HHS Public Access

Author manuscript

Am J Med Genet A. Author manuscript; available in PMC 2018 June 01.

Published in final edited form as:

Am J Med Genet A. 2017 June ; 173(6): 1531–1538. doi:10.1002/ajmg.a.38210.

Association studies of low-frequency coding variants in nonsyndromic cleft lip with or without cleft palate

Elizabeth J. Leslie[1], **Jenna C. Carlson**[1,2], **John R. Shaffer**[1,3], **Carmen J. Buxó**[4], **Eduardo E. Castilla**[5,6,7], **Kaare Christensen**[8], **Frederic W.B. Deleyiannis**[9], **L. Leigh Field**[10], **Jacqueline T. Hecht**[11], **Lina Moreno**[12], **Ieda M. Orioli**[6,13], **Carmencita Padilla**[14], **Alexandre R. Vieira**[1,3], **George L. Wehby**[15], **Eleanor Feingold**[1,2,3], **Seth M. Weinberg**[1], **Jeffrey C. Murray**[16], and **Mary L. Marazita**[1,3,17]

[1]Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, 15219 USA

[2]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

[3]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

[4]School of Dental Medicine, University of Puerto Rico, San Juan, Puerto Rico

[5]CEMIC: Center for Medical Education and Clinical Research, Buenos Aires, 1431, Argentina

[6]ECLAMC (Latin American Collaborative Study of Congenital Malformations) at INAGEMP (National Institute of Population Medical Genetics), Rio de Janeiro, 21941-617, Brazil

[7]Laboratory of Congenital Malformation Epidemiology, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, 21040-360, Brazil

[8]Department of Epidemiology, Institute of Public Health, University of Southern Denmark, Odense, DK-5000, Denmark

[9]Department of Surgery, Plastic and Reconstructive Surgery, University of Colorado School of Medicine, Denver, CO, 80045, USA

[10]Department of Medical Genetics, University of British Columbia, Vancouver, V6H 3N1, Canada

[11]Department of Pediatrics, McGovern Medical School and School of Dentistry UT Health at Houston, Houston, TX, 77030, USA

[12]Department of Orthodontics, College of Dentistry, University of Iowa, Iowa City, IA, 52242, USA

[13]Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro, Rio de Janeiro, 21941-617, Brazil

[14]Department of Pediatrics, College of Medicine; and Institute of Human Genetics, National Institutes of Health; University of the Philippines Manila, Manila, The Philippines, 1000; and Philippine Genome Center, University of the Philippines System, Manila, The Philippines 1101

Corresponding Author: Elizabeth J. Leslie, Suite 500 Bridgeside Point, 100 Technology Drive, Pittsburgh, PA 15219, Phone: 412-648-1488, Fax: 412-648-8779, ejl40@pitt.edu.

[15]Department of Health Management and Policy, College of Public Health, University of Iowa, Iowa City, IA, 52246, USA

[16]Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, Iowa, 52242, USA

[17]Clinical and Translational Science, School of Medicine, University of Pittsburgh, Pittsburgh, PA, 15213, USA

Abstract

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a group of common human birth defects with complex etiology. Although genome-wide association studies have successfully identified a number of risk loci, these loci only account for about 20% of the heritability of orofacial clefts. The “missing” heritability may be found in rare variants, copy number variants, or interactions. In this study, we investigated the role of low-frequency variants genotyped in 1995 cases and 1626 controls on the Illumina HumanCore+Exome chip. We performed two statistical tests, Sequence Kernel Association Test (SKAT) and Combined Multivariate and Collapsing (CMC) method using two minor allele frequency cutoffs (1% and 5%). We found that a burden of low-frequency coding variants in *N4BP2*, *CDSN*, *PRTG*, and *AHRR* were associated with increased risk of NSCL/P. Low-frequency variants in other genes were associated with decreased risk of NSCL/P. These results demonstrate that low-frequency variants contribute to the genetic etiology of NSCL/P.

Keywords

orofacial cleft; burden test; rare variant; association

Introduction

Clefts of the lip with or without cleft palate (CL/P) are among the most common human birth defects [Leslie and Marazita 2013; Marazita 2012]. Approximately 30% of clefts are syndromic and have co-occurring structural or cognitive anomalies and are caused primarily by chromosomal anomalies or mutations in single genes [Mossey et al. 2009]. In contrast, the remaining 70% of CL/P occur in isolation and are considered nonsyndromic [Mossey et al. 2009]. Nonsyndromic CL/P (NSCL/P) has agenetically complex etiology, caused by multiple interacting genetic and environmental risk factors. Many approaches have been used to identify genetic risk factors for NSCL/P including linkage, candidate gene, and genome-wide association studies (GWAS). Independent GWAS in several different populations and family types have revealed the significant genetic heterogeneity with at least 15 genetic loci conferring risk for NSCL/P [Beaty et al. 2010; Birnbaum et al. 2009; Grant et al. 2009; Leslie et al. 2016; Mangold et al. 2010; Sun et al. 2015; Wolf et al. 2015]. For some of these loci, such as *IRF6* and *NOG*, follow-up studies have identified functional variants that may directly influence craniofacial development [Leslie et al. 2015; Rahimov et al. 2008].

