Gene-gene interaction for nonsyndromic cleft lip with or without cleft palate in Chilean case-parent trios

José Suazo\textsuperscript{a,⁎}, José Luis Santos\textsuperscript{b}, Alicia Colombo\textsuperscript{c,d}, Rosa Pardo\textsuperscript{c,f,g}

\textsuperscript{a} Instituto de Investigación en Ciencias Odontológicas, Facultad de Odontología, Universidad de Chile, Sergio Livingstone #943, Santiago, Chile
\textsuperscript{b} Departamento de Nutrición, Diabetes y Metabolismo, Escuela de Medicina, Pontificia Universidad Católica de Chile, Lira #44, Santiago, Chile
\textsuperscript{c} Programa de Anatomía y Biología del Desarrollo, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Independencia #1027, Santiago, Chile
\textsuperscript{d} Servicio de Anatomía Patológica, Hospital Clínico de la Universidad de Chile, Santos Dumont #999, Santiago, Chile
\textsuperscript{e} Unidad de Genética, Hospital Clínico Universidad de Chile, Santos Dumont #999, Santiago, Chile
\textsuperscript{f} Unidad de Neonatología, Hospital Clínico Universidad de Chile, Santos Dumont #999, Santiago, Chile
\textsuperscript{g} Unidad de Genética, Hospital Dr. Sótero del Río, Concha y Toro #3459, Santiago, Chile

**ARTICLE INFO**

**Keywords:**
- Nonsyndromic cleft lip with or without cleft palate
- TGFB3
- MSX1
- Case-parents trios
- GxG interaction

**ABSTRACT**

Objective: Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a birth defect for which several genes susceptibility genes have been proposed. Consequently, it has been suggested that many of these genes belong to common inter-related pathways during craniofacial development gene-gene interaction. We evaluated the presence of gene-gene interaction for single nucleotide polymorphisms within interferon regulatory factor 6 (IRF6), muscle segment homeobox 1 (MSX1), bone morphogenetic protein 4 (BMP4) and transforming growth factor 3 (TGFB3) genes in NSCL/P risk in Chilean case-parent trios.

Design: From previous studies, we retrieved genotypes for 13 polymorphic variants within these four genes in 152 case-parent trios. Using the trio package (R) we evaluate the gene-gene interaction in genetic markers pairs applying a 1°-of-freedom test (1df) and a confirmatory 4°-of-freedom (4df) test for epistasis followed by both a permutation test and a Benjamini-Hochberg test for multiple comparisons adjustment.

Results: We found evidence of gene-gene interaction for rs6446693 (MSX1) and rs2268625 (TGFB3) (4df p = 0.024; permutation p = 0.015, Benjamini-Hochberg p = 0.001).

Conclusions: A significant gene-gene interaction was detected for rs6446693 (MSX1) and rs2268625 (TGFB3). This finding is concordant with research in animal models showing that MSX1 and TGFB3 are expressed in common molecular pathways acting in an epistatic manner during maxillofacial development.

1. Introduction

Expression and susceptibility to complex traits are characterized by the interaction of genes and environmental factors and by gene-gene (GxG) interaction (also known as epistasis). Thus, any study aiming to elucidate the genetic architecture of a complex trait, should consider GxG interaction analysis (Carlborg & Haley, 2004). Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a complex disorder representing the most common birth defect affecting human craniofacial development (Jugessur & Murray, 2005; Watkins, Meyer, Strauss, & Aylsworth, 2014). Consequently to its complex etiology, NSCL/P potential causal mutations have been detected within regulatory and coding regions if at least 20 candidate genes encoding products involved in different molecular pathways (Dixon, Marazita, Beatty, & Murray, 2011; Leslie & Marazita, 2013; Vieira, 2008).

