Published in final edited form as:
Nat Genet. 2010 June ; 42(6): 525–529. doi:10.1038/ng.580.

A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near *MAFB* and *ABCA4*

Terri H Beaty[1], **Jeffrey C Murray**[2], **Mary L Marazita**[3], **Ronald G Munger**[4], **Ingo Ruczinski**[1], **Jacqueline B Hetmanski**[1], **Kung Yee Liang**[1], **Tao Wu**[1], **Tanda Murray**[1], **M Daniele Fallin**[1], **Richard A Redett**[5], **Gerald Raymond**[5], **Holger Schwender**[1], **Shin C Jin**[1], **Margaret E Cooper**[3], **Martine Dunnwald**[2], **Maria A Mansilla**[2], **Elizabeth Leslie**[2], **Stephen Bullard**[6], **Andrew C Lidral**[6], **Lina M Moreno**[6], **Renato Menezes**[3], **Alexandre R Vieira**[3], **Aline Petrin**[2], **Allen J Wilcox**[7], **Rolv T Lie**[8], **Ethylin W Jabs**[9], **Yah Huei Wu-Chou**[10], **Philip K Chen**[10], **Hong Wang**[11], **Xiaoqian Ye**[9,12], **Shangzhi Huang**[13], **Vincent Yeow**[14], **Samuel S Chong**[15], **Sun Ha Jee**[16], **Bing Shi**[17], **Kaare Christensen**[18], **Doheny Kimberly**[19], **W Pugh Elizabeth**[19], **Ling Hua**[19], **E Castilla Eduardo**[20], **Andrew E Czeizel**[21], **Lian Ma**[22], **L Leigh Field**[23], **Lawrence Brody**[24], **Faith Pangilinan**[24], **James L Mills**[25], **Anne M Molloy**[26], **Peadar N Kirke**[27], **John M Scott**[26], **Mauricio Arcos-Burgos**[28], and **Alan F Scott**[5]

[1] Johns Hopkins University, School of Public Health, Baltimore, MD [2] Department of Pediatrics, University of Iowa, Iowa City, IA [3] University of Pittsburgh, School of Dental Medicine, Pittsburgh, PA [4] Utah State University, Logan, UT [5] Johns Hopkins University, School of Medicine, Baltimore, MD [6] University of Iowa, Dept. of Orthodontics, Iowa City, IA [7] NIEHS/NIH, Durham, North Carolina [8] University of Bergen, Bergen, Norway [9] Mt. Sinai Medical School, New York, NY [10] Chang Gung Memorial Hospital, Taoyuan, Taiwan [11] Peking University Health Science Center, Beijing, China [12] Wuhan University, Wuhan, China [13] Peking Union Medical College, Beijing, China [14] KK Women's & Children's Hospital, Singapore [15] National University of Singapore, Singapore [16] Yonsei University, Epidemiology & Health Promotion, Seoul, Korea [17] West China School of Stomatology, Chengdu, China [18] University of Southern Denmark, Odense, Denmark [19] Center for Inherited Disease Research, Johns Hopkins University, Baltimore MD [20] Department of Genetics, FIOCRUZ, Rio de Janeiro, Brazil [21] Foundation for the Community Control of Hereditary Diseases, Budapest, Hungary [22] School of Stomatology, Beijing University, Beijing, China [23] Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada [24] NHGRI/NIH, Bethesda MD [25] NICHD/NIH, Bethesda, MD [26] Trinity College Dublin, Dublin, Ireland [27] Health Research Board, Dublin, Ireland [28] University of Miami, Miller School of Medicine, Miami, FL

Abstract

Case-parent trios were used in a genome wide association study of cleft lip with/without cleft palate (CL/P). SNPs near two genes not previously associated with CL/P [*MAFB*: most significant SNP rs13041247, with odds ratio per minor allele OR=0.704; 95%CI=0.635,0.778; $p=2.05*10^{-11}$; and *ABCA4:* most significant SNP rs560426, with OR=1.432; 95%CI=1.292,1.587; $p=5.70*10^{-12}$] and two previously identified regions (chr. 8q24 and *IRF6*) attained genome wide significance. Stratifying trios into European and Asian ancestry groups revealed differences in statistical significance, although estimated effect sizes were similar. Replication studies from several populations showed confirming evidence, with families of European ancestry giving stronger evidence for markers in 8q24 while Asian families showed stronger evidence for *MAFB* and *ABCA4*. Expression studies support a role for *MAFB* in palate development.

Corresponding author: THB (tbeaty@jhsph.edu).

Cleft lip with or without cleft palate (CL/P) is a common human birth defect with documented genetic and environmental risk factors[1]. While CL/P can occur in many Mendelian malformation syndromes, the isolated, non-syndromic form constitutes 70% of all cases2. Evidence for genetic control of CL/P is compelling: recurrence risks are 20–30 times greater than population prevalences3,4 and both twin and family studies[5] suggest a major role for genes,

Mutations in *IRF6* cause VanderWoude syndrome, the most common Mendelian syndrome including CL/P, and markers in *IRF6* have repeatedly shown evidence of association with isolated, non-syndromic CL/P[6–9]. An allele disrupting an AP2 binding site near *IRF6* showed particularly strong evidence among European CL families, although multiple risk alleles are likely[10].

