Novel Mutations in the *IRF6* Gene on the Background of Known Polymorphisms in Polish Patients with Orofacial Clefting

Agnieszka Charzewska[1], Ewa Obersztyn[1], Dorota Hoffman-Zacharska[1,2], Jacek Lenart[1], Jarosław Poznański[3], and Jerzy Bal[1]

[1]Department of Medical Genetics, Institute of Mother and Child, Warsaw, Poland

[2] Institute of Genetics and Biotechnology, Faculty of Biology, University of Warsaw, Poland

[3]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland

Corresponding author:

Agnieszka Charzewska Ph.D.

Kasprzaka 17 A, 01-211 Warsaw, Poland

Phone: +48 32 77 313

Fax: +48 32 77 200

E-mail: agnieszka.charzewska@imid.med.pl

Running title

IRF6 novel mutations in orofacial clefting

ABSTRACT

Objective: To examine the role of the *IRF6* mutations in Polish families with Van der Woude syndrome (VWS) and popliteal pterygium syndrome (PPS) and to determine the effect of *IRF6* single nucleotide polymorphisms (rs7552506, rs2013162, and rs2235375) in CL/P susceptibility.

Design: *IRF6* mutation screening was performed by direct sequencing of all coding exons of the gene and their flanking intronic regions. Cosegregation analysis was performed to establish the relation of SNPs and CL/P phenotypes.

Patients: We screened the *IRF6* gene in 8 families with clinical recognition of VWS and PPS.

Results: In 5 families we identified pathogenic mutations, all affecting the DNA-binding or the protein-binding domain of IRF6. Two of the mutations were novel – a missense mutation Arg31Thr and a small deletion Trp40Glyfs*23. In most cases we found also a haplotype of three SNPs - rs7552506, rs2013162, and rs2235375. The association of the SNPs and CL/P susceptibility had been previously published. The variants did not cosegregate with phenotype in examined families nor did they cosegregate with pathogenic mutations. The SNPs were not causative, due to their presence in unaffected family members.

Conclusions: Two novel mutations (Arg31Thr and Trp40Glyfs*23) in the *IRF6* gene were identified to be causative for VWS and PPS syndromes. In the present study no association between the SNPs - rs7552506, rs2013162, and rs2235375 and the CL/P phenotype was found. The hypothesis, whether the haplotype of the three SNPs was correlated with *IRF6* expression level, demands further investigation.

Key words: *IRF6*, *IRF6* polymorphisms, cleft lip and palate.

INTRODUCTION

Clefts of the lip and/or palate (CLP) are common birth defects of complex genetic/ environmental etiology affecting ~1 in 700 individuals worldwide. It is estimated that 70% cases of CLP occur as isolated entities, commonly termed non-syndromic CLP (NSCLP), with no other apparent cognitive or craniofacial structural abnormalities (Dixon et al., 2011). The remaining 30% cases of orofacial clefts are composed of a wide range of malformation syndromes, including over 500 Mendelian syndromes (www.omim.org). Such CLP cases are designated as syndromic, and Van der Woude syndrome (VWS, MIM# 119300) is among the most common, accounting for ~2% of all clefts (Kondo et al, 2002). VWS is characterized by clefts of the lip and/or palate (CL/P) and congenital lower lip pits (LP). It is one of two allelic autosomal dominant disorders caused by mutations in the *IRF6* gene. The second one, popliteal pterygium syndrome (PPS, MIM# 119500) is characterized by the clinical features of VWS (CL/P and LP), with additional symptoms that include webbing of skin behind the knees (pterygia), genital anomalies, syndactyly, and other anomalies (Leslie et al., 2013). However, mutations in *IRF6* have been also shown to cause a nonsyndromic, isolated form of clefting (Rutledge et al., 2010).

