Title

Dynamic enhancer landscapes in human craniofacial development

Journal

Nature Communications, 15(1)

ISSN

2041-1723

Authors

Rajderkar, Sudha Sunil

Paraiso, Kitt

Amaral, Maria Luisa

et al.

Publication Date

2024

DOI

10.1038/s41467-024-46396-4

Dynamic Enhancer Landscapes in Human Craniofacial Development

Sudha Sunil Rajderkar[1], Kitt Paraiso[1], Maria Luisa Amaral[2], Michael Kosicki[1], Laura E. Cook[1], Fabrice Darbellay[1,3], Cailyn H. Spurrell[1], Marco Osterwalder[1,4], Yiwen Zhu[1], Han Wu[1], Sarah Yasmeen Afzal[1,5], Matthew J. Blow[6], Guy Kelman[1,7], Iros Barozzi,[1,8,9], Yoko Fukuda-Yuzawa[1,10], Jennifer A. Akiyama[1], Veena Afzal[1], Stella Tran[1], Ingrid Plajzer-Frick[1], Catherine S. Novak[1], Momoe Kato[1], Riana D. Hunter[1,11], Kianna von Maydell[1], Allen Wang[12], Lin Lin[12], Sebastian Preissl[12,13], Steven Lisgo[14,15], Bing Ren[16], 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.

Abstract

The genetic basis of human facial variation and craniofacial birth defects remains poorly understood. Distant-acting transcriptional enhancers control the fine-tuned spatiotemporal expression of genes during critical stages of craniofacial development[1–3]. However, a lack of accurate maps of the genomic locations and cell type-resolved activities of craniofacial enhancers prevents their systematic exploration in human genetics studies. Here, we combine histone modification, chromatin accessibility, and gene expression profiling of human craniofacial development with single-cell analyses of the developing mouse face to define the regulatory landscape of facial development at tissue- and single cell-resolution. We provide temporal activity profiles for 14,000 human developmental craniofacial enhancers. We find that 56% of human craniofacial enhancers share chromatin accessibility in the mouse and we provide cell population- and embryonic stage-resolved predictions of their *in vivo* activity. 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 effects[4]. Even subtle disturbances during embryogenesis can result in a range of craniofacial defects or dysfunctions[5]. 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 conception[10] and the 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 conception[11]. 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 pathways[17], 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 genes[25] 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. Several single-cell studies have been performed for the developing face in vertebrate and mammalian model systems, as well as some human face tissues[9,27,29,31,33,35,37,39,42,43,45–54]. While these studies cover several specific cell lineages or anatomical sub-regions of the face, the broad enhancer landscape of

mammalian face development at cell type resolution remains incompletely understood. 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 variation[28] and whole genome sequencing as an increasingly common clinical approach for the identification of noncoding mutations in craniofacial birth defect patients[30,32], 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 gene expression and open chromatin signatures at single cell resolution for the developing mouse 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[11,34,36]. 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)[38,40]. 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

(Supplementary Figure 1, Supplementary Data 1). To extend our compendium to earlier stages, we complemented this data with published H3K27ac peaks (ChIP-seq) from CS13-17 human face tissue and an additional available sample at CS20[26] (Supplementary Data 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 (Supplementary Data 2). We examined the correlation between H3K27ac peaks and chromatin accessibility focusing on week 7 (comprising CS18 and CS19), since the largest number of perfectly matched datasets (H3K27ac peaks and chromatin accessibility data from the same biological samples) were available for this stage. We observed that 2,225 out of 3182 (70%) of the reproducible H3K27ac peaks overlap at least one ATAC-seq peak derived from the same samples (Supplementary Data 3; Methods).

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 browser[41]. Among the 130 human craniofacial regulatory elements that have been tested in VISTA to date and that are annotated for branchial arch, facial mesenchyme, or nose, we identified 38 cases (29%) with overlaps with an enhancer predicted through the present human-derived epigenomic dataset (Supplementary Figure 2, Supplementary Data 4). A representative example of a 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 (Supplementary Data 5; Methods). Of these, a total of 28 candidate enhancers were positive for reporter activity, out of which 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, Figure 1c). Of the 16 enhancers positive for craniofacial tissues, 8 were simultaneously active in non-craniofacial structures such as the brain or limb, while the remaining 12 out of the total 28 were only positive in non-craniofacial tissues (Supplementary Data 5).

Developmental Dynamics 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 analysis[44]. 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, and Supplementary Data 6), including midface retrusion, reduced number of teeth, and abnormality of maxilla.

In a complementary assessment, we explored the putative target genes of the human reproducible enhancers with predictions from publicly available promoter-centric long-range chromatin interaction data for approximately 19,000 human promoters[55]. This interaction-based mapping strategy identified 3,005 chromatin segments containing predicted craniofacial enhancers interacting with the promoters of 2,921 nearby genes (Supplementary Data 7; Methods). Across 2,263 predicted gene-enhancer pairs with epigenomic enhancer predictions and gene expression data available from identical biological samples, we observed a positive correlation between sample-specific enhancer activity and gene

expression levels (p=0.00002; Mann-Whitney U Test; see Supplementary Figure 3, Supplementary Data 8; Methods). 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 (Supplementary Data 9; Methods). 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 with 605 unique LD SNPs. This LD SNP density represents an enrichment compared to control SNPs not implicated in craniofacial traits (OR = 1.27, $p<10^{-8}$; Methods). 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 (Supplementary Data 10).

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[25,56]. 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 (labeled “constant” in Figure 2b) as enhancers throughout all five weeks. Nearly half (6,347) showed narrow predicted activity windows limited to a single week, while another 3,749 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. We note that the analysis of temporal dynamics of subsets of enhancers may potentially be influenced by the variable number of samples or peaks per week. However, we do not observe obvious confounding effects due to these variables within the samples we have analyzed (Supplementary Figure 4, Supplementary Data 11; Methods). 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 published results for histone modifications at matched stages of mouse development[25]. 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 sequence using LiftOver (UCSC Genome Browser[57]) and that is syntenic relative to surrounding protein-coding genes. Among these conserved sequences, 8,257 (59%) of the human candidate enhancers showed H3K27ac binding 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 (Supplementary Data 12; Methods), suggesting that these regions are active enhancers in humans only 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 an enhancer showing an active enhancer signature in the human genome only. 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 face[59] (Figure 2c/d). Documented mutations in *POP1* cause Anauxetic Dysplasia with pathognomonic short stature, hypoplastic midface and hypodontia along with mild intellectual disability[61,63,64]. 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[65,66] 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).

Single-cell Transcriptomics of the Craniofacial Development

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 population-resolved activity signatures 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).

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 (Supplementary Figures 5-8, Supplementary Data 13-14). 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 (Supplementary Figures 9-10, Supplementary Data 15-16; 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* (chondrocytes)[67–69], *Msx1* (undifferentiated mesenchyme)[70–72], *Perp* (epithelial cells)[73,74], *Emcn* (endothelial cells)[75,76], *Lhx2* (sensory neurons)[77,78], *Pax6* (melanocytes)[58,60], *Tnnt1* (myocytes)[62], and *Ptn* (connective tissue)[79] (Figure 3c and 3d, Supplementary Figure 11). 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 and Gene Expression

To identify developmental enhancers at single-cell resolution, we performed single-nucleus ATAC-seq (snATAC-seq)[80] 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 has distinct open chromatin signatures (Supplementary Figure 12, Supplementary Data 17). 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%) across 11 annotated clusters for developing craniofacial cell types were retained. Clusters labeled chondrocytes, myocytes and connective tissue, and sensory neurons showed high correlation between the two data types (Figure 4a-b, Supplementary Figures 13 and 14; 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[25,82,84] (Figure 4c). Within the immediate vicinity of these two enhancer regions, we display genes with positive expression in *ScanFaceX* and those that were reported in the OMIM catalog[86,87] as human disease-causing. Both *Isl2* and *Aldh1a3* are highly expressed in sensory neurons and epithelial cell clusters, respectively, in *ScanFaceX* data (Figure 4c). *Isl2* has been shown to be selectively expressed in a subset of retinal ganglion cell axons that have important functions in binocular vision[88]. Allelic variants and mutations in *SCAPER* cause intellectual disability with retinitis pigmentosa in humans[89–91]. 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 microphthalmia[92–95]. These putative enhancer regions near *Isl2* and

Scaper, and in the intron of *Lrrk1* drive reproducible *lacZ*-reporter activity in the developing mouse face at e11.5 in anatomical regions where neuronal and epithelial cell types are expected to be found (mm2285 and mm2282, Figure 4c**)**. Notably, the spatial expression pattern of mm2285 and mm2282 is consistent with the expression of *Isl2* in cranial ganglia[96,97], and the expression of *Aldh1a3* in the retina and the nasal epithelium[98] in similar developmental windows in mice *in vivo*. 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* (Supplementary Figure 15).

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 (Supplementary Data 18-20)**.** 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*.

