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

David Cutler, Emory University
Michael Epstein, Emory University
Elizabeth Leslie, Emory University
SW Curtis, Emory University
D Chang, Emory University
MK Lee, University of Pittsburgh
JR Shaffer, University of Pittsburgh
K Indencleef, KU Leuven– University Hospital Leuven
JC Murray, University of Iowa
E Feingold, University of Pittsburgh

Only first 10 authors above; see publication for full author list.

Journal Title: Human Genetics and Genomics Advances
Volume: Volume 2, Number 2
Publisher: (publisher) | 2021-04-08
Type of Work: Article
Publisher DOI: 10.1016/j.xhgg.2021.100025
Permanent URL: https://pid.emory.edu/ark:/25593/vv6r7

Final published version: http://dx.doi.org/10.1016/j.xhgg.2021.100025
Accessed November 30, 2022 8:36 AM EST
The *PAX7* 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 Karljine 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. Leslie1,*

**Summary**

Nonsyndromic orofacial clefts (OFCs) are a common birth defect and are phenotypically heterogeneous 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 × 10⁻⁶), 300 kb downstream of *PAX7*, 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 of 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. 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.

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 HumanCore+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. These data were then phased with SHAPEIT2 and imputed with IMPUTE2 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.

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) as implemented in the GENESIS software package. 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 \( 5 \times 10^{-8} \) were considered genome-wide significant, and those with p values less than \( 1 \times 10^{-5} \) 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. Regional association plots were made with LocusZoom, where the LD blocks and recombination rates were estimated from European populations.

Modifier GWASs

We identified genetic modifiers (genetic variants that are associated with phenotype heterogeneity or expressivity) using case-control 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). 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 as implemented in GENESIS, 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. Similar regional association plots were made with LocusZoom.

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 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. 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-Quadv1_B
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). 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.

Epigenomic context of results
Topologically associated domains (TADs) were defined for significantly associated loci using the H1-ESC cell line in 3D Genome Browser. Functional enrichment was tested by first annotating all of the SNPs to the craniofacial functional regions defined by estimated ORs and their 95% CIs were calculated. 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), 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.

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 1q42.33 (lead SNP: rs61996057; p = 8.07 × 10^{-8}; Figures S3A, S4A, and S5; Table S2), have been implicated in syndromes with facial dysmorphisms. 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). 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. 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.

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
suggestive significance (Figure 2A; Figure S8A; Table S7). In 4 Human Genetics and Genomics Advances rs143865354; p 20p11 reached genome-wide significance (lead SNP: modifier analysis of CL, one locus on chromosome risk but would detect variants important for the formation 
detect variants generally important for both CL or CLP 
group comparison, this analysis would not be able 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 spe-
cific to CL.

Although the 20p11 locus had not previously been asso-
ciated with risk to OFCs, it has been associated with varia-
tion in normal facial structures. Therefore, we next investi-
gated 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

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 LINCO1432 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 mod-
ifiers, 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 spe-
cific to CL.

![Figure 1. Subtype-specific analyses](Image)

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 UCLP (light blue 
points) (A), and BCLP (dark red points) 
and UCLP (light red points) (B), and were classified in (C) by whether the 95% confi-
dence interval for the OR overlapped and 
whether the variant was identified in pre-
vious GWAS (Known) or not (Not known).
depletions were consistent throughout craniofacial development (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 mechanisms resulting in bilateral and unilateral CL and CLP will require a locus-by-locus investigation.

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 ($p = 2.63 \times 10^{-11}$), 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 ($p = 2.9 \times 10^{-5}$) 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 development:

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
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 and found SNPs in IRF6 (MIM: 607199) were associated with the formation of a unilateral CL/P compared to bilateral CL/P. 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. 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. The 14q32.33 also contains JAG2, which is part of the Notch signaling pathway and is important for craniofacial development.

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...
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 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 (R²) 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).
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 \textit{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 understanding 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**

Original data used in this study are available at dbGaP (dbGaP: phs000774.v2.p1, phs000094.v1.p1).

**Supplemental Information**

Supplemental Information can be found online at https://doi.org/10.1016/j.xhgg.2021.100025.

**Acknowledgments**

The authors thank the dedicated field staff, collaborators, and participating families for their important contributions to this study. This work was supported by grants from the National Institutes of Health (NIH), including R00-DE025060 (E.J.L.), X01-HG007485 (M.L.M., E.F.), R01-DE016148 (M.L.M., S.M.W.), U01-DE024425 (M.L.M.), R37-DE008559 (J.C.M., M.L.M.), R01-DE009886 (M.L.M.), R21-DE016930 (M.L.M.), R01-DE012472 (M.L.M.), R01-DE014581 (T.H.B.), U01-DE018993 (T.H.B.) and U01-DE020078; R01-DE027023 [S.M.W.]; funding for genotyping by the National Human Genome Research Institute (X01-HG007821); and funding for initial genomic data cleaning by the University of Washington provided by contract HHSN268201200008I from the National Institute for Dental and Craniofacial Research awarded to the Center for Inherited Disease Research.

**Declaration of interests**

The authors declare no competing interests.

Received: November 23, 2020
Accepted: January 22, 2021

**Web resources**

- GTOOL, https://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html

**References**

between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. Hum. Genet. 136, 275–286.


