

# Comparative analysis of *IRF6* variants in families with Van der Woude syndrome and popliteal pterygium syndrome using public whole-exome databases

Elizabeth J. Leslie, BA<sup>1</sup>, Jennifer Standley, BS<sup>1</sup>, John Compton, PhD<sup>2</sup>, Sherri Bale, PhD<sup>2</sup>, Brian C. Schutte, PhD<sup>3,4</sup> and Jeffrey C. Murray, MD<sup>1</sup>

**Purpose:** Mutations in the transcription factor *IRF6* cause allelic autosomal dominant clefting syndromes, Van der Woude syndrome, and popliteal pterygium syndrome. We compared the distribution of *IRF6* coding and splice-site mutations from 549 families with Van der Woude syndrome or popliteal pterygium syndrome with that of variants from the 1000 Genomes and National Heart, Lung, and Blood Institute Exome Sequencing Projects.

**Methods:** We compiled all published pathogenic *IRF6* mutations and performed direct sequencing of *IRF6* in families with Van der Woude syndrome or popliteal pterygium syndrome.

**Results:** Although mutations causing Van der Woude syndrome or popliteal pterygium syndrome were nonrandomly distributed with significantly increased frequencies in the DNA-binding domain ( $P = 0.0001$ ), variants found in controls were rare and evenly distributed in *IRF6*. Of 194 different missense or nonsense variants

described as potentially pathogenic, we identified only two in more than 6,000 controls. PolyPhen and SIFT (sorting intolerant from tolerant) reported 5.9% of missense mutations in patients as benign, suggesting that use of current *in silico* prediction models to determine function can have significant false negatives.

**Conclusion:** Mutation of *IRF6* occurs infrequently in controls, suggesting that for *IRF6* there is a high probability that disruption of the coding sequence, particularly the DNA-binding domain, will result in syndromic features. Prior associations of coding sequence variants in *IRF6* with clefting syndromes have had few false positives.

*Genet Med* 2013;15(5):338–344

**Key Words:** cleft; exome; mutation; popliteal pterygium; Van der Woude

## INTRODUCTION

The advent of massively parallel sequencing technologies has resulted in thousands of exomes and sequenced genomes, creating a catalog of genetic variation for comparison purposes that is likely to grow substantially in the coming years. The 1000 Genomes Project (1kGP) has published more than 1,000 complete human genomes representing healthy individuals.<sup>1</sup> The National Heart, Lung, and Blood Institute (NHLBI) Grand Opportunity (GO) Exome Sequencing Project (ESP5400) sequenced whole exomes of 5,400 individuals ascertained for phenotypes related to heart, lung, and blood disorders.<sup>2</sup> These genomes can act as controls for mutation significance in rare Mendelian disorders in which previous studies were often limited to sequencing a few hundred controls to determine if a detected variant is rare and normal or unique to a syndrome.

Orofacial clefts are common birth defects affecting ~1 in 1,000 individuals worldwide. Although most orofacial clefts are nonsyndromic, 30% are designated as syndromic due to the presence of additional physical or cognitive abnormalities. Online Mendelian Inheritance in Man (<http://www.ncbi.nlm.nih.gov/omim>) identifies more than 500 associations with clefting, including more than 100 Mendelian syndromes and/

or chromosomal deletion/duplication associations, as well as many sporadic or family-specific associations. Of these, Van der Woude syndrome (VWS, OMIM 119300) is among the most common, accounting for ~2% of all clefts.<sup>3,4</sup> VWS is broadly characterized by clefts of the lip and/or palate and congenital paramedian lower lip pits.<sup>5</sup> There is considerable variability in the phenotypic expression of VWS, which can range from lip pits alone to bilateral cleft lip and palate.<sup>6</sup> VWS is one of two dominant allelic disorders caused by heterozygous mutations in *IRF6*.<sup>7</sup> The popliteal pterygium syndrome (PPS, OMIM 119500) is characterized by the clinical features of VWS (clefts of the lip and/or palate and lip pits), with additional features that include webbing of skin behind the knees, genital anomalies, syndactyly, oral adhesions, and other anomalies.<sup>8</sup> The *IRF6* gene is one of a family of nine *IRF* genes that code for transcription factors that share a highly conserved helix-turn-helix DNA-binding domain (DBD) and a less conserved protein-binding domain.<sup>7</sup> Since its identification as the gene mutated in VWS and PPS, hundreds of mutations in *IRF6* have been reported.<sup>7,9–20</sup>

