High Resolution Epigenomic Atlas of Early Human Craniofacial Development

Andrea Wilderman[1,2], Jeffrey Kron[2], Jennifer VanOudenhove[2], James P. Noonan[3,4], and Justin Cotney[1,2,5,*]

[1]Graduate Program in Genetics and Developmental Biology, UConn Health; [2]Department of Genetics and Genome Sciences, UConn Health; [3]Department of Genetics, Yale University School of Medicine, [4]Kavli Institute for Neuroscience, Yale University; [5]Institute for Systems Genomics, University of Connecticut

*correspondence should be sent to cotney@uchc.edu

Abstract

Defects in embryonic patterning resulting in craniofacial abnormalities are common birth defects affecting up to 1 in 500 live births worldwide, and are mostly non-syndromic. The regulatory programs that build and shape the craniofacial complex are thought to be controlled by information encoded in the genome between genes and within intronic sequences. Early stages of human craniofacial development have not been interrogated with modern functional genomics techniques, preventing systematic analysis of genetic associations with craniofacial-specific regulatory sequences. Here we describe a comprehensive resource of craniofacial epigenomic annotations and systematic, integrative analysis with a variety of human tissues and cell types. We identified thousands of novel craniofacial enhancers and provide easily accessible genome annotations for craniofacial researchers and clinicians. We demonstrate the utility of our data to find likely causal variants for craniofacial abnormalities and identify a large enhancer cluster that interacts with *HOXA* genes during craniofacial development.

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 ten weeks of gestation[1]. Defects in the orchestration of these events result in several different congenital abnormalities including failure of features to fuse (orofacial clefting) and premature fusion of structures (craniosynostosis). Worldwide, orofacial clefting is one of the most common birth defects, affecting ~1 in 700 live births[2]. The majority of those affected with these types of clefting do not have defects in other tissues or organ systems and thus are referred to as “non-syndromic”[3]. 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[4-7]. Screening, prevention, and non-surgical therapeutic options are thus highly desirable. The high heritability of such disorders suggests a major genetic component[8,9]; however, causative genetic changes have only been identified in a fraction of those affected[10]. Candidate gene approaches have identified mutations in seven different genes that explain less than ten percent of non-syndromic orofacial clefting cases[11]. In the past decade, several genome wide association studies, copy number variant analyses, and whole exome sequencing studies have sought to identify additional genetic sources of non-syndromic orofacial clefting[11-21]. These studies identified common and rare variants associated with orofacial clefting, but most are located in non-coding portions of the genome. Our genomes are littered with gene regulatory sequences, located primarily in intronic and intergenic sequences, that are active in a small number of tissues and/or developmental stages in humans[22]. 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[23-26]. These findings, coupled with the non-syndromic nature of most orofacial clefting cases, suggest defective gene regulatory sequences may underlie much of the incidence of orofacial clefting. However, mapping of chromatin states and identification of craniofacial-specific regulatory sequences has been ignored by large functional genomics efforts such as ENCODE and Roadmap

Epigenome[22]. The lack of craniofacial-specific gene regulatory information has impeded the identification of regulatory circuitry important for human craniofacial development and has prevented accurate interpretation of clinical genetic findings in patients with craniofacial disorders. Lastly, without sufficient biological context, prioritization and developing of hypotheses to test genetic associations with craniofacial abnormalities are hindered[27-30]. Here we present a comprehensive resource of functional genomics data and predicted chromatin states for important stages of early human craniofacial development. We have profiled multiple biochemical marks of chromatin activity in developing human craniofacial tissue samples encompassing 4.5 to 8 post conception weeks. We have comprehensively compared these data with publicly available genomic and genetic data from 127 epigenomes which include a wide variety of adult and fetal tissues. We provide annotations consistent with large consortia efforts[22] in formats easily loadable into modern genome browsers to enable exploration by other researchers without large computational effort. We demonstrate how to mine this data for biological features relevant to craniofacial development and how to experimentally validate target gene interactions. In total, our analyses have identified thousands of previously unknown craniofacial enhancer sequences. These analyses will facilitate interpretation of genetic variation in the context of congenital craniofacial defects, and will enable future experimental testing of enhancer-target gene interactions in developing craniofacial tissues.

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[22]. Many of the regulatory elements identified by this method are specific to the biological context queried[31,32] (i.e. tissue type or developmental stage) and are enriched for genetic associations with disease in a relevant tissue (i.e. immune-related disorder associations in immune cell-specific enhancers)[33,34]. 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 both 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 the critical window for the formation of the human orofacial apparatus (**Fig. 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) (**Fig. 1b**)[35]. We profiled at least three biological replicates for four distinct Carnegie stages (CS) (CS13, CS14, CS15, and CS17) encompassing 4.5 post conception weeks (pcw) to 6 pcw. We also profiled single biological samples from CS20 (8 pcw) and 10 pcw embryos (**Fig. 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 (**Supplementary Table 1**). Overall the samples correlated well by mark and stage of development (**Fig. 2a and Supplementary Fig. 1**). 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 (**Fig. 2b and Supplementary Fig. 2**)[32,35]. For example, we observed very strong enrichment of H3K4me3 at promoters of genes and identified a large number of intronic or intergenic regions enriched for H3K27ac. When we examined all the samples for a given Carnegie stage, we identified thousands of enriched regions, at each stage for each mark, that were found in at least two biological replicates (**Fig. 2c**).Combined, these results indicated our ChIP-Seq data from human embryonic tissues were of high quality, reflected the previously described nature of these marks, and was 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[36-40]. 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 longer active[41-43]. 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[44,45]. These techniques can identify tissue-specific and disease-relevant regulatory information in a large cohort of tissues[35,46]. 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 (Methods)[22]. Using p-value based signals[47,48] for each of the six epigenomic marks we assayed, along with 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[49] (**Fig. 1c**). The imputed samples' signals correlated well with their primary signals and clustered generally by mark and biological function (**Fig. 3a and Supplementary Fig. 3**). Using the imputed craniofacial data, we then segmented the genome for each embryonic sample based on previously generated models of 15, 18, and 25 states of chromatin activity[22]. We identified similar numbers and proportions of segments in each state in our tissues (**Fig. 3b and Supplementary Fig. 4**).The 25-state model results showed the most similar trends across these measures and utilized all of the primary data generated in our study when compared to those previously generated by Roadmap Epigenome (**Fig. 3c,d and Supplementary Fig. 4**); therefore we focused our downstream analyses on these segmentations. Using the 25-state segmentations, we reproducibly identified 75928 segments in at least one of six enhancer categories defined by Roadmap Epigenome (EnhA1, EnhA2, EnhAF, EnhW1, EnhW2, and EnhAc). To determine if these segmentations are enriched for craniofacial enhancers, we first turned to a large catalog of experimentally validated developmental enhancers tested in mouse embryos and available in the Vista Enhancer Browser[50]. We identified over 80% of all craniofacial-positive enhancers in this database. Moreover, our enhancer annotations were significantly enriched for craniofacial enhancers versus those that lacked craniofacial activity ($p = 3.28 \times 10^{-14}$) (**Fig. 4a,b and Supplementary Fig. 5**). While these results are encouraging - namely, that our data identified craniofacial enhancers - they did not reveal any specificity for craniofacial tissues in our chromatin state annotations. To address this problem, we quantitatively compared H3K27ac signals at all enhancer segments in our data with 127 samples 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 (**Fig. 4c and Supplementary Fig. 6**). They were most similar to embryonic stem cells (ESC) and cell types derived from them (ESDR), but distinct from fetal and adult samples present in Roadmap Epigenome data. Previous analyses of Roadmap Epigenome have identified a significant number of enhancers that are tissue-specific[22]. To identify such novel enhancers in craniofacial tissue we first determined if any of our enhancer segments were ever annotated as such in the 127 samples obtained from Roadmap Epigenome. We identified 6651 enhancer segments (8.7% of total craniofacial enhancer segments) in our craniofacial epigenomic atlas that were never annotated as any type of enhancer state in all of Roadmap Epigenome (**Supplementary Table 2**). 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 novel 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[51-57] (**Fig. 4d and Supplementary Table 3**). Utilizing the Genomic Regions Enrichment of Annotations Tool (GREAT)[58], we found significant enrichment of craniofacial-specific enhancers assigned to genes associated with craniofacial abnormalities such as cleft palate in both humans and mice (**Fig. 4e and Supplementary Fig. 7**). Interestingly, we also identified more general categories of enrichment amongst the putative gene targets including general transcriptional activators (**Supplementary Table 4**). When we interrogated this list of transcription factors, we found significant enrichment for expression in both craniofacial and appendicular skeleton (**Fig. 4f**). These results suggest that many of the novel craniofacial enhancers we identified are likely to play a direct role in patterning of the 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 this uncertainty, we turned to genome wide association data obtained from the GWAS catalog related to orofacial clefting and craniofacial morphology[17,21,59-63]. We overlaid associations from these studies with each of the segmentation maps from our data, as well as data from Roadmap Epigenome, and assessed enrichment. We observed significant enrichment of orofacial clefting tag SNPs in most of our craniofacial samples and relatively few Roadmap Epigenomes (**Supplementary Fig. 8a**). These analyses identified several enhancer segments that directly contain strong genetic associations. For instance, 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[64] (**Supplementary Fig. 8b**). This particular region can directly influence *IRF6* expression and is potentially a causative allele for orofacial clefting[65].We also identified 13 other regions that are identified in craniofacial tissue and directly contain such tag SNPs, including an intronic sequence of the *TXNDC16* gene[59] (**Supplementary Fig. 8c and Supplementary Table 5**). These findings suggest that our chromatin state maps will be extremely useful in identifying and prioritizing causative variation in patients affected by craniofacial abnormalities.

