

Rare genetic variants in *SEC24D* modify orofacial cleft phenotypes

Sarah W. Curtis¹, Jenna C. Carlson^{2,3}, Terri H. Beaty⁴, Jeffrey C. Murray⁵, Seth M. Weinberg^{3,6}, Mary L. Marazita^{3,6}, Justin L. Cotney⁷, David J. Cutler¹, Michael P. Epstein¹, Elizabeth J. Leslie^{1*}

1) Department of Human Genetics, Emory University, Atlanta, GA, 30322, USA, 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, 15621, USA, 3) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, 15261, USA, 4) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA, 5) Department of Pediatrics, University of Iowa, Iowa City, IA, 52242, USA, 6) Center for Craniofacial and Dental Genetics, Department of Oral and Craniofacial Sciences, University of Pittsburgh, Pittsburgh, PA, 15261, USA, 7) Department of Genetics and Genome Sciences, University of Connecticut, CT, 06030, USA Corresponding Author: ejlesli@emory.edu

Abstract

As one of the most common structural birth defects, orofacial clefts (OFCs) have been studied for decades, and recent studies have demonstrated that there are genetic differences between the different phenotypic presentations of OFCs. However, the contribution of rare genetic variation genome-wide to different subtypes of OFCs has been understudied, with most studies focusing on common genetic variation or rare variation within targeted regions of the genome. Therefore, we used whole-genome sequencing data from the Gabriella Miller Kids First Pediatric Research Program to conduct a gene-based burden analysis to test for genetic modifiers of cleft lip (CL) vs cleft lip and palate (CLP). We found that there was a significantly increased burden of rare variants in *SEC24D* in CL cases compared to CLP cases ($p=6.86 \times 10^{-7}$). Of the 15 variants within *SEC24D*, 53.3% were synonymous, but overlapped a known craniofacial enhancer. We then tested whether these variants could alter predicted transcription factor binding sites (TFBS), and found that the rare alleles destroyed binding sites for 9 transcription factors (TFs), including Pax1 ($p=0.0009$), and created binding sites for 23 TFs, including Pax6 ($p=6.12 \times 10^{-5}$) and Pax9 ($p=0.0001$), which are known to be involved in normal craniofacial development, suggesting a potential mechanism by which these synonymous variants could have a functional impact. Overall, this study demonstrates that rare genetic variation contributes to the phenotypic heterogeneity of OFCs and suggests that regulatory variation may also contribute and warrant further investigation in future studies of genetic variants controlling risk to OFC.

Introduction:

Orofacial clefts (OFCs) are one of the most common structural birth defects, affecting 1 in 700 births (Ciminello et al. 2009; Dixon et al. 2011; Gundlach and Maus 2006; Mossey et al. 2009; Mossey and Modell 2012; Sabbagh et al. 2012). OFCs require multiple surgeries early in life, and affected individuals can experience speech, hearing, and psychosocial problems throughout life (Hameed et al. 2019; Lancaster et al. 2020; Wehby and Cassell 2010; Wehby et al. 2006). A majority (~50-70%) of OFCs cases are classified as nonsyndromic and most of these nonsyndromic cases are hypothesized to be caused by complex or multifactorial etiologies (Leslie and Marazita 2013), having both environmental (Bell et al. 2014; Candotto et al. 2019; DeRoo et al. 2016; Jamilian et al. 2017; Li et al. 2010; Martelli et al. 2015; Yin et al. 2019) and genetic risk factors (Birnbaum et al. 2009a; Birnbaum et al. 2009b; Bishop et al. 2020; Duan et al. 2020; Hao et al. 2018; Leslie et al. 2017a; Leslie et al. 2017b; Leslie et al. 2016; Leslie and Marazita 2013; Li et al. 2020; Li et al. 2017; Ludwig et al. 2016; Ludwig et al. 2017; Mangold et al. 2010; Mukhopadhyay et al. 2020; Shaffer et al. 2019).

OFCs are also phenotypically heterogeneous and can be classified as a cleft of the lip (CL), cleft of the palate (CP), or a cleft of both the lip and the palate (CLP). A majority of the previous epidemiological and genetic studies have grouped types of OFCs together for increased statistical power under a hypothesis of shared biology (Ciminello et al. 2009; Dixon et al. 2011; Leslie et al. 2016; Leslie and Marazita 2013). CL and CLP are most commonly grouped together because of a shared 2:1 male-to-female sex bias and the fact that formation of the lip precedes the formation of the palate during embryonic development (Amidei et al. 1994; Marazita 2012; Urbanova et al. 2013). However, while genetic studies have successfully identified loci that contribute to both CL and CLP (Beaty et al. 2010; Birnbaum et al. 2009a; Birnbaum et al. 2009b; Leslie et al. 2016; Mukhopadhyay et al. 2020), recent studies indicate that CL and CLP can also have distinct genetic risk factors (Carlson et al. 2019; Carlson et al. 2017a; Carlson et al. 2017b; Curtis et al. 2021a; Curtis et al. 2021b; Huang et al. 2019), or genetic modifiers, that predispose for the formation of one cleft type over another. We previously performed a targeted sequencing study of known OFC-associated regions and found evidence that common variants near 9q22, 17p22, and 20q12 are associated with the formation of CL or CLP and rare variants in 7 genes are associated with cleft type (Carlson et al. 2017b). Additional targeted studies have shown that some genotyped SNPs in *IRF6* have a stronger effect in CL compared to CLP (Carlson et al. 2019). Genome-wide, subtype specific effects have been found for CL (Huang et al. 2019), and we also showed in a genome-wide modifier analysis that common variants in 16q21 are more strongly associated with the formation of a CL over a CLP, as are low-frequency variants in *C8orf34*, *TMEM246*, and *CDC42EP3* (Carlson et al. 2017a).

These studies indicate that common and low-frequency genetic variants are associated with the phenotypic variability among OFCs. However less is known about the contribution of rare variants genome-wide to the phenotypic heterogeneity of OFCs given the relatively limited scope of previous studies of rare variants (Carlson et al. 2017b). Therefore, we set out to test whether any genes had a burden of rare variants using whole genome sequencing data from the Gabriella Kids First Pediatric Research Program, focusing on CL vs CLP cases.

Methods:

Cohort Description

This study used affected probands with cleft lip (CL) or cleft lip and palate (CLP) and their parents that were sequenced as part of the Gabriella Miller Kids First Pediatric Research Program (GMKF). Participants were recruited from the United States, Argentina, Turkey, Hungary, Spain, Colombia, and Taiwan. Participant recruitment was done at regional treatment or research centers after review and approval by each site's institutional review board (IRB) and the IRB of the affiliated US institutions (e.g., the University of Iowa, the University of Pittsburgh, and Johns Hopkins University). Because we were interested in genetic differences in CL vs CLP cases and 93.5% of the CL cases were in the samples with European ancestry, only the subset of samples with European ancestry were retained for analysis. For subtype-specific analyses (described below), 76 trios with isolated CL and 225 with CLP were used. The 20 trios (representing 6.6% of the dataset) with affected parents were retained in this analysis, but trios from families with both CL and CLP were excluded. For the modifier analysis (described below), we randomly selected a subset of unrelated individuals, with 100 having CL and 269 having CLP. For both analyses, the proportion of male and female probands were similar: 67% male, 33% female for the subtype-specific analysis and 66%, 34% female for the modifier analysis (Table S1).

