



## OPEN ACCESS

## EDITED BY

Walid D. Fakhouri,  
University of Texas Health Science Center  
at Houston, United States

## REVIEWED BY

Hong Li,  
University of Colorado Anschutz Medical  
Campus, United States  
Ariadne Letra,  
University of Pittsburgh, United States

## \*CORRESPONDENCE

Kerstin U. Ludwig,  
✉ kerstin.ludwig@uni-bonn.de

RECEIVED 07 November 2022

ACCEPTED 03 April 2023

PUBLISHED 24 April 2023

## CITATION

Siewert A, Reiz B, Krug C, Heggemann J,  
Mangold E, Dickten H and Ludwig KU  
(2023), Analysis of candidate genes for  
cleft lip  $\pm$  cleft palate using murine single-  
cell expression data.

*Front. Cell Dev. Biol.* 11:1091666.  
doi: 10.3389/fcell.2023.1091666

## COPYRIGHT

© 2023 Siewert, Reiz, Krug, Heggemann,  
Mangold, Dickten and Ludwig. This is an  
open-access article distributed under the  
terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Analysis of candidate genes for cleft lip $\pm$ cleft palate using murine single-cell expression data

Anna Siewert<sup>1</sup>, Benedikt Reiz<sup>2</sup>, Carina Krug<sup>1</sup>, Julia Heggemann<sup>1</sup>,  
Elisabeth Mangold<sup>1</sup>, Henning Dickten<sup>2</sup> and Kerstin U. Ludwig<sup>1\*</sup>

<sup>1</sup>Institute of Human Genetics, University of Bonn, School of Medicine and University Hospital Bonn, Bonn, Germany, <sup>2</sup>FASTGenomics, Comma Soft AG, Bonn, Germany

**Introduction:** Cleft lip  $\pm$  cleft palate (CL/P) is one of the most common birth defects. Although research has identified multiple genetic risk loci for different types of CL/P (i.e., syndromic or non-syndromic forms), determining the respective causal genes and understanding the relevant functional networks remain challenging. The recent introduction of single-cell RNA sequencing (scRNA-seq) has provided novel opportunities to study gene expression patterns at cellular resolution. The aims of our study were to: (i) aggregate available scRNA-seq data from embryonic mice and provide this as a resource for the craniofacial community; and (ii) demonstrate the value of these data in terms of the investigation of the gene expression patterns of CL/P candidate genes.

**Methods and Results:** First, two published scRNA-seq data sets from embryonic mice were re-processed, i.e., data representing the murine time period of craniofacial development: (i) facial data from embryonic day (E) E11.5; and (ii) whole embryo data from E9.5–E13.5 from the Mouse Organogenesis Cell Atlas (MOCA). Marker gene expression analyses demonstrated that at E11.5, the facial data were a high-resolution representation of the MOCA data. Using CL/P candidate gene lists, distinct groups of genes with specific expression patterns were identified. Among others we identified that a co-expression network including *Irf6*, *Grhl3* and *Tfap2a* in the periderm, while it was limited to *Irf6* and *Tfap2a* in palatal epithelia, cells of the ectodermal surface, and basal cells at the fusion zone. The analyses also demonstrated that additional CL/P candidate genes (e.g., *Tpm1*, *Arid3b*, *Ctnnd1*, and *Wnt3*) were exclusively expressed in *Irf6*+ facial epithelial cells (i.e., as opposed to *Irf6*- epithelial cells). The MOCA data set was finally used to investigate differences in expression profiles for candidate genes underlying different types of CL/P. These analyses showed that syndromic CL/P genes (syCL/P) were expressed in significantly more cell types than non-syndromic CL/P candidate genes (nsCL/P).

**Discussion:** The present study illustrates how scRNA-seq data can empower research on craniofacial development and disease.

## KEYWORDS

cleft lip with or without cleft palate, single-cell RNA sequencing (scRNA-seq), IRF6, craniofacial development, expression pattern, single-cell transcriptomics

## 1 Introduction

Molecular malfunctions during craniofacial development can lead to cleft lip  $\pm$  cleft palate (CL/P). CL/P represents one of the most common of all birth defects, with a global prevalence of 1 in 700 live births (Mangold et al., 2011). Importantly, CL/P can present either as an isolated, non-syndromic phenotype (nsCL/P), or within the context of more complex malformation syndromes (syCL/P), in which additional features indicative of a developmental defect are observed. Although CL/P can be caused by deleterious mutations in single high penetrance genes (Cox et al., 2018; Bishop et al., 2020), a considerable fraction of its genetic architecture is attributable to common risk variants. Research suggests that environmental factors also contribute to CL/P, as part of its multifactorial etiology (Murray, 2002).

For nsCL/P, genome-wide association studies (GWAS) have identified multiple risk loci, and positional analyses of these loci have revealed promising candidate genes. For most of these genes, however, few data are available concerning the mechanism through which they affect the underlying functional processes of craniofacial development. One of the few exceptions to this is the *IRF6-GRHL3-TFAP2A* network, which has been shown to underlie diverse types of orofacial clefting, including CL/P and cleft palate only (Kousa et al., 2019). In addition to challenges associated with attributing causality to individual variants and genes, this lack of knowledge is also explained by the limited access to molecular data from relevant time points in humans, due to technical and ethical limitations.

Recently, single-cell RNA sequencing (scRNA-seq) has been performed on tissue from embryonic mice, generating systematic transcriptomic data sets at cellular resolution. This offers new avenues for the study of the tissue-specific expression of genes that underlie developmental phenotypes, including CL/P. Two resources of particular value in terms of CL/P are the Mouse Organogenesis Cell Atlas (MOCA; Cao et al., 2019), and facial data from embryonic mice that were reported in 2019 (Li et al., 2019). While MOCA encompasses the developmental time frame embryonic day (E) 9.5–13.5, the data from Li et al. provide a deeper insight into the transcriptome of facial structures at E11.5. Two important challenges associated with the use of scRNA-seq data are data accessibility and comparability, particularly when data are generated in different labs. The data of MOCA and Li et al. vary in terms of the level of processing, output types, and usability for the research community.

The aims of the present study were to (i) aggregate these scRNA-seq data from embryonic mice and provide this as a resource for the craniofacial community; and (ii) demonstrate the value of these data in terms of the investigation of the gene expression patterns of CL/P candidate genes. First, both of the selected data sets were re-analyzed using a joint computational pipeline. Second, different CL/P candidate gene sets were used to illustrate the potential of scRNA-seq data for deciphering the CL/P etiology. In particular, the expression patterns of CL/P candidate genes were assessed across the time period of craniofacial development, with the aim of placing them in their cell type-specific context. We specifically analyzed epithelial and mesenchymal cell types, which have been previously shown to be involved in CL/P (Ji et al., 2020). As an application example, we investigated co-expression of members of the *Irf6-*

*Grhl3-Tfap2a* genetic pathway in epithelial cell sub-types and identified further genes with a potential *Irf6* interaction in these cells. Finally, potential expression differences in candidate genes for syCL/P and nsCL/P were investigated in order to test the hypothesis that during embryonic development, syCL/P candidate genes are expressed in more tissues than is the case for candidate genes for nsCL/P.

## 2 Materials and methods

### 2.1 Data sources

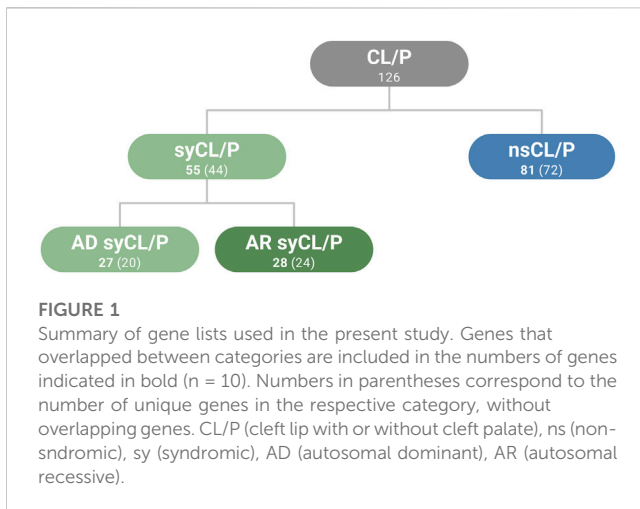
Two sets of single cell data on murine embryonic development were downloaded and analyzed using the same computational pipeline, which is described in detail in “Data analysis.” The first data set comprised single-cell gene expression data from 7,893 single cells from the lambdoidal junction, which were extracted from 4–5 mouse embryos at E11.5 (Li et al., 2019). The corresponding gene-count matrix was downloaded from the Gene Expression Omnibus (RRID:SCR\_005012; accession number: GSM3867275). The data set was then re-analyzed using our in-house pipeline. The latter included stricter filtering parameters (see below), thus reducing the number of single cells used for analysis (7,249 cells in total) compared to the original study. The final facial data set included 25 cell clusters.

