

FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate

Lina M. Moreno^{1,2,†}, Maria Adela Mansilla^{3,†}, Steve A. Bullard¹, Margaret E. Cooper⁴, Tamara D. Busch¹, Junichiro Machida¹, Marla K. Johnson³, David Brauer³, Katherine Krahn¹, Sandy Daack-Hirsch³, Jamie L'Heureux³, Consuelo Valencia-Ramirez⁵, Dora Rivera⁵, Ana Maria López⁵, Manuel A. Moreno⁶, Anne Hing⁷, Edward J. Lammer⁸, Marilyn Jones⁹, Kaare Christensen¹⁰, Rolv T. Lie¹¹, Astanand Jugessur¹², Allen J. Wilcox¹³, Peter Chines¹⁴, Elizabeth Pugh¹⁵, Kim Doheny¹⁵, Mauricio Arcos-Burgos^{16,†}, Mary L. Marazita^{4,17,†}, Jeffrey C. Murray^{3,†} and Andrew C. Lidral^{1,*,†}

¹Department of Orthodontics, College of Dentistry, University of Iowa, Iowa City, IA 52242, USA, ²Dows Institute for Dental Research, College of Dentistry, University of Iowa, Iowa City, IA 52242, USA, ³Department of Pediatrics, Roy C. Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA, ⁴Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15260, USA, ⁵College of Dentistry, University of Antioquia, Medellín, Colombia, ⁶Laboratorio de Genética Forense y Huellas Digitales del DNA, Medellín, Colombia, ⁷Department of Pediatrics, University of Washington and Craniofacial Center, Children's Hospital and Regional Medical Center, Seattle, WA 98195, USA, ⁸Departments of Genetics and Pediatrics, University of Alabama, Birmingham, AL 35294, USA, ⁹Children's Hospital Oakland Research Institute, Oakland, CA 94609, USA, ¹⁰Department of Epidemiology, University of Southern Denmark, Odense DK-5230, Denmark, ¹¹Section for Epidemiology and Medical Statistics, Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway, ¹²Craniofacial Development, Musculoskeletal Disorders, Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, VIC 3052, Australia, ¹³Epidemiology Branch, National Institute of Environmental Health Sciences (NIEHS), Durham, NC 27709, USA, ¹⁴Genome Technology Branch, Division of Intramural Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA, ¹⁵Center for Inherited Disease Research (CIDR), Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21224, USA, ¹⁶Department of Psychiatry and Behavioral Sciences, Leonard M. Miller School of Medicine, University of Miami, Miami, FL 33136, USA and ¹⁷Department of Human Genetics, Graduate School of Public Health and Clinical and Translational Sciences Institute, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received May 22, 2009; Revised and Accepted September 21, 2009

Nonsyndromic orofacial clefts are a common complex birth defect caused by genetic and environmental factors and/or their interactions. A previous genome-wide linkage scan discovered a novel locus for cleft lip with or without cleft palate (CL/P) at 9q22–q33. To identify the etiologic gene, we undertook an iterative and complementary fine mapping strategy using family-based CL/P samples from Colombia, USA and the Philippines. Candidate genes within 9q22–q33 were sequenced, revealing 32 new variants. Concurrently, 397 SNPs spanning the 9q22–q33 2-LOD-unit interval were tested for association. Significant SNP and haplotype association signals ($P = 1.45E - 08$) narrowed the interval to a 200 kb region containing: FOXE1, C9ORF156 and HEMGN. Association results were replicated in CL/P families of European descent and when all populations

*To whom correspondence should be addressed at: Department of Orthodontics, College of Dentistry, University of Iowa, 2174 Medical Laboratories, Iowa City, IA, USA. Tel: +1 3193357304; Fax: +1 3193356970; Email: andrew-lidral@uiowa.edu

†These authors contributed equally to the work.

were combined the two most associated SNPs, rs3758249 ($P = 5.01E - 13$) and rs4460498 ($P = 6.51E - 12$), were located inside a 70 kb high linkage disequilibrium block containing *FOXE1*. Association signals for Caucasians and Asians clustered 5' and 3' of *FOXE1*, respectively. Isolated cleft palate (CP) was also associated, indicating that *FOXE1* plays a role in two phenotypes thought to be genetically distinct. *Foxe1* expression was found in the epithelium undergoing fusion between the medial nasal and maxillary processes. Mutation screens of *FOXE1* identified two family-specific missense mutations at highly conserved amino acids. These data indicate that *FOXE1* is a major gene for CL/P and provides new insights for improved counseling and genetic interaction studies.

INTRODUCTION

Isolated cleft lip with or without cleft palate (CL/P) is a common complex birth defect that results from genetic variations, environmental exposures and/or their interactions. Family-based studies have estimated that CL/P genetic risk involves 3–14 loci interacting multiplicatively (1). Isolated cleft palate (CP) has often been separated in analysis from CL/P based on embryologic and recurrence-risk evidence although overlap in causal genes has been observed in some syndromic forms of clefting (2). Multicenter collaborative studies, utilizing both linkage and association approaches applied to substantial numbers of multiplex and nuclear triad families, have significantly advanced knowledge of locus, gene and environmental exposures to CL/P etiology (3–12). Discoveries include the strong effect of common *IRF6* variants (9), the collective occurrence of rare DNA variations in 20 genes: *TGFB3*, *MSX1*, *PTCH*, *FOXE1*, *GLI2*, *JAG2*, *LHX8*, *MSX1*, *MSX2*, *SATB2*, *SKI*, *SPRY2*, *TBX10*, *PVRL1*, *FGFR1*, *FGFR2*, *FGF8*, *CRISPLD2*, *BMP4* and *RYK* (13–23) and significant association or linkage to various genes (24). Most recently published genome-wide association scans for NSCL/P have identified a major locus at 8q24.21 (25,26) that confers a substantial risk in European populations.

Maternal periconceptional smoking is a well-established environmental risk factor (27); alcohol use and poor nutrition have also been suggested as risks factors for oral clefts but the data for these are weaker (28). In addition, gene-environment studies suggested that risk due to smoking may be influenced by variants in genes such as *GSTM1*, *GSTT1*, *NAT1* and *NAT2*, which are involved in the detoxification of tobacco smoking products (6,29–31).

To identify additional CL/P risk genes, we conducted a genome-wide linkage scan on 388 multiplex families from seven populations, revealing highly significant evidence of linkage to the 9q22–q33 region with an HLOD of 5.5 (11) (Fig. 1). Potential candidate genes based on gene expression or known role in human clefting syndromes mapping within the region include *ROR2* (104 cM), *FGD3*, *BARX1*, *PTCH*, *FOXE1*, *GABABR2*, *TGFBR1* and *ZNF189* (120 cM).

We describe here the detailed fine mapping of the 9q22–q33 region, using parallel positional cloning and candidate gene strategies, leading to the discovery that common *FOXE1* locus variants have a significant role in the etiology of nonsyndromic orofacial clefting.

RESULTS

Stage 1: 2-LOD-unit support interval association mapping

Analysis of the Illumina SNP panel revealed the strongest association results within *FOXE1* from Colombia and USA_1 families (for rs1443434, $P = 8.30E - 04$ and $P = 6.69E - 03$, respectively) (Fig. 2A and Supplementary Material, Appendix Table S1). Other association signals were noted within the *IARS* gene in Colombians and *PALM2* in Filipinos. Upon combining all three populations (combined data), rs1443434 located in the *FOXE1* 3'-UTR returned the most significant evidence of association ($P = 3.8E - 05$), indicating the presence of a common variant within the approximately 70 kb long haplotype block containing *FOXE1*.

Stage 2: candidate genes

Convergent association to *FOXE1* was also observed through analysis of the candidate genes within the initial linkage peak, (Fig. 2B, and Supplementary Material, Appendix Table S2). Two SNPs, rs894673 and rs3758249, within 3.7 kb 5' of *FOXE1* contributed the strongest results in both the Colombian and USA_1 populations (P -values range from $8.67E - 03$ to $1.8E - 04$). Although combining the Filipino data set with these two populations increased the significance to P -values $< E - 06$ at both rs894673 ($P = 8.25E - 07$) and rs3758249 ($P = 8.64E - 07$) the strongest signal for the Filipinos was within the *PTCH* gene at rs2297088 (97.28 Mbp, P -value = $6.49E - 03$), (Fig. 2B, and Supplementary Material, Appendix Table S2). Of note, over-transmission of the *PTCH* rs2297088–rs2236407 haplotype with borderline statistical significance ($P = 0.08$) in Filipino families with two or more affected members has been reported previously (16).

Stage 3: the critical region *XPA-HEMGN*

Given the strong association to a region at most 357 kb in size flanked by the *XPA* and *HEMGN* genes and containing *FOXE1*, additional SNPs were genotyped to further narrow the critical region. Reconstruction of linkage disequilibrium (LD) blocks in all populations (Fig. 3A–C) showed similar LD patterns for Colombia and USA, while LD blocks of smaller size were observed for the Filipinos (Fig. 3C). Association results excluded the *XPA* gene for all populations, refining the critical region to 200 kb (rs1877431–rs4618817; 99.57–99.77 Mbp) containing *FOXE1*, *C9ORF156* and

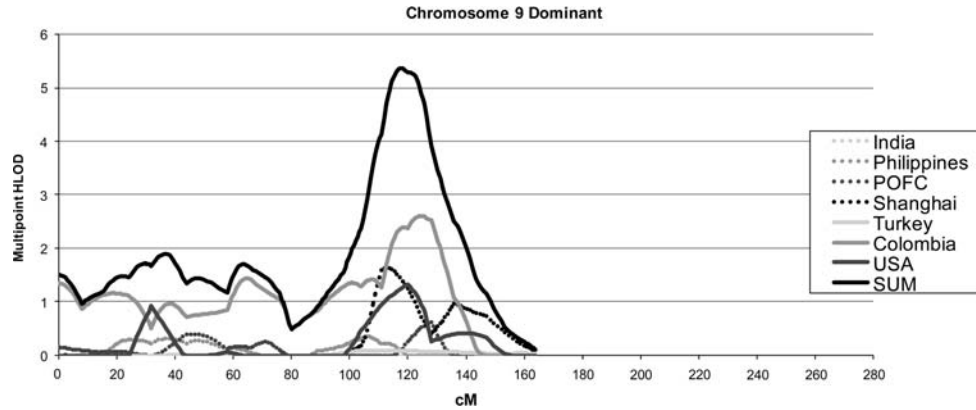


Figure 1. Maximum summed multipoint HLODs on chromosome 9 under a dominant model obtained from the 10 cM genome scan of CL/P in the CIDR-7 studies. Each population's contribution to the total HLOD is detailed elsewhere (11,68).