Despite these successes in identifying loci associated with NSCL/P, we estimate these known loci only account for a small proportion (about 20%) of the heritability in any population. GWAS typically focus on single variant analysis of common SNPs (typically with minor allele frequency (MAF) >5%). Other analyses not covered in traditional GWAS have been proposed to identify the "missing heritability" of complex disease including rare variants, copy number variation, and gene × environment and gene × gene interactions. Rare-variant studies for NSCL/P have largely been limited to resequencing of single candidate genes in small samples [Leslie and Murray 2012]. Although there are many compelling private variants [Jezewski et al. 2003; Leslie et al. 2012; Liu et al. 2016], there has been only limited support for a burden of rare variants [Al Chawa et al. 2014; Leslie et al. 2012; Suzuki et al. 2009]. We previously sequenced thirteen GWAS or candidate loci in 1,500 case-parent trios and failed to identify a burden of rare variants using several annotation categories. However, an agnostic window-based method identified few regions of over-transmission of rare variants [Leslie et al. 2015]. To date, the only genome-wide study of rare variants has been exome sequencing of 55 distant relative pairs [Bureau et al. 2014]. Therefore, a genome-scale examination of rare variants has not been conducted for NSCL/P. In this study, we sought to investigate the role of low-frequency coding variants (MAF<1% and MAF<5%) by genotyping 1995 cases and 1626 controls on an exome chip array and performing gene-based association tests.

Methods

Study Participants

The participants for this study consist of 1995 unrelated cases with NSCL/P and 1626 unrelated controls selected from a larger set of cases, controls, and families recruited from 11 countries across North America, Latin America, Asia, and Europe (Table 1) [Leslie et al. 2016]. Controls had no known history of orofacial clefts (OFCs) nor other craniofacial anomalies. Cases included probands from unrelated families and unrelated individuals recruited as singleton cases. All sites had local IRB approval as well as approval at the University of Iowa or the University of Pittsburgh.

Genotyping

Full details of the genotyping, quality control, and imputation were previously described [Leslie et al. 2016]. Briefly, all samples were genotyped for approximately 580,000 SNPs using an Illumina HumanCore+Exome chip with custom content. Quality control included filters for >2% missing call rate, deviations from Hardy-Weinberg Equilibrium ($p<0.0001$) in participants with genetically confirmed European ancestry, and monomorphic SNPs.

Principal components analysis (PCA) was performed for a subset of unrelated participants using a pruned set of 67K SNPs in low linkage disequilibrium to identify the appropriate principal components (PCs) of ancestry within these data. The resulting eigenvectors were projected onto the remaining set of participants. The resulting PCs strongly tracked global recruitment site and self-reported race/ethnicity. Further, though PCs of ancestry represent orthogonal information across the total multiethnic sample, individual PCs may be highly collinear in specific strata. Therefore, in addition to the PCs generated across all participants,

PCs of ancestry were also generated within each continental group (European, Central/South American, and Asian) for use in stratified analyses.

SNP Selection

To identify the subset of coding variants, we obtained the genomic coordinates for exons of the canonical gene transcripts from the UCSC Genome Browser. All SNPs genotyped on the Illumina HumanCore+Exome chip and located within these exons with at least one observed minor allele and a maximum minor allele frequency of 5% were considered for statistical analyses (totaling 155,361 genotyped variants).

Imputation of sporadic missingness

We imputed sporadic missing genotypes using the IMPUTE2 software and the 1000 Genomes Project (phase 3 release) as the reference panel. Imputation accuracy was assessed by masked variant analysis, demonstrating high quality imputation, with mean concordance of 0.995 for SNPs with minor allele frequency (MAF) < 0.05 and 0.960 for SNPs with MAF ≥ 0.05. The “most-likely” genotypes (i.e. genotypes with the highest probability, Q) were selected for statistical analysis if and only if the highest probability was Q > 0.9. Genotypes that could not be imputed to this degree of certainty remained missing in the final analysis.

