ELECTRONIC LETTER

Genetic evidence for the role of loci at 19q13 in cleft lip and palate

A Warrington, A R Vieira*, K Christensen, I M Orioli, E E Castilla, P A Romitti, J C Murray*

J Med Genet 2006;**43**:e26 (<http://www.jmedgenet.com/cgi/content/full/43/6/e26>). doi: 10.1136/jmg.2005.034785

Background: Clefts of the lip and palate are common birth defects, affecting approximately 1 in 700 births worldwide. The aetiology of clefting is complex, with multiple genetic and environmental influences.

Methods: Genotype based linkage disequilibrium analysis was conducted using the family based association test (FBAT) and the likelihood ratio test (LRT). We also carried out direct sequencing of the PVR and PVRL2 candidate genes based on their homology to PVRL1, a gene shown previously to cause Margarita Island clefting. Participants included 434 patients with cleft lip with or without cleft palate or cleft palate only and their mothers from eight countries in South America, 205 nuclear triads (father-mother-affected child) from Iowa, 541 nuclear triads from Denmark, and 100 patients with cleft lip and palate from the Philippines.

Results: An allelic variant in the PVR gene showed statistically significant association with both South American and Iowa populations ($p=0.0007$ and $p=0.0009$, respectively). Direct sequencing of PVR and PVRL2 yielded 26 variants, including two rare amino acid changes, one in each gene, which were not seen in controls.

Conclusions: We found an association between a common variant in a gene at 19q and isolated clefting in two heterogeneous populations. However, it is unclear from our data if rare variants in PVR and PVRL2 are sufficient to cause clefting in isolation.

Clefts of the lip and palate are common birth defects, affecting approximately 1 in 700 births worldwide. The incidence rate is widely variable and is related to geographic origin.¹ Non-syndromic cleft lip with or without cleft palate (CL/P) includes clefts in the absence of any additional physical or cognitive deficits. The aetiology of clefting is complex, with multiple genetic and environmental influences.²

Multiple lines of evidence support a role for one or more genes at 19q13 in clefting. Study of a family in which CL/P segregated in two of three generations with a balanced translocation between 2q11.2 and 19q13.3 suggested that a gene in the vicinity of the identified breakpoint might play a role in clefting. CLPTM1 (cleft lip and palate transmembrane 1) was identified as being localised to this breakpoint.³ CLPTM1 is highly conserved and widely expressed, encoding a transmembrane protein with currently unknown function. Previous evidence has also implicated BCL3, located on 19q13, in the aetiology of non-syndromic clefting, although the data have been conflicting.⁴⁻⁶ Transmission disequilibrium tests (TDT) performed in a Brazilian population yielded marginal association for BCL3 as an allele of low penetrance or as a modifier locus.⁷ Similarly, TDT analyses performed on a sample from Maryland, USA suggested an

association between BCL3 and cleft lip probands.⁸ Study of Chilean populations also showed a slight difference in BCL3 allele distribution between non-syndromic cleft cases and controls.⁹⁻¹⁰ Finally, a Chinese population showed linkage and association between markers close to APOC2 (located between PVRL2 and TOMM40) and cleft lip and palate.¹¹⁻¹²

PVR and PVRL2, located on 19q13, have recently been added to the list of candidate genes hypothesised to play a role in the aetiology of non-syndromic cleft lip.¹³ PVRL2 was originally identified in the rodent as a gene homologous to the human poliovirus receptor (PVR) and named poliovirus receptor related 2. Although the PVRL2 cellular function is not yet known, it can be speculated that this receptor could act as a cell-cell adhesion molecule.¹⁴ PVRL2 might also act as a viral receptor, since some sensitivity to viruses of the *Picornaviridae* family has been assigned to chromosome 19.¹⁵ Mutations in PVRL1, a close relation to PVRL2, cause Margarita Island ectodermal dysplasia and clefting syndrome (CLPED1).¹⁶ Data suggest that heterozygotes for the common founder mutation in CLPED1 may also show an increased risk for non-syndromic clefting.¹⁷

The current study focuses on genetic contributions to clefting, first through linkage disequilibrium analysis and then through DNA sequencing of a subset of genes on 19q13.

METHODS

Subjects

Five separate populations were used for this analysis (table 1). All samples were collected with signed consent and had local and/or University of Iowa IRB approval.

