The *PAX1* locus at 20p11 is a potential genetic modifier for bilateral cleft lip

Sarah W. Curtis,[1] Daniel Chang,[1] Myoung Keun Lee,[2] John R. Shaffer,[3] Karlijne Indencleef,[4,5] Michael P. Epstein,[1] David J. Cutler,[1] Jeffrey C. Murray,[6] Eleanor Feingold,[3,7] Terri H. Beaty,[8] Peter Claes,[4,5,9] Seth M. Weinberg,[2] Mary L. Marazita,[2,3] Jenna C. Carlson,[3,7] and Elizabeth J. Leslie[1,*]

Summary

Nonsyndromic orofacial clefts (OFCs) are a common birth defect and are phenotypically heterogenous in the structure affected by the cleft—cleft lip (CL) and cleft lip and palate (CLP)—as well as other features, such as the severity of the cleft. Here, we focus on bilateral and unilateral clefts as one dimension of OFC severity, because the genetic architecture of these subtypes is not well understood. We tested for subtype-specific genetic associations in 44 bilateral CL (BCL) individuals, 434 unilateral CL (UCL) individuals, 530 bilateral CLP individuals (BCLP), 1,123 unilateral CLP (UCLP) individuals, and unrelated control individuals (N = 1,626), using a mixed-model approach. While no novel loci were found, the genetic architecture of UCL was distinct compared to BCL, with 44.03% of suggestive loci having different effects between the two subtypes. To further understand the subtype-specific genetic risk factors, we performed a genome-wide scan for modifiers and found a significant modifier locus on 20p11 (p = 7.53×10^{-9}), 300 kb downstream of *PAX1*, that associated with higher odds of BCL versus UCL and replicated in an independent cohort (p = 0.0018) with no effect in BCLP (p > 0.05). We further found that this locus was associated with normal human nasal shape. Taken together, these results suggest bilateral and unilateral clefts may have different genetic architectures. Moreover, our results suggest BCL, the rarest form of OFC, may be genetically distinct from the other OFC subtypes. This expands our understanding of modifiers for OFC subtypes and further elucidates the genetic mechanisms behind the phenotypic heterogeneity in OFCs.

Introduction

Orofacial clefts (OFCs) are common, complex birth defects (MIM: 608864). Affecting 1 in 700 births worldwide, they are caused when one or more of the developmental programs during the first 7 weeks of pregnancy that determine the form the face do not occur properly.[1] While some OFCs present in conjunction with other congenital abnormalities, a majority of OFCs are classified as isolated, nonsyndromic OFCs (nsOFCs), which are caused by a complex combination of genetic and environmental factors and have been the focus of numerous genome-wide association studies (GWASs).[2–13] OFCs also have striking phenotypic heterogeneity. OFCs are typically categorized into three subtypes: cleft lip only (CL), cleft lip and palate (CLP), and cleft palate only (CP), where CL includes clefts confined to the lip and primary palate, CLP includes clefts that affect the lip and extend into the secondary palate (or roof of the mouth), and CP affects the secondary palate only. CL and CLP are often combined into a more general category of cleft lip with or without cleft palate (CL/P) based on the shared defect of the primary palate. OFCs affecting the primary palate can also be further subdivided based on morphological details to capture severity, including the laterality (unilateral or bilateral), the side of unilateral clefts (left or right), or the completeness of the cleft.

Population-based studies estimating recurrence risks have focused on different classifications of OFCs, and the resulting estimates can inform genetic models and the design of association studies. For example, among CLP individuals, there is no difference in the risk of either CL or CLP among their first-degree relatives; this suggests a shared genetic etiology,[14,15] contributing to the rationale of studying CL/P in genetic association studies, and many of the known risk loci show similar effects between CL and CLP.[2–4,16,17] However, less is known about severity in CL and CLP or if there is a separate genetic component to CL severity. Recurrence risk estimates based on severity are limited by sample size and have yielded mixed results. Semiquantitative measures of completeness showed no effect of severity on estimated recurrence risks.[15] However, the recurrence risk for bilateral clefts is higher than for unilateral clefts, indicating this more severe cleft type tends to recur more often in family members,[14,18] suggesting a potentially distinct genetic etiology. Previous studies examining genetic factors associated with bilateral versus unilateral clefts have been limited to targeted sequencing of a few selected candidate loci,[17] although this work has suggested the presence of a genetic

---

contribution to the different subtypes of CL. Therefore, we set out to perform a GWAS to determine if there are additional genetic variants that are either associated with cleft severity or are genetic modifiers for the cleft subtype that forms by focusing on bilateral and unilateral clefting in CL and CLP individuals.

Material and methods

Sample collection and SNP quality control

This study used samples from the Pittsburgh Orofacial Cleft (POFC) Study. The details of the sample collection and genotype quality control (QC) have been described previously.[2,19–21] Briefly, these samples came from 18 sites in 13 countries, including in the continental United States, Guatemala, Argentina, Colombia, Puerto Rico, China, Philippines, Denmark, Turkey, and Spain. All sites had institutional review board (IRB) approval, both locally and at the University of Pittsburgh or University of Iowa, with written informed consent for genomic studies and data sharing. The original study recruited individuals with OFCs, their unaffected relatives, and unrelated control individuals (individuals with no known family history of OFCs or other craniofacial anomalies; N = 1,626). For the current study, affected individuals were classified as either having a bilateral cleft lip (BCL; N = 44), a bilateral cleft lip and palate (BCLP; N = 530), a unilateral cleft lip (UCL; N = 434), or a unilateral cleft lip and palate (UCLP; N = 1,123). Although this sample was not recruited with a population-based approach, the relative frequencies of these cleft types in the POFC study are consistent with epidemiological reports of subtypes.[22] Each cleft subtype was present in each ancestry group, as defined by principal components (PCs) of genetic markers (Table S1; Figure S1). Subjects where the specific subtype of cleft was not known were excluded from this study. Related, affected individuals were retained in this study, and a genetic relatedness matrix (GRM) was used to adjust for relationships within and across families (see below).

Samples were genotyped for approximately 580,000 single-nucleotide polymorphic (SNP) markers from the Illumina Human-Core+Exome array, of which approximately 539,000 SNPs passed quality control filters recommended by the Center for Inherited Disease Research (CIDR) and the Genetics Coordinating Center (GCC) at the University of Washington.[2] These data were then phased with SHAPEIT2[23] and imputed with IMPUTE2[24] to the 1000 Genomes Project phase 3 release (September 2014) reference panel. The most-likely imputed genotypes were selected for statistical analysis if the highest probability (r^2) > 0.9. SNP markers showing deviation from Hardy-Weinberg equilibrium in European control individuals, a minor allele frequency or MAF < 5%, or imputation info scores < 0.5 were filtered out of all subsequent analyses. The information for the genotyped markers was retained after imputation, and the imputed values for these variants were only used to assess concordance. A GRM was calculated from a set of linkage disequilibrium (LD)-pruned genotyped SNPs as defined by Genome-wide Complex Trait Analysis (GCTA) using the package SNPRelate.[25]

Statistical analyses
Subtype-specific GWASs
Single-subtype genome-wide tests were done by comparing individuals from each subtype to a group of unrelated control individuals to test for genetic variants associated with each cleft subtype. The association between every genetic variant and laterality type was tested using the generalized linear mixed model (GMMAT)[26] as implemented in the GENESIS software package.[27] Sex and the estimated GRM were adjusted for under the null model to account for both population substructure and relatedness. The control group was the same for all analyses. SNPs with association p values less than 5×10^{-8} were considered genome-wide significant, and those with p values less than 1×10^{-5} were considered "suggestive" and were used for downstream enrichment and comparison analyses. The unadjusted odds ratio (OR) for each SNP was estimated for the additive model using the minor allele frequency in affected individuals compared to control individuals.[28,29] Regional association plots were made with LocusZoom, where the LD blocks and recombination rates were estimated from European populations.[30]

Modifier GWASs
We identified genetic modifiers (genetic variants that are associated with phenotypic heterogeneity or expressivity) using case-case group comparisons by directly comparing allele frequencies at each SNP between unilateral and bilateral cleft individuals. Thus, this approach has high power to identify genetic risk factors that differ between two subtypes but no power to find factors important in both groups (i.e., SNPs detected in previous GWASs of CL, CLP, or the combined CL/P group).[25] Therefore, this test has the potential to identify new loci for which there is an effect in only one subtype or where the effects are different between two groups. Such loci may be masked in an overall scan when the two groups are combined. We performed modifier analyses for severity separately in the CL and CLP subtypes (UCL versus BCL and UCLP versus BCLP) and combined as CL/P (UCL/P versus BCL/P). Similar to the subtype-specific analyses above, these tests were done using GMMAT[26] as implemented in GENESIS,[27] adjusting for sex and the GRM to account for both population substructure and relatedness. The OR for each SNP was estimated using the minor allele frequency in bilateral cleft individuals compared to unilateral cleft individuals.[28,29] Regional association plots were made with LocusZoom.[30]

Comparisons between CL and CLP analyses
The estimated ORs for suggestive SNPs (i.e., those with $p < 1 \times 10^{-5}$) in the subtype-specific analyses were compared both within a single severity subtype across cleft type (i.e., BCL versus BCLP) and across severity types within a single cleft type (e.g., UCL versus BCL). To compare whether the SNPs associated with individual subtypes were novel compared to what has already been reported in previous GWASs of CL, CLP, or CL/P, the SNPs in these analyses within 50 kb of previously associated risk SNPs[2,5,6,21] were also identified. A similar approach was done for the modifier analysis, and the suggestive loci from either the CL or CLP modifier analyses were compared to see if they either had overlapping 95% confidence intervals (CIs) or gave estimated effects in the same direction. A chi-square test was used to determine if the number of SNPs that both had similar CIs and were previously reported in the literature overlapped more than expected by chance.

Replication cohort
To replicate the statistically significant results from our modifier analysis, data from the GENEVA consortium were used, which was described previously.[2,4,21] Briefly, this cohort recruited affected individual-parent trios, where the affected individual had an oral cleft. The samples were genotyped for approximately 589,000 SNPs using the Illumina Human610-Quadv.1_B

BeadChip, phased using SHAPEIT, and imputed to the 1000 Genomes Project phase I (June 2011) reference panel using IMPUTE2. Imputed genotype probabilities were converted to most-likely genotype calls with GTOOL. This dataset was subsequently filtered to only include common SNPs with a minor allele frequency > 5%. A subset of individuals was included in both the POFC study and the GENEVA consortium, and these were removed from the replication analysis so that the two groups would be independent. Only the cases from this GENEVA cohort were selected, and they were classified as BCL (N = 28), UCL (N = 326), BCLP (N = 301), and UCLP (N = 678). PCs of ancestry were calculated using PLINK (v1.9),[31] and a majority of the cohort was of Asian (71.6%) or European (26.3%) ancestry (Figure S2). Because the replication cohort did not include related individuals, the modifier analyses (comparing BCL versus UCL and BCLP versus UCLP) were conducted using logistic regression models in PLINK (v1.9), with sex and the first four PCs as quantitative covariates, instead of the mixed-model approach that adjusts for relatedness implemented in GENESIS. Because of the small sample sizes in the replication cohort and the differences in genotyping arrays and imputation panels, only regions that were significant in the original modifier analysis were tested in this replication strategy. p values less than a Bonferroni correction for the number of SNPs in the region (0.05/ the number of SNPs tested) were considered to be evidence of significant replication.

Association with normal facial variation

The genome-wide significant modifier locus was further examined in relation to normal facial variation by reviewing the association results of SNPs in this locus in a GWAS meta-analysis of facial shape in two large cohorts (n = 8,246) from the US (MetaUS) and UK (MetaUK).[32] To analyze normal facial variation, the original study used a data-driven global-to-local facial segmentation approach, and a multivariate GWAS was then performed in each of the resulting 63 hierarchically arranged facial segments. More information on the analysis pipeline and the cohorts can be found in the initial study.[33]

Epigenomic context of results

Topologically associated domains (TADs) were defined for significantly associated loci using the H1-ESC cell line in 3D Genome Browser.[34] Functional enrichment was tested by first annotating all of the SNPs to the craniofacial functional regions defined by Wilderman et al.[35] for human embryos at CS13, CS14, CS15, CS17, and CS20 (4.5–8 weeks post conception). Enrichment tests were done using a chi-square test with the top SNPs (p < 1 × 10^{-3}) for both modifier analyses and each subtype analysis, and estimated ORs and their 95% CIs were calculated.