Variant calling and QC

The whole-genome sequencing and quality control methodology for these cohorts has been described in detail previously (Bishop et al. 2020; Mukhopadhyay et al. 2020). Briefly, DNA from either blood or saliva samples were sequenced (depending on availability) at either the McDonnell Genome Institute (MGI), Washington University School of Medicine in St. Louis or the Broad Institute. Variants were called and aligned to the GRCh38/hg38 reference genome using the GATK pipeline (Conrad et al. 2011; McKenna et al. 2010; Van der Auwera et al. 2013) at the GMKF Data Resource Center. The details of how these datasets were aligned and genotyped and harmonized has been previously published (Bishop et al. 2020; Mukhopadhyay et al. 2020). For sample-level QC, individuals with missingness or Mendelian error rare, or average read depth outside of three standard deviations from the mean were removed, as were individuals with a transition/transversion ratio (Ts/Tv), exonic Ts/Tv, silent/replacement ratio, or heterozygosity/homozygosity ratio outside expectation (Bishop et al. 2020). PLINK (v1.9) (Chang et al. 2015) was used to calculate identity-by-descent (IBD) to confirm all family relationships and X chromosome heterozygosity to confirm the sex of all participants. For variant-level QC, non-passing variants, variants with more than 2 Mendelian errors, variants with >5% missingness in individuals, variants with a Hardy-Weinberg equilibrium p-value < 1×10^{-7} in unaffected parents, and variants with a quality by depth (QD) score of <4 were removed. Additionally, calls with a genotype quality (GQ)<20 or a depth (DP)<10 were set to missing and only biallelic variants were retained using VCFTools (Danecek et al. 2011).

Annotation and filtering

All variants were annotated to publicly available databases using ANNOVAR (v201707) (Wang et al. 2010) and Variant Effect Predictor (VEP) (McLaren et al. 2016). Variants were defined as rare if their minor allele frequency (MAF) was less than 1% or 0.1% in all populations in gnomAD (v3.1.2) (Karczewski et al. 2020), and the allele count (AC) in this population was <42 (for MAF 1%) and <8 (for MAF 0.1%). Any coding (protein-altering or synonymous) variant was retained for analysis.

Statistical analyses

Subtype-specific analyses were performed by testing for a burden of rare variants genome-wide using a rare variant extension of the transmission disequilibrium test (RV-TDT) (He et al. 2014) in trios with just CL (N = 76) and trios with just CLP (N = 225) separately. Variants were collapsed by RefSeq-defined gene region, and each gene with 2 or more variants was tested. Any variant that was coding and had a MAF < 1% in gnomAD (v3.1.2) was included in the analysis. We used a Bonferroni threshold correcting for the number of genes tested to determine statistical significance (CL analysis: $p < 3.83 \times 10^{-6}$; CLP analysis: $p < 2.86 \times 10^{-6}$). The $-\log_{10}$ of the p-values from the two analyses were compared with a Pearson's correlation.

A modifier analysis for CL vs CLP was performed using RVTESTS (Zhan et al. 2016). This approach has higher power to identify genetic risk factors that differ between two subtypes, but no power to find factors important in both groups. We directly compared an unrelated subset of CL (N = 100) and CLP (N = 269) cases. Rare variants were collapsed across genes, and the association was tested using the Combined and Multivariate Collapsing method (CMC), which calculates the cumulative burden of rare variants within a region and is designed to find sets of rare variants with the same direction of effect. Any variant that was coding and had a MAF < 1% in gnomAD (v3.1.2) was included in the analysis. A Bonferroni threshold to correct for the number of genes tested was used to determine statistical significance (0.05/19,208 genes; $p < 2.60 \times 10^{-6}$). Additionally, we considered p-value < 1×10^{-5} to be suggestive and performed additional analyses to localize the signal. These analyses included restricting variants to those that were rarer (MAF < 0.1%), only protein-altering, or including coding and non-coding variants in specific regions (enhancer region definitions defined below). Sex and the first 8 principal components (PCs) were used as covariates. PCs were calculated using the GENESIS software package (Gogarten et al. 2019).

Functional annotation of results

Topologically-associated domains (TADs) were defined for significantly associated loci using the H1-ESC cell line in 3D Genome Browser (Wang et al. 2018). Regions with potential craniofacial enhancers were annotated using publicly available datasets derived from ChIP-seq data from human embryos at CS13, CS14, CS15, CS17, and CS20 (4.5-8 weeks post conception) (Attanasio et al. 2013; Wilderman et al. 2018), and enhancers that have been experimentally validated as part of the VISTA Enhancer Browser (Visel et al. 2007). Gene expression for the significant and suggestive genes from the analyses was assessed with RNA-seq data generated from human cranial neural crest cells (GEO: GSM1817212, GSM1817213, GSM1817214, GSM1817215, GSM1817216, and GSM1817217) (Edgar et al. 2002; Prescott et al. 2015), and throughout

craniofacial development (CS13-CS20) (Yankee et al. 2022). For any variants within a predicted enhancer region that was suggestive in the main analysis, transcription factor binding sites (TFBS) were predicted using TFBSTools (Tan and Lenhard 2016) and the JASPAR database (Khan et al. 2018). The 50 base pairs surrounding each variant within the region were extracted from the reference sequence, and the analysis was conducted using both the reference sequence and the alternate sequence. We then selected predicted TFBS that both overlapped with the position of the variant and differed between the reference and alternate sequence. TFBS were considered predicted if the p-value was significant after multiple test correction ($p < 0.05/\text{Number of TFBS that overlapped the variant position}$). The TFBS were considered different between reference and alternate sequence if the TFBS was significant in one sequence and not the other.

Results

We performed a subtype-specific genome-wide analysis to identify genes with a burden of rare variants by performing a RV-TDT analysis in trios with just cleft lip (CL) and trios with just cleft lip and palate (CLP), separately. This type of analysis could detect variants associated with increased risk for an OFC in general, but could also identify variants that increase the risk for just one of the subtypes of OFCs. Overall, no gene was significant in either analysis (Figure S1-S2, Table S2-S3), likely due to the small sample size. However, some of the top results were in genes known to be associated with OFCs in previous analyses of common variants (e.g., *ARHGAP29*: $p = 0.002$ in the CL analysis)(Beaty et al. 2010; Leslie et al. 2016; Yu et al. 2017). The top results for CL and CLP were different and the p-values were not correlated ($r = 0.005$, $p = 0.56$; Figure S3). Only one gene had a p-value < 0.01 in both the CL and CLP analysis (*CYP2D7*: $p = 0.007$ in CL, $p = 0.005$ in CLP), with 52 genes being $p < 0.01$ in CL alone and 108 genes being $p < 0.01$ in CLP alone. This result is consistent with the hypothesis that CL and CLP may have partially distinct genetic architectures.