Machine learning approaches to mining of activated craniofacial enhancer data.

To more comprehensively explore our data for regions likely to be important for craniofacial development and human disease, we turned to the unsupervised machine learning method known as self-organizing maps. This approach is a powerful means to identify relationships within large genomic datasets, but also allows fine-grained analysis relevant to specific biological questions[66]. We first extracted H3K27ac signals from all of our craniofacial samples and all Roadmap Epigenome samples across all enhancer segmentations, resulting in signal measurements for 425000 enhancer segments in 146 epigenomes.The resulting matrix was used to train a self-organizing toroid map with a map size of 2500 units; we selected the best scoring map from 50 map building trials. We then clustered each of the units of the map into metaclusters and found 199 that identify enhancer segments that have similar signal properties and are likely to be biologically related (**Fig. 5a**). For each enhancer segment, we assigned potential target genes and overlaid the gene assignments to each unit. Based on these gene assignments, we then determined the gene and human phenotype ontology enrichments of each unit. This resource is available for interrogation via a standard web browser, allowing for retrieval of regions, genes, and functional associations for each unit and metacluster. Inspection of this map identified several metaclusters that showed distinct H3K27ac activation in craniofacial tissues. These clusters were enriched for a number of ontologies related to craniofacial biology and abnormalities. For example, we identified a metacluster that showed significantly increased H3K27ac signal in craniofacial samples relative to other tissue types and that is enriched for potential target genes associated with various craniofacial abnormalities (**Fig. 5b**). We obtained similar types of functional enrichments when performing k-means clustering directly on the matrix of H3K27ac signals using the same number of clusters utilized for the self-organizing map (**Supplementary Fig. 9a and Supplementary Table 6**). When we assessed the sequence content of clusters most specific for craniofacial activity, we identified enrichment of motifs for the *ALX*, *DLX*, *HOX*, and *MSX* families of transcription factors (**Supplementary Fig. 9b**).

Identification of novel craniofacial locus control region and potential regulatory targets

Thus far, our analyses have focused on the annotation and activation state of individual genome segments in bulk. 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[67]. To identify such enhancer clusters, we applied 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 582 regions across the genome that demonstrate high levels of craniofacial enhancer activity (**Supplementary Table 7**). These windows had an average size of ~400kb but ranged up to 2 Mb in length. In many cell types these clusters of enhancers, sometimes referred to as super enhancers, are embedded in the genome both surrounding and within the introns of their likely tissue-specific target[68]. Indeed, most of the windows we identified contained multiple genes and were enriched for developmental genes, including multiple *Frizzled*, *WNT*, *ALX*, *DLX*, and *TBX* family members (**Supplementary Table 8**). Interestingly, we identified 37 large windows that were located entirely in intergenic space and did not overlap a promoter region for any known genes. These windows represent potentially novel large clusters of regulatory regions, but their targets and activities are difficult to interpret using linear genomic annotations and distances. Given that studies have shown that our genome can form numerous long range interactions[69,70], especially between regulatory regions, we sought to determine if any of these intergenic clusters of putative enhancers could be important for craniofacial development by identifying direct three-dimensional interactions in relevant tissues. To ensure that we were interrogating bona fide enhancer clusters, we focused our downstream efforts on regions that contained *in vivo*-validated craniofacial enhancers. We identified a single window encompassing a 450kb region located on chromosome 7 that contains five confirmed craniofacial enhancers from the Vista Enhancer Browser[50] (**Fig. 6a**). This region also contains a unique chromatin signature at its 3' end, where strongly active and repressed states are directly adjacent. This is most commonly observed at looping or topological domain boundaries. Indeed, long-range contact maps from human umbilical vein endothelial cells[69] indicate this observed chromatin state transition is a topologically associated domain (TAD) boundary (**Supplementary Fig. 10**). We tested an element annotated as a bivalent chromatin state, which is highly conserved across mammals, near this chromatin boundary for enhancer activity[71]. It displayed strong craniofacial and limb enhancer activity in the E11.5 mouse embryo (**Fig. 6a**). Inspection of chromatin data from Mouse ENCODE[72] indicate similar patterns of activation in the orthologous window, suggesting functional conservation of chromatin state in this large region (**Supplementary Fig. 11**). Having demonstrated that this region is enriched for craniofacial enhancers and active chromatin states, we sought to determine the gene(s) this region interacts with and potentially regulates. This region is located between the *NPVF* and *NFE2L3* genes, neither of which appears to be active based on observed chromatin states in developing human craniofacial tissue. The next closest target is *CBX3*, which is strongly expressed in most cell types and has similar chromatin states in both our tissues and in Roadmap Epigenome. Comparisons of the mouse and human genomes revealed this window is part of a large syntenic block between the two species which stretches nearly 10 Mb in length with the *HOXA* gene cluster at its center (**Supplementary Fig. 12**). The enrichment for craniofacial enhancer annotations, harboring of six *in vivo* validated craniofacial enhancers, potential TAD boundary, and conservation both at the sequence and epigenomic level suggest this region is an important regulatory hub.

Two control regions, the early limb control region (ELCR) and the global control region (GCR), have been identified for the *HOXD* gene cluster that are important for regulation of the cluster's expression in the developing mammalian limb. The exact coordinates of the ELCR are unknown, but they are thought to be located in the large noncoding region adjacent to the cluster, while the GCR is approximately 250kb away, beyond the *LNP* gene[73,74]. No such control regions have been identified or described for the *HOXA* cluster. Furthermore, loss of at least one gene in the cluster, *HOXA2,* has been implicated in cranial neural crest skeletal morphogenesis and results in mice born with cleft palates and other craniofacial abnormalities[57,75]. The region we have identified is located nearly 1.5 Mb from the *HOXA* cluster and contains at least seven annotated genes in the intervening genomic sequence; thus it is not clear whether this region could regulate the *HOXA* gene cluster. Utilizing circularized chromosome conformation capture with sequencing[76] (4C-seq) we assessed the interactions of four viewpoints in this window in E11.5 mouse craniofacial tissue. For two viewpoints, we identified extensive interactions within the identified window that do not cross the putative TAD boundary. When we assessed viewpoints flanking the TAD boundary, one of which contained the active enhancer HACNS50[71], we observed interactions within this identified region as well as significant interactions with the *HOXA* gene cluster (**Fig. 6b**). To confirm these interactions, we performed additional 4C-seq experiments utilizing viewpoints located directly within the *HOXA* cluster and the promoter of the *SKAP2* gene. We observed strong interactions between both of these viewpoints and the TAD boundary of the original window. Interestingly, *HOXA* made contacts with the outer limits of this window but not within the window. These findings illustrate that the region we identified in human craniofacial tissue makes strong contacts over nearly 1.5 Mb with genes of the *HOXA* cluster in developing mouse craniofacial tissue and indicate it could be a conserved global control region important for craniofacial development.

Discussion

Our understanding of the regulation of craniofacial development and the genetic changes that give rise to developmental defects has not advanced greatly in the last decade despite it being a heavily studied area of human and mouse biology and the advent of more advanced genomic technologies. Recent large consortia efforts to identify the genetics of common disease have gained traction utilizing tissue-specific annotations of the genome to identify potential regulatory regions and overlaying genetic associations[33,34]. Such genetic association data exist for craniofacial abnormalities, but the lack of craniofacial-specific annotations of regulatory function have 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. Over 6000 of the enhancer segments we identified have never been annotated previously as having enhancer activity in 127 different cell types. These regions are strongly enriched near genes implicated in craniofacial development and 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 excluded important craniofacial regulatory regions due to the lack of appropriate chromatin state annotations[77] (**Supplementary Fig. 13**). These findings illuminate that our current understanding of the regulatory information our genomes encode is incomplete and reinforces the need for more and higher resolution tissue-specific chromatin state annotations.

To illustrate the utilization of this resource we analyzed these data in many different fashions to narrow down particular regions of interest and to interrogate them for genetic and functional associations with craniofacial development. Furthermore, we demonstrated that once a region of interest has been identified, it is possible to develop a hypothesis of potential gene regulatory targets and directly test them *in vivo* in the context of both genomic and functional conservation in the mouse. Here, we chose to focus on clusters of craniofacial enhancer segments that have been functionally verified in the developing mouse embryo to ensure relevance for craniofacial biology. Our windowing approach identified an extremely dense, large array of craniofacial enhancers, suggesting we have identified an important regulatory hub. The conservation of activating histone modification signals in developing mouse craniofacial tissues indicates this region is likely important for the formation of craniofacial features in multiple species. Additionally, the identification of direct long-range interactions between portions of this unique enhancer region, including a rapidly evolving conserved non-coding sequence (HACNS50)[71], with the *HOXA* gene cluster suggest this region could be important not only for normal craniofacial development but also for evolution of the human skull. Lastly, this region has been implicated as a uniquely deleted segment in a patient with facial dysgenesis[78] (**Supplementary Figure 10**). This patient was noted to have overtly normal organs and brain activity despite lacking most features of a face resembling other non-syndromic abnormalities caused by regulatory sequence defects[23-26]. Further genetic dissection of this region in cultured human cells or in the developing mouse are needed to determine the role this region plays in regulating this conserved cluster of *HOXA* genes .

