Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a 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 in whole or in part in all forms of media, now or hereafter now, 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. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Cinderella Yang April 10, 2024

Investigating the contribution of coding *de novo* mutations to orofacial clefts

by

Cinderella Yang

Dr. Elizabeth Leslie
Adviser

Biology

Dr. Elizabeth Leslie

Adviser

Dr. Andreas Fritz

Committee Member

Dr. Michael Epstein

Committee Member

2024

Investigating the contribution of coding *de novo* mutations to orofacial clefts

By

Cinderella Yang

Dr. Elizabeth Leslie

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2024

Abstract

Investigating the contribution of coding *de novo* mutations to orofacial clefts
By Cinderella Yang

Orofacial clefts (OFCs), including cleft lip and cleft palate, are among the most common birth defects, occurring in approximately 1 in 1000 live births worldwide. OFCs have a strong genetic basis, with genetic variants of all types and frequencies increasing risk for an OFC. However, much of the genetic risk remains unaccounted for. Previous studies have highlighted a role for *de novo* mutations (DNMs) in contributing to OFC risk. Here, we analyzed coding DNMs found in whole genome sequence (WGS) data from 419 child-parent trios of Filipino ancestry with OFCs from the Gabriella Miller Kids First (GMKF) Research Program. We found a significant enrichment of predicted loss-of-function (pLOF) ($p = 9.68 \times 10^{-3}$) and protein-altering DNMs ($p = 2.61 \times 10^{-4}$). We also found a significant enrichment of pLOF DNMs in *GRHL2* ($p = 4.74 \times 10^{-6}$), a gene intolerant to pLOF variants (pLI = 1). We then combined the Filipino trios with three other previously studied GMKF cohorts with OFCs of Colombian, Taiwanese, and European ancestries, totaling 1,114 trios. In this combined cohort, we found a similar significant enrichment of pLOF (1.57×10^{-4}) and protein-altering DNMs ($p = 2.55 \times 10^{-3}$). Genes with protein-altering DNMs were found to be enriched for pathways relevant to craniofacial development. We then analyzed *de novo* structural variants (SVs) in the cohort and identified additional SVs in multiple genes with protein-altering DNMs, including *GRHL2*. Overall, we have further demonstrated that DNMs contribute to OFC risk and identified several novel candidate genes such as *GRHL2*. Further investigations are required in order to interpret variants in these genes and their association with OFCs.

Investigating the contribution of coding *de novo* mutations to orofacial clefts

By

Cinderella Yang

Dr. Elizabeth Leslie

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2024

Acknowledgements

I would like to take this opportunity to thank my advisor Dr. Elizabeth Leslie, for her guidance throughout this thesis process. I also wish to extend my gratitude to my colleagues in the Leslie Lab for their advice and help. I would like to thank Dr. Sarah Curtis in particular, for the critical support and feedback she provided. I would also like to thank Dr. Andreas Fritz and Dr. Michael Epstein for serving on my thesis committee and providing their time and guidance throughout this process. Finally, I would like to acknowledge our collaborators and all the study participants, without which this study would not have been possible.

Table of Contents

Introduction 1

Methods 4

Results 8

Discussion 21

Reference List 25

Figures

Figure 1 8

Figure 2 10

Figure 3 11

Figure 4 13

Figure 5 15

Figure 6 16

Figure 7 17

Figure 8 19

Figure 9 20

Introduction

Orofacial clefts (OFCs) are a group of birth defects usually involving the roof of the mouth (palate), the upper lip, or both. Defects in these structures are called cleft palate (CP), cleft lip (CL), or cleft lip and palate (CLP), respectively. CL and CLP are often analyzed together as a group, referred to as cleft lip with or without cleft palate (CL/P), and considered distinct from CP due to different developmental origins, as lip development occurs before secondary palate closure during embryogenesis (Watkins et al. 2014). Furthermore, OFCs are typically classified as either syndromic or isolated/nonsyndromic, with around 70% of CL/P and 50% of CP cases appearing to be isolated/nonsyndromic. Together, OFCs are among the most common craniofacial malformations in the world, occurring in around 1 in 1000 live births worldwide, although prevalence has been found to vary greatly across populations (Watkins et al. 2014). Although OFCs can be surgically repaired, affected individuals experience problems with feeding, speech, and hearing early in life, and must often undergo multiple surgeries and speech and hearing therapy (Hocevar-Boltezar, Jarc, and Kozelj 2006, James et al. 2020). Individuals may also experience lifelong psychosocial effects, with an increased incidence of mental health issues compared to individuals without clefts (Hunt et al. 2006). Lower reported quality of life and higher mortality rates have also been found in patients (Christensen et al. 2004, Marcusson, Akerlind and Paulin 2001). Moreover, these negative impacts of OFCs on affected individuals may be particularly severe in areas where access to medical care is more limited (Mossey et al. 2011). Therefore, it is important to research the causes and risk factors of OFCs to improve prevention and treatment for individuals with OFCs.

OFCs are complex and have a multifactorial etiology. Environmental risk factors such as maternal smoking have been found to be associated with an increased risk for OFCs (Khoury et al. 1987, Honein et al. 2007). OFCs also show strong familial aggregation, with first-degree relatives of OFC-affected individuals showing greatly increased relative risk compared to the general population (Grosen et al. 2009, Sivertsen et al. 2008), suggesting a strong genetic basis to OFC etiology. Affected individuals within the same family are also more likely to have the same type of cleft, although this is not always the case, suggesting that there may be pathways shared across OFC types. Monozygotic twins have also been found to have a much higher concordance rate than dizygotic twins, further demonstrating the major contribution of genetic factors to OFC etiology. Therefore, fully understanding the etiology and development of OFCs would not be possible without investigation of the genetic risk factors.

Multiple studies on the genetics of OFCs have been carried out, including linkage and association analyses. Studies have found linkage to different chromosomal regions and further identified several causal genes (Marazita et al. 2009, Moreno et al. 2009) although the complex nature of OFCs means that is difficult to obtain consistent results. Genome-wide association studies (GWAS) have been conducted to investigate single nucleotide polymorphisms (SNPs) and their contributions to various nonsyndromic OFC subtypes. More than 50 genes are reported to be associated, however, GWAS loci are likely to account for only 10-30% of the heritable risk for OFCs in any one population, although differences between populations make this fraction difficult to quantify (Leslie 2022). Advances in sequencing technologies have now made it possible to conduct whole-exome sequencing (WES) or whole-genome sequencing (WGS) studies to identify rare variants that affect OFC risk of any kind, including single nucleotide variants, insertions or deletions, and large structural variants.

Family-based studies of WES and WGS data allow for the investigation of *de novo* mutations (DNMs), which are present in the offspring but not the parents, with germline DNMs arising in the germline of the parent or during early embryonic development. On average 50 to 100 DNMs are found throughout an individual's genome, with approximately 1 DNM located in the exome (Lynch 2010). Despite their rarity, DNMs are more likely to be deleterious than inherited variants since they are not subjected to the same kinds of selective pressures (Veltman and Brunner 2012). Thus, they represent a promising avenue of research for OFC development, especially considering the sporadic nature of many OFCs. By studying DNMs in individuals affected with OFCs, we may be able to identify novel candidate genes that contribute to OFC risk and gain new insight into the pathways underlying OFC etiology, particularly by focusing our analysis on the DNMs most likely to be deleterious: missense and loss of function coding DNMs. Previous studies have investigated coding DNMs in child-parent trios with OFCs and found a significant excess of loss of function DNMs in affected individuals, providing evidence for the contribution of DNMs to OFC risk (Bishop et al. 2020).

This study aims to further analyze the contribution of coding DNMs to OFCs using WGS data from a combined cohort of 1178 child-parent trios from multiple populations and OFC types. We will be focusing our analysis primarily on *de novo* single nucleotide variants or small insertions or deletions (<50bp), while also incorporating data on *de novo* structural variants at least 50 nucleotides in length. While many of these trios have been analyzed previously, we will be analyzing a cohort of 419 Filipino trios that have not been previously studied, which will potentially provide us with novel findings. Using a combined dataset with a larger sample size will also grant us more statistical power to identify novel genes that may not have been detected before.

Methods

Cohort of OFC-affected Child-Parent Trios

We assembled a cohort of OFC-affected trios recruited from multiple domestic and international sites as part of the Gabriella Miller Kids First (GMKF) Pediatric Research Program, which was established to investigate the biological factors of childhood cancers and structural birth defects such as OFCs. The four samples consist of European, Colombian, Taiwanese, and Filipino trios, with 415 European trios recruited from sites in the United States, Argentina, Turkey, Hungary, and Spain; 275 Colombian trios recruited from Medellin, Colombia; 125 Taiwanese trios recruited from Taiwan; and 419 Filipino trios recruited from the Philippines. Recruitment and phenotypic assessment of participants occurred at regional treatment centers following review and approval by the institutional review board (IRB) for each local recruitment site.

Whole Genome Sequencing and Variant Calling

The sequencing and calling of variants for European, Colombian, and Taiwanese samples have been detailed in previous papers (Bishop et al. 2020, Mukhopadhyay et al. 2020), and the same methods were used for the new sample of Filipino trios. Blood samples were used for sequencing for the majority of trios, with saliva samples used when blood samples were not available. WGS of samples was conducted at either the McDonnell Genome Institute (MGI), Washington University School of Medicine in St. Louis, or the Broad Institute. WGS of 63 Filipino trios was conducted at the Center for Inherited Disease Research (CIDR) at Johns

Hopkins University. All samples were aligned to the human hg38 reference sequence and variant called via GATK pipelines, which was conducted at the GMKF Data Resource Center. For the discovery and annotation of structural variants, we used the GATK-SV pipeline detailed in Collins et al. 2020. Overlapping structural variant calls within the same individual were merged and counted as one structural variant.