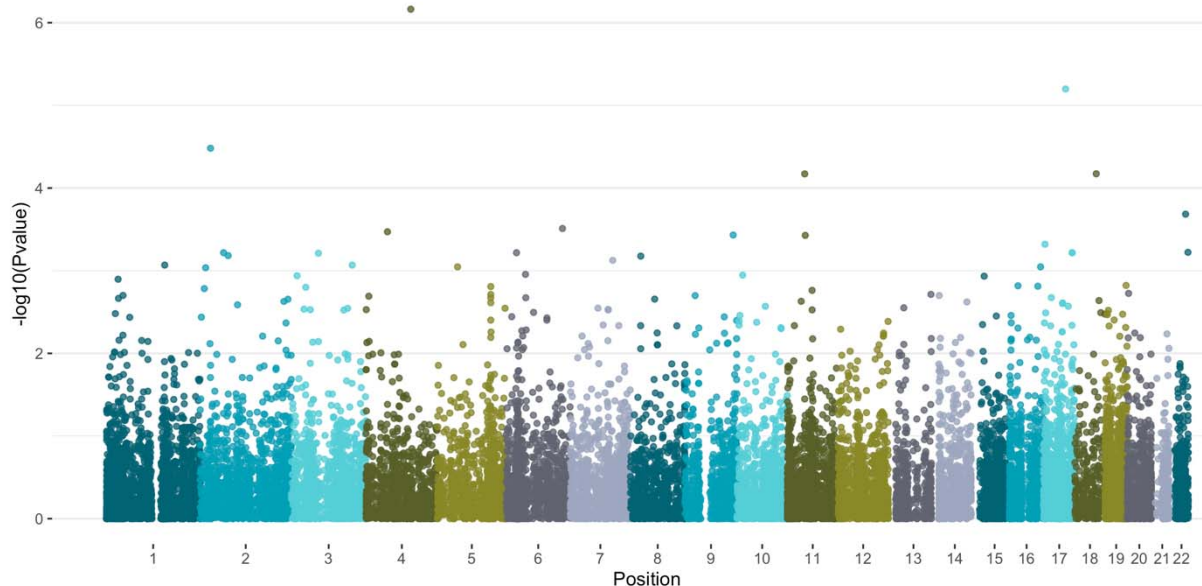


Figure 1: Manhattan plot of $-\log_{10}(p\text{-values})$ from the cleft lip vs. cleft lip and palate modifier analysis.

Because of the differences between the burden results from CL and CLP, we next performed a modifier analysis by comparing CL cases directly to CLP cases. Because this is a case-to-case group comparison, this analysis would not be able to detect variants generally important for OFC risk, but would detect variants important for the formation of one subtype over the other subtype. In the modifier analysis, *SEC24D* was genome-wide significant ($p = 6.86 \times 10^{-7}$), and *SUPT4H1* was suggestive ($p = 6.33 \times 10^{-6}$, Figure 1, Figure S4, Table 1, Table S4), with both having a burden of rare variants in CL compared to CLP (Table S5-S6). Both genes were still significant if only rarer variants (MAF < 0.1%) were included in the analysis (*SEC24D*: $p = 7.58 \times 10^{-5}$; *SUPT4H1*: $p = 0.001$) and if only protein-altering variants were included (*SEC24D*: $p = 0.003$; *SUPT4H1*: $p = 0.0008$), indicating the result is robust to different filtering and variant classification strategies.

We next further investigated these genes to better understand their role in craniofacial development by analyzing their expression during relevant timepoints, where the rare variants fell within the genes (Figure 2A, Figure S5), and the predicted topological domain (TAD) for both regions (Figure 2B, Figure S6). *SEC24D* and *SUPT4H1* were highly expressed in an *in vitro* human cranial neural crest cell model and throughout human fetal development, being in the top 10th

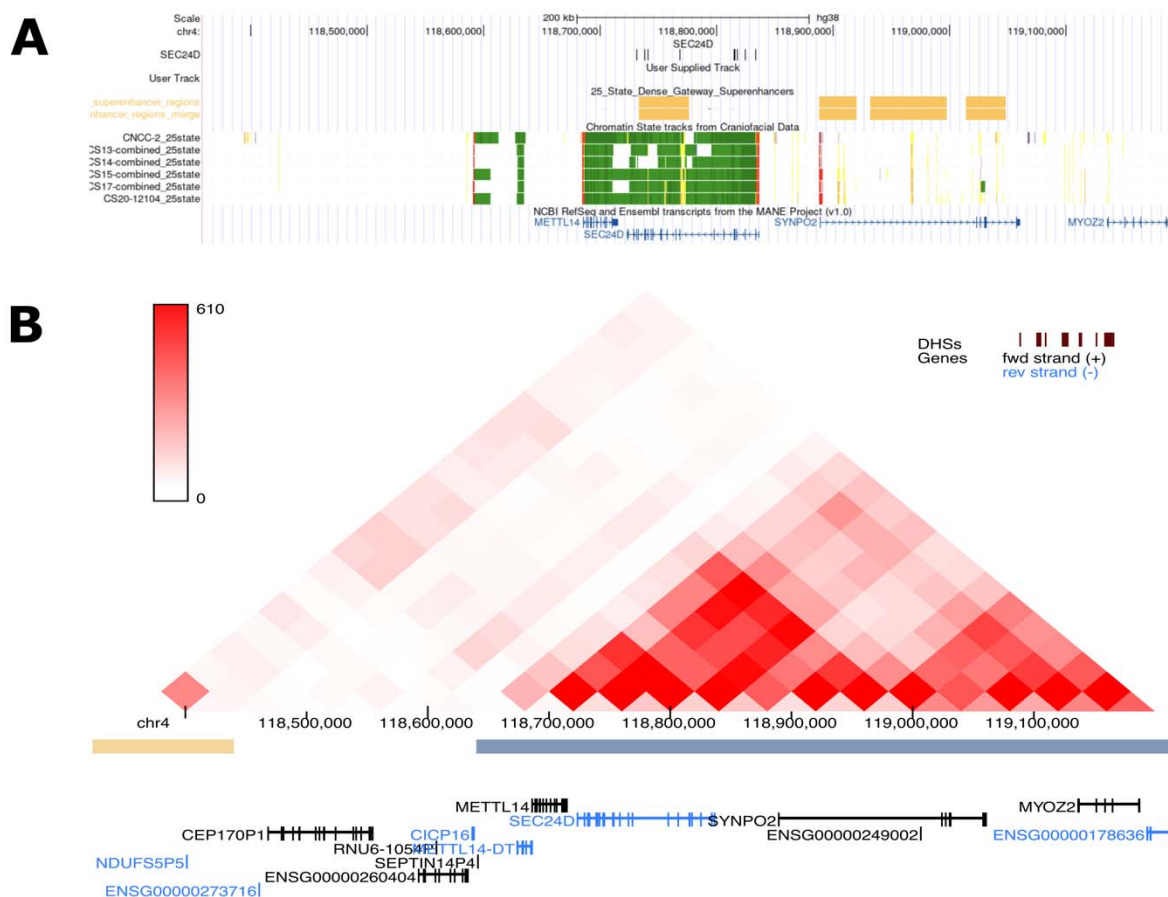


Figure 2: Rare variants within *SEC24D* fall within a predicted craniofacial super enhancer (A). The predicted topologically-associated domain (TAD) surrounding *SEC24D* (B).

Gene	Chromosome	Start (hg38)	End (hg38)	Number of rare variants	P-value
<i>SEC24D</i>	4	118722823	118838683	16	6.86×10^{-7}
<i>SUPT4H1</i>	17	58345175	58353093	2	6.33×10^{-6}
<i>C2orf16</i>	2	27537386	27582721	49	3.30×10^{-5}
<i>TXNL1</i>	18	56597209	56651600	5	6.72×10^{-5}
<i>C11orf94</i>	11	45906513	45907272	9	6.74×10^{-5}

percentile of genes expressed in cranial neural crest cell lines and in the 20th percentile of all genes expressed throughout development (Table S7), further indicating that these are both genes with a potential role for the formation of an OFC. Rare variants were distributed throughout *SEC24D*, including overlapping with a predicted craniofacial enhancer (Figure 2). Rare variants in *SUPT4H1*, however, were just in the beginning of the gene, and within a craniofacial enhancer that covered the entire gene (Supplemental Figure).

