HHS Public Access

Author manuscript

Am J Med Genet A. Author manuscript; available in PMC 2024 October 01.

Published in final edited form as:

Am J Med Genet A. 2023 October ; 191(10): 2558–2570. doi:10.1002/ajmg.a.63336.

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

Kimberly K. Diaz Perez,
Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, 30322, USA

Sydney Chung,
Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, 30322, USA

S. Taylor Head,
Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA

Michael P. Epstein,
Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, 30322, USA

Jacqueline T. Hecht,
Department of Pediatrics, McGovern Medical, School and School of Dentistry, UT Health at Houston, Houston, TX 77030, USA

George L. Wehby,
Department of Health Management and Policy, University of Iowa, Iowa City, IA, 52242, USA

Seth M. Weinberg,
Center for Craniofacial and Dental Genetics, Department of Oral and Craniofacial Sciences, University of Pittsburgh School of Dental Medicine, Pittsburgh, PA, 15213, USA

Jeffrey C. Murray,
Department of Pediatrics, University of Iowa, Iowa City, IA, 52242, USA

Mary L. Marazita,
Center for Craniofacial and Dental Genetics, Department of Oral and Craniofacial Sciences, University of Pittsburgh School of Dental Medicine, Pittsburgh, PA, 15213, USA

Elizabeth J. Leslie
Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, 30322, USA

kdiazpe@emory.edu .

Author Contributions

Kimberly K. Diaz Perez: formal analysis, writing – original draft, writing – review and editing. Sydney Chung: formal analysis, writing – review and editing. S. Taylor Head: formal analysis, writing – original draft, writing – review and editing. Michael P. Epstein: methodology, writing – review and editing. Jacqueline T. Hecht: resources, writing – review and editing. George L. Wehby: resources, writing – review and editing, Seth M. Weinberg: resources, funding acquisition, writing – review and editing, Jeffrey C. Murray: resources, funding acquisition, writing – review and editing, Mary L. Marazita: resources, funding acquisition, writing – review and editing, Elizabeth J. Leslie: conceptualization, project administration, formal analysis, methodology, writing – original draft, writing – review and editing.

Declaration of Interests

The authors declare no competing interests.

Abstract

Exome sequencing (ES) is now a relatively straightforward process to identify causal variants in Mendelian disorders. However, the same is not true for ES in families where the inheritance patterns are less clear, and a complex etiology is suspected. Orofacial clefts (OFCs) are highly heritable birth defects with both Mendelian and complex etiologies. The phenotypic spectrum of OFCs may include overt clefts and several subclinical phenotypes, such as discontinuities in the *orbicularis oris* muscle (OOM) in the upper lip, velopharyngeal insufficiency (VPI), microform clefts or bifid uvulas. We hypothesize that expanding the OFC phenotype to include these phenotypes can clarify inheritance patterns in multiplex families, making them appear more Mendelian. We performed exome sequencing to find rare, likely causal genetic variants in 31 multiplex OFC families, which included families with multiple individuals with OFCs and individuals with subclinical phenotypes. We identified likely causal variants in *COL11A2*, *IRF6*, *SHROOM3, SMC3, TBX3*, and *TP63* in six families. Although we did not find clear evidence supporting the subclinical phenotype hypothesis, our findings support a role for rare variants in the etiology of OFCs.

Keywords

orofacial clefts; exome sequencing; multiplex families; variant segregation

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 affecteds 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, Mossey, Hetmanski, Gowans, Eshete, Adeyemo, Alade, Zeng, Adamson, Naicker, 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, Ludwig, 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 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 second 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 (Supplemental Table 1).

Sequencing

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 and Classification

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), REVEL scores (Ioannidis et al., 2016), 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).

Variants of interest were classified using the American College of Medical Genetics and Genomics (ACMGG) guidelines of pathogenicity using InterVar (Li & Wang, 2017; Richards et al., 2015). We considered variants with a REVEL score ≥ 0.5 to meet the PP3 criteria while variants with a REVEL score < 0.5 met the BP4 criteria. For the pathogenicity criteria that involve allele frequencies, variants that had a maximum allele frequency on gnomAD v2 exomes < 0.001% met the PM2 criteria (PM2: Absent from controls or at an extremely low frequency) and variants that had a maximum allele frequency ≥ 0.005% met the BS1 criteria (BS1: Allele frequency is greater than expected for the disorder). The allele frequency threshold for BS1 was estimated considering a prevalence of OFCs of 1/1000, a penetrance of 50%, allelic heterogeneity of 5%, and genetic heterogeneity of 100% (Whiffin 2017). Lastly, the PP1 criteria was satisfied for variants that segregate across OFC cases in genes previously known to cause OFCs.

Single-Family Segregation Analyses

For single-family analyses, we defined individuals with either an overt cleft or a subclinical phenotype as “affected”. We categorized families as having an apparent dominant inheritance pattern as those with vertical transmission of OFCs/subclinical phenotypes in sequential generations or those with at least 50% of offspring having an OFC or subclinical phenotype. In these families, we analyzed rare, heterozygous variants shared among all

