Luciano Abreu Brito

Variantes genéticas de risco às fissuras orofaciais

Genetic risk variants for orofacial clefts

São Paulo

2016

Luciano Abreu Brito

Variantes genéticas de risco às fissuras orofaciais

Genetic risk variants for orofacial clefts

Tese apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Doutor em Ciências, na Área de Biologia/Genética.

Orientadora: Profª. Dra. Maria Rita dos Santos e Passos-Bueno

São Paulo

2016

Ficha Catalográfica

Brito, Luciano Abreu
Variantes genéticas de risco às fissuras orofaciais
164 páginas

Tese (Doutorado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Fissuras labiopalatinas 2. Sequenciamento de Exoma 3. CDH1
Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

_____	_____
Prof(a). Dr(a).	Prof(a). Dr(a).
_____	_____
Prof(a). Dr(a).	Prof(a). Dr(a).

_____

Profª. Dra. Maria Rita S. Passos-Bueno
orientadora

A todos os pacientes com os quais
tive contato ao longo deste projeto.

Education is when you read the fine print;
experience is what you get when you don't.

Pete Seeger

Agradecimentos

À minha família, em especial a meus pais e meu irmão, sem o apoio dos quais esta curta carreira já nem teria começado.

À Rita, pelo acolhimento, orientação, dedicação e disponibilidade durante todos esses anos.

Aos amigos do laboratório, que contribuíram para criar um ambiente de trabalho extremamente agradável: Gerson, Carol, Roberto, Lucas, Felipe, Van, Karina, May, Dani M, Bela, Bruno, Atique, Erika K, Joanna, Suzana, Tati, Dani B, Dani Y, Clarice, Ágatha, Camila M, Camila L, Lucas "Jr", Gabi "Jra", Cibele, Naila, Simone e Andressa.

A todos os organizadores e voluntários da Operação Sorriso, que tornaram este trabalho viável, e propiciaram momentos muito importantes de crescimento pessoal.

À equipe do Genoma, em especial ao pessoal do sequenciamento: Meire, Vanessa, Guilherme, Monize e, muito especialmente, Kátia, quem me ensinou o básico nos meus primeiros meses de laboratório.. Também aos demais colegas de departamento, em especial Vanessa S, Elaine, Michel, Natássia, Inês e Toninha.

To Eric, Christina, Antoine, Kushi, Renée, Mike, Yawei, Irving, François, Jullian, Vanessa, Maura, Chris, Leo, Sebastian, Takuya, Mary and Andy, who made my stay in Boston way easier and warmer.

Este trabalho contou com o apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e do Ministério da Ciência, Tecnologia e Inovação do Brasil.

Notas

Esta tese de doutorado compreende um trabalho desenvolvido durante os anos de 2012 a 2015 no Laboratório de Genética do Desenvolvimento do Centro de Estudos do Genoma Humano e Células Tronco, Instituto de Biociências, Universidade de São Paulo.

A tese foi redigida no modelo de artigos e capítulos, no idioma inglês. Cinco artigos foram incluídos no corpo principal da tese. Publicações em co-autoria e não relacionadas ao tema principal da tese encontram-se sumarizadas nos Apêndices, ao final da tese.

O projeto que resultou na presente tese foi cadastrado na Plataforma Brasil e contou com o parecer consubstanciado do Comitê de Ética em Pesquisa do Instituto de Biociências da Universidade de São Paulo (Número 363.876/2013).

List of Abbreviations

1kGP	1000 Genomes Project
6500ESP	Exome Variant Server database
AIM	Ancestry informative marker
CEGH60+	Centro de Estudos do Genoma Humano database
CDCV	Common disease-common variant
CDRV	Common disease-rare variant
CEU	Central Europeans from Utah
CHB	Han Chinese in Beijing, China
CI	Confidence interval
CLO	Cleft lip only
CLP	Cleft lip and palate
CL/P	Cleft lip with or without cleft palate
CPO	Cleft palate only
dpf	Days post fertilization
DSB	Double-strand break
EMT	Epithelial-mesenchymal transition
ExAC	Exome Aggregation Consortium
eQTL	Expression quantitative trait locus
FDR	False discovery rate
GTEx	Genotype-tissue expression project
GWAS	Genome-wide association studies
HDR	Homology-dependent repair
HGDC	Hereditary diffuse gastric cancer
JPT	Japanese in Tokyo, Japan
LD	Linkage disequilibrium
LoF	Loss of function
MAF	Minor allele frequency
MNE	Medionasal enhancer region
NCC	Neural crest cells
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
NSCL/P	Nonsyndromic cleft lip with or without cleft palate
NSCPO	Nonsyndromic cleft palate only
NSOFC	Nonsyndromic orofacial clefts
OFC	Orofacial clefts
OOM	*Orbicularis oris* muscle
OOMMSC	*Orbicularis oris* muscle mesenchymal stem cell
OR	Odds ratio
PCP	Planar cell polarity
RR	Relative risk
sgRNA	Single-guide RNA
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SRC	Spearman's rank correlation
TS	Target site
TSS	Transcription start site
WT	Wild type
YRI	Yoruba in Ibadan, Nigeria

Table of contents

Chapter 1

General Introduction

1. Orofacial Clefts

1.1. Clinical and Epidemiological Aspects

Orofacial clefts (OFC) are congenital defects that arise from failure during the embryological process of closure of lip and palate, resulting in the cleft of these structures. Cleft lip may be unilateral or bilateral, and either can be restricted to the lip (cleft lip only, CLO) or reach the alveolus (gum) and the pre-incisive foramen palate (cleft lip and palate, CLP; Figure 1A-B). In the most severe cases, palate is affected anteriorly (pre-incisive foramen cleft) and posteriorly (post-incisive foramen cleft), being called complete cleft palate (Figure 1C-D). Cleft palate can also occur without cleft lip (cleft palate only, CPO), and is usually restricted to the posterior palate (Figure 1E; (Schutte and Murray, 1999; Gorlin and Cohen Jr., 2001).

OFC constitute the most prevalent group of congenital craniofacial malformations, with a worldwide prevalence estimated as 1:700 liveborn babies (Mossey et al., 2009). Epidemiological findings support the division of OFC in two distinct disorders: cleft lip with or without cleft palate (CL/P) and CPO (Fogh-Andersen, 1942; Fraser, 1955; Gorlin and Cohen Jr., 2001). As it will be discussed in the following section, differences in embryonic development of lip and palate also support this division.

