Region 8q24 Is a Susceptibility Locus for Nonsyndromic Oral Clefting in Brazil

Luciano Abreu Brito,1 Lívia Máriz Ribeiro Paranaiba,2,3 Camila Fernandes Silva Bassi,1 Cibele Masotti,1 Carolina Malcher,1 David Schlesinger,1,4 Katia Maria Rocha,1 Lucas Alvizi Cruz,1 Lígia Kobayashi Bárbara,1 Nivaldo Alonso,5 Diogo Franco,6 Elizabete Bagordakis,2 Hercílio Martelli Jr.,3,7 Diogo Meyer,8 Ricardo D. Coletta,2 and Maria Rita Passos-Bueno1*

1Human Genome Research Center, Institute of Biosciences, University of São Paulo, São Paulo, Brazil
2Department of Oral Diagnosis, School of Dentistry, State University of Campinas, Piracicaba, São Paulo, Brazil
3Stomatology Clinic, Dental School, State University of Montes Claros, Montes Claros, Minas Gerais, Brazil
4Instituto do Cérebro, Instituto Israelita de Ensino e Pesquisa Albert Einstein, São Paulo, Brazil
5Department of Surgery, School of Medicine, University of São Paulo, São Paulo, Brazil
6Department of Plastic Surgery, University of Rio de Janeiro, Rio de Janeiro, Brazil
7Center for Rehabilitation of Craniofacial Anomalies, Dental School, University of Alfenas, Minas Gerais, Brazil
8Department of Genetics, Institute of Biosciences, University of São Paulo, São Paulo, Brazil

Received 28 January 2012; Revised 2 March 2012; Accepted 6 March 2012

BACKGROUND: Nonsyndromic cleft lip with or without cleft palate is a relatively common craniofacial defect with multifactorial inheritance. The association of the rs987525 single nucleotide variant, located in a gene desert at 8q24.21 region, has been consistently replicated in European populations. We performed a structured association approach combined with transcriptional analysis of the MYC gene to dissect the role of rs987525 in oral clefting susceptibility in the ethnically admixed Brazilian population. METHODS: We performed the association study conditioned on the individual ancestry proportions in a sample of 563 patients and 336 controls, and in an independent sample of 221 patients and 261 controls. The correlation between rs987525 genotypes and MYC transcriptional levels in orbicularis oris muscle mesenchymal stem cells was also investigated in 42 patients and 4 controls. RESULTS: We found a significant association in the larger sample (p = 0.0016; OR = 1.80 [95% confidence interval [CI], 1.21–2.69], for heterozygous genotype, and 2.71 [95% CI, 1.47–4.96] for homozygous genotype). We did not find a significant correlation between rs987525 genotypes and MYC transcriptional levels (p = 0.14; r = −0.22, Spearman Correlation). CONCLUSIONS: We present a positive association of rs987525 in the Brazilian population for the first time, and it is likely that the European contribution to our population is driving this association. We also cannot discard a role of rs987515 in MYC regulation, because this locus behaves as an expression quantitative locus of MYC in another tissue. Birth Defects Research (Part A) 94:464–468, 2012. © 2012 Wiley Periodicals, Inc.

Key words: rs987525; rs1476165; rs2099897; heritability; cleft lip; cleft palate; 8q24 gene desert; MYC transcriptional levels; common disease-common variant; mesenchymal stem cells

INTRODUCTION

Nonsyndromic cleft lip with or without cleft palate (NS CL/P; OMIM 119530), which represents 70% of oro-facial clefts and is the most common craniofacial anomaly, is always accompanied by physical and psychological impairments to affected children, with great family burden (Wehby and Cassell, 2010).

