Human Molecular Genetics, 2019, Vol. 28, No. 6 1038–1051

doi: 10.1093/hmg/ddy402
Advance Access Publication Date: 19 November 2018
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Genomic analyses in African populations identify novel risk loci for cleft palate

Azeez Butali[1,*], Peter A. Mossey[2], Wasiu L. Adeyemo[3], Mekonen A. Eshete[4], Lord J.J. Gowans[5], Tamara D. Busch[1], Deepti Jain[6], Wenjie Yu[7], Liu Huan[8], Cecelia A. Laurie[6], Cathy C. Laurie[6], Sarah Nelson[6,†], Mary Li[1], Pedro A. Sanchez-Lara[9], William P. Magee III[10], Kathleen S. Magee[11], Allyn Auslander[10], Frederick Brindopke[10], Denise M. Kay[12], Michele Caggana[12], Paul A. Romitti[13], James L. Mills[14], Rosemary Audu[15], Chika Onwuamah[15], Ganiyu O. Oseni[16], Arwa Owais[17], Olutayo James[3], Peter B. Olaitan[16], Babatunde S. Aregbesola[18], Ramat O. Braimah[19], Fadekemi O. Oginni[18], Ayodeji O. Oladele[18], Saidu A. Bello[20], Jennifer Rhodes[21], Rita Shiang[21], Peter Donkor[5], Solomon Obiri-Yeboah[5], Fareed Kow Nanse Arthur[5], Peter Twumasi[5], Pius Agbenorku[5], Gyikua Plange-Rhule[5], Alexander Acheampong Oti[5], Olugbenga M. Ogunlewe[3], Afisu A. Oladega[3], Adegbayi A. Adekunle[3], Akinwunmi O. Erinoso[3], Olatunbosun O. Adamson[3], Abosede A. Elufowoju[3], Oluwanifemi I. Ayelomi[3], Taiye Hailu[4], Abiye Hailu[4], Yohannes Demissie[4], Miliard Derebew[4], Steve Eliason[7], Miguel Romero-Bustillous[7], Cynthia Lo[1], James Park[1], Shaan Desai[1], Muiawa Mohammed[1], Firke Abate[4], Lukman O. Abdur-Rahman[22], Deepti Anand[23], Irfaan Saadi[24,‡], Abimibola V. Oladugba[25], Salil A. Lachke[23], Brad A. Amendt[7], Charles N. Rotimi[26], Mary L. Marazita[27], Robert A. Cornell[7], Jeffrey C. Murray[28] and Adebowale A. Adeyemo[26,*]

[1]Department of Oral Pathology, Radiology and Medicine, College of Dentistry, University of Iowa, Iowa, IA 52242, USA, [2]Department of Orthodontics, University of Dundee, Dundee DD1 4HN, UK, [3]Department of Oral and Maxillofacial Surgery, University of Lagos, Akoka, Lagos 100213, Nigeria, [4]School of Public Health, Addis Ababa University, Addis Ababa 1000, Ethiopia, [5]Kwame Nkrumah University of Science and Technology, Kumasi, Ashanti, Ghana, [6]Department of Biostatistics, Genetic Analysis Center, University of Washington,

[†]Sarah Nelson, http://orcid.org/0000-0002-2109-6465
[‡]Irfaan Saadi, http://orcid.org/0000-0002-6250-6651
Received: June 22, 2018. **Revised:** November 7, 2018. **Accepted:** November 11, 2018

Seattle, WA 98195, USA, [7]Department of Anatomy and Cell Biology, University of Iowa, Iowa, IA 52242, USA, [8]State Key Laboratory Breeding Base of Basic Science of Stomatology (Hubei-MOST, Ministry of Science and Technology) and Key Laboratory for Oral Biomedicine of Ministry of Education, School and Hospital of Stomatology, Wuhan University, Wuhan 430072, China, [9]Department of Pediatrics, Cedars-Sinai Medical Center, David Geffen School of Medicine at University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA, [10]Division of Plastic and Maxillofacial Surgery, Children's Hospital Los Angeles, Los Angeles, CA 90027, USA, [11]Operation Smile, 3641 Faculty Boulevard, Virginia Beach, VA 23453, USA, [12]Wadsworth Center, New York State Department of Health, Albany, NY 12201, USA, [13]Department of Epidemiology, College of Public Health, University of Iowa, Iowa, IA 52242, USA, [14]Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health (NIH), 12 South Drive, MSC 5635, Bethesda, MD 20892-5635, USA, [15]Department of Virology, Nigerian Institute of Medical Research, Yaba, Lagos 100001, Nigeria, [16]Department of Plastic Surgery, Ladoke Akintola University of Science and Technology, Osogbo, Oyo 210214, Nigeria, [17]Department of Pediatric Dentistry, University of Iowa, Iowa, IA 52242, USA, [18]Department of Oral and Maxillofacial Surgery, Obafemi Awolowo University, Ile-Ife, Osun A234, Nigeria, [19]Department of Oral and Maxillofacial Surgery, Usmanu Danfodiyo University Teaching Hospital, Sokoto 814212, Nigeria, [20]State House Clinic, Asokoro, Abuja 900231, Nigeria, [21]Virginia Commonwealth University School of Medicine, Richmond, VA 23298, USA, [22]Division of Pediatric Surgery, Department of Surgery, University of Ilorin, Ilorin, Kwara 240103, Nigeria, [23]Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE 19716, USA, [24]Department of Anatomy and Cell Biology, University of Kansas Medical Center, Kansas, KS 66160, USA, [25]Department of Biostatistics, University of Nigeria, Nssuka, Enugu 410001, Nigeria, [26]Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, MD 20892, USA, [27]Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine; Department of Human Genetics, Graduate School of Public Health, and Clinical and Translational Sciences, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA and [28]Department of Pediatrics, University of Iowa, Iowa, IA 52242, USA

*To whom correspondence should be addressed at: Azeez Butali, Department of Oral Pathology, Radiology and Medicine, College of Dentistry, University of Iowa, Iowa, IA 52242, USA. Tel:+319 3358980; Fax: 319-384-1169; Email: Azeez-butali@uiowa.edu; or Adebowale A. Adeyemo, Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA. Tel: (301) 594-7501; Fax: (301) 451-5426; Email: adeyemoa@mail.nih.gov

Abstract

Orofacial clefts are common developmental disorders that pose significant clinical, economical and psychological problems. We conducted genome-wide association analyses for cleft palate only (CPO) and cleft lip with or without palate (CL/P) with ~17 million markers in sub-Saharan Africans. After replication and combined analyses, we identified novel loci for CPO at or near genome-wide significance on chromosomes 2 (near *CTNNA2*) and 19 (near *SULT2A1*). *In situ* hybridization of *Sult2a1* in mice showed expression of *SULT2A1* in mesenchymal cells in palate, palatal rugae and palatal epithelium in the fused palate. The previously reported 8q24 was the most significant locus for CL/P in our study, and we replicated several previously reported loci including *PAX7* and *VAX1*.

Introduction

Orofacial clefts (OFCs) are the most common birth defects in the head and neck region, affecting 1 out of every 700 live births worldwide (1). These defects lead to significant financial, educational, medical, psychological and cultural problems for affected individuals and their families. Management of these disorders requires a multidisciplinary team of experts to restore aesthetics and function. Such expertise is often lacking in many parts of the world resulting in significant inequities in OFC care (2,3). A total of 70% of the OFCs are classified as non-syndromic with no visible recognizable structural defects other than clefts. Syndromic clefts account for 30% of the OFCs, where there is a consistently defined structural anomaly in addition to clefts. In terms of etiology, OFCs are complex traits, with genetic, environmental and stochastic factors contributing to the phenotypic expression (4). To date, 6 genome-wide association studies (GWASs) and 3 meta-analysis for cleft lip with or without cleft palate (CL/P) and 3 GWASs for cleft palate only (CPO) have been conducted, and over 40 risk loci have been identified (5–16). All of these studies have been conducted in individuals of European and Asian ancestry, with this study representing the first GWAS in Africans.

African populations represent novel and richly productive populations for genetic and environmental exposure studies for OFC because they have the greatest genetic diversity of any

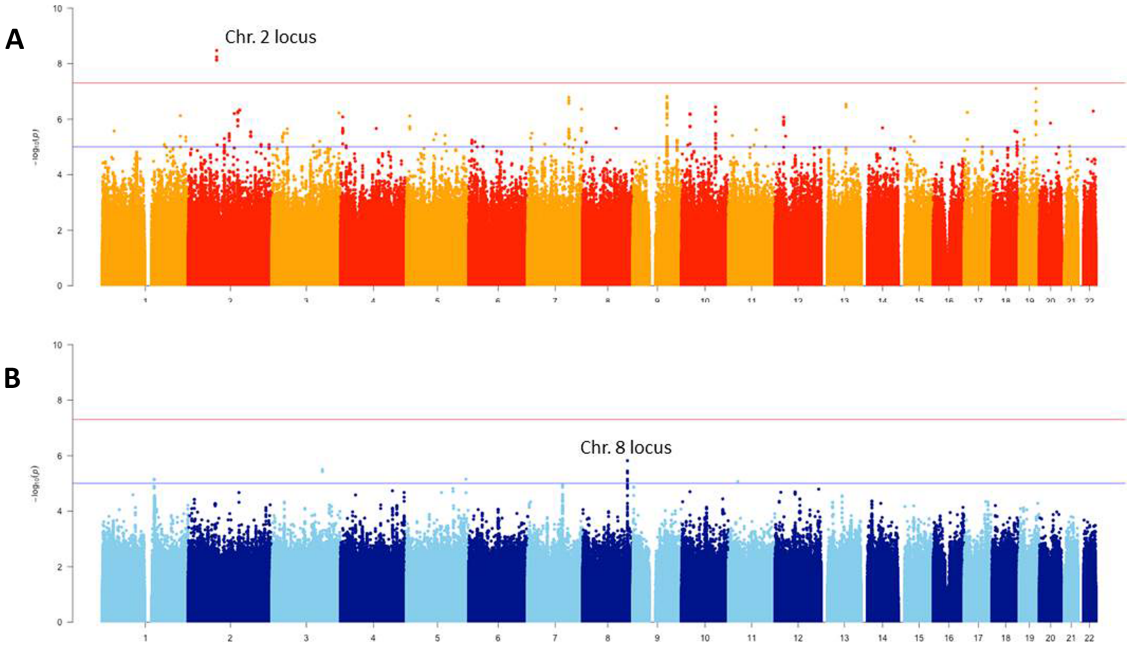

Figure 1. Manhattan plots of association statistics for CPO (**A**) and CL/P (**B**) in sub-Saharan Africa.

continental population (17,18) while residing in widely different environments. In this study involving individuals of African ancestry from Ghana, Nigeria and Ethiopia, we identified novel loci associated with CPO using data from 3178 participants (814 CL/P cases, 205 CPO cases, 2159 controls). Two of the identified novel loci were genome-wide significant after combined analysis with an independent replication sample. We also confirmed previously reported loci from GWAS of OFC in other populations, including populations of European and Asian ancestry.