We provide all our craniofacial functional genomics data and resulting chromatin state segmentations in several standard formats as well as a complete catalog of tracks that can be easily loaded into many modern genome browsers. Additionally, we provide our self organizing map of active enhancers across 146 samples as a website that can be explored by a variety of researchers to interrogate the regions, genes, and phenotypes relevant for their research without high level computational processing or expertise. (https://cotneylab.cam.uchc.edu/~jcotney/CRANIOFACIAL_HUB/Craniofacial_H3K27ac_SOM/). This will allow the craniofacial community to develop hypotheses related to craniofacial abnormalities which are rooted in craniofacial biology instead of using chromatin state annotations from other tissues not directly related to the tissue of interest. These resources stand to bring the craniofacial research world firmly into the functional genomics era, advance our understanding of these disorders, and provide tools for clinicians seeking to diagnose patients utilizing whole genome sequencing.

Acknowledgements

We would like to thank the donors to HDBR, as without them this resource would not be possible. We would also like to thank Camden Jansen and Ali Mortazavi for assistance in generation of the self-organizing map resource and Heather Adinolfi for technical assistance in the transgenic enhancer assay. Funding support was provided by an R00 award from NIDCR to J.C.

Author Contributions

All ChIP experiments were performed by J.C. Sequencing libraries and sequencing runs were performed by J.K. Data analysis was performed by J.C. 4C-Seq experiments and analysis were performed by A.W. Writing and data interpretation were performed by all authors.

Data and Code Availability

All data can be visualized in the UCSC Genome Browser using 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 and 4C-Seq data are available at GEO accessions GSE98251 and GSE97752.

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

All generic scripts for processing of 4C-Seq data from mouse are available on github: https://github.com/cotneylab/Mouse-HOXA-4C-Seq

Methods

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[79]. 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).

Antibody Specifications

Antibodies used in this study: anti-H3K27ac (ab4729, Abcam), anti-H3K4me1 (ab8895, Abcam), anti-H3K4me2 (ab7777, Abcam), anti-H3K4me3 (ab8580, Abcam), anti-H3K27me3 (07-449, EMD Millipore), anti-H3K36me3 (ab9050, Abcam).

ChIP-Seq

Fixed tissue pellets were processed for ChIP as previously described[79]. 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 listed above were added to each chromatin aliquot. 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.

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)[80] and compared visually using MultiQC (v0.9)[81]. Reads were then aligned to the human genome (hg19) using Bowtie2 (v2.2.5)[82] keeping only uniquely mapped reads. Fragment sizes of each library were estimated using PhantomPeakQualTools (v.1.14)[47]. Histone modification enriched regions were identified and annotated using HOMER (v4.8.3)[83]. 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)[84]. 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)[48]. All signal and enriched region files were converted for display in the UCSC Genome Browser using the Kent Source Tools (v329)[85]. 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)[86].

Roadmap Epigenome 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[49] using Kent Source Tools [85]. 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/).

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[49]. 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)[44] using deepTools2[86].

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[22]. 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[86]. All plots were made using tabular data generated by deepTools2 in R (v3.3.3)[87].

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[50]. 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[84]. These novel craniofacial enhancer segmentations were assessed for gene ontology and functional enrichments based on assigned target genes using GREAT (v3.0.0)[58]. Genes identified as transcriptional regulators by GREAT were assessed for enrichment of anatomical expression using default parameters in GeneORGANizer[88]. Sequence from novel craniofacial enhancer segmentations was extracted from hg19 using fastaFromBed within BEDTools[84]. The resulting sequences were assessed for transcription factor motif enrichment using HOMER[83]. 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[17,21,59-63] obtained from the GWAS Catalog (retrieved 2017-02-20)[89] using Fisher's exact test within BEDTools[84].

Self-Organizing Maps of Enhancer Activation

The self-organizing map of H3K27ac signal at all enhancer segments was generated as previously described[66]. Briefly, a union of all enhancer segmentations from craniofacial tissues and all samples in Roadmap Epigenome was generated and merged to form a consistent annotation of enhancers across the entire genome resulting in 425380 individual enhancer segments. H3K27ac signals from imputed p-value signal tracks for each of the 146 epigenomes were extracted for each of the 425380 enhancer segments. This matrix was then used to train a self-organizing map with 50 rows and 50 columns (2500 units) to allow for the possibility of small numbers of highly tissue-specific enhancers (<200) to be clustered together. We performed 50 training trials and retained the best scoring map. For this final self-organizing map we then annotated each unit with Ensembl (v75) genes based on association rules defined by GREAT[58]. Based on these unit/gene assignments we then determined enrichment of gene ontologies (http://geneontology.org/ontology/go.obo) and human phenotype ontologies from the Monarch Initiative[90] (http://purl.obolibrary.org/obo/hp.obo) as previously described[91]. Clusters of units, or metaclusters, were then determined with four separate trials testing for the presence of up to 250 metaclusters as previously described[66]. The algorithm converged on 199 clusters as optimal for the self-organizing map generated above. Metaclusters were then assessed for functional enrichments as was done for individual units above. Metaclusters identified as specific for craniofacial and brain tissues were visualized using a JavaScript web-based viewer of the self-organizing map available here: https://cotneylab.cam.uchc.edu/~jcotney/CRANIOFACIAL_HUB/Craniofacial_H3K27ac_SOM/

K-means clustering of Enhancer Activation

K-means clustering of the same H3K27ac signal matrix utilized for the self-organizing map was performed using Cluster (v3.0)[92]. Rows were centered on the mean value of the row and normalized, the number of metaclusters identified in the self-organizing map analysis above was used as the k parameter, and 100 runs were performed. The clustering result was then visualized and craniofacial-specific clusters were extracted using Java TreeView[93]. Sequences underlying the enhancers in the craniofacial-specific clusters were extracted as above for novel craniofacial enhancers. We performed motif enrichment within these sequences using a combination of multiple tools for more robust enrichment determination[94]. Functional enrichment for these enhancers was determined as above using GREAT[58].

Identification of Enhancer Clusters

To identify clusters of craniofacial enhancers we first generated overlapping 200kb windows with a 50kb step size[84]. 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[84] and determining the fraction of each window annotated as an enhancer. Overlapping windows of significant enrichment were merged into a single contiguous region. Final enriched regions were assessed for overlap with gene annotations and validated craniofacial enhancers using BEDtools[84].

Transgenic Enhancer Assay

A 2.6 kb segment centered on the conserved sequence corresponding to HACNS50[71] was amplified from human genomic DNA by polymerase chain reaction (PCR) using the following primers: HACNS50 F 5'-CACCCCATTTCTGAGGGGAAATAA-3', HACNS50 R 5'-TTATTTCCTTCAGGCCCTTG-3', and cloned into an Hsp68-lacZ reporter vector as previously described[95]. Generation of transgenic mice at the Yale University Transgenic Mouse Facility and embryo staining were carried out as previously described[95]. We required reporter gene expression in a given structure to be present in at least three independent transgenic embryos as assessed by two researchers to be considered reproducible.

Circularized Chromosome Conformation Capture with Sequencing (4C-Seq)

All animal work was done in accordance with approved University of Connecticut Health Center IACUC protocols. 4C-seq was performed according to van de Werken et al. (2012)[76] with modifications for tissue. Input mouse embryonic craniofacial and brain tissue from the same litter was fixed and nuclei isolated following homogenization with a dounce tissue grinder as described[79]. Each replicate consists of tissue from an individual litter. Subsequent digestion and ligation steps were followed from van de Werken et al. (2012)[76]. Chromatin was digested sequentially with NlaIII and DpnII. Amplification of final libraries was performed with primers selected using a primer database generated for NlaIII/DpnII digestion as previously described[76]. The sequences added to these primers were modified to allow hybridization to NextSeq 500 flow cells and split across two sets of primers to improve efficiency and allow for dense multiplexing (Table S9).

4C-seq Data Analysis

4C-seq libraries were sequenced for 75 cycles using the NextSeq500 (Illumina). Fastq files were demultiplexed by barcode yielding Fastq files for each tissue replicate. Tissue replicate Fastq files were further demultiplexed by viewpoint using Cutadapt (v1.8.3)[96]. Trimmed reads were uniquely aligned to mm9 using bowtie2[82]. Significant interactions in craniofacial tissue were assessed using r3Cseq[97] with a modification allowing a larger viewing window near the viewpoint (https://github.com/cotneylab/r3Cseq) and using brain as a control. The significant interactions are represented in the accompanying track hub as bigBed files. The location of the viewpoint and sequenced interacting fragment are denoted with thick bars. A thin bar is included to denote the connection between the viewpoint and the distal sites.