Results

Subtype-specific analysis

We performed a subtype-specific genome-wide analysis for BCL, UCL, BCLP, and UCLP individuals by comparing affected individuals of each subtype to unaffected control individuals. This approach can detect variants associated with increased risk for an OFC in general but also has the potential to identify variants that increase the risk for one or more subtypes of OFC. A single SNP in chromosome 3q28 achieved genome-wide significance in the analysis of BCL (rs72439195; p = 3.69 × 10^{-8}), and 90 regions yielded suggestive evidence, most of which have not been previously implicated in OFC formation. However, some of these regions, like 14q32.33 (lead SNP: rs61996057; p = 8.07 × 10^{-8}; Figures S3A, S4A, and S5; Table S2), have been implicated in syndromes with facial dysmorphisms.[36–38] In the analysis of UCL, two loci reached genome-wide significance (8q24 and 1q32), both of which are recognized genetic risk loci for CL/P (Figures S3B and S4B; Table S3).[2,4,7–10] Among the 21 suggestive loci, 17 have not been previously associated with OFCs, which may reflect a lack of GWASs focused specifically on CL. Some of these loci, such as 2q13 (lead SNP: rs6542368; p = 1.06 × 10^{-7}; Figure S6), are plausible candidates for craniofacial dysmorphism.[39] Both BCLP and UCLP had multiple recognized genes/regions, including 8q24 and 17p13, reach genome-wide significance (Figures S3C, S3D, S4C, and S4D; Tables S4 and S5), and 35 and 41 loci reach suggestive significance, respectively, in this analysis.[2,4,5,8,10,17]

Because of the apparent differences in suggestive and significant loci in the subtype-specific GWASs, we wanted to characterize similarity or dissimilarity of the overall genetic architectures of UCL, UCLP, BCL, and BCLP. Therefore, we performed pairwise analyses comparing the ORs and 95% CIs for SNPs identified as suggestive in the GWAS for each subtype being compared. In the comparison of BCL and UCL SNPs, we found a striking difference in estimated ORs in which 44.03% of 738 SNPs did not have overlapping CIs. A majority of these SNPs originating from the BCL analysis had an OR near 1 in the UCL analysis (Figure 1), indicating substantial differences in the genetic architecture of BCL, the more severe group. This was also seen, although to lesser degree, when the BCL subtype was compared to BCLP, where the 95% CIs for the estimated ORs did not overlap for 34.1% of 1,178 suggestive SNPs (Figure S7). In contrast, BCLP and UCLP were quite similar, with 94.7% of their 1,093 SNPs showing overlapping OR CIs (Figure 1). We also found SNPs with different effects in the subtype-specific analyses were less likely to have been previously reported in analyses of the combined group CL/P, suggesting these may be masked in traditional analyses that combine subtypes (Figure 1). For example, in the BCL-UCL comparison, 26.8% of SNPs with overlapping estimated effect sizes were recognized CL/P risk SNPs, indicating these SNPs may predispose to OFC risk but have no effect on specific subtypes. However, only 1.8% of SNPs differing in their effect sizes were previously reported, significantly less than expected by chance alone (p = 2.41 × 10^{-20}). This pattern held for all comparison groups (Table S6). We reasoned that SNPs predisposing to any type of bilateral cleft could be identified by first selecting SNPs that had non-overlapping CIs between BCL and UCL that also had overlapping CIs between BCL and BCLP. However, only 4 SNPs met these criteria, and all of them also showed nominal significance in UCLP and had overlapping CIs. We employed the same strategy to identify SNPs predisposing to any type of
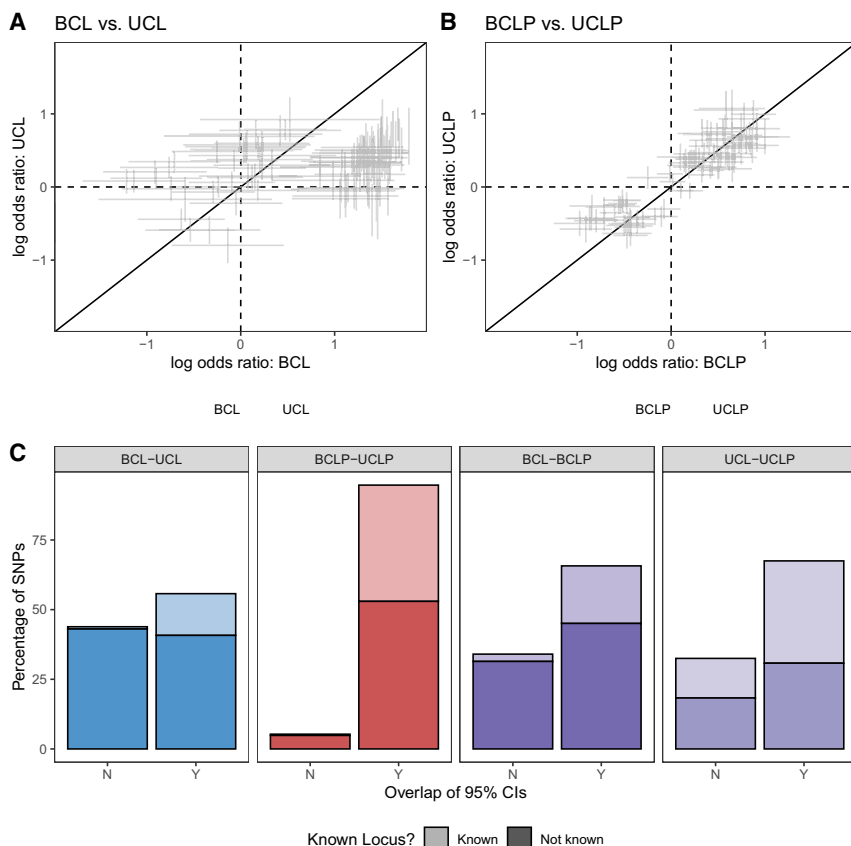
Figure 1.   Subtype-specific analyses
The log OR for SNPs that were suggestive (p < 1 × 10^{-5}) or significant (p < 5 × 10^{-8}) in the subtype-specific case-control analyses were compared between BCL (dark blue points) and UCL (light blue points) (A), and BCLP (dark red points) and UCLP (light red points) (B), and were classified in (C) by whether the 95% confidence interval for the OR overlapped and whether the variant was identified in previous GWAS (Known) or not (Not known).

unilateral cleft but were similarly unsuccessful, supporting the notion that subtype-specific risk factors are not shared between CL and CLP in this sample.
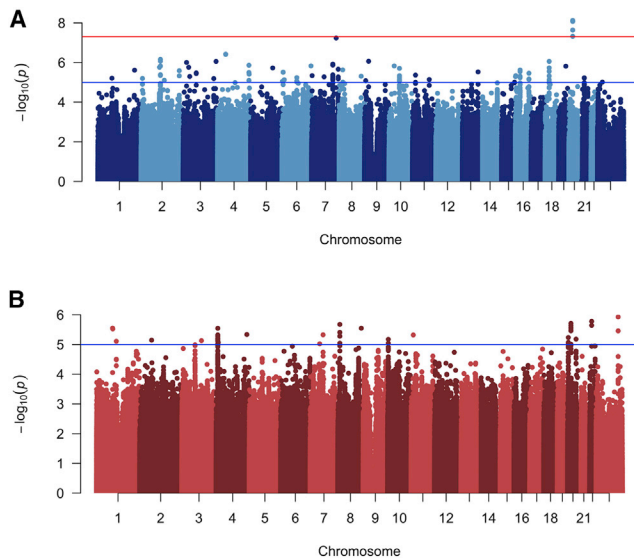
Modifier analysis

To disentangle the effects of SNPs on specific subtypes from more general effects on OFC risk, we performed a genome-wide bilateral versus unilateral modifier analysis in CL and CLP individuals. Because this is a case-to-case group comparison, this analysis would not be able to detect variants generally important for both CL or CLP risk but would detect variants important for the formation of one severity subtype compared to the other. In the modifier analysis of CL, one locus on chromosome 20p11 reached genome-wide significance (lead SNP: rs143865354; p = 7.53 × 10^{-9}), and 47 other SNPS yielded suggestive significance (Figure 2A; Figure S8A; Table S7). In the modifier analysis for CLP, no loci reached genome-wide significance, but 19 loci yielded suggestive significance (Figure 2B; Figure S8B; Table S8). Interestingly, when CL and CLP were combined (as is typical in genetic analyses of OFCs), no loci reached genome-wide significance, and only 3 loci gave suggestive significance (Figure S9; Table S9), raising the possibility that these modifiers may not be shared between CL and CLP.

The associated SNPs on 20p11 lie within *LINC01432* and are within the same topologically associated domain as *PAX1* (MIM: 167411) (Figure 3A; Figure S10). This locus

was not significant (p > 0.05) in the modifier analysis of CLP (Figure 3B). Additionally, when the OR for the lead SNP in this region was compared between CL and CLP individuals, the direction of effect was not consistent (with either a 95% CI or a 99% CI; Figure 3C). We replicated the 20p11 region in an independent sample of 28 BCL individuals, 329 UCL individuals, 306 BCLP individuals, and 685 UCLP individuals. In this 20p11 region, there were 8 SNPs passing filtering in the CL modifier analysis. While none of these SNPs were the same as those in the original analysis, one SNP (rs28970569) was also a significant modifier in the replication cohort (OR = 3.83, 95% CI = 1.64–8.95, p = 0.0018; Table S10). In the CLP modifier analysis, 9 SNPs passed our filters, but none of these were significant modifiers, consistent with the results for 20p11 in our discovery sample (p > 0.05; Table S11). Additionally, we wanted to determine the extent to which the genetic modifiers in CL were similar to the genetic modifiers in the CLP genome-wide analysis. To test this, we compared SNPs that were suggestive (p < 1 × 10^{-5}) in either the CL or CLP modifier analyses. Notably, there was no overlap between the list of suggestive SNPs in CL and the list of SNPs suggestive in CLP. Moreover, the estimated ORs were not positively correlated, all of the suggestive SNPs in the analysis of CL had no effect in CLP and vice versa (Figure 4), and a majority of the SNPs in each analysis were not near regions previously associated with CL/P (Table S12). Cumulatively, these results suggest the 20p11 modifier for bilateral versus unilateral OFCs is specific to CL.

Although the 20p11 locus had not previously been associated with risk to OFCs, it has been associated with variation in normal facial structures. Therefore, we next investigated whether the BCL modifier SNPs were also associated with normal facial variation, as that could give insights into how these SNPs might influence cleft severity. We found that rs6036034, a SNP in the 20p11 region in LD with rs143865354 (R^2 = 0.522; p = 4.75 × 10^{-8} in BCL

Figure 2. Manhattan plots for genome-wide modifier scans
Manhattan plots of $-\log_{10}(p$ values) from the bilateral versus unilateral modifier analysis in participants with (A) cleft lip, and (B) cleft lip and palate. Lines indicate suggestive (blue) and genome-wide (red) thresholds for statistical significance. The genomic inflation factors were 0.96 and 1.01, respectively.

versus UCL) was associated with normal variation in nose morphology (p = 2.63×10^{-11}), specifically projection of the nasal tip and columella and breadth of the nasal alae (Figure 5). These are the same structures disrupted by CL and are derived from the lateral nasal processes where *PAX1* is expressed.[40] Moreover, rs143865354 shows modest evidence of being an expression quantitative trait locus (eQTL) for *PAX1* in skin (p = 2.9×10^{-5}) in GTEx.

