Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An ARID3A-dependent progenitor lineage confers developmental robustness to palate morphogenesis

Huan Liu

liu.huan@whu.edu.cn

School & Hospital of Stomatology,Wuhan University/Department of Pediatrics, Cedars-Sinai Guerin Children's/TaiKang Center for Life and Medical Sciences, Wuhan University, https://orcid.org/0000-0002-9947-6687

Jie Gao

Wuhan University

Litian Han

Wuhan University

Yao Xiao

Wuhan University

Yu Liang

Wuhan Univerisity

Hongxu Tao

Institute of Hydrobiology, Chinese Academy of Sciences

Qi Shi

Wuhan University

Fei Pei

University of Southern California

Zhaoming WU

The University of Hong Kong https://orcid.org/0000-0002-0131-2423

Yonghua Sun

Institute of Hydrobiology, Chinese Academy of Sciences https://orcid.org/0000-0001-9368-6969

Robert Cornell

University of Washington https://orcid.org/0000-0003-4207-9100

Zhi Chen

Wuhan University School and Hospital of Stomatology https://orcid.org/0000-0003-2095-1324

Ophir Klein

University of California, San Francisco https://orcid.org/0000-0002-6254-7082

Article

Keywords:

Posted Date: February 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8871167/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

An ARID3A-dependent progenitor lineage confers developmental robustness to palate morphogenesis

Jie Gao[1,8], Litian Han[1,8], Yao Xiao[1,8], Yu Liang[1], Hongxu Tao[1], Qi Shi[1], Fei Pei[1], Zhaoming Wu[2], Yonghua Sun[3,4], Robert A. Cornell[5], Zhi Chen[1], Ophir D. Klein[5*], Huan Liu[1,6,7,9*]

[1] State Key Laboratory of Oral & Maxillofacial Reconstruction and Regeneration, Key Laboratory of Oral Biomedicine Ministry of Education, Hubei Key Laboratory of Stomatology, School & Hospital of Stomatology, Wuhan University, Wuhan 430079, China.

[2] Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China

[3] State Key Laboratory of Freshwater Ecology and Biotechnology, Hubei Hongshan Laboratory, Institute of Hydrobiology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Wuhan, China

[4] College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing, China

[5] Department of Oral Health Sciences, University of Washington, Seattle, WA USA

[6] Department of Pediatrics, Cedars-Sinai Guerin Children's, Los Angeles, CA 90048, USA.

[7] Frontier Science Center for Immunology and Metabolism, Wuhan University, China.

[8] TaiKang Center for Life and Medical Sciences, Wuhan University, China

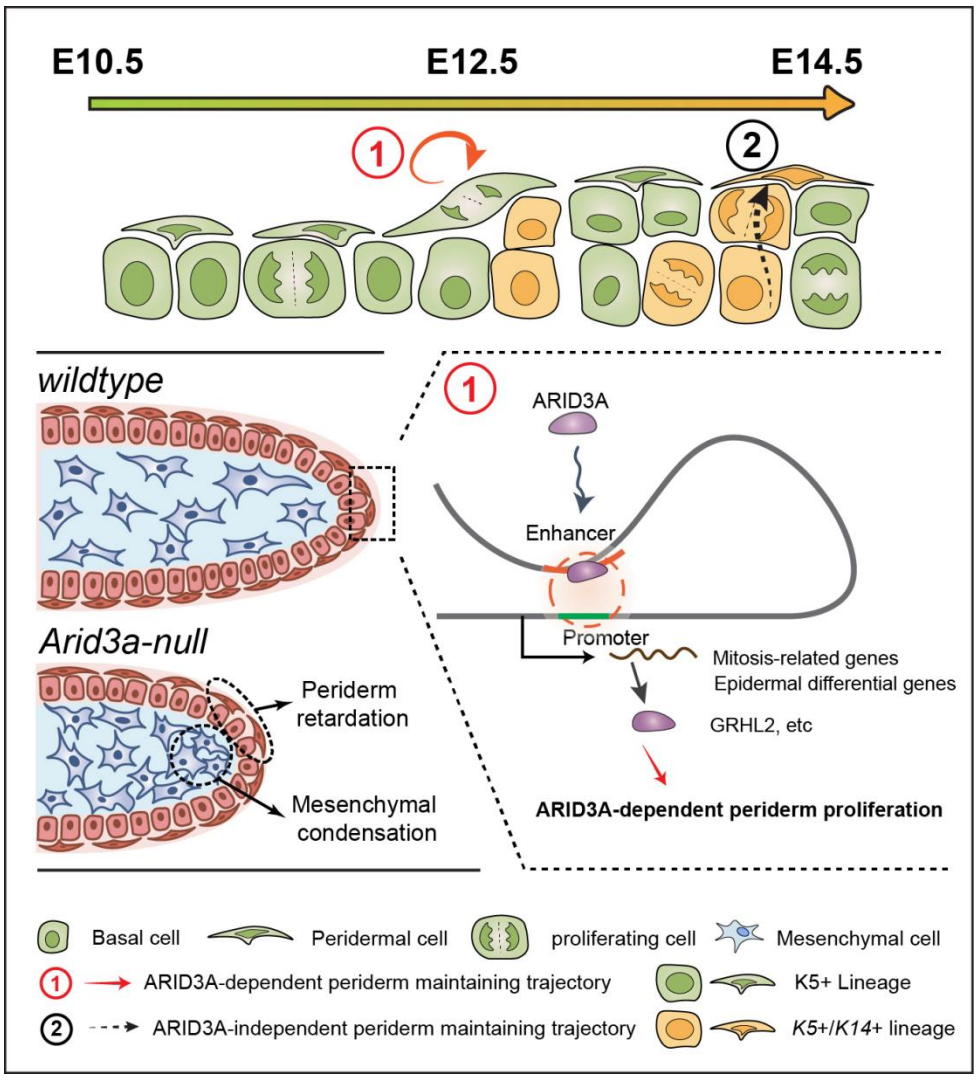
[9] These authors contributed equally

[10] Lead contact

*Correspondence: Ophir D. Klein: Ophir.Klein@cshs.org

Huan Liu: liu.huan@whu.edu.cn

Graphical abstract



Highlights

- In developing palatal shelves, a *K5*-positive progenitor pool gives rise to both the basal ($K14^+$) and superficial (periderm, $K6^+$) epithelial layers
- The transcription factor ARID3A is essential for the specification and function of a $K5^+$;$K14^-$ periderm progenitor population.
- ARID3A coordinates palate morphogenesis by controlling periderm proliferation and integrity, which supports underlying mesenchymal organization.
- Active enhancers in the ARID3A-dependent progenitor pool are enriched for genetic variants associated with risk for orofacial clefts.

Abstract

The periderm prevents pathological adhesions during orofacial development, yet its lineage origins remain unresolved. Here, we define the murine palatal epithelial hierarchy, identifying a transient, ARID3A-defined $K5^{+};K14^{-}$ progenitor subpopulation programmed for periderm differentiation. Genetic ablation of *Arid3a* impairs this lineage, causing severe delay in palate closure rather than complete failure. This "delayed-but-functional" phenotype reveals a mechanism of developmental robustness, wherein ARID3A drives a rapid lineage trajectory essential for timely morphogenesis. Mechanistically, ARID3A binds a distal enhancer to activate the key periderm regulator Grhl2. Furthermore, integrating our multi-omic atlas with human genetics demonstrates that enhancers in this ARID3A-dependent lineage are enriched for risk variants associated with orofacial clefts. Our findings identify a specific progenitor lineage that confers developmental robustness and provides a cellular address for human genetic risk.

Introduction

Congenital birth defects represent a major global health challenge, with orofacial clefts (OFCs) ranking among the most common, affecting approximately 1 in every 700 live births worldwide[1,2]. Understanding the etiology of these complex malformations is paramount for developing effective prevention and treatment strategies.

The developing secondary palate, whose failure to form leads to many cases of OFC[1], provides an *in vivo* model that undergoes a rapid and highly orchestrated sequence of fundamental morphogenetic events—including cell proliferation, tissue elevation, adhesion, and epithelial remodeling—all within a narrow and experimentally accessible timeframe[3,4]. This process relies on critical epithelial-mesenchymal interactions[4,5], and a key player in coordinating these morphogenetic events is the periderm, a transient, specialized epithelial layer that covers the developing embryo. This layer serves as a protective, "non-stick" barrier that prevents pathological adhesions between juxtaposed, adhesion-competent epithelia[6,7], such as the nascent palatal shelves and the tongue or mandible[6,7]. Beyond these physical roles, recent work has revealed the periderm to be a dynamic and signaling-rich tissue, expressing a diverse array of ligands and receptors that suggest a more active role in orchestrating development than previously appreciated[8]. Genetic studies have revealed that mutations in a core network of transcriptional regulators are essential for periderm development—including IRF6, p63, RIPK4, and the Grainyhead-like (GRHL) family—cause of a class of severe human genetic disorders such as Van der Woude syndrome (VWS)[9,10], Popliteal Pterygium Syndrome (PPS)[10], and Bartsocas-Papas Syndrome (BPS)[11] as well as non-sydromic cleft lip and palate[12,13].

In light of this functional importance, the precise lineage origins of the periderm and the mechanisms governing its expansion remain major unanswered questions. Classical studies of epidermal stratification posit that a KRT5$^+$/KRT14$^+$ (K5$^+$/K14$^+$) basal progenitor pool gives rise to all suprabasal layers[14,15], but this linear model was challenged by recent single-cell transcriptomic analyses revealing unexpected epithelial heterogeneity at these early stages[8], and definitive lineage tracing has been hindered by the fact that the widely-used *K14*-Cre driver fails to efficiently recombine in the palatal periderm[16]. Furthermore, driving *Irf6* expression with a *K14* promoter only partially rescues the periderm defects of *Irf6*-null mice, strongly suggesting that a non-*K14*-expressing population is required for proper periderm development[17]. Thus, the true lineage hierarchy within the early palatal epithelium from which progenitor(s) periderm cells arise, and what molecular mechanisms govern these cells' expansion[5], are questions that remain unresolved.

Here, by integrating inducible lineage tracing, genetic ablation, mathematical modeling, and single-cell multi-omics, we define the lineage hierarchy of the palatal epithelium. We identify a bipotent *K5*-positive progenitor pool that gives rise to both the basal layer and the periderm, and we discover a distinct, transient K5$^+$;K14$^-$ progenitor subpopulation uniquely controlled by the transcription factor ARID3A. Crucially, we show that ARID3A is not required for periderm formation per se, but rather drives a rapid lineage trajectory essential for timely palate morphogenesis. Loss of Arid3a leads to a severe delay in closure rather than complete failure, revealing a mechanism of developmental robustness where a primary, ARID3A-dependent lineage ensures rapid tissue expansion, while a secondary pathway provides functional compensation.

Mechanistically, we demonstrate that ARID3A orchestrates this process by binding distal enhancers to activate the key periderm regulator *Grhl2*. Finally, by mapping human GWAS data onto our single-cell atlas, we show that regulatory elements active in this specific progenitor population are significantly enriched for risk variants associated with orofacial clefts. These findings provide a cellular address for human genetic risk, suggesting that subtle defects in this "robustness-conferring" lineage may underlie the susceptibility to complex birth defects.

Results

A *K5*-Positive Lineage is Essential for Early Palatal Epithelial Integrity

To investigate the lineage origins of the palatal periderm, we first re-analyzed a published single-cell RNA-sequencing (scRNA-seq) dataset of the E11.5 mouse lambdoidal junction[18]. RNA velocity pseudotime analysis of the epithelial cells (*Epcam*$^{+}$) predicted two distinct trajectories originating from different progenitor populations that converged on Krt17-expressing periderm cells (Figure 1A). While the basal marker *Krt14*[14,15] was enriched in only one of the two progenitor roots, the pan-epithelial progenitor marker *Krt5* was expressed in both (Figure 1A). To determine if this dissociation between *Krt5* and *Krt14* expression was a specific feature of the developing palate, we performed a correlation assays for single cell RNA-seq profiles from multiple epithelial tissues, including human skin (HRA000395)[19], human foreskin (GSE147482)[20], human tracheal epithelium (GSE103354)[21], human mammary gland epithelium (GSE113196)[22], human buccal mucosa (GSE120654)[23], mouse E14.5 skin (E-MTAB-

11920)[8] and mouse E12.5 palate epithelium (this study). This analysis revealed that the low correlation between *Krt5* and *Krt14* expression is a distinctive feature of the early palate epithelium compared to other epithelial tissues, especially in adults (Figure 1B).

Next, we performed single molecule *in situ* expression analysis by RNAscope for *Krt5* and *Krt14* on developing palatal shelves. At E12.5, we observed a heterogenous expression pattern: *Krt5* was expressed in both the basal and superficial epithelial layers, whereas *Krt14* expression was largely restricted to the basal layer (Figure 1C, Figure S1). This resulted in a distinct $K5^{+}$;$K14^{-}$ population in the superficial layer. By E14.5, this heterogeneity resolved, with *Krt5* and *Krt14* becoming co-expressed throughout the epithelium (Figure 1D, Figure S1, schematized in Figure 1E).

The transient existence of a $K5^{+}$;$K14^{-}$ population led us to hypothesize that the $K5^{+}$ and $K14^{+}$ lineages have distinct functional roles during early palatogenesis. To test this, we generated a *K5CreERT2* knock-in allele (Figure S2A) and validated its activity by crossing to the *ROSA26-LSL-mTmG* reporter. Tamoxifen administration at E10.5 confirmed specific recombination in the $K5^{+}$ lineage throughout the craniofacial epithelium, including the palate, at E14.0 (Figure S2B). Similarly, we confirmed that an established *K14Cre* line[24] showed robust activity in the palatal epithelium (Figure S2B). We then performed a temporal ablation series using *K5CreERT2;DTA* mice and found that inducing ablation at E10.5 resulted in severe cleft palate (3/10 at E14.5), epithelial adhesions (9/10 at E14.5), and leading to embryonic lethality by E15.5, whereas induction at E12.5 produced normal embryos (Figure S3A). This established a critical requirement for the $K5^{+}$ lineage before E11.5.

