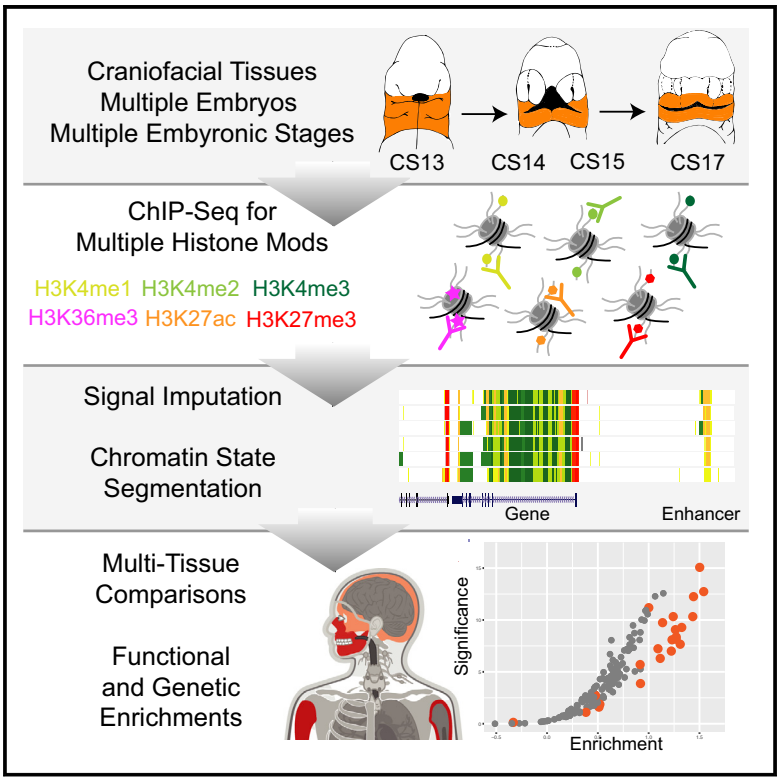
Resource

Cell Reports

# High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development

## Graphical Abstract




## Authors

Andrea Wilderman, Jennifer VanOudenhove, Jeffrey Kron, James P. Noonan, Justin Cotney

## Correspondence

cotney@uchc.edu

## In Brief

Wilderman et al. report the global identification of gene regulatory sequences active in early human craniofacial development. Systematic comparisons with over 120 different human tissues and cell types reveal shared and craniofacial-specific enhancers. Craniofacial enhancers are enriched with genetic associations for both orofacial clefting risk and face shape.

## Highlights

- Global profiling of histone modifications across early human craniofacial development
- Chromatin state segmentation reveals enhancers with craniofacial-specific activation
- Early craniofacial enhancers enriched with genetic associations for orofacial clefting
- Late craniofacial enhancers enriched with genetic associations for normal facial shape

## Data and Software Availability

GSE97752

Wilderman et al., 2018, Cell Reports *23*, 1581–1597
May 1, 2018 
https://doi.org/10.1016/j.celrep.2018.03.129

CellPress

Cell Reports
Resource



# High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development

Andrea Wilderman,[1,2] Jennifer VanOudenhove,[2] Jeffrey Kron,[2] James P. Noonan,[3,4] and Justin Cotney[2,5,6,*]
[1]Graduate Program in Genetics and Developmental Biology, UConn Health, Farmington, CT 06030, USA
[2]Department of Genetics and Genome Sciences, UConn Health, Farmington, CT 06030, USA
[3]Department of Genetics, Yale University School of Medicine, New Haven, CT 06510, USA
[4]Kavli Institute for Neuroscience, Yale University, New Haven, CT 06520, USA
[5]Institute for Systems Genomics, University of Connecticut, Storrs, CT 06269, USA
[6]Lead Contact
*Correspondence: cotney@uchc.edu
https://doi.org/10.1016/j.celrep.2018.03.129

## SUMMARY

**Defects in patterning during human embryonic development frequently result in craniofacial abnormalities. The gene regulatory programs that build the craniofacial complex are likely controlled by information located between genes and within intronic sequences. However, systematic identification of regulatory sequences important for forming the human face has not been performed. Here, we describe comprehensive epigenomic annotations from human embryonic craniofacial tissues and systematic comparisons with multiple tissues and cell types. We identified thousands of tissue-specific craniofacial regulatory sequences and likely causal regions for rare craniofacial abnormalities. We demonstrate significant enrichment of common variants associated with orofacial clefting in enhancers active early in embryonic development, while those associated with normal facial variation are enriched near the end of the embryonic period. These data are provided in easily accessible formats for both craniofacial researchers and clinicians to aid future experimental design and interpretation of noncoding variation in those affected by craniofacial abnormalities.**

## INTRODUCTION

Formation of the craniofacial complex is an intricate process of precisely timed events that occurs relatively early in vertebrate embryonic development. For example, in human embryonic development, the majority of the events that lead to the formation of the human face and skull occur during the first 10 weeks of gestation (Schoenwolf et al., 2009). Defects in the orchestration of these events result in several different congenital abnormalities, including orofacial clefting and craniosynostosis. Worldwide, orofacial clefting is one of the most common birth defects, affecting ~1 in 700 live births (World Health Organization, 2003). The majority of those affected with these types of clefting do not have defects in other tissues or organ systems, and, thus, they are referred to as non-syndromic (Mossey and Modell, 2012). While these birth defects are largely repairable through surgical means, the financial, sociological, and psychological effects have a much broader impact and represent a significant public health burden (Boulet et al., 2009; Wehby and Cassell, 2010; Wehby et al., 2011, 2012). Screening, prevention, and non-surgical therapeutic options are thus highly desirable. The high heritability of such disorders suggests a major genetic component (Grosen et al., 2010, 2011); however, causative genetic changes have only been identified in a fraction of those affected (Beaty et al., 2016; Thieme and Ludwig, 2017).

In the past decade, numerous genome-wide association studies, copy number variant analyses, and whole-exome sequencing studies have sought to identify genetic sources of risk for craniofacial defects and normal human facial variation (Beaty et al., 2010; Bureau et al., 2014; Camargo et al., 2012; Claes et al., 2018; Conte et al., 2016; Leslie et al., 2017; Letra et al., 2010; Lidral et al., 2015; Ludwig et al., 2012, 2016, 2017; Mangold et al., 2010, 2016; Mostowska et al., 2018; Yu et al., 2017; Yuan et al., 2011). These studies identified common and rare variants associated with these phenotypes, but most are located in noncoding portions of the genome preventing functional interpretation and prioritization. Our genomes are littered with gene regulatory sequences, located primarily in intronic and intergenic sequences, which are active in a small number of tissues and/or developmental stages in humans (Roadmap Epigenomics Consortium et al., 2015). While the regulatory potential of the human genome is still not completely understood, defects in regulatory sequences can cause non-syndromic developmental defects in humans and mice (Lettice et al., 2003; Petit et al., 2016; Sagai et al., 2005; Weedon et al., 2014). Of particular interest for craniofacial abnormalities, recurrent deletions of noncoding DNA near the *SOST* and *SOX9* genes have been implicated in Van Buchem disease and Pierre Robin sequence, respectively (Balemans et al., 2002; Benko et al., 2009). These findings, coupled with the non-syndromic nature of most orofacial clefting and craniosynostosis cases (Leslie and Marazita, 2015; Timberlake et al., 2016), suggest defective gene regulatory sequences may underlie much of the incidence of craniofacial abnormalities.

Despite the common nature of such birth defects and defined windows of embryonic development in which they likely occur,

mapping of chromatin states and identification of craniofacial-specific regulatory sequences have not been addressed by large functional genomics efforts such as Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenome (Roadmap Epigenomics Consortium et al., 2015). These large-scale projects have profiled chromatin states in cultured cell types derived *in vitro* from embryonic stem cells, fetal tissues from greater than 90 days of gestation, or adult post mortem samples, but they have not examined primary embryonic tissues. The embryonic period of human development, the first 8 weeks of gestation when much of craniofacial development occurs (Schoenwolf et al., 2009), has thus far been only characterized with a smaller number of functional genomics experiments in the developing limb and cortex and cultured cranial neural crest cells (CNCCs) (Prescott et al., 2015). Particularly, comparisons of epigenomic signals in human and chimp CNCCs revealed differential utilization of regulatory sequences that may play roles in the normal formation and evolution of the human face. However, it is unclear how closely these culture models recapitulate early human craniofacial development.

The lack of primary, tissue-specific genomic annotations from this critical period of human development has impeded the identification of regulatory circuitry important for human craniofacial development, and it has prevented accurate interpretation of clinical genetic findings in patients with craniofacial disorders (Thieme and Ludwig, 2017). Without sufficient biological context, prioritization and developing of hypotheses to test genetic associations with craniofacial abnormalities are hindered (Dixon et al., 2011; Khandelwal et al., 2013; Leslie and Marazita, 2015; Rahimov et al., 2012). Here, we present a comprehensive resource of functional genomics data and predicted chromatin states for critical stages of early human craniofacial development. We have profiled multiple biochemical marks of chromatin activity in developing human craniofacial tissue samples encompassing 4.5–8 post-conception weeks. We have comprehensively compared these data with publicly available functional genomics data from 127 epigenomes. We provide annotations consistent with large consortia efforts (Roadmap Epigenomics Consortium et al., 2015) in formats easily loadable into modern genome browsers to enable exploration by other researchers without large computational effort. In total, our analyses have identified thousands of previously unknown craniofacial enhancer sequences that will enable future experimental testing of enhancer-target gene interactions in developing craniofacial tissues. More importantly, this resource will facilitate future clinical interpretation of genetic variation in the context of congenital craniofacial defects that lack clear changes in the copy number or coding sequence of genes.

## RESULTS

### Profiling of Histone Modifications in Developing Human Embryonic Craniofacial Tissue

Chromatin immunoprecipitation of post-translational histone modifications coupled with next-generation sequencing (ChIP-seq) is a powerful method to identify active regulatory sequences in a global fashion from a wide variety of biological contexts (Roadmap Epigenomics Consortium et al., 2015). Many of the regulatory elements identified by this method are specific to the biological context queried (Visel et al., 2009; Zhu et al., 2013) (i.e., tissue type or developmental stage), and they are enriched for genetic associations with disease in a relevant tissue (i.e., immune-related disorder associations in immune cell-specific enhancers) (Farh et al., 2015; Pasquali et al., 2014).

To identify regulatory sequences important for human craniofacial development, we utilized ChIP-seq of six post-translational histone modifications across multiple stages and multiple biological replicates of early human craniofacial development. We focused our efforts on histone modifications profiled by large consortia and strongly associated with multiple states of chromatin activity. We performed parallel ChIP-seq experiments on craniofacial tissues obtained from 17 individual human embryos spanning a critical window for the formation of the human orofacial apparatus (Figure 1A). Specifically, we profiled marks ranging from those associated with repression (H3K27me3), promoter activation (H3K4me3), active transcription (H3K36me3), and various states of enhancer activation (H3K4me1, H3K4me2, and H3K27ac) (Figure 1B) (Ernst et al., 2011). We profiled at least three individual human embryonic craniofacial samples from each of four distinct Carnegie stages (CSs) (CS13, CS14, CS15, and CS17) encompassing 4.5 post-conception weeks (pcw) to 6 pcw (Schoenwolf et al., 2009). We also profiled single biological samples from CS20 (8-pcw) and 10-pcw embryos (Figure 1C). We obtained over 5.3 billion ChIP-seq reads across a total of 106 datasets, with mean total reads and uniquely aligned reads per sample of 50.3 and 37.3 million, respectively, meeting guidelines proposed by ENCODE (Landt et al., 2012) (Table S1).

Overall the samples correlated well by mark and stage of development (Figures 2A, S1A, and S1B). We uniformly processed these data to identify reproducibly enriched regions for each mark within each stage. The genomic features identified by each set of enriched regions closely mirror what has previously been reported for each of these post-translational marks (Figures 2B and S1C) (Ernst et al., 2011; Zhu et al., 2013). For example, we observed very strong enrichment of H3K4me3 at promoters of genes, and we identified a large number of intronic or intergenic regions enriched for H3K27ac. When we examined all the samples for a given CS, we identified thousands of enriched regions, at each stage for each mark, that were found in at least two biological replicates (Figure 2C). Combined, these results indicated that our ChIP-seq data from primary human embryonic tissues were of high quality, reflected the previously described nature of these marks, and were likely to identify tissue-specific regulatory sequences.

### Generation of Human Craniofacial Chromatin State Segmentations

Defining enriched regions for a single histone modification such as H3K27ac has been utilized to identify active regulatory sequences from a variety of tissues, biological contexts, and different species (Cotney et al., 2013; Dickel et al., 2016; Nord et al., 2013; Reilly et al., 2015; Villar et al., 2015). However, in the absence of H3K27ac, other marks can identify active regulatory sequences, and low levels of H3K27ac may be present at enhancers that are either about to become active or are no

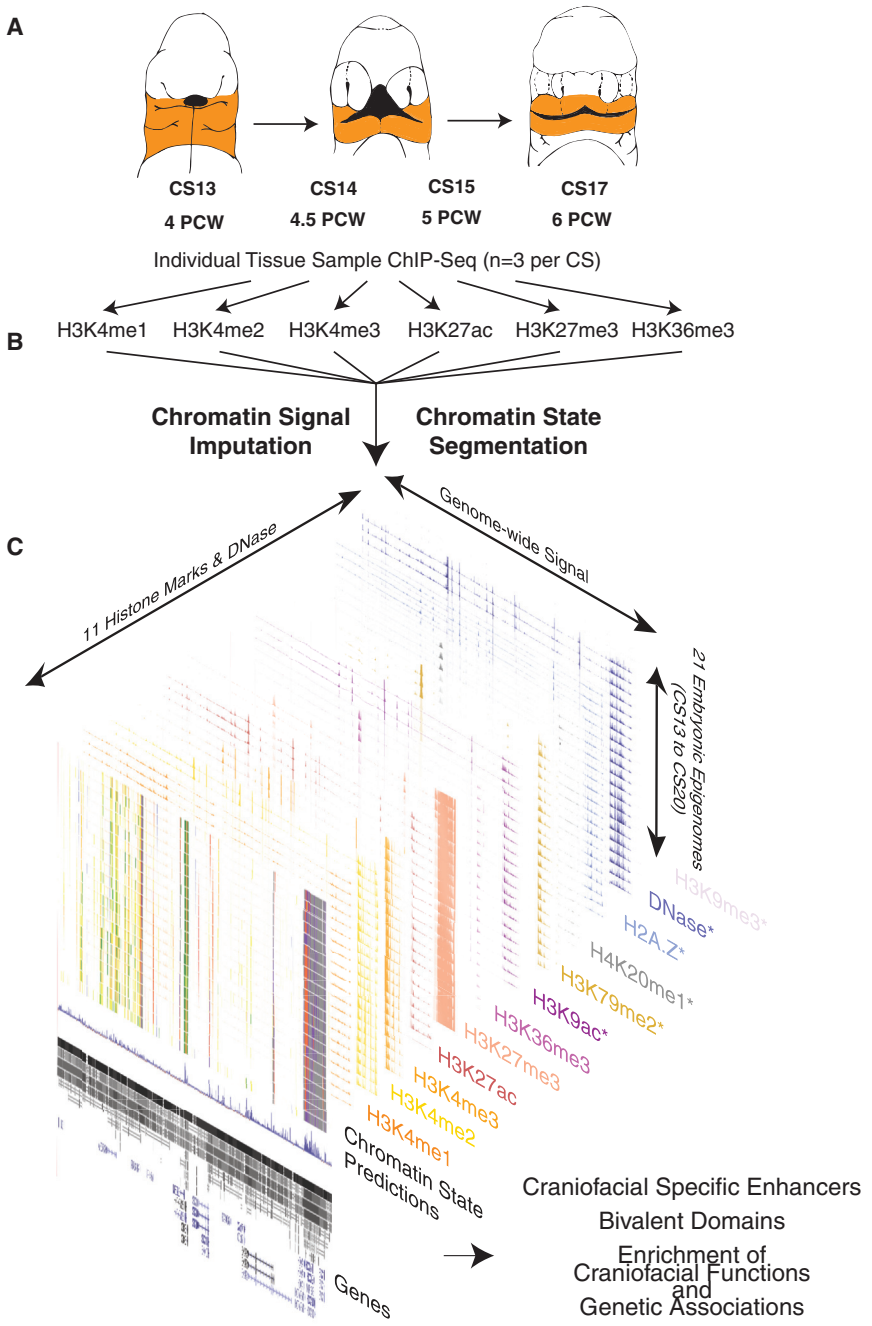



**Figure 1. Overview of Epigenomic Profiling of Early Human Craniofacial Development**

(A) Stages and craniofacial tissues (orange shading) of human embryonic development sampled in this study, indicated as individual Carnegie stages (CSs) or approximate post-conception weeks (pcw). Voids or cleavages in the embryo are indicated by black-shaded regions and do not indicate deformities.

(B) Six post-translational modifications of histones were profiled in parallel from individual human embryos via ChIP-seq.

(C) Signals from primary ChIP-seq data were imputed using ChromImpute (Ernst and Kellis, 2015) to match the 12 epigenomic signals profiled by Roadmap Epigenome (Roadmap Epigenomics Consortium et al., 2015). Asterisks indicate signals containing only imputed data. These imputed datasets were then used to predict chromatin states using a Hidden Markov Model approach (ChromHMM) (Ernst and Kellis, 2012) across the genome for each craniofacial tissue sample. These chromatin states were then used for downstream functional analyses to determine relevance for craniofacial biology and disease. (See also Figure S1 and Table S1.)

longer active (Bonn et al., 2012; Cotney et al., 2012; Kumar et al., 2016). More advanced methods, such as using machine-learning techniques and integrating multiple chromatin signals from a single tissue, allow segmentation of the genome into a more complex array of biological states (Ernst and Kellis, 2012; Hoffman et al., 2012). These techniques can identify tissue-specific and disease-relevant regulatory information in heterogenous tissues that might not be readily apparent from gene expression data or analysis of promoter activation states (Ernst et al., 2011; Hoffman et al., 2013).

To leverage such available data to identify regulatory information likely to be critical for craniofacial development, we processed our data in a uniform fashion to match those generated by Roadmap Epigenome (Roadmap Epigenomics Consortium et al., 2015). We first generated p value-based signals calculated on enrichment versus paired-input controls (Feng et al., 2012; Landt et al., 2012) for each of the six epigenomic marks we assayed. Then, using the same type of signals for 12 epigenomic marks for 127 tissues and cell types generated by Roadmap Epigenome, we imputed our data to create a uniform, directly comparable dataset of ChIP-seq signals (Ernst and Kellis, 2015). The imputed samples' signals correlated well with their primary signals and clustered generally by mark and biological function (Figures 3A, S1D, and S1E). Using the imputed craniofacial data, we then segmented the genome using ChromHMM for each embryonic sample based on previously generated models of 15, 18, and 25 states of chromatin activity (Roadmap Epigenomics Consortium et al., 2015). We identified similar numbers and proportions of segments in each state in our tissues (Figures 3B and S2). The 25-state model results showed the most similar trends across these measures, and they utilized all of the primary data generated in our study when compared to those previously generated by Roadmap Epigenome (Figures 3C, 3D, and S2);

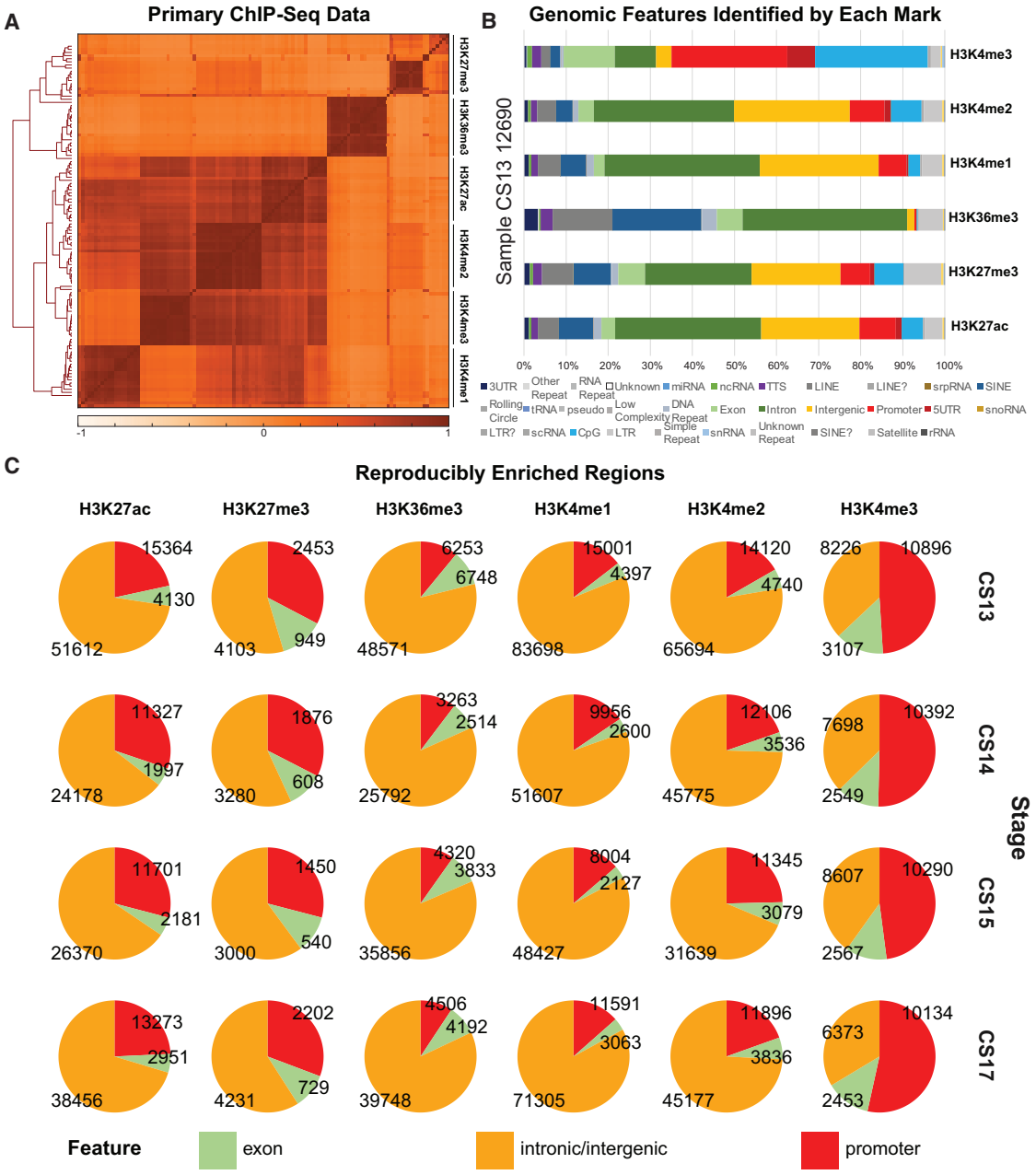



**Figure 2. Histone Modification Profiles in Human Craniofacial Development**

(A) Heatmap and hierarchical clustering of pairwise Pearson correlations for non-overlapping 10-kb bins across the human genome for 114 individual histone modification profiles from human craniofacial tissues. Relatedness of epigenomic profiles by sample is indicated by dendrogram along the vertical axes of the heatmap. Darker orange indicates positive correlation between datasets.

(B) Genomic feature annotations identified by peak calls from six histone modification profiles from the same tissue sample plotted as cumulative percentage of total peaks. Peak enrichments and genomic annotations were performed using HOMER (Heinz et al., 2010).

(C) Histone modification peaks identified in at least two separate tissue samples from the same developmental stage and annotated into three broad categories: promoter (2 kb upstream of transcription start site [TSS]), exons, and all other intronic or intergenic locations. (See also Figure S1 and Table S1.)

therefore, we focused our downstream analyses on these segmentations.