affected individuals. We defined families as having apparently recessive inheritance if there were unaffected parents and no prior family history of OFC with ~25% of offspring having an OFC or subclinical phenotype. In these families, we analyzed homozygous variants or compound heterozygous variants in the affected individuals and both parents had to be carriers for the variant(s). Because smaller families can have ambiguous inheritance patterns, most families were evaluated under both mechanisms. For the analysis of heterozygous variants, we also allowed variants to be present in unaffected relatives to allow for incomplete penetrance, which is common in familial OFC. 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 affected 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. The gene list was compiled from four sources (each downloaded September 4th, 2020): 1) Clinical synopses/genes from the Online Mendelian Inheritance in Man (OMIM) (Amberger et al., 2015) where the clinical synopses included orofacial clefts 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;”, 2) the PreventionGenetics Cleft Lip/Cleft Palate clinical genetic testing panel (PreventionGenetics, Marshfield, WI), 3) the Genomics England PanelApp cleft panel (v2.2), an expert-curated list of genes for familial cleft lip and/or cleft palate, familial isolated clefting, and syndromic clefting (Martin et al., 2019), and 4) literature-curated genes. More information on this gene list can be found in Diaz Perez et al. (2022). For the mixed model analyses, 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%) shared by affected individuals. In the 31 families, we identified an average of 195 variants shared by 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 two families (2/12, 17%) (Figure 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).The variant was not present in gnomAD and was classified as "Pathogenic" according to ACMG guidelines. 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*, and contributed to the "likely benign" ACMG classification. *SEPTIN6* and *TP63* are expressed in the mesenchyme and ectoderm of human embryonic craniofacial tissue, respectively (Yankee et al., 2022) and are only co-expressed in a small fraction of cells (Supplementary Figure 1), so it seems unlikely that these genes directly interact. It is unclear what role *SEPTIN6* could play in craniofacial development as septin6 knockout mice are viable with no obvious phenotypes (Ono et al., 2005), whereas Septin6 is involved in ciliogenesis in the developing zebrafish embryo (Zhai et al., 2014). As many questions about the function of *SEPTIN6* remain, we cannot exclude the possibility that both variants could be contributing independently to OFC risk in this family.

Family 2: We identified a heterozygous missense mutation in *SHROOM3* (NM_020859.4: c.1088A>G; NP_065910.3: p.Gln363Arg) in a non-Hispanice family of European descent 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). Although the variant was classified as a "Variant of Unknown Significance" using ACMG criteria, *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).

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 1).

Family 3: 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. Although this family could not be diagnosed with VWS based on a clinical phenotype alone, this *IRF6* variant is sufficient to render a diagnosis of VWS for this family.

Family 4: 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 have 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 and the variant was classified as a "Variant of Uncertain Significance", inactivation of *TBX3* in the neural crest in mice leads to postnatal death and a highly penetrant cleft palate (López et al., 2018).

Family 5: We identified a novel 32-base pair deletion in *SMC3* (NM_005445.4: c.2019_2050del; NP_005436.1: p.Leu676Argfs*5) in this non-Hispanic family of European descent from the United States. The deletion was classified using ACMG guidelines as "Pathogenic" as it is novel in gnomAD v2 and is predicted to cause a loss-of-function effect. 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, which are less common in SMC3-related CdL than other CdL due to mutations in other genes (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 6: 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 (Nikopensius et al., 2010). This variant was classified as "Likely Pathogenic" based on ACMGG guidelines as it was extremely rare in gnomAD v2 and is predicted to cause a loss-of-function effect on the protein.

Novel Variants in OFC Cohort

We identified various multiplex families with novel variants that were not included in our yield calculations as many are considered variants of unknown significance by ACMG criteria (Figure 2). However, we consider these variants of interest and these genes have been previously implicated in OFCs or craniofacial development and they are expressed in human craniofacial tissues (Yankee et al., 2022). For example, we found a novel missense variant (NM_024915.4: c.38C>T; NP_079191.2: p.Ala13Val) in *GRHL2* shared by two individuals with OFCs from Family 14. Although *GRHL2* has not been directly associated with OFCs, increasing evidence has pointed to the role of *GRHL2* in craniofacial development as Grhl2-deficient mice display a cleft face and split-face malformation (Rifat et al., 2010). GRHL2 is also known to directly interact with TFAP2A, and their interaction is required for AP2a activity during ectoderm lineage commitment (Collier et al., 2023).

We also identified a novel missense variant in *DLG1* (NM_001366207.1: c.175A>G; NP_001353136.1: p.Thr59Ala) in Family 17. Common variants in *DLG1* have been associated with an increased risk for non-syndromic OFCs (Mostowska et al., 2018) and a novel nonsense variant in *DLG1* was previously found in an individual with non-syndromic discontinuous CLP (Demeer et al., 2019). Lastly, we identified a missense variant of interest in *MYH9* in Family 22 (NM_002473.6: c.4745A>G; NP_002464.1: p.Glu1582Gly). Linkage and association studies of common variants have found *MYH9* to be associated with OFC susceptibility (Birnbaum, Reutter, et al., 2009; Chiquet et al., 2009).

Recurrent Genes with Variants of Interest in OFC Cohort

In this study, we also identified genes with more than one variant of interest and these included *AFDN*, *COL11A2*, *PVR*, and *SHROOM3* (Figure 2). The gene with the most variants of interest was *COL11A2*, with four variants of interest (one loss-of-function in Family 6 and missense variants in Families 10, 12, and 15). In addition, we identified two variants in *AFDN* in two separate families, including a novel missense variant in Family 8 (NM_001386888.1: c.3545A>G; NP_001373817.1: p.Asn1182Ser) and an extremely rare splice donor variant in Family 22 (NM_001386888.1: c.1222+1G>T). et al. (2022) recently reported rare heterozygous and compound heterozygous mutations in *AFDN*. *C*onditional deletion of Afdn in the palatal epithelium in mice causes a highly penetrant cleft palate (Lough et al., 2020).

We found two pedigrees (Family 19 and 26) with missense variants of interest in *PVR*. Rare *PVR* variants in OFCs have been identified but none reached statistical significance in case-control studies (Sözen et al., 2009) but there have not been familial reports of rare variants published to date. Lastly, we identified three missense variants in Families 2, 12, and 15 in *SHROOM3*, which is involved in neural tube morphogenesis and closure (Haigo et al., 2003; Hildebrand & Soriano, 1999).