Quality Control and Identification of *De Novo* Variants

Methods used for quality control and filtering of DNMs have been detailed in previous papers (Bishop et al. 2020; Mukhopadhyay et al. 2020) We filtered individuals in the cohort based on several quality control metrics, including transition/transversion ratio, silent/replacement rate, and heterozygous/homozygous ratio. Samples with values outside of 3 standard deviations from the cohort mean for missingness, Mendelian error, or average read depth were also removed. Filtering of variants was carried out via BCFtools and VCFtools (Danecek et al. 2011, 2021). Variants with a minor allele count < 2, genotype quality (GQ) < 20, read depth (DP) < 10 were removed. An allele balance (AB) filter of >= 0.30 and <=0.70 was used for variants in the offspring, and an AB filter of < 0.05 was used for corresponding variants in the parents. Annotation of variants was carried out using ANNOVAR and Variant Effect Predictor (VEP) (McLaren et al. 2016; Wang, Li and Hakonarson 2010), and variants were further filtered with only rare (minor allele frequency < 0.3% across gnomAD v3) coding DNMs retained for analysis. Quality control of structural variants was conducted using methods detailed in Collins et al. 2020. Variant classes were assigned to coding DNMs based on predicted function: synonymous, missense (including single amino acid substitutions and non-frameshifting indels), predicted loss of function (pLOF, including stop-gain, frameshifting

indels, and splicing variants). A fourth category, "protein-altering" DNMs combined missense and pLOF DNMs.

Statistical Analysis

We carried out statistical analyses using the R package DenovolyzeR (0.2.0) (Ware et al. 2015). Using the "DenovolyzerByClass" function, we determined whether there was a significant excess of coding DNMs by variant class in trios with OFCs. The enrichment value of each variant class was calculated by dividing the observed number of DNMs by the expected number of DNMs. The model to determine the expected number was developed by Samocha et al. 2014 based on sequence context and probabilities for different mutation types specific to each gene. We also used the "DenovolyzerByClass" function to conduct analyses stratified by cleft subtype, calculating the enrichment separately for trios with CL, CLP, and CP to investigate any differences in enrichment by variant class between OFC subtypes. A sex-specific analysis was also carried out using "DenovolyzerByClass" to determine if there were differences in enrichment between male and female trios. We used the "DenovolyzerByGene" function to determine whether any individual genes contained an excess of pLOF or protein-altering DNMs, by calculating the enrichment and p values for each gene with a DNM in trios with OFCs. We also used the "includeGenes" option in the "DenovolyzerByClass" function to test whether an excess of DNMs was observed in a list of clinically relevant OFC genes, which was constructed after a literature search.

We conducted a gene set enrichment analysis (GSEA) using the program Topp.Fun, available through the ToppGene Suite (Chen et al. 2009), to determine in what gene

sets/pathways DNMs were enriched in trios with OFCs. Calculated p values were adjusted for multiple testing using the Benjamini and Hochberg correction method available in ToppFun.

Gene expression of genes implicated in our analyses was assessed using bulk RNA-seq data derived from human cranial neural crest cells and craniofacial tissues throughout stages of embryonic development relevant to lip and palate development (CS13, CS14, CS15, CS17, and CS17), along with single-cell RNA-seq data derived from CS20 human craniofacial tissues (Yankee et al. 2022).

Results

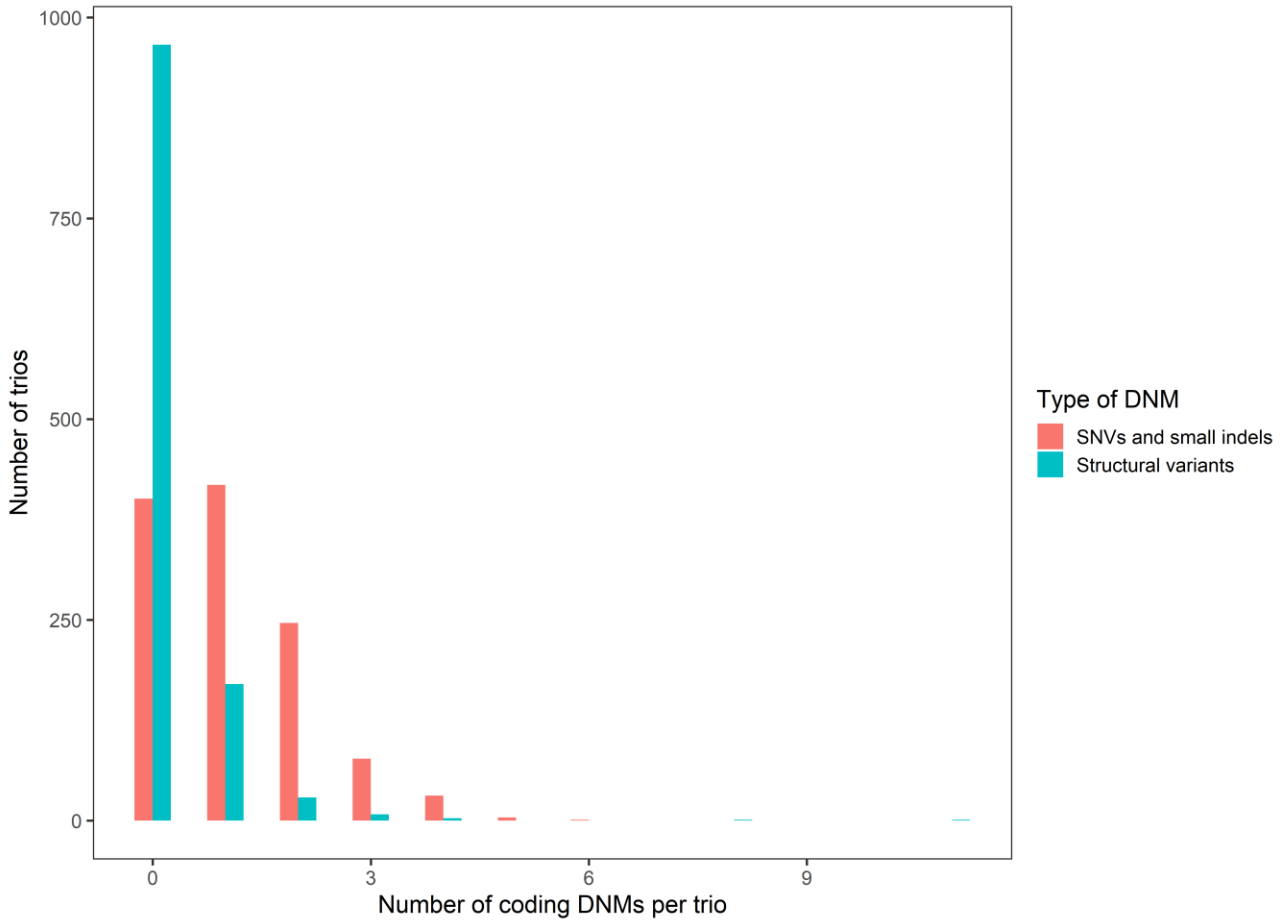

Figure 1. Distribution of coding DNMs per trio, split by type of DNM.

A set of high-confidence DNMs was generated from WGS data from 1178 complete child-parent trios with OFCs of different subtypes: 216 trios with cleft lip (CL), 840 trios with cleft lip and cleft palate (CLP), and 58 trios with cleft palate (CP). 757 of the probands were male, and 421 were female. In total, we identified 74,343 DNMs genome-wide with an average of 66.7 DNMs per proband. 1,291 coding DNMs were detected in 1,199 genes with an average

of 1.10 coding DNMs per proband, corresponding to the expected number of exonic DNMs (Lynch 2010) **(**Figure 1). We assigned variant classes to the coding DNMs based on predicted function: synonymous, missense, predicted loss of function (pLOF), and protein-altering (representing the combined group of missense and pLOF DNMs). Our final set of DNMs included 843 missense (65.3%), 149 pLOF (11.54%), and 299 synonymous (23.16%) variants (Figure 2). We further identified 283 *de novo* structural variants (SV) with an average length of 291kb, averaging 0.24 SVs per proband. We found 131 duplications (46.3%) and 129 deletions (45.6%), representing the majority of SVs. In addition, there were 15 complex SVs (5.3%), 6 insertions (2.12%), and 1 inversion and 1 translocation (0.35% each) (Figure 2).