Large active chromatin domains over the promoters of genes have been shown to identify critical regulators in developing tissues and differentiation models (Bernstein et al., 2006; Cotney et al., 2012; Rada-Iglesias et al., 2012; Whyte et al., 2013). Additionally, overlapping domains of activation and repressive-associated signals (H3K27ac, H3K4me3, and H3K27ame3) are potent identifiers of genes poised for activation in embryonic stem cells, or they display restricted domains or gradients of

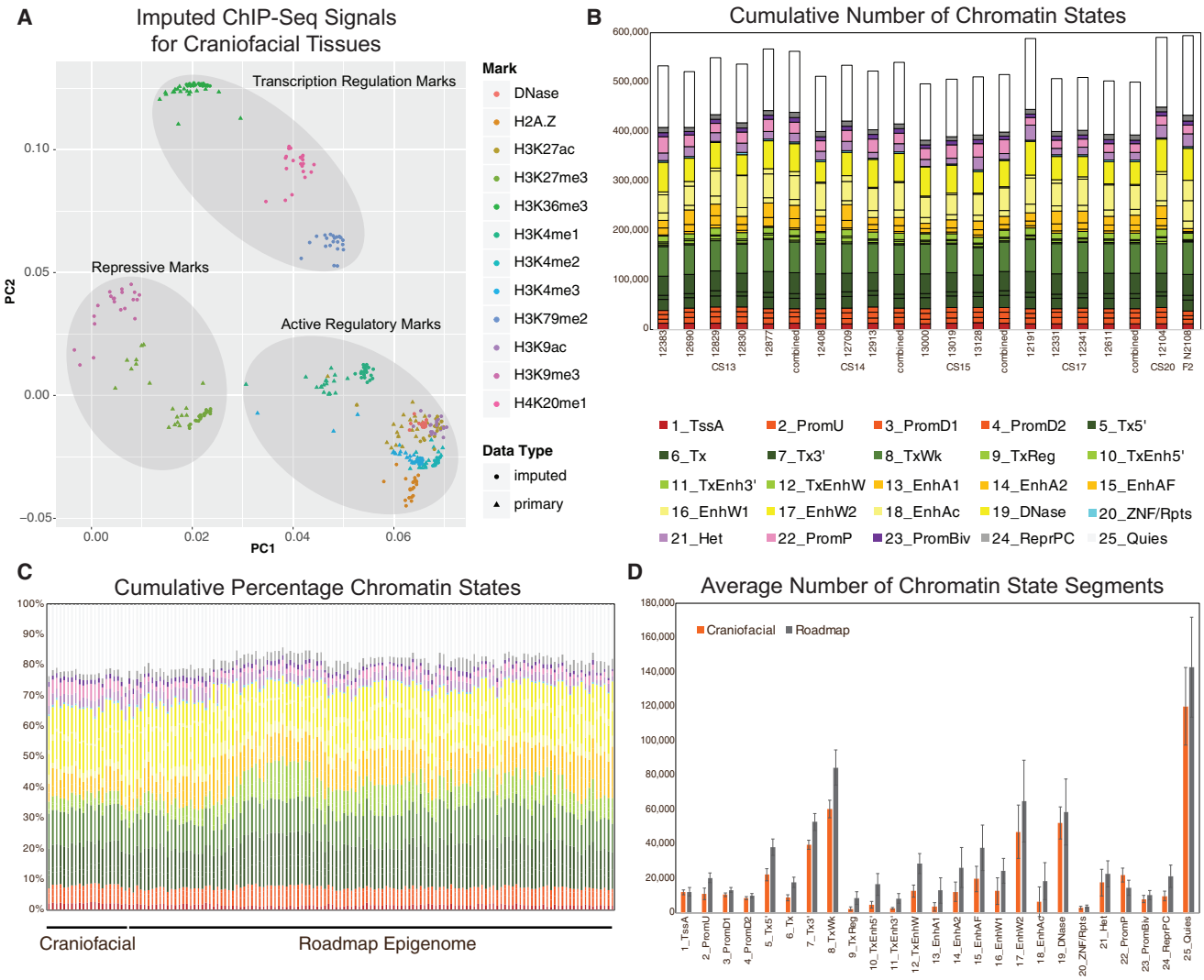



**Figure 3. Imputation of Craniofacial Epigenomic Signals and Chromatin State Segmentation**

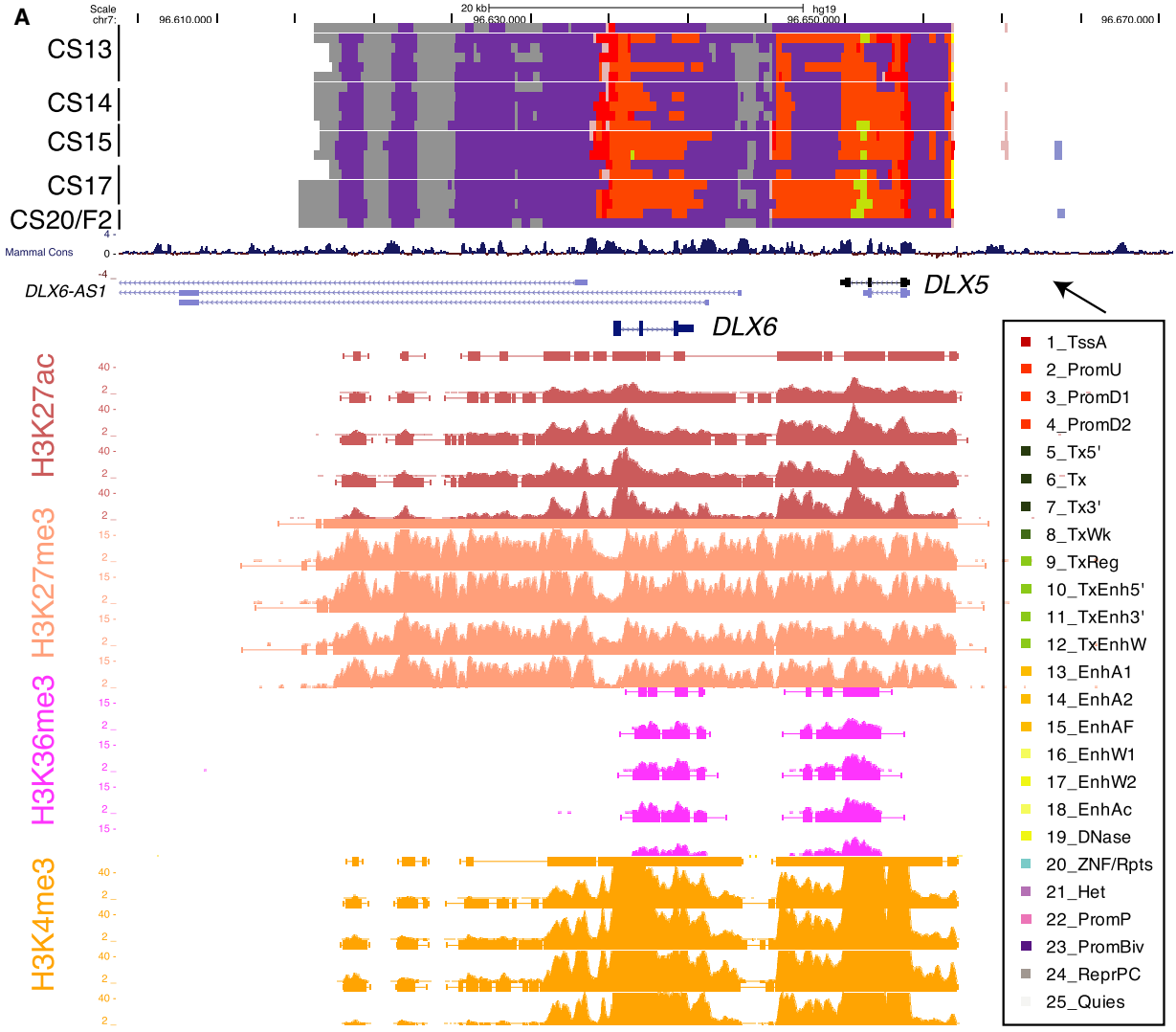
(A) Principal-component analysis projection of first two component dimensions for 252 imputed and 114 primary epigenomic profiles for human craniofacial samples across non-overlapping 10-kb bins. Samples are color coded by epigenomic mark and shapes indicate primary versus imputed data types. Samples generally cluster into three broad categories of activity: repression, regulatory element activation, and transcription regulation. (See also Figures S1D and S1E.)

(B) Numbers of individual chromatin state segments identified by each of the color-coded 25 states of chromatin activity based on imputed epigenomic signals for each of the 21 tissue samples profiled.

(C) Comparison of cumulative percentage of each chromatin state between craniofacial samples profiled here and 127 segmentations generated by Roadmap Epigenome (Roadmap Epigenomics Consortium et al., 2015).

(D) Mean numbers of segments annotated in each of the 25 states across 21 craniofacial samples (orange) and 127 Roadmap Epigenomes (gray). Error bars represent SD. Overall chromatin state segmentation in craniofacial samples identifies similar numbers and percentages of each of 25 states published by Roadmap Epigenome (Roadmap Epigenomics Consortium et al., 2015). (See also Figure S2.)

expression in heterogenous embryonic tissue samples (Bernstein et al., 2006; Cotney et al., 2012; Rada-Iglesias et al., 2012; Rada-Iglesias and Wysocka, 2011). These bivalent signals were originally identified in embryonic stem cell cultures (Bernstein et al., 2006), but they were most recently reported in developing mouse pharyngeal arch tissue as important markers for the regulation of neural crest positional identity, and they were shown to play important roles in *Drosophila* development (Kang et al., 2017; Minoux et al., 2017). These findings emphasize the biological importance of such chromatin states in developmental patterning, and they indicate that genes marked with such states are likely important for proper craniofacial development. These overlapping signals were annotated by a bivalent state in the 25-state model (23_PromBiv) and identified 957 genes with bivalent promoters. The genes identified by this bivalent state were strongly enriched for DNA-binding proteins, most significantly homeobox-containing transcription factors, and enriched for factors previously identified to play a role in embryonic




**B**

## DNA Binding Factors with Bivalent Domains in Human Craniofacial Tissue

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *ASCL2* | ***DLX2*** | *FOXD3* | *HR* | *HOXA6* | *IRX2* | ***MECOM*** | *NR5A1* | *RUNX2* | ***TBX18*** | *ZIC1* |
| *ASCL5* | ***DLX3*** | *FOXE3* | *HAND2* | *HOXA7* | ***IRX3*** | *MEIS3* | ***PAX3*** | *RUNX3* | ***TBX2*** | *ZIC2* |
| ***ALX1*** | ***DLX4*** | ***FOXF1*** | *HSF4* | *HOXA9* | *IRX4* | ***MSX1*** | *PAX6* | ***SATB2*** | *TBX21* | *ZIC4* |
| ***ALX3*** | ***DLX5*** | *FOXF2* | *HES2* | *HOXB1* | ***ISL1*** | ***MSX2*** | ***PAX7*** | *SCRT1* | *TBX4* | ***ZIC5*** |
| ***ALX4*** | ***DLX6*** | *FOXH1* | *HES4* | ***HOXB2*** | ***ISL2*** | *MSC* | ***PAX9*** | *SCRT2* | *TLX2* | *ZBTB42* |
| *ARID3C* | *DMRTA2* | *FOXJ1* | *HES5* | *HOXB3* | ***KLF2*** | *MYOD1* | *PHOX2A* | ***SIM2*** | *TIGD3* | *ZNF497* |
| *ATOH8* | ***EBF2*** | *FOXL1* | *HES7* | *HOXB4* | *LBX1* | *MYF5* | ***PITX1*** | ***SIX1*** | *TRNP1* | *ZNF574* |
| *AIRE* | ***EBF3*** | *FOXN4* | ***HEY1*** | *HOXB5* | *LBX2* | *MYF6* | ***PITX2*** | ***SIX2*** | *TCEA3* | ***ZFPM1*** |
| ***BARX1*** | *EGR1* | *FOXO6* | ***HEYL*** | *HOXB6* | *LHX3* | *NEUROD2* | ***PRRX1*** | *SIX3* | *TCF15* | *ZFR* |
| *BARX2* | *EMX1* | *FOSL1* | *HIC1* | *HOXB7* | ***LHX6*** | *NPAS1* | ***PRRX2*** | *SNAI1* | ***TFAP2A*** | |
| *BHLHA9* | ***EMX2*** | ***FOS*** | *HOXA1* | *HOXD1* | ***LHX8*** | ***NKX3-2*** | *PIF1* | *SNAI3* | *TP73* | |
| *BHLHE23* | *ETV4* | ***FOSB*** | *HOXA10* | *HOXD10* | *LMX1A* | ***NFIA*** | *POU2F2* | *SALL1* | *TWIST1* | |
| *CREB3L1* | *EVX1* | ***GBX2*** | *HOXA11* | *HOXD11* | *LMX1B* | *NFIC* | *POU2F3* | ***SOX10*** | ***TWIST2*** | |
| *CREB5* | *FLI1* | ***GATA3*** | *HOXA13* | *HOXD12* | *LRRFIP1* | *NFIX* | *POU3F2* | *SOX18* | *UNCX* | |
| *CASZ1* | *FOXA3* | *GATA5* | ***HOXA2*** | *HOXD13* | *LYL1* | ***NFATC1*** | ***POU3F3*** | *SOX2* | *MYCL* | |
| *CDX1* | ***FOXC2*** | ***GSC*** | ***HOXA3*** | *HOXD4* | *LEF1* | *NFATC2* | ***PRDM12*** | ***SOX8*** | *VENTX* | |
| *CEBPB* | *FOXD1* | *GRHL1* | *HOXA4* | *HOXD8* | *MAFA* | *NR4A2* | *PRDM16* | *TBX1* | *VAX2* | |
| ***DLX1*** | ***FOXD2*** | ***HMX1*** | *HOXA5* | *HOXD9* | *MAFF* | *NR4A3* | *RBPJL* | ***TBX15*** | *YBX2* | |

*(legend on next page)*

cranial skeleton morphogenesis, such as the gene pair *DLX5* and *DLX6* (Figures 4A and S3) (Robledo et al., 2002).

In total, we identified 189 DNA-binding factors likely to be critical regulators of craniofacial development based on chromatin states (Table S2). Because of the important role ascribed to genes with bivalent status, as mentioned above, we sought to determine if any of the bivalent marks we identified were shared with previous data or specific for craniofacial tissue. To address this, we overlapped our bivalent state calls with the same bivalent state calls from all samples profiled by Roadmap Epigenome. Of the 957 genes identified above, only 7 genes were shown to have bivalent regions within 5 kb of their transcription start site (Table S2). In particular, we identified *EGR1* and *COX7A1* with a bivalent chromatin status not previously observed in Roadmap Epigenome. *EGR1* has been implicated in cranial cartilage development in zebrafish and expressed in early cartilage in mice (McMahon et al., 1990), while *COX7A1* was shown to be differentially expressed in samples from patients with cleft lip only versus cleft lip and palate (Jakobsen et al., 2009). When we analyzed genes reported to have bivalent status in mouse craniofacial tissues (Minoux et al., 2017), 106 of 708 genes shared this status between human and mouse. Of these 106, 62 were DNA-binding proteins identified in our data, representing a significant enrichment in this class of protein among all bivalent genes (2.77-fold increase, Fisher $p < 0.0001$; Figure 4B; Table S2).

Lastly, to determine if this trend extends beyond chromatin marks to regions actively engaged by both activating and repressing complexes, we interrogated genes cobound in HUES64 human embryonic stem cells (ESCs) by BRD1, a component of the MOZ/MORF acetyltransferase complex, and RING1B, a component of Polycomb Repressive Complex 1 (PRC1) (Kang et al., 2017). We observed 134 genes of 609 cobound by these complexes had bivalent status in our data, the majority being DNA-binding factors ($n = 92$; 2.2-fold enrichment, Fisher $p < 0.0001$; Table S2). Follow-up studies of these bivalent genes, especially the DNA-binding proteins, and the complexes that generate these states in a craniofacial context may reveal previously unknown roles in normal formation of the mammalian face and skull.

## Identification of Craniofacial-Specific Enhancers and Craniofacial Super-enhancers

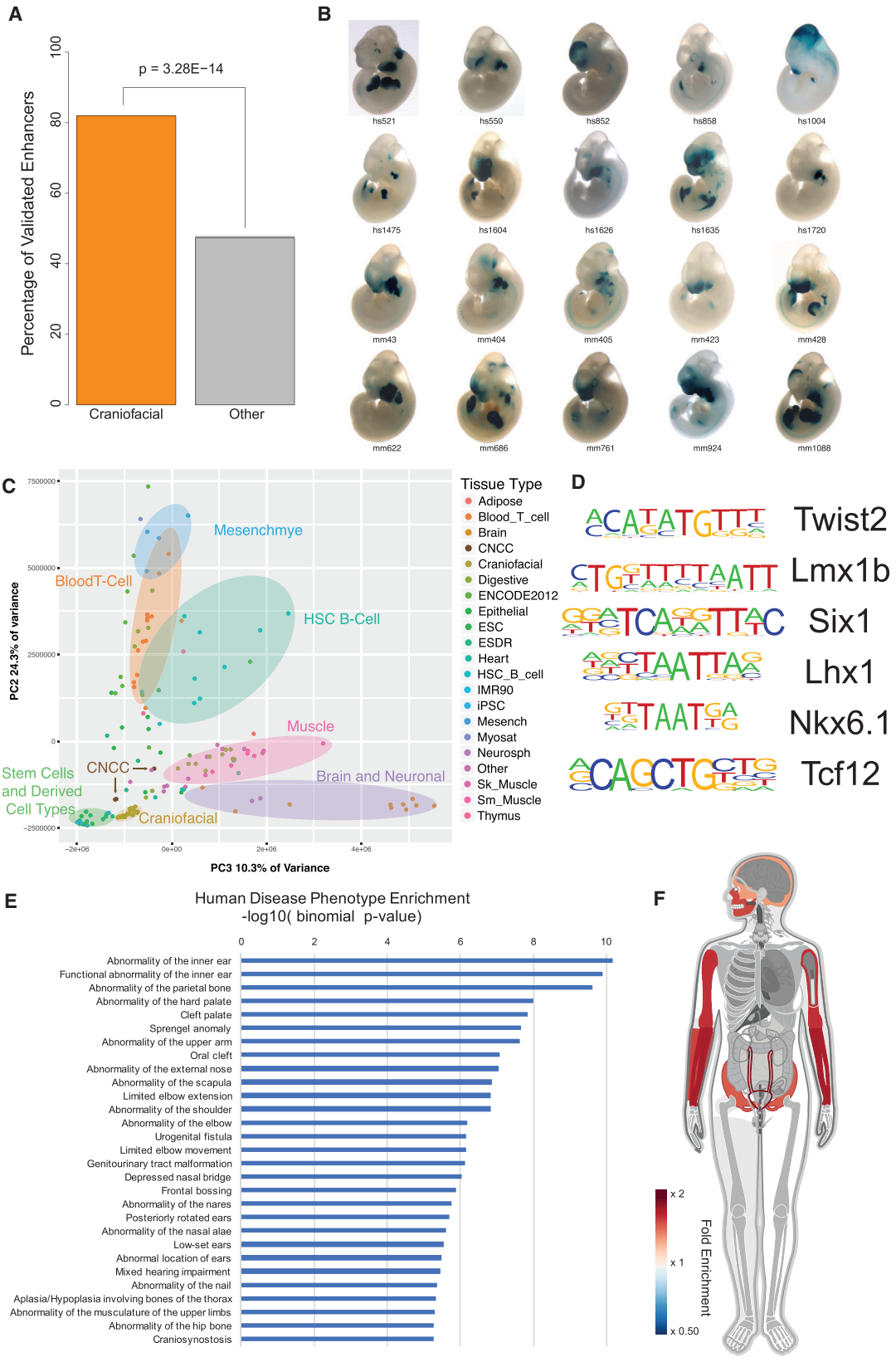
Having shown that our segmentations identify activation and/or poising of promoters important for craniofacial development, we next turned to regulatory portions of the genome. Using the 25-state segmentations, we reproducibly identified 75,928 segments in our craniofacial sample from at least one of six enhancer categories defined by Roadmap Epigenome (EnhA1, EnhA2, EnhAF, EnhW1, EnhW2, and EnhAc). To determine if our chromatin state segmentations identify bona fide craniofacial enhancers, we first compared craniofacial enhancer segments with H3K27ac ChIP-seq peak calls previously identified in cultured CNCCs (Prescott et al., 2015). We found 30-fold ($p < 10^{-4}$) and 12-fold ($p < 10^{-4}$) enrichment of overlaps with the top 5,000 active CNCC enhancers and the top 1,000 human-biased CNCC enhancers, respectively (Figures S4A and S4B).

We then turned to a large catalog of experimentally validated developmental enhancers tested in mouse embryos and available in the Vista Enhancer Browser (Visel et al., 2007). We identified over 80% of all craniofacial-positive enhancers in this database ($n = 170$). Moreover, our enhancer annotations were significantly enriched for craniofacial enhancers versus those that lacked craniofacial activity ($p = 3.28 \times 10^{-14}$) (Figures 5A, 5B, and S4C). While these results are encouraging, namely, that our data identified true craniofacial developmental enhancers, the chromatin state annotations alone do not reveal the specificity of individual regulatory regions nor do they identify target genes. To initially address this question, we quantitatively compared enhancer-associated H3K27ac signals at 425,380 enhancers from our craniofacial segmentations and 127 segmentations from Roadmap Epigenome. Both hierarchical clustering and principal-component analysis showed that our samples were well correlated with one another in this multi-tissue context (Figures 5C and S5). They were most similar to ESCs and cell types derived from them (embryonic stem cell-derived; ESDR), but they were distinct from fetal and adult samples present in Roadmap Epigenome data, suggesting our annotations harbor regulatory information not previously annotated by Roadmap Epigenome. We also uniformly processed raw epigenomic data from human CNCCs (Prescott et al., 2015) with our alignment, imputation, and segmentation pipeline. We found that the chromatin state segmentations for CNCC samples showed significant differences in the overall number and size of all chromatin states versus our data as well as Roadmap Epigenome (Figures S2I and S2J). Additionally, principal-component analysis of H3K27ac signal showed these cell types formed a cluster distinct from both the other ESDR types and our craniofacial samples (Figure 5C). These results are likely due to differences in sequencing depth, the host of marks profiled, or distinct differences in the derivation of these cell types versus primary tissues that we cannot tease apart in this study. We therefore excluded chromatin state segmentations from these cell types from downstream multi-tissue comparisons.

Due to the isolated, tissue-specific nature of many craniofacial defects, we hypothesized that enhancers identified only in developing craniofacial tissues would be enriched near genes implicated in craniofacial abnormalities. To identify such enhancers in craniofacial tissue, we determined if any of our enhancer segments were ever annotated in Roadmap Epigenome. Overlaying our segmentations on those from 127 samples identified 6,651 enhancer segments specific for craniofacial development (8.7% of total craniofacial enhancer segments) (Table S1;

**Figure 4. Large Bivalent Domains at Gene Pair *DLX5* and *DLX6***

(A) UCSC Genome Browser shot of locus encompassing the *DLX5*/*DLX6* locus. At top are chromatin state segmentations for all tissue samples. Purple states indicate bivalent regions. Imputed signals and peak calls for representative samples from each stage and for each indicated histone mark are shown below.
(B) DNA-binding proteins identified as having a bivalent promoter state in human embryonic craniofacial development. Genes in bold were also identified as having bivalent status during early mouse craniofacial development (Minoux et al., 2017). (See also Figure S3 and Table S2.)




*(legend on next page)*

Figure S6). To determine if these sites are relevant for craniofacial development or represent spurious segmentations in our data, we analyzed sequence content of these regions and functional enrichments of genes potentially regulated by these regions. When we assessed the craniofacial-specific enhancers for enrichment of transcription factor-binding sites, we identified motifs matching those of *TWIST2*, *LMX1B*, *SIX1*, *NKX6.1*, multiple members of the *LHX* and *HOX* families, and *TCF12*, all of which have been implicated in craniofacial and skeletal development (Brunskill et al., 2014; Chen et al., 1998; Gendron-Maguire et al., 1993; Laclef et al., 2003; Marchegiani et al., 2015; Sharma et al., 2013; Zhao et al., 1999; Figure 5D; Table S3). Utilizing the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010), we found significant enrichment of craniofacial-specific enhancers assigned to genes associated with craniofacial abnormalities, such as cleft palate in both humans and mice (Figure 5E). Interestingly, we also identified more general categories of enrichment among the putative gene targets, including general transcriptional activators (Table S3). When we interrogated this list of transcription factors for tissue-specific biology (Gokhman et al., 2017), we found significant enrichment for phenotypes related to craniofacial and appendicular skeleton (Figure 5F).

The above analyses focused on the annotation and activation state of individual genome segments. However, these enhancers likely do not operate in isolation, and clusters of enhancers activated in concert have been shown to be powerful regulators of important genes for a given tissue or cell type (Whyte et al., 2013). To identify such enhancer clusters, we applied the rank order of super enhancers (ROSE) (Whyte et al., 2013) algorithm for identifying super-enhancers as well as a sliding window approach to detect enrichment of craniofacial enhancer states relative to both randomly chosen sequences as well as those identified by Roadmap Epigenome. We identified 581 regions across the genome that demonstrated enrichment for craniofacial enhancers or identified as a super-enhancer region (Table S4). These windows had an average size of ~400 kb but ranged up to 2 Mb in length. In many cell types, these clusters of enhancers are embedded in the genome both surrounding and within the introns of their likely tissue-specific target (Hnisz et al., 2013). Indeed, most of the super-enhancer regions and enhancer-enriched windows we identified contained multiple genes (mean of 4.7 genes per window) and were enriched for developmental genes, including multiple *FZD*, *WNT*, *ALX*, *DLX*, and *TBX* family members (Table S4). These super-enhancer regions encompass virtually all of the same bivalent domains identified above (951 of 957), suggesting they have complex, concerted activation by regulatory elements throughout the developing craniofacial complex.

The most significantly enriched super-enhancer region based on both fractions of bases annotated as an enhancer state and H3K27ac signals across all craniofacial samples encompassed the *PRDM16* gene. The promoter of this gene is identified as a bivalent region, and both the large noncoding region upstream and large intronic sequences are littered with strongly active enhancers (Figure S7). Work in mice has identified point mutations in *Prdm16* that give rise to cleft palate, but a role for this gene in human craniofacial abnormalities has not been concretely identified (Bjork et al., 2010). The PRDM16 protein has been implicated in the methylation of H3K9, suggesting this protein could also be involved in maintaining bivalent states described above (Pinheiro et al., 2012). The strong epigenomic signals we identified in primary tissue surrounding this gene, its bivalent promoter, and reported histone modification activity suggest this gene may contribute to the regulation of many genes in craniofacial development.

## Enrichment of Orofacial Clefting and Facial Variation Genetic Associations in Craniofacial Enhancers

The results above suggest that many of the craniofacial enhancers we identified are likely to play a direct role in the patterning of bones of the face, jaws, and portions of the skull. However, it is unclear whether they are directly involved in human craniofacial abnormalities. To begin to explore whether the enhancers we identified play a role in craniofacial abnormalities, we first turned to regulatory regions previously identified as causative in two distinct craniofacial syndromes. Van Buchem disease is a rare disorder characterized by bone overgrowth in the jaws and skull (Van Hul et al., 1998). A large noncoding deletion was identified in a Dutch family between the *MEOX1* and *SOST* genes (Balemans et al., 2002). All of our chromatin state segmentations identified several strong enhancer states in this window, including at least one region previously tested in the developing mouse based on conservation (Loots et al., 2005). However, we also identified a craniofacial-specific enhancer state overlapping a deeply conserved sequence (ERC7) as well

**Figure 5. Chromatin State Segmentations Identify Craniofacial-Specific Regulatory Sequences**

(A) Percentage of *in vivo*-validated embryonic enhancers with (orange) or without (gray) craniofacial activity from the Vista Enhancer Browser (Visel et al., 2007) identified by craniofacial chromatin segments annotated as enhancer states. Significance was determined by Fisher's exact test.