Functional enrichment
We were also interested in testing whether differences in genetic architecture in BCL, UCL, BCLP, and UCLP at the SNP level were also reflected in functional elements involved in facial development. Therefore, we tested whether SNPs associated with each subtype were enriched in similar functional regions defined by epigenetic marks in human embryonic craniofacial tissues.[35] For some elements, the apparent enrichment or depletion was consistent across subtypes. For example, BCL, UCL, BCLP, and UCLP SNPs were similarly depleted in heterochromatin regions, and most were enriched in regions of strong transcription. However, there were some regions showing opposite enrichments in the different subtypes. For example, zinc finger repeat regions were enriched in both BCLP and UCLP but were depleted in BCL (Figure 6). Interestingly, the severity modifiers for both CL and CLP were depleted in regions of weak transcription and enriched in regions of low activity. Some of the suggestive modifier loci for CLP were enriched in bivalent transcription start sites, but none of the putative modifiers for risk to CL were enriched in functional domain. These enrichment/depletions were consistent throughout craniofacial devel-

opment (4.5–8 weeks post conception; Figure S11; Table S13). These observations, while not definitive, lend some support to the idea that although at the SNP level, the genetic underpinnings for cleft subtypes are distinct, this may not extend entirely to gross differences in functional element enrichments. Deciphering the true underlying mechanism(s) resulting in bilateral and unilateral CL and CLP will require a locus-by-locus investigation.

Discussion

While there have been many studies identifying genetic variants that influence overall risk to CL/P and CP only, the genetic underpinnings of specific phenotypic subtypes of CL are less studied. This report furthers our understanding of genetic variants associated with specific subtypes of OFC: BCL, UCL, BCLP, and UCLP. We used a modifier analysis, which provides more power to find genetic loci differing between two groups, and found one locus on 20p11 that replicated in an independent cohort as significantly associated with the formation of a BCL over a UCL. The associated SNPs were located in several long noncoding RNAs and within the same TAD (300 kb downstream) as the *PAX1* gene. While *PAX1* has not been associated with OFC like its paralog *PAX9*,[41] they both are transcription factors with similar DNA-binding domains regulating chondrocyte differentiation and the formation of invertebrate discs, and knockout mouse models show skeletal abnormalities.[42–44] There is also evidence that *PAX1* is upregulated by *SHH* and, in turn, upregulates *SOX5* and *BMP4*.[43–45] There is only limited literature describing *PAX1* expression in the developing face,[40] and it has not been previously associated with risk to nonsyndromic OFCs, but *PAX1* is in a pathway with other genes known to be associated with nonsyndromic OFCs.[46–51] Additionally, recent studies have shown mutations in *PAX1* cause otofaciocervical syndrome (OTFCS [MIM: 615560]), which presents with facial dysmorphisms,[52,53] and studies of normal facial variation have found this locus has also been associated with nasal width (the distance between left and right cartilaginous nasal ala) in people of European descent,[54] Latin American descent,[55] and Korean descent.[56] The link between SNPs at the *PAX1* locus and normal facial shape was further substantiated in our analysis, with effects observed in the nasal tip, columella, and alae. These anatomical structures are derived from the lateral and medial nasal processes in the embryo, which form the primary palate. Thus, it is biologically plausible that *PAX1* could affect the development of specific types of craniofacial abnormalities; however, more work is needed to investigate the underlying mechanisms.
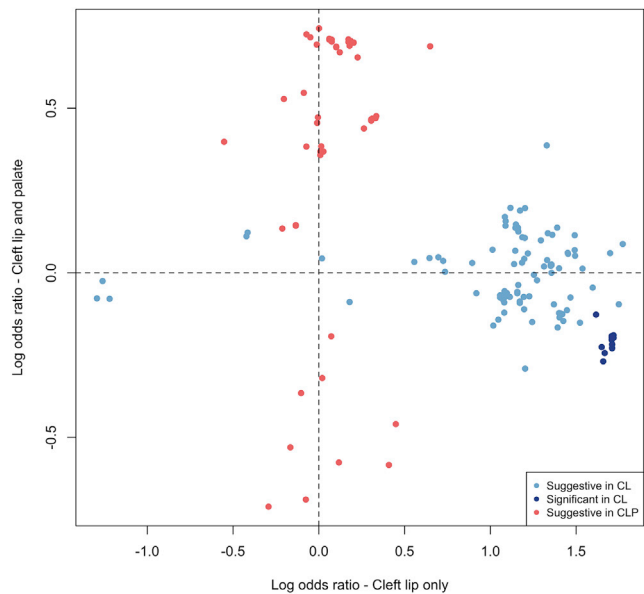
While 20p11 was the only genome-wide significant modifier found in this study, this may partly be due to limited sample size in some of the OFC subtypes. It is important to note that when a modifier analysis was conducted on all combined CL and CLP cases, fewer loci reached even suggestive significance, suggesting CL and
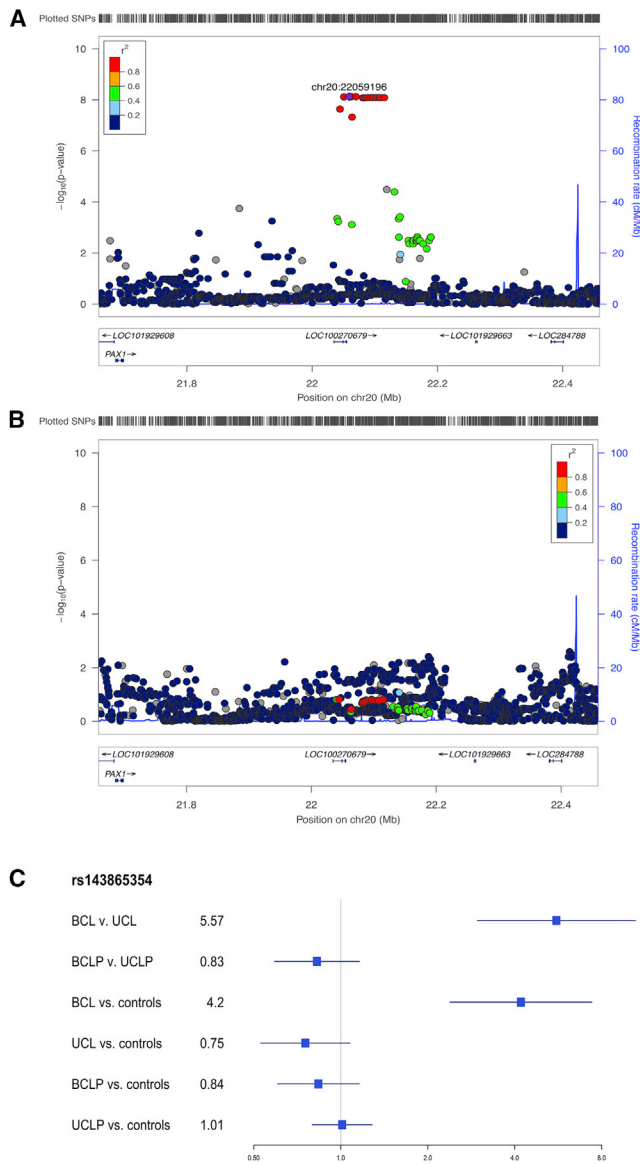
Figure 3. 20p11 associated with bilateral CL only
(A and B) Regional association plots showing $-\log_{10}$(p values) for the genome-wide significant peaks at 20p11 in the modifier analysis in (A) cleft lip and (B) cleft lip and palate. Plots were generated using LocusZoom.[30] The recombination overlay (blue line, right y axis) indicates the boundaries of the LD block. Points are color coded according to pairwise LD (r^2) with the index SNP.
(C) The OR for rs143865354 at the 20p11 locus in each of the modifier and subtype-specific analyses.

Figure 4. Distinct modifier SNPs in CL compared to CLP
The log ORs for 188 SNPs that were suggestive (p < 1×10^{-5}) or significant (p < 5×10^{-8}) in the modifier analysis in CL (blue points) or CLP (red points) were compared. No SNPs were genome-wide significant in CLP. No SNPs were significant or suggestive in both CL and CLP.

bilateral CL/P.[17] In our study, no SNPs in *IRF6* reached suggestive significance. Our study was larger than the previous study (2,339 cases versus 1,001 cases); therefore, this difference may reflect effects of modifiers for cleft subtypes in regions of genome not recognized by previous GWASs of OFCs. This is not surprising, given OFC subtypes are typically combined for GWASs, which maximizes statistical power to detect loci associated with overall risk but would mask loci with different effects in subtypes.

We also conducted analyses comparing each subtype to unrelated control individuals. This analysis should find loci associated with either overall risk or one particular cleft subtype but would have less statistical power to detect loci that differ between two subtypes. Most loci achieving genome-wide significance in these analyses were those already recognized to be associated with risk to OFCs.[2,5,6,21] There were, however, some loci yielding suggestive evidence of association for several of the subtype-specific analyses not previously reported but that could be in the causal pathway for syndromes with facial dysmorphisms. For example, SNPs in 14q32.33 gave suggestive evidence of association for BCL, with a distinct effect only seen in BCL, and 2q13 yielded suggestive evidence of association for UCL. Microdeletions in both of these regions have been associated with syndromes that include facial dysmorphisms.[36–39] The 14q32.33 also contains *JAG2*, which is part of the Notch signaling pathway and is important for craniofacial development.[58–60]

Overall, our analyses demonstrated that BCL was most distinct from the other three subtypes analyzed and that these modifiers were not shared between CL and CLP. We

CLP may have distinct modifiers. Consistent with this, the suggestive modifiers for risk in CL and CLP showed no overlap in estimated effect on risk. This suggests that the lack of overlap is not entirely due to a difference in sample size but that instead there is a biological difference in the genetics of laterality in CL compared to CLP.

This study tested for severity modifiers at a genome-wide level, but we previously tested for modifiers in 13 recognized GWAS regions known to be associated with OFCs[57] and found SNPs in *IRF6* (MIM: 607199) were associated with the formation of a unilateral CL/P compared to
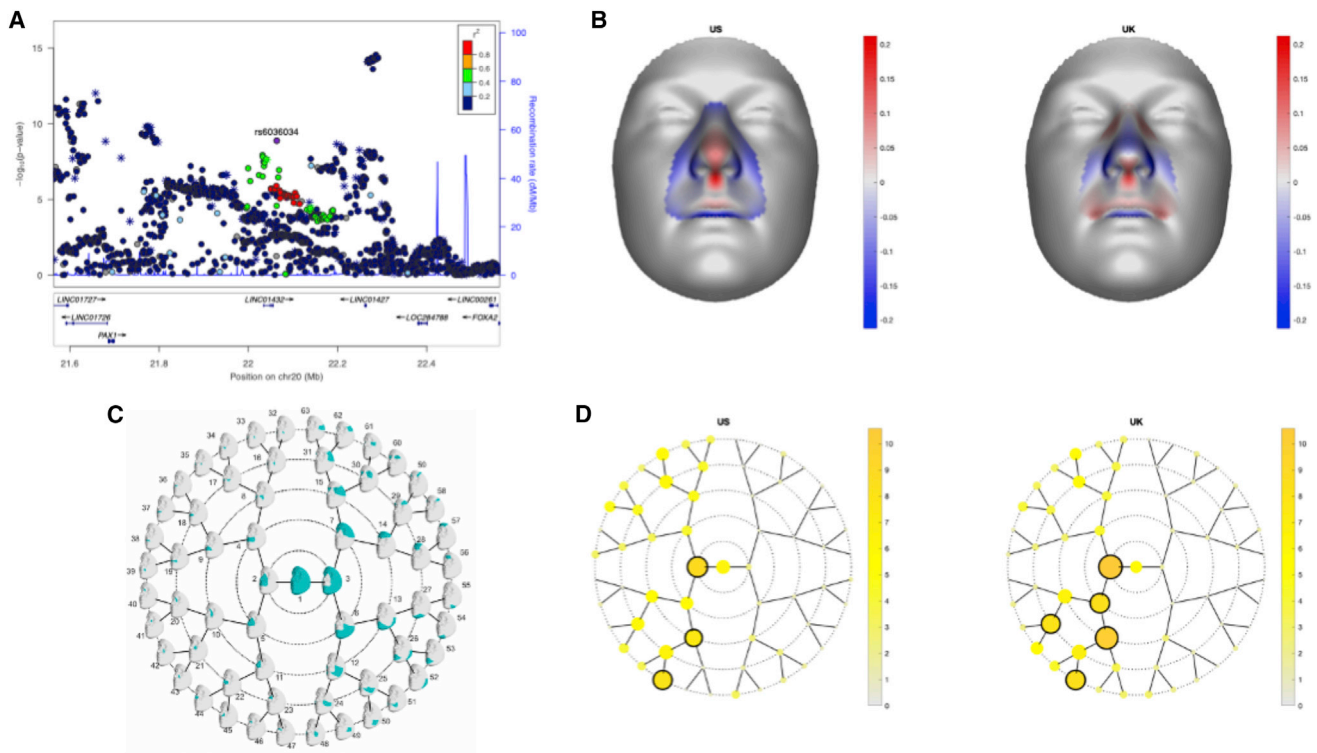
Figure 5. 20p11 is associated with normal facial variation
(A) LocusZoom plots for the association of normal facial variation and rs6036034. Points are color-coded based on linkage disequilibrium (R^2) in Europeans. The asterisks represent genotyped SNPs, and the circles represent imputed SNPs.
(B) The normal displacement (displacement in the direction locally normal to the facial surface) in each quasi-landmark of the facial segment reaching the lowest p value in MetaUS and MetaUK going from the minor to the major allele SNP variant. Blue, inward depression; red, outward protrusion.
(C) Global-to-local facial segmentation plot that shows the 63 facial segments represented in teal obtained using hierarchical spectral clustering.
(D) The $-\log10$(p value) of the meta-analysis p values per facial segment in MetaUS and MetaUK. Black-encircled facial segments have reached a genome-wide p value (p = 5.00×10^{-8}).

