5-2011

Gene Discovery in Nonsyndromic Cleft Lip with or without Cleft Palate

Brett T. Chiquet

Gene Discovery in Nonsyndromic Cleft Lip with or without Cleft Palate

by

Brett Thomas Chiquet, B.A.

APPROVED:

Supervisory Professor Jacqueline T. Hecht, Ph.D.

Stephen Daiger, Ph.D.

Richard H. Finnell, Ph.D.

Michael Gambello, M.D., Ph.D.

Karen A. Storthz, Ph.D.

APPROVED:

Dean, The University of Texas
Health Science Center at Houston
Graduate School of Biomedical Sciences

Gene Discovery in Nonsyndromic Cleft Lip with or without Cleft Palate

A
DISSERTATION

Presented to the Faculty of
The University of Texas Health Science Center at Houston and
The University of Texas M.D. Anderson Cancer Center
Graduate School of Biomedical Sciences

In Partial Fulfillment
Of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

by

Brett Thomas Chiquet, B.A.

Houston, TX
May 2011

Acknowledgements

I would first like to acknowledge my mentor, Dr. Jacqui Hecht, for having faith in me and taking a chance on this dental student who was interested in research. Her guidance and mentorship over these past years have been invaluable, and have molded every part of who I am as a scientist- from project identification and design, to grant and manuscript writing, public speaking and the art of collaboration. I am deeply appreciative of her efforts and confident that I am a better researcher because of them. To her lab (past and present) for supporting me, answering questions, helping with protocols and troubleshooting, I too am grateful, especially Drs. Shahrukh Hashmi, Karen Posey and Thomas Merritt, Francoise Coustry, Elise Bales, Huiqiu (Rachel) Wang, Tamar Powell, Candace Hurst, Elizabeth Garcia, Maria Elena Serna, Rosa Garcia, Alka Veerisetty, Qiuping Lu and Peiman Liu. I especially owe my sanity to Dr. Audrey Ester and (soon-to-be Dr.) Katelyn Weymouth. Audrey and Katelyn have been my amigos, partners in crime, sisters I never planned on, and Starbucks buddies. I would not have been able to complete this program had it not been for their support along the way.

I had numerous collaborators during this process, whose help has been amazing. I'd like to thank the patients and their families for enrolling in our studies, the surgeons who send us samples, especially the Texas Cleft-Craniofacial Team, Dr. Samuel Stal and Dr. John Mulliken, Dr. Eric Swindell, Dr. Matt Warman Dr. Yukio Nakamura, Dr. Daniel Wagner and his lab, Dr. Susan Blanton and her team of analysts and Dr. Andrew Lidral.

I owe a lot of gratitude to the GSBS faculty and staff, especially Dean Stancel, Drs. Tom Goka, John Wiener and Vicki Knutson, as well as Karen Weinberg, Bunny Perrez, Joy Lademora, Brenda Gaughn, and Lily D'Agostino, who all have been helpful in making sure

that I stayed on track and graduated on time. I also owe a lot of gratitude to the Dental Branch faculty who have been supportive of my different choice of education and for making sure that it all worked out so I can complete my research studies as well as my dental studies. I am thankful for Dr. Rena D'Souza, who recruited me to the Texas Medical Center, Dr. Karen Stortz, who suggested I talk with Dr. Hecht when I was looking for a Ph.D. mentor, and Dr. Leslie Roeder, who helped me stretch out four years of dental school over seven to ensure that I had ample time to complete both degrees.

My friends have helped me keep a balanced life and remember that there is life outside of school. From grad school friends, change ringing "teammates", members of Eight Ball Surprise, St. Paul's UMC, UTDB class of 2008 and 2011, GSBS students, and my co-officers in the Graduate Student Association- Pat Gibney, Nicole Pinaire, Chris Singh and Katelyn Weymouth- thank you all. I have been blessed to have you all in my life.

Finally, I would like to thank my family. Mom, Dad, Mrs. Jan, Kayla, Grandparents, Aunts, Uncles, Cousins- thank you for supporting me these past almost thirty-something years. Thank you for always encouraging me to do my best, reach for the stars, and letting me know that I can do anything I set off to do. To Michael- thank you for being at my side throughout this whole process and supporting me and making this whole experience more enjoyable.

Gene Discovery in Nonsyndromic Cleft Lip with or without Cleft Palate

Brett Chiquet

Advisor: Jacqueline T. Hecht, PhD

Nonsyndromic cleft lip with or without cleft palate (NSCLP), a common, complex orofacial birth defect that affects approximately 4,000 newborns each year in the United States, is caused by both genetic and environmental factors. Orofacial clefts affect the mouth and nose, causing severe deformity of the face, which require medical, dental and speech therapies. Despite having substantial genetic liability, less than 25% of the genetic contribute to NSCLP has been identified. The studies described in this thesis were performed to identify genes that contribute to NSCLP and to demonstrate the role of these genes in normal craniofacial development. Using genome scan and candidate gene approaches, novel associations with NSCLP were identified. These include MYH9 (7 SNPs, $0.009 \leq p < 0.05$), Wnt3A (4 SNPs, $0.001 \leq p \leq 0.005$), Wnt11 (2 SNPs, $0.001 \leq p \leq 0.01$) and CRISPLD2 (4 SNPs, $0.001 \leq p < 0.05$). The most interesting findings were for CRISPLD2. This gene is expressed in the fused mouse palate at E17.5. In zebrafish, crispld2 localized to the craniofacial region by one day post fertilization. Morpholino knockdown of crispld2 resulted in a lower survival rates and altered neural crest cell (NCC) clustering. Because NCCs form the tissues that populate the craniofacies, this NCC abnormality resulted in cartilage abnormalities of the jaw including fewer ceratobranchial cartilages forming the lower jaw ($\leq$ three pairs compared to five) and broader craniofacies compared to wild-type zebrafish. These findings suggest that the CRISPLD2 gene plays an important role in normal craniofacial development and perturbation of this gene in humans contributes to orofacial clefting. **Overall, these results are important because they contribute to our**

understanding of normal craniofacial development and orofacial clefting etiology, information that can be used to develop better methods to diagnose, counsel and potentially treat NSCLP patients.

Table of Contents

List of Figures

List of Tables

Chapter 1: Craniofacial development and nonsyndromic cleft lip with or without cleft palate

1.0. Introduction

Craniofacial development is a tightly regulated process that involves the complex orchestration of genetic and environmental factors, including interaction of cell growth, growth factors and receptors, the convergence and fusion of the facial and palatal processes, apoptosis, and adequate nutrient supply. An intricate system of checks and balances exists to ensure that normal development occurs [1,2]. In the following chapters, an in depth examination of orofacial clefting will be presented, including the etiology mechanism, with emphasis on understanding how genetic variations can alter normal craniofacial development and contribute to abnormal lip and palate development.

1.1. Normal craniofacial development

Craniofacial morphogenesis is a complex process involving the interaction of cell growth, growth factors and receptors, and apoptosis to guide the convergence and fusion of the facial and palatal processes [3-5]. Many genes are involved and regulate the development of the craniofacial region, including growth factors (e.g., FGFs, TGFs, PDGFs, EGFs, BMPs and respective receptors), signaling molecules (e.g., WNT family, SHH and respective receptors) and transcription factors (e.g., MSX, DLX, LHX, PRRX and BARX family genes and respective receptors) [3,6-21].

The upper lip, primary palate and secondary palate form sequentially from connective tissue derived from neuroectoderm between gestational days 20 and 56, as illustrated in Figure 1.1 [14]. Neural crest cells (NCCs) migrate and differentiate at embryonic day 17, expressing homeobox, MSX and DLX family genes, which direct craniofacial patterning [14]. Mesenchyme from the first branchial arch forms the maxillary, lateral nasal, and

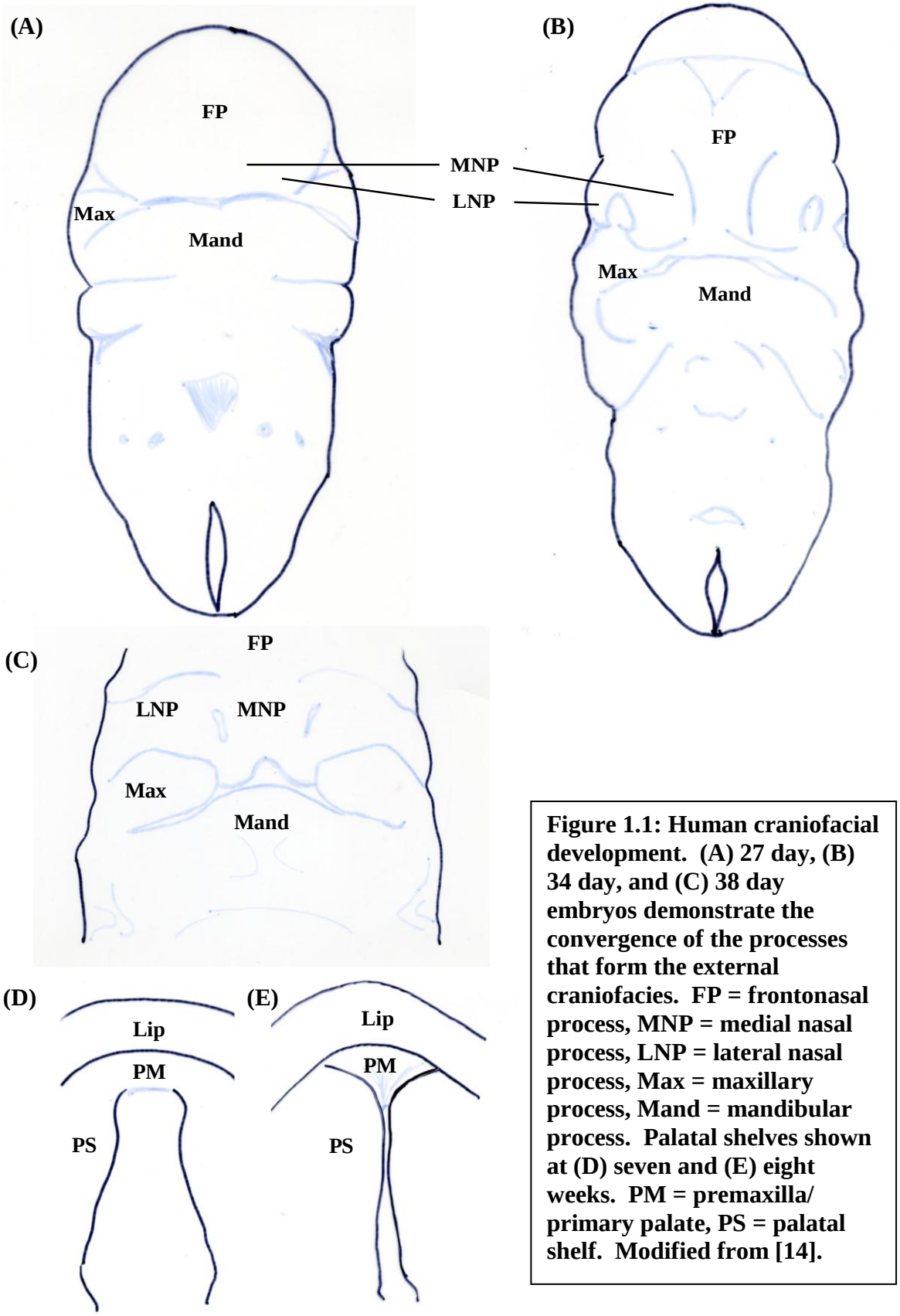


Figure 1.1: Human craniofacial development. (A) 27 day, (B) 34 day, and (C) 38 day embryos demonstrate the convergence of the processes that form the external craniofacies. FP = frontonasal process, MNP = medial nasal process, LNP = lateral nasal process, Max = maxillary process, Mand = mandibular process. Palatal shelves shown at (D) seven and (E) eight weeks. PM = premaxilla/ primary palate, PS = palatal shelf. Modified from [14].

medial nasal processes by day 32 [14]. The maxillary processes grow medially, pushing the lateral nasal process superiorly and allowing the maxillary process to fuse with both the lateral and medial nasal process. The two medial nasal processes fuse to form a continuous upper lip by gestational day 36 and external lip development is complete by gestational day 38 [14,18,24].

The primary and secondary palates are formed between gestational days 38-56, after upper lip morphogenesis is complete. The primary palate is also derived from both the maxillary and medial nasal processes. After fusion of the upper lip, proliferation of mesenchyme causes the intermaxillary segment to project into the oral cavity by day 44, forming the primary palate and dividing the oral and nasal cavities [18,24,25]. The secondary palate is formed from the downward projections of the maxillary processes that descend into the oral cavity and lie lateral to the tongue (lateral palatine processes). By day 56, the tongue descends, allowing the two lateral palatine shelves to elevate, grow medially, and meet together [14,18]. The epithelial edges can then undergo necrosis to allow fusion between the lateral palatine shelves, nasal septum and primary palate [14,18,24]. Palatal development is complete by gestational day 60 [19].

1.2. Development of cleft lip and palate

Cleft lip and palate is the result of improper fusion of the processes that form the face, caused by abnormal morphogenesis of the upper lip and primary palate either by misguided epithelial movement, disrupted epithelial-mesenchymal (EMT) transformation, or disrupted apoptosis [24]. Failure of these mechanisms result in insufficient growth, decreased nutrients and/or a diminished degradation of the epithelial seam covering the growth processes, each

predisposing cleft lip and palate [14]. A cleft of the lip and/or palate results from a failure of fusion of the (1) frontonasal prominence with the medial nasal process, (2) the primary palate with the secondary palate or (3) the lateral palatine shelves with each other [14].

1.3. Nonsyndromic cleft lip with or without cleft palate

Orofacial clefting is a common, complex birth defect that can occur as part of more than 400 syndromes or as an isolated malformation (nonsyndromic) [3]. Nonsyndromic cleft lip with or without cleft palate (NSCLP; Figure 1.2) has a birth prevalence of 1.4/1000 live births in Caucasian and Hispanic populations with males affected twice as frequently as females [3,29]. Orofacial clefting is more common in Native American and Japanese populations (3.6/1000 and 2.1/1000 live births, respectively) and less frequent in African populations (0.3/1000 live births) [3,6,30,31]. There is a higher prevalence of clefts in individuals who live in rural areas and have a lower than average socioeconomic status [19,33]. Between 20–30% of cases occur in families where clefting has previously occurred; the remaining 70–80% of cases occur in families without a history of orofacial clefts [3].

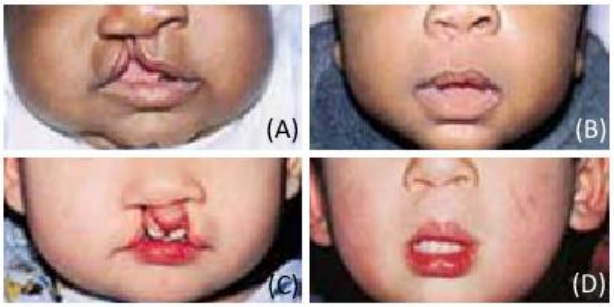


Figure 1.2: NSCLP patients (a,c) pre- and (b,d) post-surgery.

A cleft can occur on either one (unilateral, 90%) or both (bilateral, 10%) side(s) of the face, and the majority of unilateral clefts occur on the left side (66%) [19]. Clefts are classified as either complete or incomplete, depending on the extent of the cleft to the naris (a complete cleft involves the naris) and can involve either the primary or primary and secondary palate [19]. Eighty percent of bilateral clefting cases involve the palate [3].

Nonsyndromic cleft palate only (NSCPO) is considered to be a separate birth defect from NSCLP. NSCPO has a lower birth prevalence that is independent of racial background and occurs more often in females than in males [34,35]. Moreover, cleft palate occurs more often as part of a syndrome than CLP [36]. Table 1.1 summarizes these differences.

Table 1.1: Comparison of NSCLP and NSCPO

	NSCLP	NSCPO
Birth Prevalence*	**1.4**	**0.5**
Differences in birth prevalence between ethnicities	**Yes**	**No**
Male:Female Ratio	**2:1**	**2:3**
Syndromes Associated**	**292**	**554**
Nonsyndromic cases	**75%**	**50%**

*Per 1000 live births **OMIM search on July 21, 2010

Other types of clefts and malformations [cleft lip only, microform cleft, notched gums where a cleft normally occurs, discontinuous obicularis oris muscle, and dental anomalies (tooth agenesis, microdontia, supernumerary teeth, or malposition)] comprise the cleft lip and palate spectrum [37-39]. Recent studies have advocated for the complete phenotyping of affected individuals so that the dataset can be stratified by subphenotype (i.e., overt or occult) in order to reduce heterogeneity (see Section 1.4) [38,40]. This may increase the likelihood of identifying the gene(s) that contribute to the NSCLP spectrum [38].

1.4. Etiology of NSCLP

NSCLP is a complex disorder that does not follow normal Mendelian patterns of inheritance; rather NSCLP exhibits etiological heterogeneity wherein both genetic and environmental factors can predispose an orofacial cleft [41-43]. Additionally, gene-gene and/or gene-environment interactions may contribute to clefting [10,36,44]. The following sections discuss the evidence for etiologic heterogeneity of NSCLP.

1.4.1. Genetic causes of NSCLP

Evidence for a genetic etiology for NSCLP comes from studies that show (1) a heritability for NSCLP of 76%, (2) monozygotic twins are ten-fold more likely to be concordant for a cleft compared to dizygotic twins (40% vs. 4.2%), (3) siblings of affected individuals have an increased risk of having a cleft (λ_s = 30-40) compared to the general population, and (4) clefting aggregates in families [3,45]. However, NSCLP does not follow a Mendelian pattern of inheritance [19]. The multifactorial model, in which genes and environmental factors contribute to the orofacial clefting phenotype, is generally accepted as the inheritance paradigm for NSCLP [10,36]. It has been estimated that between three and fourteen genes play an etiologic role in a single population [42]; therefore, the number of total genes may be greater. Despite the high heritability associated with NSCLP, only one gene, interferon regulatory factor 6 (IRF6), has consistently been found to be associated with NSCLP in different populations [46-49]. Other gene associations have been identified, including RARA, TGFα, TGFβ, p63, MYH9, BCL3, and MSX1, but have been inconsistent and account for only a small percentage of the underlying genetic heritability [10,50-55].

1.4.2. Environmental causes of NSCLP

Evidence for an environmental cause(s) of NSCLP come from epidemiological studies that show an increased birth prevalence of clefting after *in utero* exposure to putative teratogens, such as antiepileptic medications, smoking and folic acid deficiency [10,56,57]. The only one of these environmental triggers that has consistently proven to increase the risk of NSCLP is *in utero* exposure to antiepileptic medications phenytoin and phenobarbital, both of which are folic acid antagonists that decrease folate metabolism and absorption, lowering the availability of nutrients for the developing embryo [56-58]. Maternal folic acid supplementation, which has been shown to decrease the birth prevalence of neural tube defects by up to 70%, has not had a similar effect in orofacial clefting birth prevalence [29,59-61]. However, studies have shown that higher than physiologic doses of folic acid supplementation (>6mg/day) can decrease the recurrence in women at high risk of having a child with NSCLP and NSCPO [62,63]. Studies evaluating whether exposure to either tobacco smoke or folic acid deficiency play an etiologic role in NSCLP have yielded inconsistent results [64-69].

1.4.3. Gene-environment interaction etiology of NSCLP

An increasing number of studies have focused on identifying gene-environment interactions as a cause of NSCLP. One area has been smoking metabolism genes. Cytochrome P450 proteins, epoxide hydrolase, glutathione transferases and arylamine N-acetyltransferase (NAT) genes all function in the metabolism/detoxification of smoking byproducts [13,44]. Fetal genotypes that cause a deficiency in the enzyme product are potentially at a greater risk of having an orofacial cleft if they were exposed to smoking byproducts [13,44,70,71]. Similarly, byproducts from smoking can interfere with normal development by causing growth retardation, and studies have found gene-environment

interactions between smoking and variants in transforming growth factors, muscle segment homeobox and retinoic acid receptor genes [13,41,69,72-76]. However, these results have not been consistently observed in all studies [70,71,73,74,76-80].

Maternal folate supplementation and variation in genes that regulate folate metabolism has also been studied extensively for gene-environment interactions, especially two polymorphisms in the methylenetetrahydrofolate reductase (MTHFR) gene: C677T and A1298C. Both MTHFR polymorphisms have been shown to decrease MTHFR activity and the C677T polymorphism also decreases circulating folate and increases homocysteine levels [81-84]. However, like the smoking studies, MTHFR studies have produced inconsistent results [64,65,67,68,84-96].

1.5. Approaches to gene identification in NSCLP

Section 1.4.1 presented evidence that there is a substantial genetic component to NSCLP etiology. Thus, research has been targeted to identify the genes that cause NSCLP. Two methods, genome scans and candidate gene approaches, have been utilized to identify the genetic component to NSCLP. Briefly, genome-wide scans utilize polymorphic markers that cover the genome to identify regions that are linked to and/or associated with a phenotype of interest. In contrast, candidate gene approaches are a more targeted approach that identifies specific genes that either (1) play a biologically plausible role in craniofacial development, (2) are knocked out of murine models with a clefting phenotype, (3) are deleted, translocated, or implicated in syndromes associated with an orofacial cleft phenotype, or (4) are shown to be linked and/or associated in previous NSCLP studies. The

following sections will discuss how these approaches have been used to identify genetic variants that contribute to the clefting phenotype.

1.5.1. Genome scans

NSCLP is a genetically heterogeneous disease, with different genes responsible for disease etiology in different populations [41-43]. A number of genome scans have been performed NSCLP families of different ethnicities, including Chinese, Syrian, European and European American (Caucasian), Indian, Filipino and African American [97-107]. Table 1.2 lists the regions identified by NSCLP genome scans, including those presented in Chapter four. Notably, there are six regions that were positive in five or more NSCLP genome scans; five of these contain known NSCLP genes, as shown in parentheses: 1p36.13-31 (*MTHFR*), 2p16.3-11.2 (*TGFα*), 2q32-37.2 (*SATB2*), 3p26-21.2, 6q21-27 (*T*) and 16q21-24 (*CRISPLD2*). Interestingly, the IRF6 gene (see Section 1.4.1) is found in the 1q32 region which has been identified in three of the genome scans, including the two meta-analysis studies (Table 1.2).

1.5.2. Candidate gene approach

1.5.2.a. Development genes

Normal craniofacial development, as mentioned in Section 1.1, involves the complex interaction of numerous genes, including growth factors, transcription factors, signaling molecules, and their respective receptors. Genes that direct the craniofacial processes to form, migrate and fuse together to form the face are prime candidates because perturbation of any of these genes has the potential to alter normal craniofacial morphogenesis. Such candidate genes include FGFs, TGFs, BMPs, MSXs, DLXs, WNTs and SHH [15,108-132].

Table 1.2: Chromosomal regions identified in NSCLP genome scans†

Chr.	Region	United Kingdom (2000)*	Chinese (2002) *	Syrian (2003)*	India (2004)*	Turkey (2004)*	Meta-Analysis (2004)	Filipino (2005)	Caucasian (2005)	India (2006)	Filipino (2007)	Caucasian (2009)	African American (2009)	Meta-Analysis (2009)
1	1p31-36.13	X	X		X	X				X				
	1p12-13						X							
	1q32				X		X							X
2	2p22-21										X		X	
	2p11.2-16.3	X		X		X		X	X					X
	2q32-37.2	X	X			X	X		X					
3	3p13-14.1											X		
	3p21.2-26		X	X			X					X	X	
	3q26.31-33		X							X				X
4	4p13-16.2					X		X						
	4q21.23		X											
	4q28									X				
	4q31-35.1		X	X				X						
5	5p15.31		X											
	5p13.3											X		
	5q11.2				X									
	5q22.1		X											
	5q31.1-35.1		X									X		
6	6p23-24.1	X			X		X	X						
	6p12.3									X				
	6q12-16.3		X			X								
	6q21-27	X			X	X	X				X	X		
7	7p21.1-22.2		X		X									
	7p12		X				X							
	7q21.11-22.2		X											
	7q34			X										
	7q36											X		

Chr.	Region	United Kingdom (2000)*	Chinese (2002) *	Syrian (2003)*	India (2004)*	Turkey (2004)*	Meta-Analysis (2004)	Filipino (2005)	Caucasian (2005)	India (2006)	Filipino (2007)	Caucasian (2009)	African American (2009)	Meta-Analysis (2009)
8	8p21						X				X			
	8q11.23-12.1		X											
	8q21.3-24	X					X						X	
9	9p23									X				
	9q21-22					X	X					X		X
	9q33.2				X							X		
10	10q24.32-25.3		X											
11	11p11-q14	X	X						X	X				
12	12p11-13	X					X		X					X
	12q13.13-15	X	X											
	12q21.33-24	X			X		X							
13	13q33.3									X				
14	14q21-24						X							X
	14q32.32									X				
15	15q15						X							
	15q21.2-26.1				X	X								
16	16p12.1-13.11	X	X						X					
	16q21-24	X		X	X	X		X	X					X
17	17p13.1			X										
	17q21						X							
18	18q12.1					X								
	18q21						X			X				
19	19q13.3		X		X									
20	20p12.1-13		X				X							
22	22cent-q12.3				X	X						X		
X	Xcen	X			X									
	Xq22.3			X										

†Table modified from [99]
X denotes publication reported significant association to at least part of the region indicated
*Denotes dataset also used in 2004 meta-analysis

1.5.2.b. Syndromic genes

An orofacial cleft can occur as part of the phenotype of syndromes that follow a normal Mendelian single-gene pattern of inheritance. There are 292 syndromes that are associated with the CLP phenotype (see Table 1.1, page 6) [133]. Van der Woude syndrome (VWS; OMIM: #119300) and popliteal pterygium syndrome (PPS; OMIM 119500) are two autosomal dominant syndromes that both have a clefting phenotype and are both caused by mutations in the IRF6 gene [134-142]. Lower lip pits occur in conjunction with orofacial clefting in 85% of VWS cases, with the remaining 15% of VWS cases having only an orofacial cleft phenotype (CLP and/or CP) [143]. This latter observation suggested that IRF6 may play a role in the etiology of NSCLP [144]. IRF6 coding mutations have not been found in NSCLP patients, but polymorphic markers have been shown to be in linkage disequilibrium and associated with NSCLP in multiple studies [46,47,49,145,146]. IRF6 animal and promoter studies have supported these results and are discussed in Sections 1.5.2.d and 1.6, respectively [147-149].

1.5.2.c. Chromosomal abnormalities

Chromosomal deletions, duplications, translocations or copy number variants are useful in identifying regions of the genome that may contain disease-associated genes [22,23,26-28,32]. While patients with identifiable abnormalities typically have multiple birth defects, the gene(s) in these chromosomal regions become candidate genes. A number of chromosomal abnormalities have been identified in patients with cleft lip and palate, including eight regions that have also been identified in NSCLP genome scans (Tables 1.2 and 1.3). These regions provide new areas to identify candidate genes and test for association to NSCLP.

Table 1.3: Chromosomal abnormalities in CLP syndromes.

Chromosomal Region	Type of Abnormality	Reference
1q21-25	Deletion	[22]
2p22.2*	Deletion	[23]
2q37.1*	Deletion	[23]
3p26-21*	Duplication	[26]
4p/tetrasomy 9p	Deletion/Tetrasomy	[27]
4p16.3	Deletion	[28]
4p16-15*	Deletion	[22]
4q31-35*	Deletion	[22]
der(4) t(4;20)(q35;q13.1)	Trisomy	[32]
6p24*	Deletion	[23]
7p15.3	Microdeletion	[23]
7q34-35*	Deletion	[22]
10p15-11	Duplication	[26]
11p14-11	Duplication	[26]
13q22-34*	Duplication	[26]

*Indicates region has also been identified in a NSCLP genome scan (see Table 1.2, page 11).

1.5.2.d. Animal studies that suggest NSCLP genes

Animal models have proven to be a useful tool in elucidating the genetic etiology of NSCLP [150,151]. There are three types of approaches to study NSCLP using animal models: (1) mutagenic and teratogenic screens, (2) spontaneous clefting strains and (3) targeted genetic studies of candidate genes. The mutagenic and teratogenic screens utilize chemical reagents such as N-ethyl-N-nitrosourea (ENU) to create a phenotype-driven screening model [152]. The resultant abnormal phenotypes can then be used to determine the causative genetic factor. Studies have demonstrated that ENU-injected animals have a higher frequency of orofacial clefting than control animals, providing models that can be used to identify NSCLP genes [153,154]. However, this approach has not been actively pursued.

Spontaneous clefting is a common occurrence (5-30%) in the A/WySn, A/HeJ and A/J inbred mouse strains [155-157]. Two gene loci, clf1 and clf2, have been identified as contributing to the clefting phenotype and have provided researchers with new regions of the mouse genome to interrogate for clefting genes [158,159]. The clf2 region is on mouse chromosome 11qE1, and contains two adjacent WNT genes: *Wnt3* and *Wnt9B* [160]. Both function in the canonical WNT signaling pathway and regulate midfacial development and upper lip fusion (see Chapter 4, Section 4.2.1) [9,159]. Additionally, chapter four presents data from our study interrogating variants in seven WNT genes for association with NSCLP, which was undertaken because of the results from mouse studies and because the WNT gene family plays an important role in craniofacial development (see Section 1.5.2.a).

To determine if putative NSCLP genes contribute to the clefting phenotype, targeted gene studies in mouse models have proven useful in identifying gene function(s) [161,162]. For example, mutations in IRF6 cause VWS and PPS, both of which are associated with

orofacial clefts (see Section 1.5.2.b) [46,47,49,134-142,145,146]. Two IRF6 mouse models, one with IRF6 knocked out and the other with a missense mutation that eliminates IRF6 DNA binding activity, have furthered the understanding of IRF6 protein function [147,149,163]. These studies have shown that without a functional IRF6 protein, abnormal keratinocyte differentiation and proliferation cause the keratinocytes to poorly adhere and improperly fuse together [147,149]. This results in craniofacial (cleft palate, missing external ears and a short snout), limb and skin anomalies [147,149]. The presence of non-craniofacial anomalies demonstrates that IRF6 plays a critical role in mouse development beyond the face [137,138,140,142,164]. This type of targeted gene approach has been used for other NSCLP candidate genes, including BMP4 and its receptor BMPR1α, Pax9, Sox11, and Tbx10, where knockdown of each of these genes in the mouse model has produced a clefting phenotype [151,165,166].

Other animal models have been utilized to study clefting genes, including zebrafish and chicken [167]. Zebrafish have been used to study several genes involved in craniofacial development, including the Runx, Dlx and Sox families of transcription factors [168-170]. The Dlx gene family has been shown to play a role in cartilage differentiation [169] while the Runx gene family functions in cell fate determination and chondrogenesis [168]. Two Sox9 orthologs are part of a complex signaling network that controls craniofacial skeleton development [170]. Chicken embryos have been used to study fusion of the primary palate and lip [167]. Research in chick models has shown that TGF-β3 and the bone morphogenetic proteins regulate palate migration and fusion [171,172].

1.6. Genetic variation as the cause of disease phenotype

While linkage analyses have proven to be a useful analytical tool in determining the causative gene in single-gene disorders, association studies are useful to study common, complex disorders, such as NSCLP, that do not follow Mendelian patterns of inheritance [173]. Association studies test whether sequence variants have an altered frequency in a population of interest compared to the control population [174]. Three types of sequence variants, or polymorphisms, are assessed in association studies: (1) coding, (2) promoter/regulatory, and (3) intronic.

Coding variants, or those exonic changes that are translated into protein, can be either nonsynonymous (protein-changing) or synonymous (non-protein-changing). Nonsynonymous variants alter amino acid sequence in the protein structure and therefore affect normal protein function [175]. Synonymous variants do not alter the protein structure; however, it has recently been shown that they can alter amino acid translation time, which may be detrimental to gene function by altering protein folding [176]. Therefore, both synonymous and nonsynonymous coding changes have the potential to alter gene function and disrupt normal development.

Sequence variants in the promoter or regulatory region may alter protein levels and thus protein function [177,178]. Recent research has shown that a common IRF6 risk variant, rs642961, disrupts an AP-2α binding site upstream of the IRF6 gene [148]. Transcription factor AP-2α is expressed in the craniofacial processes and limbs and mutations in AP-2α cause Branchiooculofacial syndrome (OMIM:#113620), a syndromic form of CLP that also includes a lip pits [148,179]. Together, this data gives suggests that AP-2α and IRF6 are involved in a NSCLP etiological pathway.

Intronic variants do not alter amino acid sequence, but may alter gene function if they (1) are in normally conserved splice sites or intron/exon junctions or (2) are in linkage disequilibrium with disease loci [180-182]. While intronic variants have been associated with several complex disorders, such as IRF6 variants in NSCLP, RET variants in Hirschsprung Syndrome and complement factor H in age-related macular degeneration, actually defining how these polymorphisms affect on gene function remains to be defined [46,47,49,183-185].

1.7. The role of the dentist in NSCLP treatment

The health care costs of caring for a child with an orofacial cleft is more than $100,000 and creates a significant financial and healthcare burden for these families [186]. Years of therapy by a team of specialists are required to attain satisfactory functional correction [187]. Besides the surgical interventions that are required to repair the defect, a number of other therapies play an integral role in treating NSCLP patients. Speech therapists are needed to assess and treat articulation and resonance problems that exist secondary to the altered oral anatomy [187]. Genetic healthcare professionals have the diagnostic capabilities to distinguish between syndromic and nonsyndromic forms of clefting, as well as counseling the family on recurrence risks for future pregnancies and natural histories. Dentists are needed to address a number of dental-related problems, including tooth agenesis at the cleft site, supernumerary teeth, tooth crowding, hypoplastic enamel, that often affect NSCLP patients [188]. Children with CL/P are also predisposed to dental caries; damage to the primary dentition compromises the permanent dentition and success of future surgical and orthodontic therapies [189]. Together, these groups of specialists ensure that NSCLP patients can have complete rehabilitation.

1.8. Significance of this study

NSCLP is a common birth defect caused by both genetic and environmental factors, neither of which have been fully elucidated because of the complex etiologic heterogeneity of this disorder. Therefore, the need to identify these NSCLP genes is high. Identification of genetic loci has multiple benefits, including assisting in better counseling and diagnosis, and also leading to therapeutic interventions for those at high risk for having a child with an orofacial cleft. Additionally, this research has aided in understanding craniofacial development which will in turn lead to a better understanding of the causes of orofacial clefts. The goal of this project was to further delineate the genetic etiology of NSCLP using both genomic screening and candidate gene testing approaches in our well-characterized nonHispanic white and Hispanic NSCLP families.

This thesis will detail the efforts toward the goal of gene identification in NSCLP. Chapters 3 and 4 demonstrate the usefulness of using genome scan and candidate gene approaches to identify NSCLP genes, respectively. Chapter three describes the 11 regions identified in a genome scan of ten NSCLP families. Examination of the most significant region, 22q12.2-12.3, showed that the MYH9 gene, which had previously been implicated to clefting etiology in an Italian population, was associated with NSCLP. Chapter 4 describes the interrogation of seven WNT genes and showed that variations in genes that direct normal facial development, specifically those WNT genes that regulate neural crest cell proliferation and migration, were associated with NSCLP. Chapter 5 describes studies identifying the CRISPLD2 gene which resides in the chromosome 16q24.1 region, identified in genome scan studies. CRISPLD2 is a novel gene which we showed was both associated

with NSCLP and expressed in the developing murine craniofacies. Chapter 6 discusses the results of candidate gene testing of the CRISPLD1 gene and testing of interactions between variants in folate pathway genes and both CRISPLD1 and CRISPLD2. Chapter 7 describes studies delineating the role of CRISPLD2 in craniofacial development in zebrafish. Altogether, these studies demonstrate the usefulness of different approaches to identify the genetic etiology of complex disorders and have contributed to enhancing our knowledge of NSCLP.

Chapter 2: Materials and methods

Note: Portions of this chapter are taken from previous publications where I was the first author. Guidelines from each of these journals do not require permission to reproduce manuscript content as part of a thesis (http://www.nature.com/reprints/ permission-requests.html, http://www.oxfordjournals.org/access_purchase/publication_rights.html).

2.1. Dataset

Institutional review board (IRB) approval for the use of these samples had previously been obtained (HSC-MS-03-090 from the University of Texas Health Science Center at Houston's Committee for the Protection of Human Subjects and 20010904 from the Ohio State University's Internal Review Board and the University of Antioquia-Colombia Dental School's Scientific Committee). Our primary (discovery) dataset was comprised of 568 NSCLP families (1725 individuals). These families were collected from three centers: Texas Children's Hospital in Houston, Texas, the University of Texas Craniofacial Clinic in Houston, and Children's Hospital in Boston, Massachusetts. Two hundred and forty-six families, which included 971 individuals, were collected from the Clinica Noel in Medellín, Columbia, and used as a validation population for some of our studies. The clinical information about the populations is listed in Table 2.1. All cases were examined for the presence of other malformations and for a syndrome by a clinical geneticist. Only individuals with NSCLP were included in this study. Those patients diagnosed with a syndromic form of clefting were excluded from this study. Gender, ethnicity, severity of clefting, completeness of clefting and family history (FH) of clefting (multiplex vs. simplex) (Figure 2.1) were recorded in patients charts.

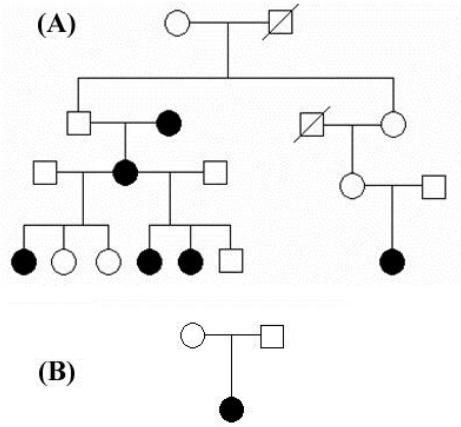


Figure 2.1: Example pedigrees of a (A) multiplex and (B) simplex trio families. Solid shapes represent an affected individual.

After obtaining informed consent, blood and/or saliva were collected from patients and parents, as well as other family members when there was a positive FH of NSCLP. DNA was

Table 2.1: NSCLP Dataset

	nonHispanic White			Hispanic			African American			Columbian***			Total		
	FH +	FH -	Total	FH +	FH -	Total	FH +	FH -	Total	FH +	FH -	Total	FH +	FH -	Total
Families	108	286	394	44	117	161	3	10	13	114	132	246	269	545	814
Individuals	474	776	1250	154	298	452	28	22	50	579	392	971	1235	1488	2723
Male	245	416	661	79	156	235	14	7	21	293	220	513	631	799	1430
Affected	122	193	315	37	73	110	6	4	10	165	89	254	330	359	689
Sidedness															
Unilateral*	24	30	54	0	7	7	1	1	2	--	--	--	25	38	63
Left Side	36	56	92	14	31	45	0	2	2	--	--	--	50	89	139
Right Side	15	29	44	8	17	25	0	0	0	--	--	--	23	46	69
Bilateral	31	38	69	11	13	24	0	1	1	--	--	--	42	52	94
Not recorded	16	40	56	4	5	9	5	0	5	--	--	--	25	45	70
Type of Cleft															
Cleft lip and palate	94	151	245	32	62	94	4	4	8	--	--	--	130	217	347
Cleft lip	24	27	51	2	7	9	0	0	0	--	--	--	26	34	60
Not recorded	4	15	19	3	4	7	2	0	2	--	--	--	9	19	28
Severity															
Incomplete	16	33	49	2	14	16	0	2	2	--	--	--	18	49	67
Unknown	17	16	33	14	21	35	0	0	0	--	--	--	31	37	68
Complete	25	82	107	17	25	42	0	2	2	--	--	--	42	109	151
Mixed**	1	11	12	0	3	3	0	0	0	--	--	--	1	14	15
Not recorded	63	51	114	4	10	14	6	0	6	--	--	--	73	61	134
Female	229	360	589	75	142	217	14	15	29	286	172	458	604	689	1293
Affected	73	93	166	22	44	66	3	6	9	107	42	149	205	185	390
Sidedness															
Unilateral*	7	22	29	1	0	1	1	1	2	--	--	--	9	23	32
Left Side	21	26	47	10	19	29	1	2	3	--	--	--	32	47	79
Right Side	7	8	15	7	9	16	0	1	1	--	--	--	14	18	32
Bilateral	19	13	32	3	9	12	0	1	1	--	--	--	22	23	45
Not recorded	19	24	43	1	7	8	1	1	2	--	--	--	21	32	53
Type of Cleft															
Cleft lip and palate	58	71	129	18	34	52	2	4	6	--	--	--	78	109	187
Cleft lip	14	14	28	4	5	9	1	1	2	--	--	--	19	20	39
Not recorded	1	8	9	0	5	5	0	1	1	--	--	--	1	14	15
Severity															
Incomplete	8	23	31	2	6	8	0	1	1	--	--	--	10	30	40
Unknown	4	11	15	6	8	14	1	1	2	--	--	--	11	20	31
Complete	17	30	47	9	20	29	0	3	3	--	--	--	26	53	79
Mixed**	1	0	1	0	1	1	0	0	0	--	--	--	1	1	2
Not recorded	43	29	72	5	9	14	2	1	3	--	--	--	50	39	89

*Cleft is unilateral but side was not recorded

**Mixed are bilateral cases with one side as complete and the other incomplete

***"--" is not recorded

extracted from the blood using the Roche DNA Isolation Kit for Mammalian Blood (Roche, Switzerland) and from the saliva using the Oragene Purifier for Saliva (DNA Genotex, Inc., Ontario, Canada) following manufacturer's protocol [190].

2.2. Polymorphism selection and genotyping

Single nucleotide polymorphisms (SNPs) were chosen based on the following criteria: (1) heterozygosity was greater than or equal to 0.3, (2) SNP must be polymorphic in the nonHispanic white (NHW) population with a minor allele frequency greater than 0.2, (3) one SNP per 10 kb, (4) exonic and coding SNPs were given priority over intronic SNPs, and (5) SNPs that "tagged" for multiple SNPs were given preference over SNPs that were not in linkage equilibrium with other SNPs. Haploview was used to identify tagSNPs [191]. For the MYH9 study, SNPs were genotyped based on a previously published study [55]. For the folate gene pathway studies, SNPs (93/96) were selected from two SNPlex™ pools designed by Dr. Richard Finnell's laboratory [192]. The three SNPs that could not be incorporated into the new SNPlex chemistries were replaced following our standard criteria. Data from three short tandem repeat (STR) markers previously genotyped in Dr. Hecht's laboratory were also included in the analysis, two for the WNT studies and one for the CRISPLD2 studies [106,193]. Information on the genes interrogated, marker type and genotyping platform used is listed in Table 2.2.