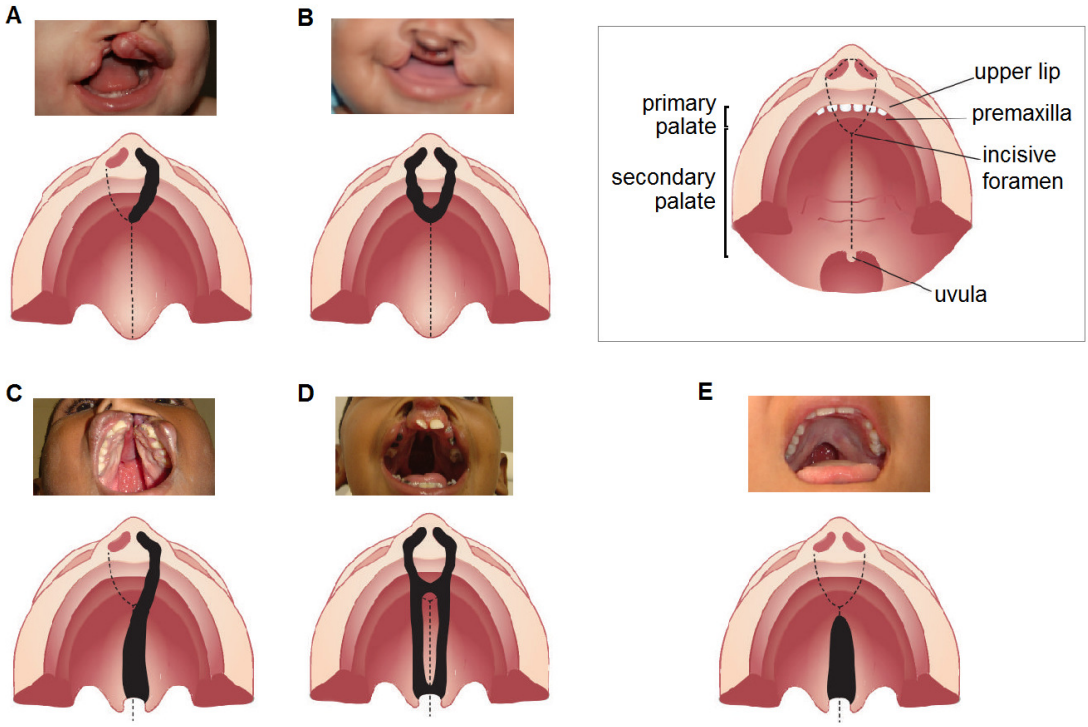

Figure 1 – Most common types of cleft affecting the palate. (A) Unilateral cleft lip with alveolar involvement; **(B)** Bilateral cleft lip with alveolar involvement; **(C)** Unilateral cleft lip with complete cleft palate; **(D)** Bilateral cleft lip with complete cleft palate; **(E)** Cleft palate only. Adapted from (Brito et al., 2012b).

The prevalence of CL/P varies substantially across populations: it is lower in Africans (~0.3:1,000), intermediate in Europeans (~1:1,000) and higher in East Asians (1.4-2.1:1,000) and Amerindians (~3.6:1,000; (Vanderas, 1987; Gorlin and Cohen Jr., 2001). In addition, low socioeconomic level is correlated with higher incidences of CL/P (Murray et al., 1997; Xu et al., 2012). On the other hand, the prevalence of CPO, frequently estimated as 1:2,000, does not show ethnic heterogeneity (Gorlin and Cohen Jr., 2001). In European populations, on which most of studies have been conducted, differences are also observed in sex ratio (with CL/P being more frequent in men – 60-80% of cases – and CPO prevailing in women), and in empirical recurrence risks among first-degree relatives (3-4% for CL/P and 2% for CPO; (Gorlin and Cohen Jr., 2001).

Based on the presence of additional malformations or comorbidities, OFC can be classified as syndromic or nonsyndromic. Nonsyndromic cases account for 70% of CL/P (nonsyndromic CL/P, NSCL/P) and 50% of CPO (nonsyndromic CPO, NSCPO; (Stanier and Moore, 2004; Jugessur et al., 2009). Generally, NSCL/P and NSCPO do not segregate

within a same family, reinforcing the etiological differences between these entities. Nevertheless, co-segregation of both forms may occur in some syndromic forms (Dixon et al., 2011). Up to date, there are more than 500 syndromes that include OFC as part of the phenotype, according to OMIM database (*Online Mendelian Inheritance in Man*).

Individuals affected by OFC often experience difficulties in feeding, which still implies in mortality, specially in developing countries (Carlson et al., 2013). Dental, speech and hearing problems may be also present. Since OFC are readily noticed facial defects, affected individuals habitually face serious adversities in social engagement, leading to a psychological burden (Marazita, 2012). Therefore, the complete rehabilitation of the patient with OFC demands reparative surgeries (multiple, starting from 3 months of age until adult life) coupled with a multidisciplinary treatment. Given the relatively high prevalence of OFC and its costly treatment (estimated as US$100,000 for a single patient; Centers for Disease Control and Prevention, http://www.cdc.gov; (Waitzman et al., 1994), these disorders represent an important problem to the health care system. Therefore, understanding the etiological factors and mechanisms that lead to OCF may, ultimately, help to treat and prevent these disorders.

1.2. Embryology

The normal closure of lip and palate comprehends a sequence of finely coordinated steps of cell growth, proliferation, migration, differentiation and apoptosis. Any punctual disturbance in the biological processes of this chain of events may perturb the subsequent events, eventually leading to OFC (Leslie and Marazita, 2013).

Lip morphogenesis starts in the 4th week of development, when the neural crest cells (NCC) delaminate from the neural folds and migrate to the developing craniofacial region through the mesenchymal tissue. NCC migration gives rise to the five facial primordia: one frontonasal, one pair of mandibular processes and one pair of maxillary processes (Figure 2A). In the following weeks, the frontonasal prominence originates, at its lower portion, the medial and lateral nasal processes (1 pair each; Figure 2B). Upper lip and primary palate are formed when the maxillary processes touch and fuse with the medial nasal processes, which occurs until the 7th week (Figure 2C; (Jiang et al., 2006). Therefore, failure during growth or fusion of these prominences results in cleft of the lip, which may reach the alveolus and primary palate (Figure 1A-D).

The morphogenesis of secondary palate begins in the 6th week, when the maxillary processes originate a pair of palatal shelves, located laterally to the developing tongue (Figure 2D). Initially, the palatal shelves grow vertically and, during the 7th week, they advance horizontally above the tongue, until they contact each other (Figure 2E). The subsequent fusion of the palatal shelves depends on the degeneration of an epithelial seam at the midline (Figure 2F), which is achieved by cell death and epithelial-mesenchymal transition (Mossey et al., 2009; Twigg and Wilkie, 2015), allowing a homogeneous mesenchyme in the palatal tissue, at the 10th week (Kerrigan et al., 2000). At this time, oral and nasal cavities are completely separated, but failures at these events will lead to cleft palate (Figure 1E).