Because some of the variants overlapped predicted craniofacial enhancers, we next investigated whether the burden test was actually detecting a non-coding signal. We tested the enhancers within the *SEC24D* region to see if there was a burden of either coding (Table S9) or any (Table S10) rare variants. A super enhancer had a burden of rare variants when considering both coding variants ($p=0.02$) and all variants ($p=0.006$), indicating that there may be a statistically significant role for non-coding variants, but to a lesser degree than coding variants in this region. We next tested whether any of the variants affected predicted transcription factor binding sites (TFBSs). For *SEC24D*, 9 binding sites for 9 transcription factors (TFs) were destroyed, including ones for PAX1, and 23 TFBS were created, including ones for Pax6 and PAX9 (Table 2, Table S8), suggesting a potential mechanism for a non-coding effect. Within the predicted TAD that contains *SEC24D*, *SEC24D* was the highest expressed gene, and thus the best candidate target gene for these enhancers, with the next highest expressed gene being *SYNPO2* (Table S7).

For the enhancers that overlapped *SUPT4H1*, the super enhancer was significant with just the coding variants ($p=6.33 \times 10^{-6}$), but not when all variants were considered ($p = 0.14$). Two variants were predicted to create binding sites for 5 TFs and destroy the 9 TFBSs, including FOXP2, which is involved in craniofacial morphology and speech (Table S11) (Xu et al. 2018). Of the other genes in the predicted TAD containing this super enhancer, *MTMR4* and *TSPOAP1* were also highly expressed throughout development (Table S7). Taken together, while it is possible that these rare variants could have non-coding effects, the association with the individual enhancers and expression of the nearby genes are not strongly associated with cleft subtypes.

Table 2: Transcription factor binding sites altered by SEC24D variants within enhancer regions				
Predicted Transcription Factor Binding Sites Destroyed				
Variant	Transcription Factor	Class	P-value (Reference)	P-value (Alternate)
chr4:118738340:C:T	RELA	Rel homology region (RHR) factors	0.00035	NA
chr4:118738340:C:T	RELB	Rel homology region (RHR) factors	6.18×10 ⁻⁵	NA
rs146307367	EHF	Tryptophan cluster factors	0.00026	NA
rs146307367	ELF4	Tryptophan cluster factors	0.00043	NA
rs146307367	PAX1	Paired box factors	0.00098	NA
rs146307367	RARA::RXRG	Nuclear receptors with C4 zinc fingers	0.00068	NA
rs146307367	SPI1	Tryptophan cluster factors	0.00011	NA
rs146307367	SPIC	Tryptophan cluster factors	0.00091	NA
rs201472807	SRF	MADS box factors	0.00017	NA
Predicted Transcription Factor Binding Sites Created				
Variant	Transcription Factor	Class	P-value (Reference)	P-value (Alternate)
chr4:118738340:C:T	HSF1	Heat shock factors	NA	0.00026
rs140046192	FOXB1	Fork head / winged helix factors	NA	0.00054
rs140046192	Pax6	Paired box factors	NA	6.12×10 ⁻⁵
rs140046192	PAX9	Paired box factors	NA	0.00012
rs140046192	TP73	p53 domain factors	NA	0.00046
rs139422490	ATF7	Basic leucine zipper factors (bZIP)	NA	0.00023
rs139422490	CREB1	Basic leucine zipper factors (bZIP)	NA	5.22×10 ⁻⁵
rs139422490	FOS::JUN(var.2)	Basic leucine zipper factors (bZIP)	NA	6.15×10 ⁻⁵
rs139422490	FOSB::JUN	Basic leucine zipper factors (bZIP)	NA	0.00022
rs139422490	FOSB::JUNB(var.2)	Basic leucine zipper factors (bZIP)	NA	0.00029
rs139422490	FOSL1::JUN(var.2)	Basic leucine zipper factors (bZIP)	NA	0.00032
rs139422490	FOSL2::JUN(var.2)	Basic leucine zipper factors (bZIP)	NA	6.27×10 ⁻⁵

rs139422490	FOSL2::JUNB(var.2)	Basic leucine zipper factors (bZIP)	NA	0.00021
rs139422490	FOSL2::JUND(var.2)	Basic leucine zipper factors (bZIP)	NA	0.00017
rs139422490	JDP2(var.2)	Basic leucine zipper factors (bZIP)	NA	0.00028
rs139422490	JUN::JUNB(var.2)	Basic leucine zipper factors (bZIP)	NA	5.51×10^{-5}
rs139422490	JUNB(var.2)	Basic leucine zipper factors (bZIP)	NA	0.00022
rs139422490	NR2F2	Nuclear receptors with C4 zinc fingers	NA	0.00019
rs139422490	PBX1	Homeo domain factors	NA	0.00016
rs139422490	POU5F1	Homeo domain factors	NA	0.00034
rs139422490	RORA	Nuclear receptors with C4 zinc fingers	NA	0.00029
rs139422490	RXRG	Nuclear receptors with C4 zinc fingers	NA	0.00046
rs146307367	NFATC2	Rel homology region (RHR) factors	NA	0.00097

Discussion

There have been many studies that have investigated the genetic architecture of OFCs but our understanding of the factors that contribute to formation of CL versus CLP is incomplete, especially for rare variants. We previously addressed this by performing modifier analyses using targeted sequencing data and in low frequency variants from genotyping array (Carlson et al. 2017a; Carlson et al. 2017b), but this is the first study to investigate rare genetic modifiers of CL and CLP genome-wide using whole genome sequence data. We found that there was a statistically significant burden of rare, coding variants in *SEC24D* in participants with CL compared to CLP, and that this association still held if we lowered the minor allele threshold or only included protein-altering variants. We also found from publicly available RNA-seq data that *SEC24D* is highly expressed in neural crest cells lines and during craniofacial development. *SEC24D* is involved in vesicle trafficking, as a member of Coat Protein Complex II (COPII) and is important for cargo selection, concentration, and shape of the vesicle (D'Arcangelo et al. 2013). In zebrafish, *sec24d* is required for the effective transportation of proteins to the extracellular matrix in cartilage and with embryos deficient for *sec24d* have a failure of chondroblast differentiation leading to severe craniofacial dysmorphology and malformed cartilage (Sarmah et al. 2010). Consistent with that, *Sec24*-null mice have early embryonic lethality (Baines et al. 2013). *SEC24D* has also been implicated in bone morphology disorders in humans before, with homozygous or compound heterozygous truncating and missense mutations causing Cole-

Carpenter Syndrome (MIM# 616294), a syndromic form of osteogenesis imperfecta that presents with disturbed ossification of the skull, leading to midface hypoplasia, macrocephaly, and micrognathia, as well as brittle bones that break easily (Garbes et al. 2015; Moosa et al. 2016; Takeyari et al. 2018; Zhang et al. 2017). Collectively, this is strong evidence that *SEC24D* is involved in bone and cartilage morphology, especially during craniofacial development. Therefore, heterozygous coding variants that subtly alter this gene's function or non-coding variants that alter its expression could modify craniofacial phenotypes, such as whether a cleft of just the lip forms compared to a cleft that also affect the bony palate. However, more work is warranted to investigate how this gene could specifically modify orofacial cleft phenotypes.

Even though it did not meet the significance threshold for this study, we also had suggestive evidence for *SUPT4H1* as a modifier of CL vs CLP. *SUPT4H1* is involved in mRNA processing and transcription elongation, and depletion of *SUPT4H1* in cell lines leads to global reduction in RNA levels; mice lacking *Supt4h1* do not survive to birth (Cheng et al. 2015; Naguib et al. 2019). Although not previously associated with orofacial phenotypes, it is highly expressed during craniofacial development and overlaps regulatory regions for craniofacial tissue. Therefore, future studies either trying to replicate this result or investigating the role of *SUPT4H1* in facial development or bone morphology are of interest.