Polymorphic markers were purchased from Applied Biosystems (AB, Foster City, California) and genotyped on our samples following the manufacturer's protocol. For Taqman® assays, the AB 7900HT Sequence Detection System was used to make genotyping calls, whereas the AB 3730 and AB GeneMapper® Software v4.0 were used to

Table 2.2: SNPs genotyped in these studies

Gene	dbSNP	chr:bp	location	Alleles*
MYH9	rs7078	22:35007860	downstream	T/C
MYH9	rs4821478	22:35020059	intron 28	A/G
MYH9	rs2009930	22:35029252	intron 19	C/T
MYH9	rs2239783	22:35035074	intron 15	C/T
MYH9	rs3752462	22:35040129	intron 13	C/T
MYH9	rs1002246	22:35044605	intron 10	T/C
MYH9	rs2071731	22:35048804	intron 5	G/A
MYH9	rs713659	22:35058638	intron 3	G/A
MYH9	rs713839	22:35063884	intron 3	A/G
MYH9	rs739097	22:35076025	intron 1	A/G
Wnt3A	rs708111	1:226257988	upstream	C/T
Wnt3A	rs708114	1:226264119	intron 1	C/T
Wnt3A	rs3094912	1:226276438	intron 1	A/T
Wnt3A	rs3121310	1:226291447	intron 2	A/G
Wnt3A	rs752107	1:226313974	downstream	C/T
Wnt3A	rs1745420	1:226318355	downstream	C/G
Wnt5A	rs566926	3:55487719	downstream	A/C
Wnt7A	rs1124480	3:13832970	upstream	C/T
Wnt7A	rs9840696	3:13840076	intron 1	A/G
Wnt7A	rs6778046	3:13846680	intron 1	C/T
Wnt7A	rs9863149	3:13858361	intron 1	C/T
Wnt7A	rs934450	3:13862730	intron 1	A/G
Wnt7A	rs1433354	3:13872246	intron 2	A/G
Wnt7A	rs6442414	3:13881839	intron 2	A/G
Wnt7A	rs11128663	3:13887851	downstream	C/G
Wnt7A	rs4685048	3:13897736	downstream	A/C
Wnt8A	rs4835761	5:137445768	upstream	A/G
Wnt8A	rs2040862	5:137447888	intron 2	C/T
Wnt8A	rs2306110	5:137455986	downstream	A/C
Wnt11	rs663746	11:75571777	upstream	C/T
Wnt11	rs1533763	11:75578175	intron 1	A/T
Wnt11	rs1533767	11:75583448	exon 3	A/G
Wnt11	rs689095	11:75591735	intron 4	A/G
Wnt11	rs596339	11:75594155	intron 4	A/G
Wnt11	rs1568507	11:75596967	downstream	A/T

Gene	dbSNP	chr:bp	location	Alleles*
Wnt3	rs142167	17:42150418	downstream	A/G
Wnt3	rs7216231	17:42170907	downstream	A/G
Wnt3	rs199525	17:42203002	intron 4	G/T
Wnt3	rs70602	17:42214876	intron 1	C/T
Wnt3	rs199501	17:42217772	intron 1	A/G
Wnt3	rs199498	17:42220763	intron 1	C/T
Wnt3	rs111769	17:42227151	intron 1	C/T
Wnt3	rs3851781	17:42246300	intron 1	C/T
Wnt3	hCV11392791	7:42248220	upstream	C/G
Wnt3	rs9890413	17:42256448	upstream	A/G
Wnt9B	rs2165846	17:42296365	intron 1	A/G
Wnt9B	rs1530364	17:42306776	intron 2	A/G
Wnt9B	rs197915	17:42345521	downstream	A/G
IRF8	rs7193803	16:84482924	upstream	G/T
IRF8	rs305082	16:84494729	intron 2	T/C
IRF8	rs305080	16:84499525	intron 2	C/T
IRF8	rs2292980	16:84502827	intron 3	A/G
IRF8	rs2280378	16:84510496	intron 7	C/T
IRF8	rs1568391	16:84513055	exon 9	T/G
IRF8	rs880365	16:84517113	downstream	C/T
CRISPLD2	rs4572384	16:83400728	upstream	T/C
CRISPLD2	rs1874014	16:83413267	intron 1	A/C
CRISPLD2	rs8051428	16:83424753	intron 1	C/T
CRISPLD2	rs1546124	16:83429802	exon 2	C/G
CRISPLD2	rs1874015	16:83435283	intron 2	T/C
CRISPLD2	rs12051468	16:83437215	exon 3	A/G
CRISPLD2	rs8061351	16:83440853	exon 4	T/C
CRISPLD2	rs2646129	16:83444995	intron 5	A/G
CRISPLD2	rs2326398	16:83460468	intron 8	A/G
CRISPLD2	rs721005	16:83463849	exon 9	C/G
CRISPLD2	rs774206	16:83477709	intron 13	A/G
CRISPLD2	rs767050	16:83480610	exon 14	G/A
CRISPLD2	rs2646112	16:83495601	intron 14	G/A
CRISPLD2	rs2641670	16:83498731	exon 15	G/A
CRISPLD2	rs4783099	16:83499080	exon 15	C/T
CRISPLD2	rs16974880	16:83500389	exon 15	T/G
CRISPLD2	rs903194	16:83502930	downstream	T/G
CRISPLD2	rs2641674	16:83510855	downstream	T/C

Gene	dbSNP	chr:bp	location	Alleles*
CRISPLD1	rs2925155	8:76048852	upstream	A/G
CRISPLD1	rs960856	8:76057462	upstream	T/A
CRISPLD1	rs7841231	8:76065317	intron 2	A/G
CRISPLD1	rs17295835	8:76069196	intron 2	T/C
CRISPLD1	rs1455809	8:76075622	intron 2	T/C
CRISPLD1	rs1455796	8:76086362	intron 2	C/G
CRISPLD1	rs10957748	8:76101956	intron 12	C/T
CRISPLD1	rs13248650	8:76109250	downstream	T/G
CRISPLD1	rs11988595	8:76115241	downstream	C/T

*Major allele as identified in NHW population listed first

SNPlex™ assays. Genotype calls were imported into Progeny Lab database management software (South Bend, Indiana). Mendelian inconsistencies and discrepancies between sample duplicates were identified and corrected using PedCheck software [194].

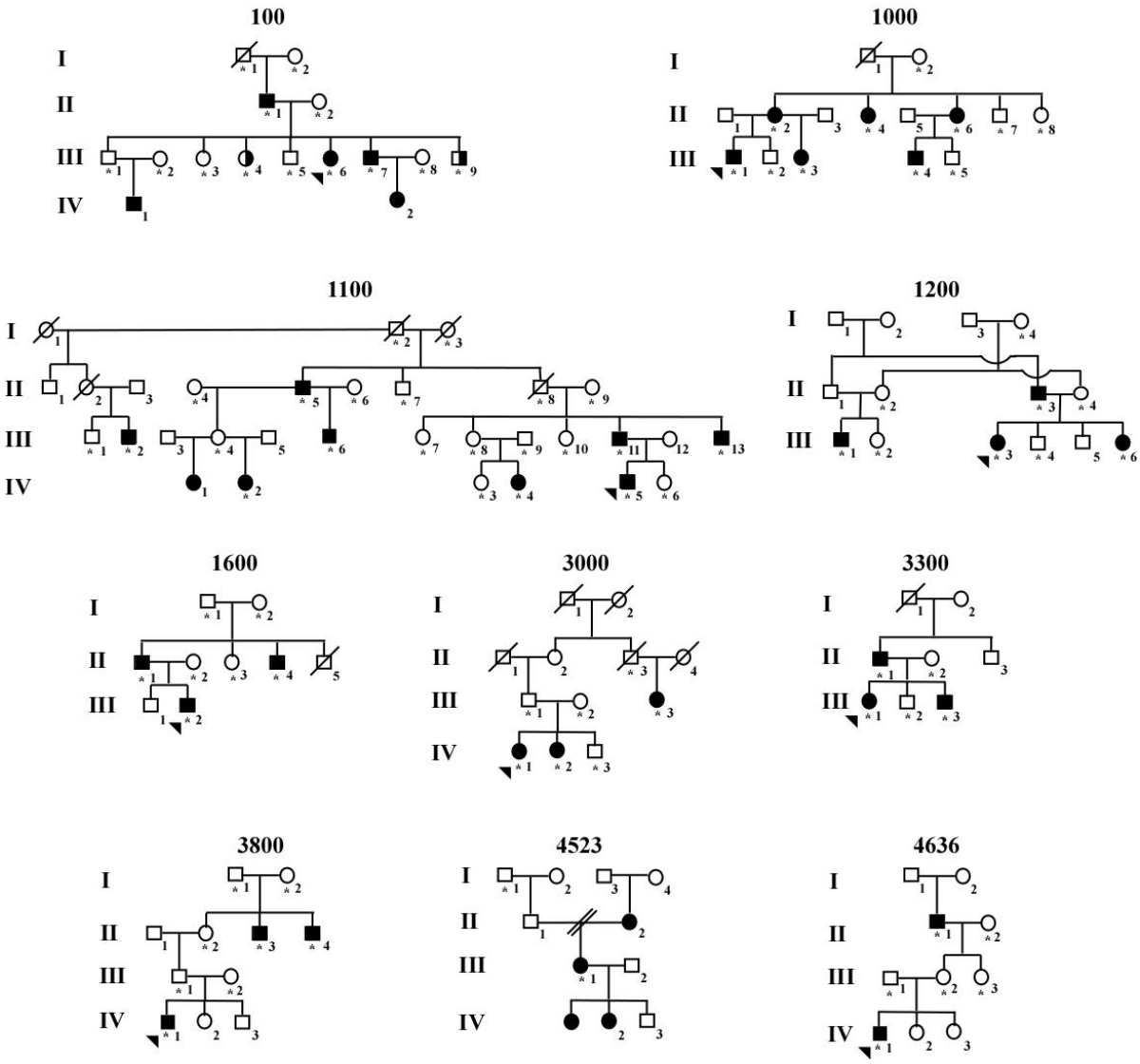
2.3. Genome scan

> Ten multiplex families were subjected to a genome scan using the 6K Illumina Linkage IVb mapping panel. These families were selected from our data set on the basis of the number of affected individuals and availability of DNA. Nine families were non-Hispanic white and one (F1100) was African American [Figure 2.2]. Under our dominant model, four of the non-Hispanic white families, 100, 1000, 1200 and 3000, can yield maximum LOD scores greater than 1 (1.6, 1.5, 1.1 and 1.2, respectively). A maximum LOD score of 2.98 is possible for the African-American family. Among the non-Hispanic white families, there were 30 affected individuals of 67 individuals. In family 100 (F100), there were two siblings (III-4 and III-9) who were reported to have a notched gum in childhood, a finding that could indicate a microform cleft lip. However, these diagnoses could not be clinically confirmed. Therefore, the linkage analysis was performed twice: first, assuming these individuals to be affected and then coding them as unknown with regard to the cleft phenotype. There were 27 genotyped individuals in the African-American family, 7 of whom were affected.
>
> The Illumina Linkage IVb mapping panel consisting of six 008 SNPs was used on a BeadStation system (Illumina Inc.). Allele detection and genotype calling were performed using the BeadStudio software (Illumina Inc.). [105]

2.4. Analyses

The analyses were performed on the population in aggregate and then on the dataset stratified by presence/absence of FH of NSCLP. SAS (Statistical Analysis Software) v9.1 was used to calculate allele frequencies and Hardy-Weinberg equilibrium (HWE). When the allele frequencies differed between the ethnicities, the analysis was performed on each

Figure 2.2: NSCLP families in genome scan*



*Solid shapes indicate individuals with NSCLP. Half-filled samples F100-III-4 and F100-III-9 had notched gums and were analyzed as affected and unknown.

population separately. GOLD (Graphical Overview of Linkage Disequilibrium) was used to calculate pairwise linkage disequilibrium (LD) and displayed using Haploview [191,195].

A panel of linkage and association testing approaches was used to analyze genotyping data. MERLIN (Multipoint Engine for Rapid Likelihood INferences) [196] and Simwalk2 [197] were utilized to perform parametric and nonparametric linkage analysis under previously defined linkage parameters [106].

Four types of analyses to detect single SNP association with NSCLP were utilized: FBAT (family based association test), PDT (pedigree diequilibrium test), geno-PDT (genotype PDT), and APL (association in the presence of linkage). Each of these analytic tools makes different assumptions; each was used to optimize the amount of information obtained from the data. FBAT has the ability to detect both linkage and association between genetic markers and disease genes [198]. PDT is an application of the traditional transmission disequilibrium test (TDT) that incorporated extended multiplex families into analysis instead of only nuclear families, and therefore allowed more information to be analyzed [199]. The geno-PDT is an extension of the PDT that tested for association using the transmission of specific genotypes to detect association [200]. PDT has more power to detect an association when a disease displays additive inheritance while geno-PDT is better suited when inheritance pattern is either dominant and recessive [200]. The APL test provided a powerful test statistic in situations when parental genotypes are missing [201]. In addition to single SNP analysis, multiple SNP haplotypes were analyzed using APL and HBAT (haplotype based association test) [198,201]. To detect if there were any gene-gene interactions between SNPs in different genes, the generalized estimating equations (GEE) as implemented in SAS was used [202].

Two approaches were utilized to correct for multiple testing. First, for FBAT, a raw P-value was obtained under normal FBAT assumptions and a "corrected" P-value was obtained using the empirical variance extension of FBAT (FBAT-e), which corrected for inflated P-values from pedigrees with multiple affected individuals [198]. Additionally, the Bonferroni correction (0.05/number of independent tests) was also used. However, the Bonferroni approach may be too conservative, as the SNPs genotyped in each gene may not be independent of each other because of linkage disequilibrium. Therefore, for our studies, when genotyping multiple genes, we used a Bonferroni correction where the number of genes tested were divided by 0.05.

2.5. Sequencing

DNA from probands from 24 multiplex NSCLP families that had one of the over-transmitted/susceptible haplotypes [203] were used to sequence the fifteen exons of the CRISPLD2 gene. Primer sets, shown in Table 2.3, were designed to capture all exonic sequences and at least 50 basepairs upstream (5') and downstream (3') of the intron/exon junctions. PCR was performed using thermal cyclers with the following conditions: 30 cycles of 95°C for 45 s, T_M°C for 90 s, and 68°C for 90 s (T_M°C listed in Table 2.3). PCR products were purified (Qiagen, Valencia, CA and Millipore, Billerica, MA) according to manufacturer's protocol. Sequencing was performed by LoneStar Labs (Houston, TX) and analyzed using Sequencher v4.8 (Gene Codes, Ann Arbor, MI). All variants were validated by either sequencing from the reverse primer or TaqMan® SNP assay.

Table 2.3: CRISPLD2 sequencing primer information

Exon*	Forward Primer Sequence**	Reverse Primer Sequence	Fragment Length	T_M***
1	gatgctcggtcctaccagcgccgcca	gaggtggcatctgtcctagttattccag	328 bp	66.5
2	agagagtcgtgtgtgctgcgttcgc	agagctggctggggaacagagcaggctggt	492 bp	66.5
3	cgttgctgtatttatggttcttatgccc	ggtctctctcactctctttgcaggcta	305 bp	61.5
4	ctcaaggataatggtattaagcacc	gcagaggctttacagatttgacaatcc	517 bp	56.5
5	tacatccaggccaaatgatatcacc	ttaaagacttcgctgctcaagccag	293 bp	56.5
6	gtgggccttggaagaggatcagtcc	aattcccctgtttcatgtaacgctg	265 bp	61.5
7	ccgctgaaatgattctttgtaaaacc	ttgcagacctaaatccactagctcc	373 bp	56.5
8	tcttagtatagcacggattggattgc	ccgaggcgaacacatgcccagacac	359 bp	61.5
9	cgactgtattgtgcttgcgtgaatc	ctcggtccagtgtgtgaatgctcag	501 bp	61.5
10	ctgagcattcacacactggaccgag	ccgtcataaacaggaaaaccgtttc	453 bp	56.5
11	ataggagcaagcgttcacttttagaagc	gtgatctcaggccaatttccagtcc	330 bp	56.5
12	cgaggagcctgctgggtcgtagtcc	gagccacctcacccagcctaaaaagctacttc	300 bp	66.5
13	tggaattcaaataaagagagccagc	cagatcttggagcttgcatgttgag	235 bp	56.5
14	ttgcccagcagtgaatgatggagcc	ttaaagggtaatgactctgggcctc	319 bp	61.5
15a	tgtctggcggcttcaaggaagtgtc	CCTCTTCCGACGCCGTGAATGCTTC	626 bp	61.5
15b	GCTATGTGTTCTTCTGTTGGTGGAG	CATCGGGTGCGTGGGGAATCTGCAC	720 bp	61.5
15c	CGTAAGGTTCCACTGAGACGAGATGTC	CAGAAGTTTCCCAGATGTTTGGTGG	846 bp	56.5
15d	CGTGCTTTGTGCATTGTAGTCTAGTC	GACAGGGCTTATCATTTATAGTCTT	812 bp	56.5
15e	ATGGACCTGGCTGTCTTTATCATCC	ctggggactggcagcattcgcatca	744 bp	61.5

*Exon number listed. Exon 15 was too large for one primer set; therefore five overlapping primer sets were designed.

**Lower case letters indicate primer located in intron; upper case letters indicate primer in exon.

***Optimum melting temperature for PCR.

2.6. Silver Staining

A 16-23 bp polythymine stretch in exon 15 of CRISPLD2 was analyzed to confirm Mendelian inheritance. The 187-bp region containing this variant was amplified using CRISPLD2-specific primers (5':CAGTGTGAGTTGACCCCATC; 3':CACTGCACGCTACCTTGGG) with the following conditions: 30 cycles of 95°C for 30 s, 57.2°C for 90 s, 68°C for 90 s The product was denatured and run on a polyacrylamide gel, fixed, and stained, as previously described [204]. Alleles were read and frequencies evaluated using the chi-square test.

2.7. Evaluation of sequence variations

Patient/proband sequences were compared to consensus sequences found at www.ncbi.nlm.nih.gov and www.ensembl.org websites using Sequencher v4.8 (Gene Codes, Ann Arbor, MI). Computer algorithms were utilized to determine the predicted effect of each sequence variant identified (see review [205]). Coding variants were analyzed using PolyPhen (http://www.bork.embl-heidelberg.de/ PolyPhen) to predict the effect of amino acid substitution on protein structure/function and SIFT (http://blocks.fhcrc.org.sift/SIFT.html) to analyze conserved protein domains for favorable or unfavorable substitution effects. Sequence variants at the intron/exon junctions were analyzed using the ESEfinder (http://www.rulai.cshl.edu/tools/ESE/) program to predict whether mRNA splicing was affected [206,207]. Transcriptional binding site prediction programs PATCH and AliBaba2 programs were used to assess untranslated and putative promoter regions sequence variants that could alter gene expression (http://www.gene-

regulation.com/pub/ programs.html) [208,209]. These bioinformatic approaches typically over-predict protein changes 20-30% of the time [210]; however, this was preferred over not detecting a putative biologic function of the variant.

2.8. *CRISPLD2* expression studies

2.8.1. Mouse

> RIKEN mouse CRISPLD2 cDNA clone, 3321402M02, extracted from the head of an E17.5 C57BL/6 mouse, was obtained from GeneService (Cambridge, UK). DNA was extracted using a Qiagen midiprep kit (Valencia, CA, USA). A CRISPLD2-specific 514-bp probe was generated by PCR amplification using the following murine specific primers: 5'CRISPLD2 (ATGAACGTCTGGGGAGACAC) and 3'CRISPLD2 (GTACCATCCCATTCCTGGTG). PCR was performed in a thermal cycler, with 30 cycles of 95°C for 30 s, 548C for 90 s, 68°C for 90 s. The probe was gel purified (Qiagen) and sequenced for verification (Lone Star Labs, Houston, TX, USA). PCR products were subcloned into pZErO-2 vector (Invitrogen, Carlsbad, CA, USA) using manufacturer's protocol. The product was linearized using NotI and SpeI restriction enzymes and sequenced (Lone Star Labs) to determine sense versus antisense orientation.
>
> Whole mouse C57/Bl6 embryos (E12.5–E17.5) were fixed in paraffin and sectioned (sagittal and frontal) 6 mm thick. Sections showing craniofacial structures were used in this study. Antisense CRISPLD2 clones were labeled with 35S and *in situ* hybridizations were carried out using standard protocols to detect CRISPLD2 expression [211]. Sense CRISPLD2 clones were used as a negative control. [203]

2.8.2. Zebrafish

Two zebrafish CRISPLD2 probes were generated from wild-type whole adult zebrafish using SuperScript™ III Reverse Transcriptase (Invitrogen, Carlsbad, CA). CRISPLD2-specific primers (5'set1: CGAGGAAAGTGGAAAAGTACAGTA; 3'set1: GATTGTCTAAAGAACAAACCATCATTA; 5'confirm:

CCCAAAATATCAAGTGTGAGAC; 3'confirm: CCATTCTTCAAGGTGCCGG) were PCR amplified using the following thermocycler conditions: 30 cycles of 95°C for 30 s, (53°C or 57°C, for set 1 and confirm, respectively) for 90 s, 68°C for 90 s. PCR products were purified (Qiagen, Valencia, CA) and sequenced (LoneStar Labs, Houston, TX). PCR products were then subcloned into Zero Blunt® TOPO® PCR Cloning Kit for Sequencing (Invitrogen, Carlsbad, CA) following the manufacturer's protocol. NotI and PstI restriction enzymes were used to linearize the probe and sequencing (LoneStar Labs) was performed to determine orientation. Antisense DIG-labeled probes (Roche Cat. No. 11277073910, Switzerland) were generated to detect CRISPLD2 expression and sense DIG-labeled probes were used as the negative control. Wild-type zebrafish were collected at the following stages for *in situ* hybridization analysis: 5-7 and 13-15 somites, 1-, 2-, 3-, 4- and 5 days post fertilization (dpf). *In situ* hybridizations were performed using standard techniques (www.zfin.org) [212]. LAS Montage Module (Leica, Wetzlar, Germany) was utilized to capture images.

2.9. Morpholino (MO) studies

Two MO antisense oligonucleotides were designed by GeneTools (Philomath, OR) to inject into zebrafish embryos: one to knockdown CRISPLD2 (*crispld2*MO: TCGATGTCAGGCGGTCTTACTTGGG) and a control MO (ControlMO: TGTAAACAGACTCACTTGTGTGTAG). Both MOs were diluted in nuclease free water to a stock concentration of 65mg/mL. The stock concentration was further diluted with nuclease free water to a working concentration of 12mg/mL. Injections (0.5-6 mg/mL) were

diluted in Danieu buffer and 0.5 uL of 2% phenol red was added to facilitate injections. One- to two-cell embryos were injected with 4pL of MO.

Embryos were stored at 28°C and observed and photographed at 1, 2, 3, 4 and 5 dpf. Gross abnormalities and survival rates were tallied. Tallies were normalized to wild-type observations. The chi-square test was utilized to assess differences between uninjected, control injected and knockdown injected embryos [213]. Embryos were harvested at 1 and 5 dpf for neural crest cell (NCC) marker and alcian blue staining, respectively.

Morpholino knockdown efficiency was assessed by extracting RNA from whole zebrafish embryos at 1- and 5 dpf using SuperScript™ III Reverse Transcriptase (Invitrogen, Carlsbad, CA) and standard protocols (www.zfin.org) [212]. Quality of RNA was assessed by gel analysis (Ambion Dentauring Agarose Gel Electrophoresis, http://www.ambion.com/techlib/append/supp/rna_gel.html) and quantity by spectrophotometer. CRISPLD2 set1 primers were used to detect CRISPLD2 expression. β-actin-specific primers (5': CAGGATCTTCATCAGGTAGTCTGTCA; 3': GTTTTCCCCTCCATTGTTGGAC) were used as a control and were PCR amplified under the following thermocycler contidions: 30 cycles of 95°C for 30 s, 57°C for 90 s, 68°C for 90 s.

2.10. Zebrafish expression studies

Zebrafish embryos were collected at either 1dpf for whole mount *in situ* hybridization studies or 5dpf for cartilage staining. Embryos collected at 1 dpf were dechorionated using 1:1000 pronase solution. Embryos were fixed in 4% paraformaldehyde and stored at -20°C in 100% methanol. *Dlx2* probe were obtained from Dr. Eric Swindell. *In situ* hybridizations

were performed as described in Section 2.8.1. Alcian blue staining was performed on 5 dpf embryos following standard protocol [212]. Images were collected using the LAS Montage Module (Leica, Wetzlar, Germany).

Chapter 3: MYH9 identified by the Genome Scan Approach

Note: The information presented in this chapter was published in 2009, in which I was the primary author: "Genomic screening identifies novel linkages and provides further evidence for a role of MYH9 in nonsyndromic cleft lip and palate." *European Journal of Human Genetics*. 17 (2): p.195-204. The *European Journal of Human Genetics* Journal does not require permission to reproduce manuscript content "in whole or in part in any printed volume (book or thesis) of which they are the author(s)" (http://www.nature.com/reprints/permission-requests.html).

3.0. Introduction

Genome scans, as discussed in section 1.5.1, use polymorphic markers to identify chromosome regions of the genome that are linked and/or associated with disease etiology. This is an important approach has been used to identify genes causing complex disorders in different populations [41-43]. Therefore, we used a genome scan to identify novel NSCLP chromosome regions in our dataset.

3.1. Materials and Methods

The details of the NSCLP family population, sample collection and DNA extraction, Linkage IVb Panel, SNP selection criteria, genotyping, and analyses (parametric and nonparametric linkage parameters and analysis, PDT, G-PDT, APL and GEE) are described in Chapter 2.

For this study, ten multiplex families (nine NHW and one African American) were subjected to Illumina Linkage IVb mapping panel (see Figure 2.2, page 28). Note that in Family 100, there are two individuals with reported notched gums, which may be part of the clefting spectrum (see Section 1.3). Therefore the data was analyzed in two ways: first with these individuals coded as "affected" and then with status codes as "unknown". The results of the analyses of the large African American family (F1100) were part of the work of another student and are not included here.

Based on the genome scan results, 439 NSCLP families (122 multiplex and 308 simplex parent-child trios) were genotyped using ten MYH9 SNPs (Table 3.1). Six of these SNPs were used in a previous study of Italian NSCLP families [55]. Four SNPs were selected using our SNP selection criteria (described in Section 2.2). Two SNPs (rs5995288 and

Table 3.1: MYH9 SNPs genotyped in this study and allele frequencies

Gene	dbSNP	Chr:bp	Location	Alleles*	NHW MAF[a]	Hispanic MAF[b]
MYH9	rs7078	22:35007860	downstream	T/C	0.30	0.25
MYH9	rs4821478	22:35020059	intron 28	A/G	0.35	0.35
MYH9	rs2009930	22:35029252	intron 19	C/T	0.35	0.37
MYH9	rs2239783	22:35035074	intron 15	C/T	0.26	0.27
MYH9	rs3752462	22:35040129	intron 13	C/T	0.36	0.37
MYH9	rs1002246	22:35044605	intron 10	T/C	0.66	0.67
MYH9	rs2071731	22:35048804	intron 5	G/A	0.63	0.66
MYH9	rs713659	22:35058638	intron 3	G/A	0.58	0.63
MYH9	rs713839	22:35063884	intron 3	A/G	0.62	0.65
MYH9	rs739097	22:35076025	intron 1	A/G	0.44	0.36

*Major allele as identified in NHW listed first

[a]Minor allele frequency of nonHispanic white dataset

[b]Corresponding frequency in Hispanic of nonHispanic white minor allele

rs2269529) were genotyped in the Italian assay [55] but were not included in this study because they were not available for TaqMan (Applied Biosystems, California) chemistries.

3.2. Results

3.2.1. Genome scan

> Ten families were subjected to the Illumina Linkage IVb SNP genome-wide panel. Nine nonHispanic white families generated 11 chromosomal regions with a LOD > 1.5 with parametric and/or nonparametric analysis [Table 3.2; Figure 3.1A-F]. The highest LOD score for either methodology was to two SNPs in the chromosome 22q12.2-q12.3 region (rs762883 and rs9862). Three of the families, 100, 1000, and 3000, had maximum LOD score greater than 1.0 in the region (1.05, 1.11 and 1.12 respectively), although not for the same set of SNPs (data not shown). [105]

3.2.2. Candidate gene testing

Ten MYH9 SNPs were genotyped in our NHW and Hispanic samples. All SNPs had a >95% call rate and were in HWE. The allele frequencies between SNPs in the NHW and Hispanic populations did not differ (Table 3.1); therefore, the data was analyzed in aggregate as well as stratified by FH [214].

> Two point parametric and nonparametric analyses in the combined dataset found evidence for linkage to a single SNP, rs1002246, in MYH9 (HLOD = 1.58 and p = 0.0006, respectively) [Table 3.3]. There was no evidence for linkage by multipoint analysis (data not shown). However, all three methods of association analysis identified evidence for altered transmission of this same SNP, rs1002246 [Table 3.4]. In addition, PDT detected evidence for an altered transmission of three additional SNPs. When the data was stratified by family history, a marginally significant association was still present for SNP rs1002246 in the simplex dataset (p = 0.03). Inspection of all two-SNP haplotypes found that most haplotypes constructed with rs1002246 and containing the "A" allele were overtransmitted, but generally did not reach significance, even when stratified by ethnicity and/or family history (data not shown). The minor allele frequency (i.e., frequency of the "A" allele) was 0.34 in the

Table 3.2: Genome Scan Results in nonHispanic white dataset*

				Maximum LOD Score			
				Parametric (hLOD)		Nonparametric (LOD)	
Chr	Region	dbSNP	cM	Aff[a]	Unk[b]	Aff[a]	Unk[b]
3	3p23	rs342758	58.08	0.35	0.85	0.92	1.64
3	3p21.1	rs11235	71.53	0.87	1.58	0.49	1.03
3	3p13-14.1**	rs736333	94.05	0.24	0.67	0.65	1.53
5	5p13.3**	rs1966983	51.74	0.94	1.64	0.72	1.19
5	5q35.2	rs1875189	192.51	0.57	0.53	1.59	1.04
6	6q27	rs727811	177.17	0.85	1.66	0.28	0.45
7	7q36.2-36.3**	rs7795368	182.8	1.7	2.59	0.57	1.16
9	9q22.2-22.32**	rs1048510	97.53	1.78	1.27	0.96	0.82
9	9q33.2-33.3	rs1984001	132.41	1.46	0.95	1.87	1.73
22	22q11.23-12.1	rs763281	21.41	1.75	1.22	1.46	0.99
22	22q12.2-12.3**	rs762883	36.94	2.66	2.12	2.49	2.17

*Modified from [105].

**Indicates novel region identified

aAff = Analysis of F100 individuals with notched gums classified as affected

bUnk = Analysis of F100 individuals with notched gums classified as unknown

Figure 3.1: LOD score plots shown by cMs for the nonHispanic white families for chromosomes (A) 2, (B) 3, (C) 6, (D) 7, (E) 9 and (F) 22. Blue line indicates parametric analysis and red line indicates nonparametric analysis. Solid lines indicates analysis with Family 100 III-4 and III-9 coded as affected while dotted lines indicates analysis with those individuals as unknown.

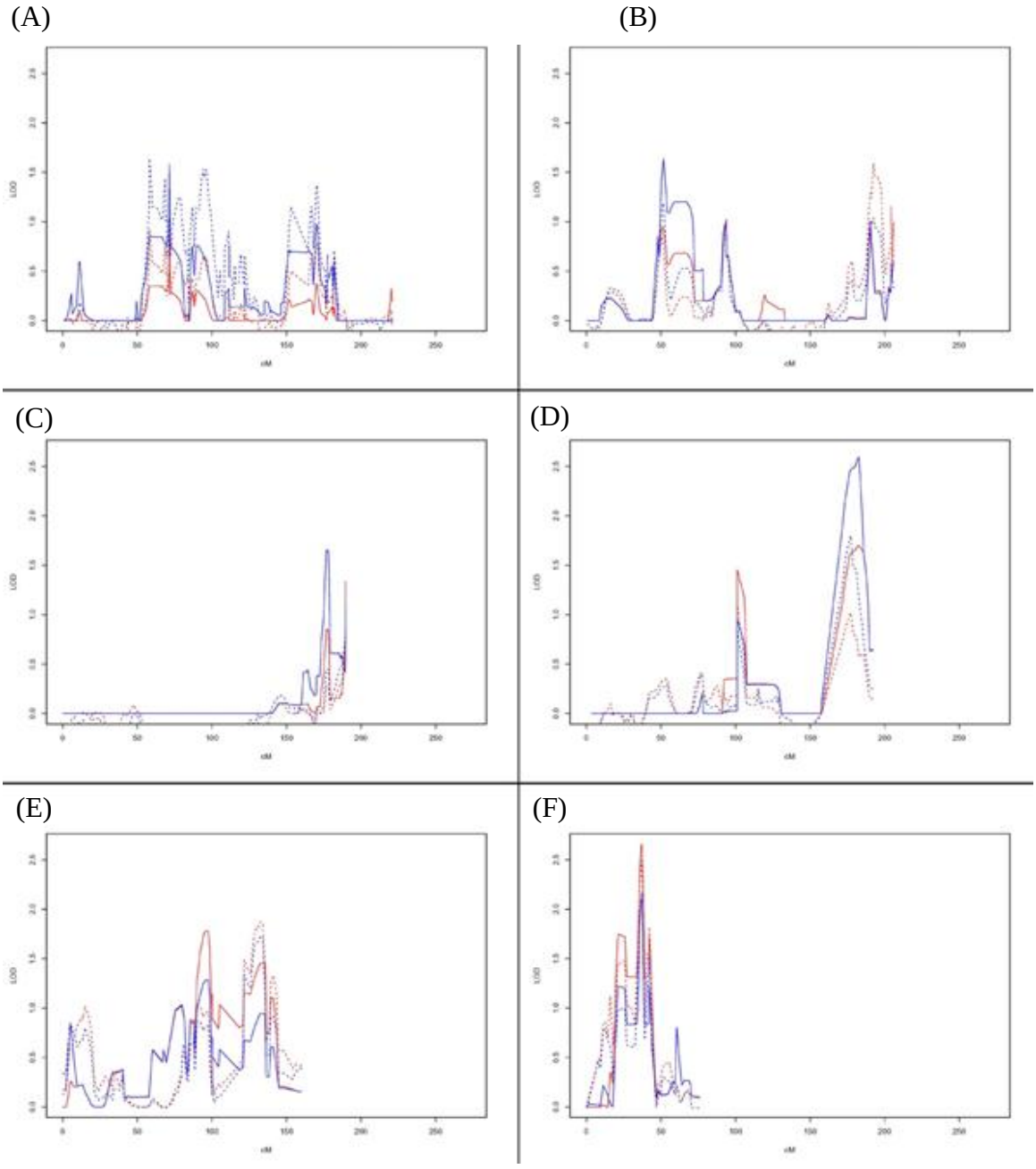


Table 3.3: MYH9 SNP linkage results*

			Parametric		Nonparametric	
dbSNP	cM	bp	HLOD	Alpha	LOD	P value
rs7078	35.008	35007860	0.018	0.083	0.01	0.4
rs4821478	35.02	35020059	0.033	0.109	0.02	0.4
rs2009930	35.029	35029252	0	0	-0.07	0.7
rs2239783	35.035	35035074	0.174	0.213	0	0.5
rs3752462	35.04	35040129	0.108	0.205	0.12	0.2
rs1002246	35.045	35044605	**1.578**	0.694	**2.26**	**0.0006**
rs2071731	35.049	35048804	0	0	0	0.5
rs713659	35.059	35058638	0	0	0	0.5
rs713839	35.064	35063884	0	0.005	0	0.5
rs739097	35.076	35076025	0	0	-0.01	0.6

*Modified from [105]

Table 3.4: MYH9 SNP association results from previous and current study*

		Italian sample	Entire sample			Multiplex families			Simplex families		
dbSNP	Location	TDT[a]	PDT	G-PDT	APL	PDT	G-PDT	APL	PDT	G-PDT	APL
rs5995288	intron 1	0.82	-	-	-	-	-	-	-	-	-
rs739097	intron 1	0.5	0.170	0.395	0.142	0.617	0.862	0.948	0.114	0.295	0.067
rs713839	intron 3	-	0.067	0.191	0.110	0.452	0.558	0.596	**0.043**	0.151	0.094
rs713659	intron 3	-	0.211	0.397	0.098	0.756	0.894	0.545	0.091	0.170	**0.014**
rs2071731	intron 5	0.88	0.120	0.186	0.165	0.607	0.663	0.456	0.065	0.144	**0.026**
rs1002246	intron 10	1	**0.020**	**0.043**	**0.009**	0.056	0.149	0.127	0.185	0.262	**0.033**
rs3752462	intron 13	**0.02**	**0.035**	0.117	0.602	0.114	0.216	0.662	0.162	0.382	0.353
rs2239783	intron 15	-	0.681	0.834	0.369	0.478	0.726	0.143	0.860	0.610	0.857
rs2009930	intron 19	0.4	**0.039**	0.131	0.177	0.195	0.386	0.718	0.087	0.257	0.065
rs4821478	intron 28	-	**0.023**	0.077	0.381	0.053	0.163	0.899	0.223	0.465	0.328
rs2269529	exon 34	0.69	-	-	-	-	-	-	-	-	-
rs7078	downstream	0.34	0.836	0.832	0.345	0.564	0.712	0.137	0.766	0.879	0.770

- = not genotyped

[a] p-value <0.05 bolded

[b] p-values have not been corrected for multiple testing

nonHispanic whites. Haplotypes with the "G" rs1002246 allele were either undertransmitted or demonstrated the expected transmission. [105]

3.3. Discussion

In this study, we performed a linkage scan to identify regions of the genome that potentially harbor NSCLP candidate genes that contribute to the etiologic heterogeneity of our NHW dataset. Eleven regions with a LOD score > 1.5 were identified, including five novel regions that have not been identified in previous NSCLP genome scans (Table 3.2). The 22q12.2-12.3 region, which was one of the novel genome scan regions, was the largest linkage peak in this study (LOD = 2.66). The MYH9 gene is in this region [160]. MYH9 is a class II nonmuscle myosin heavy chain gene involved in cell movement, cytokenesis and intracellular organization [215]. Mutations in MYH9 cause MYH9-related diseases: May-Hegglin anomally (OMIM:155100), Sebastian syndrome (OMIM:605249), Fechtner syndrome (OMIM:153640) and Epstein syndrome (OMIM:153650). These disorders are characterized by macrothrombocytopenia with inclusions of either granulocytes or leukocytes, kidney impairment, sensorineural deafness and cataracts can also be found in some of the MYH9-related diseases; however, none of these syndromes include a clefting phenotype [216]. MYH9 knockout mice are embryonic lethals and mice heterozygous for the knockout MYH9 allele are grossly normal [217,218]. MYH9 was identified as a candidate gene for NSCLP because immunohistochemistry studies show high levels of MYH9 expression in the palatal shelves prior to fusion, indicating that MYH9 plays an important role in palate development and is a potential NSCLP candidate gene [215]. Moreover, Martinelli *et al*. recently detected an association between MYH9 SNP rs3752462 and NSCLP in Italian NSCLP families (p=0.02) [55].

> To further evaluate the role of MYH9 in NSCLP, ten intragenic and flanking SNPs, which included six of the SNPs genotyped in the Italian families, were interrogated in our NSCLP sample.
>
> Linkage was found between a single SNP, rs1002246, in the MYH9 gene and our entire dataset [Table 3.3], consistent with the prior identification of this region in our genome scan. PDT, G-PDT and APL analyses all identified this SNP in our entire NSCLP sample. rs1002246 differs from the SNP found in the Italian NSCLP sample [Table 3.4]. rs1002246 is in intron 10 of the MYH9 gene but this is not in a region associated with intron- exon splicing. While intronic SNPs do not typically alter protein structure, an association with intronic variants has been reported for a number of complex diseases [46,47,49,183-185,203]. rs1002246 is in a region of reduced LD that can complicate identifying at-risk haplotypes. In fact, we were unable to identify a high-risk haplotype. Interestingly, rs3752462, the SNP identified by the Italian group, was only marginally significant in both datasets [Table 3.4]. Our finding of linkage by parametric and nonparametric analyses and association strongly suggests that MYH9 may play a causal role in NSCLP. Nevertheless, we cannot conclusively exclude the possibility that another gene in the 22q12.2-12.3 chromosomal region is contributing to the positive linkage and association. We are continuing to evaluate this region. [105]

Ten other regions were identified in this genome scan with a LOD > 1.5, including four regions that have not been previously identified in genome scans (3p14.1-13, 5p13.3, 7q36.2-36.3 and 9q22.22-22.32) (see Table 1.2, page 11). Three of the regions identified do not contain known NSCLP candidate genes (3p23, 3p14.1-13 and 5p13.3). The remaining seven regions (3p21.1, 5q35, 6q27, 7q36.2-36.3, 9q22.2-22.32, 9q33.2-33.3, and 22q11.23-12.1) have either previously been identified in other genome scans or contain known NSCLP candidate genes.

The 3p21.1 region was reported in Syrian and NHW samples [103,193]. We have previously reported an association between three short tandem repeats in this region and NSCLP and suggested that Wnt5A may play an etiologic role in NSCLP because of its proximity to these polymorphic variants [193]. Additionally, SNP interrogation of Wnt5a has shown that intronic SNP rs566926 is associated to NSCLP [118]. Wnt5A is expressed in the central nervous

system and facial processes during murine development and thus is considered a biologically plausibe candidate gene [193,219]. In Chapter 4, we describe a panel of WNT genes that were interrogated for association with NSCLP; association was detected between an intronic Wnt5A SNP and NSCLP (see Section 4.1) [118].

The 5q35 region was identified in a genome scan of Chinese multiplex families and contains 2 candidate genes: msh homolog 2 (Msx2) and fibroblast growth factor receptor 4 (FGFR4) [98,160]. Msx2 is expressed in the developing craniofacial processes. Mutations in the Msx2 gene cause midline facial clefts and a constellation of anomalies including amelogenesis imperfecta, cleft lip and palate and polycystic kidney disease [220,221]. Sequencing of Msx2 in NSCLP patients has identified four putatively benign mutations [126]. The FGF signaling pathway plays a critical role in craniofacial development and perturbations in many of the FGF pathway genes cause craniofacial abnormalities, including syndromic cleft lip and palate [24]. FGFR4 plays a significant role in myogenesis in the developing head but no association with NSCLP has been reported [111].

A genome scan study of Indian families reported linkage to chromosome region 6q27 [97]. The transcription factor T (T) gene is in this region and plays a critical role in mesoderm formation and differentiation [222]. Mutations in the T-box 22 (Tbx22) gene, which is in the same family as T, causes X-linked cleft palate and ankyloglossia syndrome (OMIM: #303400). Additionally, the T gene was considered a strong candidate for spina bifida but those results have been inconsistent [223,224]. Association studies between T and NSCLP have not been reported.

The 7q36.2-36.3 region is novel and contains sonic hedgehog homolog (SHH) and engrailed homeobox 2 (EN2), two known NSCLP candidate genes. SHH plays a critical role in the development and patterning of the craniofacial processes that give rise to the developing lip and palate [131]. Mutations in the human SHH gene cause

holoprosencephaly-3, which can be associated with cleft lip and palate (OMIM: #142945) [225]; SHH null mice can have cleft palate [110]. However, single strand conformational polymorphism analysis of the SHH gene in NSCLP patients has not identified any disease-causing mutations [132]. EN2 plays a role in central nervous system development and has also been implicated in autism spectrum disorders (OMIM: #608636) [226]. Previous RFLP testing found no association between the EN2 gene and NSCLP [227]. Although no positive findings have been found, these remain candidate genes of interest.

Two regions on chromosome 9, 9q22.2-22.32 and 9q33.2-33.3, were identified in this genomic scan. The 9q22.2-22.32 region is novel and contains both the BarH-like homeobox 1 (Barx1) and patched homolog 1 (PTCH1). Mouse Barx1 is expressed in the first and second branchial arches which form the craniofacial processes [130]. The role of Barx1 in NSCLP has not been evaluated. PTCH1, an SHH receptor, in combination with Barx1, plays a role in midface structure formation [129]. Mutations in PTCH1 cause Basal Cell Nevus syndrome (BCNS; OMIM: #109400) and five percent of BCNS patients have cleft lip and palate [228]. Interestingly, forkhead box E1 (FOXE1), which plays a role in thyroid morphogenesis, is located just outside this region in 9q22.33 [229]. The FOXE1 region has been identified in NSCLP genome scans [230]. Mutations in FOXE1 cause Bamforth-Lazarus syndrome of which cleft palate is a associated finding (OMIM: #241850). However, sequencing of the FOXE1 gene in NSCLP probands identified two missense mutations, both of which were predicted to be benign [126]. [105]

22q11.23-12.1 was previously reported in a genome study of Indian families [97]. This region is of particular interest because it contains glutathione s-transferase theta 1 and 2 (GSTT1 and GSTT2) genes that are involved in the metabolism of tobacco byproducts [71,77,79,160]. Exposure to tobacco smoke during pregnancy and fetal GSTT1-null alleles, which have been shown to interact with maternal smoking, have inconsistently been identified as a risk factor for NSCLP [71,77,79].

3.4. Summary

Here, we report six novel regions (2p22, 3p13-14.1, 5p13.3, 7q36.2-36.3, 9q22.2-22.32 and 22q12.2-12.3) that can be interrogated for NSCLP genes. We identified eight regions that were previously implicated in other genomic scans in a variety of populations. The strongest linkages were found for the 8q21.3-24.12 and 22q12.2-12.3 regions, with the MYH9 (chromosome 22) gene exhibiting the strongest

evidence for a causal role. This gene should be further evaluated in other populations for its possible role in NSCLP. The results of this study are important as they provide additional regions to search for new NSCLP genes and confirm the findings of earlier genomic scans. Our findings also demonstrate that large multiplex families with complex disorders can be successfully used in genome mapping strategies. [105]

Chapter 4: Evaluation of WNT genes identified by the Candidate Gene Approach

Note: The information presented in this chapter was published in 2008, in which I was the primary author: "Variation in WNT genes is associated with non-synromic cleft lip with or without cleft palate." *Human Molecular Genetics*. 17 (14): p.2212-2218. The *Human Molecular Genetics* Journal does not require permission to reproduce manuscript content "in full or in part in a thesis or dissertation" (http://www.oxfordjournals.org/access_purchase/publication_rights.html).

