CRISPLD2: a novel NSCLP candidate gene

Brett T. Chiquet[1,2], **Andrew C. Lidral**[3,4], **Samuel Stal**[5], **John B. Mulliken**[6], **Lina M. Moreno**[3], **Mauricio Arco-Burgos**[7], **Consuelo Valencia-Ramirez**[8], **Susan H. Blanton**[9], and **Jacqueline T. Hecht**[1,*]

[1]Department of Pediatrics, University of Texas Medical School at Houston

[2]University of Texas Dental Branch at Houston, Houston, TX 77030, USA

[3]Dows Institute for Dental Research, University of Iowa, Iowa City, IA, USA

[4]Department of Orthodontics, University of Iowa, Iowa City, IA, USA

[5]Texas Children's Hospital, Houston, TX 77030, USA

[6]Children's Hospital, Boston, MA, USA

[7]Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

[8]College of Dentistry, University of Antioquia, Medellín, Colombia, South America, Miami, FL 33101, USA

[9]University of Miami Miller School of Medicine, Miami, FL 33101, USA

Abstract

Non-syndromic cleft lip with or without cleft palate (NSCLP) results from the complex interaction between genes and environmental factors. Candidate gene analysis and genome scans have been employed to identify the genes contributing to NSCLP. In this study, we evaluated the 16q24.1 chromosomal region, which has been identified by multiple genome scans as an NSCLP region of interest. Two candidate genes were found in the region: interferon regulatory factor 8 (IRF8) and cysteine-rich secretory protein LCCL domain containing 2 (CRISPLD2). Initially, Caucasian and Hispanic NSCLP multiplex families and simplex parent–child trios were genotyped for single nucleotide polymorphisms (SNPs) in both IRF8 and CRISPLD2. CRISPLD2 was subsequently genotyped in a data set comprised of NSCLP families from Colombia, South America. Linkage disequilibrium analysis identified a significant association between CRISPLD2 and NSCLP in both our Caucasian and Hispanic NSCLP cohorts. SNP rs1546124 and haplotypes between rs1546124 and either rs4783099 or rs16974880 were significant in the Caucasian multiplex population ($P = 0.01$, $P = 0.002$ and $P = 0.001$, respectively). An altered transmission of CRISPLD2 SNPs rs8061351 ($P = 0.02$) and rs2326398 ($P = 0.06$) was detected in the Hispanic population. No association was found between CRISPLD2 and our Colombian population or IRF8 and NSCLP. *In situ* hybridization showed that CRISPLD2 is expressed in the mandible, palate and

*To whom correspondence should be addressed: Tel: +1 7135005764; Fax: +1 7135005689; ejacqueline.t.hecht@uth.tmc.edu.

nasopharynx regions during craniofacial development at E13.5–E17.5, respectively. Altogether, these data suggest that genetic variation in CRISPLD2 has a role in the etiology of NSCLP.

Introduction

Non-syndromic cleft lip with or without cleft palate (NSCLP) is a common birth defect that has a birth prevalence of 1 in 700 live births and affects 4000 newborns each year in the United States (1,2). A great deal of information is known about the demography of NSCLP but little is known about the causes of NSCLP. Numerous studies suggest that NSCLP results from the complex interaction of both genes and environmental factors. Folic acid deficiency, maternal smoking and anticonvulsant medications are some of the environmental influences that have been associated with NSCLP (3–5). Evidence for involvement of genes in NSCLP comes from family studies which show that: (1) the heritability of NSCLP in a Caucasian population is approximately 76%, (2) the rate of concordance is higher in monozygotic (25–40%) than dizygotic (3–6%) twins and (3) there is an increased relative risk to siblings ($\lambda_s = 30$–40) compared to the general population (1,6). Variation in 2–14 genes is estimated to contribute to NSCLP but this number may underestimate the true variation because there are a large number of genes contributing to craniofacial morphogenesis (7).