Birnbaum et al.[11] conducted a case-control genome wide association study (GWAS) in Germany and found significant evidence of association with markers in 8q24.21, and a US case-control GWAS confirmed this region[12], with rs987525 being the most significant marker in both studies. Here we present a GWAS using a case-parent trio design in a consortium drawing cases from Europe, the US, China, Taiwan, Singapore, Korea and the Philippines. This design has the advantage of being robust to confounding due to population stratification, which is important when cases from diverse populations are combined.

Results

Because these case-parent trios came from different populations (Table 1), we conducted a principal components analysis (PCA) on all parents to document genetic variation in our consortium (Supplementary Figure 1). Approximately 50% of parents could be classified as Asian and 45% as European, with remaining parents being of African or "other" ancestry (including mixed). Transmission disequilibrium tests (TDT) on autosomal SNPs in 1908 CL/P case-parent trios showed strong evidence of linkage and association for multiple markers (see QQ plot in Supplementary Figure 2), which clustered into specific chromosomal regions (Figure 1a). Multiple SNPs on chr. 8q24 and 4 SNPs in *IRF6* showed genome wide significance ($p<5*10^{-8}$). In addition, SNPs in two genes not previously associated with CL/P (*ABCA4* on chr. 1p22.1 and *MAFB* on 20q12) achieved genome-wide significance (Table 2), and three potential candidate genes (*PAX7* on chr. 1p36, *VAX1* on 10q25.3 and *NTN1* on 17p13) had one or more SNPs near genome-wide significance (Supplementary Table 1). We stratified these trios into 825 trios of European ancestry (Figure 1b) and 1038 of Asian ancestry (Figure 1c) as a check for consistency across racial groups (omitting 45 case-parent trios of African or "other" ancestry). Interestingly, trios of European ancestry (including European Americans) showed stronger support for chr. 8q24, while Asian trios gave the most significant evidence for both new and old candidate genes with weaker evidence for 8q24. However, p-values cannot be the only criteria when interpreting these results.

Multiple SNPs in 8q24 showed evidence at or near genome-wide significance in the allelic TDT. The strongest individual SNP was rs987525 (Table 2) in both the total sample and the European sub-group (p-value=$1.43*10^{-16}$ in the total sample), as in two previous case-control studies12,13. In our trios, rs987525 showed significant over-transmission of the A allele, giving OR(transmission)=1.78 (95%CI=1.55–2.05). Among 825 trios of European ancestry, this OR (transmission) was larger (2.01 with 95%CI=1.69–2.38); than among Asian trios (1.39 with 95%CI=1.09–1.78). Both groups were nominally significant (p-value=$5*10^{-16}$ for European trios; p-value=0.00893 for Asian trios), and both yielded similar patterns of over-transmission despite differences in p-values shown in Figures 1b and 1c.

Conditional logistic regression was used to estimate genotype relative risks under an additive model as the odds ratio of being a case, OR(case), given each additional target allele (arbitrarily defined as the minor allele among parents of European ancestry). Supplementary Figure 3 presents estimated OR(case) for 78 SNPs in a region of signal on 8q24, where multiple SNPs showed distinct over- or under-transmission. Under the additive model, all trios gave an estimated OR(case)=1.73 (95%CI=1.36–2.03) for AT heterozygotes at rs987525 and OR(case) =2.99 (95%CI=1.26–4.10) for AA homozygotes. A more general model with separate effects for heterozygotes and homozygotes yielded estimates of OR(case|AT)=1.58 (95%CI=1.30–1.94) and OR(case|AA)=3.72 (95%CI=2.36–5.87) in the total sample. When trios were stratified into European and Asian ancestry groups, the additive model gave OR(case)=1.91 (95%CI=1.57–2.33) among trios of European ancestry, and OR(case)=1.42 (95%CI=1.08–1.85) among trios of Asian ancestry, again with overlapping 95%CI. A test for heterogeneity between European and Asian trios under this model did not reach statistical significance (likelihood ratio test=3.11 with 1 df; p=0.07).

A lower minor allele frequency (MAF) at rs987525 among Asians compared to Europeans (0.078 vs. 0.260, respectively), resulting in fewer informative Asian parents, could explain differences in statistical significance. Linkage disequilibrium (LD) patterns for parents of European and Asian ancestry were similar (Supplementary Figure 4). Haplotype analysis of markers in this region strengthened evidence from Asian trios somewhat, but could not overcome limitations due to low MAF (data not shown).