The second data set was MOCA, which was generated from whole embryonic mice (Cao et al., 2019). The MOCA data set comprises the expression data of 2,058,652 single cells, as obtained from 61 mouse embryos from developmental stages E9.5–E13.5. Post-filtering, the original data set contained data on 1,331,985 cells and 38 major cell types (Supplementary Table S1). The gene-count matrix containing these 1,331,985 pre-filtered, high-quality cells was downloaded from the MOCA Website, and stored and analyzed using FASTGenomics (Scholz et al., 2018; RRID:SCR\_022898). In contrast to the original publication, the gene count matrix was split into five data sets in accordance with embryonic day in order to create a developmental time frame of gene expression: 112,269 cells (E9.5); 258,104 cells (E10.5); 449,614 cells (E11.5); 270,197 cells (E12.5); and 241,800 cells (E13.5).

### 2.2 Data analysis

#### 2.2.1 General processing

Each of the data sets was processed using the R package Seurat v4 (Hao et al., 2021; RRID:SCR\_016341). To normalize the count matrices, Log normalization (normalization.method) was applied with a Seurat default scale factor of 10,000 (scale.factor). For the selection of highly variable genes, the “vst” selection method (selection.method) was chosen, using 2,500 as the number of features (nfeatures). Scaling was performed in block sizes of 500 (block.size). For linear dimension reduction, a principal component (PC) analysis was performed. To cluster the cells, a two-step approach was used. First, for each cell, the K-nearest neighbors were calculated using the *FindNeighbors* function of Seurat, based on the first 25 PC dimensions (dims). Second, the Louvain algorithm was applied as a modularity optimization technique with a resolution of 0.5 for MOCA data and 1.1 for facial data (resolution) using the *FindClusters* function. To identify differentially expressed genes (hereafter



referred to as ‘marker genes’) for each cluster, the Wilcoxon Rank Sum test was used (`test.use`). Marker genes were obtained by comparing the expression levels of individual genes against all other clusters, and only positive markers were used. Additional parameters were a minimum fraction of 0.25 of cells expressing the tested gene in either of the populations (`min.pct`), and a threshold of a 0.25-fold change between the tested clusters (`logfc.threshold`). The uniform manifold approximation and projection algorithm (UMAP) was used as a non-linear dimension reduction method, whereby the first 25 PCs were applied as dimensions (`dims`).

### 2.2.2 Study-specific filtering

For the facial data set, additional steps were performed pre-normalization. These included the filtering-out of potential doublets by excluding cells with  $>7,500$  unique features (`nFeature_RNA`), and cells with  $>80,000$  detected RNA molecules (`nCount_RNA`). To exclude cells that were previously lysed or apoptotic, cells with the presence of the following features were excluded from the data set: (i) a percentage of  $>5\%$  of unique molecular identifiers reflecting mitochondrial genes (`percent.mt`); and/or (ii)  $< 2,300$  unique features (`nFeature_RNA`). After filtering, our data set comprised 7,249 cells. To benchmark the present pipeline, cell type annotation was performed by comparing the marker genes of each cluster with the marker genes described in the original publication.

For the pre-filtered, high-quality cells of the MOCA data, no additional filtering was required. Final cell type annotation was performed using the published marker genes of Cao et al. and the R package `scCATCH` (Shao et al., 2020). For the latter, `species` was set to “Mouse”; `match_CellMatch` was set to “TRUE”; and the tissues selected to be matched to “CellMatch” were “Brain,” “Fetal brain” and “Embryo”. Further parameters were kept at default values.

## 2.3 Curation of CL/P candidate gene lists

A literature search was performed to generate lists of genes associated with non-syndromic and syndromic forms of CL/P. The nsCL/P gene list was generated based on a recent meta-analysis of nsCL/P GWAS (Welzenbach et al., 2021). Welzenbach et al.

performed a gene-based analysis for genes located at established GWAS risk loci, which identified a set of 81 genes with an enrichment of common variants. These 81 genes were used in the present study. The syCL/P *gene list* was generated using information from a recently published study (Bishop et al., 2020), which had involved a systematic review of orofacial clefting syndromes and their associated genes. For the purposes of the present study, the list of syndromes generated by Bishop et al. was reduced using OMIM (RRID:SCR\_006437) in order to: (i) include only those syndromes whose phenotype includes CL/P, with the exclusion of other orofacial clefting phenotypes; and (ii) generate subsets of genes with autosomal dominant (AD) or autosomal recessive (AR) contributions. An overview of the gene categories is provided in Figure 1. Genes that overlapped between the syndromic and non-syndromic categories were included in an ‘overlapping genes’ list. Use of this list was restricted to the comparison of expression data between syCL/P and nsCL/P. The final numbers of unique genes for these analyses were 126 genes for CL/P overall, of which 72 genes were for nsCL/P, and 44 genes were for syCL/P (20 AD genes and 24 AR genes). Ten genes overlapped both categories.

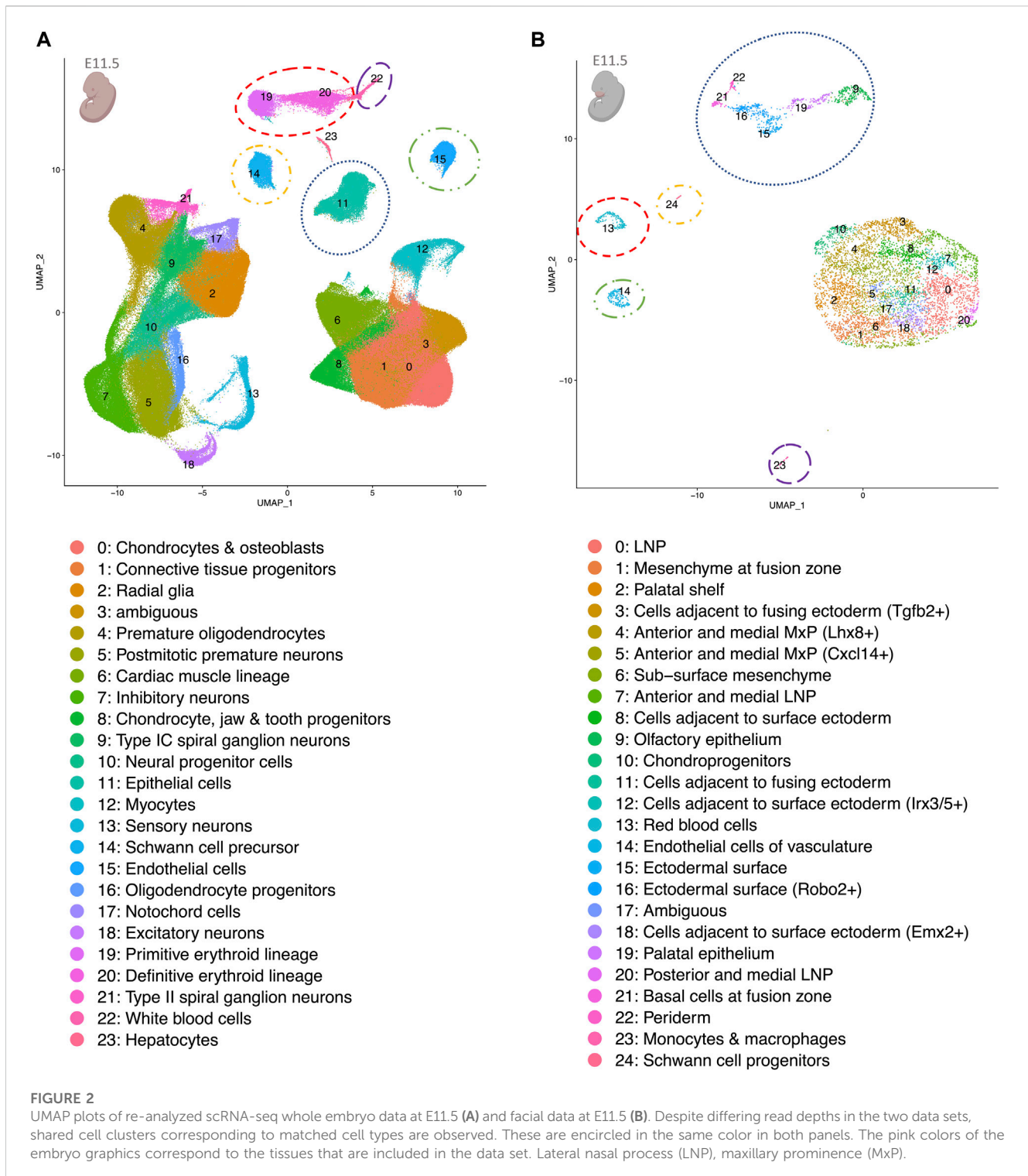
To evaluate whether the findings for CL/P are generalizable to other birth defects, gene lists were also generated for congenital heart disease (CHD). A recent publication (Nees and Chung, 2020) listed 18 genes for non-syndromic CHD (nsCHD) and 56 genes for syndromic CHD (40 AD, 16 AR). Three genes overlapped both categories. However, this group was not analyzed in the present study due to the low number of genes. All gene lists are provided in Supplementary Table S2.