Although there is a substantial genetic component to NSCLP, identifying the genes that play an etiological role in NSCLP has been challenging. There are two strategies that have been utilized to identify NSCLP genes: candidate gene linkage and association testing and genome-wide scans. Candidate gene analysis targets individual genes that have a biological basis for causing NSCLP, such as (1) genes involved in craniofacial development pathways (e.g. BMP4 and FGF8), (2) genes that underlie syndromes associated with clefting (e.g. IRF6) or (3) animal models with a clefting phenotype (e.g. Egfr, Bmp4, Bmpr1 and Folbp1) (8–18). Eight genome scans have identified regions that may potentially harbor NSCLP susceptibility genes (19–25). In 2000, Prescott *et al.* (26) identified nine chromosomal regions that are associated with orofacial clefting in their Caucasian NSCLP sib-pair population. To confirm these findings, 37 short tandem repeat (STR) markers were genotyped in 65 of our Caucasian multiplex families (27). Linkage disequilibrium (LD) analysis identified STR marker D16S3037, in the 16q24.1 chromosomal region, that was significantly associated with NSCLP ($P = 0.00063$). Results from three other NSCLP genome scans and one targeted genome scan have also suggested that a NSCLP candidate gene lies on the long arm of chromosome 16, between 16q21-24, which includes D16S3037 (19,21,25,28).

In this study, we performed candidate gene analysis of two genes within the 16q24.1 region: interferon regulatory factor 8 (IRF8), which is in the same gene family as IRF6, and cysteine-rich secretory protein LCCL domain containing 2 (CRISPLD2), a novel gene. *In situ* hybridizations were also performed to determine whether or not CRISPLD2 may have a biological role in craniofacial development.

Results

Microsatellite analysis

In a follow-up study to Prescott *et al.*'s (26) genome-wide scan, we genotyped 37 STRs in 47 multiplex NSCLP families (27). Re-analysis of the original data stratified by ethnicity with family based association test (FBAT) showed an additional microsatellite marker, D16S3037, with significant linkage and 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.

Candidate gene testing

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 (VWS) and IRF6 variants have been shown to play an etiological role in NSCLP (13,14,16,29,30). Also, IRF8 is 1 Mb downstream of D16S3037. Seven IRF8 single nucleotide polymorphisms (SNPs) (two flanking and five intragenic) were genotyped in our Caucasian and Hispanic cohorts (Fig. 1A). All SNPs were in Hardy–Weinberg equilibrium (HWE). Two SNPs had significantly different allele frequencies between the Hispanic and Caucasian populations (Table 1). Hence, families were stratified by ethnicity for all statistical analyses. There was significant LD between SNPs within IRF8 (data not shown). No evidence for linkage or association with NSCLP was found by either parametric or non-parametric linkage analysis. Previous interrogation of IRF6 in these populations showed an increased transmission of haplotypes constructed with the major allele rs2013162 ($P = 0.009$) (13). 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 (four flanking and 14 intragenic) were genotyped in the Caucasian and Hispanic multiplex families and simplex parent–child trios (Fig. 1B). 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 1). Parametric and non-parametric 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 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 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 over-transmission of the 1–2 haplotype for rs1546124 and rs16974880 (83 observed versus 66 expected, $P = 0.001$), and an overall departure from expected for all the haplotypes ($P = 0.01$).

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 ($P = 0.05$) was noted (Table 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 1). In a comparison between the Hispanic and Colombian families, two SNPs (rs4572384 and

rs1874015) had significantly different allele frequencies (Table 1). Interestingly, the Colombian sample generally had frequencies between the Caucasian and Hispanic samples. Neither did the parametric and non-parametric linkage analysis show any evidence for linkage with NSCLP nor was there evidence for altered transmission in the Colombian multiplex and simplex families.

There was significant LD between SNPs genotyped in each population (Supplementary Material, Table S1). Both Caucasian (Supplementary Material, Fig. S1) and Hispanic (Supplementary Material, Fig. S2) populations contain an LD block encompassing the region spanning rs2641670–rs16974880. While they both also exhibit strong LD in the region defined by rs2326398–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 Material, Fig. S3) reveals a pattern similar to that of the Hispanics.

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 oropharynx and nasopharynx at E13.5 (Fig. 2A), the mandible at E14.5 (Fig. 2B) and the cartilage primordia of the nasal bones, palate and tooth germs at E17.5 (Fig. 2C). CRISPLD2 is expressed in the liver at E14.5 and all time points examined (Fig. 2D). No other organ systems showed significant levels of expression.

