

Genome-wide copy number scan identifies *IRF6* involvement in Van der Woude syndrome in an Indian family

DINESH S. MANJEGOWDA^{1,2*}, MANU PRASAD³, AVINASH M. VEERAPPA⁴ AND NALLUR B. RAMACHANDRA⁴

¹Genetics Unit, Department of Anatomy, Yenepoya Medical College, Mangalore-575018, Karnataka, India

²NUCSER, K. S. Hegde Medical Academy, Nitte University, Mangalore-575018, Karnataka, India

³Centre for Craniofacial Anomaly, Yenepoya Dental College and Hospital, Mangalore-575018, Karnataka, India

⁴Genetics and Genomics Lab, Department of Studies in Zoology, University of Mysore, Manasagangotri, Mysore-570006, Karnataka, India

(Received 18 February 2014; revised 13 July 2014; accepted 12 September 2014)

Summary

Van der Woude syndrome (VWS) is an autosomal dominant developmental malformation presenting with bilateral lower lip pits related to cleft lip, cleft palate and other malformations. We performed a whole-genome copy number variations (CNVs) scan in an Indian family with members suffering from VWS using 2.6 million combined SNP and CNV markers. We found CNVs affecting *IRF6*, a known candidate gene for VWS, in all three cases, while none of the non-VWS members showed any CNVs in the *IRF6* region. The duplications and deletions of the chromosomal critical region in 1q32–q41 confirm the involvement of CNVs in *IRF6* in South Indian VWS patients. Molecular network analysis of these and other cleft lip/palate related module genes suggests that they are associated with cytokine-mediated signalling pathways and response to interferon-gamma mediated signalling pathways. This is a maiden study indicating the involvement of CNVs in *IRF6* in causing VWS in the Indian population.

1. Introduction

Van der Woude Syndrome (VWS; NCBI OMIM: 119300) is a dominantly inherited developmental disorder characterized by lower lip, cleft lip and/or cleft palate (Van der Woude, 1954; Ali *et al.*, 2009). It is the most common cleft syndrome, accounting for approximately 2% of all cases and is caused by mutations in the interferon regulatory factor 6 (*IRF6*) gene, located on 1q32.2 (Ferrero *et al.*, 2010). Nearly 100 mutations in the coding region of the gene have been reported from several populations in different geographical regions of the globe. So far, *IRF6* is the only gene shown to be associated with VWS as a dominant trait with variable penetrance and expressivity (Kayano *et al.*, 2003; Tan *et al.*, 2013). Ingraham *et al.* (2006) found that abnormalities of epithelial differentiation that resulted in cleft palate were a

consequence of adhesion between the palatal shelves and the tongue in both *Irf6*-null and homozygous missense mice. Richardson *et al.* (2009) showed that *Irf6*/*Jag2* doubly heterozygous mice displayed fully penetrant intraoral epithelial adhesions, resulting in cleft palate. However, direct sequencing of all coding regions and exon–intron boundaries of the *IRF6* gene have not shown any mutation in Indian cohorts (Moghe & Mauli, 2011). Another Indian study of a VWS pedigree conducted by Moghe *et al.* (2010) also corroborates the findings that the critical region of 1q32–q41 may not be causal for VWS in multi-ethnic and diverse Indian populations (Moghe *et al.*, 2010).

A growing body of evidence suggests that structural variations across the genome, including CNVs, are likely to contribute to human disorders (Estivill & Armengo, 2007; Conrad *et al.*, 2010). CNVs are DNA segments longer than 1 kb with >90% sequence identity that differ in the number of copies between the genomes of different individuals (Freeman *et al.*, 2006). They affect more nucleotides per genome than SNPs and contribute significantly to variations in the levels of gene expression and to phenotypes of

* Corresponding Author: Dr Dinesh S. Manjgowda, Assistant Professor, NUCSER, K. S. Hegde Medical Academy, Nitte University, Mangalore-575018, Karnataka, India. Tel: +91 824-2204490. Fax: +91 824-2204162. E-mail: dineshmgowda1@gmail.com

medical relevance among individuals (Aitman *et al.*, 2006). Many reports indicated non-involvement of *IRF6* in VWS in the Indian context; therefore, the present study was carried out to examine the involvement of CNVs in a family with members suffering from VWS in South India.

2. Materials and methods

For this study, a family with members suffering from VWS with a four generation family history and a total of seven affected members with or without cleft lip/palate were selected (Fig. 1). Subjects were between the ages of 10–55 years. From the recruited family three affected cases (4S, 5S and 6S) and two healthy controls (2C and 3C) were subjected to a genome-wide scan while the remaining samples could not be genotyped due to non-availability of the samples.

A total of 38 randomly selected non-affected Indian controls that were residing in Karnataka, India, and with an age range of 13–73 years, were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 chip. A total of 270 HapMap samples covering CEU (CEPH collection), CHB (Han Chinese in Beijing, China), JPT (Japanese in Tokyo, Japan) and YRI (Yoruba in Ibadan, Nigeria) populations and 31 Tibetan samples were selected as unrelated control subjects. A total of 5 ml EDTA blood was collected from each member of the 12 families from the Indian control group and genomic DNA was extracted using a Promega Wizard® Genomic DNA purification kit. The isolated DNA was quantified by Bio-photometer and gel electrophoresis. This research was approved by the University of Mysore-Institutional Human Ethics Review Committee (IHEC). Written informed consent was obtained from all sample donors and the IHEC approved the sample consent procedure. Written informed consent was obtained from parents/guardians in the cases of participants being minors. They also provided written informed consent for publication of photographs. A total of 38 non-cleft lip/palate control individuals from India, 31 from Tibet, 47 from New World, 269 from HapMap, 64 from Australia, 155 from China and 953 from an Ashkenazi Jewish population, totalling 1557 genomes were selected as controls to screen for CNVs in the *IRF6* gene. The data from 270 individuals from the four populations was obtained from the International HapMap Consortium (The International HapMap Consortium, 2003). The HapMap samples come from a total of 270 people: 30 both-parent-and-adult-child trios from YRI subjects, 45 unrelated JPT subjects, 45 unrelated CHB individuals, and 30 both-parent-and-adult-child trios from CEPH. The raw, unprocessed data from the

Affymetrix Genome-Wide SNP 6.0 array for the 31 individuals from Tibet, 47 from New World, 269 from HapMap, 64 from Australia, 155 from China and 953 from an Ashkenazi Jewish population, totalling up to 1557 genomes, were obtained from the ArrayExpress Archive at the European Bioinformatics Institute.

