., 2023). Being able to clearly detect and interpret these types of variants will be necessary as whole-genome sequencing will ultimately become commonplace in clinical genetics.

Lastly, future research should seek to establish larger and prospectively recruited study cohorts with broad eligibility criteria that represent the diversity of populations affected by OFCs. Deep phenotyping of study participants will be vital to help identify genotype-phenotype correlations, refine mechanisms of cleft pathogenesis, and capture other associated health outcomes that may arise over time. Having a detailed and holistic phenotyping approach may facilitate the grouping of homogeneous sets of OFC cases and improve the power for gene discovery. Deep phenotyping of both affected and unaffected individuals in pedigrees will also be needed to understand reduced penetrance and variable expressivity.

Conclusion

In conclusion, this work generated new insights into the role of rare genetic variants in the etiology of OFCs. We uncovered the extensive genetic heterogeneity of OFCs and provided

evidence that at least a subset of individuals, especially those with CP, might benefit from clinical genetic testing. Future research in OFCs needs to involve high-throughput functional validation analyses to determine the effect of the rare variants identified in this study. Future work should also develop larger study cohorts with deep phenotyping to fully assess the genetic architecture of OFCs.

References

Basha, M., Demeer, B., Revencu, N., Helaers, R., Theys, S., Bou Saba, S., . . . Vikkula, M. (2018). Whole exome sequencing identifies mutations in 10% of patients with familial non-syndromic cleft lip and/or palate in genes mutated in well-known syndromes. J Med Genet, 55(7), 449-458. doi:10.1136/jmedgenet-2017-105110

Beaty, T., Marazita, M., & Leslie, E. (2016). Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities [version 1; peer review: 2 approved]. F1000Research, 5(2800). doi:10.12688/f1000research.9503.1

Beaty, T. H., Murray, J. C., Marazita, M. L., Munger, R. G., Ruczinski, I., Hetmanski, J. B., . . . Scott, A. F. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nature Genetics, 42(6), 525-529. doi:10.1038/ng.580

Beaty, T. H., Ruczinski, I., Murray, J. C., Marazita, M. L., Munger, R. G., Hetmanski, J. B., . . . Scott, A. F. (2011). Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. Genetic Epidemiology, 35(6), 469-478. doi:10.1002/gepi.20595

Birnbaum, S., Ludwig, K. U., Reutter, H., Herms, S., Steffens, M., Rubini, M., . . . Mangold, E. (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nature Genetics, 41(4), 473-477. doi:10.1038/ng.333

Bureau, A., Parker, M. M., Ruczinski, I., Taub, M. A., Marazita, M. L., Murray, J. C., . . . Beaty, T. H. (2014). Whole exome sequencing of distant relatives in multiplex families implicates rare variants in candidate genes for oral clefts. Genetics, 197(3), 1039-1044. doi:10.1534/genetics.114.165225

Butali, A., Mossey, P. A., Adeyemo, W. L., Eshete, M. A., Gowans, L. J. J., Busch, T. D., . . . Adeyemo, A. A. (2018). Genomic analyses in African populations identify novel risk loci for cleft palate. Human Molecular Genetics, 28(6), 1038-1051. doi:10.1093/hmg/ddy402

Carlock, G., Manning, K., & Leslie, E. J. (2022). Feasibility of Social Media Recruitment for Orofacial Cleft Genetic Research. Cleft Palate Craniofac J, 59(6), 701-707. doi:10.1177/10556656211024484

Carlson, J. C., Anand, D., Butali, A., Buxo, C. J., Christensen, K., Deleyiannis, F., . . . Leslie, E. J. (2019). A systematic genetic analysis and visualization of phenotypic heterogeneity among orofacial cleft GWAS signals. Genetic Epidemiology, 43(6), 704-716. doi:10.1002/gepi.22214

Cox, L. L., Cox, T. C., Moreno Uribe, L. M., Zhu, Y., Richter, C. T., Nidey, N., . . . Roscioli, T. (2018). Mutations in the Epithelial Cadherin-p120-Catenin Complex Cause Mendelian Non-Syndromic Cleft Lip with or without Cleft Palate. Am J Hum Genet, 102(6), 1143-1157. doi:10.1016/j.ajhg.2018.04.009

