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.

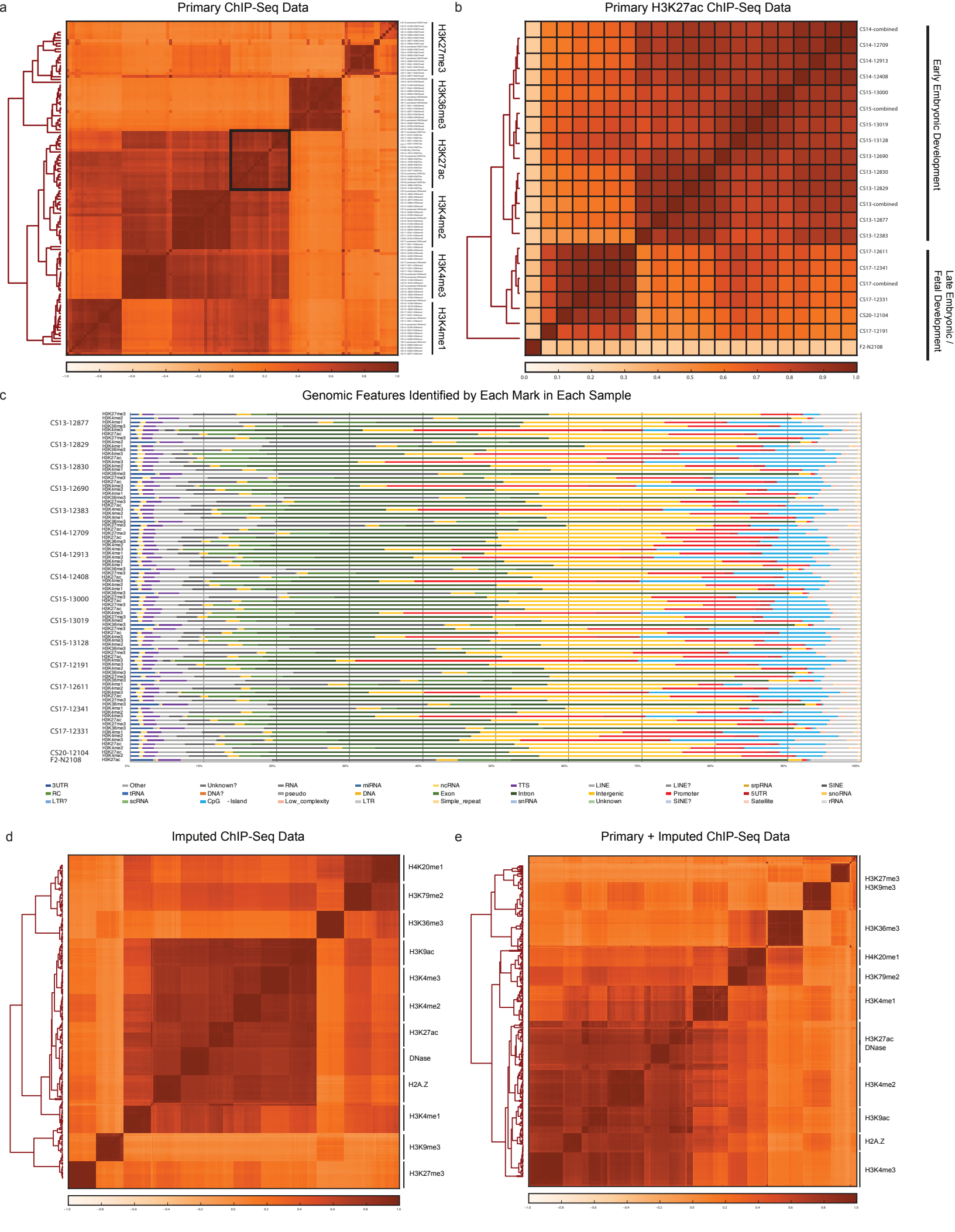
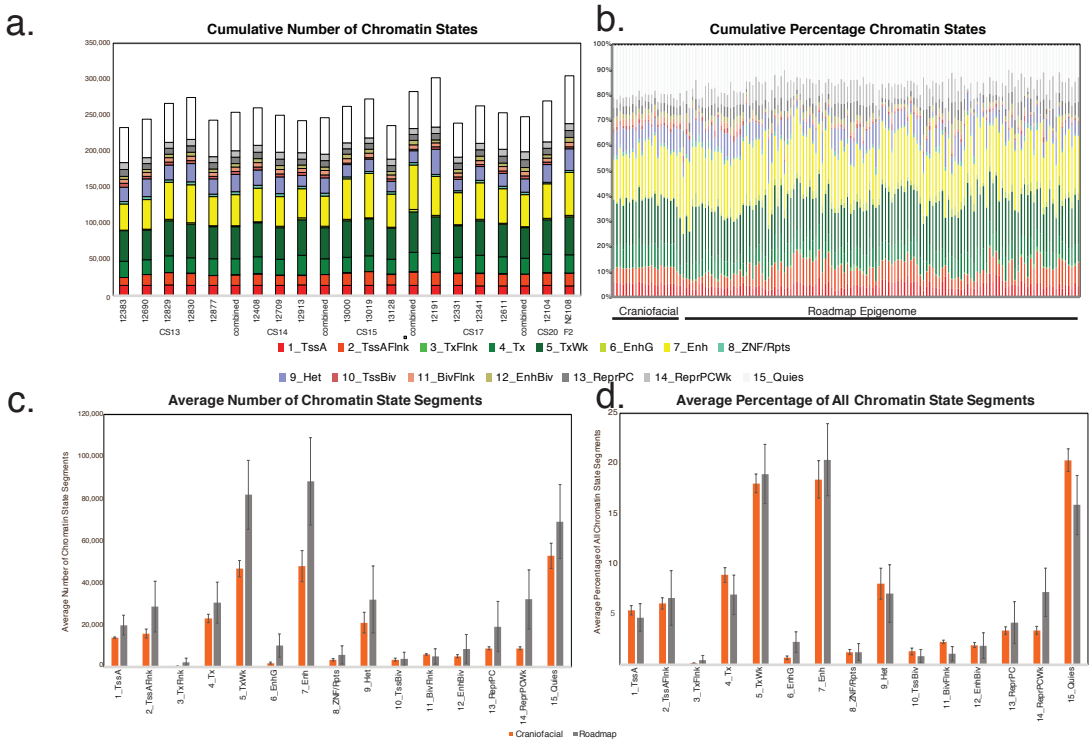
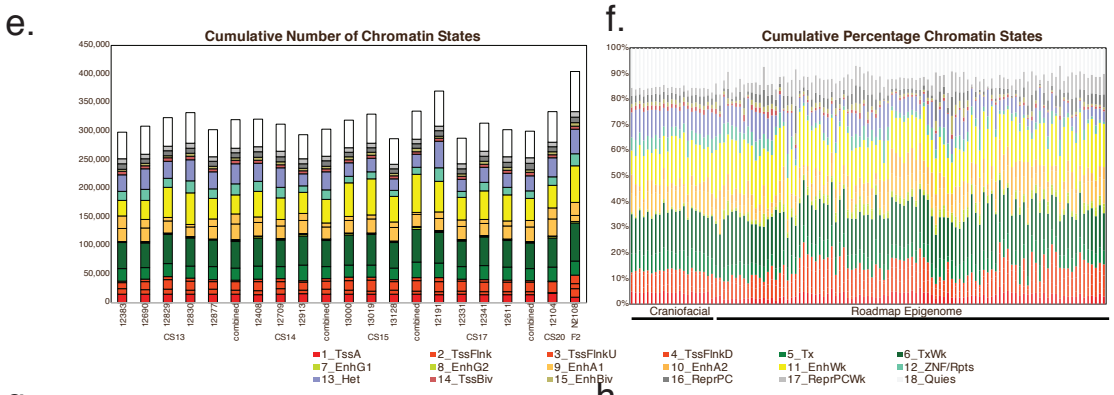

a
Primary ChIP-Seq Data
H3K27me3
H3K36me3
H3K27ac
H3K4me2
H3K4me3
H3K4me1
b
Primary H3K27ac ChIP-Seq Data
CS14-combined
CS14-12709
CS14-12913
CS14-12408
CS15-13000
CS15-combined
CS15-13019
CS15-13128
CS13-12690
CS13-12830
CS13-12829
CS13-combined
CS13-12877
CS13-12383
CS17-12611
CS17-12341
CS17-combined
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