found that the associated SNPs in all four OFC subtypes were enriched in regions associated with transcription and depleted in heterochromatin regions. This was expected because nonsyndromic OFCs form from the disruption of one of the processes involved in facial development, and thus variants associated with any subtype OFC should be enriched in regions active during facial development. It is also consistent with the study defining the functional regions, which showed enrichment in active states for SNPs involved in overall OFC risk.[35] Importantly, there were some differences in functional enrichment by subtype. For example, SNPs associated with BCLP and UCLP were enriched in zinc finger repeat regions; however, SNPs showing some evidence of association with BCL were depleted in this same region. This further emphasizes the possibility for a distinct genetic architecture associated with risk to BCL. Additionally, the modifiers for both CL and CLP were depleted in regions associated with active transcription and strongly enriched in regions of low activity. This result is somewhat surprising, given it is the opposite of what would be expected for an analysis involving craniofacial development. However, the biological mechanism by which modifiers could affect a phenotype is not known.

Therefore, this highlights the need for more studies that test how modifiers mechanistically act.

The findings from this study should also be considered in the context of its limitations. Many of the subtypes of clefting, particularly BCL, had small sample sizes. Limits of small sample sizes make it likely other subtype-specific genetic loci and modifiers may exist and we are unable to detect them in this statistical analysis. Additionally, because the subtype-specific analyses were not independent due to the shared controls group and the related individuals, a formal test for heterogeneity could not be conducted. The CIs in our analyses are less precise in the comparison involving smaller groups, and so it is likely that the estimates for different genetic effects are conservative and that the genetic heterogeneity between these subtypes is larger than we see with our current population. We were also unable to test for heterogeneity across ancestry groups while testing for subtype-specific genetic risk loci and severity modifiers. This cohort is multiethnic, including people of European, Asian, and Latin American ancestry, and previous studies have shown ancestry-specific association with risk for OFCs.[2] Studies with larger sample sizes for these clefting subtypes could lead to the discovery of more
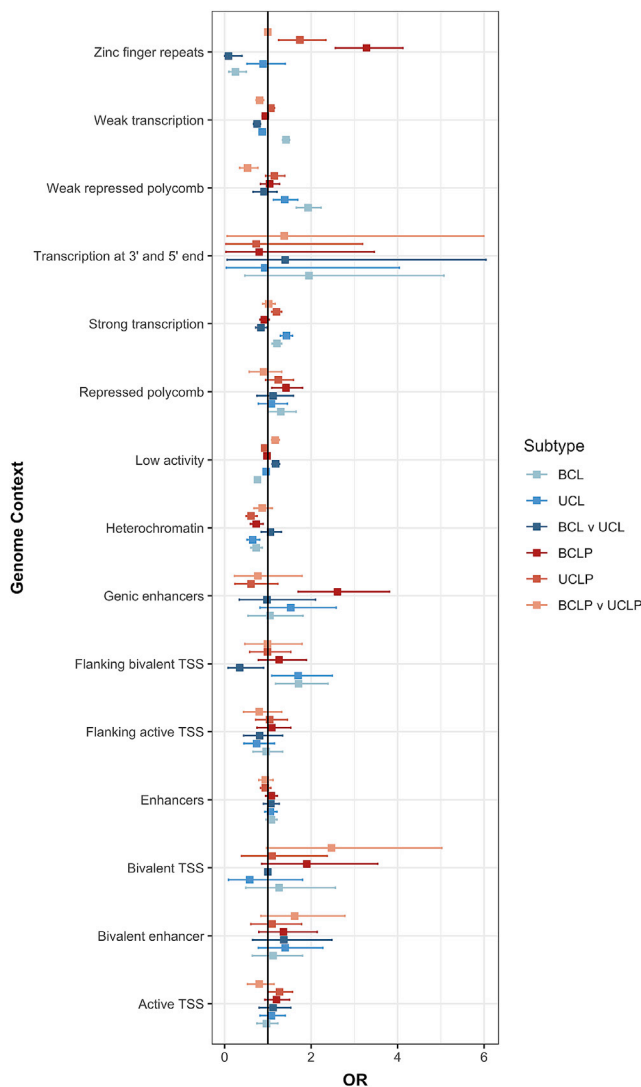
Figure 6. Functional enrichment
Enrichment of the top SNPs associated in the CL modifier analysis, CLP modifier analysis, and each subtype analysis (p < 1 × 10⁻³) were tested in each functional region defined during craniofacial development (CS15). Odds ratios and 95% confidence intervals are shown for each subtype analysis.

associated genetic loci and test for differences in associated loci between different ancestry populations.

In summary, we conducted a genome-wide scan for severity modifiers in a case-case and case-control design focused on nonsyndromic CL and CLP and found a significant modifier in 20p11 downstream of *PAX1* associated with increased risk for BCL over UCL. We also showed these modifiers for CL and CLP were distinct, with the modifiers of one cleft subtype have little to no genetic effect in the other subtypes. Furthermore, in the subtype-specific GWASs, we found several suggestive loci that had not been identified in previous GWASs that combined cleft subtypes. We also found loci associated with BCL were the most distinct from those associated with other cleft subtypes, suggesting the etiology of this rarest subtype of cleft to be unique. Overall, this study expands our understand-

ing of the genetic underpinnings of the genetic and phenotypic heterogeneity of OFCs and suggests new areas of research on cleft lip subtypes.

Data and code availability

Supplemental Information

Acknowledgments

Declaration of interests

The authors declare no competing interests.

Web resources

dbGaP, https://www.ncbi.nlm.nih.gov/gap/
GTOOL, https://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html
OMIM, https://www.omim.org/

References

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

6. Huang, L., Jia, Z., Shi, Y., Du, Q., Shi, J., Wang, Z., Mou, Y., Wang, Q., Zhang, B., Wang, Q., et al. (2019). Genetic factors define CPO and CLO subtypes of nonsyndromicorofacial cleft. PLoS Genet. *15*, e1008357.

7. Birnbaum, S., Ludwig, K.U., Reutter, H., Herms, S., de Assis, N.A., Diaz-Lacava, A., Barth, S., Lauster, C., Schmidt, G., Scheer, M., et al. (2009). IRF6 gene variants in Central European patients with non-syndromic cleft lip with or without cleft palate. Eur. J. Oral Sci. *117*, 766–769.

8. Birnbaum, S., Ludwig, K.U., Reutter, H., Herms, S., Steffens, M., Rubini, M., Baluardo, C., Ferrian, M., Almeida de Assis, N., Alblas, M.A., et al. (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat. Genet. *41*, 473–477.

9. 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.

10. Nikopensius, T., Ambrozaityte, L., Ludwig, K.U., Birnbaum, S., Jagomägi, T., Saag, M., Matuleviciene, A., Linkeviciene, L., Herms, S., Knapp, M., et al. (2009). Replication of novel susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24 in Estonian and Lithuanian patients. Am. J. Med. Genet. A. *149A*, 2551–2553.

11. Bureau, A., Parker, M.M., Ruczinski, I., Taub, M.A., Marazita, M.L., Murray, J.C., Mangold, E., Noethen, M.M., Ludwig, K.U., Hetmanski, J.B., et al. (2014). Whole exome sequencing of distant relatives in multiplex families implicates rare variants in candidate genes for oral clefts. Genetics *197*, 1039–1044.

12. Fu, J., Beaty, T.H., Scott, A.F., Hetmanski, J., Parker, M.M., Wilson, J.E., Marazita, M.L., Mangold, E., Albacha-Hejazi, H., Murray, J.C., et al. (2017). Whole exome association of rare deletions in multiplex oral cleft families. Genet. Epidemiol. *41*, 61–69.

13. Ludwig, K.U., Böhmer, A.C., Bowes, J., Nikolic, M., Ishorst, N., Wyatt, N., Hammond, N.L., Gölz, L., Thieme, F., Barth, S., et al. (2017). Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip ± cleft palate and cleft palate only. Hum. Mol. Genet. *26*, 829–842.

14. Grosen, D., Chevrier, C., Skytthe, A., Bille, C., Mølsted, K., Sivertsen, A., Murray, J.C., and Christensen, K. (2010). A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance. J. Med. Genet. *47*, 162–168.

15. Sivertsen, A., Wilcox, A.J., Skjaerven, R., Vindenes, H.A., Abyholm, F., Harville, E., and Lie, R.T. (2008). Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives. BMJ *336*, 432–434.

16. Leslie, E.J., Carlson, J.C., Shaffer, J.R., Buxó, C.J., Castilla, E.E., Christensen, K., Deleyiannis, F.W.B., Field, L.L., Hecht, J.T., Moreno, L., et al. (2017). Association studies of low-frequency coding variants in nonsyndromic cleft lip with or without cleft palate. Am. J. Med. Genet. A. *173*, 1531–1538.

17. Carlson, J.C., Taub, M.A., Feingold, E., Beaty, T.H., Murray, J.C., Marazita, M.L., and Leslie, E.J. (2017). Identifying Genetic Sources of Phenotypic Heterogeneity in Orofacial Clefts by Targeted Sequencing. Birth Defects Res. *109*, 1030–1038.

18. Mitchell, L.E., and Risch, N. (1993). Correlates of genetic risk for non-syndromic cleft lip with or without cleft palate. Clin. Genet. *43*, 255–260.

19. 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.

20. Carlson, J.C., Standley, J., Petrin, A., Shaffer, J.R., Butali, A., Buxó, C.J., Castilla, E., Christensen, K., Deleyiannis, F.W., Hecht, J.T., et al. (2017). Identification of 16q21 as a modifier of nonsyndromic orofacial cleft phenotypes. Genet. Epidemiol. *41*, 887–897.

21. Carlson, J.C., Anand, D., Butali, A., Buxo, C.J., Christensen, K., Deleyiannis, F., Hecht, J.T., Moreno, L.M., Orioli, I.M., Padilla, C., et al. (2019). A systematic genetic analysis and visualization of phenotypic heterogeneity among orofacial cleft GWAS signals. Genet. Epidemiol. *43*, 704–716.

22. Gundlach, K.K., and Maus, C. (2006). Epidemiological studies on the frequency of clefts in Europe and world-wide. J. Craniomaxillofac. Surg. *34* (*Suppl 2*), 1–2.

23. Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods *10*, 5–6.

24. 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.

25. 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.

26. 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.

27. Gogarten, S.M., Sofer, T., Chen, H., Yu, C., Brody, J.A., Thornton, T.A., Rice, K.M., and Conomos, M.P. (2019). Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics *35*, 5346–5348.

28. Bland, J.M., and Altman, D.G. (2000). Statistics notes. The odds ratio. BMJ *320*, 1468.

29. Lloyd-Jones, L.R., Robinson, M.R., Yang, J., and Visscher, P.M. (2018). Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. Genetics *208*, 1397–1408.

30. 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.

31. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

32. White, J.D., Indencleef, K., Naqvi, S., Eller, R.J., Hoskens, H., Roosenboom, J., Lee, M.K., Li, J., Mohammed, J., Richmond, S., et al. (2021). Insights into the genetic architecture of the human face. Nat. Genet. *53*, 45–53.