Given the rarity of VWS and PPS (1/35,000<sup>21</sup> and 1/300,000,<sup>22</sup> respectively), rates of mutations in functional elements of *IRF6* in any individuals sequenced as normal controls or as part of

<sup>1</sup>Department of Pediatrics, University of Iowa, Iowa City, Iowa, USA; <sup>2</sup>GeneDx, Gaithersburg, Maryland, USA; <sup>3</sup>Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, USA; <sup>4</sup>Department of Pediatrics and Human Development, Michigan State University, East Lansing, Michigan, USA. Correspondence: Jeffrey C. Murray ([jeff-murray@uiowa.edu](mailto:jeff-murray@uiowa.edu))

Submitted 2 July 2012; accepted 21 September 2012; advance online publication 15 November 2012. doi:10.1038/gim.2012.141

unrelated disease-based cohorts would be expected to be low. Therefore, the resources of the 1kGP and the ESP5400 represent a control cohort that is larger than any previously available. Approximately 300 pathogenic variants in *IRF6* in individuals with VWS or PPS have been identified. To determine if any of these variants could be extremely rare but normal variants in the general population, we compared the list of previously reported *IRF6* variants to the 1kGP and the ESP5400 databases.

The work by de Lima et al.<sup>9</sup> described the distribution of *IRF6* mutations with the goal of identifying the exons most likely to carry mutations. This was clinically useful for prioritizing mutation discovery efforts and suggested broad genotype–phenotype relationships, but categorizing the mutation distribution by exon does little to refine the regions of *IRF6* important for function. Since then, additional mutations have been reported, and we were able to carefully characterize the distribution of *IRF6* variants. This allowed us to identify the residues whose disruption is likely to be damaging (as the etiologic cause of VWS or PPS) and to further define the domains of the protein most critical for *IRF6* function in craniofacial development. This is biologically significant because it allows us to prioritize mutations for functional studies and offers insight into structure–function relationships for *IRF6* and other members of this highly conserved family of transcription factors. In addition, examining the spectrum of *IRF6* variation present in VWS or PPS and the whole-exome databases provides a benchmark for clinically interpreting *IRF6* variants from future whole-exome or whole-genome sequencing projects.

## MATERIALS AND METHODS

### Compilation of mutation data

To identify published *IRF6* mutations, we performed a PubMed search using the following terms: “*IRF6*,” “Van der Woude syndrome,” “VWS,” “popliteal pterygium syndrome,” and “PPS.” Additional mutations were obtained from the clinical sequencing database at GeneDx (Gaithersburg, MD) or reported from research sequencing in our laboratory. Control variants were obtained from the 1kGP (1,091 individuals, February 2012 data release) and the NHLBI ESP (5,379 individuals, ESP5400). Variants from the 1kGP were annotated using the SeattleSeq SNP annotation software (Build 134, <http://snp.gs.washington.edu/SeattleSeqAnnotation134/>). Several mutations have been previously reported to cause VWS in exons 1 and 2, which make up the 5′ untranslated region. These mutations create an alternate upstream start codon and are predicted to create truncated *IRF6* proteins. We categorized these with missense mutations at position M1 as mutations that “alter the start codon.” No sequencing data are available for exons 1 and 2 from whole-exome sequencing due to the limitations of the exome capture technique, so we restricted analysis to mutations in the seven protein-coding exons of *IRF6*.

### Sequencing

Twenty-three previously unreported cases of VWS or PPS had exons 1–8 and the protein-coding part of exon 9 amplified

using primers previously published by de Lima et al. PCR products were sent for sequencing using an ABI 3730XL (Functional Biosciences, Madison, WI). Chromatograms were transferred to a Unix workstation, base called with PHRED (v.0.961028), assembled with PHRAP (v. 0.960731), scanned by POLYPHRED (v. 0.970312), and viewed with the CONSED program (v. 4.0).

### Bioinformatics

Polyphen2<sup>23</sup> and sorting intolerant from tolerant (SIFT)<sup>24</sup> were used to predict the damaging effects of missense mutations. Sequences of *IRF6* orthologs from 17 species were obtained from the Ensembl database (human (NP\_006138.1), chimpanzee (ENSPTRP00000003274), gorilla (ENSGGOP00000006076), macaque (ENSMMUP0000029056), bushbaby (ENSOGAP00000009580), mouse (NP\_058547.2), rat (NP\_001102329), guinea pig (ENSCPOP0000015286), rabbit (ENSOCUP00000003494), cow (ENSBTAP00000054388), cat (ENSFCAP00000014042), dog (ENSCAFP00000017624), elephant (ENSLAFP00000001156), pig (ENSSSCP00000016552), chicken (ENSGALP000000039479), lizard (ENSACAP00000005847), frog (ENSXETP000000040424), and zebrafish (ENSDARP000000061534)). Sequences were aligned using ClustalW and viewed in Jalview (version 2.7),<sup>25</sup> which also provided a numerical conservation score based on the chemical properties of the amino acids in the alignment.