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 (Supplemental Table 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. There were 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 3). We then adjusted for affected status and relatedness and found there was no difference in the number of variants in all protein-coding genes (p=0.46) or OFC genes (p=0.64) between individuals with OFCs and subclinical phenotypes. The same was true when restricting to rare variants with a CADD score ≥ 20 (Figure 3; p=0.27 for protein-coding genes and p=0.44 for OFC genes) or a REVEL score ≥ 0.5 (Figure 3; p=0.33 for protein-coding genes and p=0.27 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 ~17–21% 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=0.33, 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. 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 the 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 and common 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. As individuals with a positive family history might have questions about risks, consideration should be given to sequencing studies to identify variants that might suggest higher recurrence risks than what epidemiologic studies alone would support.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to express our gratitude to the families involved in this study for their participation and we would like to thank our collaborators who have helped develop and produce this work. This study was supported with funding from the National Institutes of Health (X01-HG010012, T32-GM008490, R00-DE025060, R01-DE016148, R01-DE011931, R01-DE09886, R01-DE012472, P50-DE016215, R37-DE008559, R01-DE014667, R01-DD000295) and the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program. Sequencing services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268201700006I.

Data Availability

Sequence and phenotype data is available from the Database of Genotypes and Phenotypes (dbGaP), study accession: phs001675.v1.p1.

REFERENCES

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

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, & 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. 10.1093/nar/gku1205 [PubMed: 25428349]

Awotoye W, Mossey PA, Hetmanski JB, Gowans LJJ, Eshete MA, Adeyemo WL, Alade A, Zeng E, Adamson O, James O, Fashina A, Ogunlewe MO, Naicker T, Adeleke C, Busch T, Li M, Petrin A, Oladayo A, Kayali S, . . . Butali A. (2022). Damaging Mutations in AFDN Contribute to Risk of Nonsyndromic Cleft Lip With or Without Cleft Palate. Cleft Palate Craniofac J, 10556656221135926. 10.1177/10556656221135926 [PubMed: 36384317]

Awotoye W, Mossey PA, Hetmanski JB, Gowans LJJ, Eshete MA, Adeyemo WL, Alade A, Zeng E, Adamson O, Naicker T, Anand D, Adeleke C, Busch T, Li M, Petrin A, Aregbesola BS, Braimah RO, Oginni FO, Oladele AO, . . . Butali A. (2022). Whole-genome sequencing reveals de-novo mutations associated with nonsyndromic cleft lip/palate. Sci Rep, 12(1), 11743. 10.1038/s41598-022-15885-1 [PubMed: 35817949]

Basha M, Demeer B, Revencu N, Helaers R, Theys S, Bou Saba S, Boute O, Devauchelle B, Francois G, Bayet B, & 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. 10.1136/jmedgenet-2017-105110 [PubMed: 29500247]

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

Beames TG, & Lipinski RJ (2020). Gene-environment interactions: aligning birth defects research with complex etiology. Development, 147(21). 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). 10.12688/f1000research.9503.1

Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M, Baluardo C, Ferrian M, Almeida de Assis N, Alblas MA, Barth S, Freudenberg J, Lauster C, Schmidt G, Scheer M, Braumann B, Berge SJ, Reich RH, Schiefke F, . . . Mangold E. (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet, 41(4), 473–477. 10.1038/ng.333 [PubMed: 19270707]

Birnbaum S, Reutter H, Mende M, de Assis NA, Diaz-Lacava A, Herms S, Scheer M, Lauster C, Braumann B, Schmidt G, Martini M, Hemprich A, Pötzsch S, Knapp M, Nöthen MM, Kramer FJ, & Mangold E (2009). Further evidence for the involvement of MYH9 in the etiology of non-syndromic cleft lip with or without cleft palate. Eur J Oral Sci, 117(2), 200–203. 10.1111/j.1600-0722.2008.00604.x [PubMed: 19320731]

Bishop MR, Diaz Perez KK, Sun M, Ho S, Chopra P, Mukhopadhyay N, Hetmanski JB, Taub MA, Moreno-Uribe LM, Valencia-Ramirez LC, Restrepo Muñeton CP, Wehby G, Hecht JT, Deleyiannis F, Weinberg SM, Wu-Chou YH, Chen PK, Brand H, Epstein MP, . . . Leslie EJ. (2020). Genome-wide Enrichment of De Novo Coding Mutations in Orofacial Cleft Trios. The American Journal of Human Genetics. 10.1016/j.ajhg.2020.05.018

Bureau A, Parker MM, Ruczinski I, Taub MA, Marazita ML, Murray JC, Mangold E, Noethen MM, Ludwig KU, Hetmanski JB, Bailey-Wilson JE, Cropp CD, Li Q, Szymczak S, Albacha-Hejazi H, Alqosayer K, Field LL, Wu-Chou YH, Doheny KF, . . . Beaty TH. (2014). Whole exome sequencing of distant relatives in multiplex families implicates rare variants in candidate genes for oral clefts. Genetics, 197(3), 1039–1044. 10.1534/genetics.114.165225 [PubMed: 24793288]

Carlson JC, Standley J, Petrin A, Shaffer JR, Butali A, Buxó CJ, Castilla E, Christensen K, Deleyiannis FW, Hecht JT, Field LL, Garidkhuu A, Moreno Uribe LM, Nagato N, Orioli IM, Padilla C, Poletta F, Suzuki S, Vieira AR, . . . Leslie EJ. (2017). Identification of 16q21 as a modifier of nonsyndromic orofacial cleft phenotypes. Genet Epidemiol, 41(8), 887–897. 10.1002/gepi.22090 [PubMed: 29124805]

Chiquet BT, Hashmi SS, Henry R, Burt A, Mulliken JB, Stal S, Bray M, Blanton SH, & Hecht JT (2009). Genomic screening identifies novel linkages and provides further evidence for a role of MYH9 in nonsyndromic cleft lip and palate. Eur J Hum Genet, 17(2), 195–204. 10.1038/ejhg.2008.149 [PubMed: 18716610]

Collier AE, Piekos SN, Liu A, Pattison JM, Felix F, Bailetti AA, Sedov E, Gaddam S, Zhen H, & Oro AE (2023). GRHL2 and AP2a coordinate early surface ectoderm lineage commitment during development. iScience, 26(3), 106125. 10.1016/j.isci.2023.106125 [PubMed: 36843855]