We also conducted a subtype specific RV-TDT, comparing the transmission of rare variants in CL and CLP separately. Nothing in that analysis approached genome-wide significance, likely due to the small sample size and that the subtype-specific analysis would have less statistical power to detect loci that differ between two subtypes than the modifier analysis. However, some of the top results in the subtype specific analyses, like *ARHGAP29*, had been previously associated with OFCs (Leslie et al. 2016), and as cohorts increase in size, we expect to identify more robust signals that will facilitate more precise subtype comparisons.

We previously found a burden of low frequency genotyped SNPs in *C8orf34*, *TMEM246*, and *CDC42EP3* (Carlson et al. 2017a), and burden of rare variants around the loci containing *PAX7*, *ARHGAP29*, 8q24, *FOXE1*, *VAX1*, *NTN1*, and *NOG* (Carlson et al. 2017b). None of these genes/regions were significant or suggestive in this analysis, but this could be due to several factors. For example, previous studies used much larger sample sizes. A larger sample size would be better powered to detect rare variants, and therefore there may be genetic modifiers that were previously identified that this study is not powered to detect. Additionally, this lack of replication also be due to heterogeneous effects of rare variants that are undetectable when using the CMC method. There are also ancestry differences between the studies, with this study only having samples with European ancestry and previous studies being predominantly non-European; any modifier that differs by ancestry may not be detected in this study. Further studies of clefting modifiers and replication in larger, more diverse cohorts is warranted.

Although we initially focused on rare, coding variants we also found evidence that non-coding variants are important in modifying OFC subtypes. The gene bodies of *SEC24D* and *SUPT4H1* contained predicted craniofacial enhancers. The rare variants within these enhancers are predicted to alter TFBSs for many different TFs. Some, like *Pax6* and *PAX9*, have been associated with orofacial clefts before (Peters et al. 1998; Rodrigo et al. 2003; Sull et al. 2009; Vaivads et al. 2021), and others, like *PAX1*, have been associated with both normal facial morphology and as laterality modifiers of orofacial clefts (Curtis et al. 2021a; Shaffer et al. 2016). This is consistent with the hypothesis that genetic variation in regulatory regions may

have an effect by altering TFBSs and thus gene expression. However, for both genes in this study, the association with the coding variation within the gene was stronger than the association with all rare variants within the overlapping enhancer. Of the genes within their respective TADs, *SEC24D* and *SUPT4H1* were among the highest expressed genes during craniofacial development, making them likely target genes of these potential regulatory regions. Thus, while there is evidence consistent with rare, regulatory variants impacting orofacial cleft phenotypes, this effect is likely in addition to the effect by coding variation, and the effect of rare, non-coding variation genome-wide needs further investigation.

In summary, we conducted the first genome-wide scan of rare genetic modifiers in a case-case design for CL vs CLP and found that *SEC24D* and *SUPTAH1* had a significant burden of rare, coding variants in CL cases compared to CLP cases. By using publicly available datasets, we were able to determine that both genes were highly expressed in developing craniofacial tissue and contained predicted craniofacial enhancers. Additionally, we found that variants in these enhancers created or destroyed predicted binding sites for transcription factors, including *Pax9* and *Pax6*, suggesting a mechanism through which coding and noncoding variants could lead to phenotypic heterogeneity. Overall, this study expands our understanding of the genetic architecture of CL vs CLP and suggests that regulatory variation may also contribute and warrants further investigation in future studies.

Acknowledgments:

The authors thank the dedicated field staff, collaborators, and participating families for their important contributions to this study. These studies are part of the Gabriella Miller Kids First Pediatric Research Program, supported by the Common Fund of the Office of the Director of the National Institutes of Health (NIH). Sequencing of the European trios was completed at Washington University's McDonnell Genome Institute (3U54HG003079-12S1 and X01-HL132363 and the Colombian and Taiwanese trios were sequenced at the Broad Institute Sequencing Center (U24-HD090743). The sequencing centers plus the Kids First Data Resource Center, supported by the NIH Common Fund through grant U2CHL138346, provided technical and analytical support of this project. This work was supported by grants from the National Institutes of Health (NIH) including: F32-DE032260 [SWC], R00-DE025060 [EJL], R01-DE028342 [EJL], R03-DE030118 [EJL], R21- DE029698 [MPE], X01-HG007485 [MLM, EF], R01-DE016148 [MLM, SMW], U01-DE024425 [MLM], R37- DE008559 [JCM, MLM], R01-DE009886 [MLM], R21-DE016930 [MLM], R01- DE012472 [MLM], R01-DE014581 [THB], U01-DE018993 [THB], U01-DE020078; R01-DE027023 [SMW].

Declarations

The authors have no disclosures.

Conflicts of Interest

The authors have no conflicts to disclose.

Data Availability

Phenotypes and associated genome sequencing data used in this manuscript are available from the database of Genotypes and Phenotypes (dbGaP; European trios, dbGaP: phs001168.v2.p2; Colombian trios, dbGaP: phs001420.v1.p1; Taiwanese trios, dbGaP: phs000094.v1.p1) and from the Kids First Data Resource Center.

References

- Amidei RL, Hamman RF, Kassebaum DK, Marshall JA (1994) Birth prevalence of cleft lip and palate in Colorado by sex distribution, seasonality, race/ethnicity, and geographic variation. *Spec Care Dentist* 14: 233-40. doi: 10.1111/j.1754-4505.1994.tb01072.x
- Attanasio C, Nord AS, Zhu Y, Blow MJ, Li Z, Liberton DK, Morrison H, Plajzer-Frick I, Holt A, Hosseini R, Phouanavong S, Akiyama JA, Shoukry M, Afzal V, Rubin EM, FitzPatrick DR, Ren B, Hallgrimsson B, Pennacchio LA, Visel A (2013) Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* 342: 1241006. doi: 10.1126/science.1241006
- Baines AC, Adams EJ, Zhang B, Ginsburg D (2013) Disruption of the *Sec24d* gene results in early embryonic lethality in the mouse. *PLoS One* 8: e61114. doi: 10.1371/journal.pone.0061114
- Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Liang KY, Wu T, Murray T, Fallin MD, Redett RA, Raymond G, Schwender H, Jin SC, Cooper ME, Dunnwald M, Mansilla MA, Leslie E, Bullard S, Lidral AC, Moreno LM, Menezes R, Vieira AR, Petrin A, Wilcox AJ, Lie RT, Jabs EW, Wu-Chou YH, Chen PK, Wang H, Ye X, Huang S, Yeow V, Chong SS, Jee SH, Shi B, Christensen K, Melbye M, Doheny KF, Pugh EW, Ling H, Castilla EE, Czeizel AE, Ma L, Field LL, Brody L, Pangilinan F, Mills JL, Molloy AM, Kirke PN, Scott JM, Arcos-Burgos M, Scott AF (2010) A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near *MAFB* and *ABCA4*. *Nat Genet* 42: 525-9. doi: 10.1038/ng.580
- Bell JC, Raynes-Greenow C, Turner RM, Bower C, Nassar N, O'Leary CM (2014) Maternal alcohol consumption during pregnancy and the risk of orofacial clefts in infants: a systematic review and meta-analysis. *Paediatr Perinat Epidemiol* 28: 322-32. doi: 10.1111/ppe.12131
- Birnbaum S, Ludwig KU, Reutter H, Herms S, de Assis NA, Diaz-Lacava A, Barth S, Lauster C, Schmidt G, Scheer M, Saffar M, Martini M, Reich RH, Schiefke F, Hemprich A, Potzsch S, Potzsch B, Wienker TF, Hoffmann P, Knapp M, Kramer FJ, Nothen MM, Mangold E (2009a) *IRF6* gene variants in Central European patients with non-syndromic cleft lip with or without cleft palate. *Eur J Oral Sci* 117: 766-9. doi: 10.1111/j.1600-0722.2009.00680.x
- Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M, Baluardo C, Ferriani 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, Hemprich A, Potzsch S, Steegers-Theunissen RP, Potzsch B, Moebus S, Horsthemke B, Kramer FJ, Wienker TF, Mossey PA, Propping P, Cichon S, Hoffmann P, Knapp M, Nothen MM, Mangold E (2009b) Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat Genet* 41: 473-7. doi: 10.1038/ng.333
- Bishop MR, Diaz Perez KK, Sun M, Ho S, Chopra P, Mukhopadhyay N, Hetmanski JB, Taub MA, Moreno-Urbe LM, Valencia-Ramirez LC, Restrepo Muneton CP, Wehby G, Hecht JT, Deleyiannis F, Weinberg SM, Wu-Chou YH, Chen PK, Brand H, Epstein MP, Ruczinski I, Murray JC, Beaty TH, Feingold E, Lipinski RJ, Cutler DJ, Marazita ML, Leslie EJ (2020) Genome-wide Enrichment of De Novo Coding Mutations in Orofacial Cleft Trios. *Am J Hum Genet* 107: 124-136. doi: 10.1016/j.ajhg.2020.05.018