### Statistical analysis

Frequencies were calculated using the number of variants per 100 bp to account for differences in exon size (ranging from 72 to 393 bp) for only exons 2–9, which encode the *IRF6* protein. To visualize the distribution of variants, we performed a sliding window analysis using a 33–amino acid (99 bp) sliding window with a step size of 1 amino acid (3 bp). The expected number of variants per 100 bp was calculated from the total number of variants evenly distributed along the length of the *IRF6* cDNA (1,404 bp). Normalized variant counts per exon were compared with expected counts using a 1 degree of freedom  $\chi^2$  test. To account for multiple comparisons, we established a Bonferroni significance threshold of  $P = 0.007$  (i.e.,  $0.05/7$ ). The  $2 \times 7$  tables showing the distribution of variants in the exons of *IRF6* were analyzed using  $\chi^2$  or Fisher’s exact test. We also compared the distribution of mutations in the known *IRF6* protein domains (DNA binding, protein binding, and other) in the same manner. Wilcoxon rank-sum test was performed to determine the difference in conservation scores between missense variants cases and controls using STATA (version 12.0; StataCorp, College Station, TX).

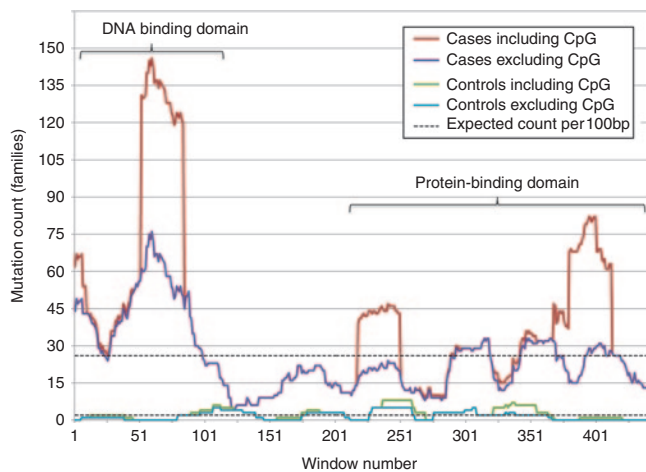
## RESULTS

### Prevalence of *IRF6* mutations

There were 295 distinct mutations identified in 549 families with VWS or PPS (Supplementary Table S1 online). Missense mutations were the most common (51.7%), whereas a large portion (40.5%) of the remaining variants resulted in a truncated

**Table 1** Distribution of *IRF6* mutations in Van der Woude syndrome/popliteal pterygium syndrome by type of mutation

	Number	%	No. of families	%
Missense				
Other codons	139	47.3%	212	38.6%
CpG	13	4.4%	119	21.7%
Nonsense				
Other codons	40	13.6%	61	11.1%
CpG	2	0.7%	44	8.0%
Frameshift insertion	26	8.8%	26	4.7%
Frameshift deletion	45	15.3%	48	8.7%
Splicing	16	5.4%	22	4.0%
In-frame insertion	1	0.3%	1	0.2%
In-frame deletion	4	1.4%	4	0.7%
In-frame indel	2	0.7%	2	0.4%
New start codon	6	2.0%	10	1.8%
Total	294		549	

**Figure 1** Sliding-window analysis of *IRF6* variants. A window size of 99 bp was used with a step size of 3 bp. Variant counts included or excluded the 15 variants at CpG sites. Red and purple represent counts in Van der Woude syndrome/popliteal pterygium syndrome families; blue and green represent counts in 1kG/ESP5400 controls. Dashed lines indicate the expected number of variants per 100 bp if the total number of variants were distributed evenly across *IRF6* (expected cases = 25; expected controls = 2).

*IRF6* protein (nonsense, frameshift, and altered start codons) (Table 1). For the families in which the syndrome was specified, we compared the types of mutations causing VWS with the types causing PPS (Supplementary Table S2 online). Every category of mutation was represented among VWS families, whereas PPS mutations were limited to missense, nonsense, and splicing mutations.

