

Cell Type- and Tissue-specific Enhancers in Craniofacial Development

Sudha Sunil Rajderkar¹, Kitt Paraiso¹, Maria Luisa Amaral², Michael Kosicki¹, Laura E. Cook¹,
Fabrice Darbellay^{1,3}, Cailyn H. Spurrell¹, Marco Osterwalder^{1,4}, Yiwen Zhu¹, Han Wu¹, Sarah
Yasmeen Afzal^{1,5}, Matthew J. Blow⁶, Guy Kelman^{1,7}, Iros Barozzi,^{1,8,9} Yoko Fukuda-
Yuzawa^{1,10}, Jennifer A. Akiyama¹, Veena Afzal¹, Stella Tran¹, Ingrid Plajzer-Frick¹, Catherine
S. Novak¹, Momoe Kato¹, Riana D. Hunter^{1,11}, Kianna von Maydell¹, Allen Wang¹², Lin Lin¹²,
Sebastian Preissl^{12,13}, Steven Lisgo^{14,15}, Bing Ren¹⁶, Diane E. Dickel^{1,17}, Len A.
Pennacchio^{1,6,18}, Axel Visel^{*;1,6,19}

¹ Environmental Genomics & System Biology Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA.

² Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA, USA

^{3#} Department of Genetic Medicine and Development, Faculty of Medicine, University of Geneva, 1211 Geneva, Switzerland

^{4#} Department for BioMedical Research, University of Bern, 3008 Bern, Switzerland

^{5#} Lucile Packard Children's Hospital, Stanford University, Stanford, CA 94304

⁶ U.S. Department of Energy Joint Genome Institute, 1 Cyclotron Road, Berkeley, CA 94720, USA.

^{7#} The Jerusalem Center for Personalized Computational Medicine, Hebrew University of Jerusalem, Jerusalem, Israel.

^{8#} Center for Cancer Research, Medical University of Vienna, Borschkegasse 8a 1090, Vienna, Austria

^{9#} Department of Surgery and Cancer, Imperial College London, London, UK.

^{10#} University Research Management Center, Tohoku University, Sendai, Miyagi, 980-8577, Japan.

^{11#} UC San Francisco, Division of Experimental Medicine, 1001 Potrero Ave, San Francisco, CA 94110

¹² Center for Epigenomics, University of California San Diego School of Medicine, La Jolla, CA, USA.

^{13#} Institute of Experimental and Clinical Pharmacology and Toxicology, Faculty of Medicine, University of Freiburg, Freiburg, Germany.

¹⁴ Human Developmental Biology Resource, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, NE1 3BZ, UK.

^{15#} Newcastle University Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle, NE1 3BZ, UK

¹⁶ Institute of Genome Medicine, Moores Cancer Center, University of California, San Diego School of Medicine, La Jolla, CA, USA.

^{17#} Octant Inc., Emeryville, CA 94608, USA.

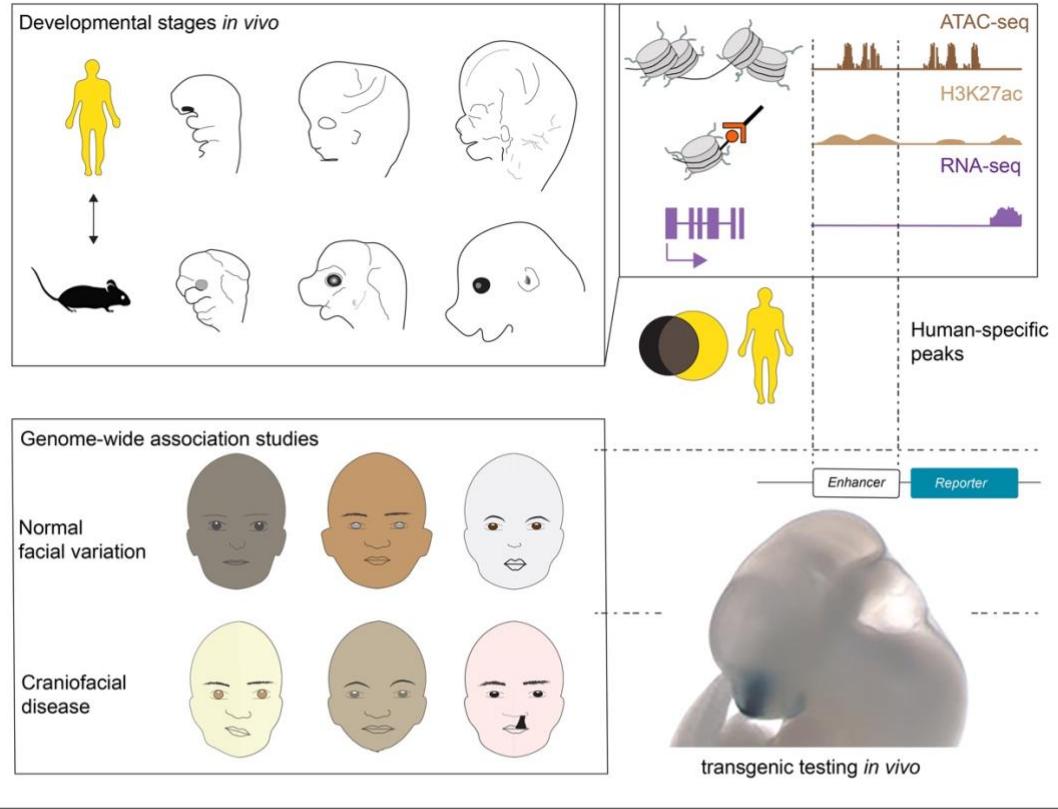
¹⁸ Comparative Biochemistry Program, University of California, Berkeley, CA 94720, USA.

¹⁹ School of Natural Sciences, University of California, Merced, Merced, California, USA.

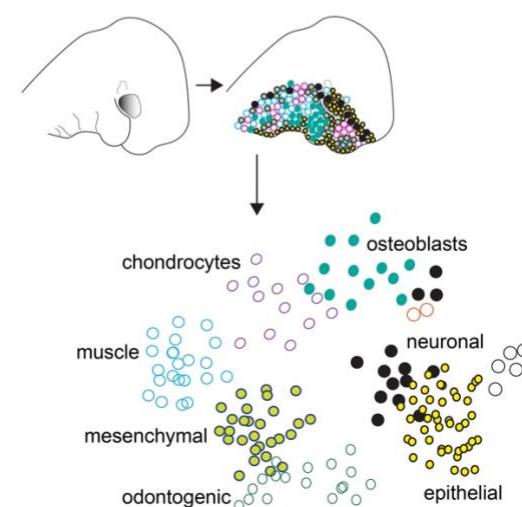
* To whom correspondence should be addressed: A.V., avisel@lbl.gov

Current address.

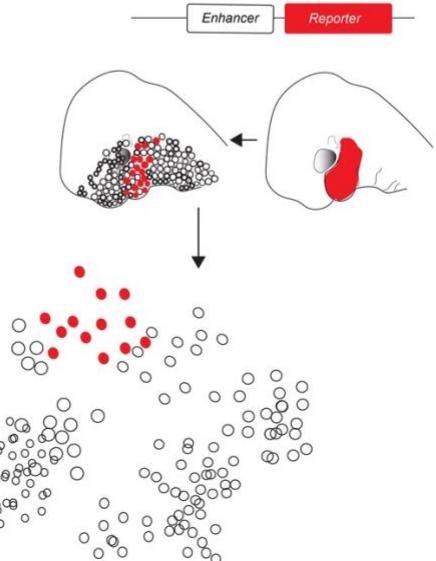
Active human developmental enhancers



Single-cell gene expression reference of the embryonic face



Cell type-specific enhancer activity



Graphical Abstract

Abstract

The genetic basis of craniofacial birth defects and general variation in human facial shape remains poorly understood. Distant-acting transcriptional enhancers are a major category of non-coding genome function and have been shown to control the fine-tuned spatiotemporal expression of genes during critical stages of craniofacial development¹⁻³. However, a lack of accurate maps of the genomic location and cell type-specific *in vivo* activities of all craniofacial enhancers prevents their systematic exploration in human genetics studies. Here, we combined histone modification and chromatin accessibility profiling from different stages of human craniofacial development with single-cell analyses of the developing mouse face to create a comprehensive catalogue of the regulatory landscape of facial development at tissue- and single cell-resolution. In total, we identified approximately 14,000 enhancers across seven developmental stages from weeks 4 through 8 of human embryonic face development. We used transgenic mouse reporter assays to determine the *in vivo* activity patterns of human face enhancers predicted from these data. Across 16 *in vivo* validated human enhancers, we observed a rich diversity of craniofacial subregions in which these enhancers are active *in vivo*. To annotate the cell type specificities of human-mouse conserved enhancers, we performed single-cell RNA-seq and single-nucleus ATAC-seq of mouse craniofacial tissues from embryonic days e11.5 to e15.5. By integrating these data across species, we find that the majority (56%) of human craniofacial enhancers are functionally conserved in mice, providing cell type- and embryonic stage-resolved predictions of their *in vivo* activity profiles. Using retrospective analysis of known craniofacial enhancers in combination with single cell-resolved transgenic reporter assays, we demonstrate the utility of these data for predicting the *in vivo* cell type specificity of enhancers. Taken together, our data provide an expansive resource for genetic and developmental studies of human craniofacial development.

Introduction

The development of the human face is a highly complex morphogenetic process. It requires the precise formation of dozens of intricate structures to enable the full complement of facial functions including food uptake, breathing, speech, major sensory functions including hearing, sight, smell, taste, and nonverbal communication through facial expression. Intriguingly, these functional constraints coincide with substantial inter-individual variation in facial morphology, which humans use as the principal means for recognizing each other. Apart from providing the basis for normal facial variation, early developmental processes underlying facial morphogenesis are highly sensitive to genetic abnormalities as well as environmental effects⁴. Even subtle disturbances during embryogenesis can result in a range of craniofacial defects or dysfunctions⁵.

In embryonic facial development, the primary germ layers as well as the neural crest contribute crucially to the formation of the pharyngeal arches, the frontonasal process and the midface, which in combination give rise to the derived structures of the face^{6–9}. The primary palate forms by the fifth week post conception¹⁰ and development of primary palate derivatives, secondary palate, and many other structures, combined with overall rapid growth, result in a discernable human-like appearance by the tenth week post conception¹¹. Genetic or environmental perturbations during these crucial developmental stages are known to result in craniofacial malformations of varying severity and of typically irreversible nature^{12–16}. Development of the mammalian face requires a conserved set of genes and signaling pathways¹⁷, which are regulated by distant-acting transcriptional enhancers that control gene expression in time and space^{1,18–24}. Together with the genes they control, these enhancers are a critical component of mammalian craniofacial morphogenesis. It is estimated that there are hundreds of thousands of enhancers in the human genome for approximately 20,000 genes²⁵ and chromatin profiling studies have identified initial sets of enhancers predicted to be active in craniofacial development^{1,25,26}. However, these data sets do not cover critical stages of human facial development, such as secondary palate formation, and provide no information about the cell type specificity of individual enhancers. In part due to the continued

incomplete annotation state of the craniofacial enhancer landscape, the number of enhancers that could be mechanistically linked to facial variation or craniofacial birth defects has remained limited^{1,18–23}. With an increasingly refined view of the genetic variation underlying human facial variation²⁷ and whole genome sequencing as an increasingly common clinical approach for the identification of noncoding mutations in craniofacial birth defect patients^{28,29}, an expanded and accurate map of human craniofacial enhancers is critical for interpretation of any noncoding findings emerging from these studies. Here we provide a comprehensive compilation of regulatory regions from the developing human face during embryonic stages critical for birth defects including orofacial clefts, along with cell type-specific gene expression and open chromatin signatures for the developing mammalian face.

Results

Epigenomic Landscape of the Human Embryonic Face

To map the epigenomic landscape of critical periods of human face development, we focused on Carnegie stages (CS) 18-23, a period coinciding with the formation of important structures including the maxillary palate, rapid overall growth, and significant changes in the relative proportions of craniofacial structures that impact on ultimate craniofacial shape^{11,30,31}. These stages are of direct clinical relevance because common craniofacial defects, including cleft palate and major facial dysmorphologies, result from disruptions within this developmental window (**Figure 1a**)^{32,33}. To determine the genomic location of enhancers, we generated genome-wide maps of the enhancer-associated histone mark H3K27ac (ChIP-seq), accessible chromatin (ATAC-seq), and gene expression (RNA-seq) from embryonic face tissue for CS18, 19, 22, and 23 (**Supplemental Table 1**). To extend our compendium to earlier stages, we complemented this data with published H3K27ac peaks (ChIP-seq) from CS13-17 human face tissue²⁶ (**Supplemental Table 1, Methods**). In total, we observed 13,983 reproducible human candidate enhancers, as defined by the presence of H3K27ac signal in at least two biological samples at any stage between CS13-23 of development (**Supplemental Table 2**). Of the 10,893 regions marked by H3K27ac at stages CS18-23, 6,718 (61.7%) also

showed accessible chromatin signal, further supporting their expected status as active enhancers (**Supplemental Table 3, Methods**).

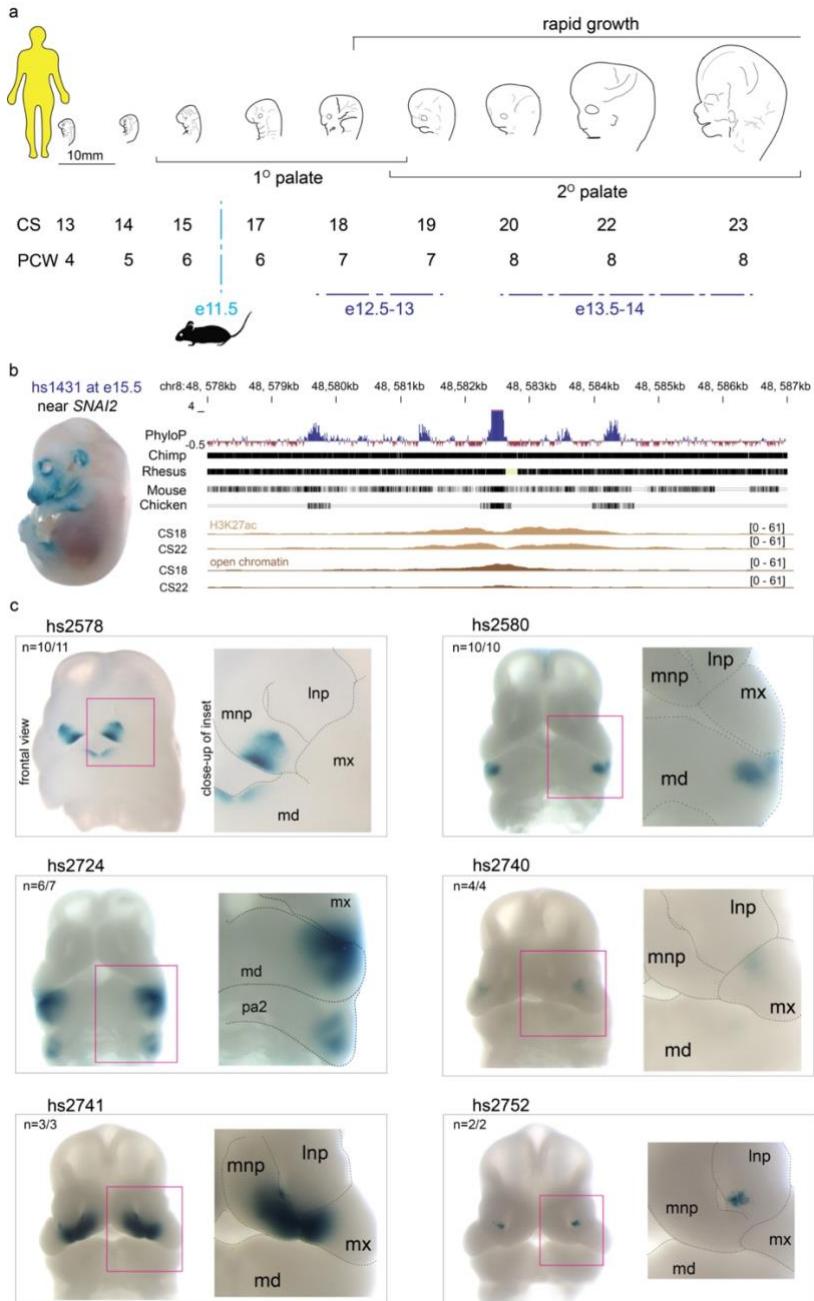


Figure 1. Developmental enhancers in human craniofacial morphogenesis. **a.** Developmental time points coinciding with critical windows of craniofacial morphogenesis are shown by Carnegie stage (CS) and post-conceptional week (PCW) in humans, and comparable embryonic (e) stages for mouse are shown in embryonic days. **b.** Representative embryo image at e15.5 for an *in vivo* validated enhancer (hs1431) shows positive *lacZ*-reporter activity in craniofacial structures (and limbs). Adjacent graphic shows the genomic context and evolutionary conservation of the region, with H3K27ac-bound and open chromatin regions located within the hs1431 element. **c.** Six examples of human craniofacial enhancers discovered in this study with *in vivo* activity validated in e11.5 transgenic mouse embryos. Enhancers hs2578, hs2580, hs2724, hs2740, hs2741 and hs2752 show *lacZ*-reporter activity in distinct subregions of the developing mouse face. Lateral nasal process (lnp), medial nasal process (mnp), maxillary process (mx), and mandibular process (md). n, reproducibility of each pattern across embryos resulting from independent transgenic integration events.