- Candotto V, Oberti L, Gabrione F, Greco G, Rossi D, Romano M, Mummolo S (2019) Current concepts on cleft lip and palate etiology. *J Biol Regul Homeost Agents* 33: 145-151 DENTAL SUPPLEMENT.
- Carlson JC, Anand D, Butali A, Buxo CJ, Christensen K, Deleyiannis F, Hecht JT, Moreno LM, Orioli IM, Padilla C, Shaffer JR, Vieira AR, Wehby GL, Weinberg SM, Murray JC, Beaty TH, Saadi I, Lachke SA, Marazita ML, Feingold E, Leslie EJ (2019) A systematic genetic analysis and visualization of phenotypic heterogeneity among orofacial cleft GWAS signals. *Genet Epidemiol* 43: 704-716. doi: 10.1002/gepi.22214
- Carlson JC, Standley J, Petrin A, Shaffer JR, Butali A, Buxo 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, Wehby GL, Weinberg SM, Beaty TH, Feingold E, Murray JC, Marazita ML, Leslie EJ (2017a) Identification of 16q21 as a modifier of nonsyndromic orofacial cleft phenotypes. *Genet Epidemiol* 41: 887-897. doi: 10.1002/gepi.22090
- Carlson JC, Taub MA, Feingold E, Beaty TH, Murray JC, Marazita ML, Leslie EJ (2017b) Identifying Genetic Sources of Phenotypic Heterogeneity in Orofacial Clefts by Targeted Sequencing. *Birth Defects Res* 109: 1030-1038. doi: 10.1002/bdr2.23605
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4: 7. doi: 10.1186/s13742-015-0047-8
- Cheng HM, Chern Y, Chen IH, Liu CR, Li SH, Chun SJ, Rigo F, Bennett CF, Deng N, Feng Y, Lin CS, Yan YT, Cohen SN, Cheng TH (2015) Effects on murine behavior and lifespan of selectively decreasing expression of mutant huntingtin allele by *supt4h* knockdown. *PLoS Genet* 11: e1005043. doi: 10.1371/journal.pgen.1005043
- Ciminello FS, Morin RJ, Nguyen TJ, Wolfe SA (2009) Cleft lip and palate: review. *Compr Ther* 35: 37-43.
- Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, Zilversmit M, Cartwright R, Rouleau GA, Daly M, Stone EA, Hurles ME, Awadalla P, Genomes P (2011) Variation in genome-wide mutation rates within and between human families. *Nat Genet* 43: 712-4. doi: 10.1038/ng.862
- Curtis SW, Chang D, Lee MK, Shaffer JR, Indencleef K, Epstein MP, Cutler DJ, Murray JC, Feingold E, Beaty TH, Claes P, Weinberg SM, Marazita ML, Carlson JC, Leslie EJ (2021a) The PAX1 locus at 20p11 is a potential genetic modifier for bilateral cleft lip. *HGG Adv* 2. doi: 10.1016/j.xhgg.2021.100025
- Curtis SW, Chang D, Sun MR, Epstein MP, Murray JC, Feingold E, Beaty TH, Weinberg SM, Marazita ML, Lipinski RJ, Carlson JC, Leslie EJ (2021b) FAT4 identified as a potential modifier of orofacial cleft laterality. *Genet Epidemiol* 45: 721-735. doi: 10.1002/gepi.22420
- D'Arcangelo JG, Stahmer KR, Miller EA (2013) Vesicle-mediated export from the ER: COPII coat function and regulation. *Biochim Biophys Acta* 1833: 2464-72. doi: 10.1016/j.bbamcr.2013.02.003
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G (2011) The

- variant call format and VCFtools. *Bioinformatics* 27: 2156-8. doi: 10.1093/bioinformatics/btr330
- DeRoo LA, Wilcox AJ, Lie RT, Romitti PA, Pedersen DA, Munger RG, Moreno Uribe LM, Wehby GL (2016) Maternal alcohol binge-drinking in the first trimester and the risk of orofacial clefts in offspring: a large population-based pooling study. *Eur J Epidemiol* 31: 1021-1034. doi: 10.1007/s10654-016-0171-5
- Dixon MJ, Marazita ML, Beaty TH, Murray JC (2011) Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet* 12: 167-78. doi: 10.1038/nrg2933
- Duan S, Shi J, Shi B, Jia Z (2020) Association analysis of GWAS hits and non-syndromic cleft lip with/without palate with cleft alveolar in Han population of western China. *Int J Clin Exp Pathol* 13: 2576-2585.
- Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30: 207-10. doi: 10.1093/nar/30.1.207
- Garbes L, Kim K, Riess A, Hoyer-Kuhn H, Beleggia F, Bevot A, Kim MJ, Huh YH, Kweon HS, Savarirayan R, Amor D, Kakadia PM, Lindig T, Kagan KO, Becker J, Boyadjiev SA, Wollnik B, Semler O, Bohlander SK, Kim J, Netzer C (2015) Mutations in SEC24D, encoding a component of the COPII machinery, cause a syndromic form of osteogenesis imperfecta. *Am J Hum Genet* 96: 432-9. doi: 10.1016/j.ajhg.2015.01.002
- Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, Rice KM, Conomos MP (2019) Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics* 35: 5346-5348. doi: 10.1093/bioinformatics/btz567
- Gundlach KK, Maus C (2006) Epidemiological studies on the frequency of clefts in Europe and world-wide. *J Craniomaxillofac Surg* 34 Suppl 2: 1-2. doi: 10.1016/S1010-5182(06)60001-2
- Hameed O, Amin N, Haria P, Patel B, Hay N (2019) Orthodontic burden of care for patients with a cleft lip and/or palate. *J Orthod* 46: 63-67. doi: 10.1177/1465312518823010
- Hao J, Gao R, Wu W, Hua L, Chen Y, Li F, Liu J, Luo D, Han J, Wang H (2018) Association between BMP4 gene polymorphisms and cleft lip with or without cleft palate in a population from South China. *Arch Oral Biol* 93: 95-99. doi: 10.1016/j.archoralbio.2018.05.015
- He Z, O'Roak BJ, Smith JD, Wang G, Hooker S, Santos-Cortez RL, Li B, Kan M, Krumm N, Nickerson DA, Shendure J, Eichler EE, Leal SM (2014) Rare-variant extensions of the transmission disequilibrium test: application to autism exome sequence data. *Am J Hum Genet* 94: 33-46. doi: 10.1016/j.ajhg.2013.11.021
- Huang L, Jia Z, Shi Y, Du Q, Shi J, Wang Z, Mou Y, Wang Q, Zhang B, Wang Q, Ma S, Lin H, Duan S, Yin B, Lin Y, Wang Y, Jiang D, Hao F, Zhang L, Wang H, Jiang S, Xu H, Yang C, Li C, Li J, Shi B, Yang Z (2019) Genetic factors define CPO and CLO subtypes of nonsyndromic orofacial cleft. *PLoS Genet* 15: e1008357. doi: 10.1371/journal.pgen.1008357
- Jamilian A, Sarkarat F, Jafari M, Neshandar M, Amini E, Khosravi S, Ghassemi A (2017) Family history and risk factors for cleft lip and palate patients and their associated anomalies. *Stomatologija* 19: 78-83.
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi 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, Walters RK, Tashman K, Farjoun Y,

- Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Genome Aggregation Database C, Neale BM, Daly MJ, MacArthur DG (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581: 434-443. doi: 10.1038/s41586-020-2308-7
- Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Cheneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, Mathelier A (2018) JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* 46: D260-D266. doi: 10.1093/nar/gkx1126
- Lancaster HS, Lien KM, Chow JC, Frey JR, Scherer NJ, Kaiser AP (2020) Early Speech and Language Development in Children With Nonsyndromic Cleft Lip and/or Palate: A Meta-Analysis. *J Speech Lang Hear Res* 63: 14-31. doi: 10.1044/2019_JSLHR-19-00162
- 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 (2017a) Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Hum Genet* 136: 275-286. doi: 10.1007/s00439-016-1754-7
- Leslie EJ, Carlson JC, Shaffer JR, Buxo CJ, Castilla EE, Christensen K, Deleyiannis FWB, Field LL, Hecht JT, Moreno L, Orioli IM, Padilla C, Vieira AR, Wehby GL, Feingold E, Weinberg SM, Murray JC, Marazita ML (2017b) Association studies of low-frequency coding variants in nonsyndromic cleft lip with or without cleft palate. *Am J Med Genet A* 173: 1531-1538. doi: 10.1002/ajmg.a.38210
- Leslie EJ, 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, Lidral AC, Valencia-Ramirez LC, Lopez-Palacio AM, Valencia DR, Arcos-Burgos M, Czeizel AE, Field LL, Padilla CD, Cutiongco-de la Paz EM, Deleyiannis F, Christensen K, Munger RG, Lie RT, Wilcox A, Romitti PA, Castilla EE, Mereb JC, Poletta FA, Orioli IM, Carvalho FM, Hecht JT, Blanton SH, Buxo CJ, Butali A, Mossey PA, Adeyemo WL, James O, Braimah RO, Aregbesola BS, Eshete MA, Abate F, Koruyucu M, Seymen F, Ma L, de Salamanca JE, Weinberg SM, Moreno L, Murray JC, Marazita ML (2016) A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. *Hum Mol Genet* 25: 2862-2872. doi: 10.1093/hmg/ddw104
- Leslie EJ, Marazita ML (2013) Genetics of cleft lip and cleft palate. *Am J Med Genet C Semin Med Genet* 163C: 246-58. doi: 10.1002/ajmg.c.31381
- Li D, Zhu G, Lou S, Ma L, Zhang C, Pan Y, Wang L (2020) The functional variant of NTN1 contributes to the risk of nonsyndromic cleft lip with or without cleft palate. *Eur J Hum Genet* 28: 453-460. doi: 10.1038/s41431-019-0549-4

- Li YH, Yang J, Zhang JL, Liu JQ, Zheng Z, Hu DH (2017) BMP4 rs17563 polymorphism and nonsyndromic cleft lip with or without cleft palate: A meta-analysis. *Medicine (Baltimore)* 96: e7676. doi: 10.1097/MD.0000000000007676
- Li Z, Liu J, Ye R, Zhang L, Zheng X, Ren A (2010) Maternal passive smoking and risk of cleft lip with or without cleft palate. *Epidemiology* 21: 240-2. doi: 10.1097/EDE.0b013e3181c9f941
- Ludwig KU, Ahmed ST, Bohmer AC, Sangani NB, Varghese S, Klamt J, Schuenke H, Gultepe P, Hofmann A, Rubini M, Aldhorae KA, Steegers-Theunissen RP, Rojas-Martinez A, Reiter R, Borck G, Knapp M, Nakatomi M, Graf D, Mangold E, Peters H (2016) Meta-analysis Reveals Genome-Wide Significance at 15q13 for Nonsyndromic Clefting of Both the Lip and the Palate, and Functional Analyses Implicate GREM1 As a Plausible Causative Gene. *PLoS Genet* 12: e1005914. doi: 10.1371/journal.pgen.1005914
- Ludwig KU, Bohmer AC, Bowes J, Nikolic M, Ishorst N, Wyatt N, Hammond NL, Golz L, Thieme F, Barth S, Schuenke H, Klamt J, Spielmann M, Aldhorae K, Rojas-Martinez A, Nothen MM, Rada-Iglesias A, Dixon MJ, Knapp M, Mangold E (2017) Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip +/- cleft palate and cleft palate only. *Hum Mol Genet* 26: 829-842. doi: 10.1093/hmg/ddx012
- Mangold E, Ludwig KU, Birnbaum S, Baluado 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, Reich RH, Hemprich A, Potzsch S, Blaumeiser B, Moebus S, Krawczak M, Schreiber S, Meitinger T, Wichmann HE, Steegers-Theunissen RP, Kramer FJ, Cichon S, Propping P, Wienker TF, Knapp M, Rubini M, Mossey PA, Hoffmann P, Nothen MM (2010) Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat Genet* 42: 24-6. doi: 10.1038/ng.506
- Marazita ML (2012) The evolution of human genetic studies of cleft lip and cleft palate. *Annu Rev Genomics Hum Genet* 13: 263-83. doi: 10.1146/annurev-genom-090711-163729
- Martelli DR, Coletta RD, Oliveira EA, Swerts MS, Rodrigues LA, Oliveira MC, Martelli Junior H (2015) Association between maternal smoking, gender, and cleft lip and palate. *Braz J Otorhinolaryngol* 81: 514-9. doi: 10.1016/j.bjorl.2015.07.011
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky 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: 1297-303. doi: 10.1101/gr.107524.110
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F (2016) The Ensembl Variant Effect Predictor. *Genome Biol* 17: 122. doi: 10.1186/s13059-016-0974-4
- Moosa S, Chung BH, Tung JY, Altmuller J, Thiele H, Nurnberg P, Netzer C, Nishimura G, Wollnik B (2016) Mutations in SEC24D cause autosomal recessive osteogenesis imperfecta. *Clin Genet* 89: 517-519. doi: 10.1111/cge.12678
- Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC (2009) Cleft lip and palate. *Lancet* 374: 1773-85. doi: 10.1016/S0140-6736(09)60695-4