References

1. Schoenwolf, G.C., Bleyl, S.B., Brauer, P.R. & Francis-West, P.H. *Larsen's Human Embryology*, 687 (Churchill Livingstone/Elsevier, Ann Arbor, MI, 2009).
2. World Health Organization. *World Atlas of Birth Defects*, 237 (World Health Organization, Geneva, Switzerland, 2003).
3. Mossey, P.A. & Modell, B. Epidemiology of oral clefts 2012: an international perspective. *Frontiers of oral biology* **16**, 1-18 (2012).
4. Wehby, G.L., Pedersen, D.A., Murray, J.C. & Christensen, K. The effects of oral clefts on hospital use throughout the lifespan. *BMC health services research* **12**, 58 (2012).
5. Wehby, G.L. *et al.* The effect of systematic pediatric care on neonatal mortality and hospitalizations of infants born with oral clefts. *BMC pediatrics* **11**, 1307-1321 (2011).
6. Boulet, S.L., Grosse, S.D., Honein, M.A. & Correa-Villaseñor, A. Children with orofacial clefts: health-care use and costs among a privately insured population. *Public health reports (Washington, D.C. : 1974)* **124**, 447-453 (2009).
7. Wehby, G.L. & Cassell, C.H. The impact of orofacial clefts on quality of life and healthcare use and costs. *Oral diseases* **16**, 3-10 (2010).
8. Grosen, D. *et al.* Risk of oral clefts in twins. *Epidemiology (Cambridge, Mass.)* **22**, 313-319 (2011).
9. Grosen, D. *et al.* 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. *Journal of Medical Genetics* **47**, 162-168 (2010).
10. Beaty, T.H., Marazita, M.L. & Leslie, E.J. Genetic factors influencing risk to orofacial clefts: today’s challenges and tomorrow’s opportunities. *F1000Research* **5**, 2800-10 (2016).
11. Camargo, M. *et al.* GWAS reveals new recessive loci associated with non-syndromic facial clefting. *European journal of medical genetics* **55**, 510-514 (2012).
12. Ludwig, K.U. *et al.* 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. *Human molecular genetics*, ddx012 (2017).
13. Ludwig, K.U. *et al.* 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 genetics* **12**, e1005914 (2016).
14. Lidral, A.C. *et al.* 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. *Human molecular genetics* **24**, 3895-3907 (2015).
15. Conte, F. *et al.* Systematic analysis of copy number variants of a large cohort of orofacial cleft patients identifies candidate genes for orofacial clefts. *Human genetics*, 1-19 (2015).
16. Bureau, A. *et al.* Whole Exome Sequencing of Distant Relatives in Multiplex Families Implicates Rare Variants in Candidate Genes for Oral Clefts. *Genetics* **197**, 1039-1044 (2014).
17. Ludwig, K.U. *et al.* Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nature genetics* **44**, 968-971 (2012).
18. Yuan, Q., Blanton, S.H. & Hecht, J.T. Association of ABCA4 and MAFB with non-syndromic cleft lip with or without cleft palate. *American journal of medical genetics Part A* **155**, 1469-1471 (2011).
19. Letra, A. *et al.* Novel cleft susceptibility genes in chromosome 6q. *Journal of dental research* **89**, 927-932 (2010).

20. Beaty, T.H. *et al.* A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nature genetics* **42**, 525-529 (2010).
21. Mangold, E. *et al.* Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nature genetics* **42**, 24-26 (2010).
22. Consortium, R.E. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
23. Petit, F. *et al.* The disruption of a novel limb cis-regulatory element of SHH is associated with autosomal dominant preaxial polydactyly-hypertrichosis. **24**, 37-43 (2015).
24. Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development (Cambridge, England)* **132**, 797-803 (2005).
25. Lettice, L.A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human molecular genetics* **12**, 1725-1735 (2003).
26. Weedon, M.N. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nature genetics* **46**, 61-64 (2014).
27. Leslie, E.J. & Marazita, M.L. Genetics of Orofacial Cleft Birth Defects. *Current Genetic Medicine Reports*, 1-9 (2015).
28. Khandelwal, K.D., Van Bokhoven, H., Roscioli, T., Carels, C.E.L. & Zhou, H. Genomic approaches for studying craniofacial disorders. *American journal of medical genetics. Part C, Seminars in medical genetics* **163C**, 218-231 (2013).
29. Rahimov, F., Jugessur, A. & Murray, J.C. Genetics of nonsyndromic orofacial clefts. *The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association* **49**, 73-91 (2012).
30. Dixon, M.J., Marazita, M.L., Beaty, T.H. & Murray, J.C. Cleft lip and palate: understanding genetic and environmental influences. *Nature reviews Genetics* **12**, 167-178 (2011).
31. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-858 (2009).
32. Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642-654 (2013).
33. Farh, K.K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-343 (2015).
34. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nature genetics* **46**, 136-143 (2014).
35. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49 (2011).
36. Reilly, S.K. *et al.* Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science (New York, NY)* **347**, 1155-1159 (2015).
37. Cotney, J. *et al.* The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* **154**, 185-196 (2013).
38. Villar, D. *et al.* Enhancer Evolution across 20 Mammalian Species. *Cell* **160**, 554-566 (2015).
39. Nord, A.S. *et al.* Rapid and Pervasive Changes in Genome-wide Enhancer Usage during Mammalian Development. *Cell* **155**, 1521-1531 (2013).
40. Barozzi, I. *et al.* Genome-wide compendium and functional assessment of in vivo heart enhancers. *Nature Communications* **7**, 1-13 (2016).

41. Kumar, V. *et al.* Comprehensive benchmarking reveals H2BK20 acetylation as a distinctive signature of cell-state-specific enhancers and promoters. *Genome research* **26**, 612-623 (2016).
42. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature genetics* **44**, 148-156 (2012).
43. Cotney, J.L. *et al.* Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome research* **22**, 1069-1080 (2012).
44. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods* **9**, 215-216 (2012).
45. Hoffman, M.M. *et al.* Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nature Methods* **9**, 473-476 (2012).
46. Hoffman, M.M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Research* **41**, 827-841 (2013).
47. Landt, S.G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome research* **22**, 1813-1831 (2012).
48. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X.S. Identifying ChIP-seq enrichment using MACS. *Nature protocols* **7**, 1728-1740 (2012).
49. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature Biotechnology* **33**, 364-376 (2015).
50. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L.A. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Research* **35**, D88-92 (2007).
51. Marchegiani, S. *et al.* Recurrent Mutations in the Basic Domain of TWIST2 Cause Ablepharon Macrostomia and Barber-Say Syndromes. *The American Journal of Human Genetics* **97**, 99-110 (2015).
52. Sharma, V.P. *et al.* Mutations in TCF12, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. *Nature genetics* **45**, 304-307 (2013).
53. Chen, H. *et al.* Multiple calvarial defects in lmx1b mutant mice. *Developmental genetics* **22**, 314-320 (1998).
54. Laclef, C., Souil, E., Demignon, J. & Maire, P. Thymus, kidney and craniofacial abnormalities in Six1 deficient mice. *Mechanisms of development* **120**, 669-679 (2003).
55. Brunskill, E.W. *et al.* A gene expression atlas of early craniofacial development. *Developmental Biology* **391**, 133-146 (2014).
56. Zhao, Y. *et al.* Isolated cleft palate in mice with a targeted mutation of the LIM homeobox gene lhx8. *Proceedings of the National Academy of Sciences* **96**, 15002-15006 (1999).
57. Gendron-Maguire, M., Mallo, M., Zhang, M. & Gridley, T. Hoxa-2 mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. *Cell* **75**, 1317-1331 (1993).
58. Mclean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology* **28**, 495-501 (2010).
59. Shi, M. *et al.* 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 (2012).
60. Birnbaum, S. *et al.* Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nature genetics* **41**, 473-477 (2009).
61. Beaty, T.H. *et al.* Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. *Genetic epidemiology* **35**, 469-478 (2011).
62. Grant, S.F.A. *et al.* A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *The Journal of pediatrics* **155**, 909-913 (2009).

63. Shaffer, J.R. *et al.* Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. *PLoS genetics* **12**, e1006149-21 (2016).
64. Zucchero, T.M. *et al.* Interferon Regulatory Factor 6 (IRF6) Gene Variants and the Risk of Isolated Cleft Lip or Palate. *dx.doi.org* (2009).
65. Rahimov, F. *et al.* Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nature genetics* **40**, 1341-1347 (2008).
66. Mortazavi, A. *et al.* Integrating and mining the chromatin landscape of cell-type specificity using self-organizing maps. *Genome research* **23**, 2136-2148 (2013).
67. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319 (2013).
68. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155**, 934-947 (2013).
69. Rao, S.S.P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665-1680 (2014).
70. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, NY)* **326**, 289-293 (2009).
71. Prabhakar, S. *et al.* Human-Specific Gain of Function in a Developmental Enhancer. *Science (New York, NY)* **321**, 1346-1350 (2008).
72. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355-364 (2014).
73. Zákány, J., Kmita, M. & Duboule, D. A Dual Role for Hox Genes in Limb Anterior-Posterior Asymmetry. *Science (New York, NY)* **304**, 1669-1672 (2004).
74. Spitz, F., Gonzalez, F. & Duboule, D. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113**, 405-417 (2003).
75. Santagati, F. Temporal requirement of Hoxa2 in cranial neural crest skeletal morphogenesis. *Development (Cambridge, England)* **132**, 4927-4936 (2005).
76. van de Werken, H.J.G. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature Methods* **9**, 969-972 (2012).
77. Leslie, E.J. *et al.* 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. *American journal of human genetics* **96**, 397-411 (2015).
78. Hoover-Fong, J.E. *et al.* Facial dysgenesis: A novel facial syndrome with chromosome 7 deletion p15.1-21.1. *American Journal of Medical Genetics* **117A**, 47-56 (2003).
79. Cotney, J.L. & Noonan, J.P. Chromatin immunoprecipitation with fixed animal tissues and preparation for high-throughput sequencing. *Cold Spring Harbor protocols* **2015**, 191-199 (2015).
80. Andrews, S. FastQC: a quality control tool for high throughput sequence data. in *Genome Biology* (2010).
81. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)* **32**, 3047-3048 (2016).
82. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359 (2012).
83. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**, 576-589 (2010).
84. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**, 841-842 (2010).
85. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome research* **12**, 996-1006 (2002).

86. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research* **42**, W187-91 (2014).
87. R Core Team. R: A language and environment for statistical computing. v3.3.3 edn (R Foundation for Statistical Computing, Vienna, Austria, 2017).
88. Gokhman, D. *et al.* Gene ORGANizer: Linking Genes to the Organs They Affect. *bioRxiv*, 106948 (2017).
89. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* **42**, D1001-6 (2014).
90. McMurry, J.A. *et al.* Navigating the Phenotype Frontier: The Monarch Initiative. *Genetics* **203**, 1491-1495 (2016).
91. Mortazavi, A., Thompson, E.C.L., Garcia, S.T., Myers, R.M. & Wold, B. Comparative genomics modeling of the NRSF/REST repressor network: From single conserved sites to genome-wide repertoire. *Genome research* **16**, 1208-1221 (2006).
92. de Hoon, M.J.L., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics (Oxford, England)* **20**, 1453-1454 (2004).
93. Saldanha, A.J. Java Treeview--extensible visualization of microarray data. *Bioinformatics (Oxford, England)* **20**, 3246-3248 (2004).
94. Kheradpour, P. & Kellis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Research* **42**, 2976-2987 (2014).
95. Visel, A. *et al.* Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nature genetics* **40**, 158-160 (2008).
96. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, pp. 10-12 (2011).
97. Thongjuea, S., Stadhouders, R., Grosveld, F.G., Soler, E. & Lenhard, B. r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Research* **41**, e132-e132 (2013).

Figure Legends

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 Carnegie Stages (CS) or approximate post-conception weeks (pcw). Voids or cleavages in the embryo are indicated by black shaded regions. **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[49] to match the 12 epigenomic signals profiled by Roadmap Epigenome[22]. Asterisks indicate signals containing only imputed data. These imputed datasets were then used to predict chromatin states using a Hidden Markov Model approach (ChromHMM)[44] 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.

Figure 2. Histone Modification Profiles in Human Craniofacial Development. a. Heatmap and hierarchical clustering of pairwise Pearson correlations for non-overlapping 10kb bins across the human genome for 114 individual histone modification profiles from human craniofacial tissues. Relatedness of epigenomic profiles by sample indicated by dendrogram along vertical axes of 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[83]. **c.** Histone modification peaks identified in at least two separate tissue samples from the same developmental stage and annotated into three broad categories: promoter (2kb upstream of TSS), exons, and all other intronic or intergenic locations.

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 10kb 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. **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[22]. **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 standard deviation. Overall chromatin state segmentation in craniofacial samples identifies similar numbers and percentages of each of 25 states published by Roadmap Epigenome[22].

Figure 4. Chromatin State Segmentations Identify Novel Craniofacial Regulatory Sequences. a. Percentage of *in vivo* validated embryonic enhancers with (orange) or without (grey) craniofacial activity from the Vista Enhancer Browser[50] identified by craniofacial chromatin segments annotated as enhancer states. Significance determined by Fisher's exact test. **b.** Selected validated enhancers with craniofacial activity identified by this study from the the Vista Enhancer Browser. **c.** Principal component analysis projection of second and third component dimensions for 146 H3K27ac profiles at 425380 regions annotated as enhancer segments in any of the samples profiled here or Roadmap Epigenome. Samples are color coded by group annotations assigned by Roadmap Epigenome or craniofacial samples from this

study. Percent of variance across samples explained by each component are indicated along each axis. **d.** Transcription factor position weight matrices identified by HOMER[83] as enriched in novel craniofacial enhancer segments. **e.** Significant enrichments of human disease phenotypes for genes assigned to novel craniofacial enhancer segments as reported by GREAT[58]. **f.** Enrichment of anatomical expression of transcription factors identified as potentially regulated by novel craniofacial enhancer segments as reported by GeneORGANizer[88]. Heat indicates fold enrichment of expression in individual anatomical region or organ. Craniofacial and appendicular skeleton showed most significant enrichments.

Figure 5. Self-Organizing Map for Biological Mining of Craniofacial Enhancers a. Flattened projections of toroid self-organizing map generated from H3K27ac signals from 146 samples across 425380 enhancer segments consisting of 2500 individual hexagonal units for four craniofacial tissues, four embryonic stem-cell and related cell types, and four adult brain tissues. Higher scoring units in a given tissue are indicated by red, lower scoring units by blue. Two selected metaclusters scoring highly for craniofacial or brain tissues are indicated by black outlines. **b**. Fold enrichment (dots) and significance (bars) of top human disease phenotypes associated with genes assigned to enhancer segments by GREAT[58] in each metacluster. A metacluster highly scoring reproducibly in craniofacial tissues is enriched for enhancers putatively assigned to genes associated with a wide variety of craniofacial abnormalities. A metacluster highly scoring across brain tissues is enriched for diverse brain and neurological diseases. While PCA and hierarchical clustering identified craniofacial tissues were more similar to ESC and ESC-derived cell types, the self-organizing map identifies distinct clusters of enhancers specific to craniofacial tissues.

Figure 6. Identification of Potential Craniofacial Locus Control Region for *HOXA* Gene Cluster. a. Large 450kb window lacking any annotated protein-coding genes with extensive enrichment of activated enhancer (yellow and orange) and transcriptionally active (green) segment annotations in human craniofacial tissue. See Figure 3b for full annotations. Multiple validated craniofacial enhancers have been identified in this window by the Vista Enhancer Browser. In this study we tested and validated the craniofacial enhancer activity of HACNS50, located within the bivalent chromatin state at the right of the displayed window. Segments interrogated by 4C-Seq indicated by vertical colored viewpoint bars **b.** Approximately 3Mb window of the human genome encompassing the window identified in panel **a** (black box) and containing the *HOXA* gene cluster. **c.** Spidergrams indicating significant interactions between color-coded viewpoints and distal sites identified by 4C-Seq in mouse E11.5 craniofacial tissue. Viewpoints 1 and 2 do not cross putative TAD boundary near HACNS50 enhancer. Viewpoints 3

and 4 make significant contacts within identified window and with the *HOXA* gene cluster. Reciprocal experiments from the *HOXA* gene cluster (viewpoint 6) indicated significant long-range interactions with both boundaries of the window in panel **a**.

Supplemental Figure and Table Legends

Supplemental figures and tables can be obtained from FigShare: **10.6084/m9.figshare.4954202**

Supplementary Figure 1. 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.

Supplementary Figure 2. Complete Histone Modification Profiles in Human Craniofacial Development 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[83].

Supplementary Figure 3. Imputed Histone Modification Profiles in Human Craniofacial Development. a. Heatmap and hierarchical clustering of pairwise Pearson correlations for imputed histone modification profiles from human craniofacial tissues. Darker orange indicates positive correlation between datasets. **b.** 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.

Supplementary Figure 4. 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[22]. **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. Error bars represent standard deviation. Overall chromatin state segmentation in craniofacial samples identifies similar numbers and percentages of each of the states published by Roadmap Epigenome[22].

Supplementary Figure 5. All Enhancers Tested for Craniofacial Activity 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.

Supplementary Figure 6. 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.

Supplementary Figure 7. Identification of Craniofacial-specific Enhancers Flanking *MSX2*. 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.

Supplementary Figure 8. Integration of CL/P GWAS Data Places SNPs within Craniofacial-specific Enhancers. a. Enrichment analysis identified orofacial cleft GWAS tag SNPs preferentially among craniofacial tissue. **b.** Enhancer state analysis permits placement of a potentially causative allele for non-syndromic CL/P (rs642961) within a predicted early development enhancer state. This enhancer state is located between *IRF6* and *DIEXF* and may influence expression of *IRF6*. **c.** The orofacial cleft-associated tag SNP rs745080 resides within the intron of *TXNDC16*, which is marked by a craniofacial-specific enhancer state.

Supplementary Figure 9. Clustering Identifies Similar Functional Enrichment to Self-Organizing Maps a. K-means clustering performed directly on the matrix of H3K27ac signals, using the same number of clusters utilized for the self-organizing map (199), showed distinct H3K27ac activation in craniofacial tissues (enlarged section on the right). **b.** Analysis of the sequence content of enhancer clusters most specific for craniofacial activity identified enrichments of motifs for the *HOX*, *LHX*, *MSX* and *DLX* families of transcription factors.

Supplementary Figure 10. Incorporation of Topologically Associated Domain Structure and Clinical Case Suggests Interaction Between Distal Enhancer and HoxA Region. Hi-C data from HUVEC visualized using the Hi-C browser (http://promoter.bx.psu.edu/hi-c/view.php) indicates an interaction between the *HOXA* cluster and an intergenic region approximately 1Mb away from the anterior side. A deletion covering this region and notably leaving the *HOXA* cluster intact has been described in a patient with facial dysgenesis[78] (indicated in purple). The intergenic region contains a 450kb enhancer state, as represented by the 15-, 18- and 25-state ChromHMM model of craniofacial data, with regions of high conservation. Four viewpoints used in 4C-seq are indicated, flanking the region of interest. Seven enhancers with craniofacial activity are located within this region, indicated by black bars and representative images. Enhancers mm403-407 and hs1600 were tested by the Vista Enhancer Browser[50], HACNS50 was tested independently.

Supplementary Figure 11. Distal Regulatory Region Shares Chromatin State in Mouse. ChIP-Seq data from Mouse Encode for embryonic day 11.5 facial prominence display a conserved set of chromatin marks in the region distal to the *HOXA* cluster suggesting a conserved function in mouse craniofacial development.

Supplementary Figure 12. Syntenic Block Near *HOXA* Cluster. A comparison of a 10 Mb window around the *HOXA* cluster on human Chromosome 7 shows synteny with mouse Chromosome 6. In addition to the preservation of gene order, there is also preservation of a large non-coding region distal to the anterior side of the *HOXA* cluster in mouse.

Supplementary Figure 13. 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.[77] 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.

Supplementary Table 1. Table showing, for each sample, for each mark, the number of total sequencing reads, the number of uniquely mapped reads, the number of multi-mapped reads, percentage of mapped reads, and percentage of uniquely mapped reads. Total numbers, in billions, and means, in millions, are displayed in bottom two rows of the table.

