Short communication
Family-based genome-wide association study in Patagonia confirms the association of the *DMD* locus and cleft lip and palate

Poletta FA, de Carvalho FM, Fonseca RF, Montaner D, Dopazo J, Mereb JC, Moreira MAM, Seuanez HN, Vieira AR, Castilla EE, Orioli IM. Family-based genome-wide association study in Patagonia confirms the association of the DMD *locus and cleft lip and palate.*
Eur J Oral Sci 2015; 00: 000–000. © 2015 Eur J Oral Sci

The etiology of cleft lip with or without cleft palate (CL±P) is complex and heterogeneous, and multiple genetic and environmental factors are involved. Some candidate genes reported to be associated with oral clefts are located on the X chromosome. At least three genes causing X-linked syndromes [midline 1 (*MID1*), oral-facial-digital syndrome 1 (*OFD1*), and dystrophin (*DMD*)] were previously found to be associated with isolated CL±P. We attempted to confirm the role of X-linked genes in the etiology of isolated CL±P in a South American population through a family-based genome-wide scan. We studied 27 affected children and their mothers, from 26 families, in a Patagonian population with a high prevalence of CL±P. We conducted an exploratory analysis of the X chromosome to identify candidate regions associated with CL±P. Four genomic segments were identified, two of which showed a statistically significant association with CL±P. One is an 11-kb region of Xp21.1 containing the *DMD* gene, and the other is an intergenic region (8.7 kb; Xp11.4). Our results are consistent with recent data on the involvement of the *DMD* gene in the etiology of CL±P. The *MID1* and *OFD1* genes were not included in the four potential CL±P-associated X-chromosome genomic segments.

Fernando A. Poletta[1,2,3,4,*], Flávia M. de Carvalho[1,2,3,*], Renata F. Fonseca[1,2,*], David Montaner[5], Joaquin Dopazo[5], Juan C. Mereb[6], Miguel A. M. Moreira[7], Hector N. Seuanez[1,7], Alexandre R. Vieira[8], Eduardo E. Castilla[2,3,4], Iêda M. Orioli[1,2]

[1]Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro, Rio de Janeiro; [2]ECLAMC (Latin-American Collaborative Study of Congenital Malformations) at INAGEMP (National Institute of Population Medical Genetics), Rio de Janeiro; [3]ECLAMC at LEMC (Laboratory of Congenital Malformation Epidemiology), Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil; [4]ECLAMC at CEMIC (Center for Medical Education and Clinical Research) and CONICET (National Council for Scientific and Technical Investigation), Buenos Aires, Argentina; [5]CIPF (Computational Genomics Department, Centro de Investigación Príncipe Felipe); and CIBERER (Centro de Investigación Biomédica en Red de Enfermedades Raras), Valencia, Spain; [6](In memoriam), ECLAMC at Hospital Zonal El Bolsón, El Bolsón, Argentina; [7]Genetics Division, National Cancer Institute, Rio de Janeiro, Brazil; [8]Departments of Oral Biology and Pediatric Dentistry and Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA

*These authors contributed equally to this work.

Dr Iêda M. Orioli, Department of Genetics, Federal University of Rio de Janeiro, PO Box 68.011, 21941-617 Rio de Janeiro, Brazil

E-mail: orioli@centroin.com.br

Key words: case–mother duo; cleft lip with or without cleft palate; *DMD* gene; oral cleft; South America

Accepted for publication August 2015

Oral clefts (OCs), cleft lip with or without cleft palate (CL±P) and cleft palate (CP), are common congenital defects that require complex management involving long-term medical, psychological, nutritional, and dental interventions, as well as speech therapy (1). The prevalence of CL±P ranges from one in 300 to one in 2,500 live births (2). Approximately 70% of CL±P cases are non-syndromic (3).

It is important to identify genes involved in the complex etiologies of OCs. Among the large number of candidate genes identified, some are located on the X chromosome. An early genome-wide association study using microsatellite markers in 92 CL±P case-sib pairs from the UK identified the following nine loci, including

one on the X chromosome: 1p36, 2p13, 6p24–23, 6q25, 8q23–24, 11p12–q14, 12p11–q24, 16q22–24, and Xcen–q21 (4). Twenty-seven genes on the X chromosome are associated with syndromic presentations of OCs (summarized in Table S1).