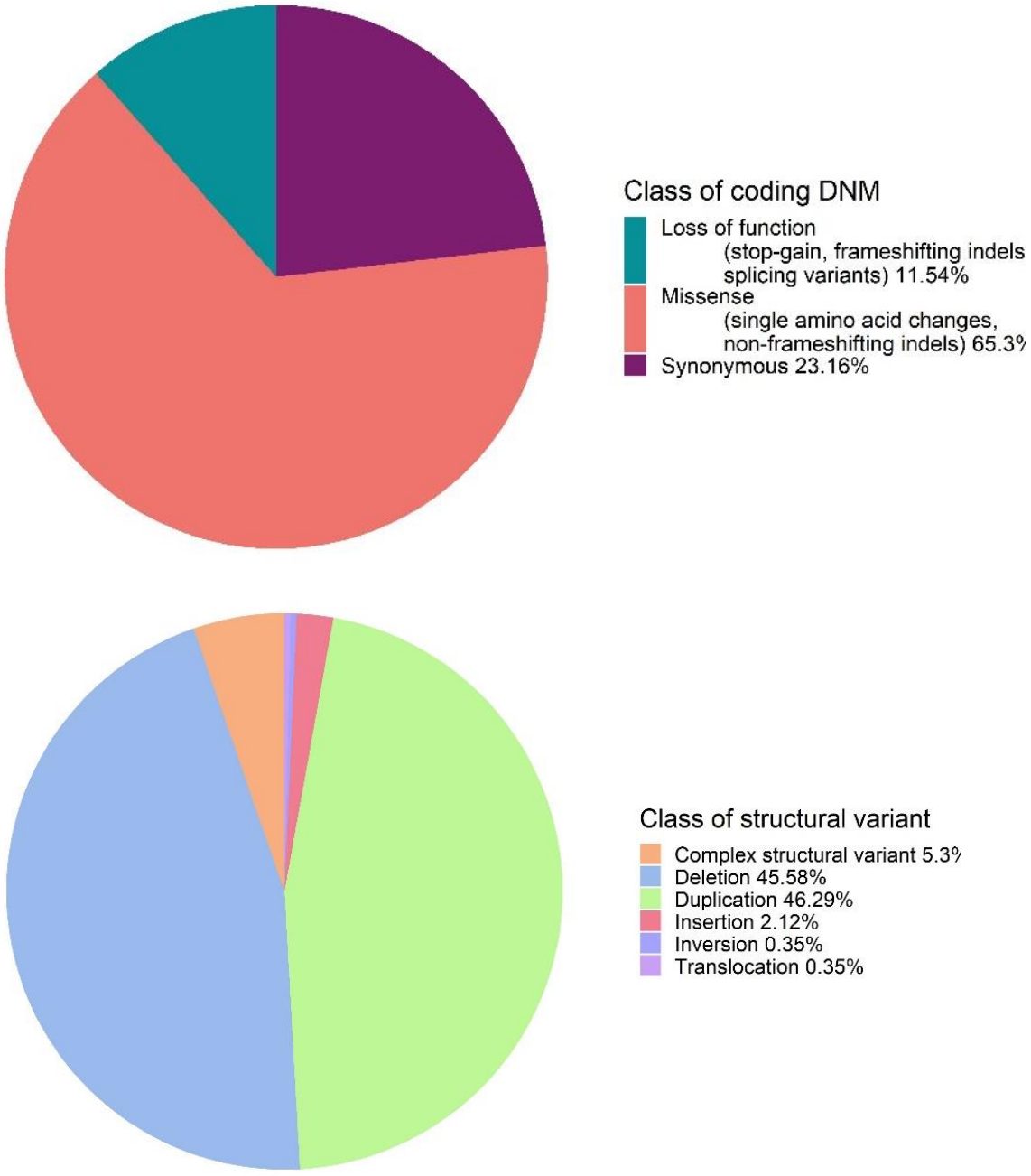

Figure 2. Distribution of coding DNMs by variant class for all OFC trios.

We initially focused on the 501 DNMs found in 419 Filipino trios that had not been analyzed previously in Bishop et al. 2020. We asked if there were more protein-altering DNMs than expected based on mutational models (Samocha et al. 2020) (Figure 3). No significant excess of synonymous DNMs was found (enrichment = 0.902, p = 0.867). We detected a significant excess of both missense (enrichment = 1.18, $p = 2.56 \times 10^{-3}$) and pLOF (enrichment = 1.42, $p = 9.68 \times 10^{-3}$) DNMs, leading to a significant excess of protein-altering DNMs as a whole (enrichment = 1.21, $p = 2.61 \times 10^{-4}$). This enrichment demonstrates that protein-altering DNMs, especially pLOF DNMs, contribute to OFC risk in OFC-affected individuals, while synonymous DNMs that were not enriched are unlikely to have an effect on OFC risk in aggregate.

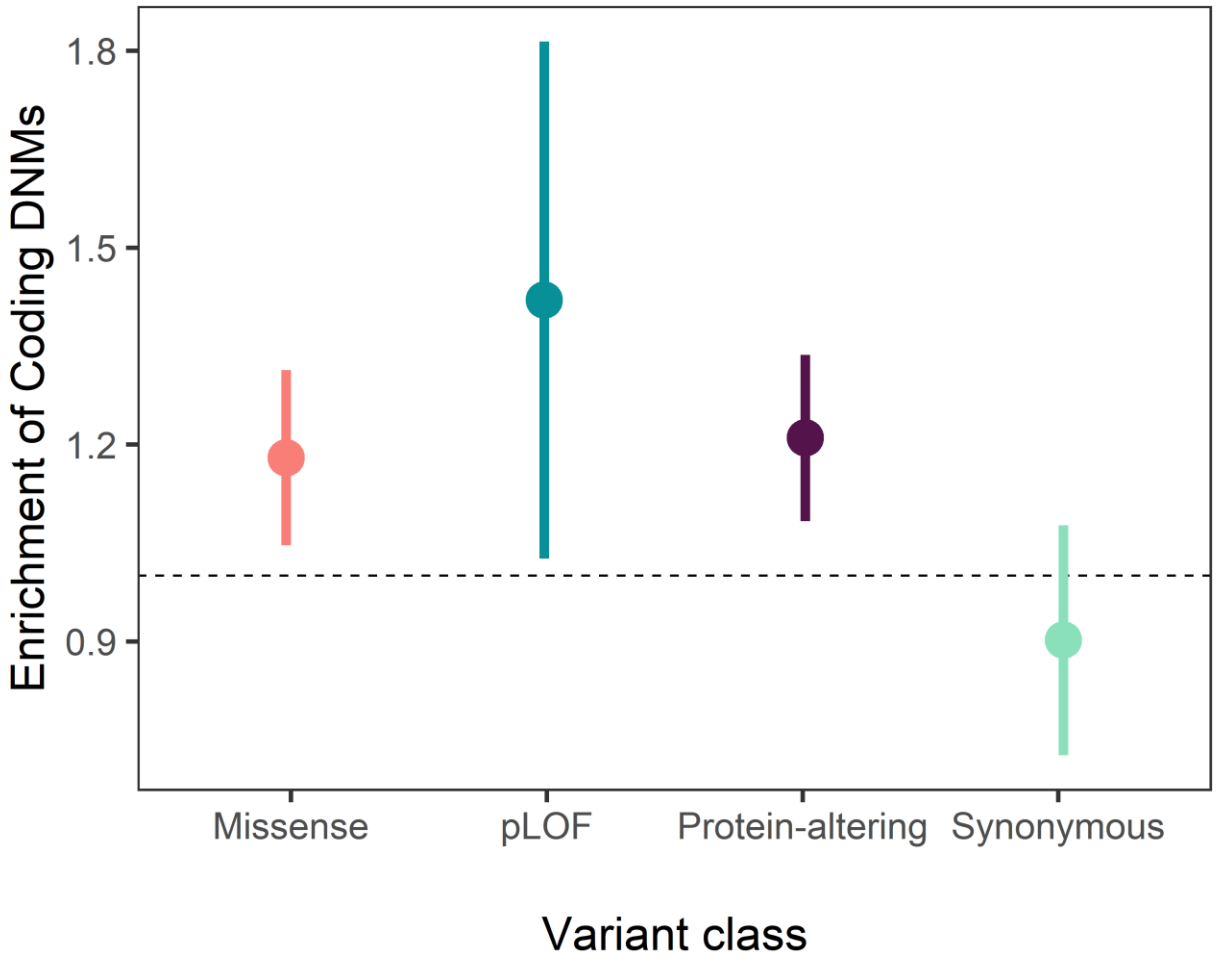

Figure 3. Enrichment of coding DNMs in Filipino trios by variant class. The error bars represent ± 2 standard errors of the enrichment value.

Next, we investigated the genes that harbor these protein-altering DNMs to identify whether there were any single genes with a significant excess of protein-altering DNMs in the cohort of Filipino trios. More protein-altering DNMs were found than expected by chance in genes *COLEC10* ($p = 3.00 \times 10^{-5}$, 2 DNMs) and *AKR1B15* ($p = 3.87 \times 10^{-5}$, 2 DNMs). *COLEC10* is a member of the C-lectin family; mutations in the gene cause the autosomal recessive 3MC syndrome, which often presents with CL/P (Munye et al. 2017, Rabin et al. 2022). *AKR1B15* is predicted to be involved in estrogen biosynthesis and has not been found to be associated with OFC risk before. When restricting the analysis to only genes with pLOF DNMs, the most significantly enriched gene was *GRHL2* ($p = 6.60 \times 10^{-6}$), with two pLOF DNMs (p.Arg75Ter, p.Pro151ArgfsTer8). *GRHL2* is one of a family of highly conserved transcription factors important for embryogenesis and craniofacial development (Carpinelli et al. 2017, Gasperoni et al 2022, Miles et al 2017,), and is predicted to be loss of function intolerant with a pLI score of 1 and a LOEUF score of 0.27.

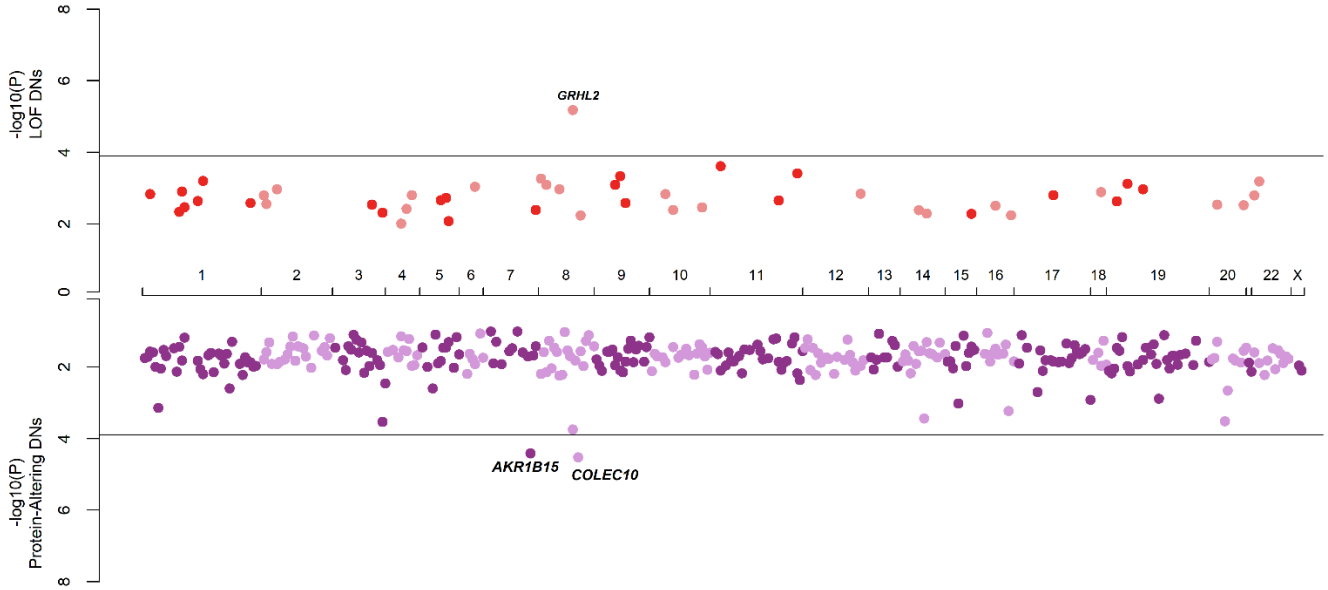

Figure 4. Single genes with an excess of pLOF DNMs (top axis) or protein-altering DNMs (missense and/or pLOF DNMs; bottom axis) in Filipino trios. The solid line indicates the significance threshold after correction for multiple tests, $p < 1.52 \times 10^{-4}$.