(B) Selected validated enhancers with craniofacial activity identified by this study from the Vista Enhancer Browser. (See also Figure S4.)

(C) Principal-component analysis projection of second and third component dimensions for 149 H3K27ac profiles at 425,568 regions annotated as enhancer segments in any of the samples profiled here or by Roadmap Epigenome. Samples are color coded by group annotations assigned by Roadmap Epigenome or craniofacial samples from this study. Percentages of variance across samples explained by each component are indicated along each axis. (See also Figure S5.)

(D) Transcription factor position weight matrices identified by HOMER (Heinz et al., 2010) as enriched in craniofacial-specific enhancer segments. (See also Table S3.)

(E) Significant enrichments of human disease phenotypes for genes assigned to craniofacial-specific enhancer segments, as reported by GREAT (McLean et al., 2010). (See also Figures S6 and S7.)

(F) Enrichment of anatomical expression of transcription factors identified as potentially regulated by craniofacial-specific enhancer segments, as reported by GeneORGANizer (Gokhman et al., 2017). Heat indicates fold enrichment of expression in individual anatomical region or organ. Craniofacial and appendicular skeleton showed the most significant enrichments.

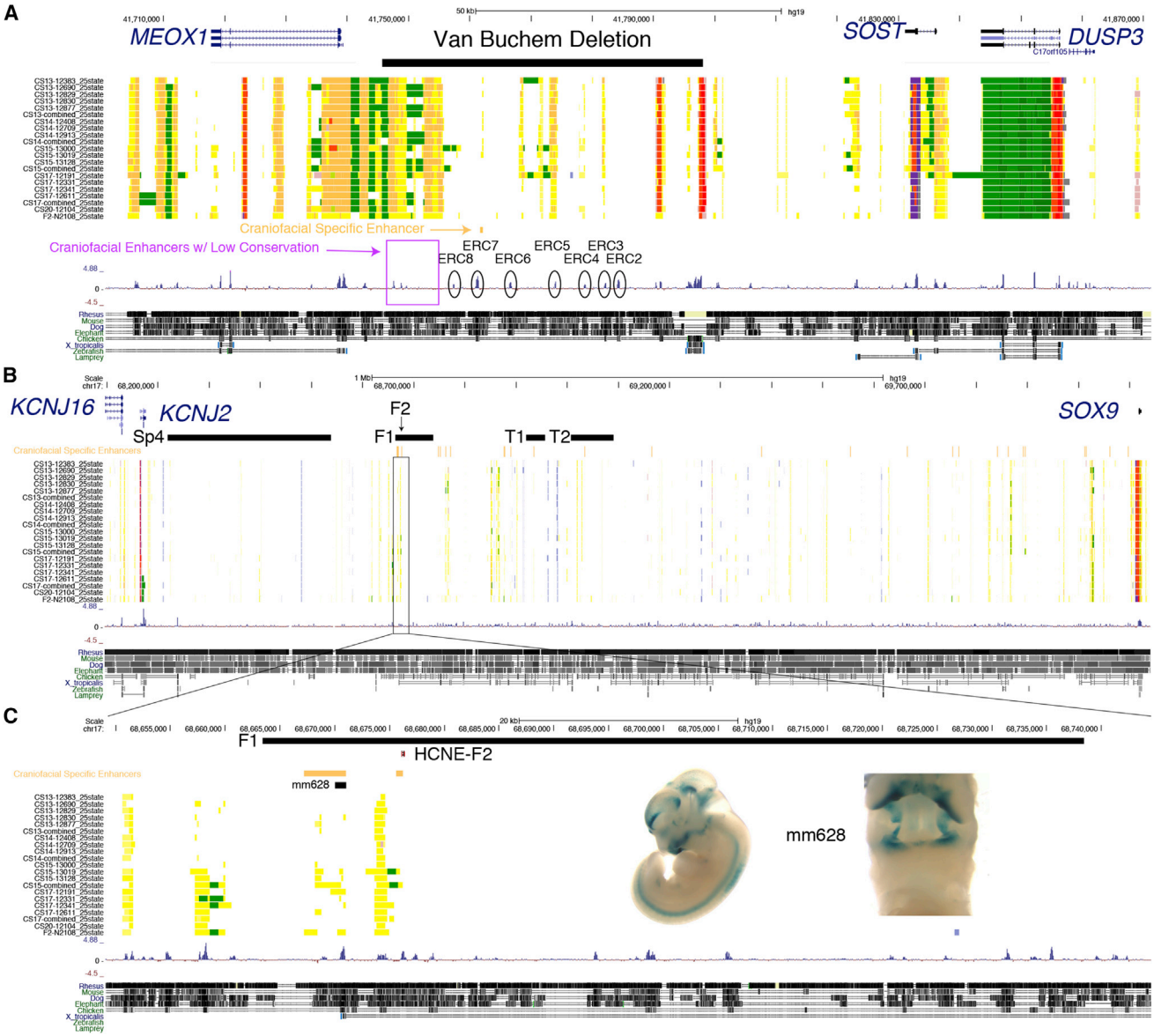



**Figure 6. Human Craniofacial Enhancers Identify Regions Important for Van Buchem Disease and Pierre Robin Sequence**

(A) UCSC Genome Browser shot of locus encompassing the large noncoding region flanked by *MEOX1* and *SOST*. Region deleted in a Dutch family affected by Van Buchem disease is indicated by the black bar. Conserved regions (ERC) tested by Loots et al. (2005) are indicated above conservation tracks. Craniofacial-specific enhancer near ERC7 is indicated in orange. Additional regions annotated with strong craniofacial enhancer states but with relatively low conservation in the Van Buchem interval are indicated by purple box.

(B) UCSC Genome Browser shot of locus encompassing the large noncoding region flanked by *KCNJ2* and *SOX9* and associated with Pierre Robin sequence (PRS). Black bars indicate intervals altered in PRS families as previously reported by Benko et al. (2009). Region encompassing the F2 mutation is highlighted in the following panel.

(C) Enlarged view of F1 deletion region and conserved noncoding element affected by F2 mutation (HCNE-F2). Craniofacial-specific enhancer regions are indicated by orange bars. A portion of one of these regions was tested in the Vista Enhancer Browser (mm628). Inset panel shows enhancer activity of mm628 in embryonic day (E) 11.5 mouse embryos from the Vista Enhancer Browser. (See also Figures S8 and S10–S17.)

as enhancers that lacked significant conservation in this interval that could play a role in Van Buchem disease (Figure 6A). We then turned to Pierre Robin sequence (PRS), a syndrome characterized by a reduced lower jaw, misplacement of the tongue, and frequent occurrence of cleft palate (Tan and Farlie, 2013). Genetic mapping in several multi-generational families affected by PRS identified recurrent deletions and translocations in a 2.46-Mb noncoding region between *KCNJ2* and *SOX9* (Benko et al., 2009; Gordon et al., 2014). Our data identified numerous shared and craniofacial-specific enhancers throughout this

region (Figure 6B). Most importantly, for two families with overlapping genetic changes, we identified craniofacial-specific enhancers within this region, including one overlapping the 200-bp sequence hypothesized to be causative in at least one of these families (F1 and F2 from Benko et al., 2009). Our craniofacial-specific segmentations also identified at least two other highly conserved regions upstream of the originally tested sequence, one of which showed craniofacial regulatory capacity (Figure 6C) (Gordon et al., 2014).

The first genome-wide association for non-syndromic cleft lip and/or palate was identified at the 8q24 locus encompassing a 640-kb noncoding region downstream of the *MYC* gene (Figure S8A) (Birnbaum et al., 2009). This region has a very significant impact on nonsyndromic cleft lip and/or palate (NSCLP) risk, the rs987525 SNP in this region yielding odds ratios ranging from 2.07 to 4.68 (Ludwig et al., 2012). While we did not annotate an enhancer directly overlapping this SNP position, we did identify a craniofacial enhancer state in all of our samples ~2 kb downstream (Figure S8B). Additionally, this large noncoding region has been studied with a number of mouse deletion lines (Uslu et al., 2014). Overlapping deletions revealed a 280-kb region dubbed a medionasal enhancer region (MNE) and a 106-kb nasal epithelial enhancer region (NEE) in the developing mouse. Deletion of the MNE resulted in lower *Myc* expression in multiple developing mouse tissues and an elevated incidence of cleft lip and palate. Inspection of these orthologous regions in the human genome revealed multiple craniofacial enhancers active across multiple states, several of which were more strongly active later in development. In particular, the only strong enhancer states annotated in each region were located near the center of both deletion intervals (Figures S8C and S8D). Our data support the findings reported for this region in the mouse for orofacial clefting, but our annotations narrow the search for causative regulatory regions within these still relatively large regions to less than 5 kb.

We next turned to the most recent genome-wide associations from several comprehensive meta-analyses focused on orofacial clefting (Leslie et al., 2017; Ludwig et al., 2017; Yu et al., 2017). We overlaid associations from these studies along with SNPs in strong linkage disequilibrium with each of the segmentation maps from our data, as well as data from Roadmap Epigenome, and we assessed enrichment using genomic regulatory elements and GWAS overlap algorithm (GREGOR) (Schmidt et al., 2015). As a control, we also interrogated associations for Crohn's disease for enrichment in our enhancers (Welter et al., 2014). We did not observe any significant enrichment of Crohn's-associated SNPs in our craniofacial segmentations. However, we did observe strong enrichment among enhancers identified in immune-related cell types and tissues. This result agreed with previous findings, and it validated our approach for assessing enrichment of genetic associations (Figure S9A) (Lee et al., 2017). When we analyzed associations for orofacial clefting, we observed significant enrichment of SNPs reported from all three studies in our craniofacial enhancers. The most significant findings in our data were consistently observed from SNPs reported by Yu et al. (2017) (Figures 7A, S9B, and S9C). However, only the credible SNPs for 24 regions based on imputation of genotyping data showed enrichment in our much smaller number of craniofacial-specific enhancers relative to total enhancers in each segmentation (Figure S9C) (Ludwig et al., 2017).

Interestingly, for all three orofacial clefting studies, enhancer segmentations from our early samples (CS13–CS15) showed the most significant enrichments. These results suggested defective patterning events during early embryonic development drive these abnormalities. For example, we identified a discrete enhancer state in the noncoding region between *IRF6* and *DIEXF* that contains a tag SNP previously associated with non-syndromic cleft lip and palate (Zucchero et al., 2004; Figure 7B). This particular enhancer region has been reported to influence *IRF6* expression, and it is potentially causative for orofacial clefting, particularly cases of cleft lip without cleft palate (Rahimov et al., 2008). In our data, we observe an enhancer state from CS13 to CS15, but not in samples from CS17 and beyond. This finding would agree with the stronger genetic associations for this region with cleft lip alone as the external features of the human face, particularly the lip, close earlier in human development than the palate (Schoenwolf et al., 2009). Our data may also be informative for other large loci previously implicated in orofacial clefting. Inspection of several of these loci, including 1p22, 1p36, 10q25, 15q13, 17q22, and 20q12, identified craniofacial enhancer states near tag SNPs previously reported (Leslie et al., 2015, 2017; Ludwig et al., 2017; Mangold et al., 2016; Yu et al., 2017). However, our annotations also identified regulatory regions some distance away, near bivalent genes or the proposed target gene in these intervals that may also be important and potentially harbor rare variation not detectable in a genome-wide association study (GWAS) (Figures S10–S17).

Finally, while the regulatory regions identified here were enriched for associations with craniofacial abnormalities, it is unclear if these regions were also informative for understating the wide variety of craniofacial shapes and appearances among humans. Recent GWAS searching for genetic associations with variability in multiple measures of facial shape identified 38 loci in those of European ancestry (Claes et al., 2018). These loci were reported to be enriched near regions with elevated H3K27ac signals in cultured human CNCCs. Indeed, when we interrogated our enhancer segmentations with the craniofacial measure SNPs using the same approach as above for orofacial clefting SNPs, we also observed significant enrichment (Figure 7C). However, unlike the orofacial clefting associations, the craniofacial measure associations were most strongly enriched in enhancers from later stages of our cohort, particularly CS17 and beyond. This would suggest regulatory events in late embryonic and early fetal development contribute to human facial variation. In summary, these results demonstrate that our chromatin state maps are enriched for genetic information necessary for normal craniofacial development and will be extremely useful in identifying and prioritizing causative variation in patients affected by craniofacial abnormalities as well as understanding normal human facial variation.

## DISCUSSION

Recent large consortia efforts to identify the genetics of common disease have gained traction utilizing tissue-specific annotations

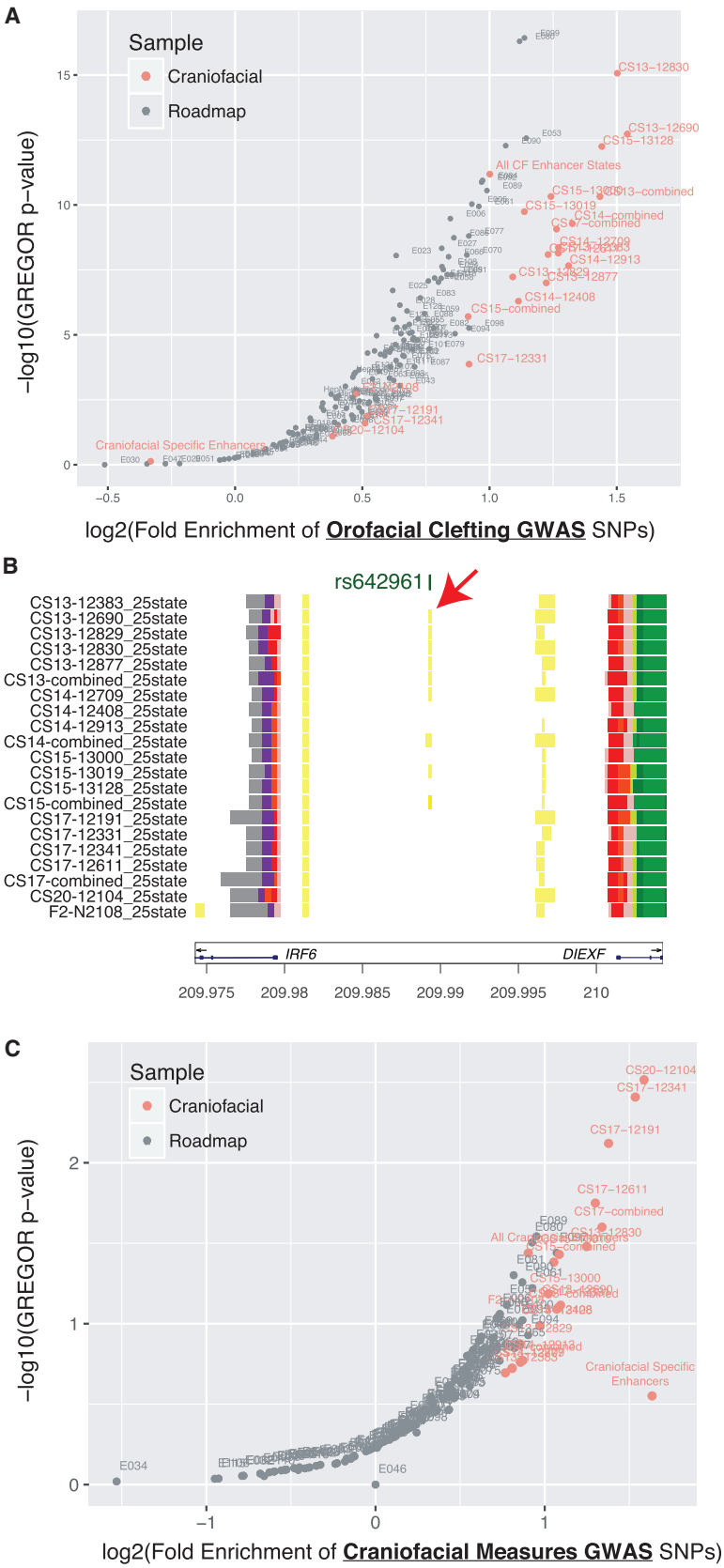



**Figure 7. Enrichment of Orofacial Clefting and Craniofacial Measure Associations in Craniofacial Enhancers**

(A) Enrichment of orofacial cleft GWAS tag SNPs identified by Yu et al. (2017) in enhancer segmentations was assessed using GREGOR (Schmidt et al., 2015). Orange circles indicate craniofacial enhancer annotations identified by a 25-state chromatin model from this study, while gray circles indicate those previously published by Roadmap Epigenome (Roadmap Epigenomics Consortium et al., 2015).

(B) Enhancer state analysis permits placement of a potentially causative allele for non-syndromic CL/P (rs642961) (Rahimov et al., 2008) within a predicted early development enhancer state. This enhancer state is located between *IRF6* and *DIEXF* and may influence expression of *IRF6*. (See also Table S4.)

(C) Same analysis as in (A) using GWAS tag SNPs reported for craniofacial measures by Claes et al. (2018). (See also Figure S9.)

of the genome to identify potential regulatory regions and overlying genetic associations (Farh et al., 2015; Pasquali et al., 2014). Such genetic association data exist for craniofacial abnormalities, but the lack of craniofacial-specific annotations of regulatory function has prevented systematic identification of causal genetic changes. We have addressed this need by generating an extensive resource of functional genomics data obtained directly from human craniofacial tissues during important stages of formation of the orofacial apparatus. We have uniformly processed our data to allow integration of these data with similarly generated signals from a variety of human tissues and developmental stages. These analyses have allowed us to generate craniofacial-specific annotations of chromatin states across the human genome. These chromatin state segmentations reveal tens of thousands of regions with potential gene regulatory activity in craniofacial development.

Our analyses identified a significant proportion of previously known genes and loci important for normal craniofacial development in mice and humans. Specifically, our bivalent promoter and super-enhancer annotations identified key developmental transcription factors as well as large noncoding regions that showed extensive activation during early craniofacial development. The bivalent promoters most likely reflect two possibilities: genes that have restricted patterns of expression in the developing craniofacial tissues or poised genes that can be rapidly activated during development. Due to the heterogeneous nature of the tissue and processing in bulk, it is difficult to determine between these two possibilities. Further efforts using gene expression data, such as single-cell RNA-seq, a battery of *in situ* hybridizations, or sequential ChIP-seq experiments, will be required to elucidate the exact nature of chromatin at these genes and determine the contributions that genes identified here make to normal craniofacial development.

The systematic localization of SNPs associated with orofacial clefting and normal facial variation in our craniofacial enhancer annotations agrees with previous work from other human fetal tissues (Maurano et al., 2012), but it extends this to a tissue/disease-specific context. Specifically, our results support the idea that common variants associated with risk for orofacial clefting, particularly in the case of the cleft lip association for an enhancer of *IRF6*, manifest early in embryonic development, likely in the first 3–5 weeks of gestation. This is in contrast to our results from analyzing craniofacial measure associations. These genetic associations were most strongly enriched in samples from 6 to 9 weeks of gestation, implicating fine-tuning of craniofacial appearance after most of the external features have been established.

Related to this idea, we processed previously published epigenomic data from cultured CNCCs (Prescott et al., 2015). We observed significant overlaps between our craniofacial enhancer segments and CNCC H3K27ac peak calls as well as human-biased CNCC H3K27ac regions. These findings suggest that we have identified regulatory information that is important for the evolution of the human face and that culture models may reflect some of the regulatory architecture of primary human craniofacial development. However, upon further interrogation with our uniform pipeline, we found that the chromatin state segmentations for CNCC samples showed distinct differences in the overall number and size of all chromatin states versus our data as well as Roadmap Epigenome (Figures S2I and S2J). This is potentially due to differences in depth of sequencing, the collection of histone modifications profiled, or the derivation of these cell types versus normal embryonic development. On this last point, enrichment of craniofacial measure associations in enhancers at later stages of development would indicate that the CNCC system generates cells much more differentiated or derived than previously appreciated. This also challenges the previous interpretation that normal variation of the human face is influenced strongly by early embryonic cell types (Claes et al., 2018). Instead, our findings support the idea that these variations result from fine-tuning of the structures after the basic structure of the face has been patterned. Further global characterization of culture models of early craniofacial development, using our epigenomic data as a guide, will be necessary to generate conditions that yield differentiation schemes more reflective of the primary tissue environment. Such systems will be necessary to perform systematic genome modification experiments (deletions, rearrangements, etc.) targeting the regions we have identified to understand their function.

Our data represent one of the most comprehensive epigenetic profiles of primary tissue from the embryonic period of human development. Over 6,000 of the enhancer segments we defined are newly identified, and, as we have shown, they reveal regions with functional contribution to genetic variation associated with disease and facial shape. These regions are strongly enriched near genes implicated in craniofacial development, and they would have remained unknown to craniofacial researchers relying solely on the current state of genome annotations. Indeed, recent targeted sequencing of GWAS intervals at 13 loci in patients affected by craniofacial abnormalities likely included many inert regions but also excluded important craniofacial regulatory regions due to the lack of appropriate chromatin state annotations (Leslie et al., 2015) (top black bars in Figures S6B, S10, S12, S13, S14, S16, and S17). These results illuminate that our current understanding of the regulatory information our genomes encode is incomplete and reinforce the need for more, higher-resolution, tissue-specific profiling of multiple epigenomic marks to yield comparable chromatin state annotations from primary tissues.

We provide all our craniofacial functional genomics data and resulting chromatin state segmentations totaling 1,978 tracks in several standard formats as well as a complete catalog of tracks that can be easily loaded into many modern genome browsers. Our data are listed in the public track hub section of the University of California, Santa Cruz (UCSC) Genome Browser as well as the Track Hub Registry (http://trackhubregistry.org/). Additionally, we have preloaded our annotations and chromatin signals at multiple genes implicated in craniofacial abnormalities in the UCSC Genome Browser that can be opened with a single click on links found at our laboratory website (https://cotney.research.uchc.edu/data/). In total, this resource will allow the craniofacial community and other developmental biologists to develop hypotheses that are rooted in human craniofacial biology instead of using chromatin state annotations from other tissues not directly related to the tissue of interest. This work brings the craniofacial research world firmly into the functional

genomics era, advances our understanding of these disorders from a regulatory perspective, and provides tools for clinicians and researchers seeking to diagnose patients utilizing whole-genome sequencing.

## EXPERIMENTAL PROCEDURES

Further details and an outline of resources used in this work can be found in the Supplemental Experimental Procedures.

### Tissue Collection and Fixation

Use of human fetal tissue was reviewd and approved by the Human Subjects Protection Program at UConn Health (UCHC 710-2-13-14-03). Human embryonic craniofacial tissue was collected, staged, and provided by the Joint MRC/Wellcome Trust Human Developmental Biology Resource (www.hdbr.org). Information describing the developmental stage, termination method, collection site, and karyotype of each embryo is found in Table S1. Tissues were flash frozen upon collection and stored at −80°C. Fixation for ChIP-seq was performed as described in Cotney and Noonan (2015). Fixed tissue pellets were stored at −80°C until batch processing for ChIP.

### Quantification and Statistical Analysis

#### *ChIP-Seq*

Fixed tissue pellets were processed for ChIP as previously described (Cotney and Noonan, 2015). Antibodies used in this study were as follows: anti-H3K27ac (ab4729, Abcam), anti-H3K4me1 (ab8895, Abcam), anti-H3K4me2 (ab7766, Abcam), anti-H3K4me3 (ab8580, Abcam), anti-H3K27me3 (07-449, EMD Millipore), and anti-H3K36me3 (ab9050, Abcam). ChIP-seq libraries were quantified by qPCR (NEBNext Library Quant Kit for Illumina), multiplexed, and sequenced for 75 cycles across multiple flow cells on an Illumina NextSeq 500 instrument.

#### *Primary ChIP-Seq Data Analysis*

ChIP-seq reads were aligned to the human genome (hg19) using Bowtie2 (version [v.] 2.2.5) (Langmead and Salzberg, 2012). Fragment sizes of each library were estimated using PhantomPeakQualTools (v.1.14) (Landt et al., 2012). Histone modification-enriched regions were identified and annotated using Hypergeometric Optimization of Motif Enrichment (HOMER, v.4.8.3) (Heinz et al., 2010). Reproducibly enriched regions were determined by creating a union of all enriched regions for a respective histone modification from all replicates of a single CS and filtering for regions identified in at least two biological replicates using BEDtools (v.2.25.0) (Quinlan and Hall, 2010). We then generated p value-based signal tracks relative to appropriate input controls based on estimated library fragment size using MACS2 (2.1.1.20160309) (Feng et al., 2012). All signal and enriched region files were converted for display in the UCSC Genome Browser using the Kent Source Tools (v.329) (Kent et al., 2002). Correlations of ChIP-seq signals and principal-component analysis across samples and marks were calculated in non-overlapping 10-kb windows using deepTools2 (v.2.5.0.1) (Ramírez et al., 2014).

## DATA AND SOFTWARE AVAILABILITY

The accession number for the ChIP-seq signals, peak calls, and chromatin state segmentations reported in this paper is GEO: GSE97752.

All data can be visualized in the UCSC Genome Browser using publicly available track hub functionality. Hub files and interesting browser examples can be found on our website (https://cotney.research.uchc.edu/data/).

All generic scripts used in processing ChIP-seq and generating chromatin states are available on GitHub (https://github.com/cotneylab/ChIP-Seq).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, 17 figures, and four tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.03.129.

