



Published in final edited form as:

*Clin Genet.* 2013 November ; 84(5): 496–500. doi:10.1111/cge.12018.

## Evaluating rare coding variants as contributing causes to nonsyndromic cleft lip and palate

Elizabeth J. Leslie and Jeffrey C. Murray\*

Department of Pediatrics, University of Iowa, Iowa City, IA 52242

### Abstract

Rare coding variants are a current focus in studies of complex disease. Previously, at least 68 rare coding variants were reported from candidate gene sequencing studies in nonsyndromic cleft lip and palate (NSCL/P), a common birth defect. Advances in sequencing technology have now resulted in thousands of sequenced exomes, providing a large resource for comparative genetic studies. We collated rare coding variants reported to contribute to NSCL/P and compared them to variants identified from control exome databases to determine if some might be rare but benign variants. 71% of the variants described as etiologic for NSCL/P were not present in the exome data, suggesting that many likely contribute to disease. Our results strongly support a role for rare variants previously reported in the majority of NSCL/P candidate genes but diminish support for variants in others. However, because clefting is a complex trait it is not possible to be definitive about the role of any particular variant for its risk for NSCL/P.

### Keywords

cleft lip; cleft palate; candidate gene; sequencing

### Introduction

The genome wide association study (GWAS), is a popular tool for identifying common alleles influencing susceptibility to complex diseases and traits. While hundreds of GWAS have been performed they collectively account for a small percentage of disease risk. As a result, there has been an increasing emphasis on the role of rare variants in response to the “missing heritability” often seen with the GWAS approach (1). Prior to GWAS, candidate gene studies often included sequencing coding regions to identify mutations which, when absent in a few hundred controls, were suggested to be etiologic. Thousands of exomes have now been sequenced across a range of phenotypes, representing a catalog of variation that can be used as controls for disease studies (2, 3).

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) affects 1/1000 individuals worldwide (4). Linkage, candidate gene, and GWAS have been performed in search of genetic risk factors for NSCL/P. Here we review the mutations reported in 25 candidate genes and compare them to variants found in 6500 exomes publicly available through the 1000 Genomes Project (1kGP) and the NHLBI Exome Sequencing Project (ESP6500) to provide evidence supporting or diminishing an etiologic role for these variants in NSCL/P. This effort parallels earlier comparisons of loci identified from candidate gene association studies where many of those have failed to replicate in subsequent GWAS (5). Reasons for

\*Corresponding author: Jeffrey C. Murray, MD, Address: 2182 Medical Laboratories, Department of Pediatrics, University of Iowa, Iowa City, Iowa 52242, Phone: 1-319-335-6897, Fax: 1-319-335-6970; jeff-murray@uiowa.edu.

The authors have no conflicts to disclose.

their failure to replicate association data include poorly matched controls, false positive results, failure to account for multiple comparison testing and “winner’s curse” where the results are positive but the overall genetic effect is overestimated causing replication attempts to use inadequate sample sizes. Some of these same phenomena could affect candidate gene sequencing studies and we sought to determine if one element of false positives—inadequate numbers of controls examined—might distinguish rare normal variants from those of etiologic significance.

## Materials and Methods

We broadly queried the PubMed database with terms related to NSCL/P including “cleft lip and palate” and “nonsyndromic cleft”. From papers (written in English) reporting candidate gene sequence we manually curated a list of all missense, nonsense, and splicing variants described as “mutations” or “rare variants” contributing to the etiology of NSCL/P. We identified 67 variants reported in 25 genes (Supplementary Table 1). Supplementary Table 2 contains the mutation positions for genes in which the RefSeq transcripts have been updated since the original publications. We compared these variants with those obtained from the 1000 Genomes Project (1,091 individuals, March 2012 data release) and the NHLBI Exome Sequencing Project (6,503 individuals drawn from a range of studies investigating cardiovascular disease). For the purposes of this paper, we will use the term “control” to describe these variants. Variants from the 1000 Genomes Project were annotated using the SeattleSeq SNP annotation software (Build 134, <http://snp.gs.washington.edu/SeattleSeqAnnotation134/>). We used Polyphen2 to predict damaging effects of missense mutations. Truncation and splice mutations were considered probably damaging.