We then combined the Filipino trios with GMKF OFC trio cohorts published in Bishop et al. 2020 to determine if the larger dataset would uncover new patterns of DNMs and genes associated with OFCs. In this combined cohort, we found a significant excess of protein-altering DNMs (enrichment = 1.1, $p = 2.55 \times 10^{-3}$) and pLOF DNMs (enrichment = 1.4, $p = 1.57 \times 10^{-4}$), consistent with the results from the Filipino trios alone and Bishop et al. 2020. However, no significant enrichment was found for missense (enrichment = 1.06, $p = 5.61 \times 10^{-2}$) or synonymous DNMs (enrichment = 0.857, p = 0.996).

Next, we conducted a gene set enrichment analysis (GSEA) to determine if the total set of genes with protein-altering DNMs in the combined cohort clustered into any specific gene sets, especially sets representing biological processes relevant to craniofacial development. As a

control, GSEA was also carried out for genes with synonymous DNMs. We found that genes with protein-altering DNMs enriched for terms broadly related to embryonic development, for example, the biological process term "embryonic morphogenesis" ($p = 5.39 \times 10^{-7}$). We also found enrichment for genes belonging to sets more specific to craniofacial development and OFCs, such as the human phenotype term "abnormal midface morphology" ($p = 4.26 \times 10^{-3}$) and human disease terms "cleft palate" ($p = 8.76 \times 10^{-5}$) and "craniofacial abnormalities" ($p = 5.31 \times 10^{-4}$). Genes that contained only synonymous DNMs were not enriched for any particular terms, as was expected as these DNMs are expected to occur randomly throughout the exome and are less likely to have an effect on OFC risk. Finally, we also carried out an analysis to determine the contribution of DNMs in a set of 418 genes derived from clinical genetic testing panels for OFCs. We identified 67 individuals with 68 DNMs in 54 genes from the list and found a significant excess of pLOF (enrichment = 3.27, $p = 7.48 \times 10^{-4}$) and protein-altering DNMs (enrichment = 2.04, $p = 1.08 \times 10^{-6}$). Notably this enrichment was over two-fold higher than the enrichment across all genes. To determine whether the enrichment in clinically relevant OFC genes accounted for the enrichment across all genes with DNMs, we removed trios that harbored DNMs in genes on the list and conducted another analysis. We still detected a significant excess of pLOF DNMs (enrichment = 1.30, $p = 2.82 \times 10^{-3}$), although no enrichment was detected for protein-altering DNMs as a whole (enrichment = 1.03, p = 0.173). Thus, it is clear that DNMs, especially pLOF DNMs, in genes currently not known to affect OFC risk are contributing to the overall enrichment of DNMs in the cohort.

We repeated our analysis of the enrichment of protein-altering DNMs in individual genes within the combined dataset (Figure 5), and identified two genes with a significant excess of protein-altering DNMs, *IRF6* ($p = 5.21 \times 10^{-10}$) and *TFAP2A* ($p = 1.06 \times 10^{-5}$). Both genes have

been implicated in OFC etiology and were previously found to be enriched for DNMs in the Bishop et al. 2020 analysis. Heterozygous mutations in *IRF6* are known to cause Van der Woude syndrome, and heterozygous mutations in *TFAP2A* have been found to cause Branchio-oculo-facial syndrome, both of which present with OFCs (de Lima et al. 2009, Kondo et al. 2002, Milunsky et al. 2008, Reiber et al. 2010). When looking only at genes with an excess of pLOF DNMs, *GRHL2* remained significantly enriched ($p = 4.65 \times 10^{-5}$), as was *TFAP2A* ($p = 2.68 \times 10^{-6}$).

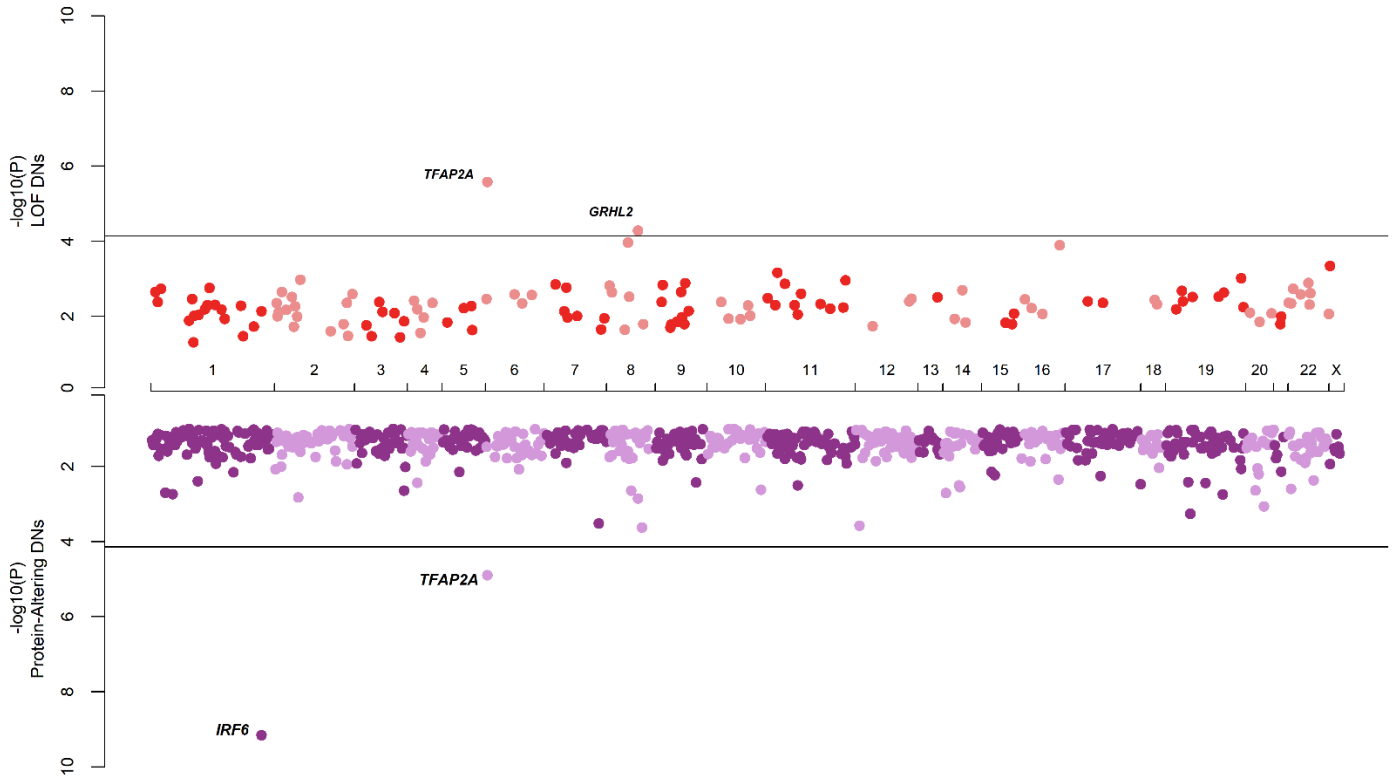

Figure 5. Single genes with an excess of pLOF DNMs (top axis) or protein-altering DNMs (missense and/or pLOF DNMs; bottom axis) in all OFC trios. The solid line indicates the significance threshold after correction for multiple tests, $p < 7.16 \times 10^{-5}$.

We next asked if there were any differences in enrichment of DNM classes between male and female probands (Figure 6). In both males and females, we detected a significant excess of pLOF DNMs, with an enrichment of 1.37 ($p = 2.20 \times 10^{-3}$) for males and an enrichment of 1.39 for females ($p = 1.54 \times 10^{-2}$). For males, we also detected a significant excess of protein-altering (enrichment = 1.1, $p = 1.05 \times 10^{-2}$), but the same was not found in females (enrichment = 1.09, $p = 5.82 \times 10^{-2}$). When we compared the enrichment of DNMs between male and female probands, no significant differences were observed for any variant class.