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 (19,21,25,26,28). 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 VWS (OMIM:119300), which is characterized by lower lip pits, CLP, CP or hypodontia (29). In addition, genetic variation in IRF6 has recently been shown to play an etiological role in the development of NSCLP (13–16). 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)] (31–34). The function of the LCCL domain is speculated to be either structural, immunological or involved in cell motility (32–34). Mutations in the LCCL domain of COCH have been identified in autosomal dominant non-syndromic sensorineural deafness disorder [DFNA9 (OMIM:601369)] (35). Akhirin, which also contains a LCCL domain in the N-terminus, is postulated to play a role in chicken retinal development (31). Recent research suggests that CLCP1 has a role in cellular motility and is regulated by ubiquitination (33). 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 (36).

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 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, an RNA polymerase binding or transcription factor activator/inhibitor binding; either could affect CRISPLD2 protein expression which would affect the developmental process (37,38). To determine if this is theoretically the case, two transcription binding site prediction programs, PATCH and AliBaba2, were used (39,40). 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 with NSCLP ($P = 0.02$ and $P = 0.06$, respectively; Table 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 (41). Also, synonymous changes have been shown to change amino acid translation time, resulting in altered protein structure and function (42). 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 (43,44). 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 (13,14,16,45–47). Together, this suggests that common genetic variation in non-coding 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 3). This suggests that these SNPs may be disease-causing variants or mutations exist that are in LD 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 1). This is not unexpected, as there is likely less admixture in the Colombian–Hispanic population compared to the Texas–Hispanic population (48). 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 (7).

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, oropharynx and nasopharynx and liver (Fig. 2). Thus, CRISPLD2 is expressed during facial development.

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 underway 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.

Materials and Methods

Study population and sample preparation

The study population consisted of 63 multiplex NSCLP families (56 Caucasian and seven Hispanic) and 287 simplex parent–child NSCLP trios (213 Caucasian and 74 Hispanic) who were ascertained at either the University of Texas Cranio-facial Clinic, Houston, Texas Children's Hospital, Houston or Children's Hospital, Boston. An additional study population consisted of 64 multiplex NSCLP families and 155 simplex parent–child NSCLP trios who were recruited from the Clinica Noel in Medellín, Colombia as previously described (28). All cases had isolated clefting and no syndromic cases of CL/P were included. This study has the approval of the Committee for the Protection of Human Subjects at the University of Texas Health Science Center at Houston (HSC-MS-03-090), the internal review boards at the Ohio State University and the University of Iowa (20010904 FWA00003007) and the scientific committee at the dental school of the University of Antioquia-Colombia.

Blood or saliva samples were obtained after informed consent. DNA was extracted from blood using the Roche DNA Isolation Kit for Mammalian Blood (Roche, Switzerland) or from saliva using Oragene Purifier (DNA Genotek Inc., Ontario, Canada) following the manufacturer's protocol.

Genotyping

Thirty-seven microsatellite markers spanning 10 autosomal regions were selected from Prescott *et al.* and genotyped using an ABI Prism 377, as previously described (26,27). Results for D16S3037 were analyzed using FBAT (49). IRF6 analysis was previously described (13).

SNPs flanking and within the IRF8 and CRISPLD2 genes were selected based on multiple criteria (1): the SNP is polymorphic within the population being tested (2), the SNP has a heterozygosity of at least 0.3 and a minor allele frequency above 0.2 and (3) the average distance between SNPs is less than 10 kb. TagSNPs were preferentially selected if they met the above criteria and were genotyped if they did not tag for other genotyped SNPs. TaqMan assays [Applied Biosystems (ABI); Foster City, CA, USA] were used to genotype SNPs

following the manufacturer's protocol and detected using ABI 7900HT Sequence Detection System. The data was exported to SDS Software v2.1 (ABI) for allele calling and then imported into Progeny Lab (South Bend, IN, USA) for data management and to check for non-Mendelian inheritance using PedCheck (50).