Recently, PATEL *et al.* (5) conducted an investigation of 14,486 single nucleotide polymorphisms (SNPs) spanning the X chromosome and found an association between the dystrophin gene (*DMD*; Xp21.2–21.1) and CL±P, mainly in Asian populations, even after correction for multiple comparisons.

Here, we analyzed independent genomic segments on the X chromosome containing polymorphic markers that may contribute to CL±P, in a homogeneous

population living in Patagonia, Argentina, by using a family-based genome-wide association study.

Material and methods

Families

The study enrolled 27 CL±P-affected children, including two siblings, and their mothers, from 26 families identified by the Latin-American Collaborative Study of Congenital Malformations (ECLAMC) (6) in Patagonia, Argentina. This region and four others were identified in a previous ECLAMC epidemiologic study as representing a high-prevalence geographic cluster for CL±P (7). Amerindian ancestry was a risk factor for OCs in this Patagonian population (7).

The clinical characteristics of the CL±P probands included in this study are summarized in Table S2. Both siblings affected were male and presented with complete CL+P on their right side. The 26 CL±P probands included 15 sporadic cases and 11 cases with a positive family history. Individuals with dysmorphic syndromes that included CL±P as a clinical feature were excluded from this study, to restrict the analysis to isolated CL±P. Eight families were from an Amerindian Mapuche Reserve, named Lago Rosario, and 18 families were from neighboring localities (Esquel, El Bolsón, Bariloche, Los Menucos, Viedma, and Maquinchao) in Patagonia, Argentina. All families carried surnames of Amerindian origin (Mapuche ethnicity).

The study protocol and informed consent form were approved by the Comité de Ética en Investigación del Centro de Educación Médica e Investigaciones Clínicas (Dr Norberto Quirno) in Buenos Aires, Argentina (IRB-1745, IORG-0001315; approval number: #238) and by the Institutional Review Board of the University of Pittsburgh, Pittsburgh, PA, USA (approval number: FWA#00006790). Prospective study patients and their families were informed about the nature of the study, and all subjects provided written informed consent before enrollment.

Genotyping

All subjects were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA), which contains 906,600 SNP probes spanning the entire human genome. Genotypes were determined using the Birdseed v2 algorithm with a 0.1 confidence threshold (http://www.broadinstitute.org/mpg/birdsuite/birdseed.html). This algorithm allows multiple-chip analysis to estimate the signal intensity for both alleles of each SNP, fitting probe-specific effects to increase precision (8). The genotyping array contains 37,902 SNP probes from the X chromosome, corresponding to 76% coverage of control population (see Data S1).

Data processing and quality control

Target data from the X chromosome (37,902 SNPs) were extracted and processed for assessment and quality control (QC), as recommended for genome-wide association studies (9). After performing QC procedures, a total of 36,446 SNPs remained for the next stage of analysis. To obtain a set of 'independent' (non-redundant) markers, neighboring SNPs in strong linkage disequilibrium (LD) were removed from the analysis. A matrix of genetic correlation (r^2) was calculated as a measure of the LD between each pair of SNPs within a window of 50 SNPs. Then, one SNP from the pair was removed if the two SNPs were in strong LD ($r^2 > 0.8$), the window was shifted five SNPs forward, and the procedure was repeated to scan the entire X chromosome. From the initial 36,446 SNPs (after QC procedures), 14,102 were removed, leaving 22,344 non-redundant SNPs for the statistical analysis.

Statistical analysis

The 22,344 non-redundant SNPs were used as 'independent' markers along the X chromosome in an exploratory analysis to identify candidate genomic regions associated with CL±P. The X-linkage transmission/disequilibrium test (X-TDT) (10) was applied to these markers. The X-TDT uses the number of transmissions of a variant allele observed in *n* pairs of heterozygous mothers and their affected children. Under the null hypothesis of no linkage between marker and disease, *n* has a binomial distribution, with a mean of $n/2$ and variance of $n/4$. The test statistic is a *Z*-score with continuity correction, for which the null hypothesis is rejected if *Z* significantly departs from zero. To identify candidate regions of the X chromosome (genomic segments) associated with CL±P, a segmentation algorithm was applied to the data from the X-TDT of all SNPs. This algorithm works by scanning the entire X chromosome, identifying regions with two or more independent SNPs with nominally significant *P* values of <0.01 and a maximum physical distance between them of 250 kb. The association of each identified genomic segment with CL±P was then evaluated through a unique combined analysis adjusted for multiple markers. The sum of log(*P*) test (11) takes into account the information on all SNPs in a genomic segment, and is adjusted for multiple comparisons by treating multiple SNPs as a set. The statistical significance of the association of each genomic segment with CL±P was evaluated by calculating a single combined *P* value based on the random-permutation distribution of empirical values, rather than the theoretical distribution derived from parametric assumptions.