Copp AJ, & Greene ND (2013). Neural tube defects--disorders of neurulation and related embryonic processes. Wiley Interdiscip Rev Dev Biol, 2(2), 213–227. 10.1002/wdev.71 [PubMed: 24009034]

Cox LL, Cox TC, Moreno Uribe LM, Zhu Y, Richter CT, Nidey N, Standley JM, Deng M, Blue E, Chong JX, Yang Y, Carstens RP, Anand D, Lachke SA, Smith JD, Dorschner MO, Bedell B, Kirk E, Hing AV, . . . 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. 10.1016/j.ajhg.2018.04.009 [PubMed: 29805042]

Das D, Zalewski JK, Mohan S, Plageman TF, VanDemark AP, & Hildebrand JD (2014). The interaction between Shroom3 and Rho-kinase is required for neural tube morphogenesis in mice. Biol Open, 3(9), 850–860. 10.1242/bio.20147450 [PubMed: 25171888]

Demeer B, Revencu N, Helaers R, Gbaguidi C, Dakpe S, François G, Devauchelle B, Bayet B, & Vikkula M (2019). Likely Pathogenic Variants in One Third of Non-Syndromic Discontinuous Cleft Lip and Palate Patients. Genes (Basel), 10(10). 10.3390/genes10100833

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, & Daly MJ (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet, 43(5), 491–498. 10.1038/ng.806 [PubMed: 21478889]

Diaz Perez KK, Curtis SW, Sanchis-Juan A, Zhao X, Head ST, Ho S, Carter B, McHenry T, Bishop MR, Valencia-Ramirez LC, Restrepo C, Hecht JT, Uribe LM, Wehby G, Weinberg SM, Beaty TH, Murray JC, Feingold E, Marazita ML, . . . Leslie EJ. (2022). Rare variants found in clinical gene panels illuminate the genetic and allelic architecture of orofacial clefting. medRxiv, 2022.2008.2015.22278795. 10.1101/2022.08.15.22278795

Dixon MJ, Marazita ML, Beaty TH, & Murray JC (2011). Cleft lip and palate: understanding genetic and environmental influences. Nat Rev Genet, 12(3), 167–178. 10.1038/nrg2933 [PubMed: 21331089]

Garland MA, Reynolds K, & Zhou CJ (2020). Environmental mechanisms of orofacial clefts. Birth defects research, 112(19), 1660–1698. 10.1002/bdr2.1830 [PubMed: 33125192]

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. 10.1136/jmg.2003.009274 [PubMed: 14757865]

Gil-Rodriguez MC, Deardorff MA, Ansari M, Tan CA, Parenti I, Baquero-Montoya C, Ousager LB, Puisac B, Hernandez-Marcos M, Teresa-Rodrigo ME, Marcos-Alcalde I, Wesselink JJ, Lusa-Bernal S, Bijlsma EK, Braunholz D, Bueno-Martinez I, Clark D, Cooper NS, Curry CJ, . . . 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. 10.1002/humu.22761 [PubMed: 25655089]

Haigo SL, Hildebrand JD, Harland RM, & Wallingford JB (2003). Shroom induces apical constriction and is required for hingepoint formation during neural tube closure. Curr Biol, 13(24), 2125–2137. 10.1016/j.cub.2003.11.054 [PubMed: 14680628]

Hamosh A, Scott AF, Amberger JS, Bocchini CA, & McKusick VA (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res, 33(Database issue), D514–D517. 10.1093/nar/gki033 [PubMed: 15608251]

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

Hildebrand JD, & Soriano P (1999). Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice. Cell, 99(5), 485–497. 10.1016/s0092-8674(00)81537-8 [PubMed: 10589677]

Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, . . . Sieh W. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet, 99(4), 877–885. 10.1016/j.ajhg.2016.08.016 [PubMed: 27666373]

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, . . . MacArthur DG. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. bioRxiv, 531210. 10.1101/531210

Khandelwal KD, van den Boogaard MH, Mehrem SL, Gebel J, Fagerberg C, van Beusekom E, van Binsbergen E, Topaloglu O, Steehouwer M, Gilissen C, Ishorst N, van Rooij I, Roeleveld N, Christensen K, Schoenaers J, Berge S, Murray JC, Hens G, Devriendt K, . . . van Bokhoven H. (2019). Deletions and loss-of-function variants in TP63 associated with orofacial clefting. Eur J Hum Genet, 27(7), 1101–1112. 10.1038/s41431-019-0370-0 [PubMed: 30850703]

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

Kline AD, Krantz ID, Sommer A, Kliewer M, Jackson LG, FitzPatrick DR, Levin AV, & 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. 10.1002/ajmg.a.31757 [PubMed: 17508425]

Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL, Daack-Hirsch S, Sander A, McDonald-McGinn DM, Zackai EH, Lammer EJ, Aylsworth AS, Ardinger HH, Lidral AC, Pober BR, Moreno L, Arcos-Burgos M, . . . Murray JC. (2002). Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. Nat Genet, 32(2), 285–289. 10.1038/ng985 [PubMed: 12219090]

Kotlar AV, Trevino CE, Zwick ME, Cutler DJ, & Wingo TS (2018). Bystro: rapid online variant annotation and natural-language filtering at whole-genome scale. Genome Biol, 19(1), 14. 10.1186/s13059-018-1387-3 [PubMed: 29409527]

Kryukov GV, Pennacchio LA, & Sunyaev SR (2007). Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am J Hum Genet, 80(4), 727–739. 10.1086/513473 [PubMed: 17357078]

Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, . . . Maglott DR. (2018). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res, 46(D1), D1062–d1067. 10.1093/nar/gkx1153 [PubMed: 29165669]