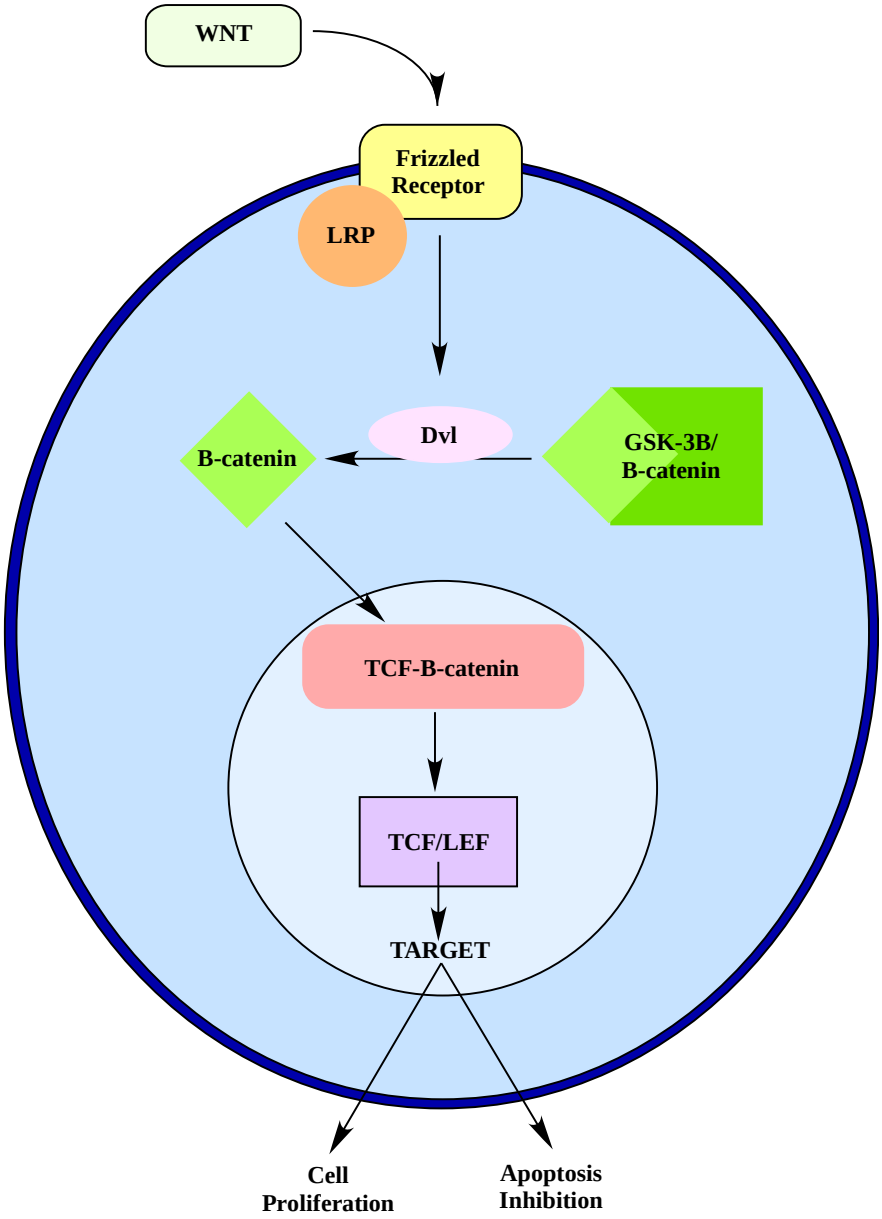
4.0. Introduction

As discussed in section 1.5.2, the candidate gene approach has provided important insights about genetic variation contributing to nonsyndromic orofacial clefts. The wingless-type (WNT) family of genes and signaling pathway play a critical role in growth and development, including embryonic induction, epithelial and mesenchymal cellular polarity, cell fate determination, cytoskeletal organization and cell proliferation [112,231,232]. WNT genes are expressed in vertebrate craniofacies throughout development, including those processes that form the lip and palate [20,21,112,114,115,117,233,234]. Moreover, craniofacial abnormalities, including orofacial clefts, are found in WNT knockout mice and zebrafish [116,235]. Based on all of these observations, we hypothesized that the WNT genes were biologically plausible candidate genes for NSCLP and undertook interrogation of this family of genes.

4.1. WNT gene family

There are 19 WNT genes which share a conserved motif of 23-24 cysteine residues and produce 350-400 amino acid palmitoylated protein products [15,112]. These genes function in either the canonical or noncanonical WNT pathway [15,113,231]. The canonical WNT pathway involves the binding of the WNT signaling molecule to a Frizzled gene receptor, activating dishevelled (Dvl) and stabilizing β-catenin, which can then enter the nucleus, form a complex with transcription factor/lymphoid enhancer-binding factor (TCF/LEF), and activate downstream target genes (i.e., *CyclinD1, c-myc*) (Figure 4.1) [15,113]. *Wnts -1*, *-2*, *-2B*, *-3*, *-3A*, *-6*, *-7B*, *-8A*, *8B* and *-9B* all function via the canonical WNT pathway. *Wnts -4*, *-5A*, and *-11* have been shown to function without β-catenin or TCF-mediated transcription

Figure 4.1: Schematic of the canonical WNT pathway.



Legend: LRP = low density lipoprotein; Dvl = disheveled; TCF/LEF = transcription factor/lymphoid enhancer-binding factor (TCF/LEF)

in noncanonical WNT pathways, including axon guidance, stem cell differentiation and planar cell polarity [119,231,236-238]. Six WNT genes (*Wnts -5b, -7a, -9a, -10a,* -10b, and -*16*) have not been assigned to either the canonical or noncanonical pathway.

4.2. Selection of WNT candidate genes

Seven WNT genes were chosen for this study: *Wnt3, Wnt9B, Wnt3A, Wnt11, Wnt8A, Wnt5A,* and *Wnt7A.* Table 4.1 lists specific details about each gene and their unique properties.

4.2.1. Wnt3 and Wnt9B

As discussed in section 1.5.2.d, the clf2 region in the A/WySn mouse strain that is associated with orofacial clefting in mice contains both the *Wnt3* and *Wnt9B* genes [158]. Recent studies have identified an inserted transposable element downstream of *Wnt9B* that inhibits normal Wnt9B expression [159]. Additionally, when A/WySn mice were bred with *Wnt9B*-/- mice, there was a higher prevalence of clefting in the progeny (67%) versus in the founders [116]. This data suggest that *Wnt9B* is responsible for the clefting phenotype in murine models and therefore an excellent candidate gene for NSCLP.

Wnt3 is a NSCLP candidate gene because a nonsynonymous *Wnt3* mutation causes Tetra-Amelia (OMIM:#273395), a rare autosomal recessive disorder in which individuals are missing all four limbs and may have an orofacial clefting phenotype [239]. Additionally, *Wnt3* has been shown to antagonize Sonic Hedgehog (SHH) signaling in chick neural plate patterning, from which neural crest cells (NCCs) migrate [16]. Notably, SHH was in one of the regions identified in the genome scan discussed in chapter three (see

Table 4.1: WNT genes interrogated in this study

Gene	Chromosomal Region	Previously Identified NSCLP Region	Predicted Function	Animal Models	Syndrome/ OMIM#
Wnt3	17q21.32	Yes [100]	Primary axis formation [240]	Knockout mouse lacks anterior-posterior patterning[240]	Tetra-Amelia (#273395) [239]
Wnt3A	1q42.13	No	Neural patterning [241]	Vestibular morphogenesis [242]	NI
Wnt5A	3p14.3	Yes [193]	Stem cell differentiation [243,244]	Chicken-mesenchymal expression in the craniofacies [120]	NI
Wnt7A	3p25.1	Yes [98,100,103]	Formation and CNS vascular differentiation [245]	Knockout mouse infertile due to abnormal mullerian duct formation/regression [246]	Al-Awadi/Raas-Rothschild syndrome (#276820) [247]; Fuhrmann syndrome (#228930) [248,249]
Wnt8A	5q31.2	Yes [98]	NCC migration [114]	Zebrafish and Xenopus- initial induction of neural crest [115,250]	NI
Wnt9B	17q21.32	Yes [100]	Midfacial morphogenesis [9]	Mouse- Cleft lip [116]	NI
Wnt11	11q13.5	Yes [98,101]	Neural patterning[241]	Zebrafish knockout- midline defects [235]	NI

NI = No syndrome identified

Section 3.3) [105]. Therefore, variation in *Wnt3* may disrupt normal craniofacial development and contribute to clefting etiology.

4.2.2. Wnt3A

> Wnt3A, in mice, has been shown to control the fate of both mesenchymal and NCCs in the craniofacial processes and to regulate palatal fusion [21]. Embryonic expression of Wnt3A is upregulated by fibroblast growth factor 8 (FGF8), which is highly expressed in the facial primordia and is also an important regulator of craniofacial development [24,251]. Additionally, Wnt3A signaling is also decreased when fibroblast growth factor receptor 1 (FGFR1), which regulates the epithelial-mesenchymal transformation necessary for normal palatal development, is inactivated [251]. [118]

4.2.3. Wnt11

> Embryogenic studies have shown that Wnt11 is necessary for directing migrating NCCs, which will later form ectomesenchyme of the developing face [14,114]. Also, bone morphogenic protein 4 (BMP4), which functions in the same pathway as Msx homeobox genes, is necessary for upper labial fusion in mice [24,252], and downregulates Wnt11 during normal urogenital development [11]. Inactivation of BMP4 causes cleft lip in mice; therefore, the relationship between BMP4 and Wnt11 suggests that variation in Wnt11 could play an etiological role in NSCLP [252]. [118]

4.2.4. Wnt8A

Wnt8A, like *Wnt3A* and *Wnt11*, has been shown to play an important role in NCC differentiation, NCC migration and craniofacial development in the mouse [14,21,114].

> NCCs are derived in the neural tube and migrate into the pharyngeal/branchial arches and differentiate into connective tissue and bone of the head and neck [14,253]. Several mouse studies have demonstrated that genes inactivated in NCCs (i.e., Tgfbr2, Tcof1 and AP-2alpha) cause orofacial clefting, as well as other malformations [254-256]. [118]

Together, this supports that *Wnt8A* is a putative NSCLP candidate gene.

4.2.5. Wnt5A

> We previously identified a region on chromosome 3p21-14 between two STR markers that may harbor a clefting gene [193]. Wnt5A is a candidate gene in this region because *in situ* hybridization studies have shown that Wnt5A is expressed in the frontonasal prominence and maxillary process, which fuse to form the primary palate [20]. [118]

4.2.6. Wnt7A

Mutations in *Wnt7A* cause Fuhrmann syndrome (OMIM:228930), which consists of bowed femurs, aplasia or hypoplasia of the fibula, and poly-, syn-, and oligodactyly [248,249]; cleft lip and palate is an occasional finding in this syndrome [257]. Mutations in *Wnt7A* also cause Al-Awadi/Raas-Rothschild syndrome (OMIM:276820), which present with hypoplastic limbs and "unusual facies", including a "high and narrow palate" [258]. The palate phenotypes in these syndromes suggest that *Wnt7A* is a potential candidate gene for NSCLP.

4.3. Materials and Methods

The details of the NSCLP family population, sample collection and DNA extraction, SNP selection criteria, genotyping, and analysis (PDT, G-PDT, APL and GEE) are completely described in chapter 2.

For this study, 566 NSCLP families (132 multiplex, 235 simplex parent-child trios and 199 simplex parent-child duos) were genotyped using 38 SNPs. This included ten in *Wnt3*, six in *Wnt3A*, one in *Wnt5A*, nine in *Wnt7A*, three in *Wnt8A*, three in *Wnt9B* and six in *Wnt11* (Table 4.2). Two previously genotyped *Wnt5A* STRs were also included in the analysis [193].

Table 4.2: WNT SNPs genotyped in this study and allele frequencies

Gene	dbSNP	Chr:bp	Location	Alleles†	NHW MAF[a]	Hispanic MAF[b,*,**,***]	p-value
Wnt3A	rs708111	1:226257988	upstream	C/T	0.483	0.524	0.290
Wnt3A	rs708114	1:226264119	intron 1	C/T	0.469	0.557*	0.034
Wnt3A	rs3094912	1:226276438	intron 1	A/T	0.454	0.500	0.228
Wnt3A	rs3121310	1:226291447	intron 2	A/G	0.308	0.382	0.055
Wnt3A	rs752107	1:226313974	downstream	C/T	0.308	0.356	0.175
Wnt3A	rs1745420	1:226318355	downstream	C/G	0.144	0.178	0.193
Wnt5A	rs566926	3:55487719	downstream	A/C	0.261	0.333	0.550
Wnt7A	rs1124480	3:13832970	upstream	C/T	0.459	0.418	0.274
Wnt7A	rs9840696	3:13840076	intron 1	A/G	0.375	0.392	0.651
Wnt7A	rs6778046	3:13846680	intron 1	C/T	0.369	0.415	0.212
Wnt7A	rs9863149	3:13858361	intron 1	C/T	0.315	0.216**	0.004
Wnt7A	rs934450	3:13862730	intron 1	A/G	0.329	0.257*	0.037
Wnt7A	rs1433354	3:13872246	intron 2	A/G	0.305	0.453***	<0.0001
Wnt7A	rs6442414	3:13881839	intron 2	A/G	0.312	0.221**	0.008
Wnt7A	rs11128663	3:13887851	downstream	C/G	0.400	0.297**	0.005
Wnt7A	rs4685048	3:13897736	downstream	A/C	0.487	0.412*	0.046
Wnt8A	rs4835761	5:137445768	upstream	A/G	0.469	0.489	0.619
Wnt8A	rs2040862	5:137447888	intron 2	C/T	0.172	0.108*	0.030
Wnt8A	rs2306110	5:137455986	downstream	A/C	0.477	0.489	0.765
Wnt11	rs663746	11:75571777	upstream	C/T	0.498	0.474	0.524
Wnt11	rs1533763	11:75578175	intron 1	A/T	0.209	0.172	0.272
Wnt11	rs1533767	11:75583448	exon 3	A/G	0.267	0.279	0.718
Wnt11	rs689095	11:75591735	intron 4	A/G	0.307	0.324	0.667
Wnt11	rs596339	11:75594155	intron 4	A/G	0.338	0.412	0.060
Wnt11	rs1568507	11:75596967	downstream	A/T	0.315	0.324	0.820
Wnt3	rs142167	17:42150418	downstream	A/G	0.258	0.523***	<0.0001
Wnt3	rs7216231	17:42170907	downstream	A/G	0.055	0.409***	<0.0001
Wnt3	rs199525	17:42203002	intron 4	G/T	0.220	0.159	0.070
Wnt3	rs70602	17:42214876	intron 1	C/T	0.219	0.149*	0.041
Wnt3	rs199501	17:42217772	intron 1	A/G	0.256	0.538***	<0.0001
Wnt3	rs199498	17:42220763	intron 1	C/T	0.241	0.489***	<0.0001
Wnt3	rs111769	17:42227151	intron 1	C/T	0.416	0.258***	<0.0001
Wnt3	rs3851781	17:42246300	intron 1	C/T	0.474	0.657***	<0.0001
Wnt3	hCV1139279	17:42248220	upstream	C/G	0.198	0.11**	0.005
Wnt3	rs9890413	17:42256448	upstream	A/G	0.363	0.215***	0.0002

Gene	dbSNP	Chr:bp	Location	Alleles†	NHW MAF[a]	Hispanic MAF[b,***]	p-value
Wnt9B	rs2165846	17:42296365	intron 1	A/G	0.421	0.720***	<0.0001
Wnt9B	rs1530364	17:42306776	intron 2	A/G	0.261	0.317	0.128
Wnt9B	rs197915	17:42345521	downstream	A/G	0.434	0.432	0.972

†Major allele as identified in NHW listed first

[a]Minor allele frequency of nonHispanic white dataset

[b]Corresponding frequency in Hispanic of nonHispanic white minor allele

*p<0.05; **p<0.01; **p<0.001

4.4. Results

All SNPs were in Hardy-Weinberg equilibrium. Comparisons of SNP allele frequencies, after Bonferroni correction, showed significant differences between European Americans and Hispanics ($p<0.001$) in 3 of the 7 genes [Table 4.2]. Therefore, in addition to an overall analysis of the complete dataset, the data were stratified by ethnicity and examined separately. There was no evidence for linkage under any of the tested parametric models or with nonparametric methods for any of the regions [(Supplemental Table 4.1)].

PDT and Geno-PDT results for the entire dataset are presented in [Table 4.3]. Out of the seven genes interrogated, 5 have SNPs that meet a nominal threshold of 0.05. If we correct the significance level for the number of genes tested, then SNPs in three genes (Wnt3A, Wnt5A and Wnt11) still yield significant results ($p\leq0.007$). [Table 4.4 (and Supplemental Tables 4.2 and 4.3)] contain the results of PDT, Geno-PDT and APL analyses when the families were grouped by ethnicity and also subdivided according to the presence or absence of family history (FH). SNPs meeting a nominal threshold (p-value <0.05) were found in each gene, generally by multiple methods. In the overall European American sample, and when by stratified by FH, associations were identified with multiple SNPs in Wnt3, Wnt5A and Wnt11 [Table 4.4]. In addition, SNPs in Wnt7A were significant in the total European American sample and in the positive FH subgroup, while a SNP in Wnt9B was significant only in the positive FH subgroup. Multiple SNPs in Wnt3A were significant in the total European American sample and in the negative FH subgroup and a SNP in Wnt8A was significant in the negative FH subgroup.

When examining the entire Hispanic sample or when stratifying by FH, associations were identified for SNPs in Wnt3A and Wnt9B. Two SNPs were significant in the total sample; one of these was significant in the negative FH subgroup. One SNP in Wnt11 was significant in the positive FH subgroup and when the sample was considered as a whole. Lastly, a SNP in Wnt3 was significant in the positive FH subgroup. Overall, across ethnicity, the same SNPs were significant in four genes: Wnt3A (2 SNPs), Wnt7A (4 SNPs), Wnt11 (1 SNP) and Wnt9B (1 SNP) and both ethnicities had significant SNPs in Wnt3, although there was no overlap.

Table 4.3: Results of WNT association analysis in complete dataset.*

Gene	dbSNP	PDT	G-PDT
Wnt3	rs142167	0.147	0.278
Wnt3	rs7216231	0.359	0.584
Wnt3	rs199525	0.353	0.636
Wnt3	rs70602	0.132	0.302
Wnt3	rs199501	0.190	0.249
Wnt3	rs199498	0.269	0.220
Wnt3	rs111769	0.498	0.331
Wnt3	rs3851781	0.914	0.219
Wnt3	hCV1139279	0.299	0.519
Wnt3	rs9890413	0.853	0.741
Wnt3A	rs708111	0.127	0.310
Wnt3A	rs708114	0.378	0.570
Wnt3A	rs3094912	0.334	0.352
Wnt3A	rs3121310	0.034	0.088
Wnt3A	rs752107	0.021	0.022
Wnt3A	rs1745420	**0.006**	0.009
Wnt7A	rs1124480	0.290	0.314
Wnt5A	D3S3719	0.320	0.340
Wnt5A	rs566926	**0.002**	0.008
Wnt5A	D3S2408	0.757	0.358
Wnt7A	rs9840696	0.272	0.483
Wnt7A	rs6778046	0.172	0.049
Wnt7A	rs9863149	0.183	0.415
Wnt7A	rs934450	0.209	0.454
Wnt7A	rs1433354	0.228	0.437
Wnt7A	rs6442414	0.665	0.091
Wnt7A	rs11128663	0.229	0.332
Wnt7A	rs4685048	0.010	0.049
Wnt8A	rs4835761	0.490	0.067
Wnt8A	rs2040862	0.306	0.045
Wnt8A	rs2306110	0.535	0.164
Wnt9B	rs2165846	0.962	0.971
Wnt9B	rs1530364	0.869	0.939

Wnt9B	rs197915	0.469	0.149
Wnt11	rs663746	0.083	**<0.001**
Wnt11	rs1533763	0.199	0.378
Wnt11	rs1533767	0.099	0.009
Wnt11	rs689095	0.386	0.151
Wnt11	rs596339	0.747	0.747
Wnt11	rs1568507	0.124	0.178

*All SNPs with p<0.01 are bolded.

Table 4.4: Results of WNT association analysis by ethnicity and family history*

A. NHW		All			Positive FH			Negative FH		
Gene	dbSNP†	PDT	G-PDT	APL	PDT	G-PDT	APL	PDT	G-PDT	APL
Wnt3	rs70602	0.154	0.308	0.041	0.309	0.585	0.009	0.299	0.426	0.365
Wnt3	rs199498	0.162	0.239	0.161	0.087	0.139	**0.006**	1.000	0.941	0.947
Wnt3	rs9890413	0.948	0.698	0.593	0.932	0.216	0.504	0.840	**0.005**	0.835
Wnt3A	rs752107	0.023	0.027	**0.003**	0.31	0.441	0.125	0.0178	0.027	0.009
Wnt5A	rs566926	**0.005**	0.016	**0.004**	0.043	0.107	0.049	0.0462	0.144	0.02
Wnt8A	rs2040862	0.352	0.055	0.061	0.581	0.372	0.085	0.009	0.016	**0.001**
Wnt9B	rs197915	0.160	0.051	0.573	0.072	0.180	0.003	0.926	0.060	0.310
Wnt11	rs663746	0.067	**0.001**	0.449	0.104	0.047	0.278	0.361	0.014	0.849
Wnt11	rs1533767	0.114	0.012	**0.001**	0.744	0.010	0.100	0.006	0.016	0.003
B. Hispanics		All			Positive FH			Negative FH		
Gene	dbSNP†	PDT	G-PDT	APL	PDT	G-PDT	APL	PDT	G-PDT	APL
Wnt3A	rs708111	NS	NS	NS	0.034	0.135	**0.007**	0.297	0.426	**0.005**
Wnt3A	rs3094912	0.343	0.572	0.031	NS	NS	NS	0.162	0.240	**<0.001**
Wnt3A	rs1745420	0.083	0.157	**<0.001**	0.096	0.264	0.009	0.346	0.376	0.008

*All SNPs with $p<0.01$ are shown and $p\leq0.007$ are bolded.

> The results of haplotype analysis are shown in [Table 4.5]. Because of the small sample size, data were not stratified by FH. In both European American and Hispanic samples, four haplotypes in Wnt3A and one in Wnt8A were significantly associated [Table 4.5]. Wnt11 haplotypes were only significant in the European American sample [Table 4.5]. While some of the SNPs within genes were in significant LD, most of the SNPs in the significantly overtransmitted haplotypes had $r^2 < 0.3$ [Supplemental Table 4.3].
>
> GEE analysis identified significant gene-gene interactions for SNPs in Wnt3A and Wnt3 (p=0.0037) in the European American sample and SNPs in Wnt3A and Wnt5A in the Hispanic sample (p=0.0388) [Table 4.6]. Models for interaction considering dominant and recessive main effects were tested in the European American sample. The best model for increased susceptibility was when at least one copy of the rarer allele was present at each locus (p=0.03, data not shown). Interestingly, when there was homozygosity for the common allele in Wnt3, the rare allele in Wnt3A was protective (p = 0.007, data not shown). The Hispanic sample was too small to evaluate the main effects. [118]

4.5. Discussion

In this study we interrogated seven WNT genes (*Wnt3*, *Wnt9B*, *Wnt3A*, *Wnt11*, *Wnt8A*, *Wnt5A*, and *Wnt7A*) for association with NSCLP. We found that (1) all seven genes reached nominal significance ($p<0.05$) in this study, (2) *Wnt3A*, *Wnt5A* and *Wnt11* remained significant after implementing a modest Bonferroni correction, and (3) only *Wnt3A* was associated with NSCLP in both populations and all analyses.

Though the seven WNT genes in this study were selected because of the roles each has in craniofacial development (see Section 4.2), it was surprising that we found association with each gene and NSCLP ($p<0.05$). These results raised issues concerning multiple testing.

Table 4.5: Overtransmitted WNT haplotypes by ethnicity

A. NHW						
Gene	SNP1	SNP order*	SNP2	SNP order	Overtransmitted Haplotype	p-value
Wnt3	rs142167	1	rs199525	3	2,1	0.049
Wnt3	rs7216231	2	rs199525	3	2,1	0.033
Wnt3	rs199525	3	rs70602	4	1,2	0.05
Wnt3A	rs708111	1	rs752107	5	2,1	0.010
Wnt3A	rs708114	2	rs752107	5	1,1	0.014
Wnt3A	rs3094912	3	rs752107	5	1,1	0.037
Wnt3A	rs3121310	4	rs752107	5	2,1	0.017
Wnt3A	rs752107	5	rs1745420	6	1,2	**0.001**
Wnt8A	rs4835761	1	rs2306110	3	1,1	0.05
Wnt11	rs663746	1	rs1533767	3	2,2	0.047
Wnt11	rs1533763	2	rs596339	5	2,2	0.012
Wnt11	rs1533767	3	rs1568507	6	2,2	0.013
B. Hispanics						
Gene	SNP1	SNP order*	SNP2	SNP order	Overtransmitted Haplotype	p-value
Wnt3	rs199498	6	rs9890413	10	2,2	0.047
Wnt3A	rs708111	1	rs3121310	4	1,1	0.022
Wnt3A	rs708111	1	rs752107	5	1,1	0.008
Wnt3A	rs708111	1	rs1745420	6	1,2	**0.007**
Wnt3A	rs708114	2	rs3094912	3	2,1	**0.004**
Wnt3A	rs708114	2	rs1745420	6	2,2	**<0.001**
Wnt3A	rs3094912	3	rs3121310	4	1,1	**<0.001**
Wnt3A	rs3094912	3	rs752107	5	1,1	**<0.001**
Wnt3A	rs3094912	3	rs1745420	6	1,2	**0.003**
Wnt3A	rs3121310	4	rs752107	5	1,1	0.027
Wnt3A	rs3121310	4	rs1745420	6	1,2	**<0.001**
Wnt3A	rs752107	5	rs1745420	6	1,2	**<0.001**
Wnt8A	rs4835761	1	rs2306110	3	2,2	0.049

*Order of SNP within candidate gene

+All SNPs with $p < 0.05$ are shown and $p \leq 0.007$ are bolded

Table 4.6: WNT gene-gene interactions*

Population	Gene1/SNP	Gene2/SNP	GEE p-value
nonHispanic whites	*Wnt3A*/rs752107	*Wnt3*/rs70602	**0.004**
Hispanic	*Wnt3A*/rs1745420	*Wnt5A*/rs566926	0.039

*All SNPs with P <0.05 are shown and P <0.007 are bolded.

> The issue of correcting for multiple tests is complicated when multiple methodologies are performed, and neither the tests nor all the SNPs are independent. Moreover, which correction factor to use is further complicated when the SNPs are in genes considered viable candidates from other scientific evidence. Therefore, we elected to report results meeting a nominal p-value of 0.05. In addition, for the association studies of the individual SNPs, we have employed a correction factor of 7 (0.05/7 = 0.007) to take into consideration the number of genes evaluated. This represents a generally acceptable correction strategy; one not as stringent as a strict Bonferroni, yet still taking the number of totally independent analyses into account. This correction is not employed in the haplotype or gene-gene interaction analyses, as these are truly exploratory in nature and are restricted by sample size. Clearly, replication of these studies is critical, as well as examining gene sequences to identify etiologic susceptibility mutations. [118].

While multiple testing remains an issue, the most striking results came from two genes that play a role in NCC differentiation and migration, *Wnt3A* and *Wnt11* (Tables 4.3 and 4.4) *Wnt3A* was significant in the complete dataset for both ethnicities, regardless of FH of NSCLP (Tables 4.3 and 4.4). *Wnt3A* was also significant in haplotypes and GEE interactions in both ethnicities (Tables 4.5 and 4.6). Of interest, each of the six *Wnt3A* SNPs genotyped was significant in at least one of the above analyses. None of the SNPs were coding variants or change transcription binding sites. rs752107 is in the 3' untranslated (UTR) region. Polymorphisms and mutations in 3'-UTRs have been associated with other

diseases, including *ITGB2* and Hirschsprung disease-associated enterocolitis, *SOX10* and complex severe neurocristopathy and *GATA4* and congenital heart disease [259-261]. The other five *Wnt3A* SNPs are intronic. Intronic SNPs are associated in several complex diseases, including NSCLP, Hirschsprung disease and age-related macular degeneration [46,47,49,183-185]; however, finding functional explanations for these intronic SNP associations is difficult. Interestingly, the *Wnt3A* haplotype of rs752107 and rs1745420 was significant in both the NHW ($p=0.001$) and Hispanic ($p<0.001$) groups (Table 4.5). GEE analysis also identified interaction between *Wnt3A* SNPs and SNPs in *Wnt3* and *Wnt5A* in our NHW and Hispanic groups, respectively (Table 4.6). The GEE test does not detect biochemical interactions; however, *Wnt3A* and *Wnt3* are both involved in NCC development [16,21] while Wnt5A and Wnt3A are both expressed in the craniofacial processes that form the upper lip and palate [20,24,251]. Therefore, detection of interactions is consistent with the biologic functions of these genes.

Overall, the results from this study suggest that variations in WNT genes that contribute to normal craniofacial development also play an etiological role in NSCLP. Specifically, those WNT genes that function in NCC differentiation and migration (*Wnt3A* and *Wnt11*) were highly associated with NSCLP. During development, NCCs form the connective tissue and skeleton that become the craniofacies [254]. The results suggest that altered regulation in any or multiple of these genes can predispose an individual to NSCLP.

Supplemental Table 4.1: WNT linkage disequilibrium results

A. *Wnt3*

dbSNP	rs142167	rs7216231	rs199525	rs70602	rs199501	rs199498	rs111769	rs3851781	hCV1139279	rs9890413
rs142167		0.471	0.101	0.160	0.797	0.738	0.015	0.018	0.037	0.000
rs7216231	0.065		0.102	0.129	0.428	0.293	0.113	0.039	0.031	0.040
rs199525	0.581	0.004		0.718	0.096	0.130	0.255	0.153	0.010	0.082
rs70602	0.638	0.003	0.767		0.149	0.144	0.243	0.124	0.002	0.126
rs199501	0.798	0.063	0.621	0.765		0.769	0.021	0.000	0.023	0.001
rs199498	0.673	0.003	0.640	0.716	0.731		0.009	0.032	0.031	0.015
rs111769	0.029	0.002	0.025	0.026	0.024	0.025		0.095	0.045	0.209
rs3851781	0.064	0.001	0.074	0.082	0.074	0.098	0.018		0.152	0.343
hCV1139279	0.001	0.001	0.006	0.005	0.003	0.004	0.128	0.163		0.037
rs9890413	0.030	0.002	0.040	0.039	0.031	0.045	0.165	0.361	0.120	

B. *Wnt3A*

dbSNP	rs708111	rs708114	rs3094912	rs3121310	rs752107	rs1745420
rs708111		0.027	0.688	0.021	0.176	0.157
rs708114	0.289		0.001	0.463	0.012	0.001
rs3094912	0.561	0.040		0.004	0.217	0.190
rs3121310	0.137	0.484	0.013		0.014	0.014
rs752107	0.195	0.030	0.282	0.054		0.005
rs1745420	0.082	0.010	0.135	0.000	0.046	

C. *Wnt5A*

STR/dbSNP	D3S3719	rs566926	D3S2408
D3S3719		0.000	0.000
rs566926	0.000		0.000
D3S2408	0.000	0.000	

D. *Wnt7A*

dbSNP	rs1124480	rs9840696	rs6778046	rs9863149	rs934450	rs1433354	rs6442414	rs11128663	rs4685048
rs1124480		0.024	0.030	0.003	0.004	0.014	0.006	0.003	0.001
rs9840696	0.087		0.731	0.111	0.155	0.239	0.023	0.008	0.014
rs6778046	0.065	0.799		0.193	0.239	0.252	0.044	0.003	0.026
rs9863149	0.037	0.214	0.244		0.679	0.001	0.001	0.013	0.002
rs934450	0.032	0.219	0.263	0.900		0.011	0.003	0.004	0.001
rs1433354	0.004	0.142	0.158	0.002	0.000		0.085	0.033	0.029
rs6442414	0.000	0.014	0.015	0.075	0.084	0.008		0.598	0.369
rs11128663	0.001	0.000	0.000	0.015	0.019	0.021	0.533		0.487
rs4685048	0.010	0.015	0.012	0.018	0.022	0.115	0.322	0.526	

E. *Wnt8A*

dbSNP	rs4835761	rs2040862	rs2306110
rs4835761		0.081	0.935
rs2040862	0.128		0.079
rs2306110	0.932	0.118	

F. *Wnt9B*

dbSNP	rs2165846	rs1530364	rs197915
rs2165846		0.018	0.062
rs1530364	0.030		0.124
rs197915	0.082	0.069	

G. *Wnt11*

dbSNP	rs663746	rs1533763	rs1533767	rs689095	rs596339	rs1568507
rs663746		0.000	0.044	0.009	0.000	0.007
rs1533763	0.026		0.004	0.094	0.296	0.092
rs1533767	0.108	0.018		0.021	0.004	0.017
rs689095	0.023	0.120	0.058		0.327	0.948
rs596339	0.027	0.445	0.017	0.232		0.318
rs1568507	0.012	0.114	0.047	0.885	0.237	

Supplemental Table 4.2: All WNT association results in NHW families

		All			Positive Family History			Negative Family History		
Gene	Marker	PDT	Geno-PDT	APL	PDT	Geno-PDT	APL	PDT	Geno-PDT	APL
Wnt3A	rs708111	0.103	0.257	0.778	0.11	0.114	0.366	0.558	0.203	0.867
Wnt3A	rs708114	0.340	0.555	0.55	0.217	0.315	0.556	0.931	0.919	0.778
Wnt3A	rs3094912	0.485	0.335	0.698	0.778	0.657	0.723	0.455	0.449	0.75
Wnt3A	rs3121310	0.014	0.033	0.065	0.051	0.073	0.499	0.137	0.319	0.08
Wnt3A	rs752107	0.023	0.027	0.003	0.31	0.441	0.125	0.0178	0.027	0.009
Wnt3A	rs1745420	0.024	0.034	0.092	0.301	0.377	0.283	0.033	0.067	0.204
Wnt7A	rs1124480	0.528	0.667	0.681	0.842	0.853	0.957	0.439	0.603	0.539
Wnt7A	rs9840696	0.229	0.473	0.958	0.040	0.097	0.047	0.288	0.584	0.149
Wnt7A	rs6778046	0.127	0.162	0.407	0.03	0.031	0.067	0.612	0.888	0.670
Wnt7A	rs9863149	0.176	0.366	0.784	0.038	0.058	0.205	0.846	0.687	0.608
Wnt7A	rs934450	0.2	0.341	0.889	0.053	0.029	0.168	0.673	0.34	0.384
Wnt7A	rs1433354	0.319	0.535	0.622	0.167	0.254	0.322	0.831	0.827	0.728
Wnt7A	rs6442414	0.733	0.168	0.933	0.371	0.34	0.833	0.598	0.278	0.720
Wnt7A	rs11128663	0.399	0.526	0.886	0.633	0.416	0.897	0.455	0.745	0.793
Wnt7A	rs4685048	0.027	0.085	0.097	0.129	0.351	0.152	0.105	0.192	0.291
Wnt5A	D3S3719	0.343	0.365	n/a	0.343	0.365	n/a	n/a	n/a	n/a
Wnt5A	rs566926	0.005	0.016	0.004	0.043	0.107	0.049	0.0462	0.144	0.02
Wnt5A	D3S2408	0.599	0.286	n/a	0.599	0.286	n/a	n/a	n/a	n/a
Wnt8a	rs4835761	0.399	0.069	0.46	0.561	0.265	0.194	0.067	0.086	0.129
Wnt8a	rs2040862	0.352	0.055	0.061	0.581	0.372	0.085	0.009	0.016	0.001
Wnt8a	rs2306110	0.409	0.188	0.818	0.749	0.679	0.052	0.117	0.113	0.197
Wnt11	rs663746	0.067	0.001	0.449	0.104	0.047	0.278	0.36139	0.01449	0.849
Wnt11	rs1533763	0.267	0.464	0.338	0.858	0.245	0.193	0.05	0.097	0.053
Wnt11	rs1533767	0.114	0.012	0.001	0.744	0.010	0.100	0.006	0.016	0.003
Wnt11	rs689095	0.374	0.127	0.540	0.943	0.217	0.652	0.135	0.219	0.289
Wnt11	rs596339	0.463	0.762	0.152	0.244	0.499	0.236	0.927	0.891	0.311
Wnt11	rs1568507	0.144	0.143	0.213	1.000	0.181	0.462	0.02	0.062	0.047
Wnt3	rs142167	0.108	0.23	0.051	0.294	0.553	0.107	0.190	0.359	0.222
Wnt3	rs7216231	0.662	0.289	0.714	0.162	0.162	0.092	0.394	0.477	0.315
Wnt3	rs199525	0.437	0.731	0.182	0.309	0.557	0.017	0.908	0.985	0.777
Wnt3	rs70602	0.154	0.308	0.041	0.309	0.585	0.009	0.299	0.426	0.365
Wnt3	rs199501	0.102	0.188	0.049	0.149	0.281	0.026	0.431	0.642	0.366
Wnt3	rs199498	0.162	0.239	0.161	0.087	0.139	0.006	1.000	0.941	0.947
Wnt3	rs111769	0.435	0.334	0.993	0.150	0.296	0.251	0.502	0.043	0.422
Wnt3	rs3851781	0.736	0.187	0.604	0.534	0.206	0.722	0.772	0.746	0.791
Wnt3	hCV1139279	0.217	0.366	0.597	0.278	0.335	0.808	0.535	0.801	0.689
Wnt3	rs9890413	0.948	0.698	0.593	0.932	0.216	0.504	0.840	0.005	0.835
Wnt9B	rs2165846	0.697	0.887	0.715	0.381	0.556	0.668	0.538	0.771	0.403
Wnt9B	rs1530364	0.574	0.828	0.644	0.083	0.146	0.298	1.000	0.821	0.856
Wnt9B	rs197915	0.160	0.051	0.573	0.072	0.180	0.003	0.926	0.060	0.310

Supplemental Table 4.3: All WNT association results in Hispanic families

		All			Positive Family History			Negative Family History		
Gene	Marker	PDT	Geno-PDT	APL	PDT	Geno-PDT	APL	PDT	Geno-PDT	APL
Wnt3A	rs708111	0.858	0.698	0.314	0.034	0.135	0.007	0.297	0.426	0.005
Wnt3A	rs708114	0.819	0.925	0.422	0.655	0.513	0.813	0.593	0.701	0.425
Wnt3A	rs3094912	0.343	0.572	0.031	0.796	0.892	0.301	0.162	0.240	0.001
Wnt3A	rs3121310	0.450	0.570	0.473	0.739	0.264	0.867	0.251	0.556	0.349
Wnt3A	rs752107	0.695	0.744	0.082	0.706	0.853	0.100	0.491	0.280	0.261
Wnt3A	rs1745420	0.083	0.157	0.001	0.096	0.264	0.009	0.346	0.376	0.008
Wnt7A	rs1124480	0.106	0.040	0.252	0.706	0.080	0.851	0.103	0.189	0.118
Wnt7A	rs9840696	0.876	0.747	0.336	1.000	0.755	0.810	0.847	0.925	0.330
Wnt7A	rs6778046	0.886	0.026	0.110	0.763	0.467	0.979	0.746	0.044	0.084
Wnt7A	rs9863149	0.858	0.576	0.911	0.796	0.801	0.100	1.000	0.625	0.133
Wnt7A	rs934450	0.866	0.429	0.531	0.439	0.714	0.090	0.655	0.456	0.017
Wnt7A	rs1433354	0.456	0.689	0.214	0.257	0.411	0.651	0.746	0.946	0.227
Wnt7A	rs6442414	0.724	0.386	0.440	0.414	0.641	0.508	1.000	0.450	0.867
Wnt7A	rs11128663	0.194	0.345	0.892	0.706	0.641	0.943	0.209	0.258	0.914
Wnt7A	rs4685048	0.131	0.182	0.337	1.000	1.000	0.800	0.096	0.122	0.330
Wnt5A	D3S3719	0.734	0.526	n/a	0.734	0.526	n/a	n/a	n/a	n/a
Wnt5A	rs566926	0.182	0.432	0.154	0.781	0.185	0.401	0.144	0.129	0.242
Wnt5A	D3S2408	0.459	0.526	n/a	0.459	0.526	n/a	n/a	n/a	n/a
Wnt8a	rs4835761	0.695	0.814	0.251	0.763	0.837	0.507	0.796	0.925	0.360
Wnt8a	rs2040862	0.637	0.783	0.636	0.157	0.157	0.260	1.000	1.000	0.850
Wnt8a	rs2306110	0.564	0.650	0.181	0.763	0.837	0.524	0.617	0.758	0.223
Wnt11	rs663746	1.000	0.031	0.245	0.025	0.104	0.282	0.336	0.080	0.081
Wnt11	rs1533763	0.366	0.608	0.622	0.564	0.564	0.323	0.480	0.683	0.924
Wnt11	rs1533767	0.637	0.701	0.900	0.414	0.414	0.997	1.000	0.585	0.867
Wnt11	rs689095	1.000	1.000	0.851	0.157	0.264	0.469	0.593	0.467	0.915
Wnt11	rs596339	0.134	0.131	0.370	1.000	1.000	0.456	0.083	0.116	0.204
Wnt11	rs1568507	0.617	0.659	0.971	0.317	0.513	0.894	0.439	0.301	0.911
Wnt3	rs142167	0.866	0.953	0.730	0.480	0.367	0.463	0.847	0.642	0.943
Wnt3	rs7216231	0.353	0.492	0.989	0.257	0.135	0.983	0.670	0.210	0.993
Wnt3	rs199525	0.531	0.678	0.997	0.317	0.317	0.194	0.670	0.627	0.943
Wnt3	rs70602	0.617	0.607	0.658	0.317	0.317	0.190	0.796	0.445	0.750
Wnt3	rs199501	0.465	0.696	0.495	0.706	0.641	0.683	0.532	0.546	0.369
Wnt3	rs199498	0.414	0.376	0.531	0.706	0.641	0.688	0.467	0.155	0.410
Wnt3	rs111769	0.847	0.946	0.358	0.317	0.513	0.073	1.000	0.798	0.442
Wnt3	rs3851781	0.414	0.758	0.676	1.000	1.000	0.037	0.394	0.741	0.440
Wnt3	hCV1139279	0.593	0.695	0.810	0.564	0.564	0.309	0.366	0.578	0.492
Wnt3	rs9890413	0.683	0.908	0.980	0.157	0.157	0.145	1.000	0.722	0.851
Wnt9B	rs2165846	0.127	0.204	0.755	0.564	0.564	0.040	0.059	0.155	0.369
Wnt9B	rs1530364	0.480	0.685	0.740	0.414	0.641	0.982	0.695	0.766	0.677
Wnt9B	rs197915	0.034	0.050	0.082	0.317	0.317	0.204	0.020	0.040	0.044

Chapter 5: Identification of association between NSCLP and a novel gene in region 16q24.1: CRISPLD2

Note: The information presented in this chapter was published in 2007, in which I was the primary author: "CRISPLD2: a novel NSCLP candidate gene." *Human Molecular Genetics*. 16 (18): p.2241-2248. The *Human Molecular Genetics* Journal does not require permission to reproduce manuscript content "in full or in part in a thesis or dissertation" (http://www.oxfordjournals.org/access_purchase/ publication_rights.html).

5.0. Introduction

Genome scans, like the one performed for chapter three, identify regions that potentially harbor a disease gene for a given population. Given the etiologic heterogeneity for NSCLP, genome scans on different populations often do not return identical results (refer to Table 1.2, page 11). Additionally, regions identified in these genome scans do not always reveal previously identified and/or biologically plausible disease genes. In this chapter, data is presented for one such region, 16q24.1, and the identification of a novel gene family, CRISPLD, that contributes to the genetic etiology of NSCLP.

5.1. Chromosome 16q24.1

In 2000, Prescott *et al.* performed a genome scan on a dataset comprised of caucasian sib-pairs from the UK and identified eleven regions across the genome that potentially harbor a clefting locus [101]. In a follow-up to that study, our laboratory genotyped the same genetic markers that showed linkage with association in their study in our large multiplex NSCLP nonHispanic white families [106]. Six regions, 2p13, 2q37, 11p12-14, 12q13, 16p13 and 16q24, were identified [106]. Significant linkage was found to STR marker d16S3037 on chromosome 16q24.1 (p=0.00063). Three other NSCLP genome scans and two meta-analyses also suggested that a clefting gene is located near 16q21-24 [97,99,100,103,104,262]. Two potential candidate genes lie within 1 Mb of this marker: interferon regulatory factor 8 (IRF8) and cysteine-rich secretory protein LCCL domain containing 2 (CRISPLD2) [160]. These genes were chosen as potential NSCLP candidate genes because (1) IRF8 is in the same gene family as IRF6, which has been associated with NSCLP, and (2) CRISPLD2 is 795 bp 5' of d16s3037 [46,49,160].