Results

Novel loci identified for CPO

The discovery analysis for CPO revealed a chromosome 2 locus with genome-wide significance (lead single nucleotide polymorphism (SNP) rs80004662, near *CTNNA2*; $P = 7.41 \times 10^{-9}$; Fig. 1). Other loci on chromosomes 7, 9 and 19 showed suggestive genome-wide significance ($5 \times 10^{-7} > P > 5 \times 10^{-8}$) on discovery analysis (Table 1; Supplementary Material, Table S1). On meta-analysis with an independent replication sample, the chromosome 2 locus remained genome-wide significant ($P = 7.29 \times 10^{-9}$; Table 2; Supplementary Material, Table S2). Genes within the same topologically associated domains (TADs), as the GWAS SNP, are potential GWAS candidates. The TAD that includes the genome-wide significant SNPs contains just three genes—*CTNNA2*, *LRRTM1* and *SUCLG1* (Fig. 2). Among these genes, *CTNNA2* is the best candidate as the chick ortholog has been implicated in control of cranial neural crest (19). *Ctnna2* has been reported to be expressed in the oral structures of the mouse embryo at E14.5 (Fig. 2).

SULT2A1 is expressed in the palate at E12.5 and E14.5

The chromosome 19 locus was near genome-wide significance (lead SNP rs62529857, *SULT2A1*; $P = 7.63 \times 10^{-8}$). We studied the expression of the ortholog of the chromosome 19 locus for CPO (*Sult2a1*) in mice. *In situ* hybridization of *Sult2a1* in mice showed expression of *SULT2A1* in mesenchymal cells in palate, palatal rugae and palatal epithelium in the fused palate (Fig. 3). We also observed expression in the tongue, mandible, maxilla and heart. SysFACE analysis also showed that SULT2A1 is expressed at low levels in the neural plate, mandible and maxilla (Supplementary Material, Table S3). The expression of *SULT2A1* in the palate and other craniofacial tissues provides a biological rationale for its role in orofacial clefting.

The 8q24 region is the most significant locus for CL/P in African populations

While the analysis for CL/P showed no genome-wide significant loci (Fig. 1; Supplementary Material, Tables S4 and S5), the most significant hit was on chromosome 8 (lead SNP, rs72728755; $P = 1.52 \times 10^{-6}$). This locus is in the 8q24 region that has been previously reported to be associated with CL/P in other populations (5–9,11). The lead SNP in our study is also one of the top-scoring SNPs in the 8q24 region in the largest meta-analysis of OFC to date (14).

Fine-mapping of the 8q24 locus for CL/P

We fine-mapped the 8q24 locus for CL/P using a number of methods. We examined haplotypes around the lead SNPs in our African sample and did a comparison with European and Asian ancestry samples from the 1000 Genomes (1KG) Project. As expected, the African sample had smaller haplotypes and finer-grained linkage disequilibrium (LD) patterns in the region (Fig. 4). Specifically, the haplotype around the lead SNP (rs72728577) is 4.084 kb in the continental African sample in contrast to 13.345 kb in European (1KG EUR), 13.477 kb in East Asian (1KG EAS) and 12.104 kb in South Asian (1KG SAS) populations (Fig. 4). Clumping analysis revealed a single clump of SNPs around the lead SNP (data not shown). Fine-mapping using

Table 1. Top hits for discovery association analysis for CPO and cleft lip/palate (CL/P)

SNP	Chr	BP (base pairs)	Effect allele	Non-effect allele	Effect allele frequency	OR (odds ratio)	95% CI (confidence intervals) (OR)	*P*-value
CPO								
rs80004662	2	82025185	A	G	0.013	7.5	3.45–16.28	7.41E-09
rs113691307	2	82028390	C	T	0.013	7.5	3.45–16.28	7.41E-09
rs62529857	19	48386473	T	C	0.023	3.5	2.16–5.68	7.84E-08
rs117381175	9	98403220	C	T	0.012	7.45	3.16–17.55	1.52E-07
rs143238378	7	119266270	G	A	0.015	4.26	2.35–7.71	1.64E-07
rs188681640	7	119146159	A	G	0.011	4.82	2.52–9.24	2.15E-07
rs150382487	7	119140602	T	A	0.011	4.81	2.52–9.21	2.16E-07
rs189675673	19	48383400	G	A	0.02	3.51	2.12–5.82	2.38E-07
rs3858092	9	98291448	A	C	0.396	1.72	1.39–2.11	2.62E-07
rs182830500	7	119161353	T	C	0.01	4.94	2.54–9.63	2.71E-07
CL/P								
rs72728755	8	129990382	T	A	0.097	1.62	1.33–1.97	1.52E-06
rs1474306	3	145361479	T	C	0.942	0.57	0.45–0.72	3.16E-06
rs6768171	3	145361918	T	G	0.942	0.57	0.45–0.73	3.76E-06
rs55658222	8	129976136	G	A	0.098	1.58	1.30–1.92	4.20E-06
rs151084002	5	172805743	C	A	0.048	1.81	1.39–2.35	7.02E-06
rs112640811	1	150097784	G	A	0.21	1.4	1.21–1.62	7.06E-06
rs13274247	8	129981468	G	A	0.42	1.32	1.17–1.50	7.06E-06
rs12090508	1	150107793	A	G	0.211	1.4	1.21–1.62	7.33E-06
rs7517537	1	150114083	C	T	0.211	1.4	1.21–1.62	7.47E-06
rs744835	8	129982547	C	T	0.478	1.32	1.17–1.48	8.36E-06

BP is base pairs; OR is odds ratio; CI is confidence intervals.

Table 2. Variants near or at genome-wide significance on combined analysis for CPO and with consistency of direction of effect

				Discovery sample		Replication sample		Combined analysis		
SNP	Gene	Chr	BP	Score	*P*-value	Z score	*P*-value	Z score	*P*-value	Direction
rs80004662	*CTNNA2*	2	82025185	9.383	7.41E-09	0.199	0.842	5.784	7.29E-09	++
rs113691307	*CTNNA2*	2	82028390	9.384	7.41E-09	−0.147	0.883	−5.783	7.33E-09	−−
rs62529857	*SULT2A1*	19	48386473	15.421	7.84E-08	0.289	0.773	5.376	7.63E-08	++
rs2325377	*DACH1*	13	71895298	15.524	3.62E-07	0.904	0.366	5.105	3.31E-07	++

a shotgun stochastic search algorithm (20) showed that the most likely configuration is a single causal variant in the region (Supplementary Material, Fig. S1).

Given that the lead SNP in the 8q24 region in our study (rs72728755) is different from the lead SNP (rs987525) reported by the most previous GWAS, we investigated this region further. SNP rs987525 is in low LD with rs72728755 (r^2 = 0.004) in our study. Reciprocal conditional analysis revealed that conditioning on rs987525 had a small effect on rs72728755 (*P*-value decreased to 1.451×10^{-5} from 1.52×10^{-6}), but conditioning in the other direction abolished the nominal significance of rs987525 (*P*-value decreased to 0.231 from 3.296×10^{-2}), suggesting that rs72728755 is driving the association in our study. We note that this finding does not exclude the possibility of more than one causal variant in the 8q24 region, given that the two SNPs are in different haplotype blocks in all 1KG Project continental ancestry populations (Supplementary Material, Fig. S2).

Characterization of chr8q.24 SNPs for enhancer elements that are active in palate formation

The 8q24 SNPs that are most strongly associated with CL/P may themselves be directly pathological (i.e. functional), or instead they may be in LD with those that are functional. We selected the lead SNP in the region (rs72728755) and two SNPs that are most strongly associated with CL/P and are in strong LD with the lead SNPs (rs17242358 and rs55658222) for further studies. To test whether these non-coding SNPs are functional by virtue of altering the function of a regulatory element, we examined the chromatin state model at each SNP based on chromatin-marked evidence from 128 cell lines from the Roadmap Epigenomics Consortium. None of the SNPs lie in chromatin-marked regions as any type of regulatory element (Fig. 5). We amplified ∼1 kb of DNA centered on each SNP, engineered the elements with either the non-risk- or risk-associated allele of the SNP (introduced by site-directed mutagenesis) into a standard firefly luciferase reporter vector and electroporated the reporters (separately) into a human fetal oral epithelial cell line (GMSM-K) (21) or primary human embryonic palate mesenchymal (HEPM) cell line (22). In both cell lines, none of the elements, whether harboring the risk or non-risk SNP variant, induced luciferase expression more than 2-fold above that in control cells electroporated with an empty firefly luciferase vector (Fig. 5). In summary, we did not find evidence that rs72728755, rs17242358 or rs55658222 reside within enhancers active in two cell types that are relevant to palate formation. It is still possible they reside