Lansdon LA, Dickinson A, Arlis S, Liu H, Hlas A, Hahn A, Bonde G, Long A, Standley J, Tyryshkina A, Wehby G, Lee NR, Daack-Hirsch S, Mohlke K, Girirajan S, Darbro BW, Cornell RA, Houston DW, Murray JC, & Manak JR (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. 10.1016/j.ajhg.2022.11.012 [PubMed: 36493769]

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, . . . MacArthur DG. (2016). Analysis of protein-coding

genetic variation in 60,706 humans. Nature, 536(7616), 285–291. 10.1038/nature19057 [PubMed: 27535533]

Leslie EJ (2022). Genetic models and approaches to study orofacial clefts. Oral Dis, 28(5), 1327–1338. 10.1111/odi.14109 [PubMed: 34923716]

Leslie EJ, Carlson JC, Shaffer JR, Butali A, Buxo CJ, Castilla EE, Christensen K, Deleyiannis FW, Leigh Field L, Hecht JT, Moreno L, Orioli IM, Padilla C, Vieira AR, Wehby GL, Feingold E, Weinberg SM, Murray JC, Beaty TH, & Marazita ML (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. 10.1007/s00439-016-1754-7 [PubMed: 28054174]

Leslie EJ, Liu H, Carlson JC, Shaffer JR, Feingold E, Wehby G, Laurie CA, Jain D, Laurie CC, Doheny KF, McHenry T, Resick J, Sanchez C, Jacobs J, Emanuele B, Vieira AR, Neiswanger K, Standley J, Czeizel AE, . . . Marazita ML. (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. 10.1016/j.ajhg.2016.02.014 [PubMed: 27018472]

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

Li Q, & Wang K (2017). InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet, 100(2), 267–280. 10.1016/j.ajhg.2017.01.004 [PubMed: 28132688]

Liu H, Busch T, Eliason S, Anand D, Bullard S, Gowans LJJ, Nidey N, Petrin A, Augustine-Akpan E-A, Saadi I, Dunnwald M, Lachke SA, Zhu Y, Adeyemo A, Amendt B, Roscioli T, Cornell R, Murray 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. 10.1002/bdra.23596 [PubMed: 28029220]

López SH, Avetisyan M, Wright CM, Mesbah K, Kelly RG, Moon AM, & Heuckeroth RO (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. 10.1016/j.ydbio.2018.09.017 [PubMed: 30292786]

Lough KJ, Spitzer DC, Bergman AJ, Wu JJ, Byrd KM, & Williams SE (2020). Disruption of the nectin-afadin complex recapitulates features of the human cleft lip/palate syndrome CLPED1. Development, 147(21). 10.1242/dev.189241

Mangold E, Ludwig KU, Birnbaum S, Baluardo C, Ferrian M, Herms S, Reutter H, de Assis NA, Chawa TA, Mattheisen M, Steffens M, Barth S, Kluck N, Paul A, Becker J, Lauster C, Schmidt G, Braumann B, Scheer M, . . . Nothen MM. (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat Genet, 42(1), 24–26. 10.1038/ng.506 [PubMed: 20023658]

Marazita ML (2012). The evolution of human genetic studies of cleft lip and cleft palate. Annu Rev Genomics Hum Genet, 13, 263–283. 10.1146/annurev-genom-090711-163729 [PubMed: 22703175]

Marosy BA, Craig BD, Hetrick KN, Witmer PD, Ling H, Griffith SM, Myers B, Ostrander EA, Stanford JL, Brody LC, & Doheny KF (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. 10.1002/cphg.27

Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, Leong IUS, Smith KR, Gerasimenko O, Haraldsdottir E, Thomas E, Scott RH, Baple E, Tucci A, Brittain H, de Burca A, Ibañez K, Kasperaviciute D, Smedley D, . . . McDonagh EM. (2019). PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet, 51(11), 1560–1565. 10.1038/s41588-019-0528-2 [PubMed: 31676867]

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, & DePristo MA (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res, 20(9), 1297–1303. 10.1101/gr.107524.110 [PubMed: 20644199]

Mostowska A, Gaczkowska A, Żukowski K, Ludwig KU, Hozyasz KK, Wójcicki P, Mangold E, Böhmer AC, Heilmann-Heimbach S, Knapp M, Zadurska M, Biedziak B, Budner M, Lasota A,

Daktera-Micker A, & Jagodziński PP (2018). Common variants in DLG1 locus are associated with non-syndromic cleft lip with or without cleft palate [10.1111/cge.13141]. Clin Genet, 93(4), 784–793. 10.1111/cge.13141 [PubMed: 28926086]

Neiswanger K, Walker K, Klotz CM, Cooper ME, Bardi KM, Brandon CA, Weinberg SM, Vieira AR, Martin RA, Czeizel AE, Castilla EE, Poletta FA, & Marazita ML (2009). Whorl patterns on the lower lip are associated with nonsyndromic cleft lip with or without cleft palate [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. Am J Med Genet A, 149A(12), 2673–2679. 10.1002/ajmg.a.33089 [PubMed: 19921634]

Neiswanger K, Weinberg SM, Rogers CR, Brandon CA, Cooper ME, Bardi KM, Deleyiannis FW, Resick JM, Bowen A, Mooney MP, de Salamanca JE, González B, Maher BS, Martin RA, & Marazita ML (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. 10.1002/ajmg.a.31760 [PubMed: 17497721]

Nikopensius T, Jagomägi T, Krjutskov K, Tammekivi V, Saag M, Prane I, Piekuse L, Akota I, Barkane B, Krumina A, Ambrozaityte L, Matuleviciene A, Kucinskiene ZA, Lace B, Kucinskas V, & 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. 10.1002/bdra.20700 [PubMed: 20672350]

Ono R, Ihara M, Nakajima H, Ozaki K, Kataoka-Fujiwara Y, Taki T, Nagata K, Inagaki M, Yoshida N, Kitamura T, Hayashi Y, Kinoshita M, & Nosaka T (2005). Disruption of Sept6, a fusion partner gene of MLL, does not affect ontogeny, leukemogenesis induced by MLL-SEPT6, or phenotype induced by the loss of Sept4. Mol Cell Biol, 25(24), 10965–10978. 10.1128/mcb.25.24.10965-10978.2005 [PubMed: 16314519]