The first of the populations was selected through the Latin American Collaborative Study of Congenital Malformations (ECLAMC). Established in 1967, ECLAMC utilises 70 different hospitals and volunteer physicians to collect data on births occurring in Latin America. From January 1998 to April 2002, ECLAMC collected blood spots on filter cards from 434 patients with CL/P or cleft palate only (CPO), and their mothers from eight countries in South America: Argentina, Brazil, Bolivia, Chile, Ecuador, Paraguay, Uruguay, and Venezuela. Patients known to have a syndrome or other major or multiple minor defects were excluded from the analysis.¹⁸

The second population consisted of 205 nuclear triads (father-mother-affected child) from the state of Iowa, USA. Blood specimens were collected from probands born between January 1987 and December 1991. The probands had CL/P and CPO.¹⁹ An additional 93 cleft lip and palate case samples from Iowa were selected for DNA sequencing.

Abbreviations: CEPH, Centre D'Etude du Polymorphisme Humain; CL/P, cleft lip with or without cleft palate; CPO, cleft palate only; ECLAMC, Latin American Collaborative Study of Congenital Malformations; ESEs, exonic splicing enhancers; FBAT, family based association test; LRT, likelihood ratio test; PVR, poliovirus receptor; SNP, single nucleotide polymorphism; TDT, transmission disequilibrium tests

Figure 1 19q13 region studied. PVR and PVRL2 gene structures are shown. Lines connecting green bars are introns. Hatched lines within introns indicate that introns are not drawn to scale with exons. Green bars depict coding regions. Yellow bars indicate untranslated regions. Blue arrows show the location of markers used in linkage disequilibrium studies. *Indicates PVR marker associated with clefts in this study. The larger pink arrows indicate the location of amino acid changes identified through direct sequencing. The purple bars over the plot are the locations of the genes in relation to the haplotype structure of the 19q13 region from the International HapMap Project. The dots denote the markers with respect to their physical location on the chromosome segment. The boxes represent the marker pair relationship and it is plotted between the two markers. The colour of the box (or the intensity of the colour) is based on the raw score for that marker pair. Bright red indicates $D' = 1$ and $LOD \geq 2$. Blue indicates $D' = 1$ and $LOD < 2$. Shades of pink/red indicate $D' < 1$ and $LOD \geq 2$. Finally white indicates $D' < 1$ and $LOD < 2$.

The third population was collected using cheek swabs from 147 triads (father-mother-affected child) between October 1999 and March 2000 according to a standard protocol. Volunteers were Danish CL/P and CPO families contacted through the University of Southern Denmark.

A fourth group consisted of 91 cleft lip and palate cases from the Philippines selected for DNA sequencing. Cases from the Philippines were studied under the auspices of Operation Smile International.²⁰ Blood samples were from

patients who were seen and examined at one of four sites within the Philippines (Cavite, Kalibo, Cebu City, and Naga City).

A fifth group consisted of controls used to verify frequencies of the variants found by direct sequencing. A total of 96 controls from the Philippines and 96 from Caucasian populations were studied. Filipino controls were collected as described above and at the same sites as the cases, and consisted of individual adults with no CL/P or

Figure 2 PVRL2 interspecies comparison. Colours are set by the variation in polarity and size of different amino acid chains, consistent with the Taylor colour scheme.²⁰ Grey bars indicate the conservation of amino acids between species at each site. No amino acids before position 40 were conserved among the species.

Table 1 Distribution of the cases by cleft type

	All	CLP	CLO	CPO	CL/P	CLP+CPO	Controls for sequencing	Controls for screening rare variants
ECLAMC*	434	241	114	69	355	310	–	–
Iowa	298	158	102	48	260	206	–	–
Denmark	536	225	170	141	395	366	–	–
Philippines	91	91	–	–	91	91	96	–
CEPH	–	–	–	–	–	–	96	1064

*For the ECLAMC population, 10 cases were syndromic or possessed an unknown cleft type and were excluded from the analysis. All, all types of clefts; CLP, cleft lip with palate; CLO, cleft lip only; CPO, cleft palate only; CL/P, cleft lip with or without cleft palate.

other recognised birth defect. Caucasian controls were samples from the Centre D'Etude du Polymorphisme Humain (CEPH).²¹ The CEPH Diversity Cell Line Panel,²² which is comprised of 1064 DNA samples from cultured lymphoblastoid cell lines derived from individuals representing 51 different human populations, was used to extend the search for point mutations not found in the first 96 individuals.

Cases with syndromic forms of clefting, as assessed by a clinical geneticist, were excluded from analysis.