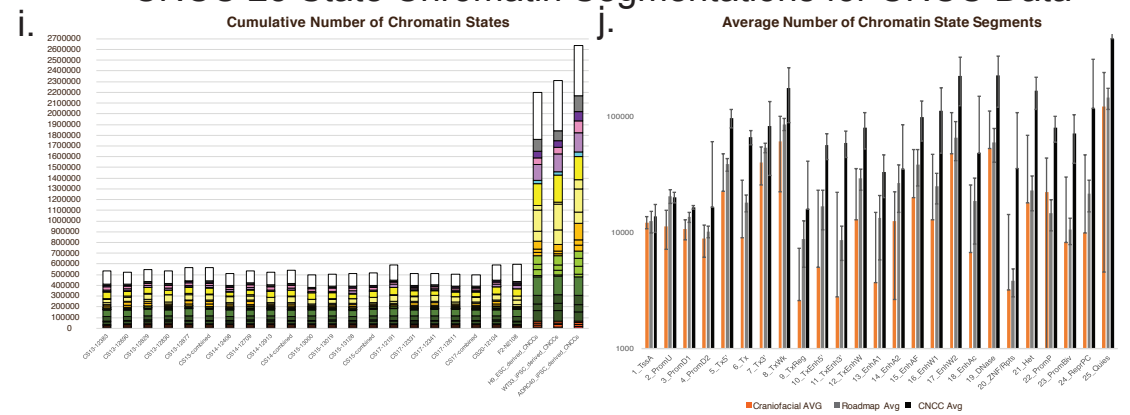


Figure S2. Imputation of Craniofacial Epigenomic Signals and Chromatin State Segmentation in the 15-State (Primary) and 18-State (Auxiliary) ChromHMM models. a. Numbers of individual chromatin state segments identified by each of the color- coded 15 states of chromatin activity based on imputed epigenomic 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.

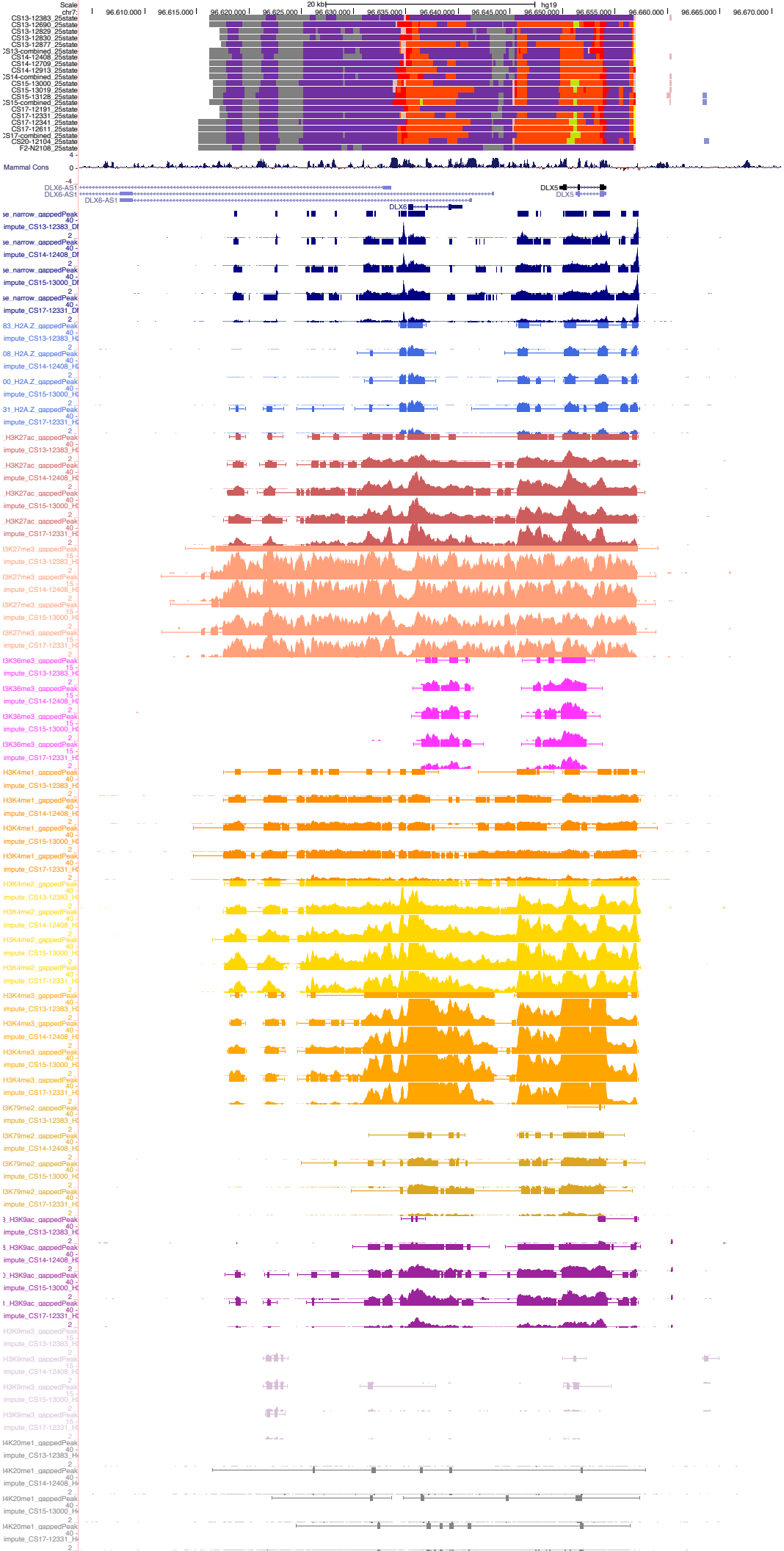
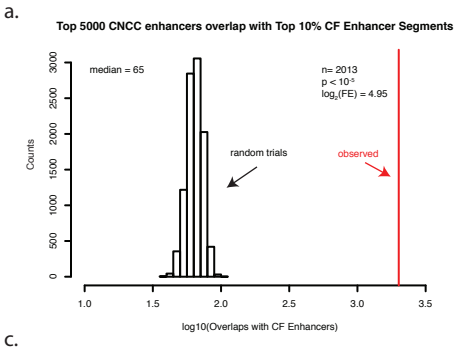
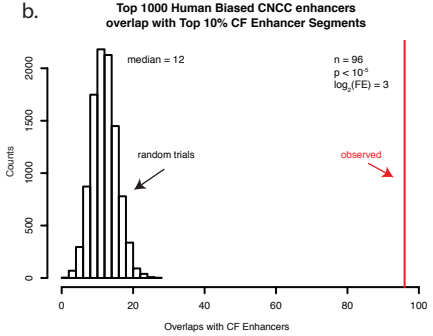

Scale
chr7:
20 kb
hg19
96.610.000
96.615.000
96.620.000
96.625.000
96.630.000
96.635.000
96.640.000
96.645.000
96.650.000
96.655.000
96.660.000
96.665.000
96.670.000
Mammal Cons
DLX6-AS1
DLX6-AS1
DLX6-AS1
DLX6
DLX5
DLX5

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.