- Mossey PA, Modell B (2012) Epidemiology of oral clefts 2012: an international perspective. *Front Oral Biol* 16: 1-18. doi: 10.1159/000337464
- Mukhopadhyay N, Bishop M, Mortillo M, Chopra P, Hetmanski JB, Taub MA, Moreno LM, Valencia-Ramirez LC, Restrepo C, Wehby GL, Hecht JT, Deleyiannis F, Butali A, Weinberg SM, Beaty TH, Murray JC, Leslie EJ, Feingold E, Marazita ML (2020) Whole genome sequencing of orofacial cleft trios from the Gabriella Miller Kids First Pediatric Research Consortium identifies a new locus on chromosome 21. *Hum Genet* 139: 215-226. doi: 10.1007/s00439-019-02099-1
- Naguib A, Sandmann T, Yi F, Watts RJ, Lewcock JW, Dowdle WE (2019) SUPT4H1 Depletion Leads to a Global Reduction in RNA. *Cell Rep* 26: 45-53 e4. doi: 10.1016/j.celrep.2018.12.004
- Peters H, Neubuser A, Kratochwil K, Balling R (1998) Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes Dev* 12: 2735-47. doi: 10.1101/gad.12.17.2735
- Prescott SL, Srinivasan R, Marchetto MC, Grishina I, Narvaiza I, Selleri L, Gage FH, Swigut T, Wysocka J (2015) Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* 163: 68-83. doi: 10.1016/j.cell.2015.08.036
- Rodrigo I, Hill RE, Balling R, Munsterberg A, Imai K (2003) Pax1 and Pax9 activate Bapx1 to induce chondrogenic differentiation in the sclerotome. *Development* 130: 473-82. doi: 10.1242/dev.00240
- Sabbagh HJ, Mossey PA, Innes NP (2012) Prevalence of orofacial clefts in Saudi Arabia and neighboring countries: A systematic review. *Saudi Dent J* 24: 3-10. doi: 10.1016/j.sdentj.2011.11.001
- Sarmah S, Barrallo-Gimeno A, Melville DB, Topczewski J, Solnica-Krezel L, Knapik EW (2010) Sec24D-dependent transport of extracellular matrix proteins is required for zebrafish skeletal morphogenesis. *PLoS One* 5: e10367. doi: 10.1371/journal.pone.0010367
- Shaffer JR, LeClair J, Carlson JC, Feingold E, Buxo CJ, Christensen K, Deleyiannis FWB, Field LL, Hecht JT, Moreno L, Orioli IM, Padilla C, Vieira AR, Wehby GL, Murray JC, Weinberg SM, Marazita ML, Leslie EJ (2019) Association of low-frequency genetic variants in regulatory regions with nonsyndromic orofacial clefts. *Am J Med Genet A* 179: 467-474. doi: 10.1002/ajmg.a.61002
- Shaffer JR, Orlova E, Lee MK, Leslie EJ, Raffensperger ZD, Heike CL, Cunningham ML, Hecht JT, Kau CH, Nidey NL, Moreno LM, Wehby GL, Murray JC, Laurie CA, Laurie CC, Cole J, Ferrara T, Santorico S, Klein O, Mio W, Feingold E, Hallgrímsson B, Spritz RA, Marazita ML, Weinberg SM (2016) Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. *PLoS Genet* 12: e1006149. doi: 10.1371/journal.pgen.1006149
- Sull JW, Liang KY, Hetmanski JB, Fallin MD, Ingersoll RG, Park J, Wu-Chou YH, Chen PK, Chong SS, Cheah F, Yeow V, Park BY, Jee SH, Jabs EW, Redett R, Scott AF, Beaty TH (2009) Maternal transmission effects of the PAX genes among cleft case-parent trios from four populations. *Eur J Hum Genet* 17: 831-9. doi: 10.1038/ejhg.2008.250
- Takeyari S, Kubota T, Miyata K, Yamamoto K, Nakayama H, Yamamoto K, Ohata Y, Kitaoka T, Yanagi K, Kaname T, Ozono K (2018) Japanese patient with Cole-carpenter syndrome

- with compound heterozygous variants of SEC24D. *Am J Med Genet A* 176: 2882-2886. doi: 10.1002/ajmg.a.40643
- Tan G, Lenhard B (2016) TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* 32: 1555-6. doi: 10.1093/bioinformatics/btw024
- Urbanova W, Kotova M, Vankova Z (2013) The incidence of cleft lip and palate in the Czech Republic in 1994-2008. *Bratisl Lek Listy* 114: 474-9. doi: 10.4149/bl_2013_099
- Vaivads M, Akota I, Pilmane M (2021) PAX7, PAX9 and RYK Expression in Cleft Affected Tissue. *Medicina (Kaunas)* 57. doi: 10.3390/medicina57101075
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43: 11 10 1-11 10 33. doi: 10.1002/0471250953.bi1110s43
- Visel A, Minovitsky S, Dubchak I, Pennacchio LA (2007) VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* 35: D88-92. doi: 10.1093/nar/gkl822
- Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38: e164. doi: 10.1093/nar/gkq603
- Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, Li D, Choudhary MNK, Li Y, Hu M, Hardison R, Wang T, Yue F (2018) The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* 19: 151. doi: 10.1186/s13059-018-1519-9
- Wehby GL, Cassell CH (2010) The impact of orofacial clefts on quality of life and healthcare use and costs. *Oral Dis* 16: 3-10. doi: 10.1111/j.1601-0825.2009.01588.x
- Wehby GL, Castilla EE, Goco N, Rittler M, Cosentino V, Javois L, McCarthy AM, Bobashev G, Litavec S, Mariona A, Dutra G, Lopez-Camelo JS, Orioli IM, Murray JC (2006) Description of the methodology used in an ongoing pediatric care interventional study of children born with cleft lip and palate in South America [NCT00097149]. *BMC Pediatr* 6: 9. doi: 10.1186/1471-2431-6-9
- Wilderman A, VanOudenhove J, Kron J, Noonan JP, Cotney J (2018) High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep* 23: 1581-1597. doi: 10.1016/j.celrep.2018.03.129
- Xu S, Liu P, Chen Y, Chen Y, Zhang W, Zhao H, Cao Y, Wang F, Jiang N, Lin S, Li B, Zhang Z, Wei Z, Fan Y, Jin Y, He L, Zhou R, Dekker JD, Tucker HO, Fisher SE, Yao Z, Liu Q, Xia X, Guo X (2018) Foxp2 regulates anatomical features that may be relevant for vocal behaviors and bipedal locomotion. *Proc Natl Acad Sci U S A* 115: 8799-8804. doi: 10.1073/pnas.1721820115
- 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*: 2022.02.28.482338. doi: 10.1101/2022.02.28.482338
- Yin X, Li J, Li Y, Zou S (2019) Maternal alcohol consumption and oral clefts: a meta-analysis. *Br J Oral Maxillofac Surg* 57: 839-846. doi: 10.1016/j.bjoms.2019.08.013
- 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, Liu Z, Yan H, Mangold E, Ruczinski I, Liu J, Marazita ML, Ludwig KU, Beaty TH, Zhang X, Sun L, Bian Z (2017) Genome-wide analyses

of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity.
Nat Commun 8: 14364. doi: 10.1038/ncomms14364

Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ (2016) RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics 32: 1423-6. doi: 10.1093/bioinformatics/btw079

Zhang H, Yue H, Wang C, Gu J, He J, Fu W, Hu W, Zhang Z (2017) Novel mutations in the SEC24D gene in Chinese families with autosomal recessive osteogenesis imperfecta. Osteoporos Int 28: 1473-1480. doi: 10.1007/s00198-016-3866-2