33. White, J.D., Indencleef, K., Naqvi, S., Eller, R.J., Roosenboom, J., Lee, M.K., Li, J., Mohammed, J., Richmond, S., Quillen, E.E., et al. (2020). Insights into the genetic architecture of the human face. bioRxiv. https://doi.org/10.1101/2020.05.12.090555.

34. Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., et al. (2018). The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol. *19*, 151.

35. Wilderman, A., VanOudenhove, J., Kron, J., Noonan, J.P., and Cotney, J. (2018). High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. Cell Rep. *23*, 1581–1597.

36. Geier, C.B., Piller, A., Eibl, M.M., Ciznar, P., Ilencikova, D., and Wolf, H.M. (2017). Terminal 14q32.33 deletion as a novel cause of agammaglobulinemia. Clin. Immunol. *183*, 41–45.

37. Holder, J.L., Jr., Lotze, T.E., Bacino, C., and Cheung, S.W. (2012). A child with an inherited 0.31 Mb microdeletion of chromosome 14q32.33: further delineation of a critical region for the 14q32 deletion syndrome. Am. J. Med. Genet. A. *158A*, 1962–1966.

38. Maurin, M.L., Brisset, S., Le Lorc'h, M., Poncet, V., Trioche, P., Aboura, A., Labrune, P., and Tachdjian, G. (2006). Terminal 14q32.33 deletion: genotype-phenotype correlation. Am. J. Med. Genet. A. *140*, 2324–2329.

39. Hladilkova, E., Barøy, T., Fannemel, M., Vallova, V., Misceo, D., Bryn, V., Slamova, I., Prasilova, S., Kuglik, P., and Frengen, E. (2015). A recurrent deletion on chromosome 2q13 is associated with developmental delay and mild facial dysmorphisms. Mol. Cytogenet. *8*, 57.

40. Tang, L.S., and Finnell, R.H. (2003). Neural and orofacial defects in Folp1 knockout mice [corrected]. Birth Defects Res. A Clin. Mol. Teratol. *67*, 209–218.

41. Peters, H., Neubüser, A., Kratochwil, K., and Balling, R. (1998). Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. Genes Dev. *12*, 2735–2747.

42. Wilm, B., Dahl, E., Peters, H., Balling, R., and Imai, K. (1998). Targeted disruption of Pax1 defines its null phenotype and proves haploinsufficiency. Proc. Natl. Acad. Sci. USA *95*, 8692–8697.

43. Takimoto, A., Mohri, H., Kokubu, C., Hiraki, Y., and Shukunami, C. (2013). Pax1 acts as a negative regulator of chondrocyte maturation. Exp. Cell Res. *319*, 3128–3139.

44. Rodrigo, I., Hill, R.E., Balling, R., Münsterberg, A., and Imai, K. (2003). Pax1 and Pax9 activate Bapx1 to induce chondrogenic differentiation in the sclerotome. Development *130*, 473–482.

45. Sivakamasundari, V., Kraus, P., Sun, W., Hu, X., Lim, S.L., Prabhakar, S., and Lufkin, T. (2017). A developmental transcriptomic analysis of *Pax1* and *Pax9* in embryonic intervertebral disc development. Biol. Open *6*, 187–199.

46. Lamb, A.N., Rosenfeld, J.A., Neill, N.J., Talkowski, M.E., Blumenthal, I., Girirajan, S., Keelean-Fuller, D., Fan, Z., Pouncey, J., Stevens, C., et al. (2012). Haploinsufficiency of SOX5 at 12p12.1 is associated with developmental delays with prominent language delay, behavior problems, and mild dysmorphic features. Hum. Mutat. *33*, 728–740.

47. Li, Y.H., Yang, J., Zhang, J.L., Liu, J.Q., Zheng, Z., and Hu, D.H. (2017). BMP4 rs17563 polymorphism and nonsyndromic cleft lip with or without cleft palate: A meta-analysis. Medicine (Baltimore) *96*, e7676.

48. Saket, M., Saliminejad, K., Kamali, K., Moghadam, F.A., Anvar, N.E., and Khorram Khorshid, H.R. (2016). BMP2 and BMP4 variations and risk of non-syndromic cleft lip and palate. Arch. Oral Biol. *72*, 134–137.

49. Hammond, N.L., Brookes, K.J., and Dixon, M.J. (2018). Ectopic Hedgehog Signaling Causes Cleft Palate and Defective Osteogenesis. J. Dent. Res. *97*, 1485–1493.

50. Lipinski, R.J., Song, C., Sulik, K.K., Everson, J.L., Gipp, J.J., Yan, D., Bushman, W., and Rowland, I.J. (2010). Cleft lip and palate results from Hedgehog signaling antagonism in the mouse: Phenotypic characterization and clinical implications. Birth Defects Res. A Clin. Mol. Teratol. *88*, 232–240.

51. Dworkin, S., Boglev, Y., Owens, H., and Goldie, S.J. (2016). The Role of Sonic Hedgehog in Craniofacial Patterning, Morphogenesis and Cranial Neural Crest Survival. J. Dev. Biol. *4*, 24.

52. Paganini, I., Sestini, R., Capone, G.L., Putignano, A.L., Contini, E., Giotti, I., Gensini, F., Marozza, A., Barilaro, A., Porfirio, B., and Papi, L. (2017). A novel PAX1 null homozygous mutation in autosomal recessive otofaciocervical syndrome associated with severe combined immunodeficiency. Clin. Genet. *92*, 664–668.

53. Pohl, E., Aykut, A., Beleggia, F., Karaca, E., Durmaz, B., Keupp, K., Arslan, E., Palamar, M., Yigit, G., Özkinay, F., and Wollnik, B. (2013). A hypofunctional PAX1 mutation causes autosomal recessively inherited otofaciocervical syndrome. Hum. Genet. *132*, 1311–1320.

54. Shaffer, J.R., Orlova, E., Lee, M.K., Leslie, E.J., Raffensperger, Z.D., Heike, C.L., Cunningham, M.L., Hecht, J.T., Kau, C.H., Nidey, N.L., et al. (2016). Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. PLoS Genet. *12*, e1006149.

55. Adhikari, K., Fuentes-Guajardo, M., Quinto-Sánchez, M., Mendoza-Revilla, J., Camilo Chacón-Duque, J., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R.B., Pérez, G.M., et al. (2016). A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nat. Commun. *7*, 11616.

56. Cha, S., Lim, J.E., Park, A.Y., Do, J.H., Lee, S.W., Shin, C., Cho, N.H., Kang, J.O., Nam, J.M., Kim, J.S., et al. (2018). Identification of five novel genetic loci related to facial morphology by genome-wide association studies. BMC Genomics *19*, 481.

57. 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.

58. Hooper, J.E., Feng, W., Li, H., Leach, S.M., Phang, T., Siska, C., Jones, K.L., Spritz, R.A., Hunter, L.E., and Williams, T. (2017). Systems biology of facial development: contributions of ectoderm and mesenchyme. Dev. Biol. *426*, 97–114.

59. Jiang, R., Lan, Y., Chapman, H.D., Shawber, C., Norton, C.R., Serreze, D.V., Weinmaster, G., and Gridley, T. (1998). Defects in limb, craniofacial, and thymic development in Jagged2 mutant mice. Genes Dev. *12*, 1046–1057.

60. Xu, J., Krebs, L.T., and Gridley, T. (2010). Generation of mice with a conditional null allele of the Jagged2 gene. Genesis *48*, 390–393.

Supplemental information

The *PAX1* locus at 20p11 is a potential genetic modifier

for bilateral cleft lip

Sarah W. Curtis, Daniel Chang, Myoung Keun Lee, John R. Shaffer, Karlijne Indencleef, Michael P. Epstein, David J. Cutler, Jeffrey C. Murray, Eleanor Feingold, Terri H. Beaty, Peter Claes, Seth M. Weinberg, Mary L. Marazita, Jenna C. Carlson, and Elizabeth J. Leslie
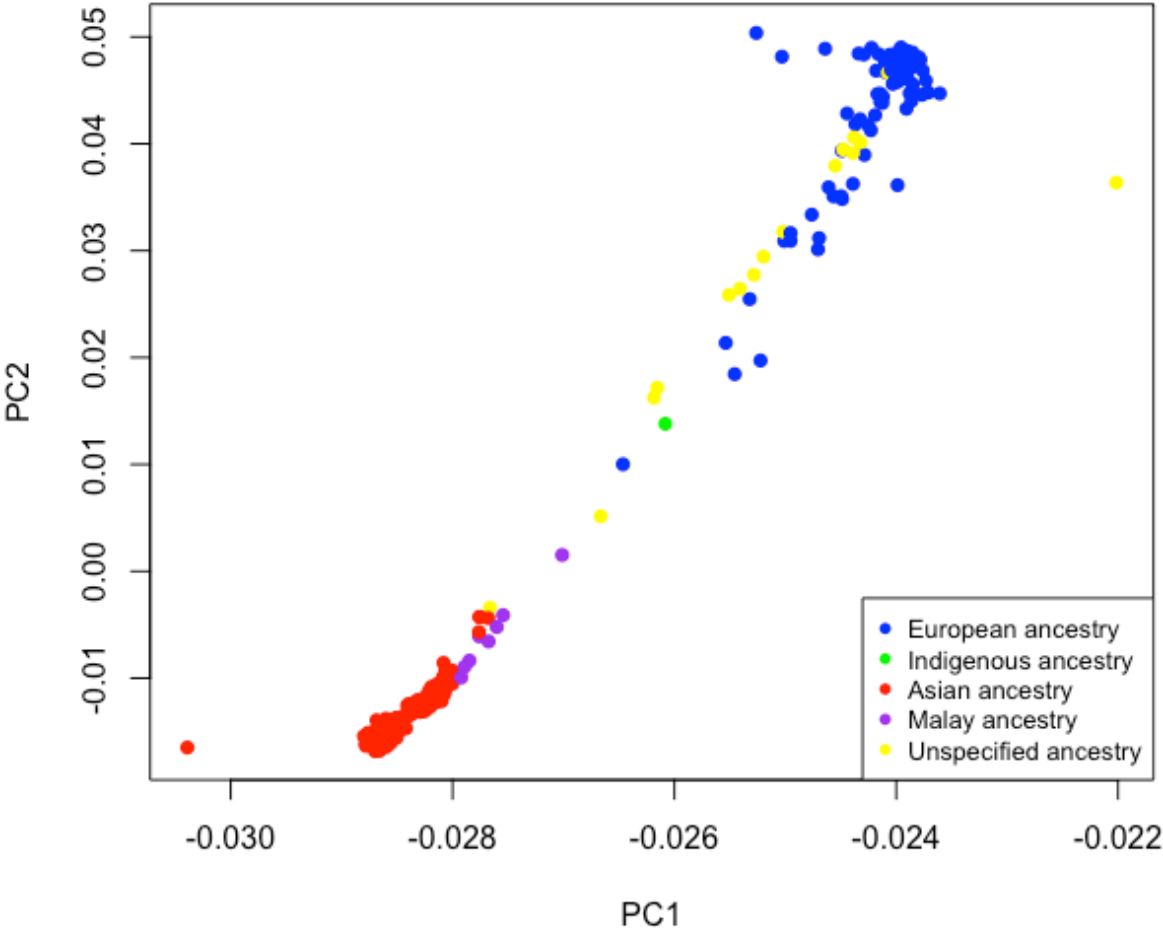
**Supplementary Figures and Tables**