SNPs in or near two other genes yielded genome wide significance: *ABCA4* on 1q22.1 and *MAFB* on 20q12 (Table 2). Among 237 SNPs mapping near *MAFB*, a group of 17 SNPs located 20–60Kb 3′ of *MAFB*'s single exon defined a region of signal including 6 SNPs with $p<5*10^{-8}$. Figure 2a shows $-\log_{10}$(p-value) of these SNPs;; Figure 2b shows estimated OR (case) and 95%CI (the null hypothesis value is always 1) and Figure 2c notes their physical location and the *MAFB* exon. Supplementary Figure 5 shows LD patterns (as r^2) for Asian and European parents.

A total of 210 SNPs mapped to the large *ABCA4* gene (with 50 exons) on 1p22.1, and a 78Kb region encompassing 97 SNPs contained two SNPs yielding genome wide significance and several approaching this level (Figure 3a). Figure 3b presents estimated OR(case) and their 95%CI and Figure 3c shows their physical position. Supplementary Figure 6 shows LD (as r^2).

Replication in independent samples focused on 5 SNPs (rs987525 in 8q24 region and 2 SNPs each in *MAFB* and *ABCA4*). Altogether 8,115 individuals from 1,965 CL/P families were drawn from several populations (Supplementary Table 2). Family-based association tests (FBAT, equivalent to the allelic TDT under an additive model in independent trios) were conducted in each population separately and pooled over all families (Supplementary Table 3). Table 3 shows each SNP was nominally significant in populations of similar ancestry to our GWAS sample. Specifically, European ancestry families (both European and European American) gave the strongest evidence for rs987525 in 8q24, while families of Asian ancestry gave stronger evidence for *MAFB* and *ABCA4*. Two SNPs near *MAFB* showed different levels of significance in families of Asian ancestry compared to families of European ancestry. Interestingly, families from Argentina and Colombia confirmed rs987525 in 8q24, while Guatemalan families (who had more Native American ancestry) did not. In Irish trios, conditional logistic regression gave an estimated OR(case)=1.75 (95%CI=1.31–2.35) for rs987525, although a nearby SNP (rs1530300) was even stronger (p=0.00008). Haplotype analysis on 11 SNPs across this 8q24 region yielded still stronger evidence from these 293 Irish trios (data not shown).

Among unrelated Irish controls, the A allele frequency at rs987525 was 0.143, substantially lower than among Irish case parents (0.247). Using allele frequencies from independent control samples from Northern Europe (Denmark, Ireland, Norway), population attributable risks (PAR) were: rs13041247 near *MAFB* gave PAR=11.1% (95%CI=6.7–15.4), and rs560426 near *ABCA4* gave PAR=9.9% (95%CI=6.7–13.2). Similar analysis on rs987525 in 8q24 in Danish and Irish controls gave PAR=10.4% (95%CI=8.4–12.5).

Supplementary Table 1 presents estimated OR(case) and allele frequencies for genes showing signal at or near genome-wide significance. These included recognized or potential candidate genes: *PAX7* on 1p36, *VAX1* on 10q25.3, plus SNPs between *NTN1* on 17p13 and a putative gene *LOC728685* (previously predicted to be a protein coding gene). Among 70 SNPs spanning 221Kb around *PAX7*, 6 had $10^{-7}<p<10^{-5}$. Among 13 SNPs in *VAX1* spanning 90Kb, two SNPs (rs7078160 and rs4752028) approached genome wide significance with TDT and conditional logistic regression (see Supplementary Table 1 for the latter model). SNP rs7078160 was among the most significant in the German case-control GWAS[11] and achieved genome wide significance in an expanded set of case-parent trios[13]. *NTN1* on 17q13.1 spanned 259Kb and included 1 SNP (rs9788972) achieving genome wide significance and 6 other SNPs yielding evidence between $10^{-8}<p<10^{-6}$. SNPs giving strong signals were clustered in the 5′ end of this gene and encompassed *LOC728867*. Supplementary Table 4 lists all SNPs with $p<10^{-5}$ among all trios (4a), trios of Asian ancestry (4b) and trios of European ancestry (4c).

We sequenced the single *MAFB* exon, plus four conserved elements 3′ of *MAFB*, and identified a rare missense variant (H131Q) which was predicted to be damaging to the protein structure (Supplementary Table 5). An additional 357 cases and 360 controls from the Philippines were genotyped, among whom 24 unrelated cases and 5 controls carried this variant. The difference in allele frequencies was significant (p=0.0002), and a TDT on Filipino families was marginally significant (p=0.08), although low MAF meant few informative trios. The H131Q variant was not present in 760 members of the CEPH diversity panel (individuals from 50 populations) nor in 180 European cases and controls. We also sequenced the 50 exons of *ABCA4*, and identified 27 missense variants, 2 of which were predicted to be damaging (R1443H and N380K, Supplementary Table 5).