Cell Population-resolved Enhancer Activity Predictions

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 *ScanFaceN* DARs from the 11 main *ScanFaceX*-matched clusters with craniofacial enhancers validated *in vivo* and curated in the VISTA Enhancer Browser[41] (Figure 5a). We observed general correlations between cluster-specific accessibility and spatial *in vivo* patterns among 77 formerly validated VISTA enhancers that showed chromatin accessibility in at least one of the 11 main clusters. For example, the predicted connective tissue-mesenchymal cluster (cluster 2) of the craniofacial snATAC-seq tends to group VISTA enhancers with activity specific to the branchial arches (Figure 5b). Despite broad correlations, we observed considerable heterogeneity of spatial patterns within most clusters. For example, the chondrocyte cluster (cluster 13) has multiple VISTA enhancers with activity in the mid-face, paranasal regions, and/or a region at the junction of the developing forebrain and nasal prominences that may constitute the developing cartilaginous regions of the face (Figure 5b). These observations underscore the spatiotemporal complexity of craniofacial morphogenesis, which relies on intricate cellular processes in combination with highly regionalized regulatory cues.

Craniofacial 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 formerly demonstrated to be required for normal facial development[1] (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 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 a cluster predicted to represent undifferentiated mesenchyme and in a subset of cells expressing *Msx1* in *ScanFaceX*, a gene shown to regulate the osteogenic lineage[99]. Similarly, based on *ScanFaceX* annotations, enhancer hs521, located near *Gbx2*, is primarily active in a subset of predicted mesenchymal cells and chondrocytes, and its activity coincides with a subset of cells expressing *Gbx2* (Figure 6c), a gene known to be active in the developing mandibular arches[9]. 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.

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. While many predicted enhancers show conserved epigenomic signatures indicating an active enhancer state in both mouse and human, we also observed elements with human-specific enhancer activity signatures, suggesting that the human but not the mouse ortholog is an active *in vivo* enhancer. We also provide additional predictions of regions with human-specific enhancer signatures that show no functional

conservation in mice that can be identified by profiling human tissues. We observed that enhancer hs2656, but not its mouse ortholog mm2280, shows craniofacial *in vivo* activity in transgenic mice. This is consistent with an epigenomic enhancer signature at this element in human, but not in mouse tissue. These lineage-specific differences in epigenomic signature and *in vivo* activity are likely due to sequence differences within the enhancer element itself, which may affect transcription factor binding sites or other functionally critical motifs embedded in the enhancer. For example, within the most conserved 425bp core sequence of enhancer hs2256, 31% of the nucleotide positions show differences between human and mouse, which include binding sites for transcription factors that are important for craniofacial development, such as TFAP2B and TCF4[100,101,103]. While human-specific signatures would need to be validated in suitable human tissue- or cell-based assays to conclusively confirm *bona fide* lineage-specific *in vivo* activity, these data suggest that profiling human tissues is an effective way to identify candidate regions with human-biased enhancer signatures. Our compendium of human craniofacial enhancers expands previously reported[26,53] human craniofacial enhancer catalogs, by approximately 5,000 newly identified—enhancers for weeks 7-8 of human craniofacial development primarily identified in this study. When comparing with craniofacial enhancers identified in previous studies, we find that our data provides independent confirmation for 37% of reported primate enhancers and 15% of human-biased enhancers[18]. Of the 13,983 reproducible human enhancers described in this study, 47% showed evidence of enhancer-associated RNA signatures in the FANTOM5 database[81,83]. In contrast, when restricting this analysis to a more differentiated craniofacial cell type available in FANTOM5 (human embryonic palatal mesenchyme)[81,83,85], we observed enhancer RNA signatures for only 3.8% of our 13,983 predicted enhancers, likely reflecting that

this cell type is only one of many that were present in our tissue samples (Supplementary Data 2 and 21). Generally, the imperfect overlap of craniofacial enhancers identified in some of these studies may be due to differences in epigenomic profiles from primary tissues comprising the entire face versus *in vitro* differentiation of a specific lineage such as neural crest or palatal mesenchyme. Additional possible sources of variation include differences in experimental modalities (H3K27ac binding versus measurements of enhancer RNA), and imperfect matching of *in vivo* developmental stages with *in vitro* models. In this study, we leveraged genome-wide profiling of H3K27ac binding for identification of enhancers. The tissue-specific validation rate we observe is comparable to that we observed in other studies using similar methods for prediction of *in vivo* enhancer activities[25]. We note that alternative experimental approaches that measure non-coding RNAs or massively-parallel reporter assays with or without mutational screens can also be used for identifying putative enhancer elements and may be useful for capturing additional craniofacial candidate enhancers[108,110].

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 gene 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 for major cell types in face morphogenesis.

While previous single-cell studies of the developing face from other animal models have described extensive annotations for ectomesenchyme, we find that the complexity of cell types in the developing mouse face poses some challenges in this respect. In particular, in comparing several mouse orthologs of the embryonic zebrafish ectomesenchymal markers[54] expressed in *ScanFaceX* that show relatively high accessibility in *ScanFaceN* in neural crest-derived populations (Supplementary Figure 16), regional identities marked by specific genes are not obviously delineated in *ScanFaceX*. These differences in cell type distributions and marker gene activity may be explained by the extent of differentiation, growth rate, evolving cell states, and developmental timing underlying craniofacial morphogenesis. One of the limitations of present methods is the ability to capture low-expressing genes or rarer cell populations among other technical and statistical challenges[112,114]. We also note that utilizing cell type annotations from *ScanFaceX* and integrating those with single-cell open chromatin data provides correlative but not definitive evidence for the target genes of a given enhancer, which requires verification through complementary experimental methods [116,118,119]. 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 activity of three craniofacial enhancers during embryonic development at single-cell resolution. This approach illustrates how these methods can be combined to determine the *in vivo* specificity of individual enhancers and relate their activity to cell type-specific expression of their putative target genes. We note that *in vivo* transgenic reporter assays can demonstrate that an enhancer is sufficient to drive expression in a tissue or cell type of interest, but integration into a safe harbor locus such as H11 removes the enhancer from the full epigenomic and three-dimensional context of its native locus[120]. Therefore, reporter expression

may not fully recapitulate the full endogenous activity of a given enhancer in its original genomic location.

All of these data are also available in FaceBase and the VISTA Enhancer Browser for community use[1,84,121]. In summary, our work provides a multifaceted and expansive resource for studies of craniofacial enhancers in human development and disease.

Methods

Ethics Statement

This research complies with all relevant ethical regulations. All aspects involving human tissue samples were reviewed and approved by the Human Subjects Committee at Lawrence Berkeley National Laboratory (LBNL) Protocol Nos. 00023126 and 00022756. All animal work was reviewed and approved by the LBNL Animal Welfare Committee.

Human embryonic face samples were obtained from the Human Developmental Biology Resource's Newcastle site (HDBR, hdbr.org), in compliance with applicable state and federal laws. The National Research Ethics Service reviewed the HDBR study under REC Ref 23/NE/0135, and IRAS project ID: 330783 in compliance with requirements from the National Health Services for research within the UK and overseas. HDBR is a non-commercial entity funded by the Wellcome Trust and Medical Research Council. Fetal tissue donation is confidential, anonymized, completely voluntary with fully informed and explicitly documented written consent, and the participants do not receive compensation. In accordance, no identifying information for human samples in this study was shared by HDBR. More information about HDBR policies and ethical approvals can be accessed at https://www.hdbr.org/ethical-approvals.

Human Samples

Primary data from embryonic whole face samples at post-conception weeks 7 and 8 were generated in this study. Whole face region excluding eyes was dissected at HDBR (Supplementary Figure 1), and all embryonic samples were shipped on dry ice and stored at −80°C until processed. Embryos of both sexes were included in the experiments. However, we did not consider embryo sex as a variable in our studies since craniofacial development is expected to show minimal differences at these early stages of development. 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. RNA-seq data for four samples at CS18, one sample at CS19, seven samples at CS22, and four samples at CS23 were generated in this study and analysis from a subset of these is presented. Processed data for CS 13-17, and CS20 was obtained from previously published studies [26] and included in our downstream integrative analyses. All datasets are listed in Supplementary Data 1.

Animal Studies and Experimental Design

Mice used for this study were housed at the LBNL Animal Care Facility, which is fully accredited by AAALAC International. Mice were housed on a 12-hour light-dark cycle in standard micro-isolator cages on hardwood bedding with enrichment consisting of crinkle cut naturalistic paper strands. Mice were maintained on ad libitum PicoLab Rodent Diet 20 (5053) and water supply with 30-70% environmental humidity and temperature of 20 – 26.2°C. All mice were health checked and monitored daily for food and water intake by trained personnel. Animals of both sexes were used in the analysis. Sample size selection and randomization strategies were followed based on our experience of

performing transgenic mouse assays for ~3000 published enhancer candidates[65,66].

Transgenic Mouse Assays *in vivo*

60 candidate human enhancer elements were selected based on a combination of criteria including overlap with ATAC-seq peaks, strength of H3K27ac active enhancer signatures, non-mouse annotated regions, and vicinity of genes with known or proposed roles in craniofacial development based on human genetics and/or mouse knockout studies (e.g., genes listed under term “abnormality of the face”; HP:0000271 in Human Phenotype Ontology[124] or “craniofacial abnormalities”; MP:0000428 in the Mammalian Phenotype Browser[125]).

