







Single-cell transcriptomics identifies divergent developmental lineage trajectories during human pituitary development

Shu Zhang ^{1,2,6}, Yueli Cui^{1,2,6}, Xinyi Ma ^{1,3,6}, Jun Yong^{1,3,4}, Liying Yan ^{1,3}, Ming Yang^{1,3,4,5}, Jie Ren^{1,2}, Fuchou Tang ^{1,2,5}, Lu Wen ^{1,2}✉ & Jie Qiao ^{1,2,3,4}✉

The anterior pituitary gland plays a central role in regulating various physiological processes, including body growth, reproduction, metabolism and stress response. Here, we perform single-cell RNA-sequencing (scRNA-seq) of 4113 individual cells from human fetal pituitaries. We characterize divergent developmental trajectories with distinct transitional intermediate states in five hormone-producing cell lineages. Corticotropes exhibit an early intermediate state prior to full differentiation. Three cell types of the PIT-1 lineage (somatotropes, lactotropes and thyrotropes) segregate from a common progenitor coexpressing lineage-specific transcription factors of different sublineages. Gonadotropes experience two multistep developmental trajectories. Furthermore, we identify a fetal gonadotrope cell subtype expressing the primate-specific hormone chorionic gonadotropin. We also characterize the cellular heterogeneity of pituitary stem cells and identify a hybrid epithelial/mesenchymal state and an early-to-late state transition. Here, our results provide insights into the transcriptional landscape of human pituitary development, defining distinct cell substates and subtypes and illustrating transcription factor dynamics during cell fate commitment.

¹ Department of Obstetrics and Gynecology, Beijing Advanced Innovation Center for Genomics, School of Life Sciences, Third Hospital, Peking University, Beijing 100871, China. ² Biomedical Pioneering Innovation Center, School of Life Sciences, Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Beijing 100871, China. ³ Key Laboratory of Assisted Reproduction, Ministry of Education, Beijing 100191, China. ⁴ Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproductive Technology, Beijing 100191, China. ⁵ Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China. ⁶ These authors contributed equally; Shu Zhang, Yueli Cui, Xinyi Ma. ✉email: wenlu@pku.edu.cn; jie.qiao@263.net

The pituitary is the central gland of the endocrine system for regulating multiple physiological processes including the stress response, body growth, reproduction and metabolism. Much of the control comes from five cell types of the anterior pituitary gland including corticotropes that secrete adrenocorticotropic hormone (ACTH), somatotropes that produce growth hormone (GH), lactotropes that release prolactin (PRL), thyrotropes that produce thyroid-stimulating hormone (TSH) and gonadotropes that produce luteinizing hormone (LH) and follicle-stimulating hormone (FSH). The development of the anterior pituitary provides an excellent model system to elucidate mechanism of organogenesis¹. The five hormone producing cell types develop in a stereotypical order from Rathke's pouch, an epithelial invagination of the oral ectoderm. Previous studies have identified various signaling pathways and transcription factors (TFs) participating in pituitary development. These include the SHH, BMP and FGF pathways for initiation of Rathke's pouch; *PITX1*, *LHX3*, *HESX1* and *PROP1* for early phase patterning; *POU1F1* (also known as PIT-1) for differentiation of somatotropes, lactotropes and thyrotropes; *TBX19* (also known as TPIT) for differentiation of corticotropes; and *NR5A1* (also known as SF-1) for differentiation of gonadotropes^{1–4}. More recently, studies have also raised interest in SOX2-positive pituitary stem cells by showing their capability of self-renewal and differentiation into all five types of endocrine cells^{5,6}.

However, our understanding of pituitary development, particularly human pituitary development, is not well defined. Genomic studies have been hampered by intermingling of different cell types in this relatively small organ^{7,8}. Recent rapid progression in single-cell RNA sequencing (scRNA-seq) technologies provides an opportunity to comprehensively understand the regulatory network and cellular heterogeneity of pituitary development. Several recent studies have reported scRNA-seq in the adult mouse and rat pituitary^{9–11}. Here, we apply scRNA-seq to human fetal pituitaries for mapping the transcriptional landscape of human pituitary development at single-cell resolution. Our results provide insights into transcriptional dynamics of progressive lineage specification of human pituitary endocrine cells, and elucidate characteristics of the pituitary stem cells, progenitor and precursor cells, and different endocrine cell types and subtypes.

Result

ScRNA-seq analysis of human pituitary development. We obtained pituitaries from 21 human fetuses from 7 to 25 weeks postfertilization (11 females and 10 males) and performed a modified STRT-seq method on a total of 5181 cells, with 4113 high-quality cells being retained after filtration (Fig. 1a and Supplementary Fig. 1a). An average of 4506 genes and 86,497 transcripts (counted as unique molecular identifiers, UMIs) were detected in each cell (Supplementary Fig. 1c). The samples were detected with similar gene numbers and *GAPDH* expression across batches (Supplementary Fig. 1b, c). The morphology of the pituitary was verified (Supplementary Fig. 1g).

We used Seurat to identify cell clusters and Uniform Manifold Approximation and Projection (UMAP) for visualization (Fig. 1b and Supplementary Fig. 1d)¹². A total of 14 cell clusters identified with known marker genes (Fig. 1b, Supplementary Fig. 1d). We identified nine clusters of anterior pituitary endocrine cells including the stem cells (Stem), cycling cells (CC), corticotropes, progenitors of the PIT-1 lineage (Pro.PIT1), somatotropes, lactotropes, thyrotropes, precursors of gonadotropes (Pre.Gonado) and gonadotropes, comprising 2,388 cells (Fig. 1b and Supplementary Fig. 1a). *PITX1* and *PITX2* were expressed in all nine clusters, and *SOX2* and *PROP1* were specifically expressed in

the stem cells (Fig. 1c and Supplementary Fig. 1f). Lineage-specific TFs, such as *POU1F1*, *TBX19* and *NR5A1*, and hormone genes, including *POMC*, *GH1*, *GH2*, *TSHB*, *PRL*, *FSHB* and *LHB*, were expressed in special clusters of the endocrine cells. Mesenchymal cells, as marked by *PDGFRA*¹³, were the second most abundant cell type after endocrine cells, comprising 1,005 cells. A cluster of 90 cells was identified as posterior pituitary pituicyte cells (P) as marked by *OTX2*, *LHX2*, *RAX* and *COL25A1*^{11,14} (Supplementary Fig. 1f). Other clusters included endothelial cells (*PECAMI*⁺), immune cells (*IMM*, *PTPRC*⁺) and red blood cells (RBC, *HBQ1*⁺). Each cell cluster was composed of multiple fetal samples, and the samples of similar stages, or different sexes, were largely mixedly distributed (Fig. 1b, Supplementary Fig. 1d, e and 2a).

We analyzed the timing of endocrine cell differentiation. The results showed that the corticotropes appeared first at 7 weeks, the earliest time point we analyzed; this was immediately followed by the somatotropes and the gonadotropes at 8 weeks, and the thyrotropes and lactotropes appeared later at approximately 10 weeks and 16 weeks, respectively (Fig. 1d). The timing of endocrine cell differentiation was validated by immunofluorescence staining for the hormone genes, which was consistent with previous studies (Supplementary Fig. 2b)¹⁵.

We analyzed a number of TFs that are mutated in human pituitary genetic diseases¹⁶, showing that these TFs were expressed in a cell type-specific manner (Supplementary Fig. 1h). The SCENIC analysis identified activation of known TFs including *SOX2*, *TBX19*, *POU1F1* and *NR5A1* (Fig. 1e and Supplementary Fig. 2c)¹⁷. Taken together, these results indicated that our data provided comprehensive and precise information on human pituitary development.

Characterizing pituitary stem cells. Human pituitary stem cells have not been comprehensively characterized. We identified stem cell-specific genes including *SOX2*, *PROP1*, *LHX3*, *HES1*, *ZFP36L1*, *ANXA1*, *NFIB*, *ZNF521* and *NR2F2* (Fig. 2a). Gene set enrichment analysis (GSEA) showed that the TGF- β , Notch, Wnt and Hedgehog signaling pathways were enriched in the stem cells, and the “tight junction”, “cell cycle” and “ECM receptor interaction” pathways were also highly enriched (Fig. 2b). We identified potential ligand-receptor genes between stem cells and mesenchymal cells, which enriched Gene Ontology (GO) terms such as “ECM receptor interaction” and the Notch, Wnt, BMP and Eph signaling pathways, suggesting that the mesenchymal cells may involve in regulation of the stem cells (Supplementary Fig. 3a, b and Supplementary Data 1). Co-immunofluorescence staining for *SOX2* and Collagen III (*COL3*) suggested that mesenchymal cells encompass stem cells in human fetal pituitaries (Supplementary Fig. 3c).

Notably, reclustering the stem cells by Seurat revealed three subpopulations (Stem1, Stem2 and Stem3, Supplement Fig. 3d). Interestingly, Stem1 cells were found to be mainly derived from the early-stage pituitaries (7 to 10 weeks), while Stem3 cells were mainly derived from the late-stage pituitaries (19 to 25 weeks), indicating that these subpopulations represented time-dependent cell state shifts in the stem cells (Fig. 2c). The clustering result remained similar after regressing out the cell cycle genes (Supplement Fig. 3d).

Examination of differentially expressed genes (DEGs) among the three subpopulations identified 114 and 165 genes that were highly expressed in Stem1 and Stem3, respectively (Supplementary Fig. 3e). Stem1 cells were enriched in genes for “stem cell proliferation” (e.g. *HMGA2*), while the Stem3 cells were strongly enriched in genes for “negative regulation of cell proliferation” (e.g., *CDKN1A*, Fig. 2d and Supplementary Fig. 3e). Reclustering