CpG dinucleotides are common mutation hotspots due to methylation and spontaneous deamination of cytosine to thymine. There were 24 CpG dinucleotides in the protein-coding

**Table 2** Distribution of variants in protein-coding exons of *IRF6*

<i>IRF6</i> exon	Exon length (bp)	VWS/PPS <sup>a</sup>		1kG/ESP5400	
		No.	Per 100 bp	No.	Per 100 bp
3	177	61	34	1	1
4	205	108	53	2	1
5	129	6	5	4	3
6	159	28	18	3	2
7	393	73	19	10	3
8	119	32	27	2	2
9 <sup>b</sup>	222	47	21	0	0
All coding <sup>c</sup>	1,404	355	25	22	2
CpG	48	163	340	11	23

1kG, 1000 Genomes Project; ESP, Exome Sequencing Project; PPS, popliteal pterygium syndrome; VWS, Van der Woude syndrome.

<sup>a</sup>Excludes splicing mutations in introns, and mutations creating new start codons. <sup>b</sup>Length of protein coding region. <sup>c</sup>All *IRF6* coding regions from exons 3 to 9, excluding CpG hotspots.

exons of *IRF6*, and C→T or G→A transitions at these dinucleotides could create 35 different missense or nonsense changes (Supplementary Table S3 online). In VWS and PPS cases, 9 of the 24 CpG dinucleotides had C→T or G→A transitions, totaling 15 different mutations. Although these mutations are a small fraction (5.3%) of the total number reported in *IRF6*, they are responsible for VWS or PPS in ~30% of families ( $P = 1.6 \times 10^{-5}$ ). Missense mutations at CpG dinucleotides were responsible for half of PPS families (Supplementary Table S2 online), and the majority of these were either R84C or R84H.

We analyzed the distribution of mutations in the seven protein-coding exons of *IRF6*, normalized to 100 bp to account for differences in exon size. The average frequency was 25 mutations per 100 bp. The frequency was significantly higher in exon 4 ( $P = 5.07 \times 10^{-8}$ ) and lower in exon 5 ( $P = 4.07 \times 10^{-5}$ ) (Figure 1, Table 2). Overall, the distribution of mutations in the protein coding exons was nonrandom ( $P = 1.37 \times 10^{-4}$ ). We observed a similar pattern when accounting for independent mutations (i.e., the same variant in two families counted once) (Supplementary Tables S4–S6 online).

Approximately 90% of the mutations causing VWS or PPS are missense or truncation mutations, and the remaining 10% include those predicted to create new start codons or affect splicing. Excluding mutations at CpG dinucleotides, the frequency of missense mutations was significantly increased in exon 4 ( $P = 6.41 \times 10^{-11}$ ) (Table 3). There was also a significant decrease in missense mutations in exon 5 ( $P = 2.27 \times 10^{-4}$ ) and exon 6 ( $P = 6.73 \times 10^{-4}$ ). Overall, missense mutations were non-randomly distributed ( $P = 7.21 \times 10^{-7}$ ). However, truncation mutations were evenly distributed across *IRF6* ( $P = 0.49$ ).

#### Distribution of mutations in *IRF6* domains

*IRF6* contains an N-terminus DBD, encoded in exons 3 and 4, and a C-terminus protein-binding domain, encoded in exons 7–9. The frequency of mutations was increased for the DBD

**Table 3** Distribution of VWS/PPS missense and truncation mutations in protein-coding exons of *IRF6*

<i>IRF6</i> exon	Exon length (bp)	Missense mutations		Truncation mutations	
		No.	Per 100 bp	No.	Per 100 bp
3	177	42	24	22	12
4	205	83	40	23	11
5	129	1	1	5	4
6	159	3	2	25	16
7	393	47	12	25	6
8	119	20	17	12	10
9	222	16	7	27	12
All coding	1,404	212	15	139	10

Excluding intronic splicing and CpG mutations.  
PPS, popliteal pterygium syndrome; VWS, Van der Woude syndrome.

**Table 4** Distribution of variants in domains of *IRF6*

<i>IRF6</i> exon	Length (bp)	VWS/PPS <sup>a</sup>		1kG/ESP5400	
		No.	Per 100 bp	No.	Per 100 bp
DNA binding	360	155	43	3	1
Protein binding	504	105	21	7	1
Other	540	95	18	12	2
All coding	1,404	355	25	22	2

1kG, 1000 Genomes Project; ESP, Exome Sequencing Project; PPS, popliteal pterygium syndrome; VWS, Van der Woude syndrome.  
<sup>a</sup>All mutations except intronic splicing and CpG mutations.