(i) Genotyping

Genome-wide genotyping was performed using an Affymetrix Genome-Wide Human SNP Array 6.0 chip and Affymetrix CytoScan® High-Density (HD) Array having 1.8 million and 2.6 million combined SNP and CNV markers, with a median inter-marker distance of 500–600 bases. These chips provide maximum panel power and the highest physical coverage of the genome (Affymetrix, Inc., Data Sheet). Genotyping quality was assessed using Affymetrix Genotyping Console Software. Briefly, all SNPs that were called using the Birdseed v2 algorithm (<http://www.broad.mit.edu/mpg/birdsuite/birdseed.html>) had a Quality Control (QC) call rate of >97% across members in families. All the members who passed SNP QC procedures were entered into the CNV analysis. The CNV calls were generated using the Canary algorithm. Contrast QC across all samples was >2.5, it was required to be >0.4. A description of the CNV methods used in this study is given below. The copy number analysis method offers two types of segmenting methods, univariate and multivariate. These methods are based on the same algorithm, but use different criteria for determining cut-points denoting CNV boundaries.

Analysing the collated data from both the BirdSuite and Canary algorithms increased the stringency on those meeting the CNV calls with a log₁₀ of odds score greater than or equal to 10 corresponding to a false-discovery rate of ~5%. All SNPs that were called using the Birdseed v2 algorithm had a QC call rate of >97% across individuals. All the subjects and members with SNPs that passed SNP QC procedures were entered into the CNV analysis. Filters were set for ID call rates for the overall SNPs to identify IDs with poor quality DNA, if any. The CNV calls were generated using the Canary algorithm. In *Affymetrix Genotyping Console Software*, contrast QC has to be >0.4 to be included in the CNV analyses. In this study, the observed contrast QC was >2.5 across all samples showing a robust strength. To control for the possibility of spurious or artifact CNVs, we used the EIGENSTRAT approach of Price *et al.* (2006). This method derives the principal components of the correlations among gene variants and corrects for those correlations in the testing. We removed one individual from the study group because they were extreme outliers on one or more significant

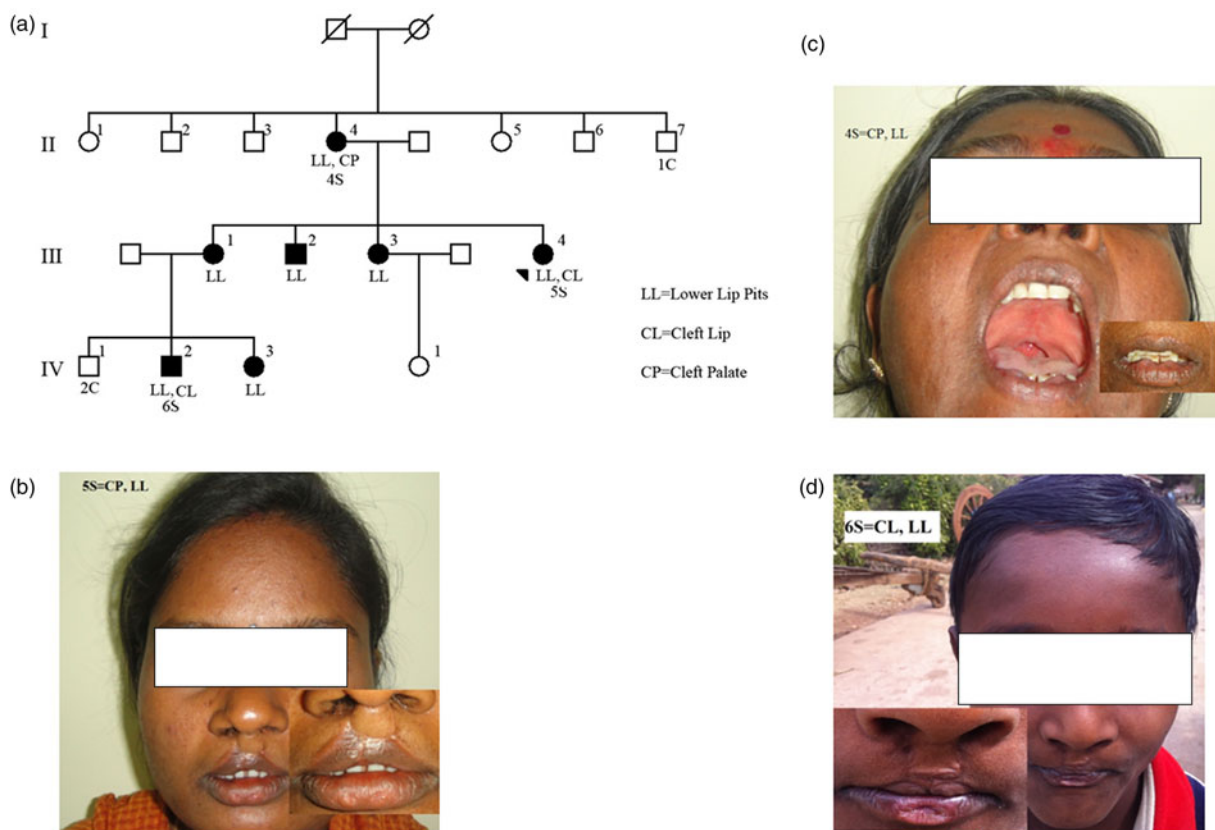


Fig. 1. Illustration of the Van der Woude pedigrees under study. All the seven affected members were found to have lower lip anomalies, among these, two individuals (5S and 6S) also had cleft lip, while only one affected member (4S) was also found to have cleft palate. (a) The pedigree under study consisted of 15 individuals, of which, seven were affected with cleft lip/palate across all the three generations. The proband of this study was a 10 year old boy diagnosed with cleft lip and lower lip anomalies. (b) Shows subject 4S who is an affected female in the second generation of the pedigree with lower lip and cleft palate. (c) Shows subject 5S who is one of the affected daughters of case 4S having lower lip and cleft palate and (d) shows case 6S, who is the affected daughter of 5S with lower lip and cleft lip.