The interferon regulatory factor-6 gene (*IRF6*, OMIM# 607199), located in chromosome 1q32.2, belongs to a family of nine transcription factors that share a highly-conserved helix-turn-helix (HTH) DNA-binding domain and a less conserved protein-binding domain. Most *IRF*s regulate the expression of interferon - alpha and - beta after viral infection. Unlike the other *IRF*s, the *IRF6* gene is involved in orofacial development, especially in cell adhesion and fusion during lip and palate formation. More precisely, the gene is a key determinant of the keratinocyte proliferation–differentiation switch, and has a crucial role in the formation of oral periderm, spatiotemporal regulation of which is essential for ensuring appropriate palatal adhesion (Dixon et al., 2011). Mutations in *IRF6* causing

haploinsufficiency of the gene have been shown to disrupt orofacial development leading to lip and/or palate clefting (Kondo et al., 2002). Common alleles of *IRF6* (Wu-Chou et al., 2013), as well as variants located within the *IRF6* enhancer (Rahimov et al., 2008), were also associated with CLP. The aim of this study was to examine the role of the *IRF6* mutations in Polish patients with CLP and to determine the effect of *IRF6* common variants in CLP susceptibility.

MATERIALS AND METHODS

We screened the *IRF6* gene in 8 families with clinical recognition of VWS and PPS. Genomic DNA was extracted from blood leukocytes using standard techniques. All DNA samples were obtained with participants' written informed consent. All coding exons of *IRF6* (exons 3-9) and their flanking intronic regions were amplified with primers (Table 1) designed using the Primer 3 software (http://frodo.wi.mit.edu). PCR amplifications were performed with 100 ng of genomic DNA template using Taq DNA Polymerase (Thermo Fisher Scientific). The PCR products were purified using ExoSAP-IT (Affymetrix) and then directly sequenced on both strands. The sequences were analyzed using the Mutation Surveyor software (Softgenetics, LLC) in comparison to the reference sequence NM_006147.2 (NCBI RefSeq; http://www.ncbi.nlm.nih.gov/nuccore). *In silico* analysis was performed using the PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2), MutPred (http://mutpred.mutdb.org) and Mutation Taster (http://www.mutationtaster.org) methods. The SNP frequencies were determined using the HapMap Project web site (http://hapmap.ncbi.nlm.nih.gov). To exclude the possibility of the newly identified mutations being common variants, their existence in public databases was checked using dbSNP – Short Genetic Variation (http://www.ncbi.nlm.nih.gov/SNP/), 1000 genomes - A Deep Catalog of Human Genetic Variation (http://browser.1000genomes.org/index.html), and EVS - Exome Variant Server (http://evs.gs.washington.edu/EVS/). The structural effect of the Arg31Thr

substitution was analyzed with the homology modeling of N-terminal 120 residue fragment of IRF6 using Yasara Structure software. Eight structures from Protein Data Bank (PDB) - 1IRF, 2IRF (Irf-2); 2PI0, 2O61, 2O6G, 3QU6, 1T2K (Irf-3); 2DLL (Irf- 4) and 3QU3 (Irf-7) were selected as the best templates for homology modeling, each of them covering over 90% of Irf-6 DNA binding domain (DBD). In order to assess possible modulation of protein-DNA interactions the modeled structure of Irf-6 was aligned with the known structure of the complex of Irf-3 with DNA (PDB record 2O6G) using Mustang algorithm implemented in Yasara Structure. The high homology between the template and target proteins (51 identical and 37 similar among 120 residues) resulted in the reasonable quality model of Irf-6 in complex with DNA duplex.

RESULTS

In 5 families from those referred to our laboratory we identified two heterozygous missense mutations and two small deletions, all affecting the DNA-binding or the protein-binding domain. The mutations we found were either known (Arg84Cys, Pro246Leufs*57) or novel (Arg31Thr, Trp40Glyfs*23).