## Results

Previous resequencing of NSCL/P candidate genes identified 43 missense, nonsense, or splicing variants in 19 genes that were not present in the controls sequenced in those studies (Table 1). An additional 24 variants in 12 genes were reported in NSCL/P cases and matched population controls or samples of the CEPH Human Diversity Panel (Supplemental Table 3). While some of these variants have been associated previously with NSCL/P, such as P153Q in *MSX1* (6) and W185X in *PVRL1* (7), we chose to focus on the 43 variants unique to NSCL/P cases in the present study. Of the 43 missense, nonsense, or splicing variants, 12 (30%) were present in the 1kG or ESP6500 datasets, suggesting that they may be rare, but benign variants. However, the remaining 31 variants were not found in over 12,000 control chromosomes and may be etiologic variants contributing to NSCL/P.

We compared the Polyphen2 predictions between rare variants found exclusively in NSCL/P cases, variants found in both datasets, and variants found only in the 1kGP and ESP6500 datasets (minor allele frequency < 1%) (Table 2). There was a higher frequency of damaging variants in the NSCL/P group (67.7%) than 1kGP/ESP6500 group (54.5%), however this difference was not statistically significant ( $p=0.14$ , Chi-square). Because not all of the sequencing variants identified in the NSCL/P resequencing studies have been reported, we could not directly compare the overall frequency of rare variants in cases to the frequency of variants in controls. These data demonstrate that Polyphen2 predictions are not sufficient to distinguish between these potentially pathogenic variants and the presumably benign variants of 1kGP and ESP6500.

## Discussion

Prior to the GWAS era, candidate genes investigations drove many studies searching for genetic etiology of NSCL/P. Genes sequenced in this approach included *BMP4*, *FGF10*,

*FGF8, FGFR1, FGFR2, FGFR3, FOXE1, GLI2, JAG2, LHX8, MSX1, MSX2, NUDT6, PAX9, PTCH1, PVR, PVRL1, PVRL2, RYK, SATB2, SKI, SPRY2, TBX10, TGFB3,* and *TP63* (4, 8). *IRF6*, the causal gene in Van der Woude syndrome (VWS) (9), has also been studied by sequencing for its role in NSCL/P, but given the phenotypic overlap between VWS and NSCL/P we have not included it in this analysis. Overall, absence of 70% of variants from current exome databases with over 7500 available controls provides support for a role of rare variants in NSCL/P in many of the published candidate gene studies. These results strongly support a role for rare variants in *MSX1*, and members of the FGF signaling pathway. In addition, neither of the two *de novo* mutations (D73H in *FGF8* and R352G in *TP63*) (10, 11) were present in the exome databases. Both mutations occur at highly conserved residues and were not present in 1000 population matched controls. With the addition of another 7000 controls, there is a statistically significant association of these mutations with NSCL/P ( $p=0.03$ ). However, assuming that each gene has 1 mutation (after sequencing 200 cases), approximately 20,000 control exomes would be needed for the variants in these candidate genes to be significant as a group.

Although most rare variants are functionally deleterious (12), the variants found in complex diseases are less penetrant than their Mendelian counterparts. The majority of the rare variants reported in individuals with NSCL/P are inherited from unaffected parents and/or are found in unaffected siblings. Because of this, it is thought that the development of NSCL/P is ultimately the result of the combined action of additional genetic, environmental or stochastic factors. Thus the missense mutations found in controls as very rare, but possibly benign variants, may be relevant risk factors that require additional factors to manifest an effect. Testing these variants in model systems may provide supportive data for their contributory role.

Comparisons of mutations causing Mendelian disorders with public databases is intuitive because Mendelian disorders are typically rare, making it less likely that these disorders are present in the individuals sequenced in the 1kGP or ESP6500. This assumption is weaker for comparisons with less rare disorders such as NSCL/P. There is no phenotype data attached to the sequencing information from 1kGP or ESP6500, however with a frequency of 1/1000 the presence of cases of NSCL/P is likely to be minimal. One exception is for the mutations in *BMP4* associated with subclinical defects of the *orbicularis oris* muscle, which are present in 3-11% of controls (13).

Although the 1kGP has dedicated tremendous effort to sequencing diverse, worldwide populations, many of the populations commonly studied for NSCL/P are not well represented. The prevalence of clefting varies by ancestry, most commonly affecting those of Asian or Amerindian descent (1/500) and less commonly those of African descent (1/2500) (4). Many of the reported variants were found in individuals from the Philippines, Chile, Uruguay, Mongolia, and Thailand, none of which are currently represented in the 1kGP. While we cannot exclude the possibility that some of these variants are rare, population specific polymorphisms, it is, however, encouraging that many of the mutations identified in the populations were not present in matched controls or in the populations available from the 1kGP or ESP6500. Even with 7500 exomes, we may still be missing rare variants of very low frequency. Recent modeling work (14, 15) suggest that the impact of rare variants on heritability may be greater than previously thought but that the frequencies, as a result of recent human population growth, may be below the thresholds detectable even by the 7500 controls used here.