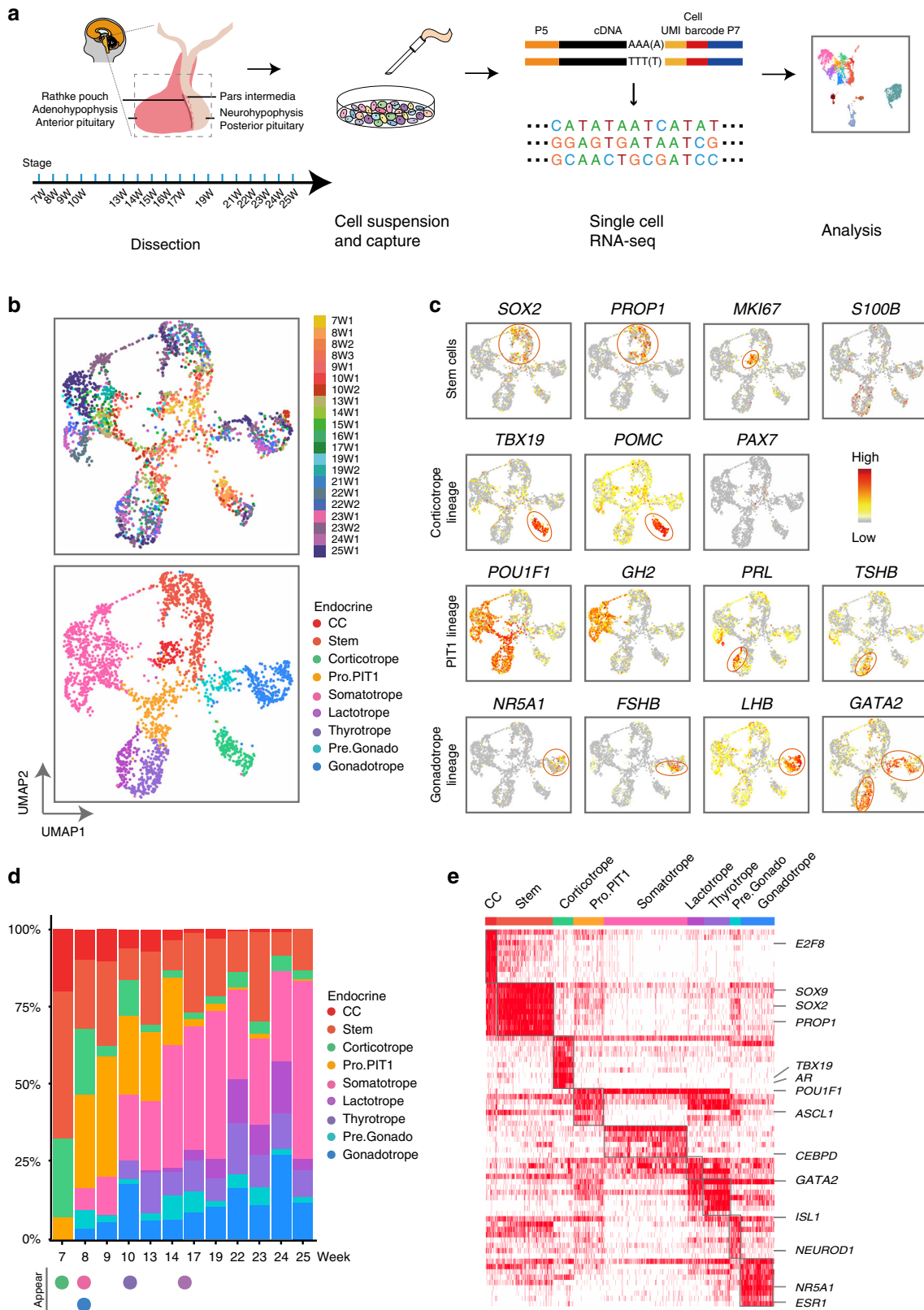
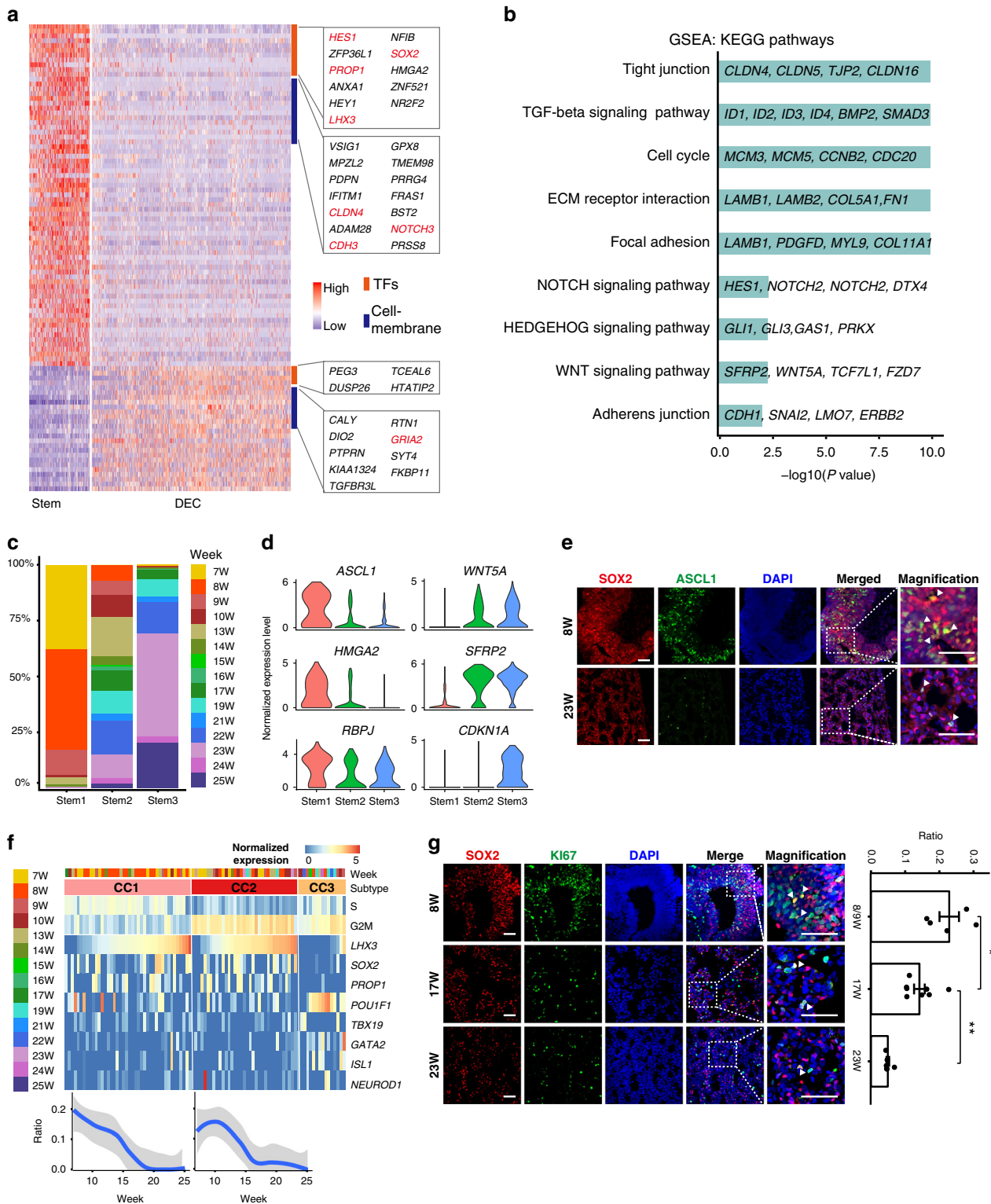


Fig. 1 Diversity of cell types in the human fetal pituitary. **a** Experimental flowchart for the scRNA-seq analysis of the human fetal pituitary. **b** UMAP plots showing the clusters of the cell cycle cells (CC), stem cells, the progenitor cells of PIT1 lineage (Pro.PIT1) or precursor cells of gonadotrope (Pre.Gonado) and the terminal endocrine cells (lower), and distribution of the fetal samples (upper). Dots: single cells. **c** Scatterplots showing expression of known markers, including TFs and hormone genes, projecting on the UMAP plot (b). Gray to red indicates no expression to high expression levels. **d** Bar plots showing the proportions of each cell type in each stage. Solid circles at the bottom indicate the earliest stages when a hormone producing cell type appears. **e** Heatmap showing the activated TFs predicted by SCENIC. For each cell type, the top 10 $-\log(P \text{ value})$ specific TFs being activated are shown, which ranked by number of cells. Columns are individual cells and rows are individual genes. White: not activated; red: activated. The expression levels of these TFs are shown in Supplementary Fig. 2c.



of the CC cluster showed that most cycling cells were *SOX2*-positive stem cells, and a small portion were *POU1F1* or *TBX19*-positive cells, which were verified by immunofluorescence staining (Fig. 2f and Supplementary Fig. 3f). The proportion of cycling stem cells in the total stem cell population gradually decreased from the early to late stages (20% in the early stages, 10% in the middle stages, and less than 5% in the late stages; Fig. 2f), which was validated by coimmunostaining for *SOX2* and *Ki67* in the pituitaries of 8-, 17- and 23-week fetuses (Fig. 2g). These results

together suggest that Stem3 cells enter a quiescent or lowly proliferative state.

Interestingly, *ASCL1* was specifically expressed in Stem1 (Fig. 2d). This expression was consistent with the expression of *Ascl1* in the early stage of pituitary development in mouse and zebrafish^{18,19}. We performed immunofluorescence staining to show that a large portion (24%) of *SOX2*-positive stem cells were also positive for *ASCL1* at 10 weeks, while there were nearly no double-positive cells at 23 weeks (Fig. 2e). *ASCL1* is required for

Fig. 2 Molecular characteristics and heterogeneity of pituitary stem cells. **a** DEGs between the stem cells and the differentiated endocrine cells shown by a z-scored heatmap. DEC: differentiated endocrine cells, which include the progenitor or precursor cells. **b** Bar plots showing differentially enriched KEGG pathways between the stem cells and the DEC detected by GSEA. Representative genes in each pathway are shown. Nominal P values are determined by two-sided Kolmogorov-Smirnov Test and adjusted by FDR. **c** Bar plots showing the stages of three subtypes of the stem cells. The Stem1 are mainly derived from the early stages (7 to 10 weeks), and the Stem3 cells are mainly derived from the late stages (19 to 25 weeks). Color: weeks postfertilization. **d** Violin plots of representative DEGs between the Stem1 and Stem3 cells. **e** Immunofluorescence staining for SOX2 and ASCL1 in the 8- and 23-week human fetal pituitaries. Triangles indicate representative cells coexpressing both genes. Scale bar, 50 μm . **f** The CC cluster are composed of different cell subtypes. The upper panel shows the average expression levels of the S and G2/M phase genes, and representative cell type markers. CC1 and CC2 are SOX2-positive stem cells with CC1 and CC2 being the S and G2/M phase, respectively. CC3 are non-stem cells expressing *POU1F1*, *TBX19* or *GATA2*. The lower panel shows the proportions of the cycling to all stem cells as a function of weeks postfertilization. Data are presented as mean \pm SEM. **g** Immunofluorescence staining of SOX2 and Ki67 in the 8-, 17- and 23-week human fetal pituitaries (left). Triangles indicate representative costaining cells. The right panel shows the ratios of double-positive and SOX2-positive cells in each sample ($n \geq 2$). *p*-Value is determined by two-sided Wilcoxon test (**P* = 0.023 which is less than 0.05, ***P* = 0.0024 which is less than 0.01). Data are presented as mean \pm SEM. Scale bar, 50 μm .

differentiation of all anterior pituitary endocrine cell types in zebrafish, while in mice, mutation of *Ascl1* affects corticotropes and gonadotropes, and maybe also thyrotropes, at different regulatory steps^{19–21}. The downregulation of *ASCL1* expression in the late pituitary stem cells may play a role in the state shift. The Wnt signaling modulator *SFRP2* and the non-canonical WNT gene *WNT5A* was prominently upregulated in Stem2 and Stem3 (Fig. 2d), which was consistent with the enrichment of the GO term “negative regulation of canonical Wnt signaling pathway” in Stem3 (Supplementary Fig. 3c).

The marker of the folliculostellate cells (FSCs) *S100B* was expressed in a small fraction (3.4%) of the stem cells of both the early and late stages (Fig. 1c). Comparison between the *S100B*-positive and *S100B*-negative stem cells revealed no DEGs relating to FSCs. The results suggested that FSCs were not present in the human fetal pituitary during 25 weeks postfertilization, which was consistent with previous immunostaining studies showing that S100-positive FSCs are basically not detected in the prenatal rat pituitary²².

In summary, we identified an embryonic state shift of the pituitary stem cells accompanied by negative regulation of cell proliferation and alternations in TFs and signaling pathways.

Pituitary stem cells in hybrid epithelial/mesenchymal state.

Previous studies have suggested that mouse pituitary stem cells undergo an epithelial-mesenchymal transition (EMT)-like process during differentiation^{23,24}. We explored EMT dynamics during differentiation of the human pituitary stem cells. Among the Kyoto Encyclopedia of Genes and Genomes (KEGG) terms enriched in the stem cells, “tight junction” is related to the epithelial state, and “ECM receptor signaling pathways” is related to the mesenchymal state (Fig. 2b). By principal component analysis (PCA), all endocrine cells were ordered in a gradient transition from the stem cells to the differentiated cells on the principal component 1 (PC1) axis, so we used the PC1 axis as a trajectory for analyzing the stem cell differentiation (Fig. 3a). Examination of the expression of epithelial and mesenchymal markers showed that both the epithelial markers, such as *CDH1*, and the mesenchymal markers, such as *VIM* and *CDH2*, were highly expressed in the stem cells but lowly expressed in the differentiated cells; *EPCAM* was highly expressed in both groups (Fig. 3b and Supplementary Fig. 4b). This pattern suggested that the stem cells existed in a state expressing both the epithelial and mesenchymal markers, fitting the term of a hybrid E/M state as previously identified during mouse organogenesis and tumor transition^{25,26}.

To clarify the correlation between the hybrid E/M state and stemness, we defined an epithelial score (E.score), a mesenchymal score (M.score) and a stemness score (S.score) (Supplementary Data 2). We found that all three scores decreased along the

differentiation axis (Fig. 3c and Supplementary Fig. 4c). This pattern was exemplified by the expression levels of known stemness markers (*SOX2*, *SOX4*, *SOX9*, *NOTCH2* and *HES1*), epithelial markers (*EPCAM*, *CDH1*, *KRT8*, *CLDN4* and *GRHL2*) and mesenchymal markers (*CDH2*, *VIM*, *COL1A1*, *COL1A2* and *SNAIL*) which decreased along the timeline (Fig. 3c). We did not observe significant EMT changes in the stem cell substates (Supplementary Fig. 4d). The stem cells also specifically expressed *CDH3*, which plays roles in both maintaining stemness and promoting the hybrid E/M state in development and cancers (Fig. 2a)²⁷.

The SCENIC analysis suggested that *CDH1*, *CDH2*, *CDH3* and *VIM* were potential targets of *SOX2* and *SOX4*; *VIM* and *CDH3* were potential targets of *PROPI* whose targeted genes enriched in the “epithelial to mesenchymal transition” term, being consistent with the previous chromatin immunoprecipitation sequencing (ChIP-seq) study²⁴ (Supplementary Fig. 4e, f).

