

Missense mutations that cause Van der Woude syndrome and popliteal pterygium syndrome affect the DNA-binding and transcriptional activation functions of IRF6

Hayley J. Little^{1,2,‡}, Nicholas K. Rorick^{3,‡}, Ling-I Su^{1,‡}, Clair Baldock¹, Saimon Malhotra¹, Tom Jowitt¹, Lokesh Gakhar⁴, Ramaswamy Subramanian⁴, Brian C. Schutte^{3,†}, Michael J. Dixon^{1,2,*} and Paul Shore¹

¹Faculty of Life Sciences, Michael Smith Building and ²Dental School, University of Manchester, Oxford Road, Manchester M13 9PT, UK, ³Department of Pediatrics and Interdisciplinary PhD Program in Genetics and ⁴Department of Biochemistry, University of Iowa, Iowa City, IA, USA

Received September 26, 2008; Revised and Accepted November 10, 2008

Cleft lip and cleft palate (CLP) are common disorders that occur either as part of a syndrome, where structures other than the lip and palate are affected, or in the absence of other anomalies. Van der Woude syndrome (VWS) and popliteal pterygium syndrome (PPS) are autosomal dominant disorders characterized by combinations of cleft lip, CLP, lip pits, skin-folds, syndactyly and oral adhesions which arise as the result of mutations in interferon regulatory factor 6 (IRF6). IRF6 belongs to a family of transcription factors that share a highly conserved N-terminal, DNA-binding domain and a less well-conserved protein-binding domain. To date, mutation analyses have suggested a broad genotype–phenotype correlation in which missense and nonsense mutations occurring throughout IRF6 may cause VWS; in contrast, PPS-causing mutations are highly associated with the DNA-binding domain, and appear to preferentially affect residues that are predicted to interact directly with the DNA. Nevertheless, this genotype–phenotype correlation is based on the analysis of structural models rather than on the investigation of the DNA-binding properties of IRF6. Moreover, the effects of mutations in the protein interaction domain have not been analysed. In the current investigation, we have determined the sequence to which IRF6 binds and used this sequence to analyse the effect of VWS- and PPS-associated mutations in the DNA-binding domain of IRF6. In addition, we have demonstrated that IRF6 functions as a co-operative transcriptional activator and that mutations in the protein interaction domain of IRF6 disrupt this activity.

INTRODUCTION

Orofacial clefting (OFC) is a common developmental genetic disorder that occurs with a prevalence which has been estimated at between 1 in 500 and 1 in 2500 live births depending on geographic origin, racial and ethnic variation, and socio-economic status (1,2). Individuals who exhibit OFC may experience problems with eating, speaking, hearing and facial appearance

which can be corrected to varying degrees by surgery, dental treatment, speech therapy and psychosocial intervention. On the basis that the lip/primary palate and the secondary palate have distinct developmental origins, OFC can be divided into cleft lip occurring either with or without cleft palate (CLP) and isolated cleft palate in which the lip is not affected (CPO). This division is validated on the basis that, under most circumstances, CLP and CPO do not segregate in the same family (3). Although

*To whom correspondence should be addressed. Tel: +44 1612755620; Fax: +44 1612755082; Email: mike.dixon@manchester.ac.uk

†Present address: Departments of Microbiology and Molecular Genetics and Pediatrics and Human Development, Michigan State University, East Lansing, MI, USA.

‡The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

OFC may occur as part of a syndrome, where structures other than the lip and palate are affected, over 70% of cases of CLP and 50% of cases of CPO arise in the absence of other abnormalities and are collectively classified as non-syndromic (4). Recent data have been demonstrated that mutations in *PVRL1*, *MSX1*, *TBX22*, *IRF6* and *FGFR1* are responsible for syndromic forms of OFC (5–9) and that variation within these genes is a contributory factor to their non-syndromic counterparts (10–16).

Van der Woude syndrome (VWS; MIM 119300) is an autosomal dominant disorder of facial development which is characterized by cleft lip, CLP and paramedian lower lip pits (17). VWS is the most common form of syndromic OFC, accounting for ~2% of all cases, and has the phenotype that most closely resembles the more common non-syndromic forms. Popliteal pterygium syndrome (PPS; MIM 119500) has a similar orofacial phenotype to VWS; however, PPS also exhibits additional anomalies that include popliteal webbing, pterygia, oral synchia, adhesions between the eyelids, syndactyly and genital anomalies (18,19). The VWS and PPS loci were initially mapped to human chromosome 1q32–q41 (20–25) and both phenotypes were subsequently demonstrated to result from mutations in the gene encoding interferon regulatory factor 6 (IRF6; Ref. 8). IRF6 belongs to a family of transcription factors that share a highly conserved N-terminal, penta-tryptophan, helix-turn-helix DNA-binding domain and a less well-conserved protein-binding domain (8).

Initially, 46 mutations in *IRF6* were identified in VWS patients, with a further 13 being detected in families with a history of PPS (8). Mutations that introduced a termination codon into IRF6 were found to be significantly more common in VWS than in PPS consistent with haploinsufficiency being the mechanism that underlies VWS (20,22,26). The missense mutations that were observed in VWS and PPS fell into two distinct categories. Whereas the missense mutations underlying VWS were almost evenly divided between the DNA-binding and protein-binding domains, the vast majority of the missense mutations found to be associated with PPS arose in the DNA-binding domain. Moreover, comparison of the sequence of IRF6 with that of IRF1 suggested that in the case of PPS every amino acid residue mutated contacted DNA directly, whereas only a small minority of the residues mutated in VWS individuals made direct contact with DNA. While this genotype–phenotype correlation has broadly been supported by subsequent studies, it is based solely on the analysis of structural models rather than on a systematic investigation of the DNA-binding properties of IRF6. Moreover, the effects of mutations in the protein interaction domain have not been investigated.

In the current investigation, we have determined the DNA-binding sequence to which wild-type IRF6 binds and used this sequence to determine the effect of VWS- and PPS-associated mutations in the DNA-binding domain of IRF6. In addition, we have demonstrated that IRF6 functions as a co-operative transcriptional activator and that VWS-causing mutations in the protein interaction domain of IRF6 disrupt this activity.

RESULTS

Identification of an IRF6 DNA-binding site

To investigate the effects of disease-causing mutations in the IRF6 DNA-binding domain (IRF6-DBD), we first sought to

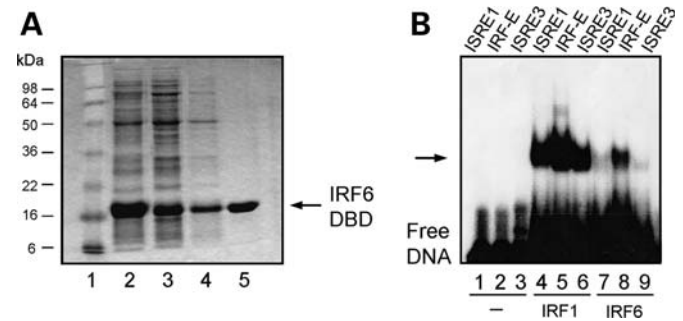


Figure 1. IRF6 DNA-binding domain binds to ISRE sites. (A) SDS–PAGE gel showing purification of the IRF6 DNA-binding domain (IRF6-DBD). The protein was purified by Nickel-affinity chromatography. Fractions from each of the purification stages are shown as indicated above the lanes. Lane 5 contains the pure protein which was subsequently used to select the IRF6 consensus sequences. (B) EMSA showing DNA binding to the three pools of degenerate ISRE sequences as indicated above the lanes. The IRF6-DBD/DNA complexes are indicated by the arrow.

identify a DNA-binding sequence to which the wild-type protein would bind. Various IRF6 constructs (amino acids 1–113; 1–225; 1–401; 1–467), all of which contained the DNA-binding domain (amino acids 13–113), were expressed as N-terminal, His-tagged fusion proteins in *E. coli* and purified to homogeneity using a single-step, Ni-affinity column (Fig. 1A). Members of the IRF family of transcription factors are known to bind to a number of consensus sites, as shown for the DNA-binding domain of IRF1 (Fig. 1B). To determine whether IRF6 could bind to a subset of these sites, electrophoretic mobility shift assays were performed using degenerate, double-stranded oligonucleotides corresponding to three potential response elements. Only the IRF6 construct composed of amino acids 1–113 caused a gel shift, demonstrating that each pool of sites contains sequences to which IRF6-DBD can bind (Fig. 1B, lanes 7–9). An identical result was obtained using *in vitro* translated proteins (data not shown).

