Published in final edited form as:

Hum Genet. 2020 October ; 139(10): 1261–1272. doi:10.1007/s00439-020-02169-9.

Evidence for craniofacial enhancer variation underlying nonsyndromic cleft lip and palate

Vershanna E. Morris^{1,2}, S. Shahrukh Hashmi^{1,2}, Lisha Zhu³, Lorena Maili^{1,2}, Christian Urbina^{1,2}, Steven Blackwell⁴, Matthew Greives⁵, Edward P. Buchanan⁶, John B. Mulliken⁷, Susan H. Blanton⁸, W. Jim Zheng³, Jacqueline T. Hecht^{1,2,4,9}, Ariadne Letra^{2,9,10}

¹Department of Pediatrics, UTHealth McGovern Medical School, Houston, Texas, 77030, USA

²Pediatric Research Center, UTHealth McGovern Medical School, Houston, Texas, 77030, USA

³UTHealth School of Biomedical Informatics, Houston, Texas, 77030, USA

⁴Shriners' Hospital for Children, Houston, Texas, 77030, USA

⁵Department of Pediatric Surgery, University of Texas Health Science Center McGovern Medical School at Houston, Houston, Texas

⁶Department of Plastic Surgery, Texas Children's Hospital, Houston, Texas, USA

⁷Department of Plastic Surgery, Boston Children's Hospital, Boston, MA, 02115, USA

⁸Dr. John T. Macdonald Foundation Department of Human Genetics, John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida, 33136, USA

⁹Department of Diagnostic and Biomedical Sciences, UTHealth School of Dentistry, Houston, Texas, 77030, USA

¹⁰Center for Craniofacial Research, UTHealth School of Dentistry, Houston, Texas, 77030, USA

Abstract

Nonsyndromic cleft lip with or without cleft palate (NSCLP) is a common birth defect for which only ~20% of the underlying genetic variation has been identified. Variants in noncoding regions have been increasingly suggested to contribute to the missing heritability. In this study, we investigated whether variation in craniofacial enhancers contributes to NSCLP. Candidate enhancers were identified using VISTA Enhancer Browser and previous publications. Prioritization was based on patterning defects in knockout mice, deletion/duplication of craniofacial genes in animal models and results of whole exome/whole genome sequencing studies. This resulted in 20 craniofacial enhancers to be investigated. Custom amplicon-based

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <https://www.springer.com/aam-terms-v1>

*Corresponding Author: Ariadne Letra, DDS, MS, PhD, Associate Professor, University of Texas Health Science Center at Houston, School of Dentistry, Department of Diagnostic and Biomedical Sciences, 1941 East Road, BBSB 4210, Houston, TX, 77054, (713) 486-4228, Ariadne.M.Letra@uth.tmc.edu.

Publisher's Disclaimer: This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

The authors declare that they have no conflict of interest.

sequencing probes were designed and used for sequencing 380 NSCLP probands (from multiplex and simplex families of nonHispanic white (NHW) and Hispanic ethnicities) using Illumina MiSeq. The frequencies of identified variants were compared to ethnically-matched European (CEU) and Los Angeles Mexican (MXL) control genomes and used for association analyses. Variants in mm427/*MSX1* and hs1582/*SPRY1* showed genome-wide significant association with NSCLP ($p = 6.4 \times 10^{-11}$). *In silico* analysis showed that these enhancer variants may disrupt important transcription factor binding sites. Haplotypes involving these enhancers and also mm435/*ABCA4* were significantly associated with NSCLP, especially in NHW ($p = 6.3 \times 10^{-7}$). Importantly, group-wise burden analysis showed several enhancer combinations significantly over-represented in NSCLP individuals revealing novel NSCLP pathways and supporting a polygenic inheritance model. Our findings support the role of craniofacial enhancer sequence variation in the etiology of NSCLP.

Keywords

enhancer; variation; nonsyndromic cleft lip/palate; network; gene

Introduction

Nonsyndromic cleft lip with or without cleft palate (NSCLP), with a prevalence of 1 in 700 newborns, is a common craniofacial malformation that imposes significant medical, psychosocial and financial burdens on patients and families (Gorlin et al. 2001; Hashmi et al. 2005; Wyszynski 2002). NSCLP is multifactorial, with genetic and environmental factors playing etiologic roles (Dixon et al. 2011). Approximately 40 genes/loci have been identified through linkage, mutation screening, candidate gene, genome-wide association studies (GWAS) and recent whole exome sequencing (WES) studies, however only approximately 20% of the underlying genetic variation in NSCLP has been identified (Dixon et al. 2011). While these studies identified a few variants in coding regions, the majority of gene variants associated with NSCLP are located in noncoding regions of the genome, many of which are known to harbor regulatory elements that control gene expression (Dixon et al. 2011; Thieme and Ludwig 2017).

Distant-acting enhancer sequences are noncoding transcriptional regulators that are among the most abundant regulatory elements in the mammalian genome; they regulate gene promoters and control cell type-specific gene expression (Wu et al. 2014). In mice, a large number of distant-acting enhancers have been identified and shown to have tissue-specific gene targets with differential expression at specific developmental stages (Dickel et al. 2018; Visel et al. 2009a). Moreover, these enhancer regions were shown to modulate the finely-tuned process of murine craniofacial development and were often significantly enriched with common variants associated with clefting (Attanasio et al. 2013; Pennacchio and Rubin 2001). Recent studies using transgenic mouse models have identified the patterns of activity of several distal-acting candidate enhancers in craniofacial tissues and their correlations with craniofacial dysmorphologies (Visel et al. 2009b). ChIP-Seq analysis on embryonic murine facial tissue revealed thousands of candidate craniofacial enhancers and transgenic reporter assays showed reproducible activity patterns in craniofacial structures (Visel et al. 2009b).

Targeted enhancer deletions showed that changes in transcript levels of the presumed target gene were associated with changes in craniofacial shape (Wu et al. 2014). In a follow-up study in humans, significant associations were identified between NSCLP and a forebrain enhancer near *FOXE1* (forkhead box E1, MIM #602617), and between a branchial arch enhancer near *FOXP1* (forkhead box P1, MIM #605515) and cleft palate only (CP, MIM #119540) (Shaffer et al. 2019). While progress has been made in identifying craniofacial enhancers in human NSCLP, there is a large knowledge gap. In this study, we investigated whether variation in craniofacial enhancers contributes to NSCLP. We describe the results of our targeted next-generation sequencing of twenty prioritized craniofacial enhancers and their association and polygenic contributions to NSCLP.

Materials and Methods

Study population

The study population has previously been described and consisted of 1,653 NSCLP families including probands, their parents and selected affected and unaffected family members from both nonHispanic white and Hispanic populations (Barba et al. 2019; Blanton et al. 2004a; Blanton et al. 2004b). Individuals were recruited at the UTHealth Craniofacial Clinics, Shriners Hospital for Children – Houston, Texas Children’s Hospital, Houston and Boston Children’s Hospital Craniofacial Clinics under appropriate institutional review board approvals and informed consent (HSC-MS-03–090). All cases had isolated nonsyndromic cleft lip and palate. Genomic DNA was obtained from peripheral blood and/or saliva samples as previously described. For this study, 380 probands with cleft lip and palate were selected from 158 multiplex families (104 nonHispanic white, NHW; and 54 Hispanic) and 222 simplex families (113 NHW and 109 Hispanic) were selected. Probands were selected based on the highest DNA sample quality/concentration and from our largest and most informative pedigrees.