For an initial assessment of the biological relevance of this genome-wide set of predicted human craniofacial enhancers, we compared it with the large collection of *in vivo*-validated enhancers available through the VISTA enhancer browser³⁴. Among the 3,193 elements that have been tested in VISTA to date, we identified 153 cases in which an enhancer predicted through the present human-derived epigenomic dataset had been previously shown to drive reproducible expression in embryonic facial structures including the nose, branchial arches, facial mesenchyme, cranial nerves, melanocytes, ear, or eye (**Extended Data Figure 1, Supplemental Table 4**). A representative example of a previously validated VISTA craniofacial enhancer is shown in **Figure 1b**.

To assess the value of these data for the discovery of additional craniofacial *in vivo* enhancers in the human genome, we tested 60 candidate human enhancers in a transgenic mouse assay (**Methods, Supplemental Table 5**). We identified 16 cases of previously unknown enhancers that showed reproducible activity in craniofacial structures. **Figure 1c** illustrates the rich diversity of craniofacial structures in which these enhancers drive reproducible *in vivo* activity. Examples include enhancers driving expression in restricted subregions of the medial nasal process and mandible (hs2578), the mandible (hs2580), the mandible and second pharyngeal arch (hs2724), the maxillary (hs2740), the medial nasal process and maxillary (hs2741), or the lateral nasal process (hs2752, **Fig. 1c**).

Functional Associations, Developmental Dynamics, and Conservation of Human Craniofacial Enhancers

To further assess the biological relevance of the human candidate enhancer sequences identified by our approach, we examined known functions of their presumptive target genes using rGREAT ontology analysis³⁵. The identified candidate enhancers are enriched near genes implicated in craniofacial human phenotypes, with 9 of the top 15 terms directly related to craniofacial or eye-associated phenotypes (**Figure 2a**), including midface retrusion, reduced number of teeth, and abnormality of maxilla.

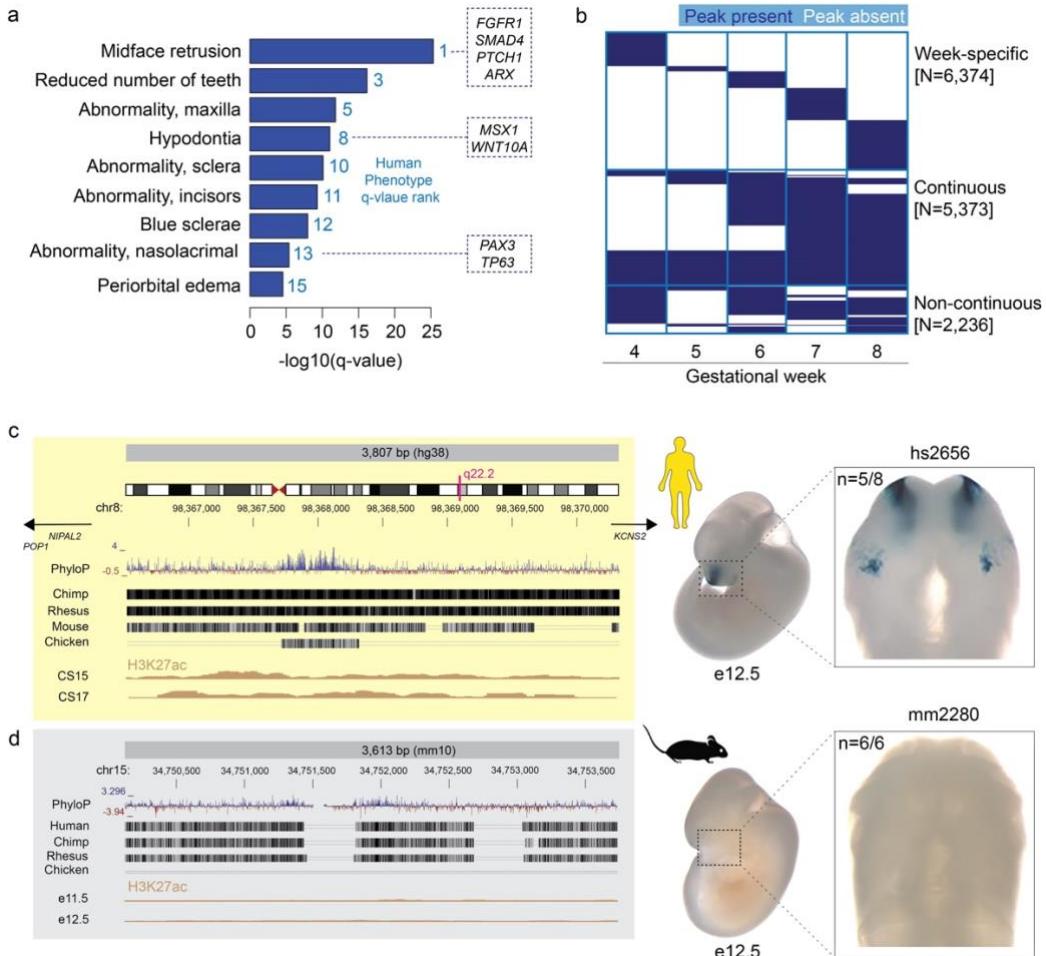


Figure 2. Developmental dynamics and conservation of human craniofacial enhancers. **a.** Results of rGREAT ontology analysis for 13,983 highly reproducible human craniofacial enhancers, ranked by Human Phenotype q-value. The ontology terms indicate that our predictions of human craniofacial enhancers are enriched near presumptive target genes known to play important roles in craniofacial development (examples in boxes). **b.** Predicted activity windows of 13,983 candidate human enhancers (rows) arranged by gestational week 4-8 of human development (columns). Blue, active enhancer signature; white, no active enhancer signature. **c/d.** Left: Genomic position and evolutionary conservation of human candidate enhancer hs2656 (**c**) and its mouse ortholog mm2280 (**d**). The human sequence, but not the orthologous mouse sequence, shows evidence of H3K27ac binding at corresponding stages of craniofacial development (beige tracks). Right: Representative embryo images at e12.5 show that human enhancer hs2656, but not its mouse ortholog mm2280, drives reproducible *lacZ*-reporter expression in the developing nasal and maxillary processes at e12.5. n, reproducibility of each pattern across embryos resulting from independent transgenic integration events.

We also examined the genome-wide set of human craniofacial candidate enhancers for the presence of noncoding variants implicated in inter-individual variation in facial shape and in craniofacial birth defects through genome-wide association studies (GWAS). We aggregated lead SNPs from 41 studies of normal facial variation and craniofacial disease (**Methods**;

Supplemental Table 6). From 1,404 lead SNPs from these studies, we identified 27,386 SNPs in linkage disequilibrium (LD; $r^2 \geq 0.8$) with the lead SNPs for the appropriate populations in the respective craniofacial GWAS. Upon intersection with H3K27ac-bound regions from bulk face tissue between stages CS13-23 (**Figure 1a**), we observed a total of 209 predicted enhancer regions overlapping 605 unique LD SNPs. This includes 43 candidate enhancer regions overlapping with 102 unique disease SNPs, and 176 candidate enhancers overlapping with 515 unique SNPs for normal facial variation (**Supplemental Table 7**).

The activity of individual enhancers can be highly dynamic across developmental stages, supporting that enhancers regulate both spatial and temporal aspects of developmental gene expression^{25,36}. To explore the temporal dynamics of human craniofacial enhancers, we determined the temporal activity profile of all 13,983 human candidate enhancers by week of development, covering gestational weeks 4 to 8 (**Figure 2b; Methods**). We found that a small proportion (1,624 elements or 11.6%) of elements were predicted to be continuously active as enhancers throughout all five weeks. Nearly half (6,347) showed narrow predicted activity windows limited to a single week, while another 3,137 showed continuous activity periods covering a subset of the five weeks. A smaller number of enhancers (2,236) with predicted non-continuous activities likely contains elements with truly discontinuous activity (e.g., in different subregions of the developing face), and elements not reaching significant signal at some stages, e.g., due to changes in relative abundance of cell types. In combination, these data sets provide an extensive catalog mapping the genomic location of human craniofacial enhancers, including their temporal activity patterns during critical stages of craniofacial development.

To assess the conservation of candidate enhancers identified from human tissues in the mouse model, we compared H3K27ac binding data from human developmental stages CS13-23 to previously published results for histone modifications at matched stages of mouse development²⁵. The majority (12,179 of 13,983; 87%) of the human candidate enhancers are conserved to the mouse genome at the sequence level, defined by the presence of alignable orthologous sequence that is syntenic relative to surrounding protein-coding genes. Among

these conserved sequences, 8,257 (59%) showed epigenomic enhancer signatures in the mouse, indicating their functional conservation. The remaining 3,922 (28%) regions were sequence-conserved but showed no evidence of enhancer activity in the mouse tissues examined (**Supplemental Table 8, Methods**), suggesting that they are human-specific and highlighting the potential value of human tissue-derived epigenomic data for human craniofacial enhancer annotation.

To assess whether the differences in epigenomic signatures between human and mouse translate into species-specific differences in *in vivo* enhancer activity, we used a transgenic mouse assay to compare the human and mouse orthologs of a predicted human-specific enhancer. We chose a candidate enhancer located near genes *POP1*, *NIPAL2* and *KCNS2*, located in the 8q22.2 region associated with non-syndromic clefts of the face³⁷ (**Figure 2c/d**). Documented mutations in *POP1* cause Anauxetic Dysplasia with pathognomonic short stature, hypoplastic midface and hypodontia along with mild intellectual disability³⁸⁻⁴⁰. We generated enhancer-*lacZ*-reporter constructs of the human and mouse orthologs of the candidate enhancer region and used CRISPR-mediated transgene insertion at the H11 safe harbor locus^{41,42} to create transgenic mice. Embryos transgenic for the human ortholog (hs2656) show reproducible activity in the developing nasal and maxillary processes at embryonic day (e) 12.5, confirming that the human tissue-derived enhancer signature correctly predicts *in vivo* activity at the corresponding stage of mouse development (**Figure 2c**). In contrast, we did not observe reproducible craniofacial enhancer activity with the mouse orthologous sequence, concordant with the absence of enhancer chromatin marks in mouse at this location (mm2280, **Figure. 2d**).

Transcriptome Landscape of the Mammalian Craniofacial Complex at Single-cell Resolution

To provide a higher-resolution view of the enhancer landscape of craniofacial development, we complemented these detailed maps of human craniofacial enhancers with single cell-resolved data, with the goal to identify the cell type specificity of individual enhancers. Given the genetic heterogeneity, limited availability, and processing challenges associated with early

human prenatal tissues, we performed these studies on mouse tissues isolated from corresponding developmental stages (**Figure 3**).

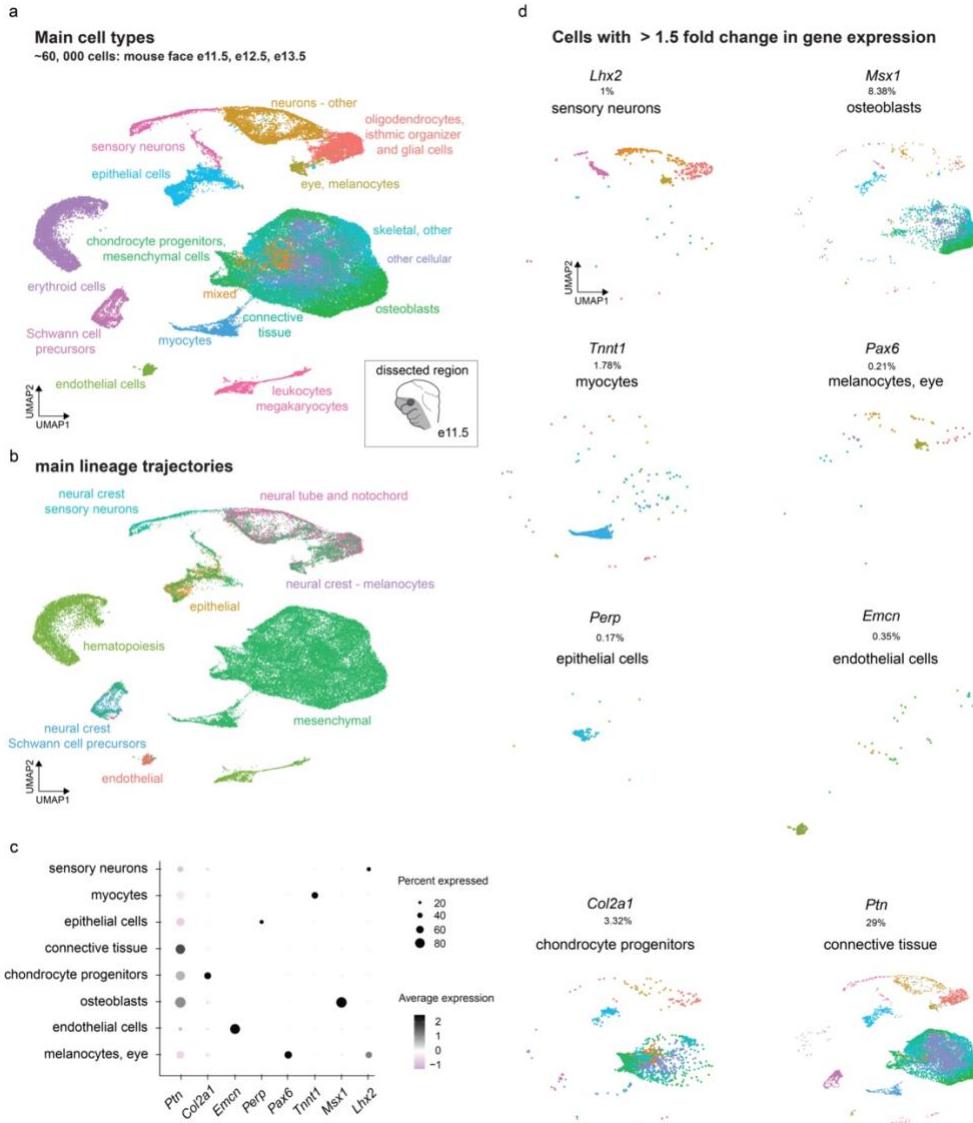


Figure 3. Gene expression in the mammalian craniofacial complex at single cell resolution. **a.** Uniform Manifold Approximation and Projection (UMAP) clustering, color-coded by inferred cell types across clusters from aggregated scRNA-seq for the developing mouse face at embryonic days 11.5–13.5, for 57,598 cells across all stages. Cartoon shows the outline of dissected region from the mouse embryonic face at e11.5, corresponding regions were excised at other stages. **b.** Same UMAP clustering, color-coded by main cell lineages. **c.** Expression of select marker genes in cell types shown in (a). **d.** UMAP plots comprising cells with >1.5-fold gene expression for marker genes representing specific cell types as shown in (a) and (c).

We generated a detailed transcriptome atlas from relevant stages of development and analyzed mouse facial tissue isolated from e11.5, e12.5, and e13.5 by single-cell RNA-seq (see **Methods**). Applying Uniform Manifold Approximation and Projection (UMAP) non-linear dimensionality reduction for unbiased clustering resulted in 42 primary detectable clusters

(Extended Data Figures 2-4, Supplemental Tables 9-10). We analyzed 57,598 cells with a median of 1,659 genes expressed per cell. We systematically assigned cell type identities to the resulting clusters (**Extended Data Figures 5 and 6, Supplemental Tables 11-12**, and **Methods**) in our final Single-cell annotated Face eXpression dataset (henceforth referred to as *ScanFaceX*), which includes 16 annotated cell types capturing the developing mammalian face and associated tissues (**Figure 3a**). Trajectory analyses using Seurat recapitulated the main lineages including epithelial, mesenchymal, endothelial, and neural crest-derived cell types including melanocytes relevant to face development (**Figure 3b**). The final annotated cell type clusters showed strong cluster-specific expression of established markers genes relevant to craniofacial development such as *Col2a1* (chondrocyte progenitors)⁴³⁻⁴⁵, *Msx1* (osteoblasts)⁴⁶⁻⁴⁸, *Perp* (epithelial cells)^{49,50}, *Emcn* (endothelial cells)^{51,52}, *Lhx2* (sensory neurons)^{53,54}, *Pax6* (melanocytes)^{55,56}, *Tnnt1* (myocytes)⁵⁷, and *Ptn* (connective tissue)⁵⁸ (**Figure 3c and 3d, Extended Data Figure 7**). These benchmarking results indicate that *ScanFaceX* provides an accurate single-cell transcriptome reference for relevant stages of craniofacial development that can serve as a foundation for integration with other chromatin data types.