Based on this, we chose E10.5 as the induction timepoint to directly compare the role of *K5*$^{+}$ and *K14*$^{+}$ lineages in palate epithelium. Inducing ablation of the *K5*$^{+}$ lineage (*K5CreERT2;DTA*) resulted in a significantly increased apoptotic indices in *K5CreERT2;DTA* embryos compared to *DTA* controls as indicated by TUNEL staining from E12.5 to E14.5 (Figure S4). We observed severe and highly penetrant phenotypes in *K5CreERT2;DTA*, including epithelial disintegration (6/12 at E14.5), adhesions (3/12 at E14.5), and cleft palate (5/12 at E14.5), culminating in near-complete embryonic lethality by E14.5 (11/12) (Figure 1F, G). In contrast, ablation of the *K14*$^{+}$ lineage at the same timepoint (*K14Cre;DTA*) produced a much milder phenotype, characterized by a thinner epithelium (3/10 at E14.5) and postnatal lethality (Figure 1F, G, and Figure S3). To confirm that the severe defect in *K5CreERT2;DTA* embryos was due to a loss of periderm identity, we performed immunostaining for key markers. While KRT6A and KRT14 were barely detectable in the palate epithelium of all genotypes at E12.5 (Figure S5A), by E13.5 the K5-ablated embryos exhibited a profound loss of KRT14, KRT6A, and E-cadherin (ECAD) expression, a phenotype not observed in *DTA* or *K14Cre;DTA* embryos, both mocked with tamoxifen at E10.5 (Figure 1H and Figure S5B, C). In *K5CreERT2;DTA* embryos, P63 staining in the remaining basal cells was diffuse and indistinct (Figure 1H), obscuring clear nuclear morphology. This observation is indicative of widespread apoptosis and is consistent with our TUNEL staining results (Figure S4).

Together, these findings demonstrate a transient heterogeneity in *Krt5* and *Krt14* expression within the early palatal epithelium. Although this heterogeneity is surprising given that these keratins are typically considered to be obligately co-expressed partners[15] in basal epithelial progenitors, it is consistent with evidence *Krt5* is expressed

in the single layer ectoderm in some part of the embryo[25] and that KRT5 expression precedes that of KRT14 expression in human skin[26]. Functionally, the findings reveal that the *K5*$^+$ lineage is required for establishing both the basal and periderm layers before E12.5, positioning it as the primary progenitor pool for the developing palatal epithelium.

A *K5*-Positive Progenitor Pool Gives Rise to Both Basal and Periderm Layers in a Temporally Defined Manner

To define the lineage relationship between *K5*$^+$ cells, the *K14*$^+$ basal layer, and the KRT6A$^+$ periderm (Figure 2A), we first performed lineage tracing using *K5CreERT2;mTmG* mice. A pulse of tamoxifen (TM) at E10.5 robustly labeled *K5*-expressing cells and their progeny (GFP$^+$). At E12.5, cells from this *K5*$^+$ lineage were already present in both the basal and superficial layers of the palatal epithelium, and many superficial GFP$^+$ cells being negative for the basal marker KRT14 (Figure 2, B and E; Figure S6A). As development proceeded to E14.5, the *K5*$^+$ lineage gave rise to nearly all KRT14$^+$ basal cells and all KRT6A$^+$ superficial periderm cells (Figure 2, C, D, F, G, I, J, L, M and N; Figure S6B and C). Thus, a common *K5*$^+$ progenitor pool exists at E10.5 and generates the two major epithelial layers of the developing palate.

This finding raised a question regarding the timing of lineage commitment: Is the *K14*$^+$ lineage specified concurrently with the *K5*$^+$ lineage from a common K5$^+$/K14$^+$ progenitor, or does it emerge later from a *K5*$^+$ progenitor that has not yet activated KRT14 expression (Figure 2O)? To resolve this, we generated a dual-recombinase reporter mouse (*K5CreERT2;K14DreERT2;H11-LSL-GFP-RSR-tdT*) capable of labeling the *K5*$^+$

lineage with GFP and the *K14*$^{+}$ lineage with tdTomato (tdT) upon TM induction (Figure 2O and Figure S7A).

When TM was administered at E10.5, only the *K5*$^{+}$ lineage (GFP^{+}) was labeled in the E14.5 palate, with no tdT^{+} cells observed (Figure 2P and Figure S6B). This early *K5*$^{+}$ lineage generated the KRT6A^{+} periderm and the majority of the KRT14^{+} basal layer (Figure 2, R, T, V and X). Induction at E10.5 resulted in labeling of only the *K5*$^{+}$ lineage, with no tdT^{+} cells detected from the *Krt14* locus at this early stage. While this could reflect a delay in DreERT2 expression, it suggested that the primary periderm contribution at this time point is K14-independent.

In contrast, a TM pulse at E12.5 resulted in the labeling of two distinct lineages by E14.5: a GFP^{+} *K5* lineage and a tdT^{+} *K14* lineage (Figure 2, Q, S, U, W and X and Figure S7B). The tdT^{+} *K14* lineage was predominantly confined to the KRT14^{+} basal layer, whereas the GFP^{+} *K5* lineage contributed to both the remaining basal cells and the entire KRT6A^{+} periderm. These *in vivo* lineage dynamics were recapitulated in an *ex vivo* palate culture system (Figure S7C), with higher genetic labeling efficiency using direct treatment of 4-Hydroxytamoxifen (4-OHT). To confirm that the observed lineage segregation was not due to differential labeling efficiency between the CreERT2 and DreERT2 systems, we examined the developing submandibular gland, a tissue where KRT5 and KRT14 are homogenously co-expressed after E12.5[27,28]. As expected, induction at E12.5 resulted in ubiquitous co-labeling (GFP^{+};tdT^{+}) of the salivary gland epithelium at E14.5, confirming the equivalent activity of the two recombinase systems under these conditions (Figure S8).

Together, these dual-tracing experiments reveal an important temporal shift in lineage potential (Figure 2Y). Before E12.5, a common *K5*$^{+}$ progenitor exists that has not yet activated *K14* expression; this K5^{+};K14^{-} population is the original progenitors of the periderm. By E12.5, this hierarchy resolves into two largely distinct lineages: the *K5*$^{+}$ lineage that continues to fuel the periderm earlier, and a newly established, committed *K14*$^{+}$ lineage that expands the basal layer and contributes to periderm later.

Single-Cell Multi-omics Defines a K5^{+};K14^{-} Progenitor that Gives Rise to the Periderm

To molecularly characterize the K5^{+};K14^{-} progenitor population identified by lineage tracing and to definitively map its fate, we performed single-cell multi-omic (scRNA-seq and scATAC-seq) analysis on the *K5*$^{+}$ lineage (GFP^{+}) isolated from the palatal epithelium of *K5CreERT2;mTmG* embryos at E12.5 and E14.5 (Figure 3A). We found that, consistent with our previous findings, clusters in scRNAseq could be clearly identified by *Krt5* and *Krt14* expression. Integrated analysis of the transcriptomic and epigenomic data revealed seven distinct epithelial states, including a “stem” population, a transient “K5^{+};K14^{-}” population with high expression of *Krt5* but low expression of *Krt14*, and a terminally differentiated “K5^{+};K6^{+}” periderm population (Figure 3B). The stem cluster, which comprised nearly 50% of epithelial cells at E12.5, was enriched for progenitor markers such as *Cxcl14*[29] and *Wnt3*[30], while the K5^{+};K6^{+} cluster expressed mature periderm markers including *Krt6a*[31] and tight junction genes[32] like *Cldn3* and *Cldn23* (Figure 3C and S9). Of note, the K5^{+};K14^{-} cluster, which expressed early periderm markers like *Grhl3*[33], was present at E12.5 but almost entirely absent by E14.5, confirming its transient nature (Figure S10A, B). The remaining clusters included a

stable “transit” population, a dynamic “Shh^{+}” signaling center, and two *K5*-negative populations (“K5-” and “Mature K5-”) that expanded over time (Figure S9 and S10C). To spatially resolve these transcriptomic states, we analyzed the correspondence between our cell type annotations and established regional markers. This mapping revealed distinct localization patterns: the “K5-” progenitors were predominantly enriched in the *Barx1*-positive posterior soft palate[34], the “Shh+” population localized to the rugae[34], and the “K5^{+};K14$^{–}$” populations were largely restricted to the periderm (Figure S11).

The chromatin accessibility landscape defined by scATAC-seq correlated well with the transcriptional states. For example, the K5^{+};K6^{+} periderm cluster showed highly accessible chromatin at terminally differentiated gene loci like *Krt6a* (either within 1kb of TSS or enhancers co-activated inferred by Signac[35]), whereas the K5^{+};K14$^{–}$ cluster displayed greater accessibility at early periderm specification genes such as *Grhl3* (Figure 3D, Table S2-3). This multi-omic approach provided a robust, high-resolution map of the cellular states within the developing palatal epithelium.

The identification of these distinct populations allowed us to resolve the basis for the low correlation between *Krt5* and *Krt14* expression observed previously (Figure 1B). While *Krt5* and *Krt14* were positively correlated in most progenitor populations, they showed no correlation in the K5^{+};K14$^{–}$ cluster at E12.5 (R^2 = -0.05) or in the K5^{+};K6^{+} periderm at both time points (R^2 = -0.07 at E12.5 and R^2 = -0.66 at E14.5), confirming that overall tissue-level heterogeneity is driven by these specific superficial cell states (Figure 3E).

To map the differentiation paths between these cell states, we performed trajectory inference using scVelo[36] and Moscot[37]. This approach integrates RNA velocity with experimental time-series data to robustly model cell type transition. These analyses

predicted two major trajectories converging on the $K5^+;K6^+$ periderm fate: one originating from the stem cluster and, critically, a second major trajectory originating directly from the $K5^+;K14^-$ population (Figure 3F, G and S12A-F). This finding provides strong computational evidence that the $K5^+;K14^-$ population is a direct and significant progenitor of the $K5^+;K6^+$ periderm.

We next defined the molecular signatures associated with the two periderm differentiation trajectories, lineage ① represents the trajectory from Stem to $K5^+;K6^+$, while lineage ② represents the trajectory from $K5^+;K14^-$ towards $K5^+;K6^+$ (Figure 3G). We identified eight distinct gene clusters with dynamic expression patterns (Figure 3H, Table S4). Notably, lineage ② was characterized by an early wave of *Sprr1a* and *Grhl3* expression, followed by the sequential activation of *Klf4*, *Krt6a*, and *Cldn3* (Figure 3H and S12G). Analysis of chromatin accessibility revealed corresponding dynamic changes in transcription factor motif activity. Consistent with the gene expression data, motifs for GRHL factors were significantly enriched in the early stages of the lineage ②, while motifs for factors like TEAD showed divergent activity between the two lineages. This was reflected in the gene activity scores of their targets; for instance, accessibility near key differentiation genes like *Sprr1a* and *Krtdap* was high early in lineage ② but only peaked late in lineage ① (Figure 3I and S12H).

Collectively, these single-cell analyses provide molecular evidence for a specific $K5^+;K14^-$ progenitor population that differentiates into the $K5^+;K6^+$ periderm and define the distinct transcriptomic and epigenetic roadmap of this developmental transition.

ARID3A Is Required for Periderm Progenitor Proliferation to Ensure Palate Morphogenesis

Having defined the K5^{+};K14^{-} population as a direct progenitor of the periderm, we next sought to identify the principal transcriptional regulators governing this specific cell fate transition. By integrating our scRNA-seq and scATAC-seq data, we performed a gene regulatory network analysis with SCENIC^{+}[38], which predicted that the *Arid3a* regulon was uniquely and highly active in the K5^{+};K14^{-} progenitor population, distinguishing it from the downstream K5^{+};K6^{+} periderm cells (Figure 4A and S13A-C). Previous publication also identified *Arid3a* as a robust marker of the periderm cluster[39]. An *in silico* perturbation screen using CellOracle[40] (Figure S13D) independently predicted that among 212 transcription factors highly expressed in our scRNA-seq profile. ARID3A, alongside known periderm regulators GRHL1 and GRHL3, was a top candidate whose loss would specifically inhibit the K5^{+};K14^{-} to K5^{+};K6^{+} periderm differentiation trajectory (Figure 4B and Figure S13E). Consistent with these predictions and previous scRNA-seq for periderm marker genes[18], we confirmed that *Arid3a* is highly and specifically expressed in the K5^{+};K14^{-} cluster at E12.5 (Figure 4C-E).

We next validated the *in vivo* expression pattern by Immunofluorescence, which confirmed that ARID3A protein is specifically localized to the superficial epithelial layer of the palatal shelves at E12.5 and is lost upon ablation of the *K5* lineage (Figure 4F and S14). We then generated *K5CreERT2;Arid3a*$^{fl/fl}$ mice to conditionally delete *Arid3a* from the early *K5* progenitor pool via tamoxifen induction at E10.5 (Figure 4G and S15A, B). While both control embryos and *K14Cre;Arid3a*$^{fl/f}$ completed palatal fusion by E15.5, the *K5CreERT2;Arid3a*$^{fl/fl}$ conditional knockout embryos exhibited a severe delay in palate closure, with a significantly wider gap between the palatal shelves at E14.5 and a persistent epithelial seam at E15.5 (Figure 4G and S15C). *Ex vivo* palatal cultures

303 treated with sg*Arid3a*-virus revealed pronounced developmental delays in the
304 experimental group compared to controls as well (Figure S16).

305 Histologically, the superficial epithelium of knockout embryos appeared disorganized,
306 with regions of abnormal multi-layering (Figure S15C). While there was no increase in
307 apoptosis (Figure S15D), we observed a significant and specific reduction in cell
308 proliferation, as measured by Ki67 staining, in the superficial epithelial layer. Notably,
309 proliferation in the underlying basal layer was unaffected (Figure 4H). This defect in
310 superficial cell proliferation had a profound non-cell-autonomous consequence on the
311 underlying tissue. The reduced expansion of the epithelial surface was associated with
312 a significant condensation of the palatal mesenchyme (Figure 4G).

313 To determine if the loss of periderm proliferation is sufficient to drive this mesenchymal
314 crowding, we developed an agent-based stochastic model simulating palatal shelf
315 growth dynamics (Figure 4I and Figure S17A; see Methods). By specifically inhibiting
316 the periderm self-renewal parameter (rself) while maintaining constant mesenchymal
317 proliferation, our simulation reproduced the in vivo phenotype: the "knockout" model
318 exhibited a significantly smaller total tissue area and a marked increase in internal
319 mesenchymal density due to spatial confinement (Figure 4J, Figure S17B, and Movie
320 S1). This crowding effect was further validated by quantifying the Nearest Neighbor
321 Distance (NND), which was significantly reduced in the proliferation-deficient model
322 (Figure S17C). These computational predictions were confirmed by histological
323 quantification, which showed a marked increase in mesenchymal cell density directly
324 beneath the epithelium in mutant embryos (Figure 4K). Thus, our modeling supports a

mechanical coupling where the expansion rate of the epithelial boundary dictates the organization of the underlying tissue.