In two families with PPS we found known mutation c. 250C>T, p. Arg84Cys affecting the DNA binding domain (DBD). In family 1 (Fig. 1) the mutation arose *de novo*. The proband was the only affected member of the family and the only one in whom the mutation was identified. He demonstrated unilateral cleft lip and palate, lip pits, cleft scrotum, partial lid synechiae, syndactyly, and popliteal pterygia. The second family with PPS (family 2, Fig. 1) exhibited the same Arg84Cys mutation. There were four affected members over three generations in this family. The mutation expression was variable – all of the patients demonstrated pterygia and syndactyly characteristic for PPS, while orofacial abnormalities ranged from none, through lip pits only, to unilateral cleft lip and palate - suggesting existence of other genetic modifiers in this family.

The other mutation we found (c. 737delC, p. Pro246Leufs*57) was located in the second functional domain of IRF6, which is engaged in protein binding. It is a known mutation that causes a frameshift and thus protein truncation. The mutation was found in a family with VWS (family 3) and there were four affected members over three generations (Fig. 1). The expression of the mutation was also variable, ranging from lip pits alone (father) to lip pits and unilateral cleft lip and palate (proband).

In the region coding DBD, in two families with VWS, we identified two novel mutations (p. Arg31Thr and p. Trp40Glyfs*23) that have not been reported in the literature so far. In family 4 (Fig. 1) we identified a mutation (c. 92G>C, p. Arg31Thr) located in exon 3 of the *IRF6* gene. There were two affected persons in this family – the mother, who demonstrated only lip pits, and her daughter, who had cleft palate. We identified the Arg31Thr mutation only in mother, because she hasn't given consent for molecular analysis in her underage child. However, *in silico* analysis (Mutation Taster, MutPred, PolyPhen-2) predicted the character of the mutation as disease causing. Additionally, the variant was absent in dbSNP, 1000 genomes and EVS databases, which excludes the possibility of it being a common one. To prove the variant as a pathogenic mutation we performed structural modeling of the N-terminal fragment of Irf-6 bound to DNA. The analysis demonstrates that Arg31 (Lys29 in Irf-3) is involved in salt bridging/hydrogen bonding with the Asp19 side chain (Asp17 in Irf-3). This interaction fixes the orientation of N-terminal helix1 Arg9-Gly21 (Arg7-Gly19 in irf-3) against the three-strand beta-sheet, and consequently fastens helix-turn-helix (HTH) DNA-binding motif (Fig. 2). It could be thus expected that substitution Arg31Thr will interfere with DNA binding.

The second novel mutation we found was also located in the region encoding DBD (exon 3). It was a small deletion (c. 117delC), which resulted in a frameshift (p. Trp40Glyfs*23) and thus protein truncation. The deletion was predicted as disease causing

(Mutation Taster) and it was absent in dbSNP, 1000 genomes and EVS databases, excluding the possibility of being a common variant. The mutation was found in family 5 (Fig. 1) and there were three affected persons over three generations. The expression of mutation was variable, ranging from lip pits in the mother (Fig. 3B), through lip pits and cleft palate in the grandfather, to the most severe phenotype in our cohort, in the proband, characterized by bilateral cleft of the lip and palate and lid synechiae (Fig. 3A).

In three patients (probands 6, 7, and 8) with recognition of VWS (CLP and lid synechiae), PPS (CL and popliteal pterygium of the left limb), and VWS (buccal synechiae and cleft of the secondary palate), respectively, we did not find any pathogenic mutations in the *IRF6* coding region. The patients had no familial history of oral clefts (sporadic cases).

Excepting for *IRF6* mutations, in most cases we found also three *IRF6* single nucleotide polymorphisms: SNP1 - c.175-5C>G (rs7552506), SNP2 - c.459G>T (rs2013162), and SNP3 - c. 667+27C>G (rs2235375). SNP 1 and 3 are intronic variants (intron 3-4 and intron 6-7, respectively), while SNP2 results in a silent polymorphism in exon 5, which does not change the protein product (Ser153Ser). We identified the SNPs in probands 6, 7, and 8, who did not demonstrate any *IRF6* mutations. The SNPs were also present in all the families with *IRF6* pathogenic mutations, but they located on a different allele than the mutation was. Inheritance and segregation of the SNPs in the examined families (Fig. 1) show that all the minor alleles have been inherited from the same parent. This allows us to assume the three SNPs as a haplotype.