Mouse enhancer elements mm2280, mm2281 mm2282, and mm2285 were selected based on conservation criteria or predicted from single-cell gene expression read outs and single-cell chromatin accessibility profiles. Transgenic enhancer-reporter assays were performed per established protocols [65,66]. Briefly, a minimal *Shh* promoter and reporter gene were integrated into a non-endogenous, safe harbor locus [66] in a site-directed transgenic mouse assay. The selected genomic region corresponding to the selected enhancer element was PCR amplified from human or mouse genomic DNA where applicable**;** the PCR amplicon was cloned into a *lacZ*-reporter vector (Addgene #139098) using Gibson assembly (New England Biolabs) [126]. 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 our pronuclear injection protocol[66]. Briefly, sgRNAs (50 ng/μl) targeting the *H11* locus and Cas9 protein (Integrated DNA Technologies catalog no. 1081058; at final concentration of 20 ng/μl) was mixed in microinjection buffer (10 mM Tris, pH 7.5; 0.1 mM EDTA). The mix was injected

into the pronuclei of single cell stage fertilized FVB/NJ (Jackson Laboratory; Strain#:001800) embryos obtained from the oviducts of super-ovulated 7–8 weeks old FVB/NJ females mated to 7-8 weeks old FVB/NJ males. The injected embryos were cultured in M16 medium supplemented with amino acids at 37 °C under 5% CO_2 for ~2 hours and transferred into the uteri of pseudo-pregnant CD-1 (Charles River Laboratories; Strain Code: 022) surrogate mothers. Embryos were collected for downstream experiments at embryonic days 10.5 through 15.5 (Theiler stages 17-23). 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 (Supplementary Figure 15) are shown in Supplementary Data 5 and 22 respectively.

For transgenic experiments demonstrating 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 using established methods in our laboratory [127]. 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 lyzed and chromatin was sonicated using a Bioruptor 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). For the ChIP-seq datasets at CS13-15, CS17 and CS20[26], publicly available and post-mapped TagAlign files were used. Peak calling was performed using MACS2 (v2.2.4; --broad flag, q-value < 0.05); upon broad peak calling and applying the FDR filter, bed files were combined and merged using bedtools[102]. A combined peak set was called by merging peaks from all samples, and overlapping peaks for each sample were counted using overlap_peaks.py. 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. For a break-down of samples as well as peaks per week, see Supplementary Data 8.

We note that the use of human embryonic tissue samples, which are typically derived from individual or a small number of fetal tissue donations, can introduce variability regarding tissue dissection and genetic heterogeneity. While some of these sources of variation are unavoidable, we tried to minimize potential batch effects. To make the analysis as comparable as possible, we down-sampled the number of input reads and the read length to a common denominator (15 million and 50 bp, respectively), and used the standard ENCODE peak-calling pipeline. To assess the possible presence of batch effects between data from these studies, we compared temporal transitions between weeks (Supplementary Figure 4). In this analysis, we did not observe discontinuities specifically associated with the transition time points between batches. While we cannot exclude the presence of some batch effects, this result suggests that study-specific batch effects do not confound our temporal dynamics analysis in major ways.

ATAC-seq

Embryonic samples were processed for ATAC-seq using standard methods [127]. In short, harvested tissues were lysed, centrifuged for 10min at 500 x g, at 4ºC, 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 (AATGATACGGCGACCACCGAGATCTACACNNNNNNNNTCGTCGGCAGCGTC) and 2 (CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTCTCGTGGGCTCGG) 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/). Briefly, reads were aligned with the Bowtie2 aligner and filtered to remove unmapped and non-primary alignments, low quality reads as well as PCR duplicates. A subsample of 15 million reads was used as input to peak-calling, adjusted for Tn5 shift reads and sets of biological samples were assembled along with pseudoreplicates. Peak calls excluded ENCODE blacklist regions[104] and peaks were assessed at an Irreproducible Discovery Rate of 0.05.

RNA-seq

Samples were processed for RNA-seq and libraries were generated with established protocols[105,127]. 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://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 hg38, and GENCODE v24 gene annotations. Mouse datasets were analyzed using mouse reference genome version mm10 and GENCODE M4 gene annotations.

rGREAT Ontology Analyses

To identify human phenotype ontology terms enriched in our list of 13,983 reproducible human craniofacial enhancers, we ran rGREAT[44] (Bioconductor version: Release 3.17) that performs GREAT[106] analysis (http://great.stanford.edu) on non-coding regions to predict their functions based on annotations of nearby genes. Following parameters were used from the GREAT tool: a default of 5kb upstream and 1kb downstream basal plus extension for proximal regulatory regions, up to 10 kb for distal regions, and curated regulatory domains were included. A background of whole genome hg38, a cut-off based on Binomial False Discovery Rate < 0.01, and Fold Enrichment > 2 was applied to retain the top terms (Supplementary Data 6).

Enhancer-Target Gene Predictions

We intersected our list of 13,983 reproducible human enhancers with publicly available long-range chromatin interaction data derived from promoter capture HiC for approximately 19,000 promoters in human embryonic stem cells[55]. Genomic coordinates of the interacting fragments were converted to hg38, the predicted target gene and extent of overlap with the human enhancers from this study are reported in Supplementary Data 7. For 3,005 chromatin segments containing predicted human craniofacial enhancers, and interacting with the promoters of 2,921 genes, we performed Spearman's Ranked Correlation

Coefficient (SRCC) analysis between enhancer signal intensities (H3K27ac ChIP-seq, Trimmed Mean of M-values normalized) and gene expression counts (RNA-seq) of the assigned target genes (Supplementary Data 7) for predicted enhancer:target gene pairs versus all other pairs. We performed this analysis for combined as well as individual activity windows shown in Figure 2b for a subset of matched samples, i.e., five instances where enhancer predictions and gene expression data were available from identical human embryonic face samples, namely CS18_12612, CS18_12695, CS19_12696, CS22_11963, and CS23_12492 (Supplementary Figure 3, Supplementary Data 8). Mann-Whitney U test statistic was used to ascertain significance between the correlated enhancer:target gene pairs of interest versus all other pairs.
We note that the correlation is highly significant but quantitatively moderate. This is likely due to technical factors including imperfect enhancer-gene associations, target gene predictions not being available for all enhancers, differences arising from comparing predictions from human embryonic stem cells versus complex primary human embryonic tissue encompassing varying stages of differentiation, not excluding cases with redundant enhancers acting on the same gene(s), and uncertainty about the expected quantitative correlation between H3K27ac signal intensity at an enhancer and the expression level of a target gene. For the correlation for class "week-specific" in Supplementary Figure 3b, the comparisons may not be significant due to the lack of capability of SRCC to detect patterns driven by one or two data points.

GWAS Data

The NHGRI-EBI Catalog of Genome-wide association studies[107] 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 Unites

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 tool[109] 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 (Supplementary Data 9-10) was intersected with 13,983 reproducible human enhancers derived from primary embryonic bulk face between CS13-23. We have partitioned a total of 14,137,504 SNPs from the dbSNP155[111,113] catalog by their association with normal face variation or human disease and overlap with reproducible fetal human face enhancers described in this work. We found that 605 out of 27386 (2.3%) of normal face variation- or human disease-associated SNPs overlapped the peaks, while only 245,727 out of 14,083,942 (1.8%) of non-associated SNPs did. The overlap was significantly different from random expectation with an odds ratio of 1.27 (Pearson's Chi-squared test with Yates' continuity correction: X-squared = 34.102, df = 1, p-value = 5.229e-09).

Intersecting VISTA Catalog with Predicted Craniofacial Enhancers

We intersected a subset of 130 human craniofacial regulatory elements (out of 3,193 total curated) in the VISTA Enhancer Browser with 13,983 reproducible human candidate enhancers for weeks 4-8 from this study requiring a minimum 100bp overlap (Supplementary Figure 2, and Supplementary Data 4). We note that VISTA enhancers are not a random sample of the genome and are intentionally picked for their high levels of evolutionary conservation, high levels of epigenomic signal in embryos, lower repeat content, and proximity to genes known to regulate embryonic development.

Single-cell RNA-seq

Both wild-type FVB/NJ crosses (ages 7-8 weeks), as well as transgenic mice harboring the *Hsp68* promoter and mCherry reporter at *H11* locus and generated as described earlier in Methods were used. 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 if applicable. 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 Multi-seq strategy[115] 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 analyses

[117,128,129]. Adhering to the standard pre-processing workflow and quality control, cells with unique feature counts between >200 and < 5%mitochondrial reads were retained. Based on the inspection of UMI/gene count plots, the UMI range which preserved the main group of cells and excluded both droplet debris and likely clumps of cells was established for each sample separately (2,000-4,000 minimum, 15,000-60,000 maximum). For scRNA-seq, samples were integrated using standard Seurat procedure; *SelectIntegrationFeatures* function was run on a list of all 9 samples to be integrated to find 3,000 most variable features. *mCherry* transcripts, genes on chromosomes X or Y (Gencode vM24) and cells expressing >5% mitochondrial genes (with names starting with *mt*) were removed from that list. *PrepSCTIntegration*, *FindIntegrationAnchors* and *IntegrateData* functions were run to obtain an integrated dataset. 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, with a median of 1,659 and a range of 500 - 8,840 genes expressed per cell (Supplementary Figure 5), and a range of 474-9,148 cells for the smallest to largest clusters (Supplementary Data 16).