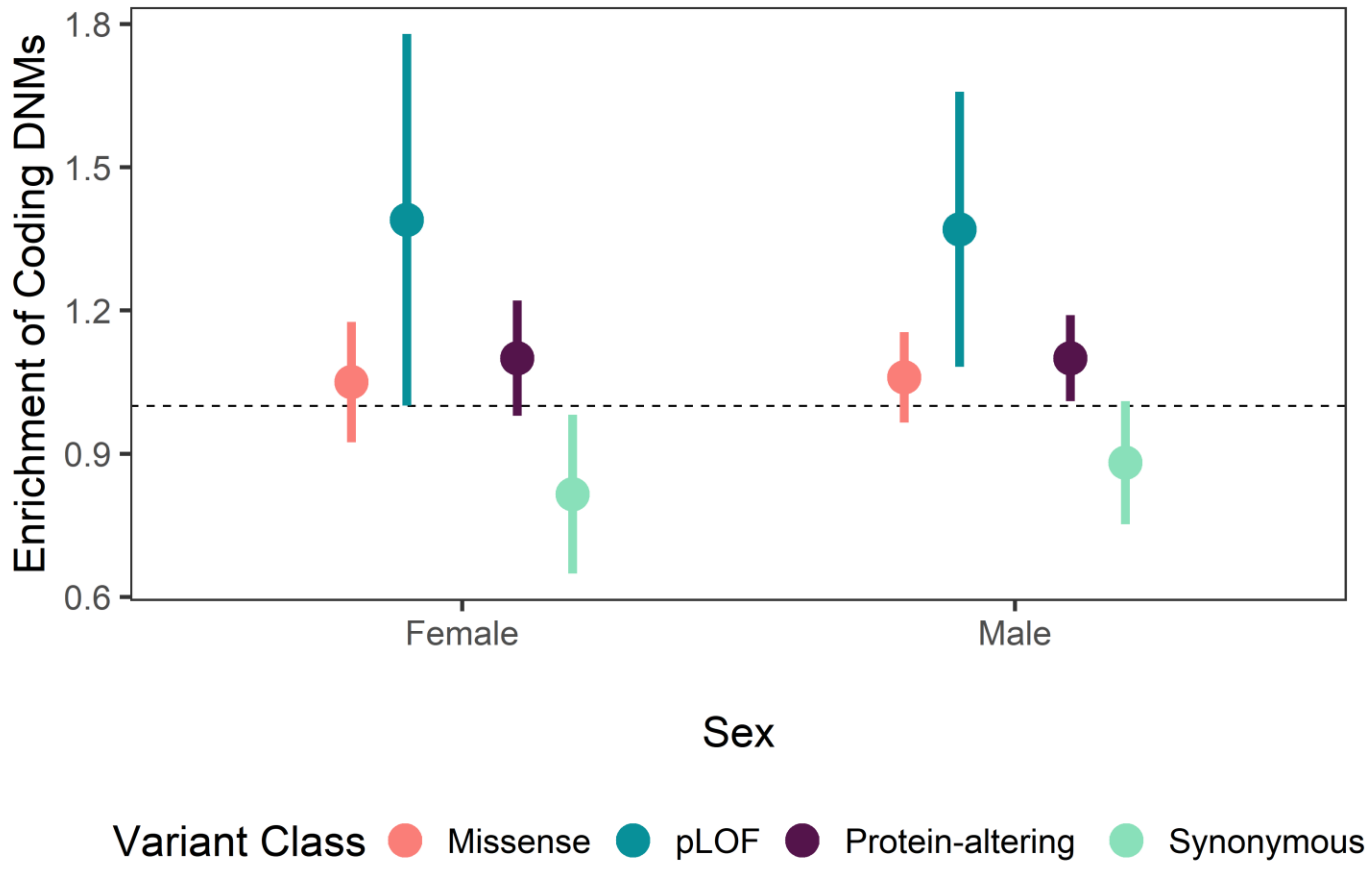

Figure 6. Enrichment of coding DNMs ± 2 standard errors by variant class for male probands and female probands

We then investigated whether there were any differences in enrichment between trios with different OFC subtypes, stratifying the analyses into CL, CLP, and CP trios. No significant excess of protein-altering DNMs was found in trios with CL (enrichment = 1.08, p = 0.156) or trios with CP (enrichment = 1.25, $p = 6.63 \times 10^{-2}$), although the latter likely reflects the small sample size of CP trios. Trios with CLP were found to harbor a significant excess of pLOF (enrichment = 1.44, $p = 1.25 \times 10^{-4}$) and protein-altering DNMs (enrichment = 1.09, $p = 1.02 \times 10^{-2}$), but there was no significant enrichment of missense (enrichment = 1.04, p = 0.150) DNMs. We note however, that the point estimates for enrichment were about the same magnitude in CL and CLP trios, but due to the small set of CL trios, this enrichment was not significant.

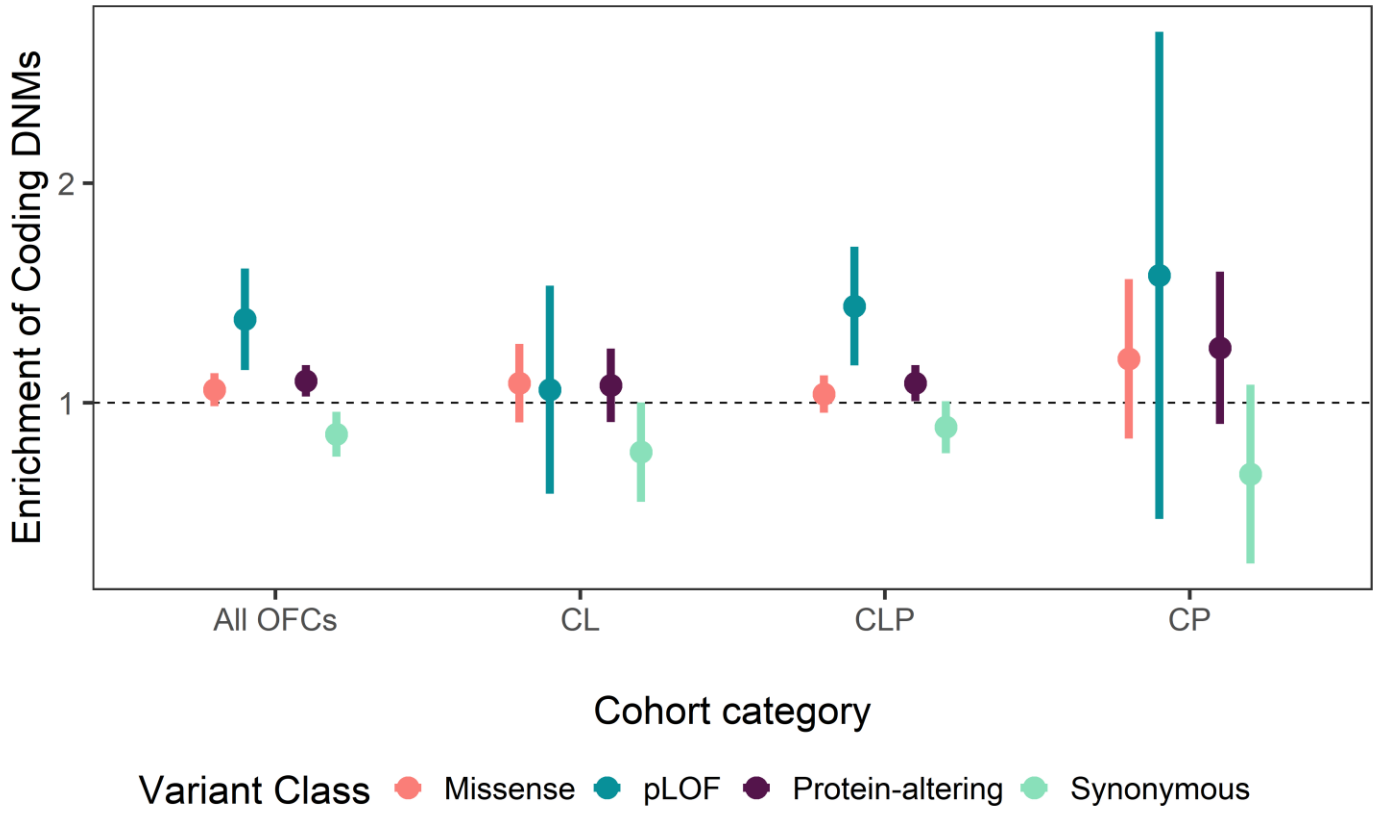

Figure 7. Enrichment of coding DNMs ± 2 standard errors by variant class for all OFC trios, trios with CL only, trios with CLP, and trios with CP only.

We then expanded our analysis to *de novo* structural variants (SVs). We found 49 SVs that were predicted to cause loss of function of 262 genes. We detected multiple SVs within the 22q11.2 chromosome region, with 1 duplication and 7 deletions affecting 62 genes in total. 22q11.2 deletion syndrome is the most common microdeletion disorder and symptoms often include palatal abnormalities, so it is unsurprising that we would detect OFC trios with a deletion in the region (McDonald-McGinn et al. 2020).

Next, we explored the 18 genes that were affected by both pLOF SVs and protein-altering DNMs. Notably, we found a pLOF SV in a gene that was also significantly enriched for DNMs: *GRHL2* with a deletion of exons 2-9 (Figure 8)**.** Additionally, we detected a 570kb long deletion including *CTNND1*, a gene we identified two protein-altering DNMs in (p.Ala678Gly, p.Ser811Ter), but which we had not calculated to be significantly enriched. Truncating mutations in *CTNND1* cause Blepharocheilodontic syndrome, which often presents with CL/P, and have also been associated with nonsyndromic CL/P (Alharatani et al. 2020, Ghoumid et al. 2017). We also identified multiple genes within the 22q11.2 region that harbored protein-altering DNMs, namely *HIRA, ZDHHC8,* and *ZNF280A*. Finally, we identified several genes in this analysis that have not been implicated in OFC risk before, and these included *AGO2*, *ATXN2L*, and *KIF3B.*