EIGENSTRAT axes and further dropped 54 CNVs in the members selected for the study as they did not meet the required QC measures. CNVs were considered validated when there was a reciprocal overlap of 50% or greater with the reference set. Though the Jaccard statistic is sensitive to the number of CNVs called by each algorithm (ideally each of the two algorithms would detect a similar number of CNV calls), the relative values between the different comparisons of algorithms/platform/site are very informative. All the overlap analyses performed have handled losses and gains separately except when otherwise stated, and were conducted hierarchically. The calls from the algorithms that were called in both were not considered; instead, they were collated so that the relative values between the different comparisons of algorithms/platform/site are still very informative. The genotyping was performed on all subjects using multiple algorithms. The details of genotyping, data analysis, breakpoint validations and other programs used in this study have been described in detail in Veerappa *et al.*, (2013 a; 2013 b; 2013 c; 2013 d).

(ii) *Weighted protein interaction network analysis*

We used weighted protein network analysis in a first attempt to identify odds ratio gene associated modules and their key constituents. Weighted protein network analysis starts from the level of hundreds of genes, identifies modules of co-expressed proteins and proteins falling under the common pathway and relates these modules to Gene Ontology information. We made use of tools such as GeneMANIA (Ward-Farley *et al.*, 2010), BIOGRID and CYTOSCAPE (Shannon *et al.*, 2003), which were developed for network studies to assess the functional consequences of the network topology.

(iii) *IRF6 Ingenuity Pathway Analysis*

Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems; www.ingenuity.com) was used to identify interactions between genes, protein–protein interactions, biological mechanisms, as well as the location and functions of target genes. Genes and chemical search was used to explore the information on protein

families, protein signalling and metabolic pathways along with normal cellular activity of proteins. Genes and their protein products are shown based on their location. An edge (line) is used to represent the relationship between two nodes. Each edge in the network was represented by obtaining the information from resource database, Ingenuity Pathways Knowledge Base and also from the cited literature. Ingenuity Pathways Knowledge Base was considered for the analysis. A cascade of protein–protein interactions or protein binding, molecular cleavage, activation, upregulation and downregulation networks were analysed.

3. Results

(i) Genome-wide copy number and phenotype assessment of VWS

The pedigree under study consisted of 15 individuals, of which, seven were affected with cleft lip/palate across all three generations. The proband of this study was a 10 year old boy diagnosed with cleft lip and lower lip anomalies. All seven affected members of the family were recruited after thorough diagnostic testing, and all seven were found to have lower lip anomalies, among these, two individuals (5S and 6S) also had cleft lip, while only one affected member (4S) was also found to have cleft palate (Fig. 1(a)). Three affected and two related unaffected members from the three generation family were screened for CNVs using Affymetrix Cytoscan HD array having 2.6 million CNV and SNP markers. Collated data from the arrays meeting the CNV calls criteria of a log₁₀ of odds score greater than or equal to 10 (corresponding to a false-discovery rate of ~5%) were selected for further investigation. We observed a total of 1198 CNV events across the genomes of both cases and controls, with 74% duplications and 26% deletions. Various CNVs ranging in size from 100 kb to several Mb were observed in multigene families of the alpha-amylase family (AMY1 and AMY2), rRNA genes, immunoglobulins, olfactory receptor genes and in the uridine glucuronosyltransferase genes. However, CNV analysis revealed the presence of rare *de novo* microduplications and deletions at the 1q32 region (Table 1). This region harbours *IRF6* and the CNVs partially disrupted the exon structure (Fig. 2). These duplication and deletion CNVs were found to have a copy number state range of 0–3 in cases, whereas a normal copy number state of 2 was observed in all unaffected individuals.

Case 4S is an affected female individual in the second generation of the pedigree (Fig. 1). This individual showed a homozygous deletion CNV of 3 kb in size from 209974442 bp to 209977378 bp at the 1q32 region, which encompasses *IRF6*. However, Case 5S, one of the affected daughters (Fig. 1) of case 4S,

showed a contiguous 53 kb heterozygous duplication CNV from 209972232 bp to 209988311 bp that encompassed upstream and exons of *IRF6*, this was followed by a 18 kb deletion CNV from 209970522 bp to 209988149 bp in the 1q32 region, but in the homozygous state with the intermediate distance of ~16 kb between the duplication and deletion CNV. The end breakpoint of the duplication CNV and the start point of the deletion CNV partially disrupted *IRF6*. These two CNVs amounted to ~71 kb in variation. Case 6S, the affected daughter of 5S (Fig. 1), also displayed a deletion CNV of 13 kb from 209974651 bp to 209988149 bp in the 1q32 region in the homozygous state. Though the deletion CNV in both cases 5S and 6S have varied start points, they both have the same end point, which encompasses *IRF6*.