To identify the consensus DNA-binding site for the IRF6-DBD, all permutations of the IRF-E pool of sequences were subsequently synthesized and one site was selected to aid isolation of additional sites using an unbiased, PCR-based, site-selection assay. After four rounds, the selected pool of DNA was cloned and 23 sequences determined (Fig. 2A and B). The sequences were subsequently used to derive the consensus binding site for the IRF6-DBD (Fig. 2C). To confirm that the selected sites were a true reflection of the IRF6 binding specificity, several individual sequences were tested in electrophoretic mobility shift assays; all sites tested were bound efficiently by IRF6 (Fig. 2D). To analyse the specificity of the identified core consensus sequence –AACCGAAAC^{C/T}, single base pair changes in the S17 sequence were investigated using EMSA (Fig. 2E). Whereas an A>T substitution at position 7 permitted binding, an A>C substitution at this position abrogated DNA binding (Fig. 2E, lanes 1–3). Similar approaches indicated that while substitution of cytosine at position 4, which was invariant in the site selection assay, could not be tolerated (Fig. 2E, lane 4), a C>G substitution at position 9 of the extended core sequence permitted DNA binding, albeit with lower affinity (Fig. 2E, lane 5). Collectively, these data indicate that the

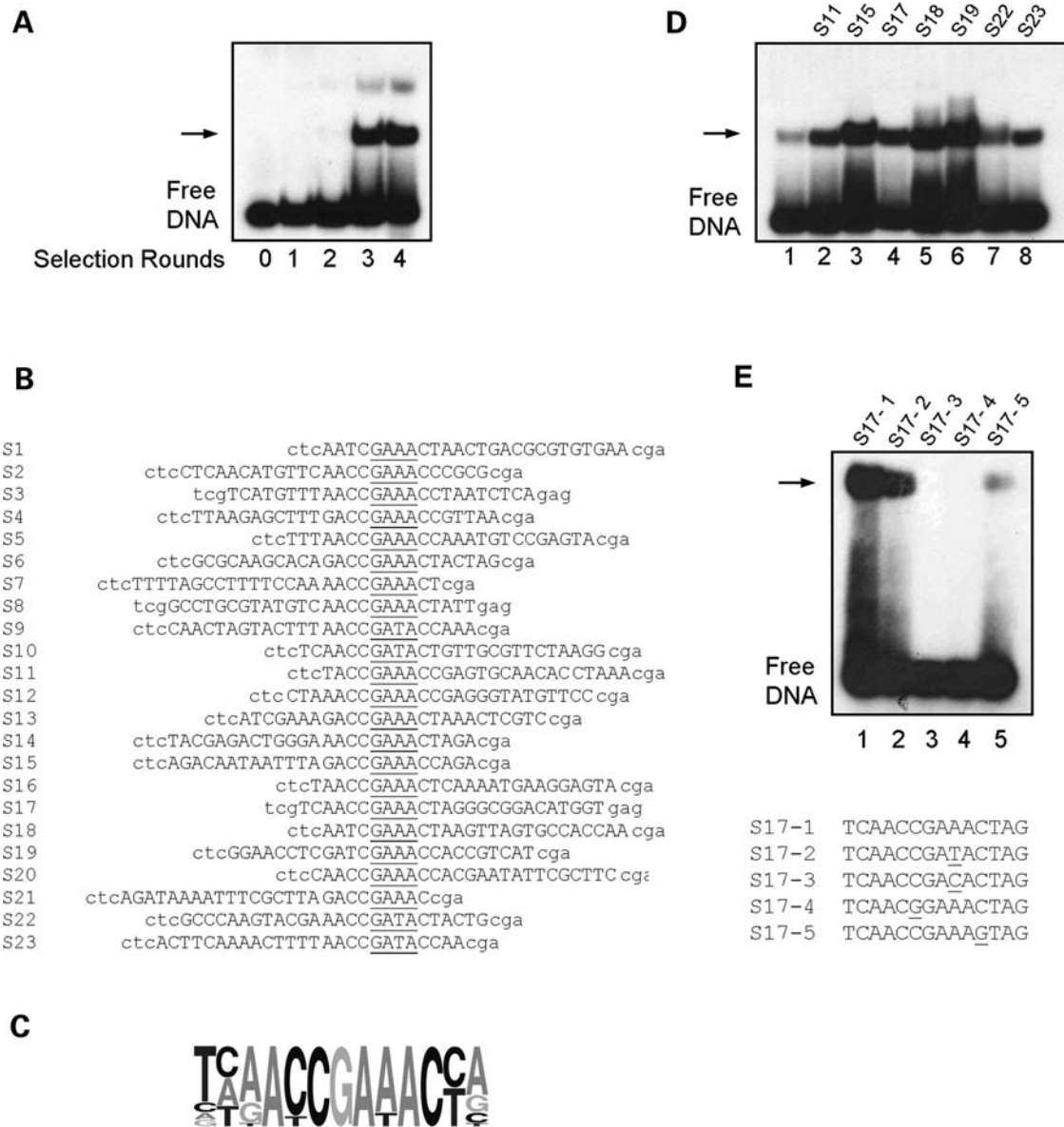


Figure 2. Selection of a consensus DNA-binding site for IRF6. (A) EMSA analysis of the selected pools of binding sites using bacterially expressed IRF6-DBD. The starting double-stranded DNA is shown in lane 0. The free DNA represents the DNA pool after the indicated number of rounds of selection. The position of the protein-DNA complex is shown (arrow). The DNA from the complexes in lane 3 was amplified and cloned for sequence analysis. (B) Sequences of the DNA-binding sites selected by IRF6 after three rounds of selection. Nucleotides derived from the random sequence (upper case) and the constant flanking primers (lower case) are indicated. The IRF core sequence is underlined in each sequence. Sites are aligned and orientated according to this IRF core sequence. (C) A schematic sequence representation for IRF6 binding sites after three rounds of selection. (D) EMSA showing IRF6-DBD binding to individual sequences obtained in the site selection. Lane 1 contains the initial sequence obtained from the IRF-E pool. The identity of each of the selected sites is shown above the lanes. (E) EMSA showing specific binding of IRF6-DBD to the S17 site and its variants containing specific point mutations. The identity of each of the sites is shown above the lanes. The core sequences are shown below the panel and the mutated bases are underlined.

IRF6-DBD possesses specific, high affinity binding to the consensus sequence AACCGAAAC_T *in vitro*.

The majority of the mutations found in the IRF6 DNA-binding domain of VWS/PPS patients inhibit DNA binding

To determine the effect of VWS- and PPS-associated mutations that arise in the DNA-binding domain of IRF6, the ability of IRF6 mutant proteins to bind to the derived IRF6 DNA-binding site was investigated. Specific disease-

causing mutations were introduced into the IRF6-DBD and their ability to affect DNA binding was determined using both purified protein and *in vitro* translated protein; an equal amount of each mutant protein, as determined by Coomassie Blue staining or phosphorimaging, respectively, being incubated with radio-labelled DNA containing the IRF6 binding site (Fig. 3B). When compared to the wild-type protein, 12 of the 13 disease-causing mutations tested abrogated DNA binding; notably, the Gly70Arg mutation, which underlies VWS, had little effect on DNA binding (Fig. 3B). These findings suggest that the majority of mutations in VWS and PPS