Whole mount *in situ* hybridization analysis of *Mafb* and immunodetection of expressed *Mafb* was carried out in mice. *Mafb* mRNA and protein were expressed in both craniofacial neuroectoderm and neural-crest derived mesoderm between embryonic (e) day 13.5–14.5 (Figure 4). Expression was strong in epithelium around the palatal shelves and in the medial edge epithelium during palatal fusion. After fusion, *Mafb* expression was stronger in oral epithelium compared to mesenchymal tissue. Similar expression studies for *Abca4* were negative for palatal expression.

Discussion

Case-parent trio designs have two important advantages: 1) as a family based design they are robust to confounding due to population stratification, a critical concern for multi-center studies; and 2) family based tests can provide greater statistical power compared to case-control designs for rare diseases[14]. Robustness becomes a primary concern when samples from multiple populations are combined, as in this consortium. Large sample sizes are required to achieve extreme levels of statistical significance demanded by GWAS, and for most birth defects this means combining samples across populations. Pooling samples can increase statistical power at the cost of increasing genetic heterogeneity, so checking for heterogeneity among sub-groups remains prudent. As seen here, there can be dramatic differences in p-values between subgroups, even when the direction and estimated magnitude of effects are similar.

In this GWAS, two recognized genes/regions were confirmed (*IRF6* and chr. 8q24) and two genes not previously associated with CL/P were identified (*ABCA4* and *MAFB*). *MAFB* is expressed in the mouse palatal shelf. A rare missense mutation in *MAFB* (H131Q) was over-represented in Filipino cases and absent in other populations. The H residue is strongly conserved across species (Supplementary Figure 7), and this change is predicted to impair protein function. *MAFB* is a transcription factor shown to play a role in development of the hindbrain structures, thymus, interneurons, pancreatic islet cells and the hematopoetic system[15]. Its expression pattern in the mouse is consistent with some role in the developing lip and palate. *ABCA4* is a member of a superfamily of transmembrane proteins, and mutations in *ABCA4* play a major role in the etiology of Stargardt disease and related retinopathies (www.ncbi.nlm.nih.gov/omim). Although there is no evidence of any relationship with clefting, more than 30 missense, frameshift and splice site variants in this large gene have been reported.

It is possible all evidence of linkage and association observed here represents indirect associations with other genes or regulatory elements outside any gene. The success of this CL/P GWAS reflects its large sample size, the robust family based approach and inclusion of samples from populations of different ancestry which confirmed previous findings and identified new genes needing further study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We sincerely thank all of the families at each recruitment site for participating in this study, and we gratefully acknowledge the invaluable assistance of clinical, field and laboratory staff who contributed to making this work possible. We are particularly grateful to Kate Durda and Jamie L'Heureux of the University of Iowa for assistance with samples and phenotype data: Laure Henkle for technical assistance; Judith Resick, Carla Brandon, and Kathy Bardi of the University of Pittsburgh for assistance with recruitment; Wendy Carricato, Kathleen Deeley, and Joseph Ruff of the University of Pittsburgh for sample handling; Poorav Patel and Margaret Rose of Johns Hopkins for assistance with data analysis; Felicia Cheah of the National University of Singapore for sample processing and management; Marcie Feldkamp and John Carey, Utah Birth Defects Network, for assistance with recruitment and case-review; Dawnya Pearce, Utah State University for sample collection and processing. Funding to support data collection, genotyping and analysis came from several sources, some to individual investigators and some to the consortium itself. The consortium for GWAS genotyping and analysis was supported by the National Institute for Dental and Craniofacial Research through U01-DE-004425; "International Consortium to Identify Genes & Interactions Controlling Oral Clefts", 2007–2009; TH Beaty, PI. Funding for individual investigators and the replication studies include: R01-DE-014581 (THB); R37-DE08559 (JCM, MLM), R01-DE09886 (MLM, LM, LLF), R01-DE012472 (MLM), R01-DE014677 (ACL, MLM), R01-DE016148 (MLM), P50-DE016215 (JCM, MLM), R21-DE016930 (MLM, EEC, AEC), R01-HD390661 and R01-DE016877 (RGM). JK02-AEO15291 (ACL), March of Dimes Basil O'Connor award #FY 98-0718 & Research Grant #6-FY01-61 (ACL); NIH R03-AR-055313 & NIH P50-DE-16215 (project #3) (MD), K99-DE-018441 (RM). Smile Train Foundation supported data collection in Chengdu (EWJ, BS). This research was supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (AJW). 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, nor the National Institutes of Health.