Supplementary Table 2. Table showing all enhancer segments in craniofacial epigenomic atlas that were never annotated as any type of enhancer state in all of Roadmap Epigenome.

Supplementary Table 3. Motifs identified for enrichment of transcription factor binding sites for novel craniofacial-specific enhancers found in this study. Consensus motif sequence, p-values, q-values, number of target sequences with the motif, percent of target sequences with the motif, number of background sequences with the motif, and percent of all background sequences with the motif are all indicated.

Supplementary Table 4. Functional categories with significant enrichment based on assignment of craniofacial-specific enhancers to the nearest gene, using Genomic Regions Enrichment of Annotations Tool (GREAT)[58].

Supplementary Table 5. Regions containing tag SNPS identified in craniofacial tissue and associated with orofacial clefting, including gene assignments where applicable. Of note is a tag SNP in the noncoding region between *IRF6* and *DEXIF*, as well as an intronic sequence within the *TXNDC16* gene.

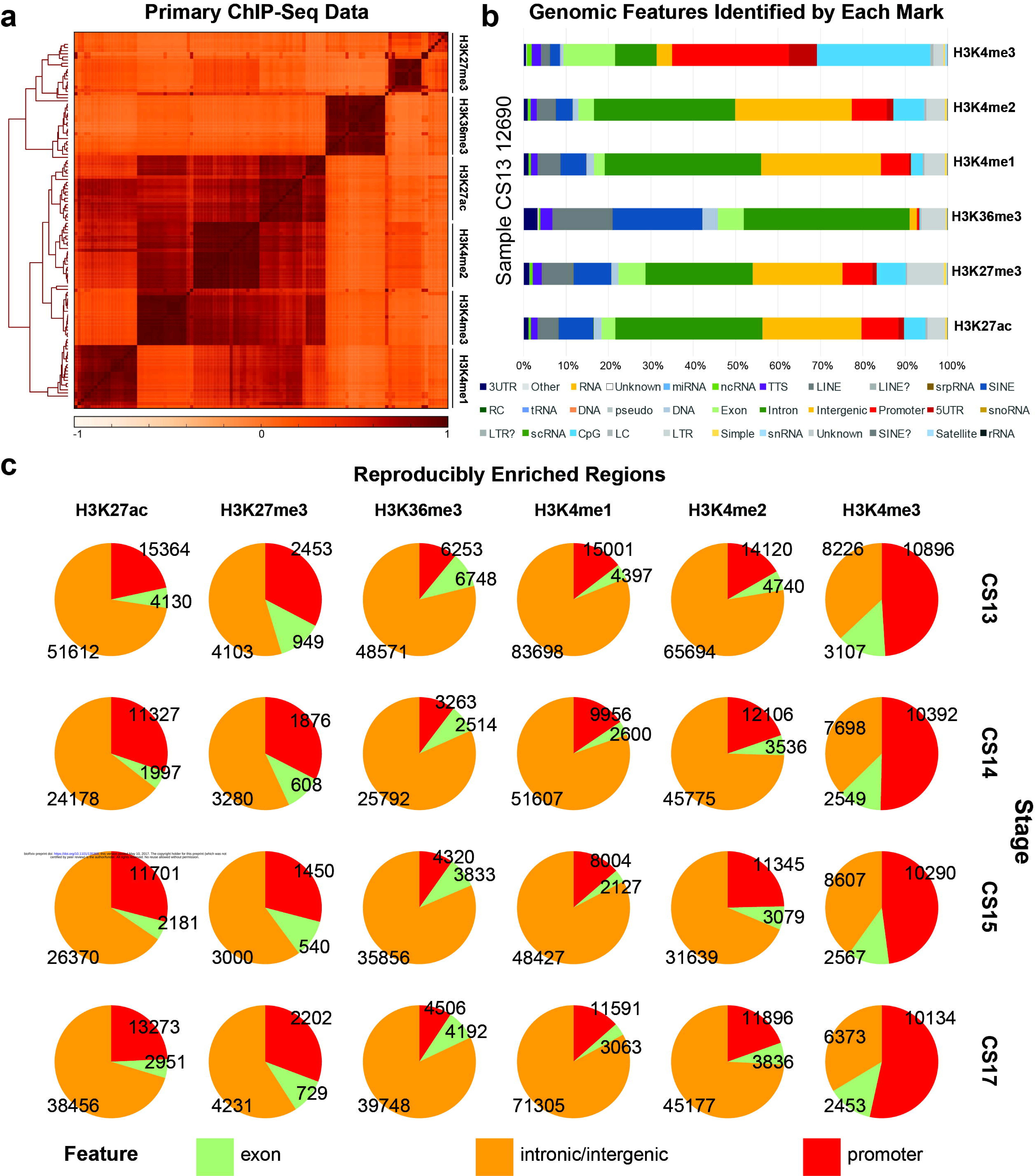
Supplementary Table 6. Functional categories with significant enrichment in the clusters of craniofacial-specific enhancers (annotated to the the nearest gene) obtained from K-means clustering on the matrix of H3K27ac signals using the same number of clusters utilized for the self-organizing map, using Genomic Regions Enrichment of Annotations Tool (GREAT)[58].

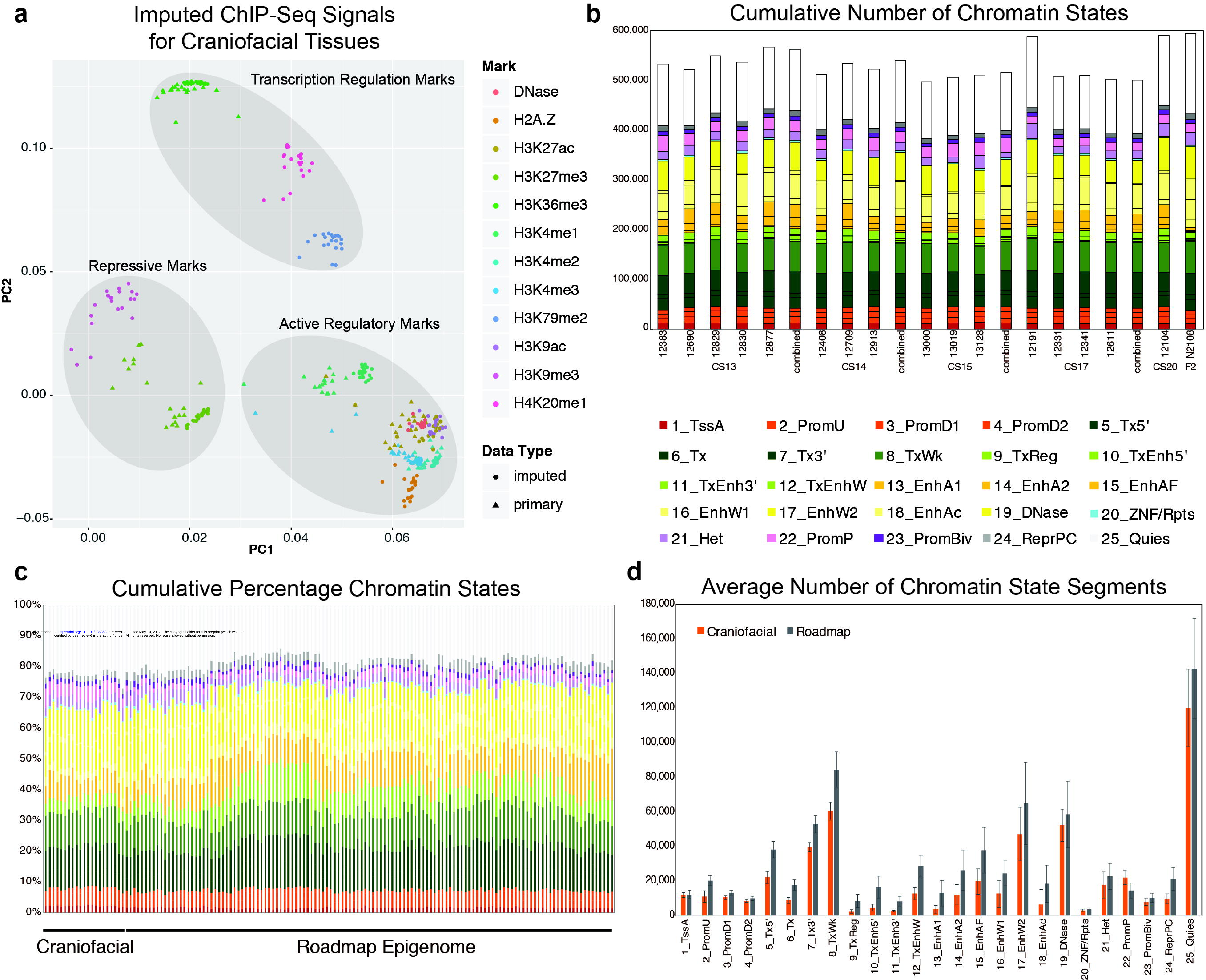
Supplementary Table 7. Sheet1: 582 regions identified across the genome as containing significant fractions of bases annotated as craniofacial enhancers; start position, end position, and fraction of bases annotated as an enhancer state are shown. The second sheet shows individual window analysis with fold enrichment versus randomized enhancer segmentations and permutation p-values.

Supplementary Table 8. Functional categories with significant enrichment based on assignment of enriched enhancer windows to the nearest gene, using Genomic Regions Enrichment of Annotations Tool (GREAT)[58].

