

## **Title**

### **The PAX1 locus at 20p11 is a modifier for bilateral cleft lip only**

## **Authors**

Sarah W. Curtis, Department of Human Genetics, Emory University, Atlanta, GA,

swhelan@emory.edu

Daniel Chang, Department of Human Genetics, Emory University, Atlanta, GA,

daniel.chang248@gmail.com

Myoung Keun Lee, Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA, myl2@pitt.edu

John R. Shaffer, Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA,

john.r.shaffer@pitt.edu

Karlijne Indencleef, Medical Imaging Research Center, University Hospitals Leuven, Leuven, Belgium; Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium,

karlijne.indencleef@kuleuven.be

Michael P. Epstein, Department of Human Genetics, Emory University, Atlanta, GA,

mpepste@emory.edu

David J. Cutler, Department of Human Genetics, Emory University, Atlanta, GA,

djcutle@emory.edu

Jeffrey C. Murray, Department of Pediatrics, University of Iowa, Iowa City, IA, Jeff-

murray@uiowa.edu

Eleanor Feingold, Department of Human Genetics & Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, feingold@pitt.edu

Terri H. Beaty, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, tbeaty1@jhu.edu

Peter Claes, Medical Imaging Research Center, University Hospitals Leuven, Leuven, Belgium;

Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium; Department of Human Genetics, KU Leuven, Leuven, Belgium, peter.claes@kuleuven.be

Seth M. Weinberg, Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA, smwst46@pitt.edu

Mary L. Marazita, Center for Craniofacial and Dental Genetics, Department of Oral Biology, and Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, marazita@pitt.edu

Jenna C. Carlson, Department of Biostatistics & Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, jnc35@pitt.edu

Elizabeth J. Leslie, Department of Human Genetics, Emory University, Atlanta, GA,

ejlesli@emory.edu

## **Acknowledgments:**

The authors thank the dedicated field staff, collaborators, and participating families for their important contributions to this study. This work was supported by grants from the National Institutes of Health (NIH) including: R00-DE025060 [EJL], 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 [TB], U01-DE018993 [TB].

National Institute of Dental and Craniofacial Research (U01-DE020078; R01-DE027023 [SMW]). Funding for genotyping by the National Human Genome Research Institute (X01-HG007821) and funding for initial genomic data cleaning by the University of Washington provided by contract HHSN268201200008I from the National Institute for Dental and Craniofacial Research awarded to the Center for Inherited Disease Research.

## **Abstract**

Nonsyndromic orofacial clefts (OFCs) are the most common craniofacial birth defect in humans and, like many complex traits, OFCs are phenotypically and etiologically heterogeneous. The phenotypic heterogeneity of OFCs extends beyond the structures affected by the cleft (e.g., cleft lip (CL) and cleft lip and palate (CLP)) to other features, such as the severity of the cleft. Here, we focus on bilateral and unilateral clefts as one dimension of OFC severity. Unilateral clefts are more frequent than bilateral clefts for both CL and CLP, but the genetic architecture of these subtypes is not well understood, and it is not known if genetic variants predispose for the formation of one subtype over another. Therefore, we tested for subtype-specific genetic associations in 44 bilateral CL (BCL) cases, 434 unilateral CL (UCL) cases, 530 bilateral CLP cases (BCLP), 1123 unilateral CLP (UCLP) cases, and unrelated controls (N = 1626), using the mixed-model approach implemented in GENESIS. While no novel loci were found in subtype-specific analyses comparing cases to controls, the genetic architecture of UCL was distinct compared to BCL, with 43.8% of suggestive loci ( $p < 1.0 \times 10^{-5}$ ) having non-overlapping confidence intervals between the two subtypes. To further understand the genetic risk factors for severity differences, we then performed a genome-wide scan for modifiers using a similar mixed-model approach and found one genome-wide significant modifier locus on 20p11 ( $p = 7.53 \times 10^{-9}$ ), 300kb downstream of *PAX1*, associated with higher odds of BCL compared to UCL, which also replicated in an independent cohort ( $p = 0.0018$ ) and showed no effect in BCLP ( $p > 0.05$ ). We further found that SNPs at this locus were associated with normal human nasal shape. Taken together, these results suggest bilateral and unilateral clefts may have differences in their genetic architecture, especially between CL and CLP. Moreover, our results suggest BCL, the rarest form of OFC, may be genetically distinct from the other OFC subtypes. This expands our understanding of genetic modifiers for subtypes of OFCs and further elucidates the genetic mechanisms behind the phenotypic heterogeneity in OFCs.

## **Introduction**

Orofacial clefts (OFCs), including cleft lip (CL), cleft lip and palate (CLP) and cleft palate (CP), are common, complex birth defects. Affecting 1 in 700 births worldwide, they are caused when one or more of the developmental programs during the first seven weeks of pregnancy that determine the form the face do not occur properly (1). While some OFCs present in conjunction with other congenital abnormalities, a majority of OFCs are classified as isolated, nonsyndromic (nsOFC), which are caused by a complex combination of genetic and environmental factors and have been the focus of numerous genome-wide association studies (2-13).

OFCs have striking phenotypic heterogeneity. OFCs are typically categorized into three subtypes: cleft lip only (CL), cleft lip and palate (CLP), and cleft palate only (CP) where CL includes clefts confined to the primary palate, CLP includes clefts that extend into the secondary palate, and CP affects the secondary palate only. CL and CLP are often combined into a more general category of cleft lip with or without cleft palate (CL/P) based on the shared defect of the primary palate. OFCs affecting the primary palate can also be further subdivided based on morphological details to capture severity, including the laterality (unilateral or bilateral), the side of unilateral clefts (left or right), or the completeness of the cleft.

Population-based studies estimating recurrence risks have focused on different classifications of OFCs, and the resulting estimates can inform genetic models and the design of association studies. For example, among CLP cases, there is no difference in the risk of either CL or CLP among their first-degree relatives; suggesting a shared genetic etiology (14, 15) contributing to the rationale of studying CL/P in genetic association studies and many of the known risk loci show similar effects between CL and CLP (2-4, 16) (17).

Less is known about severity in CL and CLP or if there is a separate genetic component to cleft lip severity. Recurrence risk estimates based on severity are limited by sample size and have yielded mixed results. Semi-quantitative measures of completeness showed no effect of severity on estimated recurrence risks (15). However, the recurrence risk for bilateral clefts is higher than for unilateral clefts, indicating this more severe cleft type tends to recur more often in family members (14). Previous studies examining genetic factors associated with bilateral vs. unilateral clefts have been limited to targeted sequencing of a few selected candidate loci (17), although this work has suggested the presence of a genetic contribution to the different subtypes of cleft lip. Therefore, we set out to perform a genome-wide association study (GWAS) to determine if there are additional genetic variants that modify cleft severity by focusing on bilateral and unilateral clefting in CL and CLP cases.

## **Methods**

### **Sample collection and SNP quality control**

This study used samples from the Pittsburgh Orofacial Cleft (POFC) Study. The details of the sample collection and genotype quality control (QC) have been described previously (2, 18-20). Briefly, these samples came from 18 sites in 13 countries, including in the continental United States, Guatemala, Argentina, Colombia, Puerto Rico, China, Philippines, Denmark, Turkey, and Spain. All sites had Institutional Review Board (IRB) approval, both locally and at the University of Pittsburgh or University of Iowa, for genomic studies and data sharing. The original study

recruited individuals with OFCs, their unaffected relatives, and unrelated controls (individuals with no known family history of OFCs or other craniofacial anomalies; N = 1626). For the current study, affected individuals were classified as either having a bilateral cleft lip (BCL; N = 44), a bilateral cleft lip and palate (BCLP; N = 530), a unilateral cleft lip (UCL; N = 434), or a unilateral cleft lip and palate (UCLP; N = 1123). Although this sample was not recruited with a population-based approach, the relative frequencies of these cleft types in the POFC study are consistent with epidemiological reports of subtypes (21). Each cleft subtype was present in each ancestry group, as defined by principal components (PCs) of genetic markers (Table S1). Subjects where the specific subtype of cleft was not known were excluded from this study. Related, affected individuals were retained in this study and a genetic relatedness matrix (GRM) was used to adjust for relationships within and across families (see below).

Samples were genotyped for approximately 580,000 single nucleotide polymorphic (SNP) markers from the Illumina HumanCore+Exome array, of which approximately 539,000 SNPs passed quality control filters recommended by the Center for Inherited Disease Research (CIDR) and the Genetics Coordinating Center (GCC) at the University of Washington (2). These data was then phased with SHAPEIT2 (22) and imputed with IMPUTE2 (23) to the 1000 Genomes Project Phase 3 release (September 2014) reference panel. The most-likely imputed genotypes were selected for statistical analysis if the highest probability ( $r^2$ ) > 0.9. SNP markers showing deviation from Hardy-Weinberg equilibrium in European controls, a minor allele frequency or MAF < 5%, or imputation INFO scores < 0.5 were filtered out of all subsequent analyses. A GRM was calculated from a set of LD-pruned genotyped SNPs as defined by GCTA using the package SNPRelate (24).