Craniofacial enhancer selection and sequencing

Twenty candidate craniofacial enhancer regions were selected for sequencing in this study (Table 1). These enhancers were prioritized for sequencing based on information obtained through the VISTA Enhancer Browser database (<http://enhancer.lbl.gov>) (Visel et al. 2007), and from craniofacial enhancers previously identified in animal models studies that were available in 2016 (Attanasio et al. 2013; Visel et al. 2009b). Initially, the list of craniofacial enhancers available in the VISTA Enhancer browser database was filtered based on evidence of positive expression in the mouse face and/or tissues involved in patterning of facial structures (i.e., in the blood vessels, somite, branchial arch, nose, facial mesenchyme, and mesenchyme from the neural crest of transgenic mouse embryos) (n=354). Next, additional filtering was based on published evidence from animal model and/or human genetic association studies suggesting that the nearest gene was associated with orofacial clefting, evidence from recent whole exome sequencing studies, and evidence that deletion/duplication of nearest genes was associated with clefting phenotypes (n=255). These enhancers were further prioritized considering the results of previous studies including transgenic animals and/or ChIP-seq data resulting in the 20 putative craniofacial enhancers to be sequenced in this study (Table 1 and Figure S1). For mouse enhancers (designated

mm), the corresponding coordinates in humans (designated hs) were retrieved and used for probe design. Probands were sequenced using an Illumina TruSeq custom panel (Illumina Inc., San Diego, CA, USA) designed to cover the 20 enhancer regions yielding 134 amplicons (45,757 bp cumulative targets) (Table 1). Library preparation was performed using TruSeq Custom Amplicon v1.5 Kit (Illumina) following the manufacturer's instructions. A bead-based protocol was used to normalize the adapter-ligated library, which was loaded directly onto the Illumina MiSeq platform for sequencing. Pooled amplicon libraries were then sequenced on the Illumina MiSeq sequencer (Illumina) which generated paired-end reads with length of 150–200 nucleotides and 100X coverage. Data analysis, including variant calling, was performed using the Illumina BaseSpace Sequence Hub (Illumina). Raw sequence reads were mapped to the human reference genome (UCSC Genome Browser hg19) using Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009). Variant calling for single nucleotide variants (SNVs) and indels was performed using the TruSeq Amplicon v2.0 application (Illumina) and the Genome Analysis Toolkit (GATK) (McKenna et al. 2010). For each proband, a variant call file (.vcf) was generated and filtered further for coverage depth, genotype quality, and variant quality using GATK according to best practice recommendations. Default variant call filters included a minimum coverage depth of 20, strand bias of -10 , and minimum genotype quality of 30. Only variants that passed quality control filters and had $>90\%$ aligned bases were analyzed.

Association analyses

All analyses were performed separately for NHW and Hispanic families, due to differences in allele frequencies among the groups (Chiquet et al. 2008). Allele frequencies for each identified variant in NSCLP probands were compared to that of 503 CEU (European) and 64 MXL (Los Angeles Mexican) genomes from the 1000 Genomes Project database (Genomes Project et al. 2015). Case-control association testing was performed for each enhancer variant using chi-square statistics. Next, haplotypes were generated and tested for association with NSCLP using Haploview v.4.2 (Barrett et al. 2005), using default settings to ignore pairwise settings of markers $>500\text{kb}$ apart and to exclude individuals with $>50\%$ missing genotypes. Bonferroni corrected p-values $2.2\text{e-}4$ ($0.05/228$ tests) were considered significant.

Group-wise burden analysis was performed to test for combined-variant effects on NSCLP risk considering each enhancer as a unit (each unit which may contain several variants within, Tables S1 and S2). We used the weighted sum statistical approach proposed by Madsen and Browning (Madsen and Browning 2009) to jointly analyse groups of 1–5 enhancer units in order to test for groupwise association with NSCLP. Analysis were performed in EPACTS (Efficient and Parallelizable Association Container Toolbox) software considering the pooled allele frequencies in each enhancer unit across NSCLP cases and controls for weighting each variant, and then applying a Wilcoxon Rank Sum Test between the binary phenotypes of 'NSCLP' and 'no NSCLP' and weighted rare variant scores. Statistical significance for the burden analysis was assumed at a multiple-comparison Bonferroni adjusted threshold of $p < 2.3\text{e-}6$. However, due to probability of the same enhancers potentially appearing in more than one combination (i.e., the posterior probability of two enhancers appearing as part of a higher order combination after they had already

occurred in a 2-enhancer combination is higher than the prior probability based on chance alone), and the risk of inflating the Type II error rate, associations with p-values between 2.3×10^{-4} and 2.3×10^{-6} were also considered and identified as suggestive associations.

Prediction of variant function

In silico prediction of associated variant function was performed using Haploreg (Ward and Kellis 2016), RegulomeDB (Boyle et al. 2012), and PROMO (Messegueur et al. 2002).

Pathway analysis

Pathway analysis was performed to infer gene-gene relationships for the genes identified as associated with NSCLP through the groupwise burden analysis. Ontology Fingerprints (OntoFin)-derived biological network (OntoFin Net) was used as source of information for gauging all gene functions, as previously described (Chen et al. 2018; Chen et al. 2015; Qin et al. 2014a; Qin et al. 2012b; Tsoi et al. 2009a).

Results

From the 380 probands, data from 369 (209 NHW and 160 Hispanics) passed the quality control filters which were included in subsequent analyses (~25,600 SNVs). The total number of variant calls across all sequenced enhancer regions and quality control filtering of variant calls is shown in Figure S1. SNVs previously not reported in the 1000 Genomes Database (or elsewhere) were excluded from analysis in order to avoid spurious associations. Similarly, indels and SNVs located at the same base pair position(s) were excluded. Indels found in cases were compared to the known indels included in the Genome Aggregation Database (gnomAD) to assess for potentially novel or over-transmitted indels. One indel located on chromosome 1 in the *ABCA4* (ATP-binding cassette subfamily A member 4, MIM #601691) gene was present in 166 cases; however, it was located in a low-complexity region of the chromosome prone to replication slippage, therefore making its interpretation difficult. The remaining 12 indels were found across 8 enhancers (Table S1).

Sequencing across the 20 craniofacial enhancer regions identified 225 and 228 variants, both common and rare ($MAF < 0.01$), in the NHW and Hispanic NSCLP probands, respectively (Tables S2 and S3). In the single variant association tests, significant associations with low-frequency variants in two enhancers, mm427 (rs1994828) and hs1582 (rs6534405), in both NHW and Hispanic NSCLP probands were identified ($P < 6.4 \times 10^{-11}$) (Table 2). The mm427 craniofacial enhancer is located 105kb upstream of *MSX1* (msh homeobox 1, MIM #142983) and is active in the facial mesenchyme and nose in the mouse embryo at E11.5. The hs1582 enhancer is located 65 kb downstream of *SPRY1* (sprouty RTK signaling antagonist 1, MIM #602465) and active in the facial mesenchyme, forebrain, and hindbrain of the mouse at E11.5 (Fig. 1). Significant haplotype associations including the associated mm427/*MSX1* variant were also found in the NHW and Hispanic populations whereas haplotypes involving the hs1582/*SPRY1* variant appear to be significantly associated with only NHW cases (Table 3). In the haplotype-based analyses, significant associations were identified for haplotypes involving the associated enhancers mm427/*MSX1* ($p = 4.11 \times 10^{-28}$

for NHW; $p=8.53 \times 10^{-9}$ for Hispanics), *hs1582/SPRY1* ($p=3.02 \times 10^{-30}$ for NHW), and also *mm435/ABCA4* ($p=6.33 \times 10^{-7}$ for NHW) (Table 3).