5.2. Materials and Methods

The details of the NSCLP family population, sample collection and DNA extraction, SNP selection criteria, genotyping, analysis (FBAT and HBAT), transcription binding site prediction algorithms, sequencing, silver staining and *in situ* hybridizations are described in Chapter 2.

For this study, 63 multiplex families (56 nonHispanic white and seven Hispanic) and 287 simplex parent-child trios or duos (213 nonHispanic white and 74 Hispanic) were genotyped with 18 SNPs (Table 5.1). Additionally, a total of 246 Columbian NSCLP families (114 multiplex and 132 simplex parent-child trios) were genotyped with 12 CRISPLD2 SNPs (Table 5.1). The discrepancy in number of SNPs genotyped in the different datasets is due to the elimination of SNPs in linkage disequilibrium with other SNPs.

5.3. Results

5.3.1. Microsatellite analysis

In a follow-up study to Prescott *et al.*'s genome wide scan, we genotyped 37 STRs in 47 multiplex NSCLP families [106]. Reanalysis of the original data stratified by ethnicity with FBAT showed an additional microsatellite marker, D16S3037, with significant linkage with association to NSCLP (p=0.00063). D16S3037 maps to chromosome 16q24.1. There are 35 known genes in this region based on UniProt, RefSeq and GenBank mRNA. [203]

5.3.2. Candidate Gene Testing

Interferon regulatory factor 8 (IRF8) was chosen as the first candidate gene because IRF8 belongs to the same family as IRF6, mutations in which are causal for Van der Woude syndrome and IRF6 variants have been shown to play an etiologic role in NSCLP [46,47,49,135,137]. Also, IRF8 is 1 Mb downstream of D16S3037. Seven IRF8 SNPs (2 flanking and 5 intragenic)

Table 5.1: 16q24.1 SNPs genotyped in this study and allele frequencies

Gene	dbSNP	Chr:bp	Location	Alleles†	NHW MAF[a]	Hispanic MAF[b,*,**,***]	Columbian MAF[c]
IRF8	rs7193803	16:84482924	upstream	G/T	0.306	0.4	NT
IRF8	rs305082	16:84494729	intron 2	T/C	0.173	0.167	NT
IRF8	rs305080	16:84499525	intron 2	C/T	0.283	0.308	NT
IRF8	rs2292980	16:84502827	intron 3	A/G	0.286	0.308	NT
IRF8	rs2280378	16:84510496	intron 7	C/T	0.418	0.423	NT
IRF8	rs1568391	16:84513055	exon 9	T/G	0.447	0.8***	NT
IRF8	rs880365	16:84517113	downstream	C/T	0.21	0.467*	NT
CRISPLD2	rs4572384	16:83400728	upstream	T/C	0.396	0.717***	0.503***
CRISPLD2	rs1874014	16:83413267	intron 1	A/C	0.443	0.261***	NT
CRISPLD2	rs8051428	16:83424753	intron 1	C/T	0.355	0.495**	NT
CRISPLD2	rs1546124	16:83429802	exon 2	C/G	0.298	0.322	0.294
CRISPLD2	rs1874015	16:83435283	intron 2	T/C	0.307	0.193*	0.751***
CRISPLD2	rs12051468	16:83437215	exon 3	A/G	0.454	0.313**	0.338
CRISPLD2	rs8061351	16:83440853	exon 4	T/C	0.284	0.32	0.393
CRISPLD2	rs2646129	16:83444995	intron 5	A/G	0.368	0.185***	NT
CRISPLD2	rs2326398	16:83460468	intron 8	A/G	0.344	0.312	0.315
CRISPLD2	rs721005	16:83463849	exon 9	C/G	0.396	0.348	0.358
CRISPLD2	rs774206	16:83477709	intron 13	A/G	0.364	0.283	0.334
CRISPLD2	rs767050	16:83480610	exon 14	G/A	0.445	0.383	0.42
CRISPLD2	rs2646112	16:83495601	intron 14	G/A	0.082	0.129	NT
CRISPLD2	rs2641670	16:83498731	exon 15	G/A	0.21	0.371***	0.266
CRISPLD2	rs4783099	16:83499080	exon 15	C/T	0.364	0.423	0.412
CRISPLD2	rs16974880	16:83500389	exon 15	T/G	0.301	0.371	0.316
CRISPLD2	rs903194	16:83502930	downstream	T/G	0.397	0.291	NT
CRISPLD2	rs2641674	16:83510855	downstream	T/C	0.373	0.303	NT

†Major allele as identified in NHW listed first

[a]Minor allele frequency of nonHispanic white dataset

[b]Corresponding frequency in Hispanic dataset of nonHispanic white minor allele

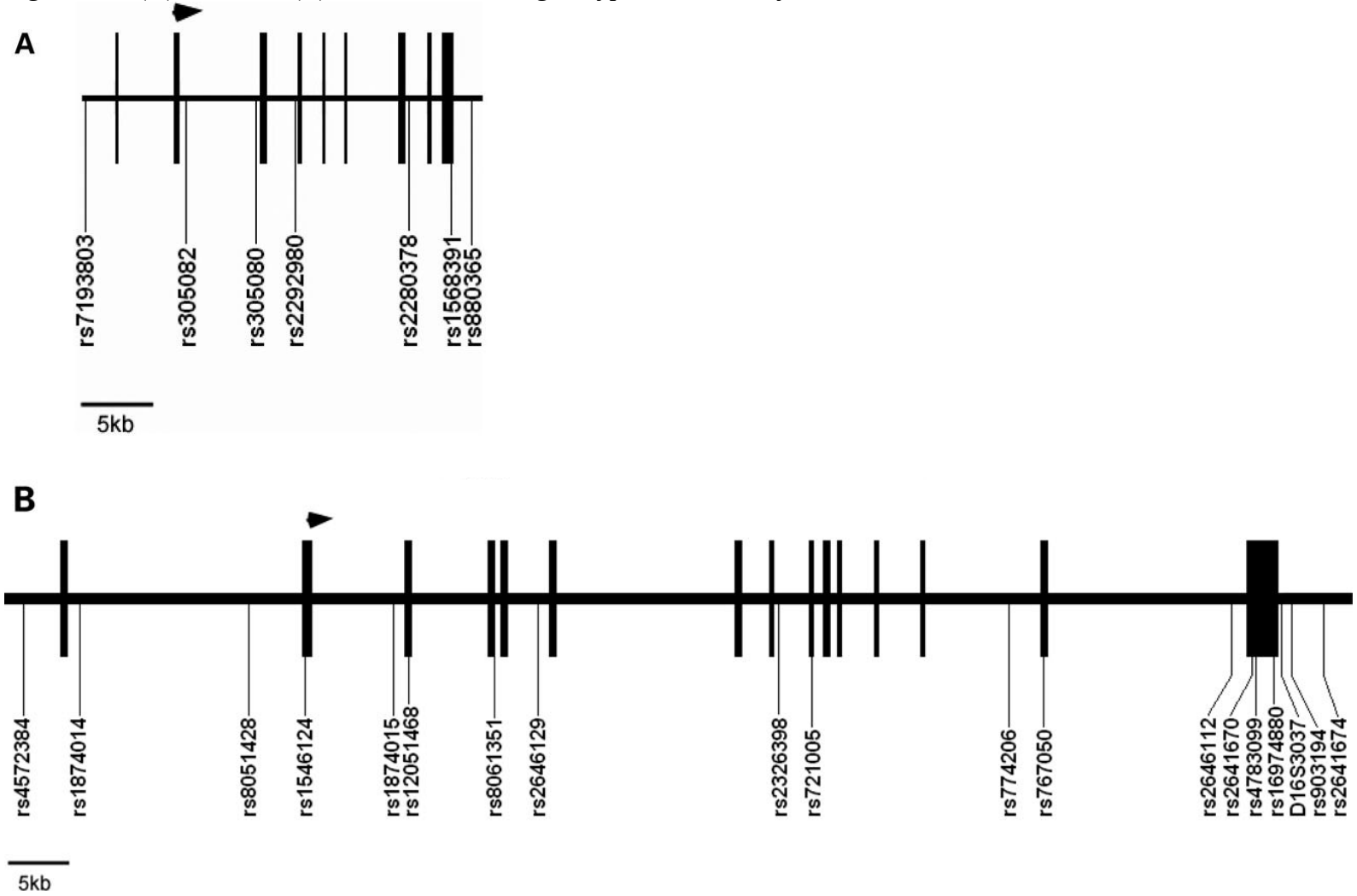
[c]Corresponding frequency in Columbian dataset of nonHispanic white minor allele; p-values represent allele frequency differences between Columbian and Hisp datasets.

*p<0.0125; **p<0.001; **p<0.0001

were genotyped in our Caucasian and Hispanic cohorts [Figure 5.1A]. All SNPs were in Hardy-Weinberg equilibrium (HWE). Two SNPs had significantly different allele frequencies between the Hispanic and Caucasian populations [Table 5.1]. Hence, families were stratified by ethnicity
for all statistical analyses. There was significant linkage disequilibrium (LD) between SNPs within IRF8 (data not shown). No evidence for linkage or association with NSCLP was found by either parametric or nonparametric linkage analysis. Previous interrogation of IRF6 in these populations showed an increased transmission of haplotypes constructed with the major allele rs2013162 (p=0.009) [46]. Using this haplotype and all IRF8 SNPs, IRF8-IRF6 gene-gene interaction was not detected. The Colombian population was not genotyped for IRF8.

Further evaluation of the 16q24.1 region for candidate genes revealed the cysteine-rich secretory protein LCCL domain containing 2 (CRISPLD2) gene, which is 795 bp upstream of D16S3037. Eighteen CRISPLD2 SNPs (4 flanking and 14 intragenic) were genotyped in the Caucasian and Hispanic multiplex families and simplex parent-child trios [Figure 5.1]. All SNPs were in HWE. Seven of the 18 SNPs had significantly different allele frequencies between the Caucasian and Hispanic groups with a Bonferroni correction applied and p-value of 0.0125 used as the criterion for significance [Table 5.1]. Parametric and nonparametric linkage analysis in the multiplex families did not demonstrate linkage in any of the ethnic groups. FBAT analysis of the simplex Caucasian trios did not detect any altered transmission. However, in the Caucasian multiplex families, altered transmission of several SNPs at the CRISPLD2 locus was noted [Table 5.2]. SNP rs1546124 (p=0.0006) yielded the largest p-value; SNPs rs4783099 and rs16974880 also yielded suggestive p-values (p=0.08 and p=0.03, respectively). When correcting for multiplex pedigrees, FBAT p-values are still significant for rs1546124 and suggestive for rs4783099 and rs16974880 (p=0.01, p=0.16, p=0.07, respectively). SNPs rs4783099 and rs16974880 are in strong LD (D'=0.869), so haplotypes comprised of rs1546124 with either rs4783099 or rs16974880 were then tested [Table 5.3]. For the rs1546124-rs4783099 haplotype, an excess transmission of the 1-1 haplotype was detected when correcting for multiplex pedigrees (86 transmitted, 69 expected; p=0.002) with an overall departure from expected observed (p=0.01). There was an overtransmission of the 1-2 haplotype for rs1546124 and rs16974880 (83 observed vs. 66 expected, p=0.001), and an overall departure from expected for all the haplotypes (p=0.01).

Figure 5.1: (A) IRF8 and (B) CRISPLD2 SNPs genotyped in this study*



*Arrow denotes ATG start site
†Taken from [203].

Table 5.2: NHW and Hispanic CRISPLD2 FBAT results†*

	NHW		Hispanic	
dbSNP	P_{raw}	$P_{corrected}$	P_{raw}	$P_{corrected}$
rs4572384	0.832	0.825	0.532	0.549
rs1874014	0.356	0.368	0.853	0.853
rs8051428	0.673	0.686	0.093	0.093
rs1546124	**0.001**	**0.012**	0.532	0.532
rs1874015	0.455	0.519	0.467	0.532
rs12051468	0.960	0.967	0.505	0.450
rs8061351	0.695	0.761	**0.028**	**0.024**
rs2646129	0.646	0.700	0.225	0.166
rs2326398	0.109	0.189	**0.048**	0.056
rs721005	0.159	0.249	0.086	0.086
rs774206	0.488	0.558	0.144	0.117
rs767050	0.774	0.777	0.411	0.423
rs2646112	0.178	0.262	0.808	0.808
rs2641670	0.215	0.120	0.465	0.450
rs4783099	0.081	0.169	0.206	0.217
rs16974880	**0.030**	0.068	0.160	0.189
rs903194	0.291	0.355	0.819	0.808
rs2641674	0.937	0.937	0.371	0.317

*Modified from [203]

†$p<0.05$ in bold

Table 5.3: CRISPLD2 Caucasian multiplex haplotypes†*

rs1546124	rs4783099	rs16974880	Observed	Expected	p-value
Allele 1	Allele 1		86	69	**0.002**
Allele 1	Allele 2		36	40	0.38
Allele 2	Allele 1		24	34	0.02
Allele 2	Allele 2		8	2	0.27
Allele 1		Allele 1	31	37	0.131
Allele 1		Allele 2	83	66	**0.001**
Allele 2		Allele 1	n/a	n/a	n/a
Allele 2		Allele 2	22	31	0.05

†Taken from [203].

*$p\leq0.001$ in bold

As the Hispanic sample is small, FBAT analysis was performed on the combined simplex and multiplex families. A slightly altered transmission of SNPs rs8061351 (p=0.03) and rs2326398 (p=0.05) vwas noted [Table 5.2]. When correcting for multiplex pedigrees, FBAT p-values remain significant for rs8061351 (p=0.02) and become suggestive for rs2326398 (p=0.06). There was no evidence for altered transmission of a haplotype consisting of either these two SNPs or the SNPs identified in the Caucasian multiplex sample.

A third NSCLP population consisting of Colombian multiplex families and simplex parent-child trios was tested. Twelve CRISPLD2 SNPs that were informative in the Caucasian and Hispanic populations were run on the Colombian NSCLP population [Table 5.1]. In a comparison between the Hispanic and Colombian families, two SNPs (rs4572384 and rs1874015) had significantly different allele frequencies [Table 5.1]. Interestingly, the Colombian sample generally had frequencies between the Caucasian and Hispanic samples. Parametric and nonparametric linkage analysis did not show any evidence for linkage with NSCLP nor was there evidence for altered transmission in the Colombian multiplex and simplex families.

There was significant linkage disequilibrium between SNPs genotyped in each population [Supplementary Table 5.1]. Both Caucasian [Supplementary Figure 5.1] and Hispanic [Supplementary Figure 5.2] populations contain an LD block encompassing the region spanning rs2641670 to rs16974880. While they both also exhibit strong LD in the region defined by rs2326398 to rs767050, this region is divided into two LD blocks in the Hispanics. In addition, the Hispanics have an additional LD block consisting of rs1546124 and rs1874015. A formal comparison of the LD between the Hispanics and the Caucasians detected significant differences based on D' and r ($p<0.0001$ for both). Because the Colombian sample was not genotyped for all of the SNPs, a statistical comparison between the LD patterns of the Colombians with either the Hispanic or the Caucasian populations was not performed. However, a visual inspection of the Colombian LD blocks found by Haploview [Supplementary Figure 5.3] reveals a pattern similar to that of the Hispanics. [203]

5.3.3. Expression Study

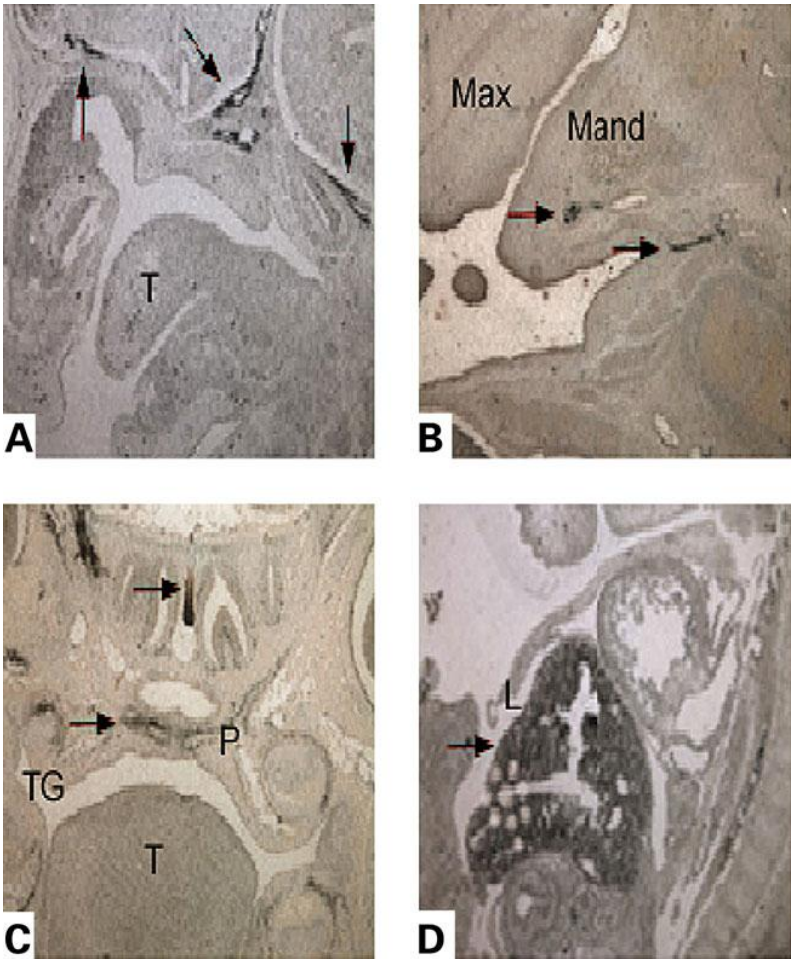
CRISPLD2 *in situ* hybridizations were performed on E12.5-E17.5 mouse sagittal and coronal sections to determine the CRISPLD2 expression in the developing embryo. CRISPLD2 was expressed in the developing oro- and nasopharynx at E13.5

> [Figure 5.2A], the mandible at E14.5 [Figure 5.2B] and the cartilage primordia of the nasal bones, palate and tooth germs at E17.5 [Figure 5.2C]. CRISPLD2 is expressed in the liver at E14.5 and all time points examined [Figure 5.2D]. No other organ systems showed significant levels of expression. [203]

5.3.4. Sequencing

Fifteen exons and approximately 100bp of the intron/exon borders were sequenced in the CRISPLD2 gene in 25 individuals with CRISPLD2 susceptible haplotype. Table 5.4 and Figure 5.3 summarizes the 20 sequence variants identified. Nine variants were identified in the 5' regions of the exons, including a previously unidentified polymorphism nine basepairs 5' of exon 12 (8:83468536; T>T/C). One variant was identified in a 3' region of an exon (exon 13); this variant has not been previously described (8:83471733; A>A/C). Ten exonic single basepair variants were found, all previously identified, including three coding changes. A stretch of 16-23 polythymines was identified in the untranslated portion of exon 15. The sequencing results from this region suggested that the length of this polythymine stretch was polymorphic. Silver stained PCR products from this region for the 25 probands identified three different sizes of polythymine stretches (Figure 5.4). To determine the normal frequency of these three alleles, silver staining was performed on the PCR products of this region on a total of 300 controls, including 100 unaffected parents of isolated NSCLP patients (CLP Parent Control), 100 unaffected parents of sporadic clubfoot cases (CF Parent Control) and 100 CEPH samples (CEPH Control). The allele frequencies of each of the groups are listed in Table 5.5. The allele frequencies significantly differed between the proband group versus each of the control groups (χ^2 p=0.005); however, when comparing the frequencies of alleles in the cases versus the controls as a collective group, no significance was detected (χ^2 p=0.32).

Figure 5.2: CRISPLD2 expression in the mouse at (A) E13.5, (B) E14.5, (C) E17.5 and (D) E13.5 *†



*Arrows denote expression. T= tongue; Max= maxilla; Mand= mandible; P= palate; TG= tooth germ; L= liver.

†Taken from [203].

Table 5.4: SNP variants identified during sequencing.

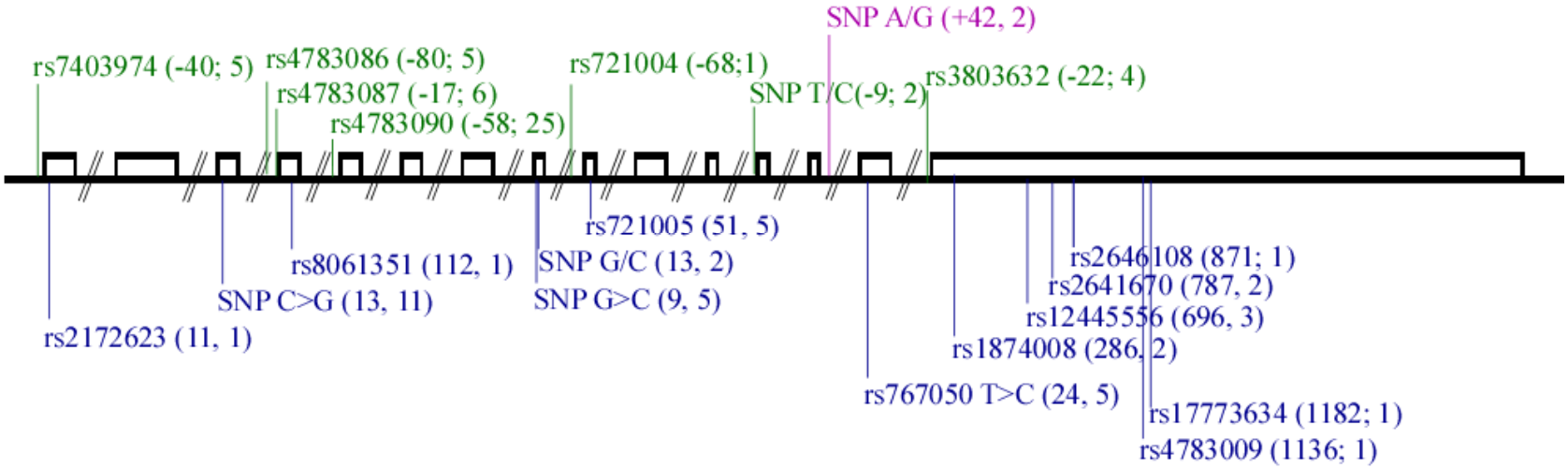
dbSNP**	bp	Location	Alleles†	Source*	Previous MAF	Sequencing MAF
rs7403974	83411052	5' exon 1	G/A	Seq	n/a	0.2
rs2172623	83411102	exon 1	C/T	Seq	n/a	0.042
rs4783086	83440412	5' exon 4	C/T	Seq	n/a	0.438
rs4783087	83440475	5' exon 4	T/C	Seq	n/a	0.458
rs8061351	83440603	exon 4	T/C	TaqMan	0.284	0.289
rs4783090	83441617	5' exon 5	G/A	NCBI	0.223	1
rs4782674	83445796	5' exon 6	C/T	NCBI	0.242	0.125
rs4782675	83445811	5' exon 6	C/T	NCBI	0.242	0.125
rs721004	83463481	5' exon 9	G/C	NCBI	0.082	0.083
rs721005	83463599	exon 9	C/G	TaqMan	0.396	0.396
SNP T>T/C	83468536	5' exon 12	T/C	Seq	n/a	0.042
SNP A>G/A	83471733	3' exon 13	A/G	Seq	n/a	0.04
rs767050	83480360	exon 14	A/G	TaqMan	0.445	0.479
rs3803632	83497673	5' exon 15	G/C; C/A	NCBI	0.033	0.091
rs1874008	83497980	exon 15	C/T	NCBI	0.227	0.292
rs12445556	83498390	exon 15	G/A	NCBI	0.008	0.063
rs2641670	83498481	exon 15	G/A	TaqMan	0.210	0.260
rs2646108	83498565	exon 15	G/A	NCBI	0.188	0.140
rs4783099	83498830	exon 15	C/A	TaqMan	0.364	0.200
rs17773634	83498876	exon 15	A/C	NCBI	0.127	0.180

†Major allele listed first

*The source for the Previous MAF is listed here. Seq = sequence data, TaqMan = SNP previously genotyped in our laboratory, NCBI = National Center for Biotechnology Information PubMed resource

**SNP = novel SNP

Figure 5.3: Schematic of variants identified during sequencing*



*Green variants are 5' of respective exon, blue variants are within the respective exon, purple variants are 3' of the exon. Numbers after the variant indicate the number of basepairs within the exon and the number of sequenced individuals that have the variant.

Figure 5.4: Alleles identified during silver staining.

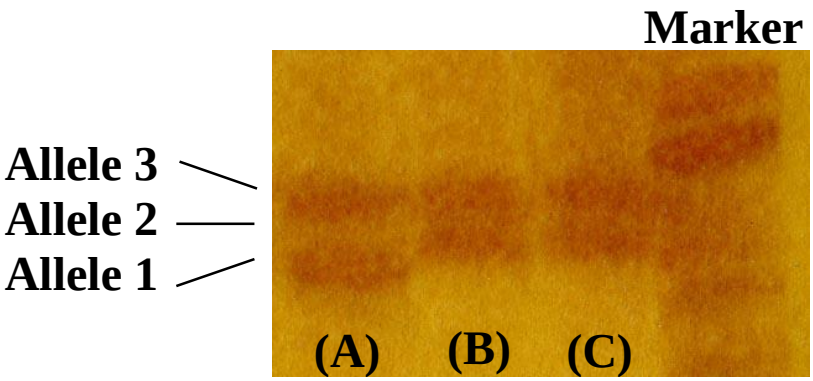


Table 5.5: Polythymine tract allele frequencies.

Allele	CLP Proband Families	CLP Parent Control	CF Parent Control	CEPH Control
1	55%	61%	67%	59.8%
2	1.7%	7%	1.5%	1.8%
3	43%	32%	31%	38%

5.4. Discussion

In this study, we evaluated the chromosome 16q24.1 region for a NSCLP genetic locus. This region was first identified by Prescott *et* al. in a genome scan of Caucasian NSCLP sib pairs, and subsequently in four other genome scans of different NSCLP populations [97,99,101,103,104]. Analysis of STR D16S3037 in our dataset provided evidence for an association with NSCLP (p=0.00063). Two candidate genes, IRF8 and CRISPLD2, were found in close proximity to this STR. IRF8 was initially evaluated because it is 1 Mb downstream of D16S3037 and because it belongs to the same gene family as IRF6. Mutations in IRF6 cause Van der Woude syndrome (VWS [OMIM:119300]), which is characterized by lower lip pits, CLP, CP or hypodontia [135]. In addition, genetic variation in IRF6 has recently been shown to play an etiological role in the development of NSCLP [46-49]. No association was found for IRF8 and no interaction with IRF6 was detected in our dataset.

CRISPLD2 is the closest gene to D16S3037, mapping 795 bp upstream. While the function of CRISPLD2 is unknown, it contains a LCCL domain, which is common to other known genes [i.e., COCH (Coagulation factor C homolog; cochlin), Akhirin and CLCP1 (CUB, LCCL-homology coagulation factor V/VIII homology domains protein)] [263-266]. The function of the LCCL domain is speculated to be either structural, immunologic, or involved in cell motility [264-266]. Mutations in the LCCL domain of COCH have been identified in autosomal dominant nonsyndromic sensorineural deafness disorder (DFNA9 [OMIM:601369]) [267]. Akhirin, which also contains a LCCL domain in the N-terminus, is postulated to play a role in chicken retinal development [263]. Recent research suggests that CLCP1 has a role in cellular motility and is regulated by ubiquitination [265]. Interestingly, a member of the ubiquitin family of genes, SUMO1 (small ubiquitin-like modifier 1), has been shown to post-translationally modify genes involved in palatal morphogenesis and haploinsufficiency of SUMO1 has recently been linked to orofacial clefting [268].

FBAT statistical analyses, which were performed without correcting for multiplex pedigrees, suggested an association with NSCLP in both our Caucasian and Hispanic populations. After correcting for multiplex pedigrees, CRISPLD2 was only significantly associated with our Caucasian cohort [Table 5.2]. In the Caucasian population, SNP rs1546124, which is in exon 2, 51 bp upstream of the ATG start codon, showed significantly altered transmission (p=0.01). This sequence change in the putative promoter region could disrupt different regulatory elements, such as, a RNA polymerase binding or transcription factor

activator/inhibitor binding; either could affect CRISPLD2 protein expression which would affect the developmental process [269,270]. To determine if this is theoretically the case, two transcription binding site prediction programs, PATCH and AliBaba2, were used [208,209]. Both identified a Sp1 site at rs1546124, which has either a C or G at this location in the DNA sequence. However, PATCH predicts a Sp1 site when G is present but not with C, whereas AliBaba2 predicts two Sp1 sites for the G allele and 1 for the C allele. Thus it is plausible that rs1546124 has an effect on CRISPLD2 expression. Future functional studies are planned. SNPs rs4783099 and rs16974880 showed suggestive p-values (p=0.08 and p=0.03, respectively) and are in the 3' UTR region of CRISPLD2 and thus were not submitted for this type of analysis.

In the Hispanic simplex population, rs8061351 and rs2326398, in exon 4 and intron 8, gave evidence of association to NSCLP (p=0.02 and p=0.06, respectively; [Table 5.2]). The exon 4 SNP is a synonymous change and should not affect the CRISPLD2 protein. However, synonymous SNPs in the medium-chain acyl-CoA dehydrogenase (MCAD) and survival of motor neuron (SMN) genes that do not code for an amino acid change in the protein have been shown to interact with regulatory elements and alter gene function [271]. Also, synonymous changes have been shown to change amino acid translation time, resulting in altered protein structure and function [176]. These studies suggest that rs8061351 may play a functional role and needs to be further investigated. It is unknown whether rs2326398, which is not a splice site or in a coding region, might cause a functional change in the protein [180,181]. Intronic SNPs have been shown to be associated with other complex diseases, such as IRF6 with NSCLP, RET (RET proto-oncogene) with Hirschsprung disease and CFH (complement factor H) with age-related macular degeneration [46,47,49,183-185]. Together, this suggests that common genetic variation in noncoding regions may be important and should not be overlooked in complex human diseases.

FBAT analysis of our Caucasian multiplex cohort demonstrated an overtransmission of haplotypes consisting of rs1546124 and either rs4783099 or rs16974880, the latter of which are in strong LD (D'=0.859, p<0.00000) [Table 5.3]. This suggests that these SNPs may be disease causing variants or mutations exist that are in linkage disequilibrium with these overtransmitted haplotypes. Sequencing of the CRISPLD2 gene in affected probands who have received one of the associated haplotypes is being performed.

The CRISPLD2 SNPs were tested in a secondary population consisting of multiplex families and simplex parent-child trios from Colombia. The SNP allele frequencies were significantly

different from the original Hispanic population [Table 5.1]. This is not unexpected, as there is likely less admixture in the Colombian-Hispanic population compared to the Texas-Hispanic population [272]. No association was found in the Colombian population with CRISPLD2 and NSCLP, in contrast to the Texas-Hispanic population. This finding supports the theory that NSCLP is an etiologically heterogeneous disease and that genetic variation in different genes underlies NSCLP in different populations [42].

To evaluate whether CRISPLD2 plays a role in craniofacial development, *in situ* hybridization was performed at various stages of development. Mouse embryos from E12.5 to E17.5, which are the critical stages of palatal development, showed that CRISPLD2 is expressed in the mandible, cartilaginous primordia of the developing nose, palate, oro- and nasopharynx and liver [Figure 5.2]. Thus, CRISPLD2 is expressed during facial development. [203]

To identify variants in CRISPLD2 that may segregate with NSCLP, 25 probands with overtransmitted susceptibility halplotypes were sequenced. 20 single basepair variants were identified, including ten exonic changes and ten intronic changes. Of the exonic changes, only three were in the coding region. Two of these three (rs8061351 and rs767050) are nonsynonymous changes and are not expected to alter the protein. SNP rs721005 is a synonymous change, substituting serine for a threonine. PolyPhen (http://www.bork.embl-heidelberg.de/ PolyPhen) and SIFT (http://blocks.fhcrc.org.sift/SIFT.html) predict this amino acid substitution to be benign and tolerant, respectively. Additionally, these three coding SNPs were previously genotyped in this sample set and were not associated to NSCLP [203].

One variant, rs7403974, was identified in the intron sequence 40 bp 5' of exon 1 (Figure 5.3). AliBaba2 and PATCH transcription binding site programs [208,209] both predict that a G>A nucleotide change results in a loss of a Sp1 binding site, potentially altering binding of regulatory elements to the CRISPLD2 gene. This is similar to the transcription binding site change associated with SNP rs1546124, as previously discussed [203], further suggesting that transcriptional control of CRISPLD2 by Sp1 might play an etiologic role in NSCLP.

A 358 bp gap of noncoding sequence in exon 15 was difficult to sequence. The chromatogram of this region suggested a stretch of 16-23 thymines. Three alleles were identified in this region in our 25 probands and 300 controls (Figure 5.4). The allele frequency of Allele 2, the rarest allele, was greater in the CLP parent control group versus the remaining groups, and is likely responsible for the overall χ^2 significance level p=0.005. However, when combining the three control groups into one, significance was not detected, suggesting that this variant does not contribute to NSCLP etiology.

> These results demonstrate that variation in the CRISPLD2 may contribute to the NSCLP phenotype and that CRISPLD2 is expressed in the craniofacial region during critical time points of palatal fusion. Variation in CRISPLD2 could affect CRISPLD2 protein levels or could affect binding sites of other transcription factors that regulate CRISPLD2 expression. This could lead to a perturbation of normal development and predisposition to orofacial clefting. These results suggest that CRISPLD2 is a novel NSCLP candidate gene and additional studies are [warranted] to determine the role that this gene has in orofacial clefting etiology. Understanding the role of CRISPLD2 will provide additional information needed to understand the complex development of lip and palate and will help further delineate the genetic factors contributing to NSCLP. [203]

Supplemental Table 5.1: Linkage disequlibrium data on dbSNPs genotyped. D' is above the diagonal, p-value is below the diagonal.

a) Caucasian population

b) Hispanic population

c) Columbian population

(A) Caucasian population LD

Caucasian	rs4572384	rs1874014	rs8051428	rs1546124	rs1874015	rs12051468	rs8061351	rs2646129	rs2326398	rs721005	rs774206	rs767050	rs2646112	rs2641670	rs4783099	rs16974880	rs903194	rs2641674
rs4572384		0.00182	0.29459	0.125	0.07151	0.40109	0.72008	0.87537	0.14341	0.81397	0.78685	0.71278	0.89542	0.46258	0.56729	0.35162	0.24071	0.74533
rs1874014	0.320		0.00000	0.31186	0.21094	0.08330	0.01434	0.38586	0.48807	0.44843	0.97689	0.47329	0.10315	0.33295	0.85089	0.67191	0.18713	0.11605
rs8051428	0.119	**0.401**		0.00031	0.72258	0.00001	0.02266	0.50033	0.03735	0.03778	0.00809	0.18745	0.07456	0.37305	0.65078	0.00656	0.13986	0.05724
rs1546124	0.109	0.127	**0.445**		0.00000	0.00000	0.05967	0.58329	0.21975	0.50437	0.74948	0.98977	0.84637	0.65200	0.31983	0.16206	0.40598	0.47144
rs1874015	0.115	0.147	0.043	**0.686**		0.00000	0.00003	0.00075	0.01884	0.26118	0.18582	0.32908	0.15462	0.57861	0.00271	0.02576	0.40594	0.04532
rs12051468	0.075	0.145	**0.452**	**0.790**	**0.693**		0.00146	0.41171	0.00106	0.00178	0.00085	0.00158	0.59907	0.16403	0.00052	0.00058	0.70742	0.34859
rs8061351	0.026	0.285	0.175	0.118	**0.292**	0.340		0.00000	0.00101	0.00003	0.00007	0.00000	0.18542	0.02845	0.00181	0.00005	0.78724	0.84022
rs2646129	0.017	0.089	0.047	0.042	**0.237**	0.079	**0.772**		0.16666	0.10510	0.38397	0.00002	0.71925	0.67054	0.33849	0.20835	0.09488	0.03767
rs2326398	0.101	0.064	0.248	0.090	0.305	0.252	0.405	0.168		0.00000	0.00000	0.00000	0.00000	0.05871	0.00000	0.00000	0.03423	0.94668
rs721005	0.016	0.060	0.221	0.050	0.129	0.212	**0.470**	0.177	**0.948**		0.00000	0.00000	0.00000	0.00217	0.00000	0.00000	0.01609	0.55548
rs774206	0.018	0.002	0.310	0.022	0.163	**0.243**	**0.481**	0.103	**0.841**	**0.970**		0.00000	0.00000	0.00288	0.00000	0.00000	0.00085	0.69520
rs767050	0.038	0.069	0.113	0.001	0.085	0.242	**0.493**	**0.366**	**0.813**	**0.861**	**0.862**		0.00003	0.00000	0.00002	0.00049	0.01887	0.00001
rs2646112	0.018	0.430	0.254	0.026	0.196	0.090	0.186	0.061	**0.838**	**1.000**	**1.000**	**0.892**		0.00226	0.02543	0.04139	0.28258	0.83028
rs2641670	0.067	0.142	0.154	0.034	0.044	0.181	0.187	0.043	0.330	0.474	0.497	**0.553**	0.355		0.00001	0.00292	0.08706	0.00001
rs4783099	0.075	0.020	0.063	0.139	0.374	0.278	0.435	0.129	**0.405**	**0.483**	**0.502**	**0.444**	0.333	**0.851**		0.00000	0.00000	0.00000
rs16974880	0.081	0.046	0.420	0.230	0.323	**0.336**	**0.643**	0.189	**0.410**	**0.462**	**0.480**	**0.437**	0.607	0.622	**0.869**		0.00000	0.00000
rs903194	0.083	0.119	0.151	0.064	0.061	0.030	0.031	0.165	0.222	0.227	**0.333**	0.206	0.283	0.257	**0.723**	**0.879**		0.00000
rs2641674	0.023	0.140	0.153	0.057	0.149	0.083	0.027	0.239	0.004	0.036	0.025	**0.404**	0.034	**0.625**	**0.654**	**0.662**	**0.458**	

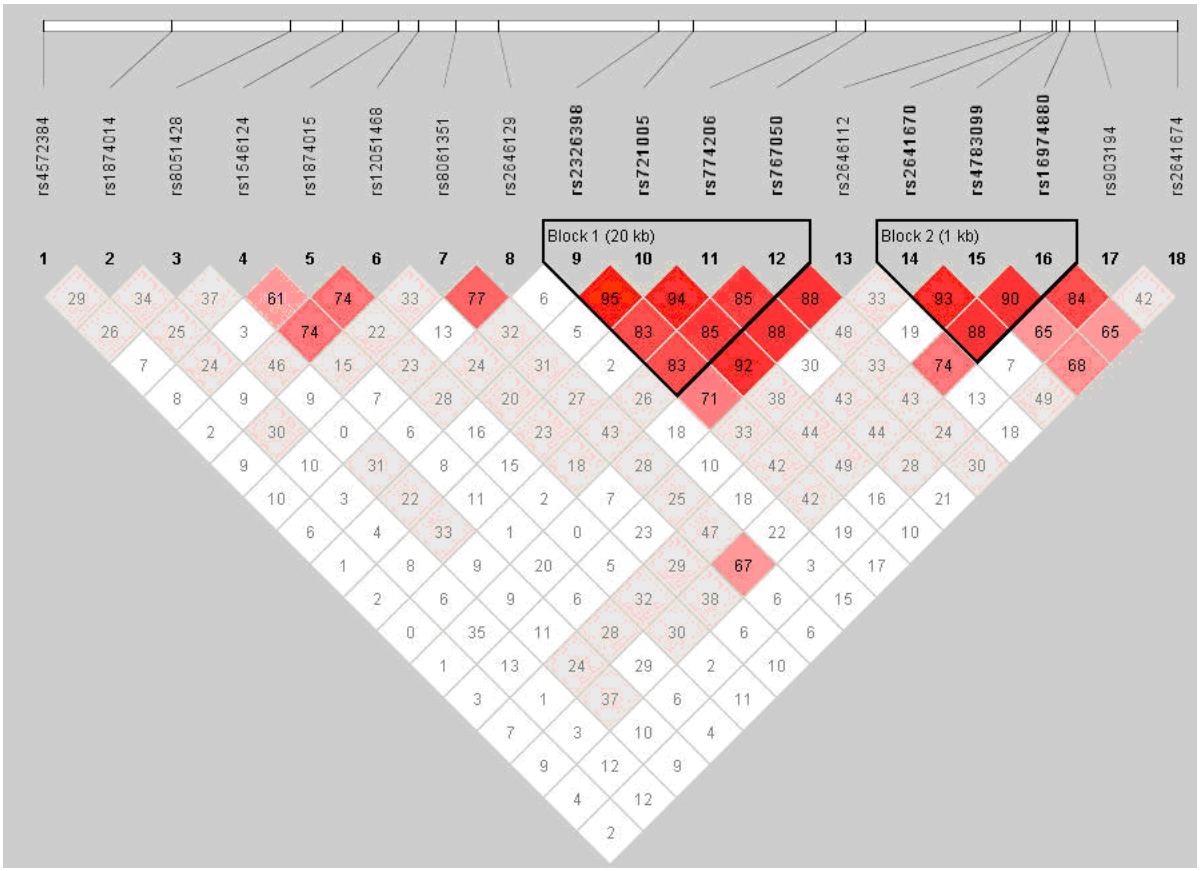
(B) Hispanic population LD

Hispanic	rs4572384	rs1874014	rs8051428	rs1546124	rs1874015	rs12051468	rs8061351	rs2646129	rs2326398	rs721005	rs774206	rs767050	rs2646112	rs2641670	rs4783099	rs16974880	rs903194	rs2641674
rs4572384		0.07314	0.24108	0.82051	0.15488	**0.00085**	0.13827	0.00231	0.06211	0.06974	0.06698	0.80926	0.45105	0.14652	0.77986	0.42520	0.00056	0.02542
rs1874014	0.247		0.44465	0.11291	0.59856	0.02904	0.87973	0.91394	0.02424	0.01503	0.03546	0.50023	0.08935	0.62938	0.42433	0.84594	0.87252	0.32876
rs8051428	0.247	0.181		**0.00018**	0.63155	0.58464	0.49559	0.48269	0.14213	0.04130	0.05408	0.41317	0.77715	0.95702	0.76440	0.11223	0.33800	0.35337
rs1546124	0.075	0.534	**0.659**		**0.00000**	0.35842	0.75856	0.11036	0.45408	0.23908	0.57703	0.01727	0.95748	0.78614	0.33405	0.36395	0.00148	0.02404
rs1874015	0.227	0.232	0.115	**1.000**		0.15532	0.28318	**0.00004**	0.94473	0.57125	0.23972	0.18893	0.31958	0.28290	0.99168	0.63995	0.00123	0.00212
rs12051468	**0.420**	0.329	0.104	0.115	0.222		0.03785	0.00171	0.02047	0.04814	0.11230	0.43475	0.65843	0.05399	0.28338	0.35543	0.00556	**0.00002**
rs8061351	0.230	0.046	0.110	0.041	0.408	0.467		0.48408	0.43031	0.20390	0.25477	0.39037	0.00223	0.49675	0.52616	0.87227	0.24865	0.89230
rs2646129	0.427	0.042	0.152	0.310	**0.455**	0.486	0.126		0.88233	0.78220	0.49193	0.20893	0.17213	0.54881	0.80530	0.55697	**0.00062**	0.03617
rs2326398	0.230	0.318	0.271	0.090	0.022	0.267	0.201	0.046		**0.00000**	**0.00000**	**0.00001**	0.06208	0.31207	0.60030	0.16343	0.05133	0.02054
rs721005	0.250	0.383	0.374	0.148	0.097	0.238	0.329	0.093	**0.962**		**0.00000**	**0.00000**	0.01848	0.18329	0.33855	0.21375	0.11794	0.04008
rs774206	0.240	0.360	0.405	0.083	0.193	0.184	0.299	0.115	**0.869**	**0.956**		**0.00000**	0.02538	0.34739	0.29749	0.22457	0.04806	0.08058
rs767050	0.056	0.190	0.123	0.445	0.329	0.153	0.194	0.261	**0.835**	**0.920**	**1.000**		**0.00086**	0.00201	**0.00064**	**0.00001**	0.24452	0.06932
rs2646112	0.160	0.256	0.097	0.024	0.857	0.081	0.611	0.997	0.353	0.499	0.447	**1.000**		0.33048	0.48766	0.05241	**0.00038**	0.00770
rs2641670	0.398	0.139	0.008	0.056	0.282	0.359	0.143	0.171	0.219	0.300	0.243	0.434	0.263		**0.00000**	**0.00014**	0.69810	0.40246
rs4783099	0.076	0.159	0.045	0.229	0.003	0.228	0.082	0.067	0.082	0.157	0.203	**0.651**	0.295	**0.942**		**0.00000**	0.01174	0.00707
rs16974880	0.221	0.058	0.261	0.113	0.075	0.215	0.036	0.169	0.193	0.168	0.192	**0.843**	0.709	**0.773**	**0.885**		0.08215	0.04195
rs903194	0.482	0.056	0.180	0.472	0.496	0.339	0.169	**0.517**	0.244	0.208	0.264	0.256	**0.648**	0.095	0.628	0.508		**0.00000**
rs2641674	0.294	0.153	0.197	0.295	0.450	**0.498**	0.018	0.277	0.254	0.243	0.213	0.365	0.481	0.180	0.586	0.499	**0.636**	