Differential Chromatin Accessibility Correlates with Cell Type-specific Signatures

To identify developmental enhancers at single-cell resolution, we performed single-nucleus ATAC-seq (snATAC-seq)⁵⁹ on mouse face embryonic tissues at select developmental time points (**Figure 4**). Across all stages analyzed, 41,483 cells that passed all quality control steps were considered in the final analysis, and their unbiased clustering resulted in 20 discernable clusters (see **Methods**). Out of a total of 115,521 open chromatin regions in the snATAC-seq data, we observed 16,564 differential accessible regions (DARs) across 20 separate clusters, indicating that each of the clusters representing identical or very similar cell types have distinct open chromatin signatures (**Extended Data Figure 8, Supplemental Table 13**). Next, we integrated our single-cell open chromatin data with the cell type annotations from *ScanFaceX* single-cell transcriptome data using Seurat-based label transfer (see **Methods**). Upon integration, a substantial subset of DARs (10,038 out of 16,564; 60%) were retained, and

developing craniofacial cell types including chondrocytes, osteoblasts, myocytes and connective tissue, epithelial cells, melanocytes, and sensory neurons showed high correlation between the two data types (**Figure 4a-b, Extended Data Figures 9 and 10; Methods**). Chromatin accessibility at putative distal enhancer regions as well as transcription start sites showed distinct cell type specificity. For example, the representative intergenic region near *Isl2* and *Scaper*, and an intronic region of *Lrrk1* differentially active in clusters representing sensory neurons and/or epithelial cells, illustrate the resolution of our data relative to previously available predictions from bulk face tissue^{25,60,61} (**Figure 4c**). These predictions of cell type specificity are well aligned with known functions of the respective target genes. *Isl2* has been shown to be selectively expressed in a subset of retinal ganglion cell axons that have important functions in binocular vision⁶². Allelic variants and mutations in *SCAPER* cause intellectual disability with retinitis pigmentosa in humans⁶³⁻⁶⁵. The *Lrrk1* intronic element is near *Aldh1a3*, a gene adjacent to *Lrrk1*; mutations in the orthologous *human ALDH1A3* cause an autosomal recessive form of isolated microphthalmia⁶⁶⁻⁶⁹. These putative enhancer regions near *Isl2* and *Scaper*, and in the intron of *Lrrk1* drive reproducible *lacZ*-reporter activity in the developing mouse face in anatomical regions that are consistent with neuronal and epithelial cell types (**Figure 4c**). Both *Isl2* and *Aldh1a3* are highly expressed in sensory neurons and epithelial cell clusters, respectively, in *ScanFaceX* data (**Figure 4c**). In an additional example, an enhancer near the promoter region of *Mymx*, which is exclusively active in the myocyte cluster, coincides with *Mymx* expression in myocytes in *ScanFaceX* (**Extended Data Figure 11**).

To facilitate utilization of the full set of genome-wide, cell type-resolved enhancer predictions, we used these mouse tissue-derived single-cell enhancer predictions in combination with our human bulk tissue-derived enhancer catalog, to generate a Single-cell annotated Face eNhancer (*ScanFaceN*) catalog of human enhancer regions with predicted activity profiles across craniofacial cell types (**Supplemental Tables 14-16**). The majority (7,899 of 13,983; 56%) of human tissue-derived facial candidate enhancers overlap with an accessible

chromatin region in at least one cluster of our *ScanFaceN* catalog, and 2,339 (30%) of these regions overlap with DARs in *ScanFaceN*.

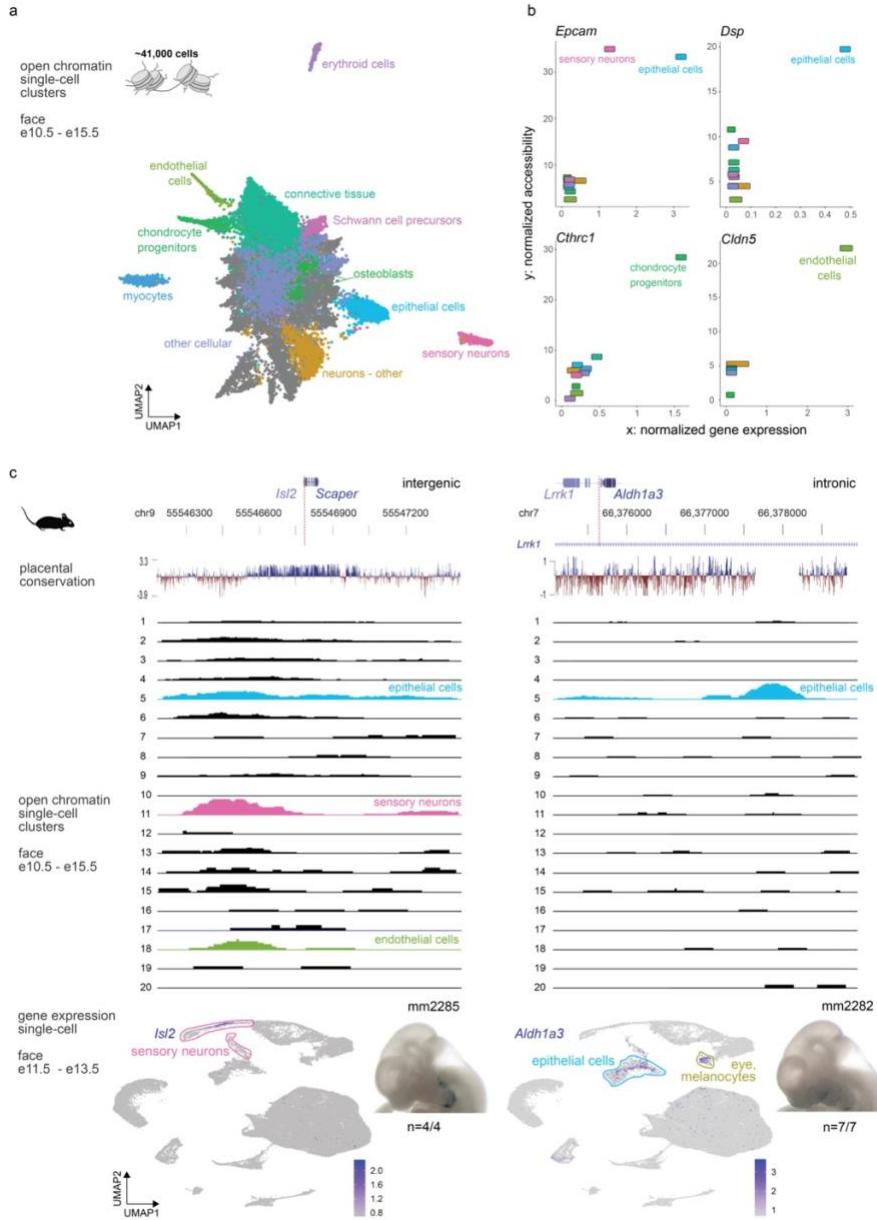


Figure 4. Differential chromatin accessibility at craniofacial *in vivo* enhancers correlates with cell type-specific expression of nearby genes. **a.** Unbiased clustering (UMAP) of open chromatin regions from snATAC-seq of the developing mouse face for stages e10.5-15.5 for approximately 41,000 cells. The cell types are assigned based on label transfer (Seurat) from cell-type annotations of the *ScanFaceX* data. **b.** Correlation between normalized gene expression (x-axis) from *ScanFaceX* and normalized accessibility (y-axis) from snATAC-seq for select genes (*Epcam*, *Dsp*, *Cthrc1*, *Cldn5*) and their transcription start sites with the highest correlation evident in relevant cell types. **c.** Genomic context and evolutionary conservation (in placentals) for corresponding regulatory regions in the vicinity of the *Isl2/Scaper* locus, and an intronic distal enhancer within *Lrrk1*. Tracks for individual snATAC-seq clusters from developing mouse face tissue (e10.5 to e15.5), with cluster-specific open chromatin signatures for relevant annotated cell types are shown for the same genomic regions. UMAP of *ScanFaceX* data shows expression of *Isl2* and *Alldh1a3* (gene adjacent to *Lrrk1*) in expected cell-types. Images for a representative mouse embryo at e11.5 for both loci show validated *in vivo* lac-Z-reporter activity of the respective regions. n, reproducibility of each pattern across embryos resulting from independent transgenic integration events.

Cell Type-resolved Enhancer Predictions and *in vivo* Enhancer Activity Patterns

To explore the relationship between predicted cell type specificities of enhancers and their respective spatial *in vivo* activity pattern during craniofacial development, we intersected the *ScanFaceNDARs* from the 11 main *ScanFaceX*-matched clusters with craniofacial enhancers validated *in vivo* and curated in the VISTA Enhancer Browser³⁴ (**Figure 5a**). We observed general correlations between cluster-specific accessibility and spatial *in vivo* patterns among 77 enhancers that showed chromatin accessibility in at least one of the 11 main clusters. For example, the connective tissue-mesenchymal cluster (cluster 2) of the craniofacial snATAC-seq tends to group VISTA enhancers with activity specific to the branchial arches, while the chondrocyte cluster (cluster 13) has multiple VISTA enhancers with activity in the mid-face, paranasal regions, and/or the region at the junction of the developing forebrain and nasal prominences that are consistent with developing cartilaginous regions of the face (**Figure 5b**). Despite these broad correlations, we observed considerable heterogeneity of spatial patterns within most clusters, underscoring the spatiotemporal complexity of craniofacial morphogenesis, which relies on cell type-specific regulatory programs in combination with highly regionalized regulatory cues.

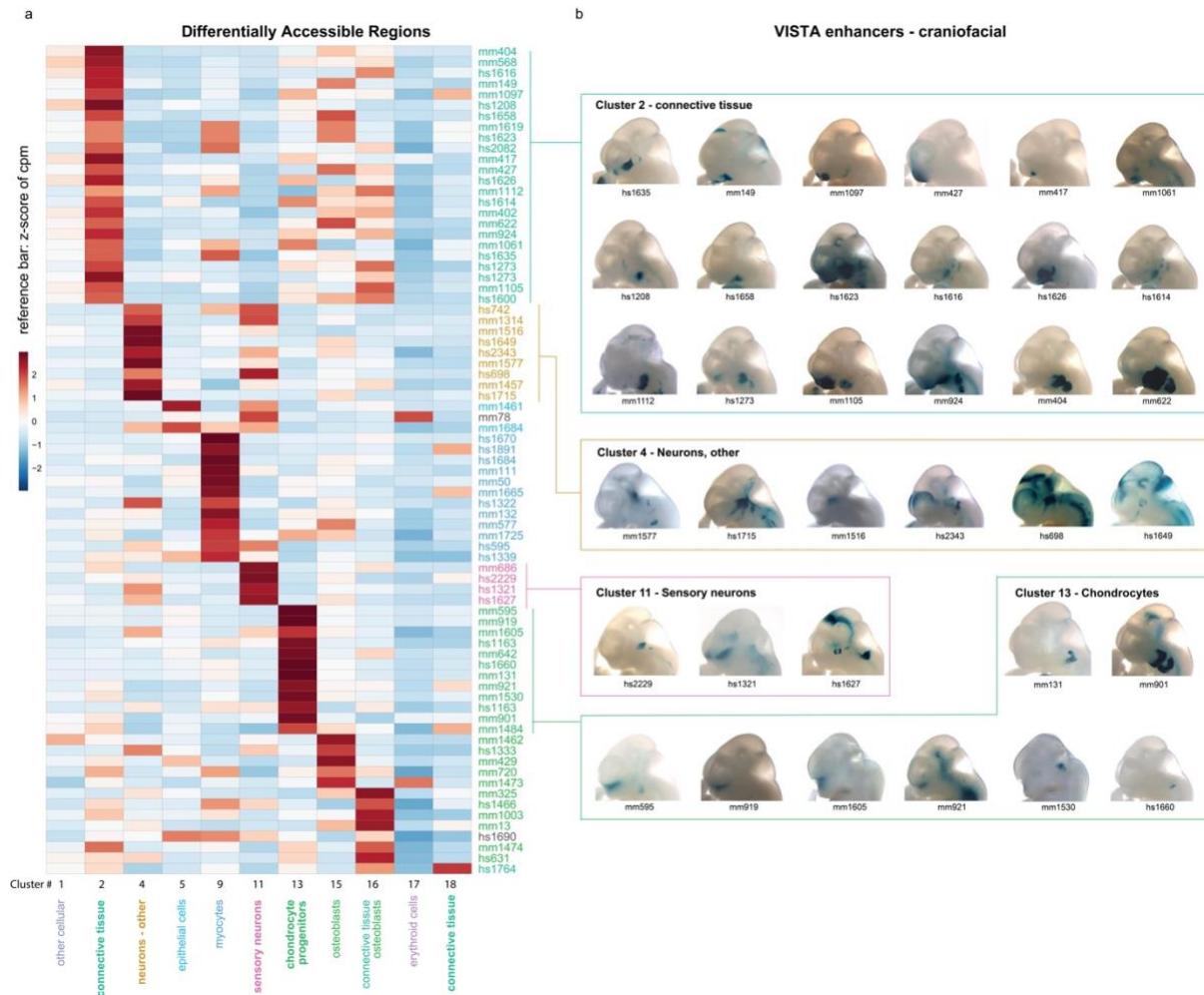


Figure 5. Cell type-specific chromatin accessibility of craniofacial *in vivo* enhancers. **a**, Heatmap indicates the chromatin accessibility of 77 craniofacial *in vivo* enhancers in 11 major cell type clusters. cpm: counts per million. **b**, Representative images of transgenic embryos from VISTA Enhancer Browser, showing *in vivo* activity pattern of 35 selected enhancers at e11.5. Embryo images are grouped by example cluster-types from (a) in this retrospective assignment.

Cell Type-specific Enhancer Activity at Single-cell Resolution

To explore whether craniofacial enhancer activity can be quantitatively assigned to specific cell types *in vivo*, we generated transgenic mice in which selected craniofacial enhancers were coupled to a fluorescent *mCherry* reporter gene (**Figure 6a**). We examined three different craniofacial enhancers (hs1431, hs746 and hs521), two of which (hs1431 and hs746) we previously demonstrated to be required for normal facial development¹ (**Figure 6b**). In all cases, we isolated craniofacial tissue from transgenic reporter embryos at e11.5 and performed scRNA-seq (**Figure 6a**). For hs1431, near *Snai2*, which is active across many regions of the developing face, *mCherry* expression is observed across almost all cell type

clusters, indicating that hs1431 is broadly active across multiple cell types during craniofacial development (**Figure 6c**). In contrast, hs746 which is in the vicinity of *Msx1*, is primarily active in osteoblasts and in a subset of cells expressing *Msx1*, a gene previously shown to regulate the osteogenic lineage⁷⁰. Enhancer hs521, located near *Gbx2*, is primarily active in a subset of mesenchymal cells and chondrocyte progenitors, and its activity coincides with a subset of cells expressing *Gbx2* (**Figure 6c**), a gene known to be active in the developing mandibular arches⁹. Together, these data illustrate how purpose-engineered enhancer-reporter mice can be used to validate and further explore the *in vivo* activity patterns of craniofacial enhancers identified through genome-wide single-cell profiling studies.

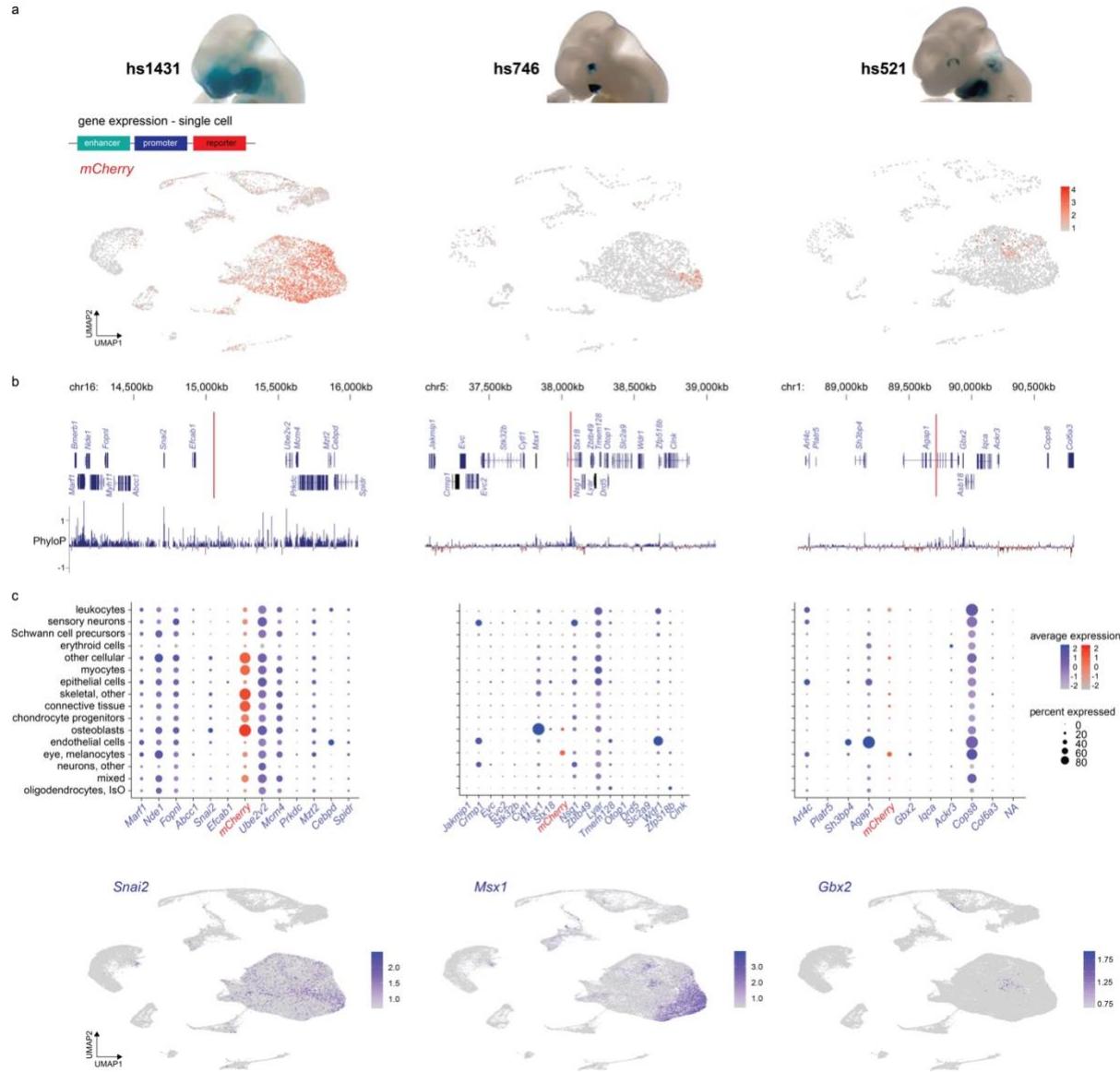


Figure 6. Cell type-specific enhancer activity at single-cell resolution. **a.** *in vivo* activity pattern of select craniofacial enhancers (hs1431, hs746, hs521) at e11.5, visualized by *lacZ*-reporter assays (top). In separate experiments, the same enhancers were coupled to an *mCherry*-fluorescent reporter gene and examined by scRNA-seq of craniofacial tissues of resulting embryos. UMAPs show enhancer-driven *mCherry* expression (see Fig. 3a for reference). **b.** Location of enhancers hs1431, hs746 and hs521 in their respective genomic context (red vertical lines), along with protein-coding genes within the genomic regions and local conservation profile (PhyloP). **c.** Average expression of genes (Seurat) in the vicinity of the respective enhancers, and proportion (percent) of cells expressing the genes in specific cell types. Enhancer-driven *mCherry* signal is plotted in the center *in lieu* of the approximate enhancer location in its endogenous genomic context. Bottom panels show expression of *Sna2*, *Msx1*, and *Gbx2* as likely candidate target genes for each of the enhancers hs1431, hs746 and hs521 across UMAPs. IsO: Isthmic Organizer Cells.