## 2.4 Creating *Irf6*+ and *Irf6*- epithelial cell sub-clusters

Based on its well-established role in both syCL/P and nsCL/P (Woude, 1954; Birnbaum et al., 2009), analyses were performed to investigate the role of *Irf6* in epithelial cells. To create *Irf6*+ and *Irf6*- epithelial sub-clusters, epithelial cell clusters in the facial data set (i.e., palatal epithelium, olfactory epithelium, ectodermal surface, ectodermal surface (*Robo2*+), periderm, and basal cells at the fusion zone) were divided into subsets according to *Irf6* expression. Previous research has shown that *Irf6*, *Grhl3*, and *Tfap2a* are part of a genetic network in which *Irf6* influences the gene expression of *Grhl3* and *Tfap2a* (Kousa et al., 2019). In order to examine if these genes are among the marker genes of the *Irf6*+ sub-clusters and to identify possible additional genes that are influenced by *Irf6*, we determined marker genes for these sub-clusters. For this purpose, the expression profiles of each sub-cluster were compared against all other cell clusters in the data set, using the parameters applied in the initial data analysis (see Data analysis; Supplementary Table S3).

## 2.5 Analysis of differences in gene expression between nsCL/P and syCL/P

The analysis of nsCL/P and syCL/P gene lists was performed in the whole embryo MOCA data sets. Two parameters were used in these comparisons: (i) the percentage of all cell types in which the respective genes were expressed; and (ii) the average expression level. For analysis (i), a cell type was considered to express a certain

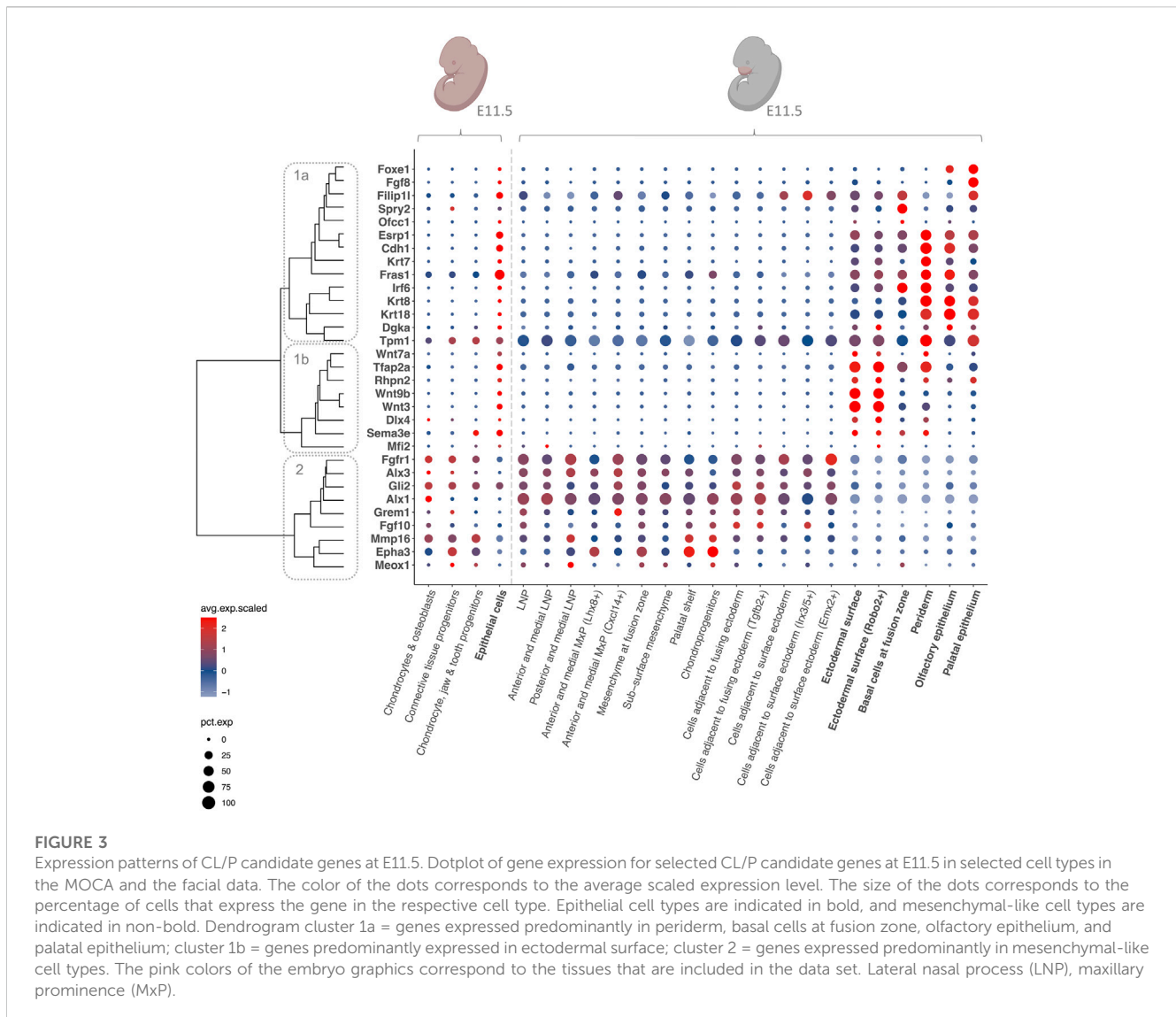


gene if the gene was expressed in at least 10% of cells. Percentages were determined for each gene in the respective list. The distributions were statistically compared using the Welch *t*-test. For analysis (ii), the average expression levels per cell type were extracted for each gene using the *AverageExpression* function from Seurat v4. The mean of these expression levels was then calculated per gene. A statistical comparison of the mean expression levels between both gene lists was performed using the Welch *t*-test.

## 3 Results

### 3.1 Facial-specific and whole embryo scRNA-seq data provide complementary insights into craniofacial development

Figure 2 shows the results generated by the UMAP algorithm for both the facial data (panel B) and the MOCA data (panel A, E11.5,



all other time points in [Supplementary Figures S1A–D](#)). The 25 cell types observed in the facial data were grouped into two main cell type clusters: (i) epithelial cells comprising periderm, basal cells at fusion zone, ectodermal surface, ectodermal surface (*Robo2+*), olfactory epithelium, and palatal epithelium; and (ii) more diverse cell types, which share a mesenchymal state, as based on the analysis of mesenchymal cell markers ([Supplementary Figure S1E](#)). Smaller cell clusters included endothelial cells and Schwann cells ([Figure 2B](#)). In the MOCA data for E11.5, a total of 24 cell types were identified, including a distinct epithelial cluster. To determine whether this at least partially represents the epithelial clusters in the facial data, the epithelial cells were sub-clustered. Three of these sub-clusters express marker genes for periderm (sub-cluster 6), basal cells at the fusion zone (sub-cluster 7) and ectodermal surface (sub-clusters 8 & 9) ([Supplementary Figure S1G](#); marker genes of the sub-clusters in [Supplementary Table S3](#)). Additional cell clusters in MOCA comprised specific cell types, such as hepatocytes, which are not represented in the facial data, as well as overlapping cell types where expected, e.g., endothelial cells, Schwann cells, and red and white blood cell types ([Figure 2A, B](#) colored circles).