## Statistical analyses

Subtype-specific genome-wide association study (GWAS): Single subtype genome-wide tests were done by comparing cases from each subtype to a group of unrelated controls to test for genetic variants associated with each cleft subtype. The association between every imputed genetic variant and laterality type was tested using the generalized linear mixed model (GMMAT) (25) as implemented in the GENESIS software package (26). Sex and the estimated GRM were adjusted for under the null model to account for both population substructure and relatedness. SNPs with association p-values less than  $5 \times 10^{-8}$  were considered genome-wide significant and those with p-values less than  $1 \times 10^{-5}$  were considered 'suggestive' and were used for downstream enrichment and comparison analyses. The unadjusted odds ratio (OR) for each SNP was estimated using the minor allele frequency in bilateral cleft cases compared to unilateral cleft cases (27, 28). Regional association plots were made with LocusZoom, where the LD blocks and recombination rates were estimated from European populations (29).

Modifier GWAS: We identified genetic modifiers using case-case group comparisons, directly comparing allele frequencies at each SNP between unilateral and bilateral cleft cases. Thus, this approach has high power to identify genetic risk factors that differ between two subtypes, but no power to find factors important in both groups (i.e., SNPs detected in previous GWAS of CL, CLP, or the combined CL/P group) (24). Therefore, this test has the potential to identify new loci for which there is an effect in only one subtype or where the effects are different between two groups; such loci may be masked in an overall scan when the two groups are combined. We performed modifier analyses for severity separately in the CL and CLP subtypes (UCL vs. BCL

and UCLP vs. BCLP) and combined as CL/P (UCL/P vs. BCL/P). Similar to the modifier analysis above, these tests were done using GMMAT (25) as implemented in GENESIS (26), adjusting for sex and the GRM to account for both population substructure and relatedness. The OR for each SNP was estimated using the minor allele frequency in cases compared to controls (27, 28). Regional association plots were made with LocusZoom(29).

Comparisons between CL and CLP analyses: The estimated ORs for suggestive SNPs (i.e. those with  $p < 1 \times 10^{-5}$ ) in the subtype-specific analyses were compared both within a single severity subtype across cleft type (i.e. BCL vs. BCLP), and across severity types within a single cleft type (e.g. UCL vs. BCL). To compare whether the SNPs associated with individual subtypes were novel compared to what has already been reported in previous GWAS of CL, CLP, or CL/P, the SNPs in these analyses within 50kb of previously associated risk SNPs (2, 5, 6, 20) were also identified. A similar approach was done for the modifier analysis, and the suggestive loci, from either the CL or CLP modifier analyses were compared to see if they either had overlapping 95% confidence intervals (CIs) or gave estimated effects in the same direction.

### **Replication cohort**

To replicate the statistically significant results from our modifier analysis, data from the GENEVA consortium was used, which was described previously (2, 4, 20). Briefly, this cohort recruited case-parent trios, where the affected individual had an oral cleft. The samples were genotyped for approximately 589K SNPs using the Illumina Human610-Quadv.1\_B BeadChip, phased using SHAPEIT, and imputed to the 1000 Genomes Project Phase I (June 2011) reference panel using IMPUTE2. Imputed genotype probabilities were converted to most-likely genotype calls with GTOOL (<http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html>). This dataset was subsequently filtered to only include common SNPs with a MAF > 5%. A subset of individuals was included in both the POFC study and the GENEVA consortium, and these were removed from the replication analysis so that the two groups would be independent. Only the cases from this GENEVA cohort were selected, and they were classified as BCL (N = 28), UCL (N = 326), BCLP (N = 301), UCLP (N = 678). PCs of ancestry were calculated using PLINK (v1.9) (30). Modifier analyses (comparing BCL vs. UCL and BCLP vs. UCLP) were conducted using logistic regression models in PLINK (v1.9), with sex and the first 4 PCs as quantitative covariates. Because of the small sample sizes in the replication cohort and the differences in genotyping arrays and imputation panels, only regions that were significant in the original modifier analysis were tested in this replication strategy. P-values less than a Bonferroni correction for the number of SNPs in the region (0.05/the number of SNPs tested) were considered to be evidence of significant replication.

### **Association with normal facial variation**

The genome-wide significant modifier locus was further examined in relation to normal facial variation by reviewing the association results of SNPs in this locus in a GWAS meta-analysis of facial shape in two large cohorts (n= 8,246) from the US (MetaUS) and UK (MetaUK) (31). To analyze normal facial variation, a data-driven global-to-local facial segmentation approach was performed. Multivariate GWAS was then performed in each of the resulting 63 hierarchically arranged facial segments. More information on the analysis pipeline and the cohorts can be found in the initial study (31).

## Epigenomic context of results

Topologically-associated domains (TADs) were defined for significantly associated loci using the H1-ESC cell line in 3D Genome Browser (32). Functional enrichment was tested by first annotating all of the SNPs to the craniofacial functional regions defined by Wilderman *et al.* (33) for human embryos at CS13, CS14, CS15, CS17, and CS20 (4.5-8 weeks post conception). Enrichment tests were done using a chi-square test with the top SNPs ( $p < 1 \times 10^{-3}$ ) for both modifier analyses and each subtype analysis, so that group sizes would be sufficient for testing, and estimated ORs and their 95% CIs were calculated.