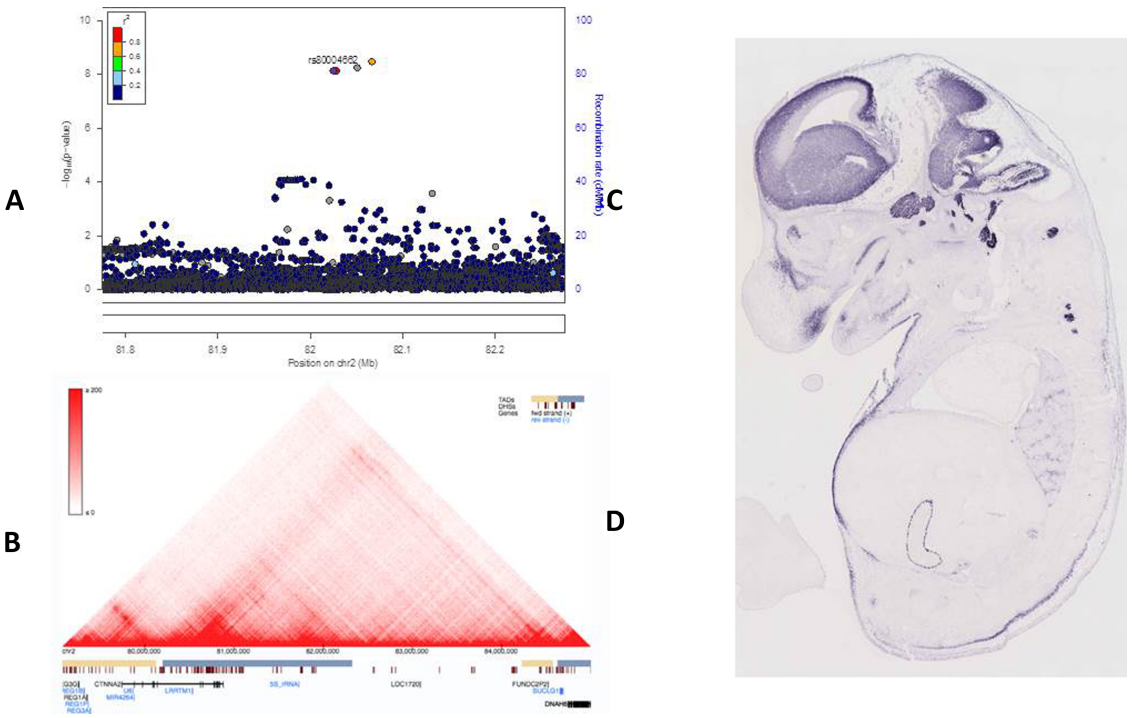

Figure 2. (**A**) Regional association plot in the chromosome 2 locus for CPO. (**B**) TAD around the chromosome 2 locus for CPO. (**C**) Ctnna2 expression in mouse embryo at 14.5 days post fertilization (dpf) (Eurexpress–A Transcriptome Atlas of the Mouse Embryo, http://www.eurexpress.org/).

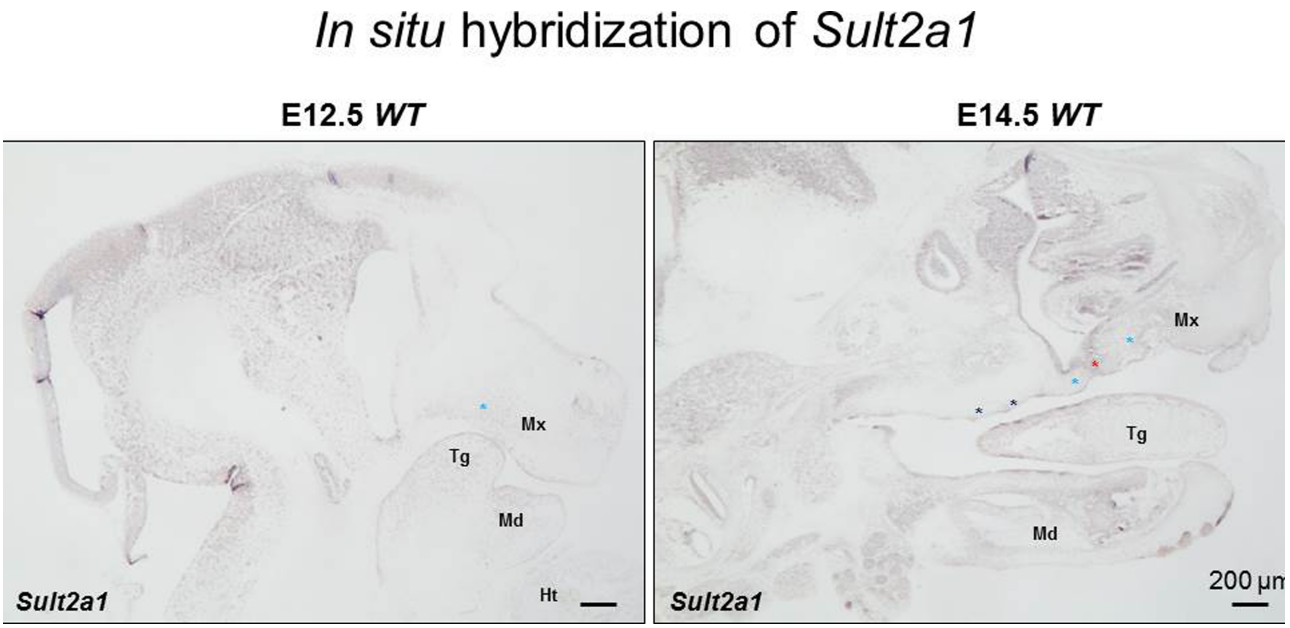

Figure 3. *In situ* hybridization of *Sult2a1* in E12.5 and E14.5 embryos. Blue asterisks show mesenchymal cells in palate; black asterisks, palatal rugae with Sult2a1 expression; red asterisk, palatal epithelium. Tg, tongue; Md, mandible; Mx, maxilla; Ht, heart. Scale bar, 200 µm.

in enhancers active in a cell type that is not represented by the cell lines we tested or by those at the Roadmap Epigenomics Consortium (http://www.roadmapepigenomics.org/). Other possibilities are that one or more of the SNPs alter the sequence, and, thereby, the functions of an unknown long non-coding RNA or the SNPs are in LD with the actual untyped functional SNPs.

Novel variants identified in known GWAS-associated genes for CL/P

We identified two novel variants (p.Gly739Ser in *DACH1* and p.Leu187Pro in *ACVR2A*) following Sanger sequencing (Table 3). These variants have not been previously reported in any genomic databases, including the gnomAD, Exome

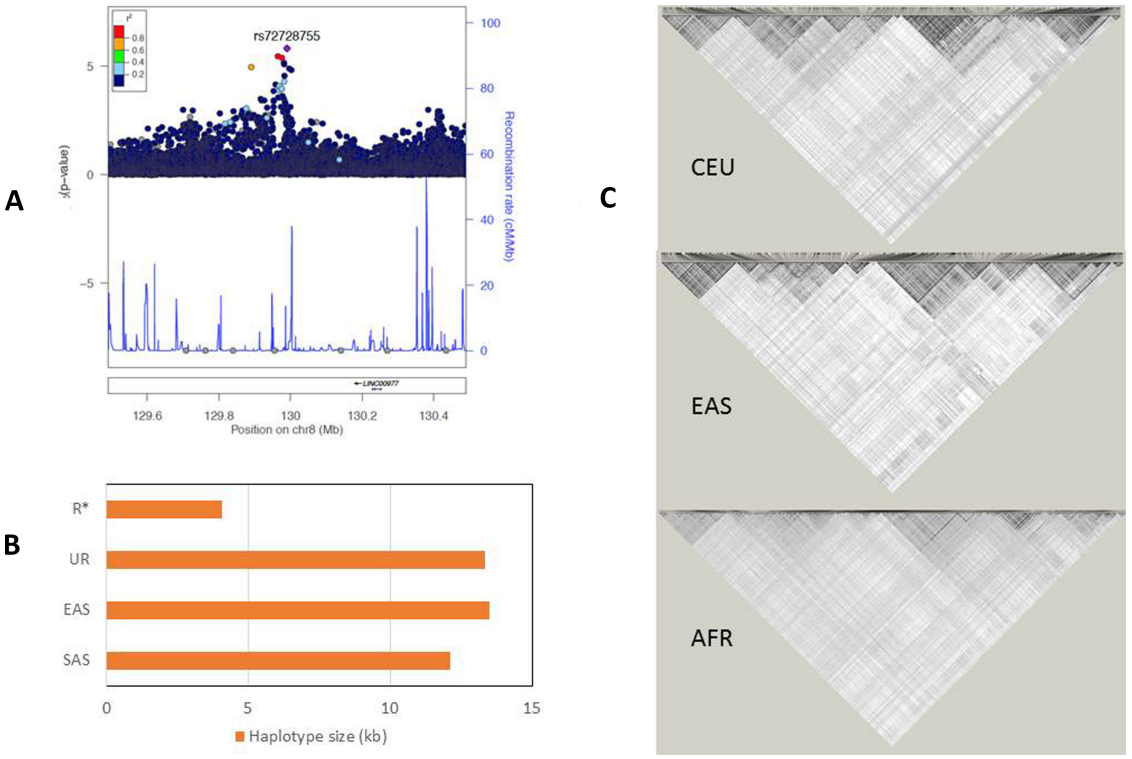

Figure 4. (**A**) Regional association plot in the chromosome 8q24 locus for CL/P. (**B**) Haplotype block sizes around the 8q24 lead SNP rs72728755 for CL/P. (**C**) LD patterns around the 8q24 locus for European (EUR), East Asian (EAS), South Asian (SAS) and continental African (AFR*) ancestries.