In sum, the data suggested that the pituitary stem cells existed in a hybrid E/M state which is associated with their stemness characters.

Reconstructing developmental trajectories. Next, we reconstructed the developmental progression of five endocrine cell lineages and identified transient precursors by applying the RNA velocity²⁸ and Slingshot²⁹ for the pseudotime trajectory analysis (Fig. 4a). The Slingshot results revealed lineage-shared and specific TFs being downregulated or upregulated during the pseudotime developmental progression, with some TFs showing peak expression in intermediated stages (Fig. 4b, Supplementary Data 3). A group of 29 TFs were identified as downregulated genes being shared among all five lineages, including *PROPI*, *SOX2*, *LHX3*, *HES1*, *TCF7L1* and *TGIF1* (Fig. 4c, d). Mutations in *TCF7L1* and *TGIF1* have been reported in patients with hypopituitarism recently^{30,31}. *POU1F1*, *TBX19* and *NR5A1* were strongly upregulated in the PIT-1, corticotrope and gonadotrope lineages, respectively. *GATA2*, *FOXJ2* and *ISL1* were shared by the thyrotrope and gonadotrope lineages. *NEUROD1* showed peak expression at the intermediated states of both the gonadotrope and corticotrope lineages (Fig. 4c).

Corticotropes experience two substates. Next, we investigated development of each cell lineage. The corticotrope is the first hormone-producing cell type that appears and *TBX19* is crucial for its development³. The pseudotime and reclustering analysis identified two subclusters: Corticotrope1, which comprised cells mainly derived from the early stage (7 to 9 weeks), and Corticotrope2, which comprised cells mainly derived from the middle or late stages (10 to 25 weeks) (Fig. 5a). *TBX19* and *POMC* were highly expressed in both clusters, which was consistent with the

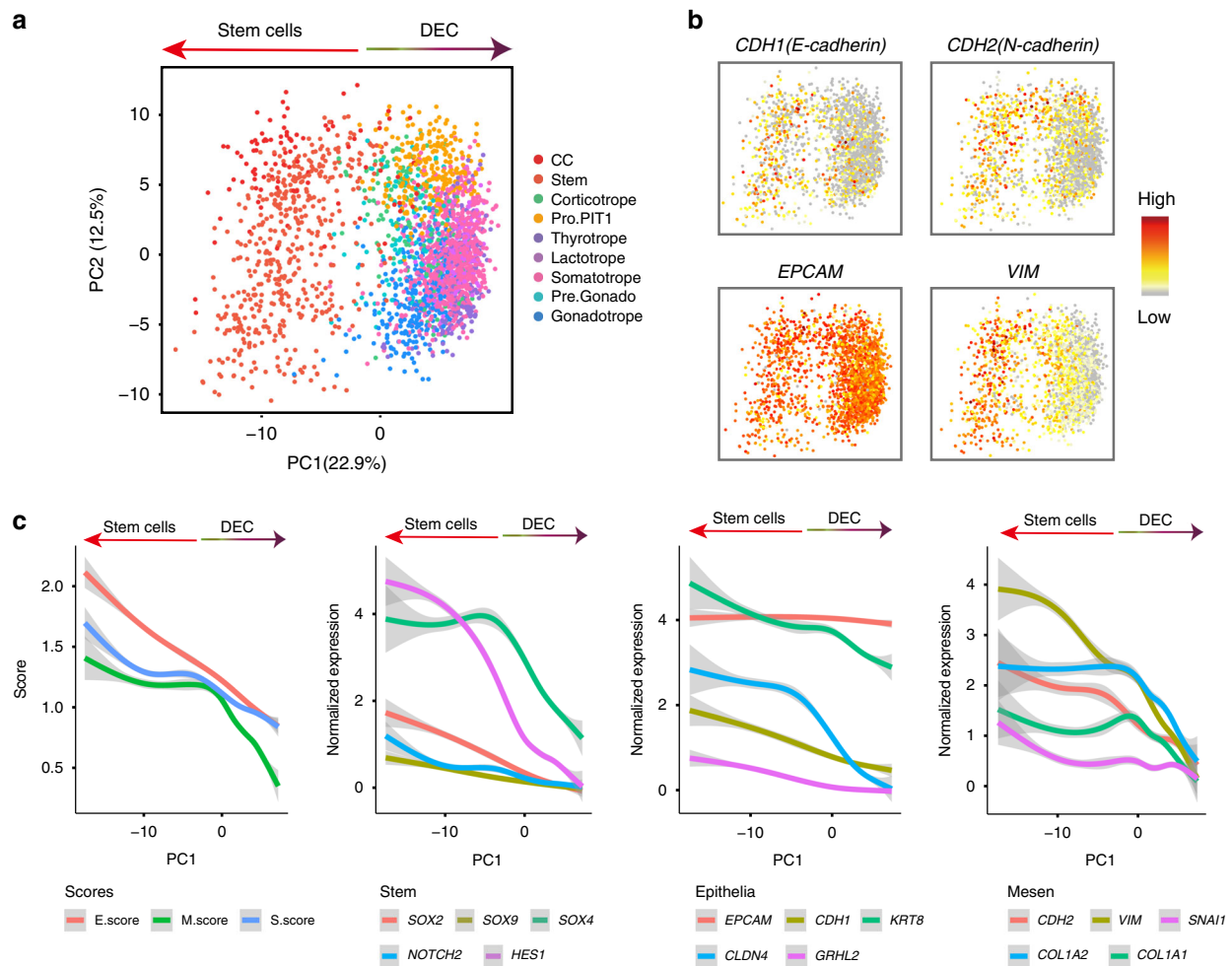


Fig. 3 Hybrid epithelial/mesenchymal state of pituitary stem cells. **a** PCA plot showing differentiation of endocrine cells along the PC1 axis. Dots: single cells; colors: cell types. **b** Scatterplots showing the expression of representative epithelial and mesenchymal markers projecting on the PCA plot. Gray to red indicate no to high expression levels. **c** The values of the epithelial score (E.score), the mesenchymal score (M.score), and the stemness score (S.score), as well as the expression levels of representative genes, decreased during differentiation. The colored lines represent loess-smoothed expression. Data are presented as mean \pm SEM.

previous immunofluorescence staining study showing that all *TBX19*-positive cells also express *POMC* (Fig. 5b, c)³². Androgen receptor (*AR*) was prominently expressed in both corticotrope clusters, which was predicted to be a potential TF regulating *POMC* by SCENIC, and the expression was verified by immunofluorescence (Figs. 1e, 5b and c). *AR* mutant mice have been shown to exhibit an increased expression of *POMC*³³.

We identified DEGs between Corticotrope1 and Corticotrope2. DEGs in Corticotrope1 enriched GO terms such as “regulation of mitotic cell cycle” (e.g., *HMG2*), suggesting that they could potentially proliferate (Fig. 5d, e, and Supplementary Data 4). Indeed, a few *TBX19*-positive cycling cells were captured (Fig. 2g). Interestingly, follistatin (*FST*), which is an antagonist of *ACTV1N*, was specifically expressed in Corticotrope1, suggesting that this intermediate cell type may also play some paracrine roles for regulating the gonadotropes. DEGs in Corticotrope2 enriched GO terms such as “response to corticosterone” (e.g., *NR4A1* and *NR4A2*), suggesting that these cells were more matured (Fig. 5d, e)³⁴.

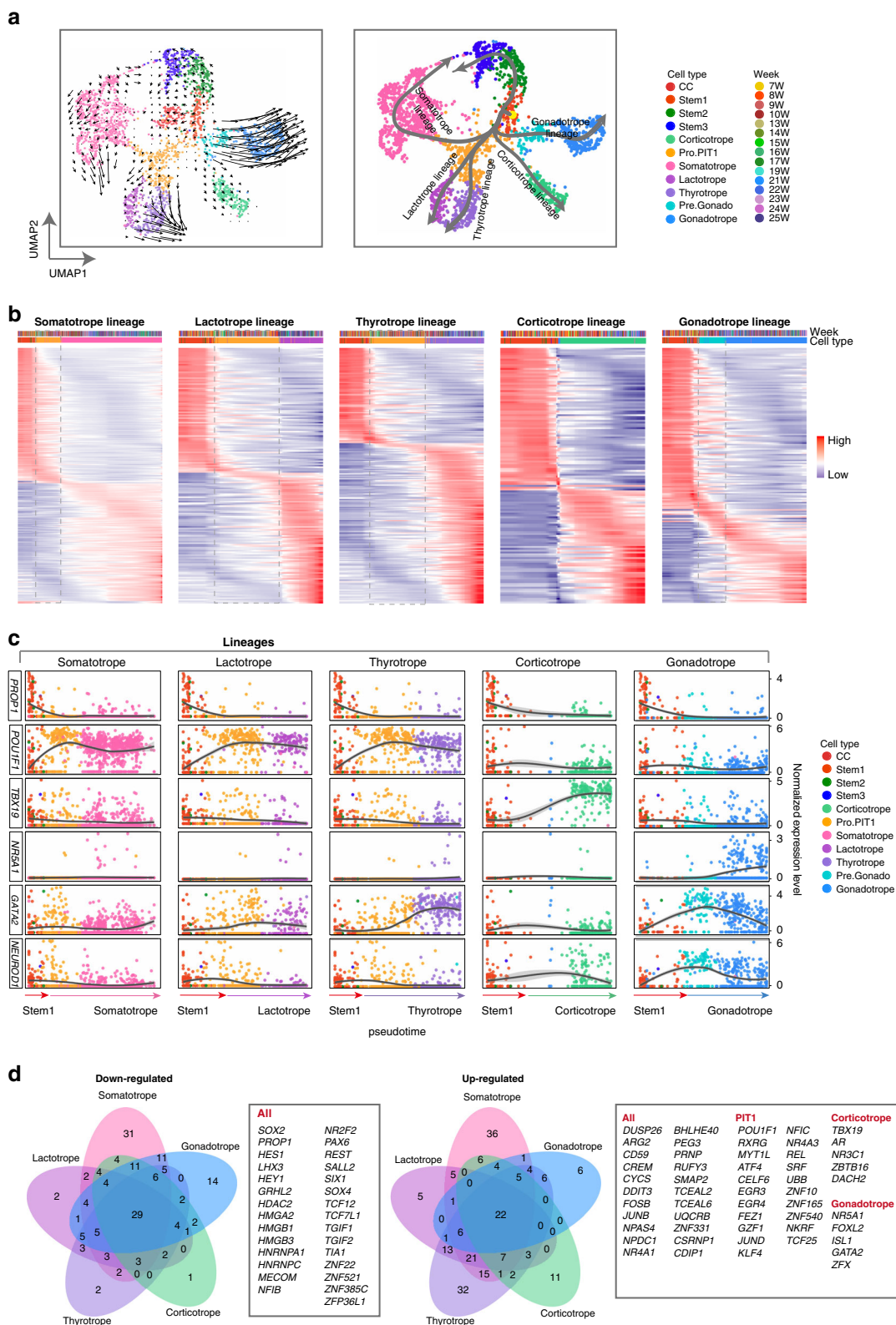
Another *POMC*-expressing lineage is the melanotrope in the intermediate lobe, which appears after 14 weeks of gestation but is scarce in the human pituitary¹⁵. The melanotrope coexpresses *TBX19* and *PAX7*, both of which are required for its differentiation³⁵. Only 4 cells, which were 7- or 8-week stem cells,

coexpressed *TBX19* and *PAX7* in our data (Fig. 1c). Comparison between these cells and other stem cells revealed only *PAX7* and three other DEGs that were not expressed in the mouse and rat melanotropes. These results suggested that we have not captured the melanotrope.

In summary, we defined corticotrope differentiation from an early intermediate state to a maturing state with upregulation of the expression of genes to establish the cortisol feedback.

PIT-1 lineages segregate from a common progenitor. The PIT-1 lineage is comprised of three endocrine cell types: the somatotrope, the lactotrope and the thyrotrope, all of which are governed by *POU1F1*^{2,36}. In addition to these three hormone producing cell types, the pseudotime trajectories analysis identified three intermediate progenitor or precursor cell populations: the Pro.PIT1_all cells as a common progenitor for all three hormone producing cell types (146 cells), the Pre.Thy as a precursor for the thyrotrope (74 cells, *GATA2*-positive) and the Pre.Som as a potential precursor for the somatotrope (19 cells, Fig. 6a, b).

