ARTICLE

A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in *GRHL3*

Elizabeth J. Leslie,[1] Huan Liu,[2,3] Jenna C. Carlson,[1,4] John R. Shaffer,[1,5] Eleanor Feingold,[1,4,5] George Wehby,[6] Cecelia A. Laurie,[7] Deepti Jain,[7] Cathy C. Laurie,[7] Kimberly F. Doheny,[8] Toby McHenry,[1] Judith Resick,[1] Carla Sanchez,[1] Jennifer Jacobs,[1] Beth Emanuele,[1] Alexandre R. Vieira,[1,5] Katherine Neiswanger,[1] Jennifer Standley,[9] Andrew E. Czeizel,[10] Frederic Deleyiannis,[11] Kaare Christensen,[12] Ronald G. Munger,[13] Rolv T. Lie,[14] Allen Wilcox,[15] Paul A. Romitti,[16] L. Leigh Field,[17] Carmencita D. Padilla,[18] Eva Maria C. Cutiongco-de la Paz,[18,19,20] Andrew C. Lidral,[21] Luz Consuelo Valencia-Ramirez,[22] Ana Maria Lopez-Palacio,[23] Dora Rivera Valencia,[24] Mauricio Arcos-Burgos,[25] Eduardo E. Castilla,[26,27,28] Juan C. Mereb,[29] Fernando A. Poletta,[26,27,28] Iêda M. Orioli,[28,30] Flavia M. Carvalho,[26,27] Jacqueline T. Hecht,[31] Susan H. Blanton,[32] Carmen J. Buxó,[33] Azeez Butali,[34] Peter A. Mossey,[35] Wasiu L. Adeyemo,[36] Olutayo James,[36] Ramat O. Braimah,[37] Babatunde S. Aregbesola,[37] Mekonen A. Eshete,[38] Milliard Deribew,[38] Mine Koruyucu,[39] Figen Seymen,[39] Lian Ma,[40] Javier Enríquez de Salamanca,[41] Seth M. Weinberg,[1] Lina Moreno,[21] Robert A. Cornell,[2] Jeffrey C. Murray,[9] and Mary L. Marazita[1,5,42,*]

Cleft palate (CP) is a common birth defect occurring in 1 in 2,500 live births. Approximately half of infants with CP have a syndromic form, exhibiting other physical and cognitive disabilities. The other half have nonsyndromic CP, and to date, few genes associated with risk for nonsyndromic CP have been characterized. To identify such risk factors, we performed a genome-wide association study of this disorder. We discovered a genome-wide significant association with a missense variant in *GRHL3* (p.Thr454Met [c.1361C>T]; rs41268753; $p = 4.08 \times 10^{-9}$) and replicated the result in an independent sample of case and control subjects. In both the discovery and replication samples, rs41268753 conferred increased risk for CP (OR = 8.3, 95% CI 4.1–16.8; OR = 2.16, 95% CI 1.43–3.27, respectively). In luciferase transactivation assays, p.Thr454Met had about one-third of the activity of wild-type GRHL3, and in zebrafish embryos, perturbed periderm development. We conclude that this mutation is an etiologic variant for

[1]Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15219, USA; [2]Department of Anatomy and Cell Biology, University of Iowa, Iowa City, IA 52242, USA; [3]State Key Laboratory Breeding Base of Basic Science of Stomatology (Hubei-MOST) and Key Laboratory for Oral Biomedicine of Ministry of Education, School and Hospital of Stomatology, Wuhan University, Wuhan 430079, China; [4]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; [5]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; [6]Department of Health Management and Policy, College of Public Health, University of Iowa, Iowa City, IA 52246, USA; [7]Department of Biostatistics, Genetic Coordinating Center, University of Washington, Seattle, WA 98195, USA; [8]Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD 21224, USA; [9]Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA; [10]Foundation for the Community Control of Hereditary Diseases, Budapest 1051, Hungary; [11]Department of Surgery, Plastic and Reconstructive Surgery, University of Colorado School of Medicine, Denver, CO 80045, USA; [12]Department of Epidemiology, Institute of Public Health, University of Southern Denmark, Odense 5230, Denmark; [13]Department of Nutrition, Dietetics, and Food Sciences, Utah State University, Logan, UT 84322, USA; [14]Department of Global Public Health and Primary Care, University of Bergen, Bergen 5020, Norway; [15]Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA; [16]Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, IA 52246, USA; [17]Department of Medical Genetics, University of British Columbia, Vancouver, BC V6H 3N1, Canada; [18]Institute of Human Genetics, National Institutes of Health, University of the Philippines Manila, Manilla 1000, the Philippines; [19]Department of Pediatrics, College of Medicine, University of the Philippines Manila, Manilla 1000, the Philippines; [20]Philippine Genome Center, University of the Philippines, Manilla 1101, the Philippines; [21]Department of Orthodontics, College of Dentistry, University of Iowa, Iowa City, IA 52242, USA; [22]Fundación Clínica Noel, Medellin 050012, Colombia; [23]Department of Basic Integrated Studies, College of Dentistry, University of Antioquia, Medellin 050001, Colombia; [24]Population Genetics and Mutacarcinogenesis Group, University of Antioquia, Medellin 050001, Colombia; [25]Genomics and Predictive Medicine, Genome Biology Department, John Curtin School of Medical Research, ANU College of Medicine, Biology & Environment, The Australian National University, Canberra, ACT 0200, Australia; [26]CEMIC: Center for Medical Education and Clinical Research, Buenos Aires 1431, Argentina; [27]Laboratory of Congenital Malformation Epidemiology, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro 21040-360, Brazil; [28]ECLAMC (Latin American Collaborative Study of Congenital Malformations) at INAGEMP (National Institute of Population Medical Genetics), Rio de Janeiro 21941-617, Brazil; [29]ECLAMC (Latin American Collaborative Study of Congenital Malformations) at Hospital de Area, El Bolson 8430, Argentina; [30]Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro, Rio de Janeiro 21941-617, Brazil; [31]Department of Pediatrics, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [32]Dr. John T. Macdonald Foundation Department of Human Genetics, Hussman Institute for Human Genomics, Mailman School of Medicine, University of Miami, Miami, FL 33124, USA; [33]School of Dental Medicine, University of Puerto Rico, San Juan 00936, Puerto Rico; [34]Department of Oral Pathology, Radiology and Medicine, Dows Institute for Dental Research, College of Dentistry, University of Iowa, Iowa City, IA 52242, USA; [35]Department of Orthodontics, University of Dundee, Dundee DD1 4HN, Scotland; [36]Department of Oral and Maxillofacial Surgery, College of Medicine, University of Lagos, Lagos, P.M.B. 12003, Nigeria; [37]Department of Oral and Maxillofacial Surgery, Obafemi Awolowo University Ile-Ife, Ife-Ife, P.M.B. 13, Nigeria; [38]Surgical Department, School of Medicine, Addis Ababa University, Addis Ababa, P.O. Box 26493, Ethiopia; [39]Department of Pedodontics, Istanbul University, Istanbul 34116, Turkey; [40]School of Stomatology, Peking University, Beijing 100081, China; [41]Hospital Infantil Universitario Niño Jesús, Unidad de Cirugía Plástica, Madrid 28009, Spain; [42]Clinical and Translational Science, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA

*Correspondence: marazita@pitt.edu

http://dx.doi.org/10.1016/j.ajhg.2016.02.014.

nonsyndromic CP and is one of few functional variants identified to date for nonsyndromic orofacial clefting. This finding advances our understanding of the genetic basis of craniofacial development and might ultimately lead to improvements in recurrence risk prediction, treatment, and prognosis.