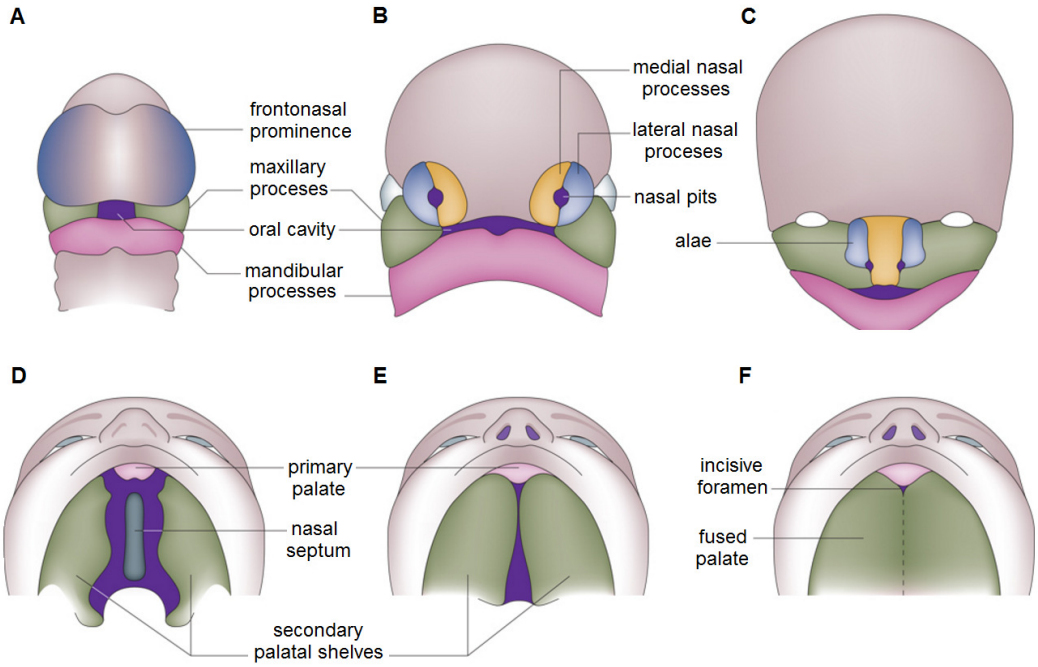

Figure 2 – Lip (A-C) and palate (D-F) embryogenesis. (A) Frontonasal prominence, maxillary processes and mandibular processes surrounding the oral cavity, at 4th week of development. **(B)** By the 5th week, medial nasal and lateral nasal processes are formed, as well as the nasal pits. **(C)** At the end of 6th week., medial nasal processes fuse with maxillary processes, giving rise to superior lip and primary palate; lateral nasal processes originate nasal alae, and mandibular processes originate the mandible. **(D)** By the 6th week, the palatal shelves originate from the maxillary processes, and grow vertically. **(E)** The elevated palatal shelves grow horizontally at the 7th week, positioned above the tongue, and reach each other. **(F)** By the 10th week, palatal shelves fuse with each other, following the degeneration of a midline epithelial layer. Adapted from (Dixon et al., 2011).

1.3. Etiology

Syndromic OFC may arise from gene mutations, chromosomal abnormalities or environmental factors, such as exposures to teratogens during the first trimester of pregnancy. NSCL/P and NSCPO, on the other hand, are complex disorders, with most of cases presenting multifactorial inheritance, where genetic and environmental susceptibility factors may play a role (Dixon et al., 2011).

Several environmental factors have been associated with increased risk of NSCL/P and NSCPO but contradictory findings are often observed. Among these factors, are maternal exposure to tobacco, alcohol consumption, obesity, infection, poor nutrition (and lack of nutrients such as folate, zinc and vitamins in general) and teratogens (as valproic acid; (Mossey et al., 2009; Dixon et al., 2011).

Evidence for a strong genetic role for NSCPO susceptibility has been obtained from studies on heritability and recurrence risk (Mitchell and Christensen, 1996; Nordstrom et al., 1996). However, probably due to the lower prevalence of NSCPO, most of studies have focused on NSCL/P, which is also our main interest.

The genetic contribution to NSCL/P has been evidenced by heritability studies in different populations. Twin studies indicate phenotypic concordance of 40-60% for monozygotic and 3-5% for dizygotic twins from Denmark (Christensen and Fogh-Andersen, 1993; Mitchell et al., 2002); high heritability has also been observed in other European countries (reaching 84% in Italy; (Calzolari et al., 1988), China (78%; (Hu et al., 1982) and Brazil (reaching 85%; (Brito et al., 2011). Another evidence for this genetic role comes from recurrence risk, which is 20-30 times higher in 1st-degree relatives of affected individuals than the population risk (Sivertsen et al., 2008; Grosen et al., 2010). Extensive research has been conducted in order to uncover the genetic basis of NSCL/P, and several susceptibility loci have emerged in the recent years.

2. Genetics of NSCL/P: Approaches and Risk Factors

2.1. Linkage and Candidate Gene Association Studies

A variety of approaches has been used to explore the genetic etiology of NSCL/P. Gene mapping strategies such as linkage and association studies has historically been the most popular. Linkage analysis relies on the co-segregation between genetic markers and the disease in families (Altshuler et al., 2008). Although several loci had been suggested by genome-wide linkage analysis, significant LOD-scores were firstly reached only in a meta-analysis, for the chromosomal regions 1q32, 2p13, 3q27-28, 9q21, 14q21-24 and 16q24 (Marazita et al., 2004).

Association analysis, under case-control or family-based design, was initially applied to candidate genes. Therefore, this approach required previous knowledge about the genes, before including them in the studies (Altshuler et al., 2008). Although many susceptibility loci were suggested by candidate gene studies, the vast majority was nonreplicable across studies (Leslie and Marazita, 2013). A single remarkable exception was *IRF6* (1q32), firstly associated with NSCL/P by (Zucchero et al., 2004), and consistently replicated thenceforth (Jugessur et al., 2008; Rahimov et al., 2008). Moreover, heterozygous loss-of-function mutations in *IRF6* lead to van der Woude syndrome (VWS1, MIM#119300), the most common syndromic form of OFC (Kondo et al., 2002).

2.2. GWAS and the Common Susceptibility Variants

The scenario dramatically changed with the genome-wide association studies (GWAS), which allowed association studies to be performed in genomic level, without bias regarding the need of *a priori* knowledge of candidate genes (Kruglyak, 2008). GWAS relies on the common disease-common variant (CDCV) hypothesis for complex diseases, which predicts that the allelic spectrum of the disease (i.e., all disease-contributing variants) is predominantly composed of frequent variants (originated from a common ancestor and maintained in the population) of low individual effects (Reich and Lander, 2001; Schork et al., 2009);. In this manner, these studies were made possible thanks to a deep characterization of the patterns of genetic variation in human

genome, provided by the Human Genome Project (Lander et al., 2001) and the HapMap Project (http://hapmap.ncbi.nlm.nih.gov).