As 22,344 markers were available for the exploratory analysis and the sample size was relatively small for a genome-wide association study, the threshold *P*-value for each individual SNP needed to be set at $0.05/22,344 = 2.2\text{E-}6$, according to the Bonferroni correction. Therefore, 1 million (1.0E+6) permutations were performed to set up empirical *P* values for each genomic segment, thereby reducing potential false-positive associations.

Data processing, QC, and exploratory analyses were carried out using the PLINK toolset (12). Mendelian inconsistencies were identified and resolved using the Ped-Check program (13). Permutation analysis and the combined sum of log(*P*) tests were performed using scripts implemented in the R project (11, 14).

Results and discussion

Our results show an association between genomic segments on the X chromosome and CL±P. We performed the X-TDT on 22,344 independent SNPs on the X chromosome, and identified four genomic segments

with at least two SNPs significantly associated with CL±P (nominally significant *P*-value of <0.01) in the Patagonian population.

From the four genomic segments identified, only two were statistically significant when a combined analysis, adjusted for multiple markers, was applied (Table 1). One genomic segment was an 11-kb region of Xp21.1 containing the *DMD* gene and the other was an 8.7-kb intergenic region of Xp11.4. The two other genomic segments identified – 117.2 kb in Xp22.31 and 32.6 kb in Xp11.23 – included the family with sequence similarity 9, member A (*FAM9A*) and synapsin I (*SYN1*) plus complement factor properdin (*CFP*) genes, respectively, but did not reach statistical significance in the combined multiple-marker test. Furthermore, the p arm of the X chromosome, mainly Xp11.23–22.31, showed stronger signs of an association with CL±P than the q arm (Table 1).

Given the predicted association of X-linked genes with syndromic forms of OC and the gender differences among OC types, we tested whether genetic variation on the X chromosome could be relevant to the etiology of isolated CL±P. Of the three X-linked genes previously reported to be associated with X-linked syndromes and isolated CL±P [midline 1 (*MID1*) (15), oral-facial-digital syndrome 1 (*OFD1*) (16), and *DMD* (5)], only the association between *DMD* and isolated CL±P was confirmed in the present study. Using the Illumina Human610-Quad platform, which includes 14,486 X-chromosome markers, PATEL *et al*. (5) identified six SNPs (rs5928207, rs5928208, rs6631759, rs5971698, rs5928214, and rs5972815) in *DMD* that were associated with OC in European and Asian populations; when the subjects were stratified by ethnicity, stronger signals were obtained for the Asian population. Using a different type of analysis, we also found an association between *DMD* and CL±P. We identified an 11-kb genomic segment containing four informative SNPs in *DMD*, which is associated with CL±P in the Patagonian population.

The *DMD* gene, which spans 2.4 Mb, is the largest found in nature. *DMD* was identified through a positional-cloning approach targeted at isolating the genes responsible for Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) (17). Duchenne muscular dystrophy is a recessive, fatal, X-linked disorder, occurring at a frequency of approximately 1 in 3,500 newborn males; BMD is a milder allelic form of muscular dystrophy. However, OCs have not been described in patients with DMD or BMD (18), and the role of the *DMD* gene in the etiology of OC has not yet been elucidated. PATEL *et al*. (5) argued against an actual role of the product of the *DMD* gene in the etiology of OC, and proposed that interactions might occur between distant SNPs within *DMD*. The authors demonstrated that the four SNPs of the *DMD* haplotype, significantly associated with OCs in the Asian population, were effectively uncorrelated with one another, and they argued against the possibility that all four SNPs were correlated with one high-risk allele of a single, as-yet-unidentified, causal gene (5). Nonetheless, we cannot dismiss the possibility that the *DMD* haplotype identified in the present study may be a chance finding, and that only one of the four SNPs is actually in LD with a causal variation. As the presence and extent of LD between any pair of SNPs vary from population to population (19), we must accept that other studies will find an association between OCs and SNPs in *DMD*, as our study did, but not necessarily for the same SNPs.