Introduction

Orofacial clefts (OFCs [MIM: 119530]) comprise any developmentally arising cleft, i.e., disruption or gap, in orofacial structures including the lips, palate, eyes, and nose.[1] Of these, cleft lip (CL), cleft palate (CP), and cleft lip with cleft palate (CLP) are among the most common birth defects in humans with prevalence between 1 in 500 and 1 in 2,500 live births.[2] Because other types of OFCs are exceedingly rare, we will use OFCs here to describe the subset affecting only the lip and palate (i.e., CL, CLP, and CP).

OFCs arise from failure of normal craniofacial developmental processes that requires coordinated cell growth, migration, differentiation, and apoptosis.[3] The upper lip develops during the 4^{th}–7^{th} weeks of embryogenesis from several events culminating in the growth, migration, and fusion of the medial and lateral nasal processes with the maxillary processes. Disruption of any of these events results in CL. Development of the secondary palate begins in the 7^{th} week of embryogenesis when the palatal shelves emerge from the maxillary processes. Growth and elevation of the palatal shelves followed by fusion at the midline results in a complete separation of the nasal and oral cavities. Failure of any of these steps can lead to CP. Because of the distinct developmental origins of the lip and palate, OFCs are commonly divided into CL with or without CP (CL/P, i.e., CL and CLP) and CP.[4,5]

In addition, OFCs are further divided into two classes—syndromic and nonsyndromic—for the purposes of clinical diagnosis, recurrence risk predictions, and identification of etiologies. Approximately 70%–80% of CL/P and 50% of CP[6] are considered nonsyndromic based on the absence of additional structural or cognitive abnormalities. There is now convincing evidence that nonsyndromic OFCs represent complex human disorders with genetic risk factors, environmental exposures, and their interactions, all contributing to susceptibility.

Although a number of the genes in which mutations cause syndromic CP have been elucidated,[7] few have been identified for nonsyndromic CP (MIM: 119540). One gene that might influence risk for CP is *FAF1* (MIM: 604460), which was initially identified from a translocation in a multiplex CP-affected family with Pierre Robin sequence (MIM: 261800).[8] Subsequently, SNPs near *FAF1* were shown to be associated with nonsyndromic CP in trios of European ancestry.[8] However, the FAF1 locus only partially accounts for the heritability of CP, and therefore, we hypothesize that other risk loci exist. Whereas five GWASs of CL/P have identified at least 15 genetic loci with compelling statistical support,[9] only one GWAS for CP has been completed with no statistically significant SNP main effects for *FAF1* or any other gene.[10] Insufficient sample sizes for genomic studies could account for the paucity of gene discovery for CP. However, several significant gene-environment interaction effects were observed upon inclusion of maternal smoking, alcohol, or vitamin exposures.[10] Moreover, many of the CL/P risk factors identified to date have also been studied in CP cohorts with mixed results, suggesting differences in the underlying genetic etiology. For example, SNPs from the four European CL/P-associated loci (1q32, 8q24, 10q25, and 17q22) were tested in CP trios but were not significantly associated.[11] Additional CL/P SNPs were tested for association with submucous CP, but again were not significantly associated.[12] However, other candidate gene studies have shown association with both CL/P and CP including a SNP within microRNA-140 (MIM: 611894)[13] and SNPs near *FOXE1* (MIM: 602617).[14] Overall there seems to be little overlap between loci associated with CL/P and the few loci associated with CP to date, supporting the notion that CL/P and CP largely have distinct genetic etiologies.

Here we report the results of a GWAS of nonsyndromic CP in a multiethnic sample. We identified a genome-wide significant association with nonsyndromic CP: a missense variant in *GRHL3* (MIM: 608317) significantly associated in case subjects of European ancestry and independently replicated. We further examined this variant in a zebrafish model and cell-based transactivation assays and conclude that it is an etiologic variant for nonsyndromic CP.

Subjects and Methods

Discovery Sample Description

Our current study was part of a collaborative effort to investigate the genetic basis of orofacial clefts in participants recruited from 13 countries across five continents. Most of the sites were part of ongoing studies led by the University of Pittsburgh Center for Craniofacial and Dental Genetics and the University of Iowa. Recruitment and study assessments were done at regional OFC treatment centers. Informed consent was obtained for all participants and all sites had both local IRB approval and approval at the University of Pittsburgh or the University of Iowa. The affection status for the current study was cleft palate (CP), defined as any disruption of the hard and/or soft palate, including submucous CP. The discovery study sample consisted of two independent subsets: (1) a set of 165 unrelated case-parent trios and (2) 78 unrelated CP case subjects and 1,700 unrelated control subjects with no family history of craniofacial birth defects (Table S1).

Genotyping and Quality Control

Participants were genotyped for approximately 580,000 SNPs on an Illumina HumanCoreExome array plus 15,890 SNPs of custom content designed to cover candidate genes and loci implicated by previous genome-wide or candidate gene studies or from animal models with craniofacial anomalies. Genetic data were extensively

cleaned and quality checked via standard analysis pipelines developed by the Genetics Coordinating Center (GCC) at the University of Washington.[15] These analyses included interrogation of samples for unexpected duplicates, genetic sex, chromosomal anomalies, familial relatedness, population structure, missingness, batch effects, duplicate sample discordance, and Mendelian errors among relatives. The quality of SNP probes was examined by inter-sample comparisons (i.e., concordance across sample duplicates and HapMap controls, Mendelian errors, etc.), missing call rates, deviations from Hardy-Weinberg equilibrium (in a genetically defined homogenous subset of European participants), concordance across duplicate probes, and separation of clusters during genotype calling. A total of 539,473 SNPs (comprising 96.74% of those attempted) passed quality filters (see Table S3). Of these, 293,633 SNPs had minor allele frequencies of 1% or greater. Principal-component analysis (PCA)-based methods were used to assess ancestry and assign participants into four genetic analysis groups reflecting continental origin: Europe, Asia, Africa, and Central/South America. Genotypes and phenotypes for this project are available on dbGaP. Further details regarding the quality assurance, ancestry analyses, imputation, and data cleaning efforts are available in the Quality Control Report issued by the University of Washington GCC for our project, available online. Results can be browsed as part of the Facebase Consortium's Human Genomics Analysis Interface (see Web Resources).

Imputation

Prior to imputation of unobserved genetic polymorphisms, SNPs passing quality filters were "pre-phased" to create haplotypes using SHAPEIT2.[16] Imputation was conducted using IMPUTE2[17] and phase 3 of the 1000 Genomes Project (comprising 2,504 individuals from 26 populations worldwide) as the reference panel. Genotype data for a total of 34,985,077 SNPs were available after imputation, including both fully imputed SNPs as well as genotyped SNPs for which sporadic missingness was imputed. Masked variant analysis, whereby genotyped SNPs were imputed in order to assess imputation accuracy, demonstrated high-quality imputation, with mean concordance of 0.995 for SNPs with minor allele frequency (MAF) < 0.05 and 0.960 for SNPs with MAF ≥ 0.05. The "most-likely" genotypes (i.e., genotypes with the highest probability [Q]) were selected for statistical analysis only if the genotype with the highest probability was greater than 0.5. Imputed SNPs with deviations from Hardy-Weinberg equilibrium were filtered out of downstream analyses.

Statistical Analyses

The European GWAS data were analyzed in the case-control subset using logistic regression while adjusting for five PCs of ancestry (defined in the European ancestry group), and in the trio subset via the transmission disequilibrium test (TDT).[18] Both the TDT and logistic regression analyses were performed as implemented in PLINK.[19] Association results for the two GWAS scans were combined by weighted odds ratio meta-analysis[20] to estimate our study-wide association results. We extended the GWAS approach to the full multiethnic sample using the same two-stage approach. For the multiethnic GWAS, 18 PCs of ancestry were included in the logistic regression model to account for population structure. Note, due to small sample sizes, stratified analyses were not performed individually in the Asian, Central/South American, or African groups.