In silico prediction of variant function for *mm427/MSX1* and *hs1582/SPRY1* suggested differential allele-specific binding affinities to transcription factors involved in important cellular functions during cell cycle regulation and craniofacial development. The *mm427/MSX1* locus was predicted to harbor binding sites for *LEF1* and *TCF4* transcription factors, whereas the *hs1582/SPRY1* locus was predicted to have binding sites for *MEF2A* (myocyte enhancer factor 2A, MIM #600660) (Table 4).

The results of the group-wise burden analysis showed novel significant and suggestive associations for 2-, 3-, 4-, and 5-enhancer variant combinations involving *RUNX2*, *ABCA4*, *BMP4* (bone morphogenetic protein 4, MIM #112262), *COL1A2* (collagen type I alpha 2 chain, MIM #120160), *MESDC1* (mesoderm development candidate 1, MIM #615466), and *MN1* (meningioma chromosome region 1, MIM #156100) ($p = 1.7 \times 10^{-5}$) (Table 5). Additional pathway analysis of these genes identified gene clusters and a novel genetic network connecting *COL1A2*, *BMP4*, and *RUNX2* with other known craniofacial development genes (e.g. *TGFB1*, *BMP2*, and *MMP13*) (Fig. 2).

Discussion

In this study, we sequenced twenty craniofacial enhancers in NSCLP probands and identified both common and rare noncoding enhancer variants, individually and in concert, as likely contributors to NSCLP susceptibility. These enhancers had been functionally validated in mice and their roles in craniofacial development previously established (Attanasio et al. 2013; Visel et al. 2009b). Our findings show that enhancer variation may contribute to NSCLP and strongly support a polygenic inheritance model long proposed to explain the heritability of NSCLP, where the effects of multiple genes affect a given trait (Carter 1969).

Variants in two enhancers, *mm427* and *hs1582*, near *MSX1* and *SPRY1*, respectively, were the most significantly associated with NSCLP. Additionally, significant haplotype associations that included the individually associated variants were also identified; haplotypes including *mm427/MSX1* were strongly associated with NSCLP in NHW and Hispanics, whereas *hs1582/SPRY1* and *mm435/ABCA4* showed significant associations in NHW probands. *MSX1*, *SPRY1*, and *ABCA4* have been identified as candidate genes for NSCLP based on animal models or genome-wide association studies (GWAS) (Beaty et al. 2016; Beaty et al. 2010; Dixon et al. 2011); their associations with NSCLP have been further and independently validated in multiple populations (Beaty et al. 2013; Chen et al. 2012; Fontoura et al. 2012; Li et al. 2017; Lin et al. 2008; Suzuki et al. 2009; Yildirim et al. 2012; Yuan et al. 2011). However, this is the first report showing the association of noncoding enhancer variation affecting these genes with NSCLP. Moreover, the associated enhancer variants are predicted to have a functional role in the regulation of their target genes, due to differential allele-specific binding affinities to transcription factors such as *LEF1*, and *TCF4*, involved in important cellular functions during cell cycle regulation and craniofacial development (Liu et al. 2010).

To overcome the challenges associated with interpretation of allelic heterogeneity and low-frequency variants in NSCLP, group-wise burden association analysis was performed and revealed novel significant and suggestive associations for 2-, 3-, 4-, and 5-enhancer variant combinations involving *RUNX2*, *ABCA4*, *BMP4* (bone morphogenetic protein 4, MIM #112262), *COL1A2* (collagen type I alpha 2 chain, MIM #120160), *MESDC1* (mesoderm development candidate 1, MIM #615466), and *MNI* (meningioma chromosome region 1, MIM #156100) ($p = 1.7 \times 10^{-5}$) (Table 5). Interestingly, whereas no single-variant associations were found between *RUNX2* or *ABCA4* enhancer variants and NSCLP, these gene variants were always significantly associated with NSCLP when combined with additional variants in other genes. These findings support a polygenic inheritance model for NSCLP (Carter 1969) and corroborate numerous studies supporting the hypothesis that multiple rare and/or low-frequency variants (each with a modest marginal effect) together explain a large proportion of the genetic basis for common complex diseases and disorders (Bodmer and Bonilla 2008; Gorlov et al. 2008). In recent years, various methods have been proposed to jointly analyze a set of variants in order to test for groupwise association with disease status (Lee et al. 2014; Li and Leal 2008; Madsen and Browning 2009; Sul et al. 2011). These methods were developed to evaluate the cumulative effects of multiple variants in a biologically relevant region, instead of testing the effects of single variants, thereby increasing power. Here, a weighted sum statistics method that has been widely used in studies of diseases in which genetic heterogeneity is a common scenario, and for diseases in which rare variants together may explain a proportion of the underlying genetic architecture was applied (Madsen and Browning 2009).

BMP4, *COL1A2*, and *RUNX2* genes are well-known for their roles in skeletal and/or craniofacial development, which when mutated, result in craniofacial abnormalities and/or NSCLP (Beaty et al. 2010; Sull et al. 2008; Suzuki et al. 2009; Wu et al. 2014; Zhou et al. 2013). Intriguingly, *ABCA4* is not expressed in the palatal tissues of mouse embryos and its potential association with NSCLP appears to be driven by variants identified within *ABCA4* that are in strong linkage disequilibrium with true causal variants in other genes (Beaty et al. 2010; Beaty et al. 2013; Fontoura et al. 2012). It has also been suggested that the presence of craniofacial enhancers within *ABCA4* intronic regions affects the expression of *ARHGAP29*, in which coding variants were also associated with NSCLP (Leslie et al. 2012). More recently, the effects of craniofacial enhancer variation on perturbing *ARHGAP29* expression and transcription factor binding properties have been shown (Liu et al. 2017).

Previous evidence also supports the involvement of *MNI* in NSCLP. Overlapping deletions in chromosome 22q12.1, which includes the *MNI* gene, has been reported as the likely cause of familial NSCLP and developmental delay, and corroborate results of mouse studies showing that *Mni* haploinsufficiency results in craniofacial defects and secondary cleft palate (Beck et al. 2015). *MESDC1* encodes a protein regulated by a microRNA (miR-574-3) and is thought to have an oncogenic function in human bladder cancer (Tatarano et al. 2012). In mouse, this gene is expressed early on during embryonic development and in the adult. Knockdown of *MESDC1* or overexpression of its regulator miR-574-3p reduced cell proliferation, migration, and invasion and induced apoptosis (Tatarano et al. 2012). While its role in craniofacial development and/or NSCLP is not yet known, the observed association

with *MESDC1* is intriguing as increasing evidence points to an important role of microRNAs in the etiology of NSCLP, and suggest that variation in microRNA processing genes may contribute to NSCLP risk (Xu et al. 2018).