Peyrard-Janvid M, Leslie EJ, Kousa YA, Smith TL, Dunnwald M, Magnusson M, Lentz BA, Unneberg P, Fransson I, Koillinen HK, Rautio J, Pegelow M, Karsten A, Basel-Vanagaite L, Gordon W, Andersen B, Svensson T, Murray JC, Cornell RA, . . . Schutte BC. . (2014). Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. Am J Hum Genet, 94(1), 23–32. 10.1016/j.ajhg.2013.11.009 [PubMed: 24360809]

Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, Domann FE, Govil M, Christensen K, Bille C, Melbye M, Jugessur A, Lie RT, Wilcox AJ, Fitzpatrick DR, Green ED, Mossey PA, Little J, Steegers-Theunissen RP, . . . Murray JC. (2008). Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. Nat Genet, 40(11), 1341–1347. 10.1038/ng.242 [PubMed: 18836445]

Ray D, Venkataraghavan S, Zhang W, Leslie EJ, Hetmanski JB, Weinberg SM, Murray JC, Marazita ML, Ruczinski I, Taub MA, & Beaty TH (2021). Pleiotropy method reveals genetic overlap between orofacial clefts at multiple novel loci from GWAS of multi-ethnic trios. PLoS Genet, 17(7), e1009584. 10.1371/journal.pgen.1009584 [PubMed: 34242216]

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

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, & Rehm HL (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. 10.1038/gim.2015.30 [PubMed: 25741868]

Rifat Y, Parekh V, Wilanowski T, Hislop NR, Auden A, Ting SB, Cunningham JM, & Jane SM (2010). Regional neural tube closure defined by the Grainy head-like transcription factors. Developmental Biology, 345(2), 237–245. 10.1016/j.ydbio.2010.07.017 [PubMed: 20654612]

Sözen MA, Hecht JT, & Spritz RA (2009). Mutation and association analysis of the PVR and PVRL2 genes in patients with non-syndromic cleft lip and palate. Genet Mol Biol, 32(3), 466–469. 10.1590/s1415-47572009000300007 [PubMed: 21637507]

Stanier P, & Moore GE (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. 10.1093/hmg/ddh052 [PubMed: 14722155]

Suzuki S, Marazita ML, Cooper ME, Miwa N, Hing A, Jugessur A, Natsume N, Shimozato K, Ohbayashi N, Suzuki Y, Niimi T, Minami K, Yamamoto M, Altannamar TJ, Erkhembaatar T, Furukawa H, Daack-Hirsch S, L'Heureux J, Brandon CA, . . . Murray JC. (2009). Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. Am J Hum Genet, 84(3), 406–411. 10.1016/j.ajhg.2009.02.002 [PubMed: 19249007]

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

van Steensel MA, Buma P, de Waal Malefijt MC, van den Hoogen FH, & Brunner HG (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. 10.1002/(sici)1096-8628(19970613)70:3<315::aid-ajmg19>3.3.co;2-y [PubMed: 9188673]

Vikkula M, Mariman EC, Lui VC, Zhidkova NI, Tiller GE, Goldring MB, van Beersum SE, de Waal Malefijt MC, van den Hoogen FH, Ropers HH, & et al. (1995). Autosomal dominant and recessive osteochondrodysplasias associated with the COL11A2 locus. Cell, 80(3), 431–437. 10.1016/0092-8674(95)90493-x [PubMed: 7859284]

Weinberg SM, Neiswanger K, Martin RA, Mooney MP, Kane AA, Wenger SL, Losee J, Deleyiannis F, Ma L, De Salamanca JE, Czeizel AE, & Marazita ML (2006). The Pittsburgh Oral-Facial Cleft study: expanding the cleft phenotype. Background and justification. Cleft Palate Craniofac J, 43(1), 7–20. 10.1597/04-122r1.1 [PubMed: 16405378]

Yankee TN, Wilderman A, Winchester EW, VanOudenhove J, & Cotney J (2022). Integrative analysis of transcriptomics in human craniofacial development reveals novel candidate disease genes. bioRxiv, 2022.2002.2028.482338. 10.1101/2022.02.28.482338

Yu Y, Zuo X, He M, Gao J, Fu Y, Qin C, Meng L, Wang W, Song Y, Cheng Y, Zhou F, Chen G, Zheng X, Wang X, Liang B, Zhu Z, Fu X, Sheng Y, Hao J, . . . 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. 10.1038/ncomms14364 [PubMed: 28232668]

Zhai G, Gu Q, He J, Lou Q, Chen X, Jin X, Bi E, & Yin Z (2014). Sept6 is required for ciliogenesis in Kupffer's vesicle, the pronephros, and the neural tube during early embryonic development. Mol Cell Biol, 34(7), 1310–1321. 10.1128/mcb.01409-13 [PubMed: 24469395]

Zieger HK, Weinhold L, Schmidt A, Holtgrewe M, Juranek SA, Siewert A, Scheer AB, Thieme F, Mangold E, Ishorst N, Brand FU, Welzenbach J, Beule D, Paeschke K, Krawitz PM, & Ludwig KU (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. 10.1016/j.xhgg.2022.100166 [PubMed: 36589413]

Zucchero TM, Cooper ME, Maher BS, Daack-Hirsch S, Nepomuceno B, Ribeiro L, Caprau D, Christensen K, Suzuki Y, Machida J, Natsume N, Yoshiura K-I, Vieira AR, Orioli IM, Castilla EE, Moreno L, Arcos-Burgos M, Lidral AC, Field LL, . . . Murray JC. (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. 10.1056/NEJMoa032909 [PubMed: 15317890]