The patient in family 1 with a *de novo* mutation was the only proband in whom the SNPs were not found. However, the SNPs were present in both his parents. We also identified the existence of the haplotype in family 2. The SNPs did not segregate with the disease phenotype, nor did they segregate with the Arg84Cys mutation - there was an affected person with mutation and without SNPs and there was an unaffected one with SNPs but without the

mutation. In this family it appears that the SNPs were insufficient to evoke the syndrome alone, but magnified the symptoms caused by the pathogenic Arg84Cys mutation. The patients with the mutation and SNPs demonstrated more severe symptoms than the patients with the mutation alone. It could have been speculated that the SNPs were genetic modifiers that can influence expression of *IRF6*, but this hypothesis requires further investigation. In family 3 in the proband and his healthy mother we identified the SNP haplotype as well. Like in the other families, the SNPs were not associated with the mutation. On the basis of existing data, it is impossible to exclude the hypothesis of magnification of the symptoms by the haplotype. Also in family 4, in the mother, we identified the haplotype, but because of the lack of molecular examination of her daughter, the hypothesis of SNP magnification can not be excluded. Finally, in family 5 we identified the same SNP haplotype. It was found in the proband and in both parents, but not in the affected grandfather. The SNP haplotype was not associated with the mutation and it did not segregate with the disease phenotype. The mother who had only the lip pits, as well as the daughter who had bilateral cleft lip and palate, demonstrated the same SNP haplotype. Thus, the SNPs cannot be assumed as genetic modifiers that differentiate disease symptoms. Moreover, the grandfather who had the mutation and no SNPs, demonstrated a more severe phenotype than the mother with both the mutation and the SNPs. On the basis of this family data, the hypothesis of SNP magnification must be excluded.

DISCUSSION

We found the *IRF6* gene as frequently mutated in the cohort of our families with CLP (0.63). In two families we found the Arg84Cys mutation, the most frequent substitution among the *IRF6* mutations. It localizes in a mutational hotspot attributed to a CpG dinucleotide in this codon, which undergoes methylation and spontaneous deamination of cytosine to thymine. Missense mutations at CpG are responsible for a half of PPS families and

the majority of these are Arg84Cys substitutions (Leslie et al., 2013). The mutation affects DBD, which is crucial in functioning of IRF6 as a transcription factor. The substitution results in a complete loss of contact with DNA, and thus makes DNA binding impossible (Kondo et al., 2002, Little et al., 2009).

We found also two novel *IRF6* mutations: Arg31Thr and Trp40Glyfs*23. The Arg31Thr substitution is located in the region encoding the DNA binding domain (amino acids 13-113) and it affects an amino acid that is highly conserved. According to structural modeling, which we performed, the mutation interferes with DNA binding. Thus, it disturbs the functioning of IRF6 as a transcription factor. This result, together with the absence of the variant in SNP databases, let us assume the Arg31Thr substitution as a pathogenic mutation leading to VWS phenotype.

The second novel mutation we found was a small deletion (c. 117delC), which resulted in a frameshift (p. Trp40Glyfs*23) and protein truncation as a consequence. According to the HGMD database (http://www.hgmd.org/), the majority of *IRF6* mutations (about 70%) are of missense type, whereas rearrangements are almost absent (2%). Small deletions, such as the one we identified, are relatively rare mutations among those found in *IRF6*. They account for about 14% of all *IRF6* mutations.

So far, about 300 different mutations in the *IRF6* gene have been described. Nonpathogenic, polymorphic variants in the *IRF6* gene have also been found to be significantly associated with oral clefting in many different populations and ethnic groups (Wu-Chou et al., 2013). Zucchero *et al.* (2004) found a higher risk of recurrence of NSCL/P for carriers of the C base (Val274) of rs2235371 polymorphism. They demonstrated that individuals heterozygous for this specific polymorphism (CT genotype) have a lower recurrence risk of CL/P than individuals homozygous for the C allele (CC genotype). In all the presented families, rs2235371 was seen as a monomorphic risk allele (CC genotype) and

can therefore contribute to the CLP phenotype, especially in patients, in whom no pathogenic *IRF6* variants have been found (probands 6, 7, and 8).