To explore how three different lineages segregate from the common progenitor cell state, we first investigated expression dynamics of known lineage-enriched genes (Fig. 6c). Previous studies have shown that mutation of *Neurod4* in mice leads to almost completely lack of *GHRHR* expression and markedly



decrease in *GH* expression^{21,37}. Mutation of *Foxo1* in mice also results in delayed somatotrope differentiation³⁸. Notably, we found that *NEUROD4* was clearly activated in Pro.PIT1_all cells in comparison with the stem cells (logFC = 1.6, $P = 1.3E-18$, Fig. 6c). *NEUROD4* was further significantly upregulated in the Pre.Som cells and the differentiated somatotropes. *FOXO1* was

not expressed in the Pro.PIT1_all cells before upregulation in the Pre.Som cells and a portion of the somatotropes. Thus, the results demonstrated that *NEUROD4* played roles earlier than *FOXO1* in the somatotropes.

ZBTB20 has been recently demonstrated to be crucial for lactotrope specification in mice^{39,40}. This gene was expressed

Fig. 4 Pseudotime developmental trajectories of hormone-producing cells. **a** Pseudotime analyses of the endocrine cells shown in the UMAP plot. Left: the RNA velocity result with the arrows predicting directions of the pseudotime. Right: the Slingshot result with the lines indicating the trajectories of lineages and the arrows indicating manually added directions of the pseudotime. Dots: single cells; colors: cell types. Yellow circle in the right panel represents the start point of the trajectories which was set as the Stem1 subcluster. Since the corticotrope was already a separate cluster in the earliest sample (7 weeks), the corticotrope trajectory may start from stem cells earlier than Stem1. The CC cluster was omitted in the Slingshot pseudotime analysis as it contained several cell types and caused wrong trajectories. **b** Heatmap showing the relative expression of TFs displaying significant changes ($P \leq 1E-5$) along the pseudotime axis of each lineage. The progenitor or precursor cells of each lineage were enclosed in frames of dashed lines. Colors: loess-smoothed expression (red, high; blue, low). The columns represent the cells being ordered along the pseudotime axis with the cell type and fetal week informations being shown above. Rows represent genes being ordered by their peak expression along the pseudotime axis. P values are determined by one way ANOVA test. **c** Scatterplots showing the expression levels of representative TFs along the pseudotime axis. Dots: single cells; colors: cell types. **d** Venn diagram of downregulated and upregulated TFs along the pseudotime axis of each lineage. Representative TFs were listed in the right boxes.

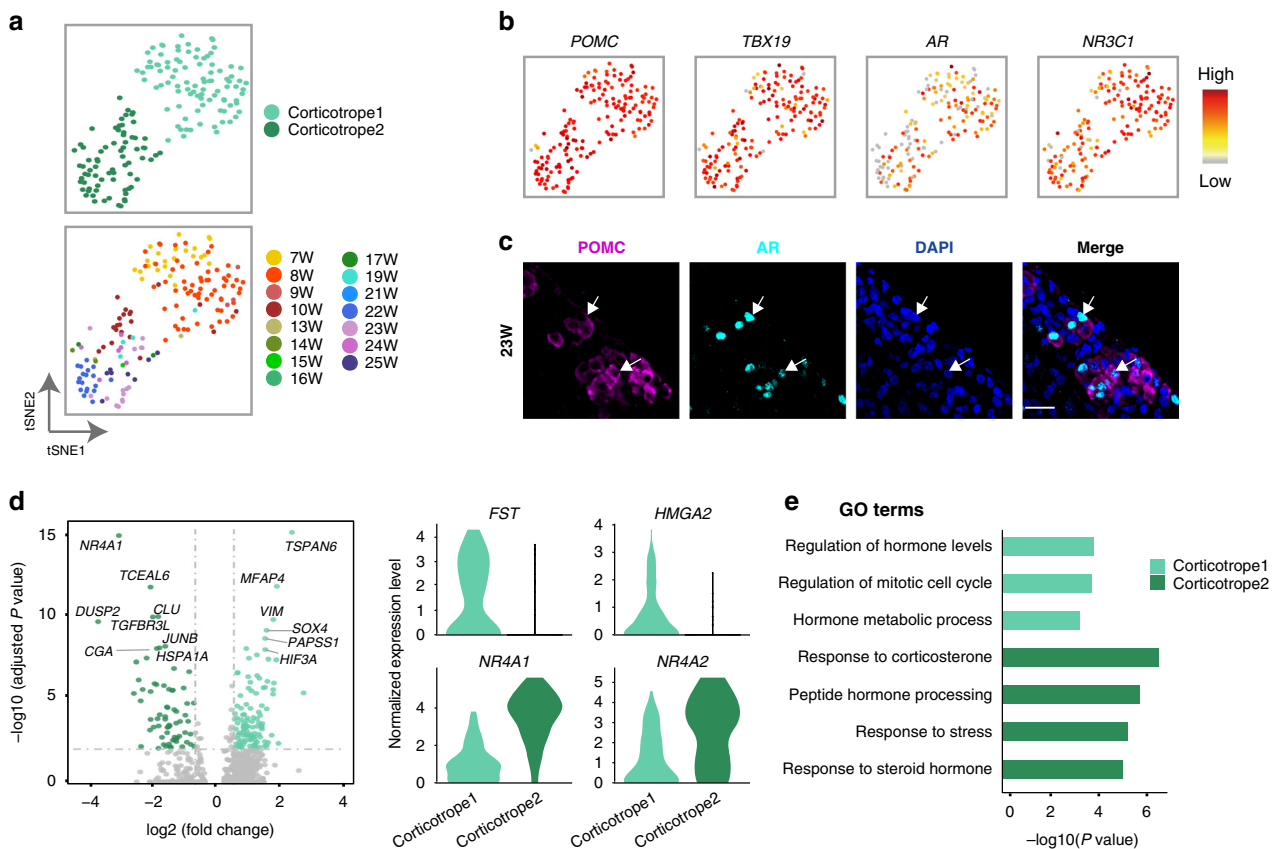


Fig. 5 Characteristics of two corticotrope substates. **a** t -SNE plots of corticotropes of different subtypes (upper) and fetal stages (lower). Dots: single cells; colors: cell types or weeks postfertilization. **b** Scatterplots showing expression of representative genes (*POMC*, *TBX19*, *AR* and *NR3C1*). Gray to red indicate no to high expression levels. **c** Immunofluorescence staining of *POMC* and *AR* (23 weeks). Scale bar, 20 μ m. **d** Volcano (left) and violin (right) plots showing DEGs between Corticotrope1 and Corticotrope2. Colored dots indicate significant DEGs in each subtype. P -values are determined by two-sided Wilcoxon test and adjusted by Bonferroni correction. **e** Bar plots showing GO terms being enriched in Corticotrope1 and Corticotrope2. P -values are determined by one-sided test.

in all pituitary cell types including the stem cells. Notably, its expression level was significantly upregulated in the Pro.PIT1_all cells in comparison with the stem cells ($\log_{2}FC = 0.6$, $P = 9.8E-08$), and the level was similar between the Pro.PIT1_all cells and the lactotropes ($P > 0.1$, Fig. 6c). Due to downregulation in other two lineages, the expression level of *ZBTB20* was slightly but significantly higher in the lactotropes comparing with the somatotropes and the thyrotropes (Lactotrope versus Somatotrope: $\log_{2}FC = 0.7$, $P = 9E-25$; Lactotrope versus Thyrotrope: $\log_{2}FC = 0.4$, $P = 1.1E-08$). This expression pattern was consistent with the essential role of *ZBTB20* in lactotrope specification.

GATA2, and possible *ASCL1*, have been implicated in thyrotrope development in mouse^{18,21,41}. In zebrafish, *sox4b*

has been shown to be required for thyrotrope development by activating *gata2a* expression⁴². We found that *ASCL1* and *SOX4* were prominently expressed in the Pro.PIT1_all cells. *GATA2* was not significantly upregulated in the Pro.PIT1_all cells ($P > 0.1$), but were prominently activated in the Pre.Thy and further upregulated in the terminal differentiated thyrotropes. Interestingly, *SOX11*, a member of SoxC family genes with *SOX4*, was significantly upregulated in the Pro.PIT1_all cells in comparison with the stem cells ($\log_{2}FC = 1.0$, $P = 4.2E-12$, Fig. 6d). The SCENIC analysis also suggested that both *SOX4* and *SOX11* bind to the regulatory region of *GATA2*. *ASCL1* displayed significantly higher expression levels in the thyrotropes than two other lineages (Thyrotrope versus

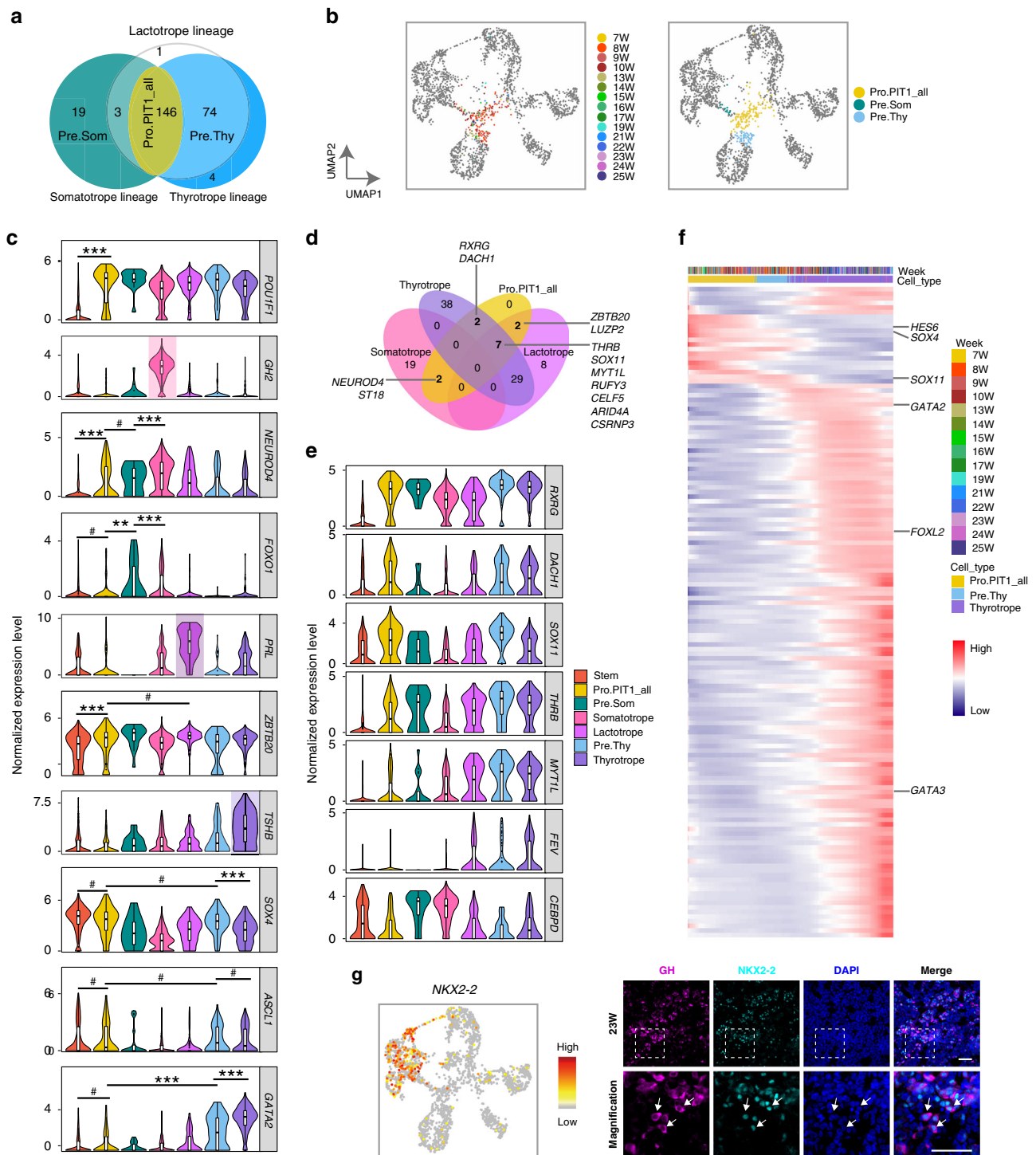


Fig. 6 Differentiation trajectories and TF dynamics of the PIT-1 Lineage. **a** Venn diagram of progenitor cells distributed in each PIT1 sublineage as shown in the pseudotime analysis. Pro.PIT1_all: the common progenitor cells of all three sublineages; Pre.Thy, the precursor cells of thyrotropes; Pre.Som: the potential precursor cells of somatotropes. **b** Distribution of the subtypes of the Pro.PIT1 on the UMAP plot. Dots: single cells; colors: weeks postfertilization (left) or cell types (right). **c** Violin and inside box plots showing the gene expression dynamics of known TFs for each PIT1 lineage. # $P > 0.1$, ** $P < 0.01$ *** $P < 0.001$. P -values are determined by two-sided Wilcoxon test and adjusted by Bonferroni correction. **d** Venn diagram showing the intersections among the differentially expressed TFs between any two of the three PIT1 lineages, and the 13 TFs that were significantly upregulated in the Pro.PIT1_all cells in comparison with the stem cells, for which the names of genes are shown. **e** Violin with inside box plots showing representative identified TFs for each PIT1 lineage. **f** Heatmap showing the relative expression levels of the significantly upregulated TFs of thyrotropes along the pseudotime axis. Colors: loess-smoothed expression (red, high; blue, low). The columns represent cells being ordered along the pseudotime axis, and cell type information is shown above the heatmap. Rows represent genes being ordered by their peak expression along the pseudotime axis. **g** The scatterplot (left) expression and immunofluorescence staining of NKX2-2 as a somatotrope-specific TF. Scale bar, 50µm. Arrows indicate representative cells co-expressing GH and NKX2-2.