Replication Sample and Genotyping

Samples utilized for replication genotyping came from five population-based OFC case-control studies (two American and three European): the Danish National Birth Cohort (DNBC),[21] the Iowa Case-Control Study (Iowa),[22] the Norway Facial Clefts Study (NCL),[23] the Norwegian Mother and Child Cohort (MoBa),[24] and the Utah Child and Family Health Study (Utah)[25] (see Table S2). Case and control subjects were identified primarily through national or state-based birth registries. Seven SNPs from three loci were genotyped using TaqMan SNP Genotyping Assays (Life Technologies) on the Fluidigm microfluidic EP1 SNP Genotyping System and GT192.24 Dynamic Array Integrated Fluidic Circuits (IFCs) (Fluidigm). Statistical analysis was performed using logistic regression, for case versus control status for each SNP, one at a time, controlling for study site.

A second, in silico, replication sample came from the GENEVA International Cleft Consortium, which previously published GWASs of CL/P and CP.[10,26] A total of 468 CP case-parent trios of varied ancestry were used from sites across the US, Europe, and Southeast and East Asia. Samples were genotyped on the Illumina Human610-Quad platform and imputed by the GENEVA Coordinating Center using the IMPUTE2 software and the 1000 Genomes Project reference panel (June 2011 release). Statistical analysis was performed with TDT, implemented in PLINK.

Sanger Sequencing

Sanger sequencing was performed using primers and methodology as previously described[27] on 84 independent case subjects with CP recruited at the University of Iowa. PCR products were sent for sequencing on an ABI 3730XL (Functional Biosciences). Chromatograms were transferred to a UNIX workstation, base-called with PHRED (v.0.961028), assembled with PHRAP (v.0.960731), scanned by POLYPHRED (v.0.970312), and viewed with the CONSED program (v.4.0). Variants were compared with the Exome Aggregation Consortium browser.

Functional Test of Human GRHL3 Mutation Variants in Zebrafish Embryos

Full-length, wild-type human *GRHL3* cDNA (GenBank: BC036890.2) was obtained from GE Healthcare. The p.Thr454Met (c.1361C>T), p.Val160Ala (c.479T>C), and p.Asn484Ser (c.1451A>G) substitutions were introduced using PCR-mediated mutagenesis.[28] cDNA coding wild-type, p.Thr454Met, p.Val160Ala, and p.Asn484Ser GRHL3 were shuttled into CS2+ vector and the corresponding capped mRNAs were generated in vitro with mMESSAGE mMACHINE SP6 kit (Ambion) and purified with RNeasy mini kit (QIAGEN). After injecting a range of quantities of mRNA, it was determined that embryos tolerated 1 ng of lacZ mRNA without a significant reduction in survival. Approximately 1 ng of mRNA was injected into wild-type zebrafish embryos (inbred from NHGRI line) at the 1-cell stage.[29] Embryos were fixed at 6 hpf and whole-mount in situ hybridization for *krt4* was performed as described previously.[30] Animal use protocols were approved by the Public Health Service Assurance.

Luciferase Reporter Constructs, Electroporation, and Luciferase Assays

For test of the function of different GRHL3 variants, a synthetic GRHL3-sensitive enhancer was generated by synthesizing an oligonucleotide containing four replicates of the *GRHL3*

Table 1. Top Hits from GWAS of Cleft Palate

Locus	SNP	Analysis	European Ancestry		Combined Multi-ethnic Sample	
			OR [99.999995% CI]	p Value	OR [99.999995% CI]	p Value
1p36 (*GRHL3*)	rs41268753	CC	8.3 (1.17,59.15)	4.08×10^{-9}	6.3 (1.04,38.86)	2.75×10^{-8}
		TDT	1.4 (0.058,34.07)	0.56	2.0 (0.13,30.53)	0.15
		Meta	5.1 (0.96, 27.13)	1.10×10^{-7}	4.5 (0.98,20.14)	6.90×10^{-8}
6p26 (*PARK2*)	rs12175475	CC	7.4 (0.77,71.21)	1.42×10^{-6}	6.6 (1.12,39.58)	8.66×10^{-9}
		TDT	1.0 (0.032,31.43)	1.00	1.3 (0.08,20.07)	0.61
		Meta	4.1 (0.61,26.88)	5.55×10^{-5}	4.1 (0.91,18.18)	5.39×10^{-6}
11q22.1 (*YAP1*)	rs117496742	CC	11.2 (1.04,121.8)	3.13×10^{-8}	9.2 (0.91,92.28)	1.64×10^{-7}
		TDT	2.0 (0.04,94.42)	0.31	2.0 (0.04,94.42)	0.31
		Meta	7.0 (0.92,52.94)	1.77×10^{-7}	6.1 (0.85,44.48)	5.88×10^{-7}

consensus binding motif[31] (sequence: 5′-CACCCCTCGAGGTC GACGGTATCGATAAGCTTGATATCCACTTGGGATATCGCCCTTA AAACCGGTTTAAGCTACGTACGAAAACCGGTTTACAGAAAACC GGTTTAACGATCGAAAACCGGTTTAAGGGCGAATTCCACATTG GTCGCTGCAGCCCGGGGGATCCACTAGTTCTAGAGCGGC-3′) and engineered into pTol2-cFos-FLuc. pTol2-cFos-FLuc was engineered from pT2cfosGW,[32] by replacing the GFP gene with FLuc. For electroporation in human oral epithelial cell line, GMSM-K[33] was electroporated with Amaxa Cell Line Nucleofector Kit V (LONZA) in Nucleofector II Device (Program: X-005) (LONZA). For dual luciferase assay, GMSM-K was co-electroporated with empty CS2+ plasmids (Gift of David Turner, U. Michigan), wild-type or different *GRHL3* mutation variants overexpression plasmids, along with GRHL3-sensitive reporter described above. The Dual-luciferase Reporter Assay System (Promega) and a luminometer were used to measure luciferase activity in cell lysates 4 days after transfection. All quantified results are presented as mean ± SEM. Three luciferase measurements were made on each of three independent biological replicates. A two-tailed unpaired Student's t test was used to determine statistical significance.

Results

GWAS of Nonsyndromic CP

We performed a GWAS in two independent subsets from a sample of nonsyndromic CP of European ancestry (Table 1). The first was a set of 38 unrelated case subjects and 835 unrelated control subjects and the second was a set of 93 unrelated case-parent trios (Table S1). A total of 455,449 genotyped SNPs and 33,669,354 imputed SNPs were analyzed separately in each subset and combined by meta-analysis (Figure 1). We observed significant associations in the case-control scan on chromosome 1 spanning *GRHL3* and nearby genes (Figures 1A and 2A; lead SNP rs41268753; $p = 4.08 \times 10^{-9}$) and on chromosome 11 near *YAP1* ([MIM: 606608]; Figure S1; rs117496742; $p = 3.13 \times 10^{-8}$). No significant associations were observed in the European TDT (Figure 1B). Both *GRHL3* and *YAP1* showed suggestive associations in a meta-analysis of European case-control and TDT scans (Figure 1C; p values = 6.90×10^{-8} and 5.88×10^{-7}, respectively). Tables S4–S6 contain results for all genotyped SNPs yielding p values less than 1.0×10^{-5} in the European case-control, TDT, and meta-analyses, respectively.