In the affected individuals we also found a haplotype of additional *IRF6* single nucleotide polymorphisms: SNP1 (rs7552506), SNP2 (rs2013162), and SNP3 (rs2235375). It was shown that the SNPs 1, 2, and 3 were associated with CLP. Scapoli et al. (2005) demonstrated the linkage disequilibrium between SNP2 and SNP3 and, in addition, an association with the NSCL/P phenotype in an Italian population. Pegelow et al. (2008) found an association between SNP1 and the CLP phenotype in a Swedish population. The extent of the contribution of the *IRF6* SNPs for oral clefting susceptibility in Polish patients is still unknown. The SNPs we found have never been tested for association in Polish patients. The SNPs are of high frequency, according to the HapMap Project (http://hapmap.ncbi.nlm.nih.gov/) each of the three SNPs accounts for equally 32% in the Italian population (which, among all the HapMap populations, is the closest to the Polish one) suggesting the existence of linkage disequilibrium between the variants. According to 1000 Genomes project, the minor allele frequencies of the three SNPs were also at the level of about 30% in five European subpopulations (Finnish, British, Iberian, Toscanini, and in Utah European ancestors). We identified the SNP haplotype in all the families that carried heterozygous *IRF6* mutations, but on a different allele. We also found the haplotype in three patients with CLP, who did not demonstrate any pathogenic *IRF6* mutations. The variants did not cosegregate with pathogenic mutations, nor with the phenotype in examined families. The SNPs were not causative, because they were present in unaffected family members. It was hypothesized that different *IRF6* polymorphic variants interfere in the proper expression of the gene in critical moments of embryonic craniofacial development and thus contribute to the occurrence of CLP (Brito et al., 2012). The expression of *IRF6* mutations found in the examined families was variable, suggesting the existence of other genetic determinants

modifying the phenotypic outcome. However, in the present study no association between the three SNPs and the CLP phenotype was found. The hypothesis, whether the haplotype of the three SNPs was correlated with *IRF6* transcriptional expression level, demands further investigation, especially functional analysis and association studies on a large cohort in the Polish population. The chance of incomplete penetrance should be also taken into account.

Recently, a new gene causing VWS, grainy-head-like 3 (*GRHL3*), has been discovered. This second gene causing VWS was found in a Finnish VWS sample, where pathogenic variants of *IRF6* have been excluded (Peyrard-Janvid et al., 2014). This finding will, in the future, contribute to the need of further investigation of the probands in our study excluded for *IRF6*.

CONCLUSIONS

The segregation of alleles found in the families with CLP does not confirm the hypothesis of symptom magnification by the haplotype of three *IRF6* SNPs (rs7552506, rs2013162, and 2235375). On the other hand, identifying of two novel mutations (p. Arg31Thr and p. Trp40Glyfs*23) will be important for proper genetic counseling, because CLP, due to an *IRF6* mutation, is not expected to be associated with serious medical complications, such as intellectual disability or developmental delay.

REFERENCES

Brito LA, Bassi CFS, Masotti C, Malcher C, Rocha KM, Schlesinger D, Bueno DF, Cruz LA, Barbara LK, Bertola DR, Meyer D, Franco D, Alonso N, Passos-Bueno MR. IRF6 is a risk factor for nonsyndromic cleft lip in the Brazilian population. Am J Med Genet Part A. 2012; 158A: 2170–2175.

Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: understanding genetic and environmental influences. Nat Rev Genet. 2011; 12: 167-178.

Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, Ferreira de Lima R, Daack-Hirsch S, Sander A, McDonald-McGinn DM, Zackai EH, Lammer EJ, Aylsworth AS, Ardinger HH, Lidral AC, Pober BR, Moreno L, Arcos-Burgos M, Palencia C, Houdayer C, Michel Bahuau15,16, Danilo Moretti-Ferreira5, Richieri-Costa A, Dixon MJ, Murray JC. Mutations in *IRF6* cause Van der Woude and popliteal pterygium syndromes. Nat Genet. 2002; 32: 285-289.