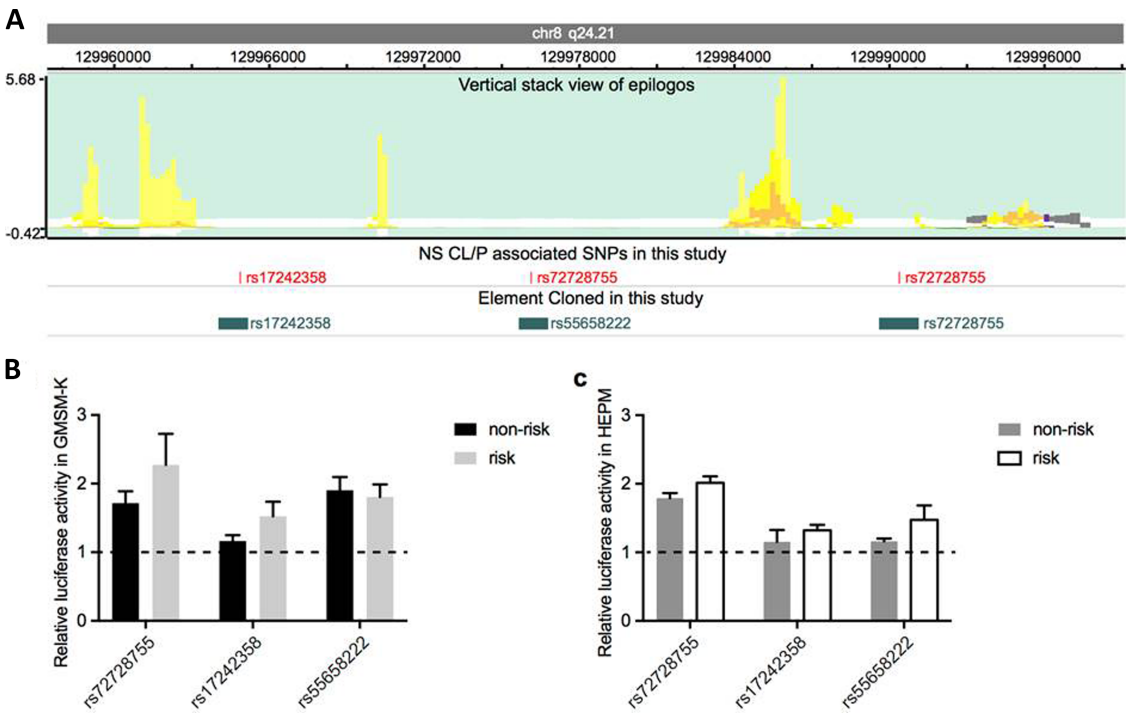

Figure 5. (**A**) Overlay of the three SNPs against chromatin marked as a regulatory element, (**B**) reporter assay in human fetal oral epithelial cell line (GMSM-K) and (**C**) primary HEPM.

Table 3. Novel variants in GWAS-identified candidate genes following Sanger sequencing

Gene	HGV^c	HGV^p	Type	Ghana	Nigeria	1KG	EVS	ExAC	P	S
ACVR2A		p.Leu187Pro	Missense	0	1	0	0	0		
DACH1		p.Gly739Ser	Missense	1	0	0	0	0	B	T

EVS, Exome Variant Server; P, PolyPhen; S SIFT; PS, PROVEAN score; B, benign; T, tolerated; PD, probably damaging; D, deleterious. c. refers to coding sequence position.

Aggregate Consortium (ExAC) and 1KG. The *DACH1* novel variant (p.Gly739Ser) was predicted to be benign and tolerated by Polymorphism Phenotyping (PolyPhen) and Sorting Intolerant from Tolerant (SIFT). However, structural analysis using the Have Your Protein Explained (HOPE) server reveals that the variant amino acid is larger than the wild type and a change in size could lead to bumps in protein folding. There may also be a loss of flexibility and torsion angles when the flexible amino acid glycine is substituted with the non-flexible serine (Supplementary Material, Fig. S3). The missense variant (p.Leu187Pro) in *ACVR2A* was predicted to be benign and tolerated by PolyPhen and SIFT.

Some previously reported OFC loci are replicated in African populations

To investigate how many previously reported loci for OFC show evidence of association in our study, we extracted all association records for terms related to 'orofacial clefts', 'cleft lip/palate', 'cleft lip' and 'cleft palate' in the National Human Genome Research Institute - European Bioinformatics Institute (NHGRI-EBI) GWAS Catalog. There were a total of 139 unique SNPs of which 121 were in our data set. However, only 39 of these SNPs (all for CL/P and/or all clefts) were genome-wide significant ($P < 5 \times 10^{-8}$) and were reported along with effect sizes. Of this subset, six variants showed significant association, i.e. $P < 0.05$, of which four SNPs also showed consistency of direction of effect for CL/P including SNPs in the chr8q24 region and in the genes *PAX7*, *VAX1* and *SOX5P1* (Table 4; Supplementary Material, Table S6). The effect sizes estimated in the present study (as indicated by the associated odds ratios) were remarkably similar to the observations in previous studies (Table 4). For CPO, only three SNPs have previously been reported to be genome-wide significant (12). These SNPs were monomorphic or near monomorphic in our data set, as they also are in other African ancestry populations in the 1KG or gnomAD databases. We also checked the association statistics for CPO in our study for the 48 SNPs and found that only 2 SNPs had a $P < 0.05$ but neither of the SNPs had consistent direction of effect with previous studies (Supplementary Material, Table 7). Given that African populations exhibit lower LD and smaller haplotype block sizes across the genome, we investigated the possibility of fine-mapping the replicated SNPs for CL/P to smaller regions that were observed in the original reports. For most of the replicated signals, African ancestry populations had the smallest haplotype blocks around the lead SNP (Fig. 6A). Fine-mapping indicated that the evidence supported one causal variant at each locus (Fig. 6B; Supplementary Material, Table S8) with the exception of one locus, rs987525 (an SNP in the 8q24 region fine-mapped above), where there was support for up to two causal variants. This finding further supports the notion that there are at least two causal variants in the 8q24 region. Clumping analysis in our study sample revealed that each of the leading association signals was consisted of a single clump of SNPs (i.e. it was unlikely that there were two or more variants explaining the association at any of the loci examined) with the exception of rs987525, which is consistent with the Finemapping (FINEMAP) analysis.

Discussion

Genomic studies of diverse populations have the potential to enrich our knowledge of the genetic architecture of many complex disorders. Here, we conducted a case-control GWAS for two OFC phenotypes, CPO and CL/P, in individuals enrolled from Ghana, Ethiopia and Nigeria. We identified two functionally plausible novel loci for CPO on chromosome 2 near *CTNNA2* and on chromosome 19 in *SULT2A1*.

CTNNA2 encodes the alpha-catenin protein that is involved in cell–cell adhesion by acting as a linker protein between cadherins and actin-containing filaments of the cytoskeleton (23). Although the role of *CTNNA2* in clefting is currently unknown, several studies have reported an association between E-cadherin and clefting (24–27). A recent GWAS for CL/P also identified a significant association near a gene involved in actin cytoskeleton (11). A recent exome sequencing study for Mendelian non-syndromic CL/P identified mutations in the epithelial cadherin-p120-catenin complex that includes CTNND1 (28). Studies in the chick embryo show that *ctnna2* is expressed in neural crest cells (19), and expression studies in the mouse embryo also demonstrate its expression in oral structures. *SULT2A1* encodes the enzyme sulfotransferase 2A1. While the gene has not previously been reported in relation to OFC, our *in situ* hybridization experiments show an expression of this gene in the palate. Knockout experiments for this gene in model organisms would further clarify its role in clefting.

Four loci showed suggestive association ($P < 5 \times 10^{-7}$) for CPO. They are near *ACVR2A* on chromosome 2, *SHH* on chromosome 7, *OPALIN* on chromosome 10 and *DACH1* on chromosome 13. *ACVR2A* encodes activin A type II receptor protein and is a member of the *TGFB* superfamily of structurally related signaling proteins (29). The *ACVR2A* mouse knockout has micrognathia and associated defects, such as cleft palate and no incisors (30). These defects are similar to the features of Pierre Robin sequence where the small mandible leads to the limited space for the tongue to descend into the mouth causing cleft palate (31). *ACVR2A* is expressed in human fetal palate suggesting that activin signaling plays a role in the development of the palate (32). *DACH1*, mouse homolog of *Drosophila* dachshund, is a transcription factor involved in the regulation of organ formation. It inhibits *TGFB* signaling by binding to *SMAD4* and *NCOR1* (33). *DACH1* is required for eye, leg and brain development. Homozygous mutants die shortly after birth due to failure to suckle, cyanosis and respiratory distress (34). The mouse *Dach2* has similar expression pattern as mouse *Dach1*, suggesting that there may be redundancy in the functions of these genes (34). Missense variations in *DACH2* have been reported in Allan–Herndon–Dudley syndrome (OMIM: 300523), Miles–Carpenter syndrome, X-linked cleft palate and/or megalocornea (35–38). These reports support a role

Table 4. Variants reported for CL/P from previous studies in NHGRI-EBI GWAS Catalog that were replicated in the present study