We next conducted a GWAS in a larger multiethnic sample of nonsyndromic CP by adding 40 unrelated case subjects, 865 unrelated control subjects, and 75 trios from Asian, African, and Central/South American continental groups to the European participants studied above (Table S1). A total of 78 unrelated cases, 1,700 unrelated controls, and 165 trios were analyzed using the same statistical approaches. As with the European sample analysis, statistically significant associations were observed in the multiethnic sample case-control analysis (Figure S2A). rs41268753 (*GRHL3*) again showed association, although the statistical significance was slightly decreased (p value = 2.75×10^{-8}), indicating that this signal was driven by the European sample. We also observed a second locus within *PARK2* (MIM: 602544) on chromosome 6 (Figure S3; lead SNP rs12175475; p value = 8.66×10^{-9}). No significant associations were observed in the multiethnic trio scan (Figure S2B), whereas the multiethnic meta-analysis showed an association for the *GRHL3* locus (Figure S2C; lead SNP rs113965554; p value = 5.03×10^{-8}), with rs41268753 showing suggestive evidence of association (p value = 6.90×10^{-8}). Tables S7–S9 contain results for all genotyped SNPs yielding p values less than 1.0×10^{-5} in the multiethnic case-control, TDT, and meta-analysis scans.

Replication

We selected seven SNPs (two SNPs each for *GRHL3* and *YAP1*, three SNPs for *PARK2*) for replication and genotyped each in 246 unrelated case subjects and 1,685 unrelated control subjects from population-based samples of European ancestry (Table 2; for description of the replication cohort see Table S2). Both *GRHL3* SNPs (rs41268753 and rs113965554) showed evidence of replication ($p = 2.8 \times 10^{-4}$ and $p = 6.8 \times 10^{-4}$, respectively). We also performed in silico replication analyses in the cohort from the previously published GWAS of nonsyndromic CP in the

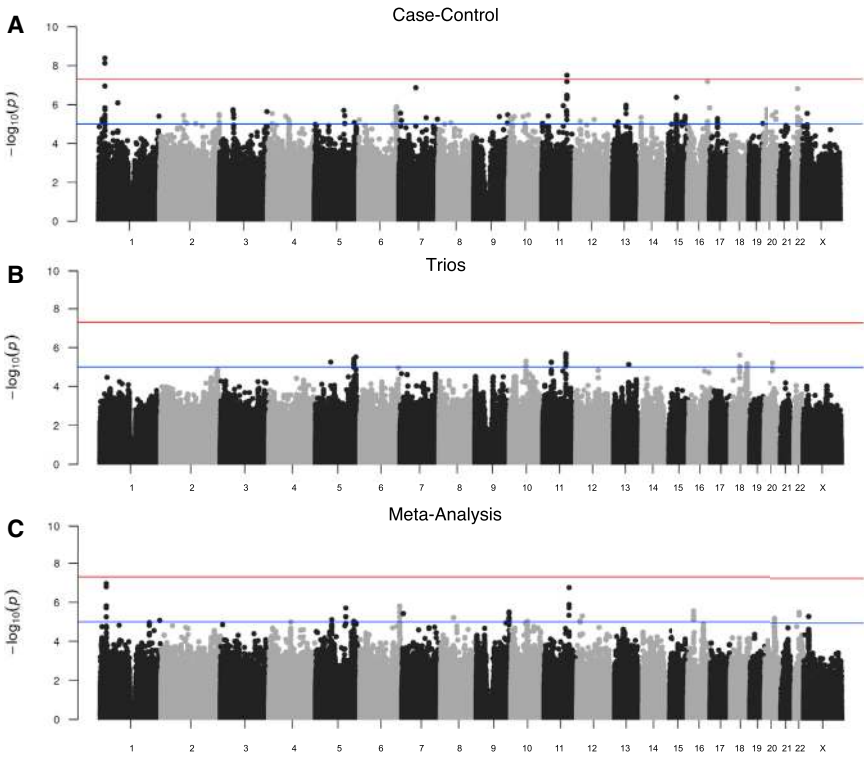

Figure 1. Manhattan Plots of $-\log_{10}$(p Values) from a European CP Cohort
Results are shown for (A) logistic regression of 38 case subjects and 835 control subjects, (B) TDT in 93 trios, and (C) meta-analysis of case-control and TDT results. Colored lines denote suggestive (blue) and genome-wide (red) thresholds for significance. The genomic inflation factor, λ, was 0.934, 1.018, and 0.926 for the case-control, trio, and meta-analysis scans, respectively, indicating negligible inflation of p values.

GENEVA Oral Clefts Consortium.[10] The SNPs we selected for replication were not among the genotyped SNPs included in the published study; two of the seven SNPs (*GRHL3* SNPs rs41268753 and rs113965554) were successfully imputed using the 1000 Genomes reference panel, and both showed evidence of association ($p = 7.6 \times 10^{-3}$ and $p = 6.8 \times 10^{-4}$, respectively; Table 2). Note that the GENEVA Oral Clefts Consortium included some of the same participants as the discovery sample, so we removed overlapping participants and re-ran analyses. After removal of overlapping individuals, evidence of replication remained ($p = 0.016$ and $p = 1.5 \times 10^{-3}$ for rs41268753 and rs113965554, respectively).

rs41268753 Is a Missense Variant in *GRHL3* that Increases Risk for CP

The risk allele for rs41268753 (*GRHL3*) results in an amino acid residue change (GenBank: NM_198173, NP_937816.1; c.1361C>T [p.Thr454Met]) in exon 11 of *GRHL3*. The minor allele frequency of rs41268753 among European controls was 2.75%, consistent with the 3.06% frequency reported in non-Finnish European populations in the ExAC database. However, among our European CP discovery case subjects, the MAF of rs41268753 was 19.74%, resulting in an 8.3-fold increased risk of CP under an additive model (OR = 8.31 [95% CI: 4.1–16.8; genome-wide 99.999995% CI: 1.17–59.15]). In contrast, the OR point estimates were more modest in the replication samples ($OR_{case\text{-}control}$ = 2.16 [1.43–3.27]; OR_{GENEVA} = 2.4 [1.48–5.019]). Although we did not detect an association with rs41268753 in the TDT analysis, the frequency was elevated in the affected probands of the trios in European and European-admixed populations from Central/South American (Table S10).

Sequencing *GRHL3* in Nonsyndromic CP Case Subjects

We previously found that rare, dominant mutations in *GRHL3* cause Van der Woude syndrome (VWS [MIM: 119300, 606713]), the most common syndromic cause of OFC, and were significantly more frequent among individuals with VWS who had CP[27] as opposed to other forms of OFC. In addition, we previously found that families with *GRHL3* mutations were significantly less likely to have lower lip pits, one of the defining features of VWS that occur in 85% of VWS cases. Based on these results, we hypothesized that dominant coding variants could be found in apparently nonsyndromic CP case subjects. We therefore explored *GRHL3* variation in CP samples by Sanger sequencing *GRHL3* exons in an independent set of 84 case subjects of European, Asian, or mixed ancestries. In these 84 case subjects, we detected only two missense variants not found in dbSNP137. The first (GenBank: NM_198173.2, NP_937816.1; c.892C>T [p.Arg298Cys]) was found in an individual with no family history of CP (Figure S4) and was also present in two individuals in the ExAC database. The second missense variant (GenBank: NM_198173.2, NP_937816.1; c.1640A>C [p.Lys547Thr]), was found in an individual in an apparently dominant pedigree with three consecutive generations with CP. However, the variant did not segregate with CP in the family and is unlikely to be pathogenic (Figure S4). In the 54 individuals reporting European ancestry, we found rs41268753 at a frequency of 10.4%, which is consistent with the frequencies of the variant among CP case subjects in our discovery and replication samples. Although it is possible that some individuals with apparently nonsyndromic CP could have coding variants such as those we previously reported in VWS, these results suggest that this is not a common occurrence.