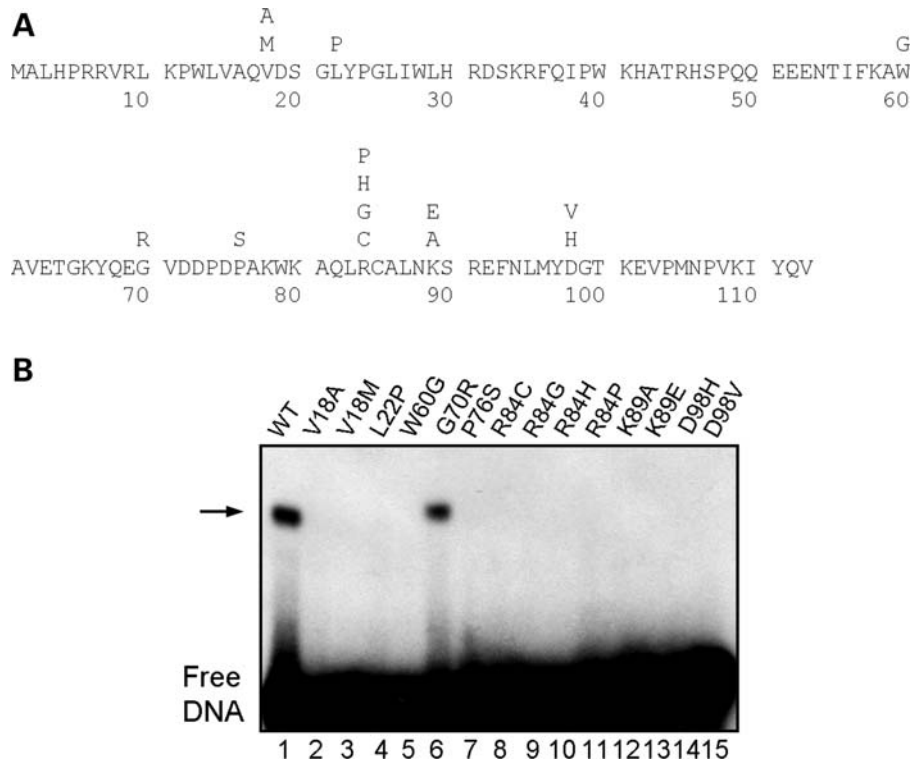


Figure 3. DNA binding of VWS and PPS mutants. (A) The amino acid sequence of the IRF6 DNA-binding domain; a subset of mutations found in VWS and PPS patients are shown above the wild-type residue. (B) EMSA showing DNA binding of the mutant IRF6-DBD proteins depicted in (A). *In vitro* translated proteins were incubated with the consensus sequence shown above the panel. Lane 1 contains the wild-type IRF-DBD protein (WT). The identity of each mutant protein is indicated above the lanes; V18A, V18M, G70R, P76S, R84G, R84P, D98H and D98V are VWS-causing mutations; L22P and W60G are PPS-causing mutations; R84C, R84H and K89E underlie both VWS and PPS.

patients exert their effect by inhibiting the DNA-binding function of IRF6; however, this appears not to be the only mechanism by which mutation of the IRF6 DNA-binding domain can cause disease since the Gly70Arg mutant binds DNA avidly.

The effect of missense mutations on the structure of the IRF6 DNA-binding domain

The distribution of missense mutations in the IRF6 DNA-binding domain differs significantly between PPS and VWS (8; unpublished data), suggesting that different mutations confer different effects on the function of IRF6. To establish whether there is a simple correlation between the disease phenotype and the extent to which a mutation disrupts the structure of the DNA-binding domain, we performed protein modelling and biochemical analyses of mutant isoforms. We performed protein modelling for the wild-type and Arg84 and Gly70 mutant isoforms of the DNA-binding domain. In wild-type IRF6, the side chain of Arg84 is involved in two hydrogen bonds with the DNA phosphate backbone (Fig. 4A). Mutating this arginine residue to glycine, proline, cysteine or histidine ablated these hydrogen bonds (Fig. 4A). In addition to the disruption of hydrogen bonding, proline, cysteine and glycine possess much smaller amino acid side-chains than arginine; therefore, the area of the DNA-protein interface was reduced with these mutations.

While the glycine, cysteine and histidine mutations were not likely to have a large effect on the tertiary structure of the DNA-binding domain, the Arg84Pro mutation has the potential to disrupt the secondary structure of helix three of IRF6. Proline residues are frequently associated with distortion of alpha-helices, as predicted above. Gly70 is not involved in the DNA-binding interface and is in loop region (L2) between helices 2 and 3 (Fig. 4B). Mutation of Gly70 to arginine made no obvious alterations to the protein structure although effects on inter- and intra-molecular interactions cannot be excluded (Fig. 4B).

As a direct test for differences in structural effects between PPS- and VWS-associated mutations, we performed circular dichroism (CD) analysis on mutant isoforms of the IRF6-DBD. Arg84Cys and Arg84His mutations are found in both VWS and PPS patients, whereas Arg84Pro and Arg84Gly occur in patients affected by VWS. The CD spectra of the Arg84Cys, Arg84His and Arg84Gly proteins were almost identical to that of the wild-type protein (Fig. 5). In contrast, the spectrum of the Arg84Pro mutant was consistent with a severely disrupted structure. To test whether the loss of structure in the Arg84Pro was due to inherent instability or simply an inability of this mutant isoform to refold *in vitro*, we performed CD analyses on a second set of IRF6-DBD proteins that were purified directly from bacteria without a denaturation/renaturation protocol. In this case, the CD spectra for

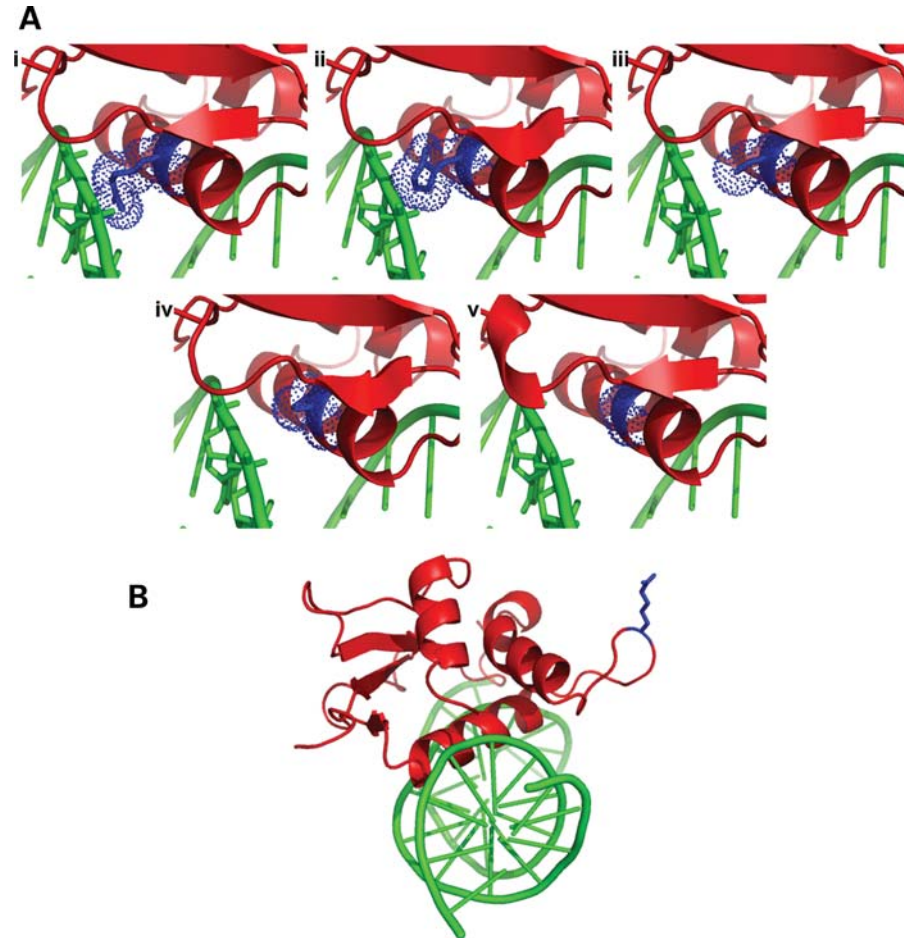


Figure 4. (A) Homology model of IRF6-DBD shown in red cartoon representation revealing a close-up of helix 3 after energy minimization. The DNA is shown in green. Panel (i) shows the position of Arginine 84 in blue; panels (ii), (iii), (iv) and (v) show the mutations R84H, R84C, R84P and R84G, respectively. In each case, the Van der Waals surface around residue 84 is indicated with blue dots. (B) The position of the G70R mutation is shown, highlighting the distance of this residue from the DNA. Colour scheme as in (A).

all four Arg84 missense mutations were nearly identical to wild-type (data not shown). To test for subtle differences in stability, we performed a thermal titration curve for each. The inflection points for the four mutants were similar, and all were slightly higher than the wild-type (Supplementary Material, Fig. S1), suggesting that all the mutant isoforms are more stable, not less stable, than the wild-type. These data suggest that the extent to which a mutation disrupts the overall structure of the IRF6 DNA-binding domain does not simply correlate with the disease phenotype. Intriguingly, the Gly70Arg mutation, which was the only mutation examined that does not abrogate DNA binding, also exhibited a similar spectrum to the wild-type protein (Fig. 5).