(Birnbaum et al., 2009) conducted the first GWAS on NSCL/P, and found association of a group of markers in a 640-kb interval at a gene desert in 8q24 region, which was confirmed shortly after by a second GWAS (Grant et al., 2009). The third GWAS came from an expansion of Birnbaum's sample, and implied two new loci (10q25 and 17q22), besides having replicated the associations of *IRF6* and 8q24 (Mangold et al., 2009). Differently from the three previous GWAS, which used case-control design and only populations of European origin, (Beaty et al., 2010) carried out a family-based GWAS with a mixed sample of European and Asian individuals. This study reported, for the first time, significant associations of 1p22.1 and 20q12, and suggested that association of these and previously reported loci may vary across populations. A meta-analysis of Mangold's and Beaty's data uncovered new associations, expanding to 12 the number of variants implicated by GWAS (Ludwig et al., 2012). Moreover, it confirmed the 8q24 locus as the strongest association in NSCL/P (Box 1). Recently, (Sun et al., 2015) conducted the fifth GWAS on NSCL/P, the first on a totally non-European sample (the Chinese population). A new susceptibility locus was revealed in this study (16p13), reinforcing the importance of testing populations other than Europeans. All loci associated by GWAS are summarized in Table 1.

Several studies have endeavored to replicate these associations in different populations. Not rarely, they failed in detecting association for some loci. As an example, 8q24 association was extensively replicated in European populations, but not in Asians or Africans (see Box 1). A drawback in many replication studies, however, is that they generally focus on testing the top-SNP at each GWAS-associated locus. In consequence, if this SNP lays in a different haplotypic block than in European populations, lack of association will probably be observed (Kruglyak, 2008). In addition, if the top-SNP is rare in a given population, the study's statistical power to detect association will dramatically decrease, as association studies are powered to detect common variant (Murray et al., 2012). Therefore, these possibilities should be considered before assuming non-association of a candidate locus in a new population.

In general, the major NSCL/P susceptibility loci uncovered by GWAS have been shown to increase only a small risk, which, collectively, do not explain a significant proportion of the populational risk to the disease (Leslie and Marazita, 2013). The arising question of where this non-explained genetic risk would be hiding was termed as

"missing heritability", and it is a common debate for most of complex disorders, (Maher, 2008; Manolio et al., 2009). Nonetheless, if in one hand GWAS have failed to explain a vast component of NSCL/P heritability, on the other, they did provide insights on new pathways involved with the disease (Visscher et al., 2012; Yang et al., 2014).

Table 1 – Genomic loci significantly associated with NSCL/P by GWAS

Region	Top SNP	Main candidate gene(s)	*P*-value	Risk (95% CI)	Associations in GWAS
1p22.1	rs560426	*ARHGAP29*	3.1×10^{-12}	RR_{het}=1.42 (1.24–1.62); RR_{hom}=1.86 (1.56–2.23) [a]	Beaty et al., 2010; Ludwig et al., 2012
1p36	rs742071	*PAX7*	7.0×10^{-9}	RR_{het}=1.316 (1.13–1.54); RR_{hom}=1.878 (1.52–2.32) [a]	Ludwig et al., 2012
1q32.2	rs861020	*IRF6*	3.2×10^{-12}	RR_{het}=1.44 (1.27–1.64); RR_{hom}=2.04 (1.60–2.60) [a]	Birnbaum et al., 2009; Mangold et al., 2009; Beaty et al., 2010; Ludwig et al., 2012; Sun et al., 2015
2p21	rs7590268	*THADA*	1.3×10^{-8}	RR_{het}=1.42 (1.23–1.64); RR_{hom}=1.98 (1.47–2.66) [a]	Ludwig et al., 2012
3p11.1	rs7632427	*EPHA3*	3.9×10^{-8}	RR_{het}=0.73 (0.64–0.83); RR_{hom}=0.61 (0.49– 0.76) [a]	Ludwig et al., 2012
8q21.3	rs12543318		1.9×10^{-8}	RR_{het}=1.27 (1.11–1.46); RR_{hom}=1.68 (1.40–2.01) [a]	Ludwig et al., 2012
8q24	rs987525	*MYC*	5.1×10^{-35}	RR_{het}=1.92 (1.66–2.22); RR_{hom}=4.38 (3.39–5.67) [a]	Birnbaum et al., 2009 Grant et al., 2009 Mangold et al., 2009; Beaty et al., 2010; Ludwig et al., 2012 Sun et al., 2015
10q25	rs7078160	*VAX1*	4.0×10^{-11}	RR_{het}=1.38 (1.21–1.58); RR_{hom}=1.94 (1.58–2.39) [a]	Mangold et al., 2010 Ludwig et al., 2012 Sun et al., 2015
13q31.1	rs8001641	*SPRY2*	2.6×10^{-10}	RR_{het}=1.31 (1.13–1.51); RR_{hom}=1.86 (1.54–2.26) [a]	Ludwig et al., 2012
15q22.2	rs1873147	*TPM1*	7.9×10^{-7}	RR_{het}=1.43 (1.23–1.67); RR_{hom}=1.65 (1.34–2.04) [a]	Ludwig et al., 2012
16p13	rs8049367	*CREBBP ADCY9*	9.0×10^{-12}	OR_{add}=0.74 (0.68-0.80) [b]	Sun et al., 2015
17p13*	rs4791774	*NTN1*	5.1×10^{-19}	OR_{add}=1.56 (0.71-0.83) [b]	Sun et al., 2015
17q22	rs227731	*NOG*	1.8×10^{-8}	RR_{het}=1.23 (1.08–1.40); RR_{hom}=1.67 (1.40– 2.0) [a]	Mangold et al., 2010 Ludwig et al., 2012
20q12	rs13041247	*MAFB*	6.2×10^{-9}	RR_{het}=0.84 (0.74–0.94); RR_{hom}=0.55 (0.45–0.66) [a]	Beaty et al., 2010; Ludwig et al., 2012; Sun et al., 2015

RR: Relative risk; **hom:** homozygous; **het:** heterozygous; **ORadd:** Odds ratio using additive model.
[a] Data retrieved from Ludwing et al. (2012).
[b] Data retrieved from Sun et al. (2015)
* Also marginally associated in Beaty et al., 2010