In the best of our knowledge, there is no genome-wide association studies for this birth defect analyzing exclusively Chilean subjects. However, our research group has been assessed the association between loci interferon regulatory factor 6 (IRF6, 1q32.2), muscle segment homeobox 1 (MSX1, 4p16.2), bone morphogenetic protein 4 (BMP4, 14q22.2) and transforming growth factor 3 (TGFB3, 14q24.3) and NSCL/P based on previous reports in other populations (Paradowska-Stolarz, 2015; Tang, Wang, Han, Guo, & Wang, 2013; Watkins et al., 2014; Wattanawong, Rattanasiri, McEvoy, Attia, & Thakkinstian, 2016). Thus, in single-gene studies we found cleft risk haplotypes composed by polymorphic variants (single nucleotide polymorphisms or SNPs) within these loci using a sample of NSCL/P case-parent trios from a Chilean population (Suazo, Santos, Jara, & Blanco, 2008; Suazo, 2011; Leslie & Marazita, 2013; Vieira, 2008).

Abbreviations: BMP4, bone morphogenetic protein 4; GxG interaction, gene-gene interaction; IRF6, interferon regulatory factor 6; MSX1, muscle segment homeobox 1; NSCL/P, nonsyndromic cleft lip with or without cleft palate; SNP, single nucleotide polymorphism; TGFB3, transforming growth factor beta 3

⁎ Corresponding author.

E-mail address: jsuazo@odontologia.uchile.cl (J. Suazo).

https://doi.org/10.1016/j.archoralbio.2018.04.009

Received 2 January 2018; Received in revised form 12 April 2018; Accepted 13 April 2018

0003-9969/ © 2018 Elsevier Ltd. All rights reserved.
There is evidence of interaction among the products of IRF6, MSX1, BMP4 and TGFβ3 during maxillofacial development in animal models. Thus, the expression of Msx1 and Bmp4 seems to be reciprocally regulated in the mesenchymal tissues of murine palatal shelves where the expression of the Bmp4 protein reverses Msx1 deficiency and rescues the cleft palate phenotype (Smith, Lozanoff, Yyyanar, & Nazarali, 2013; Zhang et al., 2002). On the other hand, a mice knock-out line for Tgfβ3 exhibits an expression reduction of Msx1 and a proliferation deficiency of the palatal shelves which possibly explain the cleft palate observed in these animals (Del Río et al., 2011). Based on these evidences and in the light of the palatal shelves which possibly explain the cleft palate observed in these animals, we have used a Wald test of 1°-of-freedom (1df) in order to identify GxG interaction between haplotypes composed by SNPs from these four genes. Unfortunately, in the best of our knowledge, there are no developed methods for haplotype interaction for case-parent trio samples. Alternatively, in a Chilean case-control sample we recently assessed haplotype-based interaction of these same loci for NSCL/P expression and found GxG interaction between haplotypes composed by SNPs of BMP4 and IRF6 (Blanco, Colombo, Pardo, & Suazo, 2017). However, population-based design (such as case-control) may generate spurious associations due to population stratification by ethnicity, a phenomenon detected in the contemporary urban Chilean population (Palomino, Palomino, Cauvi, Barton, & Chacraborty, 1997; Santos, Jara, & Blanco, 2010; Suazo, Santos, Scapoli, Jara, & Blanco, 2010).

2. Materials and methods

2.1. Subjects

The sample was composed of 152 unrelated NSCL/P Chilean case-parent trios. Cases were 38% females and 69% of them had no family history of orofacial clefts. According to the clinical classification, 92% were cataloged as cleft lip and palate whereas 8% exhibited only a cleft lip. These cases and their parents were obtained between the years 2008 and 2011 after in depth interviews of at least three family members in order to reconstruct a family history. Interviews considered maternal exposure to teratogenic substances (phenytoin, warfarin and ethanol) during pregnancy. These nuclear families were recruited at the following centers: Craniofacial Malformations Unit, School of Dentistry, Universidad de Chile; Cleft Lip/Palate Center, Hospital Exequiel Gonzalez Cortes; Dental Service, Hospital Roberto del Rio; Maxillofacial Service, Hospital San Borja-Arriaran; and Maxillofacial Service, Hospital Sótero del Rio (all of them located in the city of Santiago, Chile). Our study was approved by School of Medicine (Universidad de Chile) Institutional Review Board (Protocols #2005-11 and #018-IRF6).