Identification of the transcriptional activation domain of IRF6

Initial experiments established that the expression of IRF6 in COS-7 cells did not activate a luciferase reporter plasmid containing five copies of the IRF6 binding site (data not shown). Subsequent analysis of the sub-cellular localization of an

EGFP-IRF6 fusion protein demonstrated that it resides in the cytoplasm (Supplementary Material, Fig. S2); this is analogous to the situation with IRF3, which is also cytoplasmic and requires viral infection to induce its nuclear translocation (27). Since we have so far been unable to determine the activation signal for IRF6, we used the GAL4-DBD reporter system to investigate the functional effects of disease-causing mutations within the protein-binding domain of IRF6. IRF6 was fused to the GAL4-DBD, and its ability to regulate transcription from the LEXA-VP16/GAL4 reporter plasmid was determined (Fig. 6A). The LEXA-VP16/GAL4 reporter construct has been used previously to assess the ability of a transcription factor to activate or repress transcription; it contains multiple GAL4 and LEXA binding sites upstream of the SV40 promoter. Transcriptional activation can be determined by co-transfection with a GAL4 fusion protein alone; however, if the GAL4 fusion protein is a repressor then this can be assessed by activating the reporter with the LEXA-VP16 activator. When the expression plasmid encoding the full-length IRF6 protein fused to the GAL4-DBD, GAL-IRF6-(1–467) was co-transfected with the luciferase reporter plasmid and activation was observed (Fig. 6B);

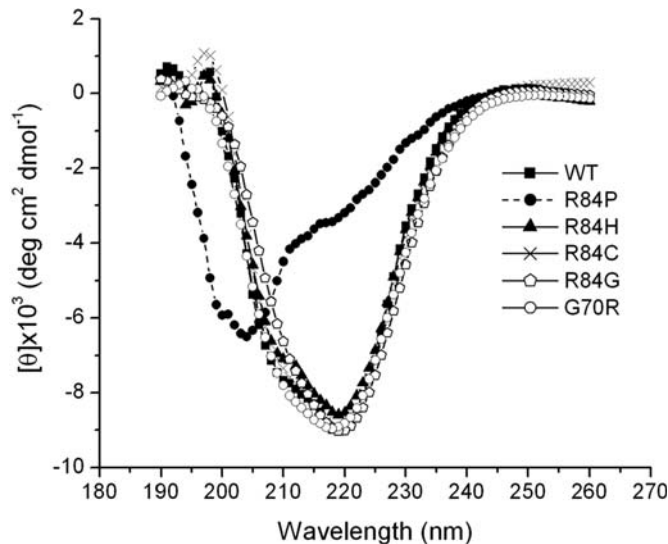


Figure 5. Circular dichroism spectra of purified recombinant wild-type (WT) and mutant IRF6 proteins. The identity of each mutant protein is indicated in the key.

however, GAL-IRF6-(1–467) did stimulate transcription in the presence of the LEXA-VP16 activator (Fig. 6B). These observations suggest that IRF6 acts as a co-operative transcriptional activator.

To map the activation domain of IRF6, a series of N-terminal deletions were generated and their ability to activate transcription determined (Fig. 6A and B). Deletion of the N-terminal region to residue 113 resulted in a 4-fold increase in transcriptional activation, while deletion to residue 226 resulted in greater than a 5-fold increase (Fig. 6B). Additional deletions into the protein-binding domain resulted in a reduction in activity. These data suggested that the transcriptional activation domain resides between residues 226 and 467; this was supported further by the observation that titration of increasing amounts of GAL-IRF6-(226–467) resulted in increased transcriptional activation (Fig. 6C).

Mutations found in the IRF6 transcriptional activation domain of VWS and PPS patients inhibit transcriptional activation

Numerous mutations have been found in the protein-binding domain of IRF6 in VWS patients. To determine the effect of a subset of these mutations on the transcriptional activation function of IRF6, specific mutations were introduced into GAL-IRF6-(226–467) and their ability to affect transcriptional activation was determined using a luciferase reporter assay (Fig. 7). Of the seven mutations tested, six (Arg250Gln, Arg250Gly, Leu294Pro, Cys374Arg and Gly376Arg) inhibited transcriptional activation completely, while one (Lys320Glu) stimulated activation above that of the wild-type (Fig. 7). The polymorphism Val274Ile, which has been demonstrated to be significantly associated with non-syndromic cleft lip and palate but which is not the disease-causing variant (14), had little effect on transcriptional activity.

DISCUSSION

The IRFs are a family of nine transcription factors that share a highly conserved, N-terminal, helix-turn-helix DNA-binding domain and a more variable protein-binding domain. The signature motif of the DNA-binding domain is a tryptophan repeat consisting of five residues spaced at 10–18 amino acid intervals (28). In the current study, we have used a site-selection assay to demonstrate that IRF6 binds to the core consensus sequence 5'-AACCGAAAC^{C/T}-3' which conforms to the IRF-E (5'-GAAAA^{G/T}/C^{G/T}GAAA^{G/T}/C^{G/T}-3'), ISRE (5'-^A/G^{NG}AAANNGAAACT-3') and minimal core (5'-AANNGAAA-3') sequences to which other IRF family members bind (29–31). Crystallization studies have shown that the DNA-binding domain of IRF1 has a helix-turn-helix motif that latches onto DNA through three of the five conserved tryptophan residues. The motif selects a short GAAA core sequence, binding to which is mediated by four amino acid residues; Arg82, Cys83, Asn86 and Ser87 (32); the equivalent residues in IRF6 are Arg84, Cys85, Asn88 and Lys89. Intriguingly, the substitution of serine by lysine at position 87 of IRF1, which is also observed in IRF4, IRF5, IRF8 and IRF9, confers the potential to reach over and mediate binding outside the GAAA core sequence (32,33). In the current study, we have demonstrated that the substitution K89S abrogates DNA binding (data not shown) confirming that IRF6 is a member of the same sub-group of IRFs as IRF4, IRF5, IRF8 and IRF9; consequently, the modelling studies reported here were based on the crystal structure of IRF4 rather than that of IRF1 (8). Interestingly, only the IRF6 construct containing amino acids 1–113 exhibited DNA binding. This situation mirrors that of IRF3 which uses an auto-inhibitory mechanism to suppress its transactivation potential, phosphorylation resulting in the alteration of the IRF3 structure which subsequently leads to unmasking of the hydrophobic active site and realignment of the DNA-binding domain for transcriptional activation (34).

Previous studies have demonstrated that mutations in IRF6 underlie VWS and PPS which are characterized by varying degrees of cleft lip, CLP, lip pits, skin-folds, syndactyly and oral adhesions (17–19). Murray and colleagues initially demonstrated that the distribution of mutations in IRF6 was non-random; for example, protein truncation mutations, while common in VWS, were rare in PPS (8). Similarly, while the missense mutations found in VWS were distributed between the DNA-binding domain and the protein-binding domain, those underlying PPS were predominantly found in the DNA-binding domain. In addition, the distribution of VWS/PPS mutations in the DNA-binding domain was skewed with residues that were predicted to contact DNA being mutated more commonly. Subsequently, these studies have been extended with 219 mutations having been identified in VWS families and 36 PPS-causing mutations being described (unpublished data). Comparison of the position and type of mutation with the clinical diagnosis in these families has supported the broad genotype–phenotype correlation; nevertheless, as outlined below, some exceptions to the general principles have been noted.