(**Table 4**) ( $P = 4.09 \times 10^{-4}$ ) but not in the protein-binding domain ( $P = 0.37$ ). The overall distribution was only marginally nonrandom when considering the number of families with mutations in each domain ( $P = 0.05$ ).

**Prevalence of *IRF6* variants in controls**

We analyzed sequences from 6,470 individuals from the 1kGP and the ESP5400 for variants in *IRF6* (**Supplementary Table S7** online). Although 22 different missense variants were identified, we excluded V274I (rs2235371) from analyses because it is commonly observed in 3% of European and 30% of Asian individuals. The remaining 21 missense variants were distributed among only 33 individuals (0.5% of controls). Seven were C→T or G→A transitions at CpG dinucleotides. The average rate was two variants per 100 bp, and these were evenly distributed throughout *IRF6* (**Table 2**,  $P = 0.91$ ). Overall, the distribution of missense variants was significantly different between cases and controls ( $P = 3.4 \times 10^{-4}$ ).

**Known pathogenic variants in controls**

We compared variants reported in VWS and PPS cases with those from the 1kGP and the ESP5400. There were two variants in common between these data sets. The first, R45Q, was reported by Kayano et al.<sup>26</sup> in a father with lip pits and his daughter, who had cleft lip and lip pits. It was found in a Luhya female from the 1kGP. This variant is highly conserved

and is predicted to be damaging by both Polyphen2 and SIFT. The second variant, D354N, was first reported by Jehee et al.<sup>19</sup> in an individual with cleft palate and her unaffected mother. D354N was also found in two individuals with VWS sequenced at GeneDx. Jehee et al.<sup>19</sup> showed that this variant reduces the GeneSplicer-predicted splicing score from 4.40 to 1.75. The aspartic acid residue is conserved among primates and placental mammals but is predicted to be benign by both Polyphen2 and SIFT. In addition, this variant was identified in three European Americans in the ESP5400 cohort.

**Bioinformatic predictions for *IRF6* missense variants**

Many computational tools have been developed to discriminate pathogenic and benign variants from sequence data.<sup>27</sup> SIFT<sup>24</sup> and Polyphen2<sup>23</sup> are two such tools. We compared a set of missense variants identified from VWS and PPS cases with whole-exome or genome-sequencing databases using both programs. Although Polyphen2 predicted that only 5.9% of the missense variants in VWS and PPS cases are benign, SIFT was more conservative and predicted a higher percentage (42.3%) to be benign (**Supplementary Figure S1** online). SIFT similarly predicted more of the 1kGP and ESP5400 variants to be benign (95.2% vs. 76.2% by Polyphen2). Amino acid conservation is another metric for evaluating pathogenicity of rare variants. Missense variants in VWS and PPS had higher conservation scores (average, 10.64) than missense variants in 1kGP/ESP5400 (average, 7.31); this was statistically significant ( $P = 8.14 \times 10^{-16}$ ).

**DISCUSSION**

To date, 237 different *IRF6* mutations have been published, not including numerous deletions of *IRF6*. Here, we report an additional 63 pathogenic variants. In 2009, de Lima et al.<sup>9</sup> described the distribution of *IRF6* mutations as nonrandom, with more mutations in exons 3, 4, 7, and 9. Not surprisingly, de Lima et al.<sup>9</sup> also reported more missense mutations in the DNA-binding and protein-binding domains, encoded by exons 3–4 and 7–9, respectively. We revisited this result with a larger sample and found that mutations causing VWS and PPS are only overrepresented in exon 4, encoding part of the DBD. There are several differences in our analysis to account for this disparity. First, we normalized frequencies to 100 bp to account for variability in exon size. Exon 7 is the largest at 393 bp, therefore it follows that more opportunities for mutation would result in more variants in this exon. Second, we excluded mutations at hypermutable CpG dinucleotides from the analysis. The increase in mutations in exons 7 and 9 observed by de Lima et al.<sup>9</sup> can be attributed to the hotspots R250X (exon 7), R400W (exon 7), and R412X (exon 9). By excluding these mutations, which can be attributed to a specific, high-frequency mechanism, we can look for patterns in distribution of the remaining mutations to gain insight into amino acid residue function for *IRF6* and the pathogenesis of VWS and PPS.