DNA extraction

ECLAMC sample DNA was extracted from filter card blood spots using modifications of published protocols. Samples from Iowa, Denmark, and the Philippines were extracted from cheek swabs or whole blood according to published protocols.

Linkage disequilibrium studies

Pre-designed TaqMan assays²³ were used for pre-selected single nucleotide polymorphisms (SNPs) selected to have high minor allele frequencies in both Asian and Caucasian populations (table 2).

These assays spanned a 310 kb region along the q arm of chromosome 19 (fig 1). The assays were genotyped on ECLAMC, Iowa, and Danish populations. Genotype data were checked, using PedCheck, for non-Mendelian transmissions as signs of data error or allele loss that could suggest the presence of microdeletions.²⁴ Each population was evaluated separately. For Iowa and Danish populations, the family based association test (FBAT)²⁵ was performed. For the ECLAMC population, the likelihood ratio test (LRT) of Weinberg²⁶ was applied under the assumption that the distribution of paternal and maternal alleles was the same. The analysis was done subdividing cleft types into “cleft lip only”, “cleft lip with cleft palate”, and “cleft palate only” cases which were evaluated separately and then in combination (cleft lip with or without cleft palate, cleft lip with cleft palate plus cleft palate only, and all cases together). Parameters R_1 and R_2 and model likelihoods were estimated.²⁷

Table 2 Markers used for linkage disequilibrium

Marker public ID*	Gene
C_1828143_10	Poliovirus receptor (PVR)
C_1846428_10	Lutheran blood group (LU) – flanking BCL3
rs1871047	Poliovirus receptor-related 2 (PVRL2)
rs3745150	Poliovirus receptor-related 2 (PVRL2)
rs157580	Translocase of outer mitochondrial membrane 40 homolog (yeast) (TOMM40)
rs204911	Cleft lip and palate associated transmembrane protein (CLPTM1)

*The ABI assay ID was used to identify the marker in the absence of a public ID.

Direct sequencing

We selected PVR and PVRL2 for sequencing based on location and homology to PVRL1.¹⁶ PVR has eight exons while PVRL2 is composed of nine exons and has two isoforms, which were screened for mutations (fig 1) in 184 cleft lip and palate cases (93 from Iowa and 91 from the Philippines). Primers (see supplemental table available at <http://www.jmedgenet.com/supplemental>) were designed to amplify overlapping regions and conserved presumed regulatory regions outside of coding sequence.

Identified variants were confirmed through resequencing the sample and DNA from the parents, and sequencing in the reverse direction. We identified PVR and PVRL2 orthologues through a BLAST search of the non-redundant database using Homo Sapiens PVR, accession BC003091, and PVRL2, accession NP_002847.1, as reference sequences. All known Pvr and PvrL2 sequences were included from the vertebrate lineage. These files, in FASTA format, were then manipulated in Jalview and submitted for remote alignment at the European Bioinformatics Institute using a Clustal algorithm.²⁸ The multiple protein sequence alignment of the PvrL2 immunoglobulin domain from different species is shown in fig 2. It is coloured with the Taylor colour scheme used within Jalview. In this scheme, the colours are set by the variation in polarity and size of different amino acid chains. Small hydrophobic amino acids are coloured green, large aromatic amino acids are green/blue, large polar/basic amino acids are purple/blue, and small polar/acidic amino acids are orange/red. Jalview provides a visual assessment of the degree of amino acid consensus at any site as per the ClustalW algorithm. The ClustalW amino acid homology quality score provided from within the Jalview program is represented by the vertical bars under the sequence alignment of the PvrL2 putative orthologues. This qualitative quality score makes it possible to estimate the significance of a given amino acid substitution within the alignment. It measures both amino acid class conservation as well as evolutionary conservation at any given site.

RESULTS

Linkage disequilibrium

All marker genotypes were in Hardy-Weinberg equilibrium and none showed strong (D' values above 0.9) pairwise linkage disequilibrium (data not shown). Bonferroni correction was applied and p values below 0.0013 were considered significant (α 0.05/36 comparisons). The PVR marker C_1828143_10 showed association with oral clefts in both populations (ECLAMC: LRT p = 0.0007; Iowa: FBAT p = 0.0009) and was typed in an independent dataset from Danish samples, but this result was not statistically significant (FBAT p = 0.4416). However, a significant value was found for the non-affected siblings of the Danish cases (p < 0.01, data not shown). We tested this same marker on 32 CEPH control triads (father-mother-child) and did not see a significant overtransmission of the same allele (p > 0.25).