BOX1: 8q24 locus

The association of a 640-kb interval at 8q24 represents the most prominent finding of GWAS on NSCL/P. The association of the top-SNP rs987525 has been consistently replicated in populations from Europe (Cura et al., 2015), Central America (Rojas-Martinez et al., 2010), Brazil (Brito et al., 2012c) and Middle-East (Aldhorae et al., 2014). Nonetheless, replication studies have failed in finding association in Asian and African populations (Beaty et al., 2010; Weatherley-White et al., 2011; Figueiredo et al., 2014). At least in Asian populations, this lack of association is thought to be consequence of low statistical power, due to low allele frequency, since larger studies find suggestive signals of association (Murray et al., 2012). In addition, (Boehringer et al., 2011) and (Liu et al., 2012) have reported association of 8q24 locus with normal variation of human facial traits. Because no known gene is present at this region, a regulatory role has been proposed since its identification (Birnbaum et al., 2009). In fact, (Uslu et al., 2014), studying the syntenic murine locus, found that a 280-kb region within the NSCL/P associated interval is enriched for long-range regulatory elements of the proximal gene *MYC*. In addition, deletions of these elements frequently led to facial dysmorphologies, including cleft lip / palate. At the cellular level, the authors verified that deletion of these regions were correlated with lower Myc expression and enriched expression of genes involved with ribosome assembly and transcriptional, suggesting that abnormal cell proliferation is a possible mechanism by which deletion of these elements causes

2.3. *Resequencing Studies and the Rare Variants*

One hypothesis that addresses to the missing heritability question relies on the role of rare variants, typically defined as <1% (Maher, 2008). Alternatively to the CDCV hypothesis, some researchers argue that the major genetic contributors to common diseases would be rare, moderate-to-high effect variants distributed in the population. According to this common disease-rare variant (CDRV) hypothesis, a combination of only few rare, high-effect variants would be necessary to cause the disease in an individual, and most of disease's phenotypic variation and expressivity observed in population would be attributed to different allele combinations, under additional influence of environmental factors (Bodmer and Bonilla, 2008; Schork et al., 2009; Gibson, 2012).

Sequencing strategies have been the most suitable approach to detect rare variants implicated with diseases (Manolio et al., 2009). Resequencing of genes associated with NSCL/P by GWAS has found possibly pathogenic rare variants in *ARGHAP29* (Leslie and Murray, 2012), *MAFB* and *PAX7* (Butali et al., 2014). In addition, the advent of next-generation sequencing (NGS) technologies stimulated a genome-wide hunt for rare variants, by means of exome and genome sequencing. With the progressive drop in NGS costs, coupled with increase in throughput, this approach has become accessible by many research groups studying common diseases (Do et al., 2012; O'Roak et al., 2012). Recently, exome sequencing in NSCL/P patients has enabled the identification of possibly pathogenic variants in new genes, such as *CDH1,* at 16q22.1, (Bureau et al., 2014) and *DLX4,* at 17q21.33 (Wu et al., 2014), among other putative candidates (Liu et al., 2015).

Nevertheless, attributing a pathogenic role for a given rare variant is not a trivial task. Firstly, classifying a given variant as rare requires the availability of large population databases; in this regard, databases such as the 1000 Genomes Project (1kGP; http://www.1000genomes.org/) and the Exome Variant Server / NHLBI Exome Sequencing Project (ESP6500; http://evs.gs.washington.edu/EVS/) are valuable starting points. However, many populations, including Brazilian, are poorly represented in these databases. Therefore, local databases are of great relevance for identifying local common variants. Secondly, incomplete penetrance and genetic heterogeneity are expected to occur within families segregating complex diseases under the CDRV model, which may represent a confounding factor in exome sequencing studies that seek for co-segregation of variants in families (Cooper et al., 2013). Even though, multiplex families still retain the best chances of finding a rare, pathogenic variant.

Objectives

Our main objective is to find the major susceptibility variants / loci underlying NSCL/P in the Brazilian population. We aimed to explore the broad spectrum of allele frequency (either rare or common variants) in NSCL/P, by means of different strategies. In this respect, our objective can be divided as follows:

a) Identify rare, moderate-to-high effect variants underlying NSCL/P in familial cases, under two main hypothesis: (i) affected relatives sharing a major causative locus, which may vary among families, and (ii) affected relatives presenting at least two moderate-effect risk variants, not necessarily the same (i.e., genetic heterogeneity of moderate-effect risk variants within a family).

b) Investigate the role of common, low-risk variants in NSCL/P etiology, by (i) characterizing the 8q24 susceptibility locus in the Brazilian population, attempting to narrow the 640-kb interval previously associated; (ii) replicating some of the GWAS hits and (iii) seek for new susceptibility factors, combining association analysis and expression quantitative trait loci mapping.

Chapter 7

General Discussion and Conclusions

Although several research groups have been engaged in the search for genetic factors underlying NSCL/P, much of its heritability is still barely understood. In this regard, the CDCV x CDRV debate in NSCL/P addresses part of this question, exploring the role of common and rare variants in NSCL/P etiology.

We successfully identified rare variants in the Epithelial-Cadherin gene leading to NSCL/P. This work consisted in one of the first publications to correlate *CDH1* variants with NSCL/P, and comprises, up to now, the largest collection of NSCL/P patients with *CDH1* mutations. A remarkable finding of this study was that a high proportion (15%) of our families with more than 2 affected individuals harbors a causal *CDH1* variant. Similar studies with different populations would be of extreme importance in order to corroborate the importance of this gene in familial cases of NSCL/P. Alternatively, a higher prevalence of *CDH1* mutations in Brazilian population might be related to founder effect of a few pathogenic alleles.

From a total of nine families with exomes sequenced, we were able to identify the causal variant in two (both harboring a mutation in *CDH1*). The success obtained for these two families was a direct consequence of the number of affected members and the availability of DNA samples. In addition, the big size of the families allowed us to sequence distantly related individuals, which dramatically reduced the variants that remained after filtering steps. That was not the situation for the majority of our families; for those, we needed to use literature data for prioritizing the best candidate variants. A

potential drawback of this strategy is that we tend to prioritize only genes with known functional data, particularly those with some relation with craniofacial structures. Broadly speaking, if the analysis is based on our current knowledge, we diminish the chances of implicating a poorly studied gene, with unknown function or relation with craniofacial development. To overcome this limitation, we are putting effort in sequencing the exome of extra relatives from some of our families (e.g., F886 and F1843). Notwithstanding, we believe that the enrichment of candidate genes related to PCP pathway, microtubules and cell adhesion are unbiased findings, since it was found after variant prioritization. Taken these results together, we suggest that genetic heterogeneity is underlying NSCL/P in at least one of our families (F886), and rare variants associated with high penetrance may explain the phenotype in six out of our nine families. In our evaluation, these results deeply encourage the application of exome sequencing in further familial cases.

In the meanwhile, we established cdh1-mutant zebrafish lines, in order to investigate how cdh1 mutations lead to OFC. Among all homozygous mutants we generated, we only observed phenotype (embryo lethality) for the frameshift deletion (double knockout). We then generated compound heterozygous with the frameshift deletion and the in-frame mutations, to explore the possibility that one of our in-frame mutations could lead to a degree of protein impairment, but without causing embryo lethality. However, none of the compound heterozygous led to a phenotype, suggesting that these variants do not compromise the function of the protein. In face of this limitation, we are currently performing alternative strategies to phenotypically evaluate the absence of E-cadherin in zebrafish, to test the two-hit model for variants in cdh1. As an example, we plan to inject a small quantity of wild-type mRNA in double-knockout embryos. We believe that, overpassing the critical period of gastrulation, we will be able to assess cdh1 knockout phenotypes in late embryo stages.