Statistical Analyses

We tested gene-based association for low-frequency variants using two tests implemented in RVTESTS [Zhan et al. 2016]. The first, the combined multivariate and collapsing (CMC) method, is a burden test that collapses variants into a single score [Li and Leal 2008]. Because burden tests make the assumption that all variants in the gene have the same direction of effect, we also utilized a second test, the sequence kernel association test (SKAT), which can detect sets of variants with opposite effects [Wu et al. 2011]. The first 18 PCs of ancestry were included in the combined analyses to protect against genomic inflation due to population structure. For stratified analyses, we adjusted for the first 5 population-specific PCs in European and Central/South American samples, and the first 3 PCs in the Asian sample. 14,664 genes with at least two variants were included in the analyses. P-values less-than 3.4×10^{-6} were considered genome-wide significant (i.e. 0.05/14,664) and p-values less than 1.0×10^{-4} were considered to be suggestive. For genes showing evidence of association, we scrutinized the quality of genotype calling by inspecting clustering in allele intensity plots. We performed functional annotation enrichment analysis on genes using ToppFun from the ToppGene Suite [Chen et al. 2009] and assessed significance using Bonferroni adjusted p-values. For comparison, we also performed analysis from a selected list of genes implicated in human OFCs by GWAS or replicated candidate gene studies, OFC syndromes, or animal models [Beaty et al. 2016; Dixon et al. 2011; Leslie and Marazita 2013] (Supplemental Table II).

Results

We performed analyses of low-frequency variants in our multiethnic population and then stratified into three ancestry groups: Europeans, Latin Americans, and Asians (Figure 1, Supplemental Figures 1-4). Two genes achieved genome-wide significance. Variants in

N4BP2 with frequencies less than 5% were associated with NSCL/P in the Latin American population ($p_{SKAT.5} = 5.04 \times 10^{-7}$) (Table II, Supplemental Figure 4). The second, *PFAS*, was associated in all populations combined and in the Latin American subgroup; however, the signal was driven by a single common variant that ultimately failed subsequent quality control checks. When we repeated the analyses excluding this SNP, there was no evidence that *PFAS* variants are associated with NSCL/P (p=0.28).

We also examined genes with p-values less than 1×10^{-4} as having suggestive evidence of association (Table II). In all populations combined, we identified an association with *PRTG* ($p_{CMC.1} = p_{CMC.5} = 3.48 \times 10^{-5}$), *DSC1* ($p_{CMC.5} = 9.03 \times 10^{-5)}$, and *ALDH5A1* ($p_{SKAT.5} = 6.68 \times 10^{-5}$). In the European subgroup, we identified associations in *CDSN* ($p_{CMC.1} = 1.29 \times 10^{-5}$), *AHRR* ($p_{CMC.1} = p_{CMC.5} = 2.65 \times 10^{-5}$), and *GTPBP3* ($p_{SKAT.5} = 8.62 \times 10^{-5}$). In the Latin American and Asian subgroups, nine and eight suggestive signals were identified, respectively (Table II). Among these, several showed consistent results between CMC and SKAT analyses including *CCDC77* in the Latin American group; and *SMIM21*, *CCDC62*, and *RNU5D* in the Asian subgroup.

SKAT and CMC both test groups of SNPs in aggregate, but it is possible that not all of the variants are truly contributing to disease. In addition, the aggregated low-frequency variants may not increase risk, but rather appear at higher frequencies in controls, suggesting a protective effect. Therefore, we examined the frequencies of the coding variants in each gene (Supplementary Table I) and performed single variant tests to determine the direction of effect and whether the association signal was driven by a minority of the variants considered in the analysis. Variants in *AHRR*, *CDSN*, *N4BP2*, and *PRTG* were found at generally higher frequencies in cases compared to controls indicating that low-frequency variants in these genes increase risk of NSCL/P. Of these, *N4BP2* was significant only in the MAF<5% analysis and two variants with MAF>1% appeared to be driving the association. *CDSN* was also notable in that all five variants had higher frequencies in cases than in controls, but no single variant dominated the signal. With the exception of *ALDH5A1*, *HEATR8*, and *YEATS2,* genes for which the directions of association were unclear, low-frequency variants in the remaining 15 genes (i.e., *SMIM21*, *CCDC62*, *DSC1*, etc.) were generally found at higher frequencies in controls and therefore appeared to decrease risk for NSCL/P.

We performed functional enrichment analysis on the group of genes surpassing our suggestive p-value threshold. We noted that this gene list was not enriched for Gene Ontology (GO) terms related to molecular functions or biological processes, nor did the genes group into particular pathways. However, two input genes (*DSC1* and *CDSN*) were annotated for the cellular component GO term “desmosome”, resulting in a statistical enrichment of the term ($p = 4.01 \times 10^{-4}$; Bonferroni adjusted p = 0.02). Interestingly, there was a slight enrichment for interactions with CEP120, a centrosomal protein associated with Joubert syndrome and other ciliopathies; three input genes were annotated for this interaction (*YEATS2*, *KIAA0586*, and *CCDC7*; $p = 8.89 \times 10^{-5}$; Bonferroni adjusted p = 0.039). There were no enrichments for terms related to disease, gene family, or domain. In contrast, when we interrogated a list of 51 genes associated with orofacial clefts in human or mouse (Supplemental Table II), enrichment analysis returned annotations including “transcription factor”, “development”, “periderm”, and “epithelium”. Because the top genes

were not enriched for these terms, we may be uncovering new biological factors underlying risk for NSCL/P.