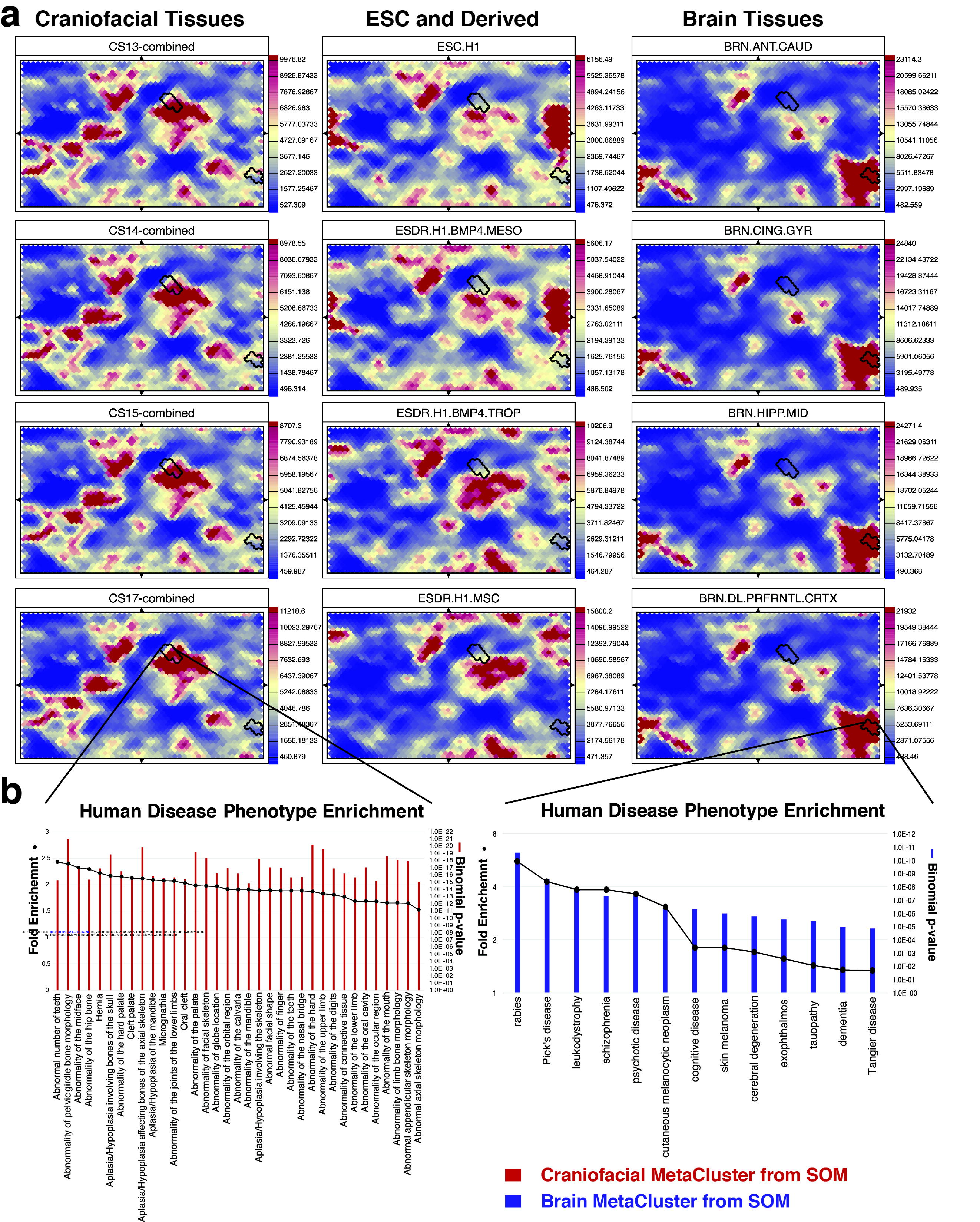
Supplementary Table 9. Primers used for 4C-Seq analysis in mouse craniofacial tissue.

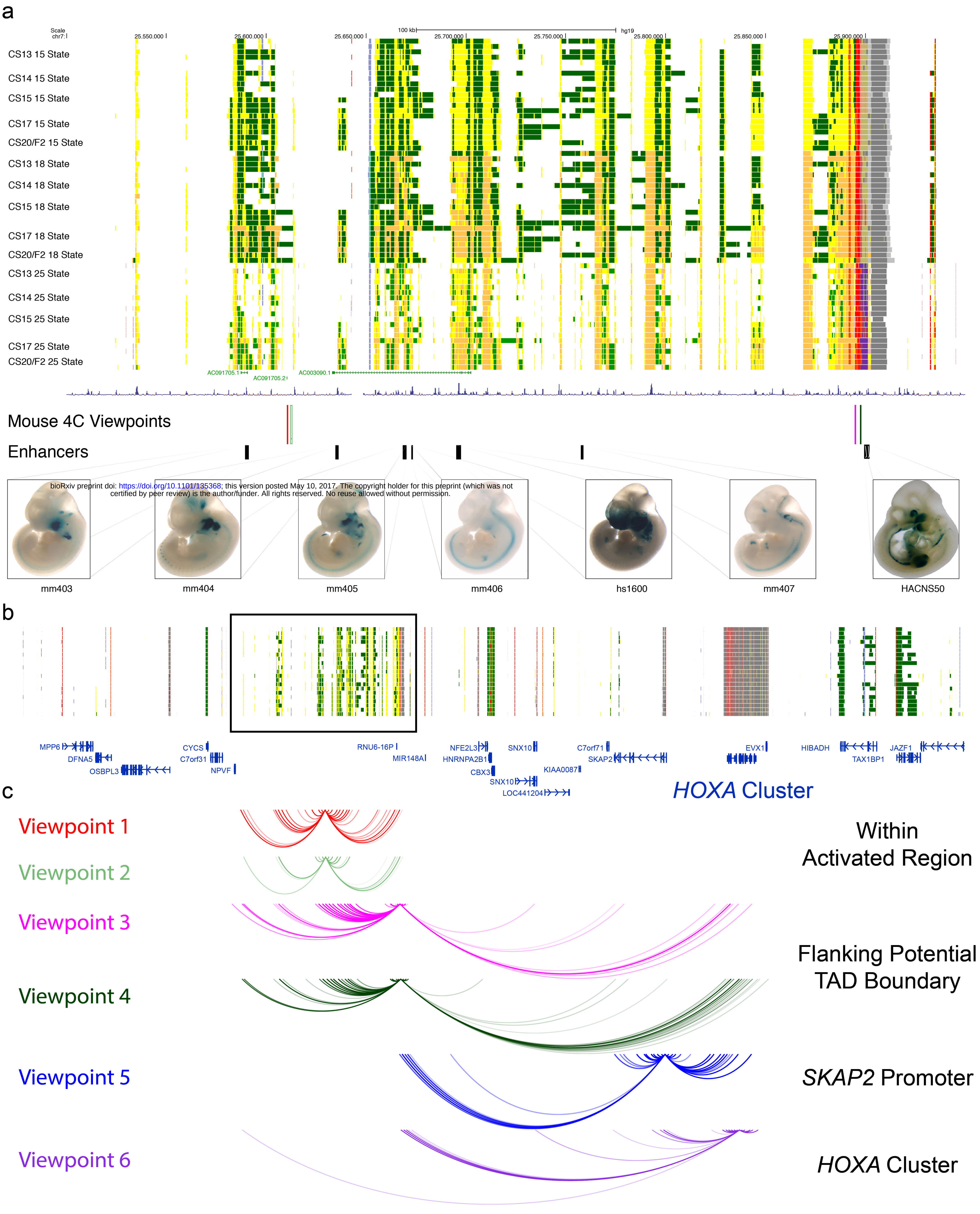
a
CS13
4 PCW
CS14-15
4.5 - 5.5 PCW
CS17
6 PCW
Individual Tissue Sample ChIP-Seq
b
H3K4me1
H3K4me2
H3K4me3
H3K27ac
H3K27me3
H3K36me3
Chromatin Signal Imputation
Chromatin State Segmentation
c
11 Histone Marks & DNase
Genome-wide Signal
21 Embryonic Epigenomes (CS13 to CS20)
H3K9me3*
DNase*
H2A.Z*
H4K20me1*
H3K79me2*
H3K9ac*
H3K36me3
H3K27me3
H3K27ac
H3K4me3
H3K4me2
H3K4me1
Chromatin State Predictions
Genes
Craniofacial Specific Enhancers
Enrichment of Craniofacial Functions and Genetic Associations
Novel Craniofacial Locus Control Region
Long-Range Interactions


a
Primary ChIP-Seq Data
H3K27me3
H3K36me3
H3K27ac
H3K4me2
H3K4me3
H3K4me1
-1
0
1
b
Genomic Features Identified by Each Mark
Sample CS13 12690
H3K4me3
H3K4me2
H3K4me1
H3K36me3
H3K27me3
H3K27ac
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
3UTR
Other
RNA
Unknown
miRNA
ncRNA
TTS
LINE
LINE?
srpRNA
SINE
RC
tRNA
DNA
pseudo
DNA
Exon
Intron
Intergenic
Promoter
5UTR
snoRNA
LTR?
scRNA
CpG
LC
LTR
Simple
snRNA
Unknown
SINE?
Satellite
rRNA
c
Reproducibly Enriched Regions
H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K4me2
H3K4me3
15364
4130
51612
2453
949
4103
6253
6748
48571
15001
4397
83698
14120
4740
65694
8226
10896
3107
CS13
11327
1997
24178
1876
608
3280
3263
2514
25792
9956
2600
51607
12106
3536
45775
7698
10392
2549
CS14
11701
2181
26370
1450
540
3000
4320
3833
35856
8004
2127
48427
11345
3079
31639
8607
10290
2567
CS15
13273
2951
38456
2202
729
4231
4506
4192
39748
11591
3063
71305
11896
3836
45177
6373
10134
2453
CS17
Stage
Feature
exon
intronic/intergenic
promoter