Pathway analysis was performed in order to better understand the relationship between the genes identified as associated with NSCLP through the groupwise burden analysis. Using Ontology Fingerprint deep-learning method (Tsoi et al. 2009b), based on integrated multi-omics data, similarity scores between two genes were correlated with their biological relevance with the assumption that the higher the score, the stronger the biological relevance between the genes (Qin et al. 2014b; Tsoi et al. 2009b). Ontology Fingerprints method has been widely used to evaluate results of GWAS, to infer active signaling pathways in cancer and identify novel pathway players (Chen et al. 2014; Qin et al. 2014b; Qin et al. 2012a). Our results identified gene clusters and a novel genetic network (Fig. 2) showing how *COL1A2*, *BMP4* and *RUNX2* are connected to each other through other genes (e.g. *TGFB1*, *BMP2* and *MMP13*) and through enhancers of other genes which are also involved in craniofacial/palatal development and clefting in humans and/or mice (Blanton et al. 2004b; Blavier et al. 2001; Fakhouri et al. 2017; Letra et al. 2012b). These observations are in agreement with previous NSCLP studies, in which *in silico* reconstruction of NSCLP networks based on integrated multi-omics data suggested likely functional relationships of 23 of the 43 curated NSCLP-implicated genes (Razaghi-Moghadam et al. 2019). Our findings also corroborate additional studies suggesting that genetic interactions involving *MSX1*, *SPRY1* and *ABCA4* and other genes as well, should be considered important factors involved in NSCLP etiology (Leslie et al. 2012; Letra et al. 2012a; Velazquez-Aragon et al. 2016).

In a recent study, more than 6,000 enhancer regions were reported as active in early human craniofacial development (Wilderman et al. 2018). Many of these enhancer regions were previously reported as strongly enriched with genetic associations for NSCLP, and many of the associated genes were identified as having bivalent promoter status (i.e., with overlapping domains of activation and repressive signals) during human and/or mouse craniofacial development or both (Wilderman et al. 2018). In contrast, despite our modest sample size and only evaluating twenty craniofacial enhancer regions, our results are in accordance with these previous findings. In this study, variants in ten genes with positive/suggestive associations to NSCLP (*MSX1*, *SIX1*, *RUNX2*, *SATB2*, *TWIST1*) or in predicted binding partners of an associated gene (*CEBPB*, *HOXA9*, *LEF1*, *MEIS3*, *TFAP2A*) were found to overlap with those previously identified as having bivalent promoter domains in human and mouse craniofacial tissues (Wilderman et al. 2018). Interestingly, *ABCA4*, *BMP4*, and *SPRY1*, long considered to be NSCLP genes and for which craniofacial enhancers were described in mouse, were not identified as having functional enhancer activity in humans in Wilderman et al. However, genetic interactions were found between these genes implicating their functional significance in NSCLP.

While additional studies are needed to confirm the present findings, our results further support the involvement of noncoding enhancer variation, individually and in concert, in the etiology of NSCLP. Several avenues for future NSCLP research appear promising. Genotyping potential targets of craniofacial enhancers could provide additional insights into

dissecting the interactions among these genomic elements and how they contribute to craniofacial development. Functional characterization of enhancer variants will also improve understanding of the role of these elements and/or their target genes in disease; the challenge is that enhancers may be located at an arbitrary distance from their target genes and/or inside other genes (Li et al. 2012). Given the known complexity in the genetic etiology of NSCLP, studies should consider combining single-gene and network-based discovery approaches, together with functional characterization of identified variants whenever possible to improve NSCLP gene discovery.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank the NSCLP families for participating in this study, Maria Elena Serna, Rosa Martinez, for recruiting patients and families and for database management. This study was supported by grants from the National Institute for Dental and Craniofacial Research (R01-DE11931, to JTH and SHB), National Center for Advancing Translational Sciences (1UL1-TR003167, to WJZ), and Cancer Prevention and Research Institute of Texas (CPRIT-RP170668, to WJZ).

References

- Attanasio C, Nord AS, Zhu Y, Blow MJ, Li Z, Liberton DK, Morrison H, Plajzer-Frick I, Holt A, Hosseini R, Phouanenavong S, Akiyama JA, Shoukry M, Afzal V, Rubin EM, FitzPatrick DR, Ren B, Hallgrímsson B, Pennacchio LA, Visel A (2013) Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* 342: 1241006. doi: 10.1126/science.1241006 [PubMed: 24159046]
- Barba A, Urbina C, Maili L, Greives MR, Blackwell SJ, Mulliken JB, Chiquet B, Blanton SH, Hecht JT, Letra A (2019) Association of IFT88 gene variants with nonsyndromic cleft lip with or without cleft palate. *Birth Defects Res* 111: 659–665. doi: 10.1002/bdr2.1504 [PubMed: 30953423]
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263–5. [PubMed: 15297300]
- Beaty TH, Marazita ML, Leslie EJ (2016) Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. *F1000Res* 5: 2800. doi: 10.12688/f1000research.9503.1 [PubMed: 27990279]
- Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Liang KY, Wu T, Murray T, Fallin MD, Redett RA, Raymond G, Schwender H, Jin SC, Cooper ME, Dunnwald M, Mansilla MA, Leslie E, Bullard S, Lidral AC, Moreno LM, Menezes R, Vieira AR, Petrin A, Wilcox AJ, Lie RT, Jabs EW, Wu-Chou YH, Chen PK, Wang H, Ye X, Huang S, Yeow V, Chong SS, Jee SH, Shi B, Christensen K, Melbye M, Doheny KF, Pugh EW, Ling H, Castilla EE, Czeizel AE, Ma L, Field LL, Brody L, Pangilinan F, Mills JL, Molloy AM, Kirke PN, Scott JM, Arcos-Burgos M, Scott AF (2010) A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nat Genet* 42: 525–9. doi: 10.1038/ng.580 [PubMed: 20436469]
- Beaty TH, Taub MA, Scott AF, Murray JC, Marazita ML, Schwender H, Parker MM, Hetmanski JB, Balakrishnan P, Mansilla MA, Mangold E, Ludwig KU, Noethen MM, Rubini M, Elcioglu N, Ruczinski I (2013) Confirming genes influencing risk to cleft lip with/without cleft palate in a case-parent trio study. *Hum Genet* 132: 771–81. doi: 10.1007/s00439-013-1283-6 [PubMed: 23512105]
- Beck M, Peterson JF, McConnell J, McGuire M, Asato M, Losee JE, Surti U, Madan-Khetarpal S, Rajkovic A, Yatsenko SA (2015) Craniofacial abnormalities and developmental delay in two families with overlapping 22q12.1 microdeletions involving the MN1 gene. *Am J Med Genet A* 167A: 1047–53. doi: 10.1002/ajmg.a.36839 [PubMed: 25810350]