### 3.2 A subset of CL/P candidate genes show convergent expression patterns

Investigation of the expression patterns of CL/P candidate genes in the scRNA-seq data sets showed that while the facial data set allowed an in-depth investigation of craniofacial structures at E11.5, the MOCA data set enabled a time course analysis over the time span of craniofacial development. Of the 126 CL/P candidate genes, all were expressed in the MOCA data sets from E9.5 - E13.5, although they varied in terms of overall expression levels and the cell types in which they were expressed. In the MOCA data, many CL/P candidate genes showed ubiquitous expression at E9.5, which became more specific at E10.5. Among the 126 CL/P candidate genes, 31 were specifically expressed in cell types of relevance to craniofacial development (i.e., epithelial cells, chondrocytes and osteoblasts, connective tissue progenitors, chondrocyte and jaw and tooth progenitors). Here, “specific expression” refers to either: (i) expression in at least one of these cell types; or (ii) expression in additional cell types, but with the highest expression levels being observed in at least one of the cell types

**TABLE 1** CL/P candidate genes with specific expression in *Irf6*+ facial epithelial cells.<sup>1</sup> adjusted *p*-value (based on Bonferroni correction using all genes in the data set);<sup>2</sup> average log<sub>2</sub> fold change in the average expression between the two tested groups (second test group: all other cell types; positive values indicate that the gene is more highly expressed in the respective cell type compared to all other cell types). nsCLO (non-syndromic cleft lip only).

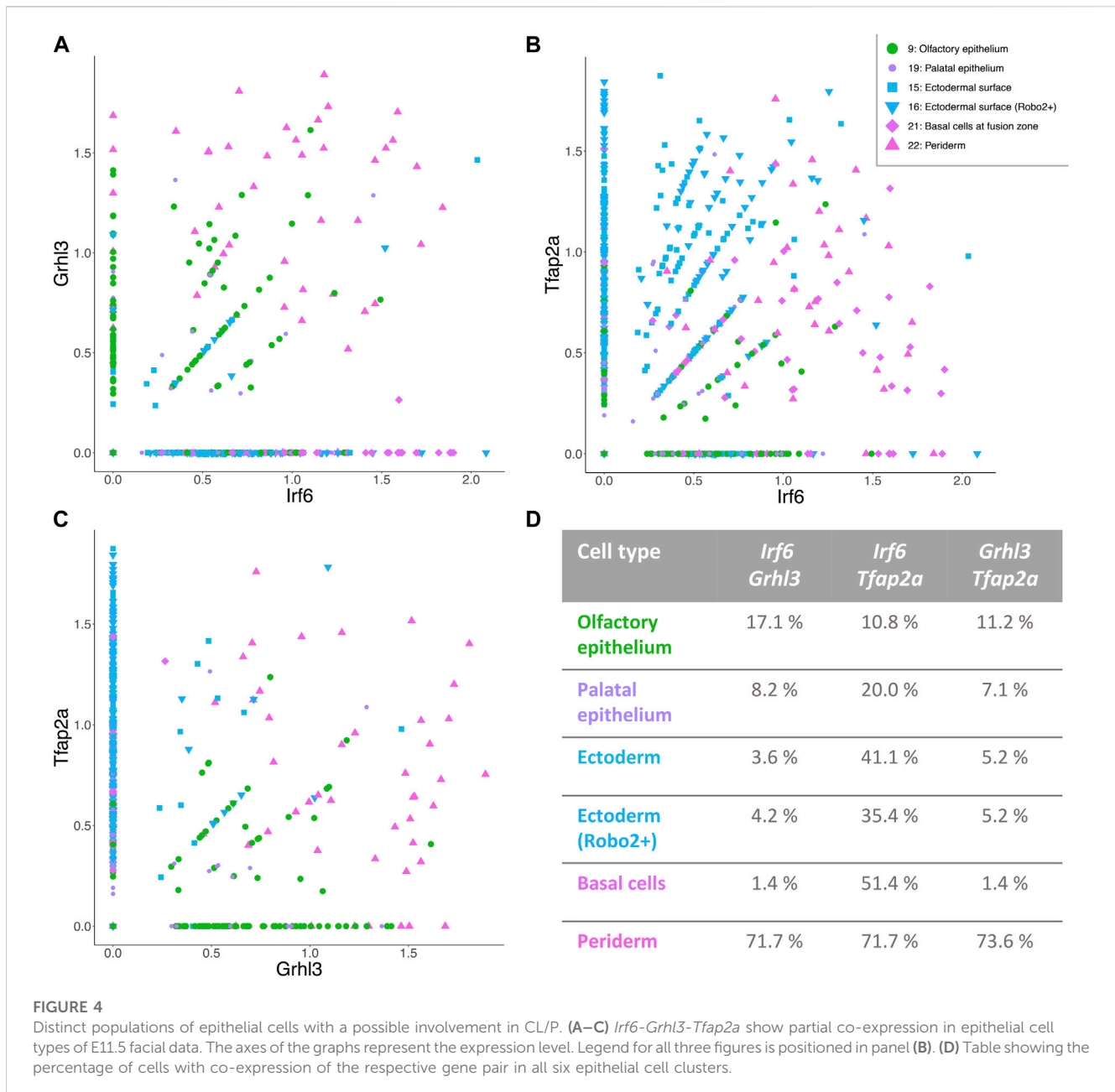
Cell type	Gene	P-val. adj. <sup>1</sup>	Log2FC <sup>2</sup>	Cleft association in humans
Periderm	<i>Tpm1</i>	1.9E-09	1.08	nsCL/P GWAS (Ludwig et al., 2012)
	<i>Pik3r1</i>	6.1E-10	0.75	nsCL/P GWAS (Leslie et al., 2017)
	<i>Tfap2a</i>	4.9E-44	1.03	Branchio-oculo-facial syndrome (Milunsky et al., 2008), nsCL/P GWAS (Ludwig et al., 2012; Leslie et al., 2017)
	<i>Wnt3</i>	1.0E-10	0.25	Tetra-amelia syndrome 1 (Niemann et al., 2004)
	<i>Ctnnd1</i>	0.0002	0.42	Blepharocheilodontic syndrome 2 (Ghoumid et al., 2017)
	<i>Fras1</i>	8.0E-10	0.69	Fraser syndrome (Fraser, 1962)
Basal cells at fusion zone	<i>Spry2</i>	4.0E-37	1.29	nsCL/P GWAS (Ludwig et al., 2012)
Ectodermal surface	<i>Arid3b</i>	7.7E-13	0.3	nsCL/P GWAS (Leslie et al., 2017)
	<i>Zfp3612</i>	0.008	0.26	nsCL/P (Lin-Shiao et al., 2019), nsCLO (Li et al., 2022)
Ectodermal surface ( <i>Robo2</i> +)	<i>Tpm1</i>	0.04	0.3	nsCL/P GWAS (Ludwig et al., 2012)
Palatal epithelium	<i>Cyb561</i>	1.6E-36	0.3	nsCL/P GWAS (Leslie et al., 2016)
	<i>Ptch1</i>	0.0001	0.33	nsCL/P GWAS (Yu et al., 2017), CPO GWAS (Butali et al., 2019), Basal cell nervous syndrome (Evans et al., 1993; Kimonis et al., 1997; Kimonis et al., 2013)
	<i>Tfap2a</i>	1.9E-09	0.27	Branchio-oculo-facial syndrome (Milunsky et al., 2008), nsCL/P GWAS (Ludwig et al., 2012; Leslie et al., 2017)
	<i>Fras1</i>	2.0E-08	0.33	Fraser syndrome (Fraser, 1962)
	<i>Ripk4</i>	5.3E-32	0.3	Popliteal pterygium syndrome, Bartsocas-Papas type 1 (Bartsocas and Papas, 1972)
Olfactory epithelium	<i>Arid3b</i>	2.5E-11	0.25	nsCL/P GWAS (Leslie et al., 2017)
	<i>Cyb561</i>	5.0E-55	0.29	nsCL/P GWAS (Leslie et al., 2016)
	<i>Xpa</i>	0.0003	0.26	nsCL/P GWAS (Welzenbach et al., 2021)

of relevance to craniofacial development. Comparison of the expression patterns of these 31 genes in the MOCA and the facial data (Figure 3) revealed that they clustered into two main groups: While 22 genes were specifically expressed in epithelial cell types (Figure 3 dendrogram cluster 1), nine genes were expressed in mesenchymal-like cell types (Figure 3 dendrogram cluster 2). Interestingly, the analyses showed that the first group (i.e., genes expressed predominantly in epithelial cell types) can be further subdivided into genes that have their highest expression levels in the ectodermal surface (Figure 3 dendrogram cluster 1b), and genes that have their highest expression levels in periderm, basal cells at fusion zone, olfactory epithelium, and palatal epithelium (Figure 3 dendrogram cluster 1a). The expression patterns of the remaining 95 CL/P candidate genes at E11.5 are shown in Supplementary Figure S4.