a
Imputed ChIP-Seq Signals for Craniofacial Tissues
Transcription Regulation Marks
Repressive Marks
Active Regulatory Marks
PC2
PC1
0.10
0.05
0.00
−0.05
0.00
0.02
0.04
0.06
Mark
DNase
H2A.Z
H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K4me2
H3K4me3
H3K79me2
H3K9ac
H3K9me3
H4K20me1
Data Type
imputed
primary
b
Cumulative Number of Chromatin States
600,000
500,000
400,000
300,000
200,000
100,000
0
12383
12690
12829
12830
12877
combined
12408
12709
12913
combined
13000
13019
13128
combined
12191
12331
12341
12611
combined
12104
N2108
CS13
CS14
CS15
CS17
CS20
F2
1_TssA
2_PromU
3_PromD1
4_PromD2
5_Tx5'
6_Tx
7_Tx3'
8_TxWk
9_TxReg
10_TxEnh5'
11_TxEnh3'
12_TxEnhW
13_EnhA1
14_EnhA2
15_EnhAF
16_EnhW1
17_EnhW2
18_EnhAc
19_DNase
20_ZNF/Rpts
21_Het
22_PromP
23_PromBiv
24_ReprPC
25_Quies
c
Cumulative Percentage Chromatin States
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Craniofacial
Roadmap Epigenome
d
Average Number of Chromatin State Segments
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
Craniofacial
Roadmap
1_TssA
2_PromU
3_PromD1
4_PromD2
5_Tx5'
6_Tx
7_Tx3'
8_TxWk
9_TxReg
10_TxEnh5'
11_TxEnh3'
12_TxEnhW
13_EnhA1
14_EnhA2
15_EnhAF
16_EnhW1
17_EnhW2
18_EnhAc
19_DNase
20_ZNF/Rpts
21_Het
22_PromP
23_PromBiv
24_ReprPC
25_Quies

a
Percentage of Validated Enhancers
0
20
40
60
80
100
p = 3.28E−14
Craniofacial
Other
b
hs521
hs550
hs852
hs858
hs1004
hs1475
hs1604
hs1626
hs1635
hs1720
mm43
mm404
mm405
mm423
mm428
mm622
mm686
mm761
mm924
mm1088
c
Mesenchmye
BloodT-Cell
HSC B-Cell
Muscle
Stem Cells and Derived Cell Types
Brain and Neuronal
Craniofacial
5000000
2500000
0
−2500000
PC2 24.1% of variance
−2e+06
0e+00
2e+06
4e+06
PC3 10.2% of Variance
Tissue Type
Adipose
Blood_T_cell
Brain
Craniofacial
Digestive
ENCODE2012
Epithelial
ES_deriv
ESC
Heart
HSC_B_cell
IMR90
iPSC
Mesench
Muscle
Myosat
Neurosph
Other
S._Muscle
Sm._Muscle
Thymus
d
Twist2
Lmx1b
Six1
Lhx1
Nkx6.1
Tcf12
e
Human Disease Phenotype Enrichment
-log10(binomial p-value)
0
2
4
6
8
10
Abnormality of the inner ear
Functional abnormality of the inner ear
Abnormality of the parietal bone
Abnormality of the hard palate
Cleft palate
Sprengel anomaly
Abnormality of the upper arm
Oral cleft
Abnormality of the external nose
Abnormality of the scapula
Limited elbow extension
Abnormality of the shoulder
Abnormality of the elbow
Urogenital fistula
Limited elbow movement
Genitourinary tract malformation
Depressed nasal bridge
Frontal bossing
Abnormality of the nares
Posteriorly rotated ears
Abnormality of the nasal alae
Low-set ears
Abnormal location of ears
Mixed hearing impairment
Abnormality of the nail
Aplasia/Hypoplasia involving bones of the thorax
Abnormality of the musculature of the upper limbs
Abnormality of the hip bone
Craniosynostosis
f
x 2
x 1
x 0.50
Fold Enrichment


a
Craniofacial Tissues
ESC and Derived
Brain Tissues
CS13-combined
9976.82
8926.87433
7876.92867
6826.983
5777.03733
4727.09167
3677.146
2627.20033
1577.25467
527.309
ESC.H1
6156.49
5525.36578
4894.24156
4263.11733
3631.99311
3000.86889
2369.74467
1738.62044
1107.49622
476.372
BRN.ANT.CAUD
23114.3
20599.66211
18085.02422
15570.38633
13055.74844
10541.11056
8026.47267
5511.83478
2997.19689
482.559
CS14-combined
8978.55
8036.07933
7093.60867
6151.138
5208.66733
4266.19667
3323.726
2381.25533
1438.78467
496.314
ESDR.H1.BMP4.MESO
5606.17
5037.54022
4468.91044
3900.28067
3331.65089
2763.02111
2194.39133
1625.76156
1057.13178
488.502
BRN.CING.GYR
24840
22134.43722
19428.87444
16723.31167
14017.74889
11312.18611
8606.62333
5901.06056
3195.49778
489.935
CS15-combined
8707.3
7790.93189
6874.56378
5958.19567
5041.82756
4125.45944
3209.09133
2292.72322
1376.35511
459.987
ESDR.H1.BMP4.TROP
10206.9
9124.38744
8041.87489
6959.36233
5876.84978
4794.33722
3711.82467
2629.31211
1546.79956
464.287
BRN.HIPP.MID
24271.4
21629.06311
18986.72622
16344.38933
13702.05244
11059.71556
8417.37867
5775.04178
3132.70489
490.368
CS17-combined
11218.6
10023.29767
8827.99533
7632.693
6437.39067
5242.08833
4046.786
2851.48367
1656.18133
460.879
ESDR.H1.MSC
15800.2
14096.99522
12393.79044
10690.58567
8987.38089
7284.17611
5580.97133
3877.76656
2174.56178
471.357
BRN.DL.PRFRNTL.CRTX
21932
19549.38444
17166.76889
14784.15333
12401.53778
10018.92222
7636.30667
5253.69111
2871.07556
488.46
b
Human Disease Phenotype Enrichment
Fold Enrichemnt
Binomial p-value
3
2.5
2
1.5
1
0.5
0
1.0E-22
1.0E-21
1.0E-20
1.0E-19
1.0E-18
1.0E-17
1.0E-16
1.0E-15
1.0E-14
1.0E-13
1.0E-12
1.0E-11
1.0E-10
1.0E-09
1.0E-08
1.0E-07
1.0E-06
1.0E-05
1.0E-04
1.0E-03
1.0E-02
1.0E-01
1.0E+00
Abnormal number of teeth
Abnormality of pelvic girdle bone morphology
Abnormality of the midface
Abnormality of the hip bone
Hernia
Aplasia/Hypoplasia involving bones of the skull
Abnormality of the hard palate
Cleft palate
Aplasia/Hypoplasia affecting bones of the axial skeleton
Aplasia/Hypoplasia of the mandible
Micrognathia
Abnormality of the joints of the lower limbs
Oral cleft
Abnormality of the palate
Abnormality of facial skeleton
Abnormality of globe location
Abnormality of the orbital region
Abnormality of the calvaria
Abnormality of the mandible
Aplasia/Hypoplasia involving the skeleton
Abnormal facial shape
Abnormality of finger
Abnormality of the teeth
Abnormality of the nasal bridge
Abnormality of the hand
Abnormality of the upper limb
Abnormality of the digits
Abnormality of connective tissue
Abnormality of the lower limb
Abnormality of the oral cavity
Abnormality of the ocular region
Abnormality of the mouth
Abnormality of limb bone morphology
Abnormal appendicular skeleton morphology
Abnormal axial skeleton morphology
Human Disease Phenotype Enrichment
Fold Enrichemnt
Binomial p-value
8
4
2
1
1.0E-12
1.0E-11
1.0E-10
1.0E-09
1.0E-08
1.0E-07
1.0E-06
1.0E-05
1.0E-04
1.0E-03
1.0E-02
1.0E-01
1.0E+00
rabies
Pick's disease
leukodystrophy
schizophrenia
psychotic disease
cutaneous melanocytic neoplasm
cognitive disease
skin melanoma
cerebral degeneration
exophthalmos
tauopathy
dementia
Tangier disease
🟥 Craniofacial MetaCluster from SOM
🟦 Brain MetaCluster from SOM


a
Scale
chr7:
100 kb
hg19
25,550,000
25,600,000
25,650,000
25,700,000
25,750,000
25,800,000
25,850,000
25,900,000
CS13 15 State
CS14 15 State
CS15 15 State
CS17 15 State
CS20/F2 15 State
CS13 18 State
CS14 18 State
CS15 18 State
CS17 18 State
CS20/F2 18 State
CS13 25 State
CS14 25 State
CS15 25 State
CS17 25 State
CS20/F2 25 State
AC091705.1
AC091705.2
AC003090.1
Mouse 4C Viewpoints
Enhancers
mm403
mm404
mm405
mm406
hs1600
mm407
HACNS50
b
MPP6
DFNA5
OSBPL3
CYCS
C7orf31
NPVF
RNU6-16P
MIR148A
NFE2L3
HNRNPA2B1
CBX3
SNX10
SNX10
LOC441204
KIAA0087
C7orf71
SKAP2
EVX1
HIBADH
TAX1BP1
JAZF1
HOXA Cluster
c
Viewpoint 1
Viewpoint 2
Viewpoint 3
Viewpoint 4
Viewpoint 5
Viewpoint 6
Within
Activated Region
Flanking Potential
TAD Boundary
SKAP2 Promoter
HOXA Cluster