- Blanton SH, Bertin T, Patel S, Stal S, Mulliken JB, Hecht JT (2004a) Nonsyndromic cleft lip and palate: four chromosomal regions of interest. *Am J Med Genet A* 125A: 28–37. doi: 10.1002/ajmg.a.20423 [PubMed: 14755463]
- Blanton SH, Bertin T, Serna ME, Stal S, Mulliken JB, Hecht JT (2004b) Association of chromosomal regions 3p21.2, 10p13, and 16p13.3 with nonsyndromic cleft lip and palate. *Am J Med Genet A* 125A: 23–7. doi: 10.1002/ajmg.a.20426 [PubMed: 14755462]
- Blavier L, Lazaryev A, Groffen J, Heisterkamp N, DeClerck YA, Kaartinen V (2001) TGF-beta3-induced palatogenesis requires matrix metalloproteinases. *Mol Biol Cell* 12: 1457–66. doi: 10.1091/mbc.12.5.1457 [PubMed: 11359935]
- Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40: 695–701. doi: 10.1038/ng.f.136 [PubMed: 18509313]
- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 22: 1790–7. doi: 10.1101/gr.137323.112 [PubMed: 22955989]
- Carter CO (1969) Polygenic inheritance and common diseases. *Lancet* 1: 1252–6. doi: 10.1016/s0140-6736(69)92130-8 [PubMed: 4182424]
- Chen G, Tsoi A, Xu H, Zheng WJ (2018) Predict effective drug combination by deep belief network and Ontology Fingerprints. *Journal of biomedical informatics*.
- Chen G, Zhao J, Cohen T, Tao C, Sun J, Xu H, Bernstam EV, Lawson A, Zeng J, Johnson AM (2015) Using Ontology Fingerprints to disambiguate gene name entities in the biomedical literature. *Database* 2015.
- Chen G, Zhao J, Cohen T, Tao C, Sun J, Xu H, Bernstam EV, Lawson A, Zeng J, Johnson AM, Holla V, Bailey AM, Meric-Bernstam F, Zheng WJ (2014) Using Ontology Fingerprints to disambiguate gene name entities in the biomedical literature. *Database* bav034: doi: 10.1093/database/bav034.
- Chen Q, Wang H, Hetmanski JB, Zhang T, Ruczinski I, Schwender H, Liang KY, Fallin MD, Redett RJ, Raymond GV, Wu Chou YH, Chen PK, Yeow V, Chong SS, Cheah FS, Jabs EW, Scott AF, Beaty TH (2012) BMP4 was associated with NSCL/P in an Asian population. *PLoS One* 7: e35347. doi: 10.1371/journal.pone.0035347 [PubMed: 22514733]
- Chiquet BT, Blanton SH, Burt A, Ma D, Stal S, Mulliken JB, Hecht JT (2008) Variation in WNT genes is associated with non-syndromic cleft lip with or without cleft palate. *Hum Mol Genet* 17: 2212–8. doi: 10.1093/hmg/ddn121 [PubMed: 18413325]
- Dickel DE, Ypsilanti AR, Pla R, Zhu Y, Barozzi I, Mannion BJ, Khin YS, Fukuda-Yuzawa Y, Plajzer-Frick I, Pickle CS, Lee EA, Harrington AN, Pham QT, Garvin TH, Kato M, Osterwalder M, Akiyama JA, Afzal V, Rubenstein JLR, Pennacchio LA, Visel A (2018) Ultraconserved Enhancers Are Required for Normal Development. *Cell* 172: 491–499 e15. doi: 10.1016/j.cell.2017.12.017 [PubMed: 29358049]
- Dixon MJ, Marazita ML, Beaty TH, Murray JC (2011) Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet* 12: 167–78. doi: 10.1038/nrg2933 [PubMed: 21331089]
- Fakhouri WD, Metwalli K, Naji A, Bakhiet S, Quispe-Salcedo A, Nitschke L, Kousa YA, Schutte BC (2017) Intercellular Genetic Interaction Between *Irf6* and *Twist1* during Craniofacial Development. *Sci Rep* 7: 7129. doi: 10.1038/s41598-017-06310-z [PubMed: 28769044]
- Fontoura C, Silva RM, Granjeiro JM, Letra A (2012) Further evidence of association of the *ABCA4* gene with cleft lip/palate. *Eur J Oral Sci* 120: 553–7. doi: 10.1111/eos.12001 [PubMed: 23167473]
- Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. *Nature* 526: 68–74. doi: 10.1038/nature15393 [PubMed: 26432245]
- Gorlin RJ, Cohen MM, Hennekam RCM (2001) *Syndromes of the head and neck*, 4th edn Oxford University Press, Oxford [England] ; New York
- Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI (2008) Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 82: 100–12. doi: 10.1016/j.ajhg.2007.09.006 [PubMed: 18179889]
- Hashmi SS, Waller DK, Langlois P, Canfield M, Hecht JT (2005) Prevalence of nonsyndromic oral clefts in Texas: 1995–1999. *Am J Med Genet A* 134: 368–72. doi: 10.1002/ajmg.a.30618 [PubMed: 15779018]

- Lee S, Abecasis GR, Boehnke M, Lin X (2014) Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 95: 5–23. doi: 10.1016/j.ajhg.2014.06.009 [PubMed: 24995866]
- Leslie EJ, Mansilla MA, Biggs LC, Schuette K, Bullard S, Cooper M, Dunnwald M, Lidral AC, Marazita ML, Beaty TH, Murray JC (2012) Expression and mutation analyses implicate ARHGAP29 as the etiologic gene for the cleft lip with or without cleft palate locus identified by genome-wide association on chromosome 1p22. *Birth Defects Res A Clin Mol Teratol* 94: 934–42. doi: 10.1002/bdra.23076 [PubMed: 23008150]
- Letra A, Fakhouri W, Fonseca RF, Menezes R, Kempa I, Prasad JL, McHenry TG, Lidral AC, Moreno L, Murray JC, Daack-Hirsch S, Marazita ML, Castilla EE, Lace B, Orioli IM, Granjeiro JM, Schutte BC, Vieira AR (2012a) Interaction between IRF6 and TGFA genes contribute to the risk of nonsyndromic cleft lip/palate. *PLoS One* 7: e45441. doi: 10.1371/journal.pone.0045441 [PubMed: 23029012]
- Letra A, Silva RM, Motta LG, Blanton SH, Hecht JT, Granjeiro JM, Vieira AR (2012b) Association of MMP3 and TIMP2 promoter polymorphisms with nonsyndromic oral clefts. *Birth Defects Res A Clin Mol Teratol* 94: 540–8. doi: 10.1002/bdra.23026 [PubMed: 22730240]
- Li B, Leal SM (2008) Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 83: 311–21. doi: 10.1016/j.ajhg.2008.06.024 [PubMed: 18691683]
- Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung WK, Snyder M, Ruan Y (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148: 84–98. doi: 10.1016/j.cell.2011.12.014 [PubMed: 22265404]
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–60. doi: 10.1093/bioinformatics/btp324 [PubMed: 19451168]
- Li YH, Yang J, Zhang JL, Liu JQ, Zheng Z, Hu DH (2017) BMP4 rs17563 polymorphism and nonsyndromic cleft lip with or without cleft palate: A meta-analysis. *Medicine (Baltimore)* 96: e7676. doi: 10.1097/MD.00000000000007676 [PubMed: 28767592]
- Lin JY, Chen YJ, Huang YL, Tang GP, Zhang L, Deng B, Li M, Ma H, Luan RS (2008) Association of bone morphogenetic protein 4 gene polymorphisms with nonsyndromic cleft lip with or without cleft palate in Chinese children. *DNA Cell Biol* 27: 601–5. doi: 10.1089/dna.2008.0777 [PubMed: 18771417]
- Liu B, Rooker SM, Helms JA (2010) Molecular control of facial morphology. *Semin Cell Dev Biol* 21: 309–13. doi: 10.1016/j.semcdb.2009.09.002 [PubMed: 19747977]
- Liu H, Leslie EJ, Carlson JC, Beaty TH, Marazita ML, Lidral AC, Cornell RA (2017) Identification of common non-coding variants at 1p22 that are functional for non-syndromic orofacial clefting. *Nat Commun* 8: 14759. doi: 10.1038/ncomms14759 [PubMed: 28287101]
- Madsen BE, Browning SR (2009) A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* 5: e1000384. doi: 10.1371/journal.pgen.1000384 [PubMed: 19214210]
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297–303. doi: 10.1101/gr.107524.110 [PubMed: 20644199]
- Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* 18: 333–4. doi: 10.1093/bioinformatics/18.2.333 [PubMed: 11847087]
- Pennacchio LA, Rubin EM (2001) Genomic strategies to identify mammalian regulatory sequences. *Nat Rev Genet* 2: 100–9. doi: 10.1038/35052548 [PubMed: 11253049]
- Qin T, Matmati N, Tsoi LC, Mohanty BK, Gao N, Tang J, Lawson AB, Hannun YA, Zheng WJ (2014a) Finding pathway-modulating genes from a novel Ontology Fingerprint-derived gene network. *Nucleic acids research* 42: e138–e138. [PubMed: 25063300]