(C) Columbian population LD

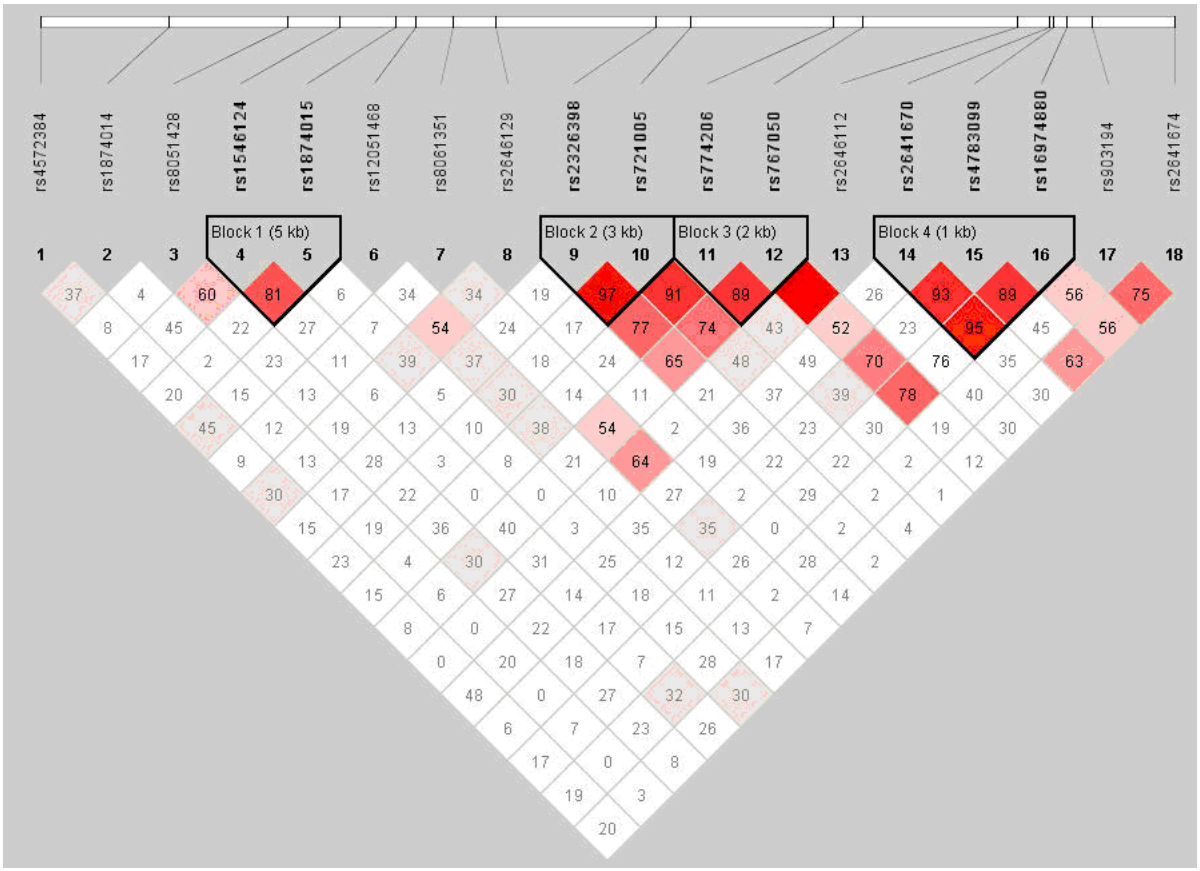
Colombian	rs4572384	rs1874014	rs8051428	rs1546124	rs1874015	rs12051468	rs8061351	rs2646129	rs2326398	rs721005	rs774206	rs767050	rs2646112	rs2641670	rs4783099	rs16974880	rs903194	rs2641674
rs4572384				0.83433	0.45756	0.82462	0.74959		0.02033	0.13929	0.46181	0.00419		0.89265	0.00060	0.00339		
rs1874014																		
rs8051428																		
rs1546124	0.022				**0.00000**	**0.00000**	0.02976		0.34113	0.43297	0.42755	0.34374		0.29150	0.16889	0.35768		
rs1874015	0.078			**0.678**		**0.00000**	0.24477		0.54369	0.89062	0.73042	0.73606		0.80120	0.01737	0.00614		
rs12051468	0.022			**0.702**	**0.782**		0.20808		0.00184	**0.00078**	0.03456	0.16341		0.02560	0.37741	0.05023		
rs8061351	0.032			0.177	0.107	0.208			**0.00082**	**0.00042**	0.01201	0.02534		0.49423	0.25406	0.71040		
rs2646129																		
rs2326398	0.263			0.149	0.107	0.247	**0.472**			**0.00000**	**0.00000**	**0.00000**		0.14360	0.01179	0.00027		
rs721005	0.160			0.114	0.021	**0.267**	**0.466**		**0.922**		**0.00000**	**0.00000**		0.09332	0.00335	**0.00038**		
rs774206	0.082			0.115	0.056	0.161	0.367		**0.748**	**0.863**		**0.00000**		0.23179	0.00857	**0.00077**		
rs767050	0.239			0.110	0.044	0.172	0.201		**0.775**	**0.836**	**0.879**			**0.00000**	**0.00000**	**0.00000**		
rs2646112																		
rs2641670	0.017			0.186	0.051	0.374	0.074		0.267	0.301	0.201	**0.559**			**0.00000**	**0.00000**		
rs4783099	0.313			0.171	0.327	0.082	0.149		0.237	0.259	0.240	**0.612**		**1.000**		**0.00000**		
rs16974880	0.367			0.150	0.469	0.150	0.039		0.290	**0.282**	**0.275**	**0.677**		**0.941**	**0.877**			
rs903194																		
rs2641674																		

Supplemental Figure 5.1: LD plot for CRISPLD2 SNPs in nonHispanic white population.†



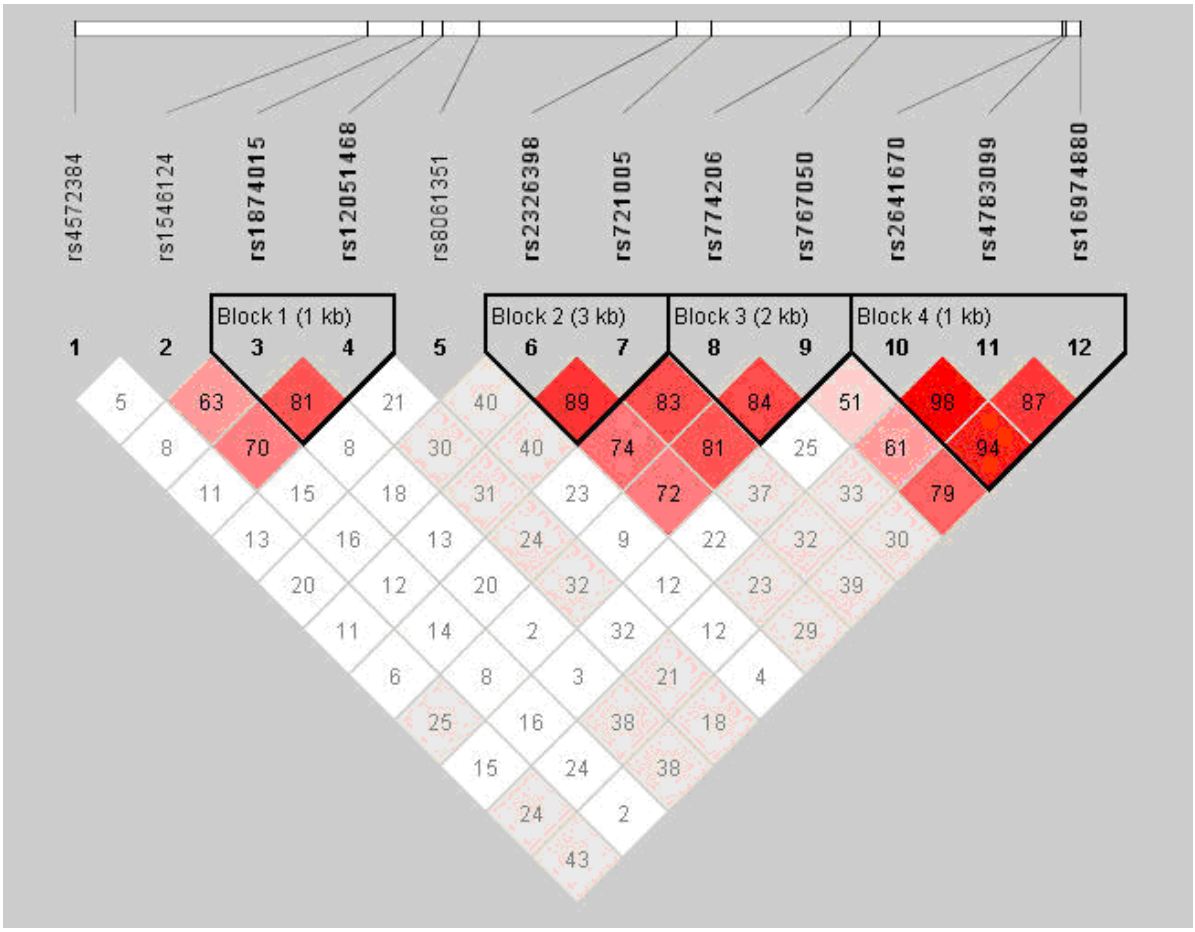
†Taken from [203].

Supplemental Figure 5.2: LD plot for CRISPLD2 SNPs in Hispanic population.†



†Taken from [203]

Supplemental Figure 5.3: LD plot for CRISPLD2 SNPs in Columbian population.†



†Taken from [203]

Chapter Six: Using previous studies to identify new candidate genes: Identification of association between NSCLP and CRISPLD1

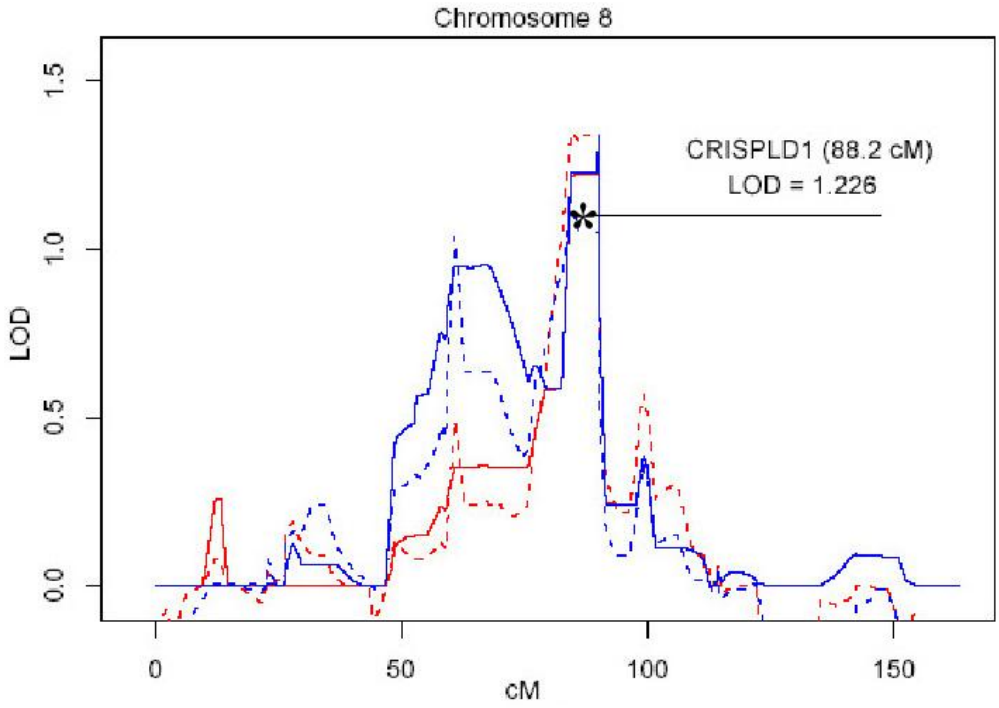
6.0. Introduction

In chapter three, the WNT gene family, which plays an important role in craniofacial development, was interrogated and we identified association between five WNT genes and NSCLP [118]. In chapter five, after identifying significance with STR marker d16s3037 which lies near the IRF6 gene family member IRF8, we tested for association between IRF8 and NSCLP [203]. IRF8 was not associated with NSCLP; however, we did find that variation in the CRISPLD2 gene, located adjacent to this STR, was associated with NSCLP and is expressed in the developing murine craniofacies [203]. CRISPLD2 is a novel gene with no known function. There is a second CRISPLD gene, CRISPLD1, that shares 70% homology at the nucleotide level and 58% at the peptide level with CRISPLD2 [160]. Recent findings from a genome scan performed in our laboratory suggest that chromosome region 8q13.2-21.13, which contains the CRISPLD1 gene, may be linked to NSCLP.

6.1. Chromosome 8q13.2-21.13

Results of our recent genome scan from nine multiplex nonHispanic white NSCLP families identified 11 chromosomal regions with LOD score ≥ 1.5 [105] (see Chapter 3). Regions with LOD score between 1.0 and 1.5 were noted but not included in the original manuscript. Of particular interest was the additional finding that the 8q13.2-21.13 chromosomal region, with a LOD score of 1.23, contained the CRISPLD1 gene (Figure 6.1). This region has not been previously detected to be associated with NSCLP using genome scans. CRISPLD1 is a member of the same family as CRISPLD2, which is a novel gene that we have recently shown to be associated with NSCLP and expressed in developing

Figure 6.1: LOD score plot shown by cMs for the NHW families on chromosome 8.



Legend: The CRISPLD1 gene (basepairs 76059309-76108097) fall between dbSNPs rs2016354 (88.14cM), rs10701 (88.20cM) and rs1464092 (88.57), all with LOD=1.226.

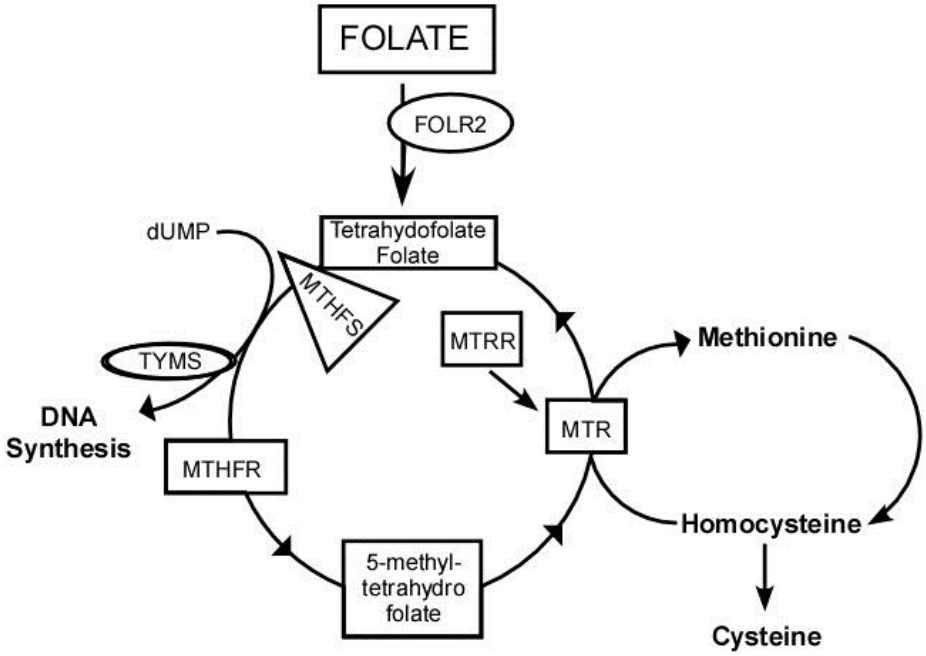
murine craniofacies [203]. Additionally, CRISPLD1 has been shown to be expressed in the brain, spinal cord, nose, alimentary system, respiratory system, skeleton and limbs (EURExpress II; www.eurexpress.org/ee/). These findings led us to assess whether there was an association between CRISPLD1 and NSCLP.

6.2. CRISPLD gene family and folic acid

Both CRISPLD1 and CRISPLD2 contain more cysteine residues (25 and 26, respectively) when compared to the average cysteine composition of comparable sized proteins (5% vs. <2%) [273]. Cysteine, a nonessential amino acid, is synthesized in the folate pathway (Figure 6.2) [274]. Genes in this pathway have been of interest in the study of birth defects because periconceptional folic acid usage decreases the birth prevalence of neural tube defects up to 70% [59,60,275-278]. Likewise, the recurrence rate of orofacial clefting has been reduced in mothers taking higher doses of periconceptional folate but the reduction of NSCLP on a population level has been modest [29,63,96,279-283]. The 5,10-methylenetetrahydrofolate reductase (MTHFR) gene regulates homocysteine levels and two nonsynonymous coding polymorphisms *C677T* (alanine to valine) and *A1298C* (glutamate to alanine) affect the enzymatic activity [88,284]. Studies assessing these MTHFR variants find minimal if any association with NSCLP [64,65,67,68,81,83,88,284-286]. However, a new study finds evidence suggesting that the folate pathway genes have a role in NSCLP [286].

Perturbation of the folate pathway could affect the production of cysteines with a resulting downstream effect on the synthesis and/or function of the CRISPLD genes, both of which require a large number of cysteine residues. Here, we asked if CRISPLD1 was

Figure 6.2: Folate gene pathway*



*dUMP = deoxyuridine monophosphate; TYMS = Tymidylate synthetase; FOLR2 = Folate receptor 2; MTHFS = 5,10-methylenetetrahydrofolate synthase; MTHFR = 5,10-methylenetetrahydrofolate reductase; MTRR = 5-methyletetrahydrofolate-homocysteine methyltransferase reductase; MTR = Methylenetetrahydrofolate dehydrogenase

associated with NSCLP and then tested whether CRISPLD1, CRISPLD2 and folate pathway genes interact to create a susceptibility to NSCLP.

6.3. Materials and Methods

Refer to chapter two for sample collection, DNA extraction, Linkage IVb Panel Genotyping, SNP criteria, genotyping, and analysis (parametric and nonparametric linkage parameters and analysis, FBAT, PDT, G-PDT, APL and GEE).

When this study was undertaken, 91 and 258 nonHispanic white and 39 and 91 Hispanic families with positive (multiplex) and negative (trios) histories of NSCLP, respectively, were genotyped with nine SNPs (Table 6.1; Figure 6.3). Additionally, 97 SNPs spanning 14 folate pathway genes were interrogated and included: Betaine-homocysteine methyltransferase (BHMT, 6 SNPs), BHMT2 (6 SNPs), Cystathionine-beta-synthase (CBS, 12 SNPs), Folate receptor 1 (FOLR1, 4 SNPs), FOLR2 (4 SNPs), Methylenetetrahydrofolate dehydrogenase 1 (MTHFD1, 12 SNPs), MTHFD2 (4 SNPs), 5,10-methylenetetrahydrofolate reductase (MTHFR, 8 SNPs), 5,10-methylenetetrahydrofolate synthase (MTHFS, 4 SNPs), Methylenetetrahydrofolate dehydrogenase (MTR, 9 SNPs), 5-methyletetrahydrofolate-homocysteine methyltransferase reductase (MTRR, 3 SNP), ATG9 autophagy related 9 homolog (NOS3, 6 SNPs), Solute carrier family 19, member 1 (SLC19A1, 10 SNPs) and Tymidylate synthetase (TYMS, 9 SNPs) [286] (Supplemental Table 6.1).

6.4. Results

Ten SNPs spanning chromosome 8q21.11-8q21.12 (Illumina Linkage IVb panel) in the nonHispanic white multiplex families gave parametric and nonparametric linkage LOD

Table 6.1: CRISPLD1 SNPs genotyped in this study and allele frequencies

Gene	dbSNP	Chr:bp	Location	Alleles†	NHW MAF[a]	Hispanic MAF[b,*]	p-value
CRISPLD1	rs2925155	8:76048852	upstream	A/G	0.263	0.317	0.033
CRISPLD1	rs960856	8:76057462	upstream	T/A	0.329	0.248*	0.002*
CRISPLD1	rs7841231	8:76065317	intron 2	A/G	0.321	0.232*	0.0005*
CRISPLD1	rs17295835	8:76069196	intron 2	T/C	0.299	0.354	0.035
CRISPLD1	rs1455809	8:76075622	intron 2	T/C	0.312	0.234*	0.002*
CRISPLD1	rs1455796	8:76086362	intron 2	C/G	0.405	0.411	0.804
CRISPLD1	rs10957748	8:76101956	intron 12	C/T	0.292	0.21*	0.001*
CRISPLD1	rs13248650	8:76109250	downstream	T/G	0.3	0.362	0.019
CRISPLD1	rs11988595	8:76115241	downstream	C/T	0.31	0.233*	0.003*

†Major allele as identified in NHW listed first

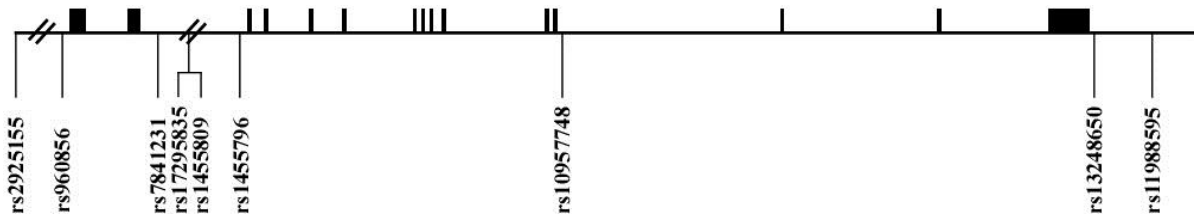
[a]Minor allele frequency of nonHispanic white dataset

[b]Corresponding frequency in Hispanic dataset of nonHispanic white minor allele

[c]Corresponding frequency in Columbian dataset of nonHispanic white minor allele; p-values represent allele frequency differences between Columbian and Hisp datasets.

*p<0.01

Figure 6.3: CRISPLD1 SNPS



scores of 1.23 and 1.07, respectively (Figure 6.1). Although these LOD scores are not considered significant, the CRISPLD1 gene is in this region.

Nine CRISPLD1 SNPs (5 intragenic, 4 intergenic; Figure 6.3, Table 6.1) were genotyped in our NHW and Hispanic families. All SNPs had >95% call rate and were in HWE. The data was stratified by ethnicity because the allele frequencies for five of the nine SNPs differed between the NHW and Hispanic groups ($p \leq 0.006$). The data was also stratified by the presence or absence of family history (FH) [214]. Significant linkage disequilibrium was found in both datasets ($r^2 > 0.95$; Supplemental Table 6.1). A maximum LOD score of 1.35 was found to CRISPLD1 SNPs in NHWs (data not shown). There was no evidence of linkage in the Hispanic dataset.

Association analysis of the single SNPs identified altered transmission for only rs1455809 in the NHW simplex families ($p=0.05$) (Supplemental Table 6.2). No single SNP associations were detected in the Hispanic group. None of the CRISPLD1 2-SNP haplotypes demonstrated altered transmission in either ethnicity (data not shown).

Three gene-gene interactions were detected between CRISPLD1 and CRISPLD2 SNPs ($0.003 \leq p < 0.01$) (Table 6.2). The same CRISPLD1 SNPs in NHW and Hispanics interacted with different CRISPLD2 SNPs in each ethnic group (Table 6.2). The most significant interactions were rs13248650–rs12051468 ($p=0.004$) and rs13248650-rs8051428 ($p=0.003$) in the NHW and Hispanic groups, respectively.

GEE analysis identified numerous interactions between SNPs in CRISPLD1 and CRISPLD2 and the folate pathway genes, even after Bonferroni correction ($p<0.0017$). In the NHW dataset, SNPs in five of the folate pathway genes interacted with SNPs in CRISPLD1. The most significant interaction was between rs7166109 in MTHFS and

Table 6.2: CRISPLD1-CRISPLD2 Gene Interactions

Ethnicity	CRISPLD1/SNP	CRISPLD2/SNP	p-value*
nonHispanic white	rs13248650	rs12051468	**0.004**
	rs2925155	rs12051468	**0.010**
	rs960856	rs903194	0.012
	rs7841231	rs903194	0.015
	rs13248650	rs8051428	0.022
	rs17295835	rs12051468	0.023
	rs10957748	rs12051468	0.039
	rs10957748	rs903194	0.043
Hispanic	rs13248650	rs8051428	**0.003**
	rs17295835	rs8051428	0.018
	rs7841231	rs2646112	0.02
	rs17295835	rs2646112	0.022
	rs13248650	rs2646112	0.024
	rs960856	rs721005	0.024
	rs7841231	rs721005	0.03

*$p<0.05$, $p<0.01$ bolded

rs10957748 (p=0.0005) (Table 6.3). rs7166109/MTHFS interacted with four other CRISPLD1 SNPs ($p \leq 0.002$). Three different CBS SNPs interacted with four different CRISPLD1 SNPs; rs1455796/CRISPLD1 interacted with two different CBS SNPs (Table 6.3). In the Hispanic dataset, the most significant CRISPLD1 interactions involved SNPs in MTR ($0.001 \leq p \leq 0.005$) (Table 6.3). Additionally, two different SNPs in TYMS interacted with different CRISPLD1 SNPS in each ethnic group.

Numerous interactions between CRISPLD2 and folate pathway genes were detected in both datasets (Table 6.4). In the NHWs, there were seventeen SNPs in eleven folate pathway genes interacting with CRISPLD2 SNPs; SNPs in two genes survive Bonferroni correction. In addition, by chance alone, there should be only one (0.05 x 14) gene interaction with $p \leq 0.05$; however, eleven genes interacted with CRISPLD2 SNPs with $p \leq 0.01$. The most significant interaction was with BHMT2 ($0.0005 \leq p \leq 0.003$). Two MTR SNPs, rs1546124 and rs1874015, also interacted with CRISPLD2 ($0.0009 \leq p \leq 0.007$). In the Hispanics, there were ten SNPs in six folate pathway genes interacting with CRISPLD2. The most significant interaction was with a SNP in TYMS. Interestingly, in both groups, rs2236222/MTHFD1 interacted with rs12051468/CRISPLD2 ($0.003 \leq p \leq 0.008$). Again, while only one gene would be expected to interact with CRISPLD2 SNPs with $p<0.05$, there are six genes with $p \leq 0.01$. Furthermore, for both ethnic groups, SNPs in TYMS, MTHFR and FOLR1 interacted with CRISPLD2 SNPs.

Table 6.3: Folate Pathway and CRISPLD1 gene interactions

Ethnicity	Gene/SNP	CRISPLD1 SNP	p-value*
nonHispanic white	CBS/rs234783	rs1455796	0.004
		rs13248650	0.005
		rs17295835	0.006
	CBS/rs2851391	rs2925155	0.007
	CBS/rs12329790	rs1455796	0.008
	MTHFS/rs7166109	rs10957748	0.0005
		rs960856	0.001
		rs7841231	0.001
		rs11988595	0.001
		rs1455809	0.002
	TYMS/rs502396	rs10957748	0.01
Hispanic	MTR/rs1266164	rs1455796	0.001
	MTR/rs12354209	rs1455796	0.004
	MTR/rs1806505	rs10957748	0.005
	TYMS/rs11540152	rs1455796	0.004

*p-value≤0.01

Table 6.4: Folate Pathway and CRISPLD2 Gene Interactions

Ethnicity	Gene/SNP	CRISPLD2 SNP	p-value*
nonHispanic white	BHMT/rs645112	rs1874014	0.009
	BHMT/rs3733890	rs16974880	0.01
	BHMT2/rs2253262	rs1874014	0.0005
	BHMT2/rs682985	rs1874014	0.001
	BHMT2/rs1422086	rs1874014	0.003
	FOLR1/rs3016432	rs2646112	0.007
	FOLR2/rs2276048	rs2646112	0.005
	MTHFD1/rs2236222	rs12051468	0.008
	MTHFD2/rs7587117	rs2641670	0.003
	MTHFR/rs1476413	rs8051428	0.007
	MTHFS/rs2586179	rs1874015	0.003
		rs767050	0.009
	MTR/rs12354209	rs1546124	0.0009
		rs1874015	0.007
	MTR/rs1266164	rs1546124	0.004
	NOS3/rs1800779	rs16974880	0.005
	NOS3/rs2373929	rs16974880	0.005
	TYMS/rs11540152	rs2646129	0.008
	TYMS/rs2853532	rs12051468	0.010
Hispanic	CBS/rs2851391	rs2646112	0.009
	FOLR1/rs2071010	rs774206	0.006
		rs721005	0.009
	MTHFD1/rs2236222	rs12051468	0.003
	MTHFD1/rs11849530	rs2641674	0.006
	MTHFR/rs1801131	rs2646112	0.003
	MTHFR/rs1476413	rs2646112	0.003
	MTHFR/rs535107	rs8061351	0.007
	SLC19A1/rs3788205	rs4783099	0.01
	TYMS/rs495139	rs8061351	0.002
	TYMS/rs1001761	rs8061351	0.003

*p-value≤0.01

6.5. Discussion

In previous studies, we reported the association of variation in CRISPLD2 with NSCLP and showed that CRISPLD2 was expressed in the craniofacies of developing mouse embryos at E13.5 [203]. This association has been confirmed in an independent dataset [287]. In our on-going studies to define the genetic loci contributing to NSCLP, 11 chromosomal regions with LOD scores ≥ 1.5 were identified in a genome scan performed on nine multiplex NHW NSCLP families [105]. A LOD score of 1.23 was found for the 8q13.2-21.13 chromosomal region which contains the CRISPLD1 gene; CRISPLD1 has significant homology to CRISPLD2 and both genes are composed of more cysteines than comparable-sized proteins (5% protein composition vs. 2%) [273]. The cysteines provide secondary structure to the protein backbone and help protect cells against the harmful effects of oxidation [288,289]. These interesting results led to the interrogation of CRISPLD1 as a candidate NSCLP gene. In complementary studies, we show that variation in different folate pathway genes individually and through interactions contribute to NSCLP [290]. Interestingly, cysteines, which could potentially be utilized by the CRISPLD genes, are produced in the homocysteine biosynthesis pathway, which is part of the folate pathway [291]. Based on these observations, we asked whether the CRISPLD genes interacted with genes in the folate pathway.

We show that only one CRISPLD1 SNP, rs1455809, had marginally altered transmission (p=0.05) suggesting that variation in CRISPLD1 alone does not play a significant etiologic role in NSCLP. Moreover, there were only three CRISPLD1-CRISPLD2 interactions (two in NHWs and one in Hispanics) suggesting that variants in these genes do not interact to create a significant susceptibility to NSCLP. Indeed, their expression patterns in mice are

different. At E14.5, CRISPLD1 is expressed in the mouse brain, spinal cord, nose, alimentary system, respiratory system, skeleton and limbs, while CRISPLD2 is expressed in the oral region, visceral organs, alimentary system and salivary gland (www.eurexpress.org/ee/) [203]. In zebrafish, CRISPLD2 is expressed in the craniofacial region and tail at all stages of development while CRISPLD1 localizes to the splanchocranium, pectoral fin, presumptive vertebrate and epiphysis (Chiquet and Hecht, unpublished results) [292].

Folic acid is important in embryogenesis and this is underscored by the 70% reduction of neural tube defects since folic acid supplementation of grain products in 1998 [277,281,293,294]. While a similar reduction has not been observed for NSCLP on a population-basis, decreased recurrence of NSCLP has been observed when high-risk mothers take high dose folic acid pre- and post conception [29,60,282,283,295]. This suggests that perturbation of folic acid levels or genes in the folate pathway could contribute to NSCLP. To evaluate this question, we first assessed whether folate pathway genes were associated with NSCLP [290]. Evidence for an association was found for SNPs in NOS3, TYMS, MTR, BHMT2, MTHFS and SLC19A1; many of the associated variants occurred in potential promoter or regulatory regions. Complete description of the results is found in Blanton et al. 2010. The endogenous folate pathway is responsible for synthesizing cysteine and cysteine is required for the CRISPLD1/2 protein structures. This led us to postulate that CRISPLD genes interact with genes in the folate pathway. Indeed, we found evidence for several interactions. In the NHWs, rs502396 in TYMS interacted with CRISPLD1 and this SNP had altered transmission in the single SNP folate pathway gene analysis [290]. Similarly, for CRISPLD2, two different SNPs in TYMS, rs11540152 and rs2853532, interacted although

these SNPs did not have altered transmission in the folate study. Additionally, rs2373929 in NOS3 interacted with CRISPLD2 and this SNP also exhibited altered transmission in the single SNP folate pathway gene analysis. In the Hispanics, three MTR SNPs interacted with CRISPLD1 but they were different from the MTR SNPs identified in the single SNP folate analysis. For CRISPLD2, rs3788203 in SLC19A1, which was associated in the single folate SNP analysis, also showed interaction. These results are particularly interesting because MTR is necessary to metabolize homocysteine to methionine [291]. NOS3 directly regulates MTR and SLC19A1 affects 5mTHF, an intermediary that is also metabolized by MTR. Additionally, interactions with both CRISPLD genes were found for CBS, BHMT and BHMT2 in the NHW dataset. These genes also participate in the methionine cycle. Other interactions were found for both CRISPLD genes and TYMS and MTHFS, both of which participate in the DNA synthesis necessary for embryonic development.

Altogether, these results suggest that CRISPLD genes play a role in NSCLP but not through a simple mechanism. We have shown that variation in CRISPLD2 is associated with NSCLP but CRISPLD1 alone is not. A more likely mechanism that is suggested by these results is that perturbation of multiple genes in a pathway affects protein function, which can have a profound effect on embryogenesis. Indeed, we have found the same etiologic model for nonsyndromic clubfoot, which is another common complex birth defect [296]. These results would also fit what is known about recurrence in NSCLP wherein high dose folic acid reduces the recurrence in high-risk families. These families may have more liability genes that could potentially respond to drug/vitamin therapy. In contrast, on a population basis, genetic heterogeneity may play a role and folate deficiency may only be a small piece of the puzzle. Thus additional studies are necessary to define all the genetic

contributions. Additionally, validation and functional studies will help determine the significance of these findings to NSCLP. These results are intriguing because they help expand our understanding of the CRISPLD gene family and its role in clefting etiology.

Supplemental Table 6.1: Folate SNPs genotyped in Blanton *et al.*, 2010 that were used for GEE interactions in this study. Modified from [286].

Gene	dbSNP	chr:bp	location	Alleles*
MTHFR	rs1889292	1:11763530	downstream	G/A
MTHFR	rs2274976	1:11773514	exon 12	G/A
MTHFR	rs1476413	1:11774887	intron 10	G/A
MTHFR	rs1801131	1:11777063	exon 8	A/C
MTHFR	rs1801133	1:11778965	exon 5	C/T
MTHFR	rs4846052	1:11780538	intron 4	C/T
MTHFR	rs3737964	1:11789631	upstream	A/G
MTHFR	rs535107	1:11812055	upstream	A/G
MTR	rs12354209	1:235025875	intron 1	A/G
MTR	rs4077829	1:235054413	intron 9	G/T
MTR	rs1806505	1:235063198	intron 13	C/T
MTR	rs6668344	1:235067949	intron 14	C/T
MTR	rs1770449	1:235104784	intron 24	A/G
MTR	rs1805087	1:235115123	exon 26	A/G
MTR	rs1266164	1:235117574	intron 27	G/A
MTR	rs11799647	1:235127544	exon 33	A/G
MTR	rs1050993	1:235128928	exon 34	G.A
MTHFD2	rs828858	2:74275702	upstream	T/A
MTHFD2	rs6758506	2:74286460	exon 2	G/A
MTHFD2	rs7587117	2:74302163	downstream	T/C
MTHFD2	rs7571842	2:74314412	downstream	A/G
MTRR	rs162029	5:7918527	upstream	G/A
MTRR	rs2303080	5:7931424	exon 5	T/A
MTRR	rs13166314	5:7955525	downstream	T/A
BHMT2	rs2253262	5:78387392	upstream	C/A
BHMT2	rs476620	5:78390202	upstream	A/G
BHMT2	rs626105	5:78405657	intron 1	G/A
BHMT2	rs682985	5:78409187	exon 2	T/C
BHMT2	rs1422086	5:78410621	intron 2	A/C

Gene	dbSNP	chr:bp	location	Alleles*
BHMT	rs645112	5:78438303	upstream	A/C
BHMT	rs567754	5:78452172	intron 4	C/T
BHMT	rs3733890	5:78457715	exon 6	G/A
BHMT	rs585800	5:78462964	exon 8	A/T
BHMT	rs617219	5:78465350	downstream	A/C
BHMT	rs1915706	5:78471967	downstream	C/T
NOS3	rs1800779	7:150320876	intron 1	A/G
NOS3	rs1800780	7:150329812	intron 12	A/G
NOS3	rs891512	7:150339022	intron 13	G/A
NOS3	rs3918211	7:150341840	exon 26	T/C
NOS3	rs2373929	7:150345745	downstream	G/A
FOLR1	rs3016432	11:71574903	upstream	C/T
FOLR1	rs2071010	11:71578612	intron 1	G/A
FOLR1	rs1540087	11:71579139	intron 1	C/T
FOLR1	rs11235462	11:71586273	downstream	T/A
FOLR2	rs651646	11:71607174	intron 1	T/A
FOLR2	rs514933	11:71607855	intron 2	A/G
FOLR2	rs2298444	11:71610062	intron 4	A/G
FOLR2	rs2276048	11:71618860	downstream	A/G
MTHFD1	rs1076991	14:63924794	upstream	T/C
MTHFD1	rs4902283	14:63954365	exon 7	C/T
MTHFD1	rs1885031	14:63957808	intron 8	A/G
MTHFD1	rs11551058	14:63963854	exon 12	G/T
MTHFD1	rs2236225	14:63978598	exon 20	C/T
MTHFD1	rs2236224	14:63978904	intron 20	C/T
MTHFD1	rs1256142	14:63980547	intron 20	T/C
MTHFD1	rs2236222	14:63984935	intron 21	T/C
MTHFD1	rs10137921	14:63985918	exon 24	C/T
MTHFD1	rs11849530	14:63988165	intron 24	A/G
MTHFD1	rs1256146	14:63990418	intron 24	G/A
MTHFD1	rs34616731	14:63999040	downstream	T/A

Gene	dbSNP	chr:bp	location	Alleles*
MTHFS	rs7166109	15:77913530	downstream	C/T
MTHFS	rs2562744	15:77961443	intron 2	T/G
MTHFS	rs2586179	15:77975061	intron 1	C/A
MTHFS	rs2115540	15:77977363	upstream	T/C
TYMS	rs2853741	18:647352	upstream	C/T
TYMS	rs502396	18:649236	intron 1	T/C
TYMS	rs1001761	18:652103	intron 2	C/T
TYMS	rs11540152	18:652215	exon 3	T/C
TYMS	rs11540153	18:659117	exon 4	C/T
TYMS	rs2853532	18:660414	intron 3	C/T
TYMS	rs495139	18:666008	downstream	C/G
CBS	rs1051319	21:43346936	exon 20	C/G
CBS	rs11700812	21:43353659	exon 12	C/T
CBS	rs2298758	21:43358596	exon 7	G/A
CBS	rs2298759	21:43359173	intron 5	A/G
CBS	rs2851391	21:43360473	intron 4	T/C
CBS	rs234713	21:43360960	intron 4	A/G
CBS	rs234714	21:43361102	intron 4	C/T
CBS	rs234783	21:43376312	upstream	C/T
CBS	rs234784	21:43376503	upstream	C/T
CBS	rs234785	21:43377074	upstream	C/G
SLC19A1	rs10483080	21:45750430	downstream	G/C
SLC19A1	rs12483377	21:45755537	downstream	G/A
SLC19A1	rs7278825	21:45760370	exon 6	G/A
SLC19A1	rs3788190	21:45761386	intron 5	G/A
SLC19A1	rs2330183	21:45777720	intron 2	T/C
SLC19A1	rs1051266	21:45782222	exon 2	G/A
SLC19A1	rs4819130	21:45782727	intron 1	T/C
SLC19A1	rs3788205	21:45788806	upstream	C/T

*Major allele as identified in NHW population listed first

Supplemental Table 6.2: CRISPLD1 Linkage Disequilibrium

	rs2925155	rs960856	rs7841231	rs17295835	rs1455809	rs1455796	rs10957748	rs13248650	rs11988595
rs2925155		0.092	0.088	0.627	0.086	0.266	0.078	0.611	0.086
rs960856	0.055		0.924	0.131	0.864	0.134	0.772	0.141	0.864
rs7841231	0.062	0.964		0.175	0.974	0.149	0.861	0.176	0.974
rs17295835	0.593	0.191	0.201		0.176	0.373	0.149	0.970	0.173
rs1455809	0.055	0.876	0.886	0.179		0.178	0.864	0.180	1.000
rs1455796	0.230	0.247	0.254	0.279	0.281		0.166	0.369	0.178
rs10957748	0.047	0.821	0.838	0.170	0.904	0.266		0.154	0.864
rs13248650	0.609	0.186	0.192	0.967	0.193	0.284	0.179		0.179
rs11988595	0.052	0.887	0.893	0.186	0.972	0.287	0.903	0.193	

*Hispanic above diagonal, nonHispanic white below. r2 values reported

yellow= $0.7 \leq r^2 < 0.8$

orange= $0.8 \leq r^2 < 0.9$

red-orange: $r^2 \geq 0.9$

Supplemental Table 6.3: CRISPLD1 association results in (A) NHW and (B) Hispanic datasets*

(A)

	All				Positive Family History				Negative Family History			
dbSNP	PDT	G-PDT	APL	FBAT-e	PDT	G-PDT	APL	FBAT-e	PDT	G-PDT	APL	FBAT-e
rs2925155	0.7404	0.9496	0.1593	0.4170	0.9449	0.6764	0.1786	0.6174	0.5211	0.3040	0.4565	0.5211
rs960856	0.4527	0.3119	0.2647	0.3440	0.7021	0.5830	0.3185	0.5446	0.4669	0.4009	0.4406	0.4669
rs7841231	0.4225	0.0817	0.1201	0.3154	0.5186	0.3379	0.2494	0.3018	0.6331	0.1838	0.2427	0.6331
rs17295835	0.7170	0.8964	0.1384	0.4214	0.7355	0.9047	0.6483	0.9452	0.3336	0.2588	0.1394	0.3336
rs1455809	0.5741	0.0544	0.1702	0.4962	0.7400	0.5096	0.2320	0.6216	0.6310	**0.0447**	0.3453	0.6310
rs1455796	0.1425	0.2988	0.5498	0.1942	0.3681	0.5477	0.7520	0.6029	0.2191	0.4790	0.6440	0.2191
rs10957748	0.7595	0.0643	0.3004	0.6535	0.9484	0.3091	0.6522	0.9857	0.5610	0.0760	0.2746	0.5640
rs13248650	1.0000	0.8348	0.2349	0.6331	0.5773	0.8800	0.9358	0.9302	0.5050	0.2904	0.1726	0.5050
rs11988595	0.7206	0.0630	0.3696	0.6712	0.7928	0.4365	0.3222	0.7066	0.8071	0.0810	0.5822	0.8071

(B)

	All				Postive Family History				Negative Family History			
dbSNP	PDT	G-PDT	APL	FBAT-e	PDT	G-PDT	APL	FBAT-e	PDT	G-PDT	APL	FBAT-e
rs2925155	0.7237	0.9488	0.4669	0.8539	0.3532	0.6146	0.7940	0.4489	0.7009	0.8381	0.5219	0.3532
rs960856	0.8728	0.9255	0.3265	0.4855	0.4838	0.4987	0.2697	0.8348	0.6733	0.6748	0.5831	0.4838
rs7841231	0.7456	0.9248	0.1243	0.4533	0.3545	0.3515	0.3497	1.0000	0.6547	0.6612	0.9022	0.3545
rs17295835	0.8618	0.7953	0.4030	0.6732	0.3304	0.6381	0.6514	0.5896	0.5529	0.7393	0.9409	0.3304
rs1455809	1.0000	0.5914	0.3373	0.3173	0.3763	0.1901	0.3053	0.6374	0.5023	0.1548	0.5954	0.3763
rs1455796	0.8788	0.4925	0.8088	0.9104	0.5994	0.7456	0.8152	0.2752	0.7653	0.9063	0.7550	0.5994
rs10957748	1.0000	0.8678	0.9729	0.1724	0.2278	0.0875	0.1434	0.5127	0.3650	0.1618	0.2152	0.2278
rs13248650	0.5775	0.7266	0.2977	0.3683	0.0641	0.1457	0.6053	0.3452	0.2855	0.5985	0.8834	0.0641
rs11988595	1.0000	0.5914	0.3108	0.3173	0.3763	0.1901	0.2639	0.6374	0.5023	0.1548	0.5640	0.3763

*$p<0.05$ in bold.