References

1. Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. Lancet 2009;374:1773–1785. [PubMed: 19747722]
2. Jugessur A, Farlie PG, Kilpatrick N. The genetics of isolated orofacial clefts: from genotypes to subphenotypes. Oral Diseases 2009;15:437–453. [PubMed: 19583827]
3. Sivertsen A, Wilcox AJ, Skjærven R, Vindenes HA. Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives. BMJ 2008;336:432–434. [PubMed: 18250102]

4. Grosen D, et al. A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance. J Med Genet. 2009 (epub).
5. Mitchell, LE. Twin studies in oral cleft research. In: Wyszynski, DF., editor. Cleft Lip and Palate. Oxford University Press; New York: 2002. p. 214-221.
6. Zuccherro TM, et al. Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. NEJM 2004;351:769–780. [PubMed: 15317890]
7. Park JW, et al. Association between IRF6 and nonsyndromic cleft lip with or without cleft palate in four populations. Genet Med 2007;9:219–227. [PubMed: 17438386]
8. Vieira AR, Cooper ME, Marazita ML, Orioli IM, Castilla EE. Interferon regulatory factor 6 (IRF6) is associated with oral-facial cleft in individuals that originate in South America. Am J Med Genet 2007;A143A:2075–2078. [PubMed: 17702008]
9. Jugessur A, et al. Genetic variants in IRF6 and the risk of facial clefts: single-marker and haplotype-based analyses in a population-based case-control study of facial clefts in Norway. Genet Epidem 2008;32:413–424.
10. Rahimov F, et al. Disruption of an AP-2α site in an IRF6 enhancer is strongly associated with cleft lip. Nat Genet 2008;40:1341–1347. [PubMed: 18836445]
11. Birnbaum S, et al. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet 2009;41:473–477. [PubMed: 19270707]
12. Grant SFA, et al. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. J Pediatr 2009;155:909–913. [PubMed: 19656524]
13. Mangold E, et al. Genome wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat Genet 2010;42:24–26. [PubMed: 20023658]
14. Laird NM, Lange C. Family based methods for linkage and association analysis. Advances Genet 2008;60:219–252. [PubMed: 18358323]
15. Yang Y, Cvekl A. Large Maf Transcription Factors: Cousins of AP-1 Proteins and Important Regulators of Cellular Differentiation. Einstein J Biol Med 2007;23:2–11. [PubMed: 18159220]

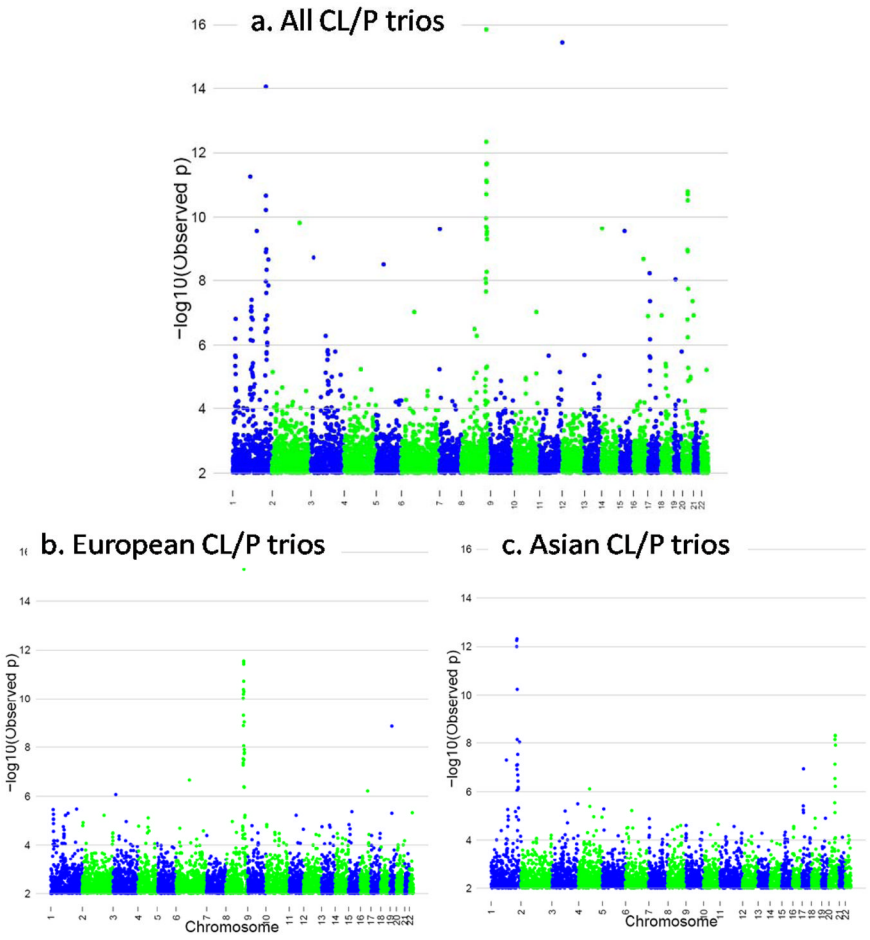

Figure 1.

Manhattan plots of $\log_{10}$(p-values) from transmission disequilibrium test (TDT) for autosomal SNPs on CL/P case-parent trios (omitting SNPs flagged for QC). (a) Results based on all 1908 CL/P trios; (b) Results based on 825 CL/P case-parent trios of European ancestry; (c) Results based on 1038 CL/P case-parent trios of Asian ancestry.