Leslie EJ, Standley J, Compton J, Bale S, Schutte BC, Murray JC. Comparative analysis of *IRF6* variants in families with Van der Woude syndrome and popliteal pterygium syndrome using public whole-exome databases. Genet Med. 2013; 15(5): 338–344.

Little HJ, Rorick NK, Su L, Baldock C, Malhotra S, Jowitt T, Gakhar L, Subramanian R, Schutte BC, Dixon MJ, Shore P. Missense mutations that cause Van der Woude syndrome and popliteal pterygium syndrome affect the DNA-binding and transcriptional activation functions of IRF6. Hum Mol Genet. 2009; 18 (3): 535–545.

Pegelow M, Peyrard-Janvid M, Zucchelli M, Fransson I, Larson O, Kere J, Larsson C, Karsten A. Familial non-syndromic cleft lip and palate--analysis of the IRF6 gene and clinical phenotypes. Eur J Orthod. 2008; 30(2):169-75.

Peyrard-Janvid M, Leslie EJ, Kousa YA, Smith TL, Dunnwald M, Magnusson M, Lentz BA, Unneberg P, Fransson I, Koillinen HK, Rautio J, Pegelow M, Karsten A, Basel-

Vanagaite L, Gordon W, Andersen B, Svensson T, Murray JC, Cornell RA, Kere J, Schutte BC. Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. Am J Hum Genet. 2014; 94(1):23-32

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

Rutledge KD, Barger C, Grant JH, Robin NH. *IRF6* mutations in mixed isolated familial clefting. Am J Med Genet Part A. 2010; 152A: 3107–3109.

Wu-Chou Y, Lo L, Chen K, Chang C, Chen Y. A combined targeted mutation analysis of IRF6 gene would be useful in the first screening of oral facial clefts. BMC Medical Genetics. 2013; 14: 37.

FIGURE LEGENDS

Fig.1

Pedigrees of examined families with orofacial clefting. uCLP, unilateral cleft lip and palate; bCLP, bilateral cleft lip and palate; LP, lip pits; LS, lid synechiae; GA, genital anomalies. The genotypes of all the patients available for molecular analyses are indicated in the pedigrees. Mutations of the *IRF6* gene are annotated at the protein level. Sequence chromatograms derived from the affected probands are shown below the pedigrees. SNPs placed in the box denote a haplotype of three single nucleotide polymorphisms in the *IRF6* gene: SNP1 - rs7552506, SNP2 - rs2013162, and SNP3 – 2235375.

Fig. 2

Structural modeling of the N-terminal fragment of Irf-6 (bottom) bound to DNA (top). Arg31 (right bottom corner) is involved in salt bridging/hydrogen bonding with the Asp19 side chain. This interaction fixes the orientation of N-terminal helix against the three-strand beta-sheet, and consequently fastens DNA-binding motif.

Fig. 3

Clinical variability in the two affected individuals in family 5. The proband (3A) presenting bilateral cleft of the lip and palate and lid synechiae, and the mother (3B) presenting lip pits only. The phenotype was evoked by a small deletion c. 117delC, p. Trp40Glyfs*23 in the *IRF6* gene.