In the current study, 12 of the 13 DNA-binding domain mutations analysed were found to abrogate DNA binding,

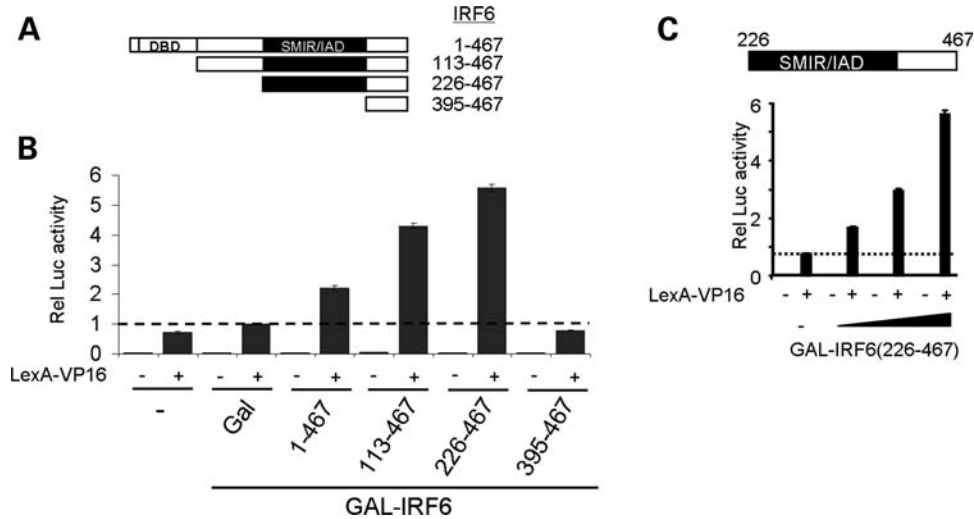


Figure 6. IRF6 activates transcription. (A) Wild-type IRF6 and a series of N-terminal deletions were fused to the C-terminus of the GAL4 DNA-binding domain as depicted. The DNA-binding and the protein interaction domains are indicated. The positions of the N- and C-terminal amino acids are shown for each of the IRF6 derivatives. (B) The graph shows activation of the luciferase reporter by the GAL-IRF6 proteins in the presence and absence of a fixed amount of LEXA-VP16 as indicated by + and -, respectively. The identity of the GAL fusion proteins in each transfection is indicated below the x-axis. All transfections were performed in triplicate; luciferase activities are presented as means with standard errors shown. All values are relative to the activity of the reporter plasmid alone. (C) Titration of increasing amounts of GAL-IRF6(226-467). The graph shows activation of the luciferase reporter by the GAL-IRF6(226-467) protein in the presence and absence of a fixed amount of LEXA-VP16 as indicated by + and -, respectively. 0, 10, 100 and 300 ng of GAL-IRF6(226-467) were transfected. All transfections were performed in triplicate; luciferase activities are presented as means with standard errors shown. All values are relative to the activity of the reporter plasmid alone.

the single exception being the mutation Gly70Arg which has been demonstrated to underlie VWS in two unrelated families and has not been reported to be a single nucleotide polymorphism (8). The CD spectra generated for the DNA-binding domain containing the Gly70Arg mutation were very similar to those of the wild-type polypeptide, indicating that no gross conformational change is induced. It is, therefore, possible that this particular mutation exerts its effect by disrupting the conformation of the full-length protein, perhaps via intra-molecular interactions between domains, or by affecting the interactions with a putative co-regulator.

The residue Arg84 is a clear mutational hotspot for PPS. Initially, mutations in Arg84, particularly Arg84Cys and Arg84His, which are predicted to result in a complete loss of contact with the core consensus sequence GAAA (8), were thought to cause solely PPS; however, recent evidence has shown that these mutations may also result in VWS (unpublished data). These combined results demonstrate that while the association between the Arg84Cys and Arg84His mutations and PPS is strong, it is not absolute. Moreover, VWS has also been shown to result from the mutations Arg84Gly and Arg84Pro (35; unpublished data). In the current study, we have demonstrated that all four mutations result in loss of DNA binding; consequently, the Arg84Gly and Arg84Pro mutations challenge the hypothesis that PPS results from a dominant negative mechanism via the formation of inactive transcription complexes (8). Nevertheless, a direct assessment of dominant-negative activity was not conducted in the current study and it is also possible that different mutations have template-specific effects. Despite these observations, it is notable that the residue Arg84 is located in the middle of helix 3 of the DNA-binding domain of IRF6 and that the amino acids proline and glycine are known to

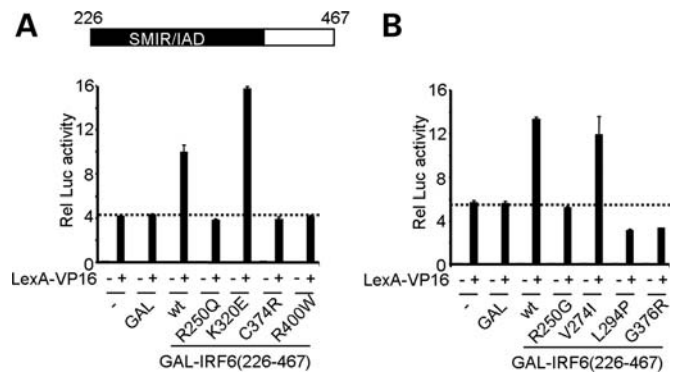


Figure 7. The effect of VWS and PPS mutations on the function of the IRF6 transactivation domain. (A and B) The graphs show activation of the luciferase reporter by the wild-type (WT) and mutant GAL-IRF6(226-467) proteins in the presence and absence of a fixed amount of LEXA-VP16 as indicated by + and -, respectively. The identity of each mutation is indicated below the x-axis. All transfections were performed in triplicate; luciferase activities are presented as means with standard errors shown. All values are relative to the activity of the reporter plasmid alone.

disrupt alpha helices (36). The tolerance of the Arg84Gly mutation in the alpha-helix, as demonstrated by CD analysis (Fig. 5), may be the consequence of the surrounding sequence; glycine is more readily tolerated in a helix than proline depending on the local sequence environment. In addition, longer helices having extensive intra-molecular hydrogen bonding networks may compensate for the tendency of glycine to break the helix. Conversely, proline as a helix terminating motif is far less ambiguous than glycine, and the Arg84Pro mutant DNA-binding domain was not able to refold; in particular, proline residues flanked by polar amino

acids have a very strong tendency to terminate helices. However, CD data generated with the native Arg84Pro mutant isoform showed CD spectra and thermal stability that were similar to wild-type. It appears that the stabilization energy when using a correctly folded domain as the starting point was too great to disrupt the domain *in silico* and *in vivo*. Consequently, if Arg84Pro and Arg84Gly mutations cause VWS by disrupting the secondary structure of IRF6 leading to a complete loss of IRF6 function, then that change is not detectable in the context of the DNA-binding domain alone; rather, future experiments must be designed to interrogate the effect of altered secondary structure in the context of the whole protein.

Despite these observations, the mutation L22P, which affects an amino acid residue that is not predicted to contact DNA, has been shown to result in VWS or PPS in members of the same family (37). These findings indicate that the genotype–phenotype correlation in man is not absolute and suggest that a clinical continuum exists with the precise phenotype being determined by a combination of the causative mutation, genetic background and stochastic factors. Intriguingly, this situation is mirrored in mice carrying either an Arg84Cys mutation, in which IRF6 protein while generated lacks the ability to bind DNA, or mice carrying a null mutation in *Irf6* (38,39). While in both lines, homozygous mutant mice display an identical phenotype comprising a hyperproliferative epidermis that fails to undergo terminal differentiation resulting in multiple soft tissue fusions, the phenotype of the *Irf6*^{+R84C} mice, which exhibited mild intra-oral adhesions, is more severe and displays greater penetrance than that reported for the *Irf6*-null mice, despite both mutations being on the same genetic background (38,39).