Since the CNVs identified here encompass *IRF6* (Fig. 2), we investigated whether *IRF6* interacts with any other known facial morphology determining genes using the protein interaction network tool, which uses known functional relationships and interactions between gene products reported in the literature. *IRF6* was found to be mapped to a single genetic network (Fig. 3) that included genes interacting with a known candidate genes of facial morphology determining genes such as *PAX3*, *PRDM16*, *COL17A1*, *TP63* and *HNRNPAB*, which are involved in cytokine-mediated signalling pathways, response to interferon-gamma and the interferon-gamma mediated signalling pathways. Protein network and pathway analysis also showed *IRF6* interacting with *PAX3*, *PRDM16*, *COL17A1*, *TP63* and *HNRNPAB* proteins directly, or sometimes with intermediate partners essential for response to interferon-gamma and interferon-gamma mediated signalling pathways, which are involved in normal facial morphology. Duplications and deletions in *IRF6* identified in this study disrupt the protein production, which in turn fails to interact with *PAX3*, *PRDM16*, *COL17A1*, *TP63* and *HNRNPAB* resulting in the blocked pathway.

(ii) Protein interaction network of VWS

Network analysis of *IRF6* (Fig. 3) has established ~24 genes with protein and genetic interactions, leading to more than 70 links (Table 2). Out of these, interactions involving some genes participating in interferon-gamma mediated signalling pathways could be accounted for by the analysis in the present study. These interactions are more specific for interferon signalling as they directly participate in cytokine-mediated signalling pathways, response to interferon-gamma and interferon-gamma mediated signalling pathways. The geometrical establishment of the *IRF6* network also shows already identified candidate genes of facial morphology determinants.

Table 1. Duplications and deletions in different regions of *IRF6* in Van der Woude subjects from a family of South India

Subjects	Size	Breakpoint	CNV type	Location	Start and end markers	Total markers
4S	3 kb	209974442 bp to 209977378 bp	Deletion	Exon	SNP_A8456813-CN_976608	2
5S	53 kb	209972232 bp to 209988311 bp	Duplication	Upstream and exon	SNP_A8768342-CN_975498	5
6S	18 kb	209970522 bp to 209988149 bp	Deletion	Upstream	CN_965373-CN_967544	3
	13 kb	209974651 bp to 209988149 bp	Deletion	Exon	CN_972117-SNP_A-8306049	26



Fig. 2. Schematic representation of copy number state of Van der Woude cases. An image of the copy number state (5S, 6S and 4S) indicates the quantitative assessments of genotyping used to determine copy number variation in the Van der Woude (VWS) cases. Blue bars indicate duplications and red indicates deletions. Case 4S shows a homozygous deletion copy number variation (CNV) of 3 kb from 209974442 bp to 209977378 bp at the 1q32 region encompassing *IRF6*. Case 5S shows a contiguous 53 kb heterozygous duplication CNV followed by a 18 kb deletion CNV from 209970522 bp to 209988149 bp in the 1q32 region, but in the homozygous state with the intermediate distance of ~16 kb between the duplication and deletion CNV. The end breakpoint of the duplication CNV and the start point of the deletion CNV partially disrupted *IRF6*. These two CNVs amounted to ~71 kb in variation. Case 6S shows a deletion CNV of 13 kb from 209974651 bp to 209988149 bp in the 1q32 region in the homozygous state. Though the deletion CNV in both the 5S and 6S cases have varied start points, they both have the same end point, which encompasses *IRF6*.

The network analysis of the above genes (Fig. 3) and their interconnecting pathways suggest that they are vital in the following three biological functions: (1) cytokine-mediated signalling pathways, (2) response to interferon-gamma and (3) interferon-gamma mediated signalling pathways. Thus, disruption of the genes involved in these pathways identifies critical regions in normal lip, palate and face morphology formation.

The observed CNV breakpoints in the 1q32 region from this study were validated by genotyping 38 individuals from India, 31 from Tibet, 47 from New World, 269 from HapMap, 64 from Australia, 155 from China and 953 from an Ashkenazi Jewish population, totalling up to 1557 genomes. None of the

members among this non-cleft lip/palate control group showed any CNV breakpoints in the 1q32 bearing *IRF6* region.

(iii) IPA

IPA revealed interactions between *IRF6* and several other transcription factors and DNA binding proteins (Fig. 4). Many of these proteins are major determinants of facial structure and morphology. *IRF6* shows interactions with diverse types of functional proteins and these processes include cell migration, signalling, development, and tissue development and morphology (Supplementary Table 1; available online).