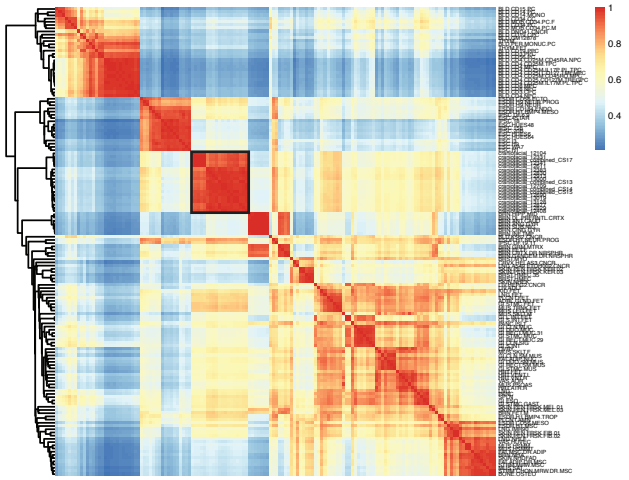

a.
Top 5000 CNCC enhancers overlap with Top 10% CF Enhancer Segments
median = 65
n= 2013
p < 10⁻⁵
log2(FE) = 4.95
random trials
observed
Counts
log10(Overlaps with CF Enhancers)


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

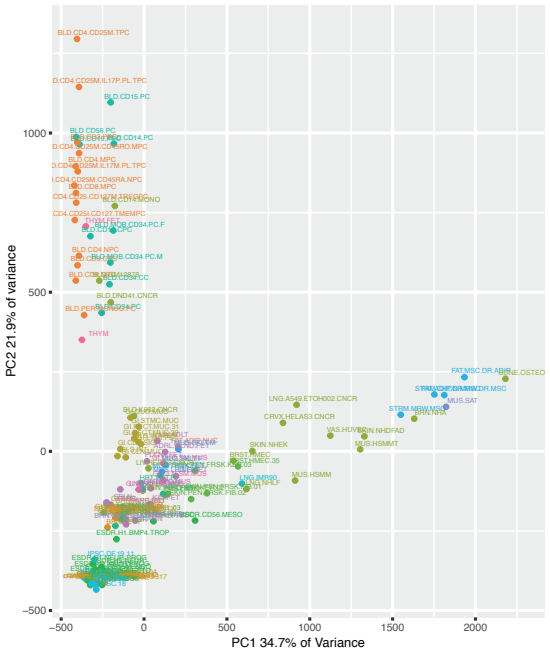
c.

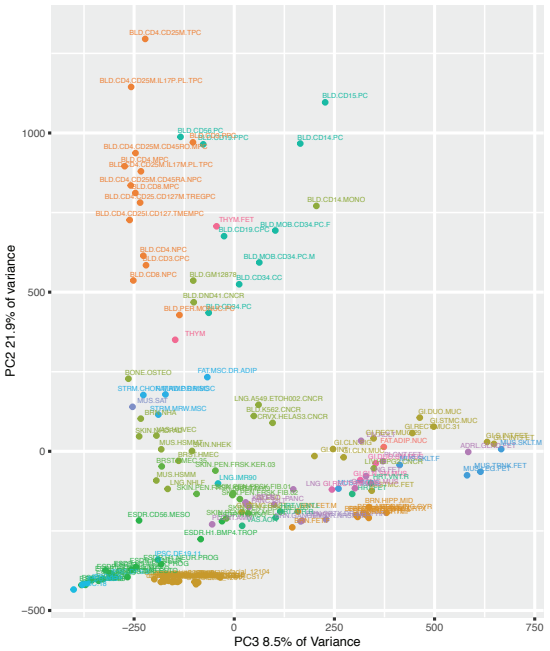
Figure S4. Overlap with CNCC enhancers and All Enhancers Tested for Craniofacial Activity a. Distribution of overlaps of CNCC active enhancer regions with 10,000 trials of random enhancer sets equal in number, length, and chromosomal distribution to the top 10% of craniofacial enhancer segments (n=7500). 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





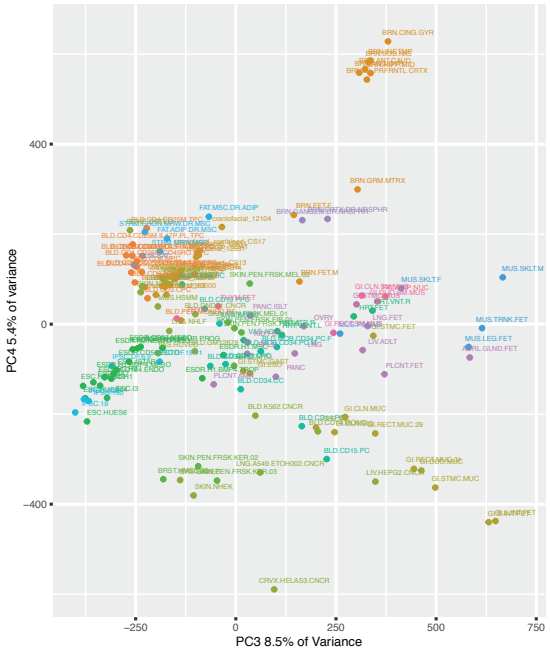


Figure S5 H3K27ac Signal at Enhancer Segments Allows for Correlation by Tissue Type. **a.** Heatmap and hierarchical clustering of pairwise comparisons of H3K27ac signals at all enhancer segments in our craniofacial data and the 127 samples from Roadmap Epigenome. Red coloring indicates positive correlation 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.

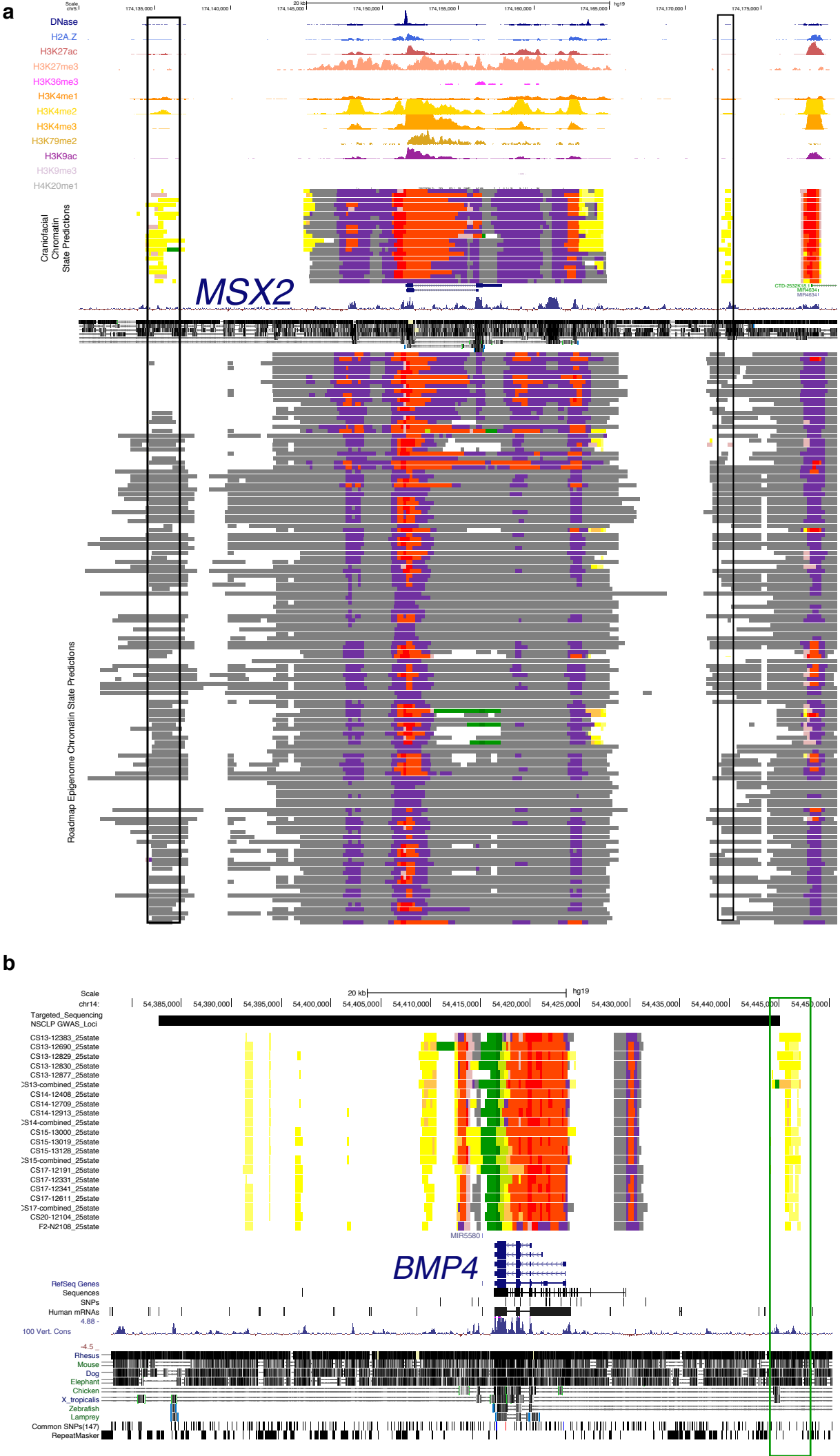

a
DNase
H2A.Z
H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K4me2
H3K4me3
H3K79me2
H3K9ac
H3K9me3
H4K20me1
Craniofacial Chromatin State Predictions
MSX2
Roadmap Epigenome Chromatin State Predictions
b
Scale
chr14:
Targeted_Sequencing
NSCLP GWAS_Loci
CS13-12383_25state
CS13-12690_25state
CS13-12829_25state
CS13-12830_25state
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.