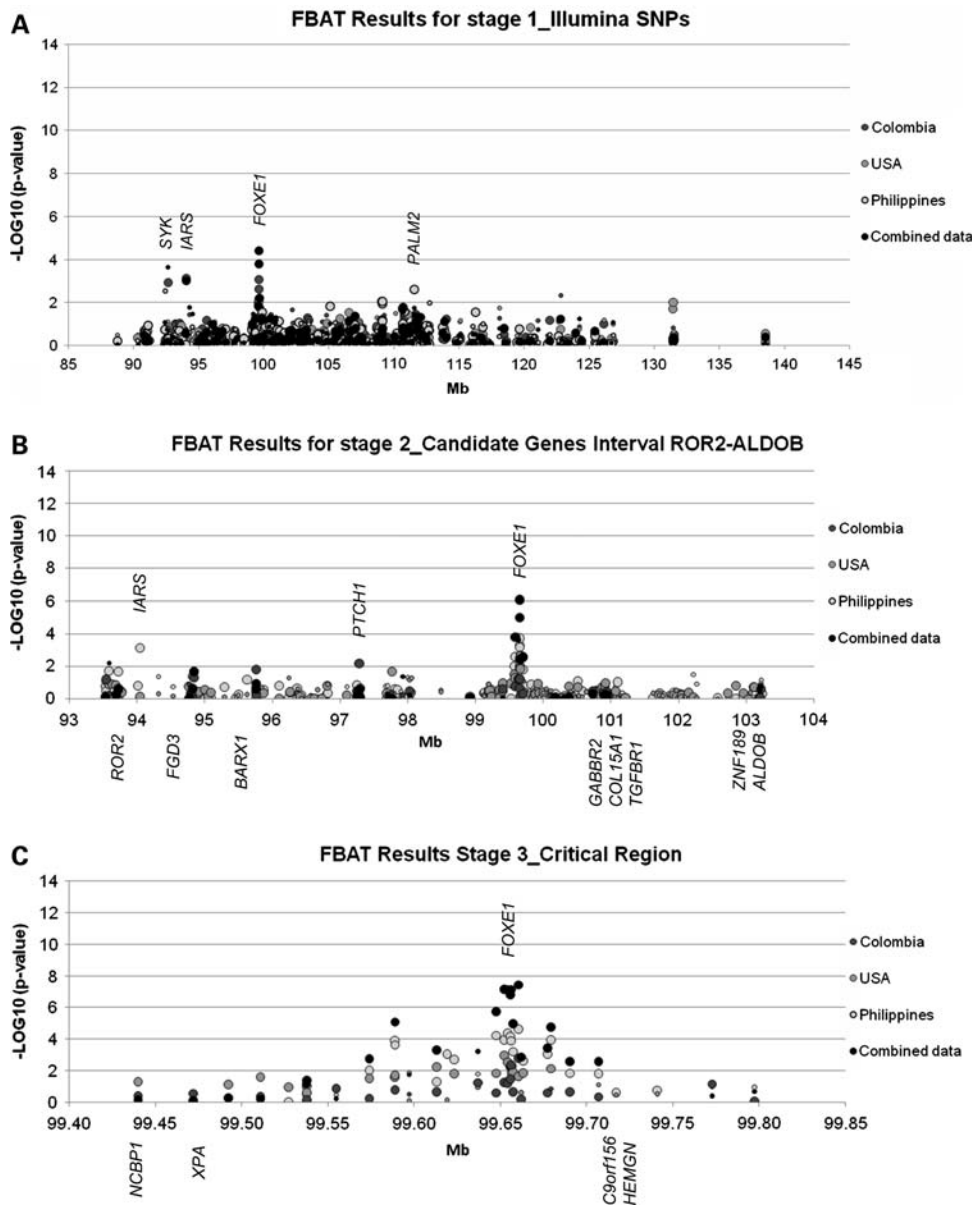
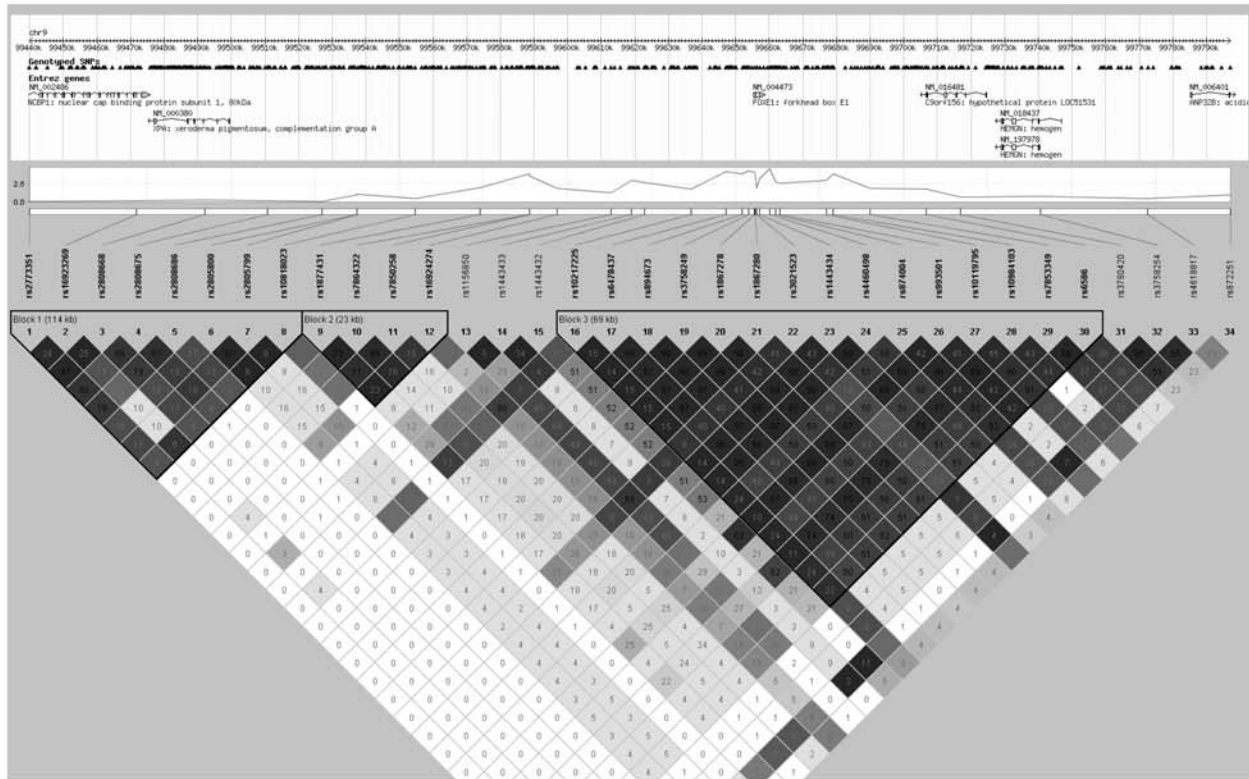


Figure 2. FBAT association results. (A) 331 Illumina markers including the 2LOD interval. (B) Candidate genes interval between *ROR2-ALDOB* (178 SNPs). (C) 357 kb critical region (34 SNPs). The size of the circles represents the over-transmission of the common (large circles) or rare allele (small circles).

A



B

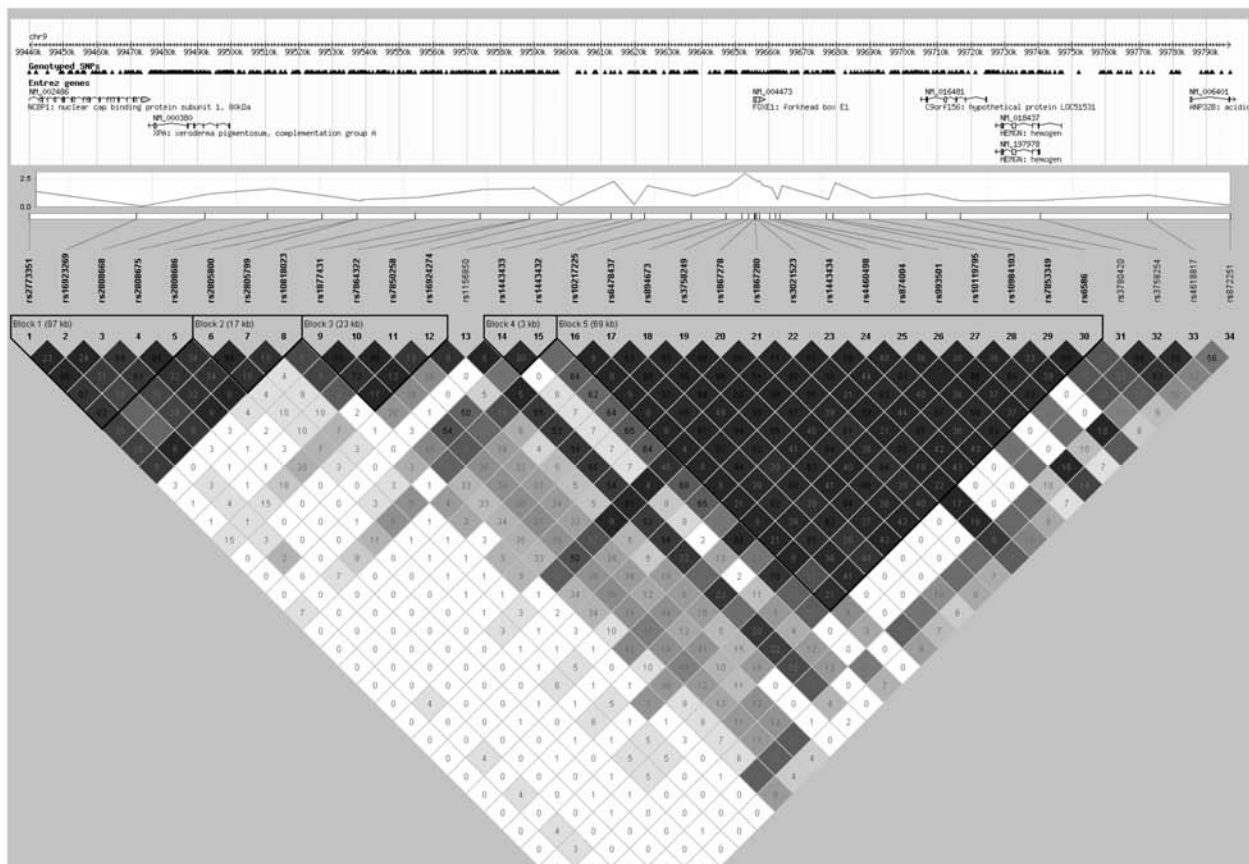


Figure 3. LD Blocks for Stage 3—357 kb critical region SNPs. FBAT association results $-\log_{10}(P\text{-value})$ are displayed by upper track. (A) Colombia, (B) the USA, (C) the Philippines.

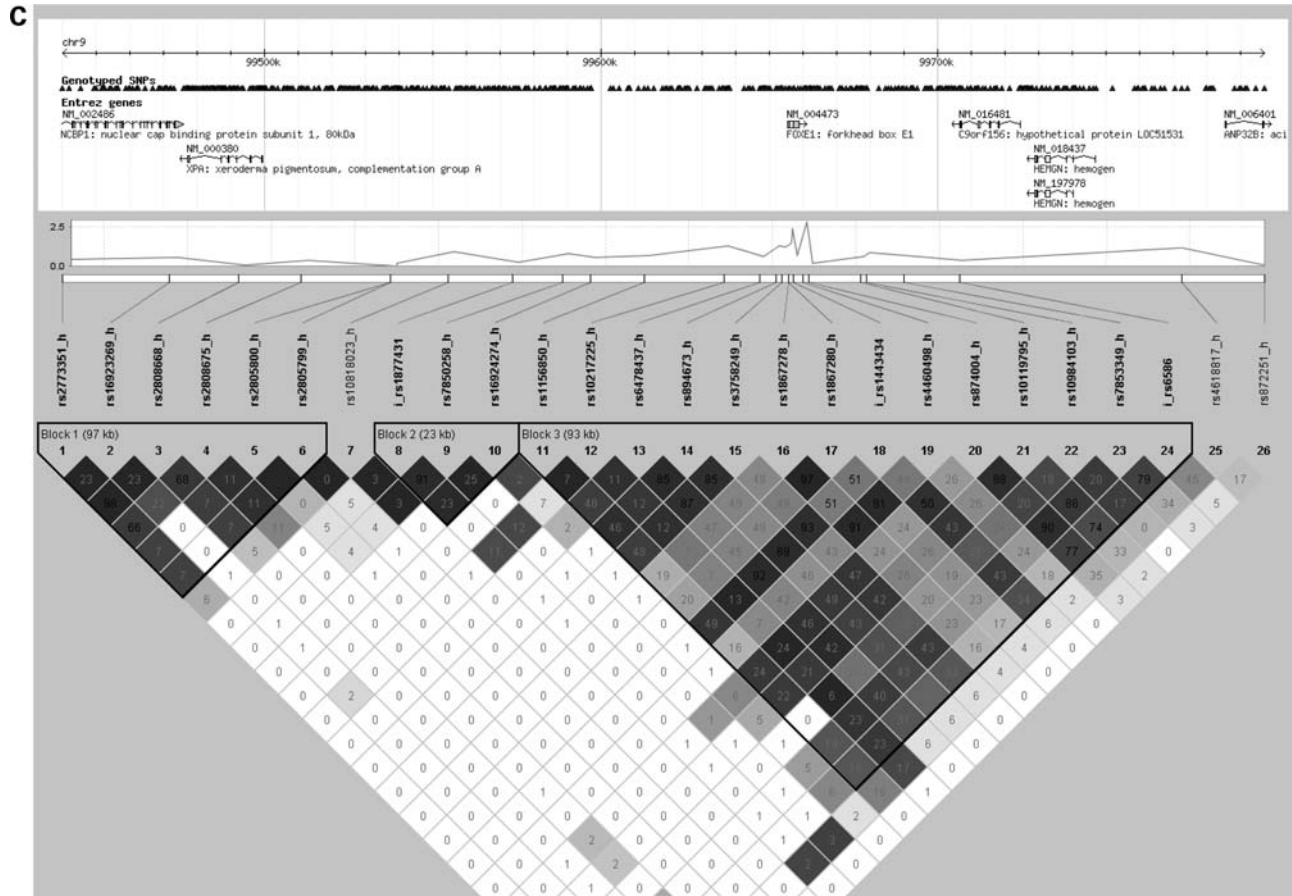


Figure 3. Continued.