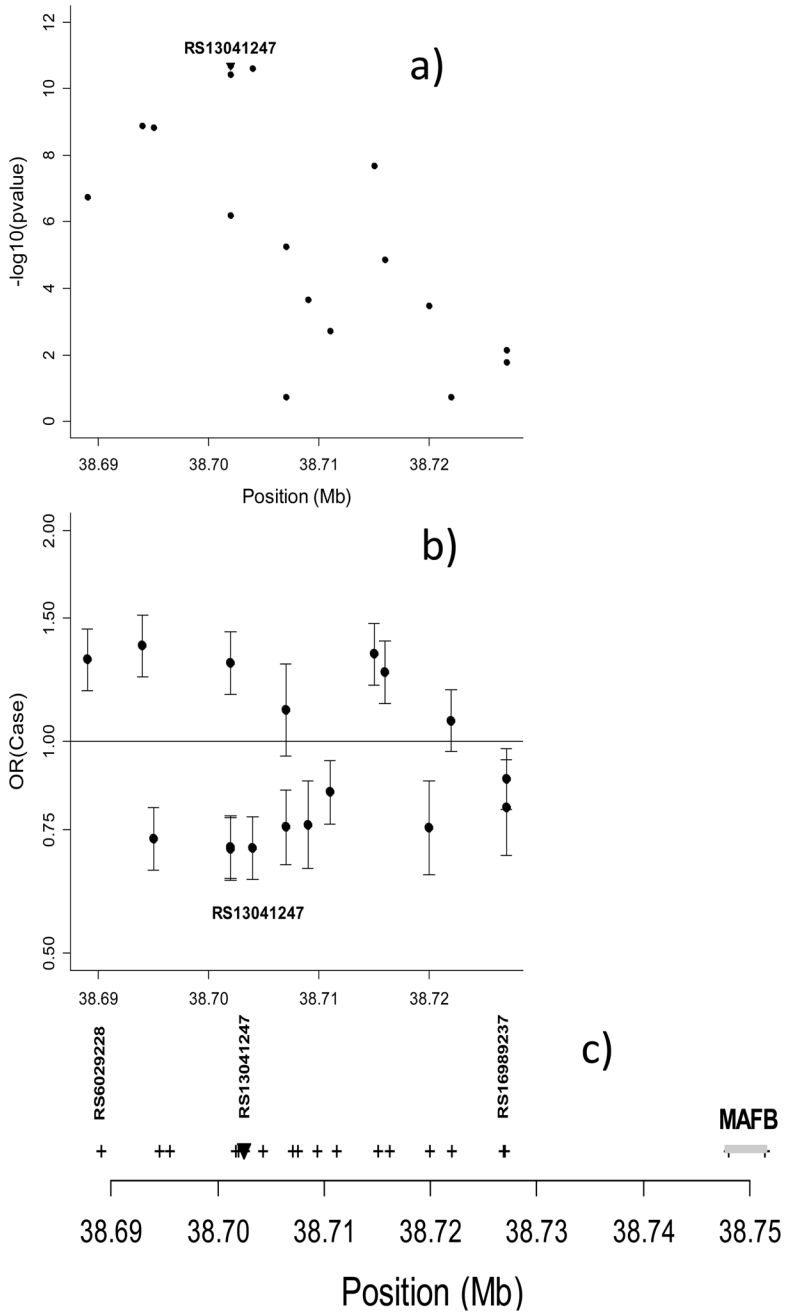

Figure 2.

Significance and effect size for SNPs near *MAFB* based on all CL/P trios. a) $-\log_{10}$(p-value) for allelic TDT for 17 SNPs near *MAFB* on chr. 20q11; b) Estimated OR(case) from a conditional logistic regression and their 95%CI under an additive model; c) Physical position of tested SNPs and the single exon of *MAFB*.