A link between *DMD* and OCs through the Walker–Warburg syndrome was in the MalaCards database, a graphical network of the top 20 diseases related to Walker–Warburg syndrome (http://www.malacards.org/card/walker_warburg_syndrome?search= walker-warburg+syndrome). Walker–Warburg syndrome is a congenital muscular dystrophy-dystroglycanopathy for which CL and CP have been described in some cases (20–22).

The other genomic segment that was significantly associated with CL±P in the present study is an 8.7-kb intergenic region (Xp11.4) that is unlikely to produce a functional protein. However, intergenic polymorphisms often occur in regulatory regions (such as gene promoters and enhancers), can ultimately affect gene function by affecting transcriptional or translational regulation, and may thus be associated with complex diseases (23). We hypothesize that the SNPs identified in this study

Table 1
X-chromosome genomic segments associated with cleft lip with or without cleft palate

Genomic segment (base-pair positions)*	Chromosome position	kb[†]	Informative SNPs (*n*)[‡]	Nominally significant SNPs at *P* < 0.05 (*n*)[§]	Gene symbol	Combined test (*P*-value)[¶]
X:8682950-8800121	Xp22.31	117.2	9	6	*FAM9A*	0.411
X:32586124-32597112	Xp.21.1	11.0	4	3	*DMD*	4.8E-04
X:39528180-39536876	Xp11.4	8.7	4	4	Intergenic region	0.020
X:47455346-47487916	Xp11.23	32.6	7	6	*SYN1*; *CFP*	0.432

CFP, complement factor properdin; *DMD*, dystrophin; *FAM9A*, family with sequence similarity 9, member A; *SYN1*, synapsin I.
*Start-to-end positions on the X chromosome (in base pairs).
[†]Genomic segment length (in kb).
[‡]Number of informative single nucleotide polymorphisms (SNPs) within the genomic segment.
[§]Number of nominally significant SNPs in the genomic segment (*P* < 0.05).
[¶]*P* values according to the combined sum of log(*P*) test adjusted for multiple markers.

may be in LD with unknown functional variants. According to DICKSON *et al.* (24), there is evidence to suggest that genome-wide associations can extend across multiple LD blocks.

The number of mother–child duos analyzed (*n* = 27) was small enough to ensure that the genes located in the CL±P-associated X-chromosome genomic segments would make an important contribution to the complex etiology of CL±P. However, an additional benefit of our study design is that the CL±P cases in the study population were more etiologically homogeneous than cases included in multicenter studies or from populations located in geographical areas with a typical CL±P prevalence.

This study confirms the association between the *DMD* gene and OCs, and identifies the Xp11.4 intergenic region as being associated with CL±P, in a South American population with a strong Amerindian ethnic background and a high prevalence of CL±P. In order to confirm the current associations and enhance their statistical analysis, further studies, including higher numbers of duos affected by CL±P, are needed.

Conflicts of interest – The authors declare no conflicts of interest.

