HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2022 March 27.

Published in final edited form as:

Cell Rep. 2022 March 08; 38(10): 110467. doi:10.1016/j.celrep.2022.110467.

Single nucleus transcriptome and chromatin accessibility of postmortem human pituitaries reveal diverse stem cell regulatory mechanisms

Zidong Zhang[1,9], **Michel Zamojski**[2,9], **Gregory R. Smith**[2,9], **Thea L. Willis**[3,9], **Val Yianni**[3,9], **Natalia Mendelev**[2], **Hanna Pincas**[2], **Nitish Seenarine**[2], **Mary Anne S. Amper**[2], **Mital Vasoya**[2], **Wan Sze Cheng**[2], **Elena Zaslavsky**[2], **Venugopalan D. Nair**[2], **Judith L. Turgeon**[4], **Daniel J. Bernard**[5], **Olga G. Troyanskaya**[1,6,7], **Cynthia L. Andoniadou**[3,8,*], **Stuart C. Sealfon**[2,*], **Frederique Ruf-Zamojski**[2,10,*]

[1]Lewis-Sigler Institute for Integrative Genomics and Graduate Program in Quantitative and Computational Biology, Princeton University, Princeton, NJ, USA

[2]Department of Neurology, Center for Advanced Research on Diagnostic Assays, Icahn School of Medicine at Mount Sinai (ISMMS), New York, NY, USA

[3]Center for Craniofacial and Regenerative Biology, King's College London, London, UK

[4]Department of Internal Medicine, University of California, Davis, Davis, CA, USA

[5]Department of Pharmacology and Therapeutics, McGill University, Montreal, QC, H3G 1Y6, Canada

[6]Department of Computer Science, Princeton University, Princeton, NJ, USA

[7]Flatiron Institute, Simons Foundation, New York, NY, USA

[8]Department of Medicine III, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[9]These authors contributed equally

[10]Lead contact

SUMMARY

*Correspondence: cynthia.andoniadou@kcl.ac.uk (C.L.A.), stuart.sealfon@mssm.edu (S.C.S.), frederique.ruf-zamojski@mssm.edu (F.R.-Z.).

AUTHOR CONTRIBUTIONS

Z.Z., M.Z., G.R.S., V.Y., and O.G.T. contributed analytic tools and analyzed data; T.L.W. contributed analytic tools, analyzed data, and performed research; N.M., V.D.N., N.S., M.A.A., and M.V. performed research; H.P. drafted the manuscript; W.C., M.Z., and E.Z. established the quality control pipeline; S.C.S. designed the study, analyzed, interpreted data, and drafted the manuscript; C.L.A., J.L.T., and D.J.B. analyzed, interpreted data, and drafted the manuscript; F.R.Z. designed the study, performed research, analyzed and interpreted data, and drafted the manuscript. All authors edited the manuscript and approved its final version.

SUPPLEMENTAL INFORMATION

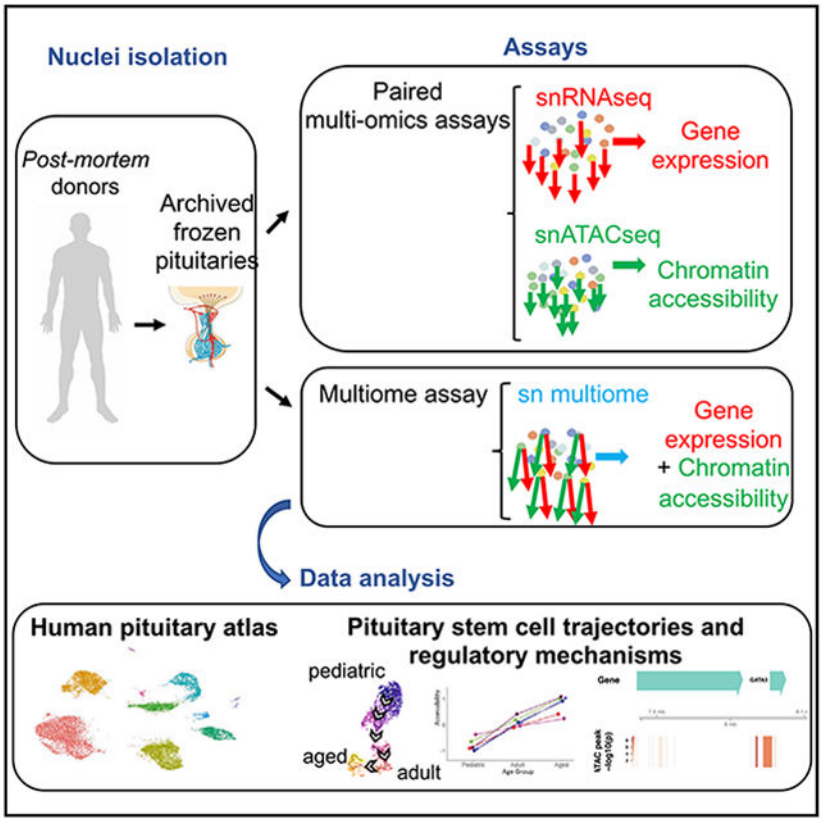
Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.110467.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Despite their importance in tissue homeostasis and renewal, human pituitary stem cells (PSCs) are incompletely characterized. We describe a human single nucleus RNA-seq and ATAC-seq resource from pediatric, adult, and aged *postmortem* pituitaries (snpituitaryatlas.princeton.edu) and characterize cell-type-specific gene expression and chromatin accessibility programs for all major pituitary cell lineages. We identify uncommitted PSCs, committing progenitor cells, and sex differences. Pseudotime trajectory analysis indicates that early-life PSCs are distinct from the other age groups. Linear modeling of same-cell multiome data identifies regulatory domain accessibility sites and transcription factors that are significantly associated with gene expression in PSCs compared with other cell types and within PSCs. We identify distinct deterministic mechanisms that contribute to heterogeneous marker expression within PSCs. These findings characterize human stem cell lineages and reveal diverse mechanisms regulating key PSC genes and cell type identity.