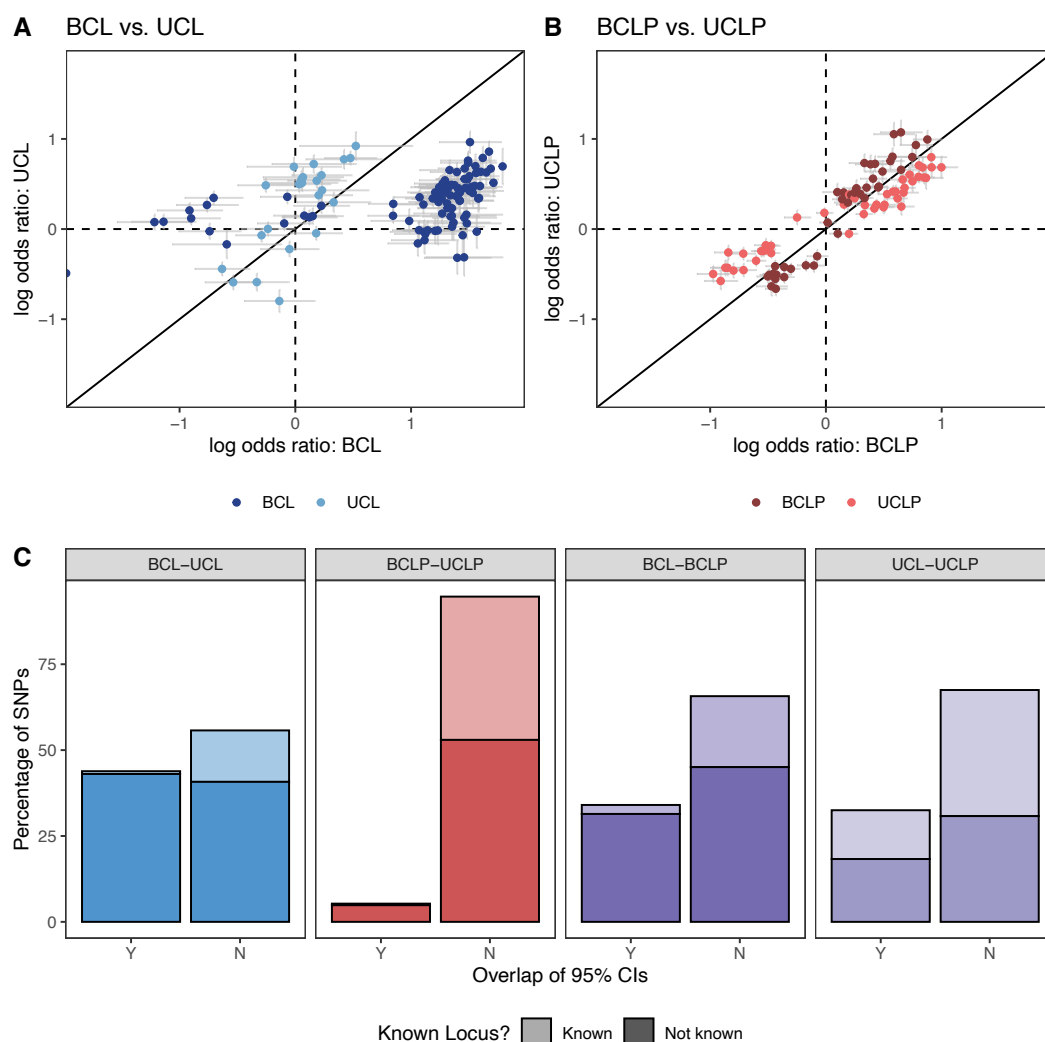
## Results

We performed a subtype-specific genome-wide analysis for BCL, UCL, BCLP, and UCLP cases by comparing cases of each subtype to unaffected controls. This approach can detect variants associated with increased risk for an OFC in general, but also has the potential to identify variants that increase the risk for one or more subtypes of OFC. A single SNP in chromosome 3q28 achieved genome-wide significance in the analysis of BCL (rs72439195;  $p = 3.69 \times 10^{-8}$ ), and 90 regions yielded suggestive evidence, most of which have not been previously implicated in OFC formation. However, some of these regions, like 14q32.33 (lead SNP: rs61996057;  $p = 8.07 \times 10^{-8}$ ; Figure S1A, Figure S2A, Figure S3, Table S2), have been implicated in syndromes with facial dysmorphisms (34-36). In the analysis of UCL, two loci reached genome-wide significance (8q24 and 1q32), both of which are recognized genetic risk loci for CL/P (Figure S1B, Figure S2B, Table S3) (2, 4, 7-10). Among the 21 suggestive loci, 17 have not been previously associated with OFCs which may reflect a lack of GWAS focused specifically on CL. Some of these loci, such as 2q13 (lead SNP: rs6542368;  $p = 1.06 \times 10^{-7}$ ; Figure S4), are plausible candidates for craniofacial dysmorphism (37). Both BCLP and UCLP have multiple recognized genes/regions, including 8q24 and 17p13, reach genome-wide significance (Figure S1C-D, Figure S2C-D, Table S4, Table S5), and 35 and 41 loci reach suggestive significance, respectively, in this analysis (2, 4, 5, 8, 10, 17).

Because of the apparent differences in suggestive and significant loci in the subtype-specific GWASs, we wanted to characterize similarity or dissimilarity of the overall genetic architectures of UCL, UCLP, BCL, and BCLP. Therefore, we performed pairwise analyses comparing the odds ratios and 95% CIs for SNPs identified as suggestive in the GWAS for each subtype being compared. In the comparison of BCL and UCL SNPs, we found a striking difference in estimated ORs in which 44.03% of 741 SNPs did not have overlapping CIs. A majority of these SNPs originating from the BCL analysis had an OR near 1 the UCL analysis (Figure 1), indicating substantial differences in the genetic architecture of BCL, the more severe group. This was also seen, although to lesser degree when the BCL subtype was compared to BCLP, where the 95% CIs for the estimated ORs did not overlap for the 34.1% of 1181 suggestive SNPs (Figure S5). In contrast, BCLP and UCLP were quite similar, with 94.7% of their 1093 SNPs showing overlapping OR confidence intervals (Figure 1). We also found SNPs with different effects in the subtype-specific analyses were less likely to have been previously reported in analyses of the combined group CL/P (Figure 1). For example, in the BCL-UCL comparison, 26.8% of SNPs with overlapping estimated effect sizes were recognized CL/P risk SNPs, indicating these SNPs may predispose to OFC risk but have no effect on specific subtypes.

However, only 1.8% of SNPs differing in their effect sizes were previously reported, significantly less than expected by chance alone ( $p = 2.41 \times 10^{-20}$ ). This pattern held for all comparison groups (Table S6). We reasoned that SNPs predisposing to any type of bilateral cleft could be identified by first selecting SNPs that had non-overlapping CIs between BCL and UCL that also had overlapping CIs between BCL and BCLP. However, only 4 SNPs met these criteria and all of them also showed nominal significance in UCLP and had overlapping CIs. We employed the same strategy to identify SNPs predisposing to any type of unilateral cleft, but were similarly unsuccessful, supporting the notion that subtype-specific risk factors are not shared between CL and CLP in this sample.

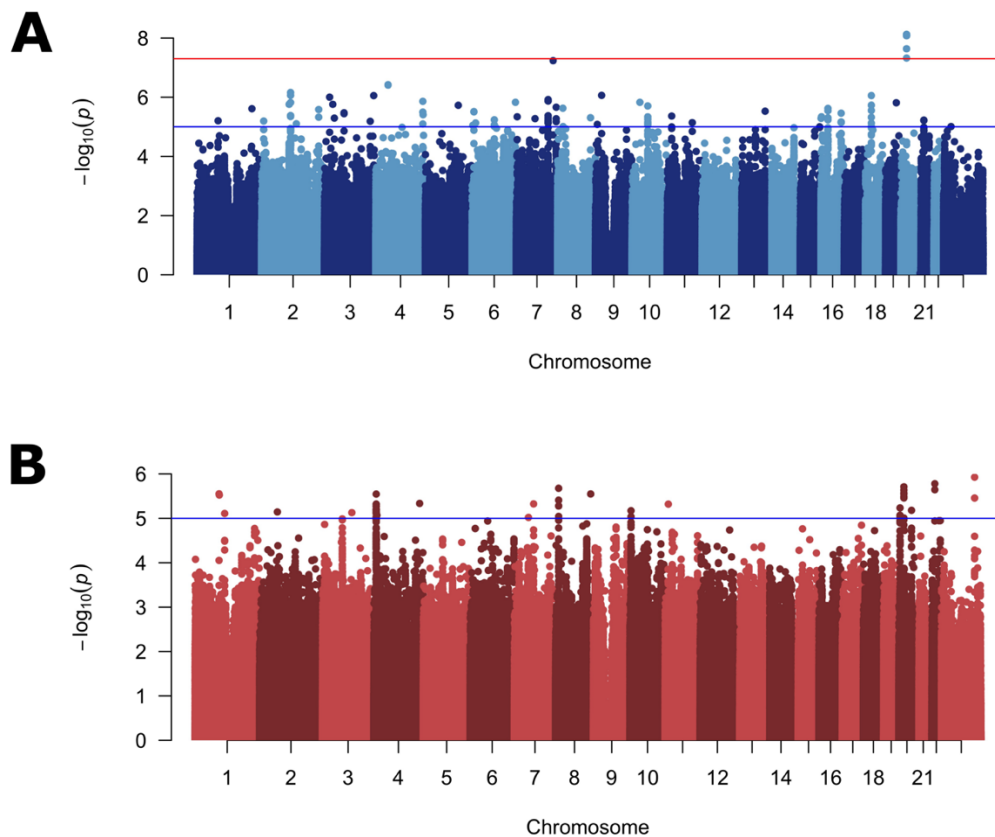
**Figure 1:** The log odds ratio for SNPs that were suggestive ( $p < 1 \times 10^{-5}$ ) or significant ( $p < 5 \times 10^{-8}$ ) in the subtype-specific case-control analyses were compared between BCL and UCL (A), BCLP and UCLP (B), and were classified by whether the confidence interval for the odds ratio overlapped and whether the variant was known (C).





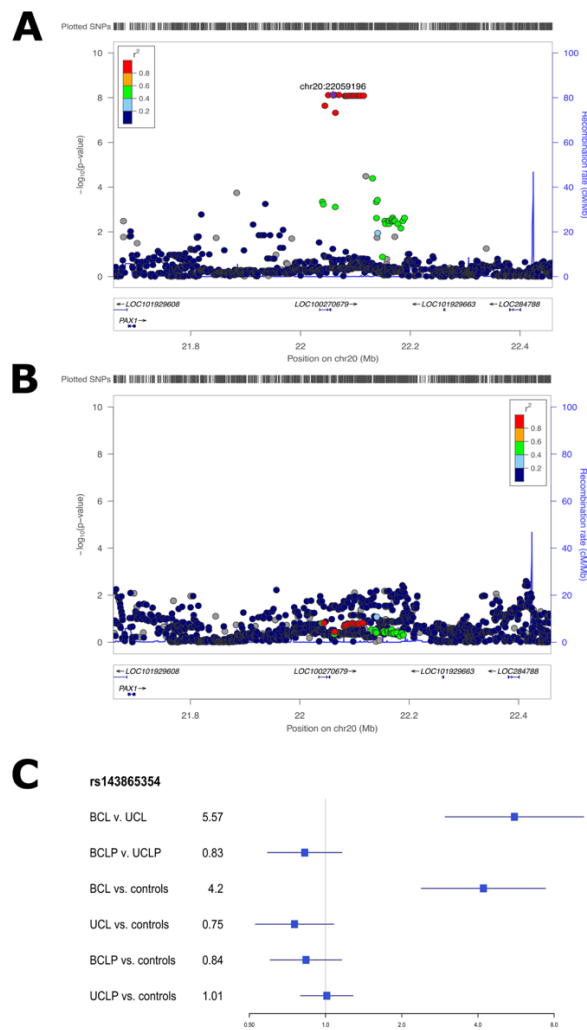
To disentangle the effects of SNPs on specific subtypes from more general effects on OFC risk, we performed a genome-wide bilateral vs. unilateral modifier analysis in CL and CLP cases. Because this is a case-to-case group comparison, this analysis would not be able to detect variants generally important for both CL or CLP risk, but would detect variants important for the formation of one severity subtype compared to the other. In the modifier analysis of CL, one locus on chromosome 20p11 reached genome-wide significance (lead SNP: rs143865354;  $p = 7.53 \times 10^{-9}$ ) and 47 other SNPs yielded suggestive significance (Figure 2A; Table S7; Figure S6A). In the modifier analysis for CLP, no loci reached genome-wide significance, but 19 loci yielded suggestive significance (Figure 2B; Table S8; Figure S6B). Interestingly, when CL and CLP were combined (as is typical in genetic analyses of OFCs), no loci reached genome-wide significance, and only 3 loci gave suggestive significance (Figure S7; Table S9), raising the possibility that these modifiers may not be shared between CL and CLP.