As a prelude to examining the effect of disease-causing mutations in the protein-binding domain of IRF6, we have analysed the ability of IRF6 to act as a transcriptional regulator. Our initial experiments demonstrated that IRF6 localized to the cytoplasm and, consequently, was unable to activate a luciferase reporter containing multiple copies of its consensus binding site. The cytoplasmic localization of IRF6 has been reported previously (40). This situation parallels that of IRF3 which also localizes in the cytoplasm as an inactive monomer maintained by auto-inhibitory domains that flank the protein-interaction domain until virus-induced phosphorylation converts IRF3 into an active monomer that enters the nucleus to activate transcription of its target genes (27,34). While we have not, as yet, identified the signal that results in phosphorylation of IRF6 with subsequent nuclear translocation, transforming growth factor β 3 is a potential candidate for this role, at least in the secondary palate where *Irf6* has been shown to be down-regulated in the medial edge epithelia of the developing palatal shelves in both *Tgfb3* and *Tgfb2* mutant mice (41,42). As a consequence of these findings, we used the GAL4-DBD reporter system to demonstrate that IRF6 acts as a co-operative transcriptional activator. In this study, we have used the minimal VP16 activation domain in a standard synergy assay; however, VP16 is unlikely to be a natural co-operative partner of IRF6 such partners remaining to be identified. This situation has also been described for the ETS family member PU.1 and IRF4 which co-operatively bind to composite elements found in the promoters and

enhancers of b-lymphoid and myeloid genes (43–45). Importantly, while all seven of the VWS-causing mutations analysed either decreased (Arg250Gln, Arg250Gly, Leu294Pro, Cys374Arg, Gly376Arg) or increased (Lys320Glu) the transcriptional activation function of IRF6, the neutral polymorphism Val274Ile had no effect on this activity. In the case of Lys320Glu, it is important to note that this mutation has been described in two unrelated families (8), the mutation arising as a *de novo* event in one of these families thereby supporting the hypothesis that Lys320Glu is pathogenic rather than a polymorphism.

One of the goals of our research is to dissect the molecular pathway in which IRF6 functions during development of the lip and palate. To date, a combination of human genetic analyses and developmental studies has led to the identification of two transcription factors which act upstream of IRF6. Recently, Murray and colleagues identified a common functional variant in an AP-2 α binding site within an *IRF6* enhancer that showed highly significant linkage disequilibrium with isolated cleft lip, but not with cleft lip and palate or isolated CLP (46). Electrophoretic mobility shift assays and chromatin immunoprecipitation subsequently demonstrated that the risk allele disrupts binding of AP-2 α to the *IRF6* enhancer (46). Importantly, *TFAP2A*, which encodes AP-2 α , maps to human chromosome 6p24 within a region where chromosomal anomalies have been associated with OFC (47,48) and mutation of *Tfap2a* in mice results in facial anomalies (49,50). Similarly, the transcription factor P63 has been demonstrated to bind to a consensus P53-response element within IRF6 while *Tp63* and *Irf6* interact genetically during development of the secondary palate (unpublished data). Although these studies have identified transcription factors that lie upstream of IRF6 during craniofacial development, the downstream targets of IRF6 remain uncharacterized; in this context, delineation of the consensus IRF6 binding site detailed in the current study will facilitate the identification of putative transcriptional targets of IRF6.

MATERIALS AND METHODS

Plasmid construction and site-directed mutagenesis

Fragments of *Irf6* were amplified from embryonic day 14 mouse cDNA using the primers listed in Supplementary Material, Table S1 and cloned into either pET-14b for electrophoretic mobility shift assays or pSG424 for luciferase assays. Disease-causing mutations were generated using the primers listed in Supplementary Material, Table S2 and the ‘Quick Change’ site-directed mutagenesis kit (Stratagene) according to manufacturer’s instructions. All constructs were verified using sequence analysis.

Electrophoretic mobility shift assays

The radio-labelled DNA was subjected to electrophoresis on a 10% non-denaturing polyacrylamide gel, excised and purified using a G-50 micro-column according to the manufacturer’s instructions (Probe Quant). The probe was incubated at room temperature for 30 min with IRF6 protein in the presence of binding buffer [20 mM Tris, pH 7.5; 50 mM NaCl;

1 mM EDTA; 5% glycerol; 5 mM dithiothreitol; 1 μ g bovine serum albumin and 100 ng poly(dI:dI)-poly(dI:dC)]. The DNA-protein complexes were resolved on a 5% polyacrylamide gel in $1 \times$ TBE for 3 h at 180 V. The resulting gels were fixed in 10% acetic acid/20% methanol solution for 20 min, dried and subjected to autoradiography.

Protein production

The coding region of the IRF6 DNA-binding domain (IRF6-DBD; amino acids 1–113) was amplified by PCR from mouse cDNA derived from embryonic day 14 using primer pairs 5'-GATCCATATGGCCCTCCACCCTCGAAGAG-3' and 5'-GATCGGATCCTCACACTTGATAGATCTT CACAGG-3'. The PCR amplified product was cloned into pET-14b in order to fuse a hexahistidine tag to the N-terminus of the expressed protein. The resultant plasmid, pET-IRF6-DBD, was used to transform *E. coli* BL21 cells and expression was induced using 1 mM IPTG for 3 h. The recombinant protein was expressed in the insoluble fraction and purified under denaturing conditions using the Ni-NTA spin kit according to the manufacturer's instructions (Qiagen). The purified protein was visualized by SDS-PAGE to determine the molecular size, protein integrity and yield. Protein refolding was achieved by dialysing 100 μ l of purified protein against solutions of decreasing denaturant using a 7000 MW Slide-A-Lyzer mini dialysis unit (Pierce). Dialysis was performed for 24 h against each of the following solutions: 2 M GuHCl, 50 mM HEPES, 500 mM KCl, 10 mM DTT; 1 M GuHCl, 50 mM HEPES, 500 mM KCl, 10 mM DTT; 0.5 M GuHCl, 50 mM HEPES, 500 mM KCl, 10 mM DTT. Final dialysis was against 50 mM HEPES, 500 mM KCl, 10 mM DTT for 12 h. *In vitro* translated IRF6 proteins were generated using a TNT kit and T7 polymerase according to the manufacturer's instructions (Promega).

Determination of IRF6 binding sequence

The putative core binding sequence for IRF members is proposed to be AANNAAAA (31). Three IRF consensus binding sites containing this sequence were synthesized; ISRE1: GAAANNAAAA; IRF-E: GAAA^{G/C}_TGAAA^{G/T}_C and ISRE3: ^{A/G}_NGAAANNAAAACT (29,30) and used in electrophoretic mobility shift assays with the His-tagged IRF6-DBD. Subsequently, all 16 possible sequence combinations of the IRF-E site were synthesized and tested for their ability to bind to IRF6-DBD. The consensus DNA-binding sequence of IRF6 was determined as described previously (51). DNA isolated from the final round of selection was ligated into the pDrive cloning vector using the PCR Cloning^{plus} kit (Qiagen) according to the manufacturer's instructions. Plasmids containing the 76 bp insert were sequenced and aligned using WebLogo (52).

Circular dichroism analysis

Far-UV CD spectra were obtained using a Jasco J-810 spectropolarimeter equipped with a temperature-controlled quartz cell with a path-length of 50 mm. Spectra were obtained between 190 and 260 nm at 20°C obtaining an average of eight scans

for IRF6-DBD and mutants at concentrations ranging from 0.4 to 0.6 mg/ml. Buffer scans were subtracted, and concentrations normalized before analysis of the data using the CDSSTR deconvolution algorithm in the software suite Dichroweb (53).

Protein modelling

The structure of the DNA-binding domain from human IRF4 (33) was used as the starting point for homology modelling of the IRF6 DNA-binding domain. All molecular modelling was performed on a Linux workstation using the Quanta2005 program (Accelrys Ltd). The model of the IRF6 DNA-binding domain was built based on the co-ordinates of the IRF4 DNA-binding domain. The model was energy minimized using steepest descents followed by the conjugate gradient algorithm to convergence, removing bad steric and electrostatic contacts. The Protein Health module was used to check the integrity of the model using a Ramachandran plot, and to identify buried hydrophilic or exposed hydrophobic residues and close contacts. Point mutations corresponding to Arg84Cys, Arg84His, Arg84Pro, Arg84Gly and Gly70Arg were inserted into the IRF6 structure and molecular dynamics performed using CHARMM.