Present study					Previous studies							
SNP	Effect allele	P value	OR	OR 95% CI	SNP-allele	P-value	OR	OR 95% CI	Reported gene	Mapped gene	Authors	PubMed ID
rs742071	T	1.25E-03	1.22	1.08–1.37	rs742071-T	7.00E-09	1.32	1.126-1.537	PAX7	PAX7	Ludwig *et al.*, 2012	22863734
rs6585429	A	8.86E-03	0.83	0.73–0.96	rs6585429-A	7.00E-13	1.23	NR	VAX1	VAX1	Yu *et al.*, 2017	28232668
rs7017252	A	2.00E-02	1.18	1.03–1.35	rs7017252-A	8.00E-16	1.6		MYC; LOC728724	LINC00824; LINC00977	Yu *et al.*, 2017	28232668
rs12543318	C	2.22E-02	0.85	0.74–0.98	rs12543318-C	9.00E-12	1.23	NR	DCAF4L2	SOX5P1; LOC100419762	Yu *et al.*, 2017	28232668
rs987525	A	3.50E-02	1.14	1.01–1.29	rs987525-A	5.00E-35	1.92	1.66-2.218	NR	LINC00824; LINC00977	Ludwig *et al.*, 2012	22863734
rs7078160	A	3.54E-02	1.16	1.01–1.33	rs7078160-A	4.00E-11	1.38	1.213-1.576	NR	KIAA1598	Ludwig *et al.*, 2012	22863734
rs6129653	A	5.85E-02	1.16	0.99–1.36	rs6129653-A	9.00E-12	1.23		MAFB	LOC102724968; LOC105372620	Yu *et al.*, 2017	28232668
rs6495117	A	7.37E-02	1.12	0.99–1.26	rs6495117-A	6.00E-11	1.2	NR	NR	LOC102723750; CLK3	Yu *et al.*, 2017	28232668
rs7552	G	7.75E-02	1.13	0.99–1.29	rs7552-G	6.00E-22	1.37	NR	FAM49A	FAM49A	Yu *et al.*, 2017	28232668
rs1838105	A	9.60E-02	0.9	0.79–1.02	rs1838105-A	1.00E-11	1.22		GOSR2	GOSR2	Yu *et al.*, 2017	28232668
rs2283487	A	1.21E-01	0.91	0.80–1.03	rs2283487-A	1.00E-10	1.2	NR	CREBBP; ADCY9	CREBBP; LOC102724927	Yu *et al.*, 2017	28232668
rs861020	A	1.32E-01	0.88	0.75–1.04	rs861020-A	3.00E-12	1.44	1.273-1.635	IRF6	IRF6	Ludwig *et al.*, 2012	22863734
rs8001641	A	1.39E-01	1.14	0.96–1.35	rs8001641-A	9.00E-11	1.35	1.141-1.607	SPRY2	LOC105370275	Ludwig *et al.*, 2012	22863734
rs9545308	A	1.40E-01	1.29	0.92–1.80	rs9545308-A	2.00E-09	1.29		SPRY2	LOC101927216	Yu *et al.*, 2017	28232668
rs2289187	G	1.50E-01	1.09	0.97–1.24	rs2289187-G	4.00E-11	1.21		NR	UBL7	Yu *et al.*, 2017	28232668
rs560426	G	1.87E-01	0.92	0.82–1.04	rs560426-G	3.00E-12	1.42	1.243-1.623	NR	ABCA4	Ludwig *et al.*, 2012	22863734
rs8049367	C	2.49E-01	1.08	0.95–1.23	rs8049367-C	9.00E-12	1.35	1.25-1.47	CREBBP; ADCY9	CREBBP; LOC102724927	Sun *et al.*, 2015	25775280
rs7148069	A	2.66E-01	1.08	0.94–1.25	rs7148069-A	2.00E-08	1.22		LOC283553	LINC00640; LOC105370496	Yu *et al.*, 2017	28232668
rs227731	C	3.29E-01	0.93	0.82–1.07	rs227731-C	9.00E-09	1.19		NOG; C17orf67	NOG; C17orf67	Yu *et al.*, 2017	28232668
rs908822	A	3.76E-01	1.2	0.80–1.81	rs908822-A	4.00E-08	1.31		LOC285419	LINC01091; LOC105377407	Yu *et al.*, 2017	28232668
rs2304269	A	3.82E-01	1.28	0.73–2.24	rs2304269-A	1.00E-12	1.23	NR	TMEM19	TMEM19	Yu *et al.*, 2017	28232668
rs13317	A	3.83E-01	1.07	0.92–1.24	rs13317-A	4.00E-08	1.18	NR	FGFR1	FGFR1	Yu *et al.*, 2017	28232668
rs287982	A	3.95E-01	0.95	0.84–1.07	rs287982-A	6.00E-09	1.22	NR	TAF1B	LOC105373421; TAF1B	Yu *et al.*, 2017	28232668
rs3741442	G	4.18E-01	0.92	0.76–1.12	rs3741442-G	4.00E-12	1.22		KRT18	KRT18; EIF4B	Yu *et al.*, 2017	28232668
rs2872615	A	4.73E-01	1.06	0.90–1.26	rs2872615-A	9.00E-12	1.22	NR	NTN1	LOC101928235; NTN1	Yu *et al.*, 2017	28232668
rs7871395	A	4.82E-01	1.05	0.92–1.20	rs7871395-A	6.00E-09	1.21		GADD45G	LOC105376137; LOC105376139	Yu *et al.*, 2017	28232668
rs957448	A	4.96E-01	0.96	0.84–1.09	rs957448-A	1.00E-12	1.23	NR	KIAA1429	KIAA1429	Yu *et al.*, 2017	28232668
rs2064163	C	5.92E-01	0.97	0.85–1.10	rs2064163-C	9.00E-19	1.3	NR	IRF6; DIEXF	DIEXF; SYT14	Yu *et al.*, 2017	28232668
rs4791774	G	6.38E-01	1.03	0.91–1.16	rs4791774-G	5.00E-19	1.56	1.42-1.72	NTN1	NTN1	Sun *et al.*, 2015	25775280
rs12681366	A	6.45E-01	1.03	0.91–1.17	rs12681366-A	2.00E-10	1.2	NR	RAD54B	RAD54B	Yu *et al.*, 2017	28232668
rs1243572	G	6.45E-01	1.04	0.88–1.22	rs1243572-G	4.00E-10	1.2		GSC	LOC107984693; LOC107984639	Yu *et al.*, 2017	28232668
rs9381107	G	6.92E-01	0.97	0.83–1.13	rs9381107-G	3.00E-09	1.2	NR	LOC100506207	LOC107986562; LOC107986563	Yu *et al.*, 2017	28232668
rs7590268	G	7.54E-01	1.02	0.88–1.18	rs7590268-G	1.00E-08	1.41	1.225-1.636	THADA	THADA	Ludwig *et al.*, 2012	22863734
rs12229892	G	7.58E-01	1.09	0.62–1.92	rs12229892-G	2.00E-10	1.2	NR	NR	PTPN11	Yu *et al.*, 2017	28232668
rs10512248	A	7.67E-01	1.02	0.90–1.15	rs10512248-A	5.00E-10	1.22	NR	PTCH1	PTCH1	Yu *et al.*, 2017	28232668
rs705704	A	7.84E-01	1.03	0.85–1.24	rs705704-A	1.00E-09	1.22		RPS26	LOC105369780	Yu *et al.*, 2017	28232668
rs481931	C	8.32E-01	0.98	0.80–1.19	rs481931-C	1.00E-12	1.25	NR	ABCA4	ABCA4	Yu *et al.*, 2017	28232668
rs1907989	G	8.84E-01	1.01	0.89–1.15	rs1907989-G	2.00E-08	1.18	NR	MSX1	LOC101928279; LINC01396	Yu *et al.*, 2017	28232668
rs13041247	T	9.60E-01	1.00	0.87–1.15	rs13041247-T	2.00E-11	1.32	1.20-1.41	MAFB	LOC102724968	Sun *et al.*, 2015	25775280

*Data from previous studies extracted from NHGRI-EBI GWAS Catalog (version 2018-09-30). 'NR' indicates 'not reported'. Effect sizes were reported with respect to the same allele across studies. Where more than one study reported the same genome-wide significant SNP, the study with the smallest *P*-value is presented in the table.

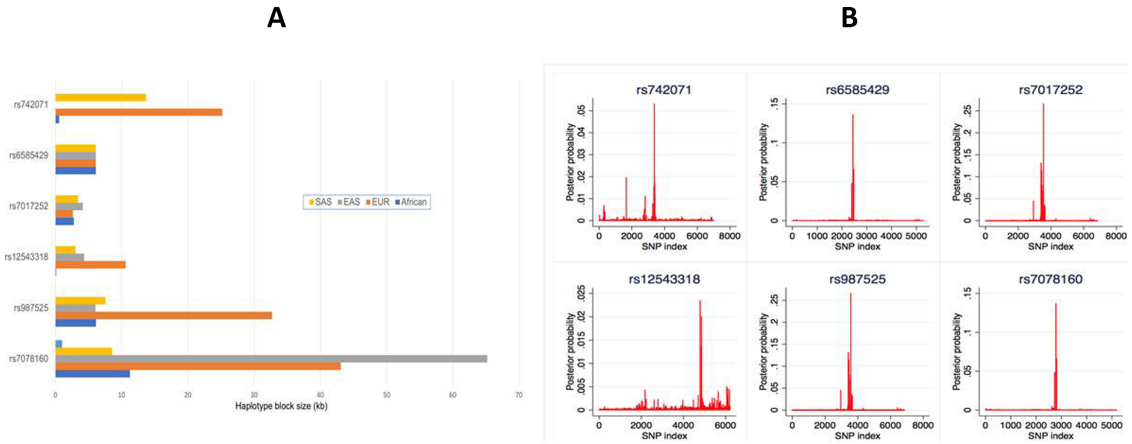

Figure 6. Haplotype blocks around the lead SNPs from previous GWAS that were replicated in the present study.

for the missense variation (p.Gly739Ser) that we found in an individual with CL/P. *OPALIN* encodes the Opalin protein and has never been reported to play a role in clefting. *SHH* encodes the sonic hedgehog protein and it plays a role in cell division and embryogenesis. Mutations in *SHH* have been implicated in holoprosencephaly (39,40). A few studies have suggested a role for *SHH* in non-syndromic CL/P (41,42). We are the first to report an association with *SHH* for isolated CPO from GWAS.

For CL/P, our most significant locus is in the 8q24 region that has been previously reported in several other studies (5–9,11) in European populations. The lead SNP in our study is different from previous reports. Our analyses suggest that the two SNPs represent distinct signals for CL/P within the 8q24 region. While the evidence in our study suggests that the lead SNP represents a single causal variant, our transfection experiments were unable to determine which of the three tightly linked lead SNPs was the causal variant. The identification of significant SNPs in the 8q24 locus in multiple populations strongly supports its role in CL/P and suggests the possibility of more than one causal locus within this region.