Previous work by de Lima et al.<sup>9</sup> described five mutational hotspots (R6, R84, R250, R400, and R412) and attributed the

high-mutation frequencies to the CpG dinucleotides in these codons. We systematically analyzed the 24 CpG dinucleotides in the coding exons of *IRF6* and found that not all of these CpG dinucleotides are mutated. From this data set, we cannot determine why some of these “hotspots” are mutated and others are not, or why some mutations occur more frequently. For example, R84C and R84H have been identified in 70 families with VWS or PPS, but V433I has not been reported. Mutations at R84 are highly associated with PPS, therefore some ascertainment bias should be taken into consideration regarding the preponderance of R84 mutations in this data set. Although V433I may not cause a phenotype, V433I was predicted to be damaging and has not been found in 500 families with VWS/PPS or in 6,500 controls. It is possible that V433I causes a particular phenotype that has yet to be observed or described. Similarly, it is also possible that this amino acid is critical to the function of IRF6 and mutation of this residue is lethal. Finally, it could be that not all of the CpG dinucleotides are methylated, and these nucleotides are no more mutable than any other nucleotide, possibly explaining its absence in more than 7,000 samples. Although the methylation status of the *IRF6* promoter has been studied in human squamous cell carcinoma cells,<sup>28</sup> it has not yet been empirically determined within the gene itself or directly from tissues obtained during craniofacial development.

We saw an increased frequency of mutations in the DBD. It has been hypothesized that VWS is caused primarily by loss-of-function mutations and PPS is caused by dominant negative mutations<sup>9,29</sup> because mutations affecting DNA binding (i.e., R84C, R84H) are highly associated with PPS, whereas truncation mutations are more commonly found in VWS. We found a similar result: 67% of the PPS families had missense mutations at R84C or R84H. However, some families with R84C mutations have VWS and some families with nonsense mutations have PPS,<sup>9</sup> suggesting a more complicated mechanism. If missense mutations causing VWS were primarily protein destabilizing, we would expect them to be distributed evenly between the DBD and protein-binding domain. However, we observed a significant increase in mutations in the DBD, highlighting the importance of this domain. Specifically, this increase is centered in exon 4, where the majority of the residues contacting DNA are located. Exclusion of mutations at residues contacting DNA did not eliminate the enrichment of missense mutations in the DBD (data not shown). The remaining missense mutations in the DBD may be protein destabilizing, but may also prevent DNA binding in another way, or have some as-yet-unknown effect on *IRF6* function.

Although our results suggest the DBD should be a focus of further biological investigation, mutation of the protein-binding domain is clearly important for the pathogenesis of VWS or PPS. Even though the exons that encode this domain are not enriched for mutations over what was expected, mutations at hotspots R250, R400, and R412 in these exons account for a significant portion (12%) of the mutations in families with VWS/PPS. Therefore, for PCR amplification-based sequencing

for mutation detection, we recommend continuing the tiered approach proposed by de Lima et al.<sup>9</sup>

Some of the criteria proposed for the classification of pathogenic variants include amino acid conservation, *in silico* prediction (i.e., Polyphen2 and SIFT), and presence or absence in databases such as dbSNP.<sup>30</sup> Polyphen2 and SIFT are two of a host of programs in use for mutation interpretation;<sup>27</sup> these programs, although popular, can have low sensitivity and specificity.<sup>31</sup> Here, we show that for missense variants in *IRF6*, Polyphen2 and SIFT predictions were not in perfect agreement. In some cases, such as the mutation L22P, previously shown to abrogate DNA binding *in vitro*,<sup>29</sup> the predictions contradict the experimental evidence. However, only a handful of mutations in *IRF6* have been functionally tested, therefore it is impossible to know the true sensitivity and specificity of these *in silico* prediction programs in this case. In the current era of whole-exome and whole-genome sequencing, it has become important to be cautious in labeling a genetic variant pathogenic, particularly in novel genes. In the case of VWS or PPS, in which it is clear that *IRF6* mutations play a role, it is still beneficial to use caution when using *in silico* programs. Similarly, although we show a significant difference in amino acid conservation between the variants causing VWS or PPS and those found in controls, conservation alone is not sufficient to determine pathogenicity of a variant.

*IRF6* variants were found in just 0.5% of controls, suggesting that mutation of *IRF6* is uncommon and variants identified in patients are likely to be truly disease causing. Furthermore, when carrier testing and/or prenatal testing are considered, families can have high confidence in the results. However, we identified two variants (R45Q and D354N) previously reported as disease causing in controls. Given the frequency with which we found D354N in controls and its relatively low conservation, it may be a rare polymorphism. Coincidentally, this variant is consistent with a deamination mechanism at this CpG site. However, R45Q is highly conserved and segregated in a family with VWS.<sup>26</sup> Samples from the 1kGP are anonymous and have no associated phenotype data, although they are presumed healthy. VWS is a rare disorder, with an estimated frequency of 1/35,000, therefore it is unlikely that one of these controls has VWS. However, VWS is a syndrome with marked variable expressivity in which 44% of affected individuals only have lower lip pits,<sup>6</sup> which could easily be overlooked without careful examination.