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 published large single-cell gene expression dataset [130] 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 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 dataset [122]. 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. We note that the label "other craniofacial" encompasses a mix of cells with the following descriptive terms retained from the auto-referencing steps: palate development, roof of mouth, mesenchyme, and premature oligodendrocytes. (Supplementary Figures 7, 9-11, Supplementary Data 14-16).

Single-nucleus ATAC-seq

Wild-type FVB/NJ crosses (ages 7-8 weeks) were used to generate mouse embryos for each of the developmental stages e10.5-15.5. Face tissue was dissected, flash frozen in liquid nitrogen (N2) 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 strategy [80]. 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 2× 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: snapATAC2 (version 1) package was used to perform read counting and cell clustering for both all-tissue clustering and tissue-level clustering [131]. 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). To stabilize the variance and reduce the impact of noise, we normalized the read coverage of all bins with log10 (count+1), applied Z-score transformation to ensure that each feature contributes equally to downstream analyses, 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, given the computational limitations at the time of analysis, and based on evidence provided by SnapATAC[131], 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 (leiden clustering resolution =1) 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. The *FindTransferAnchors* function in Seurat uses unsupervised identification of anchors representing cells from separate datasets, with the assumption that these cells are derived from shared biological states[132]. 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 marker gene.

Comparing Human Craniofacial Enhancers with Previously Reported Enhancer Catalogs

We compared human enhancers identified in this study with a set of 5,000 primate enhancers profiled from cranial neural crest cell differentiation using both chimpanzee and human cells and a list of 1,000 human-biased enhancers[18]. Genomic coordinates of these enhancers were converted to hg38 using LiftOver and intersected with our list of 13,933 reproducible human enhancers. Similarly, enhancers identified by Cap Analysis of Gene Expression (CAGE) including those from normal human embryonic palatal mesenchyme (HEPM:CNhs11894) cells were obtained from the FANTOM5 database[81,83,85]. Genomic coordinates of the enhancer lists from FANTOM5 were converted to hg38 and intersected with the 13,983 human reproducible craniofacial enhancers from this study. Results of these analysis are reported in Supplementary Data 2.

Statistics and Reproducibility

Statistical analyses are described in detail in the Methods section above. For human embryonic face samples, we performed experiments with biological replicates as follows: three at CS18, one at CS19, two at CS22 (with two technical replicates for one of two samples), one at CS19 for ChIP-seq. We performed experiments with two biological replicates at CS18, and one each at CS19, and CS22-23 for ATAC-seq; four replicates at CS18, one at CS19, seven at CS22, and four at CS23 for RNA-seq. For single-cell experiments of the mouse face, we performed experiments for eight biological replicates at E11.5, and four replicates each at E12.5 and E13.5 respectively for scRNA-seq, while single samples at each of the six mouse embryonic stages (E10.5, E11.5, E12.5, E13.5, E14.5, and E15.5) were processed for snATAC-seq. For transgenic assays primarily performed and reported in this study, we confirmed results in at least two independent animals (range 2-10 positive results) and used criteria consistent with our site-directed transgenesis pipeline established for the VISTA Enhancer Browser. Individuals who qualitatively assessed the results of *in vivo* transgenic reporter assays were blinded to genotyping information. For all other experiments, the investigators were not blinded to allocation during experiments and outcome assessment. No statistical method was used to pre-determine sample size. No data that passed quality control criteria for experiments were excluded from the analyses. The experiments were not randomized. Unless otherwise stated, default parameter settings were employed for any software tool that was used in the analyses. 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)/ R version 4.1.0.

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[84,121,133,134] Consortium (facebase.org), and can be found under the following records: RNA-seq, ChIP-seq and ATAC-seq analysis of human fetal tissue. *FaceBase Consortium* Accession: FB00001358 https://doi.org/10.25550/3C-4G62. Single-cell RNA-seq and single-nucleus ATAC-seq analysis of mouse embryonic tissue. *FaceBase Consortium* Accession: FB00001359 https://doi.org/10.25550/3C-4R98. These data are additionally deposited in NCBI's Gene Expression Omnibus[135,136] and are accessible through GEO Series Accession GSE235858 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=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 https://enhancer.lbl.gov/. Source data are provided in the Source Data File with this paper, and as a publicly accessible Seurat/R objects as applicable.

Code Availability

No previously unreported custom computer code, mathematical algorithm or software were used in the analyses of data presented in this study. Current community-accepted and benchmarked bioinformatic methods were used and are appropriately cited in the main text and Methods.

Acknowledgements

This work was supported by U.S. National Institutes of Health (NIH) grant

(R01DE028599 and U01DE024427) to A.V. 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 and fetal material was provided by the Joint MRC / Wellcome (MR/R006237/1, MR/X008304/1 and 226202/Z/22/Z) Human Developmental Biology Resource (www.hdbr.org). F.D. was supported by the Swiss National Science Foundation grant P400PB_194334.

Author Contributions Statement

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. B.R. supervised snATAC-seq experiments and integrative analysis of snATAC-seq and scRNA-seq 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.

Competing Interests Statement

Bing Ren is a co-founder of Arima Genomics, Inc, and Epigenome Technologies, Inc. The remaining authors declare no competing interests.

Figures

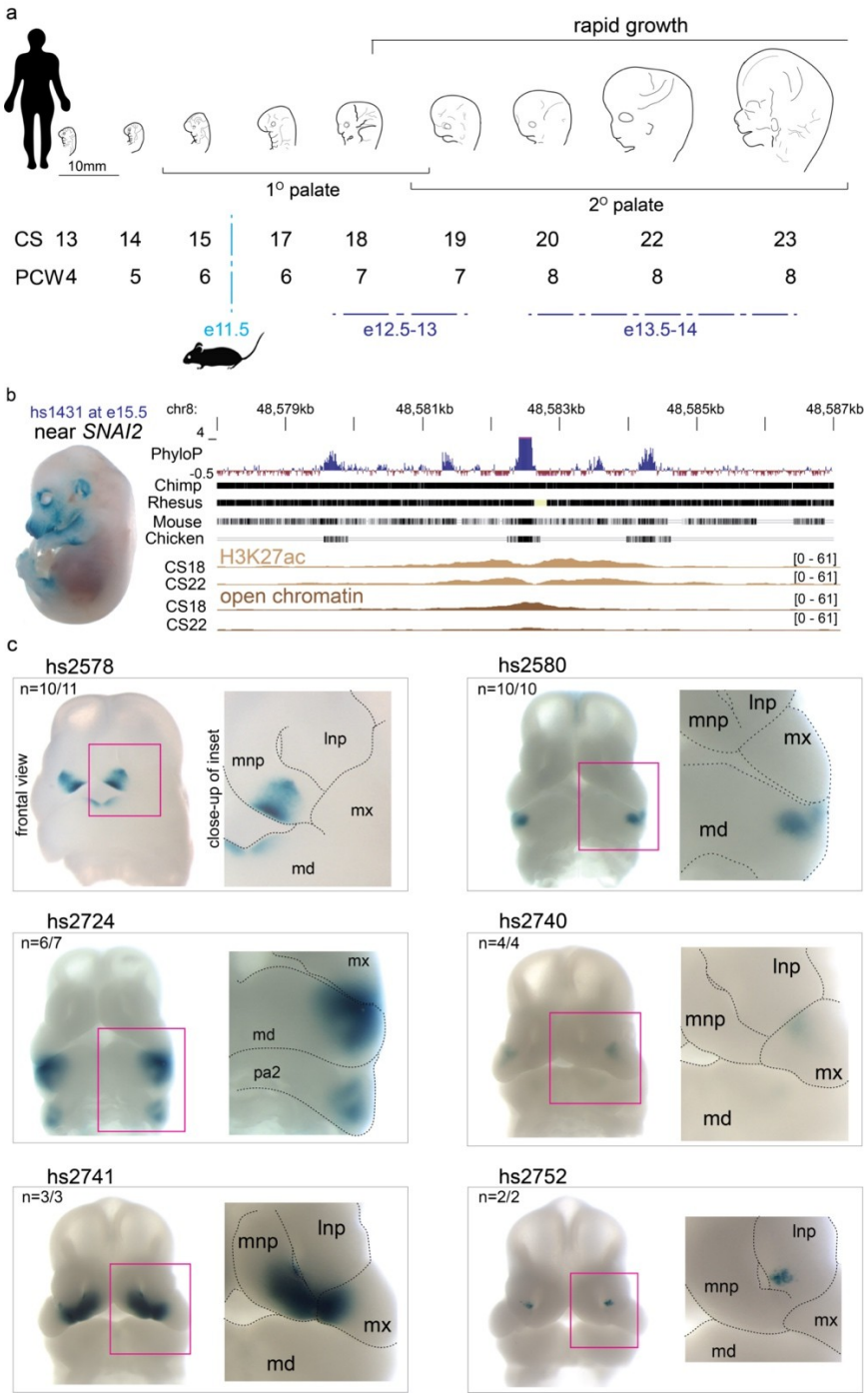

Figure 1. Developmental enhancers in human craniofacial morphogenesis. a. Developmental time points coinciding with critical windows of craniofacial morphogenesis are shown by Carnegie stage (CS) and post-conceptional week (PCW) in humans, and comparable embryonic (e) stages for mouse are shown in embryonic 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), mandibular process (md), and pharyngeal arch 2 (pa2). n, reproducibility of each pattern across embryos resulting from independent transgenic integration events.