**Figure 2:** Manhattan plots of  $-\log_{10}(p\text{-values})$  from the bilateral vs. unilateral modifier analysis in participants with cleft lip (A), and cleft lip and palate (B). Lines indicate suggestive (blue) and genome-wide (red) thresholds for statistical significance. The genomic inflation factors were 0.96 and 1.01, respectively.



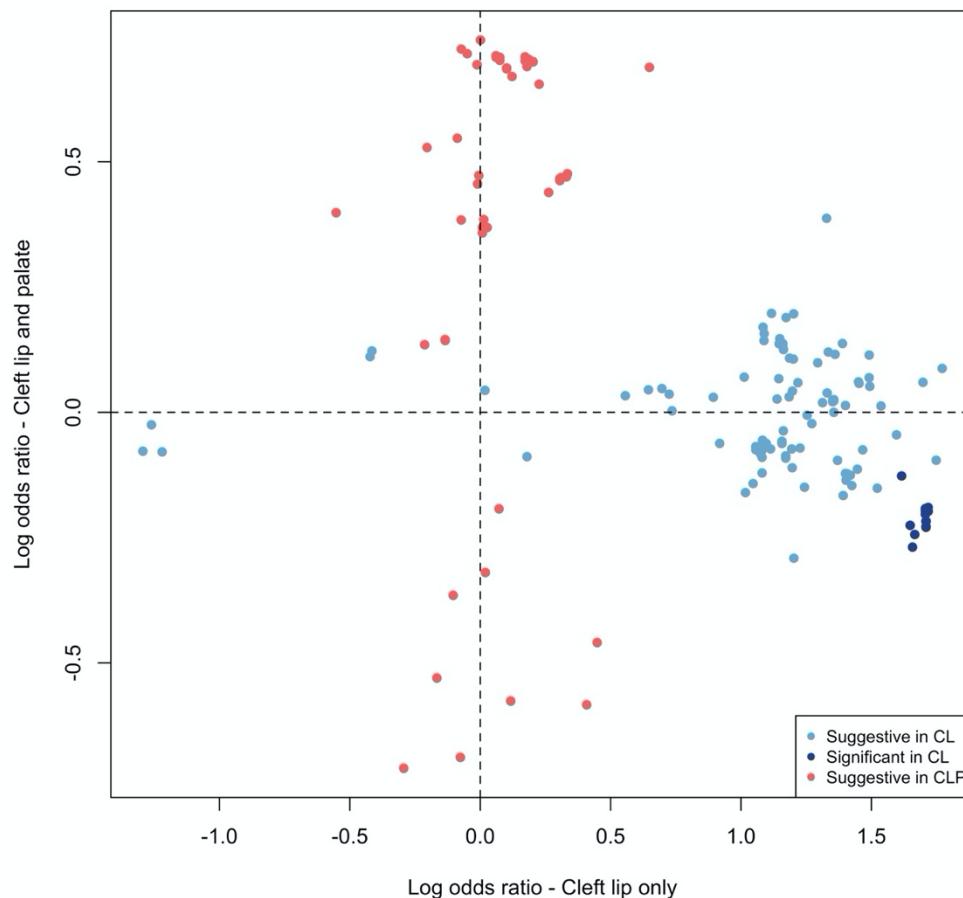
The associated SNPs on 20p11 lie within *LINC01432*, and are within the same topologically associated domain as *PAX1* (Figure 3A; Figure S8). This locus was not significant ( $p > 0.05$ ) in the modifier analysis of CLP (Figure 3B). Additionally, when the OR for the lead SNP in this region was compared between CL and CLP cases, the direction of effect was not consistent (with either a 95% CI or a 99% CI; Figure 3C). We replicated the 20p11 region in an independent sample of 28 BCL cases, 329 UCL cases, 306 BCLP cases, and 685 UCLP cases. In this 20p11 region, there were 8 SNPs passing filtering in the CL modifier analysis.

**Figure 3:** Regional association plots showing  $-\log_{10}(p\text{-values})$  for the novel genome-wide significant peaks at 20p11 in the modifier analysis in cleft lip (A) and cleft lip and palate (B). Plots were generated using LocusZoom (29). The recombination overlay (blue line, right y-axis) indicates the boundaries of the LD block. Points are color coded according to pairwise LD ( $r^2$ ) with the index SNP. The odds ratio for this locus in each of the modifier and subtype specific loci were also compared (C).



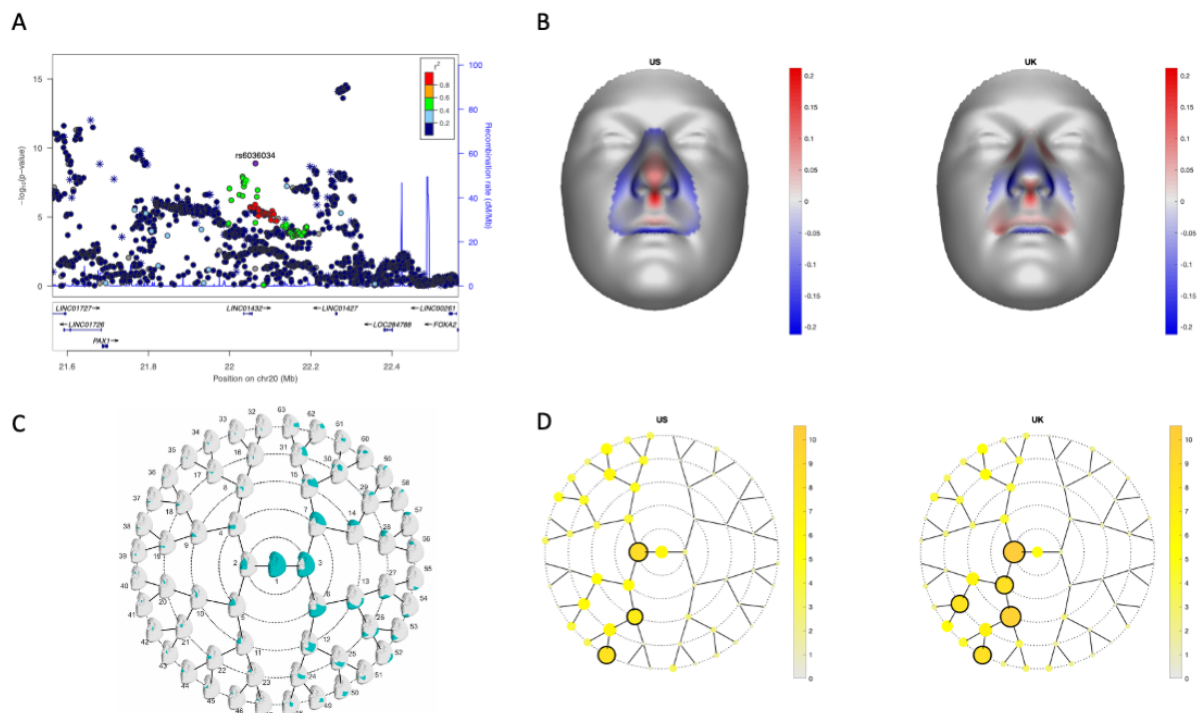
While none of these SNPs were the same as those in the original analysis, one SNP (rs28970569) was also a significant modifier in the replication cohort (OR = 3.83, 95% CI = 1.64-8.95,  $p = 0.0018$ ; Table S10). In the CLP modifier analysis, 9 SNPs passed our filters but none of were significant modifiers, consistent with the results for 20p11 in our discovery sample ( $p > 0.05$ ; Table S11). Additionally, we wanted to determine the extent to which the genetic modifiers in CL were similar to the genetic modifiers in the CLP genome-wide analysis. To test this, we compared SNPs that were suggestive ( $p < 1 \times 10^{-5}$ ) in either the CL or CLP modifier analyses. Notably, there was no overlap between the list of suggestive SNPs in CL and the list of SNPs suggestive in CLP. Moreover, the estimated ORs were not positively correlated and all of the suggestive SNPs in the analysis of CL had no effect in CLP and *vice versa* (Figure 4), and a majority of the SNPs in each analysis were not near regions previously associated with CL/P (Table S12). Cumulatively, these results suggest the 20p11 modifier for bilateral vs. unilateral OFCs is specific to CL.