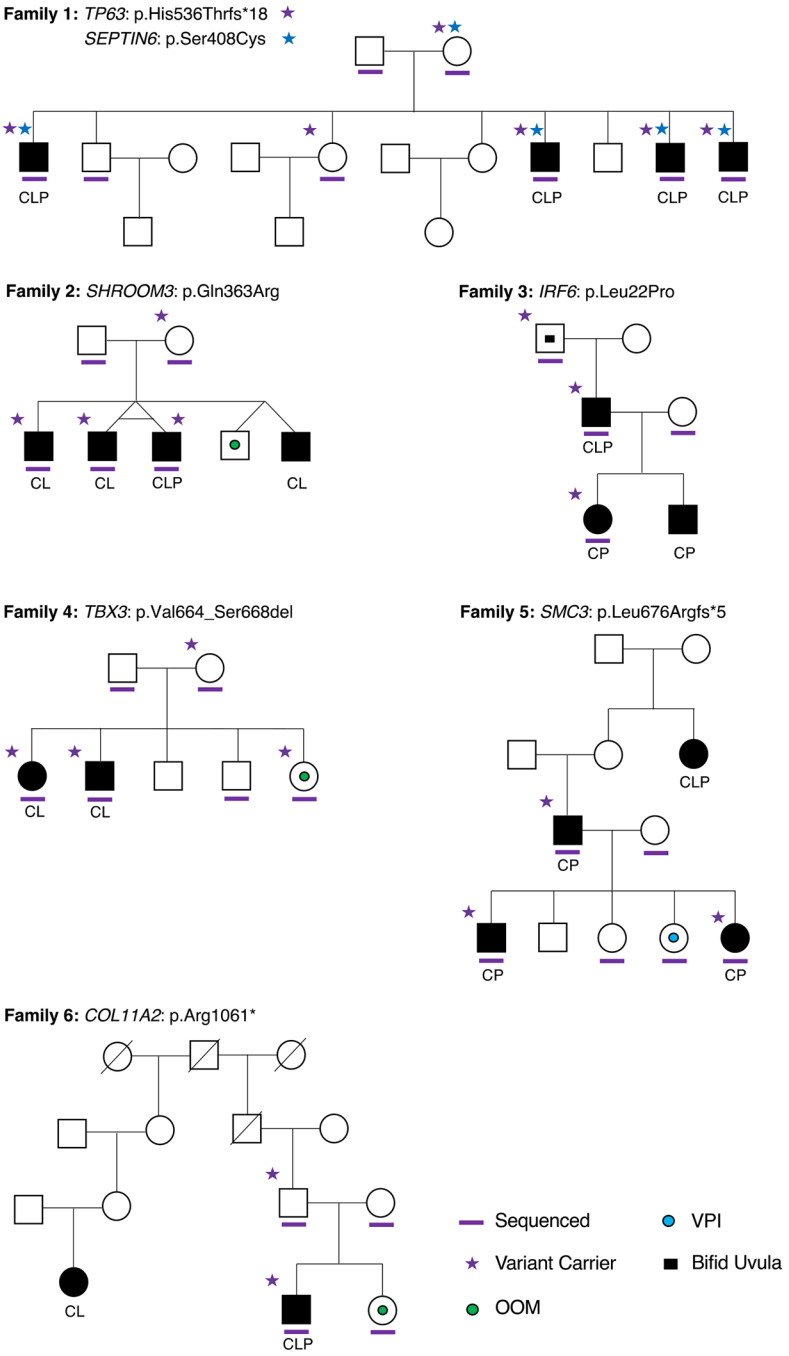

Figure 1. Likely Causal Variants in Multiplex OFC Families.

We found likely causal variants in *TP63, SEPTIN6*, *SHROOM3, IRF6*, *TBX3*, *SMC3*, and *COL11A2* in six families. Symbols that are fully shaded indicate that the individual exhibits an OFC: 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 exome sequencing data while the solid stars indicate variant carriers.