Discussion

The lack of data from primary tissues and incomplete mapping of human developmental enhancers in craniofacial morphogenesis has been a challenge in the systematic assessment of the role of enhancers in craniofacial development and disease. In the present study, we

have generated human bulk and mouse single-cell data to create a comprehensive compendium of enhancers in human and mouse development, including temporal profiles and predictions of cell type specificity. We identify major cell populations of the developing mammalian face, along with corresponding genome-wide enhancer profiles. We also show that while many enhancers are functionally conserved between human and mouse, additional human-specific enhancers that show no functional conservation in mice can be identified by profiling human tissues. Our data illustrate the considerable temporal dynamics of human craniofacial enhancers, a critical aspect for understanding the developmental timing of enhancer activity related to specific phenotypes such as clefts and mid-facial deformities. As clinical sequencing becomes increasingly common and accessible to both patients and the medical community, our data may serve as an essential resource to address the gaps in understanding the potential pathogenicity of regulatory variants.

The single-cell resources generated through this study, *ScanFaceX* for expression and *ScanFaceN* for enhancers, contain a total of 115,521 candidate enhancers as defined by chromatin accessibility, including 10,038 that show differential chromatin accessibility across annotated cell types in face morphogenesis. We demonstrated how engineered mice can be used to study these enhancers *in vivo* at single-cell resolution. Using a transgenic reporter assay coupled to single-cell RNA-seq, we defined the cell type-specific activity of three craniofacial enhancers during embryonic development. This approach illustrates how these methods can be combined to determine the *in vivo* cell type specificity of individual enhancers and relate their activity to cell type-specific expression of their putative target genes. All of these data are also available in FaceBase and the VISTA Enhancer Browser for community use^{1,61,71}. In summary, our work provides a multifaceted and expansive resource for studies of craniofacial enhancers in human development and disease.

Acknowledgements

This work was supported by U.S. National Institutes of Health (NIH) grant to A.V (R01DE028599 and U01DE024427). Research was conducted at the E.O. Lawrence Berkeley National Laboratory and performed under U.S. Department of Energy Contract DE-AC02-05CH11231, University of California (UC). The human embryonic face samples were provided by the Joint MRC / Wellcome Trust (MR/R006237/1, MR/X008304/1 and 226202/Z/22/Z) Human Developmental Biology Resource⁷² (www.hdbr.org). F.D. was supported by the Swiss National Science Foundation grant P400PB_194334.

Declaration of Interests

Bing Ren is a co-founder of Arima Genomics, Inc, and Epigenome Technologies, Inc.

Author Contributions

S.S.R., D.E.D., L.A.P., and A.V. designed the study. S.S.R., C.S., M.O., Y.Zhu, H.W., S.Y.A., J.A.A., V.A., S.T., I.P-F., C.S.N., M.Kato., R.H., K.V.M., A.W., L. L., S.P. performed experiments. J. A. A. performed imaging. S.S.R., K.P., M.L.A., M. Kosicki, L.E.C., F.D., M.B., G.K., I.B., Y.F-Y. analyzed data. S.S.R., L.P., and A.V. wrote the manuscript with input from the remaining authors. We thank Yoon Gi “Justin” Choi of the University of California, Berkeley QB3 Genomics Core for technical assistance with the 10X Genomics set up.

Methods

Human Subjects

All aspects of this study involving human tissue samples were reviewed and approved by the Human Subjects Committee at Lawrence Berkeley National Laboratory (Protocol Nos. 00023126 and 00022756). Human embryonic face samples were obtained from the Human Developmental Biology Resource at Newcastle University (HDBR, hdbr.org), in compliance with applicable state and federal laws and with fully informed consent. No identifying information for human samples was shared by HDBR. All embryonic samples for which primary data was generated in this study were verified to be karyotypically normal by HDBR. Primary data from embryonic whole face samples at post-conception weeks 7 and 8 were generated for this study. All embryonic samples were shipped on dry ice and stored at -80°C

until processed. ChIP-seq data for three samples at Carnegie stage (CS)18, one sample at CS 19, two samples at CS22 and one sample at CS23 are presented in this study, along with accompanying ATAC-seq data for two samples at CS18, one sample at CS19, one sample each at CS22 and CS23. Processed data for CS 13-16 was obtained from previously published studies²⁶ and included in our downstream integrative analyses. All datasets are listed in **Supplemental Table 1**.

Animal Studies and Experimental Design *in vivo*

All animal work was reviewed and approved by the Lawrence Berkeley National Laboratory Animal Welfare Committee. Mice used for this study were housed at the LBNL Animal Care Facility, which is fully accredited by AAALAC International. All mice were routinely health checked and monitored daily for food and water intake. Mice (*Mus musculus*; FVB strain) across developmental stages from embryonic day 10.5 through 15.5 were used in this study. Animals of both sexes were used in the analysis. Sample size selection and randomization strategies were followed based on our previous experience of performing transgenic mouse assays for ~3000 published enhancer candidates^{41,42}. Mouse embryos that lacked the reporter transgene or were not at the correct developmental stage were not included further in analysis.

Transgenic Mouse Assays *in vivo*

Transgenic enhancer-reporter assays were performed as previously described^{41,42}. Briefly, a minimal *Shh* promoter and reporter gene were integrated into a non-endogenous, safe harbor locus⁴² in a site-directed transgenic mouse assay. The selected genomic region was PCR amplified from human or mouse genomic DNA where applicable; the PCR amplicon was cloned into a *lacZ*-reporter vector (Addgene #139098) or *mCherry* reporter vector (available upon request) using Gibson assembly (New England Biolabs)⁷³. The final transgenic vector consists of the predicted enhancer–promoter–reporter sequence flanked by homology arms intended for the *H11* locus in the mouse genome. Sequence of the cloned constructs was confirmed with Sanger sequencing or MiSeq. Transgenic mice were generated using pronuclear injection, as previously described⁴². Embryos were collected for various

downstream experiments at embryonic days 10.5 through 15.5. Beta-galactosidase staining was performed in our standardized pipeline with the following modification. Embryos were fixed with 4% paraformaldehyde (PFA) for 30 minutes for E11.5 embryos, respectively, while rolling at room temperature. The embryos were genotyped for presence of the transgenic construct. Embryos positive for transgene integration into the *H11* locus and at the correct developmental stage were considered for comparative reporter gene activity across respective stages and were imaged on a Leica MZ16 microscope. Genomic coordinates for VISTA enhancer hs2656 (**Figure 2**); enhancer mm2280 (**Figure 2**), mm2282 and mm2285 (**Figure 4**), and mm2281 (**Extended Data Figure 11**) are shown in **Supplemental Table 5** and **17** respectively.

For transgenic experiments demonstrating cell type-specific enhancer activity at single-cell resolution and involving hs1431, hs746 and hs521 (**Figure 6**), a combination of *Hsp68* promoter and *mCherry* reporter were used.

ChIP-seq

Chromatin immuno-precipitations were performed as previously described ⁷⁴. Briefly, frozen and non-cross-linked face tissue was dissociated in PBS by pipetting until homogenized and cross-linked with 1% formaldehyde at room temperature. Cells were lysed and chromatin was sonicated using a Biorupto device (Diagenode) to obtain fragments with an average size ranging between 100–600 bp. Input sample was set aside and stored appropriately, Protein A and G Dynabeads (Invitrogen) were added to the sample, and chromatin was incubated for 2h at 4°C with 5 µg of anti-H3K27ac antibody (Active Motif, Cat# 39133, Lot 01613007). Immuno-complexes were sequentially washed, and the immunoprecipitated DNA complexes were eluted in an SDS buffer at 37°C for one hour. Samples were reverse-crosslinked with Proteinase K overnight at 37°C. DNA was purified with a ChIP DNA clean concentrator (D5205 Zymo Research), and a KAPA SYBR Green qPCR mix was used to assess presence of H3K27 acetylated regions versus negative control regions. DNA was quantified using Qubit, and size distribution and DNA concentration of the samples were assessed on the Agilent Bioanalyzer. Illumina TruSeq library preparation kit was used for downstream library

preparation, and libraries were sequenced as single-end 50 bp reads on an Illumina HiSeq 2500.

ChIP-seq data was analyzed using the ENCODE histone ChIP-seq Unary Control Unreplicated pipeline (<https://www.encodeproject.org/pipelines/ENCPL841HGV/>) implemented at DNAnexus (<https://www.dnanexus.com>). Briefly, reads were mapped to the human reference genome version hg38 using BWA (v0.7.7) and sorted bam file generated using samtools (v0.1.19). Peak calling was performed using MACS2 (v2.2.4; --broad flag, q-value < 0.05), and overlapping peaks identified using overlap_peaks.py. A combined peak set was called by merging peaks from all samples. Merged peaks within 1kb of transcription starts sites as defined by GENCODE were removed, resulting in 70,075 distal peaks. Of those, 13,983 peaks were present in at least two samples in each embryonic week which were retained for final analysis.

ATAC-seq

Embryonic samples were processed for ATAC-seq as previously described ⁷⁴. In short, harvested tissues were lysed, centrifuged for 10min at 500 x g, at 4°C, and the resulting cell pellet was treated with the Nextera DNA transposase Tagment DNA Enzyme (Catalog number: 20018705) and the transposed DNA was eluted using Qiagen MinElute PCR purification kit. Samples were then PCR amplified using the NEB Next High-Fidelity 2xPCR Master Mix (catalog number: NEBE6040SEA) with Nextera PCR primers 1 (AATGATACGGCGACCACCGAGATCTACACNNNNNNNTCGTCGGCAGCGTC) and 2 (CAAGCAGAAGACGGCATACGAGATNNNNNNNGTCTCGTGGGCTCGG), and DNA was purified as described above. The eluted library was analyzed for quality in a Bioanalyzer High Sensitivity assay and samples were subsequently deep sequenced on an Illumina HiSeq2500. ATAC-seq data was analyzed using the ENCODE ATAC-seq (unreplicated) pipeline (<https://www.encodeproject.org/pipelines/ENCPL344QWT/>).

RNA-seq

Samples were processed for RNA-seq and libraries were generated as previously described^{74,75}. Briefly, RNA was isolated from the dissociated face tissue using TRIzol

Reagent (Life Technologies), all samples were DNase-treated (TURBO DNA-free Kit, Life Technologies), and assessed for quality (RNA 6000 Nano Kit , Agilent) on a 2100 Agilent Bioanalyzer. TruSeq Stranded Total RNA with Ribo-Zero Human/Mouse/Rat kit (Illumina) was used to prepare RNA-seq libraries according to manufacturer's protocol. RNA-seq libraries were depleted of high molecular weight products in an Illumina Resuspension Buffer and by incubating in 60 µL Agencourt AMPure XP beads for 4 min. AMPure beads were pelleted, washed twice with 80% ethanol and the DNA was eluted per manufacturer's instructions. RNA concentration and quality of the RNAseq libraries were assessed using a 2100 Bioanalyzer with the High Sensitivity DNA Kit (Agilent), and libraries were sequenced as single-end 50 bp reads on an Illumina HiSeq 2500.

RNA-seq data was analyzed using the ENCODE RNA-Seq (Long) Pipeline-1 replicate pipeline (<https://www.encodeproject.org/pipelines/ENCPL002LSE/>) implemented at DNAnexus (<https://www.dnanexus.com>). Briefly, reads were mapped to the reference genome using STAR align (V2.12). Genome wide coverage plots were generated using bam to signals (v2.2.1). Gene expression counts were generated using RSEM (v1.4.1). Human datasets were analyzed using human reference genome version hg19, and GENCODE v24 gene annotations. Mouse datasets were analyzed using mouse reference genome version mm10 and GENCODE M4 gene annotations.

GWAS Data

The NHGRI-EBI Catalog of Genome-wide association studies⁷⁶ was mined for studies with the following keywords: craniofacial, face, cleft lip, cleft palate, microsomia, salivary, taste, and tooth. The compiled studies comprised of diverse populations and ethnicities ranging from those belonging to the United States, Europe, Taiwan, China, Singapore, Korea and the Philippines, Brazil, Spain, Latin Americas, Uyghurs as well as admixed populations. For data published in the catalog by early 2022, we aggregated 41 studies representing normal facial variation as well as dento-oro-craniofacial disease. The SNiPA tool⁷⁷ was used for querying SNPs in linkage disequilibrium ($r^2 \geq 0.8$) with the lead SNPs for the appropriate populations for the respective GWAS. This compilation of GWAS (**Supplemental Tables 6-7**) was intersected

with 13,983 highly reproducible human enhancers derived from primary embryonic bulk face between CS13-23.

Intersecting VISTA Catalog with Predicted Craniofacial Enhancers

We intersected the 3,193 curated enhancers in the VISTA Enhancer Browser with 13,983 reproducible human candidate enhancers from this study or 10,038 mouse DARs, requiring a minimum 100bp overlap. We used liftOver to obtain the genomic coordinates in species relevant assembly (**Extended Data Figure 1**, and **Supplemental Table 4**).

Single-cell RNA-seq

Transgenic embryos were harvested at the determined developmental stage, between 11.5 - 13.5 dpc (8 samples at e11.5, 1 sample at e12.5, and 4 samples at e13.5), and examined for positive *mCherry* signal. Embryos positive for *mCherry* reporter activity showed reproducible and comparable enhancer-reporter expression as seen in the *lacZ* expression patterns for VISTA enhancers hs1431, hs521 and hs746 used in this study. Embryos were consistently kept in ice-cold PBS until dissection. Upon fluorescent screening, developing face tissue was dissected with the aid of a Leica MZ16 microscope, and immediately processed for downstream experiments. Fresh mouse embryonic face tissue was mechanically dissociated by pipetting gently into a single-cell suspension using Accumax, assessed for viability of cells and cell density using Trypan Blue staining. Individual cells were quantified, spiked with 10% HEK293T/17 frozen-thawed cells, and processed using the 10X Genomics Chromium Next GEM Single Cell 3' protocol including transcript capture and library preparation for single-cell gene expression. Samples were either processed individually or pooled using a previously described Multi-seq strategy⁷⁸ upstream of the 10X Genomics Chromium protocol. The resulting libraries were sequenced on an Illumina HiSeq2500 or NovaSeq 10X. BCL files from Illumina were processed into FASTQ format, individual sample libraries were de-multiplexed as necessary, reads were aligned to mm10 reference genome where *mCherry* sequence was added as an additional chromosome. Cell Ranger 3.1.0 software was used to process the raw sequence files and generate feature-barcode matrices. After correcting for batch effects, data

from all libraries was aggregated into a single R object file using the 10X Genomics Cell Ranger 3.1.0. Seurat v3.2 guided clustering tutorial was used for formal downstream analyses⁷⁹⁻⁸¹. Adhering to the standard pre-processing workflow and quality control, cells with unique feature counts between 200-2000 and < 5%mitochondrial reads were retained. Normalization, feature selection, scaling, dimensional reduction, clustering and finding cluster biomarkers i.e., differentially expressed features were performed as guided. Our final Seurat/clustered UMAP consists of a 25,645 feature by 57,598 cell matrix.