Although many of the mutations reported in *GLI2*, *JAG2*, and *SPRY2* were also found in the exome data, diminishing a role for these rare variants in NSCL/P, these genes remain candidates through supporting biological or statistical data (8). Ultimately, the capacity of

large control data sets will provide supportive data for the role of specific mutations in either contributing to common disease or being of no or little impact. As whole exome and whole genome sequencing will continue to add to control data it will be necessary to reevaluate the role of discovered mutations for the role they play in NSCL/P and other common disorders.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank Chris Lopez for assistance with the literature review and thank the many laboratories that have contributed to the studies reported here and the many families who have participated in research protocols to make these studies possible. Our own candidate gene work has been especially enabled by collaborations with Kaare Christensen and Mary Marazita. The authors would like to thank the NHLBI GO Exome Sequencing Project and its ongoing studies which produced and provided variant calls for comparison: the Lung GO Sequencing Project [HL-102923], the WHI Sequencing Project [HL-102924], the Broad GO Sequencing Project [HL-102925], the Seattle GO Sequencing Project [HL-102926] and the Heart GO Sequencing Project [HL-103010]. This work was supported by NIH grants R37-DE008559 and U01-DE020057.

## References

1. Maher B. Personal genomes: The case of the missing heritability. *Nature*. 2008; 456:18–21. [PubMed: 18987709]
2. Exome Variant Server. Vol. 2011. Seattle, WA: NHBLI Exome Sequencing Project (ESP);
3. Consortium TGP. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–1073. [PubMed: 20981092]
4. Dixon MJ, Marazita ML, Beatty TH, et al. Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet*. 2011; 12:167–178. [PubMed: 21331089]
5. Hirschhorn JN, Lohmueller K, Byrne E, et al. A comprehensive review of genetic association studies. *Genet Med*. 2002; 4:45–61. [PubMed: 11882781]
6. Suzuki Y, Jezewski PA, Machida J, et al. In a Vietnamese population, MSX1 variants contribute to cleft lip and palate. *Genet Med*. 2004; 6:117–125. [PubMed: 15354328]
7. Sozen MA, Suzuki K, Tolarova MM, et al. Mutation of PVRL1 is associated with sporadic, non-syndromic cleft lip/palate in northern Venezuela. *Nat Genet*. 2001; 29:141–142. [PubMed: 11559849]
8. Jugessur A, Murray JC. Orofacial clefting: recent insights into a complex trait. *Curr Opin Genet Dev*. 2005; 15:270–278. [PubMed: 15917202]
9. Kondo S, Schutte BC, Richardson RJ, et al. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nat Genet*. 2002; 32:285–289. [PubMed: 12219090]
10. Leoyklang P, Siriwan P, Shotelersuk V. A mutation of the p63 gene in non-syndromic cleft lip. *J Med Genet*. 2006; 43:e28. [PubMed: 16740912]
11. Riley BM, Mansilla MA, Ma J, et al. Impaired FGF signaling contributes to cleft lip and palate. *Proc Natl Acad Sci U S A*. 2007; 104:4512–4517. [PubMed: 17360555]
12. Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet*. 2007; 80:727–739. [PubMed: 17357078]
13. Neiswanger K, Weinberg SM, Rogers CR, Brandon CA, Cooper ME, Bardi KM, Deleyiannis FWB, Resick JR, Bowen A, Mooney MP, de Salamanca JE, Gonzalez B, Maher BS, Martin RA, Marazita ML. Orbicularis Oris Muscle Defects as an Expanded Phenotypic Feature in Nonsyndromic Cleft Lip with or without Cleft Palate. *Am J Med Genet*. 2007; 143A:1143–1149. [PubMed: 17497721]
14. Keinan A, Clark AG. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*. 2012; 336:740–743. [PubMed: 22582263]

15. Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012; 337:64–69. [PubMed: 22604720]
16. Suzuki S, Marazita ML, Cooper ME, et al. Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. *Am J Hum Genet*. 2009; 84:406–411. [PubMed: 19249007]
17. Vieira AR, Avila JR, Daack-Hirsch S, et al. Medical sequencing of candidate genes for nonsyndromic cleft lip and palate. *PLoS Genet*. 2005; 1:e64. [PubMed: 16327884]
18. Salahshourifar I, Halim AS, Wan Sulaiman WA, et al. Contribution of MSX1 variants to the risk of non-syndromic cleft lip and palate in a Malay population. *J Hum Genet*. 2011; 56:755–758. [PubMed: 21866112]
19. Jezewski PA, Vieira AR, Nishimura C, et al. Complete sequencing shows a role for MSX1 in non-syndromic cleft lip and palate. *J Med Genet*. 2003; 40:399–407. [PubMed: 12807959]
20. Tongkobpetch S, Siriwan P, Shotelersuk V. MSX1 mutations contribute to nonsyndromic cleft lip in a Thai population. *J Hum Genet*. 2006; 51:671–676. [PubMed: 16868654]
21. Ichikawa E, Watanabe A, Nakano Y, et al. PAX9 and TGFB3 are linked to susceptibility to nonsyndromic cleft lip with or without cleft palate in the Japanese: population-based and family-based candidate gene analyses. *J Hum Genet*. 2006; 51:38–46. [PubMed: 16247549]
22. Mansilla MA, Cooper ME, Goldstein T, et al. Contributions of PTCH gene variants to isolated cleft lip and palate. *Cleft Palate Craniofacial Journal*. 2006; 43:21–29. [PubMed: 16405370]
23. Sozen MA, Hecht JT, Spritz RA. Mutation analysis of the PVRL1 gene in caucasians with nonsyndromic cleft lip/palate. *Genet Test Mol Biomarkers*. 2009; 13:617–621. [PubMed: 19715471]
24. Avila JR, Jezewski PA, Vieira AR, et al. PVRL1 variants contribute to non-syndromic cleft lip and palate in multiple populations. *Am J Med Genet A*. 2006; 140:2562–2570. [PubMed: 17089422]
25. Scapoli L, Palmieri A, Martinelli M, et al. Study of the PVRL1 gene in Italian nonsyndromic cleft lip patients with or without cleft palate. *Ann Hum Genet*. 2006; 70:410–413. [PubMed: 16674562]
26. Warrington A, Vieira AR, Christensen K, et al. Genetic evidence for the role of loci at 19q13 in cleft lip and palate. *J Med Genet*. 2006; 43:e26. [PubMed: 16740910]
27. Watanabe A, Akita S, Tin NT, et al. A mutation in RYK is a genetic factor for nonsyndromic cleft lip and palate. *Cleft Palate Craniofacial Journal*. 2006; 43:310–316. [PubMed: 16681403]
28. Lidral AC, Romitti PA, Basart AM, et al. Association of MSX1 and TGFB3 with nonsyndromic clefting in humans. *Am J Hum Genet*. 1998; 63:557–568. [PubMed: 9683588]