### 3.3 CL/P may involve distinct subgroups of epithelial cells

Using our data set, we first focused on the well-established CL/P risk gene *IRF6*. In the present study, *Irf6* was predominantly expressed in epithelial cells in both the MOCA and the facial

data sets, with particularly strong expression being observed in the periderm and basal cells at fusion zone in the facial data set. In MOCA, this expression was maintained throughout the developmental time period of the data set (Supplementary Figure S3). In the facial epithelial cells, considerable intra-cluster heterogeneity was observed. Cells expressing *Irf6* (denoted as *Irf6*+ cells) were observed in 58% of cells from the palatal epithelium ( $n = 71$  out of 170 cells), 40% of cells from olfactory epithelium (105/258), 44% of cells from the ectodermal surface (85/192), 44% of cells from the ectodermal surface (*Robo2*+) (86/192), 70% of cells from the basal cells at fusion zone (49/70), and 77% of the periderm cells (41/53). The six epithelial cell clusters from the facial data set were each divided into subsets according to their expression of *Irf6*, and marker genes of the *Irf6*+ cells were identified (Supplementary Table S3). A set of genes that overlapped between the marker genes of the *Irf6*+ epithelial subsets and CL/P candidate genes was identified (Table 1), which included CL/P genes that were associated with: (i) syndromic forms (e.g., *Tfap2a* (Branchio-oculo-facial syndrome, Milunsky et al., 2008), *Ctnnd1* (Blepharocheilodontic syndrome 2, Ghoumid et al., 2017) and *Fras1* (Fraser syndrome 1, Fraser, 1962); and (ii) candidate genes from GWAS loci (e.g., *Tpm1* (Ludwig et al., 2012) and *Arid3b* (Leslie et al., 2017) (Table 1,

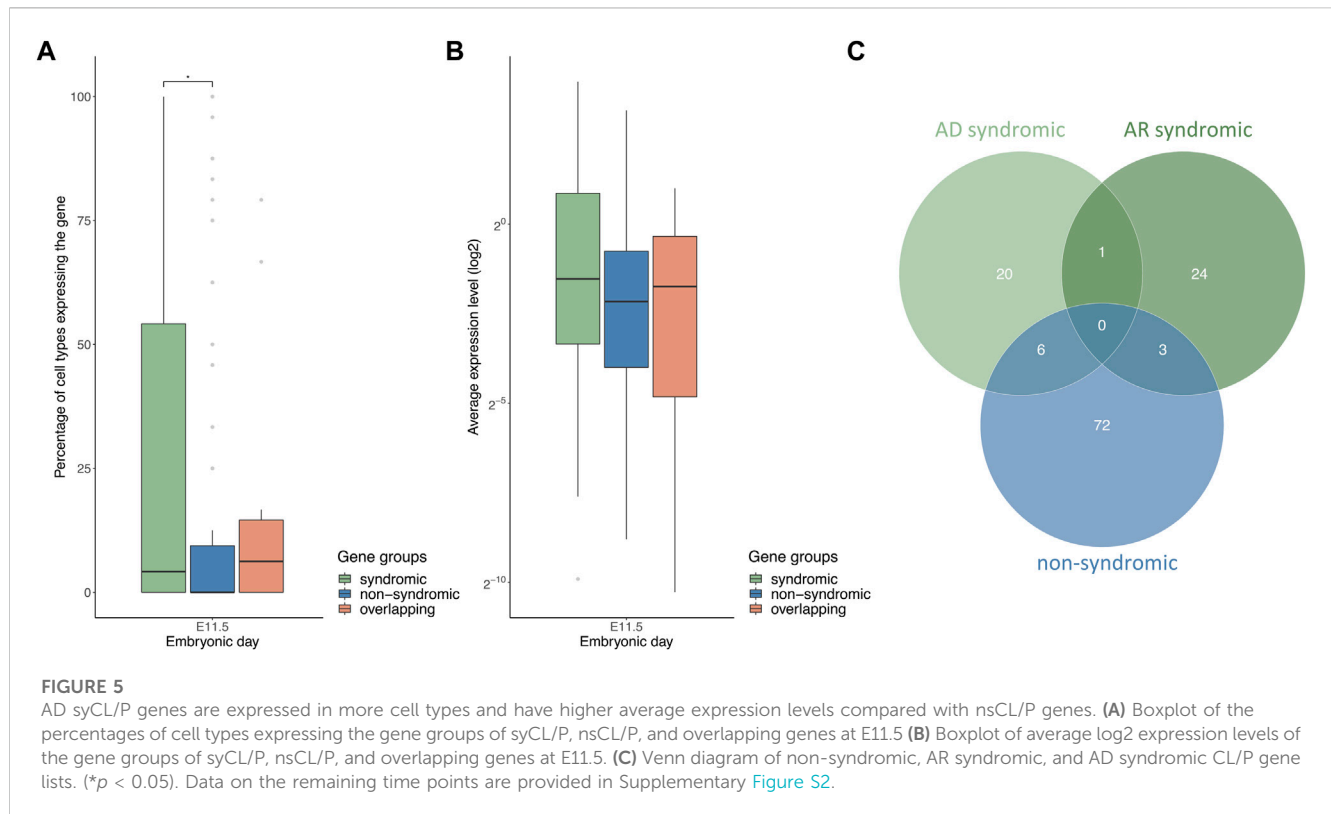


Supplementary Table S4). Interestingly, the gene Grainyhead-like 3 (*Grhl3*) was also observed among the marker genes of cells from the periderm and olfactory epithelium. As with *Irf6*, mutations in *Grhl3* cause Van der Woude syndrome. Here, however, most individuals present with a cleft palate only rather than CL/P (Mangold et al., 2016).

To elucidate the connection of *Irf6*, *Grhl3*, and *Tfp2a* in the six epithelial cell types at the transcriptomic level, the co-expression of these genes was analyzed (Figures 4A–D). Each of the *Irf6-Grhl3-Tfp2a* gene pairs showed partial co-expression, since an overlap in expression was observed in a subgroup of cells (indicated by percentage in Figure 4D). The co-expression network comprising all three genes was most abundant in the periderm, while it was reduced to only *Irf6* and *Tfp2a* in basal cells at fusion zone, the ectodermal surface clusters, and the palatal epithelium as well.

### 3.4 SyCL/P genes are expressed in more tissues compared to nsCL/P genes

To compare differences in the number of cell types between the gene lists for syCL/P and nsCLP, the analysis was restricted to the MOCA data set only, since syCL/P can affect tissues and organs outside of the craniofacial region and the MOCA data set contains more non-facial tissues. Across stages E10.5 to E11.5, the syCL/P genes were expressed in significantly more cell types than was the case for the nsCL/P genes (Figure 5A, E11.5). Comparison of the average gene expression levels of these gene sets showed that the syCL/P genes did not have significantly higher gene expression levels than the nsCL/P genes (Figure 5B E11.5). However, division of the syCL/P gene set into AD and AR genes revealed that the observed differences



in the percentage of expressing cell types between syCL/P and nsCL/P were mainly driven by the AD syCL/P genes. In comparison to the nsCL/P and AR syCL/P genes, the AD syCL/P genes: (i) were expressed in more cell types (Supplementary Figure S2A); and (ii) showed higher average expression levels (Supplementary Figure S2B).

## 4 Discussion

The present study leveraged two scRNA-seq data sets to generate insights into craniofacial development and diseases, specifically CL/P. Our reasons for selecting these data sets were threefold. First, the process of craniofacial development is largely conserved between mice and humans (Suzuki et al., 2016), which suggests that murine scRNA-seq data can be useful in terms of studying craniofacial development in the absence of human data. Second, the respective scRNA-seq samples were obtained at the time period of murine primary and secondary palate development (Miyake et al., 1996), thus increasing their suitability for studying CL/P candidate genes. Finally, research has shown that a large proportion of human embryonic scRNA-seq data from later developmental time points can be integrated with the MOCA data (Cao et al., 2020), providing further evidence for the transferability of developmental expression patterns. Although the MOCA scRNA-seq data are easily accessible via a comprehensive web browser, a systematic analysis in this setting is challenging. Of the 38 major cell clusters originally reported in MOCA, the present re-analysis identified a total of 31. This was probably attributable to

differences in processing, since in the present study, the data were first split in accordance with embryonic day (in order to reduce the size of the data set to a computable level), followed by the performance of clustering. Nevertheless, as in the original MOCA publication, less diffuse clustering of some cell types was observed over the 5 day time-period, and a joint clustering of mesenchymal-like cell types was identified, such as chondrocyte progenitors, connective tissue progenitors, chondrocytes and osteoblasts, and jaw and tooth progenitors (commencing at E10.5). With regards to the facial dataset, the present analysis identified 25 clusters as opposed to 24 main clusters reported in the original publication. While we consider the numbers of clusters similar, we observed differences in cluster annotations. On one side, our re-analysis yielded several distinct cluster annotations for four clusters that were annotated as one cluster each in Li et al. This increased the number of clusters comprising those cells. On the other hand, we also failed to identify four of the 24 original clusters, including nasolacrimal groove and dental epithelium (see Supplementary Table S1). Investigating this further, we identified marker genes for these two clusters to be predominantly expressed in some of the cells of our ectodermal surface clusters and palatal epithelium, respectively (Supplementary Figure S5). Yet, these clusters did not split further into distinct clusters when using higher resolution clustering (*data not shown*). This divergence may be attributable to the fact that the present analysis involved a stricter filtering strategy, no cell cycle regression, and high-resolution clustering of all cells together without sub-clustering (as opposed to the original study that divided the data into ectoderm and mesenchyme first, and performed sub-



clustering at different resolutions individually). Together, this contributed to a lower absolute number of overall cells (7,249 compared to 7,893 from the original) and different clusters in our re-analysis.