NS CL/P has a worldwide birth prevalence of approximately 1:700 live births (Dixon et al., 2011), varying according to ethnicity (African, 0.3:1000; European, 1.0:1000; Native American, 3.6:1000), geographic origin, and...
and socioeconomic status (Carinci et al., 2007; Jugessur et al., 2009). In Brazil, NS CL/P prevalence has ranged from 0.28:1000 to 1.54:1000 (Menegotto and Salzano, 1991; Loffredo et al., 2001; Martelli-Junior et al., 2006). NS CL/P etiology is multifactorial and is determined by numerous interacting loci with additional environmental covariates (Mossey et al., 2009). The identification of genetic risk factors for NS CL/P has been the subject of intensive research, and genome-wide scans have shown a range of plausible candidate genes or chromosomal regions (Dixon et al., 2011). Of those, the most remarkable loci are the IRF6 gene (Zucchero et al., 2004; Rahimov et al., 2008) and the rs987525 single nucleotide variant (SNV) at 8q24.21 (Birnbaum et al., 2009), as their association with NS CL/P has been largely replicated (Park et al., 2007; Jugessur et al., 2008; Grant et al., 2009; Huang et al., 2009; Nikopensius et al., 2009; Beaty et al., 2010; Blanton et al., 2010a, 2010b). Despite the several reported associations between rs987525 and NS CL/P, it is unknown whether this SNV is the cause of the association or if it is in linkage disequilibrium with still some unknown gene or region; nevertheless, a regulatory function has been credited for this region (Dixon et al., 2011). Furthermore, patients of European origin seem to be more at risk from this SNV (Beaty et al., 2010), and it is certainly important to evaluate its worldwide effect. Although the Brazilian population has a trihybrid constitution of European, African, and Native American parental populations, European contribution is as high as 80% (Santos et al., 2010), and it is possible that rs987525 contributes to oral clefting in our population. To test this hypothesis, and to avoid population stratification bias, we have conducted a structured association analysis in a large sample of patients ascertained in different localities of the country (Sample 1) and in a minor replication sample ascertained in the state of Minas Gerais (southeastern region; Sample 2). To provide new insights into the functional role of rs987525, we have also investigated the correlation between its genetic contribution and the expression levels of the MYC gene, which is an important regulator of a variety of cellular processes necessary for normal cell growth and differentiation (Secombe et al., 2004), located at approximately 1.1 Mb of rs987525.

MATERIALS AND METHODS

Ethics

All samples were collected in accordance with the rules of the Research Ethics Committee of the Institute of Biosciences, University of São Paulo, Brazil, and the Human Research Ethics Committee of the University of Alfenas, Brazil. Informed consent was obtained from the patients or legal tutors.

Patients and Controls

We ascertained two independent samples of unrelated cases and controls. Sample 1 comprises 563 NS CL/P patients from medical programs of Operation Smile (http://www.operationsmile.org) between 2007 and 2010 in five different Brazilian sites (Santarém, Fortaleza, Barbalha, Maceió, and Rio de Janeiro) and 336 controls from São Paulo (previously described by Errera et al., 2006), provided by our Biobank, because we were not able to ascertain local controls at the corresponding period. Patients affected by cleft palate only were removed from the sample. Patient ascertainment methods were previously reported by Brito et al., (2011a). DNA extraction from an oral swab followed the manufacturer’s protocols (Macherey-Nagel, Düren, Germany).

To replicate our findings, we recruited an independent sample of 221 patients and 261 controls (Sample 2) from the Center for Rehabilitation of Craniofacial Anomalies—Dental School, University of Alfenas, State of Minas Gerais (southeastern region of Brazil). Control individuals were subjects admitted as in-patients, but with conditions unrelated to clefting disorders and their endophenotypes. Genomic DNA of Sample 2 was extracted from oral mucosa cells, as described previously (Aidar and Line, 2007).

To verify the effect of rs987525 in the expression levels of the MYC gene, we established mesenchymal stem cell cultures (MSCs) from orbicularis oris muscle fragments (OMMF), discarded during surgical procedures of cheiloplasty, of an additional 42 NS CL/P patients and 4 controls submitted to surgery for other reasons than CL/P correction. These surgeries were performed in three different centers (Division of Plastic Surgery of University of São Paulo, SOBRAPAR Hospital–Campinas, and Operation Smile, Rio de Janeiro, Brazil).

Genotyping

All individuals were genotyped for the SNV rs987525 with the TaqMan method (Applied Biosystems, Foster City, CA, USA). To characterize the European, African, and Native American ancestry contributions, Samples 1 and 2 were genotyped for independent panels of 40 ancestry informative indels. FAM-labeled M13 tail was incorporated to forward primers, allowing multiplex PCR, with four to seven markers per reaction (sequences available upon request). These markers are spread across the genome and were accessed from the Marshfield database (http://research.marshfieldclinic.org), which holds a set of indel markers described by Weber et al. (2002). Sample 2 was genotyped for a panel extensively used for characterizing Brazilian populations (Bastos-Rodrigues et al., 2006), and Sample 1 was genotyped for an independent panel with matched marker selection criteria, as published previously (Brito et al., 2011a). Fragment analysis was performed with Gene Mapper software (Applied Biosystems) after capillary electrophoresis targeting the PCR products for the FAM-labeled M13 tail.