Statistical methods

Families were stratified according to reported ethnicity. In addition, for the association analyses, families were grouped according to the presence/absence of family history. Allele frequencies were calculated for the markers for each of the three ethnic groups using SAS. D' values were calculated for each ethnic group using GOLD (51) and visually inspected using the solid spine of LD method in Haploview (52). Parametric and non-parametric linkage analysis was performed using Simwalk2 and FBAT was used to test for linkage with association under a variety of models, including implementing the empirical variance option to correct for the presence of multiple nuclear families from a single pedigree. (49,53–55). Interactions between IRF8 and IRF6 were investigated using FBAT. Nominal *P*-values are reported after Bonferroni correction for multiple testing.

Two transcription binding site prediction programs, PATCH and AliBaba2, were run using preset parameters defined by these programs to determine if rs1546124 affects a putative promoter binding site (39,40).

Expression study

RIKEN mouse CRISPLD2 cDNA clone, 3321402M02, extracted from the head of an E17.5 C57BL/6 mouse, was obtained form 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, 54°C 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 *Not*I and *Spe*I 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 μm thick. Sections showing craniofacial structures were used in this study. Antisense CRISPLD2 clones were labeled with ^{35}S and *in situ* hybridizations were carried out using standard protocols to detect CRISPLD2 expression (56). Sense CRISPLD2 clones were used as a negative control.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Gorlin, RJ.; Cohen, MM.; Hennekam, RCM., editors. Syndromes of the Head and Neck. 4th. Oxford University Press; New York: 2001.

2. 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. [PubMed: 15779018]

3. 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. [PubMed: 15955936]

4. 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. [PubMed: 11283972]

5. Lidral AC, Moreno LM. Progress toward discerning the genetics of cleft lip. Curr Opin Pediatr. 2005; 17:731–739. [PubMed: 16282779]

6. 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. [PubMed: 1642234]

7. 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. [PubMed: 12454800]

8. Miettinen PJ, Chin JR, Shum L, Slavkin HC, Shuler CF, Derynck R, Werb Z. Epidermal growth factor receptor function is necessary for normal craniofacial development and palate closure. Nat Genet. 1999; 22:69–73. [PubMed: 10319864]

9. Tang LS, Finnell RH. Neural and orofacial defects in Folp1 knockout mice [corrected]. Birth Defects Res. 2003; 67:209–218.

10. MacDonald ME, Abbott UK, Richman JM. Upper beak truncation in chicken embryos with the cleft primary palate mutation is due to an epithelial defect in the frontonasal mass. Dev Dyn. 2004; 230:335–349. [PubMed: 15162512]

11. Hilliard SA, Yu L, Gu S, Zhang Z, Chen YP. Regional regulation of palatal growth and patterning along the anterior–posterior axis in mice. J Anat. 2005; 207:655–667. [PubMed: 16313398]

12. 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. [PubMed: 15716346]

13. 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. [PubMed: 16096995]

14. 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. [PubMed: 15558496]

15. Zucchero TM, Cooper ME, Maher BS, Daack-Hirsch S, Nepomuceno B, Ribeiro L, Caprau D, Christensen K, Suzuki Y, Machida J, et al. Interferon regulatory factor 6 (IRF6) is a modifier for isolated cleft lip and palate. Am J Hum Genet. 2003; 73:162. [PubMed: 12796853]

16. Zucchero TM, Cooper ME, Maher BS, Daack-Hirsch S, Nepomuceno B, Ribeiro L, Caprau D, Christensen K, Suzuki Y, Machida J, et al. Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. N Engl J Med. 2004; 351:769–780. [PubMed: 15317890]

17. 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. [PubMed: 15329828]

18. 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. [PubMed: 15690355]

19. 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. [PubMed: 15378549]

20. 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. [PubMed: 12087515]

21. 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. [PubMed: 15057975]

22. Marazita ML, Murray JC, Lidral AC, Arcos-Burgos M, Cooper ME, Goldstein T, Maher BS, Daack-Hirsch S, Schultz R, Mansilla MA, et al. 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. [PubMed: 15185170]

23. Prescott NJ, Malcolm S. Folate and the face: evaluating the evidence for the influence of folate genes on craniofacial development. Cleft Palate Craniofac J. 2002; 39:327–331. [PubMed: 12019010]

24. Riley BM, Schultz RE, Cooper ME, Goldstein-McHenry T, Daack-Hirsch S, Lee KT, Dragan E, Vieira AR, Lidral AC, Marazita ML, et al. 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. [PubMed: 17366557]