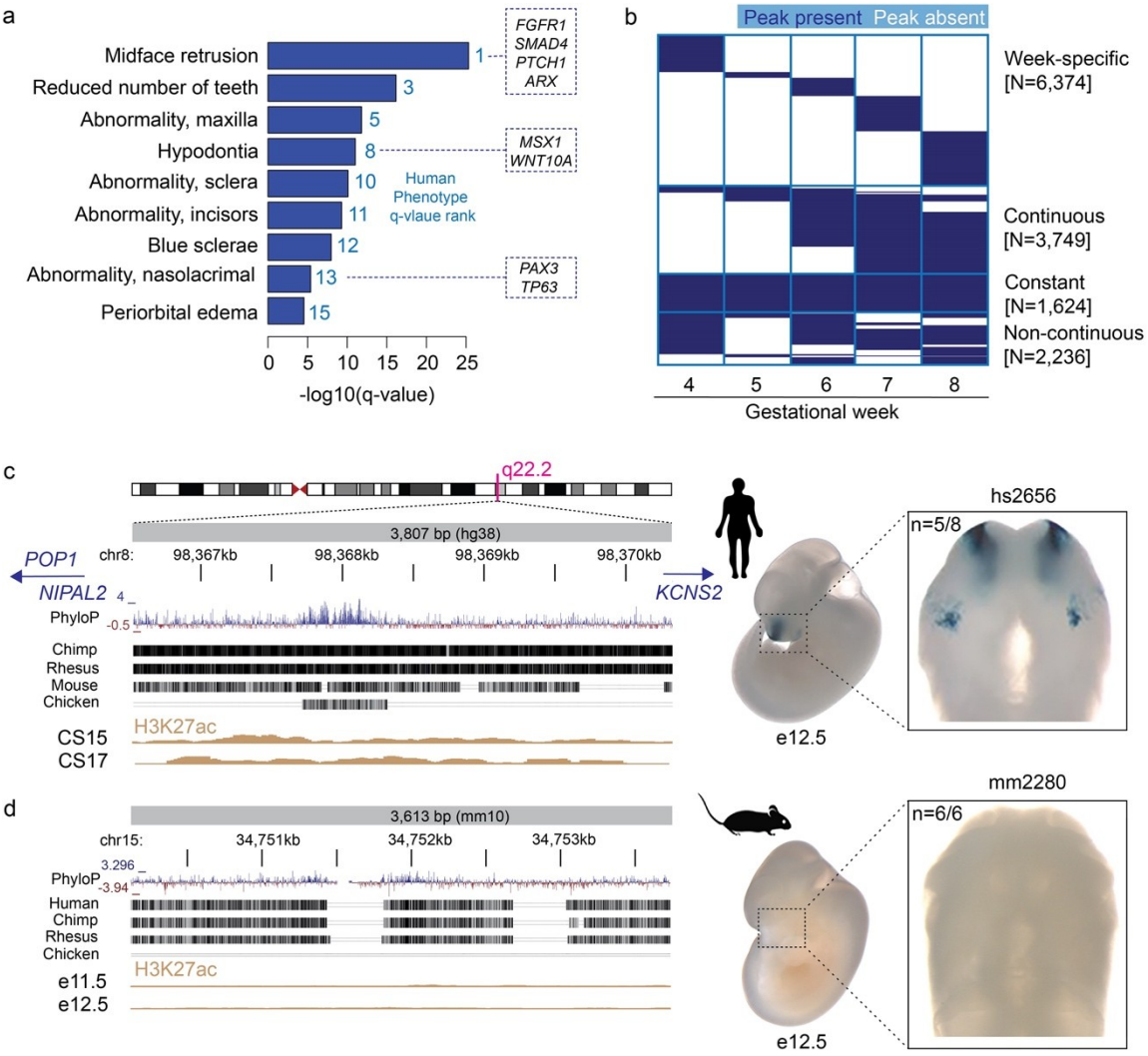

Figure 2. Developmental dynamics and conservation of human craniofacial enhancers. a. Results of rGREAT ontology analysis for 13,983 reproducible human craniofacial enhancers, ranked by Human Phenotype q-value. The ontology terms indicate that our predictions of human craniofacial enhancers are enriched 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. Source data are provided as part of Supplementary Data 2 and in Source Data file. **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.

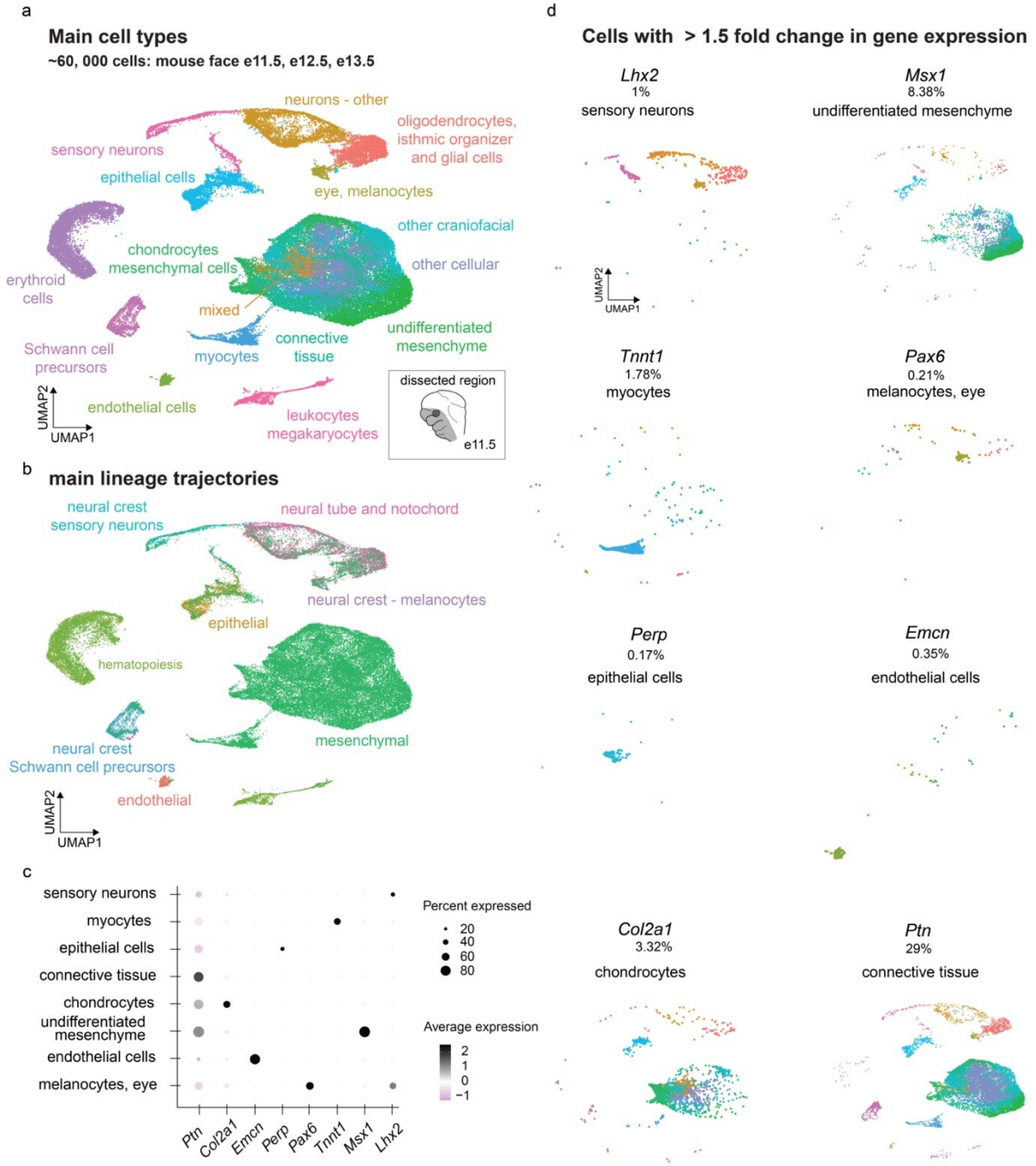

Figure 3. Gene expression in the mammalian craniofacial complex at single cell resolution. a. Uniform Manifold Approximation and Projection (UMAP) clustering, color-coded by inferred cell types across clusters from aggregated scRNA-seq for the developing mouse face at embryonic days 11.5-13.5, for 57,598 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). See Supplementary Figure 11 for additional details. **d.** UMAP plots comprising cells with >1.5-fold gene expression for marker genes representing specific cell types as

shown in (a) and (c). Source data are provided as a publicly accessible Seurat/R object file, see Data Availability Statement for details.