Luciferase assays

COS-7 cells were maintained in DMEM containing 10% foetal bovine serum. Once the cells had reached 80% confluence, they were transfected with 1 μ g DNA (total) using Lipofectamine 2000 according to the manufacturer's instructions (Invitrogen). Briefly, 300 ng pSG424-IRF6 construct, 300 ng LEXA-VP16, 300 ng firefly luciferase reporter and 100 ng *Renilla* luciferase control pRL-CMV (Promega) were transfected for 24 h. The cells were then lysed and luciferase activity measured using the dual luciferase reporter assay system following manufacturer's instructions (Promega). All transfections were performed in triplicate and standard errors calculated. To ensure that appropriate IRF6 protein had been expressed, equal amounts of COS-7 cell lysates were subjected to electrophoresis on a 10% SDS-PAGE gel and subjected to western blotting. The resulting nitrocellulose membrane was incubated with a polyclonal anti-GAL4 antibody (Santa Cruz) followed by an ECL peroxidase-labelled, anti-rabbit secondary antibody (Amersham Biosciences) for the detection of the IRF6-Gal4 complexes. Immunocomplexes were detected using the ECL plus Western Blotting Detection System according to the manufacturer's instructions (Amersham Biosciences) and exposed to x-ray film.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* Online.

ACKNOWLEDGEMENTS

The authors thank Anthony Waldschmidt and Phyllis Hemerson for technical assistance.

Conflict of Interest statement. The authors have no competing interests.

FUNDING

This work was supported by Wellcome Trust (082868) and the National Institutes of Health (P50-DE016215, DE13513).

REFERENCES

- Vanderas, A.P. (1987) Incidence of cleft lip, cleft palate, and cleft lip and palate among races: a review. *Cleft Palate J.*, **24**, 216–225.
- Murray, J.C., Daack-Hirsch, S., Buetow, K.H., Munger, R., Espina, L., Paglinawan, N., Villanueva, E., Rary, J., Magee, K. and Magee, W. (1997) Clinical and epidemiologic studies of cleft lip and palate in the Philippines. *Cleft Palate Craniofac. J.*, **34**, 7–10.
- Neilson, D.E., Brunger, J.W., Heeger, S., Bamshad, M. and Robin, N.H. (2002) Mixed clefting type in Rapp-Hodgkin syndrome. *Am. J. Med. Genet.*, **108**, 281–284.
- Jones, M.C. (1988) Etiology of facial clefts: prospective evaluation of 428 patients. *Cleft Palate J.*, **25**, 16–20.
- Suzuki, K., Hu, D., Bustos, T., Zlotogora, J., Richieri-Costa, A., Helms, J.A. and Spritz, R.A. (2000) Mutations of *PVRL1*, encoding a cell-cell adhesion molecule/herpesvirus receptor, in cleft lip/palate-ectodermal dysplasia. *Nat. Genet.*, **25**, 427–430.
- van den Boogaard, M.J., Dorland, M., Beemer, F.A. and van Amstel, H.K. (2000) *MSX1* mutation is associated with orofacial clefting and tooth agenesis in humans. *Nat. Genet.*, **24**, 342–343.
- Braybrook, C., Doudney, K., Marcano, A.C.B., Arnason, A., Bjornsson, A., Patton, M.A., Goodfellow, P.J., Moore, G.E. and Stanier, P. (2001) The T-box transcription factor gene *TBX22* is mutated in X-linked cleft palate and ankyloglossia. *Nat. Genet.*, **29**, 179–183.
- 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. et al. (2002) Mutations in *IRF6* cause Van der Woude and popliteal pterygium syndromes. *Nat. Genet.*, **32**, 285–289.
- Dode, C., Levilliers, J., Dupont, J.M., De Paepe, A., Le Du, N., Soussi-Yanicostas, N., Coimbra, R.S., Delmaghani, S., Compain-Nouaille, S., Baverel, F. et al. (2003) Loss-of-function mutations in *FGFR1* cause autosomal dominant Kallmann syndrome. *Nat. Genet.*, **33**, 463–465.
- Sozen, M.A., Suzuki, K., Tolarova, M.M., Bustos, T., Fernandez Iglesias, J.E. and Spritz, R.A. (2001) Mutation of *PVRL1* is associated with sporadic, non-syndromic cleft lip/palate in northern Venezuela. *Nat. Genet.*, **29**, 141–142.
- Suzuki, Y., Jezewski, P.A., Machida, J., Watanabe, Y., Shi, M., Cooper, M.E., Viet Le, T., Nguyen, T.D., Hai, H., Natsume, N. et al. (2004) In a Vietnamese population, *MSX1* variants contribute to cleft lip and palate. *Genet. Med.*, **6**, 117–125.
- Jezewski, P.A., Vieira, A.R., Nishimura, C., Ludwig, B., Johnson, M., O'Brien, S.E., Daack-Hirsch, S., Schultz, R.E., Weber, A., Nepomucena, B. et al. (2003) Complete sequencing shows a role for *MSX1* in non-syndromic cleft lip and palate. *J. Med. Genet.*, **40**, 399–407.
- Marcano, A.C., Doudney, K., Braybrook, C., Squires, R., Patton, M.A., Lees, M.M., Richieri-Costa, A., Lidral, A.C., Murray, J.C., Moore, G.E. and Stanier, P. (2004) *TBX22* mutations are a frequent cause of cleft palate. *J. Med. Genet.*, **41**, 68–74.
- Zuccherro, T.M., Cooper, M.E., Maher, B.S., Daack-Hirsch, S., Nepomuceno, B., Ribeiro, L., Caprau, D., Christensen, K., Suzuki, Y., Machida, J. et al. (2004) Interferon regulatory factor 6 (*IRF6*) gene variants and the risk of isolated cleft lip or palate. *N. Engl. J. Med.*, **351**, 769–780.
- Riley, B.M., Mansilla, M.A., Ma, J., Daack-Hirsch, S., Maher, B.S., Raffensperger, L.M., Russo, E.T., Vieira, A.R., Dode, C., Mohammadi, M. et al. (2007) Impaired FGF signalling contributes to cleft lip and palate. *Proc. Natl Acad. Sci. USA*, **104**, 4512–4517.
- Jugessur, A., Rahimov, F., Lie, R.T., Wilcox, A.J., Gjessing, H.K., Nilsen, R.M., Nguyen, T.T. and Murray, J.C. (2008) Genetic variants in *IRF6* and the risk of facial clefts: single-marker and haplotype-based analyses in a population-based case-control study of facial clefts in Norway. *Genet. Epidemiol.*, **32**, 413–424.
- Van der Woude, A. (1954) Fistula labii inferioris congenita and its association with cleft lip and palate. *Am. J. Hum. Genet.*, **6**, 244–256.
- Bixler, D., Poland, C. and Nance, W.E. (1973) Phenotypic variation in the popliteal pterygium syndrome. *Clin. Genet.*, **4**, 220–228.
- Froster-Iskenius, U.G. (1990) Popliteal pterygium syndrome. *J. Med. Genet.*, **27**, 320–326.
- Bocian, M. and Walker, A.P. (1987) Lip pits and deletion 1q32-41. *Am. J. Med. Genet.*, **26**, 437–443.
- Murray, J.C., Nishimura, D.Y., Buetow, K.H., Ardinger, H.H., Spence, M.A., Sparkes, R.S., Falk, R.E., Gardner, R.J., Harkness, E.M., Glinski, L.P. et al. (1990) Linkage of an autosomal dominant clefting syndrome (Van der Woude) to loci on chromosome 1q. *Am. J. Hum. Genet.*, **46**, 486–491.
- Sander, A., Schmelzle, R. and Murray, J. (1994) Evidence for a microdeletion in 1q32-41 involving the gene for Van der Woude syndrome. *Hum. Mol. Genet.*, **3**, 575–578.
- Lees, M.M., Winter, R.M., Malcolm, S., Saal, H.M. and Chitty, L. (1999) Popliteal pterygium syndrome: a clinical study of three families and report of linkage to the Van der Woude syndrome locus on 1q32. *J. Med. Genet.*, **36**, 888–892.
- Schutte, B.C., Sander, A., Malik, M. and Murray, J.C. (1996) Refinement of the Van der Woude gene location and construction of a 3.5-Mb YAC contig and STS map spanning the critical region in 1q32-q41. *Genomics*, **36**, 507–514.
- Schutte, B.C., Bjork, B.C., Coppage, K.B., Malik, M.I., Gregory, S.G., Scott, D.J., Brentzell, L.M., Watanabe, Y., Dixon, M.J. and Murray, J.C. (2000) A preliminary gene map for the Van der Woude syndrome critical region derived from 900 kb of genomic sequence at 1q32-q41. *Genome Res.*, **10**, 81–94.
- Schutte, B.C., Basart, A.M., Watanabe, Y., Laffin, J.J.S., Coppage, K., Bjork, B.C., Daack-Hirsch, S., Patil, S., Dixon, M.J. and Murray, J.C. (1999) Microdeletions at chromosome bands 1q32-q41 as a cause of Van der Woude syndrome. *Am. J. Med. Genet.*, **84**, 145–150.
- Lin, R., Mamane, Y. and Hiscott, J. (1999) Structural and functional analysis of interferon regulatory factor 3: localization of the transactivation and autoinhibitory domains. *Mol. Cell Biol.*, **19**, 2465–2474.
- Taniguchi, T., Ogasawara, K., Takaoka, A. and Tanaka, N. (2001) IRF family of transcription factors as regulators of host defense. *Annu. Rev. Immunol.*, **19**, 623–655.
- Tanaka, N., Kawakami, T. and Taniguchi, T. (1993) Recognition DNA sequences of interferon regulatory factor 1 (IRF-1) and IRF-2, regulators of cell growth and the interferon system. *Mol. Cell Biol.*, **13**, 4531–4538.
- Darnell, J.E. Jr, Kerr, I.M. and Stark, G.R. (1994) Jac-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science*, **264**, 1415–1420.
- Fujii, Y., Shimizu, T., Kusumoto, M., Kyogoku, Y., Taniguchi, T. and Hakoshima, T. (1999) Crystal structure of an IRF-DNA complex reveals novel DNA recognition and cooperative binding to a tandem repeat of core sequences. *EMBO J.*, **18**, 5028–5041.
- Escalante, C.R., Yie, J., Thanos, D. and Aggarwal, A.K. (1998) Structure of IRF-1 with bound DNA reveals determinants of interferon regulation. *Nature*, **391**, 103–106.
- Escalante, C.R., Brass, A.L., Pongubala, J.M.R., Shatova, E., Shen, L., Singh, H. and Aggarwal, A.K. (2002) Crystal structure of PU.1/IRF-4/DNA ternary complex. *Mol. Cell*, **10**, 1097–1105.
- Qin, B.Y., Liu, C., Lam, S.S., Srinath, H., Delston, R., Correia, J.J., Derynck, R. and Lin, K. (2003) Crystal structure of IRF-3 reveals mechanism of autoinhibition and virus-induced phosphoactivation. *Nat. Struct. Biol.*, **10**, 913–921.
- Item, C.B., Turhani, D., Thurnher, D., Yerit, K., Sinko, K., Wittwer, G., Adeyemo, W.L., Frei, K., Erginel-Unaltuna, N., Watzinger, F. and Ewers, R. (2005) Van Der Woude syndrome: variable penetrance of a novel mutation (p.Arg 84Gly) of the *IRF6* gene in a Turkish family. *Int. J. Mol. Med.*, **15**, 247–251.
- Gunasekaran, K., Nagarajaram, H.A., Ramakrishnan, C. and Balaram, P. (1998) Stereochemical punctuation marks in protein structures: glycine and proline containing helix stop signals. *J. Mol. Biol.*, **275**, 917–932.
- Ghassibe, M., Revencu, N., Bayet, B., Gillerot, Y., Vanwijck, R., Verellen-Dumoulin, C. and Vikkula, M. (2004) Six families with van der Woude and/or popliteal pterygium syndrome: all with a mutation in the *IRF6* gene. *J. Med. Genet.*, **41**, e15.