Curtis, S. W., Chang, D., Lee, M. K., Shaffer, J. R., Indencleef, K., Epstein, M. P., . . . Leslie, E. J. (2021). The PAX1 locus at 20p11 is a potential genetic modifier for bilateral cleft lip. HGG Adv, 2(2). doi:10.1016/j.xhgg.2021.100025

Curtis, S. W., Chang, D., Sun, M. R., Epstein, M. P., Murray, J. C., Feingold, E., . . . Leslie, E. J. (2021). FAT4 identified as a potential modifier of orofacial cleft laterality. Genetic Epidemiology, 45(7), 721-735. doi:10.1002/gepi.22420

Dixon, M. J., Marazita, M. L., Beaty, T. H., & Murray, J. C. (2011). Cleft lip and palate: understanding genetic and environmental influences. Nature Reviews: Genetics, 12(3), 167-178. doi:10.1038/nrg2933

Fu, F., Li, R., Li, Y., Nie, Z. Q., Lei, T., Wang, D., . . . Liao, C. (2018). Whole exome sequencing as a diagnostic adjunct to clinical testing in fetuses with structural abnormalities. Ultrasound in Obstetrics and Gynecology, 51(4), 493-502. doi:10.1002/uog.18915

Griesemer, D., Xue, J. R., Reilly, S. K., Ulirsch, J. C., Kukreja, K., Davis, J. R., . . . Sabeti, P. C. (2021). Genome-wide functional screen of 3'UTR variants uncovers causal variants for human disease and evolution. Cell, 184(20), 5247-5260.e5219. doi:10.1016/j.cell.2021.08.025

Huang, L., Jia, Z., Shi, Y., Du, Q., Shi, J., Wang, Z., . . . Yang, Z. (2019). Genetic factors define CPO and CLO subtypes of nonsyndromicorofacial cleft. PLoS Genetics, 15(10), e1008357. doi:10.1371/journal.pgen.1008357

Hughes, T. A. (2006). Regulation of gene expression by alternative untranslated regions. Trends in Genetics, 22(3), 119-122. doi:10.1016/j.tig.2006.01.001

Katsonis, P., Wilhelm, K., Williams, A., & Lichtarge, O. (2022). Genome interpretation using in silico predictors of variant impact. Human Genetics, 141(10), 1549-1577. doi:10.1007/s00439-022-02457-6

Khandelwal, K. D., van den Boogaard, M. H., Mehrem, S. L., Gebel, J., Fagerberg, C., van Beusekom, E., . . . van Bokhoven, H. (2019). Deletions and loss-of-function variants in TP63 associated with orofacial clefting. Eur J Hum Genet, 27(7), 1101-1112. doi:10.1038/s41431-019-0370-0

Khokha, M. K., Mitchell, L. E., & Wallingford, J. B. (2017). White paper on the study of birth defects. Birth defects research, 109(2), 180-185. doi:10.1002/bdra.23590

Lace, B., Pajusalu, S., Livcane, D., Grinfelde, I., Akota, I., Mauliņa, I., . . . Inashkina, I. (2022). Monogenic Versus Multifactorial Inheritance in the Development of Isolated Cleft Palate: A Whole Genome Sequencing Study. Front Genet, 13, 828534. doi:10.3389/fgene.2022.828534

Leslie, E. J. (2022). Genetic models and approaches to study orofacial clefts. Oral Dis, 28(5), 1327-1338. doi:10.1111/odi.14109

Leslie, E. J., Koboldt, D. C., Kang, C. J., Ma, L., Hecht, J. T., Wehby, G. L., . . . Marazita, M. L. (2016). IRF6 mutation screening in non-syndromic orofacial clefting: analysis of 1521 families. Clinical Genetics, 90(1), 28-34. doi:10.1111/cge.12675

Leslie, E. J., Liu, H., Carlson, J. C., Shaffer, J. R., Feingold, E., Wehby, G., . . . Marazita, M. L. (2016). A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. American journal of human genetics, 98(4), 744-754. doi:10.1016/j.ajhg.2016.02.014

Li, E. B., Truong, D., Hallett, S. A., Mukherjee, K., Schutte, B. C., & Liao, E. C. (2017). Rapid functional analysis of computationally complex rare human IRF6 gene variants using a novel zebrafish model. PLoS Genetics, 13(9), e1007009-e1007009. doi:10.1371/journal.pgen.1007009