Table 3 describes the results by population for each marker. Further examination of this region showed that the PVR marker is not contained by any known linkage disequilibrium block (based on the International HapMap Project data available at <http://www.hapmap.org/>) in the 19q13 region (fig 1), which suggests that a possible mutation is in or close to the gene itself.

Direct sequencing

We found seven variants in the PVR gene, five common and two rare. The direct sequencing of PVRL2 resulted in the identification of 16 variants, five common and 11 rare (see Electronic-database information for accession numbers) (table 4).

Four coding variants were found in the PVR gene, however only the one in exon 6 (I340M) was not seen in matched controls. However, the mutation segregated from the unaffected father. The BLAST results showed that the exon 6 sequence of the PVR gene is not present in any other Pvr orthologues.

Because of the conservation in other species, we decided to test the rare variant in PVRL2, E41K, located in exon 2, that resulted in an amino acid change from a glutamic acid to a lysine in an extended panel of controls. Parent samples of the affected child were not available for study. This variant was not seen in 1150 control individuals from CEPH.^{21–22} The other rare missense mutation (A355T) was found in an unaffected control and may be a rare polymorphism. The Jalview analysis indicated that the human PVRL2 sequence is aligned with the mouse, rat, green monkey, gorilla, velvet monkey, chimpanzee, brown capuchin monkey, squirrel monkey, lemur, rabbit, cow, and pig (see Electronic-database information for accession numbers) sequence using remote Clustal alignment from the European Bioinformatics Institute. The region of the gene containing the conserved sequence around the E41K mutation possesses an immunoglobulin domain according to the NCBI Conserved Domain Search.²⁹ The glutamic acid at position 41 is conserved between humans and mice, rats, cows, gorillas, green monkeys, chimpanzees, brown capuchin monkeys, and squirrel monkeys, but not in lemurs, dogs, cats, chickens, sheep, rabbits, or pigs.

Two bioinformatics experiments were performed for the E41K mutation in PVRL2, and I340M mutation in PVR, to provide further support for an aetiological role for these variants. ESEfinder software³⁰ was used to predict the presence of exonic splicing enhancers (ESEs), which may be present in most, if not all, exons.^{31–32} For the site of the

PVRL2 variant E41K, the wild type sequence appears to hold an exonic enhancer sequence CTACCCG, which is thought to serve as a binding site for a specific serine/arginine-rich protein, SRp40 (score = 3.613055; minimum threshold = 2.67). The mutated sequence destroys this exonic enhancer sequence. For the site of the PVR variant I340M, the mutant sequence appears to create a motif previously inexistent that could serve as a binding site for either serine/arginine-rich proteins SF2/ASF (sequence: GTCCCGT; score = 2.542469; minimum threshold = 1.956) or SRp55 (sequence: CATGTC; score = 3.144773; minimum threshold = 2.676). Finally, we used PolyPhen software, also available online, to predict the impact of the identified amino acid substitutions on the structure and function of the human protein.^{33–35} Both mutations were predicted to be benign.

DISCUSSION

Cleft lip and palate is a common birth defect with a multifactorial aetiology, with both genetic and environmental influences. Association studies have emerged as a useful tool in the investigation of multifactorial diseases, such as cleft lip and palate, as a way to focus on certain regions of interest in the genome. Our work provides further evidence that a locus in 19q contributes to isolated clefting in heterogeneous populations of European descent. A previously performed meta-analysis of 13 genome scans was suggestive for the role of 19q13 in clefting, which supports the current findings.³⁶ This meta-analysis combined datasets from multiple populations in a linkage analysis to detect cleft lip and palate locus candidates. In addition, study of a multiplex family in which CL/P segregated with a balanced translocation between 2q11.2 and 19q13.3 (disrupting CLPTM1) and association data from North and South American populations have implicated BCL3 in the aetiology of non-syndromic clefting.^{7–10}

We found an association between oral clefts and the PVR marker in two independent populations (that is, Iowa and South America) that remained significant even when corrected for multiple comparisons. Significant results were not seen in the Danish population, but the same PVR marker allele was more often transmitted in all populations studied.

Based on the linkage disequilibrium results, extending the mutation search to other genes in the locus or to intronic regions highly conserved and likely regulatory in the PVR gene, may disclose the specific aetiological variant for clefting suggested by the FBAT and LRT results. We did not find any

Table 3 Linkage disequilibrium results.