## ACKNOWLEDGMENTS

We would like to thank the donors to HDBR, as without them this resource would not be possible. Funding support was provided by an R00 award (DE024194-04) from NIDCR to J.C.

## AUTHOR CONTRIBUTIONS

Conceptualization, J.P.N. and J.C.; Methodology, A.W. and J.C.; Investigation, A.W. and J.C.; Visualization, A.W., J.V., and J.C.; Writing – Original Draft, A.W. and J.C.; Writing – Review & Editing, A.W., J.V., J.P.N., and J.C.; Funding Acquisition, J.C.; Resources, J.K.; Supervision, J.C.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 9, 2017
Revised: December 5, 2017
Accepted: March 28, 2018
Published: May 1, 2018

## REFERENCES

Balemans, W., Patel, N., Ebeling, M., Van Hul, E., Wuyts, W., Lacza, C., Dioszegi, M., Dikkers, F.G., Hildering, P., Willems, P.J., et al. (2002). Identification of a 52 kb deletion downstream of the SOST gene in patients with van Buchem disease. J. Med. Genet. *39*, 91–97.

Beaty, T.H., Murray, J.C., Marazita, M.L., Munger, R.G., Ruczinski, I., Hetmanski, J.B., Liang, K.Y., Wu, T., Murray, T., Fallin, M.D., et al. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat. Genet. *42*, 525–529.

Beaty, T.H., Marazita, M.L., and Leslie, E.J. (2016). Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. F1000Res. *5*, 2800.

Benko, S., Fantes, J.A., Amiel, J., Kleinjan, D.-J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C.T., et al. (2009). Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nat. Genet. *41*, 359–364.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

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

Bjork, B.C., Turbe-Doan, A., Prysak, M., Herron, B.J., and Beier, D.R. (2010). Prdm16 is required for normal palatogenesis in mice. Hum. Mol. Genet. *19*, 774–789.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat. Genet. *44*, 148–156.

Boulet, S.L., Grosse, S.D., Honein, M.A., and Correa-Villaseñor, A. (2009). Children with orofacial clefts: health-care use and costs among a privately insured population. Public Health Rep. *124*, 447–453.

Brunskill, E.W., Potter, A.S., Distasio, A., Dexheimer, P., Plassard, A., Aronow, B.J., and Potter, S.S. (2014). A gene expression atlas of early craniofacial development. Dev. Biol. *391*, 133–146.

Bureau, A., Parker, M.M., Ruczinski, I., Taub, M.A., Marazita, M.L., Murray, J.C., Mangold, E., Noethen, M.M., Ludwig, K.U., Hetmanski, J.B., et al. (2014). Whole exome sequencing of distant relatives in multiplex families

implicates rare variants in candidate genes for oral clefts. Genetics *197*, 1039–1044.

Camargo, M., Rivera, D., Moreno, L., Lidral, A.C., Harper, U., Jones, M., Solomon, B.D., Roessler, E., Vélez, J.I., Martinez, A.F., et al. (2012). GWAS reveals new recessive loci associated with non-syndromic facial clefting. Eur. J. Med. Genet. *55*, 510–514.

Chen, H., Ovchinnikov, D., Pressman, C.L., Aulehla, A., Lun, Y., and Johnson, R.L. (1998). Multiple calvarial defects in lmx1b mutant mice. Dev. Genet. *22*, 314–320.

Claes, P., Roosenboom, J., White, J.D., Swigut, T., Sero, D., Li, J., Lee, M.K., Zaidi, A., Mattern, B.C., Liebowitz, C., et al. (2018). Genome-wide mapping of global-to-local genetic effects on human facial shape. Nat. Genet. *50*, 414–423.

Conte, F., Oti, M., Dixon, J., Carels, C.E., Rubini, M., and Zhou, H. (2016). Systematic analysis of copy number variants of a large cohort of orofacial cleft patients identifies candidate genes for orofacial clefts. Hum. Genet. *135*, 41–59.

Cotney, J.L., and Noonan, J.P. (2015). Chromatin immunoprecipitation with fixed animal tissues and preparation for high-throughput sequencing. Cold Spring Harb. Protoc. *2015*, 191–199.

Cotney, J., Leng, J., Oh, S., DeMare, L.E., Reilly, S.K., Gerstein, M.B., and Noonan, J.P. (2012). Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. Genome Res. *22*, 1069–1080.

Cotney, J., Leng, J., Yin, J., Reilly, S.K., DeMare, L.E., Emera, D., Ayoub, A.E., Rakic, P., and Noonan, J.P. (2013). The evolution of lineage-specific regulatory activities in the human embryonic limb. Cell *154*, 185–196.

Dickel, D.E., Barozzi, I., Zhu, Y., Fukuda-Yuzawa, Y., Osterwalder, M., Mannion, B.J., May, D., Spurrell, C.H., Plajzer-Frick, I., Pickle, C.S., et al. (2016). Genome-wide compendium and functional assessment of in vivo heart enhancers. Nat. Commun. *7*, 12923.

Dixon, M.J., Marazita, M.L., Beaty, T.H., and Murray, J.C. (2011). Cleft lip and palate: understanding genetic and environmental influences. Nat. Rev. Genet. *12*, 167–178.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

Ernst, J., and Kellis, M. (2015). Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. Nat. Biotechnol. *33*, 364–376.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. *7*, 1728–1740.

Gendron-Maguire, M., Mallo, M., Zhang, M., and Gridley, T. (1993). Hoxa-2 mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. Cell *75*, 1317–1331.

Gokhman, D., Kelman, G., Amartely, A., Gershon, G., Tsur, S., and Carmel, L. (2017). Gene ORGANizer: linking genes to the organs they affect. Nucleic Acids Res. *45* (*W1*), W138–W145.

Gordon, C.T., Attanasio, C., Bhatia, S., Benko, S., Ansari, M., Tan, T.Y., Munnich, A., Pennacchio, L.A., Abadie, V., Temple, I.K., et al. (2014). Identification of novel craniofacial regulatory domains located far upstream of SOX9 and disrupted in Pierre Robin sequence. Hum. Mutat. *35*, 1011–1020.

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

Grosen, D., Bille, C., Petersen, I., Skytthe, A., Hjelmborg, Jv., Pedersen, J.K., Murray, J.C., and Christensen, K. (2011). Risk of oral clefts in twins. Epidemiology *22*, 313–319.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hoffman, M.M., Buske, O.J., Wang, J., Weng, Z., Bilmes, J.A., and Noble, W.S. (2012). Unsupervised pattern discovery in human chromatin structure through genomic segmentation. Nat. Methods *9*, 473–476.

Hoffman, M.M., Ernst, J., Wilder, S.P., Kundaje, A., Harris, R.S., Libbrecht, M., Giardine, B., Ellenbogen, P.M., Bilmes, J.A., Birney, E., et al. (2013). Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res. *41*, 827–841.

Jakobsen, L.P., Borup, R., Vestergaard, J., Larsen, L.A., Lage, K., Maroun, L.L., Kjaer, I., Niemann, C.U., Andersen, M., Knudsen, M.A., et al. (2009). Expression analyses of human cleft palate tissue suggest a role for osteopontin and immune related factors in palatal development. Exp. Mol. Med. *41*, 77–85.

Kang, H., Jung, Y.L., McElroy, K.A., Zee, B.M., Wallace, H.A., Woolnough, J.L., Park, P.J., and Kuroda, M.I. (2017). Bivalent complexes of PRC1 with orthologs of BRD4 and MOZ/MORF target developmental genes in *Drosophila*. Genes Dev. *31*, 1988–2002.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Khandelwal, K.D., van Bokhoven, H., Roscioli, T., Carels, C.E., and Zhou, H. (2013). Genomic approaches for studying craniofacial disorders. Am. J. Med. Genet. C. Semin. Med. Genet. *163C*, 218–231.

Kumar, V., Rayan, N.A., Muratani, M., Lim, S., Elanggovan, B., Xin, L., Lu, T., Makhija, H., Poschmann, J., Lufkin, T., et al. (2016). Comprehensive benchmarking reveals H2BK20 acetylation as a distinctive signature of cell-state-specific enhancers and promoters. Genome Res. *26*, 612–623.

Roadmap Epigenomics Consortium; Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Laclef, C., Souil, E., Demignon, J., and Maire, P. (2003). Thymus, kidney and craniofacial abnormalities in Six 1 deficient mice. Mech. Dev. *120*, 669–679.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. *22*, 1813–1831.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lee, J.C., Biasci, D., Roberts, R., Gearry, R.B., Mansfield, J.C., Ahmad, T., Prescott, N.J., Satsangi, J., Wilson, D.C., Jostins, L., et al.; UK IBD Genetics Consortium (2017). Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. Nat. Genet. *49*, 262–268.

Leslie, E.J., and Marazita, M.L. (2015). Genetics of Orofacial Cleft Birth Defects. Curr. Genet. Med. Rep. *3*, 118–126.

Leslie, E.J., Taub, M.A., Liu, H., Steinberg, K.M., Koboldt, D.C., Zhang, Q., Carlson, J.C., Hetmanski, J.B., Wang, H., Larson, D.E., et al. (2015). Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. Am. J. Hum. Genet. *96*, 397–411.

Leslie, E.J., Carlson, J.C., Shaffer, J.R., Butali, A., Buxó, C.J., Castilla, E.E., Christensen, K., Deleyiannis, F.W.B., Leigh Field, L., Hecht, J.T., et al. (2017). Genome-wide meta-analyses of nonsyndromic orofacial clefts identify

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

Letra, A., Menezes, R., Fonseca, R.F., Govil, M., McHenry, T., Murphy, M.J., Hennebold, J.D., Granjeiro, J.M., Castilla, E.E., Orioli, I.M., et al. (2010). Novel cleft susceptibility genes in chromosome 6q. J. Dent. Res. *89*, 927–932.

Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725–1735.

Lidral, A.C., Liu, H., Bullard, S.A., Bonde, G., Machida, J., Visel, A., Uribe, L.M., Li, X., Amendt, B., and Cornell, R.A. (2015). A single nucleotide polymorphism associated with isolated cleft lip and palate, thyroid cancer and hypothyroidism alters the activity of an oral epithelium and thyroid enhancer near FOXE1. Hum. Mol. Genet. *24*, 3895–3907.

Loots, G.G., Kneissel, M., Keller, H., Baptist, M., Chang, J., Collette, N.M., Ovcharenko, D., Plajzer-Frick, I., and Rubin, E.M. (2005). Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. Genome Res. *15*, 928–935.

Ludwig, K.U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A., Becker, J., Herberz, R., AlChawa, T., Nasser, E., et al. (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat. Genet. *44*, 968–971.

Ludwig, K.U., Ahmed, S.T., Böhmer, A.C., Sangani, N.B., Varghese, S., Klamt, J., Schuenke, H., Gültepe, P., Hofmann, A., Rubini, M., et al. (2016). Meta-analysis Reveals Genome-Wide Significance at 15q13 for Nonsyndromic Clefting of Both the Lip and the Palate, and Functional Analyses Implicate GREM1 As a Plausible Causative Gene. PLoS Genet. *12*, e1005914.

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

Mangold, E., Ludwig, K.U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., Reutter, H., de Assis, N.A., Chawa, T.A., Mattheisen, M., et al. (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat. Genet. *42*, 24–26.

Mangold, E., Böhmer, A.C., Ishorst, N., Hoebel, A.K., Gültepe, P., Schuenke, H., Klamt, J., Hofmann, A., Gölz, L., Raff, R., et al. (2016). Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate. Am. J. Hum. Genet. *98*, 755–762.

Marchegiani, S., Davis, T., Tessadori, F., van Haaften, G., Brancati, F., Hoischen, A., Huang, H., Valkanas, E., Pusey, B., Schanze, D., et al. (2015). Recurrent Mutations in the Basic Domain of TWIST2 Cause Ablepharon Macrostomia and Barber-Say Syndromes. Am. J. Hum. Genet. *97*, 99–110.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

McMahon, A.P., Champion, J.E., McMahon, J.A., and Sukhatme, V.P. (1990). Developmental expression of the putative transcription factor Egr-1 suggests that Egr-1 and c-fos are coregulated in some tissues. Development *108*, 281–287.

Minoux, M., Holwerda, S., Vitobello, A., Kitazawa, T., Kohler, H., Stadler, M.B., and Rijli, F.M. (2017). Gene bivalency at Polycomb domains regulates cranial neural crest positional identity. Science *355*, eaal2913.

Mossey, P.A., and Modell, B. (2012). Epidemiology of oral clefts 2012: an international perspective. Front. Oral Biol. *16*, 1–18.

Mostowska, A., Gaczkowska, A., Żukowski, K., Ludwig, K.U., Hozyasz, K.K., Wójcicki, P., Mangold, E., Böhmer, A.C., Heilmann-Heimbach, S., Knapp, M., et al. (2018). Common variants in DLG1 locus are associated with non-syndromic cleft lip with or without cleft palate. Clin. Genet. *93*, 784–793.

Nord, A.S., Blow, M.J., Attanasio, C., Akiyama, J.A., Holt, A., Hosseini, R., Phouanenavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V., et al. (2013). Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. Cell *155*, 1521–1531.

Pasquali, L., Gaulton, K.J., Rodríguez-Seguí, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, İ., Tena, J.J., Morán, I., Gómez-Marín, C., van de Bunt, M., et al. (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat. Genet. *46*, 136–143.

Petit, F., Jourdain, A.S., Holder-Espinasse, M., Keren, B., Andrieux, J., Duterque-Coquillaud, M., Porchet, N., Manouvrier-Hanu, S., and Escande, F. (2016). The disruption of a novel limb cis-regulatory element of SHH is associated with autosomal dominant preaxial polydactyly-hypertrichosis. Eur. J. Hum. Genet. *24*, 37–43.

Pinheiro, I., Margueron, R., Shukeir, N., Eisold, M., Fritzsch, C., Richter, F.M., Mittler, G., Genoud, C., Goyama, S., Kurokawa, M., et al. (2012). Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. Cell *150*, 948–960.

Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. Cell *163*, 68–83.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rada-Iglesias, A., and Wysocka, J. (2011). Epigenomics of human embryonic stem cells and induced pluripotent stem cells: insights into pluripotency and implications for disease. Genome Med. *3*, 36.

Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S.A., Swigut, T., and Wysocka, J. (2012). Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. Cell Stem Cell *11*, 633–648.

Rahimov, F., Marazita, M.L., Visel, A., Cooper, M.E., Hitchler, M.J., Rubini, M., Domann, F.E., Govil, M., Christensen, K., Bille, C., et al. (2008). Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. Nat. Genet. *40*, 1341–1347.

Rahimov, F., Jugessur, A., and Murray, J.C. (2012). Genetics of nonsyndromic orofacial clefts. Cleft Palate Craniofac. J. *49*, 73–91.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Reilly, S.K., Yin, J., Ayoub, A.E., Emera, D., Leng, J., Cotney, J., Sarro, R., Rakic, P., and Noonan, J.P. (2015). Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. Science *347*, 1155–1159.

Robledo, R.F., Rajan, L., Li, X., and Lufkin, T. (2002). The Dlx5 and Dlx6 homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. Genes Dev. *16*, 1089–1101.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development *132*, 797–803.

Schmidt, E.M., Zhang, J., Zhou, W., Chen, J., Mohlke, K.L., Chen, Y.E., and Willer, C.J. (2015). GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. Bioinformatics *31*, 2601–2606.

Schoenwolf, G.C., Bleyl, S.B., Brauer, P.R., and Francis-West, P.H. (2009). Larsen's Human Embryology, Fourth Edition (Ann Arbor, MI: Churchill Livingstone/Elsevier).

Sharma, V.P., Fenwick, A.L., Brockop, M.S., McGowan, S.J., Goos, J.A.C., Hoogeboom, A.J.M., Brady, A.F., Jeelani, N.O., Lynch, S.A., Mulliken, J.B., et al.; 500 Whole-Genome Sequences (WGS500) Consortium (2013). Mutations in TCF12, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. Nat. Genet. *45*, 304–307.

Tan, T.Y., and Farlie, P.G. (2013). Rare syndromes of the head and face-Pierre Robin sequence. Wiley Interdiscip. Rev. Dev. Biol. *2*, 369–377.

Thieme, F., and Ludwig, K.U. (2017). The Role of Noncoding Genetic Variation in Isolated Orofacial Clefts. J. Dent. Res. *96*, 1238–1247.

Timberlake, A.T., Choi, J., Zaidi, S., Lu, Q., Nelson-Williams, C., Brooks, E.D., Bilguvar, K., Tikhonova, I., Mane, S., Yang, J.F., et al. (2016). Two locus inheritance of non-syndromic midline craniosynostosis via rare*SMAD6*and common*BMP2*alleles. eLife *5*, e20125.

Uslu, V.V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N.A., Marioni, J.C., and Spitz, F. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. Nat. Genet. *46*, 753–758.

Van Hul, W., Balemans, W., Van Hul, E., Dikkers, F.G., Obee, H., Stokroos, R.J., Hildering, P., Vanhoenacker, F., Van Camp, G., and Willems, P.J. (1998). Van Buchem disease (hyperostosis corticalis generalisata) maps to chromosome 17q12-q21. Am. J. Hum. Genet. *62*, 391–399.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. Cell *160*, 554–566.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser–a database of tissue-specific human enhancers. Nucleic Acids Res. *35*, D88–D92.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

Weedon, M.N., Cebola, I., Patch, A.-M., Flanagan, S.E., De Franco, E., Caswell, R., Rodríguez-Seguí, S.A., Shaw-Smith, C., Cho, C.H.-H., Allen, H.L., et al.; International Pancreatic Agenesis Consortium (2014). Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. Nat. Genet. *46*, 61–64.

Wehby, G.L., and Cassell, C.H. (2010). The impact of orofacial clefts on quality of life and healthcare use and costs. Oral Dis. *16*, 3–10.

Wehby, G.L., Castilla, E.E., Goco, N., Rittler, M., Cosentino, V., Javois, L., Kindem, M., Chakraborty, H., Dutra, G., López-Camelo, J.S., et al. (2011). The effect of systematic pediatric care on neonatal mortality and hospitalizations of infants born with oral clefts. BMC Pediatr. *11*, 121.

Wehby, G.L., Pedersen, D.A., Murray, J.C., and Christensen, K. (2012). The effects of oral clefts on hospital use throughout the lifespan. BMC Health Serv. Res. *12*, 58.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. *42*, D1001–D1006.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

World Health Organization (2003). World Atlas of Birth Defects, Second Edition (Geneva, Switzerland: World Health Organization).

Yu, Y., Zuo, X., He, M., Gao, J., Fu, Y., Qin, C., Meng, L., Wang, W., Song, Y., Cheng, Y., et al. (2017). Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. Nat. Commun. *8*, 14364.

Yuan, Q., Blanton, S.H., and Hecht, J.T. (2011). Association of ABCA4 and MAFB with non-syndromic cleft lip with or without cleft palate. Am. J. Med. Genet. A. *155A*, 1469–1471.

Zhao, Y., Guo, Y.J., Tomac, A.C., Taylor, N.R., Grinberg, A., Lee, E.J., Huang, S., and Westphal, H. (1999). Isolated cleft palate in mice with a targeted mutation of the LIM homeobox gene lhx8. Proc. Natl. Acad. Sci. USA *96*, 15002–15006.

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell *152*, 642–654.

Zucchero, T.M., Cooper, M.E., Maher, B.S., Daack-Hirsch, S., Nepomuceno, B., Ribeiro, L., Caprau, D., Christensen, K., Suzuki, Y., Machida, J., et al. (2004). Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. N. Engl. J. Med. *351*, 769–780.

**Cell Reports, Volume *23***

## Supplemental Information

# High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development

**Andrea Wilderman, Jennifer VanOudenhove, Jeffrey Kron, James P. Noonan, and Justin Cotney**

## Supplemental Experimental Procedures

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Tissue Collection and fixation

Use of human fetal tissue was reviewed and approved by the Human Subjects Protection Program at UConn Health. Human embryonic craniofacial tissue was collected, staged and provided by the Joint MRC/Wellcome Trust Human Developmental Biology Resource (www.hdbr.org). Tissues were flash frozen upon collection and stored at -80℃. Fixation for ChIP-Seq was performed as described in Cotney and Noonan, 2015 (Cotney and Noonan, 2015). Briefly, each tissue sample was rapidly thawed in 1 mL of ice cold phosphate buffered saline (PBS) and briefly homogenized with a disposable plastic pestle in a 1.5 mL microcentrifuge tube. Samples were then fixed by the addition of formaldehyde to a final concentration of 1% and incubated at room temperature on a rotisserie for 15 minutes. Samples were then quenched with 150 mM glycine at 10 minutes at room temperature. Tissue was collected by centrifugation (5 min, 2500g, 4℃) and washed with 1 mL of fresh PBS. Fixed tissue pellets were then rapidly frozen in a dry ice/alcohol bath and stored at -80℃ until batch processing for chromatin immunoprecipitation (ChIP).

## METHOD DETAILS

### ChIP-Seq

Fixed tissue pellets were processed for ChIP as previously described (Cotney and Noonan, 2015). Briefly, samples were thawed in 1 mL of 1x Cell Lysis buffer and incubated on ice for 20 minutes. Cells were lysed with dounce homogenization and nuclei were collected by centrifugation (5 min, 2500g, 4℃). Nuclei were resuspended in 300 µL of 1x Nuclear Lysis buffer + 0.3% SDS + 2 mM sodium butyrate and incubated on ice for 20 minutes. Chromatin was sheared with a Qsonica Q800R1 sonicator system operating at amplitude 20 and 2℃ for 30 minutes (10 seconds duty, 10 seconds rest). Samples were cleared by centrifugation (5 min, 20,000g, 4℃) and soluble chromatin was transferred equally into six separate tubes with 10% reserved as an input control. SDS concentration was reduced to 0.18% with ChIP Dilution buffer. Protein G Dynabeads (ThermoFisher) separately preloaded with 2 µg of antibodies were

added to each chromatin aliquot. Antibodies used in this study: anti-H3K27ac (ab4729, Abcam), anti-H3K4me1 (ab8895, Abcam), anti-H3K4me2 (ab7766, Abcam), anti-H3K4me3 (ab8580, Abcam), anti-H3K27me3 (07-449, EMD Millipore), anti-H3K36me3 (ab9050, Abcam). Specificity of all antibodies was validated using Absurance H3 Histone Peptide Array (16-667, Millipore). ChIP samples were incubated overnight at 4℃ on a rotisserie. Chromatin was then immunoprecipitated on a magnet and supernatant was discarded. Beads were washed 8 times with 1 mL of 500 mM LiCl ChIP-Seq Wash Buffer and once with 1 mL of TE. Chromatin was eluted from the beads twice with ChIP Elution buffer at 65℃ for 10 minutes with constant agitation. Combined eluates for each ChIP were subjected to crosslink reversal overnight at 65℃. Samples were then sequentially treated with RNAse A and proteinase K, purified with a PCR Purification Kit (Qiagen), and eluted in 50 uL of EB. ChIP samples were then quantified with picoGreen (ThermoFisher) and prepared for sequencing on Illumina instruments using the Thruplex 48S Library Prep kit (Rubicon Genomics) according to manufacturer's instructions. Final libraries were quantified by QPCR (NEBNext Library Quant Kit for Illumina), multiplexed, and sequenced for 75 cycles across multiple flow cells on an Illumina NextSeq 500 instrument.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Primary ChIP-Seq Data Analysis

Sequencing data was directly retrieved from Illumina's Basespace Cloud service using Basemount command line tools provided by Illumina. Multiple FASTQs for each ChIP were combined and assessed for quality using FASTQC (v0.11.2) (Andrews, 2010) and compared visually using MultiQC (v0.9) (Ewels et al., 2016). Reads were then aligned to the human genome (hg19) using Bowtie2 (v2.2.5) (Langmead and Salzberg, 2012) keeping only uniquely mapped reads. Fragment sizes of each library were estimated using PhantomPeakQualTools (v.1.14) (Landt et al., 2012). Histone modification enriched regions were identified and annotated using HOMER (v4.8.3) (Heinz et al., 2010). Reproducibly enriched regions were determined by creating a union of all enriched regions for a respective histone modification from all replicates of a single Carnegie stage and filtering for regions identified in at least two biological replicates using BEDtools (v2.25.0) (Quinlan and Hall, 2010). We then generated p-value based signal tracks relative to appropriate input controls based on estimated library fragment size using MACS2 (2.1.1.20160309) (Feng et al., 2012). All signal and enriched region files were converted for display in the UCSC Genome Browser using the Kent Source Tools

(v329) (Kent et al., 2002). Correlations of ChIP-Seq signals and Principal Component Analysis across samples and marks were calculated in non-overlapping 10kb windows using deepTools2 (v2.5.0.1) (Ramírez et al., 2014).

### Roadmap Epigenome and cultured CNCC Data Retrieval

Aligned and consolidated primary ChIP-Seq reads in tagAlign format were retrieved from Roadmap Epigenome for eleven epigenomic signals: H2A.Z, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me3, H3K27ac, H3K27me3, H3K36me3, H3K79me2, and H4K20me1. (http://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/). To ensure the most compatible signals with our data, p-value signals were generated by MACS2 from these data based on library fragment sizes reported by Roadmap Epigenome as above. DNase p-value signals were retrieved directly from Roadmap Epigenome (http://egg2.wustl.edu/roadmap/data/byFileType/signal/consolidated/macs2signal/pval/) and converted from bigWig to bedGraph for use with ChromImpute (Ernst and Kellis, 2015) using Kent Source Tools (Kent et al., 2002). Chromatin state segmentations for 127 epigenomes and associated 15-, 18-, and 25-state model files were retrieved from Roadmap Epigenome (http://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/). Raw reads for all human CNCC ChIP-Seq from Prescott et al 2015 were retrieved from GEO accession GSE70751 and processed using procedures as above.