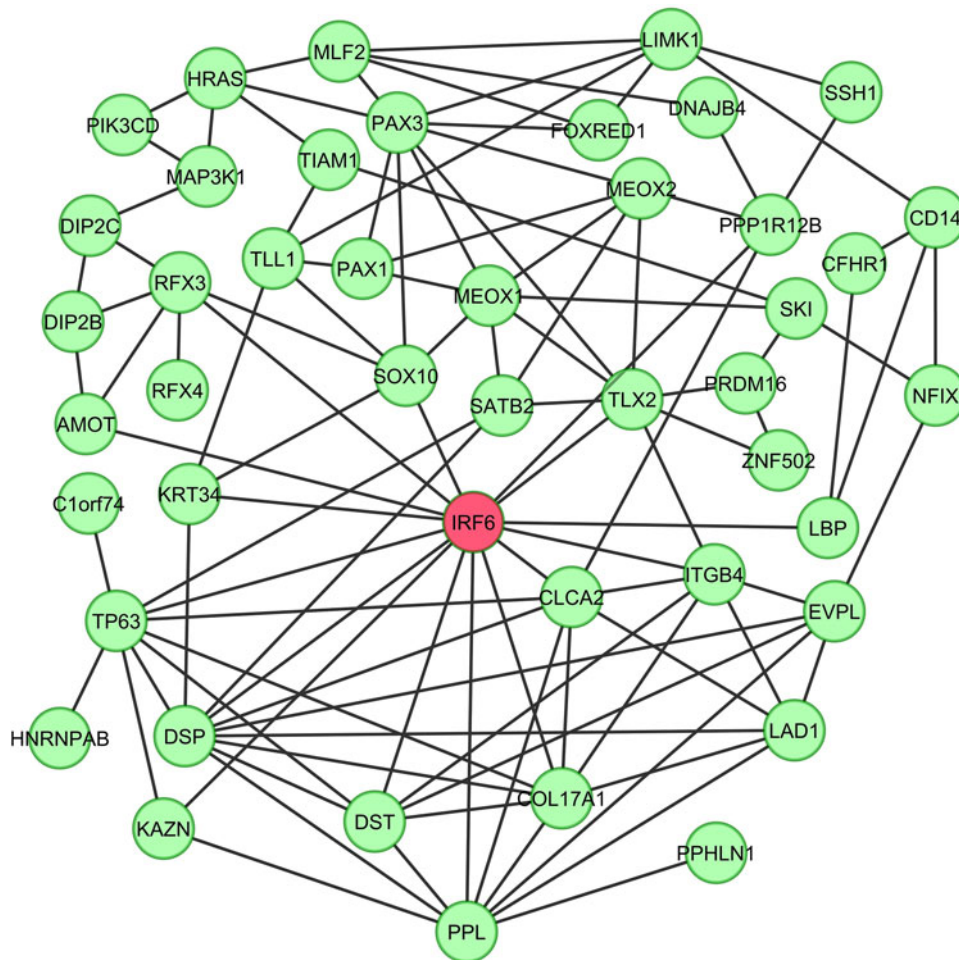


Fig. 3. *IRF6* protein interaction network establishing the associated gene modules and their key hubs. The figure shows the network of genes involved in cleft lip/palate with hub genes coloured according to their respective Gene Ontology (GO). Network analysis of candidate genes of lip/palate, including the genes identified in this study, has established ~24 genes with protein and genetic interactions that have more than >70 links. Out of these, interactions involving some genes that participate in the facial determination region could be accounted for by the analysis in the present study. These interactions are more specific for interferon signalling as they directly participate in cytokine-mediated signalling pathways, response to interferon-gamma and interferon-gamma mediated signalling pathways. The geometrical establishment of the *IRF6* network shows already identified candidate genes interacting with facial morphology genes in the present study. The smaller and larger circles depict network weights that are chosen based on the size of the input gene list. Based on the input gene list, a 'GO-based weighting, biological process based' weighting method is chosen. This weighting method assumes the input gene list is related in terms of biological processes (as defined by GO) and assigns weights based on maximizing connectivity between all input genes using the 'assigned based on query gene' strategy.

We further examined the proteins interacting with *IRF6* to determine functional relevance and disease pathways relevant to VWS. A total of 64 coding genes were included in the IPA. We ranked functional and disease categories based on enrichment for all the genes (Supplementary Table 1). The top 20 ranked categories included functional and disease categories related to cell-to-cell signalling and interaction, cellular movement, cellular growth and proliferation, tissue morphology, and cell death and survival. IPA was also used to search for biological relationships among the *IRF6* interacting genes with previously identified facial morphology genes (Fig. 4). Paths between genes represent protein-protein interactions, regulation and phosphorylation activities. *IRF6*

in the pathway is regulated by several of its partners that if altered result in cleft lip/palate. Supplementary Table 1 indicates the functional annotation of proteins, their significant p-value and the number of genes participating in that particular function.

4. Discussion

So far, globally available reports show that VWS is caused by mutations in the *IRF6* gene located on chromosome 1q32.2 (Kondo *et al.*, 2002; Ferrero *et al.*, 2010; Leslie & Marazita, 2013). Recently, Indian studies on VWS reported that the critical region on

Table 2. Annotation of the genes identified through network and pathway studies based on location, type, biomarker application and drugs

Symbol	Entrez gene name	Location	Type(s)	Biomarker application(s)	Drug(s)
<i>Clorf74</i>	Chromosome 1 open reading frame 74	Other	Other		
<i>COL17A1</i>	Collagen, type XVII, alpha 1	Extracellular space	Other		Collagenase clostridium histolyticum
<i>DSP</i>	Desmoplakin	Plasma membrane	Other		
<i>DST</i>	Dystonin	Plasma membrane	Other	Unspecified application	
<i>FOXRED1</i>	FAD-dependent oxidoreductase domain containing 1	Cytoplasm	Other		
<i>HNRNPAB</i>	Heterogeneous nuclear ribonucleoprotein A/B	Nucleus	Enzyme		
<i>HRAS</i>	Harvey rat sarcoma viral oncogene homolog	Plasma membrane	Enzyme	Prognosis, safety	
<i>IRF6</i>	Interferon regulatory factor 6	Nucleus	Transcription regulator		
<i>ITGB4</i>	Integrin, beta 4	Plasma membrane	Transmembrane receptor	Diagnosis	
<i>LAD1</i>	Ladinin 1	Extracellular space	Other		
<i>LBP</i>	Lipopolysaccharide binding protein	Plasma membrane	Transporter	Diagnosis, efficacy	
<i>LIMK1</i>	LIM domain kinase 1	Cytoplasm	Kinase	Unspecified application	Dabrafenib, dabrafenib/trametinib
<i>MEOX1</i>	Mesenchyme homeobox 1	Nucleus	Transcription regulator		
<i>MEOX2</i>	Mesenchyme homeobox 2	Nucleus	Transcription regulator		
<i>MLF2</i>	Myeloid leukemia factor 2	Nucleus	Other		
<i>PAX3</i>	Paired box 3	Nucleus	Transcription regulator		
<i>PPL</i>	Periplakin	Cytoplasm	Other	Diagnosis	
<i>PPP1R12B</i>	Protein phosphatase 1, regulatory subunit 12B	Other	Phosphatase		
<i>PRDM16</i>	PR domain containing 16	Nucleus	Transcription regulator		
<i>RFX3</i>	Regulatory factor X, 3 (influences HLA class II expression)	Nucleus	Transcription regulator		
<i>SATB2</i>	SATB homeobox 2	Nucleus	Transcription regulator		
<i>SKI</i>	v-ski avian sarcoma viral oncogene homolog	Nucleus	Transcription regulator		
<i>SOX10</i>	SRY (sex determining region Y)-box 10	Nucleus	Transcription regulator		
<i>TLX2</i>	T-cell leukemia homeobox 2	Nucleus	Transcription regulator		
<i>TP63</i>	Tumor protein p63	Nucleus	Transcription regulator	Diagnosis, efficacy, prognosis, unspecified application	