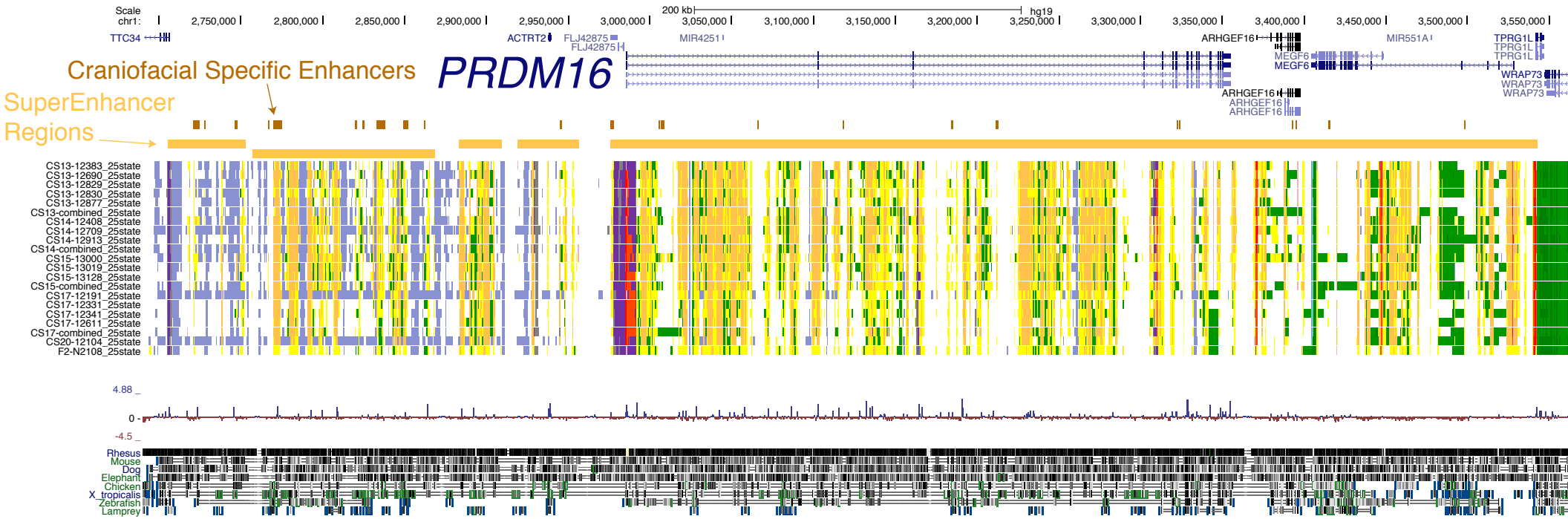

Craniofacial Specific Enhancers
PRDM16
SuperEnhancer Regions
Scale
chr1:
200 kb
hg19
2,750,000
2,800,000
2,850,000
2,900,000
2,950,000
3,000,000
3,050,000
3,100,000
3,150,000
3,200,000
3,250,000
3,300,000
3,350,000
3,400,000
3,450,000
3,500,000
3,550,000
TTC34
ACTRT2
FLJ42875
FLJ42875
MIR4251
ARHGEF16
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.

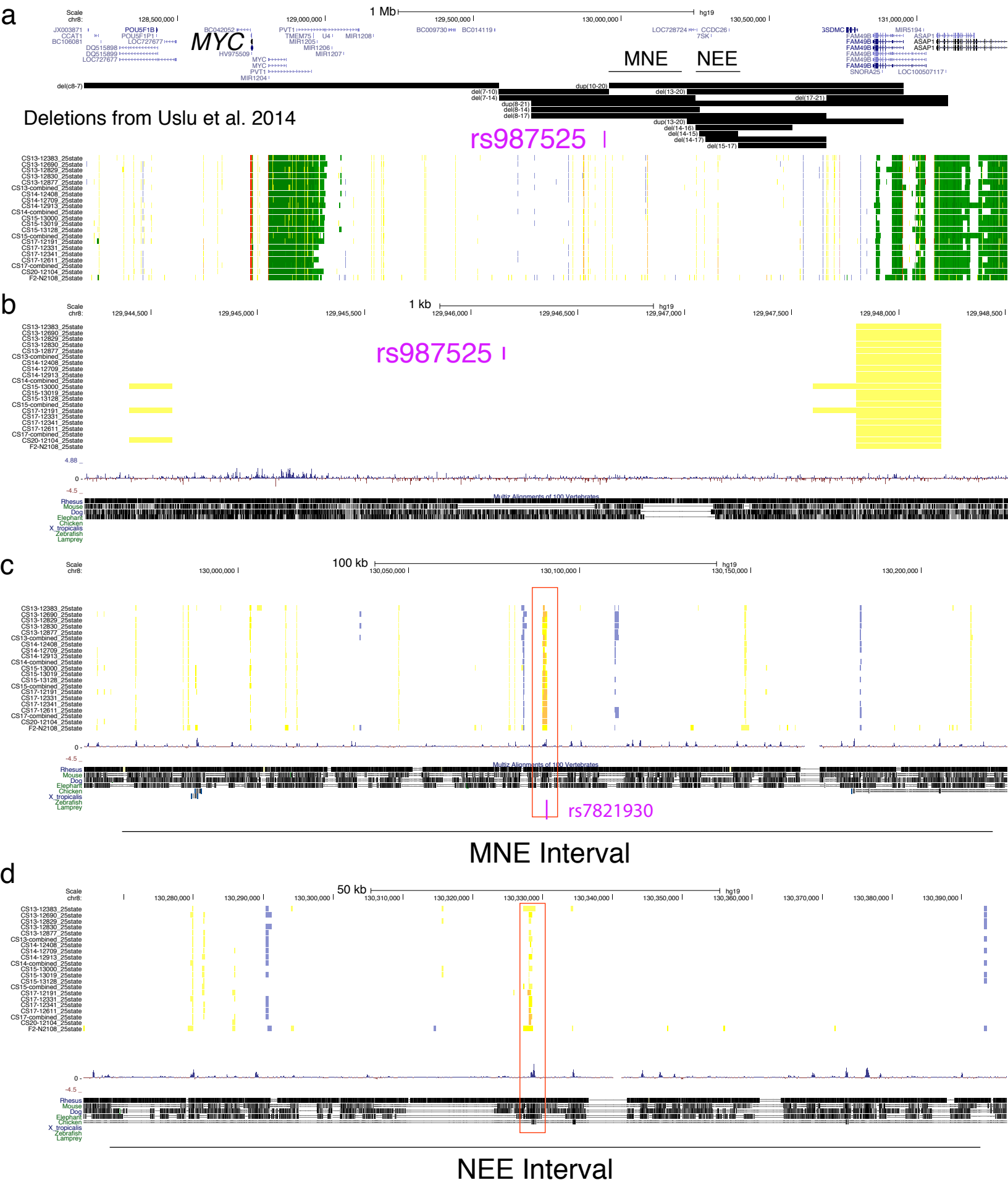

a
MYC
MNE
NEE
Deletions from Uslu et al. 2014
rs987525
b
rs987525
c
rs7821930
MNE Interval
d
NEE Interval

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.