Discussion

Since the inception of GWAS, the common disease-common variant hypothesis has dominated the search for risk loci for common diseases [Pritchard and Cox 2002; Reich and Lander 2001]. As it became clear that common variants only accounted for a portion of the genetic variance, association studies turned to other types of variants and variants with lower minor allele frequencies [Manolio et al. 2009]. This study was motivated by the fact that multiple GWAS are estimated to account for only 20% of the heritability of NSCL/P. Although there are multiple possible study designs and statistical approaches to address a rare-variant hypothesis, this study focused on tests of low-frequency coding polymorphisms from the Illumina HumanCore+Exome Chip. This approach is limited to known variants, whereas study designs based on sequencing allow the inclusion of novel variants, that may be private in NSCL/P families. We were unable to identify a large number of genes for which low-frequency variants were strongly associated with NSCL/P; only one, *N4BP2*, surpassed our genome-wide significance threshold. We identified larger numbers of genes with suggestive evidence of association, particularly within genetically-defined ancestry groups. This is likely due to increased informativeness of variants within homogenous populations which increases power to detect an association. However, these subgroups are smaller, which simultaneously decreases power. This balance of factors that influence statistical power is particularly important to consider in rare-variant studies, which are notoriously underpowered.

Interestingly, we did not identify associations with known CL/P-risk genes, including those previously implicated in rare variant studies (e.g. *MSX1*, *BMP4*, *GREM1*) [Al Chawa et al. 2014; Leslie and Murray 2012]. One possible reason for this apparent discrepancy is that previous resequencing studies focused on novel, private variants not present on commercial SNP panels while the present study is focused on low-frequency polymorphisms. Another possible reason is that many of the known risk genes are implicated in Mendelian diseases that include syndromic forms of OFCs. These genes tend to be evolutionarily constrained, so we would not expect to find deleterious coding polymorphisms in these genes. However, as OFCs are no longer considered lethal conditions in developed countries, relaxed selection may tolerate higher frequencies of deleterious alleles.

We were unable to identify a suitable replication cohort, as none of the existing GWAS consortia included the ExomeChip. As a consequence, these results should be viewed as preliminary until replicated by independent efforts. However, among our reported associated genes, there are several that could plausibly relate to CL/P risk, based on gene function or expression patterns. Two such genes are *DSC1* and *CDSN*, which were annotated as being part of desmosomes, the intercellular junctions between apposing cells that are essential for mammalian development [Garrod and Chidgey 2008]. Desmosomes are composed of the proteins plakoglobin, plakophilin, and a desmosomal cadherin (e.g. desmocollin or desmoglein). *DSC1* encodes one of the three desmocollin proteins found in humans. Its expression is confined to stratified epithelia, which includes adult skin and the tongue. In

addition, desmosomes include other associated accessory proteins that are necessary for adhesive function, such as corneodesmosin (*CDSN*), expressed in the epidermis, hair follicle, and hard palate epithelium [Jonca et al. 2011]. Inactivation of *Cdsn* in mice causes epidermal barrier dysfunction and hair follicle degeneration, resulting in postnatal death [Leclerc et al. 2009]. During palatogenesis, palatal shelves must adhere together prior to disintegration of the medial edge epithelium and formation of a complete palate; this timing coincides with an increase in desmosomes in the medial edge epithelium [Mogass et al. 2000]. Thus an association between variants in these genes and risk of NSCL/P is biologically plausible although the effects of these specific variants are still unknown.

Low-frequency variants in *PRTG*, *N4BP2*, and *AHRR* demonstrated increased risk for NSCL/P. Of these genes, a possible role in craniofacial development is only apparent for *PRTG*, which encodes protogenin, an immunoglobulin domain-containing receptor involved in survival of rostral cephalic neural crest cells that form osteogenic and chondrogenic cells in the developing face. *AHRR* (aryl hydrocarbon receptor repressor) and *N4BP2* (nedd4 binding protein) do not appear to be strongly expressed in the developing face.

Among the group of genes for which low-frequency variants decreased risk for NSCL/P, there are several interesting genes already implicated in disorders with associated craniofacial features. *EHMT1* (euchromatic histone methytransferase 1) mutations and microdeletions cause Kleefstra syndrome in humans, which includes a range of craniofacial dysmorphologies among the associated features [Kleefstra et al. 2006]. Similarly, mutations in *KIAA0586* cause a range of ciliopathies, including Joubert syndrome [Alby et al. 2015; Bachmann-Gagescu et al. 2015; Malicdan et al. 2015]. Studies of the mouse and chicken homolog, *TALPID3*, indicate that the craniofacial defects in *TALPID3* mutants are due to impaired hedgehog signaling [Buxton et al. 2004]. Finally, de novo mutations in *POLD1* (DNA polymerase delta) have been reported in mandibular hypoplasia, deafness, progeroid features, and lipodystrophy syndrome (MDPL syndrome) [Weedon et al. 2013]. In each of these cases, it appears that deleterious mutations cause craniofacial anomalies as part of syndromes. It is not clear what role the low-frequency coding variants have in decreasing risk of NSCL/P, if any.