Our study replicated several SNPs that are previously reported to be associated with OFC. Of note is the chromosome 9 locus near *PTCH1*. *PTCH1* encodes the patched homolog 1 protein, a member of the patched family that is mutated in Gorlin syndrome (whose features include OFC) (43). It is a receptor for sonic hedgehog and is involved in cell proliferation, formation of structures during embryogenesis and tumor formation (44–46). Rare and common variants in *PTCH1* have been implicated in non-syndromic CL/P (16,47,48).

This study has some limitations. There is a lack of strong evidence in the replication cohort, which is likely due to the fact that it is small in size and has limited power to detect significant associations. Other potential reasons for this observation include differences in LD, allele frequency differences and other sources of heterogeneity between population groups. Therefore, there is a need for further replication of the novel signals in larger African cohorts. Additional replication in other populations is also warranted for the new significant signals on chromosomes 2 and 19. The present study considered only common and low-frequency variants but did not consider rare variants because the genotyping tool was a GWAS SNP array with the yield boosted by imputation. A more comprehensive analysis done with whole-genome sequencing would provide a more complete association study that includes all classes of variants (including rare variants). We also noted that most of the association *P*-values in the replication sample were not small ($P < 0.05$), and those that were small often displayed inconsistency of direction of effect. For this reason, we limited the SNPs of interest to those that showed consistency of direction of effect in the replication sample in addition to being genome-wide significant in the discovery and combined analysis.

In conclusion, this first GWAS of OFC in sub-Saharan Africans identified novel loci for CPO and confirmed several findings previously reported from other ancestral populations. These findings add to the growing evidence about genetic risk factors for OFC and provide new candidate genes for functional studies.

Materials and Methods

Study population and sample information

Ethical approval was obtained from the Institutional Review Boards (IRBs) at the Lagos University Teaching Hospital (ADM/DCST/HREC/VOL.XV/321), Obafemi Awolowo University Teaching Hospital (ERC/2011/12/01), Kwame Nkrumah University of Science and Technology (CHRPE/RC/018/13), Addis Ababa University (003/10/surg), New York State Department of Health (IRB 07-007) and the NIH Office of Human Subjects Research (OHSRP 11631). We have previously reported the recruitment and sample used for the discovery study (49). In summary, eligible subjects are individuals with non-syndromic OFC and with families born to Ghanaian, Ethiopian and Nigerian parents. Births from Caucasians and Asians are excluded.

We identified eligible cases after IRB approvals through various free OFC surgical repair projects, most of which participate in the Pan-African Association of Cleft Lip and Palate network for treatment of OFC in Africa. This network is supported by cleft charities, and all use a common standardized protocol for phenotyping. For all the enrolled cases, the surgeons carried out standardized physical examinations, took clinical photographs and provided full description of OFC phenotypes and other recognizable malformations in a clinical database. We used our access to echocardiogram results to rule out cardiac defects. For both the discovery and replication samples (Supplementary Material, Table S9), controls were apparently healthy individuals without clefts enrolled at the same sites as cases. Both related (usually the mother) and unrelated controls were included in the analysis. In Nigeria, Ghana and Ethiopia, unrelated controls were recruited at infant welfare/immunization clinics at

the site of the same medical centers where the cases were enrolled and were matched for gender, age and geographical location. In the Democratic Republic of the Congo and the US sites, controls were recruited from the same medical centers as cases. Signed informed consent was obtained from all families that participated in the study. Every family recruited into the study was assigned a unique identifier (UNID) number. Data from all recruited families were remotely entered from all the centers in Africa into a secured REDCap database (50). Deidentified samples were shipped from sites in Africa to the United States.

DNA extraction and preliminary quality control

Saliva samples were labeled at the Butali laboratory in Iowa and assigned a UNID number prior to DNA extraction. The DNA extraction was done at the Butali laboratory using the Murray laboratory protocol (genetics@uiowa.edu). Every sample was quantified using Qubit (http://www.invitrogen.com/site/us/en/home/brands/Product-Brand/Qubit.html; Thermo Fisher Scientific, Grand Island, NY) and separated into a stock and several working aliquots for downstream applications. We confirmed the sex reported in the REDCap database using TaqMan XY genotyping. These were done as part of our quality control (QC) process in the laboratory to prevent sample mislabeling. We then shipped 25 µl aliquot of consented samples with confirmed genetic sex and DNA concentration of ≥50 ng/µl to the Center for Inherited Disease Research for Multi-Ethnic Genotyping Array (MEGA) genotyping.

Genotyping

The expanded Illumina MEGA v2 15070954 A2 (genome build 37) that contains over 2 million SNPs and over 60 000 rare variants selected from populations of African origin was used for genotyping. We successfully conducted genotyping on 3347 samples, which included 3198 unique samples and 70 duplicates. HapMap controls (70 unique samples and 9 duplicates) were also genotyped as part of the QC process.

Data cleaning

A detailed description of this process was recently published (51). Briefly, we checked for sex chromosome anomalies, missing call rates, batch effects, identification of large chromosomal anomalies, confirmation of relatedness (i.e. identity by descent) and establishment of continental ancestry with respect to HapMap samples using methods described in Laurie et.al (2010) (52) and implemented using R packages GWAS Tools (53), SNPRelate (54) and GENESIS (55). This process allowed the use of a high-quality genotype data set for identifying significant genotype associations with non-syndromic OFC.

Imputation and association analyses

As is usual for GWAS to conduct imputation, we did both preimputation and postimputation QC (a full report is available in dbGaP, and we present a summary here). Briefly, for preimputation genotypes, after applying technical filters, we filtered for missing call rates ≥2%, >1 discordant call in 70 study duplicates, >1 Mendelian errors in 890 duos and trios, an Hardy Weinberg equilibrium (HWE) of $P < 10^{-3}$ and a minor allele frequency (MAF) of <0.01, among others (Supplementary Material, Table S10). For the imputed SNPs, we only retained variants with a MAF of ≥0.01 and a quality metric (INFO, score is an estimated quality measurement of imputation) of ≥0.3, with the latter chosen based on the balance between stringency and inclusivity as recommended by (56). In the present study, choosing a threshold of 0.3 retained 69.5% of all imputed variants for downstream analyses, while more stringent thresholds of 0.5 and 0.8 would retain 63.5% and 49.0% of imputed variants, respectively.

Imputation was carried out using IMPUTE2 (a genotype imputation and phasing program) into the 1KG Phase 3 reference imputation panel (57). The final data set that passed QC was consisted of 3178 (1133 male and 2045 female) participants enrolled from Ethiopia (30%), Ghana (43%) and Nigeria (27%). The data set included 814 cases of CL/P, 205 cases of isolated CPO and 2159 related and unrelated controls.

The imputation yield was ∼45 million SNPs of which ∼17 million passed our QC filter and were included in the final analyses. Given the known differences in the developmental and genetic basis of isolated CL/P versus CPO, we conducted two separate GWAS analyses (one for each phenotype). Single-variant association tests were done for imputed dosage data filtered for an imputed allelic dosage frequency of <0.01 and an INFO of <0.3 using logistic mixed models as implemented in the GMAAT package (58). This approach enabled us to obtain valid association tests while adjusting for population structure (the first seven eigenvectors of the genotypes), relationships between participants (using the computed genetic relatedness matrix) and covariates (sex and study site). The Q–Q plot of the distribution of P-values did not show any residual stratification (Supplementary Material, Figure 5).

Replication

For the replication study, we included an independent sample of OFC cases and controls (300 CL/P cases, 179 CPO cases and 2523 controls) from Ghana, Nigeria, Ethiopia, Democratic Republic of Congo and African-American samples from New York and Virginia, USA. (Supplementary Material, Table S9). DNA extracted from deidentified residual dried blood spots was genotyped for NY cases (identified from the New York State Congenital Malformations Registry) and controls (identified from birth records). We selected for genotyping GWAS-significant SNPs and SNPs in LD with index SNP for a total of 48 SNPs using Fluidigm Corporation (San Francisco, CA), which allowed simultaneous genotyping of variants in samples in a multiplex, high-throughput format. Data were analyzed using PLINK2 (https://www.cog-genomics.org/plink2). For high-quality SNPs (an SNP success rate of ≥97%), association with CPO and CL/P was tested under an additive genetic model. Combined analysis of discovery and replication studies for the 48 SNPs was done as implemented in METAL (a tool for the meta-analysis of genome-wide association studies) (59). Variants that had $P < 5 \times 10^{-8}$ and had the same direction of effect in both studies were considered genome-wide significant.

Fine-mapping

Haplotypes were constructed using the confidence interval method of Gabriel et al (2002) (60). Clumping analysis of association statistics was done with PLINK (61) (Purcell *et al.*, 2007) using default parameters. Fine-mapping was done using a shotgun stochastic search algorithm as implemented in

FINEMAP (20). Reciprocal conditional analysis was done with genome-wide complex trait analysis (GCTA) (62).

Identification of GWAS candidate genes with a TAD

GWAS signals that affect enhancers most likely influence the expression of genes within the same TAD. Each region was visualized in the human reference genome (hg19) by searching for interaction domain for the index SNP ID (http://promoter.bx.psu.edu/hi-c/view.php).

Sanger sequencing

We used methods that we reported previously (49). We optimized primers for the amplification of exons in the ACVR2A1 and DACH1 genes. These genes where chosen based on their expression in the craniofacial region and the presence of mouse knockouts with cleft palate (http://www.informatics.jax.org/). A DNA concentration of 4 ng/µl of in a 10 µl reaction for the polymerase chain reaction were used. Two Yoruba HapMap samples and two water samples were added to the 96-well plates as template and non-template controls, respectively. Details of primers used and annealing temperatures are available from the Butali laboratory upon request. A total of 270 cases from Ghana, Ethiopia and Nigeria were sequenced. We sent the amplified DNA products for sequencing at Functional Biosciences, Inc., Madison, WI (http://order.functionalbio.com/seq/index).