HEMGN (Fig. 2C). The highest significance occurred within a 70 kb block (rs10217225–rs6586; 99.634–99.707 Mbp) encompassing *FOXE1* for the Colombian (rs4460498, $P = 2.3E - 05$) and USA_1 families (rs894673, $P = 1.12E - 03$) (Fig. 2C and Supplementary Material, Appendix Table S3). Less significant results were observed in the Filipinos (rs4460498, $P = 1.58E - 03$) possibly a reflection of low minor allele frequencies. Combining all three populations produced highly significant results with markers rs894673 ($P = 7.27E - 08$), rs3758249 ($P = 7.64E - 08$), rs1867280 ($P = 7.49E - 08$), and rs4460498 ($P = 3.73E - 08$).

Stage 3: haplotype analysis

For the Colombian and the USA_1 samples, the most significant haplotypes were clustered in a region spanning 9.5 kb immediately 5' of *FOXE1*, with some haplotypes extending into the gene (Fig. 4A and B). In addition, moderately significant haplotypes were found between rs1877431 and rs16924274 at –82 to –59 kb upstream and another between rs993501 and rs6586 spanning 43.5 kb starting 7.2 kb downstream of *FOXE1*. The most significant Filipino haplotypes occurred between rs874004 and rs872251, a 135.2 kb region starting 6 kb 3' of *FOXE1* (Fig. 4C). Given

several associated haplotypes overlapped between all three data sets, populations were combined yielding increased significance beyond each population alone, confirming that haplotypes immediately 5' of *FOXE1* showed the strongest association ($P = 1.45E - 08$) (Fig. 4D).

Stage 4: replication panel analysis

Markers surrounding *FOXE1* ($n = 15$), included in haplotypes with significant results for the Colombian and USA_1 populations, were genotyped in an independent set of CL/P samples that included subjects from Denmark, Norway and Iowa. Results obtained for Denmark (rs3021523 $P = 2.18E - 03$) and Norway (rs4460498 $P = 4.96E - 03$) showed independent replication of the initial *FOXE1* association signal (Supplementary Material, Appendix Table S6). Interestingly, rs4460498 was also the most strongly associated marker for the Colombian and Filipino data sets ($P = 2.3E - 05$ and $1.90E - 03$, respectively). Iowa samples were pooled with the additional USA_1 samples, and the results (USA_2) were significant for rs3758249 ($P = 5.60E - 05$; Bonferroni correction = $1.67E - 04$), replicating the association with markers in *FOXE1* (Supplementary Material, Appendix Table S6).

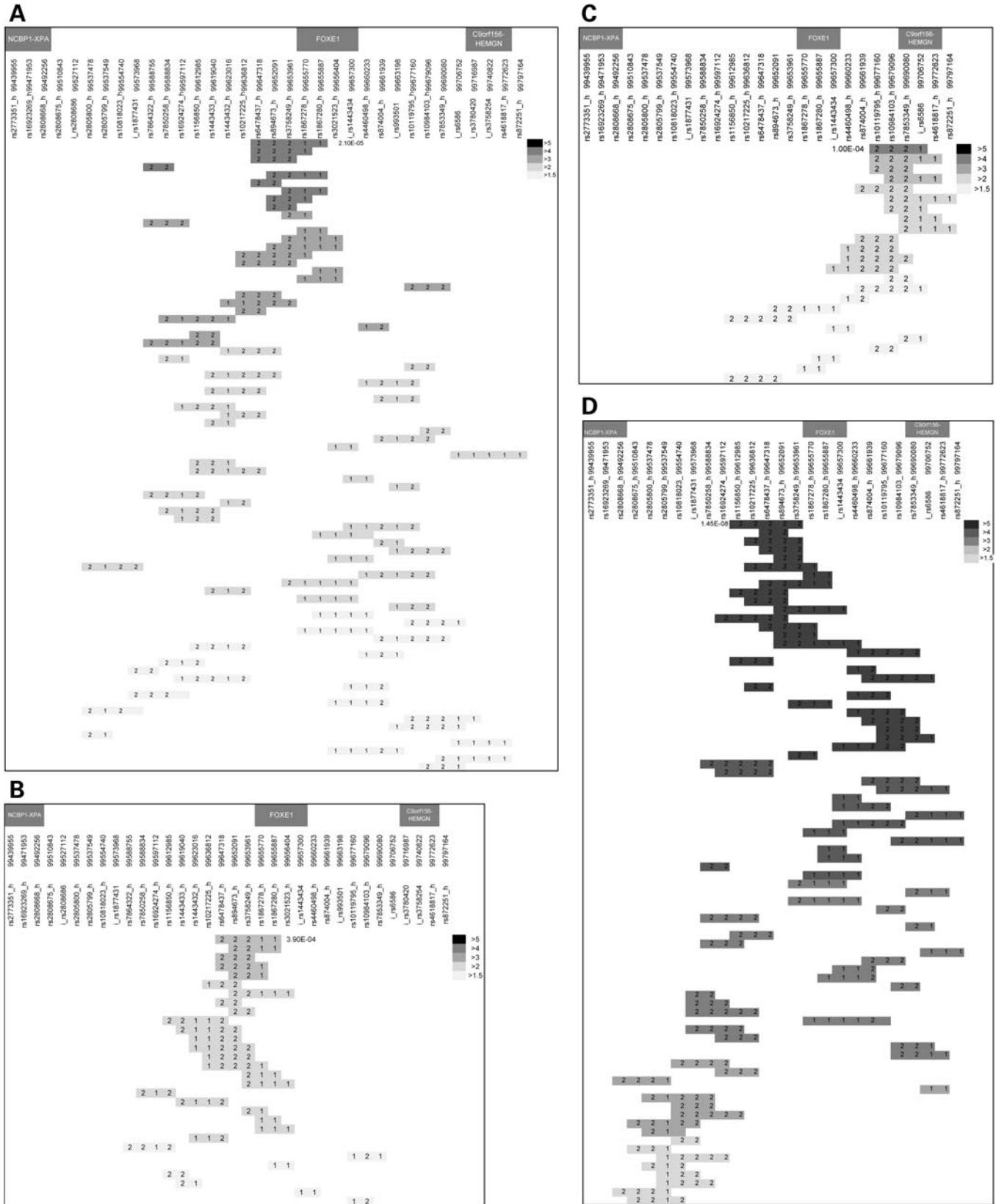


Figure 4. Haplotype window analysis results for Stage 3—357 kb critical region. Rectangles represent the size of the haplotype according to SNPs involved. Gray scale represent the $-\log_{10}(P\text{-value})$ results. (A) Colombia (B) USA (C) Philippines (D) Combined. See Supplementary Material, Appendix Table S1 for allele correspondence to specific nucleotides.

Table 1. Number of pedigrees by cleft phenotypic subgroup

Population	Clefting subgroups			CP
	<i>CLO</i>	<i>CLPO</i>	<i>CL/P</i>	
Colombia	28	214	272	0
USA_2	102	175	324	140
Philippines	69	178	298	70
Denmark	191	200	395	196
Norway	122	192	314	118

Italicized subgroups are overlapping. Clefting subgroups: *CLO*, pedigrees with cleft lip only affected individuals; *CLPO*, pedigrees with *CLP* only affected individuals; *CL/P*, pedigrees in which members are affected with at least a cleft lip; *CP*, pedigrees with at least one cleft palate only affected individual. Populations: Denmark (proband nuclear triads), Norway (proband nuclear triads, with siblings and step siblings), the USA (extended pedigrees and proband nuclear triads), Colombia (extended pedigrees and proband nuclear triads), the Philippines (proband nuclear triads). Pooled analysis: Scandinavia = Denmark + Norway; Caucasian = Scandinavia + the USA; Non-Asian = Caucasian + Colombia.

FOXE1 analysis by phenotype

To test the hypothesis that *FOXE1* has a role in isolated *CP*, DNA from cases with *CP* (Table 1) were genotyped for the 15 SNPs surrounding *FOXE1*. Association with *CP* was observed in the USA, Filipino, Norwegian and Danish families (Colombian data set did not have *CP* families) and when combining all populations, rs1867278 was the most significant ($P = 4.1E - 4$) (Fig. 5E), indicating that *FOXE1* is also involved in the etiology of isolated *CP*.

To explore whether the association for the *CL/P* sample was due to a specific subset type of cleft, the entire data set was analyzed across all cleft phenotypes (Fig. 5A–E). The *CLO* subset consists of those families in which one or more of the affected family members have cleft lip alone, while all affected family members in the *CLPO* subset have cleft lip and cleft palate. The *CL/P* subgroup is a combination of the *CL* and *CLPO* subgroups, plus families with both *CL* and *CLP*. The families in the *CP* subgroup have at least one affected individual with isolated cleft palate. *CLPO* was more significantly associated in the USA and Norwegian families, while the Colombian and Danish families showed similar association for both *CLO* and *CLPO*. Of note, the Colombian data set had 10-fold fewer *CLO* families than *CLPO* families. In contrast, no association was found with Filipino *CLO* families (Fig. 5B and C, Supplementary Material, Appendix Table S6). When all populations were combined the same markers, rs3756249, rs1867278 and rs4460498 were the most significant in each phenotypic subgroup. Combining all phenotypes and all populations, the most significant results occurred at rs3758249 ($P = 5.01E - 13$), rs1867278 ($P = 7.44E - 13$) and rs4460498 ($P = 6.51E - 12$). The over-transmission percentage was highest at rs3758249 and rs4460498 (Fig. 6) that are located 2.0 kb 5' and 4.2 kb 3' of *FOXE1*, respectively, whereas rs1867278 is within the 5'-UTR. Haplotype analysis using sliding windows in the *CL/P* families indicate the presence of two disease haplotypes, one extending 19.2 kb 5' involving markers rs10217225–rs6478437–rs894673–rs3758249–

rs1867278–rs1867280–rs3021523 (2-2-2-2-1-1-1, corresponding to T-G-T-G-A-C-C, all common alleles except rs10217225), that was the most significant in Colombian, USA, Norwegian and Danish families. Another one located 21 kb 3' and spanning 120 kb consisting of rs10119795–rs10984103–rs7853349–rs6586–rs4618817–rs872251 (2-2-2-1-1-1, corresponding to T-C-T-T-A-A) was the most significant in Filipinos (Figs 4 and 7). This latter haplotype was also observed in the other populations albeit with less significance.

Allele dosage effects

A trend towards a dosage effect was seen for the presence of two copies relative to one copy of the over-transmitted alleles for each marker (rs3758249 and rs4460498), for both associated phenotypes (*CL/P* and *CLPO*; see Supplementary Material, Appendix Table S7). However, as the overlapping confidence intervals indicate (Supplementary Material, Appendix Table S7), this difference is not statistically significant.

Parent of origin effects

Marked parent of origin effects were seen for the Colombian and Philippines population with both *CL/P* (Fig. 8 and Supplementary Material, Appendix Table S4) and *CLPO* (not shown) phenotypes and these *FOXE1* alleles. Over-transmission was seen preferentially from mothers compared with fathers. On the other hand, preferential over-transmission was observed among the fathers within the Caucasian data (USA and Scandinavian data sets) possibly indicating different roles of these alleles in different populations (Fig. 8).

Sequencing results

Sequencing of conserved non-coding regions in five candidate genes in probands from families that yielded positive LOD scores in the 9q22–33 region (11) (63 Filipino, 28 Colombians and 1 USA) identified 24 new variants. Sequencing of the coding regions in *FOXE1* found eight new variants, two are missense mutations (I59S and P208R), six are synonymous changes (Supplementary Material, Appendix Table S8). Both missense mutations are predicted to be benign by PolyPhen <http://genetics.bwh.harvard.edu/pph/>. All 32 variants are shown in Supplementary Material, Appendix Table S8 together with their allele frequencies, conservation and regulatory potential according to UCSC 5 way regulatory potential track.