In summary, here we provide evidence that a burden of low-frequency genetic variants in *N4BP2*, *CDSN, AHRR*, and *PRTG* increase risk for nonsyndromic NSCL/P; and that low-frequency variants in other genes appear to reduce risk for NSCL/P. Although many of these genes have not previously been implicated in human craniofacial disorders, our findings are corroborated by craniofacial expression patterns or previous associations with craniofacial syndromes. Additional sequencing of these genes should be performed to identify additional risk variants and to replicate these results. Furthermore, as studies of low-frequency variants are underpowered, future studies in much larger sample sizes from all ancestry groups are needed to detect low-frequency variant associations in other genes. Overall, this study demonstrates that low-frequency variants are associated with NSCL/P and may thus contribute to some of the missing heritability of this common, complex structural birth defect.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Many thanks to the participating families world-wide who made this project possible. The tireless efforts by the dedicated field staff and collaborators was similarly essential to the success of this study: Dr. Andrew Czeizel (Hungary, deceased), Dr. Juan C. Mereb (Argentina, deceased), Kathy Bardi, Rebecca DeSensi, Margaret Cooper, Joel Anderton, Wendy Carricato, Kathleen Deeley, Keri Simonette, Matthew Ford, and Dr. Joseph Losee (University of Pittsburgh); Jennifer Standley, Nichole Nidey, and Erin Brothers-Smith (University of Iowa); Ceclia Laurie, Deepti Jain (University of Washington); Kimberly F. Doheny, Marcia Adams, Marie Hurley and Mary Jewell (CIDR, Johns Hopkins University); András Kiss and Eszter Nyulászi (Hungary); Mairim Soto and Ricardo Ledesma (University of Puerto Rico); Rosemary Vance and the Children of the Americas (www.childrenoftheamericas.org, Guatemala); Elena Serna, Rosa Martinez, and Syed S. Hashmi (University of Texas, Houston); Dr. Ajit Ray (India); Dr. You-e Liu (China); Dr. Rick Martin (Missouri); Sondra Valdez (Univ. of Colorado); the Smile Train and Transforming Faces. This work was supported by grants from the National Institutes of Health (NIH) including: K99-DE025060 [EJL], X01-HG007485 [MLM, EF], R01-DE016148 [MLM, SMW], U01-DE024425 [MLM], R37-DE008559 [JCM, MLM], R01-DE009886 [MLM], R21-DE016930 [MLM], R01-DE014667 [LM], R01-DE012472 [MLM], R01-DE011931 [JTH], R01-DE011948 [KC], U01-DD000295 [GLW], R25-MD007607 [CJB], K99-DE024571 [CJB]. Genotyping and data cleaning were provided via an NIH contract to the Johns Hopkins Center for Inherited Disease Research: HHSN268201200008I. Additional support provided by: the Robert Wood Johnson Foundation, AMFDP Grant 72429 [AB]; an intramural grant from the Research Institute of the Children's Hospital of Colorado [FWD]; operating costs support in the Philippines was provided by the Institute of Human Genetics, National Institutes of Health, University of the Philippines, Manila [CP]; grants through FAPERJ, Brazil [IMO]: grant numbers: E-26/102.797/2012, E-26/110.140/2013; grants through CNPq, Brazil [IMO]: grant numbers: 481069/2012-7, 306396/2013-0, 400427/2013-3.

References

Al Chawa T, Ludwig KU, Fier H, Potzsch B, Reich RH, Schmidt G, Braumann B, Daratsianos N, Bohmer AC, Schuencke H, Alblas M, Fricker N, Hoffmann P, Knapp M, Lange C, Nothen MM, Mangold E. Nonsyndromic cleft lip with or without cleft palate: Increased burden of rare variants within Gremlin-1, a component of the bone morphogenetic protein 4 pathway. Birt Defects Res A Clin Mol Teratol. 2014

Alby C, Piquand K, Huber C, Megarbane A, Ichkou A, Legendre M, Pelluard F, Encha-Ravazi F, Abi-Tayeh G, Bessieres B, El Chehadeh-Djebbar S, Laurent N, Faivre L, Sztriha L, Zombor M, Szabo H, Failler M, Garfa-Traore M, Bole C, Nitschke P, Nizon M, Elkhartoufi N, Clerget-Darpoux F, Munnich A, Lyonnet S, Vekemans M, Saunier S, Cormier-Daire V, Attie-Bitach T, Thomas S. Mutations in KIAA0586 Cause Lethal Ciliopathies Ranging from a Hydrolethalus Phenotype to Short-Rib Polydactyly Syndrome. Am J Hum Genet. 2015; 97(2):311–318. [PubMed: 26166481]