- Qin T, Matmati N, Tsoi LC, Mohanty BK, Gao N, Tang J, Lawson AB, Hannun YA, Zheng WJ (2014b) Finding pathway-modulating genes from a novel Ontology Fingerprint-derived gene network. *Nucleic Acids Res* 42: e138. doi: 10.1093/nar/gku678 [PubMed: 25063300]
- Qin T, Tsoi LC, Sims K, Lu X, Zheng WJ (2012a) Signaling network prediction by the Ontology Fingerprint enhanced Bayesian network. *BMC Systems Biology* 6: S3.
- Qin T, Tsoi LC, Sims KJ, Lu X, Zheng WJ (2012b) Signaling network prediction by the Ontology Fingerprint enhanced Bayesian network. *BMC systems biology* 6: S3.
- Razaghi-Moghadam Z, Namipashaki A, Farahmand S, Ansari-Pour N (2019) Systems genetics of nonsyndromic orofacial clefting provides insights into its complex aetiology. *Eur J Hum Genet* 27: 226–234. doi: 10.1038/s41431-018-0263-7 [PubMed: 30254216]
- Shaffer JR, LeClair J, Carlson JC, Feingold E, Buxo CJ, Christensen K, Deleyiannis FWB, Field LL, Hecht JT, Moreno L, Orioli IM, Padilla C, Vieira AR, Wehby GL, Murray JC, Weinberg SM, Marazita ML, Leslie EJ (2019) Association of low-frequency genetic variants in regulatory regions with nonsyndromic orofacial clefts. *Am J Med Genet A* 179: 467–474. doi: 10.1002/ajmg.a.61002 [PubMed: 30582786]
- Sul JH, Han B, He D, Eskin E (2011) An optimal weighted aggregated association test for identification of rare variants involved in common diseases. *Genetics* 188: 181–8. doi: 10.1534/genetics.110.125070 [PubMed: 21368279]
- Sull JW, Liang KY, Hetmanski JB, Fallin MD, Ingersoll RG, Park J, Wu-Chou YH, Chen PK, Chong SS, Cheah F, Yeow V, Park BY, Jee SH, Jabs EW, Redett R, Jung E, Ruczinski I, Scott AF, Beaty TH (2008) Differential parental transmission of markers in RUNX2 among cleft case-parent trios from four populations. *Genet Epidemiol* 32: 505–12. doi: 10.1002/gepi.20323 [PubMed: 18357615]
- Suzuki S, Marazita ML, Cooper ME, Miwa N, Hing A, Jugessur A, Natsume N, Shimozato K, Ohbayashi N, Suzuki Y, Niimi T, Minami K, Yamamoto M, Altannamar TJ, Erkhembaatar T, Furukawa H, Daack-Hirsch S, L'Heureux J, Brandon CA, Weinberg SM, Neiswanger K, Deleyiannis FW, de Salamanca JE, Vieira AR, Lidral AC, Martin JF, Murray JC (2009) Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. *Am J Hum Genet* 84: 406–11. doi: 10.1016/j.ajhg.2009.02.002 [PubMed: 19249007]
- Tatarano S, Chiyomaru T, Kawakami K, Enokida H, Yoshino H, Hidaka H, Nohata N, Yamasaki T, Gotanda T, Tachiwada T, Seki N, Nakagawa M (2012) Novel oncogenic function of mesoderm development candidate 1 and its regulation by MiR-574–3p in bladder cancer cell lines. *Int J Oncol* 40: 951–9. doi: 10.3892/ijo.2011.1294 [PubMed: 22179486]
- Thieme F, Ludwig KU (2017) The Role of Noncoding Genetic Variation in Isolated Orofacial Clefts. *J Dent Res* 96: 1238–1247. doi: 10.1177/0022034517720403 [PubMed: 28732180]
- Tsoi LC, Boehnke M, Klein RL, Zheng WJ (2009a) Evaluation of genome-wide association study results through development of ontology fingerprints. *Bioinformatics* 25: 1314–1320. [PubMed: 19349285]
- Tsoi LC, Boehnke M, Klein RL, Zheng WJ (2009b) Evaluation of genome-wide association study results through development of ontology fingerprints. *Bioinformatics* 25: 1314–20. doi: btp158 [pii] 10.1093/bioinformatics/btp158 [PubMed: 19349285]
- Velazquez-Aragon JA, Alcantara-Ortigoza MA, Estandia-Ortega B, Reyna-Fabian ME, Mendez-Adame CD, Gonzalez-Del Angel A (2016) Gene Interactions Provide Evidence for Signaling Pathways Involved in Cleft Lip/Palate in Humans. *J Dent Res* 95: 1257–64. doi: 10.1177/0022034516647034 [PubMed: 27154735]
- Visel A, Akiyama JA, Shoukry M, Afzal V, Rubin EM, Pennacchio LA (2009a) Functional autonomy of distant-acting human enhancers. *Genomics* 93: 509–13. doi: 10.1016/j.ygeno.2009.02.002 [PubMed: 19268701]
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA (2009b) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457: 854–8. doi: 10.1038/nature07730 [PubMed: 19212405]
- Visel A, Minovitsky S, Dubchak I, Pennacchio LA (2007) VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* 35: D88–92. doi: 10.1093/nar/gkl822 [PubMed: 17130149]