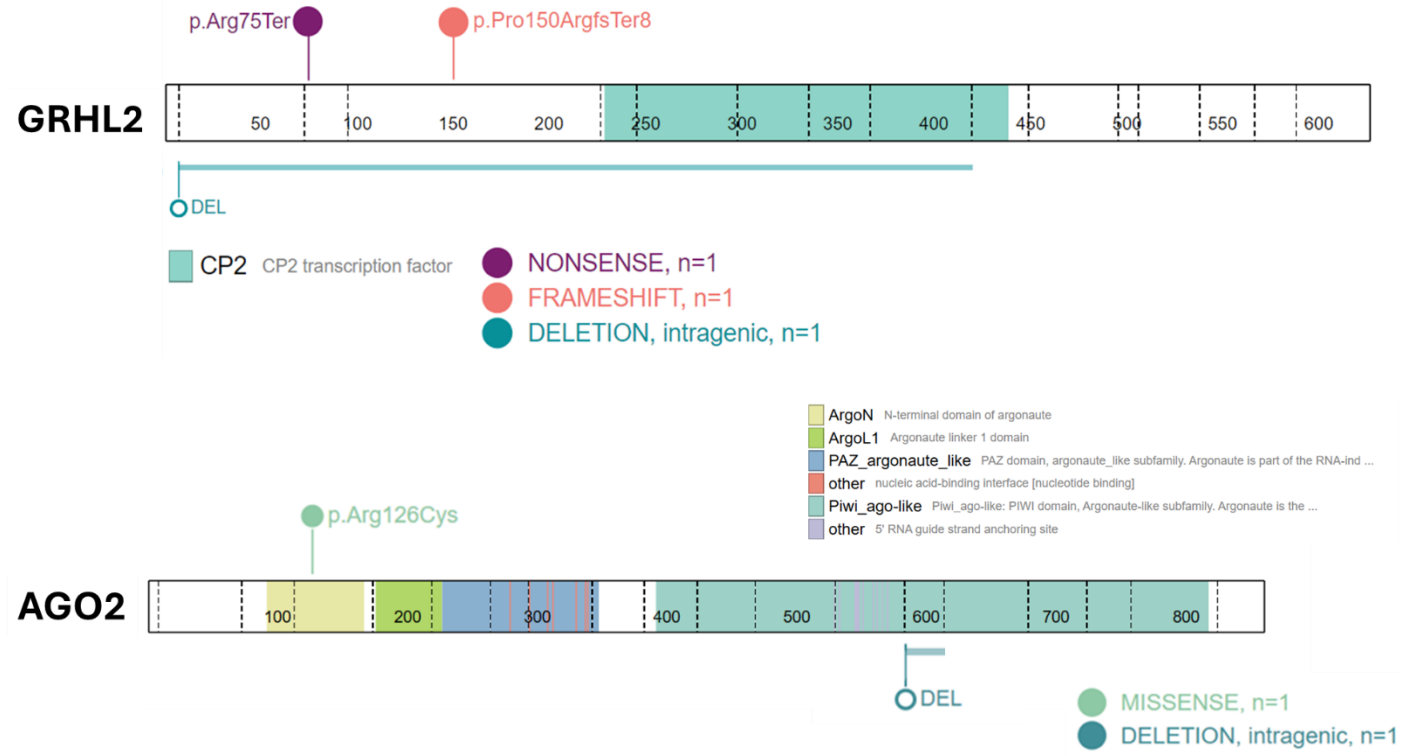

Figure 8. pLOF structural variant and DNMs in *GRHL2* and *AGO2* visualized in ProteinPaint

Finally, we assessed the expression of genes with DNMs throughout craniofacial development, using bulk RNA-seq data generated from craniofacial tissues throughout development. We found 440 genes with protein-altering DNMs (both SNVs and SVs) that were within the top 20th percentile of expressed genes throughout craniofacial development. This included many of the genes with both pLOF SVs and DNMs, including *AGO2*, *ATXN2L*, and *KIF3B*. We also assessed the expression of our top genes within CS20 craniofacial tissues using single-cell RNA-seq data and found expression throughout multiple cell types in developing craniofacial tissue for multiple genes, supporting a potential role for these genes in craniofacial development (Figure 9). *GRHL2* was strongly expressed within the ectoderm, with minimal expression in any other cell type.

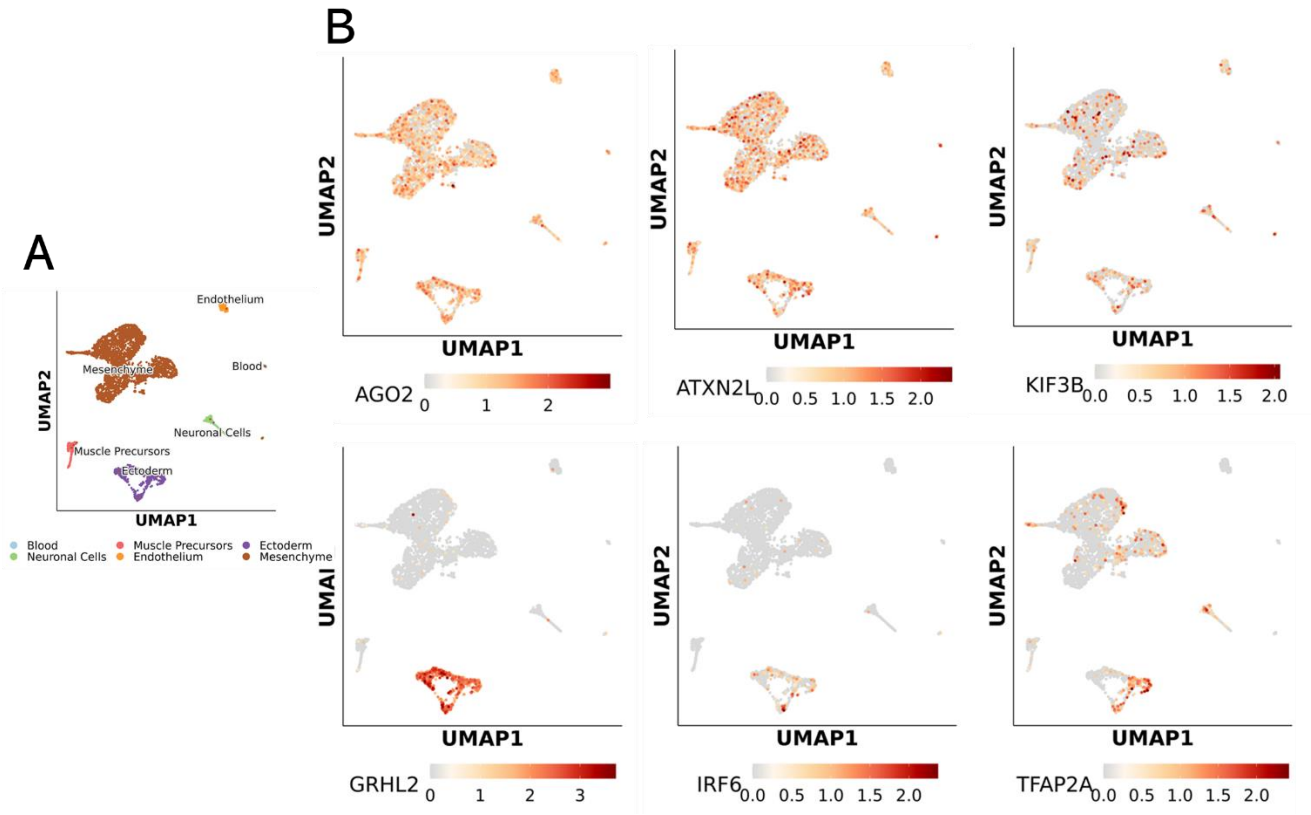

Figure 9. A) Visualization of the cells included in single-cell RNA-seq data of CS20 human craniofacial tissues, clustered by type of cell. Cell types includes blood, neuronal cells, muscle precursors, endothelium, ectoderm, and mesenchyme. B) Visualizations of expression of 6 genes in single-cell RNA-seq data of CS20 human craniofacial tissues. Cells are colored by gene expression in a white-red color scheme, with red representing higher expression. Genes shown include *AGO2*, *ATXN2L*, *KIF3B*, *GRHL2*, *IRF6*, and *TFAP2A*.

Discussion

This analysis has further demonstrated the importance of coding DNMs in the etiology of OFCs, providing insight into the role of both single nucleotide variants and large structural variants in contributing to OFC risk in humans. Our findings provide further evidence for discoveries made in previous studies, providing further insight into known OFC risk genes, as well as implicating novel genes with a potential role in OFC etiology.

The most promising new gene identified is *GRHL2,* in which we identified a total of three pLOF variants: a frameshift deletion (p.Pro151ArgfsTer8), one stopgain (p.Arg75Ter), and an intragenic deletion of exons 2-9. Since *GRHL2* is predicted to be intolerant to pLOF variants (pLI = 1, LOEUF = 0.27), these findings are particularly indicative of a potentially important role in influencing OFC risk. Another notable member of the Grainyhead-like family of transcription factors is *GRHL3,* in which mutations are known to cause Van der Woude syndrome and nonsyndromic CP (Leslie et al. 2016, Mangold et al. 2016, Peyrard-Janvid et al. 2014**).** *GRHL2* has previously been implicated in craniofacial anomalies in mice, with studies finding that Grhl2 mutations in mouse models cause significant developmental defects including neural tube defects, exencephaly, and cleft face (Crane-Smith et al. 2023, de Vries et al. 2021, Menke et al. 2015). Additionally, a novel enhancer element (mm1286) has been discovered to drive Grhl2 expression in the craniofacial primordia, with deletion causing a predisposition to palatal clefting (de Vries et al. 2020). However, this analysis represents the first time mutations in *GRHL2* have been found to contribute to OFC risk in humans. Truncating variants in *GRHL2* have also been associated with autosomal dominant progressive hearing loss and missense variants are

associated with autosomal recessive ectodermal dysplasia (Peters et al. 2002, Petrof et al. 2014, Vona et al. 2013). We also detected a truncating DNM in *GRHL1* (p.Arg263Ter) and a missense DNM in *GRHL3* (p.Pro480Ser), which further suggests that this gene family as a whole may be relevant to OFC etiology. The three pLOF variants detected here conclusively implicate *GRHL2* as an OFC risk gene, and further investigation is needed to interpret additional rare variants and their association with OFCs, plus their role in comorbidities such as deafness.

By combining our analysis of *de novo* SNVs and small indels with large structural variants, we found further evidence of the role of genes implicated in our initial analysis, namely *GRHL2* and *TFAP2A.* We also identified several genes deleted within the 22q11.2 region that harbored protein-altering DNMs, including *HIRA*, *ZDHHC8*, and *ZNF280A*. Notably, *ZDHHC8* was found to be one of the top enriched genes in our combined analysis, harboring two protein-altering DNMs (p.Leu259CysfsTer5, p.Pro702Lys), although this enrichment was not found to be significant. Considering the variability of symptoms for individuals with 22q11.2 deletion syndrome, these results could suggest a role for these genes in contributing to the clefting phenotype seen often in affected individuals, which could warrant further studies to investigate this. We also identified several novel genes that could potentially contribute to OFC risk that were not determined to be significant in our initial enrichment analysis, including *AGO2, ATXN2L,* and *KIF3B*. Knockout mice models of all three genes are embryonic lethal, suggesting that they play important roles in embryonic development (Alisch et al. 2007, Key et al. 2020, Morita et al. 2007, Nonaka et al. 1998). In *AGO2,* a gene which forms part of the RNA-induced silencing complex and predicted to be intolerant to pLOF variants (pLI = 1, LOEUF = 0.106), we identified one missense DNM (p.Arg126Cys) and one intragenic deletion of exon 14.