Figure S1: PC plot for the POFC cohort
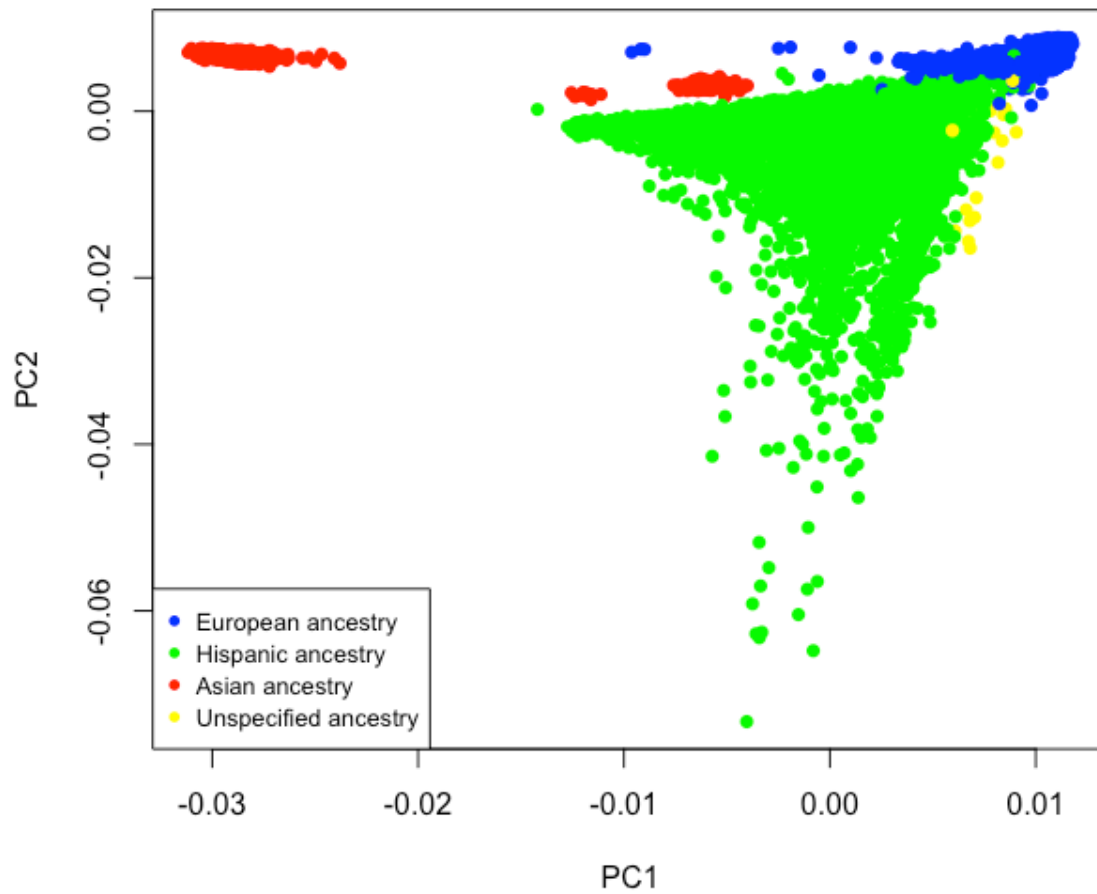
The first two principal components (PCs) for the POFC dataset used in this analysis, showing that a majority of the cohort is of European, Asian, or Hispanic ancestry.

Figure S2: PC plot for the GENEVA cohort

The first two principal components (PCs) for the GENEVA dataset used for replication in this analysis, showing that a majority of the cohort is of European and Asian ancestry.

Figure S3: QQ plots for the subtype-specific analyses
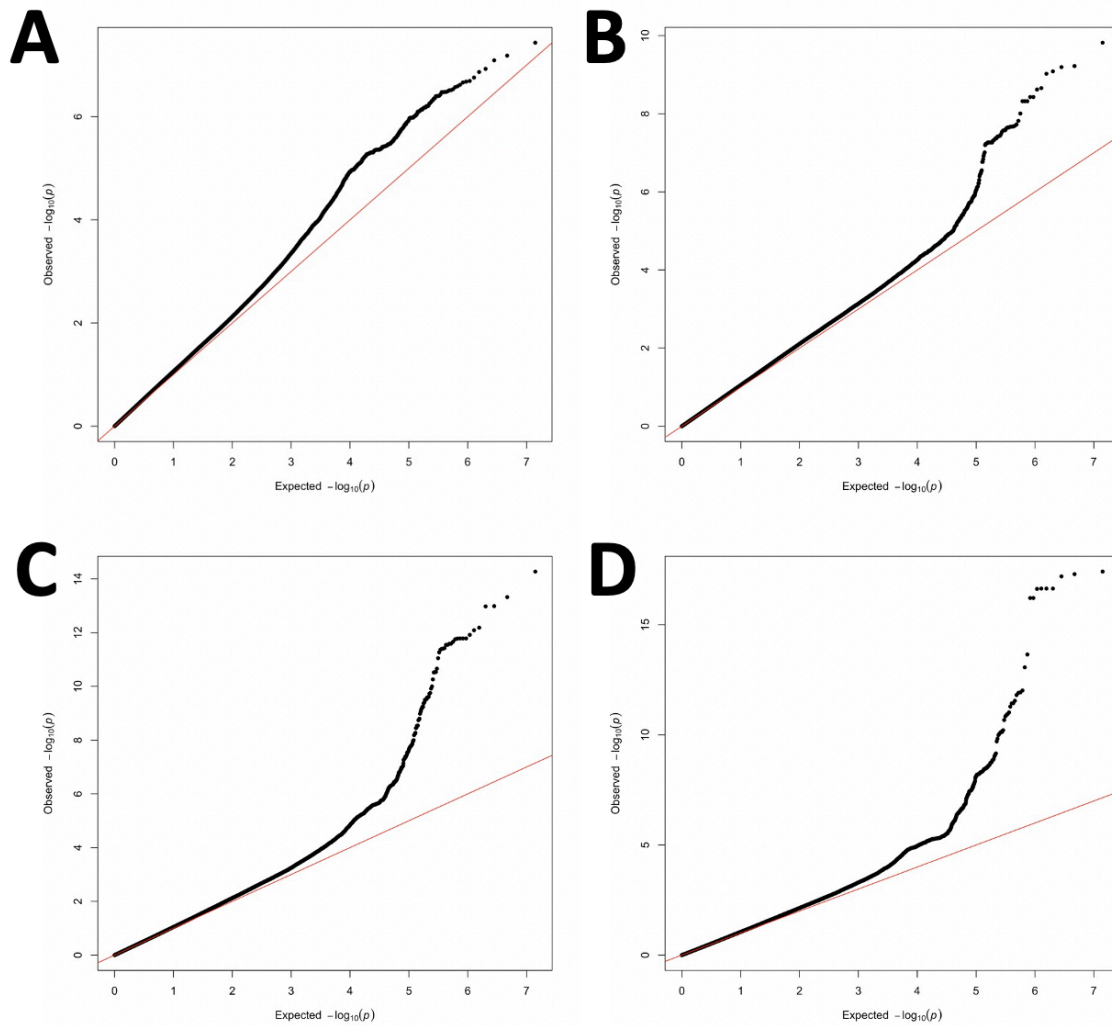
Q-Q plots of the −log10(p-values) of SNPs in the bilateral CL (A), unilateral CL (B), bilateral CLP (C), and unilateral CLP (D) subtype-specific case-control analysis. The genomic inflation factors are 1.10, 1.07, 1.05, and 1.06, respectively.

Figure S4: Manhattan plots for the subtype-specific analyses.
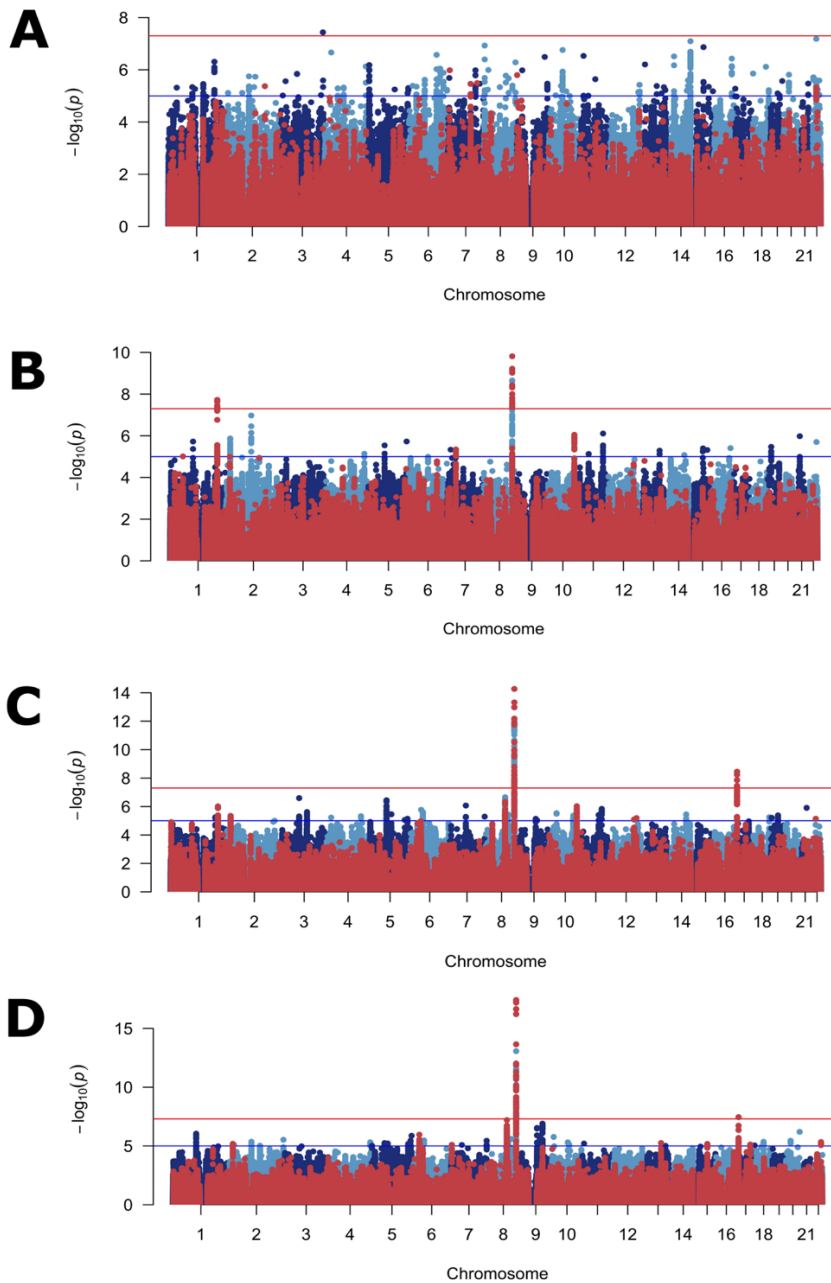
Manhattan plots of –$\log_{10}$(p-values) from the case-control analyses of bilateral cleft lip (A), unilateral cleft lip (B), bilateral cleft lip and palate (C), and unilateral cleft lip and palate (D). Lines indicate suggestive (blue) and genome-wide (red) thresholds for statistical significance. The points in red indicate SNPs within 50kb of a locus previously implicated in nonsyndromic clefting. The genomic inflation factors are 1.10, 1.07, 1.05, and 1.06 respectively.