25. Wyszynski DF, Albacha-Hejazi H, Aldirani M, Hammod M, Shkair H, Karam A, Alashkar J, Holmes TN, Pugh EW, Doheny KF, et al. A genome-wide scan for loci predisposing to non-syndromic cleftlip with or without cleft palate in two large Syrian families. Am J Med Genet. 2003; 123A:140–147. [PubMed: 14598337]

26. 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. [PubMed: 10798365]

27. Blanton SH, Bertin T, Patel S, Stal S, Mulliken JB, Hecht JT. Nonsyndromic cleft lip and palate: four chromosomal regions ofinterest. Am J Med Genet A. 2004; 125:28–37. [PubMed: 14755463]

28. 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.

29. 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. [PubMed: 14618417]

30. Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL, Daack-Hirsch S, Sander A, et al. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet. 2002; 32:285–289. [PubMed: 12219090]

31. 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. [PubMed: 15765510]

32. 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. [PubMed: 11574466]

33. 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; 26:4025–4031. [PubMed: 17213806]

34. Trexler M, Banyai L, Patthy L. The LCCL module. Eur J Biochem. 2000; 267:5751–5757. [PubMed: 10971586]

35. 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. [PubMed: 11709536]

36. 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. [PubMed: 16990542]

37. Faniello MC, Fregola A, Nistico A, Quaresima B, Crugliano T, Faraonio R, Puzzonia P, Baudi F, Parlato G, Cuda G, et al. 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. [PubMed: 16797877]

38. 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 USA. 2006; 103:6747–6752. [PubMed: 16618933]

39. Grabe N. AliBaba2: context specific identification of transcription factor binding sites. Silico Biol. 2002; 2:S1–S15.

40. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 2006; 34:D108–D110. [PubMed: 16381825]

41. Nielsen KB, Sorensen S, Cartegni L, Corydon TJ, Doktor TK, Schroeder LD, Reinert LS, Elpeleg O, Krainer AR, Gregersen N, et al. 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. [PubMed: 17273963]

42. 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. [PubMed: 17185560]

43. 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. [PubMed: 15131266]

44. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, et al. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 2005; 3:e7. [PubMed: 15630479]

45. 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. [PubMed: 14557473]

46. Li M, Atmaca-Sonmez P, Othman M, Branham KE, Khanna R, Wade MS, Li Y, Liang L, Zareparsi S, Swaroop A, et al. CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. Nat Genet. 2006; 38:1049–1054. [PubMed: 16936733]

47. 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. [PubMed: 16936732]

48. Arcos-Burgos M, Muenke M. Genetics of population isolates. Clin Genet. 2002; 61:233–247. [PubMed: 12030885]

49. 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. [PubMed: 8651310]

50. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet. 1998; 63:259–266. [PubMed: 9634505]

51. Abecasis GR, Cookson WO. GOLD–graphical overview of linkage disequilibrium. Bioinformatics. 2000; 16:182–183. [PubMed: 10842743]

52. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–265. [PubMed: 15297300]

53. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet. 1996; 58:1347–1367. [PubMed: 8651312]

54. Horvath S, Laird NM. A discordant-sibship test for disequilibrium and linkage: no need for parental data. Am J Hum Genet. 1998; 63:1886–1897. [PubMed: 9837840]

55. 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. [PubMed: 14557988]