Comparison of E11.5 transcriptome profiles between the MOCA and the facial data revealed substantial similarities at both the cell type and gene levels. For instance, red and white blood cells, endothelial cells, and Schwann cells represent distinct cell clusters that mapped at certain distances to the other clusters within the UMAP space. At the gene level, *Irf6*, *Tfap2a*, *Fras1*, *Cdh1*, and *Esrp1* exhibited similar expression patterns in epithelial cell types of both data sets. Additionally, *Tfap2a* showed expression in Schwann cell progenitors in both data sets. Together, these data suggest that the facial data set is a tissue-restricted, but high-resolution representation of the MOCA data at E11.5, and that collectively, the two datasets represent a valuable resource for genomics research into craniofacial development. However, caution is generally required when interpreting expression profiles from several scRNA-seq data sets, since scRNA-seq itself but also the combination of different sources have some limitations. These include differences in cell capture efficiency and transcript coverage, which may result in transcripts not being detected in all cells equally, and different enrichment strategies used in both studies. In addition, scRNA-seq data of tissues undergoing continuous processes during development, such as epithelial-to-mesenchymal transitions, only provide a snapshot of a possibly transient period of gene expression. Finally, varying sequencing depth adds to higher noise levels in scRNA-seq data compared to bulk RNA-seq data (Kolodziejczyk et al., 2015).

The expression patterns observed in the aggregated scRNA-seq data sets replicate previously reported and experimentally verified expression patterns. For instance, a previous study showed that *Irf6* is expressed in neural ectoderm and neural crest cells as early as E9.5 in murine embryonic development (Kousa et al., 2019). According to previous wet-lab data, *Irf6* is expressed in the ectoderm of the first and second pharyngeal arches, and in the palatal, lingual, maxillary, and mandibular epithelia, during the period E10.5–E13.5 (Kondo et al., 2002; Knight et al., 2006; Richardson et al., 2009; Goudy et al., 2013; Kousa and Schutte, 2016). In accordance, our analyses revealed the presence of *Irf6* expression in Schwann cell precursors, the palatal and olfactory epithelia, the ectodermal surface, the basal cells at the fusion zone, and the periderm in both, MOCA and facial data respectively. Of these, the highest expression was observed in the periderm and the basal cells at the fusion zone. Interestingly, only ~3% of the MOCA E11.5 epithelial cells expressed *Irf6*, as opposed to 40%–70% of those in the facial data set. This suggests that *Irf6*-expressing MOCA E11.5 epithelial cells might be derived from facial structures, while the epithelial cell cluster contains a substantial proportion of non-facial cells. Comparably, *Tfap2a* showed expression in the MOCA epithelial cells, as well as high expression levels in the facial ectodermal surface clusters and periderm. In addition, *Tfap2a* showed expression in Schwann cell precursor cells in both the MOCA and the facial data sets. Again, this expression pattern recapitulates existing data, since previous reports have demonstrated that in mice, *Tfap2a* is expressed in the ectoderm, cranial neural crest cells, the facial mesenchyme, nasal and oral epithelia, and the central and peripheral nervous system between E9–E13.5 (Mitchell et al., 1991; Chazaud et al., 1996; Moser et al., 1997). Previous studies have shown that *Esrp1* is expressed in the head region and epithelial cells, especially in cells of the ectodermal

surface as early as E9.5 in mice (Warzecha et al., 2009; Revil and Jerome-Majewska, 2013; Bebee et al., 2015; Lee et al., 2020). Similarly, our data showed a broad expression of *Esrp1* in epithelial cells of both data sets with the highest expression in the periderm in the facial data set. Furthermore, the transcription factor *Foxe1* was found to be expressed in epithelial cells of embryonic mice starting at E9.5, both in our data sets and in previous studies (Zannini et al., 1997; Dathan et al., 2002; Moreno et al., 2009). In addition, studies have shown the keratin genes *Krt8* and *Krt18* to be expressed in single-layered epithelia in embryos (Jackson et al., 1980; Owens and Birgitte Lane, 2003; Moll et al., 2008). This is also confirmed by our data, as *Krt8* and *Krt18* showed expression in epithelial cells in both data sets. However, as expected, the strongest expression in the facial data set was found in the periderm and the palatal and olfactory epithelia. In contrast to the previously described genes, *Fgfr1* has been shown to be primarily expressed in mesenchymal cell types (Bachler and Neubüser, 2001), which is also evident in our data, as *Fgfr1* was predominantly expressed in mesenchymal cell types. These similarities indicate that: (i) the data sets are reliable resources in the context of craniofacial development; and (ii) that expression patterns of genes that have not yet been experimentally validated may be characterized using scRNA-seq data.

In a first attempt to use these data in the context of CL/P, the present analyses identified two groups of CL/P candidate genes based on their expression in relevant facial cell types. Using predefined lists of CL/P candidate genes, the analyses identified distinct sets of genes that are predominantly expressed in either epithelial cells, or mesenchymal-like cells. Unsurprisingly, the first group included *Irf6*, *Tfap2a*, and *Esrp1*, which show similar expression in the six epithelial cell types of the facial data set, and which have been implicated in a regulatory network (Kousa et al., 2019; Carroll et al., 2020). A specific examination of the expression of the *Irf6*-*Grhl3*-*Tfap2a* genetic pathway revealed partial co-expression of *Irf6*, *Grhl3*, and *Tfap2a* within epithelial cells. This opens up the possibility that other CL/P candidate genes, which are among the marker genes of the *Irf6*+ epithelial cell types, or genes with an as yet unknown role in CL/P etiology, might also contribute to the *Irf6* regulatory network. We plan to follow up on this question in a future study, using more systematic co-expression network approaches (Dam et al., 2018). While the expression of *Irf6* in the periderm has already been established (Richardson et al., 2009; Richardson et al., 2014; Kousa et al., 2017), the scRNA-seq data suggest the presence of a specific sub-cell type in which *Irf6* and other CL/P candidate genes show co-expression, and that may contribute to the etiology of CL/P. Furthermore, the expression of CL/P candidate genes in adjacent facial cell types highlights CL/P candidate genes that might contribute to molecular communication between the different epithelial cell types, e.g., *Tpm1*, *Fras1*, *Krt7*, *Wnt7a*, *Rhpn2*, and *Sema3e* in the ectodermal surface clusters and periderm; and *Filip1l* in the ectodermal surface and cells adjacent to the ectodermal surface. These questions need to be addressed in the future using more sophisticated computational and experimental approaches, such as spatial transcriptomic analyses (Carangelo et al., 2022).

In a second application example, the MOCA data set was used to investigate potential differences in expressing cell types between syCL/P and nsCL/P candidate genes. In accordance with our hypothesis, syCL/P candidate genes were expressed in a larger number of cell types during the examined time period compared

to candidate genes for nsCL/P. Similar patterns were observed in the analysis of the gene lists for CHD. The AD syndromic CHD genes were expressed in significantly more cell types than the AR syndromic and the non-syndromic CHD genes (Supplementary Figure S2C). The average expression levels of the AD syndromic CHD genes were significantly higher than those of the non-syndromic CHD genes (Supplementary Figure S2D). While the precise reason for this effect requires further investigation, our analysis indicates the value of scRNA-seq data in terms of the investigation of the different genetic architectures of CL/P subtypes.

In summary, the present study involved a re-analysis of previously published scRNA-seq data. We demonstrate the value of these data using several application examples. Our processed data sets are provided in Seurat object format as an easily accessible addition to the original data (see “Data availability statement”). This resource will facilitate functional approaches to the genomics of craniofacial development and disease.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, the processed data sets can be downloaded from [https://beta.fastgenomics.org/p/Siewert\\_2023](https://beta.fastgenomics.org/p/Siewert_2023). Further inquiries can be directed to the corresponding author.