1q32–q41 and the mutations of *IRF6* may not be causal for VWS in multi-ethnic and diverse Indian populations (Moghe *et al.*, 2010; Moghe & Mauli, 2011). However, molecular studies on VWS in India are

very limited so are not able to confirm or reject these controversial reports. Therefore, a further molecular study involving the identification of CNVs has been carried out by a whole-genome scan to

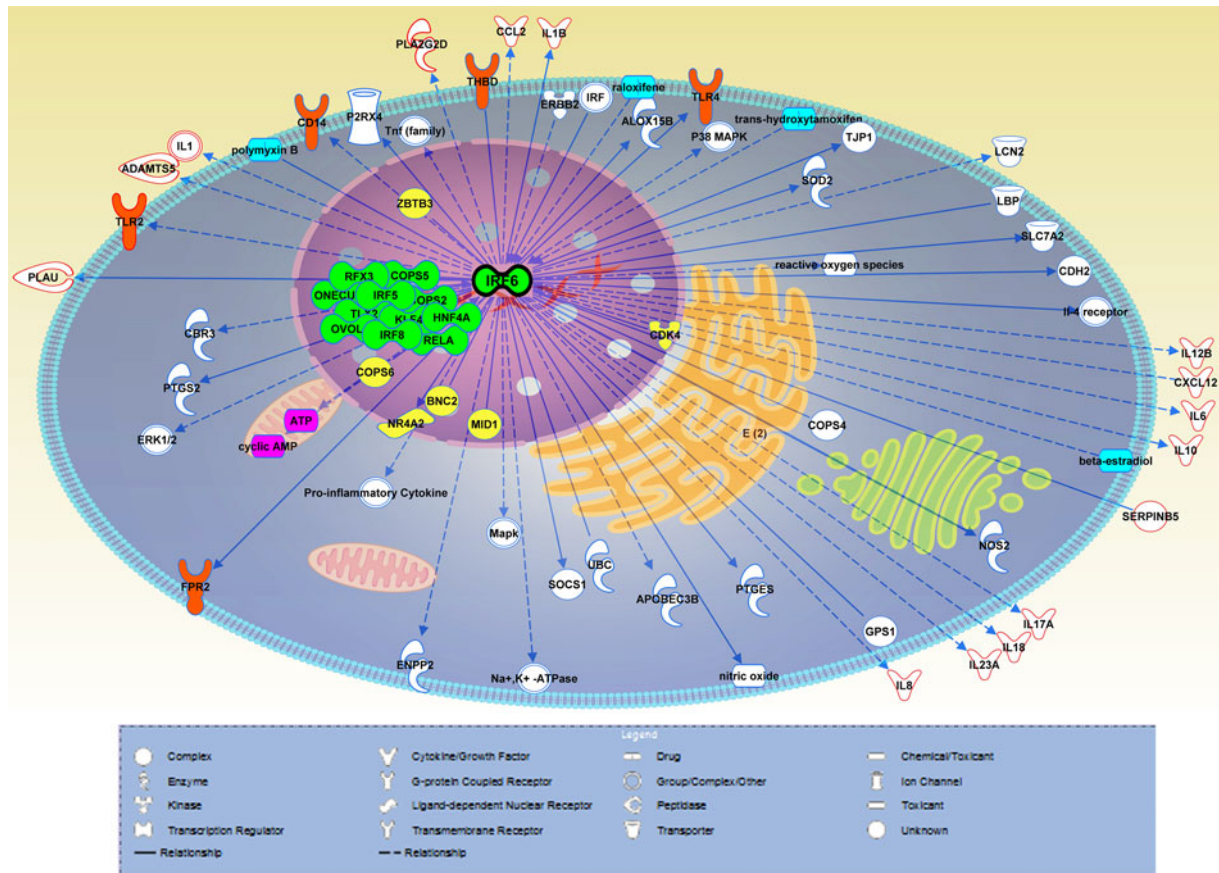


Fig. 4. Ingenuity Pathway Analysis of the *IRF6* gene and its interactions with target proteins indicating involvement in several biological functions and pathway.

analyse the susceptible region or genes, which may be involved in causing VWS in the South Indian population.

In the present study, CNV analysis in VWS cases revealed the deletion of *IRF6* in the critical region 1q32–q41. Even in studies that have shown negative reports of failing to identify mutations in *IRF6* in VWS subjects, the role of CNVs cannot be ruled out. Birnbaum *et al.* (2008) analysed *IRF6* in 63 families with what was believed to be isolated cleft lip/palate or cleft palate and identified a deletion/insertion and duplication in two families, respectively, and a history of lip pits in both family members confirmed the diagnosis of VWS. Linkage analyses in five Finnish families linked the 1q32–q41 region with VWS in three of these families (Wong *et al.*, 2001). In a recent study, five independent genetic loci were reported to be associated with different facial phenotypes, suggesting the involvement of five candidate genes: *PRDM16*, *PAX3*, *TP63*, *C5orf50* and *COL17A1* in the determination of the human face (Liu *et al.*, 2012). VWS is known to be a Mendelian disorder of variable expressivity with high penetrance. In the present study, we observed CNV breakpoints inherited inconsistently in cases 4S, 5S and 6S. We

believe this is due to the unequal crossing-over mechanism since it exhibits contiguous integration of the duplication/deletion CNVs. We put forth that these CNVs have resulted because of the unequal recombination between two 1q32.3 chromosomal regions, resulting in the gain or loss of a part of 1q32.3, which can be seen as heterozygous or homozygous duplication/deletions in the subjects.