These findings identify ARID3A as an essential, lineage-specific regulator of periderm progenitors. The delayed-but-not-halted fusion phenotype in the *Arid3a* knockout can be explained by the dual-trajectory model of periderm formation we proposed (Figures 2 and 3). We propose that ARID3A is required for the rapid proliferative expansion of the $K5^{+};K14^{-}$ progenitor lineage, which constitutes the primary source of periderm. In its absence, a second, ARID3A-independent lineage, likely originating from the basal/stem cell pool, can slowly compensate to eventually form a functional periderm, leading to delayed fusion. Supporting this, conditional deletion of *Arid3a* using a *K14Cre* driver did not cause a palate closure defect, demonstrating that ARID3A function is dispensable in the $K14^{+}$ lineage. Unlike the complete failure of palatogenesis seen in lineage ablation, the Arid3a mutant phenotype reveals a critical mechanism of developmental robustness: while the $K5^{+};K14^{-}$ lineage is essential for rapid growth, the palate retains a latent compensatory capacity to safeguard organogenesis against genetic perturbations.

ARID3A Directly Regulates a *Grhl2* Distal Enhancer to Control Periderm Development

To uncover the downstream molecular mechanisms by which ARID3A regulates periderm development, we performed ARID3A CUT&TAG (Table S5) on palate epithelium isolated from E14.5 embryos when the delayed phenotype became most significant (Figure 5A). To map the complete regulatory landscape, we integrated these data with H3K27Ac CUT&TAG (Table S6) and HiCUT-based chromatin interaction

profiles (Table S7) in E14.5 palatal epithelium, developing an Enhancer-Promoter (E-P) Annotator to link ARID3A-bound regions to their target genes (Figure 5B).

This genome-wide analysis revealed that ARID3A binds to thousands of active regulatory regions (overlapping with H3K27Ac consensus peaks), showing a strong preference for distal enhancers over promoters (2,825 vs. 826 sites, respectively) (Figure 5C). Gene Ontology (GO) analysis of the direct target genes revealed a striking functional dichotomy: ARID3A-occupied promoters were linked to genes involved in general cellular processes like transcription and protein stability, whereas ARID3A-occupied enhancers regulated a cohort of genes critical for morphogenesis, epithelial development, and cell cycle control (Figure 5D).

Among the top candidates for regulation by an ARID3A-occupied enhancer was *Grhl2*, a regulator of epithelial identity[41]. We identified a prominent ARID3A-bound enhancer located approximately 20 kb upstream of the *Grhl2* locus containing two putative ARID3A binding sites (ABS), and this is distinct from the previously reported *mm1286*[41] cleft-related element, situated about 1 kb away (Figure 5E).To functionally test the activity of this element, we utilized a zebrafish transgenesis assay as reported before[42]. This putative enhancer drove robust GFP expression specifically in the superficial epithelial layer (periderm) in a significant fraction of embryos (103/223) (Figure 5F). Critically, mutation of the two ARID3A binding sites (A2ABS) completely abolished this enhancer activity (4/205), demonstrating that its function is ARID3A-dependent (Figure 5G). To validate this regulatory link *in vivo*, we analyzed GRHL2 expression in our *K5CreERT2;Arid3a*$^{fl/fl}$ mice. Conditional deletion of *Arid3a* led to a marked reduction of GRHL2 protein in the superficial epithelium of the palate (Figure 5H).

Collectively, these data demonstrate that ARID3A directly controls a key developmental gene regulatory network by binding to distal enhancers, with the regulation of *Grhl2* representing a critical node in this network to ensure proper periderm development.

The K5^{+};K14^{-} Periderm Progenitor Lineage is a Cellular Hotspot for Human Orofacial Cleft Risk Variants

Having established the K5^{+};K14^{-} progenitor population as a critical cell type for palate morphogenesis, we sought to determine its relevance to human orofacial clefts (OFCs). We employed SNP2Cell[43], a tool designed to link polygenic traits to cell-type-specific gene regulatory networks (GRNs), by integrating published human GWAS data with our single-cell multi-omic atlas (Figure 6A).

We first curated a list of 3,439 SNPs previously associated with human OFCs (Table S8) and successfully lifted over 1,123 of these to orthologous regions of the mouse genome (Table S9). Of these, 221 SNPs overlapped with chromatin accessibility peaks present in our palatal epithelium scATAC-seq dataset (Table S10). We then used SNP2Cell to calculate an enrichment score for these OFC-associated risk variants across the GRNs of each of our seven defined epithelial cell states (Table S11).

The analysis revealed a specific enrichment of high SNP2Cell scores in only two populations: the K5^{+};K14^{-} progenitors and their immediate descendants, the K5^{+};K6^{+} periderm (Figure 6B). In contrast, the “Stem” population and other epithelial lineages showed no significant enrichment. This result points to the regulatory landscape of the transient, superficial epithelial lineage as the primary group of cells by which genetic risk variants associated with OFCs are important. By providing a "cellular address" for

previously abstract genetic risks, these findings suggest that the $K5^{+};K14^{-}$ progenitor population is a critical cellular locus of susceptibility for human orofacial clefting.

Discussion

In this study, we sought to resolve the long-standing controversy regarding the lineage origins and differentiation trajectories of the embryonic periderm. This transient epithelial layer plays an essential biological function during organogenesis by serving as a 'Teflon-like' non-stick barrier that safeguards the extension of tissues like the palatal shelves, preventing pathological adhesions during a critical window of morphogenesis[6,7]. Because of its essential role during palatogenesis, we employed the mouse palate as a model in a time window from E10.5 to E14.5, when the palatal epithelium stratifies from a simple bi-layered to a multi-layered structure. Here, we resolve the lineage origins of the embryonic periderm and propose a 'dual-trajectory' model that ensures the robustness of craniofacial morphogenesis. Unlike the classical view of a single, linear stratification process, we identify a specialized, ARID3A-dependent $K5^{+};K14^{-}$ progenitor lineage programmed for rapid periderm expansion. We demonstrate that this lineage acts as a 'fast-track' mechanism essential for timely palate closure, safeguarding the embryo against developmental delays that could lead to birth defects. This initial phase is driven by an ARID3A-GRHL2 regulatory axis, where ARID3A binds a distal enhancer to activate *Grhl2*[41,44–46], thereby coordinating the expansion of the epithelium with the morphogenesis of the underlying mesenchyme.

This model fundamentally redefines the lineage hierarchy of the developing palatal epithelium, building on the classical, linear model of stratification in which a homogenous $K5^{+};K14^{+}$ basal layer is thought to generate all suprabasal cells. Our finding that *Krt5* expression is established in a progenitor pool prior to the widespread activation of *Krt14* is consistent with foundational studies of epidermal development.

Byrne et al. [25] first reported that *Krt5* is expressed in the single-layered ectoderm in specific embryonic regions, well before stratification is complete. Similarly, KRT5 expression has been reported to precede that of KRT14 in human fetal skin[26]. Recent single cell profiling of mouse embryonic skin development has indicated an "unexpected epithelial heterogeneity at E12.5"[8], and our work describes this complexity by providing a definitive lineage map and functional hierarchy. The temporal switch we uncovered using dual-lineage tracing—from a single, common *K5*$^{+}$ progenitor pool at E10.5 to two distinct *K5*$^{+}$ and *K14*$^{+}$ lineages by E12.5—provides an *in vivo* example of progressive fate restriction. This delineation of a specific progenitor pool for the periderm, distinct from the committed basal layer, clarifies how a complex, multi-layered epithelium is constructed and explains the developmental robustness observed in the *Arid3a* knockout.

A key finding is the role of ARID3A as a specific regulator of this transient progenitor state. ARID3A has been widely reported to drive development of both myeloid and B cell lineages[47], and more recently, to regulate the proliferation of transit-amplifying progenitors in the adult intestinal epithelium, a process essential for injury-induced regeneration[48]. We found that ARID3A is not a general keratinocyte factor but rather is specially required for the proliferative expansion of the K5^{+};K14^{-} population. ARID3A appears to exert this tissue-specific control primarily through distal enhancers, consistent with our understanding of how fate decisions are regulated by long-range chromatin interactions. We identified *Grhl2* as a direct target, and proper GRHL2 dosage is critical for craniofacial development, as both loss[44,45] and gain[46] of *Grhl2* function cause severe orofacial clefting and neural tube defects in mice. This places the

ARID3A-GRHL2 axis within the known periderm gene regulatory network. Notably, the ARID3A-responsive enhancer we identified is located near other validated craniofacial enhancers, such as mm1286, whose inactivation also predisposes mice to cleft palate[41], suggesting this locus as a hub for developmental gene regulation and an interesting area for future study. Given that IRF6 also regulates GRHL family members[9], it is plausible that ARID3A functions in a parallel, indispensable pathway that converges on common downstream effectors to ensure robust periderm formation.

Perhaps the most surprising finding of our study is that palatal shelves in *Arid3a* knockouts eventually fuse, despite the initially severe defect in periderm proliferation. This observation directly touches on the question of whether the periderm expands as the embryo grows by proliferation of cells already residing within the layer or by addition of new periderm cells differentiated from the basal epithelial cells [5]. Our results provide a comprehensive answer: both processes occur, and their coordination underpins the developmental robustness of the palate. We show that ARID3A is essential for the primary, rapid wave of periderm expansion via proliferation of the $K5^+;K14^-$ lineage. The eventual fusion in its absence strongly suggests that a slower, compensatory mechanism exists. We hypothesize that this compensation arises from the established $K5^+;K14^+$ basal layer, which retains a latent plasticity to contribute to the superficial layer through an ARID3A-independent pathway. This model is supported by our lineage ablation findings, where complete ablation of the early $K5^+$ lineage results in a persistent cleft palate that never recovers, whereas specific deletion of *Arid3a* in the committed $K14^+$ lineage does not. Therefore, we propose that this 'dual-trajectory' system evolved to buffer morphogenesis: the ARID3A-dependent lineage provides the necessary

'acceleration' to match the rapid growth of the craniofacial complex, while the basal layer serves as a resilient reserve. The 'delayed fusion' phenotype thus represents the limits of this buffer—sufficient to rescue structure in a controlled mouse model, but likely rendering the system hypersensitive to additional genetic hits in complex human cases. However, the mechanisms that ensure this expansion occurs horizontally within the superficial layer, rather than driving further stratification into deeper layers, are unknown.

Finally, our integration of single-cell multi-omics with human genetics provides a precise 'cellular address' for orofacial cleft risk. Integrating our single-cell multi-omic data with human genetics, pointed to the $K5^{+}$;$K14^{-}$;$ARID3A^{+}$ progenitor as a potentially critical cell type for OFC susceptibility. This argues that the pre-E12.5 developmental window, when this progenitor population is specified and expands, is a key period of vulnerability, rather than later timepoints such as E14.5, when morphological defects become apparent. This provides a concrete, testable model for how non-coding genetic risk, acting through specific enhancers in a transient embryonic cell population, can manifest as a common and complex birth defect.

While our study resolves the primary origin of the periderm, we acknowledge certain limitations that open avenues for future research. Our model of compensation from the $KRT14^{+}$ basal layer, while compelling, requires formal lineage tracing for definitive proof. Our functional data using *K14-Cre* demonstrates that ARID3A is not required in the basal layer for this compensation to occur, strongly implicating it as the source. However, future studies employing an inducible *K14* driver to label the basal layer in an *Arid3a*-knockout background will be necessary to definitively trace the fate of these cells and formally prove they contribute to the regenerating periderm. Furthermore, the

upstream signals that activate ARID3A expression specifically within this narrow progenitor window remain to be identified, and these transient upstream and downstream factors may themselves represent candidate risk factors for orofacial clefting.

In summary, our work fundamentally redefines the lineage hierarchy of the palatal epithelium. We identify the ARID3A-GRHL2 axis not merely as a molecular switch, but as a critical 'developmental accelerator' that coordinates epithelial expansion with mesenchymal growth. By uncovering this lineage-specific mechanism of robustness, we offer a new framework for understanding how subtle deficiencies in progenitor dynamics can culminate in structural birth defects.

METHODS

Mouse models

All mice, used in the present study were maintained under specific pathogen-free conditions in the animal center attached to Wuhan University School and Hospital of Stomatology under the guideline and approval of the Institutional Animal Care and Use Committees at the School and Hospital of Stomatology attached to Wuhan University (protocol no.S07924030A). *Wild-type* (WT) mice were purchased from the Hubei Provincial Center for Disease Control and Prevention (Hubei CDC). *Krt14Cre* (*K14Cre*, #004782), *ROSA26*$^{mT/mG}$ (*mTmG*, #007676) and *ROSA26*^{DTA}(*DTA*, #009669) mice were purchased from the Jackson Laboratory. *Arid3a-flox* (#T008401) mice were purchased from GemPharmatech Co.,Ltd. *Krt5CreERT2* (*K5CreERT2*, NM-KI-200168), and Krt14-2A-DreERT2 (*K14DreERT2*) mice (#NM-KI-190125) and H11-CAG-LSL-H2B-EGFP-CAG-RSR-tdTomato (H11) mice (#NM-KI-240880) were purchased from Shanghai Model Organisms Centeusing. All mice used in the experiments were maintained under C57BL/6J background. Experiments were carried out with male and female mice. No gender-specific differences were observed. The investigators were not blinded to allocation during experiments and outcome assessment.

Zebrafish Lines

All zebrafish used in this study were raised at the China zebrafish Resource Center (CZRC) and maintained under a 14-h light/10-h dark cycle at 28.5 °C. All zebrafish experiments were performed in compliance with the ethical regulations of the University

of Chinese Academy of Sciences and the Institute of Hydrobiology, Chinese Academy of Science.

METHOD DETAILS

RNAscope In Situ Hybridization

All samples were embedded in Tissue-Plus O.C.T. Compound(Sakura Finetek, #23-730-571) and sectioned at 10 μm, frozen sections were were air-dried, rehydrated in PBS, Fluorescent in situ hybridization was performed on these sections using the RNAscope™ Multiplex Fluorescent Reagent Kit v2(ACD, Bio-Techne, #323100) according to the manufacturer's protocol. For co-detection experiments on wild-type tissue, probes for *Krt5* (ACD, #415041-C2) and *Krt14* (ACD, #422521) were used. For analysis of *K5CreERT2;Arid3a$^{f/f}$;mTmG* and *K5CreERT2;Arid3a$^{f/-}$;mTmG* mice, probes for *Krt5* and *Arid3a* (ACD, #525721) were co-hybridized. The signal was developed using tyramide signal amplification (TSA) with TSA® plus fluorophores (Akoya Biosciences), such as TSA® Vivid Fluorophore 520 (ACD, #PG-323271)(1: 1500), TSA® Vivid Fluorophore 570 (ACD, #PG-323272) and TSA® Plus Cy5 (ACD, #PG-323273)(1:1500). Slides were counterstained with DAPI and mounted with ProLong™ Gold Antifade Mountant (Thermo Fisher Scientific, #P36930). Tissues were imaged using Leica Thunder microscope and analyzed with ImageJ software (v2.14).