**Table 1**  
**Rare missense, nonsense, and splicing variants reported in NSCL/P**

Gene	Mutation	Polyphen 2 Prediction	Population	Cleft Type	NSCL/P cases	ESP6500 counts (N=13006)	1kGP counts (N=2184)	Ref.
<i>BMP4</i>	S91C	probably damaging	Spain	OO	1	2	0	(16)
<i>BMP4</i>	T102A	benign	Philippines Mongolia	CLP	2	0	2	(16)
<i>BMP4</i>	R162Q	probably damaging	Mongolia	CLP	1	0	0	(16)
<i>BMP4</i>	G168A	benign	Philippines	CLP	1	0	0	(16)
<i>BMP4</i>	R287H	benign	USA (PA)	OO	1	18	2	(16)
<i>BMP4</i>	R198×	truncation	Mongolia	CLP	1	0	0	(16)
<i>BMP4</i>	A346V	probably damaging	USA (WA)	MCL+BU	1	0	0	(16)
<i>FGF8</i>	<b>D73H</b>	<b>possibly damaging</b>	<b>USA (IA)</b>	<b>CLP, CP</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>(11)</b>
<i>FGFR1</i>	M369I	benign	Philippines	CP, CLP	2	0	0	(11)
<i>FGFR1</i>	E467K	possibly damaging	Philippines	CLP	2	0	0	(11)
<i>FGFR1</i>	R609×	truncation	USA (IA)	CLP	1	0	0	(11)
<i>FGFR2</i>	R84S	probably damaging	USA (IA)	CLP	1	0	0	(11)
<i>FGFR2</i>	D138N	possibly damaging	Philippines	CLP	5	0	0	(11)
<i>FOXE1</i>	A207V	probably damaging	Philippines	CLP	1	0	0	(17)
<i>FOXE1</i>	D285V	possibly damaging	USA (IA)	CLP	1	0	0	(17)
<i>GLI2</i>	R374H	probably damaging	USA (IA)	CLP	1	1	0	(17)
<i>GLI2</i>	R754Q	probably damaging	Philippines	CLP	1	4	0	(17)
<i>GLI2</i>	G1106D	benign	USA (IA)	CLP	1	0	0	(17)
<i>GLI2</i>	S1541Y	probably damaging	Philippines	CLP	1	0	2	(17)
<i>JAG2</i>	M559I	benign	USA (IA)	CLP	1	4	0	(17)
<i>JAG2</i>	D657H	probably damaging	Philippines	CLP	1	0	0	(17)
<i>MSX1</i>	M43L	benign	Malaysia	CLP	1	0	0	(18)
<i>MSX1</i>	G97D	benign	Philippines	CP	1	0	0	(19)
<i>MSX1</i>	V120G	benign	Denmark	CL	1	0	0	(19)
<i>MSX1</i>	G122E	benign	Uruguay	CLP	1	0	0	(19)
<i>MSX1</i>	R157S	possibly damaging	Japan	CLP	1	0	19	(19)
<i>MSX1</i>	G273C	possibly damaging	Thailand	CLP	1	0	0	(20)

Gene	Mutation	Polyphen 2 Prediction	Population	Cleft Type	NSCL/P cases	ESP6500 counts (N=13006)	1kGP counts (N=2184)	Ref.
<i>MSX1</i>	P284S	probably damaging	Thailand	CLP	1	0	0	(20)
<i>MSX2</i>	S63C	possibly damaging	USA (IA)	CLP	1	0	0	(17)
<i>NUDT6</i>	K172N	probably damaging	USA (IA)	CLP	1	0	0	(11)
<i>PAX9</i>	S214G	benign	Japan	CL/P	1	0	0	(21)
<i>PTCHI</i>	P295S	probably damaging	Philippines	CL/P	1	0	0	(22)
<i>PVRL1</i>	V89M	probably damaging	N. America	CL/P	2	0	0	(23)
<i>PVRL1</i>	S112T	benign	Philippines	CLP	1	0	0	(24)
<i>PVRL1</i>	T131A	benign	Philippines	CLP	1	0	1	(24)
<i>PVRL1</i>	R212H	probably damaging	Italy	CL/P	1	11	6	(25)
<i>PVRL1</i>	IVS4+1	splicing	Australia	CL/P	1			(23)
<i>PVRL2</i>	E41K	possibly damaging	USA (IA)	CL/P	1	2	0	(26)
<i>RYK</i>	Y456C	probably damaging	Vietnam	CLP	1	0	0	(27)
<i>SATB2</i>	T190A	possibly damaging	Philippines	CLP	1	0	0	(17)
<i>SPRY2</i>	K68N	benign	USA (IA)	CLP	1	1	0	(17)
<i>TGFB3</i>	K130R	benign	unspecified	SMCP	1	0	0	(28)
<i>TP63</i>	S90L	possibly damaging	Thailand	CL	1	0	1	(10)
<b><i>TP63</i></b>	<b>R313G</b>	<b>probably damaging</b>	<b>Thailand</b>	<b>CL</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>(10)</b>

Bolded variants arose *de novo*

Abbreviations: CL- cleft lip; CLP- cleft lip and palate; CL/P- cleft lip with or without cleft palate; SMCP- submucous cleft palate; MCL- microform cleft lip; BU- bifid uvula

**Table 2**  
**Polyphen2 predictions of rare variants (MAF<1%) in NSCL/P candidate genes**

	NSCL/P Only		NSCL/P and 1kGP/ESP6500		1kGP/ESP6500 Only	
	No.	%	No.	%	No.	%
Probably Damaging	14	45.2%	5	41.7%	318	39.7%
Possibly Damaging	7	22.6%	2	16.7%	119	14.8%
Benign	10	32.3%	5	41.7%	365	45.5%
Total	31		12		802	