Table 1. IRF6 MSX1, BMP4 and TGFβ3 SNPs considered for the current study.

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Physical Locationa</th>
<th>MAFb</th>
<th>HW P valuec</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRF6</td>
<td>rs2235371</td>
<td>Chr1: 209790735</td>
<td>0.197</td>
<td>0.870</td>
</tr>
<tr>
<td></td>
<td>rs2235375</td>
<td>Chr1: 209792242</td>
<td>0.515</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>rs2236909</td>
<td>Chr1: 20978310</td>
<td>0.498</td>
<td>0.272</td>
</tr>
<tr>
<td></td>
<td>rs764093</td>
<td>Chr1: 209809986</td>
<td>0.289</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>rs644693</td>
<td>Chr4: 4853353</td>
<td>0.473</td>
<td>0.528</td>
</tr>
<tr>
<td></td>
<td>rs3775261</td>
<td>Chr4: 4862018</td>
<td>0.403</td>
<td>0.853</td>
</tr>
<tr>
<td></td>
<td>rs12532</td>
<td>Chr4: 4863419</td>
<td>0.490</td>
<td>0.090</td>
</tr>
<tr>
<td>BMP4</td>
<td>rs2855332</td>
<td>Chr4: 53953247</td>
<td>0.318</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>rs526642</td>
<td>Chr4: 53956335</td>
<td>0.455</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>rs1957860</td>
<td>Chr4: 53962673</td>
<td>0.413</td>
<td>0.667</td>
</tr>
<tr>
<td>TGFβ3</td>
<td>rs3917201</td>
<td>Chr5: 75963212</td>
<td>0.286</td>
<td>0.853</td>
</tr>
<tr>
<td></td>
<td>rs2316625</td>
<td>Chr5: 75973291</td>
<td>0.239</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td>rs2308662</td>
<td>Chr5: 75978424</td>
<td>0.415</td>
<td>0.383</td>
</tr>
</tbody>
</table>

a Chromosome location according to GRCh38 genomic assembly at dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/).
b MAF: minor allele frequency estimated in an all parental genotypes.
c HW P value: significance for deviations from Hardy-Weinberg equilibrium for genotype frequency distribution in the parental sample.

2.2. Genotype data extraction

For GxG interaction estimation, genotype data for 13 SNPs (Table 1) were extracted from previous studies including the same Chilean case-parent trio population here described (Suazo et al., 2008; Suazo et al., 2010a, 2010b; Suazo, Santos and Scapoli et al., 2010). All SNP genotypes were obtained by a polymerase chain reaction followed by a restriction fragment length polymorphism analysis as was described in our previous reports (Suazo et al., 2008; Suazo et al., 2010a, 2010b; Suazo, Santos and Scapoli et al., 2010). As quality control, amplicons for a subset of samples were directly sequenced confirmed the genotypes obtained by polymerase chain reaction followed by a restriction fragment length polymorphism analysis and then used as internal controls (data not shown). Polymorphisms considered in the current report, their alleles frequencies and genonic positions are listed in Table 1.