A clear limitation of using the 1kGP and ESP5400 cohorts as controls is the lack of individual phenotypes or family history. As noted above, given the frequency and penetrance of VWS and PPS, it is unlikely that one of these controls would have a mutation and also be unaffected. However, for other disorders, with higher frequency and/or lower penetrance, even more caution must be used in interpreting variants from the 1kGP and ESP5400 cohorts. Another limitation to this study is the racial and ethnic heterogeneity of the VWS and PPS cases and the control cohorts. Mutations causing VWS or PPS have been reported in geographically diverse populations, but previous work found no difference in the distribution of *IRF6* mutations.<sup>9</sup> Although the 1kGP has sampled populations worldwide, the majority of

the controls used in this study come from the ESP5400, which consists of European and African-American populations.

The fact that we identified potentially disease-causing alleles in the samples from the 1kGP and the ESP5400 cohorts is not surprising. Most recently, it was estimated that these individuals carry at least 100 loss-of-function variants.<sup>32</sup> Not all mutations in *IRF6* cause VWS or PPS, *IRF6* mutations are only found in 70% of individuals with VWS/PPS,<sup>9</sup> and there are some cases of non-penetrance.<sup>21</sup> Therefore, it is important to have a large reference panel of normal controls before families make decisions on prenatal diagnosis or nonpenetrant carrier status based on finding an amino acid variant that has not been previously reported for VWS or PPS. It is also possible that there are other phenotypes resulting from mutation of *IRF6*; this is certainly possible given that *Irf6* is expressed in a variety of embryonic and adult tissues, including the placenta, liver, and lung.<sup>7</sup> Large-scale whole-exome sequencing may be able to identify mutations responsible for the remaining 30% of VWS/PPS cases and, when applied clinically, will be able to uncover any additional phenotypes resulting from *IRF6* mutation. Eventually, it will be possible to define amino acid residues that are and are not critical for IRF6 function.

By analyzing the prevalence and distribution of IRF6 variants in individuals with VWS or PPS and controls, we have demonstrated that mutation of IRF6 occurs infrequently. When mutation does occur, particularly in conserved domains, it is likely to result in VWS or PPS. Therefore, we can say that IRF6 does not tolerate a high mutational burden. Further studies of other genes and disorders will be of great importance for interpreting the variants coming from clinical whole-exome or whole-genome sequencing.

## SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

## ACKNOWLEDGMENTS

We thank the families who have participated in this study; Susie McConnell, Erin Brothers-Smith, and Nancy Davin for administrative assistance; Jenni Mancuso and Laura Henkle for technical assistance; and Kelli Ryckman for helpful discussions. We also thank the National Heart, Lung, and Blood Institute Grand Opportunity (GO) Exome Sequencing Project and its ongoing studies that produced and provided variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the Women's Health Initiative 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 National Institutes of Health (NIH) grants (DE08559 and DE020057 to J.C.M., DE013513 to B.C.S.) and the University of Iowa Institute for Clinical and Translational Science. E.J.L. was supported by NIH training grant GM008629.

## DISCLOSURE

S.B. and J.C. are employees of GeneDx. E.J.L., J.S., B.C.S., and J.C.M. declare no conflict of interest.