- Ward LD, Kellis M (2016) HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* 44: D877–81. doi: 10.1093/nar/gkv1340 [PubMed: 26657631]
- Wilderman A, VanOudenhove J, Kron J, Noonan JP, Cotney J (2018) High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep* 23: 1581–1597. doi: 10.1016/j.celrep.2018.03.129 [PubMed: 29719267]
- Wu H, Nord AS, Akiyama JA, Shoukry M, Afzal V, Rubin EM, Pennacchio LA, Visel A (2014) Tissue-specific RNA expression marks distant-acting developmental enhancers. *PLoS Genet* 10: e1004610. doi: 10.1371/journal.pgen.1004610 [PubMed: 25188404]
- Wyszynski DF (2002) Cleft lip and palate : from origin to treatment. Oxford University Press, Oxford ; New York
- Xu M, Ma L, Lou S, Du Y, Yin X, Zhang C, Fan L, Wang H, Wang Z, Zhang W, Wang L, Pan Y (2018) Genetic variants of microRNA processing genes and risk of non-syndromic orofacial clefts. *Oral Dis* 24: 422–428. doi: 10.1111/odi.12741 [PubMed: 28833944]
- Yildirim M, Seymen F, Deeley K, Cooper ME, Vieira AR (2012) Defining predictors of cleft lip and palate risk. *J Dent Res* 91: 556–61. doi: 10.1177/0022034512444928 [PubMed: 22496123]
- Yuan Q, Blanton SH, Hecht JT (2011) Association of ABCA4 and MAFB with non-syndromic cleft lip with or without cleft palate. *Am J Med Genet A* 155A: 1469–71. doi: 10.1002/ajmg.a.33940 [PubMed: 21567910]
- Zhou J, Gao Y, Lan Y, Jia S, Jiang R (2013) Pax9 regulates a molecular network involving BMP4, Fgf10, Shh signaling and the Osr2 transcription factor to control palate morphogenesis. *Development* 140: 4709–18. doi: 10.1242/dev.099028 [PubMed: 24173808]

Fig. 1. Genomic position and activity of the associated craniofacial enhancers.

a. Position of the associated enhancer mm427 variant (rs1994828) and its cis-regulatory target gene *MSX1*. The variant is located in a histone-mark rich region. **b.** The associated enhancer, mm427/*MSX1*, is active in the facial mesenchyme and nose in the mouse embryo at E11.5. **c.** Position of the associated enhancer hs1582 variant (rs6534405) and its cis-regulatory target gene *SPRY1*. **d.** The associated enhancer, hs1582/*SPRY1*, is expressed in the craniofacial region, particularly the hindbrain in the mouse embryo at E11.5.

Fig. 2. OntoFin Net-generated genetic network for inferring gene-gene and gene-pathway relationship.

Gene clusters identified from OntoFin Net showing how *COL1A2*, *BMP4* and *RUNX2* are connected to each other.

Table 1.

Craniofacial enhancers prioritized for this study

Enhancer Id. ^a	Chr.	Coordinates ^a	Nearest gene of interest	Enhancer position relative to nearest gene	Positive craniofacial tissue expression ^a	Additional evidence in craniofacial development
Human						
hs1461	6	45847301 – 45850554	<i>RUNX2</i> **	~328 kb downstream	nose	<i>RUNX2</i> variants associated with NSCLP; <i>Runx2</i> deficiency results in defective frontal bone closure and absence of secondary palate
hs1582	4	124383428 – 124386454	<i>SPRY1</i> **	~58 kb downstream	hindbrain, midbrain, facial mesenchyme	<i>SPRY1</i> variants associated with NSCLP; <i>Spry1</i> -deficient mice have craniofacial defects and cleft palate
hs1602	14	61120199 – 61122451	<i>SIX1</i>	~4 kb upstream	dorsal root ganglion, branchial arch, facial mesenchyme	<i>Six1</i> ^{-/-} mice have craniofacial malformations in the nasal cavity, ear, and jaw structures
hs1603	12	23211482 – 23214744	<i>SOX5</i>	~500 kb downstream	branchial arch, nose	Sox5-deficient embryos display patterning defects characteristic of altered BMP signaling including loss of neural crest and epidermis and lack of expansion of the neural plate
hs1626	7	93975962 – 93978880	<i>COL1A2</i>	~82 kb upstream	branchial arch, facial mesenchyme	Involved in focal adhesion and odontogenesis pathways, as well as ECM-receptor interaction pathways
hs1635	2	220771085 – 220773279	<i>EPHA4</i> *	~1.5 Mb downstream	facial and palatal mesenchyme	Deletion involving <i>EPHA4</i> reported in syndromic cleft palate case
hs2307	7	18874272 – 18875369	<i>TWIST1</i>	~270 kb downstream	branchial arch	<i>Twist1</i> ^{-/-} mice have cleft palate; <i>TWIST1</i> mutations cause Saethre-Chotzen syndrome
Mouse						
mm387	14	54690378 – 54691482	<i>Bmp4</i> ***	~220 kb upstream	branchial arch, nose, facial mesenchyme	<i>Bmp4</i> ^{-/-} mice have cleft palate; <i>BMP4</i> variants associated with NSCLP subphenotypes
mm400	7	94073344 – 94074914	<i>Col1a2</i>	~12 kb downstream	branchial arch, nose, facial mesenchyme	
mm427	4	4756237 – 4758006	<i>Msx1</i> ***	~77 kb downstream	nose, facial mesenchyme	<i>Msx1</i> ^{-/-} mice have cleft palate and dental defects; <i>MSX1</i> mutations found in NSCLP patients; enhancer deleted near <i>Msx1</i> resulted in altered craniofacial shape
mm428		4736431 – 4737277		~93 kb downstream	branchial arch, facial mesenchyme, nose	
mm435	1	94580867 – 94583522	<i>Abca4</i> **	intragenic (intron 1)	forebrain, nose, facial mesenchyme	Identified in GWAS of NSCLP; enhancers within <i>Abca4</i> shown to affect expression of <i>Arhgap29</i>
mm613	2	199557473 – 199558769	<i>Satb2</i> **	~506 kb downstream	branchial arch, ear	<i>Satb1</i> ^{-/-} knockout mice have cleft palate; <i>SATB2</i> mutations associated with cleft palate
mm729	15	81382728 – 81390271	<i>Mesdc1</i>	~65 kb downstream	branchial arch, facial mesenchyme	Modulates <i>Lrp6</i> -mediated Wnt signaling
mm646	1	210464411 – 210467294	<i>Hhat</i> *	~27 kb upstream	somite, forebrain, branchial arch, nose, facial mesenchyme	<i>Hhat</i> loss-of-function leads to holoprosencephaly and severe craniofacial defects
mm621	22	27972469 – 27973657	<i>Mn1</i> *	~120 kb downstream	branchial arch	<i>Mn1</i> ^{-/-} mice have defects in posterior palatal growth and cleft palate

Enhancer Id. ^a	Chr.	Coordinates ^a	Nearest gene of interest	Enhancer position relative to nearest gene	Positive craniofacial tissue expression ^a	Additional evidence in craniofacial development
mm622	6	27934200 – 27936053	<i>Runx2</i> ^{**}	~144 kb downstream	branchial arch	<i>RUNX2</i> variants associated with NSCLP; <i>Runx2</i> deficiency results in defective frontal bone closure and absence of secondary palate
mm921		27816201 – 27817687		~218 kb downstream	facial mesenchyme	
mm924		45400539 – 45402148		intragenic (intron 4)	branchial arch, facial mesenchyme	
mm657		45400539 – 45402148		intragenic (intron 2)	branchial arch	

[illegible]

* gene is associated with orofacial clefting in animal studies

** gene is associated with orofacial clefting human studies

gene is associated with orofacial clefting in both human and animal studies

Table 2.