### Chromatin Imputation

Bedgraph files for all p-value signals from primary ChIP-Seq data were converted to 25 bp resolution and processed for model training and generation of imputed signals for all samples using ChromImpute (v1.0.1) as previously described (Ernst and Kellis, 2015). Resulting imputed signal tracks were converted to bigWig format for display in UCSC genome browser and converted to combined signal format at 200 bp resolution for use with ChromHMM (v1.12) (Ernst and Kellis, 2012) using deepTools2 (Ramírez et al., 2014).

### Chromatin State Segmentation

Signal files for individual chromosomes for each craniofacial epigenome were binarized and segmentation was performed using previously published joint 15-, 18-, and 25-state chromatin models using ChromHMM as previously described(Roadmap Epigenomics et al., 2015). Following segmentation, annotation of states and generation of genome browser files was

performed based on annotations provided by Roadmap Epigenome. Individual models of 15, 18 and 25 chromatin states were also learned for each craniofacial epigenome using default settings in ChromHMM. Pearson Correlations and Principal Component Analyses were performed on total H3K27ac signals extracted observed in all imputed p-value signal tracks for craniofacial and Roadmap Epigenome samples from the union of all enhancer state segmentations (EnhA1, EnhA2, EnhAF, EnhW1, EnhW2, and EnhAc) using deepTools2 (Ramírez et al., 2014). All plots were made in R (v3.3.3) (R Core Team, 2017) using tabular data generated by deepTools2.

## Functional Enrichments in Craniofacial Epigenomes

Craniofacial enhancer state segmentations (EnhA1, EnhA2, EnhAF, EnhW1, EnhW2, and EnhAc) were interrogated for tissue activity in the developing mouse embryo from the Vista Enhancer Browser (Visel et al., 2007). Significance of overlap of enhancers identified in human craniofacial tissue and shown to be active in mouse craniofacial tissue relative to all other tissue annotations was determined using Fisher's exact test. To identify totally novel craniofacial enhancers, enhancer state segmentations for all craniofacial segmentations were interrogated for single base overlap with the same states from all Roadmap Epigenomes using BEDtools (Quinlan and Hall, 2010). These novel craniofacial enhancer segmentations were assessed for gene ontology and functional enrichments based on assigned target genes using GREAT (v3.0.0) (McLean et al., 2010). Genes identified as transcriptional regulators by GREAT were assessed for enrichment of anatomical expression using default parameters in GeneORGANizer (Gokhman et al., 2017). Sequence from novel craniofacial enhancer segmentations was extracted from hg19 using fastaFromBed within BEDTools (Quinlan and Hall, 2010). The resulting sequences were assessed for transcription factor motif enrichment using HOMER (Heinz et al., 2010). Enhancer state segmentations from craniofacial epigenomes and all Roadmap epigenomes were interrogated for significance of overlap with GWAS tag SNPs associated with orofacial clefting and craniofacial morphology (Beaty et al., 2011; Birnbaum et al., 2009; Grant et al., 2009; Ludwig et al., 2012; Mangold et al., 2010; Shaffer et al., 2016; Shi et al., 2012) obtained from the GWAS Catalog (retrieved 2017-02-20)(Welter et al., 2014) using Fisher's exact test within BEDTools (Quinlan and Hall, 2010). Adjusted p-value thresholds for 154 comparisons were determined using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995).

### Identification of Enhancer Clusters

Enhancers have been shown to cluster spatially over long distances and within topological domains (Ing-Simmons et al., 2015), thus to identify clusters of craniofacial enhancers we first generated overlapping 200kb windows (median contact domain size from high resolution Hi-C (Rao et al., 2014)) with a 50kb step size using BEDtools (Quinlan and Hall, 2010). Next, we intersected these windows with all enhancer chromatin state segmentations from craniofacial tissues. We then calculated the fraction of each window annotated as an enhancer state. We tested for enrichment of enhancers in each window using permutation testing by randomly shuffling the craniofacial enhancer segments across the genome 1000 times using BEDtools (Quinlan and Hall, 2010) and determining the fraction of each window annotated as an enhancer. Permutation p-values were corrected using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995) for 60739 200kb windows and additionally filtered for a minimum fraction of enhancer states of 0.15 to ensure strong enhancer activation. Overlapping windows passing these criteria were merged into a single contiguous region. Final enriched regions were assessed for overlap with gene annotations and validated craniofacial enhancers using BEDtools (Quinlan and Hall, 2010). We also identified super-enhancer regions using H3K27ac ChIP-Seq reads at all craniofacial enhancer segments with default parameters in ROSE (Whyte et al. 2013).

## DATA AND SOFTWARE AVAILABILITY

All data can be visualized in the UCSC Genome Browser using publically available track hub functionality. Hub files and interesting browser examples can be found on our website: http://cotney.research.uchc.edu/data/

ChIP-Seq signals, peak calls, chromatin state segmentations are available at GEO accession GSE97752.

All generic scripts used in processing ChIP-Seq and generating chromatin states are available on github: https://github.com/cotneylab/ChIP-Seq

**RESOURCES TABLE**

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Rabbit Polyclonal Anti-H3K27ac | Abcam | Cat# ab4729; RRID: AB_2118291 |
| Rabbit Polyclonal Anti-H3K4me1 | Abcam | Cat# ab8895; RRID: AB_306847 |
| Rabbit Polyclonal Anti-H3K4me2 | Abcam | Cat# ab7766; RRID: AB_2560996 |
| Rabbit Polyclonal Anti-H3K4me3 | Abcam | Cat# ab8580; RRID: AB_306649 |
| Rabbit Polyclonal Anti-H3K27me3 | EMD Millipore | Cat# 07-449; RRID: AB_310624 |
| Rabbit Polyclonal Anti-H3K36me3 | Abcam | Cat# ab9050; RRID: AB_306966 |
| Biological Samples | | |
| Human embryonic craniofacial tissue | Joint MRC/Wellcome Trust Human Developmental Biology Resource | www.hdbr.org |
| Critical Commercial Assays | | |
| Absurance H3 Histone Peptide Array | Millipore | 16-667 |
| Thruplex 48S Library Prep kit | Takara Bio | R400427 |
| NextSeq 500/550 High Output v2 kit (75 cycles) | Ilumina | FC-404-2005 |
| NEBNext Library Quant Kit for Illumina | New England Biolabs | E7630S |
| | | |
| | | |
| Deposited Data | | |
| ChIP-Seq signals, peak calls, chromatin state segmentations | This Paper | GEO: GSE97752 |
| UCSC Genome Browser Track Hubs | This Paper | https://cotneylab.cam.uchc.edu/~jcotney/CRANIOFACIAL_HUB/Craniofacial_Data_Hub.txt |
| | | |
| Software and Algorithms | | |
| Basemount | Illumina | https://basemount.basespace.illumina.com/ |
| FASTQC (v0.11.2) | Andrews, 2010 | http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ RRID:SCR_014583 |
| MultiQC (v0.9) | Ewels et al., 2016 | http://multiqc.info/ RRID:SCR_014982 |
| Bowtie2 (v2.2.5) | Langmead and Salzberg, 2012 | http://bowtie-bio.sourceforge.net/bowtie2/index.shtml |

| | | |
|---|---|---|
| PhantomPeakQualTools (v.1.14) | Landt et al., 2012 | https://code.google.com/p/phantompeakqualtools/ RRID:SCR_005331 |
| HOMER (v4.8.3) | Heinz et al., 2010 | http://homer.ucsd.edu/homer/ RRID:SCR_010881 |
| BEDtools (v2.25.0) | Quinlan and Hall, 2010 | https://github.com/arq5x/bedtools2 RRID:SCR_006646 |
| MACS2 (2.1.1.20160309) | Feng et al., 2012 | https://github.com/taoliu/MACS/ |
| Kent Source Tools (v329) | Kent et al., 2002 | https://github.com/ENCODE-DCC/kentUtils |
| deepTools2 (v2.5.0.1) | Ramírez et al., 2014 | https://github.com/fidelram/deepTools |
| ChromImpute (v1.0.1) | Ernst and Kellis, 2015 | http://www.biolchem.ucla.edu/labs/ernst/ChromImpute/ |
| ChromHMM (v1.12) | Ernst and Kellis, 2012 | http://compbio.mit.edu/ChromHMM/ |
| R Project for Statistical Computing (v3.3.3) | R Core Team, 2017 | http://www.r-project.org/ RRID:SCR_001905 |
| GREAT (v3.0.0) | McLean et al., 2010 | http://great.stanford.edu/public/html/splash.php RRID:SCR_005807 |
| GeneORGANizer | Gokhman et al., 2017 | http://geneorganizer.carmelab.huji.ac.il/ |
| GWAS Catalog (retrieved 2017-02-20) | Welter et al., 2014 | http://www.ebi.ac.uk/gwas/home RRID:SCR_012745 |
| SOMatic | Mortazavi et al., 2013 | https://github.com/csjansen/SOMatic |
| Cluster (v3.0) | de Hoon et al., 2004 | http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm RRID:SCR_013505 |
| Java TreeView | Saldanha, 2004 | http://jtreeview.sourceforge.net/ |
| Motif Enrichment | Kheradpour and Kellis, 2014 | http://compbio.mit.edu/encode-motifs/ |
| Cutadapt (v1.8.3) | Martin, 2011 | https://cutadapt.readthedocs.io/en/stable/ RRID:SCR_011841 |
| GREGOR (v1.4.0) | Schmidt et al., 2015 | https://genome.sph.umich.edu/wiki/GREGOR |

## Supplemental Figures

**Figure S1. Detailed Histone Modification Profiles in Human Craniofacial Development. a.** Heatmap and hierarchical clustering of pairwise Pearson correlations for 114 individual histone modification profiles from human craniofacial tissues. Darker orange indicates positive correlation between datasets. Enlarged from **Fig. 2a** to include sample details, showing samples cluster closely by histone mark. **b.** Correlation of only H3K27ac data contained in the area boxed in black in part **a**. Heatmap and hierarchical clustering show that the samples cluster well into groups by early or late stage of development. **c.** Genomic feature annotations identified by peak calls from six histone modification profiles from all craniofacial samples, across all Carnegie stages, plotted as cumulative percentage of total peaks. Peak enrichments and genomic annotations were performed using HOMER (Heinz et al., 2010). **d.** Heatmap and hierarchical clustering of pairwise Pearson correlations for imputed histone modification profiles from human craniofacial tissues. Darker orange indicates positive correlation between datasets. **e.** Heatmap and hierarchical clustering of pairwise Pearson correlations for imputed and primary histone modification profiles from human craniofacial tissues. Darker orange indicates positive correlation between datasets. Related to Figure 2.

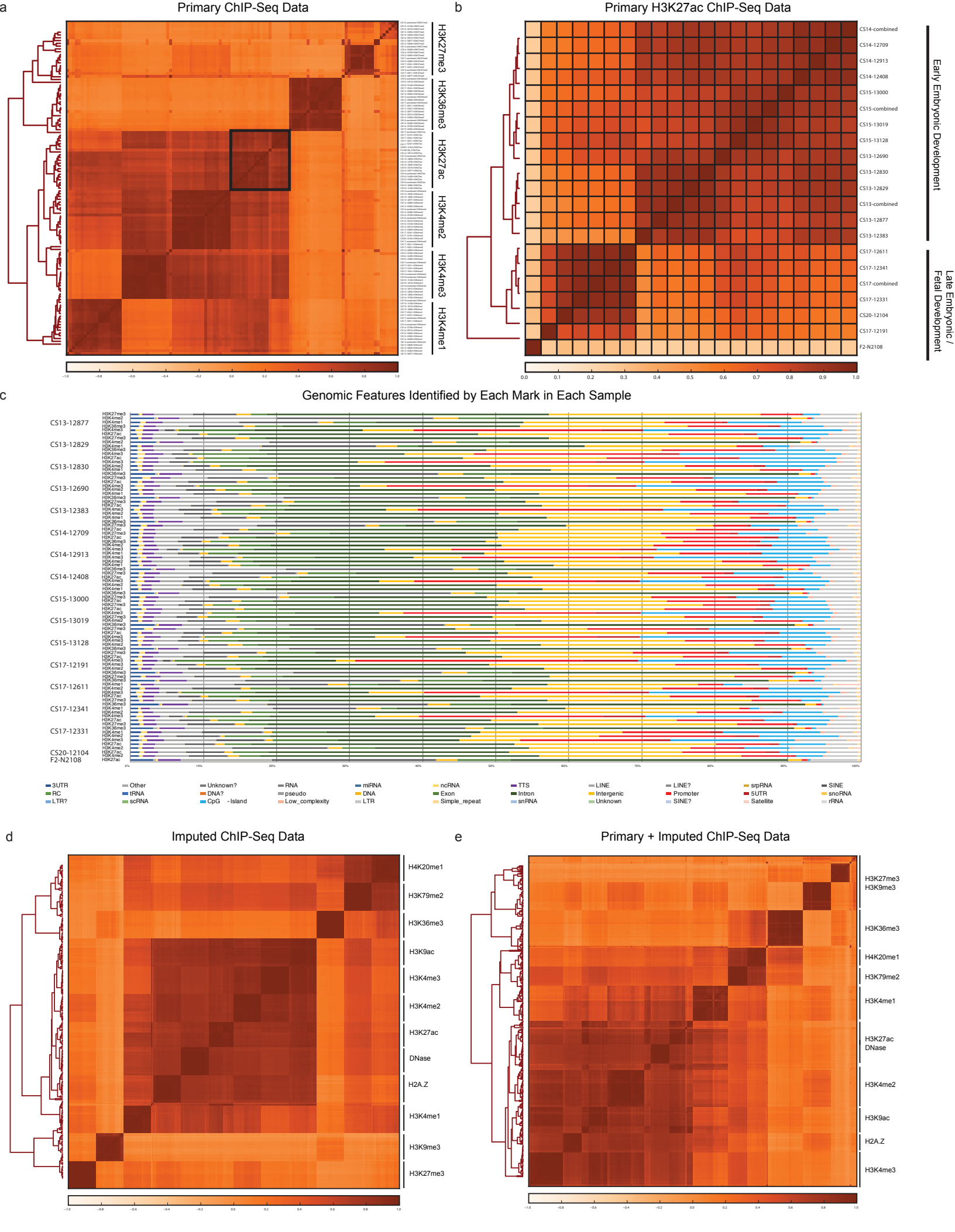
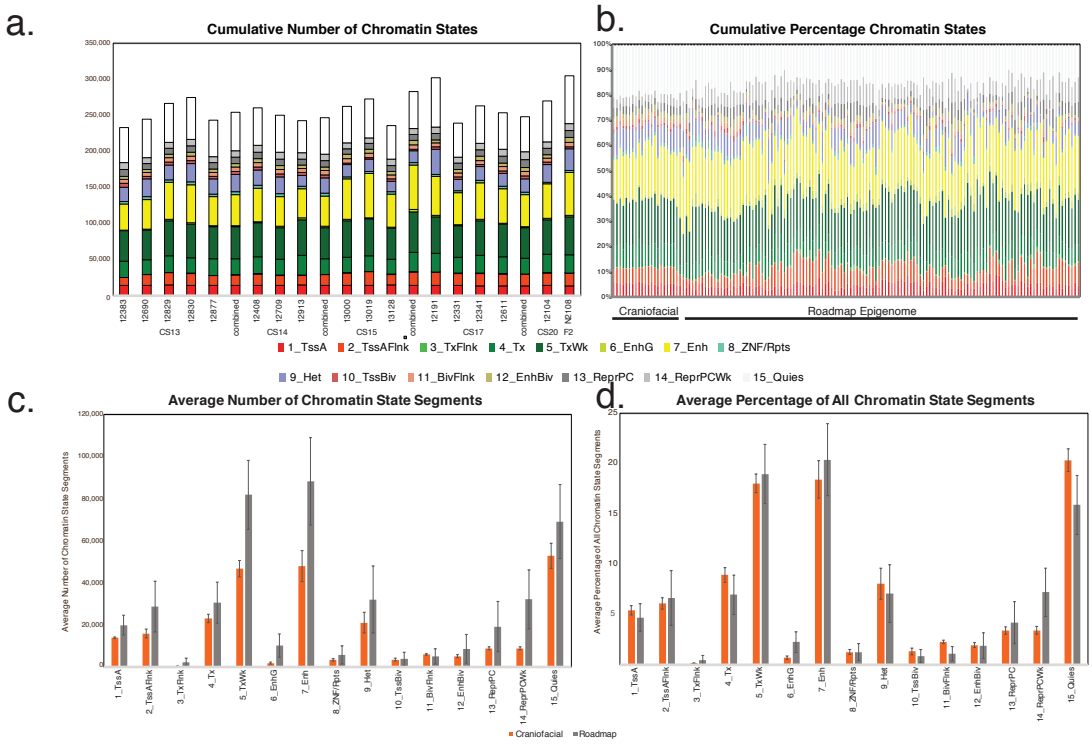
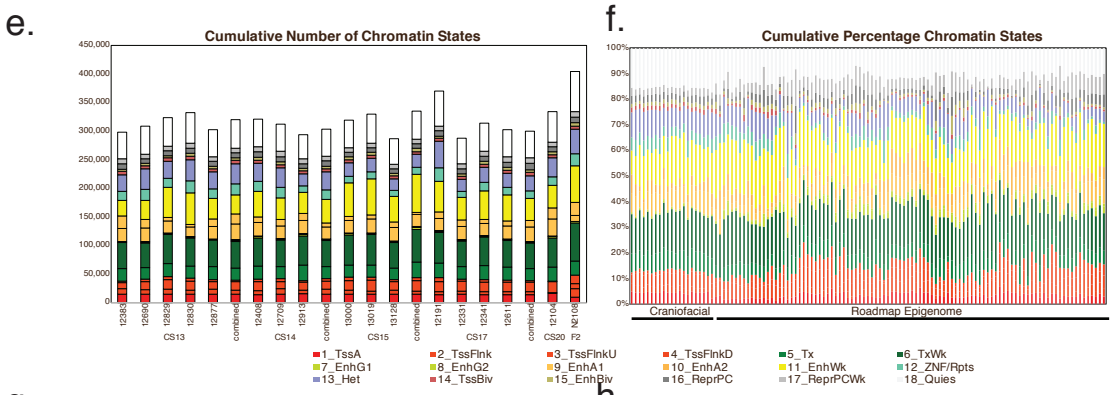

a
Primary ChIP-Seq Data
H3K27me3
H3K36me3
H3K27ac
H3K4me2
H3K4me3
H3K4me1
b
Primary H3K27ac ChIP-Seq Data
CS14-combined
CS14-12709
CS14-12913
CS14-12408
CS15-13000
CS15-combined
CS15-13019
CS15-13128
CS13-12690
CS13-12830
CS13-12829
CS13-combined
CS13-12877
CS13-12383
CS17-12611
CS17-12341
CS17-combined
CS17-12331
CS20-12104
CS17-12191
F2-N2108
Early Embryonic Development
Late Embryonic / Fetal Development
c
Genomic Features Identified by Each Mark in Each Sample
CS13-12877
CS13-12829
CS13-12830
CS13-12690
CS13-12383
CS14-12709
CS14-12913
CS14-12408
CS15-13000
CS15-13019
CS15-13128
CS17-12191
CS17-12611
CS17-12341
CS17-12331
CS20-12104
F2-N2108
3UTR
Other
Unknown?
RNA
miRNA
ncRNA
TTS
LINE
LINE?
srpRNA
SINE
RC
tRNA
DNA?
pseudo
DNA
Exon
Intron
Intergenic
Promoter
5UTR
snoRNA
LTR?
scRNA
CpG - Island
Low_complexity
LTR
Simple_repeat
snRNA
Unknown
SINE?
Satellite
rRNA
d
Imputed ChIP-Seq Data
H4K20me1
H3K79me2
H3K36me3
H3K9ac
H3K4me3
H3K4me2
H3K27ac
DNase
H2A.Z
H3K4me1
H3K9me3
H3K27me3
e
Primary + Imputed ChIP-Seq Data
H3K27me3
H3K9me3
H3K36me3
H4K20me1
H3K79me2
H3K4me1
H3K27ac
DNase
H3K4me2
H3K9ac
H2A.Z
H3K4me3

**Figure S1. Detailed Histone Modification Profiles in Human Craniofacial Development. a.** Heatmap and hierarchical clustering of pairwise Pearson correlations for 114 individual histone modification profiles from human craniofacial tissues. Darker orange indicates positive correlation between datasets. Enlarged from **Fig. 2a** to include sample details, showing samples cluster closely by histone mark. **b.** Correlation of only H3K27ac data contained in the area boxed in black in part **a**. Heatmap and hierarchical clustering show that the samples cluster well into groups by early or late stage of development. **c.** Genomic feature annotations identified by peak calls from six histone modification profiles from all craniofacial samples, across all Carnegie stages, plotted as cumulative percentage of total peaks. Peak enrichments and genomic annotations were performed using HOMER (Heinz et al., 2010). **d.** Heatmap and hierarchical clustering of pairwise Pearson correlations for imputed histone modification profiles from human craniofacial tissues. Darker orange indicates positive correlation between datasets. **e.** Heatmap and hierarchical clustering of pairwise Pearson correlations for imputed and primary histone modification profiles from human craniofacial tissues. Darker orange indicates positive correlation between datasets. Related to Figure 2.

# 15 Chromatin State Model




# 18 Chromatin State Model




# CNCC 25 State Chromatin Segmentations for CNCC Data

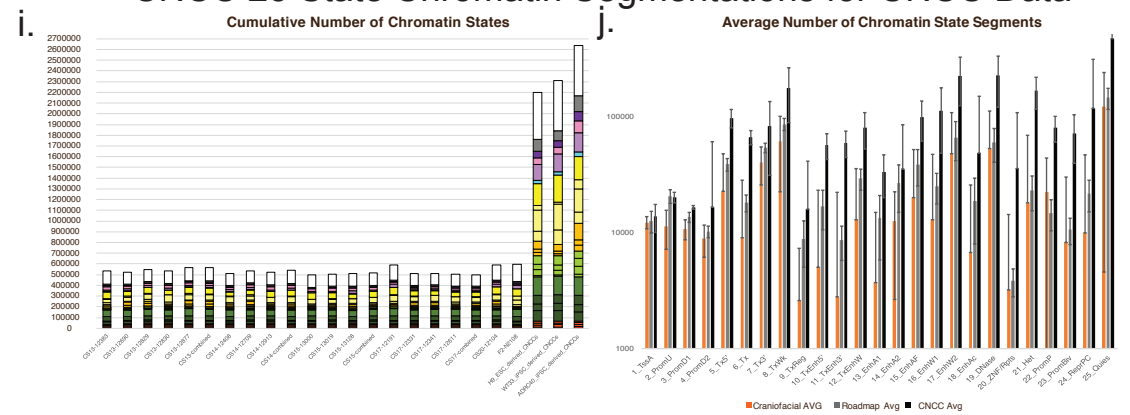


**Figure S2. Imputation of Craniofacial Epigenomic Signals and Chromatin State Segmentation in the 15-State (Primary) and 18-State (Auxiliary) ChromHMM models. a.** Numbers of individual chromatin state segments identified by each of the color- coded 15 states of chromatin activity based on imputed epigenomic signals for each of the 21 tissue samples profiled. **b**. Comparison of cumulative percentage of each chromatin state between craniofacial samples profiled here and 127 segmentations generated by Roadmap Epigenome (Roadmap Epigenomics et al., 2015). **c**. Mean numbers of segments annotated in each of the 15 states across 21 craniofacial samples (orange) and 127 Roadmap Epigenomes (gray). **d.** Mean percentages of segments annotated in each of the 15 states across 21 craniofacial samples (orange) and 127 Roadmap Epigenomes (gray). **e.** Same as in panel **a**, but for 18-State Model. **f.** Same as in panel **b**, but for 18-State Model. **g.** Same as in panel **c**, but for 18-State Model. **h.** Same as in panel **d**, but for 18-State Model. Overall chromatin state segmentation in craniofacial samples identifies similar numbers and percentages of each of the states published by Roadmap Epigenome (Roadmap Epigenomics et al., 2015). **i.** Numbers of individual chromatin state segments for each of the colored 25 states in human craniofacial tissues and CNCCs. **j.** Mean numbers of segments annotated in each of the 25 states across craniofacial samples (orange), Roadmap (grey), and CNCCs (black). Error bars in all plots represent standard deviation. Related to Figure 3.