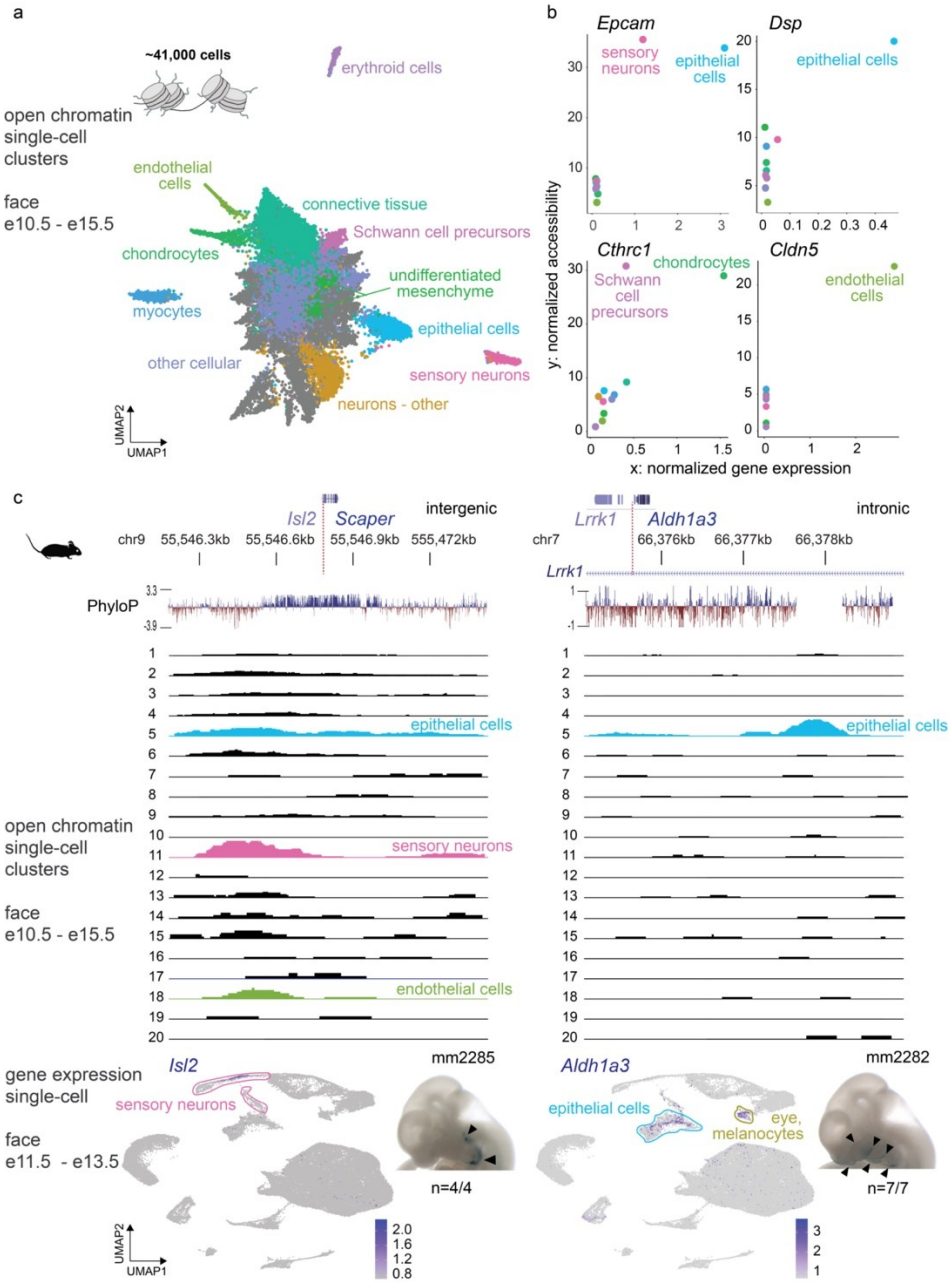

Figure 4. Differential chromatin accessibility at craniofacial *in vivo* enhancers correlates with expression of nearby genes. a. Unbiased clustering (UMAP) of open chromatin regions from snATAC-seq of the developing mouse face for stages e10.5-15.5 for approximately 41,000 cells. The cell types are 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. Colors in (**4b**) and the individual snATAC-seq tracks in (**4c**) correspond to the color code used in (**4a**). UMAP of *ScanFaceX* data shows expression of *Isl2* and *Aldh1a3* (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; black arrowheads point towards stained regions. n, reproducibility of each pattern across embryos resulting from independent transgenic integration events. Source data for 4b are provided as a Source Data file.

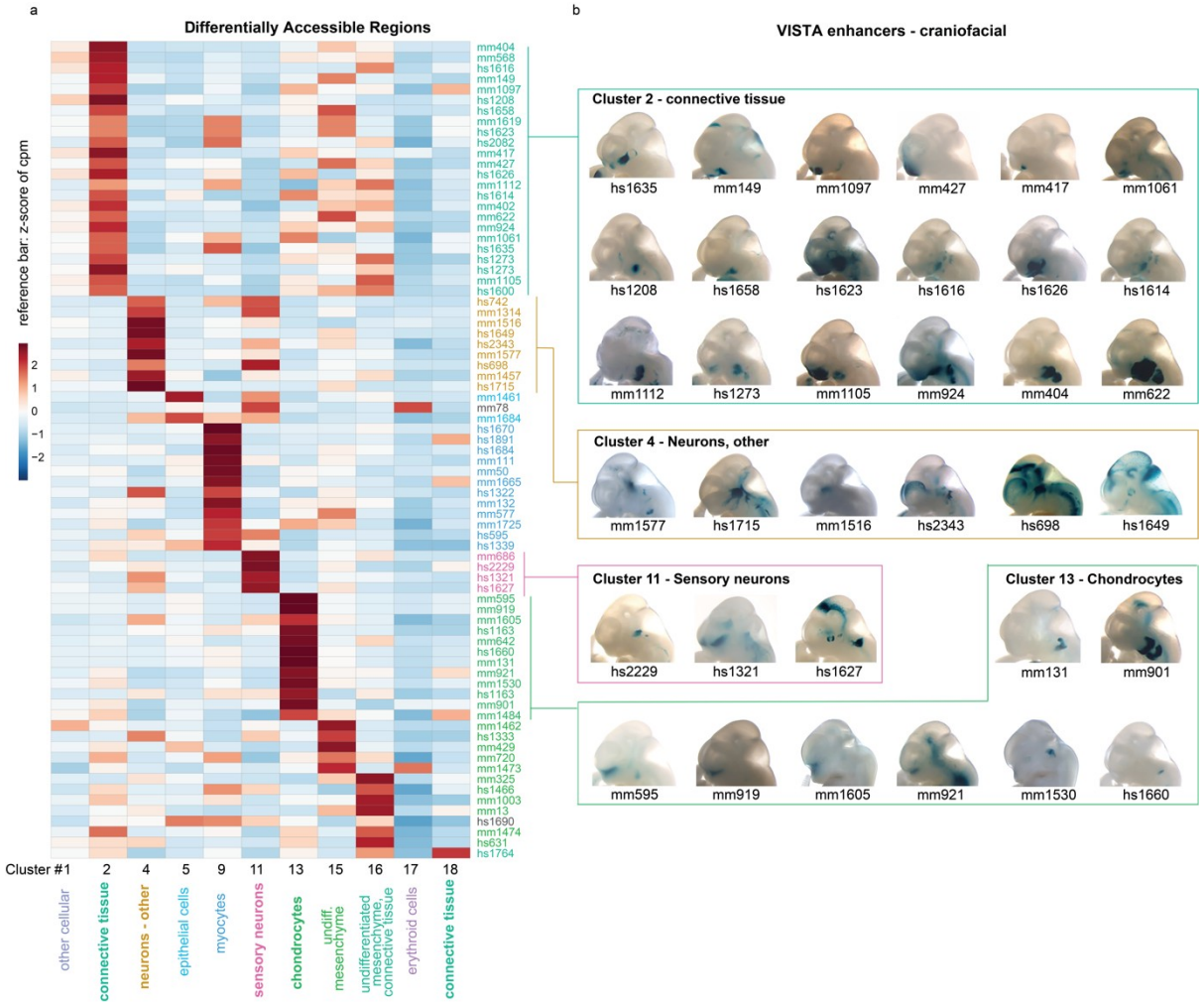

Figure 5. Correlating Cell Population-Resolved Enhancer Signatures with Enhancer *in vivo* Activity Patterns. a. Heatmap indicates the chromatin accessibility of 77 craniofacial *in vivo* VISTA enhancers in 11 major clusters representing predicted cell types. cpm: counts per million. **b.** Representative 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. Source data for 5a are provided as a Source Data file.