**Figure 4:** The log odds ratios for 188 SNPs that were suggestive ( $p < 1 \times 10^{-5}$ ) or significant ( $p < 5 \times 10^{-8}$ ) in the modifier analysis in CL or CLP were compared. No SNPs were genome-wide significant in CLP. No SNPs were significant or suggestive in both CL and CLP.



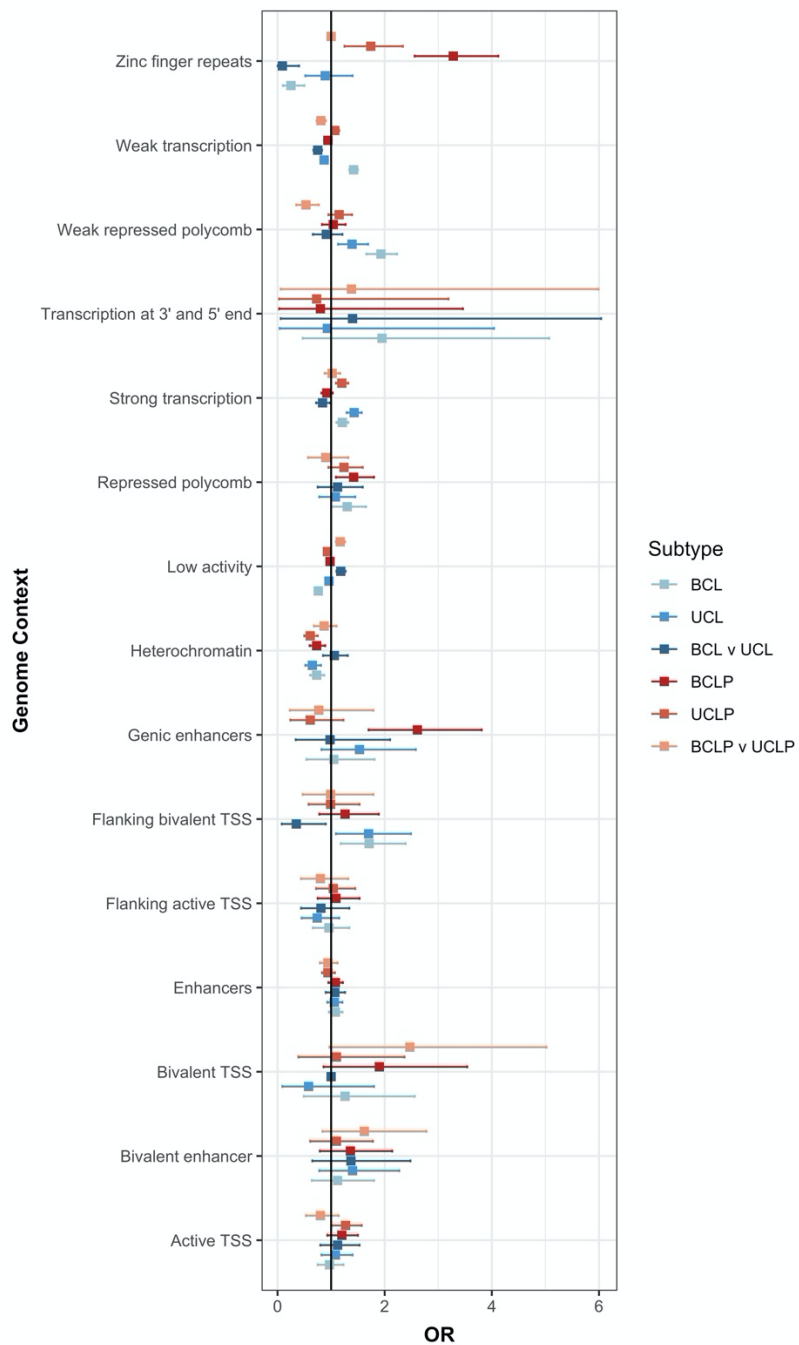
Although the 20p11 locus had not previously been associated with risk to OFCs, it has been associated with variation in normal facial structures. Therefore, we next investigated whether the BCL modifier SNPs were also associated with normal facial variation as that could give insights into how these SNPs might influence cleft severity. We found that rs6036034, a SNP in the 20p11 region in LD with rs143865354 ( $R^2 = 0.522$ ;  $p = 4.75 \times 10^{-8}$  in BCL vs. UCL), was associated with normal variation in nose morphology ( $p = 2.63 \times 10^{-11}$ ), specifically projection of the nasal tip and columella and breadth of the nasal alae (Figure 5). These are the same structures disrupted by cleft lip and are derived from the lateral nasal processes where *PAX1* is expressed (38). Moreover, rs143865354 shows modest evidence of being an eQTL for *PAX1* in skin ( $p = 2.9 \times 10^{-5}$ ) in GTEx.

**Figure 5:** LocusZoom plots for the association of normal facial variation and rs6036034 (A). Points are color coded based on linkage disequilibrium ( $R^2$ ) in Europeans. The asterisks represent genotyped SNPs, the circles represent imputed SNPs. The normal displacement (displacement in the direction locally normal to the facial surface) in each quasi-landmark of the facial segment reaching the lowest p-value in MetaUS and MetaUK going from the minor to the major allele SNP variant (B). Blue, inward depression; red, outward protrusion. Global-to-local facial segmentation plot that shows the 63 facial segments represented in teal obtained using hierarchical spectral clustering (C). The  $-\log_{10}(p\text{-value})$  of the meta-analysis p-values per facial segment in MetaUS and MetaUK. Black-encircled facial segments have reached a genome-wide p-value ( $p = 5.00 \times 10^{-8}$ ) (D).



We were also interested in testing whether differences in genetic architecture in BCL, UCL, BCLP, and UCLP at the SNP level were also reflected in functional elements involved in facial development. Therefore, we tested whether SNPs associated with each subtype were enriched in similar functional regions defined by epigenetic marks in human embryonic craniofacial tissues (33). For some elements, the apparent enrichment or depletion was consistent across subtypes. For example, BCL, UCL, BCLP, and UCLP SNPs were similarly depleted in heterochromatin regions, and most were enriched in regions of strong transcription. However, there were some regions showing opposite enrichments in the different subtypes. For example, zinc finger repeat regions were enriched in both BCLP and UCLP, but were depleted in BCL (Figure 6). Interestingly, the severity modifiers for both CL and CLP were depleted in regions of weak transcription, and enriched in regions of low activity. Some of the suggestive modifier loci for CLP were enriched in bivalent transcription start sites but none of the putative modifiers for risk to CL were enriched in functional domain. These enrichment/depletions were consistent throughout craniofacial development (4.5-8 weeks post conception; Figure S9; Table S13). These observations, while not definitive, lend some support the idea that although at the SNP level, the genetic underpinnings for cleft subtypes are distinct, this may not extend entirely to gross differences in functional element enrichments. Deciphering the true underlying mechanism(s) resulting in bilateral and unilateral CL and CLP will require a locus-by-locus investigation.

**Figure 6:** Enrichment of the top SNPs associated in either the CL modifier analysis, CLP modifier analysis, and each subtype analysis ( $p < 1 \times 10^{-3}$ ) were tested in each functional region defined during craniofacial development (CS15).