(i) Molecular network of *IRF6*

We propose that the *IRF6* gene network model (Fig. 3) controls three major functions: (1) cytokine-mediated signalling pathways, (2) response to interferon-gamma and (3) interferon-gamma mediated signalling pathways. We also identified disease-related networks (modules) and disease-related hub genes and found large protein–protein interaction hubs associated with facial morphology determinants (Fig. 3). We used minimal cut sets (MCSs) in this network study (Klamt, 2006), which helps in studying the structural fragility and to identify knock out strategies in cellular networks. When the MCSs are computed with respect to “physical and genetic” interactions in realistic networks, the experimental block of essential

genes will inevitably lead to mutants. Thus, the candidate genes identified in this study are the indicators of major MCSs. The immediate interacting protein partners represents primary level MCSs and the immediate interacting protein partners of the primary level MCSs represent secondary level MCSs. These MCSs are distributed in a fragile interconnected network. This can be demonstrated as follows: *IRF6* interacts with *TP63* and *PAX3* proteins, which are essential for a proper interferon-gamma mediated signalling pathway, whereas *PRDM16* and *COL17A1* are essential for efficient signalling. Duplication and deletion in *IRF6* identified in this study disrupts the protein interaction with *TP63*, *PAX3*, *PRDM16* and *COL17A1* resulting in the blocked pathway. This type of structural fragility in the *IRF6* network is given by the size distribution over the total set of MCSs. In such networks, a dysfunction can be caused in higher probability, as a malfunction requires only one or a few network elements to fail. IPA revealed *IRF6* involvement in cell migration, signalling, development, and tissue development and morphology. This probably explains the non-migration of certain cells to the lip and palate region and failure to develop facial morphology.

The present study is a maiden report indicating the involvement of CNVs in *IRF6* in causing VWS. Genome-wide analysis of CNVs confirms the involvement of the critical 1q32–q41 region and *IRF6* in South Indian VWS patients and the CNVs found in the patients were associated with facial morphology. Various studies have shown that direct sequencing of all coding regions and exon–intron boundaries of *IRF6* have not shown any mutation in Indian cohorts (Moghe *et al.*, 2010; Moghe & Mauli, 2011) of VWS in multi-ethnic and diverse Indian populations, indicating the presence of a secondary causal gene that can cause VWS. Based on these reports, we hypothesized a role for the *PRDM16*, *PAX3*, *TP63*, *C5orf50* and *COL17A1* genes (which are known to interact physically with *IRF6*) in VWS. These genes have not been formally validated but have been found to play a significant role in pathways involving cytokine-mediated signalling, response to interferon-gamma and interferon-gamma mediated signalling pathways. Based on our study of the pathway analysis of *IRF6*, we identified various forms of interactions with *IRF6* proteins that play a key role in VWS manifestation. Hence while considering studies that failed to identify mutations in *IRF6*, it is necessary to focus on the above-mentioned probable genes. Since CNVs in *IRF6*/the 1q32.2 region were identified in living cleft lip and cleft palate subjects, and since this gene is not expressed in leukocytes it cannot be analysed from blood; therefore, identifying the changes at expression level was a possible limitation of this study.

This work was supported by the Yenepoya University funding seed grant (YU/Seed Grant/2011-011). The Authors thank participants and their family for participating in this study.

5. Declaration of interest

None

6. Authors' contributions

DSM and AMV performed experiments and analysis. DSM, AMV, MP and NBR conceived of the study, designed the experiments, interpreted the results and prepared the manuscript. DSM, MP, AMV and NBR read and approved the final manuscript.

7. Supplementary material

The online supplementary material can be found available at <http://journals.cambridge.org/GRH>

References

- Affymetrix, Inc. (2009). Data Sheet: Genome Wide Human SNP Array 6-0 (http://www.affymetrix.com/estore/catalog/131533/AFFY/Genome-Wide+Human+SNP+Array+6.0#1_3). Accessed 14 February 2013.
- Aitman, T.J., Dong, R., Vyse, T.J., Norsworthy, P.J., Johnson, M.D., Smith, J., Mangion, J., Robertson-Lowe, C., Marshall, A.J., Petretto, E., Hodges, M.D., Bhangal, G., Patel, S.G., Sheehan-Rooney, K., Duda, M., Cook, P.R., Evans, D.J., Domin, J., Flint, J., Boyle, J.J., Pusey, C.D. & Cook, H.T. (2006). Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851–855.
- Ali, A., Singh, S.K. & Raman, R. (2009). Coding region of *IRF6* gene may not be causal for Van Der Woude syndrome in cases from India. *The Cleft Palate-Craniofacial Journal* **46**, 541–544.
- Birnbaum, S., Reutter, H., Lauster, C., Scheer, M., Schmidt, G., Saffar, M., Martini, M., Hemprich, A., Henschke, H., Kramer, F.J. & Mangold, E. (2008). Mutation screening in the *IRF6*-gene in patients with apparently nonsyndromic orofacial clefts and a positive family history suggestive of autosomal-dominant inheritance. *American Journal of Medical Genetics. Part A*. **146A**, 787–790.
- Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., Macarthur, D.G., Macdonald, J.R., Onyiah, I., Pang, A.W., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Wellcome Trust Case Control Consortium, Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W. & Hurles, M.E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712.
- Estivill, X. & Armengo, L. (2007). Copy number variants and common disorders: filling the gaps and exploring