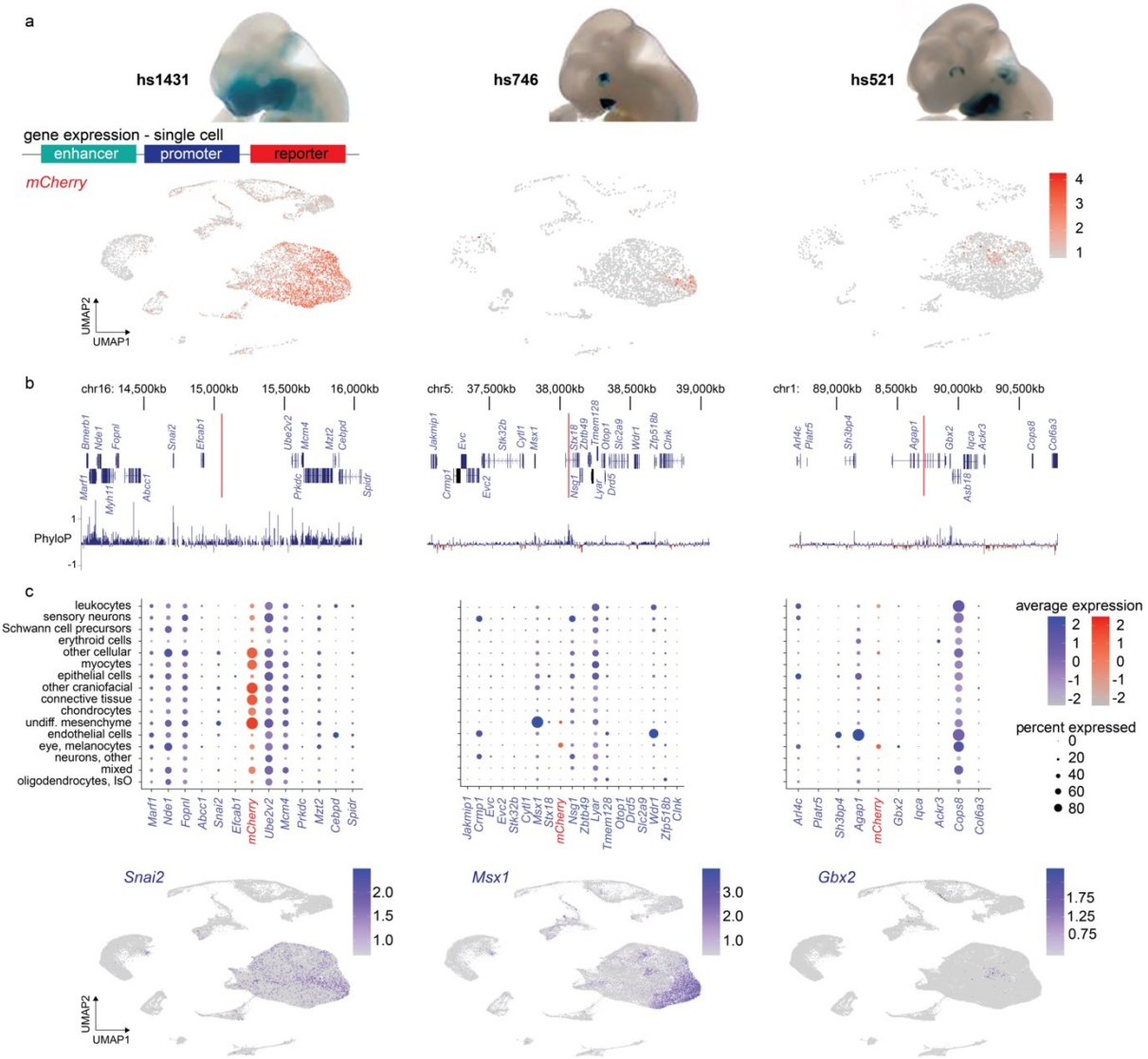

Figure 6. Enhancer activity at single-cell resolution. a. *in vivo* activity pattern of select craniofacial enhancers (hs1431, hs746, hs521) at e11.5, visualized by *lacZ*-reporter assays (top). In separate experiments, the same enhancers were coupled to an *mCherry*-fluorescent reporter gene and examined by scRNA-seq of craniofacial tissues of resulting embryos. UMAPs show enhancer-driven *mCherry* expression (see **Figure 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.** Seurat-based average expression of genes in the vicinity of the respective enhancers, and proportion (percent) of cells expressing those genes in annotated cell types. Enhancer-driven *mCherry* signal is plotted in the center in between the names of the two genes whose promoters are closest to its location within the genome. For example, for hs1431, mCherry is highly expressed (indicated by red color intensity) in clusters labeled “other cellular”,

"myocytes", "skeletal, other", "connective tissue", and "undifferentiated mesenchyme", while it is also expressed in a larger proportion of cells (indicated by greater diameter of the circles) in those same clusters. In the same plot, *Snai2* is highly expressed (indicated by blue color intensity) in a subset of cells (indicated by lesser diameter of circles) in identical clusters as compared to mCherry. Bottom panels show expression of *Snai2*, *Msx1,* and *Gbx2* as likely candidate target genes for each of the enhancers hs1431, hs746 and hs521 across UMAPs. undiff.: undifferentiated IsO: Isthmic Organizer Cells. Source data are provided as a publicly accessible Seurat/R object file, see Data Availability Statement for details.

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. Morrison, J.A., McLennan, R., Teddy, J.M., Scott, A.R., Kasemeier-Kulesa, J.C., Gogol, M.M., and Kulesa, P.M. (2021). Single-cell reconstruction with spatial context of migrating neural crest cells and their microenvironments during vertebrate head and neck formation. Development *148*. 10.1242/dev.199468.

28. 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.

29. Morrison, J.A., McLennan, R., Wolfe, L.A., Gogol, M.M., Meier, S., McKinney, M.C., Teddy, J.M., Holmes, L., Semerad, C.L., Box, A.C., et al. (2017). Single-cell transcriptome analysis of avian neural crest migration reveals signatures of invasion and molecular transitions. Elife *6*, e28415.

30. 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.

31. Saunders, L.M., Mishra, A.K., Aman, A.J., Lewis, V.M., Toomey, M.B., Packer, J.S., Qiu, X., McFaline-Figueroa, J.L., Corbo, J.C., Trapnell, C., et al. (2019). Thyroid hormone regulates distinct paths to maturation in pigment cell lineages. Elife *8*. 10.7554/eLife.45181.

32. 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.

33. Zalc, A., Sinha, R., Gulati, G.S., Wesche, D.J., Daszczuk, P., Swigut, T., Weissman, I.L., and Wysocka, J. (2021). Reactivation of the pluripotency program precedes formation of the cranial neural crest. Science *371*. 10.1126/science.abb4776.

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

35. Krivanek, J., Soldatov, R.A., Kastriti, M.E., Chontorotzea, T., Herdina, A.N., Petersen, J., Szarowska, B., Landova, M., Matejova, V.K., Holla, L.I., et al. (2020). Dental cell type atlas reveals stem and differentiated cell types in mouse and human teeth. Nat. Commun. *11*, 4816.

36. 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.

37. Soldatov, R., Kaucka, M., Kastriti, M.E., Petersen, J., Chontorotzea, T., Englmaier, L., Akkuratova, N., Yang, Y., Häring, M., Dyachuk, V., et al. (2019). Spatiotemporal structure of cell fate decisions in murine neural crest. Science *364*, eaas9536.

38. 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.

39. Krivanek, J., Soldatov, R.A., Kastriti, M.E., Chontorotzea, T., Herdina, A.N., Petersen, J., Szarowska, B., Landova, M., Matejova, V.K., Holla, L.I., et al. (2020). Dental cell type atlas reveals stem and differentiated cell types in mouse and human teeth. Nat. Commun. *11*, 4816.

40. 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.

41. 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.

42. Li, H., Jones, K.L., Hooper, J.E., and Williams, T. (2019). The molecular anatomy of mammalian upper lip and primary palate fusion at single cell resolution. Development *146*. 10.1242/dev.174888.

43. Lencer, E., Prekeris, R., and Artinger, K.B. (2021). Single-cell RNA analysis identifies pre-migratory neural crest cells expressing markers of differentiated derivatives. Elife *10*, e66078.

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

45. Xu, J., Liu, H., Lan, Y., Adam, M., Clouthier, D.E., Potter, S., and Jiang, R. (2019). Hedgehog signaling patterns the oral-aboral axis of the mandibular arch. Elife *8*, e40315.

46. Elliott, K.H., Balchand, S.K., Bonatto Paese, C.L., Chang, C.-F., Yang, Y., Brown, K.M., Rasicci, D.T., He, H., Thorner, K., Chaturvedi, P., et al. (2023). Identification of a heterogeneous and dynamic ciliome during embryonic development and cell differentiation. Development *150*. 10.1242/dev.201237.