Bachmann-Gagescu R, Phelps IG, Dempsey JC, Sharma VA, Ishak GE, Boyle EA, Wilson M, Marques Lourenco C, Arslan M, University of Washington Center for Mendelian G. Shendure J, Doherty D. KIAA0586 is Mutated in Joubert Syndrome. Hum Mutat. 2015; 36(9):831–835. [PubMed: 26096313]

Beaty TH, Marazita ML, Leslie EJ. Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. F1000Res. 2016; 5:2800. [PubMed: 27990279]

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):525–529. [PubMed: 20436469]

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):473–477. [PubMed: 19270707]

Bureau A, Parker MM, Ruczinski I, Taub MA, Marazita ML, Murray JC, Mangold E, Noethen MM, Ludwig KU, Hetmanski JB, Bailey-Wilson JE, Cropp CD, Li Q, Szymczak S, Albacha-Hejazi H, Alqosayer K, Field LL, Wu-Chou YH, Doheny KF, 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):1039–1044. [PubMed: 24793288]

Buxton P, Davey MG, Paton IR, Morrice DR, Francis-West PH, Burt DW, Tickle C. Craniofacial development in the talpid3 chicken mutant. Differentiation. 2004; 72(7):348–362. [PubMed: 15554946]

Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 2009; 37(Web Server issue):W305–311. [PubMed: 19465376]

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

Garrod D, Chidgey M. Desmosome structure, composition and function. Biochim Biophys Acta. 2008; 1778(3):572–587. [PubMed: 17854763]

Grant SF, Wang K, Zhang H, Glaberson W, Annaiah K, Kim CE, Bradfield JP, Glessner JT, Thomas KA, Garris M, Frackelton EC, Otieno FG, Chiavacci RM, Nah HD, Kirschner RE, Hakonarson H. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. J Pediatr. 2009; 155(6):909–913. [PubMed: 19656524]

Jezewski PA, Vieira AR, Nishimura C, Ludwig B, Johnson M, O'Brien SE, Daack-Hirsch S, Schultz RE, Weber A, Nepomucena B, Romitti PA, Christensen K, Orioli IM, Castilla EE, Machida J, Natsume N, Murray JC. Complete sequencing shows a role for MSX1 in non-syndromic cleft lip and palate. J Med Genet. 2003; 40(6):399–407. [PubMed: 12807959]

Jonca N, Leclerc EA, Caubet C, Simon M, Guerrin M, Serre G. Corneodesmosomes and corneodesmosin: from the stratum corneum cohesion to the pathophysiology of genodermatoses. Eur J Dermatol. 2011; 21(2):35–42. [PubMed: 21628128]

Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A, Genevieve D, Cormier-Daire V, van Esch H, Fryns JP, Hamel BC, Sistermans EA, de Vries BB, van Bokhoven H. Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. Am J Hum Genet. 2006; 79(2):370–377. [PubMed: 16826528]

Leclerc EA, Huchenq A, Mattiuzzo NR, Metzger D, Chambon P, Ghyselinck NB, Serre G, Jonca N, Guerrin M. Corneodesmosin gene ablation induces lethal skin-barrier disruption and hair-follicle degeneration related to desmosome dysfunction. J Cell Sci. 2009; 122(Pt 15):2699–2709. [PubMed: 19596793]

Leslie E, Murray J. Evaluating rare coding variants as contributing causes to non-syndromic cleft lip and palate. Clin Genet. 2012:n/a–n/a.

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

Leslie EJ, Mansilla MA, Biggs LC, Schuette K, Bullard S, Cooper M, Dunnwald M, Lidral AC, Marazita ML, Beaty TH, Murray JC. Expression and mutation analyses implicate ARHGAP29 as the etiologic gene for the cleft lip with or without cleft palate locus identified by genome-wide association on chromosome 1p22. Birth Defects Research Part A: Clinical and Molecular Teratology. 2012; 94(11):934–942. [PubMed: 23008150]

Leslie EJ, Marazita ML. Genetics of cleft lip and cleft palate. Am J Med Genet C Semin Med Genet. 2013; 163C(4):246–258. [PubMed: 24124047]

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, Naji 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

Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet. 2008; 83(3):311–321. [PubMed: 18691683]

Liu H, Leslie EJ, Jia Z, Smith T, Eshete M, Butali A, Dunnwald M, Murray J, Cornell RA. Irf6 directly regulates Klf17 in zebrafish periderm and Klf4 in murine oral epithelium, and dominant-negative KLF4 variants are present in patients with cleft lip and palate. Hum Mol Genet. 2016; 25(4):766–776. [PubMed: 26692521]

Malicdan MC, Vilboux T, Stephen J, Maglic D, Mian L, Konzman D, Guo J, Yildirimli D, Bryant J, Fischer R, Zein WM, Snow J, Vemulapalli M, Mullikin JC, Toro C, Solomon BD, Niederhuber JE, Program NCS, Gahl WA, Gunay-Aygun M. Mutations in human homologue of chicken talpid3 gene (KIAA0586) cause a hybrid ciliopathy with overlapping features of Jeune and Joubert syndromes. J Med Genet. 2015; 52(12):830–839. [PubMed: 26386044]