Heterozygous mutations in *AGO2* have been found to impair neurological development and cause craniofacial abnormalities (Lessel et al. 2020).

Along with identifying novel candidate genes such as *GRHL2,* by combining the previously analyzed dataset with a new cohort of Filipino trios and new data on structural variants, we also found further evidence for several of the genes previously identified in Bishop et al. 2020. We identified additional variants in these genes: one frameshift insertion in *IRF6* (p.Arg6ProfsTer15), one frameshift deletion in *ZFHX3* (p.Ser518AlafsTer18), and one deletion of *CTNND1*. We also identified DNMs in genes previously known to be associated with OFCs, finding a very high enrichment of DNMs in clinically relevant OFC genes. These results support the findings of prior studies, further reinforcing the role these genes have in contributing to OFC risk.

In general, it is difficult to identify specific genes that may play a role in OFC development through analyzing enrichment of DNMs, as the etiologically heterogeneous nature of OFCs means that the genetic factors that influence OFC risk are spread throughout the genome. Therefore, many of the genes with DNMs that influence OFC risk will not reach significance for enrichment and may then be missed in our analysis. Further studies with larger sample sizes are thus necessary to increase power and pinpoint OFC candidate genes. There was also insufficient consideration of the effect of missense variants on gene function, as within this study all missense DNMs were considered equally as protein-altering when calculating enrichment of variant classes, despite the likelihood of many of these DNMs being benign. Future work on DNMs could incorporate the pathogenicity of missense DNMs into the analysis, potentially using different statistical tools, to prioritize the variants most likely to have a deleterious effect on gene function and thus identify the genes most likely to contribute to OFCs.

Further analyses of structural variation within this dataset could also be conducted, as we mainly considered the impact of SVs predicted to cause loss of function. It would be valuable to analyze the other SVs with different effects on gene function, especially of duplications, which represented the majority of the SVs. Continued analyses of the WGS dataset in general would be valuable, as we have thus far analyzed a small subset of the genetic variants influencing OFC risk across the genome. We could combine our findings with analyses of other types of variation such as noncoding variation and rare inherited variants, to further identify genetic regions likely to contribute to OFC risk.

In our study we have demonstrated the important role of rare coding DNMs in influencing OFC risk in 1178 child-parent trios. By incorporating multiple types of DNMs including SNVs, small indels, and SVs, we have made discoveries that would otherwise have been missed, demonstrating the value of investigating different types of genetic variation. Through our analyses we have identified novel candidate OFC risk genes, the most important of which is *GRHL2*, and found further evidence to support the role of previously reported genes. Future studies are still necessary in order to elucidate the function of the variants discovered, as well as to identify additional variants within these genes to provide further evidence for their role in OFC etiology.

Reference List

Alharatani, Reham, Athina Ververi, Ana Beleza-Meireles, Weizhen Ji, Emily Mis, Quinten T. Patterson, John N. Griffin, et al. "Novel truncating mutations in CTNND1 cause a dominant craniofacial and cardiac syndrome." *Human Molecular Genetics* 29, no.11 (2020): 1900-1921. https://doi.org/10.1093/hmg/ddaa050

Alisch, Reid S., Peng Jin, Michael Epstein, Tamara Caspary, and Stephen T. Warren. "Argonaute2 Is Essential for Mammalian Gastrulation and Proper Mesoderm Formation." *PLoS Genetics* 3, no.12 (2007): e227. https://doi.org/10.1371%2Fjournal.pgen.0030227

Bishop, Madison R., Kimberley K. Diaz Perez, Miranda Sun, Samantha Ho, Pankaj Chopra, Nandita Mukhopadhyay, Jacqueline B. Hetmanski, et al. "Genome-wide enrichment of *de novo* coding mutations in orofacial cleft trios." *American Journal of Human Genetics* 107, no.1 (2020): 124-136. https://doi.org/10.1016/j.ajhg.2020.05.018

Carpinelli, M.R., M.E. de Vries, S.M. Jane, and S. Dworkin. "Grainyhead-like Transcription Factors in Craniofacial Development." *Journal of Dental Research* 96, no.11 (2017): 1200-1209. https://doi.org/10.1177/0022034517719264

Chen, Jing, Eric E. Bardes, Bruce J. Aronow, and Anil G. Jegga. "ToppGene Suite for gene list enrichment analysis and candidate gene prioritization." *Nucleic Acids Research* 37, no. suppl_2 (2009): W305-W311. https://doi.org/10.1093/nar/gkp427

Christensen, Kaare, Knud Juel, Anne M. Herskind, and Jeffrey C. Murray. "Long term follow up study of survival associated with cleft lip and palate at birth." *BMJ (Clinical research ed.)* 328, no.7453 (2004): 1405. https://doi.org/10.1136%2Fbmj.38106.559120.7C

Collins, Ryan L., Harrison Brand, Konrad J. Karczewski, Xuefang Zhao, Jessica Alföldi, Laurent C. Francioli, Amit V. Khera, et al. "A structural variation reference for medical and population genetics." *Nature* 581 (2020): 444-451. https://doi.org/10.1038/s41586-020-2287-8

Crane-Smith, Zoe, Sandra C.P. De Castro, Evanthia Nikolopoulou, Paul Wolujewicz, Damian Smedley, Yunping Lei, Emma Mather, et al. "A non-coding insertional mutation of Grhl2 causes gene over-expression and multiple structural anomalies including cleft palate, spina bifida and encephalocele." *Human Molecular Genetics* 32, no.17 (2023): 2681-2692. https://doi.org/10.1093/hmg/ddad094

Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, et al. "Twelve years of SAMtools and BCFtools." *Gigascience* 10, no.2 (2021): giab008. https://doi.org/10.1093/gigascience/giab008

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. "The variant call format and VCFtools." *Bioinformatics* 27, no.15 (2011): 2156-2158. https://doi.org/10.1093/bioinformatics/btr330

de Lima, Renata L., Sarah A. Hoper, Michella Ghassibe, Margaret E. Cooper, Nicholas K. Rorick, Shinji Kondo, Lori Katz, et al. "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." *Genetics in Medicine* 11, no.4 (2009): 241-247. https://doi.org/10.1097/gim.0b013e318197a49a

de Vries, Michael, Harley G. Owens, Marina R. Carpinelli, Darren Partridge, Ariena Kersbergen, Kate D. Sutherland, Alana Auden, et al. "Delineating the roles of Grhl2 in craniofacial development through tissue-specific conditional deletion and epistasis approaches in mouse." *Developmental Dynamics* 250, no.8 (2021): 1191-1209. https://doi.org/10.1002/dvdy.322

de Vries, Michael, Marina Carpinelli, Emilie Rutland, Aaron Hatzipantelis, Darren Partridge, Alana Auden, Peter J. Anderson, et al. "Interrogating the Grainyhead-like 2 (Grhl2) genomic locus identifies an enhancer element that regulates palatogenesis in mouse." *Developmental Biology* 459, no.2 (2020): 194-203. https://doi.org/10.1016/j.ydbio.2019.11.015

Gasperoni, Jemma G., Jarrad N. Fuller, Charbel Darido, Tomasz Wilanowski, and Sebastian Dworkin. "Grainyhead-like (Grhl) Target Genes in Development and Cancer." *International Journal of Molecular Sciences* 23, no.5 (2022): 2735. https://doi.org/10.3390%2Fijms23052735

Ghoumid, Jamal, Morgane Stichelbout, Anne-Sophie Jourdain, Frederic Frenois, Sophie Lejeune-Dumoulin, Marie-Pierre Alex-Cordier, Marine Lebrun, et al. "Blepharocheilodontic syndrome is a CDH1 pathway-related disorder due to mutations in CDH1 and CTNND1." *Genetics in Medicine* 19, no. 9 (2017): 1013-1021. https://doi.org/10.1038/gim.2017.11

Grosen, Dorthe, Cecile Chevrier, Axel Skytthe, Camilla Bille, Kirsten Mølsted, Åse Sivertsen, Jeffrey C. Murray, and Kaare Christensen. "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 (2010): 162-168. https://doi.org/10.1136/jmg.2009.069385

Hocevar-Boltezar, I, A Jarc, and V Kozelj. "Ear, Nose and Voice Problems in Children with Orofacial Clefts." *The Journal of Laryngology & Otology* 120, no. 4 (2006): 276–81. https://doi.org/10.1017/S0022215106000454.

Honein, Margaret A., Sonja A. Rasmussen, Jennita Reefhuis, Paul A. Romitti, Edward J. Lammer, Lixian Sun, and Adolfo Correa. "Maternal smoking and environmental tobacco smoke exposure and the risk of orofacial clefts." *Epidemiology* 18, no.2 (2007): 226-233. https://doi.org/10.1097/01.ede.0000254430.61294.c0

Hunt, Orlagh, Donald Burden, Peter Hepper, Mike Stevenson, and Chris Johnston. "Self-reports of psychosocial functioning among children and young adults with cleft lip and palate." *The Cleft palate-craniofacial journal* 43, no.5 (2006): 598-605. https://doi.org/10.1597/05-080

James, Olutayo, Adegbayi A. Adekunle, Olawale O. Adamson, Olushola F. Agbogidi, Wasiu L. Adeyemo, Azeez Butali, Akinola L. Ladeinde, Mobolanle O. Ogunlewe. "Management of Orofacial Cleft in Nigeria - A Retrospective Study." *Annals of Maxillofacial Surgery* 10, no.2 (2020): 434-438. https://doi.org/10.4103%2Fams.ams_104_20

Key, Jana, Patrick N. Harter, Nesli-Ece Sen, Elise Gradhand, Georg Auburger, and Suzana Gispert. "Mid-Gestation lethality of Atxn2l-Ablated Mice." *International Journal of Molecular Sciences* 21, no.14 (2020): 5124. https://doi.org/10.3390/ijms21145124

Khoury, M.J., A. Weinstein, S. Panny, N.A. Holtzman, P.K. Lindsay, K. Farrel, and M. Eisenberg. "Maternal cigarette smoking and oral clefts: A population-based study." *American Journal of Public Health* 77, no.5 (1987): 623-625. https://doi.org/10.2105/ajph.77.5.623