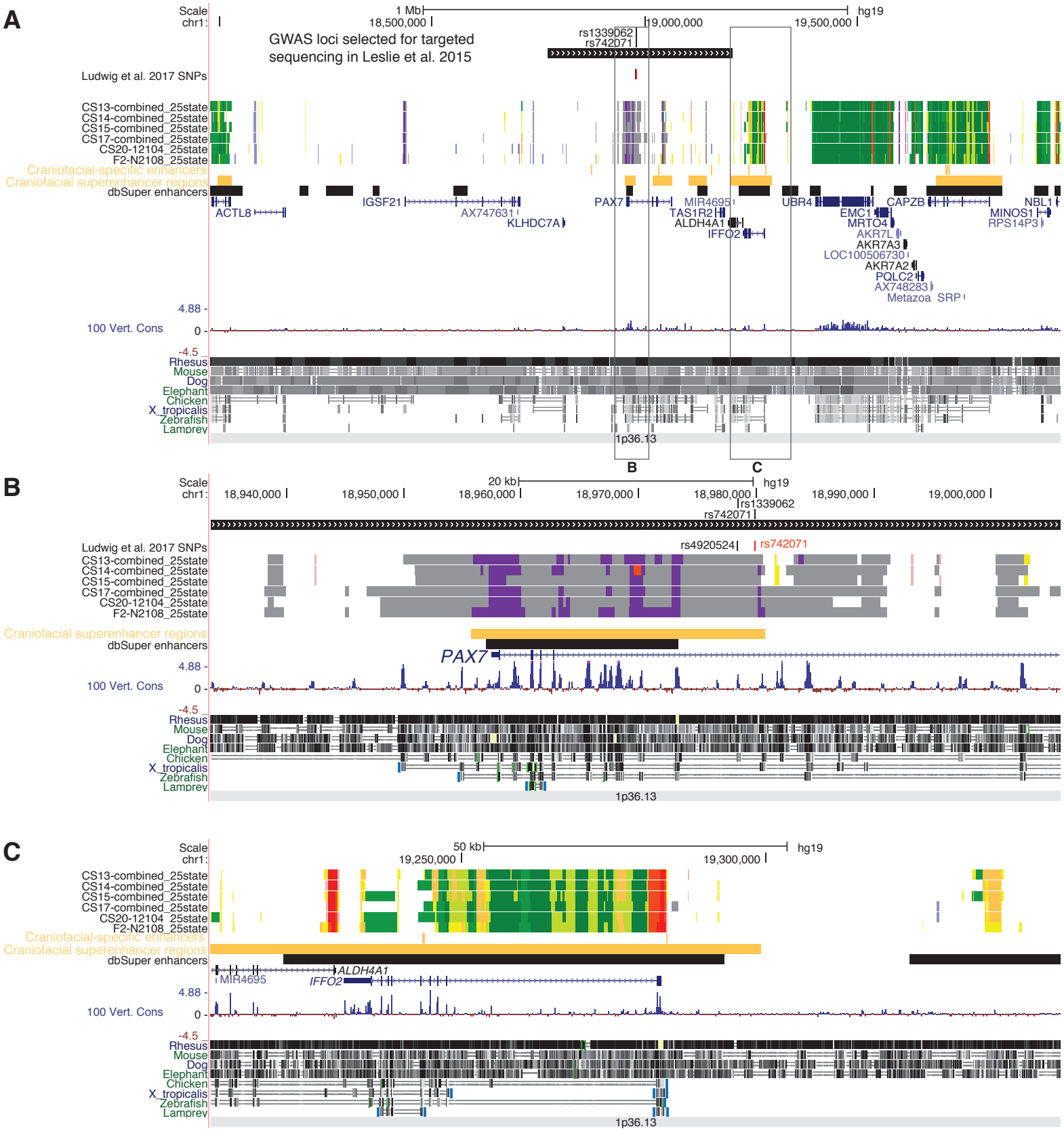

A
Scale
chr1:
1 Mb
18,500,000
19,000,000
19,500,000
hg19
GWAS loci selected for targeted sequencing in Leslie et al. 2015
rs1339062
rs742071
Ludwig et al. 2017 SNPs
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific 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.

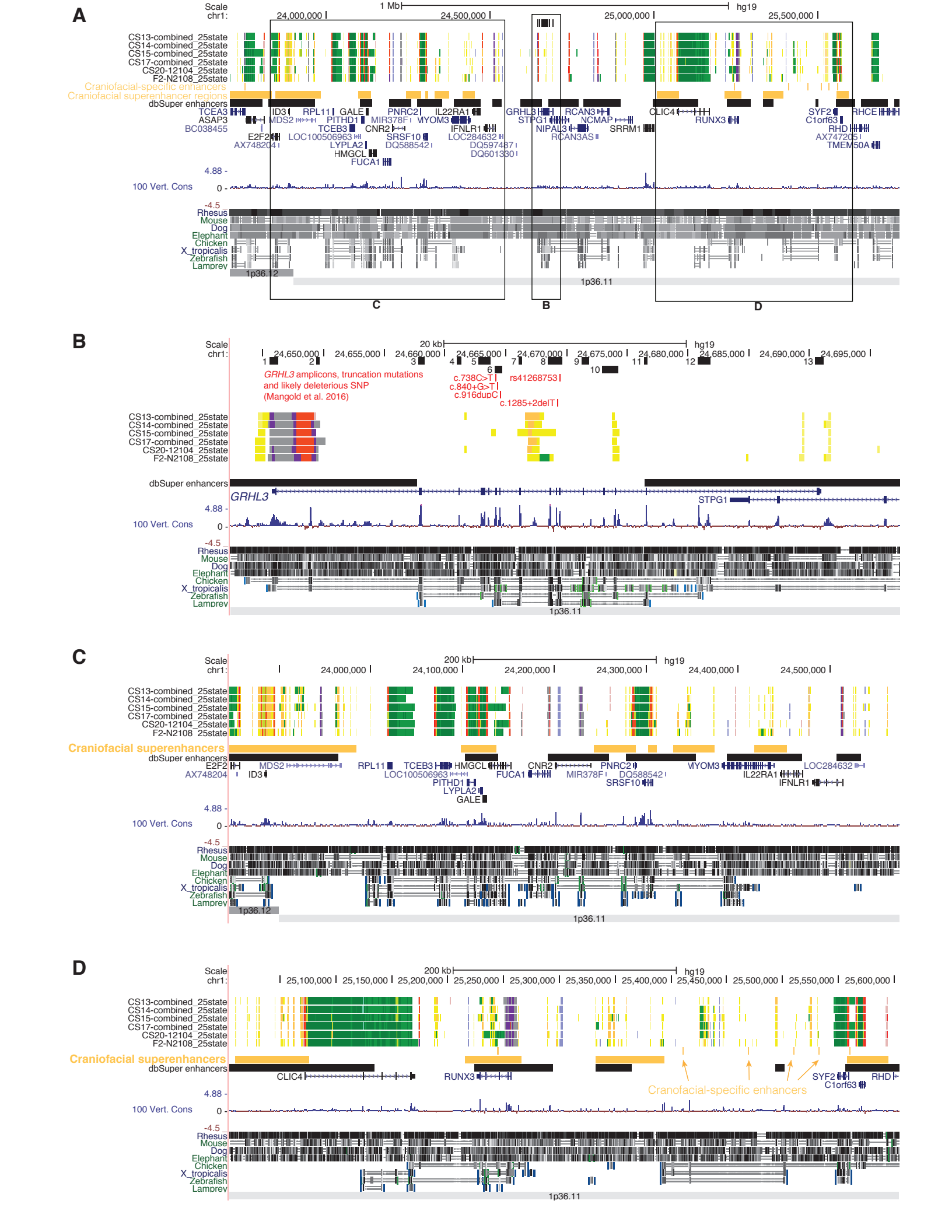

A
Scale
chr1:
1 Mb
hg19
24,000,000
24,500,000
25,000,000
25,500,000
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
Craniofacial-specific enhancers
Craniofacial superenhancer regions
dbSuper enhancers
TCEA3
ASAP3
BC038455
E2F2
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.