Lineage Ablation

To achieve temporal and cell-type-specific lineage ablation, we utilized a Cre-loxP-based system in which targeted cells express the Diphtheria Toxin Receptor (DTA) upon Cre-mediated recombination. *K14Cre;DTA* and *K5CreERT;DTA* mice were generated by crossing homozygous transgenic lines, with noon of vaginal plug detection designated as E0.5. Constitutive Cre activity in *K14Cre;DTA* embryos initiated immediate lineage ablation while Tamoxifen-dependent nuclear translocation of CreERT2 in *K5CreERT;DTA* embryos enabled temporally controlled ablation. For temporal induction of Cre recombinase activity, Tamoxifen (MCE, #HY-13757A) dissolved in corn oil (MCE, HY-Y1888) was administered at one single dose (75 µl per 10 g body weight, 10 mg/ml) intraperitoneally to pregnant *K5CreERT;DTA* mice at different embryonic dates from E10.5 to E12.5. Control groups included Cre-negative;DTA-positive, Cre-positive;DTA-negative, and untreated littermates. Both *K14cre;DTA* and *K5CreERT;DTA* embryos were harvested at E12.5 to E15.5 post-induction for histological and immunohistochemical analyses, to assess the extent of lineage loss and subsequent developmental consequences, using markers specific to the ablated cell population. Apoptosis of the DTA+ lineage was confirmed using TUNEL assay.

Genetic Lineage tracing

Inducible lineage tracing was performed using *K5CreERT2;mTmG, K14;mTmG(as control); K5CreERT2;K14DreERT2;H11*(*K5;K14;H11*) *mice*. For temporal induction of Cre or Dre recombinase activity, the same dose of Tamoxifen as before was administered intraperitoneally to pregnant mice at different embryonic dates at E10.5 or

E12.5. Tissues were harvested at E14.5 to E15.5 post-induction, After observing the overall morphology under a ZEISS Axio Zoom.V16 microscope, these tissue were embedded in OCT, and cryosectioned.

Lentivirus Production and Knockout Verification

The lentiviral expressing short guide RNA (sgRNA) targeting the sequence of *Arid3a* gene (CAGATGCCACCACCGGACCA, LV-Arid3a-sgRNA lentivirus) and negative control (CGCTTCCGCGGCCCGTTCAA, control lentivirus) were synthesized and cloned into GV708 (lentiCRISPRv2vector with BsmBI sites (purchased from Shanghai Genechem Co., Ltd.),recombinant vector was detected by DNA sequencing. The viral vector were transfected into 293T cells using Lipofectamine 2000 (Invitrogen; Thermo Fisher Scientific, Inc.) together with two helper plasmids psPAX2 and pMD2.G. Infectious lentiviruses were harvested 72 h post-transfection, rapid centrifugation to remove cell debris, and then filtered through 0.45 μm cellulose acetate filters. Virus titer was determined by fluorescence-activated cell sorting analysis of GFP positive 293T cells and was approximately 1.4×10^9 transducing units (TU)/mL medium, and stored at −80 C for further use.

Design primers that flank the genomic target site (upstream and downstream) to enable amplification and sequencing of the edited region. Use DNA from the mixed-cell population infected with LV-A-sgRNA lentivirus and from control lentivirus. Amplify the target locus with the designed primers and perform sequencing (e.g., Sanger) on the

amplicons. Align sequencing reads to the reference sequence and inspect the target locus and nearby bases for editing signatures.

Palatal Explant Cultures

Once embryos were dissected from the uterus in PBS, limb bud morphology was immediately examined to urther verify the age of embryos[49]. We used E10.5 and E12.5 *K5;K14;H11* and *WT* embryos for palatal explant cultures and removed the the brain, tongue and mandible of their heads. The remaining maxilla and palatal shelves (the explant) were placed on the Nuclepore™ Track-Etch membrane (Cytiva, #10417106), filled with 3 mL of BGJb medium containing 1x GlutaMAX (Gibco, #35050061). Penicillin (50 unit/mL) and streptomycin (50 µg/mL) were added into the culture medium, but serum was not added. After 24 hour treated with 1mg/ml 4-Hydroxytamoxifen (4-OHT) for *K5;K14;H11* explants or LV-Arid3a-sgRNA lentivirus and control lentivirus for WT explants, the explants were transferred into a 15-mL tube filled with 7 mL of BGJb culture medium with compliment mentioned above. The culture medium was flushed for 3 min with oxygen followed by 3 sec with carbon dioxide and left in air for several minutes. Explants were cultured with slow rotation (10-20 rpm) on a horizontal rotator in incubator. After 24 hour to 96 hour culture, explants were photographed and then were either fixed with 4% PFA or continually cultured in the fresh culture medium. At least 5 explants were prepared for culture at each developmental stage (E10.5 and E12.5 *K5;K14;H11* embryos for lineage tracing *ex vivo*; E12.5 *WT* embryos for *Arid3a* knockout exoeriences ex vivo).

Mathematical Modeling

We constructed an agent-based stochastic model to simulate the growth dynamics of the palate during developmental stages E13.5–E14.5. The average size of periderm, basal cells and mesenchyme cells were measured. The model couples three dynamic processes: periderm growth, geometric spatial expansion, and mesenchymal filling. In our simulation, the palate is simplified as a semi-circular geometry where the periderm cells (N_P) form the boundary, and the mesenchymal cells (N_M) fill the internal space. Periderm Growth Dynamics: Periderm cell population growth is driven by two distinct mechanisms: differentiation from basal cells and self-renewal. We assumed the basal cell population is proportional to the periderm surface area (NBasal ≈NP). The rate of change in periderm cell number is described by the differential equation:

$$\frac{dN_P}{dt} = \underbrace{\alpha \bullet N_P(t)}_{\text{Differentiation}} + \underbrace{\beta \bullet N_P(t)}_{\text{Self-Proliferation}}$$

where α (BASAL_TO_PERIDERM_RATE) represents the rate of differentiation from basal to periderm cells, and β (PERIDERM_PROLIF_RATE) represents the periderm self-renewal rate. To model the biological phenotypes, we defined the periderm with proliferation (PwP) condition with β_{PwP} >0, resulting in compound growth of the boundary. In contrast, the periderm without proliferation (PwoP) condition is modeled with β_{PwoP} ≈0, where the periderm loses self-renewal capacity and growth relies solely on passive replenishment from the basal layer. Geometric-Spatial Coupling A key assumption of our model is that the periderm size physically constrains tissue boundaries. The radius R(t) of the palate is derived directly from the periderm count:

$$L(t) = N_P(t) \bullet \delta$$

$$R(t) = \frac{L(t)}{\pi}$$

where δ (PERI_CELL_LEN) is the physical length of a single periderm cell. This defines a tissue space expansion ratio λ:

$$\lambda(t) = \frac{R(t)}{R(t-1)}$$

This parameter λ is applied to all internal cells to simulate passive biomechanical stretching during tissue growth, updated as $x_i(t) = x_i(t-1) \cdot \lambda(t)$.

Mesenchymal Dynamics and Density Evolution: Mesenchymal cells proliferate exponentially within the defined boundary, governed by $dN_M/dt = \gamma \cdot N_M(t)$, where γ (MESENCHYME_PROLIF_RATE) is the proliferation rate. We assumed γ remains constant between PwP and PwoP conditions. A boundary constraint was enforced such that for any mesenchymal cell i, its position vector x_i satisfies $|x_i| \leq \eta \cdot R(t)$, where η (MESENCHYME_LIMIT) represents the relative position coefficient of the basement membrane.

Finally, we derived the evolution of cell density ρ(t) by comparing spatial expansion versus cellular proliferation. The tissue capacity (Area A(t)) scales with $e^{2(\alpha+\beta)t}$, while mesenchymal content scales with $e^{\gamma t}$. The density evolution is thus described by:

$$\rho(t) = \frac{N_M(t)}{A(t)} \propto \frac{e^{\gamma t}}{e^{2(\alpha+\beta)t}} = e^{[\gamma - 2(\alpha+\beta)]t}$$

In the PwP model, active periderm proliferation ensures $2(\alpha+\beta) \approx \gamma$, maintaining homeostatic density. Conversely, in the PwoP model ($\beta \approx 0$), spatial expansion lags behind internal proliferation ($2\alpha \ll \gamma$), resulting in an exponential increase in cell density and the observed "crowding" phenotype.

The simulation was initialized at t=0 and evolved for a total duration of 48 hours. To account for the stochastic nature of the model and ensure statistical robustness, we performed 10 independent simulation runs (n=10) for each experimental condition (WT and KO). The final statistical results were derived from the ensemble data of these repeated realizations. Detail parameters and scripts for simulation could be found in github.

Immunofluorescence

Tissue sections were left to equilibrate at 37℃ temperature for 2 hour and washed in PBS. Tissue sections were permeabilized in 0.5% Triton-X in PBS for 10 min and blocked for 2 hour at 37°C temperature in 2.5% BSA (Sigma-Aldrich) in PBS. Section was incubated overnight at 4°C in primary antibody in 0.05% PBS-Tween-20. The following primary antibodies were used: rabbit anti-Cytokeratin5 (1:400, Abcam, ab52635), rabbit anti-Cytokeratin14 (1:400, Abcam, ab181595), rabbit anti-Cytokeratin17 (1:400, Abcam, ab109725), rabbit anti-keratin6a (1:800, Biolegend, 905701), Anti-GFP(1:400, Abcam, ab290), rat anti-Ecadherin (1:300, Life Technologies, 13-1900), Anti-p63(1:400, Abcam, ab735), ARID3A Polyclonal antibody(proteintech, 14068-1-AP), Ki-67 (D3B5) Rabbit mAb(1:200, CST, 9129), Anti Grhl2(1:200, Abcam, ab271023), TUNEL staining was performed using the One-step TUNEL In Situ Apoptosis Kit (Red, Elab Fluor® 594, Elabscience®). Immunohistochemistry staining was perforemed usig the Ready-to-use immunohistochemical UltraSensitiveTM SP kit (MXB, KIT-9710) and tissues were imaged using OLYMPUS confocal microscopes (SPINSR or FV1200), and analyzed with ImageJ software (v2.14).

H&E staining of frozen sections

Frozen sections that had undergone immunostaining were prepared for H&E staining by removing the coverslips and washing off the mountant under running water for 1 minutes. Then slides were sequentially immersed in hematoxylin solution (Servicebio, G1005-1) for 10 seconds, running water for 10 seconds, PBS for 5 minutes, following eosin solution (Servicebio, G1005-2) for 1 minutes and running water for 10 seconds. Sections were dehydrated at room temperature, and then cleaned with three changes of xylene substitute (Servicebio, G1138-1L). Finally, slides were mounted in neutral mounting medium (Servicebio, G8593). Images were acquired on a pannoramic MIDI microscope (3D HISTECH).

Dissociation of mouse palatal epithelium for downstream sequencing

The protocol for single cell dissociation was modified from previous studies[42]. Palatal epithelial cells were isolated through pooled samples from multiple litters. In ice-cold PBS, heads were removed from embryos. The top of the head was cut flat, and the head was laid ventral. The secondary palate was exposed by removing the mandible and tongue. Paired palatal shelves were washed twice in PBS and incubated with 10 mg/ml Dispase II (Millipore, Cat#04942078001) at 37°C and swirled at 100 rpm for 32 min to enzymatically separate the epithelium from the mesenchyme. Following epithelial separation using forceps in PBS, tissues were centrifuged at 400g for 5 minutes. The pellets were resuspended in 0.25% Trypsin-EDTA for 10 minutes at 37C with frequent trituration. The trypsinization procedure was repeated 3-5 times until no visible tissue

aggregates remained. Trypsinization was stopped with addition of DME/f12 + 10% FBS. Cells were pipetted up and down 20 times to assist in cell dispersion, pelleted at 300 g for 3 min, and resuspended in PBS with 5% FBS before being filtered through a 70uM cell strainer and maintained under standardized culture conditions prior to downstream sequencing applications. Cell numbers and viability were calculated for dissociated cells, which showed greater than 90% viability in all samples.

Preparation of Sequencing Libraries for scRNA-seq

Palatal epithelial cells were isolated from E12.5 (n=12) and E14.5 (n=14) *K5CreERT2*;*mTmG* conceptuses respectively for scRNAseq. The single cell library was generated using the 10x Genomics Chromium Next GEM platform (v3.1). Captured cells underwent reverse transcription to generate barcoded cDNA libraries, followed by enzymatic fragmentation, adapter ligation, and dual-indexed paired-end sequencing (Illumina NovaSeq 6000, PE150).

scRNA-seq data Analysis

Raw sequencing data were demultiplexed and aligned to the reference genome (mm10) using Cell Ranger (v7.1.0) to generate feature-count matrices with unique molecular identifier (UMI) correction. Initial QC filters excluded cells with <500 or >6,000 detected genes, mitochondrial gene content >10%, or ribosomal RNA content <5%. Doublets were identified via Scrublet (v0.2.3; threshold: doublet score >0.25). Raw counts were

normalized using SCTransform (Seurat v5.0) to correct for sequencing depth and stabilize variance.

Datasets from E12.5 and E14.5 *K5CreERT2;mTmG* embryos were integrated via scvi (v1.0) [50] to correct batch effects while preserving biological variation. Both the interated datas and the scRNA-seq data of E11.5 mouse lambdoidal junction from a published dataset[18] were analysed by the following procedures. Clustering was performed using the Leiden algorithm (resolution=0.5) within Scanpy (v1.11.1). Uniform Manifold Approximation and Projection (UMAP) was employed for non-linear dimensionality reduction. Cluster identities were annotated via marker gene detection (Wilcoxon rank-sum test, $\log_2$ FC >0.5, adjusted P<0.01). To validate cell identidy transformation between E12.5 to E14.5, we ultilized scvelo (dynamical model) (v0.25.0)[51] and moscot[52]. The former was ultilized spliced information to comput RNA velocity, the latter ultilized experimental time information. Pseudotime was inferred using CellOracle(v0.18.0) [53]. Driver genes along trajectories were identified via TradeSeq (v0.99.46)[54] . GO enrichment analysis was conducted using clusterProfiler.