To explore the role of common variants in NSCL/P etiology, we choose to characterize the best-associated locus 8q24, and search for new loci through an eQTL mapping-based association analysis. The high admixture degree of Brazilian population introduces a powerful confounding factor that needs to be accounted for. For this reason, we have used a structured association approach, which takes advantage of information on individual ancestry components of cases and controls before performing the association tests. Differently from our previous reports (Brito et al., 2012a; Brito et

al., 2012c) we used here an AIM panel composed of biallelic SNPs and a larger sample set.

Our association study narrowed the previously associated 8q24 locus to a 310-kb interval, composed of multiple linkage disequilibrium blocks. The most significant SNP reported here (rs987525; $p=4.8x10^{-8}$; $OR_{het}=2.10$ [1.65-2.68 95%CI]; $OR_{hom}=3.23$ [2.19-4.79 95%CI]) coincided with the same found in GWAS. Our odds ratio estimates corroborate the moderate effect previously suggested (Birnbaum et al., 2009); nevertheless, we should bear in mind that they are based exclusively in allele frequencies, without any correction for population structure, and thus subjected to stratification bias, which could lead to under or overestimated values. Our associated interval overlaps with a putative regulatory element, hs1877, in the previously defined *MYC* Medionasal Enhancer Region. These results may indicate that this regulatory element confers the most critical risk among the regulatory elements implicated with facial morphogenesis of this gene desert. On the other hand, we diminish the relevance of *IRF6* SNP rs642961 in susceptibility to NSCL/P in our population. Our results are consistent with recent findings that suggest that the association observed for *IRF6* may be driven by other variants (Sun et al., 2015), even though a functional role has been attributed to this variant (Rahimov et al., 2008). Finally, we also reported, for the first time in our population, the association of a marker in 20q12 region. It has been suggested that *MAFB* gene may be driving this association (Beaty et al., 2010). Accordingly, we have recently included *MAFB* gene in a targeted NGS panel and we will be able, in the near future, to analyze its coding region in ~200 NSCL/P patients.

Lastly, we applied an eQTL-based association study to seek for new NSCL/P susceptibility genes. We opted for this approach as an alternative to GWAS, since it directly investigates regulatory variants. We revealed, for the first time, the association of eQTLs for *MRPL53* gene (2p13.1). To endorse the validity of this association (and, in consequence, of this approach), we have tried to find a second evidence of an etiological role for this gene. Nevertheless, we have failed in finding association in a re-analysis of the meta-analysis study (Ludwig et al., 2012) with imputed SNPs at 2p13.1 locus (Ludwig, personal communication); furthermore, no pathogenic variant was found in the resequencing of *MRPL53* in Brazilian NSCL/P patients. Therefore, the emerging challenge is to establish a functional link between this gene and the disease.

In conclusion, the present work evidences the role of rare variants in NSCL/P etiology, suggesting *CDH1* as a major contributor for moderate-to-high effect variants. In

addition, we also provide insights into the association of major susceptibility locus at 8q24, and report association of 2p13.1 locus, possibly implicated with *MRPL53* gene.

References

Aldhorae KA, Bohmer AC, Ludwig KU, Esmail AH, Al-Hebshi NN, Lippke B, Golz L, Nothen MM, Daratsianos N, Knapp M*et al*. 2014. Nonsyndromic cleft lip with or without cleft palate in arab populations: genetic analysis of 15 risk loci in a novel case-control sample recruited in Yemen. Birth Defects Res A Clin Mol Teratol 100:307-313.

Altshuler D, Daly MJ, Lander ES. 2008. Genetic mapping in human disease. Science 322:881-888.

Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Liang KY, Wu T, Murray T, Fallin MD*et al*. 2010. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat Genet 42:525-529.

Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M, Baluardo C, Ferrian M, Almeida de Assis N, Alblas MA*et al*. 2009. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet 41:473-477.

Bodmer W, Bonilla C. 2008. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 40:695-701.

Boehringer S, van der Lijn F, Liu F, Gunther M, Sinigerova S, Nowak S, Ludwig KU, Herberz R, Klein S, Hofman A*et al*. 2011. Genetic determination of human facial morphology: links between cleft-lips and normal variation. Eur J Hum Genet 19:1192-1197.

Brito LA, Bassi CF, Masotti C, Malcher C, Rocha KM, Schlesinger D, Bueno DF, Cruz LA, Barbara LK, Bertola DR *et al*. 2012a. IRF6 is a risk factor for nonsyndromic cleft lip in the Brazilian population. Am J Med Genet A 158A:2170-2175.

Brito LA, Cruz LA, Rocha KM, Barbara LK, Silva CB, Bueno DF, Aguena M, Bertola DR, Franco D, Costa AM *et al*. 2011. Genetic contribution for non-syndromic cleft lip with or without cleft palate (NS CL/P) in different regions of Brazil and implications for association studies. Am J Med Genet A 155A:1581-1587.

Brito LA, Meira JG, Kobayashi GS, Passos-Bueno MR. 2012b. Genetics and management of the patient with orofacial cleft. Plast Surg Int 2012:782821.

Brito LA, Paranaiba LM, Bassi CF, Masotti C, Malcher C, Schlesinger D, Rocha KM, Cruz LA, Barbara LK, Alonso N *et al*. 2012c. Region 8q24 is a susceptibility locus for nonsyndromic oral clefting in Brazil. Birth Defects Res A Clin Mol Teratol 94:464-468.

Bureau A, Parker MM, Ruczinski I, Taub MA, Marazita ML, Murray JC, Mangold E, Noethen MM, Ludwig KU, Hetmanski JB *et al*. 2014. Whole exome sequencing of distant relatives in multiplex families implicates rare variants in candidate genes for oral clefts. Genetics 197:1039-1044.

Butali A, Mossey P, Adeyemo W, Eshete M, Gaines L, Braimah R, Aregbesola B, Rigdon J, Emeka C, Olutayo J *et al*. 2014. Rare functional variants in genome-wide association identified candidate genes for nonsyndromic clefts in the African population. Am J Med Genet A 164A:2567-2571.

Calzolari E, Milan M, Cavazzuti GB, Cocchi G, Gandini E, Magnani C, Moretti M, Garani GP, Salvioli GP, Volpato S. 1988. Epidemiological and genetic study of 200 cases of oral cleft in the Emilia Romagna region of northern Italy. Teratology 38:559-564.