- complexity in genome-wide association studies. *PLoS Genetics* **3**, 1787–1799.
- Ferrero, G. B., Baldassarre, G., Panza, E., Valenzise, M., Pippucci, T., Mussa, A., Pepe, E., Seri, M. & Silengo, M. C. (2010). A heritable cause of cleft lip and palate – Van der Woude syndrome caused by a novel *IRF6* mutation. *Review of the literature and of the differential diagnosis*. *European Journal of Pediatrics* **169**, 223–228.
- Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Altshuler, D. M., Aburatani, H., Jones, K. W., Tyler-Smith, C., Hurles, M. E., Carter, N. P., Scherer, S. W. & Lee, C. (2006). Copy number variation: new insights in genome diversity. *Genome Research* **16**, 949–961.
- Ingraham, C. R., Kinoshita, A., Kondo, S., Yang, B., Sajan, S., Trout, K. J., Malik, M. I., Dunnwald, M., Goudy, S. L., Lovett, M., Murray, J. C. & Schutte, B. C. (2006). Abnormal skin, limb and craniofacial morphogenesis in mice deficient for interferon regulatory factor 6 (*Irf6*). *Nature Genetics* **38**, 1335–1340.
- Kayano, S., Kure, S., Suzuki, Y., Kanno, K., Aoki, Y., Kondo, S., Schutte, B. C., Murray, J. C., Yamada, A. & Matsubara, Y. (2003). Novel *IRF6* mutations in Japanese patients with Van der Woude syndrome: two missense mutations (R45Q and P396S) and a 17-kb deletion. *Journal of Human Genetics* **48**, 622–628.
- Klamt, S. (2006). Generalized concept of minimal cut sets in biochemical networks. *Biosystems* **83**, 233–247.
- Kondo, S., Schutte, B. C., Richardson, R. J., Bjork, B. C., Knight, A. S., Watanabe, Y., Howard, E., de Lima, R. L., Daack-Hirsch, S., Sander, A., McDonald-McGinn, D. M., Zackai, E. H., Lammer, E. J., Aylsworth, A. S., Ardinger, H. H., Lidral, A. C., Pober, B. R., Moreno, L., Arcos-Burgos, M., Valencia, C., Houdayer, C., Bahuau, M., Moretti-Ferreira, D., Richieri-Costa, A., Dixon, M. J. & Murray, J. C. (2002). Mutations in *IRF6* cause Van der Woude and popliteal pterygium syndromes. *Nature Genetics* **32**, 285–289.
- Leslie, E. J. & Marazita, M. L. (2013). Genetics of cleft lip and cleft palate. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics* **163C**, 246–258.
- Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G., Wollstein, A., Lao, O., de Bruijne, M., Ikram, M. A., van der Lugt, A., Rivadeneira, F., Uitterlinden, A. G., Hofman, A., Niessen, W. J., Homuth, G., de Zubicaray, G., McMahon, K. L., Thompson, P. M., Daboul, A., Puls, R., Hegenscheid, K., Bevan, L., Pausova, Z., Medland, S. E., Montgomery, G. W., Wright, M. J., Wicking, C., Boehringer, S., Spector, T. D., Paus, T., Martin, N. G., Biffar, R. & Kayser, M. (2012). A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genetics* **8**, e1002932.
- Moghe, G. A., Kaur, M. S., Thomas, A. M., Raseswari, T., Swapna, M. & Rao, L. (2010). The role of 9qh+ in phenotypic and genotypic heterogeneity in a Van der Woude syndrome pedigree. *Journal of the Indian Society of Pedodontics and Preventive Dentistry* **28**, 104–109.
- Moghe, G. A. & Mauli, S. (2011). *IRF6* mutations may not be a major cause of Van der Woude syndrome in India. *European Journal of Pediatrics* **170**, 129.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904–909.
- Richardson, R. J., Dixon, J., Jiang, R., Dixon, M. J. (2009). Integration of *IRF6* and *Jagged2* signalling is essential for controlling palatal adhesion and fusion competence. *Human Molecular Genetics* **18**, 2632–2642.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498–2504.
- Tan, E. C., Lim, E. C. & Lee, S. T. (2013). *De novo* 2.3 Mb microdeletion of 1q32.2 involving the Van der Woude Syndrome locus. *Molecular Cytogenetics* **6**, 31.
- The International HapMap Consortium. (2003). The International HapMap Project. *Nature* **426**, 789–796.
- Van der Woude, A. (1954). Fistula labii inferioris congenita and its association with cleft lip and palate. *American Journal of Human Genetics* **6**, 244–256.
- Veerappa, A. M., Padakannaya, P. & Ramachandra, N. B. (2013a). Copy number variation-based polymorphism in a new pseudoautosomal region 3 (PAR3) of a human X-chromosome-transposed region (XTR) in the Y chromosome. *Functional & Integrative Genomics* **13**, 285–293.
- Veerappa, A. M., Saldanha, M., Padakannaya, P. & Ramachandra, N. B. (2013b). Family-based genome-wide copy number scan identifies five new genes of dyslexia involved in dendritic spinal plasticity. *Journal of Human Genetics* **58**, 539–547.
- Veerappa, A. M., Saldanha, M., Padakannaya, P. & Ramachandra, N. B. (2013c). Genome-wide copy number scan identifies disruption of *PCDH11X* in developmental dyslexia. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics* **162B**, 889–897.
- Veerappa, A. M., Vishweswaraiyah, S., Lingaiah, K., Murthy, M., Manjgowda, D. S., Nayaka, R. & Ramachandra, N. B. (2013d). Unravelling the complexity of human olfactory receptor repertoire by copy number analysis across population using high resolution arrays. *PLoS ONE* **8**, e66843.
- Wardle-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C. T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D. & Morris, Q. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research* **1**, 38–45.
- Wong, F. K., Koillinen, H., Rautio, J., Teh, B. T., Ranta, R., Karsten, A., Larson, O., Linder-Aronson, S., Huggare, J., Larsson, C. & Kere, J. (2001). Genetic heterogeneity and exclusion of a modifying locus at 17p11.2-p11.1 in Finnish families with van der Woude syndrome. *Journal of Medical Genetics* **38**, 198–202.