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):24–26. [PubMed: 20023658]

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–753. [PubMed: 19812666]

Marazita ML. The Evolution of Human Genetic Studies of Cleft Lip and Cleft Palate. Annu Rev Genomics Hum Genet. 2012; 13:263–283. [PubMed: 22703175]

Mogass M, Bringas P Jr, Shuler CF. Characterization of desmosomal component expression during palatogenesis. Int J Dev Biol. 2000; 44(3):317–322. [PubMed: 10853828]

Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. Lancet. 2009; 374(9703): 1773–1785. [PubMed: 19747722]

Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant…or not? Hum Mol Genet. 2002; 11(20):2417–2423. [PubMed: 12351577]

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

Reich DE, Lander ES. On the allelic spectrum of human disease. Trends Genet. 2001; 17(9):502–510. [PubMed: 11525833]

Sun Y, Huang Y, Yin A, Pan Y, Wang Y, Wang C, Du Y, Wang M, Lan F, Hu Z, Wang G, Jiang M, Ma J, Zhang X, Ma H, Ma J, Zhang W, Huang Q, Zhou Z, Ma L, Li Y, Jiang H, Xie L, Jiang Y, Shi B, Cheng J, Shen H, Wang L, Yang Y. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun. 2015; 6:6414. [PubMed: 25775280]

Suzuki S, Marazita ML, Cooper ME, Miwa N, Hing A, Jugessur A, Natsume N, Shimozato K, Ohbayashi N, Suzuki Y, Niimi T, Minami K, Yamamoto M, Altannamar TJ, Erkhembaatar T, Furukawa H, Daack-Hirsch S, L'Heureux J, Brandon CA, Weinberg SM, Neiswanger K, Deleyiannis FW, de Salamanca JE, Vieira AR, Lidral AC, Martin JF, Murray JC. Mutations in

BMP4 are associated with subepithelial, microform, and overt cleft lip. Am J Hum Genet. 2009; 84(3):406–411. [PubMed: 19249007]

Weedon MN, Ellard S, Prindle MJ, Caswell R, Lango Allen H, Oram R, Godbole K, Yajnik CS, Sbraccia P, Novelli G, Turnpenny P, McCann E, Goh KJ, Wang Y, Fulford J, McCulloch LJ, Savage DB, O'Rahilly S, Kos K, Loeb LA, Semple RK, Hattersley AT. An in-frame deletion at the polymerase active site of POLD1 causes a multisystem disorder with lipodystrophy. Nat Genet. 2013; 45(8):947–950. [PubMed: 23770608]

Wolf ZT, Brand HA, Shaffer JR, Leslie EJ, Arzi B, Willet CE, Cox TC, McHenry T, Narayan N, Feingold E, Wang X, Sliskovic S, Karmi N, Safra N, Sanchez C, Deleyiannis FW, Murray JC, Wade CM, Marazita ML, Bannasch DL. Genome-wide association studies in dogs and humans identify ADAMTS20 as a risk variant for cleft lip and palate. PLoS Genet. 2015; 11(3):e1005059. [PubMed: 25798845]

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011; 89(1):82–93. [PubMed: 21737059]

Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: An Efficient and Comprehensive Tool for Rare Variant Association Analysis Using Sequence Data. Bioinformatics. 2016