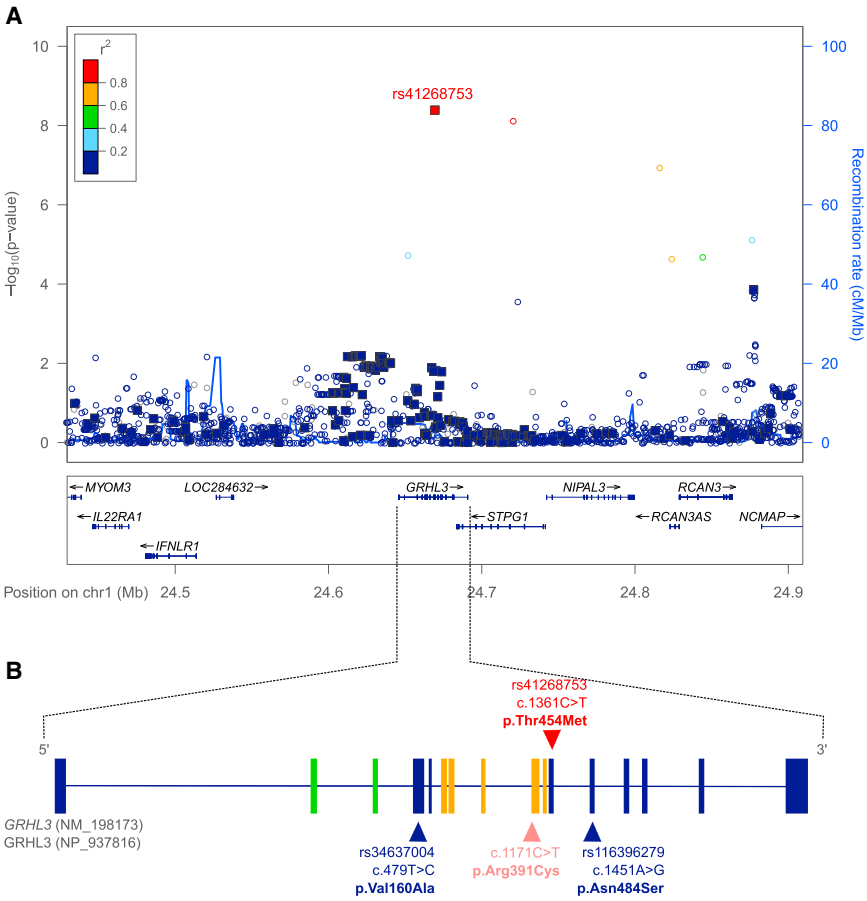

Figure 2. *GRHL3* Is Associated with Non-syndromic CP

(A) Regional association plot[34] showing $-\log_{10}$(p values) for genotyped (filled squares) and imputed (open circles) SNPs for the 1p36 locus near *GRHL3*. The recombination overlay (blue line, right y axis) indicates the boundaries of the LD-block. Points are color coded according to pairwise linkage disequilibrium (r^2) with the index SNP, rs41267835.

(B) Schematic representation of *GRHL3* (GenBank: NM_198173). Exons are colored according to the three known protein domains: the transactivation (green), the DNA binding (orange), and the dimerization (light blue) domains. The position of each of the variants tested in zebrafish assays are indicated and described at the nucleotide and protein level. rs41268753 (p.Thr454Met) (red) is associated with CP in the current study; rs34637004 (p.Val160Ala) and rs116396279 (p.Asn484Ser) (navy) were not associated with CP; and p.Arg391Cys (salmon) is a mutation previously reported to cause Van der Woude syndrome.[27]

Functional Analysis of rs41268753

In our previous work, we tested the activity of GRHL3 protein variants detected in VWS-affected families by injecting synthetic mRNA encoding them into single-cell zebrafish embryos.[27] At late blastula stage (5 hr post fertilization [hpf]), embryos injected with wild-type human *GRHL3* mRNA exhibited ectopic expression of *krt4*, a marker of the most superficial cell layer (the periderm), in deep blastomeres. In contrast, ectopic *krt4* expression was absent in embryos injected with the mRNA containing *GRHL3* mutations detected in VWS-affected families. Further, such embryos exhibited loss of endogenous *krt4* expression in the periderm and ruptured through the animal pole at around 6 hpf, consistent with a failure of periderm differentiation.[27,30] These results signify that in zebrafish embryos, *GRHL3* variants detected in VWS-affected families lack transactivation activity and appear to have the potential to interfere with the activity of endogenous Grhl3.

Here, we utilized the same in vivo zebrafish assay to determine whether p.Thr454Met (rs41268753) alters the function of GRHL3.[27,35] Because common variants associated with disease are expected to have a smaller effect size compared to those mutations causing Mendelian diseases,[36] we anticipated that the common p.Thr454Met variant would have lower-than-wild-type ability to induce ectopic *krt4* but would less efficiently disrupt periderm development than the rare mutations previously identified in VWS-affected families (e.g., p.Arg391Cys [c.1171C>T]). We further anticipated that two other coding polymorphisms that are not associated with CP would have wild-type activity in these assays; we selected two such variants (rs116396279 [p.Asn484Ser] and rs34637004 [p.Val160Ala]) that were not associated with CP in our GWAS ($p > 0.2$).

We engineered full-length *GRHL3* cDNA to encode each of these variants (depicted in Figure 2B), synthesized mRNA in vitro from each construct, and injected it into wild-type zebrafish embryos at the 1- to 2-cell stage. At shield stage (6 hpf), a majority of embryos injected with mRNA encoding beta-galactosidase (as a negative control) (Figures 3A and 3E), wild-type GRHL3 (Figures 3B and 3E), p.Asn484Ser GRHL3 (Figures 3C and 3E), and p.Val160Ala GRHL3 (not shown) all developed normally, except for a slight delay in epiboly, and only a few had ruptured. In contrast, virtually all embryos injected with mRNA encoding the p.Arg391Cys VWS variant had ruptured by 6 hpf (Figure 3E), consistent with our earlier observations.[27] The majority of embryos injected with mRNA encoding p.Thr454Met GRHL3 remained intact at 6 hpf, although they appeared to have stalled at mid-epiboly stage; by 8 hpf, however, virtually all had ruptured (Figures 3D and 3E). We repeated the injections, fixed embryos at 5 hpf, and processed embryos to reveal expression of *krt4* mRNA. As expected, embryos injected with mRNA encoding wild-type GRHL3 (Figure S5B) and p.Val160Ala (Figure S5C) and p.Asn484Ser (Figure S5D) GRHL3 exhibited

Table 2. Results of Genetic Association in the Replication Cohort of 246 Case Subjects and 1,685 Unrelated Control Subjects for 7 SNPs in Loci Observed in the Discovery GWAS

Locus	SNP	Minor Allele	p Value	OR	95% CI
1p36 (*GRHL3*)	rs41268753	T	2.81×10^{-4}	2.16	1.43–3.27
	rs113965554	A	6.82×10^{-4}	1.97	1.33–2.91
6p26 (*PARK2*)	rs12175475	T	0.97	0.99	0.52–1.87
	rs9347594	T	0.97	1.00	0.82–1.21
	rs9365380	T	0.91	1.01	0.83–1.23
11q22.1 (*YAP1*)	rs11225351	T	0.81	1.03	0.83–1.28
	rs117496742	A	0.20	0.60	0.27–1.31

intense ectopic *krt4* expression in deep cells; endogenous *krt4* expression in periderm was present, further confirming that the p.Val160Ala and p.Asn484Ser variants are indistinguishable from wild-type GRHL3 in this assay. A minority of embryos injected with mRNA encoding p.Thr454Met GRHL3 exhibited small patches of low-level ectopic *krt4* expression (not shown), whereas the majority had large patches of superficial cells lacking endogenous *krt4* expression (Figure S5E).