Somatotrope: $\log_{FC} = 1.3$, $P = 6.6E-15$; Thyrotrope versus Lactotrope: $\log_{FC} = 0.7$, $P = 3.7E-09$).

Then, we investigated a comprehensive view of lineage-enriched genes and identified a total of 1,277 DEGs, including 107 TFs, between any two of three lineages ($\log_{FC} \geq 0.5$ and adjusted $P \leq 0.01$, Fig. 6d, Supplementary Fig. 5a, b and Supplementary Data 5). Thirteen TFs were significantly upregulated in the Pro.PIT1_all cells comparing with the stem cells. Among these TF, remarkably, *NEUROD4* and *ZBTB20* were identified as the primary lineage-enriched TFs for the somatotropes and the lactotropes, respectively (Fig. 6c, d). *RXRG* and *DACH1* were identified as two thyrotrope-enriched genes, both of which were prominently upregulated in the Pro.PIT1_all cells and kept the high expression levels in the thyrotropes while downregulated in the somatotropes and the lactotropes (Fig. 6c, d).

A total of 19 somatotrope-enriched TFs, including *FOXO1*, *CEBPD* and *NKX2-2*, were upregulated during the differentiating process of the somatotrope. *NKX2-2*, which has been shown to be essential for development of neuroendocrine, gastrointestinal tract and pancreas⁴³, was specifically upregulated in a portion of terminal somatotropes (Fig. 6g), and it is highly expressed in the human adult pituitary in the GTEx database. We validated the coexpression of *NKX2-2* with GH by immunofluorescence (Fig. 6g). A total of 8 lactotrope-enriched TFs (e.g. *SIX6*) and 38 thyrotrope-enriched (e.g., *GATA2*, *ISL1* and *FOXL2*) were upregulated during the differentiation of the lactotropes and the thyrotropes, respectively (Fig. 6d, e). The pseudotime analysis of thyrotrope-lineage differentiation showed that expression of *SOX4* and *SOX11* peaked before activation of *GATA2* in the Pre.Thy cells (Fig. 6f). *SOX4* and *SOX11* were significantly downregulated in the thyrotrope comparing with the Pre.Thy (*SOX4*: $\log_{FC} = 1.0$, $P = 2.7E-7$; *SOX11*: $\log_{FC} = 1.3$, $P = 4.9E-13$).

A surprising finding was that the lactotropes and the thyrotropes were close to each other relative to the somatotropes. This was displayed in UMAP, and fewer DEGs were identified between these two lineages comparing with the somatotropes, and the transcriptomes of the lactotropes and the thyrotropes were closely correlated (Figs. 4a, 6d and Supplementary Fig. 5a, b, c). In contrast, the somatotropes and lactotropes were more close to each other comparing with the thyrotropes in the mouse and rat adult pituitaries (Supplementary Fig. 5c). We speculated that this may be due to that both the lactotropes and the thyrotropes were in less matured states comparing with the somatotropes before 25 weeks. In the human fetus, GH begins to secrete before 10 weeks and peaks at approximate 22 weeks, while PRL begins to secrete at 25 weeks and peaks at birth⁴⁴. Supporting this, we found that *TRHR* and *ESR1* were lowly expressed in the fetal thyrotropes and lactotropes, respectively.

Furthermore, despite of many evidences suggesting the existence of mammosomatotropes coexpressing GH and PRL in the human and rodent pituitary^{45,46}, our analysis did not identify a distinct cell cluster that corresponds to a common precursor of the somatotropes and the lactotropes, even after removing potential batch effects of the somatotropes and the lactotropes; the result was similar to that of the rat scRNA-seq study⁹ (Supplementary Fig. 5d).

Taken together, our results characterized the transcriptome dynamics during specification of the PIT-1 lineages, in which a common progenitor coexpresses lineage-enriched TFs prior to activation of divergent TF networks.

Gonadotropes exhibit two developmental trajectories.

Gonadotropes mainly secrete two types of hormones, LH and FSH, which are essential for reproduction in both sexes.

The pseudotime and reclustering analyses interestingly identified five subclusters (Fig. 7a). There was a clear intermediate precursor cell state, Pre.Gonado, which expressed *GATA2* and *FOXL2*, but not *NR5A1*. The other four clusters (Gonadotrope1, 2, 3 and 4) comprised cells that expressed *NR5A1* and *GNRHR* with different expression patterns of *LHB*, *FSHB* and the primate-specific hormone chorionic gonadotropin (CGBs). *NR5A1*, *GNRHR* and *LHB* were expressed in the four cell clusters at a similar level, while CGBs were expressed solely in Gonadotrope2 (*LHB*^{high}*CGB*^{high}*FSHB*^{low}), and *FSHB* was more highly expressed in Gonadotrope4 (*LHB*^{high}*CGB*^{low}*FSHB*^{high}).

The pseudotime analysis revealed two developmental trajectories: one trajectory was from the Pre.Gonado to the Gonadotrope1 and terminated at the Gonadotrope2 (Type I trajectory), while the other trajectory was from the Pre.Gonado to the Gonadotrope3 and Gonadotrope4 (Type II trajectory, Fig. 7a, b). For the Type I trajectory, the intermediate Gonadotrope1 was solely in the early stages (8 to 14 weeks) while Gonadotrope2 was comprised of both early and late stages; all Type II trajectory cells were in the late stages (15 to 25 weeks). These results indicated that the Type I and II trajectories represented an early and a late gonadotrope lineage, respectively. Among all five clusters, Gonadotrope2 had the most DEGs, and comparing between Gonadotrope2 and Gonadotrope4 identified 265 and 30 DEGs, respectively (Fig. 7c, Supplementary Fig. 6 and Supplementary Data 6). GO analysis showed that Gonadotrope2 DEGs enriched “regulated exocytosis” and “C21 – steroid hormone biosynthetic process” (e.g. the steroidogenic enzyme *CYP11A1*), suggesting that Gonadotrope2 actively secreted hormones (Fig. 7c). *WNT4* and *GATA2* were more highly expressed in Gonadotrope2 and Gonadotrope4, respectively (Fig. 7d). Other DEGs included *HIF3A* and *MC2R* in the Type I trajectory and folate receptor *FOLR1* and secretoglobin *SCGB2A1* in the Type II trajectory (Fig. 7d and Supplementary Fig. 6).

Together, these data determined two gonadotrope developmental trajectories including a previously unappreciated *LHB*^{high}*CGB*^{high}*FSHB*^{low} subcluster.

Species comparison between human and rodent pituitaries.

Next, we compared our scRNA-seq data with two recently published scRNA-seq datasets of mouse and rat adult pituitaries^{9,10}. The human scRNA-seq data, which used a plate-based method, recovered higher number of genes per cell comparing with the rodent data using the droplet-based 10X genomic method (Supplementary Fig. 7a). Integrating three datasets showed that most pituitary cell types were conserved among human and rodent (Fig. 8a). Anterior pituitary known markers and new markers, including *SOX2*, *POMC*, *GH1*, *PRL*, *TSHB*, *GNRHR*, *ALDH1A2*, *NR3C1*, *DLK1*, *OLFM1*, *DIO2* and *KCNK3*, showed similar cell-type-specific expression patterns (Fig. 8b, Supplementary Fig. 7b).

Notably, no progenitor or precursor cell types were found in the rodent datasets, indicating that differentiation rarely occurs in these adult pituitaries. Also, in the human fetal pituitaries, the proliferating cells were mainly the stem cells, while in the adult rodent pituitaries, the proliferating cells were mainly the somatotropes and the lactotropes. The rat FSCs were clustered closely to human and mouse stem cells; all expressed *SOX2*, but only the FSCs expressed *S100B*.

We then attempted to find species-specific genes, despite that any differences between the fetal human dataset and the adult rodent datasets could reflect the species or stage differences. We recognized cell-type-specific genes for each species and then identified human (or fetal)-specific genes and rodent (or adult)-specific genes (Supplementary Data 7). Among the identified genes, *CGB* and *GH2* are primate-specific and thus do not exist in

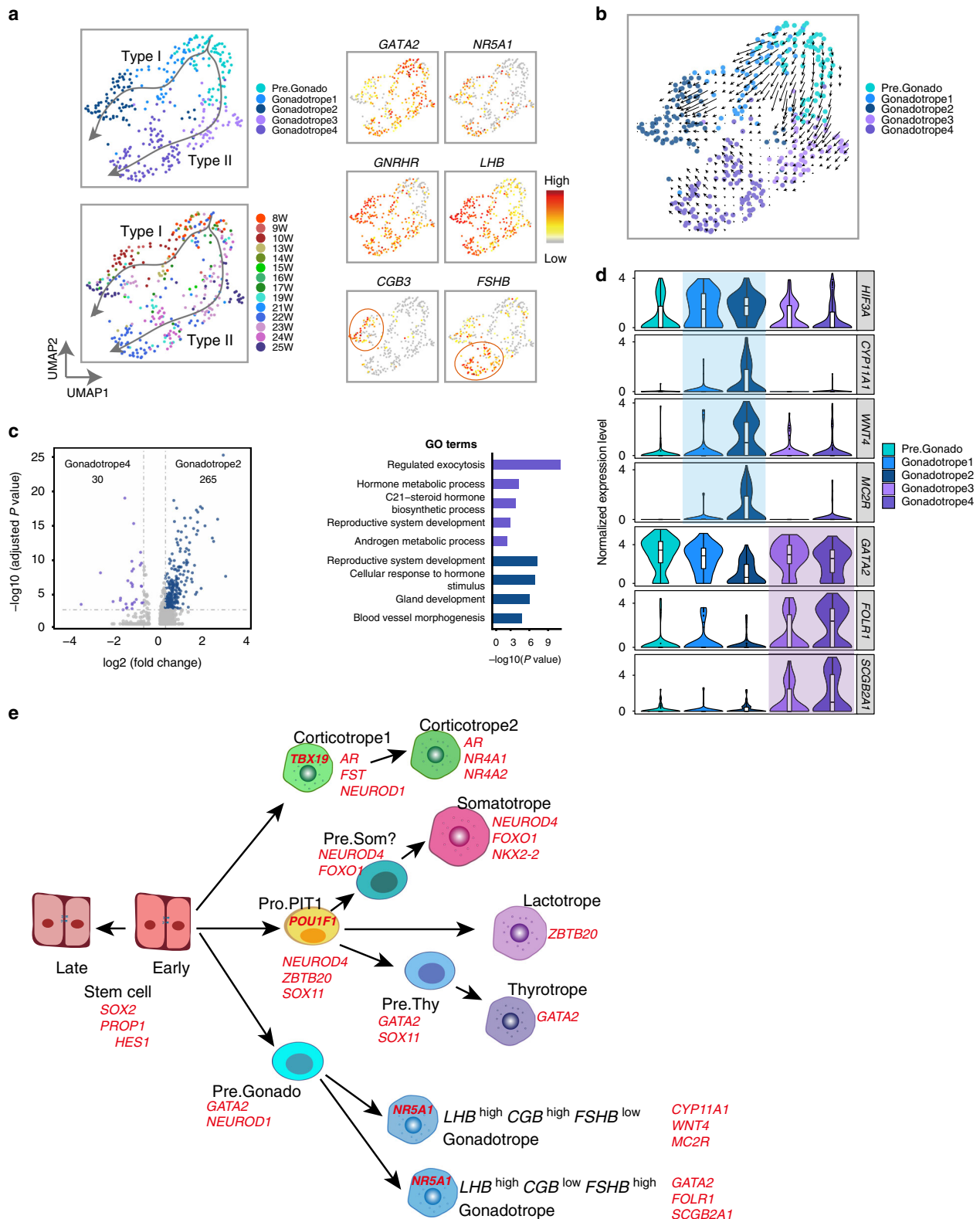


Fig. 7 Two developmental trajectories of gonadotropes. **a** UMAP plots (left) showing that Gonadotropes are comprised of two sublineages with scatterplots (right) showing the expression patterns of representative marker genes. Dots: single cells; colors: cell types (upper) or weeks postfertilization (lower). Gray to red indicate no to high expression levels. **b** RNA velocity analysis of gonadotropes projected onto the UMAP plot showing two sublineages. Dots: single cells; colors: cell types; arrows: predicted directions of the pseudotime. **c** Volcano plots (left) of DEGs between Gonadotrope2 and Gonadotrope4, with bar plots (right) showing GO terms enriched in each subtype. Colored dots indicate significant DEGs in each subtype. *P*-values are determined by two-sided Wilcoxon test and adjusted by Bonferroni correction. **d** Violin with inside box plots showing representative DEGs of two sublineages. **e** Diagram of the development of human fetal anterior pituitary. For a given stage, representative enriched TFs are highlighted. The Pre.Som is labeled with a question mark as this intermediated substate is supported by sufficient number of cells.