## Author contributions

AS and KUL conceived and designed the study. BR and HD provided computational and infrastructural resources and assisted with the setup of the analytical environment. EM contributed genetic data. AS performed the data analysis and statistical testing, with the support of CK and JH. AS and KUL interpreted the data and drafted the first version of the manuscript. All authors have reviewed the final version of the article and approved its submission for publication.

## References

- Bachler, M., and Neubüser, A. (2001). Expression of members of the fgf family and their receptors during midfacial development. *Mech. Dev.* 100 (2), 313–316. doi:10.1016/S0925-4773(00)00518-9
- Bartsocas, C. S., and Papas, C. V. (1972). Popliteal pterygium syndrome. Evidence for a severe autosomal recessive form. *J. Med. Genet.* 9 (2), 222–226. doi:10.1136/jmg.9.2.222
- Bebee, T. W., Juw Won Park, K. I., Alex, M., Xing, Y., and Carstens, R. P. (2015). The splicing regulators *Esrp1* and *Esrp2* direct an epithelial splicing program essential for mammalian development. *ELife* 4, 1–27. doi:10.7554/eLife.08954
- Birnbaum, S., Ludwig, K. U., Reutter, H., Herms, S., Rubini, M., Baluardo, C., et al. (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat. Genet.* 41 (4), 473–477. doi:10.1038/ng.333
- Bishop, M. R., Diaz Perez, K. K., Sun, M., Ho, S., Chopra, P., Mukhopadhyay, N., et al. (2020). Genome-wide enrichment of de novo coding mutations in orofacial cleft trios. *Am. J. Hum. Genet.* 107 (1), 124–136. doi:10.1016/j.ajhg.2020.05.018
- Butali, A., Mossey, P. A., Adeyemo, W. L., Eshete, M. A., Gowans, L. J. J., Busch, T. D., et al. (2019). Genomic analyses in african populations identify novel risk loci for cleft palate. *Hum. Mol. Genet.* 28 (6), 1038–1051. doi:10.1093/hmg/ddy402
- Cao, J., Diana, R., Pliner, H. A., Paul, D. K., Deng, M., Riza, M., et al. (2020). A human cell Atlas of fetal gene expression. *Sci. (New York, N.Y.)* 370 (6518), 7721. doi:10.1126/science.aba7721
- Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., et al. (2019). The single-cell transcriptional landscape of mammalian Organogenesis. *Nature* 566 (7745), 496–502. doi:10.1038/s41586-019-0969-x
- Carangelo, G., Magi, A., and Semeraro, R. (2022). From multitude to singularity: An up-to-Date overview of ScRNA-seq data generation and analysis. *Front. Genet.* 13, 994069–994116. doi:10.3389/fgene.2022.994069
- Carroll, S. H., Macias Trevino, C., Li, E. B., Kawasaki, K., Myers, N., Hallett, S. A., et al. (2020). An *irf6-esrp1/2* regulatory Axis controls midface Morphogenesis in vertebrates. *Dev. Camb.* 147, dev194498. doi:10.1242/dev.194498
- Chazaud, C., Oulad-Abdelghani, M., Bouillet, P., Décimo, D., Chambon, P., and Dollé, P. (1996). AP-2.2, a novel gene related to AP-2, is expressed in the forebrain, limbs and face during mouse embryogenesis. *Mech. Dev.* 54 (1), 83–94. doi:10.1016/0925-4773(95)00463-7
- Cox, L. L., CoxMoreno, T. C. L. M. U., Zhu, Y., Richter, C. T., Nidey, N., Standley, J. M., et al. (2018). Mutations in the epithelial cadherin-P120-catenin complex cause mendelian non-syndromic cleft lip with or without cleft palate. *Am. J. Hum. Genet.* 102 (6), 1143–1157. doi:10.1016/j.ajhg.2018.04.009
- Dam, S. V., Vösa, U., Franke, L., and Pedro de Magalhães, J. (2018). Gene Co-expression analysis for functional classification and gene-disease predictions. *Briefings Bioinforma.* 19 (4), 575–592. doi:10.1093/bib/bbw139
- Dathan, N., Parlato, R., Rosica, A., De Felice, M., and Di Lauro, R. (2002). Distribution of the *tif2/foxe1* gene product is consistent with an important role in the development

## Funding

KUL is supported by grants from the German Research Council (Deutsche Forschungsgemeinschaft (DFG), LU-1944/3-1).

## Acknowledgments

The authors thank Christine Fischer for providing the mouse embryo graphics, Friederike David for technical support during data analysis, and Prof. Maximilian Billmann for helpful discussions.

## Conflict of interest

BR and HD were employed by FASTGenomics (Comma Soft AG).

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2023.1091666/full#supplementary-material>