Upon performing TESS searches, we found that 29 of these variants either disrupt or create a transcription factor binding site (TFBS). These analyses predicted 63 disrupted TFBS and 61 created TFBS (Supplementary Material, Appendix Table S9). From those we investigate further the top 25% according to a log likelihood/length of TFBS of 2 and log likelihood greater than 12 (Supplementary Material, Appendix Table S10). Four out of the 31 top TFBS signals are strong candidates for involvement in craniofacial development:

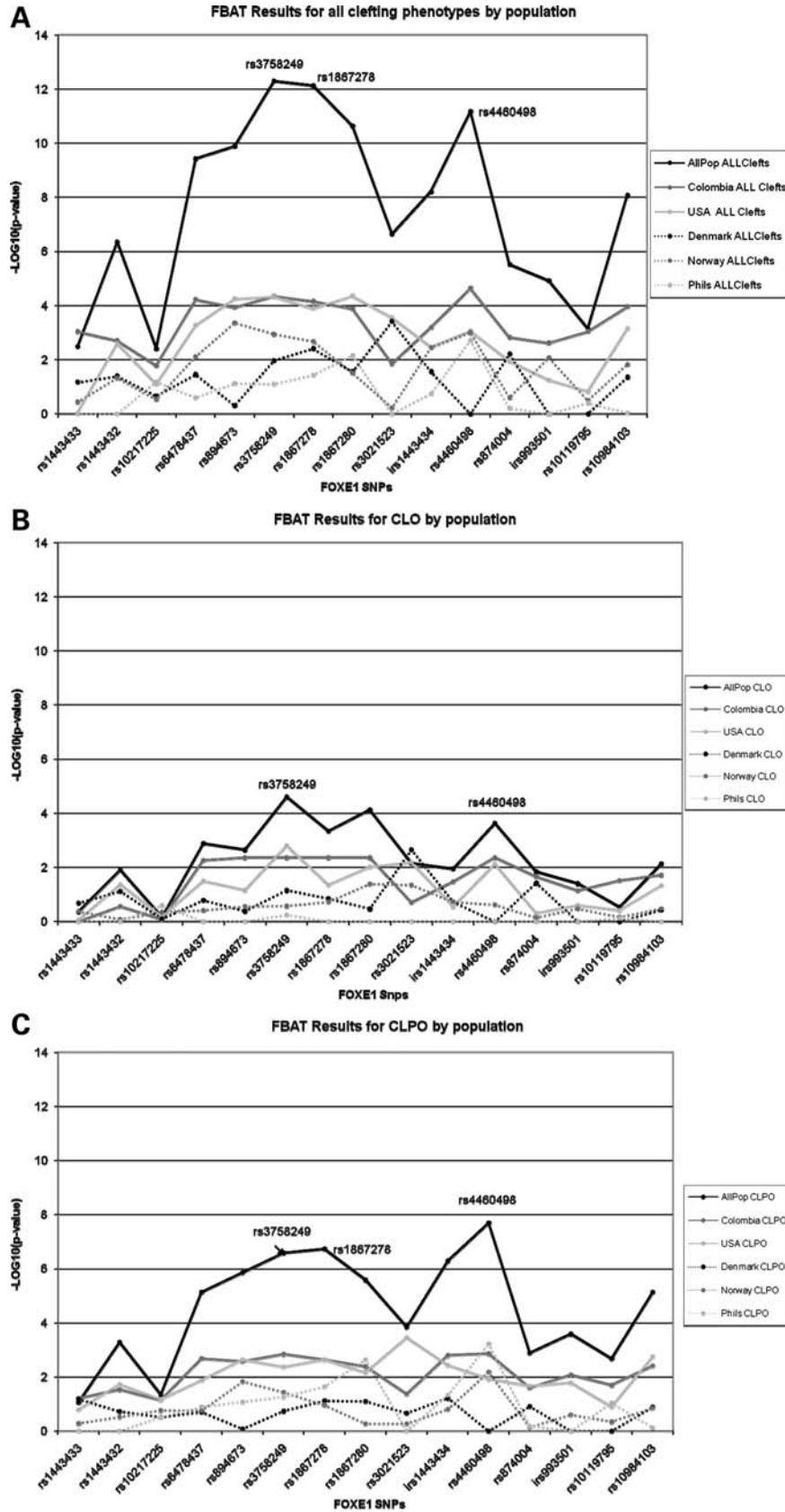


Figure 5. FBAT results. (A) All cleft phenotypes. (B) CLO by population. (C) CLPO by population. (D) CL/P by population and (E) CP by population.

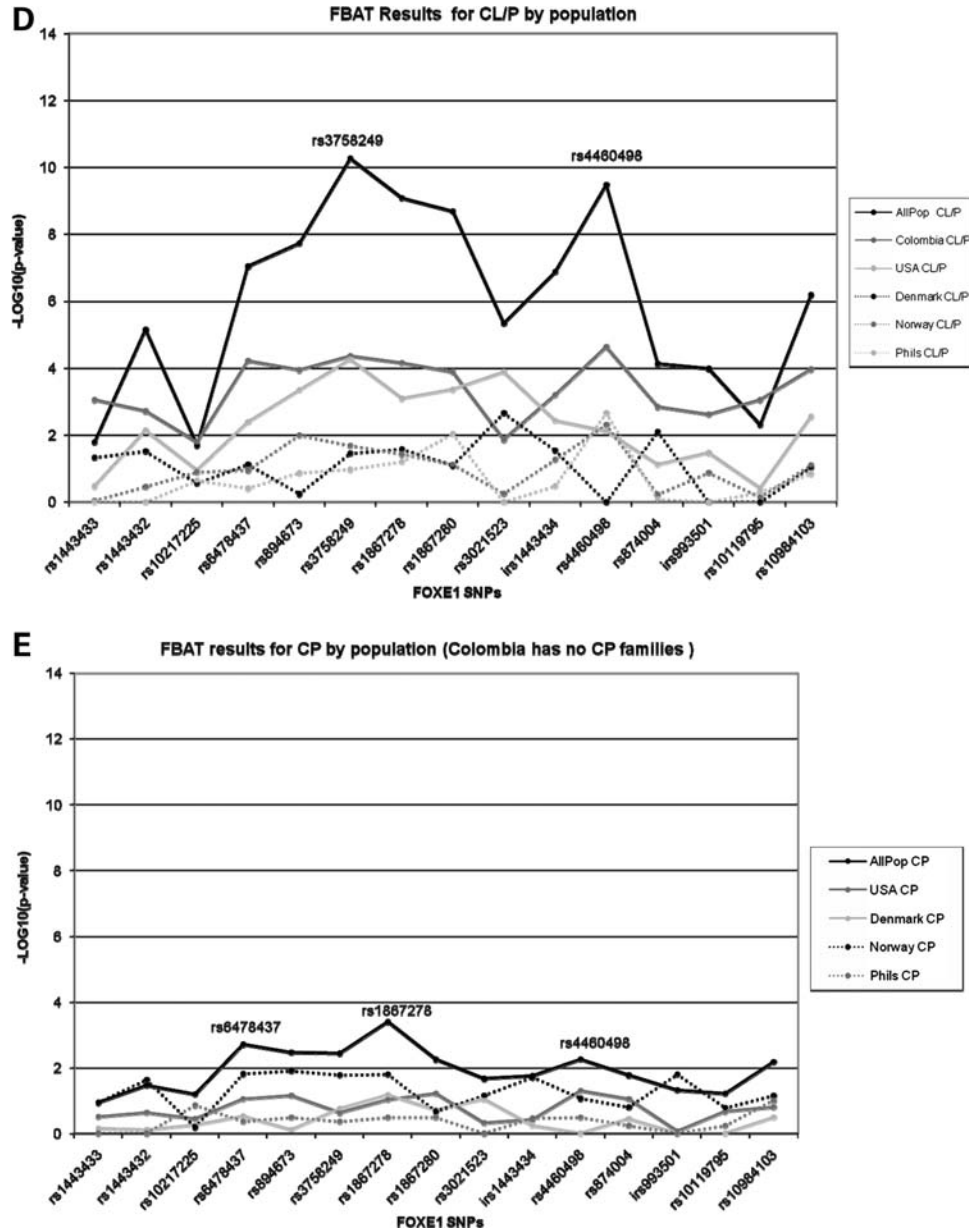


Figure 5. Continued

SP1, *AP3*, *AP1* and *POU1F1* (Supplementary Material, Appendix Tables S9 and S10).

Expression results

Foxe1 expression was observed at E10.5 and E11.5 in the rostral epithelium of the oral pharynx, including the caudal epithelium of the medial nasal and maxillary processes (Fig. 9). In particular, *Foxe1* is expressed in the epithelium involved in the fusion between the medial nasal and maxillary processes. Neither *C9orf156* nor *Hemgn* were observed during primary palatogenesis (data not shown). These data are consistent with *Foxe1* playing a role in lip development.

DISCUSSION

Through multiple stages involving positional cloning, candidate gene sequencing and developmental gene expression analysis, we have identified *FOXE1* as a major disease gene within the previously published 9q22–33 linkage peak. An initial panel of 300 SNPs narrowed the 20 Mb 2-LOD interval to 357 Kb. Fine mapping data that excludes the two genes (*XPA* and *HEMGN*) flanking this interval and expression data, negative for *C9orf156* and positive for *Foxe1*, converge on *FOXE1* as the etiologic gene.

Within a 200 kb critical region, a 20 kb disease haplotype with greater effect in Colombian and Scandinavian samples was identified that extends 19.2 kb upstream of the *FOXE1*

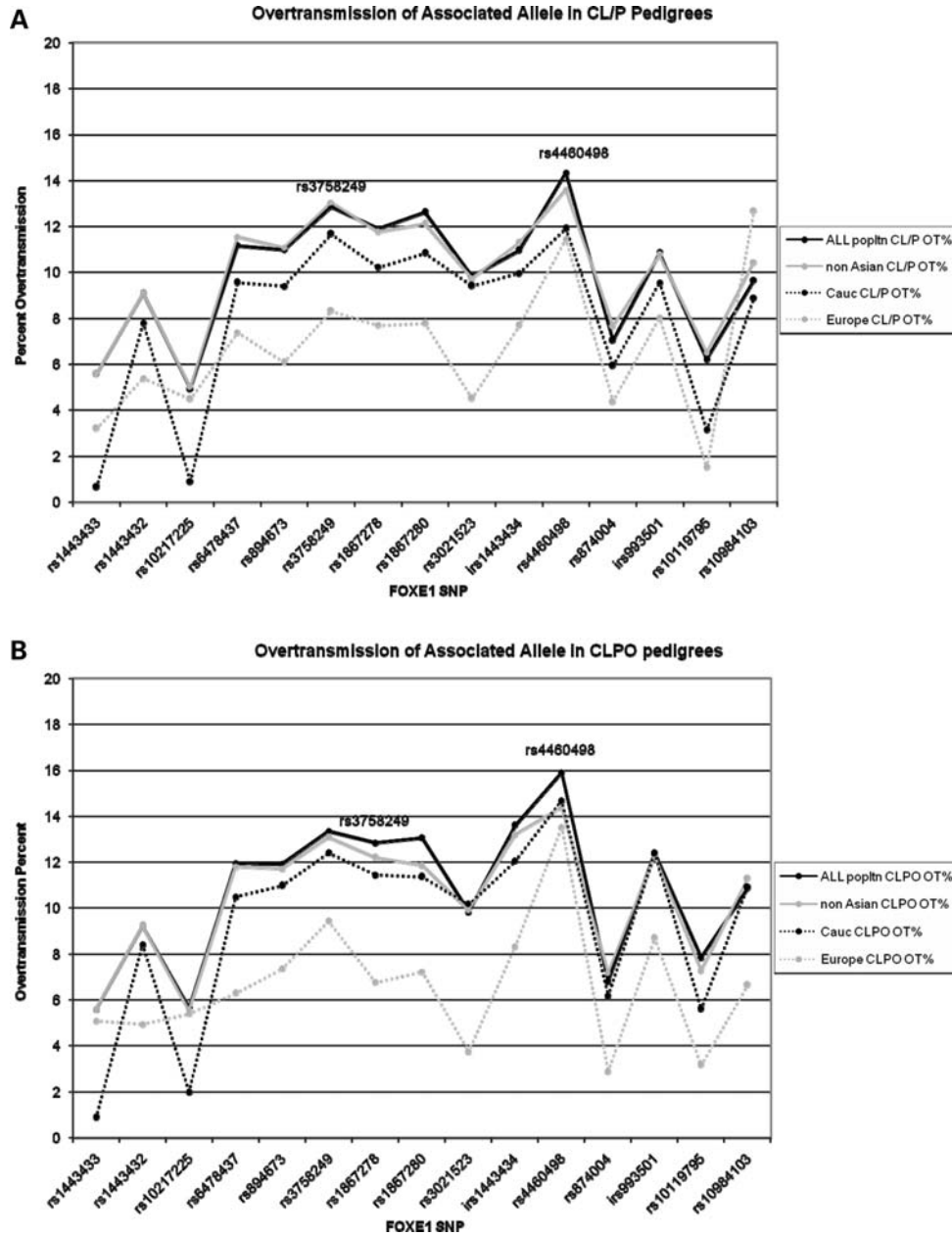


Figure 6. Over-transmission of the associated allele. (A) CL/P and (B) CLPO.

start codon (Fig. 7). Although all populations analyzed in this study contributed to the *FOXE1* association signal, SNP and haplotype frequencies found within each population indicate that this predisposing variant is more common in populations of Caucasian descent. This is in contrast to the *IRF6* predisposing SNP rs642961 G/A (9,32) that is more common in populations of Asian descent. Our data indicate two additional *FOXE1* risk alleles. Independent signals from a haplotype spanning 95.5 kb starting 21.1 kb 3' of *FOXE1* in our Filipino population indicate an Asian-specific allele (Figs 4C and 7). Previous reports of a translocation breakpoint 55 kb 3' of *FOXE1* in two siblings with bilateral cleft lip and palate lend credence to this hypothesis (33). Filipinos have demonstrated

similar evidence for two or more variants contributing to clefting at the *IRF6* locus (9,32). Finally, the presence of a third allele in a separate LD block is supported by significant SNPs and haplotypes containing rs1443432 and rs780258 67 kb upstream of *FOXE1* (Figs 3–6). Finding multiple risk alleles are to be expected for a complex disease such as orofacial clefting.