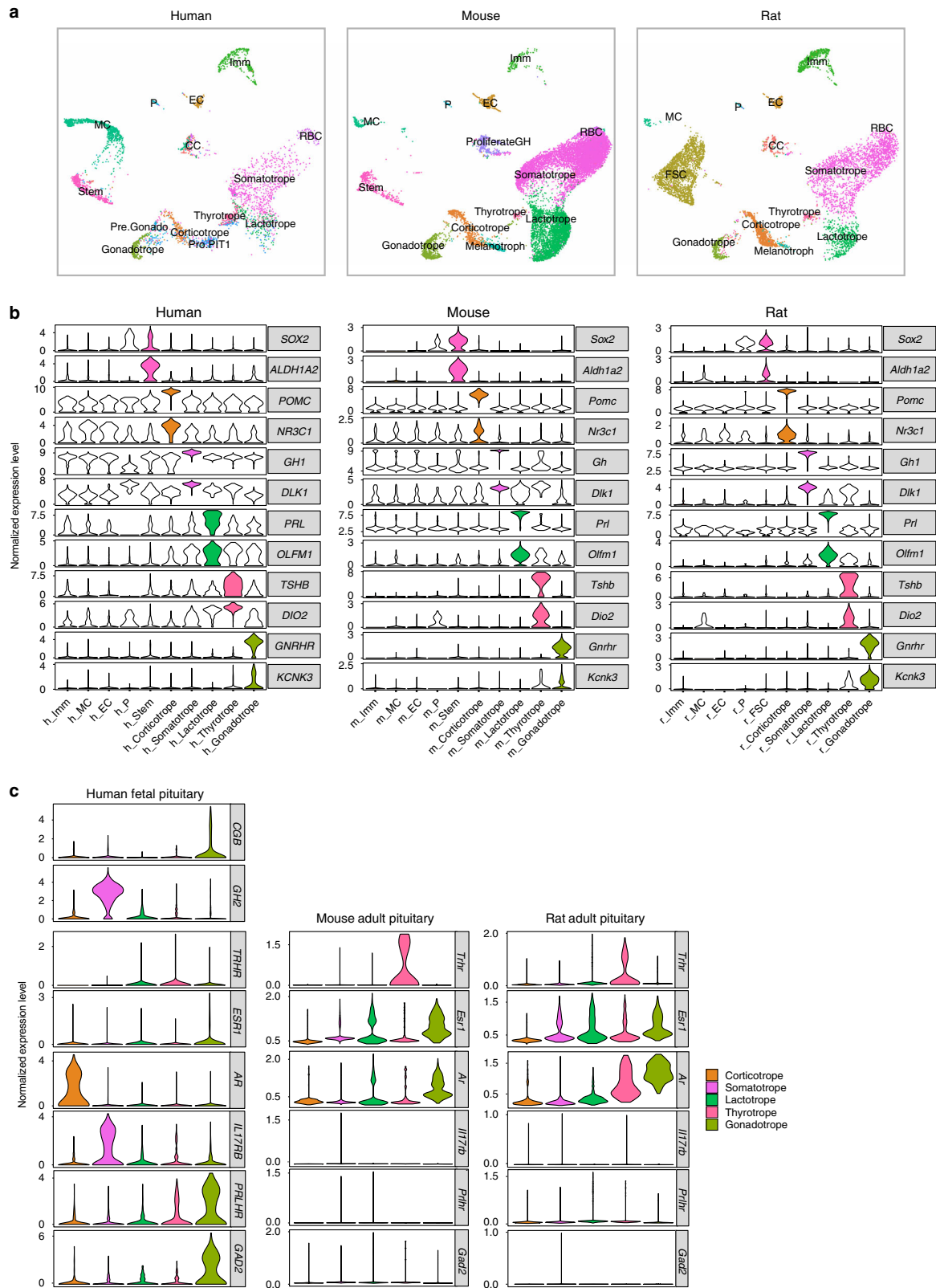


Fig. 8 Comparison among human and rodent pituitaries. **a** UMAP plots of the pituitary cell types by integrating human and rodent pituitary scRNA-seq data sets, with the cell types identified in each data set being shown separately. P, pituicytes; MC, mesenchymal cells; EC, endothelial cells; Imm, immune cells; RBC, red blood cells. **b** Violin plots of the representative identified markers shared among human, mouse and rat. **c** Violin plots of the representative distinct cell type-specific genes between the human fetal pituitary and rodent adult pituitaries.

the rodent data (Fig. 8c). *Trhr* and *Esr1* were highly expressed in the rodent thyrotropes and lactotropes, respectively, but lowly expressed in the human fetal pituitary, thus reflecting stage differences. Interestingly, *AR* was highly expressed in the human fetal corticotropes while lowly expressed in the rodent adult corticotropes instead with high expression in the gonatropes (Fig. 8c). Also, *PRLHR* was highly expressed in the human fetal gonatropes while nearly not expressed in rodent⁴⁷. *IL17RB* and *GAD2* also showed specific expression in the human fetal pituitary. Together, these results revealed a small number of potential species-specific genes.

Discussion

We applied scRNA-seq to elucidate transcriptomic heterogeneity and dynamics in the human fetal pituitary. First, we characterized divergent developmental trajectories with distinct transitional intermediate states in five pituitary hormone-producing cell lineages (Figs. 4a and 7e). For the PIT-1 lineage, integrating our scRNA-seq data with previous mouse genetic studies provides insights into how three distinct cell types are specified from a common progenitor cell. We found that lineage-enriched TFs, including *NEUROD4*, *ZBTB20*, *ASCL1*, *SOX4* and *SOX11*, were coexpressed in the progenitor cells (*NEUROD4*^{mid}/*ZBTB20*^{high}/*ASCL1*^{high}). During differentiation, the cells further segregate into the *NEUROD4*^{high}/*ZBTB20*^{mid}/*ASCL1*^{low} somatotropes, the *NEUROD4*^{mid}/*ZBTB20*^{high}/*ASCL1*^{low} lactotropes and the *NEUROD4*^{low}/*ZBTB20*^{mid}/*ASCL1*^{high} thyrotropes. This finding is consistent with the results of recent scRNA-seq studies demonstrating coactivation of alternative programs preceding two-cell fate commitment in several developmental systems^{48–50}. A recent mouse genetic study interestingly showed that the lactotropes were significantly increased in the *Neurod4*-null mice, with a more prominently increase in the *Ascl1*;*Neurod4* double knockout mice²¹. Thus, it seems that there is competition among the three programs in the progenitor cells with the lactotrope lineage being produced in a situation of high *ZBTB20* expression without high activity of alternative lineage-enriched *NEUROD4* and *ASCL1*.

How thyrotropes are specified in mammals has not been sufficiently clarified. We found that two SoxC family genes, *SOX4* and *SOX11*, were prominently expressed in the Pro.PIT1_{all} cells before activation of *GATA2* and *GATA3* in the Pre.Thy (Fig. 6f), suggesting that the SOX genes play roles in thyrotrope commitment by regulating the GATA genes as *sox4b* does in zebrafish; these two SoxC genes may act redundantly as in other developmental processes⁵¹. Another interesting candidate gene is *DACH1*, which was thyrotrope-enriched and clearly upregulated in the progenitor cells. A previous study has suggested that *Six6* interacts with *Dach1* for regulating cell proliferation in pituitary; however, the function of *DACH1* on the thyrotrope development has not been addressed^{52,53}. It should be noted that *DACH1* was also prominently expressed in Pre.Gonado.

The developmental trajectories of the corticotrope and gonadotrope lineages were different from the developmental trajectory of the PIT1 lineage. Corticotrope development was relatively straightforward. The terminal *POMC* gene was coexpressed with *TBX19* even in the early intermediate state. We did not capture a *TBX19*⁺/*POMC*⁻ state, suggesting that this state may be very transient or much earlier than cells we collected. In contrast, the gonadotrope lineage was characterized by two multistep developmental trajectories that clearly contained a *GATA2*⁺/*POU1F1*⁻/*NR5A1*⁻/*LHB*⁻ intermediate state before terminal differentiation. Interestingly, an interaction between the corticotrope and the gonadotrope lineages seems to exist during the fetal stage. The *LHB*^{high}/*CGB*^{high}/*FSHB*^{low} fetal gonadotrope subtype specifically expresses *MC2R*, which is the ACTH receptor. Additionally, the early

intermediate corticotrope subtype (Corticotrope1) specifically expressed *FST*, which functions as an activin antagonist to inhibit FSH release, and consistently, *FSHB* was expressed at low levels in this fetal gonadotrope subtype (Figs. 5d and 7a).

Second, we identified a fetal gonadotrope subtype (the *LHB*^{high}/*CGB*^{high}/*FSHB*^{low} cells of the Type I trajectory). The hypothalamic-pituitary-gonadal (HPG) axis is activated in the mid-gestational human fetus⁵⁴. Our data suggested that the *LHB*^{high}/*CGB*^{high}/*FSHB*^{low} cells matured at 10 weeks post-fertilization, and thus the developmental timing of these cells matches the activation of the fetal HPG axis. The development of the Type II trajectory occurs after 15 weeks postfertilization, lagging behind the Type I trajectory. In addition, our data suggested that the *LHB*^{high}/*CGB*^{high}/*FSHB*^{low} cells are matured with a feature of actively secreting hormones. These results together indicate that the *LHB*^{high}/*CGB*^{high}/*FSHB*^{low} cells play a role in establishing the early fetal HPG axis. The fetal HPG axis is essential for development of the male genitalia, yet other biological significances are not fully understood⁵⁵. After mid-gestation, the HPG axis is silenced towards the end of gestation, and reactivated at birth (also called minipuberty), and suppressed throughout childhood until reactivation at puberty⁵⁶. It is possible that the *LHB*^{high}/*CGB*^{high}/*FSHB*^{low} cells are also involved in these regulations. Interestingly, it seems that this fetal gonadotrope subtype does not exist in mice since CGB are a primate-specific hormone and the gonadotrope cell type is the last cell type to reach maturation in mice¹; in contrast, we found that this subtype is among the earliest generated cell types in human.

Third, we characterized a hybrid E/M state in the human fetal pituitary stem cells. Previous studies have suggested that pituitary stem cells undergo an EMT-like process for cell migration and differentiation^{23,24,57,58}. Particularly, mice deficient of *Prop1* reveal impaired migration of stem cells, and *Prop1* has been shown to directly target both the epithelial and mesenchymal genes^{24,58}. Our results showed that most typical EMT TFs are not expressed in the stem cells, except weak expression of *SNAIL1*, *SNAIL2* and *ZEB1*, indicating that the cells are not undergoing full EMT, and this is consistent with our previous study on mouse organogenesis²⁵. It is noteworthy that the coexpression of the epithelial and mesenchymal markers is also detected in mouse adult stem cells and rat FSCs (Supplementary Fig. 7b). However, it is possible that, since we only analyzed limited number of cells from 7 weeks on, we have not captured all E/M substates of the pituitary stem cells including the early *SOX2*⁺/*PROPI*⁻ cells and other transitioning cells.

Collectively, this study provides key insights into the transcriptional landscape of human pituitary development, defining distinct cell substates and subtypes, and illustrating transcriptional implementation during major cell fate decisions. The data may also help identify disease genes of congenital hypopituitarism and provide a reference for human pluripotent stem cells-generated anterior pituitary tissue for therapeutic application and disease modeling⁵⁹.