- of foregut endoderm, palate, and hair. *Dev. Dyn.* 224 (4), 450–456. doi:10.1002/dvdy.10118
- Evans, D. G. R., Ladusans, E. J., Rimmer, S., Burnell, L. D., Thakker, N., and Farrndon, P. A. (1993). Complications of the naevoid basal cell carcinoma syndrome: Results of a population based study. *J. Med. Genet.* 30 (6), 460–464. doi:10.1136/jmg.30.6.460
- Fraser, G. R. (1962). Our genetical 'load': A review of some aspects of genetical variation. *Ann. Hum. Genet.* 25, 387–415. doi:10.1111/j.1469-1809.1962.tb01774.x
- Ghoumid, J., Stichelbout, M., Frenois, F., Lejeune-Dumoulin, S., Alex-Cordier, M., Lebrun, M., et al. (2017). Blepharochelodontic syndrome is a CDH1 pathway-related disorder due to mutations in CDH1 and CTNND1. *Genet. Med.* 19 (9), 1013–1021. doi:10.1038/gim.2017.11
- Goudy, S., Angel, P., Jacobs, B., Hill, C., Mainini, V., Smith, A. L., et al. (2013). Cell-autonomous and non-cell-autonomous roles for Irf6 during development of the tongue. *PLoS ONE* 8 (2), e56270. doi:10.1371/journal.pone.0056270
- Hao, Y., Stephanie, H., Erica Andersen-Nissen, W. M. M., Zheng, S., Butler, A., Lee, M. J., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell.* 184 (13), 3573–3587. doi:10.1016/j.cell.2021.04.048
- Jackson, B. W., Christine, G., Erika, S., Kurt, B., Werner, W., and Karl, I. (1980). Formation of cytoskeletal elements during mouse embryogenesis: Intermediate filaments of the cytokeratin type and desmosomes in preimplantation embryos. *Differentiation* 17 (1–3), 161–179. doi:10.1111/j.1432-0436.1980.tb01093.x
- Ji, Y., Garland, M. A., Sun, B., Zhang, S., Reynolds, K., McMahon, M., et al. (2020). Cellular and developmental basis of orofacial clefts. *Birth Defects Res.* 112 (19), 1558–1587. doi:10.1002/bdr2.1768
- Kimonis, V. E., Goldstein, A. M., Pastakia, B., Yang, M. L., Kase, R., Digiovanna, J. J., et al. (1997). Clinical manifestations in 105 persons with nevoid basal cell carcinoma syndrome. *Am. J. Med. Genet.* 69 (3), 299–308. doi:10.1002/(SICI)1096-8628(19970331)69:3<299::AID-AJMG16>3.0.CO;2-M
- Kimonis, V. E., Singh, K. E., Zhong, R., Pastakia, B., Digiovanna, J. J., and Bale, S. J. (2013). Clinical and radiological features in young individuals with nevoid basal cell carcinoma syndrome. *Genet. Med.* 15 (1), 79–83. doi:10.1038/gim.2012.96
- Knight, A. S., Schutte, B. C., Jiang, R., and Dixon, M. J. (2006). Developmental expression analysis of the mouse and chick orthologues of IRF6: The gene mutated in van der Woude syndrome. *Dev. Dyn.* 235 (5), 1441–1447. doi:10.1002/dvdy.20598
- Kolodziejczyk, A. A., Svensson, V., Marioni, J. C., and Teichmann, S. A. (2015). The technology and Biology of single-cell RNA sequencing. *Mol. Cell.* 58 (4), 610–620. doi:10.1016/j.molcel.2015.04.005
- Kondo, S., Schutte, B. C., Richardson, R. J., Bjork, B. C., Knight, A. S., Watanabe, Y., et al. (2002). Mutations in IRF6 cause van der Woude and popliteal pterygium syndromes. *Nat. Genet.* 32 (2), 285–289. doi:10.1038/ng985
- Kousa, Y. A., Roushangar, R., Patel, N., Walter, A., Marangoni, P., Krumlauf, R., et al. (2017). IRF6 and SPRY4 signaling interact in periderm development. *J. Dent. Res.* 96 (11), 1306–1313. doi:10.1177/0022034517119870
- Kousa, Y. A., and Schutte, B. C. (2016). Toward an orofacial gene regulatory network. *Dev. Dyn.* 245 (3), 220–232. doi:10.1002/dvdy.24341
- Kousa, Y. A., Zhu, H., Fakhouri, W. D., Lei, Y. A. K., Roushangar, R. R., Patel, N. K., et al. (2019). The tfap2a-IRF6-GRHL3 genetic pathway is conserved in neurulation. *Hum. Mol. Genet.* 28 (10), 1726–1737. doi:10.1093/hmg/ddz010
- Lee, S. K., Sears, M. J., Zhang, Z., Hong, L., Salhab, I., Krebs, P., et al. (2020). Cleft lip and cleft palate in *Esrp1* knockout mice is associated with alterations in epithelial-mesenchymal crosstalk. *Dev. Camb.* 147 (21), dev187369. doi:10.1242/dev.187369
- Leslie, E. J., Carlson, J. C., Eleanor Feingold, J. R. S., George, W., Laurie, C. A., Jain, D., et al. (2016). A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p 24.2, 17q23 and 19q13. *Hum. Mol. Genet.* 25 (13), 2862–2872. doi:10.1093/hmg/ddw104
- Leslie, E. J., Carlson, J. C., Shaffer, J. R., Butali, A., Carmen, J., Castilla, E. E., et al. (2017). Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Hum. Genet.* 136 (3), 275–286. doi:10.1007/s00439-016-1754-7
- 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. *Dev. Camb.* 146 (12), dev174888. doi:10.1242/dev.174888
- Li, M. J., JiaS, Y., Zhu, S., Shi, B., and Zhong, L. (2022). Targeted Re-sequencing of the 2p21 locus identifies non-syndromic cleft lip only novel susceptibility gene ZFP36L2. *Front. Genet.* 13, 802229. doi:10.3389/fgene.2022.802229
- Lin-Shiao, E., Lan, Y., Welzenbach, J., Alexander, K. A., Zhang, Z., Knapp, M., et al. (2019). P63 establishes epithelial enhancers at critical craniofacial development genes. *Sci. Adv.* 5 (5), eaaw0946–15. doi:10.1126/sciadv.aaw0946
- Ludwig, K. U., Mangold, E., Herms, S., Nowak, S., Paul, A., Becker, J., et al. (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Palate Identify Six. New Risk Loci* 44 (9), 968–971. doi:10.1038/ng.2360
- Mangold, E., Anne, C., Hoebel, A. K., and Gültepe, P. (2016). Sequencing the GRHL3 coding region reveals rare truncating mutations and a common susceptibility variant for nonsyndromic cleft palate. *Am. J. Hum. Genet.* 98 (4), 755–762. doi:10.1016/j.ajhg.2016.02.013
- Mangold, E., Ludwig, K. U., and Nöthen, M. M. (2011). Breakthroughs in the genetics of orofacial clefting. *Trends Mol. Med.* 17 (12), 725–733. doi:10.1016/j.molmed.2011.07.007
- Milunsky, J. M., Maher, T. A., Roberts, A. E., Stalker, H. J., Zori, R. T., Burch, M. N., et al. (2008). TFAP2A mutations result in branchio-oculo-facial syndrome. *Am. J. Hum. Genet.* 82 (5), 1171–1177. doi:10.1016/j.ajhg.2008.03.005
- Mitchell, P. J., Timmons, P. M. J. M., Tjian, R., and Rigby, P. W. (1991). Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis. *Genes. Dev.* 5 (1), 105–119. doi:10.1101/gad.5.1.105
- Miyake, T., Cameron, A. M., and Hall, B. K. (1996). Detailed staging of inbred C57BL/6 mice between Theiler's [1972] stages 18 and 21 (11–13 days of gestation) based on craniofacial development. *J. Craniofac. Genet. Dev. Biol.* 16, 1–31.
- Moll, R., Divo, M., and Lutz, L. (2008). The human keratins: Biology and pathology. *Histochem. Cell. Biol.* 129 (6), 705–733. doi:10.1007/s00418-008-0435-6
- Moreno, L. M., Maria Adela Mansilla, S. A. B., Cooper, M. E., Busch, T. D., Johnson, M. K., Busch, T. D., et al. (2009). FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. *Hum. Mol. Genet.* 18 (24), 4879–4896. doi:10.1093/hmg/ddp444
- Moser, M., Rüschoff, J., and Buettner, R. (1997). Comparative analysis of AP-2 $\alpha$  and AP-2 $\beta$  gene expression during murine embryogenesis. *Dev. Dyn.* 208 (1), 115–124. doi:10.1002/(SICI)1097-0177(199701)208:1<115::AID-AJA11>3.0.CO;2-5
- Murray, J. C. (2002). Gene/environment causes of cleft lip and/or palate. *Clin. Genet.* 61 (9), 248–256. doi:10.1034/j.1399-0004.2002.610402.x
- Nees, S. N., and Chung, W. K. (2020). Genetic basis of human congenital heart disease. *Cold Spring Harb. Perspect. Biol.* 12 (9), 0367499–a36840. doi:10.1101/cshperspect.a036749
- Niemann, S., Zhao, C., Pascu, F., Stahl, U., Aulepp, U., Lee, N., et al. (2004). Homozygous WNT3 mutation causes tetra-amelia in a large consanguineous family. *Am. J. Hum. Genet.* 74 (3), 558–563. doi:10.1086/382196
- Owens, D. W., and Birgitte Lane, E. (2003). The quest for the function of simple epithelial keratins. *BioEssays* 25 (8), 748–758. doi:10.1002/bies.10316
- Revil, T., and Jerome-Majewska, L. A. (2013). During embryogenesis, *Esrp1* expression is restricted to a subset of epithelial cells and is associated with splicing of a number of developmentally important genes. *Dev. Dyn.* 242 (3), 281–290. doi:10.1002/dvdy.23918
- Richardson, R. J., Nigel, L., Hammond, P. A. C., Saloranta, C., Nousiainen, H. O., Salonen, R., et al. (2014). Periderm prevents pathological epithelial adhesions during embryogenesis. *J. Clin. Investigation* 124 (9), 3891–3900. doi:10.1172/JCI71946
- Richardson, R. J., Dixon, J., Jiang, R., and Dixon, M. J. (2009). Integration of IRF6 and Jagged2 signalling is essential for controlling palatal adhesion and fusion competence. *Hum. Mol. Genet.* 18 (14), 2632–2642. doi:10.1093/hmg/ddp201
- Scholz, C., Biernat, P., Becker, M., Bassler, K., Günther, P., Balfer, J., et al. (2018). FASTGenomics: An analytical ecosystem for single-cell RNA sequencing data partial. *BioRxiv*.
- Shao, X., Liao, J., Lu, X., Xue, R., Ni, A., and Fan, X. (2020). ScCATCh: Automatic annotation on cell types of clusters from single-cell RNA sequencing data. *IScience* 23 (3), 100882. doi:10.1016/j.isci.2020.100882
- Suzuki, A., Sangani, D. R., Ansari, A., and Iwata, J. (2016). Molecular mechanisms of midfacial developmental defects. *Dev. Dyn.* 245 (3), 276–293. doi:10.1002/dvdy.24368
- Warzecha, C. C., Sato, T. K., Nabet, B., Hogenesch, J. B., and Carstens, R. P. (2009). ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol. Cell.* 33 (5), 591–601. doi:10.1016/j.molcel.2009.01.025
- Welzenbach, J., Hammond, N. L., Nikolić, M., Thieme, F., Ishorst, N., Leslie, E. J., et al. (2021). Integrative approaches generate insights into the architecture of non-syndromic cleft lip  $\pm$  cleft palate. *Hum. Genet. Genomics Adv.* 2 (3), 100038–100114. doi:10.1016/j.xhgg.2021.100038
- Woude, A. V. D. (1954). Fistula labii inferioris congenita and its association with cleft lip and palate. *Am. J. Hum. Genet.* 6 (2), 244–256.
- Yu, Y., Zuo, X., He, M., Gao, J., Fu, Y., Qin, C., et al. (2017). Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat. Commun.* 8 (2), 14364–14411. doi:10.1038/ncomms14364
- Zannini, M., Avantageggiato, V., Biffali, E., Sato, K., Pischetola, M., Taylor, B. A., et al. (1997). TTF-2, a new forkhead protein, shows a temporal expression in the developing thyroid which is consistent with a role in controlling the onset of differentiation. *EMBO J.* 16 (11), 3185–3197. doi:10.1093/emboj/16.11.3185