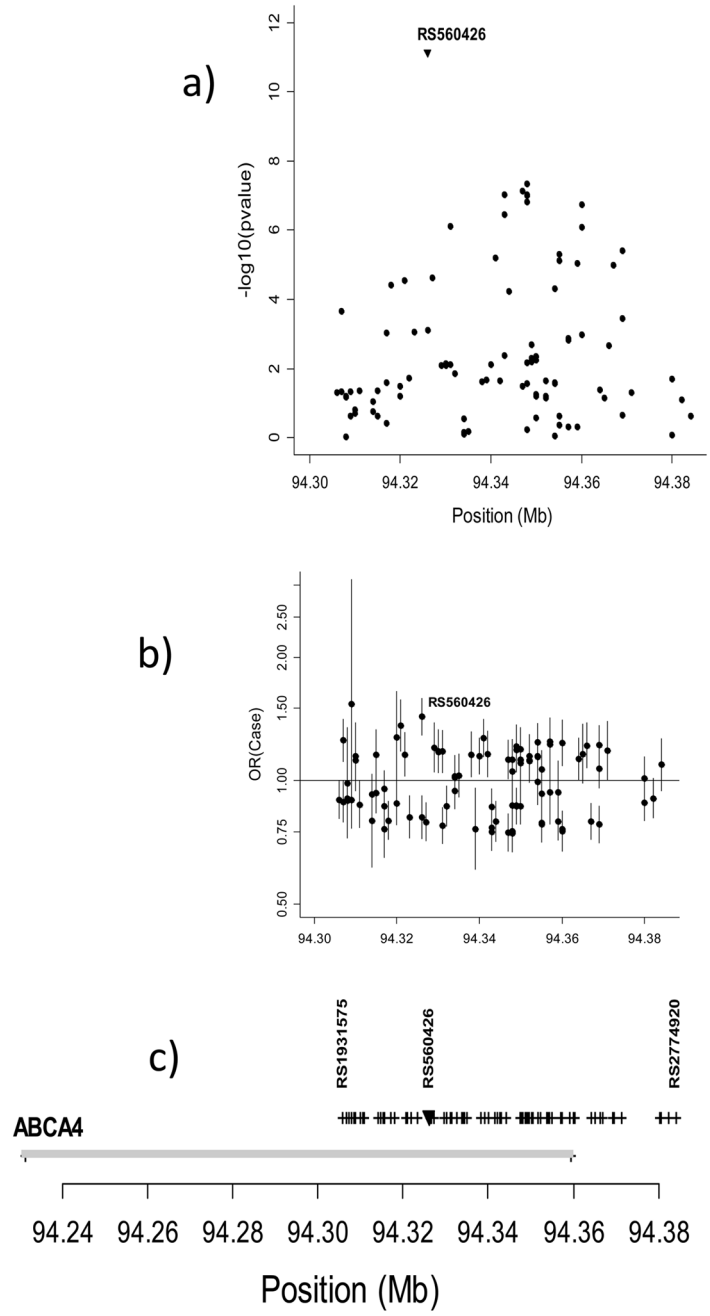

Figure 3.

Significance and effect size for SNPs in and near *ABCA4* based on all CL/P trios. a) $-\log_{10}$(p-value) for allelic TDT for 98 SNPs in or near *ABCA4* on chr. 1p22.1; b) Estimated OR(case) from a conditional logistic regression and their 95%CI under an additive model fit to 1908 CL/P case-parent trios; c) Physical position of tested SNPs and combined exons of the *ABCA4* gene.

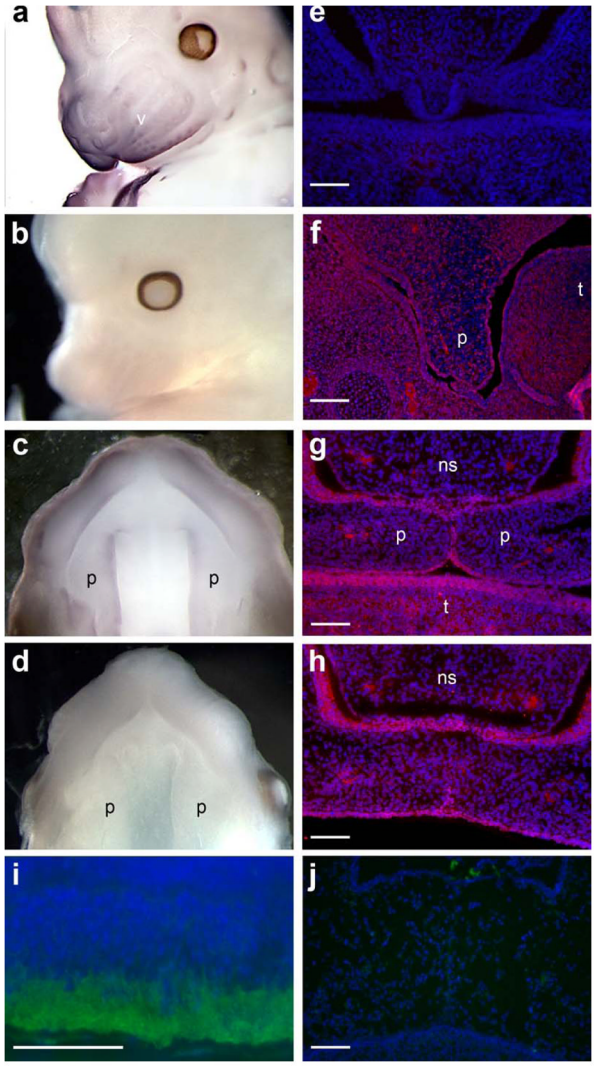

Figure 4.