Marker public ID (gene)	Population	Minor allele frequency	Cleft types (p values)					
			All	CLP	CLO	CPO	CL/P	CLP+CPO
C_1828143_10 (PVR)	ECLAMC	0.11	0.0007	0.12	0.03	0.01	0.01	0.008
	Iowa	0.19	0.0009	0.08	0.007	0.02	0.006	0.02
	Danish	0.16	0.44	0.43	0.37	0.17	0.94	0.76
	Iowa+Danish	–	0.008	0.74	0.01	0.009	0.09	0.07
C_1846428_10 (LU/BCL3)	ECLAMC	0.16	0.63	0.54	0.03	0.28	0.37	0.94
	Iowa	0.28	0.01	0.02	0.02	0.08	0.01	0.08
rs1871047 (PVRL2)	ECLAMC	0.35	0.57	0.16	0.18	0.78	0.50	0.29
	Iowa	0.39	0.80	0.75	1.0	0.91	0.80	0.91
rs3745150 (PVRL2)	ECLAMC	0.48	0.60	0.85	0.26	0.85	0.61	0.80
	Iowa	0.41	0.59	0.53	0.46	0.57	0.45	0.57
rs157580 (TOMM40)	ECLAMC	0.27	0.009	0.14	0.17	0.15	0.05	0.05
	Iowa	0.40	0.93	0.56	0.90	0.58	0.90	0.43
rs204911 (CLPTM1)	ECLAMC	0.07	0.004	0.03	0.33	0.22	0.01	0.01
	Iowa	0.02	0.20	0.56	1.0	0.56	0.65	0.41

All, any clefts; CLP, cleft lip with cleft palate; CLO, cleft lip only; CPO, cleft palate only; CL/P, cleft lip with or without cleft palate.

ECLAMC data were analysed using the likelihood ratio test (LRT) of Weinberg.²² Iowa and Danish data were analysed using FBAT. Statistically significant results appear in bold, less common allele overtransmitted.

Table 4 PVRL2 and PVR variants identified in the cleft cases studied

Gene region	Variant location	Frequency of the rare allele		
		Iowa	Philippines	Public ID
PVRL2				
Exon 1	C78A (L26L)	0.007	0.0	
Intron 1	88+52 C>T	0.007	0.0	
	88+92 C>A	0.57	0.41	rs2306149
	89-209 T>C	0.52	0.42	rs419010
	89-105 T>C	0.52	0.07	rs394221
Exon 2	G121A (E41K)	0.008	0.0	
Intron 2	479-101 G>A	0.018	0.0	
	479-72 G>A	0.012	0.0	
Intron 3	776-75 C>A	0.1	0.04	rs11672399
Intron 4	893+43 G>A	0.28	0.42	rs393584
	893+176 G>A	0.25	0.29	rs2075642
Exon 6	G1044A (E348L)	0.06	0.006	rs283814
	C1063T (A355T)	0.0	0.011	
Intron 6	1197-20 A>T	0.085	0.15	
Intron 8	1260+6 G>A	0.0	0.012	
	1260+20 G>A	0.0	0.47	
PVR				
5' UTR	-57 G>T	0.39	0.23	
	-1 C>T	0.0	0.10	
Exon 2	G195A (A67Y)	0.12	0.07	
Exon 3	G512A (G171D)	0.06	0.0	
Exon 5	G902A (V302M)	0.01	0.0	
Exon 6	A1017G (I340M)	0.01	0.0	
Intron 6	1152+30 A>C	0.49	0.0	

variants of obvious functional significance that correlate to the linkage disequilibrium results, which make us hypothesise that “the PVR mutation” is yet to be found. Previous evidence for aetiological variants 19q13 includes evidence for the BCL3 and CLPTM1 genes suggesting that a regulatory element in this region in linkage disequilibrium may be affected. Multi-species comparisons³⁷ are affording opportunities to identify such regions informatively. We provided statistical support that the E41K mutation is aetiological by testing a large number of controls. Our multi-species comparison also showed that the position 41 is strongly conserved and a change from an acidic charged polar to a basic charged polar amino acid potentially disrupts PVRL2 function. The deletion of a possible exonic enhancer with the E41K mutation may also contribute to loss of function in PVRL2. Exonic enhancers are thought to serve as binding sites for specific serine/arginine-rich proteins, a family of structurally related and highly conserved splicing factors characterised by one or two RNA-recognition motifs and by a distinctive C-terminal domain highly enriched in arginine/serine dipeptides (RS domain). However, PolyPhen did not predict any damaging consequence for this mutation. It is difficult to draw any conclusions from the I340M mutation analysis, but the presence of a new exonic enhancer with the mutation could also alter splicing.