2.3. Statistical analyses

We calculated allele frequencies based on simple proportions using parental genotypes. Departures from the Hardy-Weinberg equilibrium in genotype distributions in parents were evaluated through a chi-square goodness-of-fit test implemented in STATA 12 statistical package. In order to select candidate SNPs pairs for GxG interaction, we performed a conditional logistic regression (1df Wald test). This conditional model is based in the assumption of an additive mode of inheritance for a pair of SNPs randomly selecting one among three pseudo-controls and, therefore, containing one interaction parameter (Li et al., 2010), providing a rapid analysis of pairwise GxG interactions (Xiao et al., 2016). As we previously mentioned, 1df Wald test may be considered as evidence of epistasis but also may reflect a violation of interaction model (Cordell, 2002; Xiao et al., 2016). Thus, these results were confirmed applying the 4df likelihood ratio test which considers the genotypes for the case and three possible genotypes (pseudo-controls) based on both parental genotypes (Cordell, 2002; Schaid, 1999; Xiao et al., 2016). Due to the multiple hypotheses evaluated in this study, it is necessary to adjust the significant results by a multiple comparison correction. The most popular method, named Bonferroni correction, can increase the false-positive rate when SNPs pairs are not independent (i.e. they are in linkage disequilibrium) reducing the statistical power (Nyholt, 2004). Therefore, we have opted for a
permutation test following the criteria of Xiao et al. (2016). There are 45 pairs of SNPs for interaction, hence the adjusted p-value must be less than 0.05/45 = 0.0011, being the reason why we considered 1000 permutations as adequate for our aim. To confirm the interaction signal, a random 1000 permutation data sets was originated for each significant pair for 4df test and the permutation p-value was calculated by the proportion of permuted 4df statistic exceeding the observed among the 1000 permuted data (Xiao et al., 2016). All of these tests (1df, 4df and permutations) were included in the trio package (Schwender et al., 2014) and performed in RStudio software version 1.0.136. Due to small sample sizes may affect the statistical power of permutations tests (Gadbury, Page, Heo, Mountz, & Allison, 2003), we decide to apply the Benjamini-Hochberg test (Benjamini & Hochberg, 1995) as a complementary multiple comparison correction test. This test is based on the control of false discovery rate to avoid Type-I errors. A rank from the more significant p-value to the less significant one is constructed. Then a corrected value is assigned for each individual p-value computed by means of the quotient of its rank position and the total number of tests and multiplied by the recommended false discovery rate of 0.05. Thus, if a corrected p-value is < 0.05, it is considered as significant after multiple comparisons correction (Benjamini & Hochberg, 1995)

3. Results

Table 1 shows the list of SNPs used in this report, their genomic position and allele frequencies in the parental sample. Two of them, rs7640933 (IRF6) and rs2855532 (BMP4) were excluded for interaction analysis due to their deviation of Hardy-Weinberg equilibrium (Table 1). Therefore, GxG interaction assessment was performed using 11 SNPs given 45 possible pairs of markers. Table 2 shows the significant results for GxG interaction for NSCL/P risk in this stepwise study. Thus, for the screening step using 1df Wald statistic, we found four SNPs pairs with evidence of interaction. This is the case of rs2236909 (IRF6) and rs1957860 (BMP4) (p = 0.004), rs2235375 (IRF6) and rs1957860 (BMP4) (p = 0.019), rs6446693 (MSX1) and rs2268625 (TGFB3) (p = 0.027), rs3775261 (MSX1) and rs2268625 (TGFB3) (p = 0.038), and rs1957860 (BMP4) and rs3917201 (TGFB3) (p = 0.034) (Table 2). After the confirmatory step (4df Cordell statistic) only one pair remains significant, rs6446693-rs2268625 (p = 0.024), markers within MSX1 and TGFB3 genes, respectively (Table 1). In addition, these SNPs pair showed a significant P value extracted for 1000 permutations (p = 0.015) and after Benjamini-Hochberg correction (p = 0.001), confirming the evidence of GxG interaction for these markers in NSCL/P risk (Table 2).