Preparation of Sequencing Libraries for scATAC-seq

Palatal epithelial cells were isolated from E12.5 (n=16) and E14.5 (n=14) wildtype embryos for scATAC seq. Libraries were prepared using the Chromium Next GEM Single Cell ATAC Kit (10x Genomics). Nuclei were encapsulated in GEMs and subjected to Tn5 transposition to tag accessible chromatin with cell barcodes. Transposed fragments were then PCR amplified to add Illumina adapters and dual

indices (i7/i5), followed by purification and size selection. Libraries were sequenced on the NovaSeq Plus with paired-end reads and dual-index reads as recommended. Data can be processed with Cell Ranger ATAC.

scATAC-seq data Analysis

The scATAC-seq data analysis was performed as follows. Raw sequencing reads were aligned using Cell Ranger ATAC (v1.2.0). The aligned data underwent quality control using the Signac package. Subsequently, dimensionality reduction was performed with SnapATAC2 (v2.4.0) [55]. We first integrated the scRNA-seq and scATAC-seq datasets using the scglue (v0.3.3) [56], which enabled the transfer of cell type annotations from the scRNA-seq to the scATAC-seq data. Following this integration, ArchR (v1.0.2) [57] was used to infer gene scores, calculate motif accessibility scores, and identify differentially accessible regions (DARs) between cell clusters. Finally, CellOracle was utilized for trajectory inference and tradeSeq was used to identify genomic regions with dynamic accessibility along these inferred lineages.

GRN inference:

To construct a comprehensive gene regulatory network (GRN), The integrated single cell data was analyzed with the SCENIC+ pipeline(v1.0) [58] to predict transcription factors (TFs), their putative target genes, and associated regulatory regions. The workflow began with pseudo-pairing of scRNA and scATAC profiles using figR(v0.1.0) [59]. Subsequently, pycisTopic (v1.0.2) was applied to the scATAC-seq fragment counts to

perform topic modeling, identifying consensus accessible regions that serve as candidate TF binding sites. A custom cisTarget database was created based on these consensus regions. Regulons were then constructed by identifying enriched TF motifs within regions linked to target genes, thereby establishing both TF-region and TF-gene relationships. Finally, the activity of each regulon was scored using AUCell based on the expression of target genes and the accessibility of target regions. Regulon specificity scores were subsequently derived from these activity scores to identify cell-type-specific regulators.

We performed in silico TF perturbation analysis using CellOracle to predict the effects of TFs on cell fate. The GRN previously predicted by SCENIC+ was imported into CellOracle as the base network. Leveraging the differentiation trajectory described in the previous section, we derived pseudotime-based differentiation vectors from pseudotime gradients. The perturbation scores were then computed as the cross-product of these differentiation vectors and simulated perturbation vectors. These scores indicate whether the perturbation of a given TF (overexpression or knockout) promotes or opposes differentiation along the osteogenic lineage. This simulation was performed systematically for all TFs within the GRN to comprehensively map their regulatory impact.

SNP2CELL Analysis

To identify disease-relevant regulatory networks, we employed the SNP2CELL(v0.2.0) [43]framework to integrate human GWAS data with our mouse single-cell analyses. First, we obtained GWAS summary statistics for traits related to cleft lip and palate from the GWAS Catalog, comprising an initial set of 3,437 SNPs. These human genome-referenced SNPs (hg38) were subsequently lifted over to the mouse genome (mm10) using the UCSC Genome Browser's liftOver tool, resulting in 1,123 mappable SNPs. For the single-cell component, we used two inputs: a gene regulatory network (GRN) inferred from our scRNA-seq data via the SCENIC+ pipeline, and cell-type specificity scores derived from cluster marker genes. Finally, the processed GWAS data, the SCENIC+-inferred GRN, and the cell-type scores were provided as input to the SNP2CELL pipeline to prioritize disease-relevant and cell-cluster-specific modules within our transcription factor network.

Preparation of Sequencing Libraries for H3K27Ac HiCUT

Palatal epithelial cells were isolated from E14.5 wildtype embryos(n=70) for HiCUT seq. Genomic DNA was sheared and then subjected to proximity ligation to join interacting DNA fragments. Following biotinylation and enrichment of ligated junctions, libraries were amplified by PCR. The final libraries were sequenced on an Illumina NovaSeq™ instrument with dual-indexed paired-end reads.

HiCUT-seq data Analysis

Paired-end reads were aligned independently to the mm10 genome using bowtie2 (global parameters:—very-sensitive–L 30 –score-min L,-0.6,-0.2 –end-to-end—reorder; local parameters:—very-sensitive–L 20 –score-min L,-0.6,-0.2 –end-to-end—reorder) through the HiC-Pro software[60]. The valid pair file was used for downstream analysis. Trans and < 1kb interactions were filtered out. For comparative analyses, we identified chromatin interactions that fall between a 20Kb to 2Mb window that have at least one anchor of the interacting pair falling in the Arid3a CUT&TAG peak region. Following the Arid3a CUT&TAG peak overlap and distance thresholding, we consider each valid pair as an individual interaction. If two overlapping valid pairs were present, we considered them as two interaction points.

Preparation of Sequencing Libraries for CUT&Tag

Palatal epithelial cells were isolated from E14.5 wildtype embryos (n=14) for CUT&TAG seq. The CUT&Tag assay was conducted using Hyperactive Universal CUT&Tag Assay Kit for Illumina (Vazyme, TD903). Epithelial cells that digested from palate shelves (1×10^6) were used for the CUT&Tag experiment. Cells were incubated in the H3K27Ac antibody (Abcam, #ab4729) or the Arid3a antibody (proteintech,14068-1-AP) overnight. PA-Tn5 transposase was used to cut the genome and add a special adaptor sequence to build a library. Libraries were constructed using Hyperactive Universal CUT&Tag Assay Kit (Vazyme, TD903) for Illumina' and TruePrep Index Kit V2(Vazyme, TD202) for Illumina.

CUT&Tag-seq data Analysis

Raw paired-end sequenced reads were first cut for adaptor sequences and trimmed using trim_galore (v0.6.6). Then, cleaned fastq data were mapped using Bowtie 2 (v2.4.2) to the mm10 reference genome. Picard was used for marking and removing duplication. Samtools (v1.11) was used for converting and sorting the SAM files into BAM format. Macs2 (2.2.7.1) was used for peak calling. A total of 32,347 and 31,810 peaks were obtained from H3K27Ac and Arid3a CUT&Tag, respectively. Deeptools was used for normalizing signals from sorted BAM files via bamCoverage, and computeMatrix was used for calculating the overall signal distribution around the peak center called by macs2 using the 'reference-point' model. Data were visualized using plotProfile and plotHeatmap in deeptools.

Plasmid constructs and transient reporter analysis of periderm enhancer in zebrafish

The candidate enhancer elements investigated in this study (chr15:37,202,744-37,203,504) were amplified by PCR using the KOD-Plus-Mutagenesis Kit (Toyobo, #SMK-101), with primer sequences showed as follows. Following sequence verification by Sanger sequencing, each element was inserted into the reporter vector pSK-MSC-GFP-Tol2 via homologous recombination using PCR-generated homology arms (20 bp each) for in vivo enhancer assays in zebrafish. For each assay, a minimum of 200 one-cell stage zebrafish embryos were injected, with each embryo receiving a co-injection of the reporter construct (~20 pg) and *tol2* transposase mRNA (~25 pg)[61]. This procedure

was performed in three independent biological replicates on different days. At approximately 11 hpf, injected embryos were screened for enhancer-driven activity in the epidermis via epifluorescence microscopy (Leica). Subsequently, the detailed epithelial expression patterns from transgenic lines that displayed consistent GFP expression in at least 20% of the injected individuals were documented.

Tg(Arid3a_miniF_GFP)_Pla_R	GAGTTACAGtaccctgttatccctagta
Tg(Arid3a_miniF_GFP)_Pla_F	GGTGTGTTACccagtgacgtaggaagtcc

QUANTIFICATION AND STATISTICAL ANALYSIS

Panels generally represent multiple independent experiments performed on different days with different mice. No formal randomization techniques were used but mice were allocated to experiments randomly and samples were processed in an arbitrary order. No blinding was performed. Power calculations were used to ensure that the number of animals used per experiment was high enough to ensure reproducibility. Male and female mice were used equally and were grouped together in the analyses. Variation was always indicated using standard deviation (SD).

Statistical analysis was performed with GraphPad Prism software (v 9.2.0), following its statistics guide. For analysis of the statistical significance of differences between two groups we generally performed two-tailed Student's t-tests. For analysis of the statistical significance of differences among more than two groups, we performed repeated measures one-way ANOVAs and Tukey's multiple comparisons tests. To assess the statistical significance of differences among multiple groups at multiple time points, we

performed repeated measures two-way ANOVAs and Tukey's multiple comparisons tests. Variability within groups was tested using Levine's median test prior to performing t-tests or ANOVAs.

Data and code availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive[62] in National Genomics Data Center[63], China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA030477) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa. Scripts for data analysis and figure generation are available on GitHub (https://github.com/GilbertHan1011/palateDevelopManuscript and https://github.com/Badgerliu/EP_annotator).

Reference

1. Dixon, M. J., Marazita, M. L., Beaty, T. H. & Murray, J. C. Cleft lip and palate: understanding genetic and environmental influences. *Nat. Rev. Genet.* **12**, 167–178 (2011).
2. Salari, N. *et al.* Global prevalence of cleft palate, cleft lip and cleft palate and lip: A comprehensive systematic review and meta-analysis. *J. Stomatol. Oral Maxillofac. Surg.* **123**, 110–120 (2022).

885 3. Lough, K. J., Byrd, K. M., Spitzer, D. C. & Williams, S. E. Closing the Gap:
886 Mouse Models to Study Adhesion in Secondary Palatogenesis. *J. Dent. Res.* **96**,
887 1210–1220 (2017).
888 4. Bush, J. O. & Jiang, R. Palatogenesis: morphogenetic and molecular mechanisms of
889 secondary palate development. *Dev. Camb. Engl.* **139**, 231–243 (2012).
890 5. Yu Lan & Rulang Jiang. Mouse models in palate development and orofacial cleft
891 research: Understanding the crucial role and regulation of epithelial integrity in facial
892 and palate morphogenesis. *Curr. Top. Dev. Biol.*
893 https://doi.org/10.1016/bs.ctdb.2021.12.003 (2022)
894 doi:10.1016/bs.ctdb.2021.12.003.
895 6. Richardson, R. J. *et al.* Periderm prevents pathological epithelial adhesions during
896 embryogenesis. *J. Clin. Invest.* **124**, 3891–3900 (2014).
897 7. Hammond, N. L., Dixon, J. & Dixon, M. J. Periderm: Life-cycle and function
898 during orofacial and epidermal development. *Semin. Cell Dev. Biol.* **91**, 75–83
899 (2019).
900 8. Jacob, T. *et al.* Molecular and spatial landmarks of early mouse skin development.
901 *Dev. Cell* **58**, 2140–2162.e5 (2023).
902 9. Peyrard-Janvid, M. *et al.* Dominant Mutations in GRHL3 Cause Van der Woude
903 Syndrome and Disrupt Oral Periderm Development. *Am. J. Hum. Genet.* **94**, 23–32
904 (2014).

905 10. Kondo, S. *et al.* Mutations in IRF6 cause Van der Woude and popliteal pterygium
906 syndromes. *Nat. Genet.* **32**, 285–289 (2002).
907 11. Oberbeck, N. *et al.* The RIPK4–IRF6 signalling axis safeguards epidermal differentiation
908 and barrier function. *Nature* **574**, 249–253 (2019).
909 12. Thomason, H. A. *et al.* Cooperation between the transcription factors p63 and IRF6 is
910 essential to prevent cleft palate in mice. *J. Clin. Invest.* **120**, 1561–1569 (2010).
911 13. Zucchero, T. M. *et al.* Interferon Regulatory Factor 6 (IRF6) Gene Variants and the
912 Risk of Isolated Cleft Lip or Palate. *N Engl J Med* (2004).
913 14. Gonzales, K. A. U. & Fuchs, E. Skin and Its Regenerative Powers: An Alliance
914 between Stem Cells and Their Niche. *Dev. Cell* **43**, 387–401 (2017).
915 15. Fuchs, E. & Green, H. Changes in keratin gene expression during terminal
916 differentiation of the keratinocyte. *Cell* **19**, 1033–1042 (1980).
917 16. Lane, J. *et al.* Control elements targeting Tgfb3 expression to the palatal epithelium
918 are located intergenically and in introns of the upstream Ift43 gene. *Front. Physiol.* **5**,
919 (2014).
920 17. Kousa, Y. A., Moussa, D. & Schutte, B. C. IRF6 expression in basal epithelium
921 partially rescues Irf6 knockout mice. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **246**,
922 670–681 (2017).

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

19. Zou, Z. *et al.* A Single-Cell Transcriptomic Atlas of Human Skin Aging. *Dev. Cell* **56**, 383-397.e8 (2021).

20. Wang, S. *et al.* Single cell transcriptomics of human epidermis identifies basal stem cell transition states. *Nat. Commun.* **11**, 4239 (2020).

21. Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018).

22. Chen, W. *et al.* Single-cell landscape in mammary epithelium reveals bipotent-like cells associated with breast cancer risk and outcome. *Commun. Biol.* **2**, 306 (2019).

23. Jones, K. B. *et al.* Quantitative Clonal Analysis and Single-Cell Transcriptomics Reveal Division Kinetics, Hierarchy, and Fate of Oral Epithelial Progenitor Cells. *Cell Stem Cell* **24**, 183-192.e8 (2019).

24. Dassule, H. R., Lewis, P., Bei, M., Maas, R. & McMahon, A. P. Sonic hedgehog regulates growth and morphogenesis of the tooth. *Development* **127**, 4775–4785 (2000).

25. Byrne, C., Tainsky, M. & Fuchs, E. Programming gene expression in developing epidermis. *Development* **120**, 2369–2383 (1994).

26. Moll, R., Moll, I. & Wiest, W. Changes in the Pattern of Cytokeratin Polypeptides in Epidermis and Hair Follicles During Skin Development in Human Fetuses. *Differentiation* **23**, 170–178 (1982).

27. Aure, M. H., Symonds, J. M., Mays, J. W. & Hoffman, M. P. Epithelial Cell Lineage and Signaling in Murine Salivary Glands. *J. Dent. Res.* **98**, 1186–1194 (2019).

28. Hauser, B. R., Aure, M. H., Kelly, M. C., Hoffman, M. P. & Chibly, A. M. Generation of a Single-Cell RNAseq Atlas of Murine Salivary Gland Development. *iScience* **23**, 101838 (2020).

29. Cheng, C.-L. *et al.* CXCL14 Maintains hESC Self-Renewal through Binding to IGF-1R and Activation of the IGF-1R Pathway. *Cells* **9**, 1706 (2020).

30. Sun, Z. *et al.* Joint single-cell multiomic analysis in Wnt3a induced asymmetric stem cell division. *Nat. Commun.* **12**, 5941 (2021).