The association patterns suggest the phenotypic spectrum of *FOXE1* variants involves all combinations of primary and secondary palatal clefts. These findings differ from those of the *IRF6* rs642961 variant disrupting an AP-2 α -binding site that was not associated with CP in Filipinos. Furthermore, in contrast to *IRF6*, *FOXE1* markers could not distinguish CLPO or

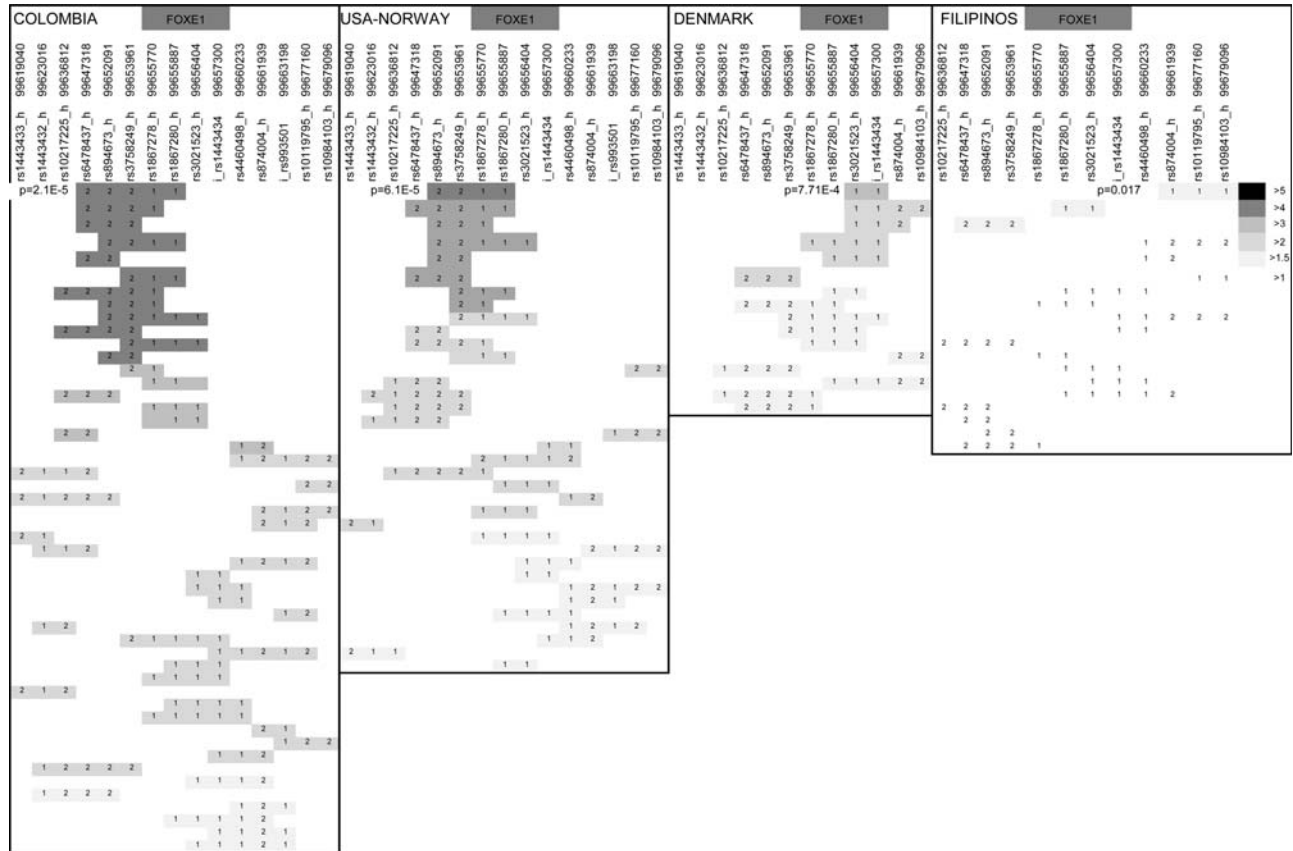


Figure 7. Haplotype window analysis results for replication panel markers and the CL/P phenotype. Rectangles represent the size of the haplotype according to SNPs involved. Colors represent the $-\log_{10}(P\text{-value})$ results. *i*_rs#, markers typed by CIDR; *rs#_h*, markers genotyped in our laboratories. See Supplementary Material, Appendix Table S11 for allele correspondence to specific nucleotides.

CL/P from CLO in our populations. Both isolated cleft lip and isolated cleft palate can be separated from, or lumped together with patients having CLPO based on embryology, recurrence and syndromic gene findings. *FOXE1* appears to present one unifying path for these phenotypes.

Of note is the maternal over-transmission of *FOXE1* alleles in the Colombian and Filipino samples compared with all other Caucasian samples as well as the paternal over-transmission observed in families from the USA and Scandinavia. This may reflect a parent of origin effect such as imprinting, although there is no existing evidence to suggest imprinting for the *FOXE1* locus. Preferential maternal over-transmission could also indicate a maternal genetic effect. There is dissimilarity between the maternal and paternal ancestral lineages of Colombian families with a predominance of Native American mitochondrial haplotypes (87.5%) and Caucasian Y chromosome haplotypes (57.4%), indicating admixture events between Caucasian immigrant men and Native American women (34); a common finding amongst South American populations (35,36). A Brazilian orofacial cleft study not only demonstrated similar differences between ancestral maternal and paternal contributions, but also showed significantly higher frequency of Native American mitochondrial and Y haplotypes in cases compared with controls (37). Furthermore, evidence of genetic heterogeneity due to Native American ancestry was underscored by the

presence of association between CL/P and *IRF6* ($P = 0.023$) and of CL with *RFC1* ($P = 0.017$) in non-carriers of the Native American haplogroup D (38).

Similar parent of origin effects for orofacial clefting have been described for *CBS*, *IRF6*, *MTHFR*, *MTR*, *PAX3*, *PAX7*, *RUNX2*, *TGFA* and *TGFB3* (6,39–45). Most interesting were the results reported by Shi *et al.* (6), indicating a maternal genetic effect for *IRF6* with the fetal risk allele being protective when the mother is a risk allele carrier.

Only two *FOXE1* coding mutations, I59S and P208R, were identified in our screen of over 200 patients. Neither mutation is present in 24 control individuals from this study or in 186 controls in a candidate gene mutation screen (15). P208R is immediately adjacent to the A207V identified in a patient with CL/P (15). I59S is located in the first alpha helix of the forkhead DNA-binding domain within which two mutations, S57N and A65V, have been described in families with features of the Bamforth–Lazarus syndrome (BLS) (46,47). The S57N and A65V mutations have been shown to significantly reduce DNA binding and transcription activation. Furthermore, an I84S mutation at the corresponding *FOXL2* amino acid as *FOXE1* I59 has been reported in a family with blepharophimosis-ptosis-epicanthus inversus syndrome (48). Therefore, the I59S mutation in our study may have similar effects on *FOXE1* function. The lack of a more severe phenotype in our proband may be due to only one mutated *FOXE1*

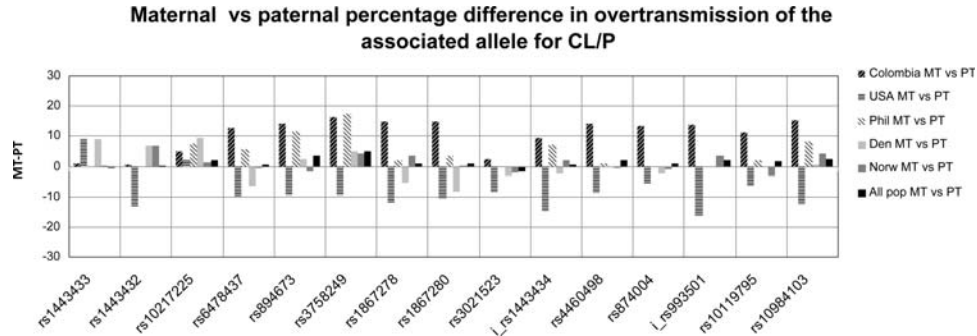


Figure 8. CL/P% maternal (MT) versus paternal (PT). Difference values above 10 or below -10 are considered significant at an α of 0.05. See Supplementary Material, Appendix Table S11 for allele correspondence to specific nucleotides and transmission direction of each allele.

copy, whereas BLS is recessive. The phenotypic spectrum of *FOXE1* coding mutations may include features less severe than BLS, such as hypothyroidism. Of note, *FOXE1* variants confer a 5.7-fold for thyroid cancer (49).

The variants identified by sequencing the noncoding regions of *FOXE1*, *PTCH*, *TGFBR1*, *GABBR2*, *ZNF189*, *FGD3* could presumably have etiologic effects. Although none were associated with CL/P, they occurred at conserved nucleotides and some were predicted to affect TFBSs, including *SP1*, *API*, *AP3* and *POU1F1* that are involved in facial development.

FOXE1 is a member of a transcription factor family that contains a DNA-binding forkhead domain and that are involved in embryonic pattern formation. Newborn mice null for *Foxe1* exhibit cleft palate and thyroid anomalies (50). *Foxe1* is expressed in the secondary palate epithelium in both mice at E13.5–E15.5 (51) and humans at week 11 of gestation (52). The specific expression pattern of *Foxe1* at the point of fusion between maxillary and nasal processes, described for the first time in this study, strongly suggests *FOXE1* as a key player in primary palatogenesis.

Foxe1 is a downstream target of the *Shh/Gli* pathway in hair follicle morphogenesis (53). Thus, it is plausible that there are two disease genes in the 9q22–q33 region since the *SHH* receptor *PTCH* was associated with CL/P in our Filipino data set and there is ample evidence that *SHH* signaling is involved during primary palatogenesis (54). Furthermore, there is overlapping *Foxe1* and *Shh* expression in the caudal epithelium of the mesial nasal and maxillary processes.

The paucity of *FOXE1* mutations despite exhaustive searches (15) and in light of the highly significant association described here indicate that causal mutations are in nearby noncoding regions that regulate *FOXE1* expression, most likely within the 70 kb high LD block containing the 5'-UTR and promoter. Recently, a variant affecting MYF-5 DNA binding in a cluster of *GLI*-binding sites 1.2 kb upstream of the transcription start site has been described in 11 of 25 patients with NSCLP (55). We found this variant in 9 of 184 case and 2 of 186 control chromosomes. While this is coincident with etiologic effects, this variant is too infrequent to explain the strong association results unless it splits a common haplotype into two disease alleles.

Efforts to find other causal variants will require in depth sequencing and characterization of regulatory function in highly conserved regions within our putative disease haplotypes.

Utilizing the Vertebrate Multiz Alignment (28 species) and the PhastCons track of the UCSC browser (56), we performed a preliminary screening of the 200 kb critical region with a set threshold score above 350. Fourteen regions with high PhastCons scores (Table 2) were identified. One region that contains the most significantly associated SNP also harbors five *GLI2*-binding sites (57). The first of these *GLI2*-binding sites is located 20 bp downstream of rs3758249, suggesting further a connection of *FOXE1* to the *SHH/GLI* pathway for primary palatogenesis.