We also examined the ability of p.Thr454Met and p.Arg391Cys, alone or together with wild-type GRHL3, to activate reporter expression in vitro. We engineered four replicates of a consensus GRHL3 binding motif[31] into a luciferase reporter vector and electroporated this construct together with GRHL3 expression plasmids into an oral keratinocyte cell line (GMSM-K). In cells transfected with cDNA encoding wild-type GRHL3, the level of reporter activity was 80-fold higher than in cells transfected with an empty expression plasmid (control-transfected cells) (Figure 3F). In cells transfected with cDNA encoding p.Arg391Cys GRHL3, the level of reporter activity was lower than in control-transfected cells (Figure 3F). By contrast, in cells transfected with cDNA encoding p.Thr454Met GRHL3, the level of reporter activity was about 30-fold higher than in control cells (Figure 3F). In cells cotransfected with equal amounts of cDNA encoding wild-type GRHL3 and either p.Arg391Cys GRHL3 or p.Thr454Met GRHL3 variants, the level of reporter activity was intermediate between the levels in cells transfected with either variant alone (Figure 3F). In summary, whereas the VWS-associated p.Arg391Cys variant lacks transactivation activity, the CP-associated p.Thr454Met variant retains about four-tenths normal activity. Both variants have the ability to interact with GRHL3, presumably by forming dimers with it, lessening its transactivation activity.

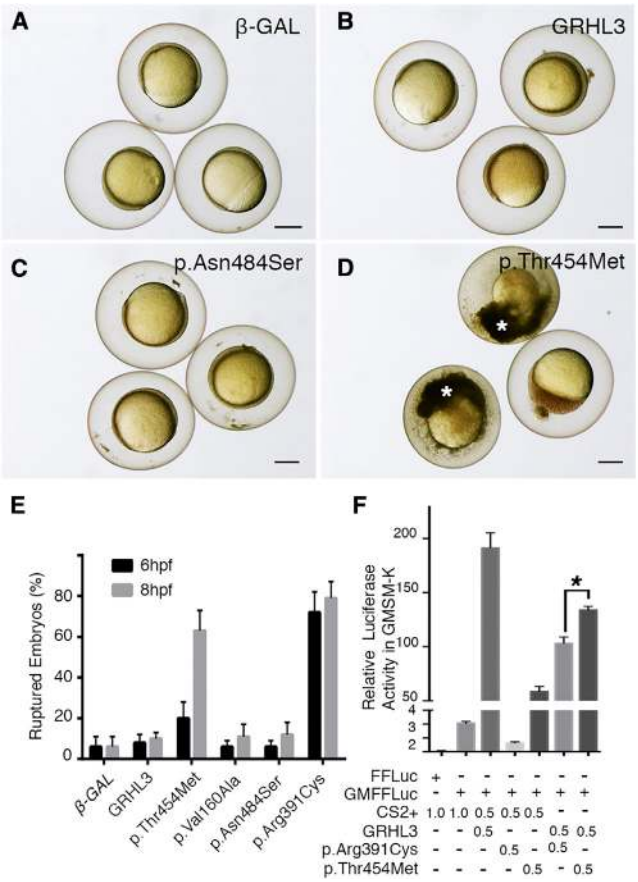

Figure 3. p.Thr454Met, a CP-Associated Variant of *GRHL3*, Disrupts Development of Periderm upon Overexpression in Zebrafish Embryos

(A–D) Lateral views of live embryos derived from a single clutch at about 8 hpf, injected at the 1- to 2-cell stage with mRNAs encoding the indicated proteins. Asterisks (in D) indicate rupturing embryos. Scale bars represent 250 μm.

(E) Bar chart representing percentage of embryos, from 3 experiments of at least 40 embryos each, that ruptured after injection of mRNA encoding the indicated protein. Error bars indicate standard error.

(F) Bar chart representing normalized luciferase levels in GMSM-K oral epithelium cells co-electroporated with the indicated amount (in μg) of the indicated plasmids. Abbreviations are as follows: FFLuc, empty firefly luciferase vector; GMFFLuc, multimerized GRHL3 binding site in firefly luciferase vector; CS2+, empty plasmid used to make overexpression constructs; GRHL3, wild-type; p.Arg391Cys; a GRHL3 variant identified in a family with Van der Woude syndrome;[27] p.Thr454Met, a GRHL3 variant associated with CP in the present study. Luciferase signal is the ratio of firefly and renilla luciferase measurements, normalized to the signal in empty firefly transfection group (FFLuc). The data represent results from three separate experiments, and error bars represent standard error. Asterisk indicates $p < 0.05$.

Discussion

GRHL3 encodes a transcription factor in the grainy-head family with multiple conserved roles in developmental processes including neural tube closure,[37] epidermal barrier formation,[38] wound healing,[39] and craniofacial development.[27,40] In zebrafish, embryos injected with morpholinos targeting *grhl3* exhibit craniofacial hypoplasia,[40] while those injected simultaneously with morpholinos targeting *grhl3* and *grhl1*, or those injected with RNA encoding Grhl1 DNA binding domain, expected to block function of both orthologs, rupture and die during

gastrulation with loss of periderm markers.[35] Analysis of *grhl3* mutants has not been reported. In mouse, loss of *Grhl3* results in oral adhesions and aberrant elevation of the palatal shelves, leading to cleft palate.[27] In both organisms, Grhl3 is required for development of the periderm, Grhl3 acting downstream of Irf6[35] in zebrafish, but no genetic interaction between Irf6 and Grhl3 was observed in mouse.[35] Dominant mutations in *GRHL3* and *IRF6* cause VWS, the most common Mendelian OFC syndrome.[27,41]

VWS is characterized by OFCs and lower lip pits but has considerable variable expressivity (see GeneReviews in Web Resources). Because 12%–15% of individuals with VWS present with isolated OFCs, VWS was hypothesized to serve as a clinical model for nonsyndromic OFCs.[42] Mutations in *IRF6* (MIM: 607199) were discovered to be one of the causes of VWS in 2002[41] and its locus on 1q32 was subsequently implicated in nonsyndromic CL/P in studies using linkage analysis, candidate gene, and GWAS approaches.[9,43] The recent discovery of *GRHL3* mutations as a second cause of VWS established genetically determined subclasses.[27] We reported that whereas VWS-affected individuals carrying *IRF6* mutations are most likely to have CL/P and lip pits, those carrying *GRHL3* mutations are more likely to have CP and less likely to have lip pits,[27] leading to the hypothesis that genetic variation at this locus also confers risk for nonsyndromic CP.

The association of CP with rs41268753 suggested three possible identities of the etiologic variant(s) at this locus: (1) rs41268753 is the etiologic variant; (2) the etiologic variant is an unobserved common variant in linkage disequilibrium with rs41268753; or (3) rs41268753 tags a common haplotype carrying a burden of incompletely penetrant *GRHL3* mutations, such as those found in VWS. In vitro luciferase assays and in vivo zebrafish experiments allowed us to determine that rs41268753 disrupts the normal transcriptional activation function of GRHL3 and is a likely etiologic variant for CP in humans. Furthermore, Sanger sequencing of 84 additional CP case subjects revealed only one novel coding variant, so it is unlikely that rs41268753 is a surrogate for a burden of private mutations. However, we acknowledge that this study included a small number of CP case subjects and we cannot exclude the possibility that some fraction of individuals with nonsyndromic CP will have dominant *GRHL3* mutations. We recently estimated the frequency of such variants in *IRF6* to be ~0.3% based on sequencing of 2,472 CL/P case subjects.[44] Similar studies for CP are required to completely characterize the spectrum of *GRHL3* variants.

By identifying an association between the *GRHL3* locus and nonsyndromic CP, our present study extends the genotype-phenotype correlation observed in VWS to nonsyndromic OFCs. In short, in both syndromic and nonsyndromic OFCs, *IRF6* variants underlie risk for CL/P and *GRHL3* variants do so for CP. Ostensibly, it extends a simplistic paradigm in which damaging mutations with large effects cause Mendelian disorders whereas less damaging coding variants (e.g., hypomorphic or inhibitory) or regulatory polymorphisms confer risk of common forms of the same disorder.[45,46] However, it is clear that the relationship between genotype and phenotype at the *GRHL3* locus is not so simplistic. In our previous work, we demonstrated that dominant GRHL3 mutations cause VWS. However, these mutations are incompletely penetrant; 10% of individuals with mutations were asymptomatic and 46% presented with apparently isolated OFCs. Although these individuals did not have lower lip pits, it remains possible that future generations might present with this characteristic feature of VWS.