Chapter Seven: The role of CRISPLD2 during zebrafish development

7.0. Introduction

In chapters five and six, we showed an association of CRISPLD2 with NSCLP, which suggests that perturbation of CRISPLD2 expression plays a role in this birth defect. However, the function of this gene in craniofacial development is unknown. As previously discussed in Section 1.5.2.d, animal studies have provided important insights about the function of NSCLP genes. A number of different organisms, including mouse, zebrafish and chicken, have been utilized to study the role of the IRF6 gene in the etiology of orofacial clefting. Two studies discussed in Section 1.5.2.d showed that mice lacking *Irf6* had, among other anomalies, a clefting phenotype. This resulted from abnormal keratinocyte proliferation and differentiation [147,149]. The zebrafish model has also been used to determine the function of z*irf6* and identify the role it plays in NSCLP. Studies have shown that Irf6 is expressed in the developing zebrafish mouth and pharyngeal arches [297]. Morpholino knockdown of Irf6 in zebrafish embryos failed to disrupt development; however, deletion of the maternal Irf6 transcripts from these embryos arrested development at the blastula stage [298]. Further investigation showed that these embryos had abnormal and weak superficial epithelium [298]. These findings and those in the mouse supports the role of IRF6 in normal epithelium development. Expression studies of Irf6 in chick embryos have found it to also be expressed in the developing craniofacial processes, including the maxillary process, medial nasal process, nasal pit and medial epithelial edge of the lateral palatine shelves [299,300]. Altogether, these three animal studies indicate that IRF6 is an important regulator of epithelial growth and differentiation.

In previous studies, we have shown that *CRISPLD2* is expressed in the developing orofacial region and liver of mouse embryos (see Section 5.3.3 and Figure 5.2, page 79) [203].

As shown in Table 7.1, Crispld2 is expressed in many tissues (Genomics Institute of the Novartis Research Foundation GNF, http://symatlas.gnf.org and T1Dbase, http://t1dbase.org [301]). Interestingly, while neither database lists expression in the fetal face or head, the clone used for *in situ* analysis was derived from the head of an E17.5 mouse, indicating that CRISPLD2 is expressed in that region (Riken CRISPLD2 clone 3321402M02, GeneService, Cambridge, UK).

Table 7.1. Expression of murine CRISPLD2

Prostate	Testis
Placenta	Uterus
Adipocyte	Spinal Cord
Ovary	CD 14+ Monocytes
Fetal and Adult Lung	CD33+ Myeloid
Whole Blood	Olfactory Bulb

As shown in Table 7.1, CRISPLD2 is expressed in the fetal and adult lung. In 1999, Kaplan *et al.* reported their findings of a novel CRISP family gene, late-gestation lung I (*Lgl1*), that was differentially expressed in rat lung tissue [302]. Lgl1 was later classified as a CAP (cysteine-rich secretory proteins, antigen 5 and pathogenesis-related 1) superfamily gene and named CRISPLD2 [303]. Although the HUGO Gene Nomenclature Committee at the European Bioinformatics Institute still lists the gene name as CRISPLD2 (http://www.genenames.org), the Lgl1 nomenclature remains in use in pulmonary research. Lgl1 is a glycoprotein that participates in fetal lung and ureteric branching [304-306]. Crispld2/Lgl1 null mice are embryonic lethals while Crispld2/Lgl1 heterozygous mice are grossly normal [307]. Closer inspection of the lungs from Crispld2/Lgl1 heterozygous mice found delayed maturation; Crispld2 appears to play an important role in epithelial to mesenchymal transition (EMT) in lung tissue [307]. EMT has previously been shown to play an important role in the transition of neuroepithelial cells to neural crest cells (NCCs) prior

to migration and formation of craniofacies, suggesting a potential role of CRISPLD2 in craniofacial development [308,309].

Here, zebrafish was used to define the expression and function of zCrispld2. The zebrafish model was chosen because it has been used to study many genes that regulate craniofacial development, including the Runx, Dlx and Sox gene families (see Section 1.5.2.d) [168-170,297].

7.1. Materials and Methods

The details of the expression, morpholino (MO) and *in situ hybridization* studies are completely described in Chapter 2. Wildtype zebrafish (Danio rerio) were raised and housed following standard protocols [310].

7.4. Results

7.4.1. Zebrafish and human CRISPLD2 sequences are homologous

Amino acid alignment of CRISPLD2 in seven vertebrate species identified homology between all species (Figure 7.1). The z*crispld2* is 55% homologous to human CRISPLD2 at the peptide level. The zebrafish CRISPLD2 protein is 407 amino acids in length compared 461-507 of other species. This difference is reflected in the 108 amino acid residue gap in the alignment field that maps to human amino acid residue numbers 337-444. This difference is species-specific for the zebrafish.

7.4.2. CRISPLD2 is expressed during zebrafish development.

As shown in Figure 7.2A, CRISPLD2 was detected early in development at 3, 4, 27 and 72 hpf. Whole mount *in situ* hybridization was performed to define expression at: 5-7

Figure 7.1: CRISPLD2 amino acid alignment across vertebrate species. Black boxes indicate >50% of sequences have the identical amino acid; Grey boxes indicate >50% of sequences have a conserved substitution.

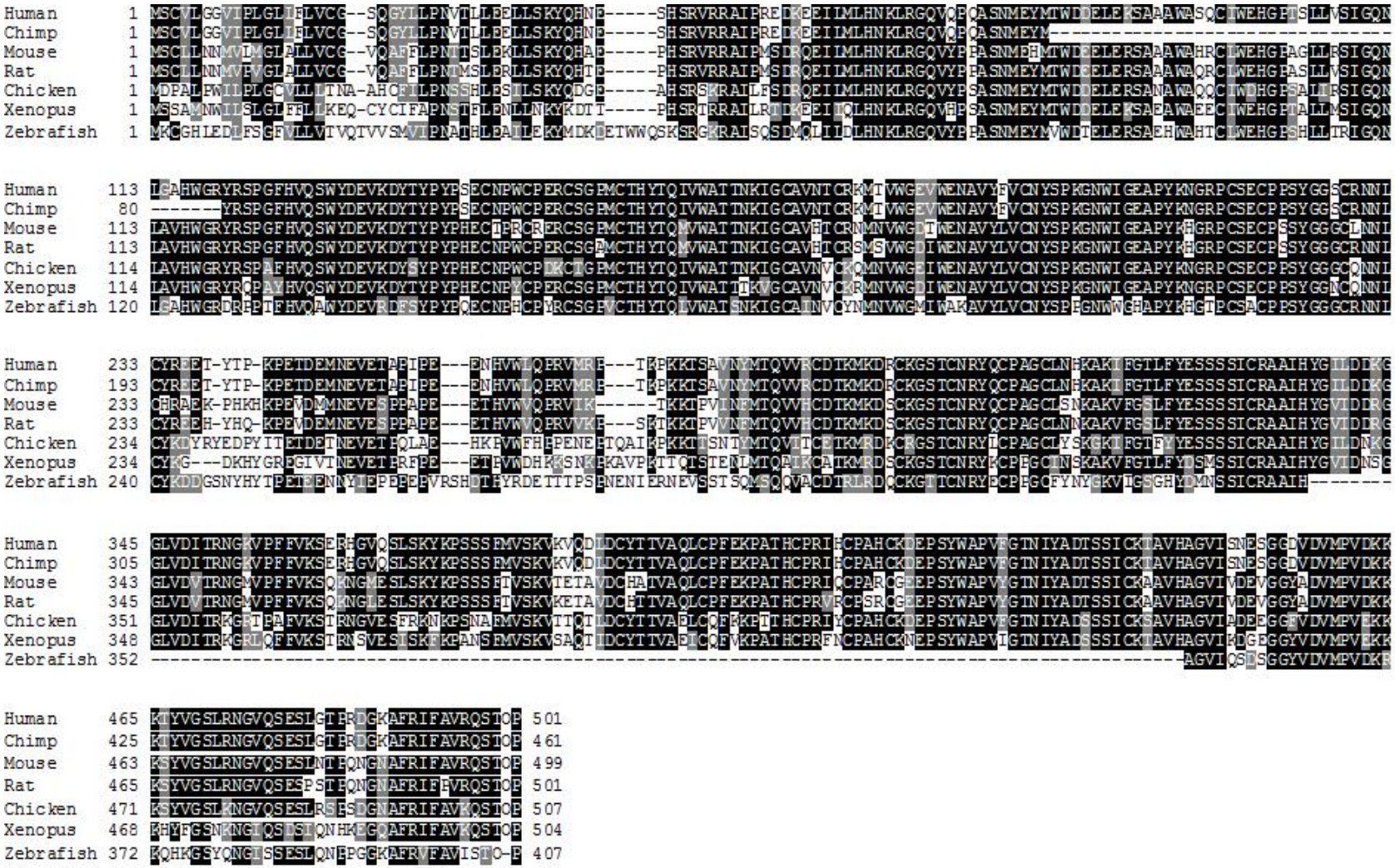
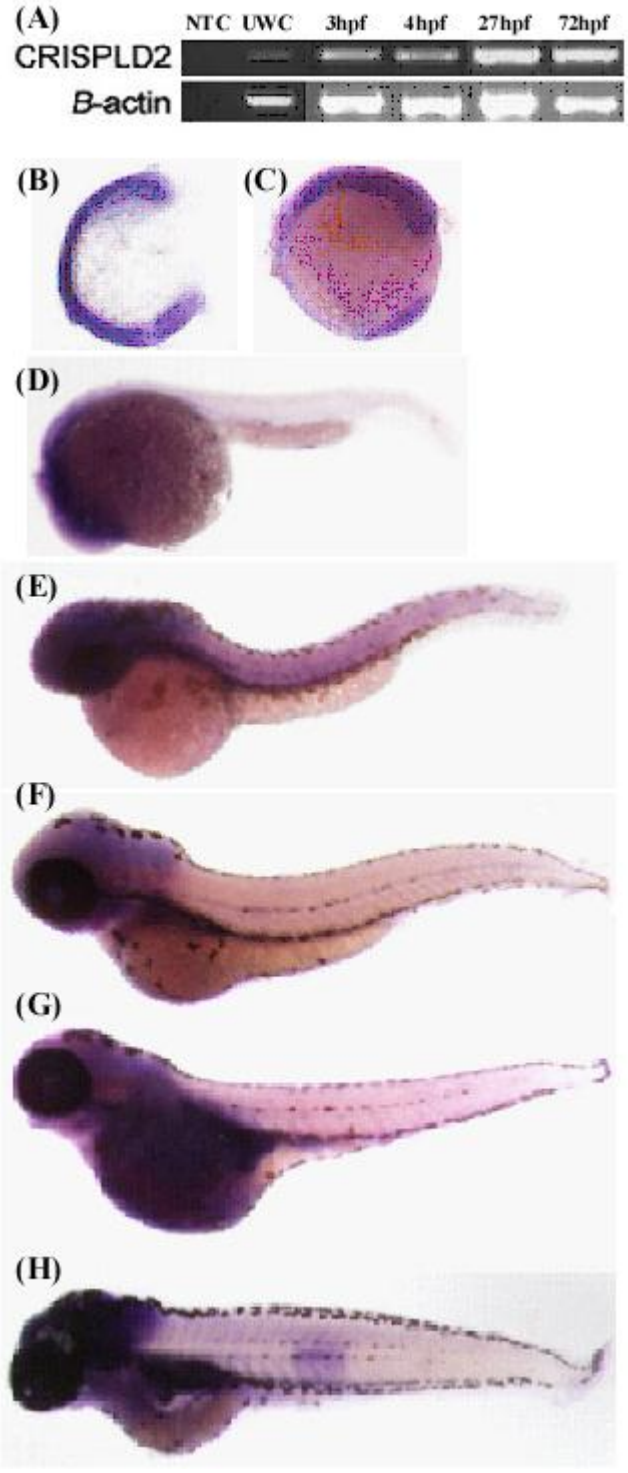


Figure 7.2. CRISPLD2 is expressed in zebrafish during development. (A) Detection of zCRISPLD2 mRNA by RTPCR during development (NTC = no template control, UWT = unfertilized wild type embryo, hpf = hrs post fertilization). (B-H) zCRISPLD2 expression is shown by *in situ* hybridization in purple at (B) 5-7 somite, (C) 13-15 somite, (D) 1 dpf, (E) 2 dpf, (F) 3 dpf, (G) 4 dpf, and (H) 5 dpf.



somite, 13-15 somite, 1dpf, 2dpf, 3dpf, 4dpf and 5dpf using two antisense probes. As shown in Figs. 1B and C, CRISPLD2 is ubiquitously expressed during early development. CRISPLD2 localizes to the anterior half of the embryo including the head region between 1dpf to 5dpf (Figure 7.2D-H). These results were confirmed using multiple probes (data not shown).

7.4.3. Knockdown of CRISPLD2 adversely affects survival rate and causes abnormal phenotypes.

We first tested the efficacy of two z*crispld2* MOs; one directed at exon 1 (designated controlMO) and the other at exon 2 (Figure 7.3A). The controlMO failed to block gene translation (Figure 7.3B, lanes 1-4). *crispld2*MO knocked down gene translation at concentrations greater or equal to 1ng at 1 dpf (Figure 7.3B, lanes 5-6); gene expression was detected at 5 dpf and at concentrations less than 1ng per injection (Figure 7.3B, lanes 7-11).

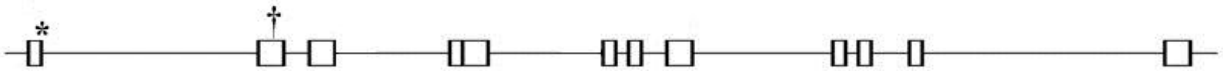
*Crispld2*MO knockdown embryos had lower survival rates, at 1 and 5 dpf, when compared to control embryos ($p<0.05$ and $p<0.005$, respectively) (Figure 7.4A). A higher percentage of abnormal phenotypes were also found in these embryos ($p<0.05$) (Figure 7.4B). At 1 dpf, these abnormal phenotypes included failure to develop (Figure 7.4D), delayed development (Figure 7.4E-F) and grossly normal embryos when compared to wild type (Figure 7.4C). At 5 dpf, of those embryos that survived, a range of phenotypes was observed including severely truncated body (Figure 7.4H), shortened and curved tail with cardiac edema (Figure 7.4I) and grossly normal when compared to wild type (Figure 7.4G).

7.4.4. Knockdown of CRISPLD2 disrupts NCC formation.

NCCs originate in the hindbrain and migrate into the craniofacial region to form the maxillary and mandibular jaw structures [311]. Dlx2 is expressed in these migrating NCCs

Figure 7.3: Morpholino knockdown of CRISPLD2. (A) Gene showing controlMO (*) and *crispld2*MO (†) targets (boxes = exons). (B) RTPCR results following MO injection at various timepoints. CRISPLD2 expression is knocked down at c*rispld2*MO concentrations greater or equal to 1 ng. β-Actin was used as a control.

(A)



(B)

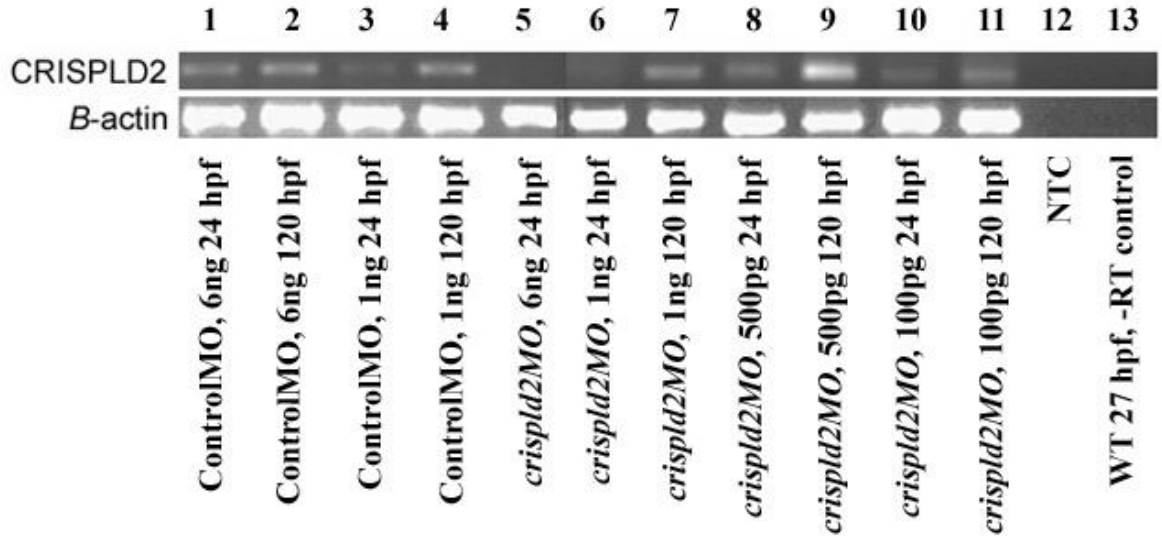
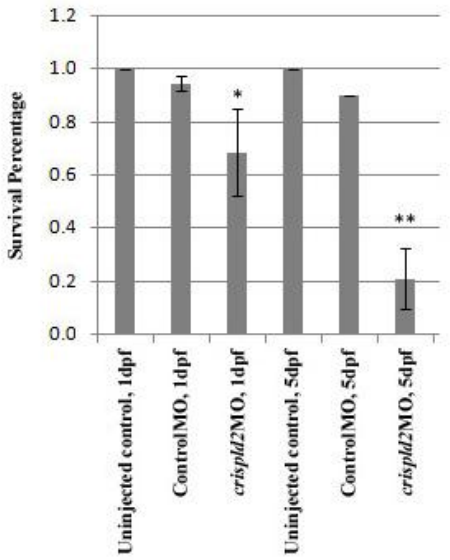
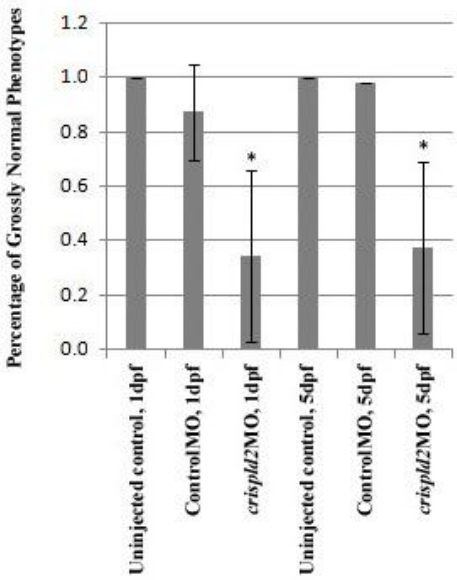


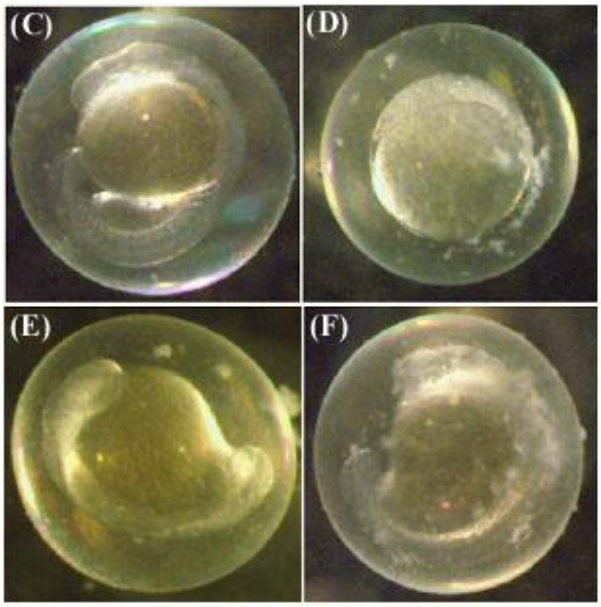
Figure 7.4: CRISPLD2 knockdown causes decreased survival and abnormal phenotypes. Quantification of survival and abnormal phenotypes after MO knockdown. (A) Embryos with CRISPLD2 knockdown have significantly lower survival rates compared to control embryos (T-test: * $p<0.05$; ** $p<0.005$). (B) Embryos with CRISPLD2 knockdown have significantly more abnormal phenotypes compared to control embryos. Examples of phenotypes are shown in C-F at 1dpf and G-I at 5 dpf. (C) Uninjected embryo at 1 dpf. (D-F) Various MO phenotypes at 1 dpf. (G) Uninjected embryo at 5 dpf. (H-I) MO phenotypes at 5 dpf.

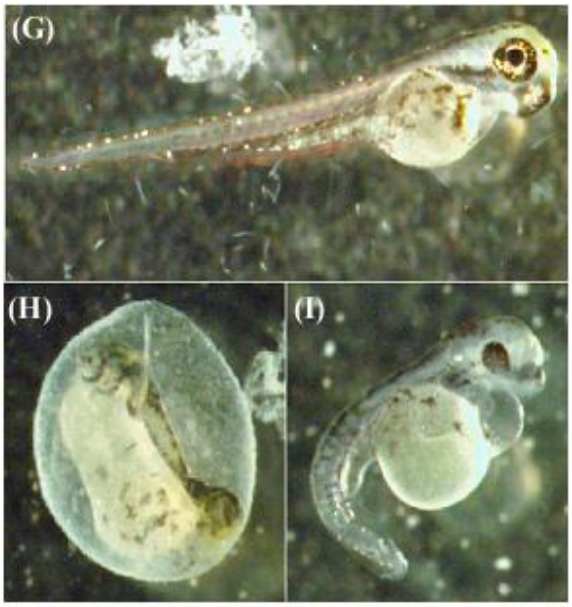


(A)



(B)





and is useful as a marker to visualize normal NCC formation and migration [312,313]. Knockdown of CRISPLD2 altered the pattern of Dlx2 clustering compared to control embryos at 1 dpf (Figure 7.5). Four distinct clusters of NCCs were seen in the *crispld2*MO injected embryos compared to controls. Additionally, the anterior NCCs clustering appeared less dense and clusters 3 and 4 were indistinct from each other. ControlMO injected (data not shown) and uninjected embryos had identical Dlx2 expression patterns (Figure 7. 5).

7.4.5. Knockdown of CRISPLD2 disrupts normal craniofacial cartilage.

Alcian blue staining was use to visualize the cartilage structures of the craniofacies. Numerous anomalies were found in the *crispld2*MO injected embryos including fewer ceratobranchial cartilages forming the lower jaw (≤ three pairs compared to five) and broader Meckel's and palatoquadrate cartilages that are more "U" shaped than "V" shaped (Figure 7.6).

7.5. Discussion

Previously, we showed that CRISPLD2 is associated with NSCLP in humans and expressed during specific stages of mouse craniofacial development [203,314]. In these studies, zebrafish were used to further define expression and function of CRISPLD2 during craniofacial development. We found that zCrispld2 is expressed throughout development from 3hpf to 5dpf (Figure 7.2). zCrispld2 was diffusely expressed throughout all tissues during early development (5-15 somite stages) and later localized to the craniofacial tissues (1-5 dpf). This expression pattern suggests that z*crispld2* might play an important role in overall early development as well as craniofacial development. This is supported by the observation that the Crispld2/Lgl1 null mouse resulted in embryonic lethality by E9.5, and

Figure 7.5: Dlx2 expression is altered by knockdown of CRISPLD2. Dlx2 expression is visualized by whole mount *in situ* staining at 1 dpf. (A, B) Dlx2 expression in wild type embryos and (C, D) Dlx2 expression in zCRISPLD2 knock down embryos (125X). Arrows denote Dlx2 clustering.

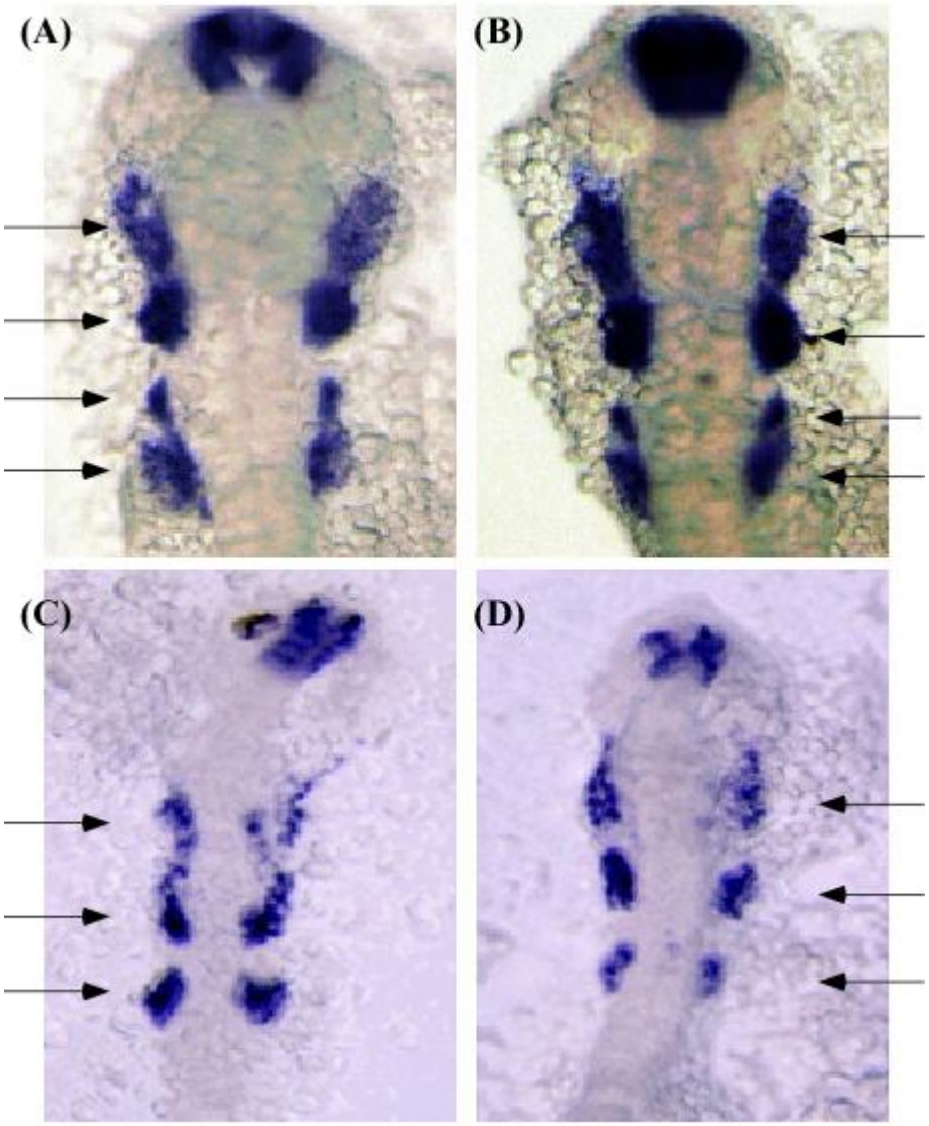
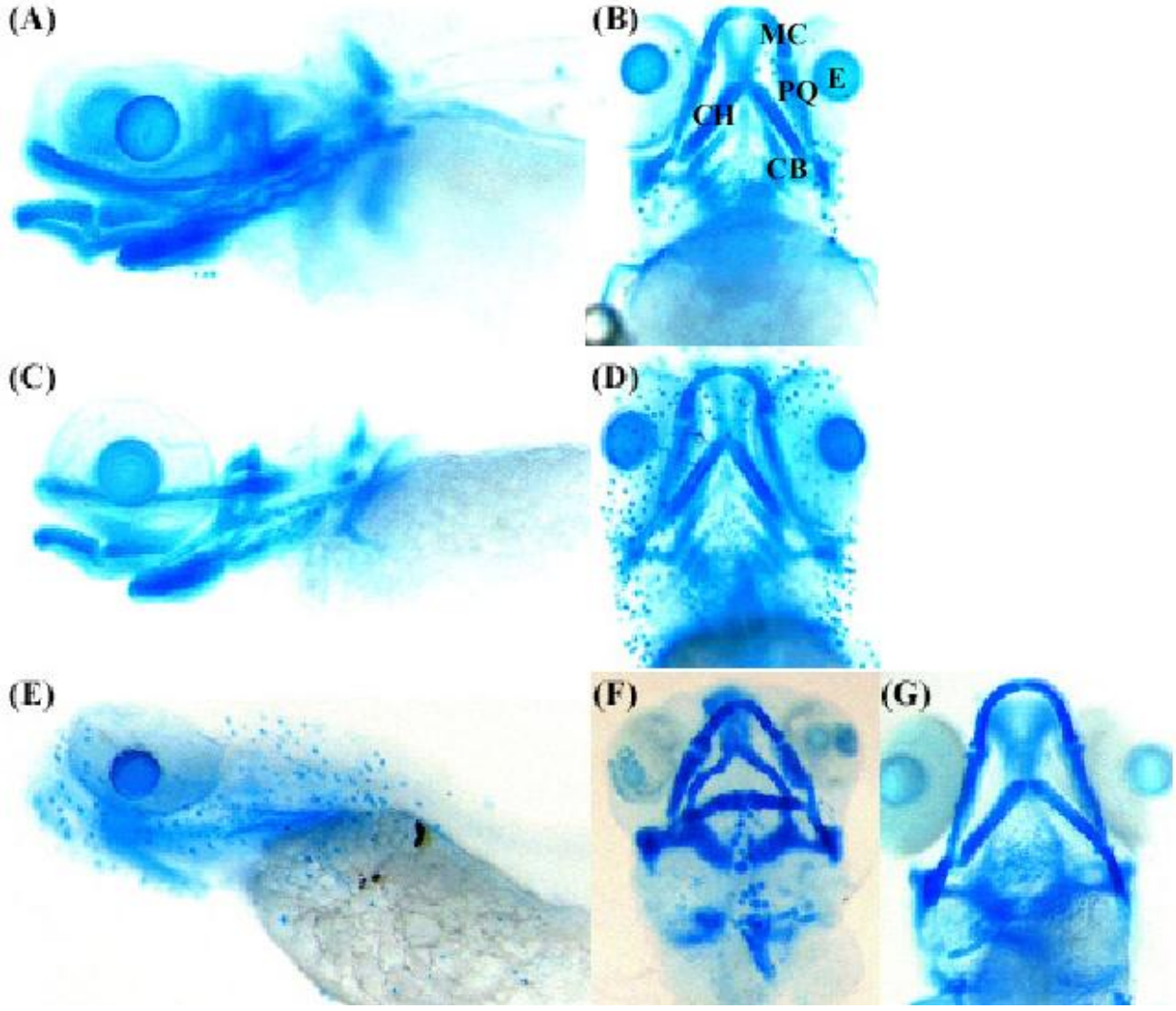


Figure 7.6: Knockdown of CRISPLD2 causes craniofacial abnormalities. Alcian blue staining at 5 dpf shows structural craniofacial anomalies in *crispld2*MO knockdown zebrafish (100X). (A and B) Lateral and ventral view of wild type embryos. (C and D) control MO injected embryos. (E, F and G) *crispld2*MO injected embryos had fewer pairs of ceratobranchial cartilages (CB) and broader Meckel's (MC) and palatoquadrate (PQ) cartilages. (E = eye, CH = ceratohyal)



by our finding that loss of z*crispld2* at 1dpf resulted in lower survival rates and an array of abnormal phenotypes in surviving embryos (Figure 7.4).

In the Crispld2/Lgl1 mouse model, heterozygous mice were grossly normal. However, close examination revealed abnormal lung development with delayed alveolar maturation and disorganization of lung elastin fibers that were trapped in the interstitium, both of these findings resolved with age [307]. Lan *et al*. concluded that Crispld2/Lgl1 functions to regulate epithelial-mesenchymal interactions, similar to lung alveolarization [307]. Interestingly, epithelial-mesenchymal interactions play an important role in gastrulation, craniofacial and cardiac development and cancer progression and metastasis [309,315-317]. In craniofacial development, neuroepithelial cells programmed to become NCCs undergo EMT before they can migrate into the craniofacies [308,309]. This suggests that Crispld2/Lgl1 plays a role in regulating NCC development and migration. This is supported by our observations in zebrafish, where loss of z*crispld2* resulted in altered NCC clustering at 1dpf, including less dense first and second NCC clusters (Figure 7.5). These clusters will form the structures of the first and second pharyngeal arches, including the Meckel's and palatoquadrate cartilages, which were abnormal in zebrafish lacking z*cripsld2* (Figure 7.6). The ceratobranchial cartilages forming the lower jaw are derived from the third, fourth, fifth, sixth and seventh pharyngeal arches [318]. Interestingly, the first three pairs of ceratobranchial cartilages, which were missing in *crispld2*MO embryos, are derived from NCC clusters and pharyngeal arches three and four, which were indistinct from each other at 1 dpf (Figures 7.5 and 7.6). Although the oral facial malformations found in *crispld2*MO zebrafish are different from human NSCLP, this may reflect species-specific differences [318]. Our findings in both zebrafish and mouse show that CRISPLD2 plays an important role in craniofacial

development, and specifically in NCC regulation [203] (Hecht 2010, unpublished data). While we found that loss of *zcrispld2* caused perturbation of NCC migration, we did not identify whether *zCrispld2* inhibits normal EMT in the craniofacies, as it does in the lung [307]. To test whether *zCrispld2* specifically affects craniofacial EMT, future studies should compare the expression of neuroepithelial cell markers, such as *crestin* or *snail1b*, in *crispld2*MO and wild-type embryos [319-321].

The Crispld2/Lgl1 heterozygote mouse has provided some insights about the potential factors that regulate Crispld2/Lgl1. In mouse mesenchymal tissue, Crispld2/Lgl1 has two distinct functions. First, during early lung and kidney organogenesis, Crispld2/Lgl1 aids in branching morphogenesis, a process that is stimulated by retinoic acid. Then, during later gestation events, Crispld2/Lgl1 appears to participate in mesenchymal maturation, which is regulated by glucocorticoids [305,306]. Retinoic acid appears to play a dichotomous role in Crispld2/Lgl1 regulation; while it stimulates branching during early gestation, retinoic acid downregulates Crispld2/Lgl1 expression that is induced by glucocorticoids during late gestation [322]. In mouse studies, retinoic acid disrupts normal pharyngeal arch development by altering NCC differentiation and patterning. As such, it is a strong teratogen, inducing cleft palate when administered at critical times in development [150,323-328]. Retinoic acid regulates gene expression of *Dlx2* and *Hoxb2*, two genes that play critical roles in proximodistal and dorsoventral NCC patterning, respectively [328-331]. In zebrafish, retinoic treatment results in the loss of Dlx2 and specific craniofacial dysmorphogenesis [328]. Loss of Dlx2 in mouse caused multiple defects in first and second pharyngeal arch derived tissues, including cleft palate, while loss of Hoxb2 does not result in orofacial clefting [330,332]. Our findings in zebrafish also show that z*cripld2* alters Dlx2 expression (Figure 7.5). This data,

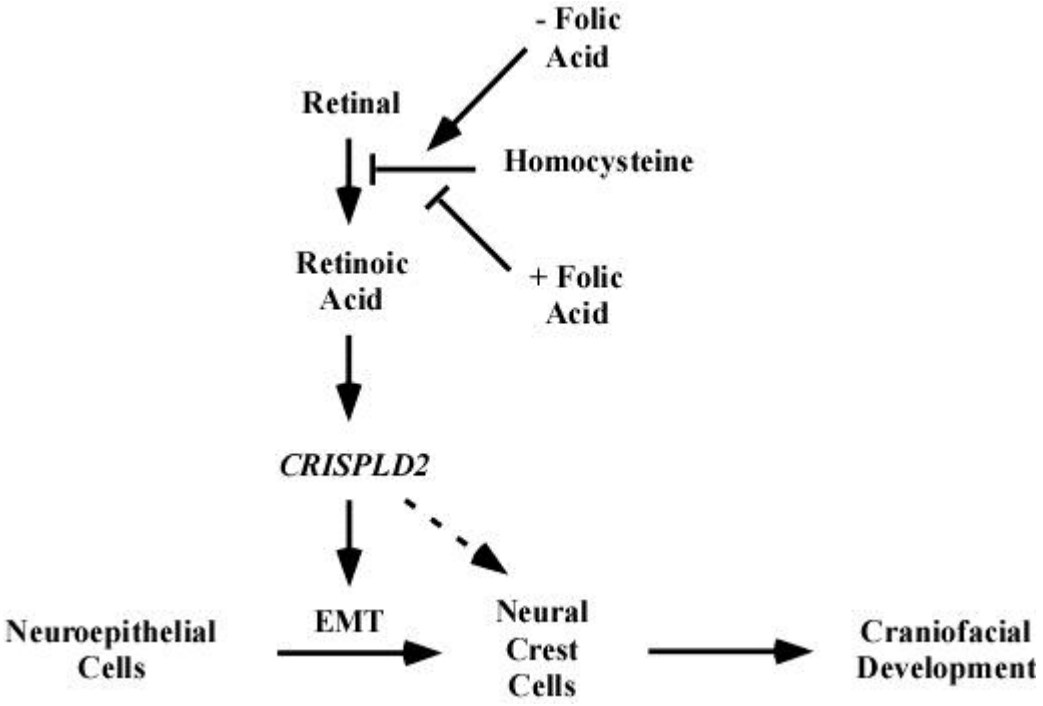
combined with those of Nadeau *et al.* in the Crispld2/Lgl1 mouse, suggests that CRISPLD2 mediates Dlx2 regulation by retinoic acid [322]. Future studies should focus on treating wild-type zebrafish embryos with retinoic acid to determine whether Crispld2 levels are altered. Crispld2 levels should decrease with retinoic acid treatment, while treatment of zebrafish embryos with a retinoic acid receptor antagonist should cause Crispld2 levels to increase. The Dlx2 expression pattern of retinoic acid treated embryos and *crispld2*MO embryos should be compared, and, if similar, wild-type z*crispld2* mRNA can be injected into embryos that are treated with retinoic acid to determine if normal Dlx2 patterning can be rescued. These studies will define the relationship between retinoic acid, Crispld2 and Dlx2.

In vivo mouse studies have shown that retinoic acid receptor null embryos had midline craniofacial anomalies in tissues of NCC origin [325]. Similar phenotypes were also seen in retinoic acid receptor null chick embryos [333]. Studies in the chick embryo have shown that the conversion of retinal to retinoic acid is inhibited by homocysteine and that chick embryos treated with homocysteine have similar defects to those embryos lacking the retinoic acid receptor [333]. Homocysteine and cysteine are endogenously produced in the methionine arm of the folate gene pathway [94,274,334]. Elevated homocysteine levels, which can be caused by low levels of folic acid, has been shown to decrease both NCC numbers and distance migrated [335,336]. Moreover, hyperhomocysteinemia, caused by reduced folate intake, has been implicated as a cause of orofacial clefting [337-341]. High doses of folic acid, six to ten times greater than what is found in most prenatal supplements, result in decreased homocysteine levels which have been shown to reduce the recurrence of orofacial clefting. Recently, we have shown that variants in genes in the folate gene pathway are associated with NSCLP, with our most significant associations found in genes in the methionine arm

[286]. Variants in five of the genes in the methionine arm interacted with variants in *CRISPLD2*, suggesting that homocysteine and genes that regulate homocysteine could also regulate CRISPLD2 and that CRISPLD2 could mediate the effects of homocysteine on NCCs [342]. This can be tested by treating wild-type zebrafish embryos ewith homocysteine to determine whether Crispld2 expression changes. If the change is identical to retinoic acid treated embryos, a relationship between homocysteine, retinoic acid and Crispld2would be established. Alternatively, Crispld2 could be regulated by additional mechanisms.

Craniofacial morphogenesis is a finely tuned process consisting of cell growth, growth factors and receptors, apoptosis and adequate nutrient supply [14]. Perturbation of any of these processes can result in craniofacial dysmorphology. The number of genes that contribute to NSCLP is growing and studies are now focusing on determining the function of these genes in craniofacial morphogenesis. Results from our studies suggest that perturbations in CRISPLD2 expression contribute to NSCLP. However, as shown in the model depicted in Figure 7.7, the process is likely to be complex; CRISPLD2, under the influence of retinoic acid, folic acid and homocysteine, directs EMT of neuroepithelial cells to become NCCs and form the craniofacies. Additionally, *CRISPLD2* may also regulate NCCs by other mechanisms. Although loss of z*crispld2* did not result in a cleft, our findings have important implications that suggest CRISPLD2 function should be further evaluated. For example, the CRISPLD2-homocysteine-retinoic acid connection needs to be further defined. Our findings are important because they demonstrate a role of CRISPLD2

Figure 7.7: Proposed model for CRISPLD2 regulation of NCCs during craniofacial development. *CRISPLD2* is required for EMT of neuroepithelial cells to NCCs. When there are insufficient levels of folic acid, homocysteine prevents the conversion of retinal to retinoic acid. Retinoic acid is an upstream regulator of *CRISPLD2*. *CRISPLD2* may also regulate NCCs by other mechanisms (dotted line).



in the formation of NCC and early craniofacial development in zebrafish e Moreover, we suggest that perturbations in genes that regulate NCCs, such as CRISPLD2, can have a detrimental effect on NCC-derived tissues and alter normal craniofacial development and thereby contribute to NSCLP.

Chapter 8: Summary and Future Studies

8.1. Summary

Cleft lip and palate, a common, complex birth disorder, affects one out of every 700 live births in Texas and approximately 4000 newborns each year in the United States. Patients with cleft lip and palate face significant healthcare challenges related to surgical, dental and speech therapies. The exact cause of this disorder is unknown, but multiple genetic and environmental influences are known to be involved. The purpose of this study was to identify genetic causes of NSCLP and to understand how these genes contribute to normal craniofacial development. The ultimate goal of these studies is to gain a better understanding of the causes of NSCLP in order to better diagnose and counsel NSCLP families and to develop potential therapeutic interventions.

As discussed in Section 1.5 (page 9), there are two methods of gene identification for genetic diseases: genome scan and candidate gene approach. In chapter three, we demonstrated the effectiveness of a genome scan to identify genomic regions that were linked to NSCLP in our dataset. Eleven regions in our NHW dataset were associated with NSCLP. Interestingly, five of these regions, including our most significant findings in the nonHispanic dataset, from chromosome 22q12.2-12.3, were novel regions not previously identified in NSCLP genome scans (see Table 3.2, page 41). Candidate gene interrogation of myosin heavy chain 9 (MYH9) gene identified an association and confirmed the association reported in an Italian NSCLP dataset [55,105]. These results demonstrate the effectiveness of genome scans in identifying chromosome regions that may harbor genes contributing to disease etiology.

In chapter four, we interrogated the WNT gene family. WNT genes are expressed in the craniofacies throughout development and knockout WNT animal models have orofacial clefts [20,21,112,114-117,233-235]. Interestingly, the WNT genes that were most significantly associated with our NSCLP dataset, Wnt3A and Wnt11, were those that function in NCC differentiation and migration [118]. NCCs play a critical role during craniofacial development and form the tissues that populate the maxillary and mandibular jaw structures [311]. These results demonstrated the effectiveness of a candidate gene approach in identifying genes that have a known role in normal development, that when perturbed, may contribute to disease etiology.