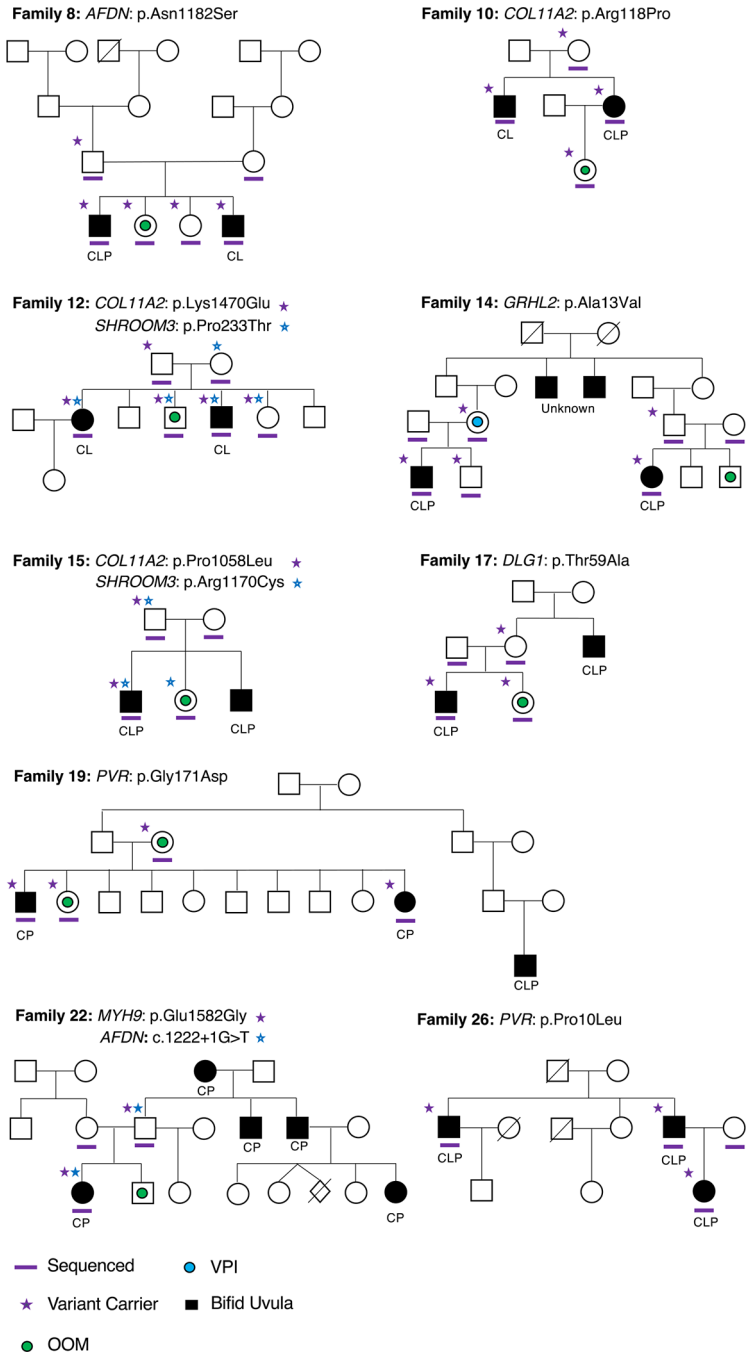

Figure 2. Additional Variants of Interest in Multiplex OFC Families.

Symbols that are fully shaded indicate that the individual exhibits an OFC: CL (cleft lip), CP (cleft palate), and CLP (cleft lip and palate). The symbol with a green circle indicates individuals with discontinuities in the orbicularis oris muscle (OOM), the blue circle indicates individuals with velopharyngeal insufficiency (VPI), and the black solid square inside the symbol indicates the sample had a bifid uvula. The purple solid lines highlight individuals with available exome sequencing data while the solid stars represent variant carriers.

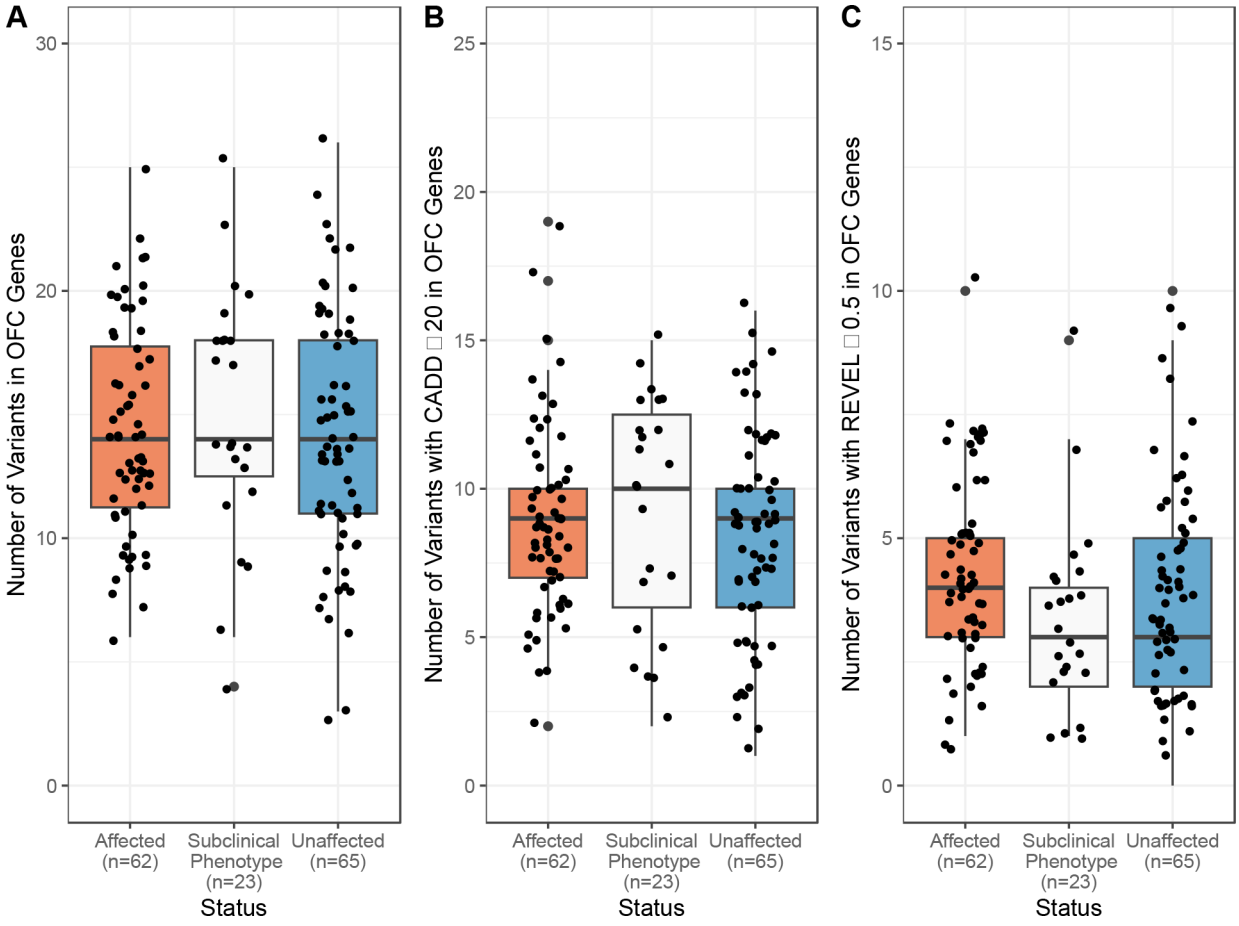

Figure 3. Number of Variants in OFC Genes per Person Within Groups.

The number of variants in genes associated with OFCs per person (A) overall, (B) variants with a CADD ≥ 20, and (C) variants with a REVEL score ≥ 0.5 across affected status, including affected individuals (n=62, orange), individuals with subclinical phenotypes (n=23, gray) and unaffected (n=65, blue) individuals.