## REFERENCES

1. Consortium TGP. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–1073.
2. Exome Variant Server. <http://snp.gs.washington.edu/EVS/>. Accessed September, 2011.
3. Cohen MM Jr, Bankier A. Syndrome delineation involving orofacial clefting. *Cleft Palate Craniofac J* 1991;28:119–120.
4. Murray JC, Daack-Hirsch S, Buetow KH, et al. Clinical and epidemiologic studies of cleft lip and palate in the Philippines. *Cleft Palate Craniofac J* 1997;34:7–10.
5. Van Der Woude A. Fistula labii inferioris congenita and its association with cleft lip and palate. *Am J Hum Genet* 1954;6:244–256.
6. Burdick AB, Bixler D, Puckett CL. Genetic analysis in families with van der Woude syndrome. *J Craniofac Genet Dev Biol* 1985;5:181–208.
7. 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.
8. Gorlin RJ, Sedano HO, Cervenka J. Popliteal pterygium syndrome. A syndrome comprising cleft lip-palate, popliteal and intercrural pterygia, digital and genital anomalies. *Pediatrics* 1968;41:503–509.
9. de Lima RL, Hoper SA, Ghassibe M, et al. Prevalence and nonrandom distribution of exonic mutations in interferon regulatory factor 6 in 307 families with Van der Woude syndrome and 37 families with popliteal pterygium syndrome. *Genet Med* 2009;11:241–247.
10. Brosch S, Baur M, Blin N, Reinert S, Pfister M. A novel IRF6 nonsense mutation (Y67X) in a German family with Van der Woude syndrome. *Int J Mol Med* 2007;20:85–89.
11. Rutledge KD, Barger C, Grant JH, Robin NH. IRF6 mutations in mixed isolated familial clefting. *Am J Med Genet A* 2010;152A:3107–3109.
12. Birkeland AC, Larrabee Y, Kent DT, et al. Novel IRF6 mutations in Honduran Van der Woude syndrome patients. *Mol Med Report* 2011;4:237–241.
13. Minones-Suarez L, Mas-Vidal A, Fernandez-Toral J, Llano-Rivas I, Gonzalez-Garcia M. A novel mutation in the IRF6 gene associated with facial asymmetry in a family affected with Van der Woude Syndrome. *Pediatr Dermatol* 2011; epub ahead of print 13 October 2011.
14. Salahshourifar I, Halim AS, Sulaiman WA, Ariffin R, Naili Muhamad Nor N, Zifalil BA. De novo interstitial deletion of 1q32.2-q32.3 including the entire IRF6 gene in a patient with oral cleft and other dysmorphic features. *Cytogenet Genome Res* 2011;134:83–87.
15. Desmyter L, Ghassibe M, Revencu N, et al. IRF6 screening of syndromic and a priori non-syndromic cleft lip and palate patients: identification of a new type of minor VWS sign. *Mol Syndromol* 2010;1:67–74.
16. Scioletti AP, Brancati F, Gatta V, et al. Two novel mutations affecting splicing in the IRF6 gene associated with van der Woude syndrome. *J Craniofac Surg* 2010;21:1654–1656.
17. Matsuzawa N, Kondo S, Shimozato K, et al. Two missense mutations of the IRF6 gene in two Japanese families with popliteal pterygium syndrome. *Am J Med Genet A* 2010;152A:2262–2267.
18. Yeetong P, Mahatumarat C, Siriwan P, Rojvachiranonda N, Suphapeetiporn K, Shotelersuk V. Three novel mutations of the IRF6 gene with one associated with an unusual feature in Van der Woude syndrome. *Am J Med Genet A* 2009;149A:2489–2492.
19. Jehsee FS, Burin BA, Rocha KM, et al. Novel mutations in IRF6 in nonsyndromic cleft lip with or without cleft palate: when should IRF6 mutational screening be done? *Am J Med Genet A* 2009;149A:1319–1322.
20. Malik S, Kakar N, Hasnain S, Ahmad J, Wilcox ER, Naz S. Epidemiology of Van der Woude syndrome from mutational analyses in affected patients from Pakistan. *Clin Genet* 2010;78:247–256.
21. Burdick AB. Genetic epidemiology and control of genetic expression in van der Woude syndrome. *J Craniofac Genet Dev Biol Suppl* 1986;2:99–105.
22. Froster-Iskenius UG. Popliteal pterygium syndrome. *J Med Genet* 1990;27:320–326.
23. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–249.
24. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31:3812–3814.
25. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25:1189–1191.
26. Kayano S, Kure S, Suzuki Y, et al. Novel IRF6 mutations in Japanese patients with Van der Woude syndrome: two missense mutations (R45Q and P396S) and a 17-kb deletion. *J Hum Genet* 2003;48:622–628.

27. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat* 2011;32:358–368.
28. Botti E, Spallone G, Moretti F, et al. Developmental factor IRF6 exhibits tumor suppressor activity in squamous cell carcinomas. *Proc Natl Acad Sci USA* 2011;108:13710–13715.
29. Little HJ, Rorick NK, Su LI, et al. Missense mutations that cause Van der Woude syndrome and popliteal pterygium syndrome affect the DNA-binding and transcriptional activation functions of IRF6. *Hum Mol Genet* 2009;18:535–545.
30. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med* 2008;10:294–300.
31. Flanagan SE, Patch AM, Ellard S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers* 2010;14:533–537.
32. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012;335:823–828.



**This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>**