31. Richardson, R. J. *et al.* Periderm prevents pathological epithelial adhesions during embryogenesis. *J. Clin. Invest.* **124**, 3891–3900 (2014).

32. Huang, W. *et al.* A Single-cell Atlas of Developing Mouse Palates Reveals Cellular and Molecular Transitions in Periderm Cell Fate. *Genomics Proteomics Bioinformatics* **23**, qzaf013 (2025).

33. Jacob, T. *et al.* Molecular and spatial landmarks of early mouse skin development. *Dev. Cell* **58**, 2140–2162.e5 (2023).

962 34. Im, H. *et al.* Molecular Regulation of Palatogenesis and Clefting: An Integrative
963 Analysis of Genetic, Epigenetic Networks, and Environmental Interactions. *Int. J. Mol.*
964 *Sci.* **26**, 1382 (2025).
965 35. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell
966 chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
967 36. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA
968 velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**,
969 1408–1414 (2020).
970 37. Klein, D. *et al.* Mapping cells through time and space with moscot. *Nature* **638**,
971 1065–1075 (2025).
972 38. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of
973 enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023).
974 39. Li, H., Jones, K. L., Hooper, J. E. & Williams, T. The molecular anatomy of
975 mammalian upper lip and primary palate fusion at single cell resolution. *Development*
976 dev.174888 (2019) doi:10.1242/dev.174888.
977 40. Kamimoto, K. *et al.* Dissecting cell identity via network inference and in silico gene
978 perturbation. *Nature* **614**, 742–751 (2023).
979 41. De Vries, M. *et al.* Interrogating the Grainyhead-like 2 (Grhl2) genomic locus
980 identifies an enhancer element that regulates palatogenesis in mouse. *Dev. Biol.* **459**,
981 194–203 (2020).

982 42. Liu, H. *et al.* Analysis of zebrafish periderm enhancers facilitates identification of a
983 regulatory variant near human KRT8/18. *eLife* **9**, (2020).
984 43. To, K. *et al.* A multi-omic atlas of human embryonic skeletal development. *Nature*
985 **635**, 657 - 667 (2024).
986 44. De Vries, M. E. *et al.* Grainyhead-like 2 interacts with noggin to regulate tissue
987 fusion in mouse. *Development* **151**, (2024).
988 45. Collier, A. E. *et al.* GRHL2 and AP2a coordinate early surface ectoderm lineage
989 commitment during development. *iScience* **26**, 106125 (2023).
990 46. Crane-Smith, Z. *et al.* A non-coding insertional mutation of*Grhl2*causes gene over-
991 expression and multiple structural anomalies including cleft palate, spina bifida and
992 encephalocele. *Hum. Mol. Genet.* **32**, 2681 - 2692 (2023).
993 47. Ratliff, M. L., Mishra, M., Frank, M. B., Guthridge, J. M. & Webb, C. F. The
994 Transcription Factor ARID3a Is Important for In Vitro Differentiation of Human
995 Hematopoietic Progenitors. *J. Immunol.* **196**, 614 - 623 (2016).
996 48. Angelis, N. *et al.* Loss of ARID3A perturbs intestinal epithelial proliferation-
997 differentiation ratio and regeneration. *J. Exp. Med.* **221**, e20232279 (2024).
998 49. Yu, K. & Yonemitsu, M. A. In Vitro Analysis of Palatal Shelf Elevation During
999 Secondary Palate Formation. *Anat. Rec.* **302**, 1594 - 1604 (2019).
1000 50. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative
1001 modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053 - 1058 (2018).

51. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).

52. Klein, D. *et al.* Mapping cells through time and space with moscot. *Nature* **638**, 1065–1075 (2025).

53. Kamimoto, K. *et al.* Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742–751 (2023).

54. Van den Berge, K. *et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020).

55. Zhang, K., Zemke, N. R., Armand, E. J. & Ren, B. A fast, scalable and versatile tool for analysis of single-cell omics data. *Nat. Methods* **21**, 217–227 (2024).

56. Cao, Z.-J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* **40**, 1458–1466 (2022).

57. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

58. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023).

59. Kartha, V. K. *et al.* Functional inference of gene regulation using single-cell multi-omics. *Cell Genomics* **2**, 100166 (2022).

60. Liu, H. *et al.* Analysis of zebrafish periderm enhancers facilitates identification of a regulatory variant near human KRT8/18. *eLife* **9**, (2020).

61. Xiao, Y. *et al.* Chromatin conformation of human oral epithelium can identify orofacial cleft missing functional variants. *Int. J. Oral Sci.* **14**, 43 (2022).

62. Zhang (张思思), S. *et al.* The GSA Family in 2025: A Broadened Sharing Platform for Multi-omics and Multimodal Data. *Genomics Proteomics Bioinformatics* **23**, qzaf072 (2025).

63. CNCB-NGDC Members and Partners *et al.* Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2025. *Nucleic Acids Res.* **53**, D30 - D44 (2025).

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (82322014 and 82270948) and The Interdisciplinary Research Project of School of Stomatology Wuhan University (No. XNJC202306) to Huan Liu; National Natural Science Foundation of China (Key Program No. 82230029) to Zhi Chen.

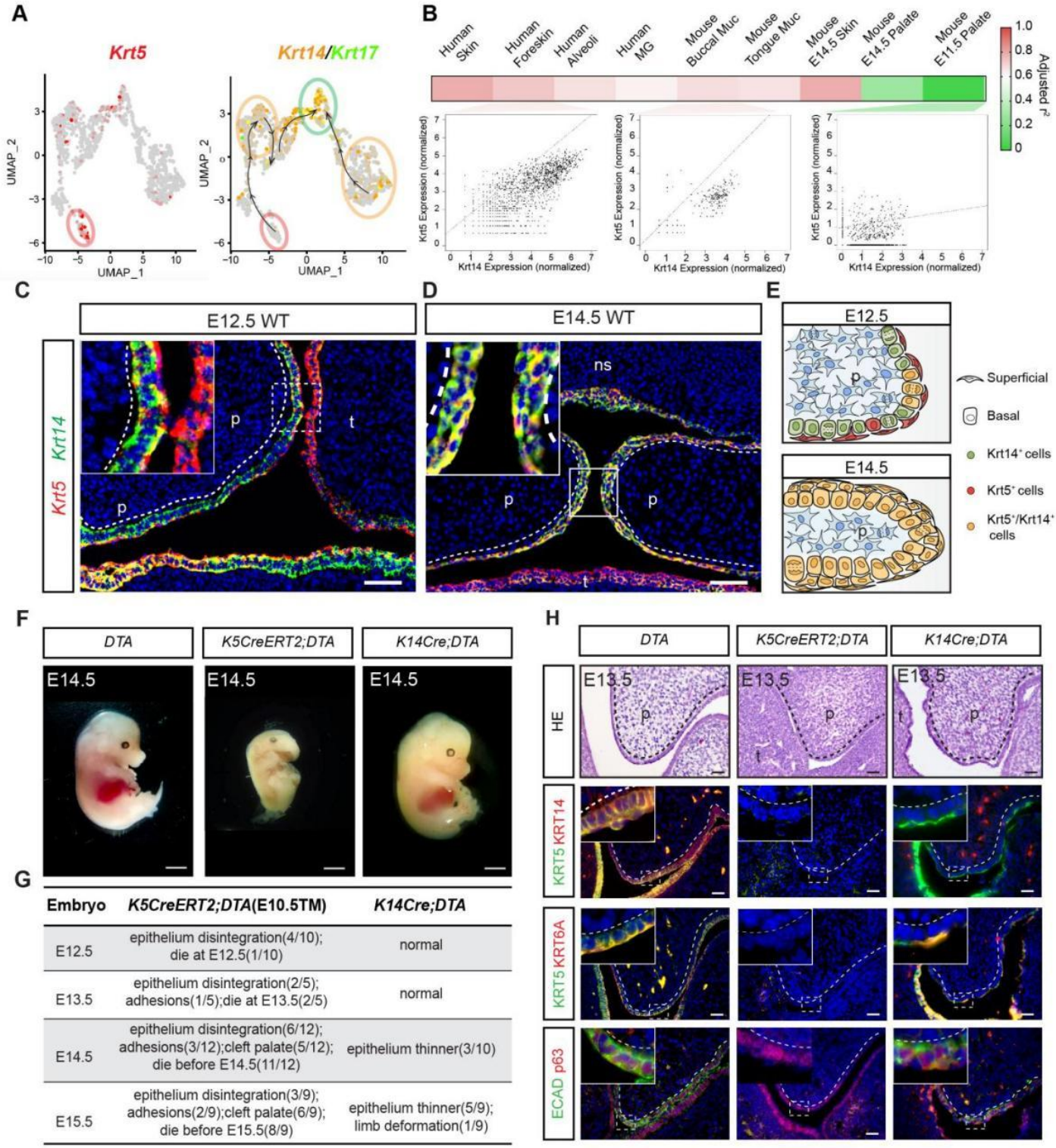
AUTHOR CONTRIBUTIONS

H.L., J.G. and L.H. designed the study and wrote the manuscript with critical input from O.D.K , Z.C., Y.S. R.A.C. and Z.W.; Y.S., H.L. and Z.C. contributed critical resources;

H.L., J.G., L.H., Y.X., Q.S.,Y.L., H.T. and F.P. conducted experiments, acquired data; H.L., J.G. and L.H. analyzed data.

COMPETING INTERESTS

The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.



Embryo	*K5CreERT2;DTA*(E10.5TM)	*K14Cre;DTA*
E12.5	epithelium disintegration(4/10); die at E12.5(1/10)	normal
E13.5	epithelium disintegration(2/5); adhesions(1/5);die at E13.5(2/5)	normal
E14.5	epithelium disintegration(6/12); adhesions(3/12);cleft palate(5/12); die before E14.5(11/12)	epithelium thinner(3/10)
E15.5	epithelium disintegration(3/9); adhesions(2/9);cleft palate(6/9); die before E15.5(8/9)	epithelium thinner(5/9); limb deformation(1/9)

1047 **Figure 1. The heterogeneity roles of *K5*⁺ and *K14*⁺ lineages during palatogenesis.**

(A)UMAP visualization of *Krt5*(Red), *Krt14*(Yellow) and *Krt17*(Green) and trajectory analysis of whole epithelial cells from E11.5 murine palatal tissues (GSE132462). Red circles mark *K5*$^{+}$ enriched population. (B) Correlation analysis of single-cell datasets from epithelial tissues reported in public databases including human skin(HRA000395)[19], human foreskin (GSE147482)[20], human tracheal epithelium (GSE103354)[21], human mammary gland epithelium (GSE113196)[22], human buccal mucosa (GSE120654)[23], mouse E14.5 skin (E-MTAB-11920)[8] and mouse E12.5 palate epithelium (this study). (C-E) Representative Multi-spectral RNAscope analysis of E12.5(C) and E14.5(D) palatal shelves of wild-type embryos and the corresponding schematic diagrams(E). Each staining was performed at three different animals (N⩾3 mice from ⩾3 independent experiments).The white dotted line represented the location of the basement membrane. Green fluorescence, *Krt14*; magenta fluorescence, *Krt5*; counterstaining with DAPI, blue nuclear fluorescence. ns, nasal septum; p, palatal shelf; t, tongue. Scale bar, 50μm. (F) Representative morphologies of E14.5 DTA (i), *K5CreERT2;DTA* (ii) and *K14Cre;DTA* (iii) mice. Scale bar, 1mm. (N = 6 mice per condition) (G)Tables of embryo phenotypes at E12.5-E15.5 *K5CreERT2;DTA*(E10.5TM) and *K14Cre;DTA* mice. N⩾3 mice in each timepoints were used. (H)H&E staining and dual immunostaining of KRT5 with KRT14, KRT6A or ECAD at E13.5 DTA, *K5CreERT2;DTA* and *K14Cre;DTA* mice palates. Each staining was performed at three different animals (N⩾3 mice from ⩾3 independent experiments). p, palatal shelf. Scale bar, 20μm.

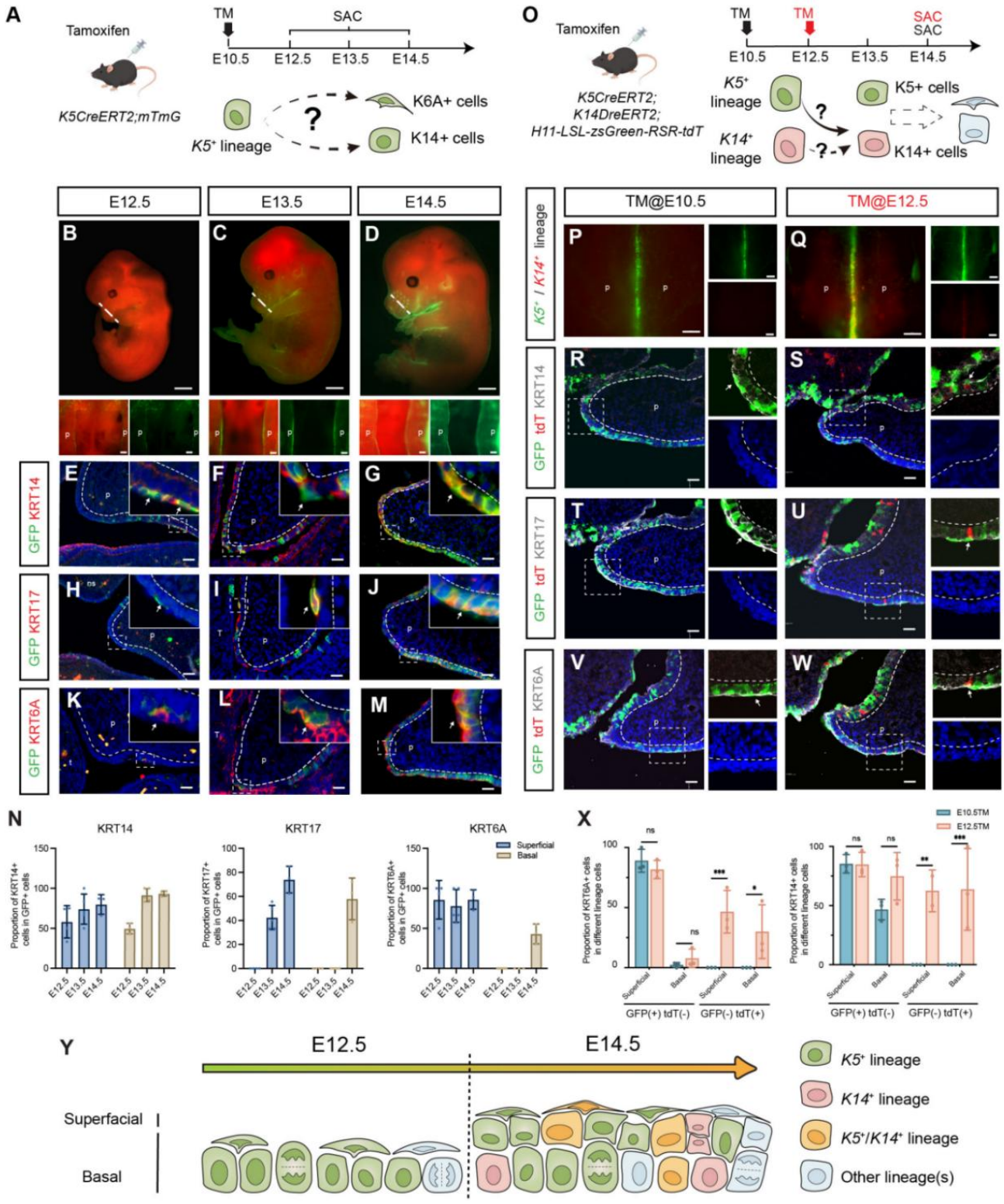


Figure 2. Lineage tracing distinguished the fate of K5⁺Krt14⁻ cells which contributed to periderm.