We compared the identified novel variations with variations in the 1KG database (http://www.1000genomes.org/), Exome Variant Server database (http://snp.gs.washington.edu/EVS/) and ExAC database (http://exac.broadinstitute.org/). The variants were also compared to over 5200 African and African-American control exomes in these databases. We also sequenced population-matched controls for each novel variant in order to validate novel variants. We predicted the functional effects of novel variants using bioinformatics tools such as PolyPhen (http://genetics.bwh.harvard.edu/pph2/) (63), SIFT (http://sift.jcvi.org/) (64) and HOPE (http://www.cmbi.ru.nl/hope) (65). Segregation analyses were performed to determine if variants are *de novo* or inherited by sequencing samples from parents, when available.

In situ hybridization of Sult2a1 in mice at E12.5 and E14.5

The *in situ* hybridization method used in this study was adapted from our Sox2 paper (66). In summary, we used formalin-fixed paraffin embedded tissue sections for *in situ* hybridization. Mouse palatal samples were processed following the typical paraffin embedding process. Sagittal sections were cut in 8 µm, and we used the standard *in situ* hybridization method listed in Gregorieff's protocol (67). Digoxigenin-labeled probe was made from DIG RNA Labeling Kit (Roche Diabetes Care, Inc., Indianapolis, IN, USA #11175025910). Primers used for *Sult2a1* are the following: *Sult2a1*-F: 5′-ATGATGTCAGACTATAATTGGTT-3′, and *Sult2a1-SP6*-R: 5′-ATTTAGGTGACACTATAGTTATTCCCATGGGAAAATCCCTGGG-3′.

Luciferase experiments to determine the functional role of SNPs at the 8q24 locus

Plasmid Construct. We used RP11-976D7 as a template to clone all three candidate elements in the 8q24 locus. The entire products were cloned into pENTR/D-TOPO plasmid (Life Technologies, Carlsbad, CA) for validation using Sanger sequencing. Site-directed mutagenesis was employed to get either non-risk or risk allele into the elements. We then shuttled all the candidate elements into cFos-FFLuc plasmid for *in vitro* luciferase assay.

Cell culture, electroporation and dual luciferase assay GMSM-K human embryonic oral epithelial cell line 6 (a kind gift from Dr Daniel Grenier) were maintained in keratinocyte serum-free medium (Life Technologies) supplemented with epithelial growth factor (EGF) and bovine pituitary extract (Life Technologies). All cells were incubated at 37°C in 5% CO_2. HEPM 7 were purchased from American Type Culture Collection (ATCC) (ATCC®, Manassas, VA, USA, CRL-1486™) and maintained in ATCC-formulated Eagle's Minimum Essential Medium (ATCC) supplemented with 10% fetal bovine serum (Life Technologies) and 1% antibiotic-antimycotic (Life Technologies). For dual luciferase activity assay, each reporter construct was cotransfected with Renilla luciferase plasmid for three biological replicates. Briefly, plasmids were electroporated into GMSM-K cells with Amaxa™ Cell Line Nucleofector® Kit V (Lonza, Cologne, Germany) using Nucleofector™ II (Lonza) (program: X-005), and plasmids were electroporated into HEPM cells with Amaxa™ Basic Nucleofector™ Kit for primary mammalian fibroblasts (Lonza) using Nucleofector™ II (Lonza) (program: U-020). The Dual-Luciferase Reporter Assay System (Promega, Madison, WI) and 20/20n Luminometer (Turner BioSystems, Sunnyvale, CA) were employed to evaluate the luciferase activity 72 h posttransfection. Relative luciferase activities were calculated by the ratio between the value for firefly and Renilla luciferase activities. Three measurements were made for the lysate from each transfection group. All quantified results are presented as mean ± scanned electron microscope (SEM). Student's *t*-test was used to determine statistical significance.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We owe a lot of gratitude to families in Ghana, Ethiopia and Nigeria who participated in this research. The Center for Research on Genomics and Global Health (CRGGH) (1ZIAHG200362) is supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, the Center for Information Technology and the Office of the Director at the National Institutes of Health.

Conflict of Interest statement. None declared.

Funding

National Institute of Dental and Craniofacial Research (NIDCR) K99/R00 (DE022378); Robert Wood Johnson Foundation (72429 to A.B.), National Institute of Health (NIH) (R01 DE023575 to R.A.C.); R37 (DE-08559 and DE-016148 to J.C.M.); Ghana Cleft Foundation (G001 to L.J.J.G.); Intramural Research Program of the National Institutes of Health in the CRGGH; Intramural Research Program of the National Institutes of Health; Eunice Kennedy Shriver National Institute of Child Health and Human Development (HHSN275201100001I and HHSN27500005).

References

1. Mossey, P.A., Little, J., Munger, R.G., Dixon, M. and Shaw, W.C. (2009) Cleft lip and palate. *Lancet*, **374**, 1773–1785.
2. Adetayo, O., Ford, R. and Martin, M. (2012) Africa has unique and urgent barriers to cleft care: lessons from practitioners at the Pan-African Congress on cleft lip and palate. *Pan Afr. Med. J.*, **12**, 15.
3. Awoyale, T.A., Onajole, A.T., Ogunnowo, B.E., Adeyemo, W.L., Wanyonyi, K.L. and Butali, A. (2015) Quality of life of family caregivers of children with orofacial clefts in Nigeria: a mixed-methods study. *Oral Dis.*, **22**, 116–122.
4. 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.
5. 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.* (2011) Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat. Genet.*, **41**, 473–477.
6. 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 non-syndromic cleft lip with or without cleft palate on 8q24. *J. Pediatr.*, **155**, 909–913.
7. 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.* (2009) Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat. Genet.*, **42**, 24–26.
8. 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.
9. Ludwig, K.U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A., Becker, J., Herberz, R., AIChawa, 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.
10. 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.
11. Leslie, E.J., Carlson, J.C., Shaffer, J.R., Feingold, E., Wehby, G., Laurie, C.A., Jain, D., Laurie, C.C., Doheny, K.F., McHenry, T. *et al.* (2016) A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. *Hum. Mol. Genet.*, **25**, 2862–2872.
12. Leslie, E.J., Liu, H., Carlson, J.C., Shaffer, J.R., Feingold, E., Wehby, G., Laurie, C.A., Jain, D., Laurie, C.C., Doheny, K.F. *et al.* (2016) A genome-wide association study of nonsyndromic cleft palate identifies an etiologic missense variant in GRHL3. *Am. J. Hum. Genet.*, **98**, 744–754.
13. Mangold, E., Böhmer, A.C., Ishorst, N., Hoebel, A.K., Gültepe, P., Schuenke, H., Klamt, J., Hofmann, A., Gölz, L., Raff, R. *et al.* (2016) Sequencing the GRHL3 coding region reveals rare truncating mutations and a common susceptibility variant for nonsyndromic cleft palate. *Am. J. Hum. Genet.*, **98**, 755–762.
14. Leslie, E.J., Carlson, J.C., Shaffer, J.R., Butali, A., Buxó, C.J., Castilla, E.E., Christensen, K., Deleyiannis, F.W., Leigh Field, L., Hecht, J.T. *et al.* (2017) Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Hum. Genet.*, **136**, 275–286.
15. Ludwig, K.U., Ahmed, S.T., Böhmer, A.C., Sangani, N.B., Varghese, S., Klamt, J., Schuenke, H., Gültepe, P., Hofmann, A., Rubini, M. *et al.* (2016) Meta-analysis reveals genome-wide significance at 15q13 for nonsyndromic clefting of both the lip and the palate, and functional analyses implicate GREM1 as a plausible causative gene. *PLoS Genet.*, **12**, e1005914.
16. Yu, Y., Zuo, X., He, M., Gao, J., Fu, Y., Qin, C., Meng, L., Wang, W., Song, Y., Cheng, Y. *et al.* (2017) Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat. Commun.*, **8**, 14364.
17. Cavalli-Sforza, L.L. and Feldman, M.W. (2003) The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.*, **33**, 266–275.
18. Ramsay, M., Tiemessen, C.T., Choudhury, A. and Soodyall, H. (2011) Africa: the next frontier for human disease gene discovery. *Hum. Mol. Genet.*, **20**, R214–R220.
19. Jhingory, S., Wu, C.Y. and Taneyhill, L.A. (2010) Novel insight into the function and regulation of αN-catenin by Snail2 during chick neural crest cell migration. *Dev. Biol.*, **344**, 896–910.
20. Benner, C., Spencer, C.C., Havulinna, A.S., Salomaa, V., Ripatti, S. and Pirinen, M. (2016) FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*, **32**, 1493–1501.
21. 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.
22. Yoneda, T. and Pratt, R.M. (1981) Interaction between glucocorticoids and epidermal growth factor in vitro in the growth of palatal mesenchymal cells from the human embryo. *Differentiation*, **19**, 194–198.
23. Cooper, G.M. (2000) Figure 11.14: model of attachment of actin filaments to catenin–cadherin complexes. In *The Cell: A Molecular Approach*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.
24. Bureau, A., Parker, M.M., Ruczinski, I., Taub, M.A., Marazita, M.L., Murray, J.C., Mangold, E., Noethen, M.M., Ludwig, K.U. and Hetmanski, J.B. (2014) Whole exome sequencing of distant relatives in multiplex families implicates rare variants in candidate genes for oral clefts. *Genetics*, **197**, 1039–1044.
25. Brito, L.A., Yamamoto, G.L., Melo, S., Malcher, C., Ferreira, S.G., Figueiredo, J., Alvizi, L., Kobayashi, G.S., Naslavsky, M.S., Alonso, N. *et al.* (2015) Rare variants in the epithelial cadherin gene underlying the genetic etiology of nonsyndromic cleft lip with or without cleft palate. *Hum. Mutat.*, **36**, 1029–1033.
26. Machado, R.A., de Freitas, E.M., de Aquino, S.N., Martelli, D.R., Swerts, M.S., Reis, S.R., Persuhn, D.C., Moreira, H.S., Dias, V.O., Coletta, R.D. *et al.* (2017) Clinical relevance of breast and gastric cancer-associated polymorphisms as potential susceptibility markers for oral clefts in the Brazilian population. *BMC Med. Genet.*, **18**, 39.
27. Song, H., Wang, X., Yan, J., Mi, N., Jiao, X., Hao, Y., Zhang, W. and Gao, Y. (2017) Association of single-nucleotide polymorphisms of CDH1 with nonsyndromic cleft lip with or without cleft palate in a northern Chinese Han population. *Medicine (Baltimore)*, **96**, e5574.