## Discussion

While there have been many studies identifying genetic variants that influence overall risk to CL/P and CP, the genetic underpinnings of specific phenotypic subtypes of cleft lip is less studied. This report furthers our understanding of genetic variants associated with specific subtypes of OFC: BCL, UCL, BCLP, and UCLP. We used a modifier analysis which provides more power to find genetic loci differing between two groups, and found one locus on 20p11 that replicated in an independent cohort as significantly associated with the formation of a BCL over a UCL. The associated SNPs were located in several long, noncoding RNAs and within the same TAD (300kb downstream) as the *PAX1* gene. While *PAX1* has not been associated with OFC like its paralog *PAX9* (39), they both are transcription factors with similar DNA-binding domains regulating chondrocyte differentiation and the formation of invertebrate discs, and knock-out mouse models show skeletal abnormalities (40-42). There is also evidence that *PAX1* is upregulated by *SHH*, and in turn, upregulates *SOX5* and *BMP4* (41-43). There is only a limited literature describing *PAX1* expression in the developing face (38) and it has not been previously associated with risk to nonsyndromic OFCs, but *PAX1* is in a pathway with other genes known to be associated with nonsyndromic OFCs (44-49). Additionally, recent studies have shown mutations in *PAX1* cause otofaciocervical syndrome (OTFCS, MIM#615560) which presents with facial dysmorphisms (50, 51), and studies of normal facial variation have found this locus has also been associated with nasal width (the distance between left and right cartilaginous nasal ala) in people of European descent (52) and in people of Latin American descent (53). The link between SNPs at the *PAX1* locus and normal facial shape was further substantiated in our analysis, with effects observed in the nasal tip, columella and alae. These anatomical structures are derived from the lateral and medial nasal processes in the embryo, which form the primary palate. Thus, it is biologically plausible that *PAX1* could affect the development of specific types of craniofacial abnormalities, however, more work is needed to investigate the underlying mechanisms.

While 20p11 was the only genome-wide significant modifier found in this study, this may partly be due to limited sample size in some of the OFC subtypes. It is important to note that when a modifier analysis was conducted on all combined CL and CLP cases, fewer loci reached even suggestive significance, suggesting CL and CLP may have distinct modifiers. Consistent with this, the suggestive modifiers for risk in CL and CLP showed no overlap in estimated effect on risk. This suggests that the lack of overlap is not entirely due to a difference in sample size, but that instead that there is a biological difference in the genetics of laterality in CL compared to CLP.

This is the first study to test for severity modifiers at a genome-wide level, but we previously tested for modifiers in 13 recognized GWAS regions known to be associated with OFCs (54) and found SNPs in *IRF6* were associated with the formation of a unilateral CL/P compared to bilateral CL/P (17). In our study, no SNPs in *IRF6* reached suggestive significance. Our study was larger than the previous study (2339 cases vs. 1001 cases), therefore, this difference may reflect effects of modifiers for cleft subtypes in regions of genome not recognized by previous GWAS of OFCs. This is not surprising given OFC subtypes are typically combined for GWAS, which maximizes statistical power to detect loci associated with overall risk, but would mask loci with different effects in subtypes.

We also conducted analyses comparing each subtype to unrelated controls. This analysis should find loci associated with either overall risk or one particular cleft subtype, but would have less statistical power to detect loci that differ between two subtypes. Most loci achieving genome-wide significance in these analyses were those already recognized to be associated with risk to OFCs (2, 5, 6, 20). There were, however, some loci yielding suggestive evidence of association for several of the subtype-specific analyses not previously reported, but could be in the causal pathway for syndromes with facial dysmorphisms. For example, SNPs in 14q32.33 gave suggestive evidence of association for BCL, with a distinct effect only seen in BCL, and 2q13 yielded suggestive evidence of association for UCL. Microdeletions in both of these regions have been associated with syndromes that include facial dysmorphisms (34-37). The 14q32.33 also contains *JAG2* which is part of the Notch signaling pathway and is important for craniofacial development (55-57).

Overall, our analyses demonstrated that BCL was most distinct from the other three subtypes analyzed and that these modifiers were not shared between CL and CLP. We found that the associated SNPs in all four OFC subtypes were enriched in regions associated with transcription and depleted in heterochromatin regions. This was expected because nonsyndromic OFCs form from the disruption of one of the processes involved in facial development and thus variants associated with any subtype OFC should be enriched in regions active during facial development. It is also consistent with the study defining the functional regions, which showed enrichment in active states for SNPs involved in overall OFC risk (33). Importantly, there were some differences in functional enrichment by subtype. For example, SNPs associated with BCLP and UCLP were enriched in zinc finger repeat regions, however, SNPs showing some evidence of association with BCL were depleted in this same region. This further emphasizes the possibility for a distinct genetic architecture associated with risk to BCL. Additionally, the modifiers for both CL and CLP were depleted in regions associated with active transcription and strongly enriched in regions of low activity. This result is somewhat surprising given it is the opposite of what would be expected for an analysis involving craniofacial development. However, the biological mechanism by which modifiers could affect a phenotype is not known. Therefore, this highlights the need for more studies that test how modifiers mechanistically act.

The findings from this study should also consider its limitations. Many of the subtypes of clefting, particularly BCL, had small sample sizes. Limits of small sample sizes make it likely other subtype-specific genetic loci and modifiers may exist and we are unable to detect them in this statistical analysis. Additionally, we were unable to test for heterogeneity across ancestry groups while testing for subtype-specific genetic risk loci and severity modifiers. This cohort is multiethnic, including people of European, Asian, and Latin American ancestry, and previous studies have shown ancestry-specific association with risk to OFC (2). Studies with larger sample sizes for these clefting subtypes could lead to the discovery of more associated genetic loci and test for differences in associated loci between different ancestry populations.

In summary, we conducted the first genome-wide scan for severity modifiers in a case-case and case-control design focused on nonsyndromic CL and CLP and found a significant modifier in 20p11 downstream from *PAX1* associated with increased risk for BCL over UCL. We also showed these modifiers for CL and CLP were distinct, with the modifiers of one cleft subtype have little to no genetic effect in the other subtypes. Furthermore, in the subtype-specific



GWASs, we found several suggestive loci that had not been previously identified in previous GWASs that combined cleft subtypes. We also found loci associated with BCL were the most distinct from those associated with other cleft subtypes, suggesting the etiology of this rarest subtype of cleft to be unique. Overall, this study expands our understanding of the genetic underpinnings of the genetic and phenotypic heterogeneity of OFCs and suggests new areas of research on cleft lip subtypes.

## **References**

1. Leslie EJ, Marazita ML. Genetics of cleft lip and cleft palate. *Am J Med Genet C Semin Med Genet.* 2013;163C(4):246-58. Epub 2013/10/15. doi: 10.1002/ajmg.c.31381. PubMed PMID: 24124047; PMCID: PMC3925974.
2. 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. 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.* 2016;25(13):2862-72. Epub 2016/04/02. doi: 10.1093/hmg/ddw104. PubMed PMID: 27033726; PMCID: PMC5181632.
3. 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. 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.* 2017;136(3):275-86. Epub 2017/01/06. doi: 10.1007/s00439-016-1754-7. PubMed PMID: 28054174; PMCID: PMC5317097.
4. 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. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nat Genet.* 2010;42(6):525-9. Epub 2010/05/04. doi: 10.1038/ng.580. PubMed PMID: 20436469; PMCID: PMC2941216.
5. 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. Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat Commun.* 2017;8:14364. Epub 2017/02/25. doi: 10.1038/ncomms14364. PubMed PMID: 28232668; PMCID: PMC5333091.
6. 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. Genetic factors define CPO and CLO subtypes of nonsyndromic orofacial cleft. *PLoS Genet.* 2019;15(10):e1008357. Epub 2019/10/15. doi: 10.1371/journal.pgen.1008357. PubMed PMID: 31609978; PMCID: PMC6812857.

7. 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. IRF6 gene variants in Central European patients with non-syndromic cleft lip with or without cleft palate. *Eur J Oral Sci*. 2009;117(6):766-9. Epub 2010/02/04. doi: 10.1111/j.1600-0722.2009.00680.x. PubMed PMID: 20121942.
8. 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, 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. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat Genet*. 2009;41(4):473-7. Epub 2009/03/10. doi: 10.1038/ng.333. PubMed PMID: 19270707.
9. 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, 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. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat Genet*. 2010;42(1):24-6. Epub 2009/12/22. doi: 10.1038/ng.506. PubMed PMID: 20023658.
10. Nikopensius T, Ambrozaityte L, Ludwig KU, Birnbaum S, Jagomagi T, Saag M, Matuleviciene A, Linkeviciene L, Herms S, Knapp M, Hoffmann P, Nothen MM, Kucinskas V, Metspalu A, Mangold E. Replication of novel susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24 in Estonian and Lithuanian patients. *Am J Med Genet A*. 2009;149A(11):2551-3. Epub 2009/10/20. doi: 10.1002/ajmg.a.33024. PubMed PMID: 19839039.
11. Bureau A, Parker MM, Ruczinski I, Taub MA, Marazita ML, Murray JC, Mangold E, Nothen 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, Ling H, Scott AF, Beaty TH. Whole exome sequencing of distant relatives in multiplex families implicates rare variants in candidate genes for oral clefts. *Genetics*. 2014;197(3):1039-44. Epub 2014/05/06. doi: 10.1534/genetics.114.165225. PubMed PMID: 24793288; PMCID: PMC4096358.
12. Fu J, Beaty TH, Scott AF, Hetmanski J, Parker MM, Wilson JE, Marazita ML, Mangold E, Albacha-Hejazi H, Murray JC, Bureau A, Carey J, Cristiano S, Ruczinski I, Scharpf RB. Whole exome association of rare deletions in multiplex oral cleft families. *Genet Epidemiol*. 2017;41(1):61-9. Epub 2016/12/03. doi: 10.1002/gepi.22010. PubMed PMID: 27910131; PMCID: PMC5154821.
13. 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. 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*. 2017;26(4):829-42. Epub 2017/01/15. doi: 10.1093/hmg/ddx012. PubMed PMID: 28087736; PMCID: PMC5409059.