4. Discussion

It has been well documented that family-based designs, such as case-parent trios, show advantages over the population-based ones, especially regarding the insensitivity to population stratification and the possibility of parent-of-origin assessment (Santos et al., 2002; Weinberg, 1999). Despite the advantages of family-based designs, the majority of methods for GxG interaction detection has been developed for case-control and cohort designs (Li et al., 2010). Based on a stepwise method, we found evidence of interaction between a SNPs pair located within MSX1 (rs6446693) and TGFB3 (rs2268625) genes in the increase of NSCL/P risk (Table 2). Although these SNPs have no individual effect on the association with this malformation, they are part of the haplotypes which showed transmission distortions in this same sample of trios (Suazo et al., 2010a; Suazo, Santos and Scapoli et al., 2010). These markers are not coding or regulatory variants and therefore, the interaction herein detected may reflect the presence of interacting cleft causal mutations in linkage disequilibrium with these SNPs, which must be detected in future studies. A report of Vieira et al. (2003), described interaction between polymorphic variants (a different set of markers to those considered in the current study) from MSX1 and TGFB3 in nonsyndromic orofacial cleft risk using trios from eight Latin-American countries including Chilean nuclear families. Although this last report apply a different statistical method, can be considered supporting our results in Chile.

There is evidence that the products of MSX1 and TGFB3 belong to common pathways in murine maxillofacial development. Independent studies have demonstrated that knock-out mice for Msx1 and Tgfb3 exhibit cleft palate among other craniofacial abnormalities (Satokata & Maas, 1994; Kaartinen et al., 1995). In addition, Tgfb3-/- mouse showed a significant decrease in the proliferation of the palatal mesenchymal which might explain the absence of contact and fusion of these structures during palatal development (Del Rio et al., 2011). Associated to the absence of Tgfb3 in these mice, there is a lack of the normal antero-posterior and medial-lateral gradient of Msx1 expression in the palatal shelves mesenchymal but no alteration of other genes pattern, such as Bmp2, Bmp4 or Shh (Del Rio et al., 2011). In accordance with these last results, when epidermal growth factor pathway is blocked Tgfb3-/- in palate culture, the mesenchymal pattern of Msx1 expression is restored and palate fusion is rescued (Barrio et al., 2014). Bioinformatic analyses have demonstrated that TGFβ3 and MSX1 products belong to the same molecular network modulating cell proliferation not only in craniofacial territories but also in other developmental process (Dai, Yu, Si, Fang, & Shen, 2015). Taking together, the results of these reports demonstrated an epistatic interaction among Tgfb3 and Msx1 in craniofacial development in animal models.

Using the same SNPs set considered for the current report, our group reported evidence of interaction between haplotypes composed by markers within IRF6 and BMP4 genes in the risk of NSCL/P in the Chilean population (Blanco et al., 2017). This latter report was based on a case-control design where the cases are the same subjects who have been herein analyzed together with their progenitors. Due to ancestral origin, the contemporary urban Chilean population shows a gradient among American admixture and socioeconomic generating a population stratification (Palomino et al., 1997; Valenzuela, 1988) which may generate spurious results of association (Santos et al., 2002). This is the reason why it has been suggested that case-control results may be ideally confirmed by a case-parent trio design study (Kazeem & Farrall, 2005). This previous study (Blanco et al., 2017) was performed as a haplotype-based interaction method which is more powerful than
single-marker methods due to its capability to capture the variability of a wide genomic region (Liu, Zhang, & Zhao, 2008). In the best of our knowledge there are no haplotype-based interaction methods for case-parents trio design. In conclusion, we have considered that both studies of our group have valid results and cannot be considered as complementary each other.