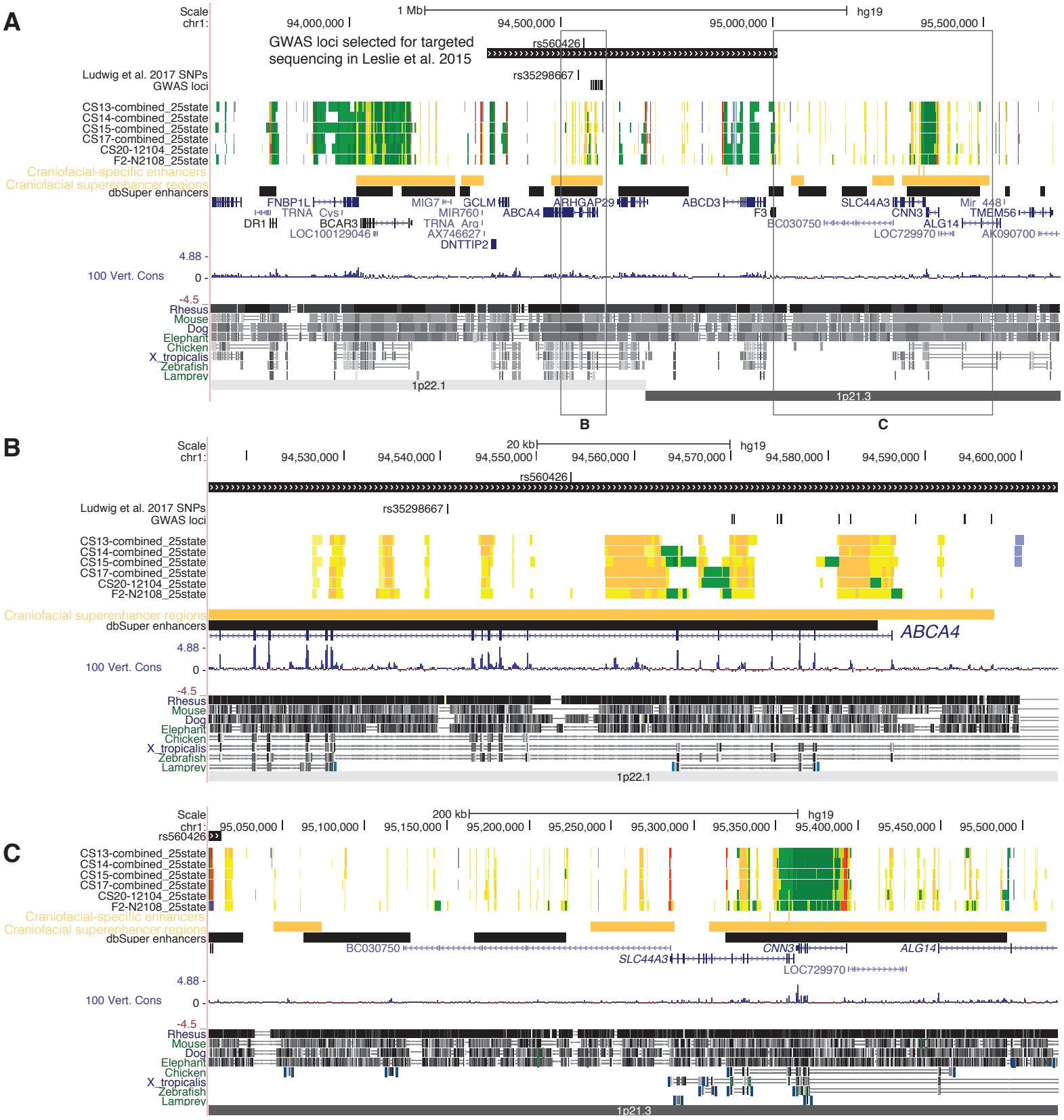

A
Scale
chr1:
1 Mb
hg19
94,000,000
94,500,000
95,000,000
95,500,000
GWAS loci selected for targeted sequencing in Leslie et al. 2015
rs560426
rs35298667
Ludwig et al. 2017 SNPs
GWAS loci
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
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.

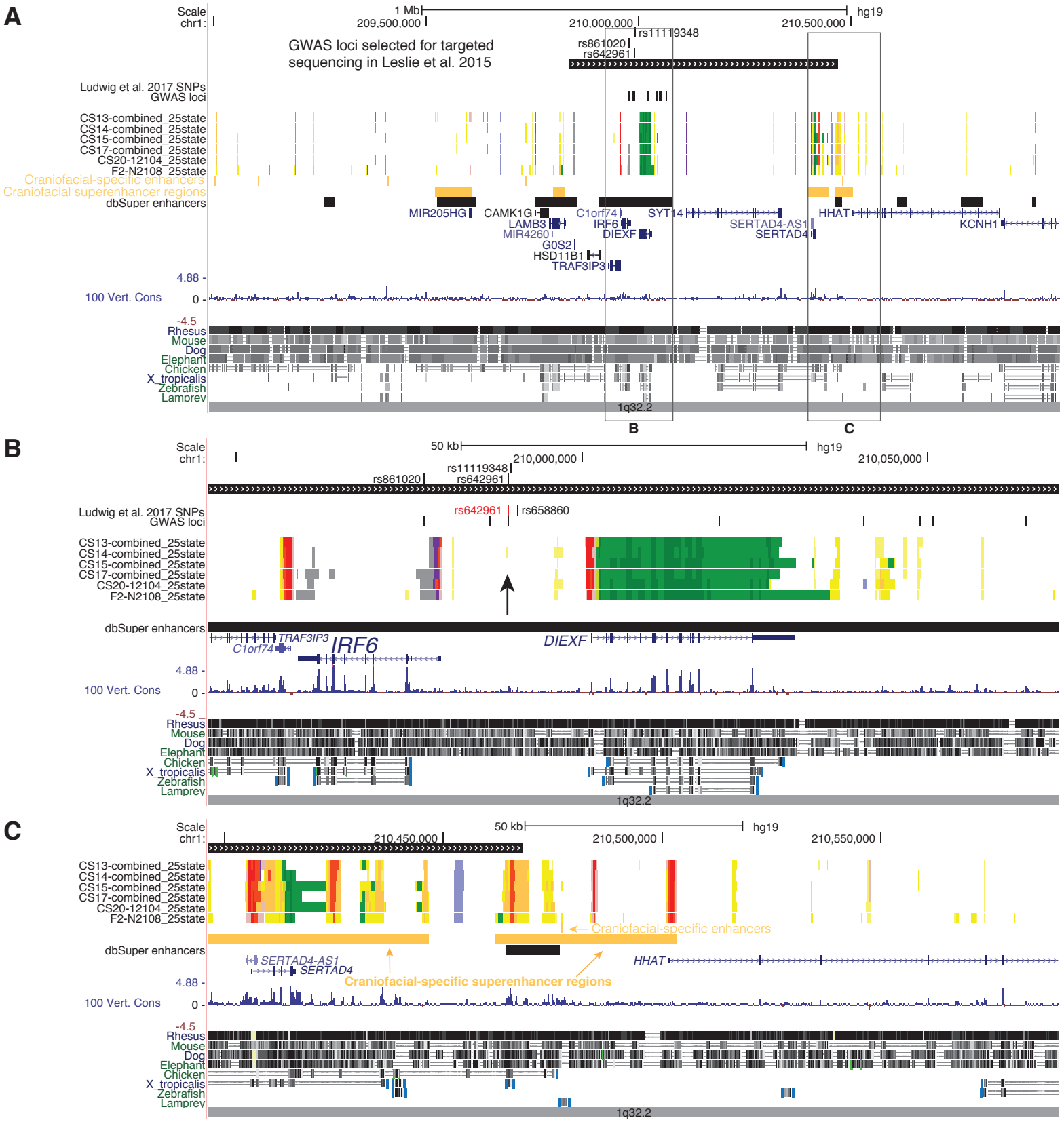

A
Scale
chr1:
1 Mb
209,500,000
210,000,000
210,500,000
hg19
GWAS loci selected for targeted sequencing in Leslie et al. 2015
rs11119348
rs861020
rs642961
Ludwig et al. 2017 SNPs
GWAS loci
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
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.