Further examination of the 19q13 region shows that the PVR marker is not contained within any described linkage disequilibrium block (fig 1), which suggests that a possible mutation is in or close to the gene itself or awaits more definitive mapping when the blocks are defined at higher resolution. Our future effort will focus on intronic regions of PVR and the closest genes in the area: CEAL1 (carcinoembryonic antigen-like 1) and zinc finger proteins 180, 228, and 285 (ZNF180, ZNF228, and ZNF285). These studies add to our knowledge of the causes of cleft lip and palate and suggest genes and paths on 19q13 for additional investigations.

ACKNOWLEDGEMENTS

The authors thank all families for enthusiastically participating in this project. We are also indebted to Sandra Daack-Hirsch and Buena Nepomuceno and her team for sample collection and to Carla Nishimura, Ross Avila, Diana Caprau, and Jane Kimani for technical support. Mary Marazita and Margaret Cooper provided support for the use of the LRT statistics. Peter Jezewski helped with the use of the Jalview program.

ELECTRONIC-DATABASE INFORMATION

The supplemental table is available at <http://www.imedgenet.com/supplemental>. PVRL2 sequence segments were deposited at the NCBI, accession numbers AY331688, AY331689, AY331690, AY331691, AY331692, AY331693, and AY331694. PVRL2 orthologs: NCBI accession numbers BAC35982.1 (mouse), XP_218427.1 (rat), BAA02136.1 (green monkey), BAC41509.1 (gorilla), BAC41503.1 (velvet monkey), BAC41513.1 (chimpanzee), BAC41505.1 (brown capuchin monkey), BAC41506.1 (squirrel monkey), BAC41712.1 (lemur), BAC41713.1 (rabbit), CN790540 (cow), and AF308632.1 (pig). ESEfinder software is available online at <http://rulai.cshl.edu/tools/ESE/>. PolyPhen software is available online at <http://www.bork.embl-heidelberg.de/PolyPhen/>. The International HapMap Project site is at <http://www.hapmap.org/>.

Authors' affiliations

A Warrington, A R Vieira, J C Murray, Department of Pediatrics, University of Iowa, Iowa City, IA, USA
K Christensen, Centre for the Prevention of Congenital Malformations, University of Southern Denmark, DK-5000 Odense C, Denmark
I M Orioli, Latin American Collaborative Study of Congenital Malformations (ECLAMC), Department of Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil
E E Castilla, Latin American Collaborative Study of Congenital Malformations (ECLAMC), Department of Genetics, Oswaldo Cruz Institute, Rio de Janeiro, RJ, Brazil
P A Romitti, Department of Epidemiology, University of Iowa, Iowa City, IA, USA

This work is supported by NIH Grants DE 08559, 5 D43 TW05503, and P50 DE016215 (JCM); R01 DE 11948 and the Egmont Foundation (KC); and The March of Dimes Foundation Grant FY02-212; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil 522059/96-1; Agencia Nacional de Promoción Científica y Tecnológica, Consejo Nacional de Investigaciones Científicas y Tecnológicas, Argentina; and Grant PICT 495 (EEC)

Competing interests: none declared

*These authors equally contributed to the organisation and development of this work

Correspondence to: Jeffrey C Murray, Department of Pediatrics, University of Iowa, 2182 ML, Iowa City, IA 52242, USA; jeff-murray@uiowa.edu

Revised version received 26 July 2005
 Accepted for publication 10 August 2005

REFERENCES

- Vanderas AP**. Incidence of cleft lip, cleft palate, and cleft lip and palate among races: a review. *Cleft Palate J* 1987;**24**:216–25.
- Murray JC**. Gene/environment causes of cleft lip and/or palate. *Clin Genet* 2002;**61**:248–56.
- Yoshiura K, Machida J, Daack-Hirsch S, Patil SR, Ashworth LK, Hecht JT, Murray JC**. Characterization of a novel gene disrupted by a balanced chromosomal translocation t(2;19)(q11.2;q13.3) in a family with cleft lip and palate. *Genomics* 1998;**54**:231–40.
- Stein J, Mulliken JB, Stal S, Gasser DL, Malcolm S, Winter R, Blanton SH, Amos C, Seemanova E, Hecht JT**. Nonsyndromic cleft lip with or without cleft palate: evidence of linkage to BCL3 in 17 multigenerational families. *Am J Hum Genet* 1995;**57**:257–72.
- Amos C, Gasser D, Hecht JT**. Nonsyndromic cleft lip with or without cleft palate: new BCL3 information. *Am J Hum Genet* 1996;**59**:743–4.