28. Cox, L.L., Cox, T.C., Moreno Uribe, L.M., Zhu, Y., Richter, C.T., Nidey, N., Standley, J.M., Deng, M., Blue, E., Chong, J.X. *et al.* (2018) Mutations in the epithelial cadherin-p120-catenin complex cause Mendelian non-syndromic cleft lip with or without cleft palate. *Am. J. Hum. Genet.*, **102**, 1143–1157.
29. Attisano, L., Cárcamo, J., Ventura, F., Weis, F.M., Massagué, J. and Wrana, J.L. (1993) Identification of human activin and TGF beta type I receptors that form heteromeric kinase complexes with type II receptors. *Cell*, **75**, 671–680.
30. Matzuk, M.M., Kumar, T.R. and Bradley, A. (1995) Different phenotypes for mice deficient in either activins or activin receptor type II. *Nature*, **374**, 356–360.
31. Tan, T.Y., Kilpatrick, N. and Farlie, P.G. (2013) Developmental and genetic perspectives on Pierre Robin sequence. *Am. J. Med. Genet. C Semin. Med. Genet.*, **163C**, 295–305.
32. Lambert-Messerlian, G., Eklund, E., Pinar, H., Tantravahi, U. and Schneyer, A.L. (2007) Activin subunit and receptor expression in normal and cleft human fetal palate tissues. *Pediatr. Dev. Pathol.*, **10**, 436–445.
33. Wu, K., Yang, Y., Wang, C., Davoli, M.A., D'Amico, M., Li, A., Cveklova, K., Kozmik, Z., Lisanti, M.P., Russell, R.G. *et al.* (2003) DACH1 inhibits transforming growth factor-beta signaling through binding Smad4. *J. Biol. Chem.*, **278**, 51673–51684.
34. Davis, R.J., Shen, W., Sandler, Y.I., Amoui, M., Purcell, P., Maas, R., Ou, C.N., Vogel, H., Beaudet, A.L. and Mardon, G. (2001) Dach1 mutant mice bear no gross abnormalities in eye, limb, and brain development and exhibit postnatal lethality. *Mol. Cell. Biol.*, **21**, 1484–1490.
35. Chen, J.D., Mackey, D., Fuller, H., Serravalle, S., Olsson, J. and Denton, M.J. (1989) X-linked megalocornea: close linkage to DXS87 and DXS94. *Hum. Genet.*, **83**, 292–294.
36. Miles, J.H. and Carpenter, N.J. (1991) Unique X-linked mental retardation syndrome with fingertip arches and contractures linked to Xq21.31. *Am. J. Med. Genet.*, **38**, 215–223.
37. Bialer, M.G., Lawrence, L., Stevenson, R.E., Silverberg, G., Williams, M.K., Arena, J.F., Lubs, H.A. and Schwartz, C.E. (1992) Allan–Herndon–Dudley syndrome: clinical and linkage studies on a second family. *Am. J. Med. Genet.*, **43**, 491–497.
38. Forbes, S.A., Richardson, M., Brennan, L., Arnason, A., Bjornsson, A., Campbell, L., Moore, G. and Stanier, P. (1995) Refinement of the X-linked cleft palate and ankyloglossia (CPX) localisation by genetic mapping in an Icelandic kindred. *Hum. Genet.*, **95**, 342–346.
39. Aguinaga, M., Llano, I., Zenteno, J.C. and Kofman, A.S. (2011) Novel sonic hedgehog mutation in a couple with variable expression of holoprosencephaly. *Case Rep. Genet.*, 703497.
40. Mercier, S., Dubourg, C., Garcelon, N., Campillo-Gimenez, B., Gicquel, I., Belleguic, M., Ratié, L., Pasquier, L., Loget, P. *et al.* (2011) New findings for phenotype–genotype correlations in a large European series of holoprosencephaly cases. *J. Med. Genet.*, **48**, 752–760.
41. Orioli, I.M., Vieira, A.R., Castilla, E.E., Ming, J.E. and Muenke, M. (2002) Mutational analysis of the sonic hedgehog gene in 220 newborns with oral clefts in a South American (ECLAMC) population. *Am. J. Med. Genet.*, **108**, 12–15.
42. de Araujo, T.K., Secolin, R., Félix, T.M., de Souza, L.T., Fontes, M.Í., Monlleó, I.L., de Souza, J., Fett-Conte, A.C., Ribeiro, E.M. *et al.* (2016) A multicentric association study between 39 genes and nonsyndromic cleft lip and palate in a Brazilian population. *J. Craniomaxillofac. Surg.*, **44**, 16–20.
43. Johnson, R.L., Rothman, A.L., Xie, J., Goodrich, L.V., Bare, J.W., Bonifas, J.M., Quinn, A.G., Myers, R.M., Cox, D.R., Epstein, E.H. Jr. *et al.* (1996) Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science*, **272**, 1668–1671.
44. Hahn Christiansen, J., Wicking, C., Zaphiropoulos, P.G., Chidambaram, A., Gerrard, B., Vorechovsky, I., Bale, A.E., Toftgard, R., Dean, M. *et al.* (1996) A mammalian patched homolog is expressed in target tissues of sonic hedgehog and maps to a region associated with developmental abnormalities. *J. Biol. Chem.*, **271**, 12125–12128.
45. Villavicencio, E.H., Walterhouse, D.O. and Iannaccone, P.M. (2000) The sonic hedgehog–patched–gli pathway in human development and disease. *Am. J. Hum. Genet.*, **67**, 1047–1054.
46. Corcoran, R.B. and Scott, M.P. (2002) A mouse model for medulloblastoma and basal cell nevus syndrome. *J. Neurooncol.*, **53**, 307–318.
47. Mansilla, M.A., Cooper, M.E., Goldstein, T., Castilla, E.E., Lopez Camelo, J.S., Marazita, M.L. and Murray, J.C. (2017) Contributions of PTCH gene variants to isolated cleft lip and palate. *Cleft Palate Craniofac. J.*, **43**, 21–29.
48. Moreno, L.M., Mansilla, M.A., Bullard, S.A., Cooper, M.E., Busch, T.D., Machida, J., Johnson, M.K., Brauer, D., Krahn, K., Daack-Hirsch, S. *et al.* (2009) FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. *Hum. Mol. Genet.*, **18**, 4879–4896.
49. Gowans, L.J., Adeyemo, W.L., Eshete, M., Mossey, P.A., Busch, T., Aregbesola, B., Donkor, P., Arthur, F.K., Bello, S.A., Martinez, A. *et al.* (2016) Association studies and direct DNA sequencing implicate genetic susceptibility loci in the etiology of nonsyndromic orofacial clefts in sub-Saharan African populations. *J. Dent. Res.*, **95**, 1245–1256.
50. Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N. and Conde, J.G. (2009) Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.*, **42**, 377–381.
51. Oseni, G.O., Jain, D., Mossey, P.A., Busch, T.D., Gowans, L.J.J., Eshete, M.A., Adeyemo, W.L., Laurie, C.A., Laurie, C.C., Owais, A. *et al.* (2018) Identification of paternal uniparental disomy on chromosome 22 and a de novo deletion on chromosome 18 in individuals with orofacial clefts. *Mol. Genet. Genomic. Med.*, doi:10.1002/mgg3.459.
52. 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.* (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiol.*, **34**, 591–602.
53. Gogarten, S.M., Bhangale, T., Conomos, M.P., Laurie, C.A., CP, M.H., Painter, I., Zheng, X., Crosslin, D.R., Levine, D., Lumley, T. *et al.* (2012) GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics*, **28**, 3329–3331.
54. Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C. and Weir, B.S. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.
55. Conomos, M.P. and Thornton, T. (2016) GENESIS: GENetic EStimation and Inference in Structured samples (GENESIS): statistical methods for analyzing genetic data from samples with population structure and/or relatedness. R package version 2.4.0.

56. de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S. and Voight, B.F. (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.*, **17**, R122–R128.
57. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. and Abecasis, G.R. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.*, **44**, 955–959.
58. Chen, H., Wang, C., Conomos, M.P., Stilp, A.M., Li, Z., Sofer, T., Szpiro, A.A., Chen, W., Brehm, J.M., Celedón, J.C. *et al.* (2016) Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am. J. Hum. Genet.*, **98**, 653–666.
59. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics*, **26**, 2190–2191.
60. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science.*, **296**, 2225–2229.
61. 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. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
62. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
63. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R. *et al.* (2010) A method and server for predicting damaging missense mutations. *Nat. Methods.*, **7**, 248–249.
64. Kumar, P., Henikoff, S. and Ng, P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.*, **4**, 1073–1081.
65. Venselaar, H., Te Beek, T.A., Kuipers, R.K., Hekkelman, M.L. and Vriend, G. (2010) Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics*, **11**, 548.
66. Sun, Z., Yu, W., Sanz Navarro, M., Sweat, M., Eliason, S., Sharp, T., Liu, H., Seidel, K., Zhang, L., Moreno, M. *et al.* (2016) Sox2 and Lef-1 interact with Pitx2 to regulate incisor development and stem cell renewal. *Development*, **143**, 4115–4126.
67. Gregorieff, A. and Clevers, H. (2015) In situ hybridization to identify gut stem cells. *Curr. Protoc. Stem Cell Biol.*, **34**, 2F.1.1–2F.1.11.