Figure S5: Regional association plots showing –log$_{10}$(p-values) for the novel genome-wide suggestive peaks at 14q32 in the modifier analysis in bilateral CL (A), unilateral CL (B), bilateral CLP (C), and unilateral CLP (D). Plots were generated using LocusZoom. The recombination overlay (blue line, right y-axis) indicates the boundaries of the LD block. Points are color coded according to pairwise LD (r^2) with the index SNP.
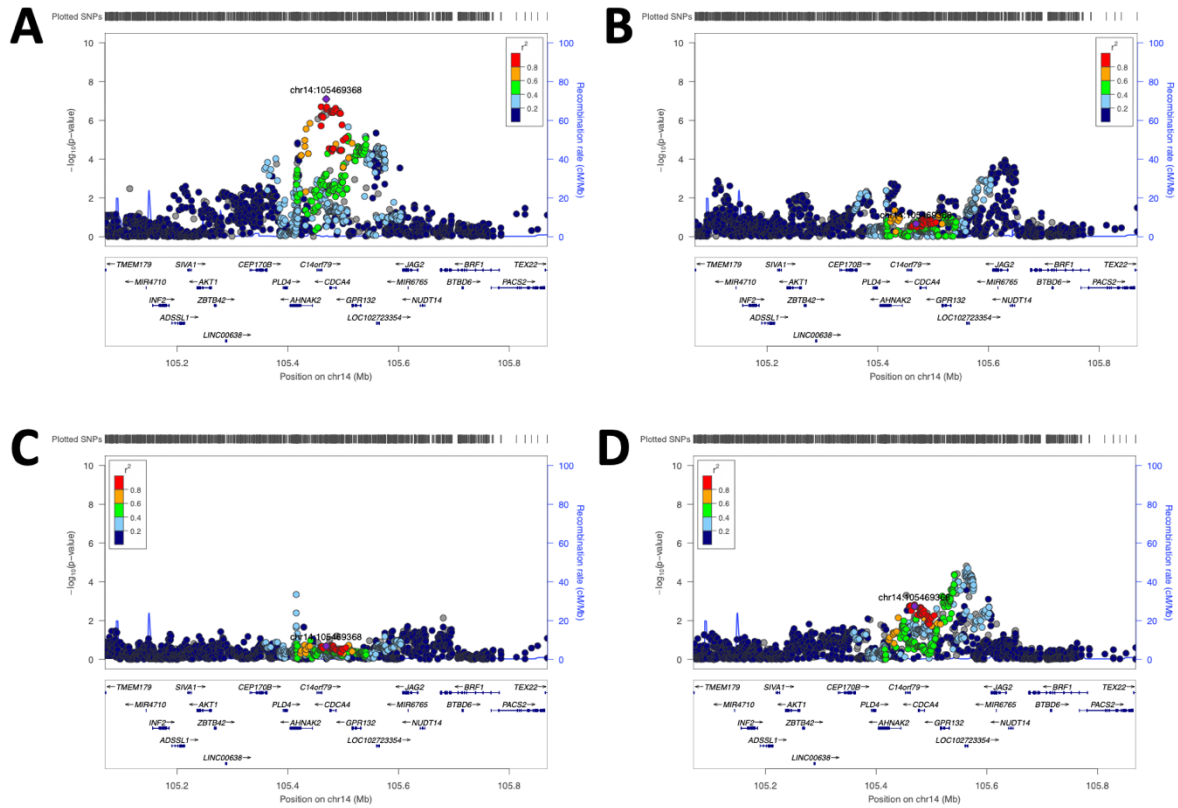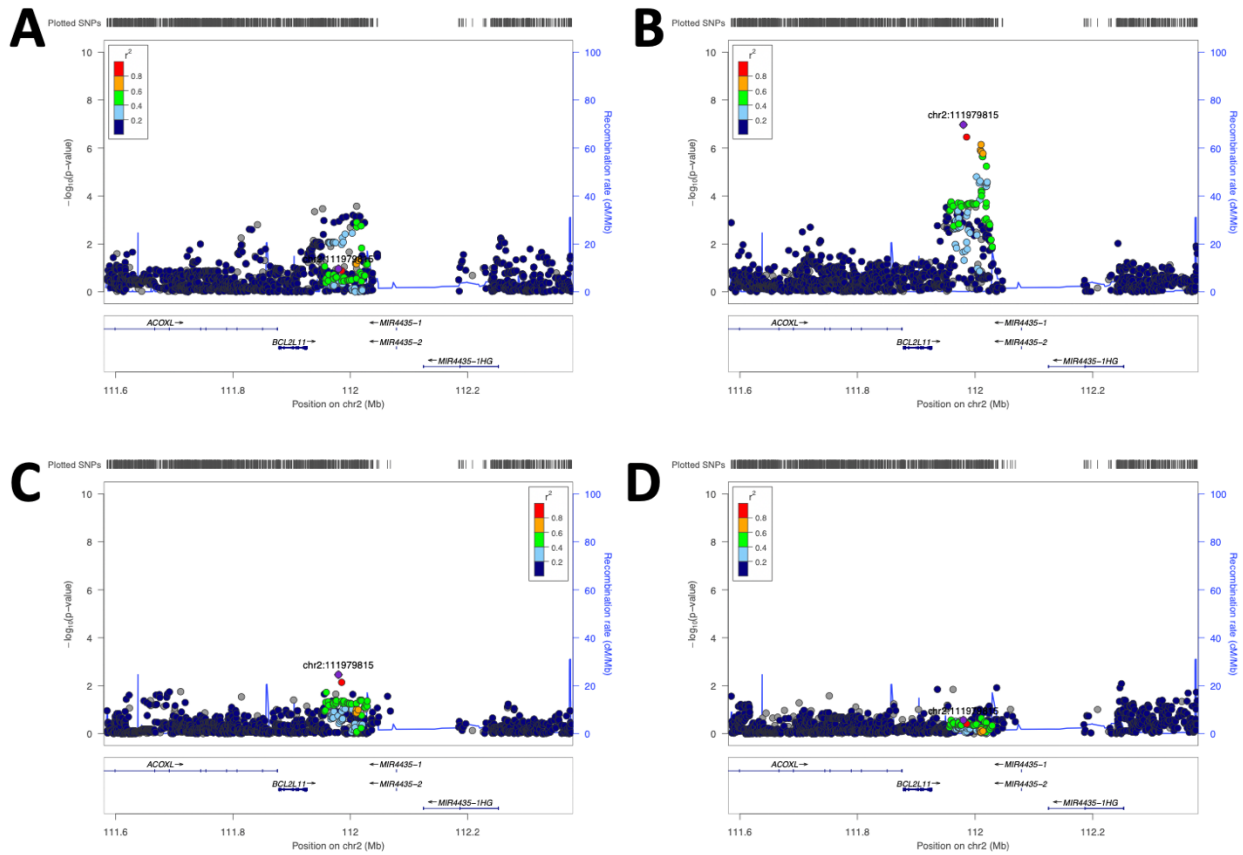
Figure S6: Regional association plots for subtype-specific analyses.

Regional association plots showing –log₁₀(p-values) for the novel genome-wide suggestive peaks at 2q13 in the modifier analysis in bilateral CL (A), unilateral CL (B), bilateral CLP (C), and unilateral CLP (D). Plots were generated using LocusZoom. The recombination overlay (blue line, right y-axis) indicates the boundaries of the LD block. Points are color coded according to pairwise LD (r²) with the index SNP.
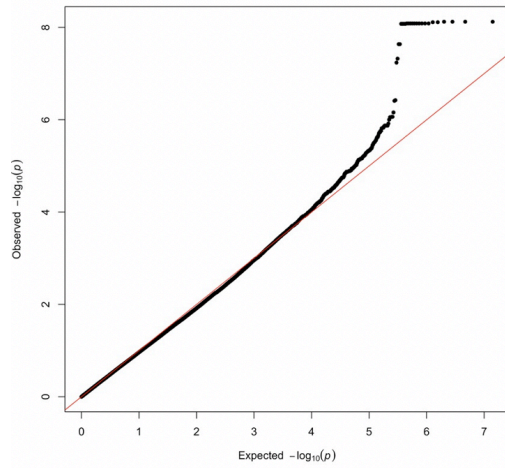
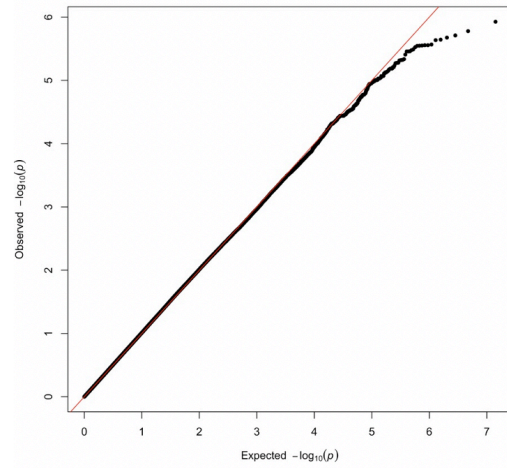**Figure S7: Comparison of subtypes across cleft type**
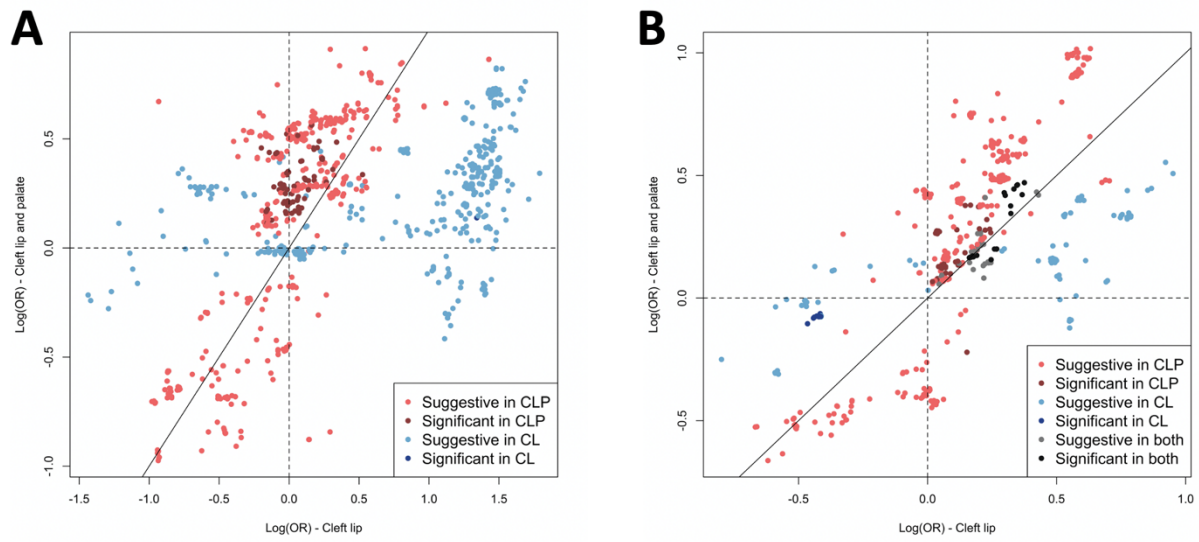
The log odds ratio for SNPs that were suggestive (p < 1 × 10^{-5}) or significant (p < 5 × 10^{-8}) in the subtype-specific case-control analyses in were compared between BCL and BCLP (A), UCL and UCLP (B)
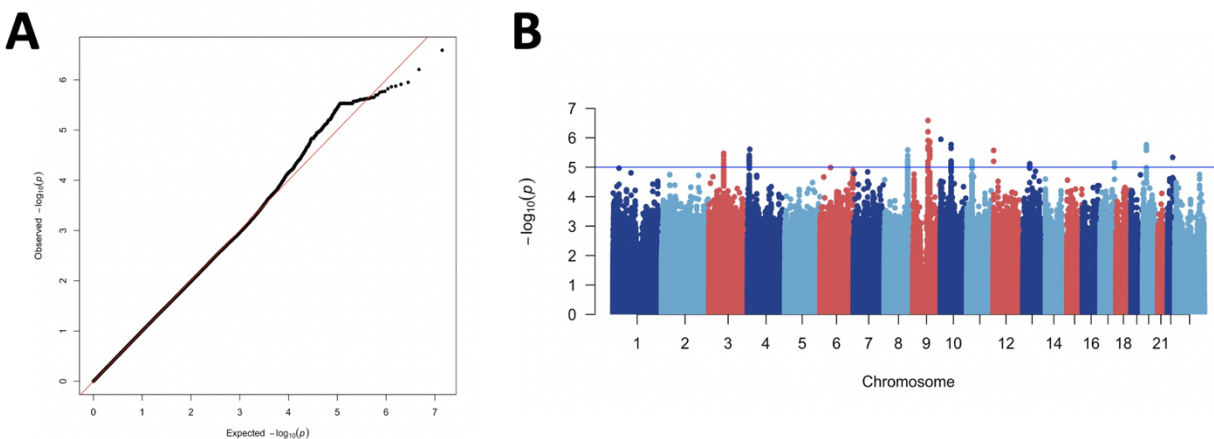
Figure S8: Q-Q plot of modifier analyses

Q-Q plots of the −log10(p-values) of SNPs in the CL (A) and CLP (B) modifier analysis. For the analysis in CL, the genomic inflation factor was 0.96; in CLP, the genomic inflation factor was 1.01.

Figure S9: Analysis of CL/P

Genome-wide results for the bilateral vs unilateral modifier analysis with CL and CLP combined. Q-Q plots of the −log10(p-values) of SNPs in the CL/P analysis (A). The genomic inflation factor is 1.00. Manhattan plots of −$\log_{10}$(p-values) from the modifier analysis in participants with cleft lip with or without cleft palate (B). Lines indicate suggestive (blue) and genome-wide (red) thresholds for statistical significance.