Assigning cell-type identity to scRNA-seq clusters: We systematically assigned cell type identities to the clusters in our craniofacial scRNA-seq dataset using two computational methods. (i) Using our primary single cell dataset as query, we assigned cell type identities by Seurat-based automated reference mapping to a previously published large single-cell gene expression dataset⁸² of whole mouse embryonic development for stages e9.5-13.5, the reference was down sampled to 100K cells for efficient processing and retained all 38 broad cell types originally described. 27 cell types from the reference were summarily mapped in our craniofacial scRNA-seq dataset by Seurat's label transfer; the referenced cell types showed a good overall correlation with the cell types associated with the top 20 marker genes in most clusters in our *ScanFaceX* dataset. (ii) In parallel, we used the scoreMarkers wrapper function previously described in the *scran* package which uses effect sizes (Cohen's *d* statistic) to perform differential expression to list marker genes for each of the clusters in a scRNA-seq dataset⁸³. These marker gene sets were tested for enrichment of Gene Ontology (GO) biological process terms by performing a hypergeometric test to identify GO terms overrepresented in our *ScanFaceX* dataset. Cell-type annotations from methods (i) and (ii) described above were compared and resulted in each cluster in the *ScanFaceX* dataset having one or more cell-type annotations. Finally, cell clusters that showed similar or close cell-type specific signatures were manually merged to reflect 16 formal annotations for definitive cell types capturing craniofacial development and morphology (**Supplemental Tables 11-12**).

Single-nucleus ATAC-seq

Face tissue was dissected from mouse embryos for each of the developmental stages e10.5-15.5, and flash frozen and stored at -80°C until ready to process. Tissue was transported to the Center for Epigenomics, University of California, San Diego School of Medicine, La Jolla, CA for processing using a combinatorial indexing-assisted single nucleus ATAC-seq previously described⁵⁹. Briefly, nuclei were isolated and permeabilized in optimized conditions, pelleted and suspended in resuspended in 500µL high salt tagmentation buffer. Nuclei were counted using a hemocytometer and 2,000 nuclei were dispensed into each well of a 96-well plate per sample. A BenchSmart™ 96 (Mettler Toledo) was used to add 1µL barcoded Tn5 transposomes to each of the wells in the 96-well plate, the mix was incubated for 60 min at 37 °C with shaking (500 rpm). EDTA at a final concentration of 20mM was then added to each well for incubation at 37 °C for 15 min with shaking (500 rpm) to terminate the Tn5 reaction. Next, nuclei were suspended in 20 µL of 2x sorting buffer (2 % BSA, 2 mM EDTA in PBS), wells for each sample were combined and stained with Draq7 at 1:150 dilution (Cell Signaling). 20 nuclei per sample were sorted per well into eight 96-well plates (total of 768 wells) in 10.5 µL of Elution Buffer (25 pmol primer i7, 25 pmol primer i5, 200 ng BSA (Sigma) using a Sony SH800. A Biomek i7 Automated Workstation (Beckman Coulter) was used for performing downstream steps. Samples were incubated at 55 °C for 7 min with shaking (500 rpm) in 1 µL 0.2% SDS, followed by addition of 12.5% Triton-X to quench the SDS. Samples were PCR-amplified (12.5 µL NEBNext High-Fidelity 2x NEB PCR Master Mix; [72 °C 5 min, 98 °C 30 s, (98 °C 10 s, 63 °C 30 s, 72°C 60 s) × 12 cycles, held at 12 °C]). Wells were combined post-PCR. A manual MinElute PCR Purification Kit (Qiagen) along with a vacuum manifold (QIAvac 24 plus, Qiagen) was used for library purification, and size selection was performed with SPRISelect reagent (Beckmann Coulter, 0.55x and 1.5x). A Qubit fluorimeter (Life Technologies) was used to quantify the libraries and the nucleosomal pattern of fragment size distribution was verified on a High Sensitivity D1000 Tapestation (Agilent). Libraries were sequenced on a NextSeq500 or HiSeq4000 (Illumina) using custom sequencing primers.

Reads were aligned to mm10 reference genome using bowtie2 with default parameters and cell barcodes were added as a BX tag in the bam file. Only primary alignments were kept. Duplicated read pairs were removed with Picard, and proper read pairs with insert size less than 2000 were kept for further analysis.

Clustering and cell-type annotation: snapATAC package was used to perform read counting and cell clustering for both all-tissue clustering and tissue-level clustering⁸⁴. First, we removed nuclei with less than 400 fragments or TSS enrichment < 4 for all tissues and calculated a cell-by-bin matrix at 5000-bp resolution for every sample independently, binarized the matrices and subsequently merged them for each clustering task. Next, we filtered out any bins overlapping with ENCODE blacklist (mm10, <http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10-mouse/mm10.blacklist.bed.gz>). We then normalized the read coverage of all bins with $\log_{10}(\text{count}+1)$ and Z-score transformation, and only removed bins with absolute Z scores higher than 2. After these filtering steps, we calculated Jaccard Index and performed dimensional reduction using the runDiffusionMaps function on similarity matrices. The memory usage of the matrices scales quadratically with the number of nuclei. Therefore, we sampled a subset of 30,000 “landmark” nuclei to compute the matrices and then extended to the rest of the cells. After dimensional reduction, we selected top 20 eigenvectors based on the variance explained by each eigenvector and computed 20 nearest neighbors for each nucleus and applied the Leiden algorithm to define 20 clusters.

To perform label transfer from the scRNA-seq to the corresponding snATAC-seq data we first created a gene activity matrix from the snATAC-seq data using accessibility in TSS and gene bodies with the SnapATAC package. We then converted our gene activity matrix into a Seurat object and used default parameters for the Seurat function FindTransferAnchors to perform canonical correlation analysis on the gene activity matrix along with the gene expression quantification from the scRNA-seq data. Finally, we used the TransferData function to annotate the snATAC-seq data via label transfer.

For the scatter plots showing normalized accessibility versus gene expression (**Figure 4b**), we used a gene by cell matrix which has counts for reads at the TSS and the gene body of each gene. To explore the accessibility for VISTA elements hs1431 and hs746, we generated UMAP plots using accessibility at the corresponding open chromatin region (overlapping peak call). For hs521, we assigned accessibility at the nearest peak as there was no overlapping open chromatin region called. The plotted colors are scaled by RPM and smoothed.

Statistical Analyses

Statistical analyses are described in detail in the Methods sections above. Whenever a p-value is reported in the text, the statistical test is also indicated. All statistics were estimated, and plots were generated using the statistical computing environment R (www.r-project.org).

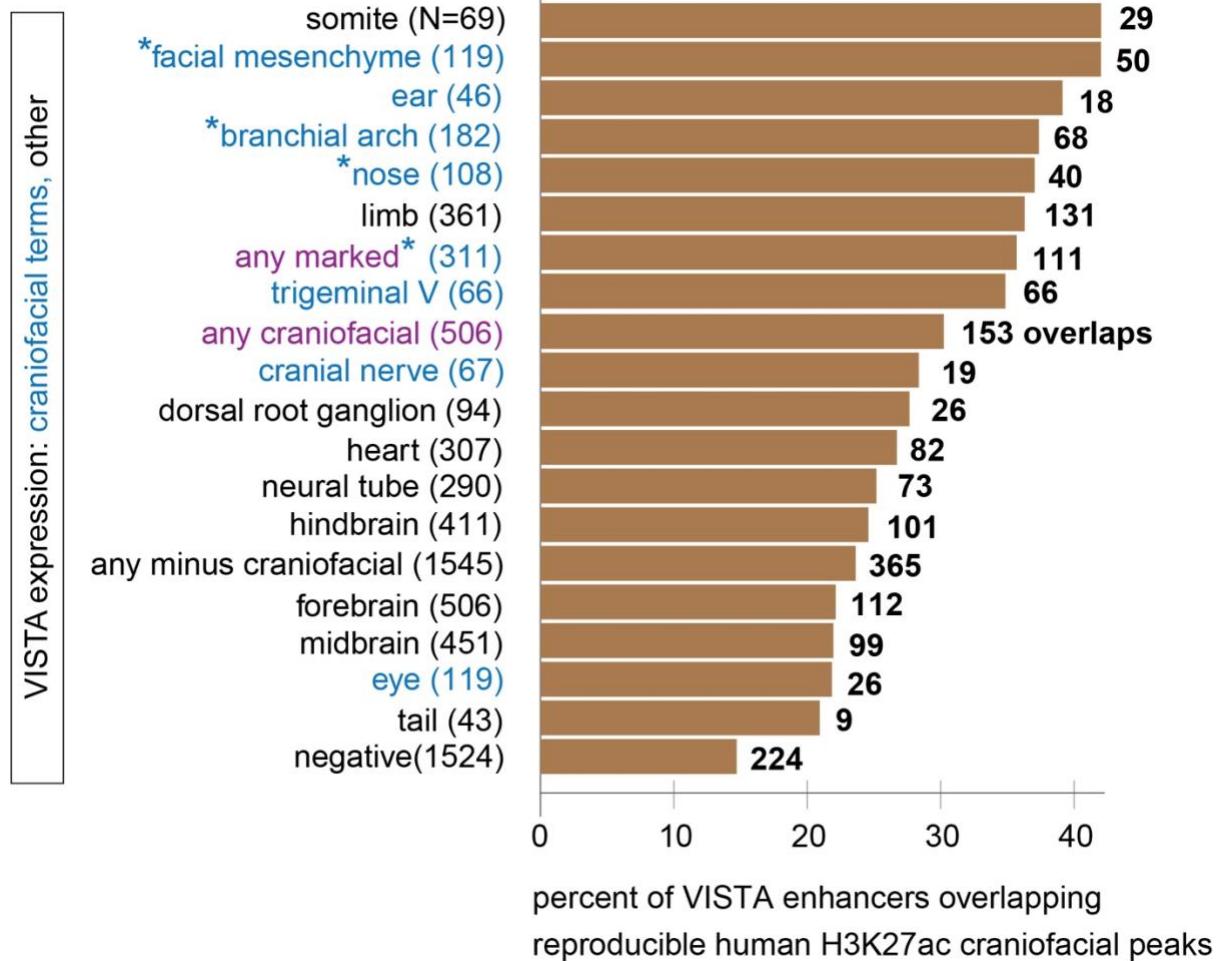
Imaging

For both brightfield and fluorescent images, all embryos were imaged with a Leica MZ16 microscope and a Leica DFC420 digital camera using identical lighting conditions.

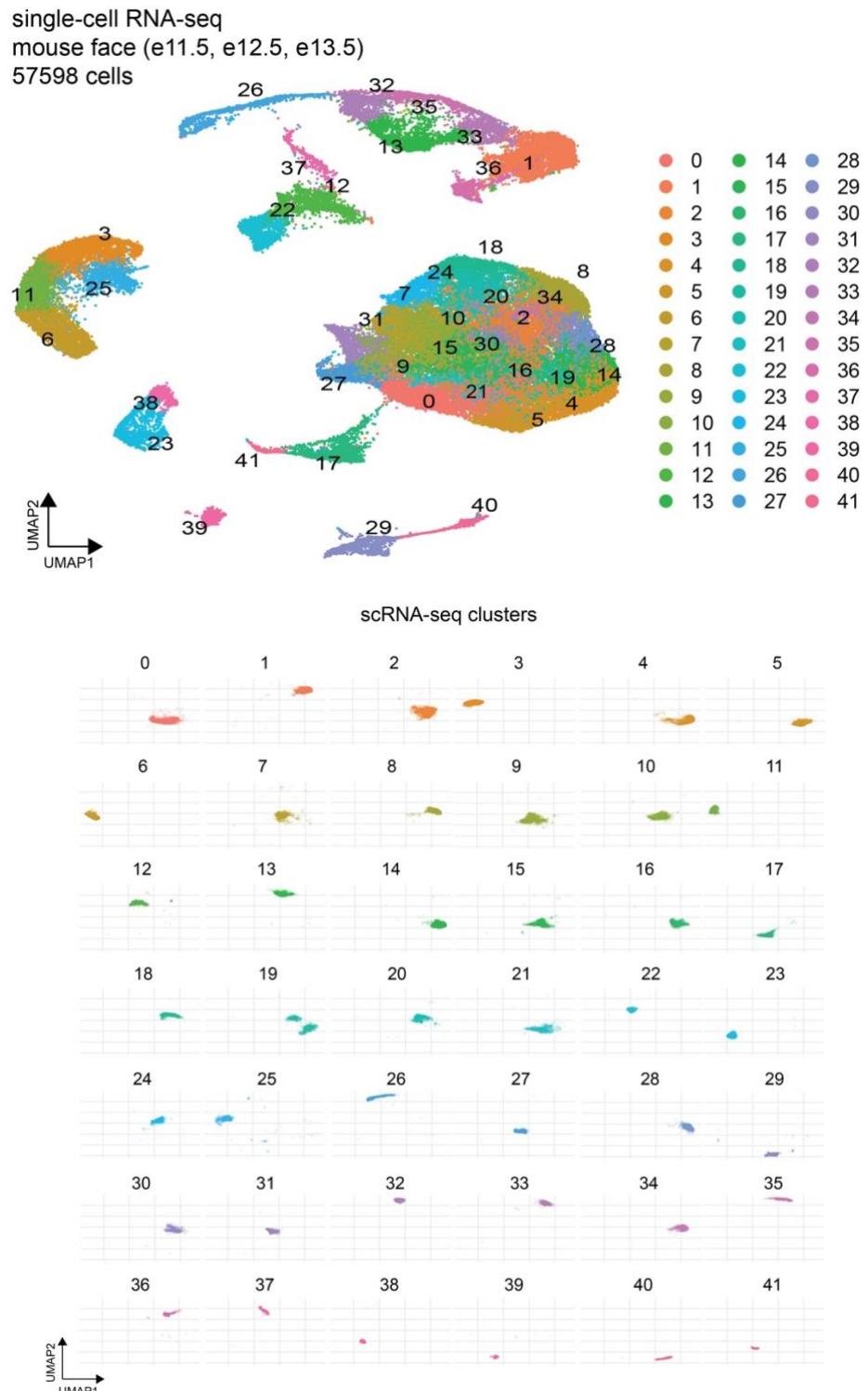
Data Availability

The ChIP-seq, ATAC-seq, RNA-seq as well as scRNA-seq and snATAC-seq data presented in this publication, and generated as part of this study are accessible at the National Institute of Dental and Craniofacial Research's FaceBase^{61,71,86,88} Consortium (facebase.org), Record ID 3B-Y34G; DOI 10.80001/3B-Y34G. These data are additionally deposited in NCBI's Gene Expression Omnibus^{85,87} and are accessible through GEO Series accession number GSE235858. Additional data supporting the findings of this study are available from the corresponding author upon reasonable request. Images of embryos with *lacZ*-reporter activity are available from the VISTA Enhancer Browser (enhancer.lbl.gov), and raw images are available on request from the lead authors.

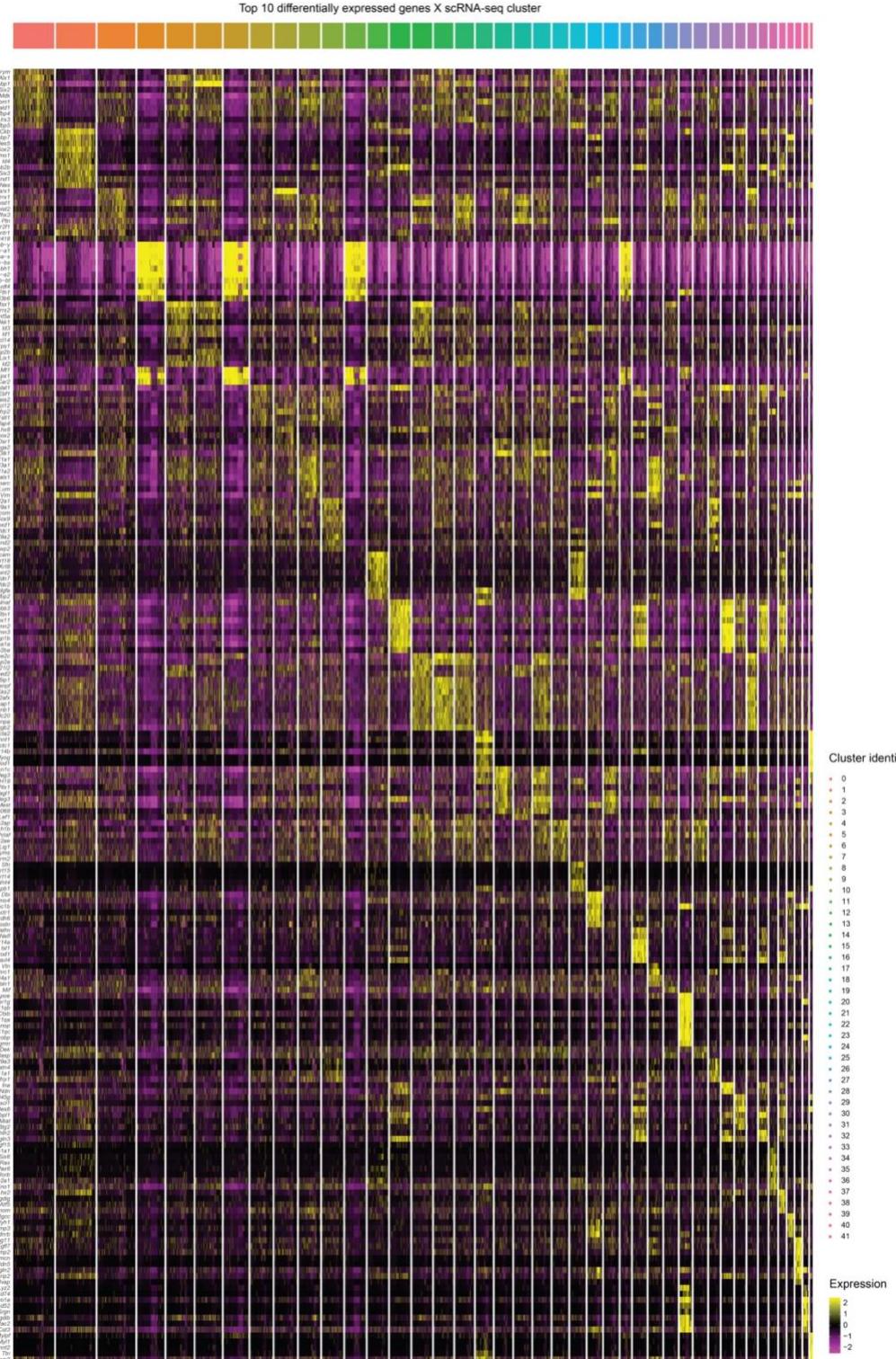
Extended Data



Extended Data Figure 1. Comparison of human face enhancers and VISTA craniofacial enhancers. Y-axis shows terms for recorded expression of enhancer driven *lacZ* for elements reported in the VISTA Enhancer Browser, numbers in parentheses denote the total number (N) of such observations in the VISTA catalog. Specific craniofacial terms are denoted in blue, “any craniofacial” comprises the seven craniofacial terms shown here plus “melanocytes”. Terms with fewer than 40 elements, incl. melanocytes, are not shown individually. Bars on the right indicate the percentage of VISTA enhancers (x-axis) for the relevant expression category that overlap human developmental face enhancers in this study (n=13,983). The number in bold by each bar denotes the absolute number of such overlaps out of the total N for each expression category. See **Supplemental Table 4** for a full list of elements positive for craniofacial terms.