- 6 **Amos C**, Stein J, Mulliken JB, Stal S, Malcolm S, Winter R, Blanton SH, Seemanova E, Gasser DL, Hecht JT. Nonsyndromic cleft lip with or without cleft palate: erratum. *Am J Hum Genet* 1996;**59**:744.
- 7 **Gaspar DA**, Matioli SR, Pavanello RC, Araujo BC, Andre M, Steman S, Otto PA, Passos-Bueno MR. Evidence that BCL3 plays a role in the etiology of nonsyndromic oral clefts in Brazilian families. *Genet Epidemiol* 2002;**23**:364–74.
- 8 **Maestri NE**, Beaty TH, Hetmanski J, Smith EA, McIntosh I, Wyszynski DF, Liang KY, Duffy DL, VanderKolk C. Application of transmission disequilibrium tests to nonsyndromic oral clefts: including candidate genes and environmental exposures in the models. *Am J Med Genet* 1997;**73**:337–44.
- 9 **Carreño HZ**, Suazo JS, Paredes A, Sola JA, Valenzuela JB, Blanco RC. Association study of the nonsyndromic cleft lip/palate phenotype and microsatellite markers located in 6p, 17q and 19q. *Rev Med Chil* 2002;**130**:35–44.
- 10 **Blanco R**, Suazo J, Santos JL, Paredes M, Sung H, Carreño H, Jara L. Association between 10 microsatellite markers and nonsyndromic cleft lip palate in the Chilean population. *Cleft Palate Craniofac J* 2004;**41**:163–7.
- 11 **Marazita ML**, Field LL, Cooper ME, Tobias R, Maher BS, Peanchitlertkajorn S, Liu YE. Nonsyndromic cleft lip with or without cleft palate in China: assessment of candidate regions. *Cleft Palate Craniofac J* 2002;**39**:149–56.
- 12 **Marazita ML**, Field LL, Cooper ME, Tobias R, Maher BS, Peanchitlertkajorn S, Liu YE. Genome scan for loci involved in cleft lip with or without cleft palate, in Chinese multiplex families. *Am J Hum Genet* 2002;**71**:349–64.
- 13 **Murray JC**, Schutte BC. Cleft palate: players, pathways, and pursuits. *J Clin Invest* 2004;**113**:2–4.
- 14 **Eberlé F**, Dubreuil P, Mattei MG, Devillard E, Lopez M. The human PRR2 gene, related to the human poliovirus receptor gene (PVR), is the true homolog of the murine MPH gene. *Gene* 1995;**159**:267–72.
- 15 **Couillin P**, Huyghe F, Grisard M, Cong NV, Louis M, Hofmann-Radvanyi H, Boué A. Echovirus 6, 11, 19; coxsackie virus B3 sensitivities and poliovirus 1, II, III sensitivities are on chromosome 19 respectively on 19pter-q133 and 19q131-133. *Cytogenet Cell Genet* 1987;**46**:599.
- 16 **Suzuki K**, Hu D, Bustos T, Zlotogora J, Richieri-Costa A, Helms JA, Spritz RA. Mutations of PVRL1, encoding a cell-cell adhesion molecule/herpesvirus receptor, in cleft lip/palate-ectodermal dysplasia. *Nat Genet* 2000;**25**:427–9.
- 17 **Sözen M**, Suzuki K, Tolarova M, Bustos T, Iglesias J, Spritz RA. Nonsense mutation of PVRL1 (W185X) is associated with non-syndromic cleft lip/palate in northern Venezuela. *Nat Genet* 2001;**29**:141–2.
- 18 **Orioli IM**, Vieira AR, Amorim MR, Castilla EE. Feasibility of using molecular genetic tools in a congenital malformation registry. *Front Fetal Health* 2001;**3**:221.
- 19 **Romitti PA**, Munger RG, Murray JC, Daack-Hirsch S, Hanson JW, Burns TL. The effect of follow-up on limiting non-participation bias in genetic epidemiologic investigations. *Eur J Epidemiol* 1998;**63**:557–68.
- 20 **Murray JC**, Daack-Hirsch S, Buetow KH, Munger R, Espina L, Paglinawan N, Villanueva E, Rary J, Magee K, Magee W. Clinical and epidemiologic studies of cleft lip and palate in the Philippines. *Cleft Palate Craniofac J* 1997;**34**:7–10.
- 21 **Dausset J**, Cann H, Cohen D, Lathrop M, Lalouel JM, White R. Centre D'Etude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* 1990;**6**:575–7.