Methods

Human fetal pituitary dissection. The donors in this study were pregnant women who could not continue pregnancy because of their own diseases (such as cervical insufficiency, inevitable abortion, infection, eclampsia, as examples). All patients voluntarily donated the fetal tissues and signed informed consents. This study was approved by the Reproductive Study Ethics Committee of Peking University Third Hospital (2017SZ-043).

We collected 21 human fetal pituitaries from fetuses at 7 to 25 week postfertilization (corresponding to 9 to 27 weeks of gestation), including 11 female fetuses and 10 male fetuses. The pituitary tissues were dissected under a dissecting microscope. For fetuses earlier than 14 weeks, the whole pituitaries were analyzed; for fetuses later than 14 weeks, we separated the anterior and posterior pituitaries, and only analyzed the anterior pituitaries except two fetuses (15W1 and 17W1) of

which we collected cells from both parts. A total of 58 high-quality cells were obtained from the posterior pituitaries of these two fetuses, most of which ($n = 43$) were pituicytes and other cells included 12 immune cells and 3 red blood cells. The tissues were washed twice with Dulbecco's phosphate buffered saline (DPBS) and then the minced tissues were digested with 1 mg/ml type II (17101015, GIBCO) and type IV collagenase (17104019, GIBCO) at 37 °C for 15–30 min. After filtration through 40- μ m nylon mesh, the cells were washed once with 10% FBS DMEM, and a single-cell suspension was obtained.

Immunohistochemistry and immunofluorescence assays. After the whole pituitary was dissected and washed three times with DPBS, the tissue was fixed with 4% paraformaldehyde at 4 °C overnight.

For histological analysis, fixed tissue was stained with H&E. For IF, after washing and dehydration, fixed tissue was embedded in Tissue-Tek O.C.T. Compound (#4583, Sakura) and sectioned at a thickness of 10 μ m. Then, the sections were washed, permeabilized, blocked and incubated with commercial primary antibodies (1:50, Mouse anti-Sox2 antibody, sc365823, Santa Cruz Biotechnology; 1:200, Rabbit Anti-MASH1/Achaete-scute homolog 1 (ASCL1) antibody, ab211327; 1:200, Rabbit Anti-KI67 antibody, ab15580, Abcam; 1:75, Goat Anti-POMC antibody, ab32893, Abcam; 1:50, Mouse Anti-AR antibody, sc-7305, Santa Cruz Biotechnology; Mouse Anti-GH antibody, sc-374266, Santa Cruz Biotechnology; 1:200, Rabbit Anti-NKX2-2 antibody, ab191077, Abcam; 1:500, Rabbit Anti-Collagen III antibody, ab7778, Abcam; 1:50, Mouse Anti-PIT1 antibody, sc-25258, Santa Cruz Biotechnology; 1:200, Rabbit Anti-TSH β antibody, ab155958, Abcam; 1:50, Mouse Anti-PRL antibody, sc-46698, Santa Cruz Biotechnology; 1:100, Rabbit Anti-FSH β antibody, ab180489, Abcam; 1:100, Rabbit Anti-LH β antibody, ab150416, Abcam) at 4 °C overnight. We used commercial secondary antibodies that were incubated for 2 h at room temperature. Finally, the sections were counterstained with DAPI in an antifade solution (P36931, Invitrogen) and then mounted. The samples were imaged by using an AIRSi+ Nikon confocal microscope (Nikon, Japan).

Statistics and reproducibility. For each representative immunohistochemistry and immunofluorescence assay, we took the nearby 1–2 weeks as biological replicates ($n \geq 2$) due to sampling limitations. The positive cells in different sections ($n \geq 3$) were counted automatically by ImageJ software or manually.

scRNA-seq library construction and sequencing. We used a mouth pipette to randomly pick single cells and used a modified STRT-seq protocol to construct a scRNA-seq library^{25,60}. The cells were lysed in a lysis buffer with an 8-nt cell barcode and 8-nt UMI. Then, mRNA was reverse transcribed into cDNA with SuperScript™ II Reverse Transcriptase (18064014, Invitrogen). After preamplification, the samples with different cell barcodes were pooled together, and the mixture was labeled with a biotin modification by 4 cycles of PCR. Then, the full-length cDNA was sheared into fragments with approximately 300-bp lengths by Covaris (S2), and the 3' cDNA was enriched by Dynabeads MyOne Streptavidin C1 (65002, Invitrogen) to construct a library according to the Kapa Hyper Prep Kit protocol (KK8505, Kapa Biosystems). Cleaned libraries were sequenced as paired-end 150-base reads on an Illumina HiSeq platform (sequenced by Novogene). As individual fetal samples were collected at different time points, they were subjected to different sequencing runs.

scRNA-seq data processing. Cells were split by the first 8-bp barcode in Reads2, and then the next 8-bp UMIs in Reads2 were added to the header in Reads1. After trimming the template switch oligo (TSO) and polyA sequences and removing the short reads (length < 37 bp) and low-quality reads ($N > 10\%$) in Reads1, the clean reads were aligned by Tophat (version 2.0.12) to the hg19 genome downloaded from UCSC. Then, HTSeq was applied to count the uniquely mapped reads⁶¹, and the number of different UMIs for each gene in each cell was considered the transcriptional count.

To filter out low-quality cells and multiple cells sequenced as one cell, we selected only cells with a gene number ≥ 2000 , an initial reads number $\leq 1E7$, and a mapping ratio $\geq 20\%$ and genes with at least one count in at least three cells for the following analysis. The filtered gene expression data set of transcriptional counts was analyzed using the Seurat (Version 2.3.4) package¹². As most of the single cells in our data sets were around 100,000 transcriptional counts, the scale factor was set to 100,000. The resolution of “FindClusters” was set as 1 for all cells and merged subclusters of MC with the same known markers (Supplementary Fig. 1d, f), 1.5 for endocrine cells and merged subclusters of Stem, Somatotrope and Gonadotrope with the same known markers (Fig. 1b, c), 0.3 for stem cells (Fig. 2c), 0.4 for gonadotropes (Fig. 7a). Main cell types were identified by the combination of known markers for each cluster.

Identification of DEGs and enrichment analysis. DEGs were identified by the function “FindAllMarkers” or “FindMarkers” in Seurat packages using “wilcox” test methods and Bonferroni correction. Significant DEGs were selected from genes with adjusted P value $p_{val_adj} \leq 0.01$ and log processed average fold change $avg_logFC \geq 0.5$ for further analysis and visualization. In venn diagram of Fig. 6d, TFs were selected from significantly upregulated DEGs ($p_{val_adj} \leq 0.01$ and

$avg_logFC \geq 0.5$) between each two of lineages. TFs of each lineage were the union of upregulated genes ($P \leq 1E-4$) compared to the other two lineages, and then the intersection of upregulated genes ($P \leq 1E-4$) compared to the other two lineages were selected as the specific TFs of that lineage. GO analysis and KEGG pathway enrichment analysis of these significant DEGs were performed by Metascape (<http://metascape.org>)⁶². Pathway enrichment comparisons of each combination of two clusters were analyzed by GSEA^{63,64}.

Remove cell cycle effect. To remove cell cycle effect in the non-proliferative stem cell, we firstly ran a PCA on cell cycle genes in Seurat package (s.genes, g2/m.genes) of stem cells and observed a little cell cycle effect in some stem cells. Then, we regressed out cell cycle scores (S.Score and G2/M.Score), and re-ran a PCA on cell cycle genes and found no cells were separated by these genes. Next, we found clusters using the newly scaled data after regressing out cell cycle scores, and revealed that the newly identified clusters were same as the originally found stem subtypes.

Gene score definitions. We defined the E.score, the M.score and the S.score by averaging the expression levels of curated epithelial markers, mesenchymal markers collected from previous studies, and stemness genes in the GO term “stem cell population maintenance” (GO: 0019827) respectively^{25,26}. Markers of these scores were listed in Supplementary Data 2.

Prediction of activated TFs. The modules of TFs were identified by the SCE-NIC¹⁷ python workflow (version 0.9.1) using default parameters (<http://scenic.aertslab.org>). A human TF gene list was collected from the resources of pySCENIC (<https://github.com/aertslab/pySCENIC/tree/master/resources>), animal TFDB^{65,66} (<http://bioinfo.life.hust.edu.cn/HumanTFDB#!/download>) and the Human Transcription Factors⁶⁷ database (<http://humantfs.ccrb.utoronto.ca/download.php>). Activated TFs were identified in the AUC matrix, and differentially activated TFs were selected using “FindAllMarkers” of the Seurat package. The top 10 enriched activated TFs were ranked by $-\log_{10}(p_value)$ and demonstrated using the binary matrix (1 activated; 0 not activated). Networks of the modules with TFs and their target genes were visualized by the R package igraph (version 1.2.5) (<https://igraph.org/>).

Construction of lineage trajectories. Lineage trajectories were constructed by Slingshot²⁹ with a UMAP or PCA plot as the dimensionality reduction results. For Fig. 4a, the start cluster was set as Stem1, and end clusters were mature hormone producing cell types. For Fig. 7a, the start cluster was set as Pre.Gonado. The trajectories were considered the developmental pseudotime of each lineage. TF dynamics along the pseudotime axis were identified by the R package gam (version 1.16)⁶⁸, and significantly changed TFs were selected from TFs with P -values $\leq 1E-5$. In Fig. 4d, upregulated and downregulated TFs were two clusters of candidate TFs identified by hierarchical clustering of genes in Fig. 4b.

RNA velocity analysis. The directions of pseudotime were predicted by RNA velocity²⁸ using exonic and intronic gene expression levels. After alignment by Tophat, mapped bam files were processed by the python package velocityo (version 0.9.1) to produce loom files with spliced and unspliced gene counts. Then, the loom files were merged and analyzed to predict directions following the analysis pipeline with a k -nearest neighbor ($k = 10$). The directions of RNA velocity were projected in a UMAP plot.

Cell-cell interaction analysis. Interactions between pairwise cell clusters were inferred by CellPhoneDB v.2.0⁶⁹, which includes a public repository of curated ligands, receptors and their interactions. We ran the CellPhoneDB framework using a statistical method and detected L-R pairs that were expressed in more than 20% of cells. Significant L-R pairs (P -value ≤ 0.05 and mean value ≥ 0.5) were demonstrated using igraph and heatmap. Cell types expressing ligands were considered as active cell types sending signals, while cell types expressing the corresponding receptors were considered as target cell types receiving signals.

Integrating human and rodent datasets. Mouse and rat data sets were downloaded from GEO website GSE120410 and GSE132224, respectively. Cell types were identified by Seurat in each data set independently, and then DEGs were identified by the function “FindAllMarkers” of Seurat packages using “wilcox” test methods and Bonferroni correction. Significant DEGs of human data were selected from genes with $p_{val_adj} \leq 0.01$ and $avg_logFC \geq 0.5$. For identifying the potential species-specific genes, we firstly identified cell-type-specific genes (the corticotropes, somatotropes, lactotropes, thyrotropes and gonadotropes) for each species and then made comparison between human and rodent. The RESCUE⁷⁰ (version 1.0.1) method was applied for imputing the dropouts for the rodent data. The rodent genes required to be both the mouse and the rat genes. As gene number and gene expression level were much lower in rodent data sets, significant DEGs of rodent data sets were selected from genes with $p_{val_adj} \leq 0.01$ and $avg_logFC \geq 0.25$.

To compare the three data sets further, we integrated them by Seurat3 (Version 3.1.1) standard workflow. We only used genes shared in all three data sets. Top 2000 variable features were selected by variance stabilizing transformation (“vst”), and then anchors were identified and passed to the “IntegrateData” function to integrate them, which also removed batch effect among these data sets. After scaling the integrated data and running PCA and UMAP, we clearly observed the relationships among them in Fig. 8a. We also used this workflow to remove batch effect in somatotropes and lactotropes from late stages with more than 10 cells.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The authors declare that all data supporting the findings of this study are available within the article and its supplementary information files or from the corresponding author upon reasonable request.

The raw data have been deposited in the GSA (Genome Sequence Archive) databases of the National Genomics Data Center (NGDC, <https://bigd.big.ac.cn/>) under the BioProject accession code: PRJCA003249. The gene expression matrix data have been deposited in both the GSA and the Gene Expression Omnibus (GEO) under accession code: GSE142653. There are no restrictions on access to these data.

Gene expression patterns of the endocrine cells are also available on the shiny webpage: https://tanglab.shinyapps.io/Human_Fetal_Pituitary_Endocrine_Cells/.