Lowther, C., Valkanas, E., Giordano, J. L., Wang, H. Z., Currall, B. B., O'Keefe, K., . . . Talkowski, M. E. (2020). Systematic evaluation of genome sequencing as a first-tier diagnostic test for prenatal and pediatric disorders. bioRxiv, 2020.2008.2012.248526. doi:10.1101/2020.08.12.248526

Ludwig, K. U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A., . . . Nöthen, M. M. (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nature Genetics, 44(9), 968-971. doi:10.1038/ng.2360

Luo, J., Zhou, T., You, X., Zi, Y., Li, X., Wu, Y., . . . Shi, Y. (2019). Assessing concordance among human, in silico predictions and functional assays on genetic variant classification. Bioinformatics, 35(24), 5163-5170. doi:10.1093/bioinformatics/btz442

Marazita, M. L., Lidral, A. C., Murray, J. C., Field, L. L., Maher, B. S., Goldstein McHenry, T., . . . Arcos-Burgos, M. (2009). Genome scan, fine-mapping, and candidate gene analysis of non-syndromic cleft lip with or without cleft palate reveals phenotype-specific differences in linkage and association results. Human Heredity, 68(3), 151-170. doi:10.1159/000224636

Mone, F., Eberhardt, R. Y., Morris, R. K., Hurles, M. E., McMullan, D. J., Maher, E. R., . . . Kilby, M. D. (2021). COngenital heart disease and the Diagnostic yield with Exome sequencing (CODE) study: prospective cohort study and systematic review. Ultrasound in Obstetrics and Gynecology, 57(1), 43-51. doi:10.1002/uog.22072

Morris, V. E., Hashmi, S. S., Zhu, L., Maili, L., Urbina, C., Blackwell, S., . . . Letra, A. (2020). Evidence for craniofacial enhancer variation underlying nonsyndromic cleft lip and palate. Human Genetics. doi:10.1007/s00439-020-02169-9

Petrovski, S., Aggarwal, V., Giordano, J. L., Stosic, M., Wou, K., Bier, L., . . . Wapner, R. J. (2019). Whole-exome sequencing in the evaluation of fetal structural anomalies: a prospective cohort study. Lancet, 393(10173), 758-767. doi:10.1016/s0140-6736(18)32042-7

Rahimov, F., Marazita, M. L., Visel, A., Cooper, M. E., Hitchler, M. J., Rubini, M., . . . Murray, J. C. (2008). Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. Nature Genetics, 40(11), 1341-1347. doi:10.1038/ng.242

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med, 17(5), 405-424. doi:10.1038/gim.2015.30

Shaffer, J. R., LeClair, J., Carlson, J. C., Feingold, E., Buxó, C. J., Christensen, K., . . . Leslie, E. J. (2019). Association of low-frequency genetic variants in regulatory regions with nonsyndromic orofacial clefts. Am J Med Genet A, 179(3), 467-474. doi:10.1002/ajmg.a.61002

Suzuki, S., Marazita, M. L., Cooper, M. E., Miwa, N., Hing, A., Jugessur, A., . . . Murray, J. C. (2009). Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. Am J Hum Genet, 84(3), 406-411. doi:10.1016/j.ajhg.2009.02.002

Thusberg, J., Olatubosun, A., & Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. Hum Mutat, 32(4), 358-368. doi:10.1002/humu.21445

Whiffin, N., Karczewski, K. J., Zhang, X., Chothani, S., Smith, M. J., Evans, D. G., . . . Genome Aggregation Database, C. (2020). Characterising the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. Nature Communications, 11(1), 2523. doi:10.1038/s41467-019-10717-9

Yu, Y., Zuo, X., He, M., Gao, J., Fu, Y., Qin, C., . . . Bian, Z. (2017). Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. Nature Communications, 8, 14364-14364. doi:10.1038/ncomms14364

Zieger, H. K., Weinhold, L., Schmidt, A., Holtgrewe, M., Juranek, S. A., Siewert, A., . . . Ludwig, K. U. (2023). Prioritization of non-coding elements involved in non-syndromic

cleft lip with/without cleft palate through genome-wide analysis of de novo mutations. HGG Adv, 4(1), 100166. doi:10.1016/j.xhgg.2022.100166