SNV association results

SNV	Enhancer Id.	Nearest Gene	Ethnicity ^a	Associated Allele	Ratio Case:Control	Frequency Case:Control ^b	p-value	Odds Ratio (95% CI) ^{***}
rs1994828	mm427	<i>MSX1</i>	NHW	C	390:28, 663:343	0.933 : 0.659	7.68×10^{-27} *	7.2 (4.8–11.2)
			Hispanic	C	295:25, 87:41	0.922 : 0.680	6.41×10^{-11} *	5.6 (3.1–10.1)
rs6534405	hs1582	<i>SPRY1</i>	NHW	A	56:362, 0:1006	0.134 : 0.000	9.76×10^{-32} **	149.6 (21.7–1029) [†]

^a NHW, non-Hispanic white^b Frequency in CEU and MXL control populations* Chi-square test/Fisher exact test; threshold for statistical significance is $p < 2.2 \times 10^{-4}$

** p-value based on Fisher exact test

*** CI = confidence interval

[†] 95% confidence interval calculated based on Agresti's adjustment for cells with frequency of 0

Table 3.

Haplotype analysis results

Enhancer Id.	Nearest Gene	Haplotype ^a	Ethnicity ^b	Probability ratio Case:Control	Frequency in Case:Control ^c	p-value	Odds ratio (95% CI)
mm435	<i>ABCA4</i>	TCCTATTTGGACACCAGGGACCCGTCCGCC	NHW	10.0 : 398.0, 0.0 : 1005.0	0.025 : 0.000	$6.33 \times 10^{-7}^{**}$	27.7 (4.0– 1194.9) [†]
mm427	<i>MSX1</i>	CGGGACT	NHW	121.4 : 296.6, 71.7 : 934.3	0.290 : 0.071	$4.11 \times 10^{-28}^{*}$	5.3 (3.8– 7.4)
			Hispanic	122.9 : 197.1, 13.7 : 114.3	0.384 : 0.107	$8.53 \times 10^{-9}^{*}$	(2.7– 10.0)
hs1582	<i>SPRY1</i>	AAAGTACTATG <u>A</u> GTGA	NHW	51.3 : 358.3, 0.0 : 1005.1	0.125 : 0.000	$3.02 \times 10^{-30}^{**}$	145.3 (24.8– 5870) [†]

^aSignificant SNP in individual SNP analysis is underlined^bNHW, non-Hispanic white^cFrequency in CEU and MXL control populations^{*}Haplotype-based association test;^{**}p-values based on a Fisher exact test; Bonferroni corrected significance threshold = $p < 2.2 \times 10^{-4}$ ^{***}CI = confidence interval[†]95% confidence interval calculated based on Agresti's adjustment for cells with frequency of 0

Table 4.*In silico* predicted transcription binding sites

SNP	Enhancer Id.	Nearest Gene	Allele ^a		Haploreg	RegulomeDB	PROMO
			Ref	Alt			
rs1994828	mm427	<i>MSX1</i>	T	<u>C</u>	<i>PBX1, SPDEF, TCF4</i>	<i>LEF1</i>	<i>CEBPB, LEF1, TCF-4, TFAP2A, SRY</i>
rs6534405	hs1582	<i>SPRY1</i>	C	<u>A</u>	<i>ARID3A, MEF2A, MEIS::HOXA9, RXRA</i>	<i>ELF3, MEF2A</i>	<i>TBP, FOXP3</i>

^aRef, Reference allele; Alt, Alternate allele; risk-associated allele is underlined

Table 5.

Variant burden analysis of enhancer combinations among NHW cases and controls

Enhancer Combinations	Proportion rare SNV (MAF<0.05) ^a	All SNVs ^b	SNVs passing threshold ^c	Singletons among SNVs passing threshold ^d	P-value ^e
1-enhancer					
<i>RUNX2</i> (mm924)	0.0987	9	5	1	1.2×10^{-3}
2-enhancers					
<i>RUNX2</i> (mm657, mm924)	0.1293	12	6	1	7.2×10^{-5}
<i>RUNX2</i> (mm924), <i>MESDC1</i> (mm729)	0.1975	41	30	8	1.6×10^{-4}
3-enhancers					
<i>BMP4</i> (mm387), <i>RUNX2</i> (mm657, mm924)	0.1321	14	7	1	1.7×10^{-5}
<i>ABCA4</i> (mm435), <i>RUNX2</i> (mm657, mm924)	0.2225	37	14	3	2.4×10^{-5}
<i>ABCA4</i> (mm435), <i>RUNX2</i> (mm924), <i>MESDC1</i> (mm729)	0.2767	66	38	10	3.1×10^{-5}
4-enhancers					
<i>ABCA4</i> (mm435), <i>RUNX2</i> (mm924), <i>COL1A2</i> (hs1626), <i>MESDC1</i> (mm729)	0.3310	79	48	14	5.3×10^{-6}
<i>ABCA4</i> (mm435), <i>RUNX2</i> (mm657, mm924), <i>BMP4</i> (mm387)	0.2253	39	15	3	6.1×10^{-6}
<i>ABCA4</i> (mm435), <i>RUNX2</i> (mm657, mm924), <i>MESDC1</i> (mm729)	0.2990	69	39	10	6.8×10^{-6}
5-enhancers					
<i>ABCA4</i> (mm435), <i>RUNX2</i> (mm657, mm924), <i>COL1A2</i> (hs1626), <i>MESDC1</i> (mm729)	0.3504	82	49	14	1.2×10^{-6}
<i>ABCA4</i> (mm435), <i>RUNX2</i> (mm924), <i>COL1A2</i> (hs1626), <i>MESDC1</i> (mm729), <i>MNI</i> (mm622)	0.3324	85	50	16	2.0×10^{-6}
<i>ABCA4</i> (mm435), <i>RUNX2</i> (mm924), <i>COL1A2</i> (hs1626), <i>BMP4</i> (mm387), <i>MESDC1</i> (mm729)	0.3324	81	49	14	2.3×10^{-6}
<i>ABCA4</i> (mm435), <i>RUNX2</i> (mm657, mm924), <i>BMP4</i> (mm387), <i>MESDC1</i> (mm729)	0.3004	71	40	10	3.1×10^{-6}

^aProportion rare SNV (MAF < 0.05): Fraction of individual carrying rare variants below --max-maf, 0.05 threshold.^bAll SNVs: Number of variants within group.^cSNVs passing threshold: number of variants passing the --max-maf 0.05 threshold.^dSingletons among SNVs passing threshold: number of singletons among variants in SNVs passing threshold.^eMadsen test; significant if p-value $< 2.3 \times 10^{-6}$; p-values between 2.3×10^{-4} and 2.3×10^{-6} were considered suggestive associations