Figure S10: Topologically-associated domain (TAD) of the 20p11 locus.
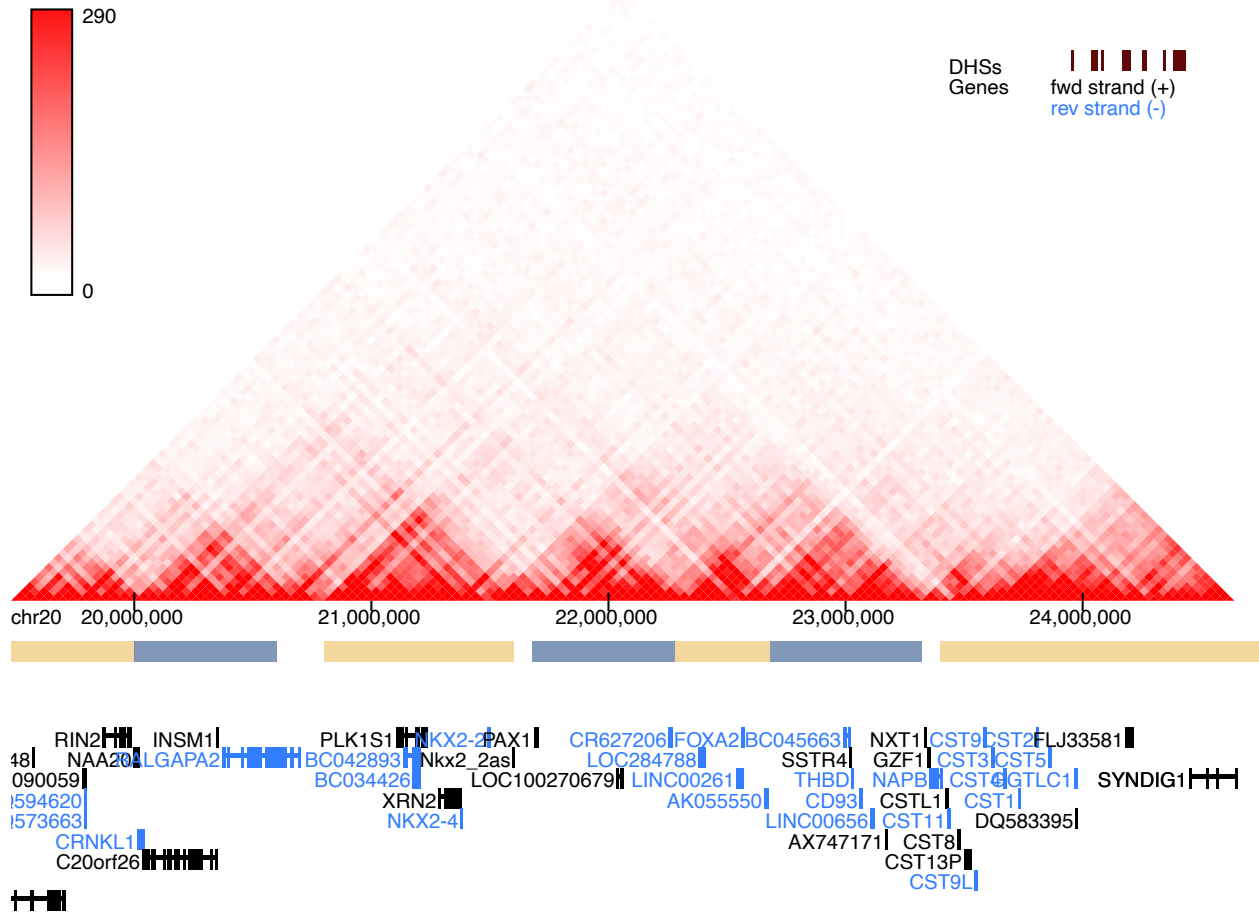
Topologically-associated domain (TAD) surrounding the 20p11 locus that was significantly associated with bilateral CL. The associated SNPs were in the same TAD as *Pax1*.

Figure S11: Functional enrichment
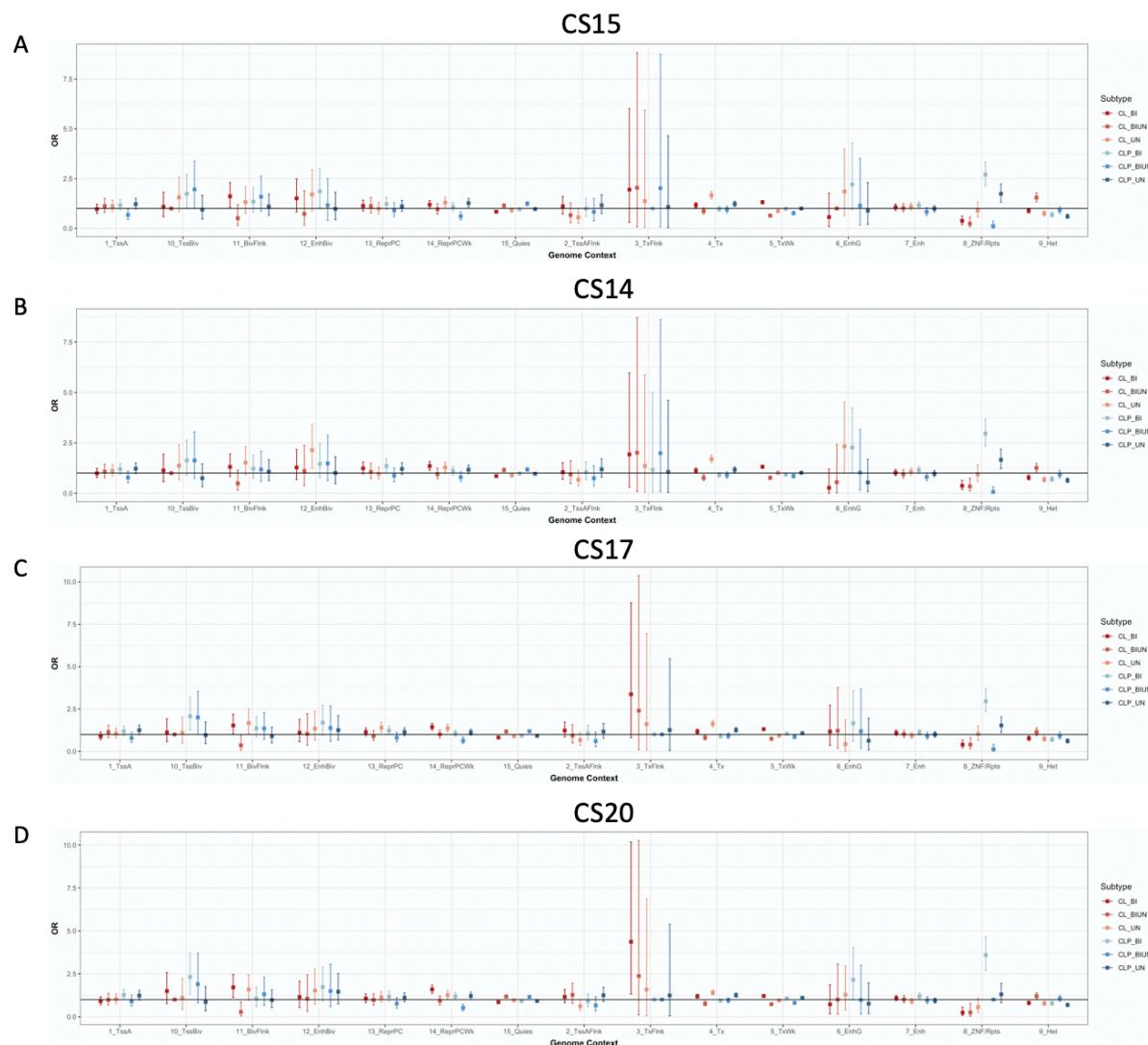
Enrichment of the top SNPs associated in either the CL modifier analysis, CLP modifier analysis, and each subtype analysis (p < 1 × 10^{-3}) were tested in each genomic region defined during craniofacial development.

Table S1: Participant demographics based on genetically defined ancestry groups

	Cleft lip only		Cleft lip and palate		Controls
	Bilateral	Unilateral	Bilateral	Unilateral	
Number of participants	44 (8.1%)	502	572 (31.9%)	1221	1626
Number of families	44	434	530	1123	1626
Number with Asian ancestry	15 (34.1%)	184 (36.7%)	121 (21.4%)	205 (16.9%)	161 (9.9%)
Number with European ancestry	13 (29.5%)	151 (30.1%)	159 (28.1%)	315 (26.0%)	835 (51.4%)
Number with Latin/South American ancestry	15 (34.1%)	167 (33.3%)	285 (50.4%)	692 (57.0%)	626 (38.5%)
Number with Unspecified ancestry	1 (2.3%)	0 (0%)	1 (0.2%)	1 (0.1%)	4 (0.2%)

Table S6: Comparison of subtype-specific analyses

Comparison	Percent of SNPs with same effect size that are known	Percent of SNPs with different effect sizes that are known	Odds ratio	95% CI	P-value
BCL-UCL	1.8%	26.8%	0.05	0.02-0.11	2.41×10^{-20}
BCLP-UCLP	8.6%	44.0%	0.12	0.04-0.28	1.04×10^{-7}
BCL-BCLP	7.7%	31.4%	0.18	0.12-0.26	6.89×10^{-20}
UCL-UCLP	43.7%	54.3%	0.65	0.47-0.88	0.005

Table S9: Loci with suggestive significance in the modifier analysis of cleft lip with or without cleft palate

Locus	Top variant	Reference allele	Alternate allele	AF[1]	OR[2]	95%CI	P
20p11.23	rs12480480	G	A	0.17	1.49	1.37-1.62	1.72E-06
4p15.33	rs11373830	C	CA	0.07	1.99	1.73-2.3	2.47E-06
4p16.1	rs149050104	C	T	0.06	1.82	1.61-2.05	3.94E-06

[1]Allele frequency of alternate allele

[2]Risk of BCL over UCL

Table S10: Replication results for 20p11 in cleft lip

SNP	CHR	BP	Reference allele	Alternate allele	AF[1]	OR[2]	95% CI	P-value
rs6113495	20	22061487	G	A	0.46	1.05	0.57-1.92	0.8735
rs6113496	20	22061632	T	G	0.40	0.74	0.40-1.35	0.33
rs6047869	20	22062683	A	C	0.40	0.74	0.40-1.35	0.33
20:22062945	20	22062945	G	A	0.15	1.14	0.34-3.81	0.8314
rs115747027	20	22063265	T	C	0.40	0.74	0.40-1.35	0.33
20:22063347	20	22063347	T	C	0.13	3.83	1.64-8.95	0.001897
rs113999308	20	22063421	C	T	0.28	1.17	0.40-3.43	0.763
rs117160568	20	22063660	A	G	0.46	1.05	0.57-1.93	0.8575

[1]Allele frequency of alternate allele
[2]Risk of BCL over UCL

Table S11: Replication results for 20p11 in cleft lip and palate

SNP	CHR	BP	Reference allele	Alternate allele	AF[1]	OR[2]	95% CI	P-value
rs6113495	20	22061487	G	A	0.44	1.03	0.84-1.26	0.7304
rs6113496	20	22061632	T	G	0.39	0.90	0.73-1.11	0.3611
rs6047869	20	22062683	A	C	0.39	0.91	0.74-1.12	0.3846
20:22062934	20	22062934	T	G	0.05	0.94	0.55-1.60	0.8225
20:22062945	20	22062945	G	A	0.16	1.34	0.92-1.94	0.1232
rs115747027	20	22063265	T	C	0.39	0.90	0.73-1.11	0.3611
20:22063347	20	22063347	T	C	0.18	1.12	0.80-1.57	0.476
rs113999308	20	22063421	C	T	0.28	1.30	0.96-1.75	0.08084
rs117160568	20	22063660	A	G	0.44	1.03	0.84-1.26	0.7304

[1]Allele frequency of alternate allele
[2]Risk of BCLP over UCLP

Table S12: Comparison of known loci with modifier analyses

Comparison	Known SNPs	Odds ratio	95% CI	P-value
BCL v UCL	rs200656853; rs35640367	0.26	0.04-0.84	0.03
BCLP v UCLP	rs71776173; rs134124	0.53	0.08-1.71	0.32