Fig. 1

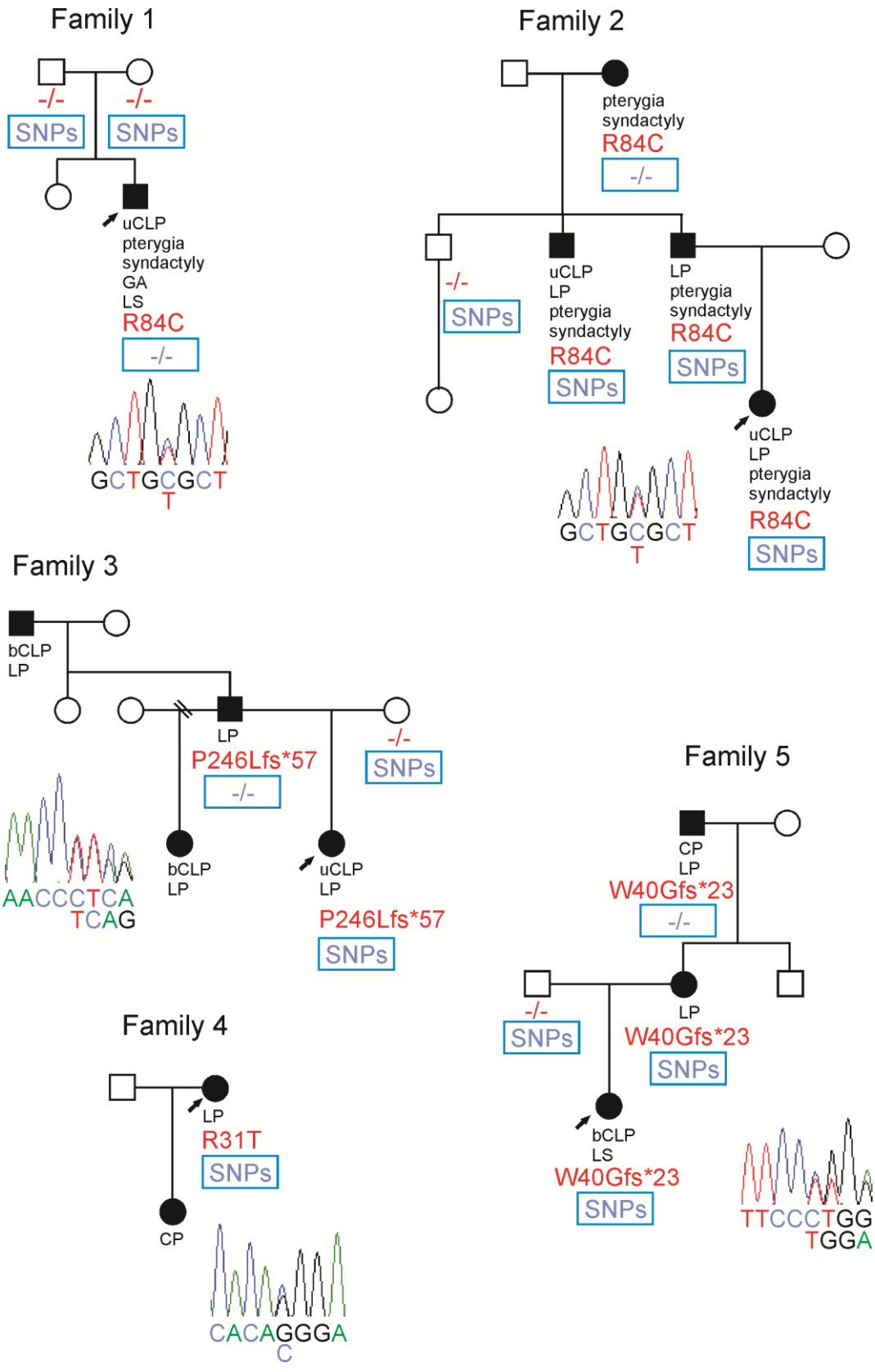
Family 1
-/-
SNPs
-/-
SNPs
uCLP
pterygia
syndactyly
GA
LS
R84C
-/-
GCTGCGCT
T
Family 2
pterygia
syndactyly
R84C
-/-
-/-
SNPs
uCLP
LP
pterygia
syndactyly
R84C
SNPs
LP
pterygia
syndactyly
R84C
SNPs
uCLP
LP
pterygia
syndactyly
R84C
SNPs
GCTGCGCT
T
Family 3
bCLP
LP
LP
P246Lfs*57
-/-
-/-
SNPs
bCLP
LP
uCLP
LP
P246Lfs*57
SNPs
AACCCTCA
TCAG
Family 4
LP
R31T
SNPs
CP
CACAGGGA
C
Family 5
CP
LP
W40Gfs*23
-/-
-/-
SNPs
LP
W40Gfs*23
SNPs
bCLP
LS
W40Gfs*23
SNPs
TTCCCTGG
TGGA