Mafb, and not *Abca4*, is expressed during the development of the secondary palate in the mouse. In situ hybridization for *Mafb* on whole mount e13.5 embryos (a–d) shows expression in craniofacial ectoderm, vibrissae, and neural-crest derived mesoderm in murine embryos. Signal was also detected in the elevated palatal shelves (b – view of the roof of the mouth). Immunofluorescence staining for *Mafb* (red) on e13.5 palatal sections shows *Mafb* localized in the epithelium of the palatal shelves (f) and in the medial edge epithelium during palatal fusion on e14.5 tissue sections (g, h). Expression is also detected at the base of the nasal septum and in the tongue epithelium (g). Note the absence of signal in the sense probe (b, d) and no primary control (e). Immunofluorescence staining for *Abca4* (green) on adult murine retina (i) and e14.5 palatal sections (j) show the presence of *Abca4* in the rim of rods photoreceptor cells of the retina and its absence in orofacial structures. Nuclei were counterstained with DAPI (blue). v, vibrissae; p, palatal shelf; t, tongue, ns, nasal septum. (Scale bar = 100 μm panels e–h; = 50 μm panel i).

Table 1

Number of trios by recruitment site noting complete and incomplete trios (those with 1 parent missing).

Recruitment Site	CL Trios Complete (Incomplete)	CLP Trios Complete (Incomplete)	Total Trios Complete (Incomplete)
Utah	68(16)	96(20)	164(36)
Norway	106(4)	174(8)	280(12)
Korea	19(0)	40(2)	59(2)
Maryland	19(12)	71(42)	90(54)
Pittsburgh	26(2)	70(28)	96(30)
Singapore	15(1)	45(7)	60(8)
Taiwan	42(4)	176(11)	218(15)
Iowa	16(9)	29(11)	45(20)
Denmark	6(15)	15(12)	21(27)
Philippines	0(0)	94(4)	94(4)
WuHan	39(3)	136(9)	175(12)
Shandong Prov.	54(21)	129(70)	183(91)
Western China	43(3)	63(3)	106(6)
Total	453(90)	1138(227)	1591(317)

* total includes probands of indeterminate cleft type: 2 in WuHan; 3 in Shandong Prov.

Table 2

Estimated OR(case) for SNPs showing genome wide significance in 4 regions under an additive model plus minor allele and its frequency among all parents and among parents of European and Asian CL/P cases.

SNP	Location	OR(case)*	95%CI	P	MA**	Overall MAF	Euro. MAF	Asian MAF
Chr 8q24								
rs987525	130015336	1.781	(1.550,2.047)	1.11E-16	A	0.167	0.276	0.079
IRF6								
rs2073485	208029417	0.689	(0.615,0.771)	1.07E-10	A	0.300	0.169	0.403
rs2013162	208035307	0.705	(0.636,0.782)	2.29E-11	A	0.422	0.335	0.491
rs861020	208043734	1.432	(1.274,1.609)	1.20E-09	A	0.245	0.246	0.244
rs10863790	208054670	0.580	(0.504,0.667)	1.11E-14	C	0.198	0.015	0.342
MAFB								
rs6072081	38694468	1.369	(1.237,1.515)	1.05E-09	A	0.466	0.446	0.398
rs6065259	38695393	0.728	(0.656,0.807)	1.15E-09	A	0.387	0.384	0.389
rs17820943	38701930	0.707	(0.638,0.784)	2.76E-11	T	0.395	0.375	0.414
rs13041247	38702488	0.704	(0.635,0.778)	1.44E-11	C	0.396	0.375	0.414
rs11696257	38704230	0.705	(0.636,0.781)	1.75E-11	T	0.396	0.375	0.414
rs6102085	38715043	1.332	(1.205,1.473)	1.76E-08	G	0.465	0.372	0.462
ABCA4								
rs4147811	94347644	0.745	(0.670,0.828)	3.80E-08	A	0.344	0.358	0.333
rs481931	94342604	0.750	(0.675,0.834)	8.14E-08	A	0.339	0.351	0.332
rs560426	94326026	1.432	(1.292,1.587)	5.01E-12	G	0.399	0.471	0.342

*Under an additive model;

**In the conditional logistic regression model, the minor allele (MA) among Europeans was set to be the target allele and the OR(case|each MA) was estimated.

Table 3

P-values for replication of 5 SNPs showing genome wide significance in GWAS using independent families from various populations.

Source Population	Pedigrees	8q24 region	*MAFB*		*ABCA4*	
		rs987525	rs13041247	rs11696257	rs560426	rs481931
East Asian	331	0.6964	**0.0161**	**0.0009**	**0.0003**	**0.0290**
South Asian	51	0.1172	0.0638	0.1675	0.8299	**0.0382**
European & Euro.American	1,149	**$1.1^{*}10^{-16}$**	**0.0002**	**0.0231**[*]	**0.0058**	0.4418
South/Central American	434	**0.0013**	0.3375	0.9344	0.4487	0.2378
Total	**1965**	**$4.4^{*}10^{-16}$**	**0.0001**	**0.0013**	**$3.3^{*}10^{-5}$**	**0.0487**

[*] Omitting Irish samples (see Supplementary Table 3)