In chapters five through seven, we describe studies that identified an association of CRISPLD2 with NSCLP, CRISPLD2 had not previously been associated with NSCLP or any craniofacial anomalies. CRISPLD2 had previously been identified in mice and rats and is called Lgl1 (late gestational lung 1) [302]. Studies in mice suggest that CRISPLD2 (Lgl1) aids in branching morphogenesis during kidney and lung organogenesis (Quinlan, Nadeau). The chromosomal 16q21-24 region has previously been identified as potentially harboring a NSCLP candidate gene in four genome scans and 2 meta-analyses of genome scan data [97,99,100,103,104,262]. CRISPLD2 was identified after we confirmed that polymorphic variants in that chromosome region were associated with NSCLP. While we did not find any coding mutations in our probands with CRISPLD2 susceptible haplotypes, we did find that CRISPLD2 is expressed during mouse craniofacial morphogenesis.

As demonstrated in chapter four, a candidate gene approach can be used to assess genes in the same family. Therefore, we interrogated the CRISPLD1 gene for association with NSCLP. No significant association was identified for SNPs in the CRISPLD1 gene and

NSCLP. However, we did find that variants in CRISPLD1 and CRISPLD2 interacted together and that variants in both genes interacted with variants in genes in the folate pathway. We next used zebrafish to define expression and timing of CRISPLD2 in craniofacial morphogenesis. Zebrafish ubiquitously express CRISPLD2 throughout development from 5-7 somite stage to 5dpf; gene expression localizes to the craniofacial region by 1dpf. Knockdown of CRISPLD2 resulted in lower survival rates and a higher number of phenotypic abnormalities. At 1dpf, embryos showed altered NCC clustering which translated into abnormal cartilage in the jaw by 5dpf. While we did not see a cleft, these results suggest that CRISPLD2 is still required for normal craniofacial development in zebrafish.

8.2. Future Studies

8.2.1. Genome scan

While association was found to the MYH9 gene in chromosome 22q12.2-12.3, the results do not exclude the possibility that other genes in 22q12.2-12.3 chromosomal region play a role in NSCLP. This genome scan was only conducted on only nine of our more than 100 NHW multiplex families. While these nine families were some of the larger multiplex families, the results do not necessarily represent clefting etiology in the entire sample since NSCLP is genetically heterogeneous. Additionally, of the 11 regions identified in this study, only one region was interrogated in any depth. Additional studies are needed to follow up these initial results. First, the significant SNPs in each of the eleven regions identified in the NHW population should be tested in the entire dataset. This may narrow the regions into more defined peaks and reduce the number of genes requiring interrogation.

Additionally, candidate genes can be interrogated or all regions can be resequenced to determine which genes/variants contribute to NSCLP.

Besides nine NHW families, the genome scan was also performed on a large African American family, where we identified three other regions (2p22, 3p26, and 8q21.3-24.12), none of which showed linkage in the NHW family dataset. These results and subsequent follow-up of the chromosome 8 region, which had the highest LOD score, were not part of my studies [105]. Briefly, two candidate genes were interrogated, syndecan 2 (SDC2) and growth differentiation factor 6 (GDF6), but no association was identified [105]. However, this was not unexpected because the candidate gene analysis was performed using our NHW and Hispanic populations, as clefting is rare in the African American population. To best interrogate these three regions, additional African American families must be collected and a similar approach to the NHW studies should be performed.

8.2.2. WNT

NCCs play a critical role in normal craniofacial development; therefore, it is not surprising that variation in genes that regulate NCCs contribute to NSCLP etiology [118]. Future NSCLP genetic studies should focus on other genes that regulate NCC function, especially in light of our findings that knockdown of CRISPLD2 in zebrafish alters NCC clustering.

8.2.3. CRISPLD

As mentioned in Section 1.4 (page 7), NSCLP is a genetically heterogeneous birth defect with individual genes being shown to play a role in clefting etiology in different populations [41,42]. Interestingly, the association between CRISPLD2 and NSCLP have recently been replicated in NHW and Brazilian datasets, supporting our finding that variation in

CRISPLD2 contributes to NSCLP [314]. The gene interaction results are interesting but are based on statistical testing [202]. It is important to test whether these variants interact on a biochemical level as well. If an interaction between the folic acid pathway and CRISPLD genes is detected at the biochemical level, it would be interesting to test if decreased maternal folate levels during fetal development correlated with NSCLP-susceptible CRISPLD2 genotypes, and if so, whether these pregnant women had an increased risk of having a child with NSCLP. These studies will further help assess the role of the CRISPLD gene family in clefting etiology.

The zebrafish studies provided insight on the role of CRISPLD2 during early development. We show that CRISPLD2 plays a role in normal NCC clustering, and further research is warranted to delineate the CRISPLD2-homocysteine-RA connection and determine how these genes contribute to clefting etiology. The zebrafish model could be used to for these experiments. Zebrafish embryos can be treated with RA and/or homocysteine to determine if Dlx2 or CRISPLD2 expression is decreased or altered, as suggested by previous homocysteine-RA and CRISPLD2-RA studies in other zebrafish tissues [326-328,333].

Additional animal studies are necessary to further delineate the role of CRISPLD2 in craniofacial development because the zebrafish model is only useful as a simple vertebrate model. To do this, a transgenic mouse is an appropriate model. To accomplish this goal, a CRISPLD2 ES cell line has been purchased from the Knock Out Mouse Project (KOMP) and the conditional CRISPLD2 knockout mice will be generated. The conditional approach is being used because the CRISPLD2 traditional knockout mouse is embryonic lethal (ref). Additionally, our lab has recently performed whole mount *in situ* hybridizations on mouse

embryos that show that CRISPLD2 is expressed in the NCC clusters during early mouse craniofacial development (Yuan and Hecht, 2010 unpublished data), further supporting our findings that CRISPLD2 plays a role in normal NCC development and migration. Additional studies are warranted to determine how CRISPLD2 variation can lead to detrimental effects of NCC-derived tissues.

References

1. Kang P, Svoboda KK: Epithelial-mesenchymal transformation during craniofacial development. *J Dent Res* 2005; **84:** 678-690.

2. Kulesa PM, Schnell S, Rudloff S, Baker RE, Maini PK: From segment to somite: segmentation to epithelialization analyzed within quantitative frameworks. *Developmental dynamics : an official publication of the American Association of Anatomists* 2007; **236:** 1392-1402.

3. Gorlin RJ, M.M. Cohen, and R.C.M. Hennekam: *Syndromes of the Head and Neck,* Fourth Edition edn. New York: Oxford University Press, 2001.

4. Johnston MC, Hassell JR, Brown KS: The embryology of cleft lip and cleft palate. *Clin Plast Surg* 1975; **2:** 195-203.

5. Trasler DG, Fraser FC: *Handbook of Teratology*. New York: Plenum Press, 1977.

6. Fraser FC: The genetics of cleft lip and palate. *Am J Hum Genet* 1970; **22:** 336-352.

7. Johnston MC, Bronsky PT: Prenatal craniofacial development: new insights on normal and abnormal mechanisms. *Crit Rev Oral Biol Med* 1995; **6:** 368-422.

8. Koo SH, Cunningham MC, Arabshahi B, Gruss JS, Grant JH, 3rd: The transforming growth factor-beta 3 knock-out mouse: an animal model for cleft palate. *Plast Reconstr Surg* 2001; **108:** 938-948; discussion 949-951.

9. Lan Y, Ryan RC, Zhang Z, Bullard SA, Bush JO, Maltby KM, Lidral AC, Jiang R: Expression of Wnt9b and activation of canonical Wnt signaling during midfacial morphogenesis in mice. *Dev Dyn* 2006; **235:** 1448-1454.

10. Lidral AC, Moreno LM: Progress toward discerning the genetics of cleft lip. *Curr Opin Pediatr* 2005; **17:** 731-739.

11. Miyazaki Y, Oshima K, Fogo A, Hogan BL, Ichikawa I: Bone morphogenetic protein 4 regulates the budding site and elongation of the mouse ureter. *J Clin Invest* 2000; **105:** 863-873.

12. Murray JC: Face facts: genes, environment, and clefts. *Am J Hum Genet* 1995; **57:** 227-232.

13. Murray JC: Gene/environment causes of cleft lip and/or palate. *Clin Genet* 2002; **61:** 248-256.

14. Nanci A: *Ten Cate's Oral Histology - Development, Structure, and Function*, 6th edn. Los Angelos: Mosby, 2003.

15. Nusse R: Wnt signaling in disease and in development. *Cell Res* 2005; **15:** 28-32.

16. Robertson CP, Braun MM, Roelink H: Sonic hedgehog patterning in chick neural plate is antagonized by a Wnt3-like signal. *Dev Dyn* 2004; **229:** 510-519.

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

18. Senders CW, Peterson EC, Hendrickx AG, Cukierski MA: Development of the upper lip. *Arch Facial Plast Surg* 2003; **5:** 16-25.

19. Wyszynski D (ed): *Cleft Lip and Palate: From Origin to Treatment*. Oxford: Oxford University Press, 2002.

20. Yamaguchi TP, Bradley A, McMahon AP, Jones S: A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* 1999; **126:** 1211-1223.

21. Yamaguchi TP, Takada S, Yoshikawa Y, Wu N, McMahon AP: T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes Dev* 1999; **13:** 3185-3190.

22. Brewer C, Holloway S, Zawalnyski P, Schinzel A, FitzPatrick D: A chromosomal deletion map of human malformations. *Am J Hum Genet* 1998; **63:** 1153-1159.

23. Shi M, Mostowska A, Jugessur A, Johnson MK, Mansilla MA, Christensen K, Lie RT, Wilcox AJ, Murray JC: Identification of microdeletions in candidate genes for cleft lip and/or palate. *Birth defects research Part A, Clinical and molecular teratology* 2009; **85:** 42-51.

24. Jiang R, Bush JO, Lidral AC: Development of the upper lip: Morphogenetic and molecular mechanisms. *Dev Dyn* 2006; **235:** spc1.

25. Hinrichsen K: The early development of morphology and patterns of the face in the human embryo. *Adv Anat Embryol Cell Biol* 1985; **98:** 1-79.

26. Brewer C, Holloway S, Zawalnyski P, Schinzel A, FitzPatrick D: A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality--and tolerance of segmental aneuploidy--in humans. *Am J Hum Genet* 1999; **64:** 1702-1708.

27. Kobayashi J, Kimijima Y, Yamada S, Amagasa T, Saito-Ohara F: 4p- syndrome and 9p tetrasomy mosaicism with cleft lip and palate. *Journal of cranio-maxillo-facial surgery : official publication of the European Association for Cranio-Maxillo-Facial Surgery* 2000; **28:** 165-170.

28. Gandelman KY, Gibson L, Meyn MS, Yang-Feng TL: Molecular definition of the smallest region of deletion overlap in the Wolf-Hirschhorn syndrome. *Am J Hum Genet* 1992; **51:** 571-578.

29. Hashmi SS, Waller DK, Langlois P, Canfield M, Hecht JT: Prevalence of nonsyndromic oral clefts in Texas: 1995-1999. *Am J Med Genet A* 2005; **134:** 368-372.

30. Altemus LA: The incidence of cleft lip and palate among North American negroes. *Cleft Palate J* 1966; **3:** 357-361.

31. Bonaiti C, Briard ML, Feingold J, Pavy B, Psaume J, Migne-Tufferand G, Kaplan J: An epidemiological and genetic study of facial clefting in France. I. Epidemiology and frequency in relatives. *J Med Genet* 1982; **19:** 8-15.

32. Plotner PL, Smith JL, Northrup H: Trisomy 20q caused by der(4) t(4;20) (q35;q13.1): report of a new patient and review of the literature. *Am J Med Genet* 2002; **111:** 71-75.

33. Chung CS, Mi MP, Beechert AM: Genetic epidemiology of cleft lip with or without cleft palate in the population of Hawaii. *Genet Epidemiol* 1987; **4:** 415-423.

34. Murray JC: Face facts: genes, environment, and clefts.[comment]. *Am J Hum Genet* 1995; **57:** 227-232.

35. Wyszynski DF, Beaty TH, Maestri NE: Genetics of nonsyndromic oral clefts revisited. *Cleft Palate Craniofac J* 1996; **33:** 406-417.

36. Lidral AC, Murray JC: Genetic approaches to identify disease genes for birth defects with cleft lip/palate as a model. *Birth Defects Res A Clin Mol Teratol* 2004; **70:** 893-901.

37. Letra A, Menezes R, Granjeiro JM, Vieira AR: Defining subphenotypes for oral clefts based on dental development. *J Dent Res* 2007; **86:** 986-991.

38. Marazita ML: Subclinical features in non-syndromic cleft lip with or without cleft palate (CL/P): review of the evidence that subepithelial orbicularis oris muscle defects are part of an expanded phenotype for CL/P. *Orthodontics & craniofacial research* 2007; **10:** 82-87.

39. Menezes R, Vieira AR: Dental anomalies as part of the cleft spectrum. *The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association* 2008; **45:** 414-419.

40. Harville EW, Wilcox AJ, Lie RT, Vindenes H, Abyholm F: Cleft lip and palate versus cleft lip only: are they distinct defects? *Am J Epidemiol* 2005; **162:** 448-453.

41. Maestri NE, Beaty TH, Hetmanski J, Smith EA, McIntosh I, Wyszynski DF, Liang KY, Duffy DL, VanderKolk C: Application of transmission disequilibrium tests to nonsyndromic oral clefts: including candidate genes and environmental exposures in the models. *Am J Med Genet* 1997; **73:** 337-344.

42. Schliekelman P, Slatkin M: Multiplex relative risk and estimation of the number of loci underlying an inherited disease. *Am J Hum Genet* 2002; **71:** 1369-1385.

43. Wyszynski DF, Duffy DL, Beaty TH: Maternal cigarette smoking and oral clefts: a meta-analysis. *Cleft Palate Craniofac J* 1997; **34:** 206-210.

44. Shi M, Wehby GL, Murray JC: Review on genetic variants and maternal smoking in the etiology of oral clefts and other birth defects. *Birth defects research Part C, Embryo today : reviews* 2008; **84:** 16-29.

45. Mitchell LE, Risch N: Mode of inheritance of nonsyndromic cleft lip with or without cleft palate: a reanalysis. *Am J Hum Genet* 1992; **51:** 323-332.

46. Blanton SH, Cortez A, Stal S, Mulliken JB, Finnell RH, Hecht JT: Variation in IRF6 contributes to nonsyndromic cleft lip and palate. *Am J Med Genet A* 2005; **137:** 259-262.

47. Scapoli L, Palmieri A, Martinelli M, Pezzetti F, Carinci P, Tognon M, Carinci F: Strong evidence of linkage disequilibrium between polymorphisms at the IRF6 locus and nonsyndromic cleft lip with or without cleft palate, in an Italian population. *Am J Hum Genet* 2005; **76:** 180-183.

48. Zucchero TM, Cooper ME, Maher BS, Daack-Hirsch S, B. N, Ribeiro L, Caprau D, Christensen K, Suzuki Y, Machida J, Natsume N, Yoshiura K, Vieira AR, Orioli IM, Castilla EE, Moreno L, Arcos-Burgos M, Lidral AC, Field LL, Liu Y, Ray A, Goldstein TH, Schultz RE, Shi M, Kondo S, Schutte BC, Marazita M, Murray J: Interferon Regulatory Factor 6 (IRF6) is a Modifier for Isolated Cleft Lip and Palate. *Am J Human Genet* 2003; **73:** 162.

49. Zucchero TM, Cooper ME, Maher BS, Daack-Hirsch S, Nepomuceno B, Ribeiro L, Caprau D, Christensen K, Suzuki Y, Machida J, Natsume N, Yoshiura K, Vieira AR, Orioli IM, Castilla EE, Moreno L, Arcos-Burgos M, Lidral AC, Field LL, Liu YE, Ray A, Goldstein TH, Schultz RE, Shi M, Johnson MK, Kondo S, Schutte BC, Marazita ML, Murray JC: Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. *N Engl J Med* 2004; **351:** 769-780.

50. Shaw D, Ray AK, Marazita ML, Field L: Further evidence of a relationship between the retinoic acid receptor alpha locus and nonsyndromic cleft lip with or without cleft palate (CL+ P). *Am J Hum Genet* 1993; **53:** 1156-1157.

51. Suzuki Y, Jezewski PA, Machida J, Watanabe Y, Shi M, Cooper ME, Viet le T, Nguyen TD, Hai H, Natsume N, Shimozato K, Marazita ML, Murray JC: In a Vietnamese population, MSX1 variants contribute to cleft lip and palate. *Genet Med* 2004; **6:** 117-125.

52. Amos C, Gasser D, Hecht JT: Nonsyndromic cleft lip with or without cleft palate: new BCL3 information. *Am J Hum Genet* 1996; **59:** 743-744.

53. Stein J, Mulliken JB, Stal S, Gasser DL, Malcolm S, Winter R, Blanton SH, Amos C, Seemanova E, Hecht JT: Nonsyndromic cleft lip with or without cleft palate: evidence of linkage to BCL3 in 17 multigenerational families. *Am J Hum Genet* 1995; **57:** 257-272.

54. Marazita ML, Mooney MP: Current concepts in the embryology and genetics of cleft lip and cleft palate. *Clin Plast Surg* 2004; **31:** 125-140.

55. Martinelli M, Di Stazio M, Scapoli L, Marchesini J, Di Bari F, Pezzetti F, Carinci F, Palmieri A, Carinci P, Savoia A: Cleft lip with or without cleft palate: implication of the heavy chain of non-muscle myosin IIA. *J Med Genet* 2007.

56. Artama M, Auvinen A, Raudaskoski T, Isojarvi I, Isojarvi J: Antiepileptic drug use of women with epilepsy and congenital malformations in offspring. *Neurology* 2005; **64:** 1874-1878.

57. Azarbayjani F, Danielsson BR: Phenytoin-induced cleft palate: evidence for embryonic cardiac bradyarrhythmia due to inhibition of delayed rectifier K+ channels resulting in hypoxia-reoxygenation damage. *Teratology* 2001; **63:** 152-160.

58. Hernandez-Diaz S, Werler MM, Walker AM, Mitchell AA: Folic acid antagonists during pregnancy and the risk of birth defects. *N Engl J Med* 2000; **343:** 1608-1614.

59. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet* 1991; **338:** 131-137.

60. Berry RJ, Li Z, Erickson JD, Li S, Moore CA, Wang H, Mulinare J, Zhao P, Wong LY, Gindler J, Hong SX, Correa A: Prevention of neural-tube defects with folic acid in China. China-U.S. Collaborative Project for Neural Tube Defect Prevention. *N Engl J Med* 1999; **341:** 1485-1490.

61. Ray JG, Meier C, Vermeulen MJ, Wyatt PR, Cole DE: Association between folic acid food fortification and congenital orofacial clefts. *J Pediatr* 2003; **143:** 805-807.

62. Czeizel AE, Tâimâar L, Sâarkèozi A: Dose-dependent effect of folic acid on the prevention of orofacial clefts. *Pediatrics* 1999; **104:** e66.

63. Tolarova M, Harris J: Reduced recurrence of orofacial clefts after periconceptual supplementation with high dose folic acid and multivitamins. *Teratology* 1995; **51:** 71-78.

64. Blanton SH, Kolle BS, Hecht JT, Mulliken JB, Martin ER: No evidence supporting MTHFR as a risk factor in the development of familial NSCLP. *Am J Med Genet* 2000; **92:** 370-371.

65. Hecht JT, Patel S, Mulliken JB, Blanton SH: MTHFR is not a risk factor for the development of isolated NSCLP. *Am J Med Genet* 2002; **110:** 404-405.

66. Shaw GM, Lammer EJ, Zhu H, Baker MW, Neri E, Finnell RH: Maternal periconceptional vitamin use, genetic variation of infant reduced folate carrier (A80G), and risk of spina bifida. *Am J Med Genet* 2002; **108:** 1-6.

67. Shaw GM, Rozen R, Finnell RH, Todoroff K, Lammer EJ: Infant C677T mutation in MTHFR, maternal periconceptional vitamin use, and cleft lip. *Am J Med Genet* 1998; **80:** 196-198.

68. Shaw GM, Todoroff K, Finnell RH, Rozen R, Lammer EJ: Maternal vitamin use, infant C677T mutation in MTHFR, and isolated cleft palate risk. *Am J Med Genet* 1999; **85:** 84-85.

69. Shaw GM, Wasserman CR, Lammer EJ, O'Malley CD, Murray JC, Basart AM, Tolarova MM: Orofacial clefts, parental cigarette smoking, and transforming growth factor-alpha gene variants. *Am J Hum Genet* 1996; **58:** 551-561.

70. Hartsfield JK, Jr., Hickman TA, Everett ET, Shaw GM, Lammer EJ, Finnell RA: Analysis of the EPHX1 113 polymorphism and GSTM1 homozygous null polymorphism and oral clefting associated with maternal smoking. *Am J Med Genet* 2001; **102:** 21-24.

71. van Rooij IA, Wegerif MJ, Roelofs HM, Peters WH, Kuijpers-Jagtman AM, Zielhuis GA, Merkus HM, Steegers-Theunissen RP: Smoking, genetic polymorphisms in biotransformation enzymes, and nonsyndromic oral clefting: a gene-environment interaction. *Epidemiology* 2001; **12:** 502-507.

72. Dejmek J, Solansk y I, Podrazilovâa K, Srâam RJ: The exposure of nonsmoking and smoking mothers to environmental tobacco smoke during different gestational phases and fetal growth. *Environ Health Perspect* 2002; **110:** 601-606.

73. Beaty TH, Hetmanski JB, Zeiger JS, Fan YT, Liang KY, VanderKolk CA, McIntosh I: Testing candidate genes for non-syndromic oral clefts using a case-parent trio design. *Genet Epidemiol* 2002; **22:** 1-11.

74. Christensen K, Olsen J, Norgaard-Pedersen B, Basso O, Stovring H, Milhollin-Johnson L, Murray JC: Oral clefts, transforming growth factor alpha gene variants, and maternal smoking: a population-based case-control study in Denmark, 1991-1994. *Am J Epidemiol* 1999; **149:** 248-255.

75. Mitchell LE, Murray JC, O'Brien S, Christensen K: Evaluation of two putative susceptibility loci for oral clefts in the Danish population. *Am J Epidemiol* 2001; **153:** 1007-1015.

76. Romitti PA, Lidral AC, Munger RG, Daack-Hirsch S, Burns TL, Murray JC: Candidate genes for nonsyndromic cleft lip and palate and maternal cigarette smoking and alcohol consumption: evaluation of genotype-environment interactions from a population-based case-control study of orofacial clefts. *Teratology* 1999; **59:** 39-50.

77. Shi M, Christensen K, Weinberg CR, Romitti P, Bathum L, Lozada A, Morris RW, Lovett M, Murray JC: Orofacial cleft risk is increased with maternal smoking and specific detoxification-gene variants. *Am J Hum Genet* 2007; **80:** 76-90.

78. Ramirez D, Lammer EJ, Iovannisci DM, Laurent C, Finnell RH, Shaw GM: Maternal smoking during early pregnancy, GSTP1 and EPHX1 variants, and risk of isolated orofacial clefts. *The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association* 2007; **44:** 366-373.

79. Lammer EJ, Shaw GM, Iovannisci DM, Finnell RH: Maternal smoking, genetic variation of glutathione s-transferases, and risk for orofacial clefts. *Epidemiology* 2005; **16:** 698-701.

80. Beaty TH, Maestri NE, Hetmanski JB, Wyszynski DF, Vanderkolk CA, Simpson JC, McIntosh I, Smith EA, Zeiger JS, Raymond GV, Panny SR, Tifft CJ, Lewanda AF, Cristion CA, Wulfsberg EA: Testing for interaction between maternal smoking and TGFA genotype among oral cleft cases born in Maryland 1992-1996. *Cleft Palate Craniofac J* 1997; **34:** 447-454.

81. Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP: A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* 1995; **10:** 111-113.

82. Molloy AM, Daly S, Mills JL, Kirke PN, Whitehead AS, Ramsbottom D, Conley MR, Weir DG, Scott JM: Thermolabile variant of 5,10-methylenetetrahydrofolate

reductase associated with low red-cell folates: implications for folate intake recommendations. *Lancet* 1997; **349:** 1591-1593.

83. van der Put NM, Gabreels F, Stevens EM, Smeitink JA, Trijbels FJ, Eskes TK, van den Heuvel LP, Blom HJ: A second common mutation in the methylenetetrahydrofolate reductase gene: an additional risk factor for neural-tube defects? *Am J Hum Genet* 1998; **62:** 1044-1051.

84. van Rooij IA, Vermeij-Keers C, Kluijtmans LA, Ocke MC, Zielhuis GA, Goorhuis-Brouwer SM, van der Biezen JJ, Kuijpers-Jagtman AM, Steegers-Theunissen RP: Does the interaction between maternal folate intake and the methylenetetrahydrofolate reductase polymorphisms affect the risk of cleft lip with or without cleft palate? *Am J Epidemiol* 2003; **157:** 583-591.

85. Martinelli M, Scapoli L, Pezzetti F, Carinci F, Carinci P, Stabellini G, Bisceglia L, Gombos F, Tognon M: C677T variant form at the MTHFR gene and CL/P: a risk factor for mothers? *Am J Med Genet* 2001; **98:** 357-360.

86. Wyszynski DF, Diehl SR: Infant C677T mutation in MTHFR, maternal periconceptional vitamin use, and risk of nonsyndromic cleft lip. *Am J Med Genet* 2000; **92:** 79-80.

87. Gaspar DA, Pavanello RC, Zatz M, Passos-Bueno MR, Andre M, Steman S, Wyszynski DF, Matiolli SR: Role of the C677T polymorphism at the MTHFR gene on risk to nonsyndromic cleft lip with/without cleft palate: results from a case-control study in Brazil. *Am J Med Genet* 1999; **87:** 197-199.

88. Jugessur A, Wilcox AJ, Lie RT, Murray JC, Taylor JA, Ulvik A, Drevon CA, Vindenes HA, Abyholm FE: Exploring the effects of methylenetetrahydrofolate reductase gene variants C677T and A1298C on the risk of orofacial clefts in 261 Norwegian case-parent triads. *Am J Epidemiol* 2003; **157:** 1083-1091.

89. Prescott NJ, Winter RM, Malcolm S: Maternal MTHFR genotype contributes to the risk of non-syndromic cleft lip and palate. *J Med Genet* 2002; **39:** 368-369.

90. Shotelersuk V, Ittiwut C, Siriwan P, Angspatt A: Maternal 677CT/1298AC genotype of the MTHFR gene as a risk factor for cleft lip. *J Med Genet* 2003; **40:** e64.

91. Gaspar DA, Pavanello RC, Zatz M, Passos-Bueno MR, Andre M, Steman S, Wyszynski DF, Matiolli SR: Role of the C677T polymorphism at the MTHFR gene on risk to nonsyndromic cleft lip with/without cleft palate: results from a case-control study in Brazil. *Am J Med Genet* 1999; **87:** 197-199.

92. Pezzetti F, Martinelli M, Scapoli L, Carinci F, Palmieri A, Marchesini J, Carinci P, Caramelli E, Rullo R, Gombos F, Tognon M: Maternal MTHFR variant forms

increase the risk in offspring of isolated nonsyndromic cleft lip with or without cleft palate. *Hum Mutat* 2004; **24:** 104-105.

93. Zhu J, Ren A, Hao L, Pei L, Liu J, Zhu H, Li S, Finnell RH, Li Z: Variable contribution of the MTHFR C677T polymorphism to non-syndromic cleft lip and palate risk in China. *American Journal of Medical Genetics Part A* 2006; **140:** 551-557.

94. Boyles AL, Wilcox AJ, Taylor JA, Meyer K, Fredriksen A, Ueland PM, Drevon CA, Vollset SE, Lie RT: Folate and one-carbon metabolism gene polymorphisms and their associations with oral facial clefts. *Am J Med Genet A* 2008; **146A:** 440-449.

95. Blanton SH, Patel S, Hecht JT, Mulliken JB: MTHFR is not a risk factor in the development of isolated nonsyndromic cleft lip and palate. *Am J Med Genet* 2002; **110:** 404-405.

96. van Rooij IA, Vermeij-Keers C, Kluijtmans LA, Ocke MC, Zielhuis GA, Goorhuis-Brouwer SM, van der Biezen JJ, Kuijpers-Jagtman AM, Steegers-Theunissen RP: Does the interaction between maternal folate intake and the methylenetetrahydrofolate reductase polymorphisms affect the risk of cleft lip with or without cleft palate? *Am J Epidemiol* 2003; **157:** 583-591.

97. Field LL, Ray AK, Cooper ME, Goldstein T, Shaw DF, Marazita ML: Genome scan for loci involved in nonsyndromic cleft lip with or without cleft palate in families from West Bengal, India. *Am J Med Genet A* 2004; **130:** 265-271.

98. Marazita ML, Field LL, Cooper ME, Tobias R, Maher BS, Peanchitlertkajorn S, Liu YE: Genome scan for loci involved in cleft lip with or without cleft palate in Chinese multiplex families. *Am J Hum Genet* 2002; **71:** 349-364.

99. Marazita ML, Field LL, Tuncbilek G, Cooper ME, Goldstein T, Gursu KG: Genome-scan for loci involved in cleft lip with or without cleft palate in consanguineous families from Turkey. *Am J Med Genet A* 2004; **126:** 111-122.

100. Marazita ML, Murray JC, Lidral AC, Arcos-Burgos M, Cooper ME, Goldstein T, Maher BS, Daack-Hirsch S, Schultz R, Mansilla MA, Field LL, Liu YE, Prescott N, Malcolm S, Winter R, Ray A, Moreno L, Valencia C, Neiswanger K, Wyszynski DF, Bailey-Wilson JE, Albacha-Hejazi H, Beaty TH, McIntosh I, Hetmanski JB, Tuncbilek G, Edwards M, Harkin L, Scott R, Roddick LG: Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32-35. *Am J Hum Genet* 2004; **75:** 161-173.

101. Prescott NJ, Lees MM, Winter RM, Malcolm S: Identification of susceptibility loci for nonsyndromic cleft lip with or without cleft palate in a two stage genome scan of affected sib-pairs. *Hum Genet* 2000; **106:** 345-350.

102. Radhakrishna U, Ratnamala U, Gaines M, Beiraghi S, Hutchings D, Golla J, Husain SA, Gambhir PS, Sheth JJ, Sheth FJ, Chetan GK, Naveed M, Solanki JV, Patel UC, Master DC, Memon R, Antonarakis GS, Antonarakis SE, Nath SK: Genomewide scan for nonsyndromic cleft lip and palate in multigenerational Indian families reveals significant evidence of linkage at 13q33.1-34. *Am J Hum Genet* 2006; **79:** 580-585.

103. Wyszynski DF, Albacha-Hejazi H, Aldirani M, Hammod M, Shkair H, Karam A, Alashkar J, Holmes TN, Pugh EW, Doheny KF, McIntosh I, Beaty TH, Bailey-Wilson JE: A genome-wide scan for loci predisposing to non-syndromic cleft lip with or without cleft palate in two large Syrian families. *Am J Med Genet* 2003; **123A:** 140-147.

104. Schulz RE, Cooper ME, Daack-Hirsch S, Nepomucena B, Graf KA, O'Brien EK, Narazita ML, Murray JC: A targeted scan for loci linked to nonsyndromic cleft lip and palate in Filipino families. *Am J Med Genet* 2003; **10:** 1-5.

105. Chiquet BT, Hashmi SS, Henry R, Burt A, Mulliken JB, Stal S, Bray M, Blanton SH, Hecht JT: Genomic screening identifies novel linkages and provides further evidence for a role of MYH9 in nonsyndromic cleft lip and palate. *Eur J Hum Genet* 2009; **17:** 195-204.

106. Blanton SH, Bertin T, Patel S, Stal S, Mulliken JB, Hecht JT: Nonsyndromic cleft lip and palate: four chromosomal regions of interest. *Am J Med Genet A* 2004; **125:** 28-37.

107. Riley BM, Schultz RE, Cooper ME, Goldstein-McHenry T, Daack-Hirsch S, Lee KT, Dragan E, Vieira AR, Lidral AC, Marazita ML, Murray JC: A genome-wide linkage scan for cleft lip and cleft palate identifies a novel locus on 8p11-23. *Am J Med Genet A* 2007; **143:** 846-852.

108. Riley BM, Mansilla MA, Ma J, Daack-Hirsch S, Maher BS, Raffensperger LM, Russo ET, Vieira AR, Dode C, Mohammadi M, Marazita ML, Murray JC: Impaired FGF signaling contributes to cleft lip and palate. *Proc Natl Acad Sci U S A* 2007; **104:** 4512-4517.

109. Riley BM, Murray JC: Sequence evaluation of FGF and FGFR gene conserved non-coding elements in non-syndromic cleft lip and palate cases. *Am J Med Genet A* 2007.

110. Rice R, Spencer-Dene B, Connor EC, Gritli-Linde A, McMahon AP, Dickson C, Thesleff I, Rice DP: Disruption of Fgf10/Fgfr2b-coordinated epithelial-mesenchymal interactions causes cleft palate. *J Clin Invest* 2004; **113:** 1692-1700.

111. Nie X, Luukko K, Kettunen P: FGF signalling in craniofacial development and developmental disorders. *Oral Dis* 2006; **12:** 102-111.

112. Cadigan KM, Nusse R: Wnt signaling: a common theme in animal development. *Genes Dev* 1997; **11:** 3286-3305.

113. Logan CY, Nusse R: The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* 2004; **20:** 781-810.

114. De Calisto J, Araya C, Marchant L, Riaz CF, Mayor R: Essential role of non-canonical Wnt signalling in neural crest migration. *Development* 2005; **132:** 2587-2597.

115. Lewis JL, Bonner J, Modrell M, Ragland JW, Moon RT, Dorsky RI, Raible DW: Reiterated Wnt signaling during zebrafish neural crest development. *Development* 2004; **131:** 1299-1308.

116. Juriloff DM, Harris MJ, McMahon AP, Carroll TJ, Lidral AC: Wnt9b is the mutated gene involved in multifactorial nonsyndromic cleft lip with or without cleft palate in A/WySn mice, as confirmed by a genetic complementation test. *Birth Defects Res A Clin Mol Teratol* 2006; **76:** 574-579.

117. Brugmann SA, Goodnough LH, Gregorieff A, Leucht P, ten Berge D, Fuerer C, Clevers H, Nusse R, Helms JA: Wnt signaling mediates regional specification in the vertebrate face. *Development* 2007; **134:** 3283-3295.

118. Chiquet BT, Blanton SH, Burt A, Ma D, Stal S, Mulliken JB, Hecht JT: Variation in WNT genes is associated with non-syndromic cleft lip with or without cleft palate. *Hum Mol Genet* 2008; **17:** 2212-2218.

119. Vijayaragavan K, Szabo E, Bossâe M, Ramos-Mejia V, Moon RT, Bhatia M: Noncanonical Wnt signaling orchestrates early developmental events toward hematopoietic cell fate from human embryonic stem cells. *Cell stem cell* 2009; **4:** 248-262.

120. Geetha-Loganathan P, Nimmagadda S, Antoni L, Fu K, Whiting CJ, Francis-West P, Richman JM: Expression of WNT signalling pathway genes during chicken craniofacial development. *Dev Dyn* 2009; **238:** 1150-1165.

121. Lidral AC, Murray JC, Buetow KH, Basart AM, Schearer H, Shiang R, Naval A, Layda E, Magee K, Magee W: Studies of the candidate genes TGFB2, MSX1, TGFA, and TGFB3 in the etiology of cleft lip and palate in the Philippines. *Cleft Palate Craniofac J* 1997; **34:** 1-6.

122. Shiang R, Lidral AC, Ardinger HH, Buetow KH, Romitti PA, Munger RG, Murray JC: Association of transforming growth-factor alpha gene polymorphisms with nonsyndromic cleft palate only (CPO). *Am J Hum Genet* 1993; **53:** 836-843.

123. Vieira AR, Meira R, Modesto A, Murray JC: MSX1, PAX9, and TGFA contribute to tooth agenesis in humans. *J Dent Res* 2004; **83:** 723-727.

124. Lidral AC, Romitti PA, Basart AM, Doetschman T, Leysens NJ, Daack-Hirsch S, Semina EV, Johnson LR, Machida J, Burds A, Parnell TJ, Rubenstein JL, Murray JC: Association of MSX1 and TGFB3 with nonsyndromic clefting in humans. *Am J Hum Genet* 1998; **63:** 557-568.

125. Jezewski P, Vierira A, Schultz R, Machida J, Suzuki K, Liudwig B, Daack-Hirsch S: Mutations in MSX1 are associated with non-syndromic orofacial clefting. *Am J Hum Genet* 2001; **69:** 558.

126. Vieira AR, Avila JR, Daack-Hirsch S, Dragan E, Felix TM, Rahimov F, Harrington J, Schultz RR, Watanabe Y, Johnson M, Fang J, O'Brien SE, Orioli IM, Castilla EE, Fitzpatrick DR, Jiang R, Marazita ML, Murray JC: Medical sequencing of candidate genes for nonsyndromic cleft lip and palate. *PLoS Genet* 2005; **1:** e64.

127. Tongkobpetch S, Siriwan P, Shotelersuk V: MSX1 mutations contribute to nonsyndromic cleft lip in a Thai population. *J Hum Genet* 2006; **51:** 671-676.

128. Park J, Park BY, Kim HS, Lee JE, Suh I, Nam CM, Kang DR, Kim S, Yun JE, Go EN, Jee SH, Beaty TH: MSX1 polymorphism associated with risk of oral cleft in Korea: evidence from case-parent trio and case-control studies. *Yonsei Med J* 2007; **48:** 101-108.

129. Cox TC: Taking it to the max: the genetic and developmental mechanisms coordinating midfacial morphogenesis and dysmorphology. *Clin Genet* 2004; **65:** 163-176.

130. Tissier-Seta JP, Mucchielli ML, Mark M, Mattei MG, Goridis C, Brunet JF: Barx1, a new mouse homeodomain transcription factor expressed in cranio-facial ectomesenchyme and the stomach. *Mech Dev* 1995; **51:** 3-15.

131. Hu D, Helms JA: The role of sonic hedgehog in normal and abnormal craniofacial morphogenesis. *Development* 1999; **126:** 4873-4884.

132. Orioli IM, Vieira AR, Castilla EE, Ming JE, Muenke M: Mutational analysis of the Sonic Hedgehog gene in 220 newborns with oral clefts in a South American (ECLAMC) population. *Am J Med Genet* 2002; **108:** 12-15.

133. Online Mendelian Inheritance in Man OT: The Human Genome Data Base Project, Johns Hopkins University, Baltimore, MD World Wide Web<URL:http//gdbwww.gdb.org/omim/docs/omimtop.html>,. 1995.

134. Item T, Thurnher, Yerit, Sinko, Wittwer, Adeyemo, Frei, Erginel-Unaltuna, Watzinger: Van Der Woude syndrome: variable penetrance of a novel mutation (p.Arg84Gly) of the IRF6 gene in a Turkish family. *Int J Mol Med* 2005; **15:** 247-251.

135. Kayano S, Kure S, Suzuki Y, Kanno K, Aoki Y, Kondo S, Schutte BC, Murray JC, Yamada A, Matsubara Y: Novel IRF6 mutations in Japanese patients with Van der Woude syndrome: two missense mutations (R45Q and P396S) and a 17-kb deletion. *J Hum Genet* 2003; **48:** 622-628.

136. Kim Y, Park JY, Lee TJ, Yoo HW: Identification of two novel mutations of IRF6 in Korean families affected with Van der Woude syndrome. *Int J Mol Med* 2003; **12:** 465-458.

137. Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL, Daack-Hirsch S, Sander A, McDonald-McGinn DM, Zackai EH, Lammer EJ, Aylsworth AS, Ardinger HH, Lidral AC, Pober BR, Moreno L, Arcos-Burgos M, Valencia C, Houdayer C, Bahuau M, Moretti-Ferreira D, Richieri-Costa

A, Dixon MJ, Murray JC: Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nat Genet* 2002; **32:** 285-289.

138. Peyrard-Janvid P, Koillinen, Larsson, Frannson, Rautio, Hukki, Larson, Karsten, Kere: Novel and de novo mutations of the IRF6 gene detected in patients with Van der Woude or popliteal pterygium syndrome. *Eur J Hum Genet* 2005; **13:** 1261-1267.

139. Schutte BC, Basart AM, Watanabe Y, Laffin JJ, Coppage K, Bjork BC, Daack-Hirsch S, Patil S, Dixon MJ, Murray JC: Microdeletions at chromosome bands 1q32-q41 as a cause of Van der Woude syndrome. *Am J Med Genet* 1999; **84:** 145-150.

140. Shotelersuk V, Srichomthong C, Yoshiura K, Niikawa N: A novel mutation, 1234del(C), of the IRF6 in a Thai family with Van der Woude syndrome. *Int J Mol Med* 2003; **11:** 505-507.

141. Wang LJ, Zhang H, Xiao M, Li J, Yang C, Lin X, Wu Z, Hu L, Kong X: Novel mutations in the IRF6 gene for Van der Woude syndrome. *Hum Genet* 2003; **113:** 382-386.

142. Ye J, Shi, Fan, Song, Fan, Bian: Identification of novel mutations of IRF6 gene in Chinese families with Van der Woude syndrome. *Int J Mol Med* 2005; **16:** 851-856.

143. Burdick AB, Bixler D, Puckett CL: Genetic analysis in families with van der Woude syndrome. *J Craniofac Genet Dev Biol* 1985; **5:** 181-208.

144. Woolf CM: Congenital cleft lip: A genetic study of 496 propositi. *J Med Genet* 1971; **8:** :65-72.

145. Ghassibe M, Bayet B, Revencu N, Verellen-Dumoulin C, Gillerot Y, Vanwijck R, Vikkula M: Interferon regulatory factor-6: a gene predisposing to isolated cleft lip with or without cleft palate in the Belgian population. *Eur J Hum Genet* 2005; **13:** 1239-1242.

146. Srichomthong C, Siriwan P, Shotelersuk V: Significant association between IRF6 820G->A and non-syndromic cleft lip with or without cleft palate in the Thai population. *J Med Genet* 2005; **42:** e46.

147. Ingraham CR, Kinoshita A, Kondo S, Yang B, Sajan S, Trout KJ, Malik MI, Dunnwald M, Goudy SL, Lovett M, Murray JC, Schutte BC: Abnormal skin, limb and craniofacial morphogenesis in mice deficient for interferon regulatory factor 6 (Irf6). *Nat Genet* 2006; **38:** 1335-1340.

148. Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, Domann FE, Govil M, Christensen K, Bille C, Melbye M, Jugessur A, Lie RT, Wilcox AJ, Fitzpatrick DR, Green ED, Mossey PA, Little J, Steegers-Theunissen RP,

Pennacchio LA, Schutte BC, Murray JC: Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nat Genet* 2008; **40:** 1341-1347.

149. Richardson RJ, Dixon J, Malhotra S, Hardman MJ, Knowles L, Boot-Handford RP, Shore P, Whitmarsh A, Dixon MJ: Irf6 is a key determinant of the keratinocyte proliferation-differentiation switch. *Nat Genet* 2006; **38:** 1329-1334.

150. Johnston MC, Bronsky PT: Animal models for human craniofacial malformations. *J Craniofac Genet Dev Biol* 1991; **11:** 277-291.

151. Juriloff DM, Harris MJ: Mouse genetic models of cleft lip with or without cleft palate. *Birth defects research Part A, Clinical and molecular teratology* 2008; **82:** 63-77.

152. Brown SD: Mouse models of genetic disease: new approaches, new paradigms. *J Inherit Metab Dis* 1998; **21:** 532-539.

153. Nagao T, Fujikawa K: Male-mediated teratogenesis: spectrum of congenital malformations in the offspring of A/J male mice treated with ethylnitrosourea. *Teratog Carcinog Mutagen* 1996; **16:** 301-305.

154. Yamada T, Fujiwara K, Mishima K, Sugahara T: Effect of ENU (ethylnitrosourea) mutagenesis in cleft lip and/or palate pathogenesis in mice. *Int J Oral Maxillofac Surg* 2005; **34:** 74-77.

155. Juriloff DM: Differences in frequency of cleft lip among the A strains of mice. *Teratology* 1982; **25:** 361-368.

156. Kalter H: The history of the A family of inbred mice and the biology of its congenital malformations. *Teratology* 1979; **20:** 213-232.

157. Gong SG: Phenotypic and molecular analyses of A/WySn mice. *The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association* 2001; **38:** 486-491.

158. Juriloff DM, Harris MJ, Dewell SL: A digenic cause of cleft lip in A-strain mice and definition of candidate genes for the two loci. *Birth Defects Res A Clin Mol Teratol* 2004; **70:** 509-518.

159. Juriloff DM, Harris MJ, Dewell SL, Brown CJ, Mager DL, Gagnier L, Mah DG: Investigations of the genomic region that contains the clf1 mutation, a causal gene in multifactorial cleft lip and palate in mice. *Birth Defects Res A Clin Mol Teratol* 2005; **73:** 103-113.

160. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: The human genome browser at UCSC. *Genome Res* 2002; **12:** 996-1006.

161. Fleck MW, Hirotsune S, Gambello MJ, Phillips-Tansey E, Suares G, Mervis RF, Wynshaw-Boris A, McBain CJ: Hippocampal abnormalities and enhanced excitability in a murine model of human lissencephaly. *J Neurosci* 2000; **20:** 2439-2450.

162. Zhu H, Wlodarczyk BJ, Scott M, Yu W, Merriweather M, Gelineau-van Waes J, Schwartz RJ, Finnell RH: Cardiovascular abnormalities in Folr1 knockout mice and folate rescue. *Birth Defects Research* 2007; **79:** 257-268.

163. Matsuzawa N, Yoshiura K, Machida J, Nakamura T, Niimi T, Furukawa H, Toyoda T, Natsume N, Shimozato K, Niikawa N: Two missense mutations in the IRF6 gene in two Japanese families with Van der Woude syndrome. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology & Endodontics* 2004; **98:** 414-417.

164. Koillinen H, Wong FK, Rautio J, Ollikainen V, Karsten A, Larson O, Teh BT, Huggare J, Lahermo P, Larsson C, Kere J: Mapping of the second locus for the Van der Woude syndrome to chromosome 1p34. *Eur J Hum Genet* 2001; **9:** 747-752.

165. Hamachi T, Sasaki Y, Hidaka K, Nakata M: Association between palatal morphogenesis and Pax9 expression pattern in CL/Fr embryos with clefting during palatal development. *Arch Oral Biol* 2003; **48:** 581-587.

166. Peters H, Neubuser A, Kratochwil K, Balling R: Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes Dev* 1998; **12:** 2735-2747.

167. Jugessur A, Murray JC: Orofacial clefting: recent insights into a complex trait. *Curr Opin Genet Dev* 2005; **15:** 270-278.

168. Flores MV, Lam EY, Crosier P, Crosier K: A hierarchy of Runx transcription factors modulate the onset of chondrogenesis in craniofacial endochondral bones in zebrafish. *Dev Dyn* 2006; **235:** 3166-3176.

169. Kaji T, Artinger KB: dlx3b and dlx4b function in the development of Rohon-Beard sensory neurons and trigeminal placode in the zebrafish neurula. *Dev Biol* 2004; **276:** 523-540.

170. Yan YL, Willoughby J, Liu D, Crump JG, Wilson C, Miller CT, Singer A, Kimmel C, Westerfield M, Postlethwait JH: A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. *Development* 2005; **132:** 1069-1083.

171. Gato A, Martinez ML, Tudela C, Alonso I, Moro JA, Formoso MA, Ferguson MW, Martâinez-Alvarez C: TGF-beta(3)-induced chondroitin sulphate proteoglycan mediates palatal shelf adhesion. *Dev Biol* 2002; **250:** 393-405.

172. Ashique AM, Fu K, Richman JM: Endogenous bone morphogenetic proteins regulate outgrowth and epithelial survival during avian lip fusion. *Development* 2002; **129:** 4647-4660.

173. Freimer N, Sabatti C: The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology. *Nat Genet* 2004; **36:** 1045-1051.

174. Burton PR, Tobin MD, Hopper JL: Key concepts in genetic epidemiology. *Lancet* 2005; **366:** 941-951.

175. Lehninger AL: *Biochemistry : the molecular basis of cell structure and function*, 2d edn. New York: Worth Publishers, 1975.

176. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM: A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 2007; **315:** 525-528.

177. Vidyarani M, Selvaraj P, Prabhu Anand S, Jawahar MS, Adhilakshmi AR, Narayanan PR: Interferon gamma (IFNgamma) & interleukin-4 (IL-4) gene variants & cytokine levels in pulmonary tuberculosis.[see comment]. *Indian J Med Res* 2006; **124:** 403-410.

178. Ben Nasr H, Chahed K, Bouaouina N, Chouchane L: PTGS2 (COX-2) -765 G > C functional promoter polymorphism and its association with risk and lymph node metastasis in nasopharyngeal carcinoma. *Mol Biol Rep* 2007.

179. Milunsky JM, Maher TA, Zhao G, Roberts AE, Stalker HJ, Zori RT, Burch MN, Clemens M, Mulliken JB, Smith R, Lin AE: TFAP2A mutations result in branchio-oculo-facial syndrome. *Am J Hum Genet* 2008; **82:** 1171-1177.

180. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D: Ultraconserved elements in the human genome. *Science* 2004; **304:** 1321-1325.

181. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJ, Cooke JE, Elgar G: Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* 2005; **3:** e7.

182. Daly AK, Day CP: Candidate gene case-control association studies: advantages and potential pitfalls. *Br J Clin Pharmacol* 2001; **52:** 489-499.

183. Li M, Atmaca-Sonmez P, Othman M, Branham KE, Khanna R, Wade MS, Li Y, Liang L, Zareparsi S, Swaroop A, Abecasis GR: CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. *Nat Genet* 2006; **38:** 1049-1054.

184. Maller J, George S, Purcell S, Fagerness J, Altshuler D, Daly MJ, Seddon JM: Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet* 2006; **38:** 1055-1059.

185. McWhinney SR, Boru G, Binkley PK, Peczkowska M, Januszewicz AA, Neumann HP, Eng C: Intronic single nucleotide polymorphisms in the RET protooncogene are associated with a subset of apparently sporadic pheochromocytoma and may modulate age of onset. *J Clin Endocrinol Metab* 2003; **88:** 4911-4916.

186. CDC: Economic Costs of Birth Defects and Cerebral Palsy -- United States. *MMWR CDC Surveill Summ* 1992; **44:** 694-699.

187. Kasten EF, Schmidt SP, Zickler CF, Berner E, Damian LA, Christian GM, Workman H, Freeman M, Farley MD, Hicks TL: Team care of the patient with cleft lip and palate. *Current problems in pediatric and adolescent health care* 2008; **38:** 138-158.

188. Rivkin CJ, Keith O, Crawford PJ, Hathorn IS: Dental care for the patient with a cleft lip and palate. Part 1: From birth to the mixed dentition stage. *Br Dent J* 2000; **188:** 78-83.

189. Cheng LL, Moor SL, Ho CT: Predisposing factors to dental caries in children with cleft lip and palate: a review and strategies for early prevention. *The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association* 2007; **44:** 67-72.

190. Sambrook J, Fritsch EF, Maniatis T: *Molecular cloning. A laboratory manual, 2nd ed.* NY: Cold Spring Harbor Laboratory Press, 1989.

191. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21:** 263-265.

192. Shaw GM, Lu W, Zhu H, Yang W, Briggs FB, Carmichael SL, Barcellos LF, Lammer EJ, Finnell RH: 118 SNPs of folate-related genes and risks of spina bifida and conotruncal heart defects. *BMC medical genetics* 2009; **10:** 49.

193. Blanton SH, Bertin T, Serna ME, Stal S, Mulliken JB, Hecht JT: Association of chromosomal regions 3p21.2, 10p13, and 16p13.3 with nonsyndromic cleft lip and palate. *Am J Med Genet A* 2004; **125:** 23-27.

194. O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63:** 259-266.

195. Abecasis GR, Cookson WO: GOLD--graphical overview of linkage disequilibrium. *Bioinformatics* 2000; **16:** 182-183.

196. Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin--rapid analysis of dense genetic maps using sparse gene flow trees.[see comment]. *Nat Genet* 2002; **30:** 97-101.

197. Sobel E, Lange K: Descent graphs in pedigree analysis: Applications to haplotyping location scores and marker sharing statistics. *Am J Hum Genet* 1996; **58:** 1323-1337.

198. Van Steen K, McQueen MB, Herbert A, Raby B, Lyon H, Demeo DL, Murphy A, Su J, Datta S, Rosenow C, Christman M, Silverman EK, Laird NM, Weiss ST, Lange C: Genomic screening and replication using the same data set in family-based association testing. *Nat Genet* 2005; **37:** 683-691.

199. Martin ER, Monks SA, Warren LL, Kaplan NL: A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 2000; **67:** 146-154.

200. Martin ER, Bass MP, Gilbert JR, Pericak-Vance MA, Hauser ER: Genotype-based association test for general pedigrees: the genotype-PDT. *Genet Epidemiol* 2003; **25:** 203-213.

201. Chung RH, Hauser ER, Martin ER: The APL test: extension to general nuclear families and haplotypes and examination of its robustness. *Hum Hered* 2006; **61:** 189-199.

202. Hancock DB, Martin ER, Li YJ, Scott WK: Methods for interaction analyses using family-based case-control data: conditional logistic regression versus generalized estimating equations. *Genet Epidemiol* 2007; **8:** 883-893.

203. Chiquet BT, Lidral AC, Stal S, Mulliken JB, Moreno LM, Arco-Burgos M, Valencia-Ramirez C, Blanton SH, Hecht JT: CRISPLD2: A Novel NSCLP Candidate Gene. *Hum Mol Genet* 2007; **16:** 2241-2248.

204. Hecht JT, Hogue D, Strong LC, Hansen MF, Blanton SH, Wagner M: Hereditary multiple exostosis and chondrosarcoma: linkage to chromosome II and loss of heterozygosity for EXT-linked markers on chromosomes II and 8. *Am J Hum Genet* 1995; **56:** 1125-1131.

205. Shen J, Deininger PL, Zhao H: Applications of computational algorithm tools to identify functional SNPs in cytokine genes. *Cytokine* 2006; **35:** 62-66.

206. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR: ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003; **31:** 3568-3571.

207. Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR: An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet* 2006; **15:** 2490-2508.

208. Grabe N: AliBaba2: context specific identification of transcription factor binding sites. *In Silico Biol* 2002; **2:** S1-15.

209. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E: TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 2006; **34:** D108-110.

210. Mooney S: Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. *Briefings in Bioinformatics* 2005; **6:** 44-56.

211. Shen MM: Identification of differentially expressed genes in mouse development using differential display and in situ hybridization. *Methods* 2001; **24:** 15-27.

212. Sprague J, Bayraktaroglu L, Clements D, Conlin T, Fashena D, Frazer K, Haendel M, Howe DG, Mani P, Ramachandran S, Schaper K, Segerdell E, Song P, Sprunger B, Taylor S, Van Slyke CE, Westerfield M: The Zebrafish Information Network: the zebrafish model organism database. *Nucleic Acids Res* 2006; **34:** D581-585.

213. Rosner B: *Fundamentals of Biostatistics*, 5th edn.: Duxbury, 2000.

214. Lewis SW, Reveley AM, Reveley MA, Chitkara B, Murray RM: The familial/sporadic distinction as a strategy in schizophrenia research. *Br J Psychiatry* 1987; **151:** 306-313.

215. Marigo V, Nigro A, Pecci A, Montanaro D, Di Stazio M, Balduini CL, Savoia A: Correlation between the clinical phenotype of MYH9-related disease and tissue distribution of class II nonmuscle myosin heavy chains. *Genomics* 2004; **83:** 1125-1133.

216. Pecci A, Panza E, Pujol-Moix N, Klersy C, Di Bari F, Bozzi V, Gresele P, Lethagen S, Fabris F, Dufour C, Granata A, Doubek M, Pecoraro C, Koivisto PA, Heller PG, Iolascon A, Alvisi P, Schwabe D, De Candia E, Rocca B, Russo U, Ramenghi U, Noris P, Seri M, Balduini CL, Savoia A: Position of nonmuscle myosin heavy chain IIA (NMMHC-IIA) mutations predicts the natural history of MYH9-related disease. *Hum Mutat* 2007.

217. Matsushita T, Hayashi H, Kunishima S, Hayashi M, Ikejiri M, Takeshita K, Yuzawa Y, Adachi T, Hirashima K, Sone M, Yamamoto K, Takagi A, Katsumi A, Kawai K, Nezu T, Takahashi M, Nakashima T, Naoe T, Kojima T, Saito H: Targeted disruption of mouse ortholog of the human MYH9 responsible for macrothrombocytopenia with different organ involvement: hematological, nephrological, and otological studies of heterozygous KO mice. *Biochem Biophys Res Commun* 2004; **325:** 1163-1171.

218. Mhatre AN, Li Y, Bhatia N, Wang KH, Atkin G, Lalwani AK: Generation and characterization of mice with Myh9 deficiency. *Neuromolecular Med* 2007; **9:** 205-215.

219. Gavin BJ, McMahon JA, McMahon AP: Expression of multiple novel Wnt-1/int-1-related genes during fetal and adult mouse development. *Genes Dev* 1990; **4:** 2319-2332.

220. Alappat S, Zhang ZY, Chen YP: Msx homeobox gene family and craniofacial development. *Cell Res* 2003; **13:** 429-442.

221. Suda N, Kitahara Y, Ohyama K: A case of amelogenesis imperfecta, cleft lip and palate and polycystic kidney disease. *Orthod Craniofac Res* 2006; **9:** 52-56.

222. Edwards YH, Putt W, Lekoape KM, Stott D, Fox M, Hopkinson DA, Sowden J: The human homolog T of the mouse T(Brachyury) gene; gene structure, cDNA sequence, and assignment to chromosome 6q27. *Genome Res* 1996; **6:** 226-233.

223. Trembath D, Sherbondy AL, Vandyke DC, Shaw GM, Todoroff K, Lammer EJ, Finnell RH, Marker S, Lerner G, Murray JC: Analysis of select folate pathway genes, PAX3, and human T in a Midwestern neural tube defect population. *Teratology* 1999; **59:** 331-341.

224. Morrison K, Papapetrou C, Attwood J, Hol F, Lynch SA, Sampath A, Hamel B, Burn J, Sowden J, Stott D, Mariman E, Edwards YH: Genetic mapping of the human homologue (T) of mouse T(Brachyury) and a search for allele association between human T and spina bifida. *Hum Mol Genet* 1996; **5:** 669-674.

225. Verlinsky Y, Rechitsky S, Verlinsky O, Ozen S, Sharapova T, Masciangelo C, Morris R, Kuliev A: Preimplantation diagnosis for sonic hedgehog mutation causing familial holoprosencephaly. *N Engl J Med* 2003; **348:** 1449-1454.

226. Benayed R, Gharani N, Rossman I, Mancuso V, Lazar G, Kamdar S, Bruse SE, Tischfield S, Smith BJ, Zimmerman RA, Dicicco-Bloom E, Brzustowicz LM, Millonig JH: Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus. *Am J Hum Genet* 2005; **77:** 851-868.

227. Chenevix-Trench G, Jones K, Green AC, Duffy DL, Martin NG: Cleft lip with or without cleft palate: associations with transforming growth factor alpha and retinoic acid receptor loci. *Am J Hum Genet* 1992; **51:** 1377-1385.

228. Hahn H, Wicking C, Zaphiropoulous PG, Gailani MR, Shanley S, Chidambaram A, Vorechovsky I, Holmberg E, Unden AB, Gillies S, Negus K, Smyth I, Pressman C, Leffell DJ, Gerrard B, Goldstein AM, Dean M, Toftgard R, Chenevix-Trench G, Wainwright B, Bale AE: Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome. *Cell* 1996; **85:** 841-851.

229. Trueba SS, Auge J, Mattei G, Etchevers H, Martinovic J, Czernichow P, Vekemans M, Polak M, Attie-Bitach T: PAX8, TITF1, and FOXE1 gene expression patterns during human development: new insights into human thyroid development and thyroid dysgenesis-associated malformations. *J Clin Endocrinol Metab* 2005; **90:** 455-462.

230. J. Machida LMM, M.A. Mansilla, S.B. Bullard, T.D. Busch, M.K. Johnson, T. McHenry, M.E. Cooper, C. Valencia-Ramirez, M. Arcos-Burgos, A. Hing, E.J. Lammer, M. Jones, K. Christensen, J.C. Murray, M.L. Marazita, and A.C. Lidral: The Role of FOXE1 in the Etiology of Cleft Lip: *57th Annual Meeting of the American Society of Human Genetics*. San Diego, California, 2007, p 470.

231. Bejsovec A: Wnt pathway activation: new relations and locations. *Cell* 2005; **120:** 11-14.

232. Dale RM, Sisson BE, Topczewski J: The Emerging Role of Wnt/PCP Signaling in Organ Formation. *Zebrafish* 2009; **6:** 9-14.

233. Girotti M, Zingg HH: Gene expression profiling of rat uterus at different stages of parturition. *Endocrinology* 2003; **144:** 2254-2265.

234. Kemp C, Willems E, Abdo S, Lambiv L, Leyns L: Expression of all Wnt genes and their secreted antagonists during mouse blastocyst and postimplantation development. *Dev Dyn* 2005; **233:** 1064-1075.

235. Heisenberg CP, Tada M, Rauch GJ, Saude L, Concha ML, Geisler R, Stemple DL, Smith JC, Wilson SW: Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature* 2000; **405:** 76-81.

236. Liu Y, Wang X, Lu CC, Kerman R, Steward O, Xu XM, Zou Y: Repulsive Wnt signaling inhibits axon regeneration after CNS injury. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2008; **28:** 8376-8382.

237. Louis I, Heinonen KM, Chagraoui J, Vainio S, Sauvageau G, Perreault C: The signaling protein Wnt4 enhances thymopoiesis and expands multipotent

hematopoietic progenitors through beta-catenin-independent signaling. *Immunity* 2008; **29:** 57-67.

238. Andersson ER, Prakash N, Cajanek L, Minina E, Bryja V, Bryjova L, Yamaguchi TP, Hall AC, Wurst W, Arenas E: Wnt5a regulates ventral midbrain morphogenesis and the development of A9-A10 dopaminergic cells in vivo. *PLoS ONE* 2008; **3:** e3517.

239. Niemann S, Zhao C, Pascu F, Stahl U, Aulepp U, Niswander L, Weber JL, Muller U: Homozygous WNT3 mutation causes tetra-amelia in a large consanguineous family. *Am J Hum Genet* 2004; **74:** 558-563.

240. Liu P, Wakamiya M, Shea MJ, Albrecht U, Behringer RR, Bradley A: Requirement for Wnt3 in vertebrate axis formation. *Nat Genet* 1999; **22:** 361-365.

241. Nordstrèom U, Jessell TM, Edlund T: Progressive induction of caudal neural character by graded Wnt signaling. *Nat Neurosci* 2002; **5:** 525-532.

242. Riccomagno MM, Takada S, Epstein DJ: Wnt-dependent regulation of inner ear morphogenesis is balanced by the opposing and supporting roles of Shh. *Genes Dev* 2005; **19:** 1612-1623.

243. Yang DH, Yoon JY, Lee SH, Bryja V, Andersson ER, Arenas E, Kwon YG, Choi KY: Wnt5a is required for endothelial differentiation of embryonic stem cells and vascularization via pathways involving both Wnt/beta-catenin and protein kinase Calpha. *Circ Res* 2009; **104:** 372-379.

244. Lange C, Mix E, Rateitschak K, Rolfs A: Wnt signal pathways and neural stem cell differentiation. *Neurodegener Dis* 2006; **3:** 76-86.

245. Stenman JM, Rajagopal J, Carroll TJ, Ishibashi M, McMahon J, McMahon AP: Canonical Wnt signaling regulates organ-specific assembly and differentiation of CNS vasculature. *Science* 2008; **322:** 1247-1250.

246. Parr BA, McMahon AP: Sexually dimorphic development of the mammalian reproductive tract requires Wnt-7a. *Nature* 1998; **395:** 707-710.

247. Woods CG, Stricker S, Seemann P, Stern R, Cox J, Sherridan E, Roberts E, Springell K, Scott S, Karbani G, Sharif SM, Toomes C, Bond J, Kumar D, Al-Gazali L, Mundlos S: Mutations in WNT7A cause a range of limb malformations, including Fuhrmann syndrome and Al-Awadi/Raas-Rothschild/Schinzel phocomelia syndrome. *Am J Hum Genet* 2006; **79:** 402-408.

248. Kumar D, Duggan MB, Mueller RF, Karbani G: Familial aplasia/hypoplasia of pelvis, femur, fibula, and ulna with abnormal digits in an inbred Pakistani Muslim

family: a possible new autosomal recessive disorder with overlapping manifestations of the syndromes of Fuhrmann, Al-Awadi, and Raas-Rothschild. *Am J Med Genet* 1997; **70:** 107-113.

249. Woods CG, Stricker S, Seemann P, Stern R, Cox J, Sherridan E, Roberts E, Springell K, Scott S, Karbani G, Sharif SM, Toomes C, Bond J, Kumar D, Al-Gazali L, Mundlos S: Mutations in WNT7A cause a range of limb malformations, including Fuhrmann syndrome and Al-Awadi/Raas-Rothschild/Schinzel phocomelia syndrome. *Am J Hum Genet* 2006; **79:** 402-408.

250. LaBonne C, Bronner-Fraser M: Neural crest induction in Xenopus: evidence for a two-signal model. *Development* 1998; **125:** 2403-2414.

251. Ciruna B, Rossant J: FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Dev Cell* 2001; **1:** 37-49.

252. Liu W, Sun X, Braut A, Mishina Y, Behringer RR, Mina M, Martin JF: Distinct functions for Bmp signaling in lip and palate fusion in mice. *Development* 2005; **132:** 1453-1461.

253. Chai Y, Maxson RE, Jr.: Recent advances in craniofacial morphogenesis. *Dev Dyn* 2006; **235:** 2353-2375.

254. Brewer S, Feng W, Huang J, Sullivan S, Williams T: Wnt1-Cre-mediated deletion of AP-2alpha causes multiple neural crest-related defects. *Dev Biol* 2004; **267:** 135-152.

255. Dixon J, Jones NC, Sandell LL, Jayasinghe SM, Crane J, Rey JP, Dixon MJ, Trainor PA: Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proc Natl Acad Sci U S A* 2006; **103:** 13403-13408.

256. Ito Y, Yeo JY, Chytil A, Han J, Bringas P, Jr., Nakajima A, Shuler CF, Moses HL, Chai Y: Conditional inactivation of Tgfbr2 in cranial neural crest causes cleft palate and calvaria defects. *Development* 2003; **130:** 5269-5280.

257. Pfeiffer RA, Stoss H, Voight HJ, Wundisch GF: Absence of fibula and ulna with oligodactyly, contractures, right-angle bowing of femora, abnormal facial morphology, cleft lip/palate and brain malformation in two sibs: a possibly new lethal syndrome. *Am J Med Genet* 1988; **29:** 901-908.

258. Camera G, Ferraiolo G, Leo D, Spaziale A, Pozzolo S: Limb/pelvis-hypoplasia/aplasia syndrome (Al-Awadi/Raas-Rothschild syndrome): report of two Italian sibs and further confirmation of autosomal recessive inheritance. *J Med Genet* 1993; **30:** 65-69.

259. Inoue K, Ohyama T, Sakuragi Y, Yamamoto R, Inoue NA, Yu LH, Goto Y, Wegner M, Lupski JR: Translation of SOX10 3' untranslated region causes a complex severe neurocristopathy by generation of a deleterious functional domain. *Hum Mol Genet* 2007; **16:** 3037-3046.

260. Moore SW, Sidler D, Zaahl MG: The ITGB2 immunomodulatory gene (CD18), enterocolitis, and Hirschsprung's disease. *J Pediatr Surg* 2008; **43:** 1439-1444.

261. Reamon-Buettner SM, Cho SH, Borlak J: Mutations in the 3'-untranslated region of GATA4 as molecular hotspots for congenital heart disease (CHD). *BMC medical genetics* 2007; **8:** 38.

262. Marazita ML, A.C. Lidral, J.C. Murray, L.L. Field, B.S. Maher, T.G. McHenry, M.E. Cooper, M. Govil, S. Daack-Hirsch, B. Riley, A. Jugessur, T. Felix, L. Morene, M.A. Mansilla, A.R. Vieira, K. Doheny, E. Pugh, C. Valencia-Ramirez, and M. Arcos-Burgos: Genome Scan, Fine-Mapping, and Candidate Gene Analysis of Non-Syndromic Cleft Lip with or without Cleft Palate Reveals Phenotype-Specific Differences in Linkage and Association Results. *Hum Hered* 2009; **68:** 151-170.

263. Ahsan M, Ohta K, Kuriyama S, Tanaka H: Novel soluble molecule, Akhirin, is expressed in the embryonic chick eyes and exhibits heterophilic cell-adhesion activity. *Dev Dyn* 2005; **233:** 95-104.

264. Liepinsh E, Trexler M, Kaikkonen A, Weigelt J, Banyai L, Patthy L, Otting G: NMR structure of the LCCL domain and implications for DFNA9 deafness disorder. *EMBO J* 2001; **20:** 5347-5353.

265. Nagai H, Sugito N, Matsubara H, Tatematsu Y, Hida T, Sekido Y, Nagino M, Nimura Y, Takahashi T, Osada H: CLCP1 interacts with semaphorin 4B and regulates motility of lung cancer cells. *Oncogene* 2007.

266. Trexler M, Banyai L, Patthy L: The LCCL module. *Eur J Biochem* 2000; **267:** 5751-5757.

267. Robertson NG, Resendes BL, Lin JS, Lee C, Aster JC, Adams JC, Morton CC: Inner ear localization of mRNA and protein products of COCH, mutated in the sensorineural deafness and vestibular disorder, DFNA9. *Hum Mol Genet* 2001; **10:** 2493-2500.

268. Alkuraya FS, Saadi I, Lund JJ, Turbe-Doan A, Morton CC, Maas RL: SUMO1 haploinsufficiency leads to cleft lip and palate. *Science* 2006; **313:** 1751.

269. Faniello MC, Fregola A, Nistico A, Quaresima B, Crugliano T, Faraonio R, Puzzonia P, Baudi F, Parlato G, Cuda G, Morrone G, Venuta S, Costanzo F: Detection and functional analysis of an SNP in the promoter of the human ferritin H gene that modulates the gene expression. *Gene* 2006; **377:** 1-5.

270. Law AJ, Lipska BK, Weickert CS, Hyde TM, Straub RE, Hashimoto R, Harrison PJ, Kleinman JE, Weinberger DR: Neuregulin 1 transcripts are differentially expressed in schizophrenia and regulated by 5' SNPs associated with the disease. *Proc Natl Acad Sci U S A* 2006; **103:** 6747-6752.

271. Nielsen KB, Sorensen S, Cartegni L, Corydon TJ, Doktor TK, Schroeder LD, Reinert LS, Elpeleg O, Krainer AR, Gregersen N, Kjems J, Andresen BS: Seemingly neutral polymorphic variants may confer immunity to splicing-inactivating mutations: a synonymous SNP in exon 5 of MCAD protects from deleterious mutations in a flanking exonic splicing enhancer. *Am J Hum Genet* 2007; **80:** 416-432.

272. Arcos-Burgos M, Muenke M: Genetics of population isolates. *Clin Genet* 2002; **61:** 233-247.

273. White SH: Amino acid preferences of small proteins. Implications for protein stability and evolution. *J Mol Biol* 1992; **227:** 991-995.

274. Aguilar B, Rojas JC, Collados MT: Metabolism of homocysteine and its relationship with cardiovascular disease. *J Thromb Thrombolysis* 2004; **18:** 75-87.

275. Lumley J, Watson L, Watson M, Bower C: Periconceptional supplementation with folate and/or multivitamins for preventing neural tube defects. *Cochrane Database Syst Rev* 2001**:** CD001056.

276. Czeizel AE, Dudas I: Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *N Engl J Med* 1992; **327:** 1832-1835.

277. Sayed AR, Bourne D, Pattinson R, Nixon J, Henderson B: Decline in the prevalence of neural tube defects following folic acid fortification and its cost-benefit in South Africa. *Birth Defects Res A Clin Mol Teratol* 2008; **82:** 211-216.

278. Bower C, D'Antoine H, Stanley FJ: Neural tube defects in Australia: Trends in encephaloceles and other neural tube defects before and after promotion of folic acid supplementation and voluntary food fortification. *Birth Defects Res A Clin Mol Teratol* 2009.

279. Badovinac RL, Werler MM, Williams PL, Kelsey KT, Hayes C: Folic acid-containing supplement consumption during pregnancy and risk for oral clefts: a meta-analysis. *Birth Defects Res A Clin Mol Teratol* 2007; **79:** 8-15.

280. Bille C, Olsen J, Vach W, Knudsen VK, Olsen SF, Rasmussen K, Murray JC, Andersen AM, Christensen K: Oral clefts and life style factors--a case-cohort study based on prospective Danish data. *Eur J Epidemiol* 2007; **22:** 173-181.

281. Canfield MA, Collins JS, Botto LD, Williams LJ, Mai CT, Kirby RS, Pearson K, Devine O, Mulinare J: Changes in the birth prevalence of selected birth defects after grain fortification with folic acid in the United States: findings from a multi-state population-based study. *Birth Defects Res A Clin Mol Teratol* 2005; **73:** 679-689.

282. Czeizel AE, Toth M, Rockenbauer M: Population-based case control study of folic acid supplementation during pregnancy. *Teratology* 1996; **53:** 345-351.

283. Shaw GM, Lammer EJ, Wasserman CR, O'Malley CD, Tolarova MM: Risks of orofacial clefts in children born to women using multivitamins containing folic acid periconceptionally. *Lancet* 1995; **346:** 393-396.

284. Weisberg I, Tran P, Christensen B, Sibani S, Rozen R: A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity. *Mol Genet Metab* 1998; **64:** 169-172.

285. Goyette P, Sumner JS, Milos R, Duncan AM, Rosenblatt DS, Matthews RG, Rozen R: Human methylenetetrahydrofolate reductase: isolation of cDNA, mapping and mutation identification. *Nat Genet* 1994; **7:** 195-200.

286. Blanton SH, R. R. Henry, J. B. Mulliken, S. Stal, R. H. Finnell and J. T. Hecht: The Folate Pathway and Nonsyndromic Cleft Lip and Palate. *Birth Defects Research Part A: Clinical and Molecular Teratology (revisions submitted)* 2010.

287. Letra A, Menezes R, Cooper ME, Fonseca RF, Tropp S, Govil M, Granjeiro JM, Mansilla MA, Murray JC, Castilla EE, I.M. O, Czeizel AE, Ma L, Chiquet BT, Hecht JT, Vieira AR, Marazita ML: CRISPLD2 variants including a C471T silent mutation may contribute tononsyndromic cleft lip with or without palate. *Cleft Palate J* 2010 in press.

288. Saitão N: Principles of protein architecture. *Adv Biophys* 1989; **25:** 95-132.

289. Benard O, Balasubramanian KA: Effect of oxidant exposure on thiol status in the intestinal mucosa. *Biochem Pharmacol* 1993; **45:** 2011-2015.

290. Blanton SH, Henry RR, Mulliken JB, Stal S, Burt A, Finnell RH, Hecht JT: The Folate Pathway and Nonsyndromic Cleft Lip and Palate. *Birth Defects Res Part A: Clin and Mol Terat (submitted)* 2010.

291. Ramakrishnan S, Sulochana KN, Lakshmi S, Selvi R, Angayarkanni N: Biochemistry of homocysteine in health and diseases. *Indian J Biochem Biophys* 2006; **43:** 275-283.

292. Thisse B, Thisse, C.: Fast Release Clones: A High Throughput Expression Analysis. *ZFIN Direct Data Submission (http://zfinorg)* 2004.

293. Honein MA, Paulozzi LJ, Mathews TJ, Erickson JD, Wong LY: Impact of folic acid fortification of the US food supply on the occurrence of neural tube defects. *JAMA* 2001; **285:** 2981-2986.

294. MMWR: Spina bifida and anencephaly before and after folic acid mandate--United States, 1995-1996 and 1999-2000. *MMWR Morb Mortal Wkly Rep* 2004; **53:** 362-365.

295. Tolorova M, Harris J: Reduced recurrence of orofacial clefts after periconceptual supplementation with high dose folic acid and multivitamins. *Teratology* 1995; **51:** 71-78.

296. Ester AR, Weymouth KS, Burt A, Wise CA, Scott A, Gurnett CA, Dobbs MB, Blanton SH, Hecht JT: Altered transmission of HOX and apoptotic SNPs identify a potential common pathway for clubfoot. *Am J Med Genet A* 2009; **149A:** 2745-2752.

297. Ben J, Jabs EW, Chong SS: Genomic, cDNA and embryonic expression analysis of zebrafish IRF6, the gene mutated in the human oral clefting disorders Van der Woude and popliteal pterygium syndromes. *Gene Expr Patterns* 2005; **5:** 629-638.

298. Sabel JL, d'Alenðcon C, O'Brien EK, Van Otterloo E, Lutz K, Cuykendall TN, Schutte BC, Houston DW, Cornell RA: Maternal Interferon Regulatory Factor 6 is required for the differentiation of primary superficial epithelia in Danio and Xenopus embryos. *Dev Biol* 2009; **325:** 249-262.

299. Knight AS, Schutte BC, Jiang R, Dixon MJ: Developmental expression analysis of the mouse and chick orthologues of IRF6: the gene mutated in Van der Woude syndrome. *Developmental dynamics : an official publication of the American Association of Anatomists* 2006; **235:** 1441-1447.

300. Washbourne BJ, Cox TC: Expression profiles of cIRF6, cLHX6 and cLHX7 in the facial primordia suggest specific roles during primary palatogenesis. *BMC developmental biology* 2006; **6:** 18.

301. Hulbert EM, Smink LJ, Adlem EC, Allen JE, Burdick DB, Burren OS, Cavnor CC, Dolman GE, Flamez D, Friery KF, Healy BC, Killcoyne SA, Kutlu B, Schuilenburg H, Walker NM, Mychaleckyj J, Eizirik DL, Wicker LS, Todd JA, Goodman N: T1DBase: integration and presentation of complex data for type 1 diabetes research. *Nucleic Acids Res* 2007; **35:** D742-746.

302. Kaplan F, Ledoux P, Kassamali FQ, Gagnon S, Post M, Koehler D, Deimling J, Sweezey NB: A novel developmentally regulated gene in lung mesenchyme:

homology to a tumor-derived trypsin inhibitor. *Am J Physiol* 1999; **276:** L1027-1036.

303. Gibbs GM, Roelants K, O'Bryan MK: The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins--roles in reproduction, cancer, and immune defense. *Endocr Rev* 2008; **29:** 865-897.

304. Oyewumi L, Kaplan F, Gagnon S, Sweezey NB: Antisense oligodeoxynucleotides decrease LGL1 mRNA and protein levels and inhibit branching morphogenesis in fetal rat lung. *Am J Respir Cell Mol Biol* 2003; **28:** 232-240.

305. Oyewumi L, Kaplan F, Sweezey NB: Lgl1, a mesenchymal modulator of early lung branching morphogenesis, is a secreted glycoprotein imported by late gestation lung epithelial cells. *Biochem J* 2003; **376:** 61-69.

306. Quinlan J, Kaplan F, Sweezey N, Goodyer P: LGL1, a novel branching morphogen in developing kidney, is induced by retinoic acid. *Am J Physiol Renal Physiol* 2007; **293:** F987-993.

307. Lan J, Ribeiro L, Mandeville I, Nadeau K, Bao T, Cornejo S, Sweezey NB, Kaplan F: Inflammatory cytokines, goblet cell hyperplasia and altered lung mechanics in Lgl1+/- mice. *Respir Res* 2009; **10:** 83.

308. Berndt JD, Clay MR, Langenberg T, Halloran MC: Rho-kinase and myosin II affect dynamic neural crest cell behaviors during epithelial to mesenchymal transition in vivo. *Dev Biol* 2008; **324:** 236-244.

309. Kang P, Svoboda KK: Epithelial-mesenchymal transformation during craniofacial development. *J Dent Res* 2005; **84:** 678-690.

310. Westerfield M: *The zebrafish book : a guide for the laboratory use of zebrafish (Danio rerio)*, Ed. 3. edn. Eugene, OR: M. Westerfield, 1995.

311. Graham A: Jaw development: chinless wonders. *Curr Biol* 2002; **12:** R810-812.

312. Akimenko MA, Ekker M, Wegner J, Lin W, Westerfield M: Combinatorial expression of three zebrafish genes related to distal-less: part of a homeobox gene code for the head. *J Neurosci* 1994; **14:** 3475-3486.

313. Ellies DL, Stock DW, Hatch G, Giroux G, Weiss KM, Ekker M: Relationship between the genomic organization and the overlapping embryonic expression patterns of the zebrafish dlx genes. *Genomics* 1997; **45:** 580-590.

314. Letra A, Menezes R, Cooper ME, Fonseca RF, Tropp S, Govil M, Granjeiro JM, Imoehl SR, Mansill MA, Murray JC, Castilla EE, Orioli IM, Czeizel AE, Ma L, Chiquet BT, Hecht JT, Vieira AR, Marazita ML: *CRISPLD2* variants including a

C471T silent mutation may contribute to nonsyndromic cleft lip with or without cleft palate. *Cleft Palate Craniofacial Journal (In Press)* 2010.

315. Feng Q, Di R, Tao F, Chang Z, Lu S, Fan W, Shan C, Li X, Yang Z: PDK1 regulates vascular remodeling and promotes epithelial-mesenchymal transition in cardiac development. *Mol Cell Biol* 2010.

316. Micalizzi DS, Farabaugh SM, Ford HL: Epithelial-Mesenchymal Transition in Cancer: Parallels Between Normal Development and Tumor Progression. *J Mammary Gland Biol Neoplasia* 2010.

317. Sanders EJ, Prasad S: Invasion of a basement membrane matrix by chick embryo primitive streak cells in vitro. *J Cell Sci* 1989; **92 (Pt 3):** 497-504.

318. Schilling TF, Le Pabic P: Fishing for the signals that pattern the face. *J Biol* 2009; **8:** 101.

319. Luo R, An M, Arduini BL, Henion PD: Specific pan-neural crest expression of zebrafish Crestin throughout embryonic development. *Dev Dyn* 2001; **220:** 169-174.

320. Rubinstein AL, Lee D, Luo R, Henion PD, Halpern ME: Genes dependent on zebrafish cyclops function identified by AFLP differential gene expression screen. *Genesis* 2000; **26:** 86-97.

321. Thisse C, Thisse B, Postlethwait JH: Expression of snail2, a second member of the zebrafish snail family, in cephalic mesendoderm and presumptive neural crest of wild-type and spadetail mutant embryos. *Dev Biol* 1995; **172:** 86-99.

322. Nadeau K, Montermini L, Mandeville I, Xu M, Weiss ST, Sweezey NB, Kaplan F: Modulation of lgl1 by steroid, retinoic Acid, and vitamin d models complex transcriptional regulation during alveolarization. *Pediatr Res* 2010; **67:** 375-381.

323. Campbell JL, Jr., Smith MA, Fisher JW, Warren DA: Dose-response for retinoic acid-induced forelimb malformations and cleft palate: a comparison of computerized image analysis and visual inspection. *Birth Defects Res B Dev Reprod Toxicol* 2004; **71:** 289-295.

324. Padmanabhan R, Ahmed I: Retinoic acid-induced asymmetric craniofacial growth and cleft palate in the TO mouse fetus. *Reprod Toxicol* 1997; **11:** 843-860.

325. Lohnes D, Mark M, Mendelsohn C, Dolle P, Dierich A, Gorry P, Gansmuller A, Chambon P: Function of the retinoic acid receptors (RARs) during development (I). Craniofacial and skeletal abnormalities in RAR double mutants. *Development* 1994; **120:** 2723-2748.

326. Li N, Kelsh RN, Croucher P, Roehl HH: Regulation of neural crest cell fate by the retinoic acid and Pparg signalling pathways. *Development* 2010; **137:** 389-394.

327. Reijntjes S, Rodaway A, Maden M: The retinoic acid metabolising gene, CYP26B1, patterns the cartilaginous cranial neural crest in zebrafish. *Int J Dev Biol* 2007; **51:** 351-360.

328. Ellies DL, Langille RM, Martin CC, Akimenko MA, Ekker M: Specific craniofacial cartilage dysmorphogenesis coincides with a loss of dlx gene expression in retinoic acid-treated zebrafish embryos. *Mech Dev* 1997; **61:** 23-36.

329. Yan YL, Jowett T, Postlethwait JH: Ectopic expression of hoxb2 after retinoic acid treatment or mRNA injection: disruption of hindbrain and craniofacial morphogenesis in zebrafish embryos. *Dev Dyn* 1998; **213:** 370-385.

330. Davenne M, Maconochie MK, Neun R, Pattyn A, Chambon P, Krumlauf R, Rijli FM: Hoxa2 and Hoxb2 control dorsoventral patterns of neuronal development in the rostral hindbrain. *Neuron* 1999; **22:** 677-691.

331. Depew MJ, Simpson CA, Morasso M, Rubenstein JL: Reassessing the Dlx code: the genetic regulation of branchial arch skeletal pattern and development. *J Anat* 2005; **207:** 501-561.

332. Qiu M, Bulfone A, Martinez S, Meneses JJ, Shimamura K, Pedersen RA, Rubenstein JL: Null mutation of Dlx-2 results in abnormal morphogenesis of proximal first and second branchial arch derivatives and abnormal differentiation in the forebrain. *Genes Dev* 1995; **9:** 2523-2538.

333. Limpach A, Dalton M, Miles R, Gadson P: Homocysteine inhibits retinoic acid synthesis: a mechanism for homocysteine-induced congenital defects. *Exp Cell Res* 2000; **260:** 166-174.

334. Dalton ML, Gadson PF, Jr., Wrenn RW, Rosenquist TH: Homocysteine signal cascade: production of phospholipids, activation of protein kinase C, and the induction of c-fos and c-myb in smooth muscle cells. *FASEB J* 1997; **11:** 703-711.

335. Rosenquist TH, Bennett GD, Brauer PR, Stewart ML, Chaudoin TR, Finnell RH: Microarray analysis of homocysteine-responsive genes in cardiac neural crest cells in vitro. *Dev Dyn* 2007; **236:** 1044-1054.

336. Tierney BJ, Ho T, Reedy MV, Brauer PR: Homocysteine inhibits cardiac neural crest cell formation and morphogenesis in vivo. *Dev Dyn* 2004; **229:** 63-73.

337. Knott L, Hartridge T, Brown NL, Mansell JP, Sandy JR: Homocysteine oxidation and apoptosis: a potential cause of cleft palate. *In Vitro Cell Dev Biol Anim* 2003; **39:** 98-105.

338. Wong WY, Eskes TK, Kuijpers-Jagtman AM, Spauwen PH, Steegers EA, Thomas CM, Hamel BC, Blom HJ, Steegers-Theunissen RP: Nonsyndromic orofacial clefts: association with maternal hyperhomocysteinemia. *Teratology* 1999; **60:** 253-257.

339. Rubini M, Brusati R, Garattini G, Magnani C, Liviero F, Bianchi F, Tarantino E, Massei A, Pollastri S, Carturan S, Amadori A, Bertagnin E, Cavallaro A, Fabiano A, Franchella A, Calzolari E: Cystathionine beta-synthase c.844ins68 gene variant and non-syndromic cleft lip and palate. *Am J Med Genet A* 2005; **136A:** 368-372.

340. Verkleij-Hagoort A, Bliek J, Sayed-Tabatabaei F, Ursem N, Steegers E, Steegers-Theunissen R: Hyperhomocysteinemia and MTHFR polymorphisms in association with orofacial clefts and congenital heart defects: a meta-analysis. *Am J Med Genet A* 2007; **143A:** 952-960.

341. Ozerol E, Ozerol I, Gokdeniz R, Temel I, Akyol O: Effect of smoking on serum concentrations of total homocysteine, folate, vitamin B12, and nitric oxide in pregnancy: a preliminary study. *Fetal Diagn Ther* 2004; **19:** 145-148.

342. Chiquet BT, R. R. Henry, A. Burt, J. B. Mulliken, S. Stal, S. H. Blanton and J. T. Hecht: Nonsyndromic cleft lip and palate: CRISPLD Genes and the Folate Gene Pathway Connection. *Birth Defects Research Part A: Clinical and Molecular Teratology (revisions submitted)* 2010.

VITA

Brett Thomas Chiquet was born in Thibodeaux, Lousiana to Carol LaBorde and Bryan Chiquet on October 5, 1981. Brett grew up in Donaldsonville, Louisiana, attended and graduated as valedictorian from Ascension Catholic High School in May 1999. Brett graduated Magna Cum Laude from the Louisiana Scholars' College at Northwestern State University in Natchitoches, Louisiana, with a BA in Liberal Arts with a Concentration in Scientific Inquiry. After graduation he worked as a research technician at UT Southwestern Medical Center in Dallas, Texas. In May 2004 he entered the University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences and in August 2004 he entered the Dental Branch. Both DDS and PhD were completed in May 2011. In July 2011, Brett will start a residency program in Pediatric Dentistry at the University of Texas Health Science Center in Houston, Texas, where he wishes to pursue a career in pediatric dentistry research.