Received: 23 December 2019; Accepted: 23 September 2020;

Published online: 19 October 2020

References

- Zhu, X., Gleiberman, A. S. & Rosenfeld, M. G. Molecular physiology of pituitary development: signaling and transcriptional networks. *Physiol. Rev.* **87**, 933–963 (2007).
- Li, S. et al. Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1. *Nature* **347**, 528–533 (1990).
- Lamolet, B. et al. A pituitary cell-restricted T box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins. *Cell* **104**, 849–859 (2001).
- Zhao, L. P. et al. Steroidogenic factor 1 (SF1) is essential for pituitary gonadotrope function. *Development* **128**, 147–154 (2001).
- Andoniadou, C. L. et al. Sox2(+) stem/progenitor cells in the adult mouse pituitary support organ homeostasis and have tumor-inducing potential. *Cell Stem Cell* **13**, 433–445 (2013).
- Rizzoti, K., Akiyama, H. & Lovell-Badge, R. Mobilized adult pituitary stem cells contribute to endocrine regeneration in response to physiological demand. *Cell Stem Cell* **13**, 419–432 (2013).
- Brinkmeier, M. L. et al. Discovery of transcriptional regulators and signaling pathways in the developing pituitary gland by bioinformatic and genomic approaches. *Genomics* **93**, 449–460 (2009).
- Ma, Y. et al. Identification of candidate genes for human pituitary development by EST analysis. *BMC Genomics* **10**, 109 (2009).
- Fletcher, P. A. et al. Cell type- and sex-dependent transcriptome profiles of rat anterior pituitary cells. *Front. Endocrinol.* **10**, 623 <https://doi.org/10.3389/fendo.2019.00623> (2019).
- Cheung, L. Y. M. et al. Single-cell RNA sequencing reveals novel markers of male pituitary stem cells and hormone-producing cell types. *Endocrinology* **159**, 3910–3924 (2018).
- Chen, Q., Leshkowitz, D., Blechman, J. & Levkowitz, G. Single-cell molecular and cellular architecture of the mouse neurohypophysis. *eNeuro* **7**, <https://doi.org/10.1523/ENEURO.0345-19.2019> (2020).
- Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e1821 (2019).
- Farahani, R. M. & Xaymardan, M. Platelet-derived growth factor receptor alpha as a marker of mesenchymal stem cells in development and stem cell biology. *Stem Cells Int.* **2015**, 362753 (2015).
- Mortensen, A. H., Vanessa, S., Thomas, L. & Camper, S. A. Deletion of OTX2 in neural ectoderm delays anterior pituitary development. *Hum. Mol. Genet.* **24**, 939–953 (2015).
- Baker, B. L. & Jaffe, R. B. The genesis of cell types in the adenohypophysis of the human fetus as observed with immunocytochemistry. *Am. J. Anat.* **143**, 137–161 (1975).
- Fang, Q. et al. Genetics of combined pituitary hormone deficiency: roadmap into the genome era. *Endocr. Rev.* **37**, 636–675 (2016).
- Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
- Zhang, F. et al. Enhancer-bound LDB1 regulates a corticotrope promoter-pausing repression program. *Proc. Natl Acad. Sci. USA* **112**, 1380–1385 (2015).
- Pogoda, H. M. et al. The proneural gene ascl1a is required for endocrine differentiation and cell survival in the zebrafish adenohypophysis. *Development* **133**, 1079–1089 (2006).
- Prince, K. L., Walvoord, E. C. & Rhodes, S. J. The role of homeodomain transcription factors in heritable pituitary disease. *Nat. Rev. Endocrinol.* **7**, 727–737 (2011).
- Ando, M. et al. The proneural bHLH genes Mash1, Math3 and NeuroD are required for pituitary development. *J. Mol. Endocrinol.* **61**, 127–138 (2018).
- Allaerts, W. & Vankelecom, H. History and perspectives of pituitary folliculostellate cell research. *Eur. J. Endocrinol.* **153**, 1–12 (2005).
- Cheung, L. Y., Davis, S. W., Brinkmeier, M. L., Camper, S. A. & Pérezmillán, M. I. Regulation of pituitary stem cells by epithelial to mesenchymal transition events and signaling pathways. *Mol. Cell. Endocrinol.* **445**, 14 (2016).
- Inés, P. M. M., Brinkmeier, M. L., Mortensen, A. H. & Camper, S. A. PROP1 triggers epithelial-mesenchymal transition-like process in pituitary stem cells. *Elife* **5**, <https://doi.org/10.7554/eLife.14470> (2016).
- Dong, J. et al. Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. *Genome Biol.* **19**, 31 (2018).
- Pastushenko, I. et al. Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463–468 (2018).
- Vieira, A. F. & Paredes, J. P-cadherin and the journey to cancer metastasis. *Mol. Cancer* **14**, 178 (2015).
- La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
- Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
- Gaston-Massuet, C. et al. Transcription factor 7-like 1 is involved in hypothalamo-pituitary axis development in mice and humans. *Proc. Natl Acad. Sci. USA* **113**, E548–E557 (2016).
- Tatsi, C. et al. Pituitary stalk interruption syndrome and isolated pituitary hypoplasia may be caused by mutations in holoprosencephaly-related genes. *J. Clin. Endocrinol. Metab.* **98**, E779–E784 (2013).
- Davis, S. W., Mortensen, A. H. & Camper, S. A. Birthdating studies reshape models for pituitary gland cell specification. *Dev. Biol.* **352**, 215–227 (2011).
- Miyamoto, J. et al. The pituitary function of androgen receptor constitutes a glucocorticoid production circuit. *Mol. Cell. Biol.* **34**, 4807–4814 (2007).
- Helbling, J. C., Minni, A. M., Pallet, V. & Moisan, M. P. Stress and glucocorticoid regulation of NR4A genes in mice. *J. Neurosci. Res.* **92**, 825–834 (2014).
- Budry, L. et al. The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. *Genes Dev.* **26**, 2299–2310 (2012).
- Tatsumi, K. et al. Cretinism with combined hormone deficiency caused by a mutation in the PIT1 gene. *Nat. Genet.* **1**, 56–58 (1992).
- Zhu, X. Y. et al. Sustained Notch signaling in progenitors is required for sequential emergence of distinct cell lineages during organogenesis. *Genes Dev.* **20**, 2739–2753 (2006).
- Kapali, J. et al. Foxo1 is required for normal somatotrope differentiation. *Endocrinology* **157**, 4351–4363 (2016).
- Cao, D. M. et al. ZBTB20 is required for anterior pituitary development and lactotrope specification. *Nat. Commun.* **7**, 11121 <https://doi.org/10.1038/ncomms11121> (2016).
- Dong, Q., Chen, X. Y. & Li, G. M. Effect of transcription factor ZBTB20 on mouse pituitary development. *Genet. Mol. Res.* **14**, 17622–17629 (2015).
- Dasen, J. S. et al. Reciprocal interactions of Pit1 and GATA2 mediate signaling gradient-induced determination of pituitary cell types. *Cell* **97**, 587–598 (1999).
- Quiroz, Y. et al. The HMG-Box transcription factor Sox4b is required for pituitary expression of gata2a and specification of thyrotrope and gonadotrope cells in zebrafish. *Mol. Endocrinol.* **26**, 1014–1027 (2012).
- Balderes, D. A., Magnuson, M. A. & Sussel, L. Nkx2.2:Cre knock-in mouse line: a novel tool for pancreas- and CNS-specific gene deletion. *Genesis* **51**, 844–851 (2013).
- Kaplan, S. L., Grumbach, M. M. & Aubert, M. L. The ontogenesis of pituitary hormones and hypothalamic factors in the human fetus: maturation of central nervous system regulation of anterior pituitary function. *Recent Prog. Horm. Res.* **32**, 161–243 (1976).
- Frawley, L. S. & Boockfor, F. R. Mammosomatotropes: presence and functions in normal and neoplastic pituitary tissue. *Endocr. Rev.* **12**, 337–355 (1991).
- Luque, R. M. et al. Reporter expression, induced by a growth hormone promoter-driven Cre recombinase (rGHP-Cre) transgene, questions the developmental relationship between somatotropes and lactotropes in the adult mouse pituitary gland. *Endocrinology* **148**, 1946–1953 (2007).
- Prazienskova, V., Popelova, A., Kunes, J. & Maletinska, L. Prolactin-releasing peptide: physiological and pharmacological properties. *Int. J. Mol. Sci.* **20**, <https://doi.org/10.3390/ijms20215297> (2019).

48. Treutlein, B. et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
49. Petropoulos, S. et al. Single-cell RNA-Seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **167**, 285 (2016).
50. Soldatov, R. et al. Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* **364**, <https://doi.org/10.1126/science.aas9536> (2019).
51. Kavyanifar, A., Turan, S. & Lie, D. C. SoxC transcription factors: multifunctional regulators of neurodevelopment. *Cell Tissue Res.* **371**, 91–103 (2018).
52. Li, X., Perissi, V., Liu, F., Rose, D. W. & Rosenfeld, M. G. Tissue-specific regulation of retinal and pituitary precursor cell proliferation. *Science* **297**, 1180–1183 (2002).
53. Backman, M., Machon, O., Van Den Bout, C. J. & Krauss, S. Targeted disruption of mouse Dach1 results in postnatal lethality. *Dev. Dyn.* **226**, 139–144 (2003).
54. Kuirri-Hanninen, T., Sankilampi, U. & Dunkel, L. Activation of the hypothalamic-pituitary-gonadal axis in infancy: minipuberty. *Horm. Res. Paediatr.* **82**, 73–80 (2014).
55. Zambrano, E., Guzman, C., Rodriguez-Gonzalez, G. L., Durand-Carbajal, M. & Nathanielsz, P. W. Fetal programming of sexual development and reproductive function. *Mol. Cell Endocrinol.* **382**, 538–549 (2014).
56. Abreu, A. P. & Kaiser, U. B. Pubertal development and regulation. *Lancet Diabetes Endocrinol.* **4**, 254–264 (2016).
57. Saishu, Y., Takako, K. & Yukio, K. EMT Involved in Migration of Stem/Progenitor Cells for Pituitary Development and Regeneration. *J. Clin. Med.* **5**, <https://doi.org/10.3390/jcm5040043> (2016).
58. Himes, A. D. & Raetzman, L. T. Premature differentiation and aberrant movement of pituitary cells lacking both Hes1 and Prop1. *Dev. Biol.* **325**, 151–161 (2009).
59. Ozone, C. et al. Functional anterior pituitary generated in self-organizing culture of human embryonic stem cells. *Nat. Commun.* **7**, 10351 (2016).
60. Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
61. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166 (2015).
62. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
63. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
64. Mootha, V. K. et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
65. Zhang, H. M. et al. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res.* **43**, D76–D81 (2015).
66. Hu, H. et al. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **47**, D33–D38 (2019).
67. Lambert, S. A. et al. The human transcription factors. *Cell* **175**, 598–599 (2018).
68. Hastie, T. & Tibshirani, R. Generalized additive models for medical research. *Stat. Methods Med. Res.* **4**, 187–196 (1995).
69. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).
70. Tracy, S., Yuan, G. C. & Dries, R. RESCUE: imputing dropout events in single-cell RNA-sequencing data. *BMC Bioinformatics* **20**, 388 (2019).

Acknowledgements

The authors thank Dr Hongbo Yang at Department of Endocrinology, Peking Union Medical College Hospital for insightful discussion. We thank Chunyan Shan from the Core Facilities at the School of Life Sciences of Peking University for assistance with confocal microscopy, as well as High Performance Computing Platform of the Center for Life Science at Peking University for assistance with computing analysis. This work is supported by the National Key R&D Program of China (2018YFC1003101, 2018YFA0107601 and 2017YFA0103402), the National Natural Science Foundation of China (31871457, 81521002), the Research Units of Comprehensive Diagnosis and Treatment of Oocyte Maturation Arrest (2018RU001, Chinese Academy of Medical Sciences), the Beijing Municipal Science and Technology Commission (Z181100001318001), and Beijing Advanced Innovation Center for Genomics (ICG) at Peking University.

Author contributions

L.W., J.Q., T.F. and S.Z. conceived the project. L.W. and S.Z. wrote the manuscript with help from all of the authors. Y.C., J.Y., M.Y. and J.R. performed scRNA-seq. Y.C. and X.M. performed immunofluorescence, immunohistochemistry and imaging. S.Z. conducted the bioinformatics analyses. All of the authors edited and proofread the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-19012-4>.

Correspondence and requests for materials should be addressed to L.W. or J.Q.

Peer review information *Nature Communications* thanks Sally Camper and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020