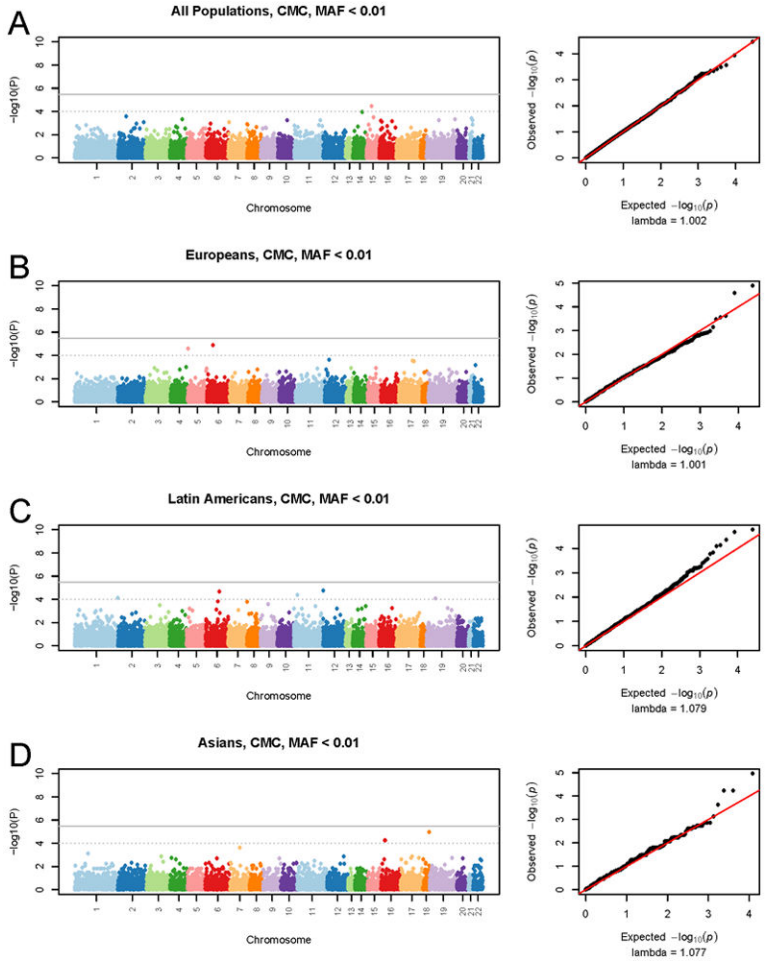

Figure 1. Association results for gene-based tests (CMC) of variants less than 1%

Manhattan plots and QQ plots are shown for the all populations (A), Europeans (B), Asians (C), and Latin Americans (D). The grey line shows Bonferroni-corrected genome-wide significance (i.e. $p<3.4 \times 10^{-6}$) and the dashed grey line shows the suggestive threshold (i.e. $p<1 \times 10^{-4}$).

Table I
Demographic details of cases and controls

Population	Recruitment Site	Controls	NSCL/P Cases
Europeans	Denmark	0	46
	Hungary	253	105
	Spain	0	33
	Turkey	171	172
	United States	411	220
Latin American	Argentina	30	111
	Colombia	277	681
	Guatemala	208	102
	Puerto Rico	106	84
	United States	5	72
Asian	China	27	157
	India	38	51
	Philippines	96	159
Unspecified (included in combined analysis)		4	2
Total		**1626**	**1995**

Table II

Association results for low-frequency variants

Gene	Population	CMC		SKAT		Direction of Association
		1%	5%	1%	5%	
AHRR	Europeans	*2.65E-05*	*2.65E-05*	4.38E-04	4.38E-04	Increase Risk
ALDH5A1	All Populations	0.04	0.42	4.92E-03	*6.68E-05*	Unknown
SMIM21	Asians	*1.09E-05*	*1.09E-05*	*3.34E-05*	*3.34E-05*	Decrease Risk
CCDC62	Asians	NV	*1.09E-05*	NV	*1.15E-05*	Decrease Risk
CCDC77	Latin Americans	*1.70E-05*	*1.70E-05*	6.23E-04	6.23E-04	Decrease Risk
CDSN	Europeans	*1.29E-05*	0.97	7.51E-04	0.04	Increase Risk
DOPEY1	Latin Americans	*2.12E-05*	0.17	2.87E-03	0.06	Decrease Risk
DSC1	All Populations	0.53	*9.03E-05*	0.62	1.64E-03	Decrease Risk
EHMT1	Asians	5.90E-03	*2.51E-05*	0.02	4.38E-04	Decrease Risk
FLYWCH1	Asians	*5.79E-05*	6.41E-03	3.34E-04	0.02	Decrease Risk
GTPBP3 *	Europeans	0.15	1.82E-04	0.56	*8.62E-05*	
HEATR8	Latin Americans	0.05	0.24	*4.95E-05*	*5.99E-05*	Unknown
KIAA0586	Asians	0.02	*1.21E-05*	0.03	1.54E-04	Decrease Risk
N4BP2	Latin Americans	6.56E-03	0.70	1.60E-03	***5.04E-07***	Increase Risk
NLN	Latin Americans	0.08	*2.55E-05*	0.09	9.27E-04	Decrease Risk
OR2T6	Latin Americans	*7.47E-05*	1.21E-03	5.52E-04	0.02	Decrease Risk
OR52E4	Latin Americans	*4.43E-05*	*4.43E-05*	4.32E-04	4.32E-04	Decrease Risk
POLD1	Latin Americans	5.23E-03	*6.52E-05*	0.29	0.04	Decrease Risk
PRTG	All Populations	*3.48E-05*	*3.48E-05*	0.10	0.10	Increase Risk
RAD23A	Latin Americans	*8.23E-05*	0.04	5.07E-04	0.02	Decrease Risk
RNU5D	Asians	0.05	*1.08E-05*	0.14	*2.74E-05*	Decrease Risk
THOC6	Asians	*5.79E-05*	*5.79E-05*	1.11E-03	1.11E-03	Decrease Risk
YEATS2	Latin Americans	0.99	2.47E-03	0.02	*6.00E-05*	Unknown

NV: no variants; bold: $p < 3.4 \times 10^{-6}$; italics: $p < 1.0 \times 10^{-4}$

* Driven by single SNP (MAF = 3.7%)