In the aggregate, our findings place *FOXE1* as the third locus along with *IRF6* and 8q24 in which common variants have a substantial impact on the occurrence of cleft lip and palate. It affords a new pathway to investigate biologically and new tools to improve genetic counseling.

MATERIALS AND METHODS

Subjects

Inclusion criteria for participating subjects were a diagnosis of nonsyndromic clefting (58). Subjects were clinically examined, medical and family histories reviewed and gestational environmental exposures were analyzed to rule out syndromes and known teratogens. The Colombian families comprise a collection of extended families and nuclear triads from the province of Antioquia in Northwest Colombia and were recruited through the Clinica Noel in Medellin (273 affected pedigrees plus 90 control triads). The USA Caucasian families were recruited as part of a Multicenter Collaborative Project and include families from the states of California, Washington, Ohio and Iowa (464 pedigrees). Filipino pedigrees (577 total) were ascertained through the Operation Smile Organization (59). Independent replication panels comprised of families from Denmark, Norway and Iowa. Samples from Denmark were 592 case triads consisting of affected individuals (born 1981–1990), their parents and some unaffected sibling(s). A Danish cohort of unaffected control individuals was also included, yielding 499 genotyped children (born 1997–2003), 444 typed mothers, and 313 adults. Samples from Norway (60) were triads recruited as part of a case-cohort study of facial clefts (born 1996–2001), including 452 case triads and 762 control triads. Finally, the samples from Iowa consisted of a collection of 249 case triads and 207 unrelated controls

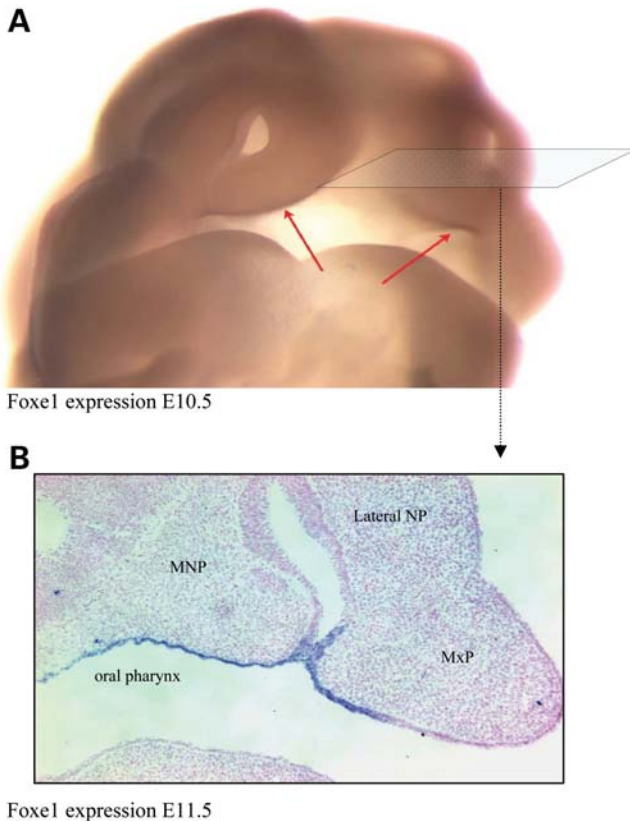


Figure 9. *Foxe1* expression at 45 somites. (A) Whole mount *in situ*. (B) Section *in situ*. Expression in caudal epithelium of medial nasal and maxillary processes.

(genotyped parents of unaffected individuals) (61). For analysis purposes, the Iowa case sample was pooled with all other USA samples (Table 3).

Cleft phenotypic subgroups were determined from the cleft history of first-, second- and third-degree relatives of the proband. Analysis of markers in *FOXE1* was stratified by cleft phenotype in subsets of pedigrees of the following cleft phenotypes: cleft lip only families (CLO) where all affected family members present a cleft lip only. Families with cleft lip and palate only (CLPO) are those in which all affected family members have a CLP phenotype. The CL/P phenotypic group includes all pedigrees in which individuals had either a CLO or CLPO. Finally, CP pedigrees are those where there is at least one individual affected with a cleft palate only among other possible cleft phenotypes. An overall group of families (ALL group) represents the combination of CL/P and CP groups, and also includes those few pedigrees with a nonsyndromic cleft of unknown type (Table 1).

All samples utilized in our study were obtained after individuals and/or their parents gave signed informed consent under a protocol approval by institutional review boards at the Scientific Committee at the Dental School of University of Antioquia-Colombia, the Ohio State University, the University of Iowa, Children's Hospital and Regional Medical Center in Seattle, the Hope Foundation (Bacolod City, Negros-Philippines), the University of Pittsburgh, and the ethics committees in Denmark and Norway.

Genotyping and association analysis

Parallel fine mapping and candidate gene approaches were employed iteratively using association and sequencing methods to identify the locus responsible for the 9q22–q33 linkage signal.

In the first approach (Stage 1), the region between 102 and 140 cM including the 2-LOD-unit support interval was evaluated with 331 SNPs genotyped in 571 pedigrees with CL/P from Colombia, USA_1 and the Philippines (Table 3). The genotyping was performed by the Center for Inherited Disease Research (CIDR) using a custom Illumina SNP panel. The goal of the initial fine mapping was to saturate all known genes in the 1-LOD-unit support interval (Fig. 1) with 2–6 SNPs per gene while filling intergenic gaps of >100 kb with 1 SNP/20 kb. The -1 to 2-LOD-unit support interval was covered in a similar fashion; however, intergenic regions had approximately 1 SNP/50 kb. To cover the second Colombian peak (120–140 cM) (9), clusters of 2–4 SNPs every 200–400 kb were seeded across this interval. This saturation provided adequate SNP resolution in the Paisa population from Colombia (62). SNPs were chosen using the information from the Hapmap Project (63), <http://www.hapmap.org>, based on the CEPH (Utah residents with ancestry from northern and western Europe) (CEU), Japanese in Tokyo, Japan (JPT) and Han Chinese in Beijing, China (CHB) populations. Tagging SNPs were evaluated in regards to minor allele frequency, deviations from Hardy–Weinberg equilibrium (HWE), inter-marker distance, as well as LD patterns and haplotype block structures using HAPLOVIEW version 2.05 (56). Tagging SNPs were selected to include at least 85% of the haplotypes in the LD blocks and MAF was set to 20% minimum (for complete list of SNPs, see Supplementary Material, Appendix Table S1).

In parallel, candidate genes within the 2-LOD-unit support interval (Stage 2) (*ROR2*, *FDG3*, *BARX1*, *PTCH*, *FOXE1*, *GPR51*, *COL15A1*, *TGFBR1*, *ZNF189* and *ALDOB*) were tested for association by genotyping 44 additional SNPs genotyped with TaqMan assays from Applied Biosystems (Foster City, CA, USA) in the same panel of both Colombian and USA pedigrees used in Stage 1 plus 26 additional pedigrees from USA and 307 Filipino triads. Genotypes obtained from 134 Illumina SNPs (from Stage 1) located within these selected candidate genes were included in the association analysis (for complete list of SNPs, see Supplementary Material, Appendix Table S2).

In Stage 3, a critical region of 357 kb (99.44–99.8 Mbp) containing the genes *XPA*, *FOXE1*, *C9ORF156* and *HEMGN* was fine mapped by another round of genotyping and association analysis using a total of 34 SNPs (22 new Taqman assays, 5 Taqman assays genotyped in stage 2 and 7 Illumina SNPs from stage 1) (Supplementary Material, Appendix Table S3).

Single SNP association was tested using family-based association methods (FBAT) (64) utilizing the *-e* option to account for multiple sibs in a family or multiple nuclear families in a pedigree. HBAT haplotype analysis was used for genotypic data on Stage 3 (critical region, *XPA-FOXE1-C9ORF156-HEMGN*) using a sliding window approach with a maximum of five SNPs per window.

Table 2. Regions harboring potentially regulatory elements

Priority	Name relative to ATG	Position start (UCSC March 2006 Assembly)	Position end (UCSC March 2006 Assembly)	Size (bp)	PhastCons placental mammal scores	Distance relative to ATG (bp)
1	CRS +20.7	99 676 550	99 677 150	600	615	20 550
2	CRS -67	99 588 450	99 588 900	450	604	-67 550
3	CRS +9.4	99 665 300	99 665 750	450	514	9300
4	CRS -50	99 605 600	99 605 950	350	507	-50 400
5	CRS -51.5	99 604 125	99 604 900	775	440	-51 875
6	CRS +23.0	99 678 800	99 679 700	900	398 390 396	22 800
7	CRS +4	99 660 000	99 660 400	400	525	4000
8	CRS -82	99 573 800	99 574 700	900	425	-82 200
9	CRS -44.2	99 611 750	99 612 050	300	492	-44 250
10	CRS -22.4	99 633 350	99 634 050	700	474	-22 650
11	CRS -9.0	99 647 200	99 647 450	250	460	-8800
12	CRS -72.0	99 583 600	99 584 000	400	379 357 360	-72 400
13	CRS -2.2	99 653 800	99 654 450	650	4 <i>Gli</i> binding sites	-2200
14	CRS -0.45	99 654 950	99 655 350	400	419	-1050

Selected regions for searches of potential regulatory elements based on PhastCons placental mammal scores above 350, 7X regulatory potential, transcription binding sites. Region 13 shows no PhastCons placental mammal scores but contains four *Gli* binding sites and association results. CRS, conserved regulatory sequence.

Table 3. Number of pedigrees and markers included in the study

A

	Stage 1	Stage 2			Stage 3		Stage 4
	331 SNPs	178 SNPs			34 SNPs		15 SNPs
	2LOD INTERVAL	ROR2-ALDOB			XPA-HEMGN		FOX E1
	331 Illumina	134 Illumina	44 Taqman	7 Illumina	27 Taqman	15 Taqman	
Case Populations	Pedigrees	Pedigrees	Pedigrees	Pedigrees	Pedigrees	Pedigrees	Pedigrees
Colombia	232	232	232	232	273	273	
USA_1 California, Washington and Ohio	130	130	156	130	184		
USA_2 California, Washington, Ohio and Iowa							464
Philippines multiplex families_1	209						0
Philippines multiplex families and trios_2			307		307		368
Denmark							592
Norway							452
TOTAL	571	362	695	362	764	2149	

B

Control Populations	Control Diad/Triads	Control Individuals included in analysis	15 SNPs FOX E1
Colombia	90	170 genotyped parents	15
Iowa-USA	109	207 genotyped parents	15
Denmark	840	499 genotyped children	12
Norway	762	761 genotyped children	15
Total		1637	

USA_1 and USA_2 are collectively referring to in the text as the USA Caucasian families. (A) Cases (B) Controls.

In Stage 4, association with alleles of *FOXE1* was replicated by further genotyping a cluster of 15 SNPs surrounding *FOXE1* in an independent CL/P sample panel consisting of cohorts from Denmark ($n = 395$ peds with CL/P) and Norway ($n = 314$ peds with CL/P) as well as 139 additional CL/P affected triads and 207 unrelated controls from USA (Tables 1 and 3). To further evaluate the role of *FOXE1* in different cleft phenotypes, the same 15 SNPs mentioned above were genotyped in additional CP pedigrees (Table 1, Supplementary Material, Appendix Table S6). All *FOXE1* genotypic data were tested for association by performing SNP and haplotype TDT analyses as implemented in the FBAT software and subsetting by cleft phenotype (Table 1, Supplementary Material, Appendix Table S6) (64).

Correction for multiple testing was done by adjusting the alpha level of significance by means of Bonferroni corrections. *P*-values obtained in Stage 1 (2-LOD-unit support interval) analysis were not adjusted in an effort not to miss variants of lower effect. However *P*-values obtained in Stages 2 (178 SNPs) and 3 (34 SNPs) were adjusted using Bonferroni corrections with alpha levels of significance calculated as $0.05/178 = 2.81E - 04$ and $0.05/34 = 1.47E - 03$, respectively. In addition, *P*-values obtained for *FOXE1* analyses by cleft phenotype were adjusted using a Bonferroni correction with an alpha level for significance calculated as $0.05/300$ (15 SNPs \times 5 populations \times 4 cleft phenotypes, CLO, CLPO, CL/P, CP) = $1.67E - 04$.