Cell Culture

The MSCs were established according to previously published protocol (Bueno et al., 2009) and expanded to 80% confluence. RNA and DNA extraction followed the manufacturer’s protocols (Macherey-Nagel). We confirmed the homogeneous mesenchymal status of 10 cell cultures using five mesenchymal cell markers (CD29, CD90, CD105, SH3, and SH4), one hematopoietic (CD45), and one endothelial (CD31) through flow cytometric analysis (EasyCyte Flow Cytometer; Guava Technologies, Hayward, CA) according to standard protocol in our laboratory (Bueno et al., 2010).

Statistical Analyses

We used Structure 2.3.3 to infer ancestry components of each individual (Pritchard et al., 2000a; Falush et al., 2007).
2003, 2007), in a run assuming K = 3 parental populations, based on the trihybrid origin of Brazilians; we also incorporated “learning samples” in the run, with prespecified population of origin (European, African, and Native American reference populations from Marshfield Clinic collection) to assist the software to estimate ancestry of the admixed individuals. Following the ancestry inference, STRAT was used to test the SNV for association, conditioning on the individual ancestry proportions (Pritchard et al., 2000b).

We obtained MYC expression values from microarray gene expression data (Human Gene 1.0 ST microarray chips; Affymetrix, Santa Clara-CA, USA), with the three-step robust multiarray average preprocessing method implemented in the software Expression Console (Affymetrix). We assessed the correlation between the rs987525 genotype and MYC expression levels with a Spearman correlation test (95% confidence interval [CI]).

RESULTS

Association Analyses

After excluding individuals with low genotype assignment (<80%), Sample 1 comprised 471 patients (18% presenting cleft lip only [CLO] and 82% presenting cleft lip and cleft palate [CLP]) and 326 controls. Average ancestry contributions were estimated at 22.2% African, 60.3% European, and 17.5% Native American for patients, and 18.8% African, 70.9% European, and 10.3% Native American for controls (individual ancestry profile is shown in Fig. 1). Ancestry analysis stratifying the patients for region of origin is shown in Table 1. Minor allele frequency (MAF) of rs987525 was similar in cases (34%) and controls (35%), and genotype distributions followed the Hardy-Weinberg expected proportions in both groups (p > 0.48). Despite the similar MAF between cases and controls, a significant association with the A allele was observed (p = 0.006) in the structured association test (Table 1). We further investigated whether the European ancestry would be driving this association by stratifying the analysis into two groups: a European group with the 361 patients presenting more than 50% of European ancestry, and a non-European group with 110 patients with less than 50% of European ancestry, and then comparing both groups with the same 326 controls. We found a stronger association in the European group (p = 0.008) than in the non-European group (p = 0.21).

Sample 2 consisted of 196 patients (28% CLO and 72% CLP) and 263 controls, after excluding low-genotype assignment individuals. Average ancestry contributions of this sample were estimated at 13.4% African, 84.1% European, and 2.5% Native American for patients and 8.1% African, 89.0% European, and 2.9% Native American for controls (Fig. 1). Minor allele frequency was higher in patients (39.5%) than in controls (27.4%), and genotypes matched Hardy-Weinberg proportions in both groups (p > 0.31). A significant association between rs987525 and NS CL/P group was also observed in this sample (p = 0.0016). Because rs987525 allele frequencies vary greatly among continents, and differences in ancestry contribution between cases and controls are more pronounced in Sample 1, we considered that we would obtain a more reliable odds ratio (OR) estimates only for Sample 2, from where cases and controls come from a single location. For this group, OR was estimated at 1.80 (95% CI, 1.21–2.69) for the heterozygous genotype (AC versus CC) and 2.70 (95% CI, 1.47–4.96) for the homozygous genotype (AA versus CC), under an additive model.

MYC Expression Analyses

We found a nonsignificant correlation between rs987525 genotypes (Supporting Information Table S1) and MYC transcriptional levels in MSCs (Supporting Information Fig. S1; p = 0.14; r = −0.22, Spearman correlation).