- 22 **Cann HM**, Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL. A human genome diversity cell line panel. *Science* 2002;**296**:261–2.
- 23 **Ranade K**, Chang MS, Ting CT, Pei D, Hsiao CF, Olivier M, Pesich R, Hebert J, Chen YD, Dzau VJ, Curb D, Olshen R, Risch N, Cox DR, Botstein D. High-throughput genotyping with single nucleotide polymorphisms. *Genome Res* 2001;**11**:1262–8.
- 24 **O'Connell JR**, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998;**63**:259–66.
- 25 **Horvath S**, Xu X, Laird N. The family based association test method: strategies for studying general genotype-phenotype associations. *Eur J Hum Gen* 2001;**9**:301–6.
- 26 **Weinberg CR**. Allowing for missing parents in genetic studies of case-parent triads. *Am J Hum Genet* 1999;**64**:1186–93.
- 27 **Vieira AR**, Orioli IM, Castilla EE, Cooper ME, Marazita ML, Murray JC. MSX1 and TGFB3 contribute to clefting in South America. *J Dent Res* 2003;**82**:289–92.
- 28 **Clamp M**, Cuff J, Searle SM, Barton GJ. The Jalview Java alignment editor. *Bioinformatics* 2004;**20**:426–7.
- 29 **Marchler-Bauer A**, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, Liebert CA, Liu C, Madej T, Marchler GH, Mazumder R, Nikolskaya AN, Panchenko AR, Rao BS, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Vasudevan S, Wang Y, Yamashita RA, Yin JJ, Bryant SH. CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* 2003;**31**:383–7.
- 30 **Cartegni L**, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acid Res* 2003;**31**:3568–71.
- 31 **Blencowe BJ**. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* 2000;**25**:106–10.
- 32 **Graveley BR**. Sorting out the complexity of SR protein functions. *RNA* 2000;**6**:1197–1211.
- 33 **Sunyaev S**, Ramensky V, Bork P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet* 2000;**16**:198–200.
- 34 **Sunyaev S**, Ramensky V, Koch I, Lathe III W, Kondrashov AS, Bork P. Prediction of deleterious human alleles. *Hum Mol Genet* 2001;**10**:591–7.
- 35 **Ramensky V**, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002;**30**:3894–3900.
- 36 **Marazita ML**, Murray JC, Lidral AC, Arcos-Burgos M, Cooper ME, Goldstein T, Maher BS, Daack-Hirsch S, Schultz R, Field LL, Liu YE, Prescott N, Malcolm S, Winter R, Ray A, Moreno L, Valencia C, Neiswanger K, Wyszynski DF, Bailey-Wilson JE, Albach-Hejazi H, Beaty TH, McIntosh I, Hetmanski JB, Tuncbilek G, Edwards M, Harkin L, Scott R, Roddick LG. Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32–35. *Am J Hum Genet* 2004;**75**:161–73.
- 37 **Thomas JW**, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, Blanchette M, Siepel AC, Thomas PJ, McDowell JC, Maskeri B, Hansen NF, Schwartz MS, Weber RJ, Kent WJ, Karolchik D, Bruen TC, Bevan R, Cutler DJ, Schwartz S, Elnitski L, Idol JR, Prasad AB, Lee-Lin SQ, Maduro VV, Summers TJ, Portnoy ME, Dietrich NL, Akhter N, Ayele K, Benjamin B, Cariaga K, Brinkley CP, Brooks SY, Granite S, Guan X, Gupta J, Haghighi P, Ho SL, Huang MC, Karlins E, Laric PL, Legaspi R, Lim MJ, Maduro QL, Masiello CA, Mastrian SD, McCloskey JC, Pearson R, Stantripop S, Tiongson EE, Tran JT, Tsurgeon C, Vogt JL, Walker MA, Wetherby KD, Wiggins LS, Young AC, Zhang LH, Osoegawa K, Zhu B, Zhao B, Shu CL, De Jong PJ, Lawrence CE, Smit AF, Chakravarti A, Haussler D, Green P, Miller W, Green, eds. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* 2003;**424**:788–93.