56. Shen MM. Identification of differentially expressed genes in mouse development using differential display and *in situ* hybridization. Methods. 2001; 24:15–27. [PubMed: 11327798]
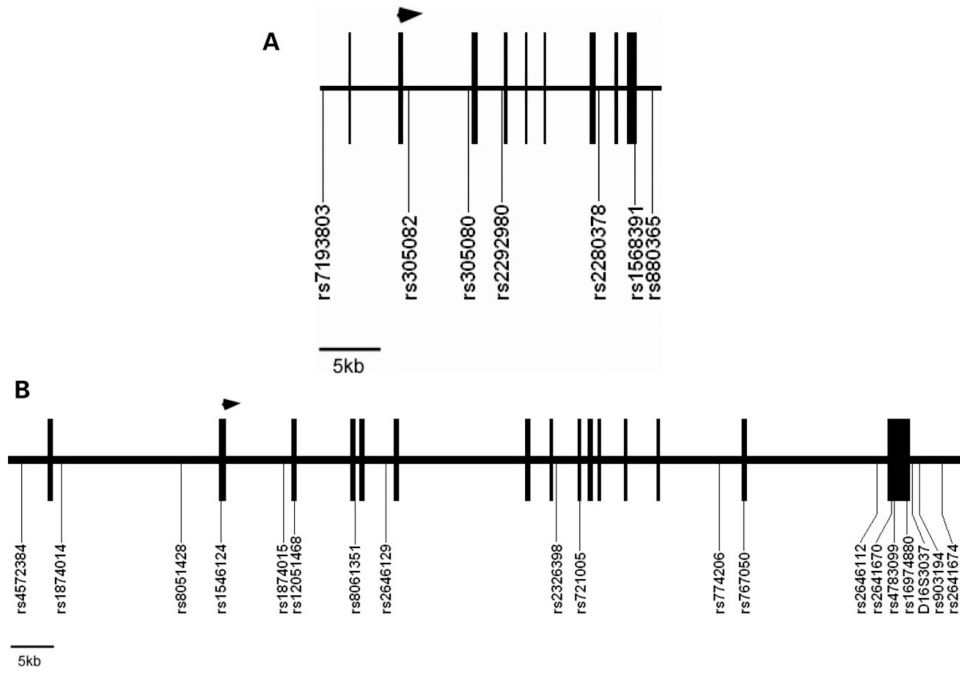
Figure 1.
SNPs genotyped in candidate genes. (**A**) IRF8. (**B**) CRISPLD2. STR D16S3037 is shown at 3′ of the CRISPLD2. Arrows denote transcription start site.
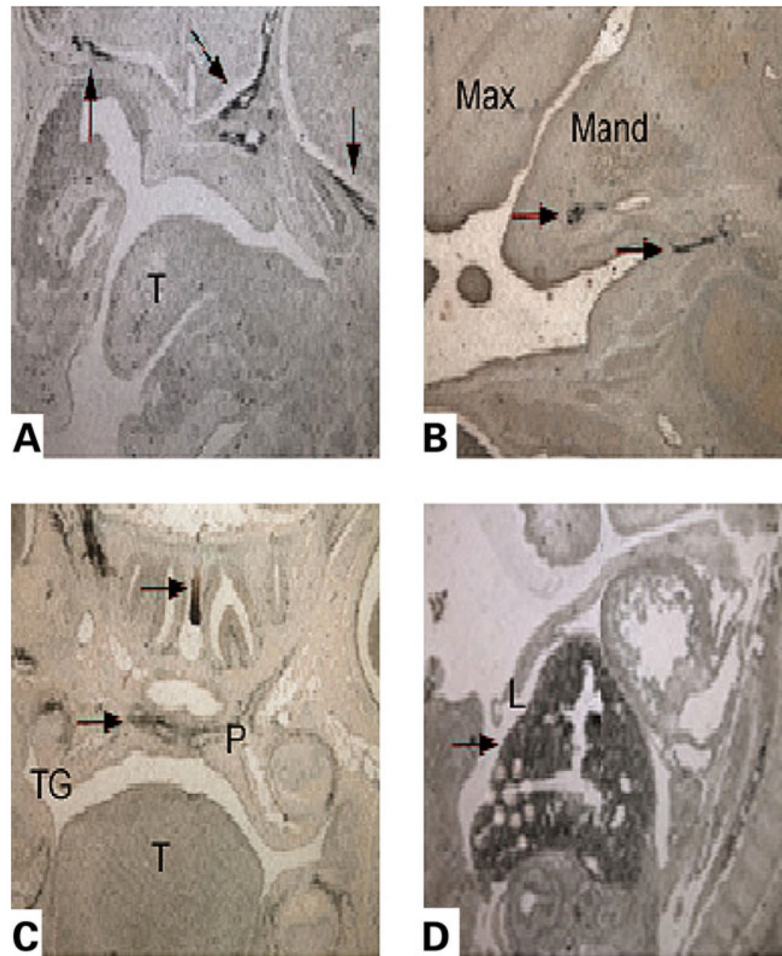
Figure 2.
CRISPLD2 expression mouse *in situ* hybridizations show CRISPLD2 expression in (**A**) the naso- and oropharynx at E13.5, (**B**) the mandible at E14.5, (**C**) the palate and cartilage primordium of the nasal septum at E17.5 and (**D**) the liver at all stages tested (E14.5 shown). Expression is denoted by arrows. T, tongue; Max, maxilla; Mand, mandible; P, palate; TG, tooth germ; L, liver.