14. Grosen D, Chevrier C, Skytthe A, Bille C, Molsted K, Sivertsen A, Murray JC, Christensen K. A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance. *J Med Genet*. 2010;47(3):162-8. Epub 2009/09/16. doi: 10.1136/jmg.2009.069385. PubMed PMID: 19752161; PMCID: PMC2909851.
15. Sivertsen A, Wilcox AJ, Skjaerven R, Vindenes HA, Abyholm F, Harville E, Lie RT. Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives. *BMJ*. 2008;336(7641):432-4. Epub 2008/02/06. doi: 10.1136/bmj.39458.563611.AE. PubMed PMID: 18250102; PMCID: PMC2249683.
16. 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. Association studies of low-frequency coding variants in nonsyndromic cleft lip with or without cleft palate. *Am J Med Genet A*. 2017;173(6):1531-8. Epub 2017/04/21. doi: 10.1002/ajmg.a.38210. PubMed PMID: 28425186; PMCID: PMC5444956.
17. Carlson JC, Taub MA, Feingold E, Beaty TH, Murray JC, Marazita ML, Leslie EJ. Identifying Genetic Sources of Phenotypic Heterogeneity in Orofacial Clefts by Targeted Sequencing. *Birth Defects Res*. 2017;109(13):1030-8. Epub 2017/08/02. doi: 10.1002/bdr2.23605. PubMed PMID: 28762674; PMCID: PMC5549861.
18. 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, Deleyiannis F, Christensen K, Munger RG, Lie RT, Wilcox A, Romitti PA, Field LL, Padilla CD, Cutiongco-de la Paz EM, Lidral AC, Valencia-Ramirez LC, Lopez-Palacio AM, Valencia DR, Arcos-Burgos M, 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, Deribew M, Koruyucu M, Seymen F, Ma L, de Salamanca JE, Weinberg SM, Moreno L, Cornell RA, Murray JC, Marazita ML. A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. *Am J Hum Genet*. 2016;98(4):744-54. Epub 2016/03/29. doi: 10.1016/j.ajhg.2016.02.014. PubMed PMID: 27018472; PMCID: PMC4833215.
19. 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. Identification of 16q21 as a modifier of nonsyndromic orofacial cleft phenotypes. *Genet Epidemiol*. 2017;41(8):887-97. Epub 2017/11/11. doi: 10.1002/gepi.22090. PubMed PMID: 29124805; PMCID: PMC5728176.
20. 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. A systematic genetic analysis and visualization of phenotypic heterogeneity among orofacial cleft GWAS signals. *Genet Epidemiol*. 2019;43(6):704-16. Epub 2019/06/07. doi: 10.1002/gepi.22214. PubMed PMID: 31172578; PMCID: PMC6687557.
21. Gundlach KK, Maus C. Epidemiological studies on the frequency of clefts in Europe and world-wide. *J Craniomaxillofac Surg*. 2006;34 Suppl 2:1-2. Epub 2006/10/31. doi: 10.1016/S1010-5182(06)60001-2. PubMed PMID: 17071381.

22. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013;10(1):5-6. Epub 2012/12/28. doi: 10.1038/nmeth.2307. PubMed PMID: 23269371.
23. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529. Epub 2009/06/23. doi: 10.1371/journal.pgen.1000529. PubMed PMID: 19543373; PMCID: PMC2689936.
24. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76-82. Epub 2010/12/21. doi: 10.1016/j.ajhg.2010.11.011. PubMed PMID: 21167468; PMCID: PMC3014363.
25. Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T, Szpiro AA, Chen W, Brehm JM, Celson JC, Redline S, Papanicolaou GJ, Thornton TA, Laurie CC, Rice K, Lin X. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *Am J Hum Genet*. 2016;98(4):653-66. Epub 2016/03/29. doi: 10.1016/j.ajhg.2016.02.012. PubMed PMID: 27018471; PMCID: PMC4833218.
26. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, Rice KM, Conomos MP. Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics*. 2019;35(24):5346-8. Epub 2019/07/23. doi: 10.1093/bioinformatics/btz567. PubMed PMID: 31329242.
27. Bland JM, Altman DG. Statistics notes. The odds ratio. *BMJ*. 2000;320(7247):1468. Epub 2000/05/29. doi: 10.1136/bmj.320.7247.1468. PubMed PMID: 10827061; PMCID: PMC1127651.
28. Lloyd-Jones LR, Robinson MR, Yang J, Visscher PM. Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. *Genetics*. 2018;208(4):1397-408. Epub 2018/02/13. doi: 10.1534/genetics.117.300360. PubMed PMID: 29429966; PMCID: PMC5887138.
29. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-7. Epub 2010/07/17. doi: 10.1093/bioinformatics/btq419. PubMed PMID: 20634204; PMCID: PMC2935401.
30. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. Epub 2015/02/28. doi: 10.1186/s13742-015-0047-8. PubMed PMID: 25722852; PMCID: PMC4342193.
31. White JD, Indencleef K, Naqvi S, Eller RJ, Roosenboom J, Lee MK, Li J, Mohammed J, Richmond S, Quillen EE, Norton HL, Feingold E, Swigut T, Marazita ML, Peeters H, Hens G, Shaffer JR, Wysocka J, Walsh S, Weinberg SM, Shriver MD, Claes P. Insights into the genetic architecture of the human face. *bioRxiv*. 2020:2020.05.12.090555. doi: 10.1101/2020.05.12.090555.
32. 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. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol*. 2018;19(1):151. Epub 2018/10/06. doi: 10.1186/s13059-018-1519-9. PubMed PMID: 30286773; PMCID: PMC6172833.