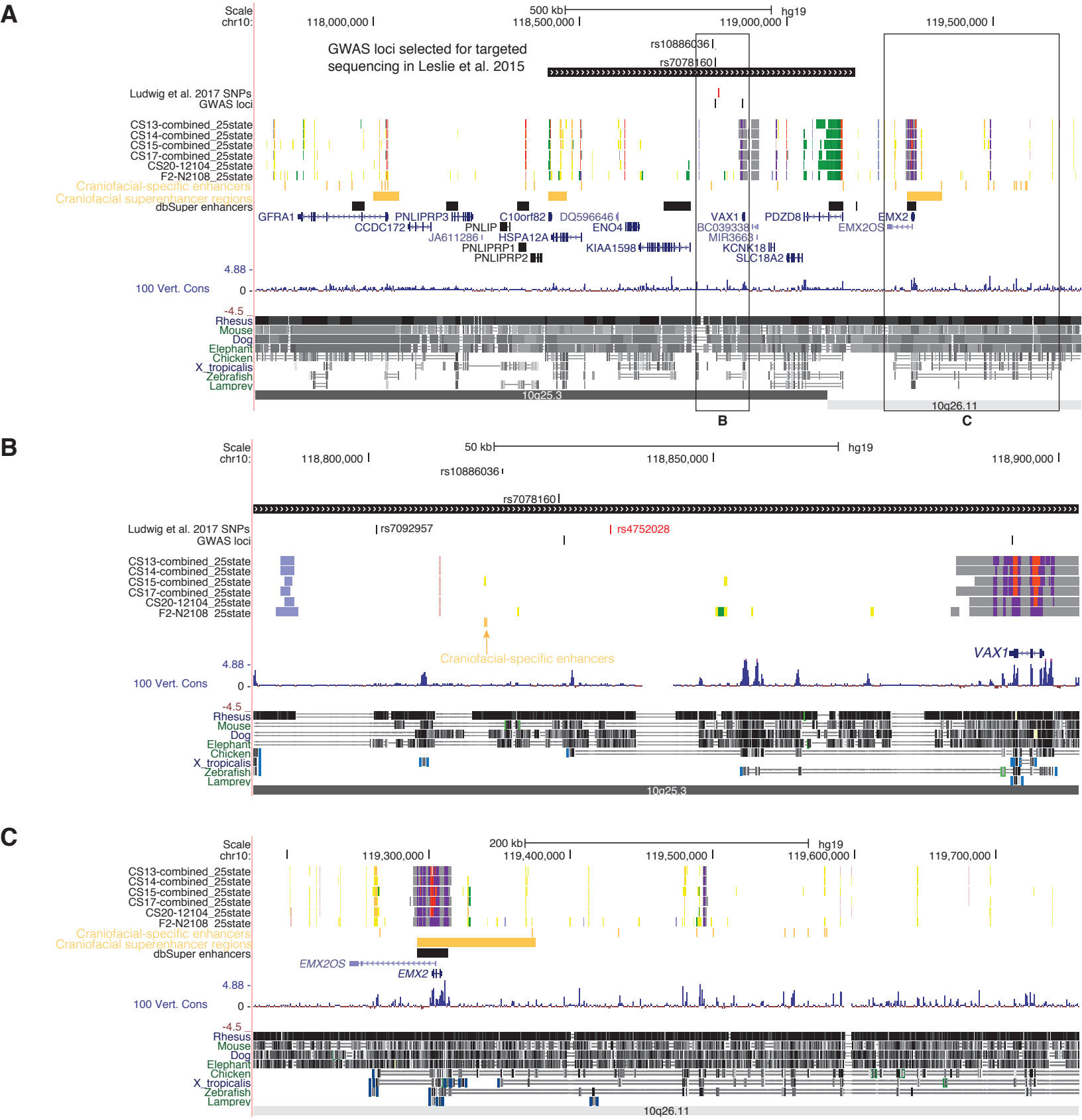

A
Scale
chr10:
500 kb
hg19
118,000,000
118,500,000
119,000,000
119,500,000
GWAS loci selected for targeted sequencing in Leslie et al. 2015
rs10886036
rs7078160
Ludwig et al. 2017 SNPs
GWAS loci
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
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.

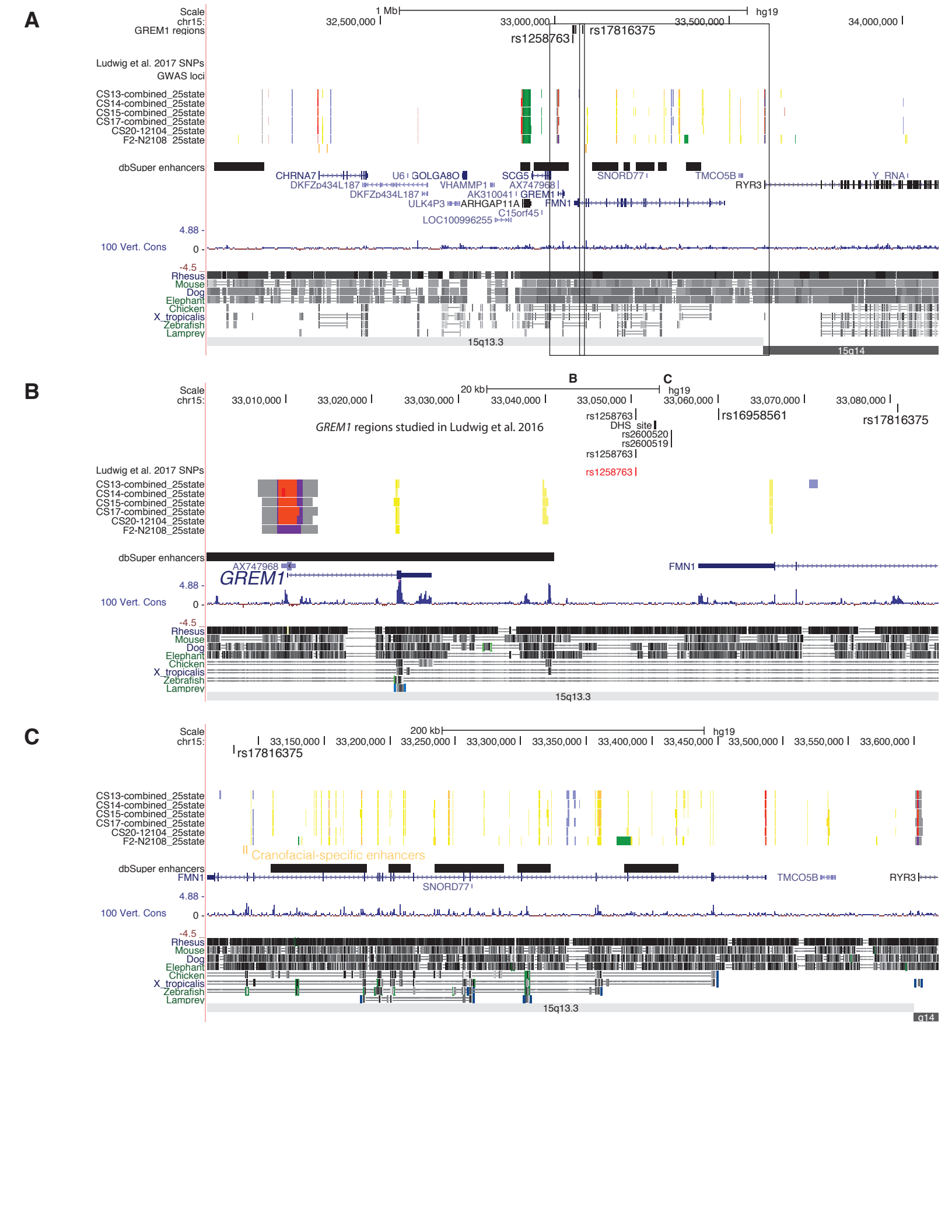

A
Scale
chr15:
GREM1 regions
1 Mb
hg19
32,500,000
33,000,000
33,500,000
34,000,000
rs1258763
rs17816375
Ludwig et al. 2017 SNPs
GWAS loci
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
F2-N2108_25state
dbSuper enhancers
CHRNA7
U6
GOLGA8O
SCG5
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.