Parental over-transmission

To investigate parent of origin effects within *FOXE1*, we performed overall parental TDT and parent-specific TDT calculations as implemented in S.A.G.E., which provides the exact McNemar test statistics for the familial data within the CLPO and CL/P phenotypic groups. Furthermore, the overall parental TDT transmission counts were utilized to calculate an over-transmission percentage rate. Assuming a null hypothesis of 50% transmission rate for any allele from heterozygous parents, the over-transmission rate was the ratio of the count of over-transmitted allele to the total count of transmitted alleles minus 50% (0.50 in decimal format). Ninety-five percent confidence intervals were calculated (assuming the over-transmission rate follows a binomial distribution) to determine if these over-transmission rates significantly differed among populations or cleft sub-groups. The over-transmission rates represent an alternative approach to the PAR% calculations to assess the overall impact of the associated allele.

Associated allele dosage effects

To test for dosage effects of the associated allele, the Weinberg LRT method was used to calculate the relative risk (RR) for heterozygotes and homozygotes for the over-transmitted allele for the most significantly associated SNP within cleft groups driving the association (65). Log-linear models with genotype and imprinting effects were fit to the data using the SAS version of the LRT program (65). Ninety-five percent confidence intervals were calculated to determine if the RR for carrying one copy of the associated allele differed significantly from the RR of two copies of the

associated allele. Using the parameter estimate (β) and β 's 95% CI as the arguments to the exponential function, the RR and its 95% CI are calculated: $RR (95\% CI) = e^{\beta_{estimate}}$ ($e^{\beta_{estimate}^{lower}}$, $e^{\beta_{estimate}^{upper}}$).

Sequencing

Previously, re-sequencing of coding regions in 9q22–33 candidate genes, *PTCH*, *FOXE1* and *TGFBR1*, revealed only six private missense mutations (3, 2 and 1, respectively) in 180 CL/P patients (15,16). Therefore, in parallel with the association analysis above and in an effort to identify potential regulatory variants that could explain the linkage signal, 41 conserved non-coding regions between these and other clefting candidate genes (*GABABR2*, *ZNF189*, *FGD3*) were re-sequenced in 63 Filipino, 28 Colombian and 1 USA affected probands from families yielding positive 9q22–33 LOD scores. Selection criteria for noncoding regions to sequence were: at least 70% conservation in nucleotide identity over a minimum of 100 bp in five species (dog, mouse, rat, chicken and frog). A total of 34 kb of sequence was covered with 63 primer pairs (Supplementary Material, Appendix Table S5). Variants found were evaluated using the UCSC browser for 5X regulatory potential and conservation (<http://genome.ucsc.edu/>). In addition, to investigate the potential of creating or deleting TFBS, both the wild-type and variant allele were queried in the Transcription Element Search Software (TESS) (66) (<http://www.cbil.upenn.edu/cgi-bin/tess/tess>).

Upon narrowing the disease locus, *FOXE1* 5'-UTR, coding and 3'-UTR regions were also sequenced in the affected probands from families yielding positive 9q22–33 LOD scores described above. All sequencing reactions were performed as described previously (16).

Expression analysis

Expression of *Foxe1*, *Hemgn* and *C9orf156* during primary palatogenesis was characterized by *in situ* hybridization. C57BL/6J mouse embryos at E9.5, E10.5 and E11.5 were evaluated by both whole mount and section *in situ* with gene-specific digoxigenin-labeled RNA probes (67). *Foxe1* (NM183298) probe is a 0.8 kb *Bam*HI/*Not*I cDNA clone in pBluescript (generously provided by Roberto DiLauro). *Hemgn* (NM.053149) probe is a 529 bp cDNA clone in pCRII-TOPO. *C9orf156* homolog clone 5830415F09 (BC092543) is a 560 bp cDNA clone in pCRII-TOPO. Coordinates according to the UCSC Mm July 2007 and primer sequences for both *Hemgn* and *C9orf156* probes are as follows: mmHemgn-ex34F, AGA-TGT-CGC-TGA-AGG-CTG-TC 46409176–46409195; mmHemgn-ex34R, TTC-CGG-ATC-TTG-ATC-TGC-TT 46407541–46407560; mmC9orf156-ex34F, CAA-GCT-GGA-GAA-GGT-GGA-AG 46398966–46398985; mmC9orf156-ex34R, GGA-AGC-ATC-CCA-CTG-TGT-TT 46395064–46395083.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are especially grateful to the families that participated in this study and also the staff at each recruiting site. Special thanks to the Universidad de Antioquia Paternity testing lab for providing access to anonymous Colombian control samples. We would like to thank Xiaoyan Qin, Peng Shi, Sandy Imoehl, Scott Pate and Jason Dierdorff for their help with genotyping experiments and Kristin Oleson for sample preparation. Thanks to Susie Mc Connell, Nancy Davin and Daniel Benton for their administrative support.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the National Institutes of Health [RO1-DE014667-8, KO2-AEO15291, a March of Dimes Basil O'Connor award # FY 98-0718 and Research Grant #6-FY01-616, a Cleft Palate Foundation Grant and an American Association of Orthodontics Foundation Faculty Development Award to A.C.L., DE-08559 to J.C.M., DE-11948 to K.C., P50-DE-16215 to M.L.M. and J.C.M., R01-DE016148 to M.L.M., R21-DE016930 to M.L.M., R01-DE09886 to M.L.M., R01-DE012472 to M.L.M. and Fogarty 1RO3-TW-007644 to J.C.M.]; and also by the Intramural Research Program of the National Institute of Environmental Health Sciences, National Institutes of Health. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, Contract Number N01-HG-65403. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Dental and Craniofacial Research, or the National Institutes of Health.

REFERENCES

- Schliekelman, P. and Slatkin, M. (2002) Multiplex relative risk and estimation of the number of loci underlying an inherited disease. *Am. J. Hum. Genet.*, **71**, 1369–1385.
- Kondo, S., Schutte, B.C., Richardson, R.J., Bjork, B.C., Knight, A.S., Watanabe, Y., Howard, E., de Lima, R.L., Daack-Hirsch, S., Sander, A. *et al.* (2002) Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nat. Genet.*, **32**, 285–289.
- Field, L.L., Ray, A.K., Cooper, M.E., Goldstein, T., Shaw, D.F. and Marazita, M.L. (2004) 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*, **130A**, 265–271.
- Prescott, N.J., Lees, M.M., Winter, R.M. and Malcolm, S. (2000) 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.*, **106**, 345–350.
- Marazita, M.L., Field, L.L., Cooper, M.E., Tobias, R., Maher, B.S., Peanchitlertkajorn, S. and Liu, Y.E. (2002) Genome scan for loci involved in cleft lip with or without cleft palate, in Chinese multiplex families. *Am. J. Hum. Genet.*, **71**, 349–364.
- Shi, M., Christensen, K., Weinberg, C.R., Romitti, P., Bathum, L., Lozada, A., Morris, R.W., Lovett, M. and Murray, J.C. (2007) Orofacial cleft risk is increased with maternal smoking and specific detoxification-gene variants. *Am. J. Hum. Genet.*, **80**, 76–90.
- Zeiger, J.S., Hetmanski, J.B., Beaty, T.H., VanderKolk, C.A., Wyszynski, D.F., Bailey-Wilson, J.E., de Luna, R.O., Perandones, C., Tolarova, M.M., Mosby, T. *et al.* (2003) Evidence for linkage of nonsyndromic cleft lip with or without cleft palate to a region on chromosome 2. *Eur. J. Hum. Genet.*, **11**, 835–839.
- Wyszynski, D.F., Albacha-Hejazi, H., Aldirani, M., Hammoud, M., Shkair, H., Karam, A., Alashkar, J., Holmes, T.N., Pugh, E.W., Doheny, K.F. *et al.* (2003) 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. A*, **123A**, 140–147.
- Zucchero, T.M., Cooper, M.E., Maher, B.S., Daack-Hirsch, S., Nepomuceno, B., Ribeiro, L., Caprau, D., Christensen, K., Suzuki, Y., Machida, J. *et al.* (2004) Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. *N. Engl. J. Med.*, **351**, 769–780.
- Marazita, M.L. and Mooney, M.P. (2004) Current concepts in the embryology and genetics of cleft lip and cleft palate. *Clin. Plast. Surg.*, **31**, 125–140.
- Marazita, M.L., Murray, J.C., Lidral, A.C., Arcos-Burgos, M., Cooper, M.E., Goldstein, T., Maher, B.S., Daack-Hirsch, S., Schultz, R., Mansilla, M.A. *et al.* (2004) Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32–35. *Am. J. Hum. Genet.*, **75**, 161–173.
- Choi, S.J., Marazita, M.L., Hart, P.S., Sulima, P.P., Field, L.L., McHenry, T.G., Govil, M., Cooper, M.E., Letra, A., Menezes, R. *et al.* (2008) The PDGF-C regulatory region SNP rs28999109 decreases promoter transcriptional activity and is associated with CL/P. *Eur. J. Hum. Genet.*, **17**, 774–784.
- Riley, B.M., Mansilla, M.A., Ma, J., Daack-Hirsch, S., Maher, B.S., Raffensperger, L.M., Russo, E.T., Vieira, A.R., Dode, C., Mohammadi, M. *et al.* (2007) Impaired FGF signaling contributes to cleft lip and palate. *Proc. Natl Acad. Sci. USA*, **104**, 4512–4517.
- Jezewski, P.A., Vieira, A.R., Nishimura, C., Ludwig, B., Johnson, M., O'Brien, S.E., Daack-Hirsch, S., Schultz, R.E., Weber, A., Nepomucena, B. *et al.* (2003) Complete sequencing shows a role for MSX1 in non-syndromic cleft lip and palate. *J. Med. Genet.*, **40**, 399–407.
- Vieira, A.R., Avila, J.R., Daack-Hirsch, S., Dragan, E., Felix, T.M., Rahimov, F., Harrington, J., Schultz, R.R., Watanabe, Y., Johnson, M. *et al.* (2005) Medical sequencing of candidate genes for nonsyndromic cleft lip and palate. *PLOS Genet.*, **1**, e64.
- Mansilla, M.A., Cooper, M.E., Goldstein, T., Castilla, E.E., Lopez Camelo, J.S., Marazita, M.L. and Murray, J.C. (2006) Contributions of PTCH gene variants to isolated cleft lip and palate. *Cleft Palate Craniofac. J.*, **43**, 21–29.
- Avila, J.R., Jezewski, P.A., Vieira, A.R., Orioli, I.M., Castilla, E.E., Christensen, K., Daack-Hirsch, S., Romitti, P.A. and Murray, J.C. (2006) PVRL1 variants contribute to non-syndromic cleft lip and palate in multiple populations. *Am. J. Med. Genet. A*, **140**, 2562–2570.
- Chiquet, B.T., Lidral, A.C., Stal, S., Mulliken, J.B., Moreno, L.M., Arcos-Burgos, M., Valencia-Ramirez, C., Blanton, S.H. and Hecht, J.T. (2007) CRISPLD2: a novel NSCLP candidate gene. *Hum. Mol. Genet.*, **16**, 2241–2248.
- Lidral, A.C., Murray, J.C., Buetow, K.H., Basart, A.M., Scheerer, H., Shiang, R., Naval, A., Layda, E., Magee, K. and Magee, W. (1997) 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.*, **34**, 1–6.
- Lidral, A.C., Romitti, P.A., Basart, A.M., Doetschman, T., Leysens, N.J., Daack-Hirsch, S., Semina, E.V., Johnson, L.R., Machida, J., Burds, A. *et al.* (1998) Association of MSX1 and TGFB3 with nonsyndromic clefting in humans. *Am. J. Hum. Genet.*, **63**, 557–568.
- Watanabe, A., Akita, S., Tin, N.T., Natsume, N., Nakano, Y., Niikawa, N., Uchiyama, T. and Yoshiura, K. (2006) A mutation in RYK is a genetic factor for nonsyndromic cleft lip and palate. *Cleft Palate Craniofac. J.*, **43**, 310–316.
- Suzuki, Y., Jezewski, P.A., Machida, J., Watanabe, Y., Shi, M., Cooper, M.E., Viet le, T., Nguyen, T.D., Hai, H., Natsume, N. *et al.* (2004) In a Vietnamese population, MSX1 variants contribute to cleft lip and palate. *Genet. Med.*, **6**, 117–125.
- Suzuki, S., Marazita, M.L., Cooper, M.E., Miwa, N., Hing, A., Jugessur, A., Natsume, N., Shimozato, K., Ohbayashi, N., Suzuki, Y. *et al.* (2009) Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. *Am. J. Hum. Genet.*, **84**, 406–411.
- Vieira, A.R. (2008) Unraveling human cleft lip and palate research. *J. Dent. Res.*, **87**, 119–125.