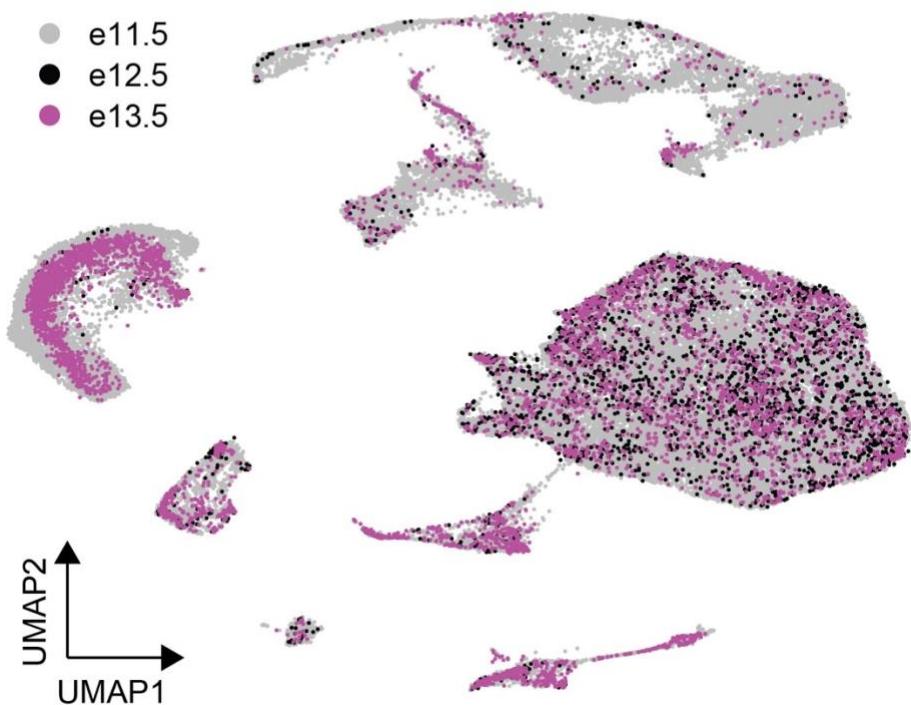


Extended Data Figure 2. Unbiased clustering of single-cell gene expression data of the developing mouse face. For clarity, individual UMAPs are shown to demonstrate the spatial extent along x-y coordinates for each cluster (0-41) that comprises the final UMAP shown in **Figure 3**.

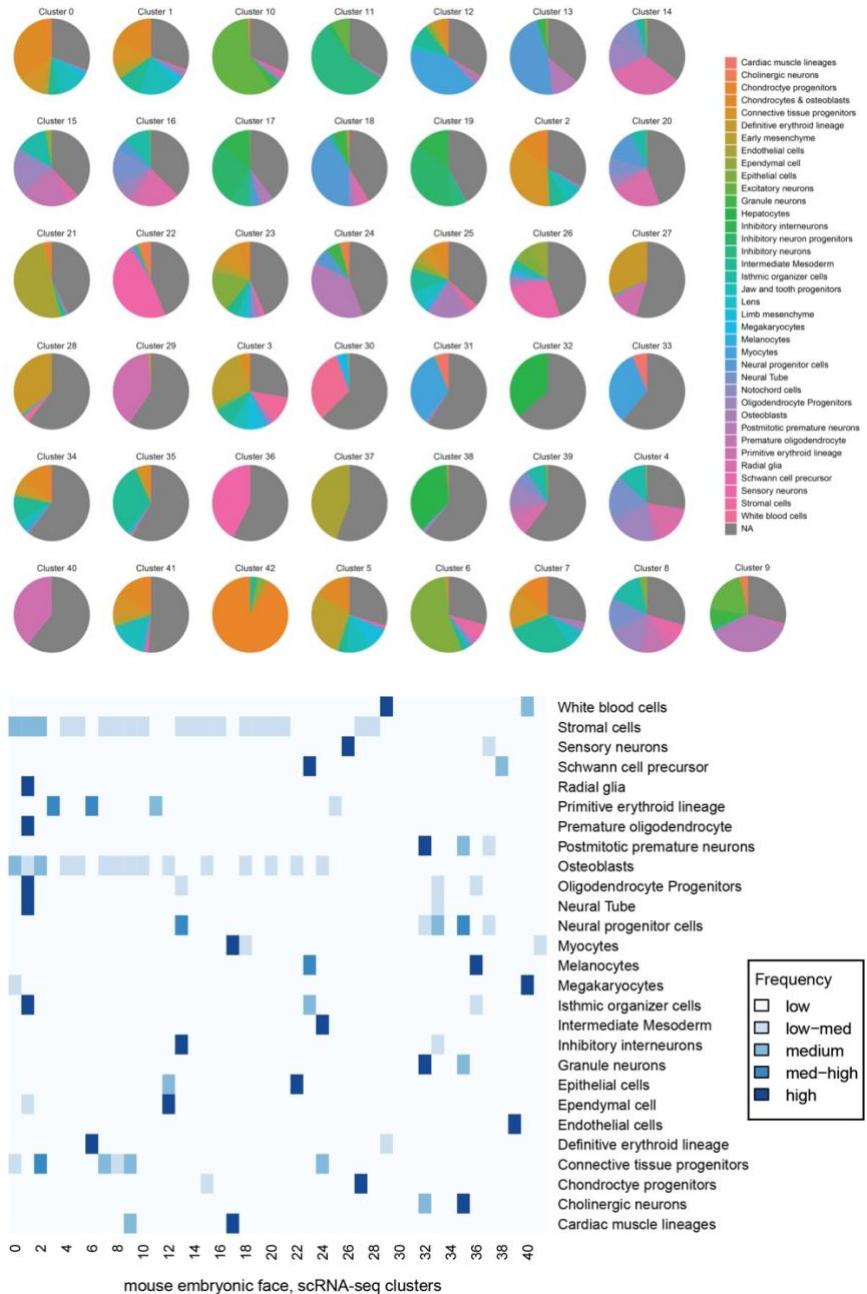


Extended Data Figure 3. Cluster-wise marker genes in single-cell (sc) gene expression data of the developing mouse face. Heatmap shows top 10 marker genes, i.e., genes highly enriched for expression in a cluster over all clusters (y-axis) for each of the 42 original clusters (x-axis) defined in the single-cell gene expression data.

mouse scRNA-seq, unbiased clustering by embryonic stage

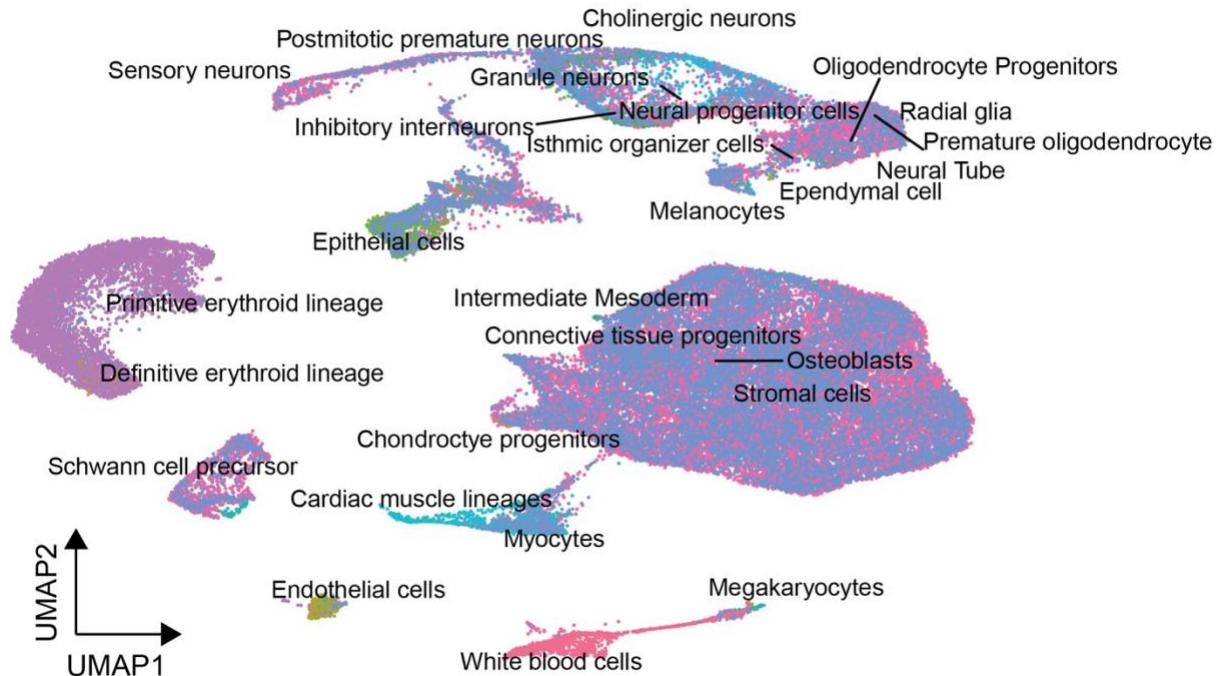


Extended Data Figure 4. Distribution of cells in the single-cell gene expression data by mouse embryonic stage. UMAP shows distribution of cells from respective mouse embryonic stages e11.5 (gray: 49,882 cells), e12.5 (black: 2,340 cells) and e13.5 (magenta: 5,376).

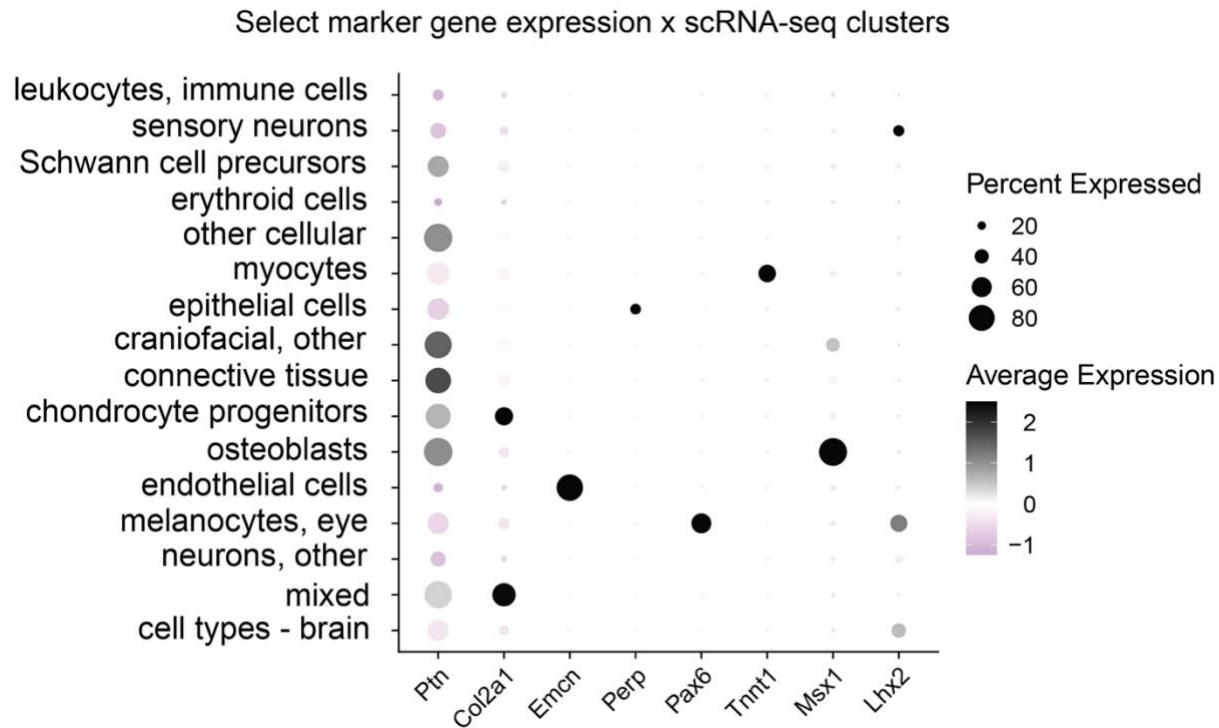


Extended Data Figure 5. Cluster-wise proportion of cell-types in mouse single-cell gene expression data. Our single-cell gene expression data was queried to a previously published scRNA-seq large dataset of whole embryo developmental timepoints (Cao et al., 2019) using Seurat-based auto referencing as a first step for assigning cell-type identities in an unbiased manner. Heatmap shows frequency (low to high) of cell types from the reference (n=27, y-axis) that are reflected in each of the 42 clusters (x-axis). These collectively informed the first broad annotations of cell types for our ScanFaceX data.