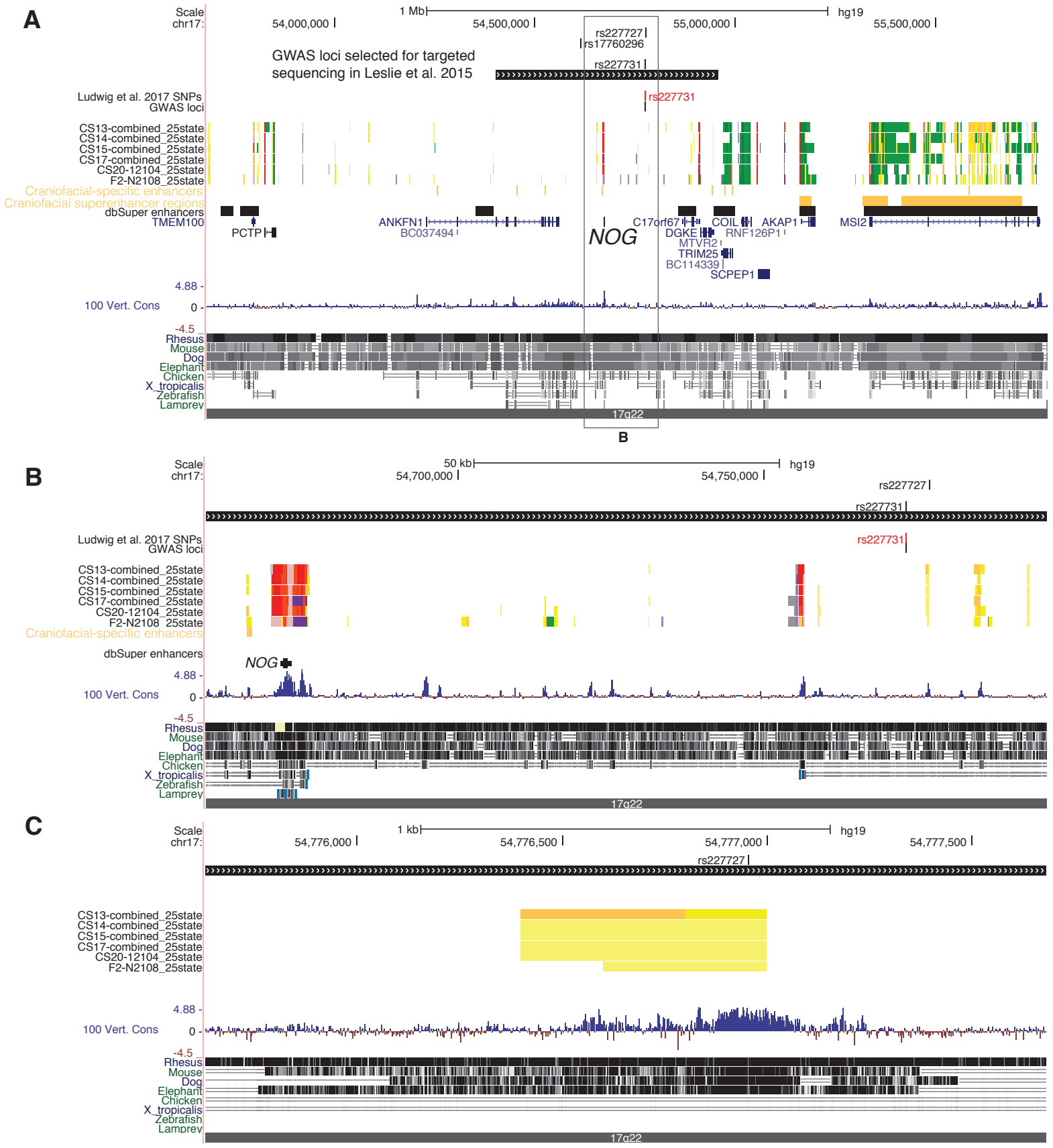

A
Scale
chr17:
1 Mb
hg19
54,000,000
54,500,000
55,000,000
55,500,000
rs227727
rs17760296
GWAS loci selected for targeted
sequencing in Leslie et al. 2015
rs227731
Ludwig et al. 2017 SNPs
GWAS loci
rs227731
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
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.

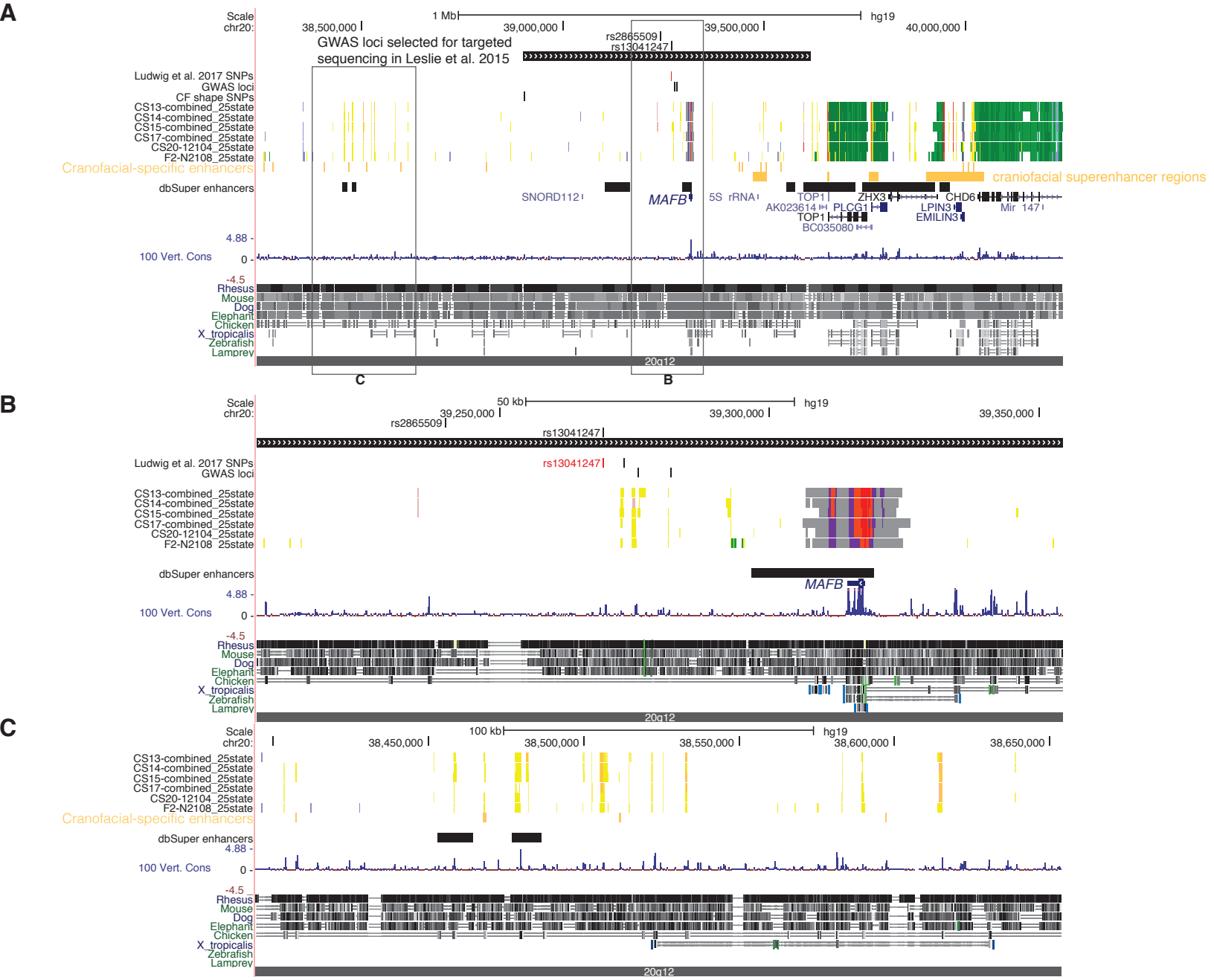

A
Scale
chr20:
1 Mb
hg19
38,500,000
39,000,000
39,500,000
40,000,000
GWAS loci selected for targeted sequencing in Leslie et al. 2015
rs2865509
rs13041247
Ludwig et al. 2017 SNPs
GWAS loci
CF shape SNPs
CS13-combined_25state
CS14-combined_25state
CS15-combined_25state
CS17-combined_25state
CS20-12104_25state
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.