25. Birnbaum, S., Ludwig, K.U., Reutter, H., Herms, S., Steffens, M., Rubini, M., Baluardo, C., Ferrian, M., Almeida de Assis, N., Alblas, M.A. *et al.* (2009) Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat. Genet.*, **41**, 473–477.
26. Grant, S.F., Wang, K., Zhang, H., Glaberson, W., Annaiah, K., Kim, C.E., Bradfield, J.P., Glessner, J.T., Thomas, K.A., Garris, M. *et al.* (2009) A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J. Pediatr.*, in press.
27. Little, J., Cardy, A. and Munger, R.G. (2004) Tobacco smoking and oral clefts: a meta-analysis. *Bull. World Health Organ.*, **82**, 213–218.
28. Bille, C., Olsen, J., Vach, W., Knudsen, V.K., Olsen, S.F., Rasmussen, K., Murray, J.C., Andersen, A.M. and Christensen, K. (2007) Oral clefts and life style factors—a case-cohort study based on prospective Danish data. *Eur. J. Epidemiol.*, **22**, 173–181.
29. Lammer, E.J., Shaw, G.M., Iovannisci, D.M., Van Waes, J. and Finnell, R.H. (2004) Maternal smoking and the risk of orofacial clefts: Susceptibility with NAT1 and NAT2 polymorphisms. *Epidemiology*, **15**, 150–156.
30. van Rooij, I.A., Wegerif, M.J., Roelofs, H.M., Peters, W.H., Kuijpers-Jagtman, A.M., Zielhuis, G.A., Merkus, H.M. and Steegers-Theunissen, R.P. (2001) Smoking, genetic polymorphisms in biotransformation enzymes, and nonsyndromic oral clefting: a gene–environment interaction. *Epidemiology*, **12**, 502–507.
31. Lammer, E.J., Shaw, G.M., Iovannisci, D.M. and Finnell, R.H. (2005) Maternal smoking, genetic variation of glutathione S-transferases, and risk for orofacial clefts. *Epidemiology*, **16**, 698–701.
32. Rahimov, F., Marazita, M.L., Visel, A., Cooper, M.E., Hitchler, M.J., Rubini, M., Domann, F.E., Govil, M., Christensen, K., Bille, C. *et al.* (2008) Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nat. Genet.*, **40**, 1341–1347.
33. Davies, S.J., Wise, C., Venkatesh, B., Mirza, G., Jefferson, A., Volpi, E.V. and Ragoussis, J. (2004) Mapping of three translocation breakpoints associated with orofacial clefting within 6p24 and identification of new transcripts within the region. *Cytogenet. Genome Res.*, **105**, 47–53.
34. Moreno Lina, M. (2005) Mapping genetic loci involved in non syndromic cleft lip with or without cleft palate. Doctoral Thesis Oral Science. University of Iowa, Iowa City, IA.
35. Carvajal-Carmona, L.G., Soto, I.D., Pineda, N., Ortiz-Barrientos, D., Duque, C., Ospina-Duque, J., McCarthy, M., Montoya, P., Alvarez, V.M., Bedoya, G. *et al.* (2000) Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am. J. Hum. Genet.*, **67**, 1287–1295.
36. Mesa, N.R., Mondragon, M.C., Soto, I.D., Parra, M.V., Duque, C., Ortiz-Barrientos, D., Garcia, L.F., Velez, I.D., Bravo, M.L., Munera, J.G. *et al.* (2000) Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. *Am. J. Hum. Genet.*, **67**, 1277–1286.
37. Vieira, A.R., Karras, J.C., Orioli, I.M., Castilla, E.E. and Murray, J.C. (2002) Genetic origins in a South American clefting population. *Clin. Genet.*, **62**, 458–463.
38. Vieira, A.R., Orioli, I.M., Castilla, E.E., Cooper, M.E., Marazita, M.L. and Murray, J.C. (2003) Variants in SKI, IRF6 and RFC1 are associated with cleft lip/palate in a South American population. *ASHG Annual Meeting*, Los Angeles, USA.
39. Rubini, M., Brusati, R., Garattini, G., Magnani, C., Liviero, F., Bianchi, F., Tarantino, E., Massei, A., Pollastri, S., Carturan, S. *et al.* (2005) Cystathionine beta-synthase c.844ins68 gene variant and non-syndromic cleft lip and palate. *Am. J. Med. Genet. A*, **136A**, 368–372.
40. Mostowska, A., Hozyasz, K.K. and Jagodzinski, P.P. (2006) Maternal MTR genotype contributes to the risk of non-syndromic cleft lip and palate in the Polish population. *Clin. Genet.*, **69**, 512–517.
41. Reutter, H., Birnbaum, S., Mende, M., Lauster, C., Schmidt, G., Henschke, H., Saffar, M., Martini, M., Lauster, R., Schiefke, F. *et al.* (2008) TGFB3 displays parent-of-origin effects among central Europeans with nonsyndromic cleft lip and palate. *J. Hum. Genet.*, **53**, 656–661.
42. Sull, J.W., Liang, K.Y., Hetmanski, J.B., Fallin, M.D., Ingersoll, R.G., Park, J., Wu-Chou, Y.H., Chen, P.K., Chong, S.S., Cheah, F. *et al.* (2009) Maternal transmission effects of the PAX genes among cleft case-parent trios from four populations. *Eur. J. Hum. Genet.*, **17**, 831–839.
43. Sull, J.W., Liang, K.Y., Hetmanski, J.B., Fallin, M.D., Ingersoll, R.G., Park, J., Wu-Chou, Y.H., Chen, P.K., Chong, S.S., Cheah, F. *et al.* (2008) Differential parental transmission of markers in RUNX2 among cleft case-parent trios from four populations. *Genet. Epidemiol.*, **32**, 505–512.
44. Sull, J.W., Liang, K.Y., Hetmanski, J.B., Wu, T., Fallin, M.D., Ingersoll, R.G., Park, J.W., Wu-Chou, Y.H., Chen, P.K., Chong, S.S. *et al.* (2009) Evidence that TGFA influences risk to cleft lip with/without cleft palate through unconventional genetic mechanisms. *Hum. Genet.*, **126**, 385–394.
45. Jugessur, A., Wilcox, A.J., Lie, R.T., Murray, J.C., Taylor, J.A., Ulvik, A., Drevon, C.A., Vindenes, H.A. and Abyholm, F.E. (2003) 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.*, **157**, 1083–1091.
46. Clifton-Bligh, R.J., Wentworth, J.M., Heinz, P., Crisp, M.S., John, R., Lazarus, J.H., Ludgate, M. and Chatterjee, V.K. (1998) Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. *Nat. Genet.*, **19**, 399–401.
47. Castanet, M., Park, S.M., Smith, A., Bost, M., Leger, J., Lyonnet, S., Pelet, A., Czernichow, P., Chatterjee, K. and Polak, M. (2002) A novel loss-of-function mutation in TTF-2 is associated with congenital hypothyroidism, thyroid agenesis and cleft palate. *Hum. Mol. Genet.*, **11**, 2051–2059.
48. Dollfus, H., Stoetzel, C., Riehm, S., Lahlou Boukoffa, W., Bediard Boulaneb, F., Quillet, R., Abu-Eid, M., Speeg-Schatz, C., Francfort, J.J., Flament, J. *et al.* (2003) Sporadic and familial blepharophimosis-ptosis-epicanthus inversus syndrome: FOXL2 mutation screen and MRI study of the superior levator eyelid muscle. *Clin. Genet.*, **63**, 117–120.
49. Gudmundsson, J., Sulem, P., Gudbjartsson, D.F., Jonasson, J.G., Sigurdsson, A., Bergthorsson, J.T., He, H., Blondal, T., Geller, F., Jakobsdottir, M. *et al.* (2009) Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat. Genet.*, **41**, 460–464.
50. De Felice, M., Ovitt, C., Biffali, E., Rodriguez-Mallon, A., Arra, C., Anastassiadis, K., Macchia, P.E., Mattei, M.G., Mariano, A., Scholer, H. *et al.* (1998) A mouse model for hereditary thyroid dysgenesis and cleft palate. *Nat. Genet.*, **19**, 395–398.
51. Dathan, N., Parlato, R., Rosica, A., De Felice, M. and Di Lauro, R. (2002) Distribution of the titf2/foxe1 gene product is consistent with an important role in the development of foregut endoderm, palate, and hair. *Dev. Dyn.*, **224**, 450–456.
52. Trueba, S.S., Auge, J., Mattei, G., Etchevers, H., Martinovic, J., Czernichow, P., Vekemans, M., Polak, M. and Attie-Bitach, T. (2005) 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.*, **90**, 455–462.
53. Brancaccio, A., Minichiello, A., Grachtchouk, M., Antonini, D., Sheng, H., Parlato, R., Dathan, N., Dlugosz, A.A. and Missero, C. (2004) Requirement of the forkhead gene Foxe1, a target of sonic hedgehog signaling, in hair follicle morphogenesis. *Hum. Mol. Genet.*, **13**, 2595–2606.
54. Hu, D. and Helms, J.A. (1999) The role of sonic hedgehog in normal and abnormal craniofacial morphogenesis. *Development*, **126**, 4873–4884.
55. Venza, M., Visalli, M., Venza, I., Torino, C., Tripodo, B., Melita, R. and Teti, D. (2009) Altered binding of MYF-5 to FOXE1 promoter in non-syndromic and CHARGE-associated cleft palate. *J. Oral Pathol. Med.*, **38**, 18–23.
56. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
57. Eichberger, T., Regl, G., Ikram, M.S., Neill, G.W., Philpott, M.P., Aberger, F. and Frischauf, A.M. (2004) FOXE1, a new transcriptional target of GLI2 is expressed in human epidermis and basal cell carcinoma. *J. Invest. Dermatol.*, **122**, 1180–1187.
58. Moreno, L.M., Arcos-Burgos, M., Marazita, M.L., Krahn, K., Maher, B.S., Cooper, M.E., Valencia-Ramirez, C.R. and Lidral, A.C. (2004) Genetic analysis of candidate loci in non-syndromic cleft lip families from Antioquia-Colombia and Ohio. *Am. J. Med. Genet. A*, **125A**, 135–144.
59. Murray, J.C., Daack-Hirsch, S., Buetow, K.H., Munger, R., Espina, L., Paglinawan, N., Villanueva, E., Rary, J., Magee, K. and Magee, W. (1997) Clinical and epidemiologic studies of cleft lip and palate in the Philippines. *Cleft Palate Craniofac. J.*, **34**, 7–10.
60. Nguyen, R.H., Wilcox, A.J., Moen, B.E., McConaughy, D.R. and Lie, R.T. (2007) Parent's occupation and isolated orofacial clefts in Norway: a population-based case–control study. *Ann. Epidemiol.*, **17**, 763–771.
61. Romitti, P.A., Lidral, A.C., Munger, R.G., Daack-Hirsch, S., Burns, T.L. and Murray, J.C. (1999) 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*, **59**, 39–50.
62. Service, S., DeYoung, J., Karayiorgou, M., Roos, J.L., Pretorius, H., Bedoya, G., Ospina, J., Ruiz-Linares, A., Macedo, A., Palha, J.A. *et al.* (2006) Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat. Genet.*, **38**, 556–560.
63. (2003) The International HapMap Project. *Nature*, **426**, 789–796.
64. Laird, N.M., Horvath, S. and Xu, X. (2000) Implementing a unified approach to family-based tests of association. *Genet Epidemiol*, **19** (Suppl 1), S36–S42.
65. Weinberg, C.R. (1999) Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. *Am. J. Hum. Genet.*, **65**, 229–235.
66. Schug, J. and Overton, G.C. (1998) *Current Protocols in Bioinformatics*, Wiley.
67. Jiang, R., Bush, J.O. and Lidral, A.C. (2006) Development of the upper lip: morphogenetic and molecular mechanisms. *Dev. Dyn.*, **235**, spc1.
68. Marazita, M.L., Lidral, A.C., Murray, J.C., Field, L.L., Maher, B.S., Goldstein McHenry, T., Cooper, M.E., Govil, M., Daack-Hirsch, S., Riley, B. *et al.* (2009) 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.*, **68**, 151–170.