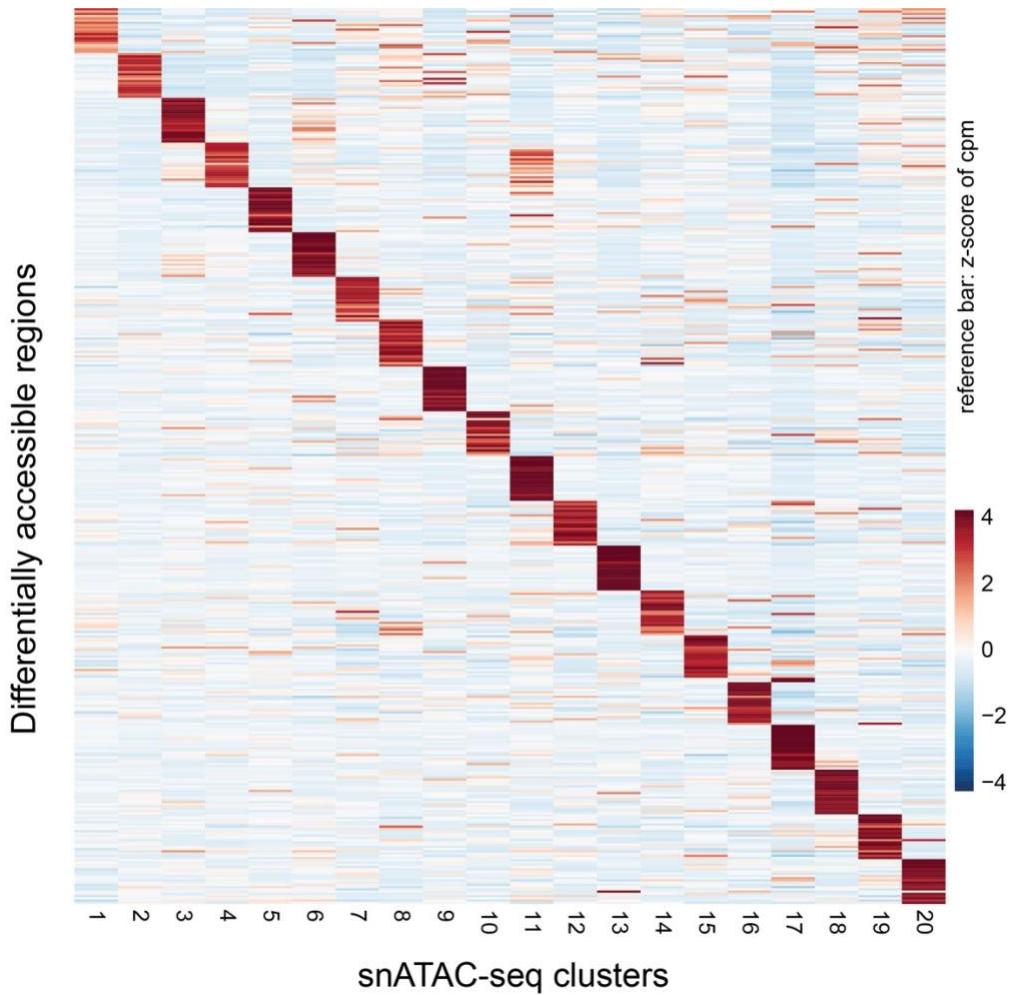
Automated referencing results



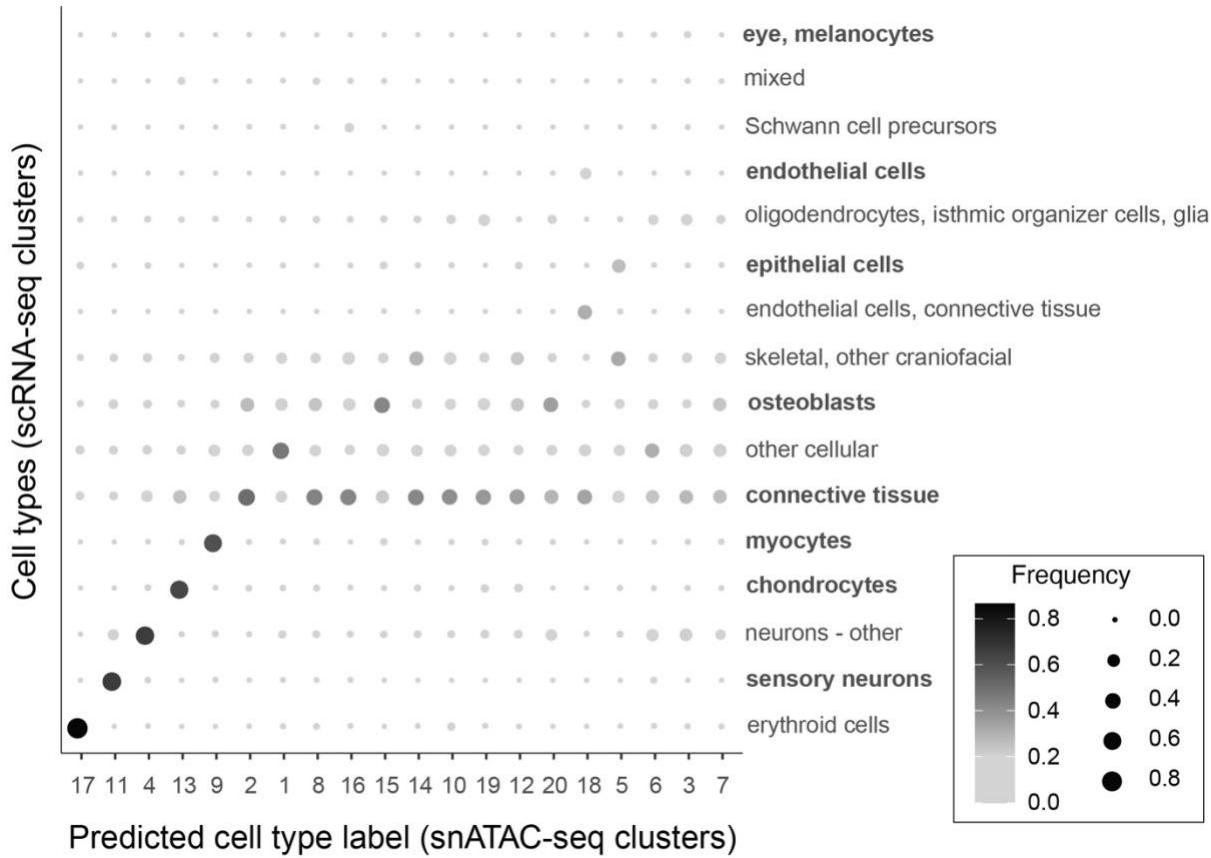
Extended Data Figure 6. Assignment of cell-type identities in *ScanFaceX*. UMAP shows raw results from Seurat-based automated referencing and cell type annotations for *ScanFaceX* data. A down-sampled (100K cells) subset from Cao et al., 2019 whole-embryo single cell data was used as reference.



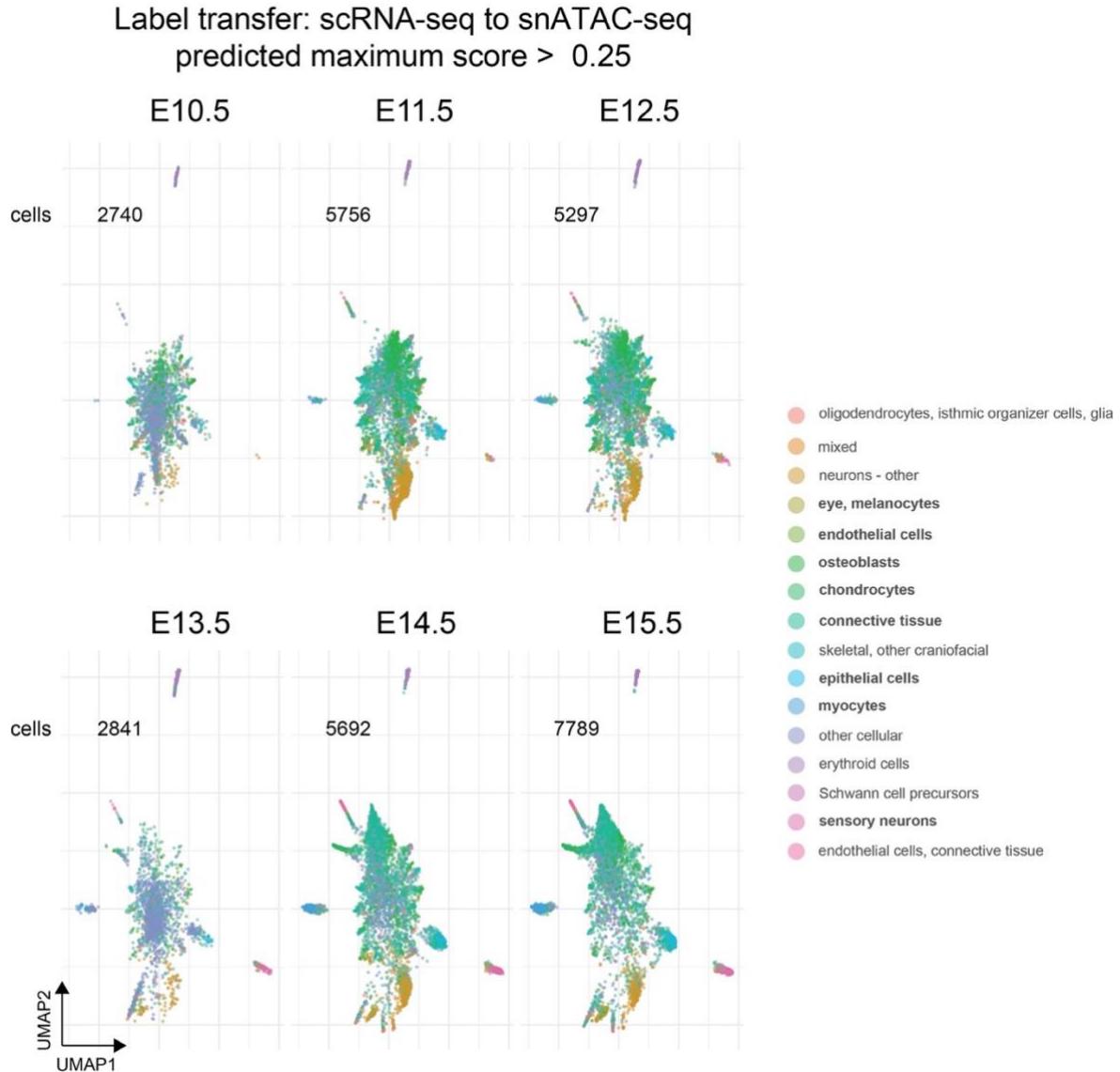
Extended Data Figure 7. Expression of select marker genes across all clusters in ScanFaceX.
Dot plot shows expression of select marker genes across all original clusters (consolidated per final annotations in **Supplemental Table 12**) of ScanFaceX, a subset of this plot is shown in main **Figure 3**. Color scale denotes low (light grey) to high (black) expression while increasing circle diameters denote corresponding higher proportion of cells within respective clusters.



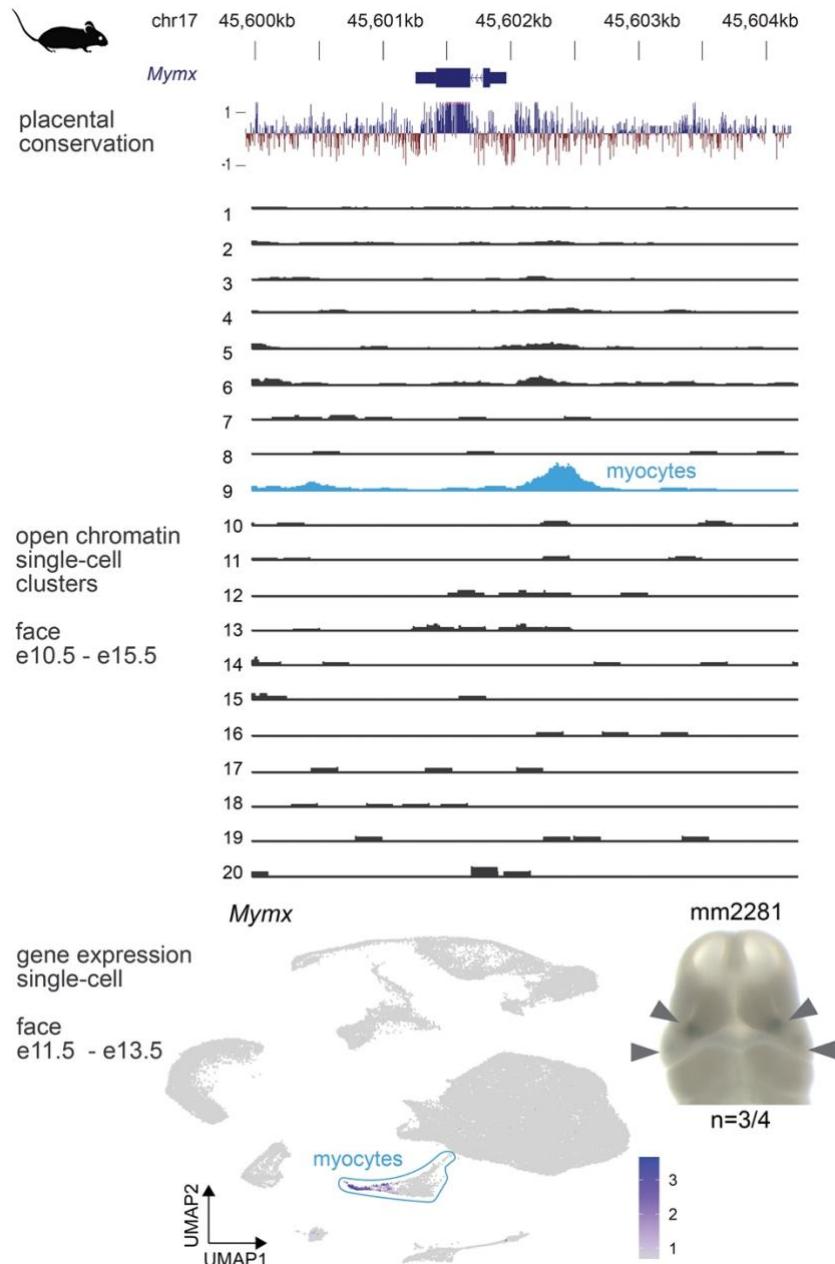
Extended Data Figure 8. Differentially accessible regions in snATAC-seq, mouse face. Heatmap shows the top 20 DARs map exclusively to each of the 20 clusters in snATAC-seq (mouse face).



Extended Data Figure 9. Correlation of scRNA-seq and snATAC-seq face data. Dot plot shows correlation, i.e., strength of label transfer between gene expression quantification (scRNA-seq; y-axis; n=16) and accessibility in TSS and gene bodies (snATAC-seq; x-axis; n=20) for integrated gene expression and open chromatin data for final annotated cell-types. Color scale denotes low (light grey) to high (black) degree of correlation while increasing circle diameters denote corresponding higher proportion of cells within correlated cell types for the respective clusters. Cell types in bold are cell types shown in **Figures 3-5**.



Extended Figure 10. Developmental stage-wise correlation of scRNA-seq and snATAC-seq face data.
Individual UMAPs show the total number of cells in our snATAC-seq assay that pass the >0.25 threshold for the predicted maximum score for label transfer between the integrated scRNA-seq and snATAC-seq datasets for 16 final cell-type annotations (key on the right).



Extended Data Figure 11. Differentially accessible regulatory regions correlate with cell-type specific signatures. The genomic context and placental conservation scores for a regulatory region near *Mymx* promoter are shown, followed by tracks for individual snATAC-seq clusters from developing mouse face tissue (e10.5 – e15.5). This region shows distinct open chromatin signature in the myocyte-specific cluster. UMAP of ScanFaceX shows expression of *Mymx* in myocytes. Image for a representative mouse embryo at e11.5 shows validated *in vivo* *lacZ*-reporter activity (grey arrowheads) of this enhancer. *MYMX* (*Myomixer*) encodes an integral membrane protein that regulates myoblast fusion, is conserved across vertebrates and *MYMX* mutations underlie an autosomal recessive disorder, Carey-Fineman-Ziter syndrome-2 (CFZS2) in humans that is characterized by weakness of the facial musculature, hypomimic facies, micrognathia, and facial dysmorphism among a range of other defects. n, reproducibility of each pattern across embryos resulting from independent transgenic integration events.

References

1. Attanasio, C., Nord, A.S., Zhu, Y., Blow, M.J., Li, Z., Liberton, D.K., Morrison, H., Plajzer-Frick, I., Holt, A., Hosseini, R., et al. (2013). Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* **342**, 1241006.
2. Fakhouri, W.D., Rahimov, F., Attanasio, C., Kouwenhoven, E.N., Ferreira De Lima, R.L., Felix, T.M., Nitschke, L., Huver, D., Barrons, J., Kousa, Y.A., et al. (2014). An etiologic regulatory mutation in IRF6 with loss- and gain-of-function effects. *Hum. Mol. Genet.* **23**, 2711–2720.
3. Rahimov, F., Marazita, M.L., Visel, A., Cooper, M.E., Hitchler, M.J., Rubini, M., Domann, F.E., Govil, M., Christensen, K., Bille, C., et al. (2008). Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nat. Genet.* **40**, 1341–1347.
4. Richmond, S., Howe, L.J., Lewis, S., Stergiakouli, E., and Zhurov, A. (2018). Facial Genetics: A Brief Overview. *Front. Genet.* **9**, 462.
5. Maden, M. (2001). Vitamin A and the developing embryo. *Postgrad. Med. J.* **77**, 489–491.
6. Le Douarin, N.M., Brito, J.M., and Creuzet, S. (2007). Role of the neural crest in face and brain development. *Brain Res. Rev.* **55**, 237–247.
7. Som, P.M., and Naidich, T.P. (2013). Illustrated review of the embryology and development of the facial region, part 1: Early face and lateral nasal cavities. *AJNR Am. J. Neuroradiol.* **34**, 2233–2240.
8. Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H., et al. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81.
9. Yuan, Y., Loh, Y.-H.E., Han, X., Feng, J., Ho, T.-V., He, J., Jing, J., Groff, K., Wu, A., and Chai, Y. (2020). Spatiotemporal cellular movement and fate decisions during first pharyngeal arch morphogenesis. *Science Advances* **6**, eabb0119.
10. Li, C., Lan, Y., and Jiang, R. (2017). Molecular and Cellular Mechanisms of Palate Development. *J. Dent. Res.* **96**, 1184–1191.
11. Diewert, V.M. (1985). Development of human craniofacial morphology during the late embryonic and early fetal periods. *Am. J. Orthod.* **88**, 64–76.
12. Twigg, S.R.F., and Wilkie, A.O.M. (2015). New insights into craniofacial malformations. *Hum. Mol. Genet.* **24**, R50-9.
13. Zalc, A., Rattenbach, R., Auradé, F., Cadot, B., and Relaix, F. (2015). Pax3 and Pax7 play essential safeguard functions against environmental stress-induced birth defects. *Dev. Cell* **33**, 56–66.
14. Fitriasari, S., and Trainor, P.A. (2021). Diabetes, Oxidative Stress, and DNA Damage Modulate Cranial Neural Crest Cell Development and the Phenotype Variability of Craniofacial Disorders. *Front Cell Dev Biol* **9**, 644410.