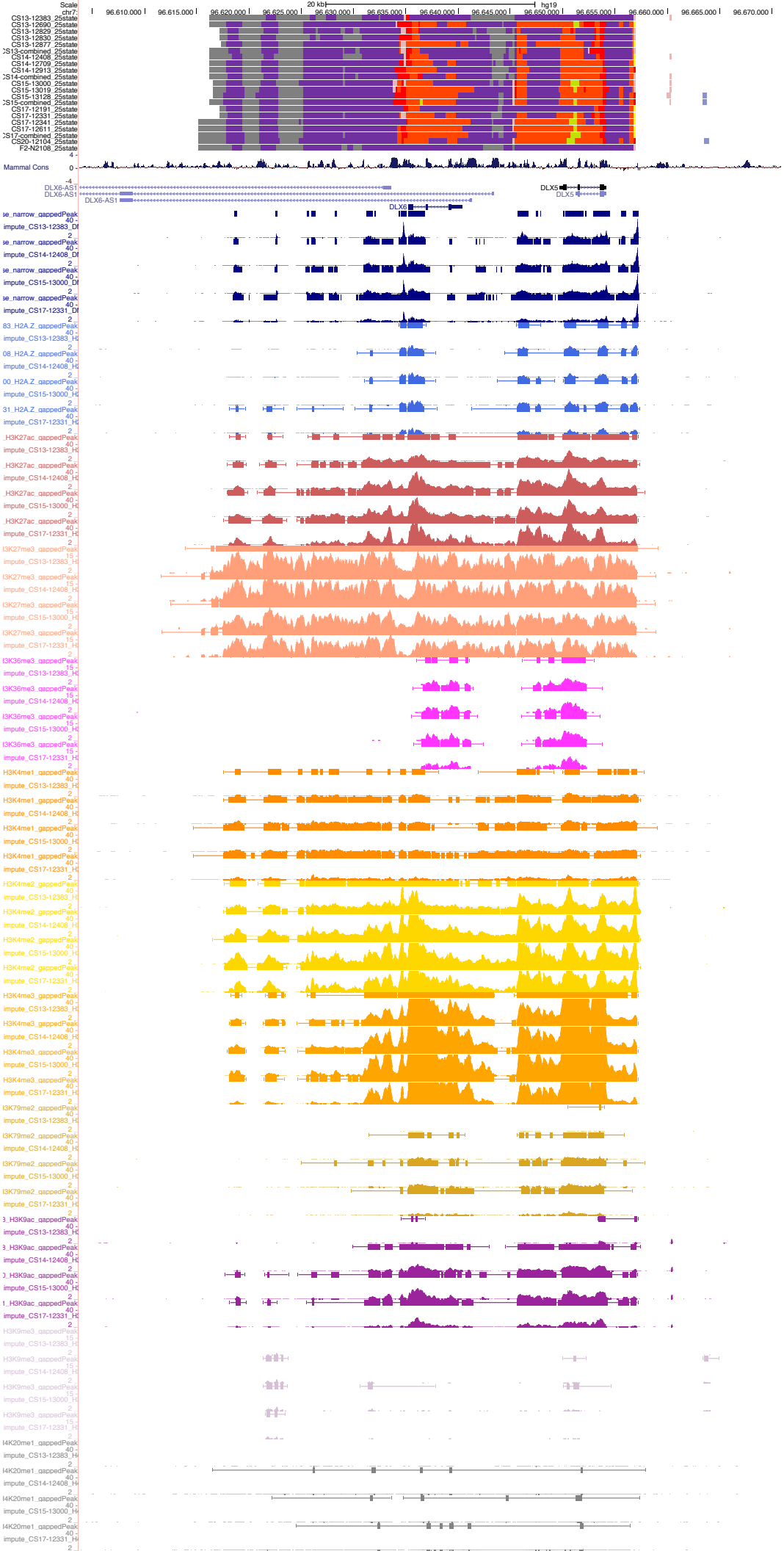
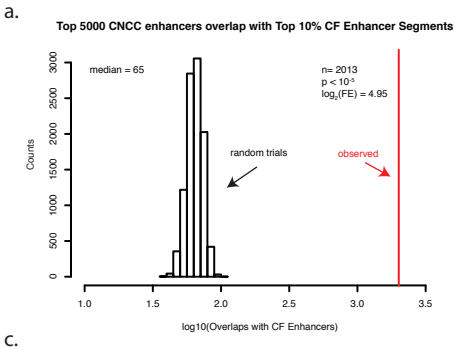
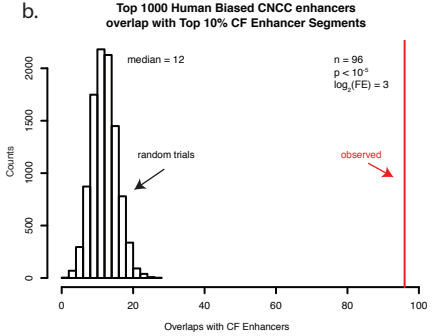

Scale
chr7:
20 kb
hg19
96.610.000
96.615.000
96.620.000
96.625.000
96.630.000
96.635.000
96.640.000
96.645.000
96.650.000
96.655.000
96.660.000
96.665.000
96.670.000
Mammal Cons
DLX6-AS1
DLX6-AS1
DLX6-AS1
DLX6
DLX5
DLX5

**Figure S3. Large Bivalent Domains at Gene Pair *DLX5* and *DLX6*.** UCSC Genome Browser shot of locus encompassing the *DLX5/DLX6* locus. At top are chromatin state segmentations for all tissue samples. See Figure S2 for color code. Purple states indicate bivalent regions. Imputed signals and peak calls for representative samples from each stage and for each mark are shown below. Related to Figure 4.

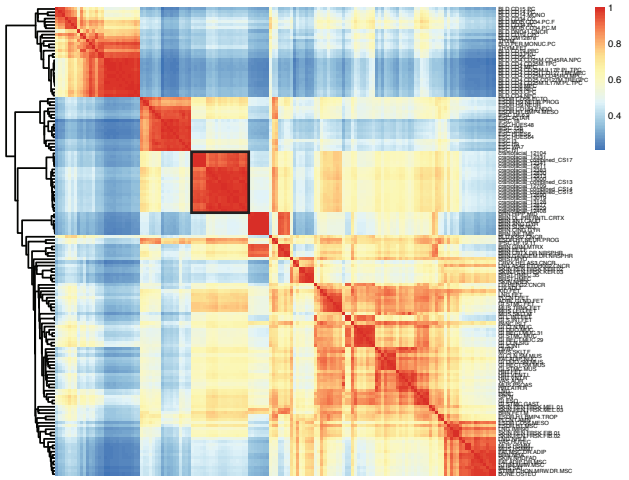

a.
Top 5000 CNCC enhancers overlap with Top 10% CF Enhancer Segments
median = 65
n= 2013
p < 10⁻⁵
log2(FE) = 4.95
random trials
observed
Counts
log10(Overlaps with CF Enhancers)


b.
Top 1000 Human Biased CNCC enhancers
overlap with Top 10% CF Enhancer Segments
median = 12
n = 96
p < 10⁻⁵
log2(FE) = 3
random trials
observed
Counts
Overlaps with CF Enhancers

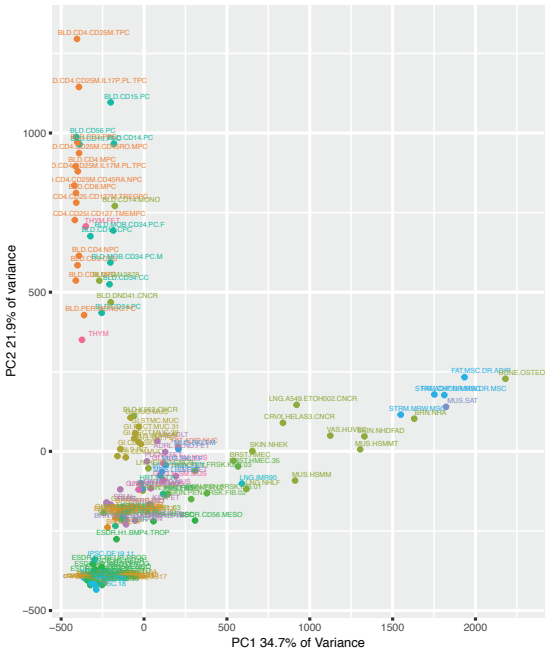
c.

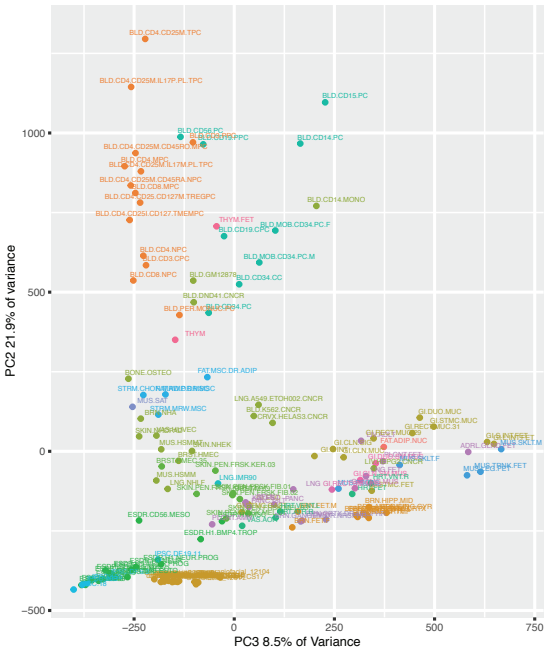
**Figure S4. Overlap with CNCC enhancers and All Enhancers Tested for Craniofacial Activity a.** Distribution of overlaps of CNCC active enhancer regions with 10,000 trials of random enhancer sets equal in number, length, and chromosomal distribution to the top 10% of craniofacial enhancer segments (n=7500). Red vertical line indicates observed number of overlaps between CNCC enhancers and craniofacial enhancers. **b.** Distribution of overlaps of human biased CNCC enhancer regions with 10,000 trials of random enhancer sets equal in number, length, and chromosomal distribution to the top 10% of craniofacial enhancer segments (n=7500). Red vertical line indicates observed number of overlaps between human biased CNCC enhancers and craniofacial enhancers. **c.** All enhancers identified and tested by this study from the Vista Enhancer Browser. Enhancers with hs prefix indicated the human genomic sequence was tested while those with the mm prefix indicate that the orthologous sequence from mouse identified by this study was tested. Related to Figure 5.

a Pairwise Correlation of H3K27ac Signals




b Principal Component Analysis of H3K27ac Signals



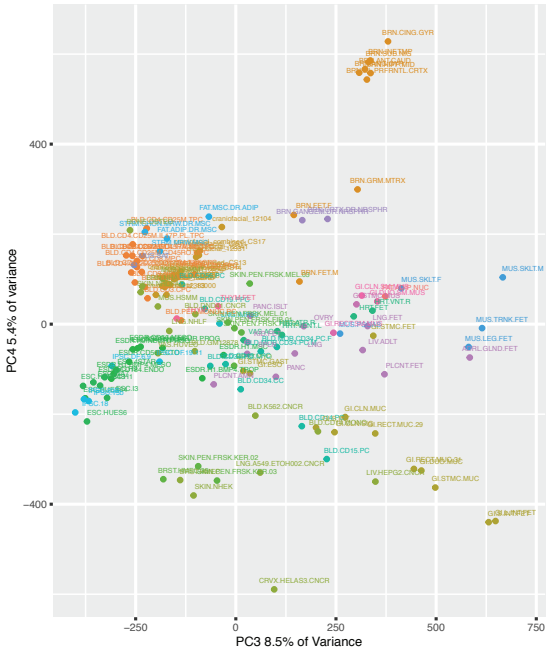




**Figure S5 H3K27ac Signal at Enhancer Segments Allows for Correlation by Tissue Type.** **a.** Heatmap and hierarchical clustering of pairwise comparisons of H3K27ac signals at all enhancer segments in our craniofacial data and the 127 samples from Roadmap Epigenome. Red coloring indicates positive correlation between datasets, blue indicates less correlation. **b.** Principal component analyses of the first four component dimensions of H3K27ac signals in a serial progressive fashion (i.e PC1 vs PC2, PC2 vs PC3, etc.). Samples are color coded by tissue type. Related to Figure 5.

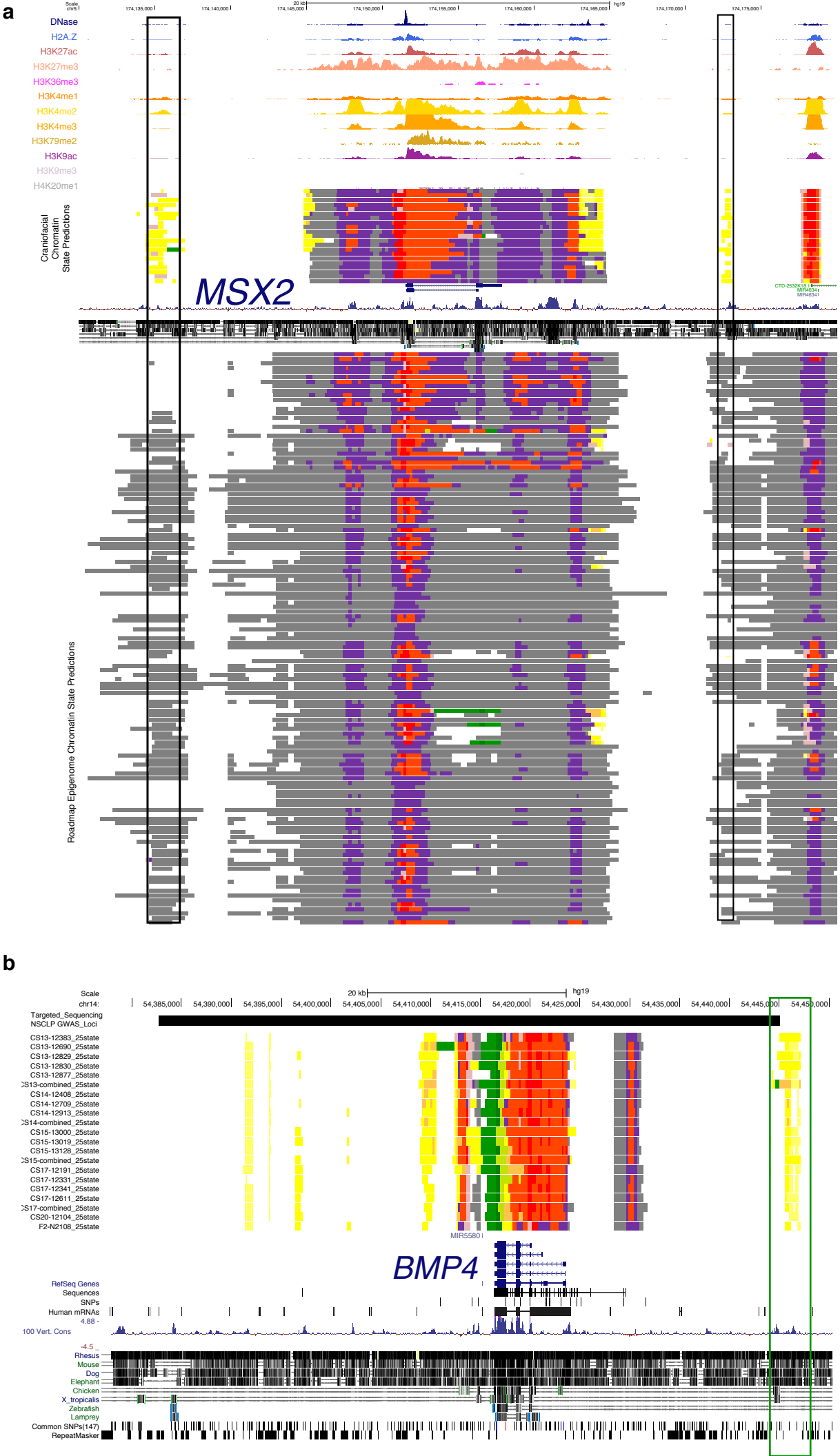

a
DNase
H2A.Z
H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K4me2
H3K4me3
H3K79me2
H3K9ac
H3K9me3
H4K20me1
Craniofacial Chromatin State Predictions
MSX2
Roadmap Epigenome Chromatin State Predictions
b
Scale
chr14:
Targeted_Sequencing
NSCLP GWAS_Loci
CS13-12383_25state
CS13-12690_25state
CS13-12829_25state
CS13-12830_25state
CS13-12877_25state
CS13-combined_25state
CS14-12408_25state
CS14-12709_25state
CS14-12913_25state
CS14-combined_25state
CS15-13000_25state
CS15-13019_25state
CS15-13128_25state
CS15-combined_25state
CS17-12191_25state
CS17-12331_25state
CS17-12341_25state
CS17-12611_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
MIR5580
BMP4
RefSeq Genes
Sequences
SNPs
Human mRNAs
100 Vert. Cons
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Common SNPs(147)
RepeatMasker

**Figure S6. Identification of Craniofacial-specific Enhancers Flanking *MSX2* and *BMP4*. a.** Enhancer states annotated by the 25-state model that are found only in craniofacial tissue but not the 127 samples from Roadmap Epigenome are located upstream and downstream of *MSX2*, a gene implicated in multiple craniofacial abnormalities. The enhancer states fall within a region of conservation and are supported at top by ChIP signals from a single human craniofacial tissue sample. **b.** Targeted sequencing of 13 Loci Identified by GWAS studies to be important In craniofacial development misses a regulatory region in *BMP4***.** The study by Leslie et al. (Leslie et al., 2015) performed targeted sequencing of a region of ~60 kb surrounding the BMP4 gene (black bar at top of figure). This region excluded a region immediately adjacent (outlined by green box) identified as an enhancer by the 25-State, Imputed ChromHMM model in all 21 craniofacial tissues analyzed. Related to Figure 5.

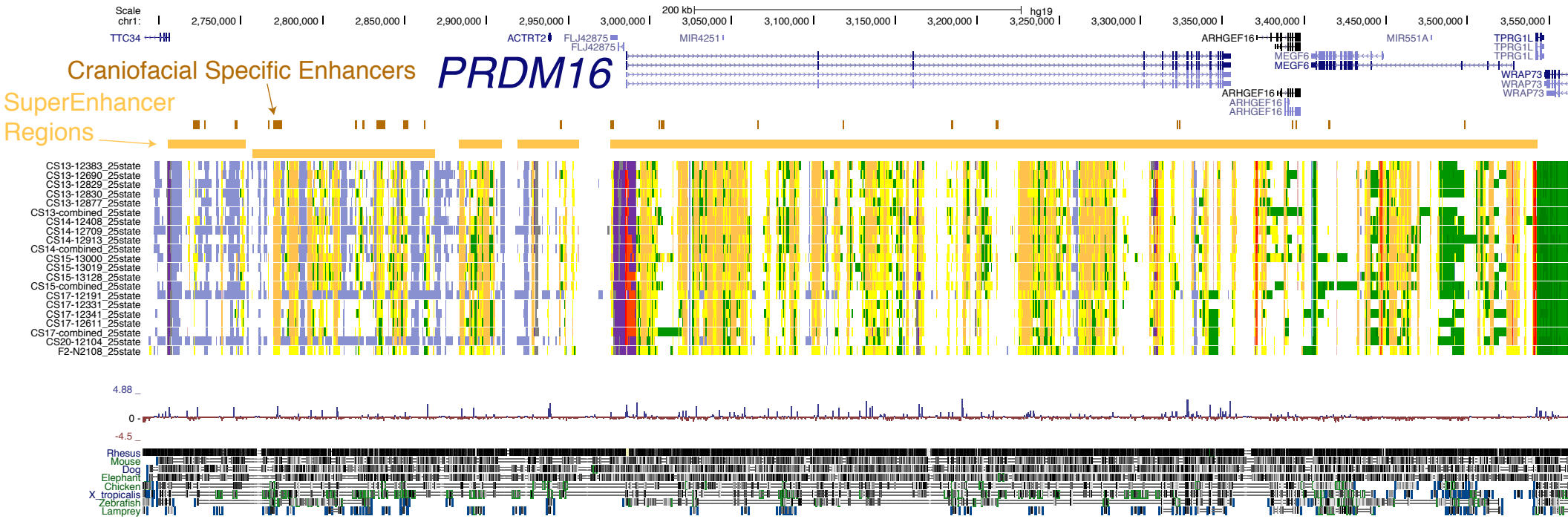

Craniofacial Specific Enhancers
PRDM16
SuperEnhancer Regions
Scale
chr1:
200 kb
hg19
2,750,000
2,800,000
2,850,000
2,900,000
2,950,000
3,000,000
3,050,000
3,100,000
3,150,000
3,200,000
3,250,000
3,300,000
3,350,000
3,400,000
3,450,000
3,500,000
3,550,000
TTC34
ACTRT2
FLJ42875
FLJ42875
MIR4251
ARHGEF16
MIR551A
TPRG1L
TPRG1L
TPRG1L
MEGF6
MEGF6
WRAP73
WRAP73
WRAP73
ARHGEF16
ARHGEF16
ARHGEF16
CS13-12383_25state
CS13-12690_25state
CS13-12829_25state
CS13-12830_25state
CS13-12877_25state
CS13-combined_25state
CS14-12408_25state
CS14-12709_25state
CS14-12913_25state
CS14-combined_25state
CS15-13000_25state
CS15-13019_25state
CS15-13128_25state
CS15-combined_25state
CS17-12191_25state
CS17-12331_25state
CS17-12341_25state
CS17-12611_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
4.88
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey

**Figure S7. The *PRDM16* locus is a Super-enhancer region**. UCSC Genome Browser shot of locus encompassing the *PRDM16* locus and the noncoding region upstream. Super-enhancer regions as identify by ROSE are indicated by orange bars (Whyte et al., 2013). Craniofacial specific enhancers indicated by darker orange bars. Numerous craniofacial enhancer segments are annotated throughout this regions in all samples profiled. Related to Figure 7.

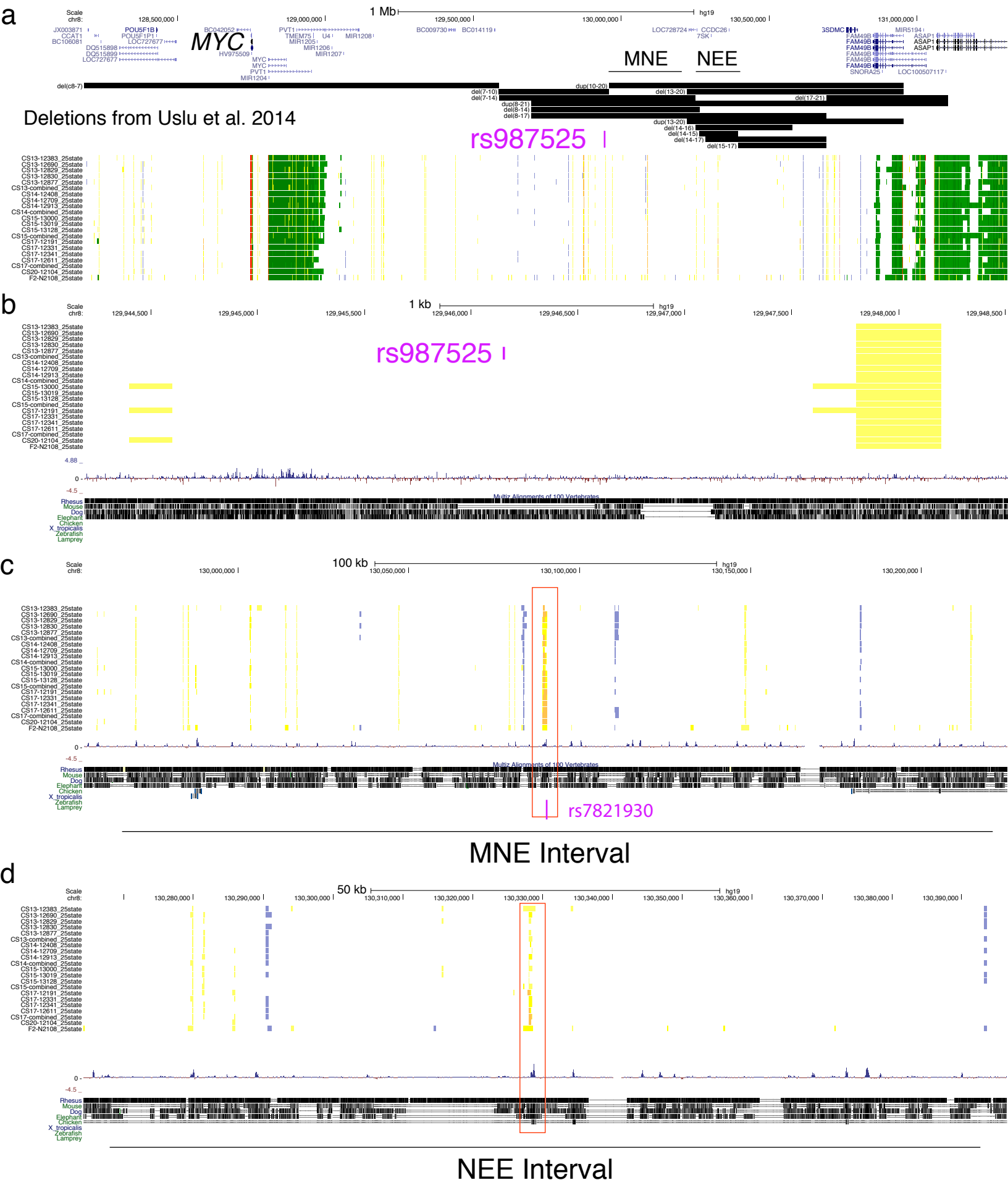

a
MYC
MNE
NEE
Deletions from Uslu et al. 2014
rs987525
b
rs987525
c
rs7821930
MNE Interval
d
NEE Interval

**Figure S8. Human Craniofacial Enhancer States prioritize regions within the 8q24 Clefting locus. a.** UCSC Genome Browser shot of 8q24 locus implicated in orofacial clefting in humans and mice. This region has been examined using multiple mouse deletion lines as indicated by black bars and deletion number (Uslu et al., 2014). Two regions implicated as harboring important regulatory elements are indicated by MNE and NEE. Strongest risk allele position from orofacial clefting GWAS indicated in purple (Birnbaum et al., 2009). **b.** Enlarged region encompassing the rs987525 position. A consistent craniofacial enhancer segment is identified less than 2kb downstream of this risk position. **c.** Enlarged region encompassing the MNE interval. A reproducibly strong human craniofacial enhancer state is located in the center of this interval as indicated by red box. A single common human SNP is located in this region rs7821930. **d.** Enlarged region encompassing the NEE interval. A reproducibly strong human craniofacial enhancer state is located in the center of this interval as indicated by red box. Related to Figure 7.


a
Sample
Craniofacial
Roadmap
−log10(GREGOR p−value)
log2(Fold Enrichment of Crohn's GWAS SNPs)
Craniofacial Specific Enhancers
BLD.CD19.PPC
BLD.CD4.MPC
BLD.CD56.PC
BLD.CD8.MPC
BLD.CD4.NPC
Thymus
BLD.CD19.CPC
BLD.K562.CNCR
E102
E075
E101
E079
E094
E098
b
Sample
Craniofacial
Roadmap
−log10(GREGOR p−value)
log2(Fold Enrichment of GWAS SNPs from Leslie et al 2017)
E079
CS14−12913
CS15−combined
CS13−12830
CS13−combined
CS14−12709
All Craniofacial Enhancers
Craniofacial Specific Enhancers
E050
c
Sample
Craniofacial
Roadmap
−log10(GREGOR p−value)
log2(Fold Enrichment of Credible SNPs from Ludwig et al 2017)
CS13−12830
CS13−12383
Craniofacial Specific Enhancers
H1.ESC
iPSC.18
All Craniofacial Enhancers
ESDR.H9.NEUR
ESC.HUES64
ESC.4STAR
ESDR.CD184.ENDO
BRN.GANGEM.DR.NRSPHR
CS13−combined
E068
E029

**Figure S9. Significant enrichment of orofacial GWAS SNPs but not Crohn's GWAS SNPs in human craniofacial enhancers. a.** Enrichment of Crohn's GWAS tag SNPs retrieved from the GWAS Catalog in enhancer segmentations assessed using GREGOR (Schmidt et al., 2015). Orange circles indicate craniofacial enhancer annotations identified by a 25 State chromatin model from this study while grey circles indicate those previously published by Roadmap Epigenome (Roadmap Epigenomics et al., 2015). No significant enrichment was detected for craniofacial segmentations,but was observed for multiple immune cell types. **b**. Same analysis as in a using GWAS tag SNPs reported for orofacial clefting by Leslie et al 2017. **c.** Same analysis as in **a** using GWAS tag SNPs reported for 24 regions by Ludwig et al 2017. Related to Figure 7.