Carlson L, Hatcher KW, Vander Burg R. 2013. Elevated infant mortality rates among oral cleft and isolated oral cleft cases: a meta-analysis of studies from 1943 to 2010. Cleft Palate Craniofac J 50:2-12.

Christensen K, Fogh-Andersen P. 1993. Cleft lip (+/- cleft palate) in Danish twins, 1970-1990. Am J Med Genet 47:910-916.

Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. 2013. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum Genet 132:1077-1130.

Cura F, Bohmer AC, Klamt J, Schunke H, Scapoli L, Martinelli M, Carinci F, Nothen MM, Knapp M, Ludwig KU*et al*. 2015. Replication analysis of 15 susceptibility loci for nonsyndromic cleft lip with or without cleft palate in an italian population. Birth Defects Res A Clin Mol Teratol.

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

Do R, Kathiresan S, Abecasis GR. 2012. Exome sequencing and complex disease: practical aspects of rare variant association studies. Hum Mol Genet 21:R1-9.

Figueiredo JC, Ly S, Raimondi H, Magee K, Baurley JW, Sanchez-Lara PA, Ihenacho U, Yao C, Edlund CK, van den Berg D*et al*. 2014. Genetic risk factors for orofacial clefts in Central Africans and Southeast Asians. Am J Med Genet A 164A:2572-2580.

Fogh-Andersen P. 1942. Inheritance of Harelip and Cleft Palate. Copenhagen: Munksgaard.

Fraser FC. 1955. Thoughts on the etiology of clefts of the palate and lip. Acta Genet Stat Med 5:358-369.

Gibson G. 2012. Rare and common variants: twenty arguments. Nat Rev Genet 13:135-145.

Gorlin RJ, Cohen Jr. MMH, R. C. M. 2001. Syndromes of the Head and Neck. New York, NY, US: Oxford University Press. p 905-907.

Grant SF, Wang K, Zhang H, Glaberson W, Annaiah K, Kim CE, Bradfield JP, Glessner JT, Thomas KA, Garris M*et al*. 2009. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. J Pediatr 155:909-913.

Grosen D, Chevrier C, Skytthe A, Bille C, Molsted K, Sivertsen A, Murray JC, Christensen K. 2010. A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance. J Med Genet 47:162-168.

Hu DN, Li JH, Chen HY, Chang HS, Wu BX, Lu ZK, Wang DZ, Liu XG. 1982. Genetics of cleft lip and cleft palate in China. Am J Hum Genet 34:999-1002.

Jiang R, Bush JO, Lidral AC. 2006. Development of the upper lip: morphogenetic and molecular mechanisms. Dev Dyn 235:1152-1166.

Jugessur A, Rahimov F, Lie RT, Wilcox AJ, Gjessing HK, Nilsen RM, Nguyen TT, Murray JC. 2008. Genetic variants in IRF6 and the risk of facial clefts: single-marker and haplotype-based analyses in a population-based case-control study of facial clefts in Norway. Genet Epidemiol 32:413-424.

Jugessur A, Shi M, Gjessing HK, Lie RT, Wilcox AJ, Weinberg CR, Christensen K, Boyles AL, Daack-Hirsch S, Trung TN*et al*. 2009. Genetic determinants of facial clefting: analysis of 357 candidate genes using two national cleft studies from Scandinavia. PLoS One 4:e5385.

Kerrigan JJ, Mansell JP, Sengupta A, Brown N, Sandy JR. 2000. Palatogenesis and potential mechanisms for clefting. J R Coll Surg Edinb 45:351-358.

Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL, Daack-Hirsch S, Sander A*et al*. 2002. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet 32:285-289.

Kruglyak L. 2008. The road to genome-wide association studies. Nat Rev Genet 9:314-318.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W*et al*. 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.

Leslie EJ, Marazita ML. 2013. Genetics of cleft lip and cleft palate. Am J Med Genet C Semin Med Genet 163C:246-258.

Leslie EJ, Murray JC. 2012. Evaluating rare coding variants as contributing causes to non-syndromic cleft lip and palate. Clin Genet 84:496-500.

Liu F, van der Lijn F, Schurmann C, Zhu G, Chakravarty MM, Hysi PG, Wollstein A, Lao O, de Bruijne M, Ikram MA*et al*. 2012. A genome-wide association study identifies five loci influencing facial morphology in Europeans. PLoS Genet 8:e1002932.

Liu YP, Xu LF, Wang Q, Zhou XL, Zhou JL, Pan C, Zhang JP, Wu QR, Li YQ, Xia YJ*et al*. 2015. Identification of susceptibility genes in non-syndromic cleft lip with or without cleft palate using whole-exome sequencing. Med Oral Patol Oral Cir Bucal 20:e763-770.

Ludwig KU, Mangold E, Herms S, Nowak S, Reutter H, Paul A, Becker J, Herberz R, AlChawa T, Nasser E*et al*. 2012. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat Genet 44:968-971.

Maher B. 2008. Personal genomes: The case of the missing heritability. Nature 456:18-21.

Mangold E, Ludwig KU, Birnbaum S, Baluardo C, Ferrian M, Herms S, Reutter H, de Assis NA, Chawa TA, Mattheisen M*et al*. 2009. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat Genet 42:24-26.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A*et al*. 2009. Finding the missing heritability of complex diseases. Nature 461:747-753.

Marazita ML. 2012. The evolution of human genetic studies of cleft lip and cleft palate. Annu Rev Genomics Hum Genet 13:263-283.

Mitchell LE, Beaty TH, Lidral AC, Munger RG, Murray JC, Saal HM, Wyszynski DF. 2002. Guidelines for the design and analysis of studies on nonsyndromic cleft lip and cleft palate in humans: summary report from a Workshop of the International Consortium for Oral Clefts Genetics. Cleft Palate Craniofac J 39:93-100.

Mitchell LE, Christensen K. 1996. Analysis of the recurrence patterns for nonsyndromic cleft lip with or without cleft palate in the families of 3,073 Danish probands. Am J Med Genet 61:371-376.

Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. 2009. Cleft lip and palate. Lancet 374:1773-1785.

Murray JC, Daack-Hirsch S, Buetow KH, Munger R, Espina L, Paglinawan N, Villanueva E, Rary J, Magee K, Magee W. 1997. Clinical and epidemiologic studies of cleft lip and palate in the Philippines. Cleft Palate Craniofac J 34:7-10.

Murray T, Taub MA, Ruczinski I, Scott AF, Hetmanski JB, Schwender H, Patel P, Zhang TX, Munger RG, Wilcox AJ*et al*. 2012. Examining markers in 8q24 to explain differences in evidence for association with cleft lip with/without cleft palate between Asians and Europeans. Genet Epidemiol 36:392-399.

Nordstrom RE, Laatikainen T, Juvonen TO, Ranta RE. 1996. Cleft-twin sets in Finland 1948-1987. Cleft Palate Craniofac J 33:340-347.