Whereas the families with rare coding mutations have dominant inheritance patterns with some nonpenetrance, the nonsyndromic CP-affected families with rs41268753 have sporadic occurrences of CP and many of the individuals with the risk allele are unaffected. This is consistent with the results of the functional assays that indicated both the CP-associated p.Thr454Met variant and the VWS-associated p.Arg391Cys variant have an inhibitory effect on wild-type GRHL3, but that the p.Thr454Met variant appears to be comparatively less disruptive. Moreover, in contrast to the p.Arg391Cys variant, the p.Thr454Met variant retains about 40% of normal transactivation activity. It is noteworthy that although each minor allele of rs41268753 elevates ones' risk of CP, the homozygous condition is nonetheless compatible with life, as shown by the fact that 48 homozygotes for the risk allele are documented in the ExAC database and several in our study population. We also note that OFCs might not have been an exclusion criterion for the studies contributing to ExAC and that 0.04%–0.2%, or 24–120 of the 60,706 individuals in ExAC, might have an OFC. The effects of both rare VWS mutations and rs41268753 are likely to be modified by the presence of other genetic risk factors, environmental exposures, epigenetic marks, or stochastic events, leading to the wide spectrum of OFC affection severity present in study subjects and the general population.

We employed a two-stage study design consisting of case and control subjects and case-parent trios, which were analyzed separately and combined by meta-analysis. However, our significant results were limited to the case-control arm of the study. This was probably due to the small number of trios and the relatively low MAF of associated SNPs. The risk allele at rs41268753, which occurs at a frequency of 3% in populations of European ancestry, is extremely rare or absent in non-European populations and thus unlikely to account for much of their genetic variance in CP. In fact, the frequency of CP across different populations is less variable than for CL/P, so we would predict fewer such population-specific signals for CP than have been observed for CL/P.[26,47,48] Small sample sizes might have also contributed to our failure to replicate a previously reported association with *FAF1*.[8] In multiple cohorts of European ancestry, the TT genotype of rs1149795 (near *FAF1*) was associated with a 1.47-fold increased risk of CP.

However, our present study was insufficiently powered to detect associations with modest effect. Larger sample sizes of multiple ancestral groups will be required to more comprehensively identify the common CP risk alleles.

In summary, we performed a GWAS of nonsyndromic CP and identified and independently replicated an association between rs41268753 and this common congenital malformation. Moreover, we demonstrated that this variant impairs *GRHL3* function and periderm development in zebrafish and is likely to be the etiologic variant for this CP risk locus. This is one of few associations observed via GWASs to occur in the coding region of a gene and one of few examples of functional variants identified to date for nonsyndromic OFCs.[46,49]

Accession Numbers

The accession numbers for the individual genotype and phenotype data analyzed in this paper are dbGaP: phs000774.v1.p1 and phs000094.v1.p1.

Supplemental Data

Supplemental Data include Supplemental Acknowledgments, five figures, and ten tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.02.014.

Acknowledgments

We dedicate this paper to the memory of our colleagues Dr. Andrew Czeizel and Dr. Juan C. Mereb; they will be missed. Many thanks to the participating families who made the studies summarized here possible and who provided on-going inspiration to the investigators. The tireless effort by the dedicated field staff and collaborators was similarly essential to the success of this study. This work was supported by grants from the NIH including X01-HG007845 and R01-DE016148 (to M.L.M.), and genotyping and data cleaning were provided via an NIH contract to the Johns Hopkins Center for Inherited Disease Research (HHSN2682012000008I). The replication studies were funded by NIH grants R01-DE020895 (to G.L.W.) and U01-DE018993 to Dr. Terri Beaty at Johns Hopkins University School of Public Health. See the Supplemental Acknowledgments for additional grant support and full acknowledgments.

Received: December 2, 2015
Accepted: February 17, 2016
Published: March 24, 2016

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://browser.1000genomes.org
dbGaP, http://www.ncbi.nlm.nih.gov/gap
ExAC Browser, http://exac.broadinstitute.org/
Fundación Clínica Noel, http://www.clinicanoel.org.co/
GenBank, http://www.ncbi.nlm.nih.gov/genbank/
GeneReviews, Schutte et al. (1993). IRF6-Related Disorders, http://www.ncbi.nlm.nih.gov/books/NBK1407/
Human Genetic Analysis Interface, http://facebase.sdmgenetics.pitt.edu/
OMIM, http://www.omim.org/
QC Report for Genotypic Data, http://www.ccdg.pitt.edu/docs/Marazita_ofc_QC_report_feb2015.pdf

References

1. Tessier, P. (1976). Anatomical classification facial, cranio-facial and latero-facial clefts. J. Maxillofac. Surg. *4*, 69–92.
2. Mossey, P.A., Little, J., Munger, R.G., Dixon, M.J., and Shaw, W.C. (2009). Cleft lip and palate. Lancet *374*, 1773–1785.
3. Dixon, M.J., Marazita, M.L., Beaty, T.H., and Murray, J.C. (2011). Cleft lip and palate: understanding genetic and environmental influences. Nat. Rev. Genet. *12*, 167–178.
4. Fogh-Andersen, P. (1942). Inheritance of Harelip and Cleft Palate (Copenhagen: Munksgaard).
5. Fraser, F.C. (1955). Thoughts on the etiology of clefts of the palate and lip. Acta Genet. Stat. Med. *5*, 358–369.
6. Jones, M.C. (1988). Etiology of facial clefts: prospective evaluation of 428 patients. Cleft Palate J. *25*, 16–20.
7. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. *43*, D789–D798.
8. Ghassibe-Sabbagh, M., Desmyter, L., Langenberg, T., Claes, F., Boute, O., Bayet, B., Pellerin, P., Hermans, K., Backx, L., Mansilla, M.A., et al. (2011). FAF1, a gene that is disrupted in cleft palate and has conserved function in zebrafish. Am. J. Hum. Genet. *88*, 150–161.
9. Leslie, E.J., and Marazita, M.L. (2013). Genetics of cleft lip and cleft palate. Am. J. Med. Genet. C. Semin. Med. Genet. *163C*, 246–258.
10. Beaty, T.H., Ruczinski, I., Murray, J.C., Marazita, M.L., Munger, R.G., Hetmanski, J.B., Murray, T., Redett, R.J., Fallin, M.D., Liang, K.Y., et al. (2011). Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. Genet. Epidemiol. *35*, 469–478.
11. Mangold, E., Ludwig, K.U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., Reutter, H., de Assis, N.A., Chawa, T.A., Mattheisen, M., et al. (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat. Genet. *42*, 24–26.
12. Reiter, R., Brosch, S., Goebel, I., Ludwig, K.U., Pickhard, A., Högel, J., Schlömer, G., Mangold, E., Kubisch, C., and Borck, G. (2015). A post GWAS association study of SNPs associated with cleft lip with or without cleft palate in submucous cleft palate. Am. J. Med. Genet. A. *167A*, 670–673.
13. Li, L., Meng, T., Jia, Z., Zhu, G., and Shi, B. (2010). Single nucleotide polymorphism associated with nonsyndromic cleft palate influences the processing of miR-140. Am. J. Med. Genet. A. *152A*, 856–862.
14. Ludwig, K.U., Böhmer, A.C., Rubini, M., Mossey, P.A., Herms, S., Nowak, S., Reutter, H., Alblas, M.A., Lippke, B., Barth, S., et al. (2014). Strong association of variants around FOXE1 and orofacial clefting. J. Dent. Res. *93*, 376–381.
15. Laurie, C.C., Doheny, K.F., Mirel, D.B., Pugh, E.W., Bierut, L.J., Bhangale, T., Boehm, F., Caporaso, N.E., Cornelis, M.C., Edenberg, H.J., et al.; GENEVA Investigators (2010). Quality control and quality assurance in genotypic data for genome-wide association studies. Genet. Epidemiol. *34*, 591–602.

16. Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods *10*, 5–6.
17. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
18. Spielman, R.S., and Ewens, W.J. (1996). The TDT and other family-based tests for linkage disequilibrium and association. Am. J. Hum. Genet. *59*, 983–989.
19. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
20. Kazeem, G.R., and Farrall, M. (2005). Integrating case-control and TDT studies. Ann. Hum. Genet. *69*, 329–335.
21. Olsen, J., Melbye, M., Olsen, S.F., Sørensen, T.I., Aaby, P., Andersen, A.M., Taxbøl, D., Hansen, K.D., Juhl, M., Schow, T.B., et al. (2001). The Danish National Birth Cohort–its background, structure and aim. Scand. J. Public Health *29*, 300–307.
22. Munger, R.G., Romitti, P.A., Daack-Hirsch, S., Burns, T.L., Murray, J.C., and Hanson, J. (1996). Maternal alcohol use and risk of orofacial cleft birth defects. Teratology *54*, 27–33.
23. Wilcox, A.J., Lie, R.T., Solvoll, K., Taylor, J., McConnaughey, D.R., Abyholm, F., Vindenes, H., Vollset, S.E., and Drevon, C.A. (2007). Folic acid supplements and risk of facial clefts: national population based case-control study. BMJ *334*, 464.
24. Magnus, P., Irgens, L.M., Haug, K., Nystad, W., Skjaerven, R., and Stoltenberg, C.; MoBa Study Group (2006). Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). Int. J. Epidemiol. *35*, 1146–1150.
25. Munger, R.G., Tamura, T., Johnston, K.E., Feldkamp, M.L., Pfister, R., Cutler, R., Murtaugh, M.A., and Carey, J.C. (2011). Oral clefts and maternal biomarkers of folate-dependent one-carbon metabolism in Utah. Birth Defects Res. A Clin. Mol. Teratol. *91*, 153–161.
26. Beaty, T.H., Murray, J.C., Marazita, M.L., Munger, R.G., Ruczinski, I., Hetmanski, J.B., Liang, K.Y., Wu, T., Murray, T., Fallin, M.D., et al. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat. Genet. *42*, 525–529.
27. Peyrard-Janvid, M., Leslie, E.J., Kousa, Y.A., Smith, T.L., Dunnwald, M., Magnusson, M., Lentz, B.A., Unneberg, P., Fransson, I., Koillinen, H.K., et al. (2014). Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. Am. J. Hum. Genet. *94*, 23–32.
28. Heckman, K.L., and Pease, L.R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. Nat. Protoc. *2*, 924–932.
29. Westerfield, M. (1993). The Zebrafish Book (Eugene, OR: University of Oregon Press).
30. Sabel, J.L., d'Alençon, C., O'Brien, E.K., Van Otterloo, E., Lutz, K., Cuykendall, T.N., Schutte, B.C., Houston, D.W., and Cornell, R.A. (2009). Maternal Interferon Regulatory Factor 6 is required for the differentiation of primary superficial epithelia in *Danio* and *Xenopus* embryos. Dev. Biol. *325*, 249–262.
31. Boglev, Y., Wilanowski, T., Caddy, J., Parekh, V., Auden, A., Darido, C., Hislop, N.R., Cangkrama, M., Ting, S.B., and Jane, S.M. (2011). The unique and cooperative roles of the Grainy head-like transcription factors in epidermal development reflect unexpected target gene specificity. Dev. Biol. *349*, 512–522.
32. Fisher, S., Grice, E.A., Vinton, R.M., Bessling, S.L., and McCallion, A.S. (2006). Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science *312*, 276–279.
33. Gilchrist, E.P., Moyer, M.P., Shillitoe, E.J., Clare, N., and Murrah, V.A. (2000). Establishment of a human polyclonal oral epithelial cell line. Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod. *90*, 340–347.
34. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.
35. de la Garza, G., Schleiffarth, J.R., Dunnwald, M., Mankad, A., Weirather, J.L., Bonde, G., Butcher, S., Mansour, T.A., Kousa, Y.A., Fukazawa, C.F., et al. (2013). Interferon regulatory factor 6 promotes differentiation of the periderm by activating expression of Grainyhead-like 3. J. Invest. Dermatol. *133*, 68–77.
36. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.
37. Ting, S.B., Wilanowski, T., Auden, A., Hall, M., Voss, A.K., Thomas, T., Parekh, V., Cunningham, J.M., and Jane, S.M. (2003). Inositol- and folate-resistant neural tube defects in mice lacking the epithelial-specific factor Grhl-3. Nat. Med. *9*, 1513–1519.
38. Ting, S.B., Caddy, J., Hislop, N., Wilanowski, T., Auden, A., Zhao, L.L., Ellis, S., Kaur, P., Uchida, Y., Holleran, W.M., et al. (2005). A homolog of *Drosophila* grainy head is essential for epidermal integrity in mice. Science *308*, 411–413.
39. Caddy, J., Wilanowski, T., Darido, C., Dworkin, S., Ting, S.B., Zhao, Q., Rank, G., Auden, A., Srivastava, S., Papenfuss, T.A., et al. (2010). Epidermal wound repair is regulated by the planar cell polarity signaling pathway. Dev. Cell *19*, 138–147.
40. Dworkin, S., Simkin, J., Darido, C., Partridge, D.D., Georgy, S.R., Caddy, J., Wilanowski, T., Lieschke, G.J., Doggett, K., Heath, J.K., and Jane, S.M. (2014). Grainyhead-like 3 regulation of endothelin-1 in the pharyngeal endoderm is critical for growth and development of the craniofacial skeleton. Mech. Dev. *133*, 77–90.
41. 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.
42. Murray, J.C., Nishimura, D.Y., Buetow, K.H., Ardinger, H.H., Spence, M.A., Sparkes, R.S., Falk, R.E., Falk, P.M., Gardner, R.J., Harkness, E.M., et al. (1990). Linkage of an autosomal dominant clefting syndrome (Van der Woude) to loci on chromosome Iq. Am. J. Hum. Genet. *46*, 486–491.
43. 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.
44. Leslie, E.J., Koboldt, D.C., Kang, C.J., Ma, L., Hecht, J.T., Wehby, G.L., Christensen, K., Czeizel, A.E., Deleyiannis, F.W., Fulton, R.S., et al. (2015). IRF6 mutation screening in non-syndromic orofacial clefting: analysis of 1521 families.

Clin. Genet. Published online October 1, 2015. http://dx.doi.org/10.1111/cge.12675.
45. Haller, G., Alvarado, D., McCall, K., Yang, P., Cruchaga, C., Harms, M., Goate, A., Willing, M., Morcuende, J.A., Baschal, E., et al. (2016). A polygenic burden of rare variants across extracellular matrix genes among individuals with adolescent idiopathic scoliosis. Hum. Mol. Genet. *25*, 202–209.
46. 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.
47. Ludwig, K.U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A., Becker, J., Herberz, R., AlChawa, T., Nasser, E., et al. (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat. Genet. *44*, 968–971.
48. Sun, Y., Huang, Y., Yin, A., Pan, Y., Wang, Y., Wang, C., Du, Y., Wang, M., Lan, F., Hu, Z., et al. (2015). Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat. Commun. *6*, 6414.
49. Leslie, E.J., Taub, M.A., Liu, H., Steinberg, K.M., Koboldt, D.C., Zhang, Q., Carlson, J.C., Hetmanski, J.B., Wang, H., Larson, D.E., et al. (2015). Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. Am. J. Hum. Genet. *96*, 397–411.