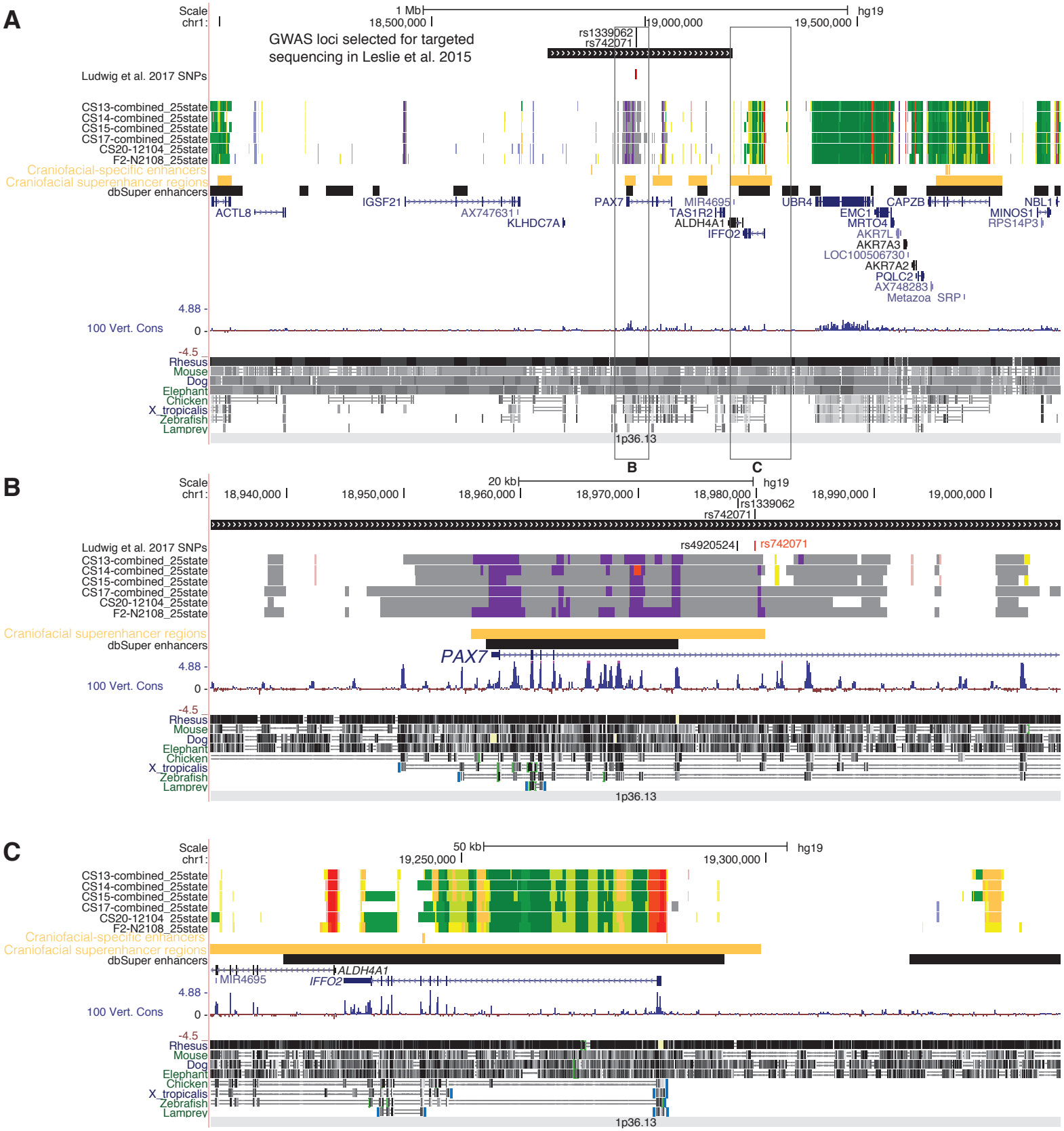

A
Scale
chr1:
1 Mb
18,500,000
19,000,000
19,500,000
hg19
GWAS loci selected for targeted sequencing in Leslie et al. 2015
rs1339062
rs742071
Ludwig et al. 2017 SNPs
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
Craniofacial superenhancer regions
dbSuper enhancers
ACTL8
IGSF21
AX747631
KLHDC7A
PAX7
MIR4695
TAS1R2
ALDH4A1
IFFO2
UBR4
CAPZB
EMC1
MRTO4
AKR7L
AKR7A3
LOC100506730
AKR7A2
PQLC2
AX748283
Metazoa SRP
NBL1
MINOS1
RPS14P3
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1p36.13
B
C
B
Scale
chr1:
20 kb
18,940,000
18,950,000
18,960,000
18,970,000
18,980,000
18,990,000
19,000,000
hg19
rs1339062
rs742071
Ludwig et al. 2017 SNPs
rs4920524
rs742071
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial superenhancer regions
dbSuper enhancers
PAX7
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1p36.13
C
Scale
chr1:
50 kb
19,250,000
19,300,000
hg19
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
Craniofacial superenhancer regions
dbSuper enhancers
MIR4695
ALDH4A1
IFFO2
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1p36.13

**Figure S10 Orofacial Clefting Locus 1p36.13 PAX7** Approximately 2Mb window surrounding *PAX7* (**A**). The region selected for targeted sequencing (Leslie et al. 2015). The lead SNP rs742071 and the associated SNP by sequencing, rs1339062, are shown. The *PAX7* promoter region was also investigated in Ludwig et al. (2017) where rs4920524 was identified as contributing the largest posterior probability risk. Bivalent states (purple) are present at the *PAX7* promoter and the promoter of *IGSF21*. A craniofacial superenhancer overlaps the *PAX7* bivalent region (**B**) and SNPs including rs1339062, rs742071 and rs4920524 are present within a portion of the superenhancer region that has not been identified as a superenhancer in any tissues or cell types in the dbSuper database (Khan and Zhang, 2016). Other large craniofacial superenhancers carrying craniofacial-specific enhancers are within 1Mb of *PAX7*. The region near *IFFO2* (**C**) is shown as an example. Related to Figure 7.

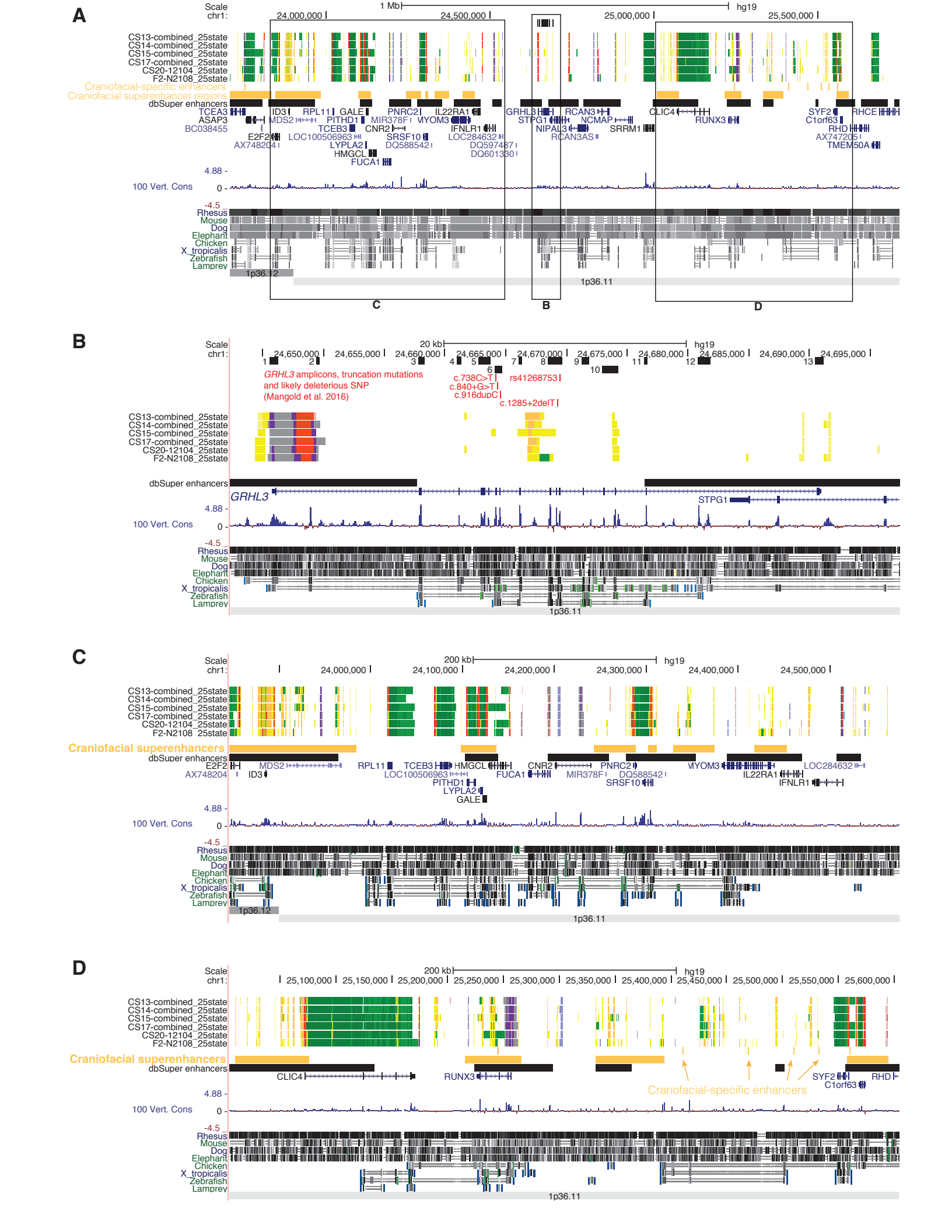

A
Scale
chr1:
1 Mb
hg19
24,000,000
24,500,000
25,000,000
25,500,000
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
Craniofacial superenhancer regions
dbSuper enhancers
TCEA3
ASAP3
BC038455
E2F2
AX748204
ID3
MDS2
RPL11
PITHD1
TCEB3
LOC100506963
LYPLA2
HMGCL
FUCA1
GALE
PNRC2
MIR378F
CNR2
SRSF10
IL22RA1
VMYOM3
IFNLR1
LOC284632
DQ588542
DQ597487
DQ601380
GRHL3
STPG1
NIPAL3
RCAN3
NCMAP
RCAN3AS
CLIC4
SRRM1
RUNX3
SYF2
C1orf63
RHD
RHCE
AX747205
TMEM50A
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1p36.12
1p36.11
C
B
D
B
Scale
chr1:
20 kb
hg19
24,650,000
24,655,000
24,660,000
24,665,000
24,670,000
24,675,000
24,680,000
24,685,000
24,690,000
24,695,000
GRHL3 amplicons, truncation mutations
and likely deleterious SNP
(Mangold et al. 2016)
c.738C>T
rs41268753
c.840+G>T
c.916dupC
c.1285+2delT
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
dbSuper enhancers
GRHL3
STPG1
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1p36.11
C
Scale
chr1:
200 kb
hg19
24,000,000
24,100,000
24,200,000
24,300,000
24,400,000
24,500,000
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial superenhancers
dbSuper enhancers
E2F2
AX748204
ID3
MDS2
RPL11
PITHD1
TCEB3
LOC100506963
LYPLA2
HMGCL
FUCA1
GALE
CNR2
MIR378F
PNRC2
DQ588542
SRSF10
VMYOM3
IL22RA1
IFNLR1
LOC284632
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1p36.12
1p36.11
D
Scale
chr1:
200 kb
hg19
25,100,000
25,150,000
25,200,000
25,250,000
25,300,000
25,350,000
25,400,000
25,450,000
25,500,000
25,550,000
25,600,000
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial superenhancers
dbSuper enhancers
CLIC4
RUNX3
SYF2
C1orf63
RHD
Cranofacial-specific enhancers
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1p36.11

**Figure S11 Orofacial Clefting Locus 1p36.11 GRHL3** Approximately 2Mb region surrounding *GRHL3*, a gene implicated in Van der Woude syndrome and non-syndromic cleft palate only (Peyrard-Janvid et al., 2014). (**A**). *GRHL3* itself shows a poised promoter state. Amplicons for targeted sequencing in cases of individuals with non-syndromic orofacial clefts by Mangold et al. (Mangold et al., 2016), truncations mutations and the likely deleterious SNP rs41268753 are shown in panel **B.** Additionally, an intronic region between amplicons 7 and 8 contains enhancer states in all craniofacial timepoints surveyed. Regions to either side of *GRHL3* contain craniofacial superenhancers and many enhancer states, including craniofacial-specific enhancers suggesting this area may contain several genes important in early craniofacial development (**C,D**). Related to Figure 7.

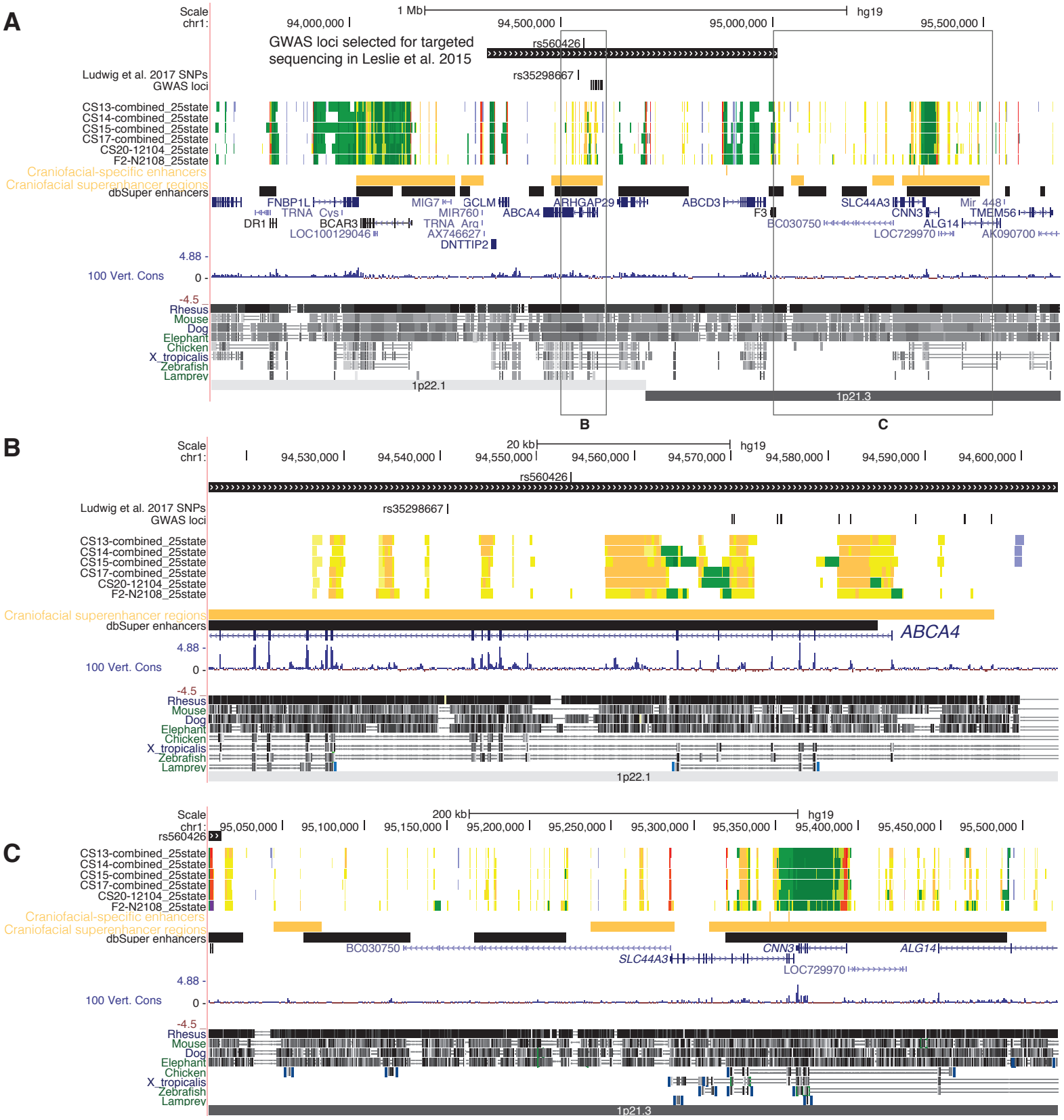

A
Scale
chr1:
1 Mb
hg19
94,000,000
94,500,000
95,000,000
95,500,000
GWAS loci selected for targeted sequencing in Leslie et al. 2015
rs560426
rs35298667
Ludwig et al. 2017 SNPs
GWAS loci
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
Craniofacial superenhancer regions
dbSuper enhancers
FNBP1L
TRNA Cys
DR1
BCAR3
LOC100129046
MIG7
GCLM
MIR760
TRNA Arg
AX746627
DNTTIP2
ARHGAP29
ABCA4
ABCD3
F3
BC030750
SLC44A3
CNN3
ALG14
LOC729970
Mir 448
TMEM56
AK090700
4.88
0
100 Vert. Cons
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1p22.1
1p21.3
B
C
B
Scale
chr1:
20 kb
hg19
94,530,000
94,540,000
94,550,000
94,560,000
94,570,000
94,580,000
94,590,000
94,600,000
rs560426
rs35298667
Ludwig et al. 2017 SNPs
GWAS loci
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial superenhancer regions
dbSuper enhancers
ABCA4
4.88
0
100 Vert. Cons
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1p22.1
C
Scale
chr1:
rs560426
200 kb
hg19
95,050,000
95,100,000
95,150,000
95,200,000
95,250,000
95,300,000
95,350,000
95,400,000
95,450,000
95,500,000
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
Craniofacial superenhancer regions
dbSuper enhancers
BC030750
SLC44A3
CNN3
ALG14
LOC729970
4.88
0
100 Vert. Cons
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1p21.3

**Figure S12 Orofacial Clefting Locus 1p22.1 ARHGAP29** Approximately 2Mb region surrounding rs560426 (**A**). The region was selected for targeted sequencing (Leslie et al. 2015). The region also contains rs35298667, identified as a likely significant contributor to nsCL/P (Ludwig et al., 2017) and multiple SNPs identified in GWAS for non-syndromic orofacial clefting (Yu et al. 2017 and Leslie et al. 2017 reported loci) are present within craniofacial superenhancer within the region of targeted resequencing (**B**). Additional craniofacial superenhancers including craniofacial-specific superenhancer regions neighboring the region of targeted resequencing (**C**). Related to Figure 7.

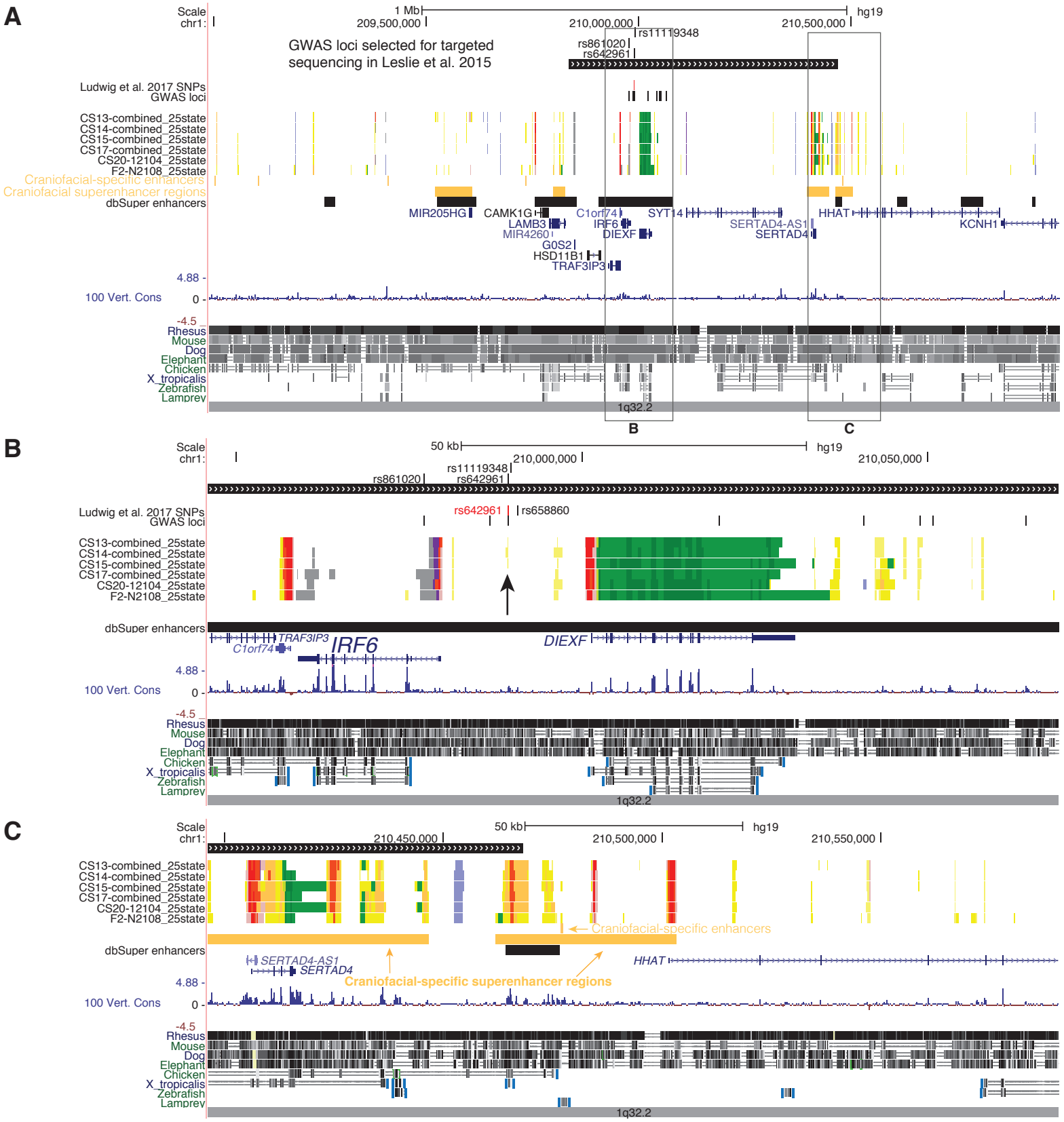

A
Scale
chr1:
1 Mb
209,500,000
210,000,000
210,500,000
hg19
GWAS loci selected for targeted sequencing in Leslie et al. 2015
rs11119348
rs861020
rs642961
Ludwig et al. 2017 SNPs
GWAS loci
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
Craniofacial superenhancer regions
dbSuper enhancers
MIR205HG
CAMK1G
C1orf74
SYT14
HHAT
LAMB3
IRF6
SERTAD4-AS1
KCNH1
MIR4260
DIEXF
SERTAD4
G0S2
HSD11B1
TRAF3IP3
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1q32.2
B
C
B
Scale
chr1:
50 kb
hg19
210,000,000
210,050,000
rs861020
rs11119348
rs642961
rs658860
Ludwig et al. 2017 SNPs
GWAS loci
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
dbSuper enhancers
TRAF3IP3
C1orf74
IRF6
DIEXF
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1q32.2
C
Scale
chr1:
50 kb
hg19
210,450,000
210,500,000
210,550,000
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
dbSuper enhancers
SERTAD4-AS1
SERTAD4
HHAT
Craniofacial-specific superenhancer regions
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
1q32.2

**Figure S13 Orofacial Clefting Locus 1q32.2 IRF6** Approximately 2Mb region surrounding *IRF6*. (**A**) The lead SNP rs642961 falls within a region marked as an active enhancer in embryonic craniofacial tissue (arrow, **B**). Additional SNPs identified through targeted sequencing (Leslie et al., 2015) or GWAS (Yu et al., 2017; Leslie et al., 2017) are found near enhancers active in craniofacial tissue. (**B**) Craniofacial-specific enhancers and craniofacial-specific superenhancers located adjacent to the region of targeted sequencing (**C**). Related to Figure 7.