Table 1

Allele frequencies

Gene	dbSNP	Alleles[a]	TagSNP[b]	Major allele frequencies		
				Caucasian	Hispanic	Colombian
IRF8	rs7193803	G/T	Yes	0.694	0.600	NT
IRF8	rs305082	T/C	Yes	0.827	0.833	NT
IRF8	rs305080	C/T	No	0.717	0.692	NT
IRF8	rs2292980	A/G	Yes	0.714	0.692	NT
IRF8	rs2280378	C/T	Yes	0.582	0.577	NT
IRF8	rs1568391	T/G	Yes	0.553	0.200 ***	NT
IRF8	rs880365	C/T	Yes	0.790	0.533 *	NT
CRISPLD2	rs4572384	T/C	No	0.604	0.283 ***	0.497 ***
CRISPLD2	rs1874014	A/C	Yes	0.557	0.739 ***	NT
CRISPLD2	rs8051428	C/T	No	0.645	0.505 **	NT
CRISPLD2	rs1546124	C/G	No	0.702	0.678	0.706
CRISPLD2	rs1874015	T/C	No	0.693	0.807 *	0.249 ***
CRISPLD2	rs12051468	A/G	Yes	0.546	0.687 **	0.662
CRISPLD2	rs8061351	T/C	Yes	0.716	0.680	0.607
CRISPLD2	rs2646129	A/G	Yes	0.632	0.815 ***	NT
CRISPLD2	rs2326398	A/G	Yes	0.656	0.688	0.685
CRISPLD2	rs721005	C/G	Yes	0.604	0.652	0.642
CRISPLD2	rs774206	A/G	Yes	0.636	0.717	0.666
CRISPLD2	rs767050	G/A	Yes	0.555	0.617	0.580
CRISPLD2	rs2646112	G/A	No	0.918	0.871	NT
CRISPLD2	rs2641670	G/A	Yes	0.790	0.629 ***	0.734
CRISPLD2	rs4783099	C/T	No	0.636	0.577	0.588
CRISPLD2	rs16974880	T/G	No	0.699	0.629	0.684
CRISPLD2	rs903194	T/G	No	0.603	0.709	NT
CRISPLD2	rs2641674	T/C	Yes	0.627	0.697	NT

NT denotes not typed.

[a] Most common Caucasian allele listed first.

[b] TagSNP status.

* *P*-value 0.0125.

** *P*-value <0.001.

*** *P*-value <0.0001.

Table 2
Caucasian and Hispanic CRISPLD2 FBAT results

dbSNP	Caucasian		Hispanic	
	P_{raw}	$P_{corrected}$	P_{raw}	$P_{corrected}$
rs4572384	0.83222	0.82495	0.53162	0.54851
rs1874014	0.35645	0.36803	0.85268	0.85268
rs8051428	0.67260	0.68556	0.09345	0.09345
rs1546124	0.00064	0.01236	0.53162	0.53162
rs1874015	0.45468	0.51915	0.46685	0.53162
rs12051468	0.95995	0.96676	0.50499	0.44969
rs8061351	0.69489	0.76134	0.02799	0.02364
rs2646129	0.64587	0.69972	0.22525	0.16552
rs2326398	0.10863	0.18851	0.04819	0.05551
rs721005	0.15880	0.24937	0.08635	0.08635
rs774206	0.48794	0.55750	0.14413	0.11666
rs767050	0.77444	0.77671	0.41108	0.42334
rs2646112	0.17793	0.26233	0.80837	0.80837
rs2641670	0.21527	0.12032	0.46521	0.44969
rs4783099	0.08149	0.16929	0.20590	0.21704
rs16974880	0.03014	0.06766	0.15985	0.18926
rs903194	0.29128	0.35525	0.81855	0.80837
rs2641674	0.93689	0.93669	0.37109	0.31731

Table 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