15. Van Otterloo, E., Williams, T., and Artinger, K.B. (2016). The old and new face of craniofacial research: How animal models inform human craniofacial genetic and clinical data. *Dev. Biol.* **415**, 171–187.
16. Wilkie, A.O., and Morriss-Kay, G.M. (2001). Genetics of craniofacial development and malformation. *Nat. Rev. Genet.* **2**, 458–468.
17. Roosenboom, J., Hens, G., Mattern, B.C., Shriver, M.D., and Claes, P. (2016). Exploring the Underlying Genetics of Craniofacial Morphology through Various Sources of Knowledge. *Biomed Res. Int.* **2016**, 3054578.
18. Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83.
19. Crane-Smith, Z., Schoenebeck, J., Graham, K.A., Devenney, P.S., Rose, L., Ditzell, M., Anderson, E., Thomson, J.I., Klenin, N., Kurrasch, D.M., et al. (2021). A Highly Conserved Shh Enhancer Coordinates Hypothalamic and Craniofacial Development. *Front Cell Dev Biol* **9**, 595744.
20. Xu, P., Yu, H.V., Tseng, K.-C., Flath, M., Fabian, P., Segil, N., and Crump, J.G. (2021). Foxc1 establishes enhancer accessibility for craniofacial cartilage differentiation. *Elife* **10**, e63595.
21. Lin-Shiao, E., Lan, Y., Welzenbach, J., Alexander, K.A., Zhang, Z., Knapp, M., Mangold, E., Sammons, M., Ludwig, K.U., and Berger, S.L. (2019). p63 establishes epithelial enhancers at critical craniofacial development genes. *Science Advances* **5**, eaaw0946.
22. Charité, J., McFadden, D.G., Merlo, G., Levi, G., Clouthier, D.E., Yanagisawa, M., Richardson, J.A., and Olson, E.N. (2001). Role of Dlx6 in regulation of an endothelin-1-dependent, dHAND branchial arch enhancer. *Genes Dev.* **15**, 3039–3049.
23. Guo, C., Sun, Y., Zhou, B., Adam, R.M., Li, X., Pu, W.T., Morrow, B.E., Moon, A., and Li, X. (4 2011). A Tbx1-Six1/Eya1-Fgf8 genetic pathway controls mammalian cardiovascular and craniofacial morphogenesis. *J. Clin. Invest.* **121**, 1585–1595.
24. Kvon, E.Z., Waymack, R., Gad, M., and Wunderlich, Z. (2021). Enhancer redundancy in development and disease. *Nat. Rev. Genet.* **22**, 324–336.
25. Gorkin, D.U., Barozzi, I., Zhao, Y., Zhang, Y., Huang, H., Lee, A.Y., Li, B., Chiou, J., Wildberg, A., Ding, B., et al. (2020). An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583**, 744–751.
26. Wilderman, A., VanOudenhove, J., Kron, J., Noonan, J.P., and Cotney, J. (2018). High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep.* **23**, 1581–1597.
27. White, J.D., Indencleef, K., Naqvi, S., Eller, R.J., Hoskens, H., Roosenboom, J., Lee, M.K., Li, J., Mohammed, J., Richmond, S., et al. (2021). Insights into the genetic architecture of the human face. *Nat. Genet.* **53**, 45–53.
28. Bishop, M.R., Diaz Perez, K.K., Sun, M., Ho, S., Chopra, P., Mukhopadhyay, N., Hetmanski, J.B., Taub, M.A., Moreno-Uribe, L.M., Valencia-Ramirez, L.C., et al. (2020). Genome-wide enrichment of DE Novo coding mutations in orofacial cleft trios. *Am. J. Hum. Genet.* **107**, 124–136.

29. Morris, V.E., Hashmi, S.S., Zhu, L., Maili, L., Urbina, C., Blackwell, S., Greives, M.R., Buchanan, E.P., Mulliken, J.B., Blanton, S.H., et al. (2020). Evidence for craniofacial enhancer variation underlying nonsyndromic cleft lip and palate. *Hum. Genet.* *139*, 1261–1272.

30. Frisdal, A., and Trainor, P.A. (2014). Development and evolution of the pharyngeal apparatus. *Wiley Interdiscip. Rev. Dev. Biol.* *3*, 403–418.

31. Murillo-Rincón, A.P., and Kaucka, M. (2020). Insights Into the Complexity of Craniofacial Development From a Cellular Perspective. *Front Cell Dev Biol* *8*, 620735.

32. Piest, K.L. (2002). Embryology and Anatomy of the Developing Face. In *Pediatric Oculoplastic Surgery*, J. A. Katowitz, ed. (Springer New York), pp. 11–29.

33. Burdi, A.R. (2006). Developmental Biology and Morphogenesis of the Face, Lip and Palate. In *Cleft Lip and Palate*, S. Berkowitz, ed. (Springer Berlin Heidelberg), pp. 3–12.

34. Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res.* *35*, D88-92.

35. Gu, Z., and Hübschmann, D. (2022). rGREAT: an R/Bioconductor package for functional enrichment on genomic regions. *Bioinformatics*. *10.1093/bioinformatics/btac745*.

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

37. Prescott, N.J., Lees, M.M., Winter, R.M., and Malcolm, S. (2000). Identification of susceptibility loci for nonsyndromic cleft lip with or without cleft palate in a two stage genome scan of affected sib-pairs. *Hum. Genet.* *106*, 345–350.

38. Barraza-García, J., Rivera-Pedroza, C.I., Hisado-Oliva, A., Belinchón-Martínez, A., Sentchordi-Montané, L., Duncan, E.L., Clark, G.R., Del Pozo, A., Ibáñez-Garikano, K., Offiah, A., et al. (2017). Broadening the phenotypic spectrum of POP1-skeletal dysplasias: identification of POP1 mutations in a mild and severe skeletal dysplasia. *Clin. Genet.* *92*, 91–98.

39. Elalaoui, S.C., Laarabi, F.Z., Mansouri, M., Mrani, N.A., Nishimura, G., and Sefiani, A. (2016). Further evidence of POP1 mutations as the cause of anauxetic dysplasia. *Am. J. Med. Genet. A* *170*, 2462–2465.

40. Glazov, E.A., Zankl, A., Donskoi, M., Kenna, T.J., Thomas, G.P., Clark, G.R., Duncan, E.L., and Brown, M.A. (2011). Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia. *PLoS Genet.* *7*, e1002027.

41. Kvon, E.Z., Zhu, Y., Kelman, G., Novak, C.S., Plajzer-Frick, I., Kato, M., Garvin, T.H., Pham, Q., Harrington, A.N., Hunter, R.D., et al. (2020). Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. *Cell* *180*, 1262-1271.e15.

42. Osterwalder, M., Tran, S., Hunter, R.D., Meky, E.M., von Maydell, K., Harrington, A.N., Godoy, J., Novak, C.S., Plajzer-Frick, I., Zhu, Y., et al. (2022). Characterization of Mammalian In Vivo Enhancers Using Mouse Transgenesis and CRISPR Genome Editing. *Methods Mol. Biol.* *2403*, 147–186.

43. Zhou, G., Lefebvre, V., Zhang, Z., Eberspaecher, H., and de Crombrugghe, B. (1998). Three high mobility group-like sequences within a 48-base pair enhancer of the Col2a1 gene are required for cartilage-specific expression in vivo. *J. Biol. Chem.* 273, 14989–14997.

44. Lefebvre, V., Li, P., and de Crombrugghe, B. (1998). A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J.* 17, 5718–5733.

45. Saito, T., Yano, F., Mori, D., Ohba, S., Hojo, H., Otsu, M., Eto, K., Nakauchi, H., Tanaka, S., Chung, U.-I., et al. (2013). Generation of Col2a1-EGFP iPS cells for monitoring chondrogenic differentiation. *PLoS One* 8, e74137.

46. Nassif, A., Senussi, I., Meary, F., Loiodice, S., Hotton, D., Robert, B., Bensidhoum, M., Berdal, A., and Babajko, S. (2014). Msx1 role in craniofacial bone morphogenesis. *Bone* 66, 96–104.

47. Orestes-Cardoso, S., Nefussi, J.-R., Lezot, F., Oboeuf, M., Pereira, M., Mesbah, M., Robert, B., and Berdal, A. (2002). Msx1 Is a Regulator of Bone Formation During Development and Postnatal Growth: In Vivo Investigations in a Transgenic Mouse Model. *Connect. Tissue Res.* 43, 153–160.

48. Han, J., Ishii, M., Bringas, P., Maas, R.L., Maxson, R.E., and Chai, Y. (2007). Concerted action of Msx1 and Msx2 in regulating cranial neural crest cell differentiation during frontal bone development. *Mech. Dev.* 124, 729–745.

49. Ihrie, R.A., Marques, M.R., Nguyen, B.T., Horner, J.S., Papazoglu, C., Bronson, R.T., Mills, A.A., and Attardi, L.D. (2005). Perp is a p63-regulated gene essential for epithelial integrity. *Cell* 120, 843–856.

50. Franke, W.W., Heid, H., Zimbelmann, R., Kuhn, C., Winter-Simanowski, S., Dörflinger, Y., Grund, C., and Rickelt, S. (2013). Transmembrane protein PERP is a component of tessellate junctions and of other junctional and non-junctional plasma membrane regions in diverse epithelial and epithelium-derived cells. *Cell Tissue Res.* 353, 99–115.

51. Morgan, S.M., Samulowitz, U., Darley, L., Simmons, D.L., and Vestweber, D. (1999). Biochemical characterization and molecular cloning of a novel endothelial-specific sialomucin. *Blood* 93, 165–175.

52. Kuhn, A., Brachtendorf, G., Kurth, F., Sonntag, M., Samulowitz, U., Metze, D., and Vestweber, D. (2002). Expression of endomucin, a novel endothelial sialomucin, in normal and diseased human skin. *J. Invest. Dermatol.* 119, 1388–1393.

53. Kolterud, A., Alenius, M., Carlsson, L., and Bohm, S. (2004). The Lim homeobox gene Lhx2 is required for olfactory sensory neuron identity. *Development* 131, 5319–5326.

54. Berghard, A., Hägglund, A.-C., Bohm, S., and Carlsson, L. (2012). Lhx2-dependent specification of olfactory sensory neurons is required for successful integration of olfactory, vomeronasal, and GnRH neurons. *FASEB J.* 26, 3464–3472.

55. Kubic, J.D., Young, K.P., Plummer, R.S., Ludvik, A.E., and Lang, D. (2008). Pigmentation PAX-ways: the role of Pax3 in melanogenesis, melanocyte stem cell maintenance, and disease. *Pigment Cell Melanoma Res.* 21, 627–645.

56. Raviv, S., Bharti, K., Rencus-Lazar, S., Cohen-Tayyar, Y., Schyr, R., Evantal, N., Meshorer, E., Zilberman, A., Idelson, M., Reubinoff, B., et al. (2014). PAX6 regulates

melanogenesis in the retinal pigmented epithelium through feed-forward regulatory interactions with MITF. *PLoS Genet.* *10*, e1004360.

57. Wei, B., and Jin, J.-P. (2016). TNNT1, TNNT2, and TNNT3: Isoform genes, regulation, and structure-function relationships. *Gene* *582*, 1–13.

58. Mitiadis, T.A., Salmivirta, M., Muramatsu, T., Muramatsu, H., Rauvala, H., Lehtonen, E., Jalkanen, M., and Thesleff, I. (1995). Expression of the heparin-binding cytokines, midkine (MK) and HB-GAM (pleiotrophin) is associated with epithelial-mesenchymal interactions during fetal development and organogenesis. *Development* *121*, 37–51.

59. Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D.U., Zhang, Y., Sos, B.C., Afzal, V., Dickel, D.E., et al. (2018). Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* *21*, 432–439.

60. Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shores, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., Kaul, R., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* *583*, 699–710.

61. Brinkley, J.F., Fisher, S., Harris, M.P., Holmes, G., Hooper, J.E., Jabs, E.W., Jones, K.L., Kesselman, C., Klein, O.D., Maas, R.L., et al. (2016). The FaceBase Consortium: a comprehensive resource for craniofacial researchers. *Development* *143*, 2677–2688.

62. Pak, W., Hindges, R., Lim, Y.-S., Pfaff, S.L., and O'Leary, D.D.M. (2004). Magnitude of binocular vision controlled by islet-2 repression of a genetic program that specifies laterality of retinal axon pathfinding. *Cell* *119*, 567–578.

63. Jauregui, R., Thomas, A.L., Liechty, B., Velez, G., Mahajan, V.B., Clark, L., and Tsang, S.H. (2019). SCAPER-associated nonsyndromic autosomal recessive retinitis pigmentosa. *Am. J. Med. Genet. A* *179*, 312–316.

64. Tatour, Y., Sanchez-Navarro, I., Chervinsky, E., Hakonarson, H., Gawi, H., Tahsin-Swafiri, S., Leibu, R., Lopez-Molina, M.I., Fernandez-Sanz, G., Ayuso, C., et al. (2017). Mutations in SCAPER cause autosomal recessive retinitis pigmentosa with intellectual disability. *J. Med. Genet.* *54*, 698–704.

65. Kahrizi, K., Huber, M., Galetzka, D., Dewi, S., Schröder, J., Weis, E., Kariminejad, A., Fattah, Z., Ropers, H.-H., Schweiger, S., et al. (2019). Homozygous variants in the gene SCAPER cause syndromic intellectual disability. *Am. J. Med. Genet. A* *179*, 1214–1225.

66. Yahyavi, M., Abouzeid, H., Gawdat, G., de Preux, A.-S., Xiao, T., Bardakjian, T., Schneider, A., Choi, A., Jorgenson, E., Baier, H., et al. (2013). ALDH1A3 loss of function causes bilateral anophthalmia/microphthalmia and hypoplasia of the optic nerve and optic chiasm. *Hum. Mol. Genet.* *22*, 3250–3258.

67. Lin, S., Harlalka, G.V., Hameed, A., Reham, H.M., Yasin, M., Muhammad, N., Khan, S., Baple, E.L., Crosby, A.H., and Saleha, S. (2018). Novel mutations in ALDH1A3 associated with autosomal recessive anophthalmia/microphthalmia, and review of the literature. *BMC Med. Genet.* *19*, 160.

68. Fares-Taie, L., Gerber, S., Chassaing, N., Clayton-Smith, J., Hanein, S., Silva, E., Serey, M., Serre, V., Gérard, X., Baumann, C., et al. (2013). ALDH1A3 mutations cause recessive anophthalmia and microphthalmia. *Am. J. Hum. Genet.* *92*, 265–270.

69. Mory, A., Ruiz, F.X., Dagan, E., Yakovtseva, E.A., Kurolap, A., Parés, X., Farrés, J., and Gershoni-Baruch, R. (2013). A missense mutation in ALDH1A3 causes isolated microphthalmia/anophthalmia in nine individuals from an inbred Muslim kindred. *Eur. J. Hum. Genet.* **22**, 419–422.

70. Satokata, I., Ma, L., Ohshima, H., Bei, M., Woo, I., Nishizawa, K., Maeda, T., Takano, Y., Uchiyama, M., Heaney, S., et al. (2000). Msx2 deficiency in mice causes pleiotropic defects in bone growth and ectodermal organ formation. *Nat. Genet.* **24**, 391–395.

71. Samuels, B.D., Aho, R., Brinkley, J.F., Bugacov, A., Feingold, E., Fisher, S., Gonzalez-Reiche, A.S., Hacia, J.G., Hallgrímsson, B., Hansen, K., et al. (2020). FaceBase 3: analytical tools and FAIR resources for craniofacial and dental research. *Development* **147**, 10.1242/dev.191213.

72. Gerrelli, D., Lisgo, S., Copp, A.J., and Lindsay, S. (2015). Enabling research with human embryonic and fetal tissue resources. *Development* **142**, 3073–3076.

73. Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345.

74. Spurrell, C.H., Barozzi, I., Kosicki, M., Mannion, B.J., Blow, M.J., Fukuda-Yuzawa, Y., Slaven, N., Afzal, S.Y., Akiyama, J.A., Afzal, V., et al. (2022). Genome-wide fetalization of enhancer architecture in heart disease. *Cell Rep.* **40**, 111400.

75. Rajderkar, S., Barozzi, I., Zhu, Y., Hu, R., Zhang, Y., Li, B., Fukuda-Yuzawa, Y., Kelman, G., Akeza, A., Blow, M.J., et al. (2021). Topologically Associating Domain Boundaries are Commonly Required for Normal Genome Function. *bioRxiv*, 2021.05.06.443037. 10.1101/2021.05.06.443037.

76. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2018). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012.

77. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K., and Kastenmüller, G. (2015). SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334–1336.

78. McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V., Hu, J.L., Murrow, L.M., Weissman, J.S., Werb, Z., et al. (2019). MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626.

79. Stuart, T., and Satija, R. (2019). Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257–272.

80. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29.

81. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420.

82. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* *566*, 496–502.

83. Amezquita, R.A., Lun, A.T.L., Becht, E., Carey, V.J., Carpp, L.N., Geistlinger, L., Marini, F., Rue-Albrecht, K., Risso, D., Soneson, C., et al. (2020). Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* *17*, 137–145.

84. Fang, R., Preissl, S., Li, Y., Hou, X., Lucero, J., Wang, X., Motamed, A., Shiau, A.K., Zhou, X., Xie, F., et al. (2021). Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat. Commun.* *12*, 1337.

85. Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* *30*, 207–210.

86. Hochheiser, H., Aronow, B.J., Artinger, K., Beaty, T.H., Brinkley, J.F., Chai, Y., Clouthier, D., Cunningham, M.L., Dixon, M., Donahue, L.R., et al. (2011). The FaceBase Consortium: a comprehensive program to facilitate craniofacial research. *Dev. Biol.* *355*, 175–182.

87. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillip, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* *41*, D991-5.

88. Schuler, R.E., Bugacov, A., Hacia, J.G., Ho, T.V., Iwata, J., Pearlman, L., Samuels, B.D., Williams, C., Zhao, Z., Kesselman, C., et al. (2022). FaceBase: A Community-Driven Hub for Data-Intensive Research. *J. Dent. Res.* *101*, 1289–1298.