47. Farmer, D.T., Mlcochova, H., Zhou, Y., Koelling, N., Wang, G., Ashley, N., Bugacov, H., Chen, H.-J., Parvez, R., Tseng, K.-C., et al. (2021). The developing mouse coronal suture at single-cell resolution. Nat. Commun. *12*, 1–14.

48. Piña, J.O., Raju, R., Roth, D.M., Winchester, E.W., Chattaraj, P., Kidwai, F., Faucz, F.R., Iben, J., Mitra, A., Campbell, K., et al. (2023). Multimodal spatiotemporal transcriptomic resolution of embryonic palate osteogenesis. Nat. Commun. *14*, 1–15.

49. Holmes, G., Gonzalez-Reiche, A.S., Saturne, M., Motch Perrine, S.M., Zhou, X., Borges, A.C., Shewale, B., Richtsmeier, J.T., Zhang, B., van Bakel, H., et al. (2021). Single-cell analysis identifies a key role for Hhip in murine coronal suture development. Nat. Commun. *12*, 1–16.

50. Kolla, L., Kelly, M.C., Mann, Z.F., Anaya-Rocha, A., Ellis, K., Lemons, A., Palermo, A.T., So, K.S., Mays, J.C., Orvis, J., et al. (2020). Characterization of the development of the mouse cochlear epithelium at the single cell level. Nat. Commun. *11*, 1–16.

51. Ji, Y., Zhang, S., Reynolds, K., Gu, R., McMahon, M., Islam, M., Liu, Y., Imai, T., Donham, R., Zhao, H., et al. (2021). Single cell transcriptomics and developmental trajectories of murine cranial neural crest cell fate determination and cell cycle progression. bioRxiv, 2021.05.10.443503. 10.1101/2021.05.10.443503.

52. Gu, R., Zhang, S., Saha, S.K., Ji, Y., Reynolds, K., McMahon, M., Sun, B., Islam, M., Trainor, P.A., Chen, Y., et al. (2022). Single-cell transcriptomic signatures and gene regulatory networks modulated by Wls in mammalian midline facial formation and clefts. Development *149*. 10.1242/dev.200533.

53. Yankee, T.N., Oh, S., Winchester, E.W., Wilderman, A., Robinson, K., Gordon, T., Rosenfeld, J.A., VanOudenhove, J., Scott, D.A., Leslie, E.J., et al. (2023). Integrative analysis of transcriptome dynamics during human craniofacial development identifies candidate disease genes. Nat. Commun. *14*, 4623.

54. Fabian, P., Tseng, K.-C., Thiruppathy, M., Arata, C., Chen, H.-J., Smeeton, J., Nelson, N., and Crump, J.G. (2022). Lifelong single-cell profiling of cranial neural crest diversification in zebrafish. Nat. Commun. *13*, 1–13.

55. Jung, I., Schmitt, A., Diao, Y., Lee, A.J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S., et al. (2019). A compendium of promoter-centered long-range chromatin interactions in the human genome. Nat. Genet. *51*, 1442–1449.

56. 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.

57. Nassar, L.R., Barber, G.P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., et al. (2023). The UCSC Genome Browser database: 2023 update. Nucleic Acids Res. *51*, D1188–D1195.

58. 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.

59. 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.

60. Raviv, S., Bharti, K., Rencus-Lazar, S., Cohen-Tayar, Y., Schyr, R., Evantal, N., Meshorer, E., Zilberberg, 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.

61. 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.

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

63. 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.

64. 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.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

70. 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.

71. 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.

72. 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.

73. 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.

74. 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.

75. 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.

76. 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.

77. 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.

78. 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.

79. Mitsiadis, 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.

80. 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.

81. Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S., et al. (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. *16*, 22.

82. Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, 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.

83. Noguchi, S., Arakawa, T., Fukuda, S., Furuno, M., Hasegawa, A., Hori, F., Ishikawa-Kato, S., Kaida, K., Kaiho, A., Kanamori-Katayama, M., et al. (2017). FANTOM5 CAGE profiles of human and mouse samples. Sci Data *4*, 170112.

84. 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.

85. Yoneda, T., and Pratt, R.M. (1981). Mesenchymal cells from the human embryonic palate are highly responsive to epidermal growth factor. Science *213*, 563–565.

86. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). Omim.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. *43*, D789-98.

87. Amberger, J.S., Bocchini, C.A., Scott, A.F., and Hamosh, A. (2018). Omim.org: leveraging knowledge across phenotype–gene relationships. Nucleic Acids Res. *47*, D1038–D1043.

88. 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.

89. 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.

90. 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.

91. Kahrizi, K., Huber, M., Galetzka, D., Dewi, S., Schröder, J., Weis, E., Kariminejad, A., Fattahi, 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.

92. 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.

93. 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.

94. 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.

95. 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.

96. Vermeiren, S., Bellefroid, E.J., and Desiderio, S. (2020). Vertebrate Sensory Ganglia: Common and Divergent Features of the Transcriptional Programs Generating Their Functional Specialization. Front Cell Dev Biol *8*, 587699.

97. Thaler, J.P., Koo, S.J., Kania, A., Lettieri, K., Andrews, S., Cox, C., Jessell, T.M., and Pfaff, S.L. (2004). A postmitotic role for Isl-class LIM homeodomain proteins in the assignment of visceral spinal motor neuron identity. Neuron *41*, 337–350.

98. Dupé, V., Matt, N., Garnier, J.-M., Chambon, P., Mark, M., and Ghyselinck, N.B. (2003). A newborn lethal defect due to inactivation of retinaldehyde dehydrogenase type 3 is prevented by maternal retinoic acid treatment. Proceedings of the National Academy of Sciences *100*, 14036–14041.

99. 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.

100. Knight, R.D., Javidan, Y., Zhang, T., Nelson, S., and Schilling, T.F. (2005). AP2-dependent signals from the ectoderm regulate craniofacial development in the zebrafish embryo. Development *132*, 3127–3138.

101. Van Otterloo, E., Milanda, I., Pike, H., Thompson, J.A., Li, H., Jones, K.L., and Williams, T. (2022). AP-2α and AP-2β cooperatively function in the craniofacial surface ectoderm to regulate chromatin and gene expression dynamics during facial development. Elife *11*, e70511.

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

103. Brugmann, S.A., Goodnough, L.H., Gregorieff, A., Leucht, P., ten Berge, D., Fuerer, C., Clevers, H., Nusse, R., and Helms, J.A. (2007). Wnt signaling mediates regional specification in the vertebrate face. Development *134*, 3283–3295.

104. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci. Rep. *9*, 1–5.

105. 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.

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

107. 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.

108. Coppola, C.J., C Ramaker, R., and Mendenhall, E.M. (2016). Identification and function of enhancers in the human genome. Hum. Mol. Genet. *25*, R190–R197.

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

110. Inoue, F., and Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. Genomics *106*, 159–164.

111. Sherry, S.T., Ward, M., and Sirotkin, K. (1999). dbSNP—Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. Genome Res. *9*, 677–679.

112. Hon, C.-C., Shin, J.W., Carninci, P., and Stubbington, M.J.T. (2018). The Human Cell Atlas: Technical approaches and challenges. Brief. Funct. Genomics *17*, 283–294.

113. Sayers, E.W., Agarwala, R., Bolton, E.E., Brister, J.R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., Hefferon, T., et al. (2019). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *47*, D23–D28.

114. Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D.J., Hicks, S.C., Robinson, M.D., Vallejos, C.A., Campbell, K.R., Beerenwinkel, N., Mahfouz, A., et al. (2020). Eleven grand challenges in single-cell data science. Genome Biol. *21*, 31.

115. 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.

116. Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. Nat. Genet. *51*, 1664–1669.

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

118. Schoenfelder, S., Javierre, B.M., Furlan-Magaril, M., Wingett, S.W., and Fraser, P. (2018). Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. J. Vis. Exp. 10.3791/57320.

119. Hoellinger, T., Mestre, C., Aschard, H., Le Goff, W., Foissac, S., Faraut, T., and Djebali, S. (2023). Enhancer/gene relationships: Need for more reliable genome-wide reference sets. Front Bioinform *3*, 1092853.

120. Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell *167*, 1170–1187.

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

122. 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.

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

124. Köhler, S., Gargano, M., Matentzoglu, N., Carmody, L.C., Lewis-Smith, D., Vasilevsky, N.A., Danis, D., Balagura, G., Baynam, G., Brower, A.M., et al. (2021). The Human Phenotype Ontology in 2021. Nucleic Acids Res. *49*, D1207–D1217.

125. Blake, J.A., Baldarelli, R., Kadin, J.A., Richardson, J.E., Smith, C.L., Bult, C.J., and Mouse Genome Database Group (2021). Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology. Nucleic Acids Res. *49*, D981–D987.

126. 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.

127. 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.

128. 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.

129. 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.

130. 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.

131. Fang, R., Preissl, S., Li, Y., Hou, X., Lucero, J., Wang, X., Motamedi, 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.

132. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21.

133. 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.

134. 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.

135. 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.

136. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, 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.

1548