33. Wilderman A, VanOudenhove J, Kron J, Noonan JP, Cotney J. High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep.* 2018;23(5):1581-97. Epub 2018/05/03. doi: 10.1016/j.celrep.2018.03.129. PubMed PMID: 29719267; PMCID: PMC5965702.
34. Geier CB, Piller A, Eibl MM, Ciznar P, Ilencikova D, Wolf HM. Terminal 14q32.33 deletion as a novel cause of agammaglobulinemia. *Clin Immunol.* 2017;183:41-5. Epub 2017/07/15. doi: 10.1016/j.clim.2017.07.003. PubMed PMID: 28705765.
35. Holder JL, Jr., Lotze TE, Bacino C, Cheung SW. A child with an inherited 0.31 Mb microdeletion of chromosome 14q32.33: further delineation of a critical region for the 14q32 deletion syndrome. *Am J Med Genet A.* 2012;158A(8):1962-6. Epub 2012/04/11. doi: 10.1002/ajmg.a.35289. PubMed PMID: 22488736.
36. Maurin ML, Brisset S, Le Lorc'h M, Poncet V, Trioche P, Aboura A, Labrune P, Tachdjian G. Terminal 14q32.33 deletion: genotype-phenotype correlation. *Am J Med Genet A.* 2006;140(21):2324-9. Epub 2006/10/06. doi: 10.1002/ajmg.a.31438. PubMed PMID: 17022077.
37. Hladilkova E, Baroy T, Fannemel M, Vallova V, Miscio D, Bryn V, Slamova I, Prasilova S, Kuglik P, Frengen E. A recurrent deletion on chromosome 2q13 is associated with developmental delay and mild facial dysmorphisms. *Mol Cytogenet.* 2015;8:57. Epub 2015/08/04. doi: 10.1186/s13039-015-0157-0. PubMed PMID: 26236398; PMCID: PMC4521466.
38. Tang LS, Finnell RH. Neural and orofacial defects in Fop1 knockout mice [corrected]. *Birth Defects Res A Clin Mol Teratol.* 2003;67(4):209-18. Epub 2003/07/12. doi: 10.1002/bdra.10045. PubMed PMID: 12854656.
39. Peters H, Neubuser A, Kratochwil K, Balling R. Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes Dev.* 1998;12(17):2735-47. Epub 1998/09/10. doi: 10.1101/gad.12.17.2735. PubMed PMID: 9732271; PMCID: PMC317134.
40. Wilm B, Dahl E, Peters H, Balling R, Imai K. Targeted disruption of Pax1 defines its null phenotype and proves haploinsufficiency. *Proc Natl Acad Sci U S A.* 1998;95(15):8692-7. Epub 1998/07/22. doi: 10.1073/pnas.95.15.8692. PubMed PMID: 9671740; PMCID: PMC21138.
41. Takimoto A, Mohri H, Kokubu C, Hiraki Y, Shukunami C. Pax1 acts as a negative regulator of chondrocyte maturation. *Exp Cell Res.* 2013;319(20):3128-39. Epub 2013/10/02. doi: 10.1016/j.yexcr.2013.09.015. PubMed PMID: 24080012.
42. Rodrigo I, Hill RE, Balling R, Munsterberg A, Imai K. Pax1 and Pax9 activate Bapx1 to induce chondrogenic differentiation in the sclerotome. *Development.* 2003;130(3):473-82. Epub 2002/12/20. doi: 10.1242/dev.00240. PubMed PMID: 12490554.
43. Sivakamasundari V, Kraus P, Sun W, Hu X, Lim SL, Prabhakar S, Lufkin T. A developmental transcriptomic analysis of Pax1 and Pax9 in embryonic intervertebral disc development. *Biol Open.* 2017;6(2):187-99. Epub 2016/12/25. doi: 10.1242/bio.023218. PubMed PMID: 28011632; PMCID: PMC5312110.
44. Lamb AN, Rosenfeld JA, Neill NJ, Talkowski ME, Blumenthal I, Girirajan S, Keelean-Fuller D, Fan Z, Pouncey J, Stevens C, Mackay-Loder L, Terespolsky D, Bader PI, Rosenbaum K, Vallee SE, Moeschler JB, Ladda R, Sell S, Martin J, Ryan S, Jones MC, Moran R, Shealy A, Madan-Khetarpal S, McConnell J, Surti U, Delahaye A, Heron-Longe B, Pipiras E, Benzacken B, Passemard S, Verloes A, Isidor B, Le Caignec C, Glew GM, Opheim KE, Descartes M, Eichler EE,

- Morton CC, Gusella JF, Schultz RA, Ballif BC, Shaffer LG. Haploinsufficiency of SOX5 at 12p12.1 is associated with developmental delays with prominent language delay, behavior problems, and mild dysmorphic features. *Hum Mutat.* 2012;33(4):728-40. Epub 2012/02/01. doi: 10.1002/humu.22037. PubMed PMID: 22290657; PMCID: PMC3618980.
45. Li YH, Yang J, Zhang JL, Liu JQ, Zheng Z, Hu DH. BMP4 rs17563 polymorphism and nonsyndromic cleft lip with or without cleft palate: A meta-analysis. *Medicine (Baltimore).* 2017;96(31):e7676. Epub 2017/08/03. doi: 10.1097/MD.00000000000007676. PubMed PMID: 28767592; PMCID: PMC5626146.
46. Saket M, Saliminejad K, Kamali K, Moghadam FA, Anvar NE, Khorram Khorshid HR. BMP2 and BMP4 variations and risk of non-syndromic cleft lip and palate. *Arch Oral Biol.* 2016;72:134-7. Epub 2016/10/30. doi: 10.1016/j.archoralbio.2016.08.019. PubMed PMID: 27591802.
47. Hammond NL, Brookes KJ, Dixon MJ. Ectopic Hedgehog Signaling Causes Cleft Palate and Defective Osteogenesis. *J Dent Res.* 2018;97(13):1485-93. Epub 2018/07/06. doi: 10.1177/0022034518785336. PubMed PMID: 29975848; PMCID: PMC6262265.
48. Lipinski RJ, Song C, Sulik KK, Everson JL, Gipp JJ, Yan D, Bushman W, Rowland IJ. Cleft lip and palate results from Hedgehog signaling antagonism in the mouse: Phenotypic characterization and clinical implications. *Birth Defects Res A Clin Mol Teratol.* 2010;88(4):232-40. Epub 2010/03/10. doi: 10.1002/bdra.20656. PubMed PMID: 20213699; PMCID: PMC2922848.
49. Dworkin S, Boglev Y, Owens H, Goldie SJ. The Role of Sonic Hedgehog in Craniofacial Patterning, Morphogenesis and Cranial Neural Crest Survival. *J Dev Biol.* 2016;4(3). Epub 2016/08/03. doi: 10.3390/jdb4030024. PubMed PMID: 29615588; PMCID: PMC5831778.
50. Paganini I, Sestini R, Capone GL, Putignano AL, Contini E, Giotti I, Gensini F, Marozza A, Barilaro A, Porfirio B, Papi L. A novel PAX1 null homozygous mutation in autosomal recessive otofaciocervical syndrome associated with severe combined immunodeficiency. *Clin Genet.* 2017;92(6):664-8. Epub 2017/06/29. doi: 10.1111/cge.13085. PubMed PMID: 28657137.
51. Pohl E, Aykut A, Beleggia F, Karaca E, Durmaz B, Keupp K, Arslan E, Palamar M, Yigit G, Ozkinay F, Wollnik B. A hypofunctional PAX1 mutation causes autosomal recessively inherited otofaciocervical syndrome. *Hum Genet.* 2013;132(11):1311-20. Epub 2013/07/16. doi: 10.1007/s00439-013-1337-9. PubMed PMID: 23851939.
52. 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, Hallgrimsson B, Spritz RA, Marazita ML, Weinberg SM. Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. *PLoS Genet.* 2016;12(8):e1006149. Epub 2016/08/26. doi: 10.1371/journal.pgen.1006149. PubMed PMID: 27560520; PMCID: PMC4999139.
53. Adhikari K, Fuentes-Guajardo M, Quinto-Sanchez M, Mendoza-Revilla J, Camilo Chacon-Duque J, Acuna-Alonzo V, Jaramillo C, Arias W, Lozano RB, Perez GM, Gomez-Valdes J, Villamil-Ramirez H, Hunemeier T, Ramallo V, Silva de Cerqueira CC, Hurtado M, Villegas V, Granja V, Gallo C, Poletti G, Schuler-Faccini L, Salzano FM, Bortolini MC, Canizales-Quinteros S, Cheeseman M, Rosique J, Bedoya G, Rothhammer F, Headon D, Gonzalez-Jose R, Balding D, Ruiz-Linares A. A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. *Nat Commun.* 2016;7:11616. Epub 2016/05/20. doi: 10.1038/ncomms11616. PubMed PMID: 27193062; PMCID: PMC4874031.

54. Leslie EJ, Taub MA, Liu H, Steinberg KM, Koboldt DC, Zhang Q, Carlson JC, Hetmanski JB, Wang H, Larson DE, Fulton RS, Kousa YA, Fakhouri WD, Najj A, Ruczinski I, Begum F, Parker MM, Busch T, Standley J, Rigdon J, Hecht JT, Scott AF, Wehby GL, Christensen K, Czeizel AE, Deleyiannis FW, Schutte BC, Wilson RK, Cornell RA, Lidral AC, Weinstock GM, Beaty TH, Marazita ML, Murray JC. Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. *Am J Hum Genet.* 2015;96(3):397-411. Epub 2015/02/24. doi: 10.1016/j.ajhg.2015.01.004. PubMed PMID: 25704602; PMCID: PMC4375420.
55. Hooper JE, Feng W, Li H, Leach SM, Phang T, Siska C, Jones KL, Spritz RA, Hunter LE, Williams T. Systems biology of facial development: contributions of ectoderm and mesenchyme. *Dev Biol.* 2017;426(1):97-114. Epub 2017/04/02. doi: 10.1016/j.ydbio.2017.03.025. PubMed PMID: 28363736; PMCID: PMC5530582.
56. Jiang R, Lan Y, Chapman HD, Shawber C, Norton CR, Serreze DV, Weinmaster G, Gridley T. Defects in limb, craniofacial, and thymic development in Jagged2 mutant mice. *Genes Dev.* 1998;12(7):1046-57. Epub 1998/05/09. doi: 10.1101/gad.12.7.1046. PubMed PMID: 9531541; PMCID: PMC316673.
57. Xu J, Krebs LT, Gridley T. Generation of mice with a conditional null allele of the Jagged2 gene. *Genesis.* 2010;48(6):390-3. Epub 2010/06/10. doi: 10.1002/dvg.20626. PubMed PMID: 20533406; PMCID: PMC3114645.