DISCUSSION

It has been postulated that part of the susceptibility alleles for complex disorders might have a common origin...
### Table 1

Estimates of Average Ancestry Contributions of Parental Populations (European, African, and Native American), rs987525 Genotype Frequencies, Risk Allele (A) Frequency, and p Value of the Structured Association Tests

<table>
<thead>
<tr>
<th>Subsamples</th>
<th>N</th>
<th>European</th>
<th>African</th>
<th>Native American</th>
<th>AA</th>
<th>AC</th>
<th>CC</th>
<th>f(A)</th>
<th>p value (H-W test)</th>
<th>p value (structured association)</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Patients</td>
<td>471</td>
<td>60</td>
<td>22</td>
<td>18</td>
<td>12</td>
<td>45</td>
<td>44</td>
<td>34</td>
<td>0.92</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>Santaréém</td>
<td>82</td>
<td>54</td>
<td>14</td>
<td>32</td>
<td>7</td>
<td>38</td>
<td>55</td>
<td>26</td>
<td>0.99</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>Barbalha</td>
<td>52</td>
<td>64</td>
<td>24</td>
<td>12</td>
<td>12</td>
<td>58</td>
<td>30</td>
<td>41</td>
<td>0.09</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Fortaleza</td>
<td>163</td>
<td>64</td>
<td>20</td>
<td>16</td>
<td>12</td>
<td>42</td>
<td>46</td>
<td>33</td>
<td>0.42</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>Maceió</td>
<td>62</td>
<td>59</td>
<td>23</td>
<td>19</td>
<td>12</td>
<td>44</td>
<td>44</td>
<td>34</td>
<td>0.92</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>Rio de Janeiro</td>
<td>112</td>
<td>58</td>
<td>31</td>
<td>11</td>
<td>14</td>
<td>48</td>
<td>40</td>
<td>38</td>
<td>0.72</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>Controls</td>
<td>326</td>
<td>71</td>
<td>19</td>
<td>41</td>
<td>13</td>
<td>44</td>
<td>42</td>
<td>35</td>
<td>0.48</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sample 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients</td>
<td>196</td>
<td>84</td>
<td>13</td>
<td>3</td>
<td>16</td>
<td>47</td>
<td>37</td>
<td>40</td>
<td>0.69</td>
<td>0.0016</td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>263</td>
<td>89</td>
<td>8</td>
<td>3</td>
<td>9</td>
<td>37</td>
<td>57</td>
<td>21</td>
<td>0.31</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The results are presented for Sample 1 as a whole and stratified by location of origin of the patients (Santaréém, Brabalha, Fortaleza, Maceió, and Rio de Janeiro).

The odds ratio was only estimated for Sample 2, since cases and controls from this sample are less stratified. The odds ratios for heterozygous (AC; \( p = 0.004 \)) and homozygous (AA; \( p = 0.001 \)) genotypes were estimated by comparing with CC genotype, under an additive model.

CA, frequency of the risk allele (A); OR, odds ratio; CI, confidence interval; H-W, Hardy-Weinberg.

---

"The results are presented for Sample 1 as a whole and stratified by location of origin of the patients (Santaréém, Brabalha, Fortaleza, Maceió, and Rio de Janeiro). The odds ratio was only estimated for Sample 2, since cases and controls from this sample are less stratified. The odds ratios for heterozygous (AC; \( p = 0.004 \)) and homozygous (AA; \( p = 0.001 \)) genotypes were estimated by comparing with CC genotype, under an additive model."
sample with the highest heritability, which comes from Barbalha (Brito et al., 2012). We have therefore stratified the present association analysis by location of origin, and we have verified that the Barbalha group reached the highest significance ($p = 0.005$), compared with samples from Santarém ($p = 0.07$), Fortaleza ($p = 0.33$), Maceió ($p = 0.21$), and Rio de Janeiro ($p = 0.19$; Table 1). These results provide confirm heritability as a confounding factor in association analyses.

In summary, our results provided further evidence that rs987525 is involved with NS CL/P susceptibility in patients of European descent, and heritability seemed to be an important confounding factor. In addition, we did not find evidence of rs987525 modulating MYC expression.

ACKNOWLEDGMENTS

We thank the Operation Smile team and its executive director Clovis Brito, Constância Urbani and Márcia Teixeira for secretarial assistance, and the Brazilian Ministry of Health, FAPESP, and CNPq for financial support.

REFERENCES