References

1. STRAUSS RP. The organization and delivery of craniofacial health services: the state of the art. *Cleft Palate Craniofac J* 1999; **36**: 189–195.
2. WYSZYNSKI DF, BEATY TH, MAESTRI NE. Genetics of nonsyndromic oral clefts revisited. *Cleft Palate Craniofac J* 1996; **33**: 406–417.
3. JONES MC. Etiology of facial clefts: prospective evaluation of 428 patients. *Cleft Palate J* 1988; **25**: 16–20.
4. PRESCOTT NJ, LESS MM, WINTER RM, MALCOLM S. Identification of susceptibility loci for nonsyndromic cleft lip with or without cleft palate in a two stage genome scan of affected sib-pairs. *Hum Genet* 2000; **106**: 345–350.
5. PATEL PJ, BEATY TH, RUCZINSKI I, MURRAY JC, MARAZITA ML, MUNGER RG, HETMANSKI JB, WU T, MURRAY T, ROSE M, REDETT RJ, JIN SC, LIE RT, WU-CHOU YH, WANG H, YE X, YEOW V, CHONG S, JEE SH, SHI B, SCOTT AF. X-linked markers in the Duchenne muscular dystrophy gene associated with oral clefts. *Eur J Oral Sci* 2013; **121**: 63–68.
6. CASTILLA EE, ORIOLI IM. ECLAMC: the Latin-American collaborative study of congenital malformations. *Community Genet* 2004; **7**: 76–94.
7. POLETTA FA, CASTILLA EE, ORIOLI IM, LOPEZ-CAMELO JS. Regional analysis on the occurrence of oral clefts in South America. *Am J Med Genet A* 2007; **143A**: 3216–3227.
8. KORN JM, KURUVILLA FG, MCCARROLL SA, WYSOKER A, NEMESH J, CAWLEY S, HUBBELL E, VEITCH J, COLLINS PJ, DARVISHI K, LEE C, NIZZARI MM, GABRIEL SB, PURCELL S, DALY MJ, ALTSHULER D. Integrated genotype calling and associa-tion analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 2008; **40**: 1253–1260.
9. ANDERSON CA, PETTERSSON FH, CLARKE GM, CARDON LR, MORRIS AP, ZONDERVAN KT. Data quality control in genetic case-control association studies. *Nat Protoc* 2010; **5**: 1564–1573.
10. HO GY, BAILEY-WILSON JE. The transmission/disequilibrium test for linkage on the X chromosome. *Am J Hum Genet* 2000; **66**: 1158–1160.
11. SHI M, UMBACH DM, WEINBERG CR. Identification of risk-related haplotypes with the use of multiple SNPs from nuclear families. *Am J Hum Genet* 2007; **81**: 53–66.
12. PURCELL S, NEALE B, TODD-BROWN K, THOMAS L, FERREIRA MA, BENDER D, MALLER J, SKLAR P, DE BAKKER PI, DALY MJ, SHAM PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
13. O'CONNELL JR, WEEKS DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
14. *R project for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org/. Accessed on January 18, 2015.
15. SCAPOLI L, MARTINELLI M, ARLOTTI M, PALMIERI A, MASIERO E, PEZZETTI F, CARINCI F. Genes causing clefting syndromes as candidates for non-syndromic cleft lip with or without cleft palate: a family-based association study. *Eur J Oral Sci* 2008; **116**: 507–511.
16. JUGESSUR A, SKARE O, LIE RT, WILCOX AJ, CHRISTENSEN K, CHRISTIANSEN L, NGUYEN TT, MURRAY JC, GJESSING HK. X-linked genes and risk of orofacial clefts: evidence from two population-based studies in Scandinavia. *PLoS ONE* 2012; **7**: e39240.
17. KOENIG M, MONACO AP, KUNKEL LM. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell* 1988; **53**: 219–228.
18. MUNTONI F, TORELLI S, FERLINI A. Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurol* 2003; **2**: 731–740.
19. SVED JA, MCRAE AF, VISSCHER PM. Divergence between human populations estimated from linkage disequilibrium. *Am J Hum Genet* 2008; **83**: 737–743.
20. BURTON BK, DILLARD RG, WEAVER RG. Walker-Warburg syndrome with cleft lip and cleft palate in two sibs. *Am J Med Genet* 1987; **27**: 537–541.
21. PRATAP A, AGRAWAL A, TIWARI A, LAKSHMI R, RAJBANSHI S. The Walker-Warburg syndrome with cleft lip and palate. *Singapore Med J* 2007; **48**: e66–e67.
22. VAJSAR J, BASKIN B, SWOBODA K, BIGGAR DW, SCHACHTER H, RAY PN. Walker-Warburg Syndrome with *POMT1* mutations can be associated with cleft lip and cleft palate. *Neuromuscul Disord* 2008; **18**: 675–677.
23. MOTTAGUI-TABAR S, FAGHIHI MA, MIZUNO Y, ENGSTRÖM PG, LENHARD B, WASSERMAN WW, WAHLESTEDT C. Identification of functional SNPs in the 5-prime flanking sequences of human genes. *BMC Genom* 2005; **6**: 18.
24. DICKSON SP, WANG K, KRANTZ I, HAKONARSON H, GOLDSTEIN DB. Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010; **8**: e1000294.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. X-linked genes causing syndromes that present with oral clefts.

Table S2. Probands with cleft lip with or without cleft palate included in this study.

Data S1. Genome-wide Human SNP array 6.0 data set.