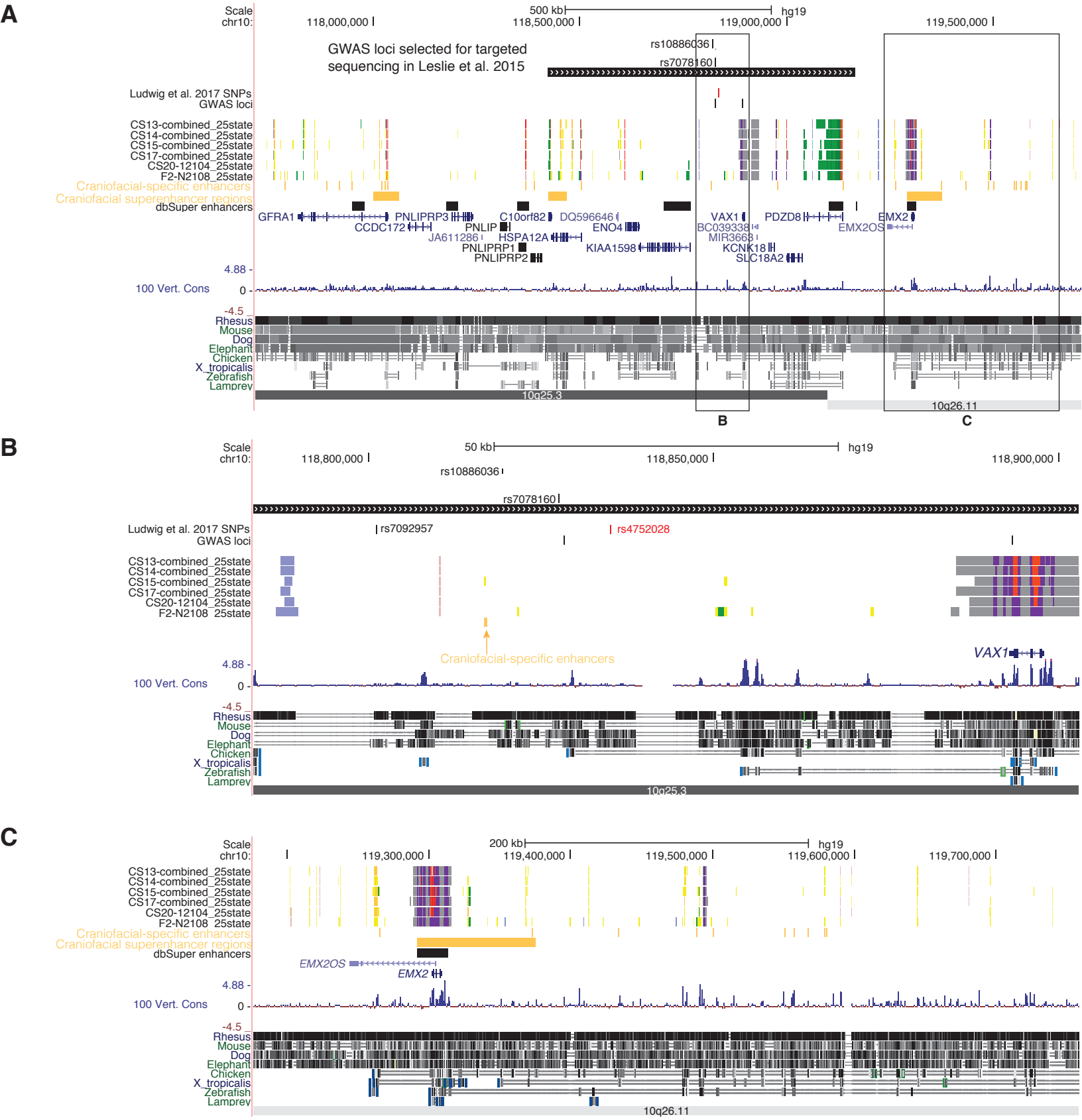

A
Scale
chr10:
500 kb
hg19
118,000,000
118,500,000
119,000,000
119,500,000
GWAS loci selected for targeted sequencing in Leslie et al. 2015
rs10886036
rs7078160
Ludwig et al. 2017 SNPs
GWAS loci
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
Craniofacial superenhancer regions
dbSuper enhancers
GFRA1
PNLIPRP3
C10orf82
DQ596646
VAX1
PDZD8
EMX2
CCDC172
PNLIP
ENO4
BC039338
EMX2OS
JA611286
HSPA12A
MIR3663
PNLIPRP1
KIAA1598
KCNK18
PNLIPRP2
SLC18A2
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
10q25.3
10q26.11
B
C
B
Scale
chr10:
50 kb
hg19
118,800,000
118,850,000
118,900,000
rs10886036
rs7078160
Ludwig et al. 2017 SNPs
GWAS loci
rs7092957
rs4752028
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
VAX1
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
10q25.3
C
Scale
chr10:
200 kb
hg19
119,300,000
119,400,000
119,500,000
119,600,000
119,700,000
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
Craniofacial superenhancer regions
dbSuper enhancers
EMX2OS
EMX2
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
10q26.11

**Figure S14 Orofacial Clefting Locus 10q25.3 VAX1** Approximately 1.5Mb region surrounding *VAX1* (**A**). The *VAX1* promoter region contains a bivalent chromatin state and SNPs associated with non-syndromic oral clefts are located near a craniofacial-specific enhancer (**B**). The neighboring gene *EMX2* also shows a bivalent chromatin state as well as proximity to a craniofacial-specific enhancer (**C**). Related to Figure 7.

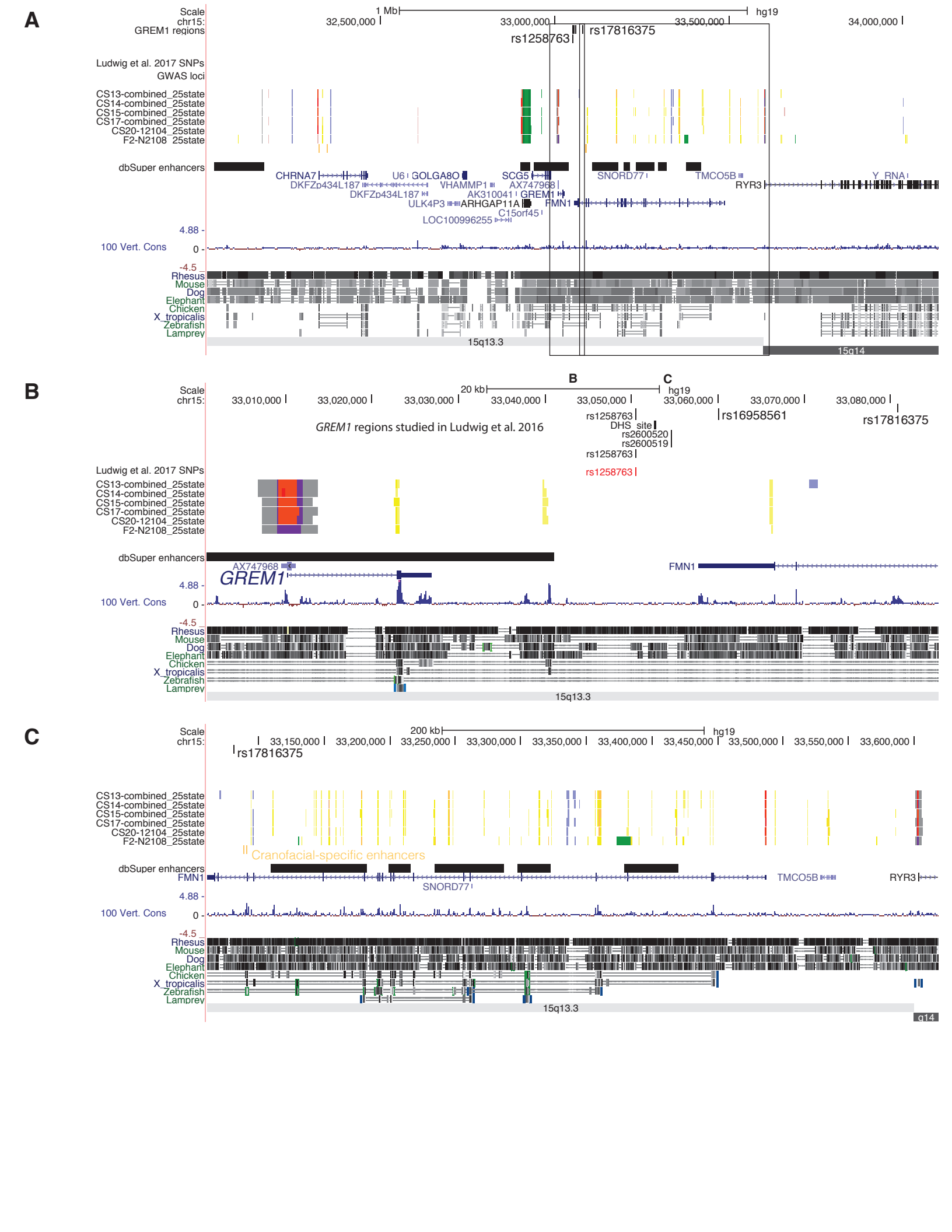

A
Scale
chr15:
GREM1 regions
1 Mb
hg19
32,500,000
33,000,000
33,500,000
34,000,000
rs1258763
rs17816375
Ludwig et al. 2017 SNPs
GWAS loci
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
dbSuper enhancers
CHRNA7
U6
GOLGA8O
SCG5
SNORD77
TMCO5B
Y_RNA
RYR3
DKFZp434L187
VHAMMP1
AX747968
DKFZp434L187
AK310041
GREM1
ULK4P3
ARHGAP11A
FMN1
C15orf45
LOC100996255
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
15q13.3
15q14
B
Scale
chr15:
20 kb
B
C
hg19
33,010,000
33,020,000
33,030,000
33,040,000
33,050,000
33,060,000
33,070,000
33,080,000
rs1258763
rs16958561
rs17816375
GREM1 regions studied in Ludwig et al. 2016
DHS site
rs2600520
rs2600519
rs1258763
rs1258763
Ludwig et al. 2017 SNPs
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
dbSuper enhancers
AX747968
GREM1
FMN1
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
15q13.3
C
Scale
chr15:
200 kb
hg19
33,150,000
33,200,000
33,250,000
33,300,000
33,350,000
33,400,000
33,450,000
33,500,000
33,550,000
33,600,000
rs17816375
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Cranofacial-specific enhancers
dbSuper enhancers
FMN1
SNORD77
TMCO5B
RYR3
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
15q13.3
q14

**Figure S15 Orofacial Clefting Locus 15q13.3 GREM1** Approximately 2Mb region containing the intergenic region between *GREM1* and *FMN1*, found to have a strong association with nsCL/P and to be predominant in a rare form of clefting with lip and soft palate cleft but intact hard palate (Ludwig et al. 2016) (**A**). SNPs identified in the intergenic region and in the promoter and intronic region of FMN1 are shown in panel **B.** The SNP rs17816375, found to have the strongest eQTL effect by Ludwig et al. is near craniofacial-specific enhancers in the intron of *FMN1*. Additionally, *FMN1* introns contain many enhancer states with patterns that suggest differences in enhancer activity between embryonic and fetal craniofacial development (**C**). Related to Figure 7.

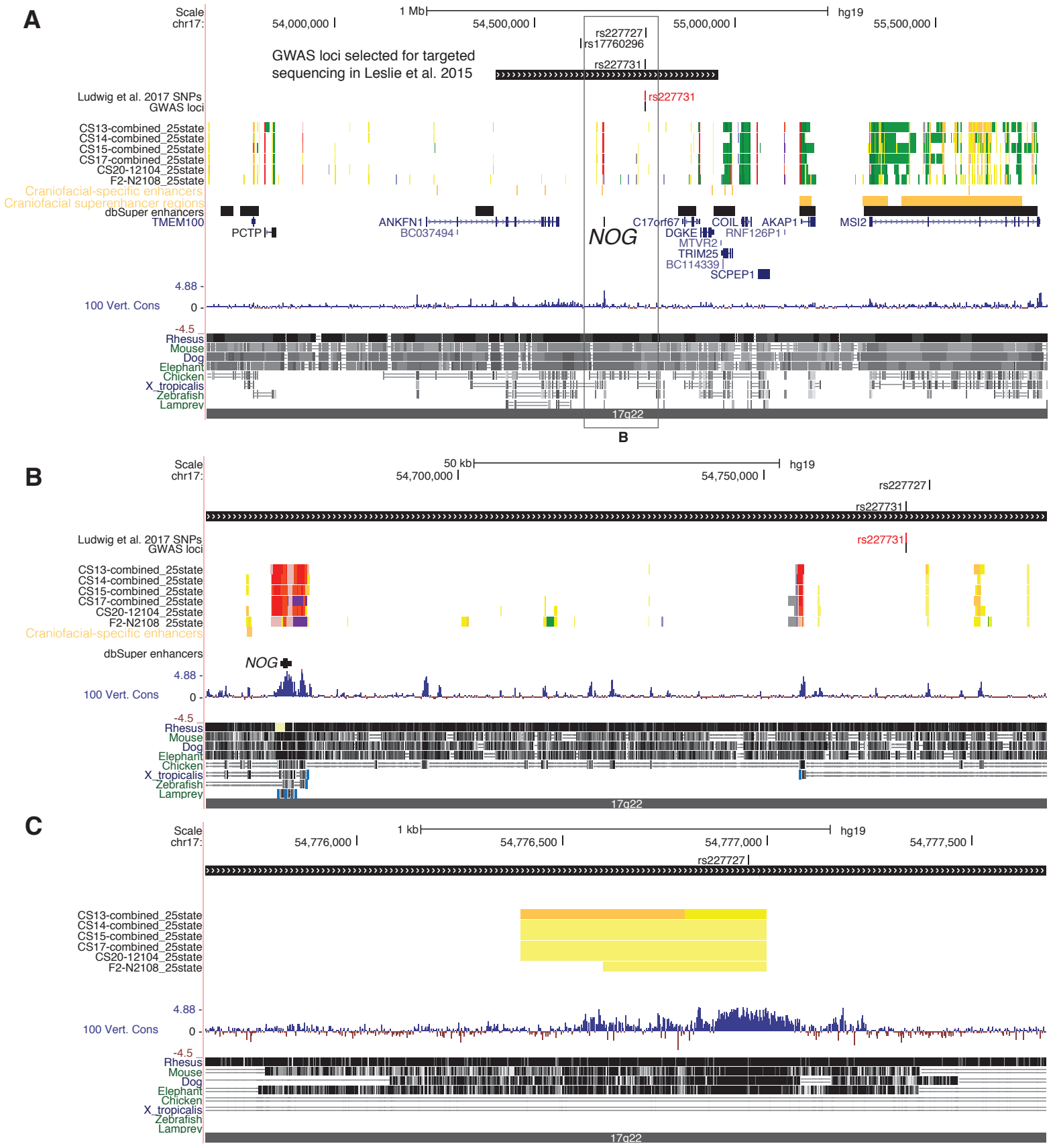

A
Scale
chr17:
1 Mb
hg19
54,000,000
54,500,000
55,000,000
55,500,000
rs227727
rs17760296
GWAS loci selected for targeted
sequencing in Leslie et al. 2015
rs227731
Ludwig et al. 2017 SNPs
GWAS loci
rs227731
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
Craniofacial superenhancer regions
dbSuper enhancers
TMEM100
PCTP
ANKFN1
BC037494
NOG
C17orf67
DGKE
COIL
AKAP1
RNF126P1
MTVR2
TRIM25
BC114339
SCPEP1
MSI2
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
17q22
B
Scale
chr17:
50 kb
hg19
54,700,000
54,750,000
rs227727
rs227731
Ludwig et al. 2017 SNPs
GWAS loci
rs227731
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
dbSuper enhancers
NOG
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
17q22
C
Scale
chr17:
1 kb
hg19
54,776,000
54,776,500
54,777,000
54,777,500
rs227727
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
17q22

**Figure S16 Orofacial Clefting Locus 17q22 NOG** Approximately 2Mb region surrounding the gene for the BMP antagonist *NOG* (**A**). The active promoter and bivalent states of *NOG* at different stages of embryonic and fetal development reflect the expression of *NOG* in early craniofacial development (Matsui and Klingensmith, 2014). SNPs associated with non-syndromic oral clefts reside near or in enhancer states active in early craniofacial development (**B,C**). Related to Figure 7.

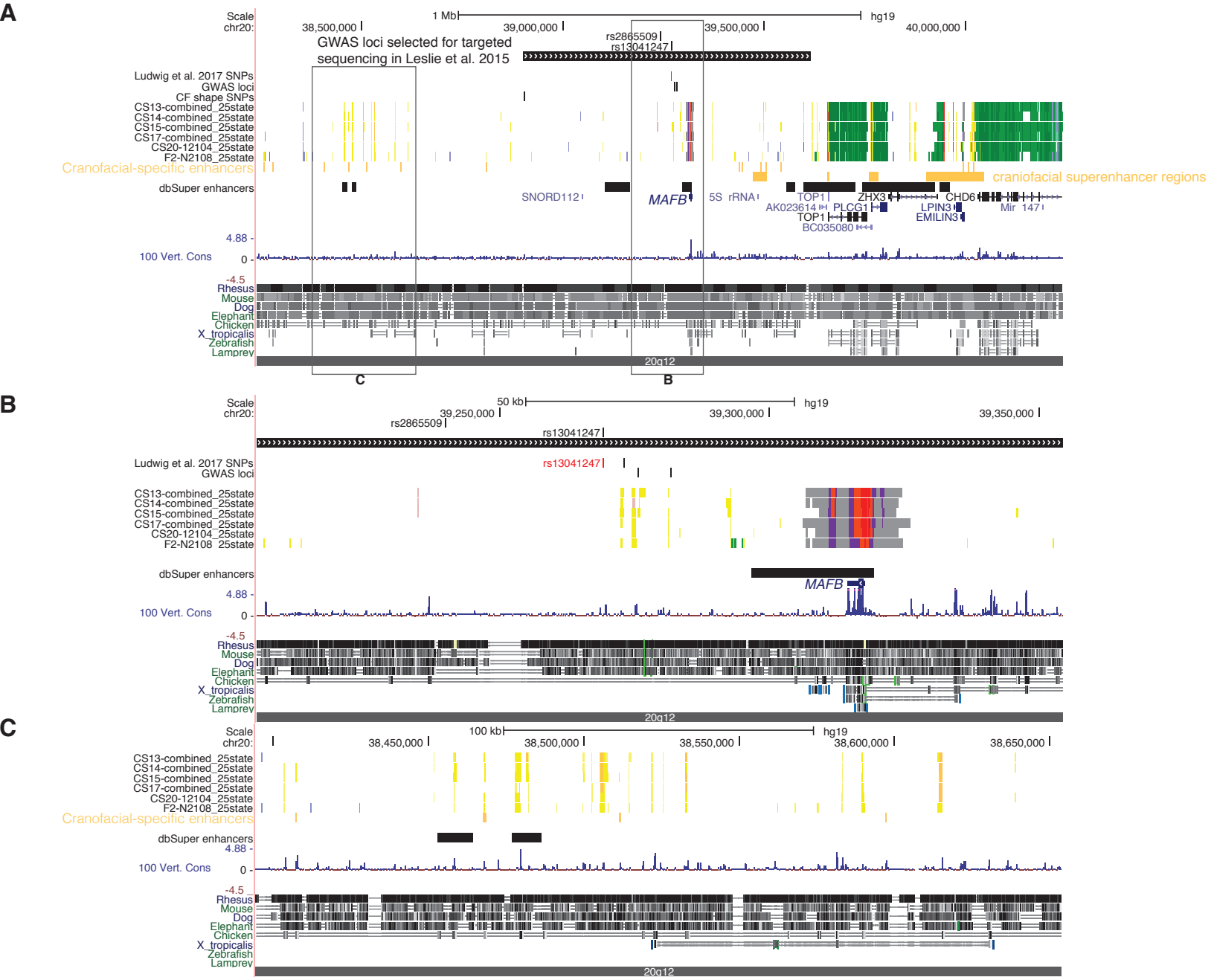

A
Scale
chr20:
1 Mb
hg19
38,500,000
39,000,000
39,500,000
40,000,000
GWAS loci selected for targeted sequencing in Leslie et al. 2015
rs2865509
rs13041247
Ludwig et al. 2017 SNPs
GWAS loci
CF shape SNPs
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Cranofacial-specific enhancers
craniofacial superenhancer regions
dbSuper enhancers
SNORD112
MAFB
5S rRNA
TOP1
ZHX3
CHD6
AK023614
PLCG1
LPIN3
Mir 147
TOP1
EMILIN3
BC035080
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
20q12
C
B
B
Scale
chr20:
50 kb
hg19
39,250,000
39,300,000
39,350,000
rs2865509
rs13041247
Ludwig et al. 2017 SNPs
GWAS loci
rs13041247
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
dbSuper enhancers
MAFB
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
20q12
C
Scale
chr20:
100 kb
hg19
38,450,000
38,500,000
38,550,000
38,600,000
38,650,000
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Cranofacial-specific enhancers
dbSuper enhancers
4.88
100 Vert. Cons
0
-4.5
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
20q12

**Figure S17 Orofacial Clefting Locus 20q12 MAFB** Approximately 2Mb region around *MAFB* (**A**). SNPs associated with non-syndromic oral clefts reside in or near enhancer states active in early craniofacial development (**B**). A region <1Mb away from *MAFB* contains several craniofacial-specific enhancers and patterns of enhancer state that may suggest differences in early vs. late craniofacial development (**C**). Related to Figure 7.

**Supplemental References**


Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. In Genome Biology.

Beaty, T.H., Ruczinski, I., Murray, J.C., Marazita, M.L., Munger, R.G., Hetmanski, J.B., Murray, T., Redett, R.J., Fallin, M.D., Liang, K.Y.*, et al.* (2011). Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. Genet Epidemiol *35*, 469-478.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological *57*, 289-300.

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

Cotney, J.L., and Noonan, J.P. (2015). Chromatin immunoprecipitation with fixed animal tissues and preparation for high-throughput sequencing. Cold Spring Harb Protoc *2015*, 191-199.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat Methods *9*, 215-216.

Ernst, J., and Kellis, M. (2015). Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. Nat Biotechnol *33*, 364-376.

Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047-3048.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat Protoc *7*, 1728-1740.

Gokhman, D., Kelman, G., Amartely, A., Gershon, G., Tsur, S., and Carmel, L. (2017). Gene ORGANizer: linking genes to the organs they affect. Nucleic Acids Res, 106948.

Grant, S.F., Wang, K., Zhang, H., Glaberson, W., Annaiah, K., Kim, C.E., Bradfield, J.P., Glessner, J.T., Thomas, K.A., Garris, M.*, et al.* (2009). A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. J Pediatr *155*, 909-913.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Ing-Simmons, E., Seitan, V.C., Faure, A.J., Flicek, P., Carroll, T., Dekker, J., Fisher, A.G., Lenhard, B., and Merkenschlager, M. (2015). Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. Genome Res *25*, 504-513.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res *12*, 996-1006.

Khan, A., and Zhang, X. (2016). dbSUPER: a database of super-enhancers in mouse and human genome. Nucleic Acids Res *44*, D164-171.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P.*, et al.* (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res *22*, 1813-1831.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Leslie, E.J., Taub, M.A., Liu, H., Steinberg, K.M., Koboldt, D.C., Zhang, Q., Carlson, J.C., Hetmanski, J.B., Wang, H., Larson, D.E.*, et al.* (2015). Identification of functional variants for

cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. Am J Hum Genet *96*, 397-411.
Ludwig, K.U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A., Becker, J., Herberz, R., AlChawa, T., Nasser, E.*, et al.* (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat Genet *44*, 968-971.
Mangold, E., Bohmer, A.C., Ishorst, N., Hoebel, A.K., Gultepe, P., Schuenke, H., Klamt, J., Hofmann, A., Golz, L., Raff, R.*, et al.* (2016). Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate. Am J Hum Genet *98*, 755-762.
Mangold, E., Ludwig, K.U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., Reutter, H., de Assis, N.A., Chawa, T.A., Mattheisen, M.*, et al.* (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat Genet *42*, 24-26.
Matsui, M., and Klingensmith, J. (2014). Multiple tissue-specific requirements for the BMP antagonist Noggin in development of the mammalian craniofacial skeleton. Dev Biol *392*, 168-181.
McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.
Peyrard-Janvid, M., Leslie, E.J., Kousa, Y.A., Smith, T.L., Dunnwald, M., Magnusson, M., Lentz, B.A., Unneberg, P., Fransson, I., Koillinen, H.K.*, et al.* (2014). Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. Am J Hum Genet *94*, 23-32.
Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England) *26*, 841-842.
R Core Team (2017). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing).
Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Research *42*, W187-191.
Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S.*, et al.* (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665-1680.
Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J.*, et al.* (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330.
Schmidt, E.M., Zhang, J., Zhou, W., Chen, J., Mohlke, K.L., Chen, Y.E., and Willer, C.J. (2015). GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. Bioinformatics (Oxford, England) *31*, 2601-2606.
Shaffer, J.R., Orlova, E., Lee, M.K., Leslie, E.J., Raffensperger, Z.D., Heike, C.L., Cunningham, M.L., Hecht, J.T., Kau, C.H., Nidey, N.L.*, et al.* (2016). Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. PLoS genetics *12*, e1006149-1006121.
Shi, M., Murray, J.C., Marazita, M.L., Munger, R.G., Ruczinski, I., Hetmanski, J.B., Wu, T., Murray, T., Redett, R.J., Wilcox, A.J.*, et al.* (2012). Genome wide study of maternal and parent-of-origin effects on the etiology of orofacial clefts. American journal of medical genetics Part A *158A*, 784-794.
Uslu, V.V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N.A., Marioni, J.C., and Spitz, F. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. Nature genetics.
Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Research *35*, D88-92.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L.*, et al.* (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Research *42*, D1001-1006.
Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.