38. Richardson, R.J., Dixon, J., Malhotra, S., Hardman, M.J., Knowles, L., Boot-Handford, R.P., Shore, P., Whitmarsh, A. and Dixon, M.J. (2006) IRF6 is a key determinant of the keratinocyte proliferation/differentiation switch. *Nat. Genet.*, **38**, 1329–1334.
39. Ingraham, C.R., Kinoshita, A., Kondo, S., Yang, B., Sajan, S., Trout, K.J., Malik, M.I., Dunnwald, M., Goudy, S.L., Lovett, M. *et al.* (2006) Abnormal skin, limb and craniofacial morphogenesis in mice deficient for interferon regulatory factor 6 (Irf6). *Nat. Genet.*, **38**, 1335–1340.
40. Bailey, C.M., Abbott, D.E., Margaryan, N.V., Khalkhali-Ellis, Z. and Hendrix, M.J. (2008) Interferon regulatory factor 6 promotes cell cycle arrest and is regulated by the proteasome in a cell cycle-dependent manner. *Mol. Cell Biol.*, **28**, 2235–2243.
41. Knight, A.S., Schutte, B.C., Jiang, R. and Dixon, M.J. (2006) Developmental expression analysis of the mouse and chick orthologues of *IRF6*; the gene mutated in Van der Woude syndrome. *Dev. Dyn.*, **235**, 1441–1447.
42. Xu, X., Han, J., Ito, Y., Bringas, P. Jr, Urata, M.M. and Chai, Y. (2006) Cell autonomous requirement for Tgfbr2 in the disappearance of the medial edge epithelium during palatal fusion. *Dev. Biol.*, **297**, 238–248.
43. Brass, A.L., Kehrl, E., Eisenbeis, C.F., Storb, U. and Singh, H. (1996) Pip, a lymphoid-restricted IRF, contains a regulatory domain that is important for autoinhibition and ternary complex formation with the Ets factor PU.1. *Genes Dev.*, **10**, 2335–2347.
44. Brass, A.L., Zhu, A.Q. and Singh, H. (1999) Assembly requirements of PU.1-Pip (IRF-4) activator complexes: inhibiting function in vivo using fused dimers. *EMBO J.*, **18**, 977–991.
45. Marecki, S., Riendeau, C.J., Liang, M.D. and Fenton, M.J. (2001) PU.1 and multiple IFN regulatory factor proteins synergize to mediate transcriptional activation of the human IL-1 beta gene. *J. Immunol.*, **166**, 6829–6838.
46. Rahimov, F., Marazita, M.L., Visel, A., Cooper, M.E., Hitchler, M.J., Rubini, M., Domann, F.E., Govil, M., Christensen, K., Bille, C. *et al.* (2008) Disruption of an AP-2 α binding site in an *IRF6* enhancer is associated with cleft lip. *Nat. Genet.*, **40**, 1341–1347.
47. Davies, A.F., Stephens, R.J., Olavesen, M.G., Heather, L., Dixon, M.J., Magee, A., Flinter, F. and Ragoussis, J. (1995) Evidence of a locus for orofacial clefting on human chromosome 6p24 and STS content of the region. *Hum. Mol. Genet.*, **4**, 121–128.
48. Davies, A.F., Imaizumi, K., Mirza, G., Stephens, R.S., Kuroki, Y., Matsuno, M. and Ragoussis, J. (1998) Further evidence for the involvement of human chromosome 6p24 in the aetiology of orofacial clefting. *J. Med. Genet.*, **35**, 857–861.
49. Schorle, H., Meier, P., Buchert, M., Jaenisch, R. and Mitchell, P.J. (1996) Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature*, **381**, 235–238.
50. Zhang, J., Hagopian-Donaldson, S., Serbedzija, G., Elsemore, J., Plehn-Dujowich, D., McMahon, A.P., Flavell, R.A. and Williams, T. (1996) Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2. *Nature*, **381**, 238–241.
51. Shore, P. and Sharrocks, A.D. (1995) The ETS-domain transcription factors Elk-1 and SAP-1 exhibit differential DNA binding specificities. *Nucleic Acids Res.*, **23**, 4698–4706.
52. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
53. Whitmore, L. and Wallace, B.A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.*, **32**, W668–W673.