(A) Schematic of *K5*$^{+}$ cell lineage tracing: Tamoxifen (TM) was administered to pregnant *K5CreERT2;mTmG* mice at E10.5, with embryos harvested at E12.5, E13.5, E14.5. (B-D) Gross phenotypes of *K5CreERT2;mTmG* mice (morphology and mandible-removed profile views) at E12.5 (B), E13.5 (C), and E14.5 (D). White dashed lines indicate the region selected for subsequent immunofluorescence staining. Scale bars, 1mm. (E-M) Palatal sagittal sections co-stained for GFP (green, *K5*$^{+}$ lineage) and keratin markers (red, KRT14, KRT17, KRT6A) at indicated stages. Scale bars, 20μm. (N) Quantification of GFP^{+} cells co-expressing differentiation markers (brown: basal cells; blue: superficial cells). This metric (double positive / total GFP+) represents the proportion of the *K5*$^{+}$ lineage that has differentiated. (O) Dual-recombinase lineage tracing schematic: TM was given to pregnant K5CreERT2;K14DreERT2;H11-LSL-GFP-RSR-tdT(*K5;K14;H11*) dams at E10.5/E12.5, embryos harvested at E14.5. (P-Q) Fluorescence morphology of E14.5 palates with mandible removed from *K5;K14;H11* embryos following TM induction at E10.5 (P) or E12.5 (Q). (R-W) Sections of the *K5;K14;H11* embryos at E14.5 injected at E10.5 (R, T, V) or E12.5 (S, U, W) were co-stained for GFP (K5-only lineage), tdT (Krt14-only lineage), and keratin markers (gray ,KRT14, KRT17, KRT6A). Scale bars, 20μm. (X) Quantification of K5^{+}K14^{-} (GFP^{+}RFP^{-}) contribution to K14^{+} basal/KRT6A^{+} suprabasal cells (green: basal; pink: suprabasal). (Y) The schematic diagrams of the fate of K5^{+}K14^{-} cells. By E12.5, the *K5*$^{+}$ lineage that continues to fuel the periderm earlier, and a newly established, committed *K14*$^{+}$ lineage that expands the basal layer and contributes to periderm later.

All experiments: N⩾3 mice from ⩾3 independent experiments; data presented as mean ± SEM; statistical significance (two-way ANOVA): ns=not significant, **p<0.01, ***p<0.001, ****p<0.0001. Nuclei counterstained with DAPI (blue); white arrows indicate periderm cells. ns, nasal septum; p, palatal shelf; t, tongue.

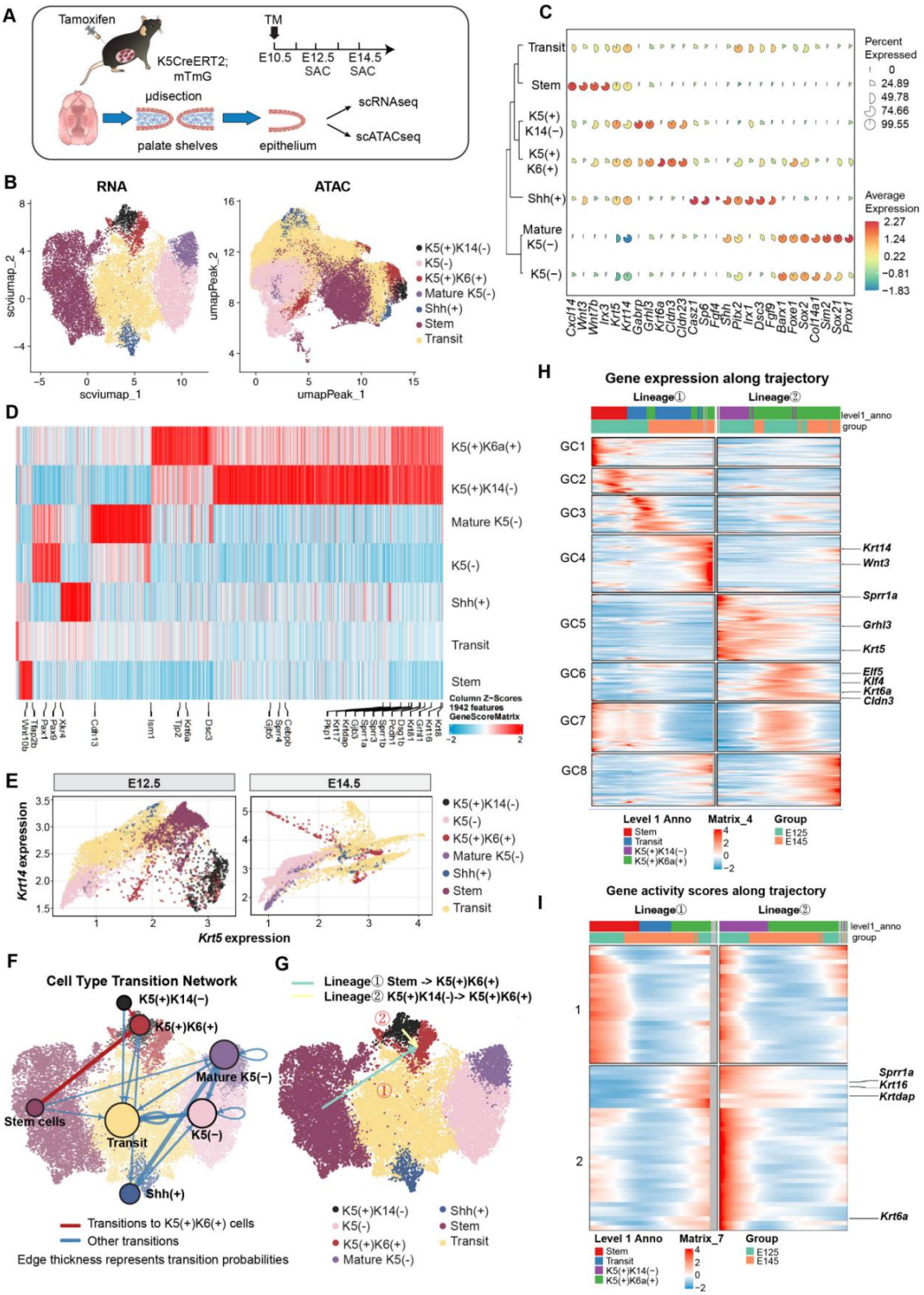
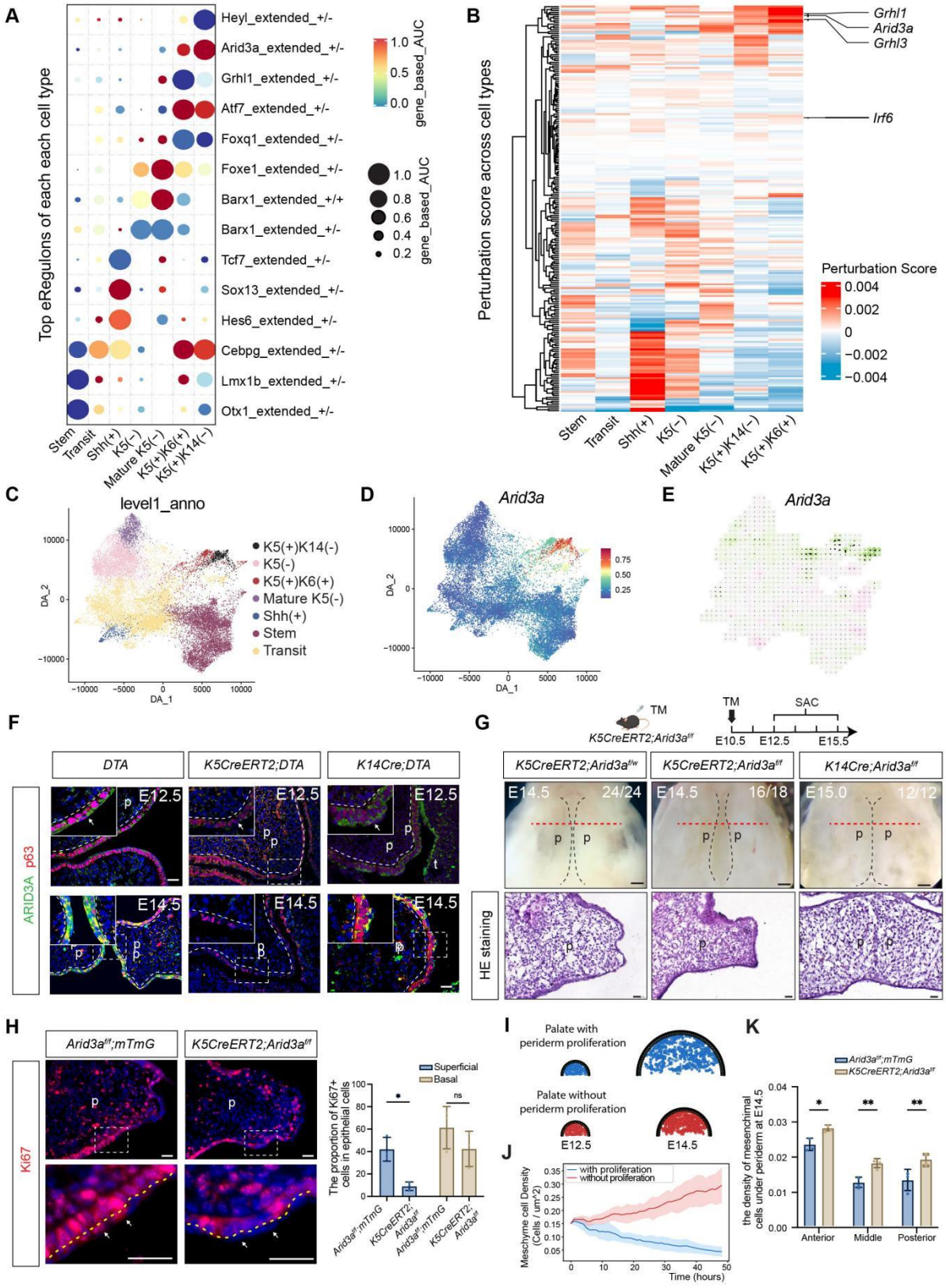


Figure 3. scRNA-seq and scATAC-seq of the periderm revealed the cellular identity of K5^{+};K14^{-} cells.

(A) Schematic overview of the single cell multi-omics experimental design. (B) Uniform Manifold Approximation and Projection (UMAP) visualization of the integrated scRNA-seq and scATAC-seq datasets. Cells are colored by their assigned cluster. (C) Dot plot showing the average expression level (color intensity) and the percentage of expressing cells (dot shape) of marker genes for each cell cluster. (D)

Heatmap displaying the accessibility of differentially accessible chromatin regions across each cell cluster, derived from scATAC-seq. (E) Scatter plot showing the expression patterns of *Krt5* and *Krt14* gene expression pattern. The expression of these two genes effectively distinguishes cell lineages and exhibits distinct patterns at different developmental timepoints. (F) RNA velocity analysis using scVelo, projecting cellular transitions and differentiation trajectories onto the UMAP embedding. (G) Trajectory inference using Moscot, showing predicted cell-state transitions and developmental flows. (H) Heatmap of dynamic gene expression along the reconstructed pseudotemporal trajectory of the two $K6^+$ Periderm differentiation lineages. (I) Heatmap showing the corresponding dynamic changes in gene activity scores along the $K6^+$ Periderm lineages, as inferred from scATAC-seq peak accessibility


A
Top eRegulons of each each cell type
Heyl_extended_+/-
Arid3a_extended_+/-
Grhl1_extended_+/-
Atf7_extended_+/-
Foxq1_extended_+/-
Foxe1_extended_+/-
Barx1_extended_+/+
Barx1_extended_+/-
Tcf7_extended_+/-
Sox13_extended_+/-
Hes6_extended_+/-
Cebpg_extended_+/-
Lmx1b_extended_+/-
Otx1_extended_+/-
Stem
Transit
Shh(+)
K5(−)
Mature K5(−)
K5(+)K6(+)
K5(+)K14(−)
gene_based_AUC
B
Perturbation score across cell types
Grhl1
Arid3a
Grhl3
Irf6
Perturbation Score
C
level1_anno
K5(+)K14(-)
K5(-)
K5(+)K6(+)
Mature K5(-)
Shh(+)
Stem
Transit
DA_1
DA_2
D
Arid3a
E
Arid3a
F
DTA
K5CreERT2;DTA
K14Cre;DTA
ARID3A p63
E12.5
E14.5
G
TM
K5CreERT2;Arid3a f/f
SAC
E10.5
E12.5
E15.5
K5CreERT2;Arid3a f/w
K5CreERT2;Arid3a f/f
K14Cre;Arid3a f/f
E14.5 24/24
E14.5 16/18
E15.0 12/12
HE staining
H
Arid3a f/f;mTmG
K5CreERT2;Arid3a f/f
Ki67
Superficial
Basal
The proportion of Ki67+ cells in epithelial cells
I
Palate with periderm proliferation
Palate without periderm proliferation
E12.5
E14.5
J
with proliferation
without proliferation
Meschyme cell Density (Cells / um^2)
Time (hours)
K
the density of mesenchimal cells under periderm at E14.5
Anterior
Middle
Posterior

1093

1094

Figure 4. *Arid3a* regulates cell proliferation in the periderm progenitors $K5^{+};K14^{-}$ cells.