Fig. 2

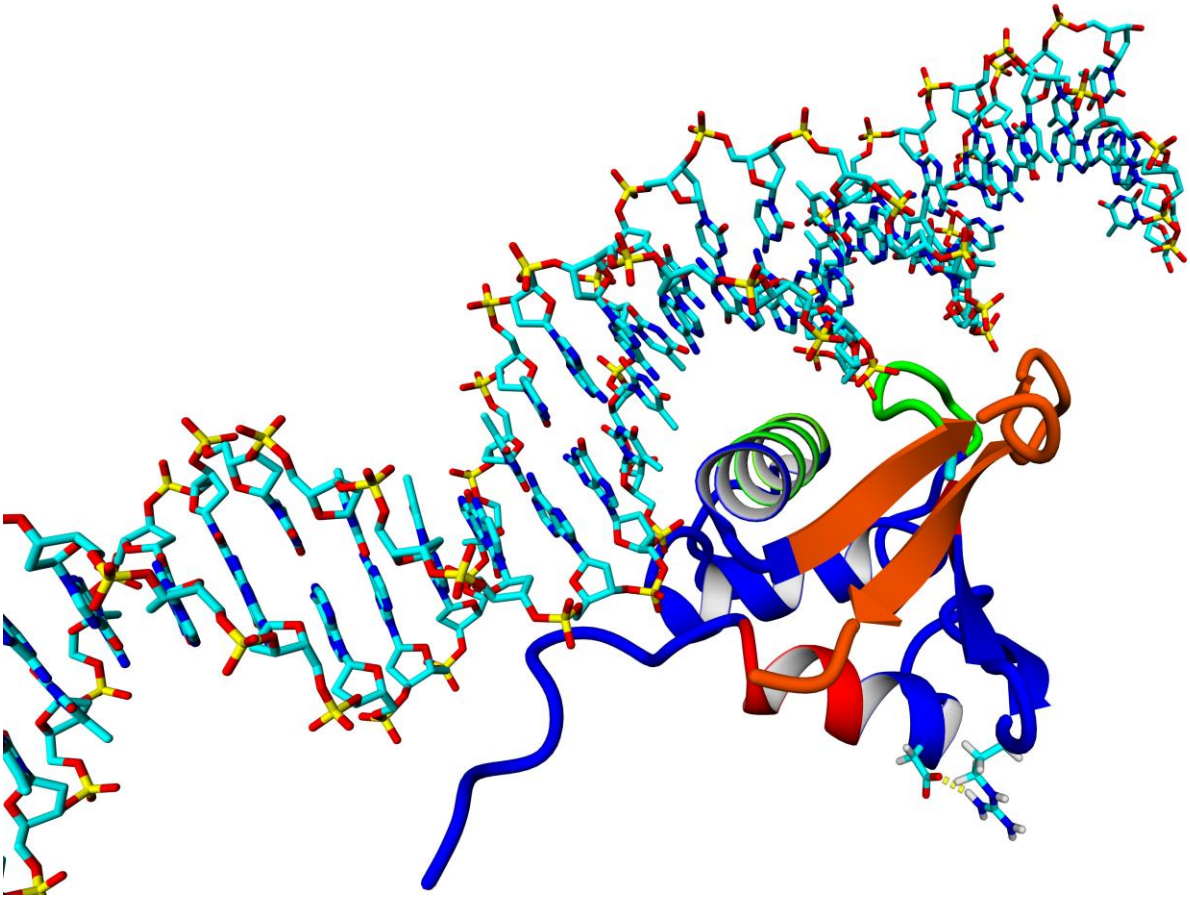

Fig. 3