O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C*et al*. 2012. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 43:585-589.

Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, Domann FE, Govil M, Christensen K, Bille C*et al*. 2008. Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. Nat Genet 40:1341-1347.

Reich DE, Lander ES. 2001. On the allelic spectrum of human disease. Trends Genet 17:502-510.

Rojas-Martinez A, Reutter H, Chacon-Camacho O, Leon-Cachon RB, Munoz-Jimenez SG, Nowak S, Becker J, Herberz R, Ludwig KU, Paredes-Zenteno M*et al*. 2010. Genetic risk factors for nonsyndromic cleft lip with or without cleft palate in a Mesoamerican population: Evidence for IRF6 and variants at 8q24 and 10q25. Birth Defects Res A Clin Mol Teratol 88:535-537.

Schork NJ, Murray SS, Frazer KA, Topol EJ. 2009. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 19:212-219.

Schutte BC, Murray JC. 1999. The many faces and factors of orofacial clefts. Hum Mol Genet 8:1853-1859.

Sivertsen A, Wilcox AJ, Skjaerven R, Vindenes HA, Abyholm F, Harville E, Lie RT. 2008. Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives. Bmj 336:432-434.

Stanier P, Moore GE. 2004. Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. Hum Mol Genet 13 Spec No 1:R73-81.

Sun Y, Huang Y, Yin A, Pan Y, Wang Y, Wang C, Du Y, Wang M, Lan F, Hu Z*et al*. 2015. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun 6:6414.

Twigg SR, Wilkie AO. 2015. New insights into craniofacial malformations. Hum Mol Genet.

Uslu VV, Petretich M, Ruf S, Langenfeld K, Fonseca NA, Marioni JC, Spitz F. 2014. Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. Nat Genet 46:753-758.

Vanderas AP. 1987. Incidence of cleft lip, cleft palate, and cleft lip and palate among races: a review. Cleft Palate J 24:216-225.

Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. Am J Hum Genet 90:7-24.

Waitzman NJ, Romano PS, Scheffler RM. 1994. Estimates of the economic costs of birth defects. Inquiry 31:188-205.

Weatherley-White RC, Ben S, Jin Y, Riccardi S, Arnold TD, Spritz RA. 2011. Analysis of genomewide association signals for nonsyndromic cleft lip/palate in a Kenya African Cohort. Am J Med Genet A 155A:2422-2425.

Wu D, Mandal S, Choi A, Anderson A, Prochazkova M, Perry H, Gil-Da-Silva-Lopes VL, Lao R, Wan E, Tang PL*et al*. 2014. DLX4 is associated with orofacial clefting and abnormal jaw development. Hum Mol Genet 24:4340-4352.

Xu MY, Deng XL, Tata LJ, Han H, Chen XH, Liu TY, Chen QS, Yao XW, Tang SJ. 2012. Case-control and family-based association studies of novel susceptibility locus 8q24 in nonsyndromic cleft lip with or without cleft palate in a Southern Han Chinese population located in Guangdong Province. DNA Cell Biol 31:700-705.

Yang T, Jia Z, Bryant-Pike W, Chandrasekhar A, Murray JC, Fritzsch B, Bassuk AG. 2014. Analysis of PRICKLE1 in human cleft palate and mouse development demonstrates rare and common variants involved in human malformations. Mol Genet Genomic Med 2:138-151.

Zucchero TM, Cooper ME, Maher BS, Daack-Hirsch S, Nepomuceno B, Ribeiro L, Caprau D, Christensen K, Suzuki Y, Machida J*et al*. 2004. Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. N Engl J Med 351:769-780.

Chapter 8

Abstract

Orofacial clefts (or cleft lip / palate) are congenital malformations with high prevalence in population (~1:700 births). Among the orofacial cleft types, an etiologically distinct group is composed by cleft lip with or without cleft palate, which, in 70% of cases, is not accompanied by other malformations (nonsyndromic cleft lip with or without cleft palate, NSCL/P). NSCL/P presents complex etiology, often with multifactorial inheritance. Although important, the genetic contribution to NSCL/P is still poorly comprehended, and the susceptibility loci that have been associated with NSCL/P do not explain the totality of the disease's heritability. In light of this, our aim was to investigate risk variants for NSCL/P by means of different strategies. With exome sequencing for NSCL/P familial cases, we report that the epithelial cadherin-encoding gene contributes with rare, moderate-to-high risk variants to NSCL/P etiology. In addition, we suggest an etiological contribution of genes laying in planar cell polarity pathway, or involved with epithelial-mesenchymal transition, cell adhesion, cell cycle regulation, and interaction with microtubules. Using structured association approach, we narrowed the associated interval of 8q24 region in a Brazilian population, and also validated the association for 20q12. Finally, by combining association analysis with eQTL mapping, we found association of regulatory variants of *MRPL53,* in 2p13, with NSCL/P. In conclusion, this study contributes with a deeper comprehension of the etiological role of rare and common variants for NSCL/P.

Resumo

As fissuras orofaciais, ou fissuras labiopalatinas, são malformações prevalentes na população mundial, presente em cerca de um a cada 700 nascimentos. Dentro das fissuras orofaciais, um grupo etiologicamente distinto é composto pelas fissuras de lábio com ou sem fissura de palato, que, em 70% dos casos, não estão associadas a nenhuma comorbidade (fissuras de lábio com ou sem palato não sindrômicas, FL/P NS). A etiologia das FL/P NS é complexa, e em muitos casos apresenta herança multifatorial. A contribuição genética para as FL/P NS, embora sabidamente relevante, ainda é pouco conhecida. Ainda, os loci de suscetibilidade consistentemente associados às FL/P NS, não conferem um risco que explique a herdabilidade total da doença. O objetivo do presente trabalho foi investigar, por meio de diferentes estratégias, variantes de risco às FL/P NS. Utilizando sequenciamento de exoma em casos familiais, verificamos que o gene codificante da caderina epitelial, *CDH1,* contribui importantemente com variantes raras de efeito moderado a alto na etiologia das FL/Ps. Além disso, propusemos que também podem ter relevância etiológica genes envolvidos na via de polaridade planar celular, transição epitélio-mesênquima, adesão celular, regulação de ciclo celular ou de interação com microtúbulos. Por meio de um estudo de associação com correção para estratificação populacional, caracterizamos o intervalo de associação da região 8q24, o principal locus de suscetibilidade às FL/P, e identificamos associação significativa também para a região 20q12. Por fim, combinando o estudo de associação com mapeamento de eQTLs, encontramos pela primeira vez a associação entre marcadores na região 2p13, que regulam *MRPL53*, em FL/P NS. Em conclusão, este trabalho contribui para o melhor entendimento da relevância de variantes raras, de efeito moderado a alto, e comuns, de efeito pequeno, na etiologia das FL/P NS.