Dot heatmap showing the top three eRegulons of each palatal cell type. Explain Regeion_based_AUC, explain gene_based_AUC. (B) Heatmap showing perturbation score, quantitative measurement to evaluate the effects of the top TF in silico gene perturbation on different palatal epithelial populations. Warm color indicates the promotion effect of the TF in related cell population, and cold color indicates the inhibitory effect of TF in the related cell population. (C) Force-directed graph embeddings of the integrated scRNA-seq from E12.5 and E14.5 palate epithelium. (D) Normalized expression level and distribution of *Arid3a* in palatal epithelium. (E) In-silico perturbation of *Arid3a* in palatal epithelium profile. Arrows in the plots indicate the reverse differentiation potential caused by loss of *Arid3a*. Green indicate negative effect, pink indicate positive effect on differentiation, grey indicate neutral effect. (F) Dual immunostaining of ARID3A(green) with p63(red) at E12.5(Top) and E14.5(Bottom) DTA, *K5CreERT2;DTA* and *K14Cre;DTA* mice middle palates. (G) Schematic of conditional knockout *Arid3a* strategy and sample collection timeline. TM was administered to pregnant *K5CreERT2;Arid3a$^{f/f}$;mTmG* mices at E10.5. Embryos were harvested at E12.5 to E15.5 .(H) Proliferation ability of *K5CreERT2;Arid3a$^{f/w}$* and *K5CreERT2;Arid3a$^{f/f}$* embryos at E14.5 was detected by staining with Ki67(red). Sections were showed at low (top) and high (bottom) magnification. Quantification of the quantity of $Ki67^{+}$ positive cells in superficial and basal epithelial cells. (I)Stochastic agent-based mathematical modeling simulating palatal growth dynamics (E12.5 – E14.5). The model couples periderm proliferation with mesenchymal expansion. (Top) In the "Palate with periderm proliferation" model (simulating the Control condition), periderm cells undergo self-renewal (black dots) while differentiating from basal cells (green squares), allowing the boundary to expand coordinately with the proliferating mesenchyme (blue circles), resulting in normal palatal outgrowth. (Bottom) In the "Palate without periderm proliferation" model (simulating the *Arid3a* cKO condition), periderm cells lose self-renewal capacity, resulting in insufficient boundary expansion and mesenchymal crowding. (J) Temporal dynamics of simulated mesenchymal cell density (cells/μm²) over the 48-hour simulation period (corresponding to E12.5–E14.5). The red curve(*Arid3a* cKO model) shows a progressive increase in cell density due to restricted boundary expansion, while the blue curve (Control model) shows stabilization. Solid lines indicate the mean, and shaded areas represent the standard deviation (SD) from N=10 independent simulation runs. (K) Quantification of mesenchymal cell density(cells/μm²) in Arid3af/f;mTmG and *K5CreERT2;Arid3a$^{f/f}$;mTmG* palates within 50 μm × 50 μm sub-epithelial regions. Bar graph showing mesenchymal cell density at E14.5. All experiments: N⩾3 mice from ⩾3 independent experiments; data presented as mean±SEM; statistical significance (two-way ANOVA): ns=not significant, **p<0.01,***p<0.001, ****p<0.0001. Nuclei counterstained with DAPI (blue); white arrows indicate periderm cells. ns, nasal septum; p, palatal shelf; t, tongue.

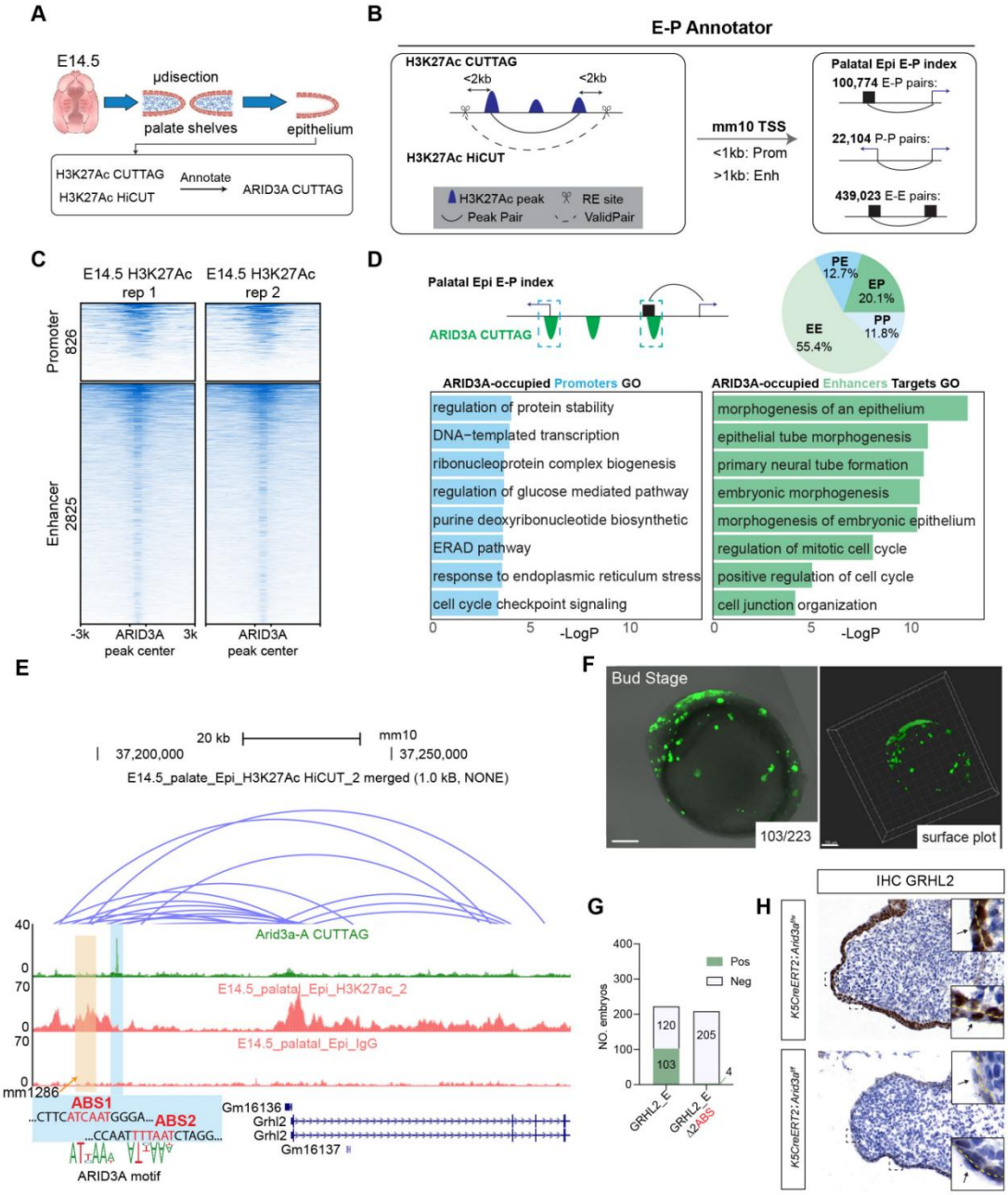


Figure 5. ARID3A Binds a Distal Enhancer to Directly Regulate *Grhl2* in the Periderm

(A) Schematic of the experimental workflow for genomic profiling of palate epithelium isolated from E14.5 mouse embryos. (B) Diagram illustrating the E-P (Enhancer-Promoter) Annotator pipeline. H3K27Ac CUT&TAG peaks are classified as promoters (Prom) or enhancers (Enh) based on proximity to a transcription start site (TSS). H3K27Ac HiCUT data identifies chromatin loops, which are filtered to define high-confidence E-P, E-E (Enhancer-Enhancer), and P-P (Promoter-Promoter) pairs. (C) Heatmap showing ARID3A and H3K27Ac CUT&TAG signal intensity centered on ARID3A binding peaks. Peaks are classified as promoter-associated (n=826) or enhancer-associated (n=2,825). (D) Pie chart showing the distribution of ARID3A-mediated chromatin interactions as

defined by the E-P Annotator. Below, Gene Ontology (GO) analysis reveals distinct functions for genes targeted by ARID3A-occupied promoters (left, general cellular processes) versus ARID3A-occupied enhancers (right, developmental morphogenesis and cell cycle). (E) Genome browser view of the *Grhl2* locus. Tracks show HiCUT-defined chromatin interactions (arched lines), ARID3A CUT&TAG binding, and H3K27Ac CUT&TAG signals from E14.5 palate epithelium. A prominent ARID3A-bound enhancer containing two ARID3A binding sites (ABS1, ABS2) interacts with the *Grhl2* promoter. Notably, this element is distinct from and located proximal to the previously reported *Grhl2* craniofacial enhancer, mm1286. The ARID3A binding motif is shown below. (F) Zebrafish transgenesis assay showing the activity of the putative *Grhl2* enhancer. The element drives robust GFP expression in the superficial periderm at the bud stage (left, brightfield; right, fluorescence surface plot). Quantification of embryos with enhancer activity is shown (103/223). Scale bar, 50 μm. (G) Quantification of the zebrafish assay using the same enhancer element but with the two ARID3A binding sites mutated (A2ABS), showing a near-complete loss of enhancer activity. (H) Immunohistochemical staining for GRHL2 (brown) in the palatal epithelium of control (*K5CreERT2;Arid3a*$^{f/w}$*;mTmG*) and *Arid3a* conditional knockout (*K5CreERT2;Arid3a*$^{f/f}$*;mTmG*) mice at E14.5. Loss of ARID3A results in a marked reduction of GRHL2 protein in the superficial periderm layer. Black arrows pointed to the periderm cells. N=6 mice were used. Scale bar, 20 μm.

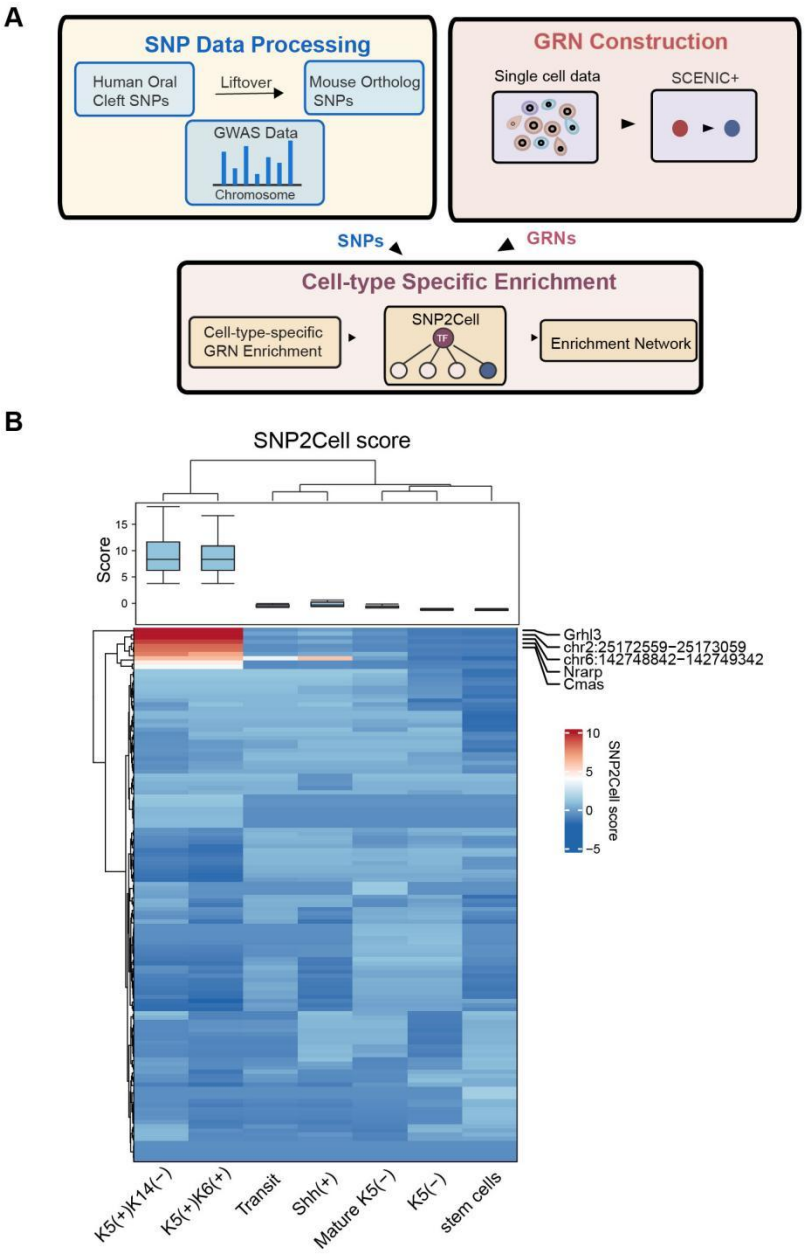


Figure 6. Orofacial Cleft GWAS Variants are Specifically Enriched in the K5^{+};K14^{-} Periderm Progenitor Lineage

(A) Schematic of the SNP2Cell workflow used to link human orofacial cleft (OFC) risk variants to specific cell types. Human GWAS SNPs were lifted over to the mouse genome. Concurrently, a gene regulatory network (GRN) was constructed from the palatal epithelium single-cell data using SCENIC+. SNP2Cell was then used to integrate these two data streams to calculate a cell-type-

specific enrichment score for the OFC-associated SNPs. (B) Heatmap displaying the SNP2Cell enrichment scores for OFC-associated SNPs across the seven identified palatal epithelial cell states. A total of 3,439 human OFC GWAS SNPs were lifted over to 1,123 orthologous mouse SNPs, of which 221 overlapped with epithelial chromatin accessibility peaks. The analysis reveals a striking and significant enrichment of high SNP2Cell scores specifically within the $K5^{+}K14^{-}$ progenitor and $K5^{+}K6^{+}$ periderm populations. The boxplot at the top summarizes the score distribution, highlighting the specific enrichment in these two cell types compared to all others. Loci for top-scoring SNPs, including the known clefting gene *Grhl3*, are indicated.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MovieS1.mp4
- TableS3scATACpeakFigure3D.csv
- TableS8OFCrelatedGWASSNPhg38.txt
- TableS5ARID3ACUTTAGPeakE145PalatalEpi.xlsx
- TableS9OFCrelatedGWASSNPliftovermm10.txt
- TableS11Fig6BrelatedSNP2Cell.xlsx
- TableS1scRNAmarkers.csv
- TableS10SNPsoverlappedpalatalepipeaks.xlsx
- Supplementaryfigures.pdf
- TableS6H3K27AcCUTTAGConsensusPeakE14.5PalatalEpi.xlsx
- TableS2scATACpeak.csv
- TableS4Fig3HGCgene.csv
- TableS7E145EH3K27acHiCUTunion.txt