We have found three previous studies applying the step-wise method here performed for GxG interaction assessment in the risk of NSCL/P. The first one described interaction of SNPs belonging to wingless-type mmtv integration gene family and other cleft genes in samples of Asian and European trios (Li et al., 2015). Later, Xiao et al. (2016) evaluated the possible interaction between 13 NSCL/P susceptibility loci in European and Asian trios. These authors reported several pairs of SNPs from these loci showing significant evidence of interaction (Xiao et al., 2016). Finally, a third report described evidence of interaction for SNPs from genes within 16p13.3 region in Chinese trios (Liu et al., 2017). The authors of all just mentioned articles have highlighted the advantages of the statistical model for GxG interaction assessment in triads. Thus, in comparison to other methods, this one perform an effective screening approach in order to detect possible interacting pairs of SNPs which is later confirmed by the 4df Cordell likelihood ratio test model (Cordell, 2002; Li et al., 2015, Xiao et al., 2017). The 4df likelihood ratio test method has the advantage that can detect evidence of epistasis even for markers or genes with low effects on a phenotype associated to permutations for multiple comparisons, a method more effective than Bonferroni correction in presence of linkage disequilibrium (Xiao et al., 2016). In addition, in our study, all included SNPs have a minor allele frequency ≥ 0.2 which increase the power to detect interaction (Xiao et al., 2016). In this context, Emily, Mailund, Hein, Schauers, & Schierup (2009) have estimated that for loci with allelic frequency lower than 0.1, sample sizes greater than 300 individuals may be necessary to detect two-locus interaction. Unfortunately, the fact that this method considers all possible parental genotypes does not allow identifying interacting risk allele. Thus, additional functional/molecular methods may be useful for detecting a combined allele effect from the significant SNPs pair. This further study can complemented our results in order to get closer to translational research.

Although the advantages of the method here used and above mentioned, we need to consider our findings with caution due to the main limitation related to our modest sample size and the large number of test performed for all possible SNPs pairs, which can contribute decreasing the statistical power. Thus, we applied a second method for multiple comparisons correction (Benjamini & Hochberg, 1995) more appropriated for small or modest sample sizes in comparison to Bonferroni method (Streiner, 2015). Another limitation is the number of genes selected, which may limit the probability to detect interaction. As we previously mentioned, there is evidence that the products of these genes belong to common molecular networks in the maxillofacial development in animal models. However, it is possible that many other unknown proteins are acting in these networks in a manner not detected in the current study. Thus, it is necessary to amplify the number of genes analyzing large scale groups as those derived from genome-wide studies, for example. As we previously mentioned, there is no genome-wide association studies for orofacial clefts in Chile, which could be considered a disadvantage for the current report. However, a genome-wide association studies combined analyzed samples from Argentina, Bolivia, Brazil, Chile, Ecuador, Paraguay, Uruguay, and Venezuela was reported (Leslie et al., 2016) which did not show evidence of association for the 4 loci reported by our group. These differences may be explained by genetic heterogeneity. In this context, the contemporary Central and South American populations share a common Amerindian–Caucasian–African admixture but each country show certain particularities in its genetic background. At Spanish arrival, Amerindian populations were sparsely distributed in Argentina and Brazil in comparison to Andean countries. In addition, African descendants were mainly concentrated in Central America and in the northern region of South America (Vieira, Karras, Orioli, Castilla, & Murray, 2002). Therefore, our contemporary population is the product of the admixture of Amerindian and Caucasian genes with a slight contribution of African alleles (Eyheramendy, Martínez, Manevy, Vial, & Repetto, 2015).

In summary, we found evidence based on a step-wise method and confirmed by two independent methods for multiple comparisons, of gene–gene interaction for polymorphic variants belonging to TGFBS and MSXI genes in the expression of nonsyndromic cleft lip with or without cleft palate in the Chilean population. Although our results may be considered with caution due to the modest sample size of our trio population, they may be supported by a previous report including Chilean samples and by findings in animal models showing that these are expressed in common molecular pathways in an epistatic manner during maxillofacial development. Further studies with larger sample size and/ or in independent samples may confirm these results. In addition, molecular studies in both animal and human cell models may help us to elucidate the direct interaction (epistasis) among these two genes or the presence of other intermediary proteins and to identify specific risk allele combination, getting closer to a translational approach.

Conflict of interest

All authors declare no conflicts of interest.

Acknowledgements

The authors wish to thank to the patients, their families and health individuals who voluntarily cooperated with us and the staff members of all cleft children rehabilitation centers and of the Blood Bank of the Hospital San José. This study was partially supported by Fondo de Nacional de Desarrollo Científico y Tecnológico- Chile (FONDECYT) grants 1061078 and 11090105.

References


indicates defects of epithelial-mesenchymal interaction. Nature Genetics, 11, 415–421.