Kondo, Shinji, Brian C. Schutte, Rebecca J. Richardson, Bryan C. Bjork, Alexandra S. Knight, Yoriko Watanabe, Emma Howard, et al. "Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes." *Nature Genetics* 32, no.2 (2002): 285-289. https://doi.org/10.1038/ng985

Leslie, Elizabeth J. "Genetic models and approaches to study orofacial clefts." Oral Diseases 28, no.5 (2021): 1327–1338, https://doi.org/10.1111/odi.14109

Leslie, Elizabeth J., Huan Liu, Jenna C. Carlson, John R. Shaffer, Eleanor Feingold, George Wehby, Cecelia A. Laurie, et al. "A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3." *American Journal of Human Genetics* 98, no.4 (2016): 744-754. https://doi.org/10.1016/j.ajhg.2016.02.014

Lynch, Michael. "Rate, molecular spectrum, and consequences of human mutation." *Proceedings of the National Academy of Sciences of the United States of America* 107, no.3 (2010): 961-968. https://doi.org/10.1073%2Fpnas.0912629107

Mangold, Elisabeth, Anne C. Böhmer, Nina Ishorst, Ann-Kathrin Hoebel, Pinar Gültepe, Hannah Schuenke, Johanna Klamt, et al. "Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate." *American Journal of Human Genetics* 98, no.4 (2016): 755-762. https://doi.org/10.1016/j.ajhg.2016.02.013

Marazita, Mary L., Andrew C. Lidral, Jeffrey C. Murray, L. Leigh Field, Brion S. Maher, Toby Goldstein McHenry, Margaret E. Cooper, et al. "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, no.3 (2009): 151-170. https://doi.org/10.1159%2F000224636

Marcusson, Agneta, Ingemar Akerlind, and Gunnar Paulin. "Quality of Life in Adults with Repaired Complete Cleft Lip and Palate." *The Cleft Palate Craniofacial Journal* 38, no.4 (2001): 379-385. https://doi.org/10.1597/1545-1569_2001_038_0379_qoliaw_2.0.co_2

McDonald-McGinn, Donna M., Heather S. Hain, Beverley S. Emanuel, and Elaine H. Zackai. "22q11.2 Deletion Syndrome." *GeneReviews*, Feb 27, 2020. https://www.ncbi.nlm.nih.gov/books/NBK1523/

McLaren, William, Laurent Gil, Sarah E. Hunt, Harpreet Singh Riat, Graham R.S. Ritchie, Anja Thormann, Paul Flicek and Fiona Cunningham. "The Ensembl Variant Effect Predictor." *Genome Biology* 17, no.122: (2016). https://doi.org/10.1186/s13059-016-0974-4

Menke, Chelsea, Megan Cionni, Trevor Siggers, Martha L. Bulyk, David R. Beier, and Rolf W. Stottmann. "Grhl2 is required in non-neural tissues for neural progenitor survival and forebrain development." *Genesis* 53, no.9 (2015): 573-582. https://doi.org/10.1002%2Fdvg.22875

Miles, Lee B., Charbel Darido, Jan Kaslin, Joan K. Heath, Stephen M. Jane, and Sebastian Dworkin. "Mis-expression of grainyhead-like transcription factors in zebrafish leads to defects in enveloping layer (EVL) integrity, cellular morphogenesis and axial extension." *Scientific Reports* 7, no. 17607 (2017). https://doi.org/10.1038/s41598-017-17898-7

Milunsky, Jeff M., Tom A. Maher, Geping Zhao, Amy E. Roberts, Heather J. Stalker, Roberto T. Zori, Michelle N. Burch, et al. "TFAP2A mutations result in branchio-oculo-facial syndrome." *American Journal of Human Genetics* 82 no. 5 (2008): 1171-1177. https://doi.org/10.1016/j.ajhg.2008.03.005

Moreno, Lina M., Maria Adela Mansilla, Steve A. Bullard, Margaret E. Cooper, Tamara D. Busch, Junichiro Machida, et al. "FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate." *Human Molecular Genetics* 18, no.24 (2009): https://doi.org/10.1093%2Fhmg%2Fddp444

Morita, Sumiyo, Takuro Horii, Mika Kimura, Yuji Goto, Takahiro Ochiya, and Izuho Hatada. "One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation." *Genomics* 89, no.6 (2007): 687-696. https://doi.org/10.1016/j.ygeno.2007.01.004

Mossey PA, WC Shaw, RG Munger, JC Murray, J Murthy, and J Little. "Global Oral Health Inequalities: Challenges in the Prevention and Management of Orofacial Clefts and

Potential Solutions." *Advances in Dental Research* 23, no.2 (2011): 247-258. https://doi.org/10.1177/0022034511402083

Mukhopadhyay, Nandita, Madison Bishop, Michael Mortillo, Pankaj Chopra, Jacqueline B. Hetmanski, Margaret A. Taub, Lina M. Moreno, et al. "Whole genome sequencing of orofacial cleft trios from the Gabriella Miller Kids First Pediatric Research Consortium identifies a new locus on chromosome 21." *Human Genetics* 139 (2020): 215-226. https://doi.org/10.1007/s00439-019-02099-1

Munye, Mustafa M., Anna Diaz-Font, Louise Ocaka, Maiken L. Henriksen, Melissa Lees, Angela Brady, Dagan Jenkins, et al. "COLEC10 is mutated in 3MC patients and regulates craniofacial development." *PLoS Genetics* 13, no.3 (2017): e1006679. https://doi.org/10.1371/journal.pgen.1006679

Nonaka, Shigenori, Yosuke Tanaka, Yasushi Okada, Sen Takeda, Akihiro Harada, Yoshimitsu Kanai, Mizuho Kido, and Nobutaka Hirokawa. Randomization of Left–Right Asymmetry due to Loss of Nodal Cilia Generating Leftward Flow of Extraembryonic Fluid in Mice Lacking KIF3B Motor Protein." *Cell* 95, no.6 (1998): 829-837. https://doi.org/10.1016/S0092-8674(00)81705-5

Peters, Linda M., David W. Anderson, Andrew J. Griffith, Kenneth M. Grundfast, Theresa B. San Agustin, Anne C. Madeo, Thomas B. Friedman, Robert J. Morell. "Mutation of a transcription factor, TFCP2L3, causes progressive autosomal dominant hearing loss, DFNA28." *Human Molecular Genetics* 11, no.23 (2002): 2877-2885. https://doi.org/10.1093/hmg/11.23.2877

Petrof, Gabriela, Arti Nanda, Jake Howden, Takuya Takeichi, James R. McMillan, Sophia Aristodemou, Linda Ozoemena, et al. "Mutations in GRHL2 result in an autosomal-recessive ectodermal Dysplasia syndrome." *American Journal of Human Genetics* 95, no.3 (2014): 308-314. https://doi.org/10.1016/j.ajhg.2014.08.001

Peyrard-Janvid, Myriam, Elizabeth J. Leslie, Youssef A. Kousa, Tiffany L. Smith, Martine Dunnwald, Måns Magnusson, Brian A. Lentz, et al. "Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development." *American Journal of Human Genetics* 94, no. 1 (2014): 23-32. https://doi.org/10.1016/j.ajhg.2013.11.009

Rabin, Rachel, Yoel Hirsch, Wendy K. Chung, Josef Ekstein, Ephrat Levy-Lahad, Shachar Zuckerman, Hagar Mor-Shaked, et al. "Expanding the phenotypic spectrum of COLEC10-Related 3MC syndrome: A glimpse into COLEC10-Related 3MC syndrome in the Ashkenazi Jewish population." *American Journal of Medical Genetics Part A* 188, no. 10 (2022): 3110-3117. https://doi.org/10.1002/ajmg.a.62943

Reiber, Judith, Yves Sznajer, Elena G. Posteguillo, Dietmar Müller, Stanislas Lyonnet, Clarisse Baumann, and Walter Just. "Additional clinical and molecular analyses of TFAP2A in

patients with the branchio-oculo-facial syndrome." *American Journal of Medical Genetics Part A* 152A, no.4 (2010): 994-999. https://doi.org/10.1002/ajmg.a.33331

Samocha, Kaitlin E., Elise B. Robinson, Stephan J. Sanders, Christine Stevens, Aniko Sabo, Lauren M. McGrath, Jack A. Kosmicki, et al. "A framework for the interpretation of de novo mutation in human disease." *Nature Genetics* 46 (2014): 944-950. https://doi.org/10.1038/ng.3050

Veltman, Joris A. and Han G. Brunner. "De novo mutations in human genetic disease." *Nature reviews. Genetics* 13, no.8 (2012): 565-575. https://doi.org/10.1038/nrg3241

Vona, Barbara, Indrajit Nanda, Cordula Neuner, Tobias Müller, and Thomas Haaf. "Confirmation of GRHL2 as the gene for the DFNA28 locus." *American Journal of Medical Genetics Part A* 161, no.8 (2013): 2060-2065. https://doi.org/10.1002/ajmg.a.36017

Wang, Kai, Mingyao Li, and Hakon Hakonarson. "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data." *Nucleic Acids Research* 38, no.16 (2010): e164. https://doi.org/10.1093/nar/gkq603

Ware, James S., Kaitlin E. Samocha, Jason Homsy, and Mark J. Daly. "Interpreting *de novo* Variation in Human Disease Using denovolyzeR." *Current Protocols in Human Genetics* 87, no.1 (2015). https://doi.org/10.1002/0471142905.hg0725s87

Watkins, Stephanie E., Robert E. Meyer, Ronald P. Strauss, and Arthur S. Aylsworth. "Classification, Epidemiology, and Genetics of Orofacial Clefts." *Clinics in Plastic Surgery* 41, no.2 (2014):149-163. https://doi.org/10.1016/j.cps.2013.12.003

Yankee, Tara N., Andrea Wilderman, Emma Wentworth Winchester, Jennifer VanOudenhove, and Justin Cotney. "Integrative analysis of transcriptomics in human craniofacial development reveals novel candidate disease genes." *bioRxiv* (2022). https://doi.org/10.1101/2022.02.28.482338