Graphical abstract



In brief

This study profiles the gene expression and chromatin accessibility landscapes in *postmortem* male and female pituitaries of different ages using single nucleus multiomics technologies. Zhang et al. characterize the pituitary stem cell population and develop computational methods, which allow us to elucidate regulatory mechanisms underlying pituitary stem cell identity.

INTRODUCTION

Tissues are composed of several cell types that can assume different gene expression states in response to environmental cues (Morris, 2019). Major objectives of current biological research include resolving cellular heterogeneity within tissues and elucidating the regulatory mechanisms determining cell types and states. With the recent development of single-cell (sc) omics technologies, researchers have refined the characterization of cell types in many tissues (Griffiths et al., 2018; Tanay and Regev, 2017).

The pituitary gland secretes hormones that control crucial physiological processes, including reproduction, metabolism, and the stress response. The adenohypophysis represents the main portion of the pituitary gland and contains five hormone-producing cell lineages. Despite the physiological relevance of the pituitary in health and disease, human sc RNA-sequencing (RNA-seq) studies to date have omitted the postnatal pituitary (Han et al., 2020; Zhang et al., 2020). Furthermore, mapping of the pituitary epigenome landscape has not been included in the ENCODE project (ENCODE Project Consortium et al., 2020; Davis et al., 2018), and no chromatin accessibility profiling of the human pituitary at sc resolution has been reported.

Of particular interest is the insight into pituitary stem cells (PSCs) to be obtained from sc analyses. PSCs are a heterogeneous population containing both uncommitted and committing stem cells, and thus provide an ideal target to study the transcriptional and chromatin accessibility mechanisms underlying the expression of specific lineage-defining genes. Furthermore, the study of PSCs is motivated by the need to develop stem cell therapies that could restore lost or damaged endocrine cell populations in the pituitary. Pituitary hormone deficiencies, which include congenital hypopituitarism (combined pituitary hormone deficiencies), acquired hypopituitarism (secondary to trauma, surgery, chemotherapy, or radiotherapy), as well as pituitary tumors such as adenomas, result in a severe disruption of endocrine systems that could be addressed with PSC therapy (Castinetti et al., 2011). Previous mouse studies demonstrated the existence of PSCs and their ability to self-renew and differentiate into all five endocrine cell types (Andoniadou et al., 2013; Rizzoti et al., 2013), thus opening potential therapeutic avenues for human pituitary deficiencies and pituitary tumors (Russell et al., 2018; Vankelecom, 2016). Little is known about the epigenetic landscape (chromatin changes as measured by chromatin accessibility assays) and the dynamics of human PSCs during postnatal life, which is critical information for realizing their therapeutic potential. Sc resolution studies of the human pituitary are important for resolving cellular identity and revealing the regulatory mechanisms of this key cell type.

One impediment to conducting sc resolution studies of the human pituitary is the technical difficulty in generating high-quality datasets from the frozen *postmortem* pituitary samples provided by tissue banks, which are the only potential source for normal pituitary tissue. We recently developed an integrated single nucleus (sn) multiomics analysis using frozen adult murine pituitary (Ruf-Zamojski et al., 2021). We have now successfully employed a similar procedure to characterize all major cell types in the human pituitary with a particular focus on PSCs. Archived frozen *postmortem* pituitaries from pediatric, adult, and elderly

male and female subjects were jointly analyzed by snRNA-seq and sn assay for transposase-accessible chromatin using *sequencing* (ATAC-seq) (sn multiomics). Importantly, we also generated same-cell human pituitary sn multiome data. Analyses of these data enabled us to characterize the transcriptome and chromatin accessibility landscapes of pituitary cell types. We further refined the identification of human PSC subtypes and their changes during aging and provide insight into the diverse gene regulatory mechanisms underlying stem cell identity and commitment and heterogeneous marker expression within the PSC cell type.

RESULTS

Sn multiomics profiling of human pituitaries

To construct cell-type-specific genome-wide maps of gene expression and open chromatin in the human pituitary, we conducted same-sample multiomics assays of sn transcriptome (snRNA-seq) and sn chromatin accessibility (snATAC-seq) in frozen *postmortem* pituitaries from pediatric, adult, and aged males and females. These six pituitaries had been stored in tissue banks at −80 °C for an average of ~10 years since donation (range 4–20; Table 1). Of note, nuclei were isolated from the same pituitary fragment and were processed for both snRNA-seq and snATAC-seq. Hence, although the paired datasets were not derived from the same nuclei, they were sampled from the same population of nuclei in each analyzed pituitary. In addition, to test for tissue heterogeneity and accuracy of cell-type-specific mapping across assays, and to improve inference of regulatory mechanisms, the remaining sample comprising nearly the entire pituitary from one female donor was pulverized and the isolated nuclei were then used to carry out 1) same-sample analysis of sn transcriptome and sn chromatin accessibility, and 2) same-cell sn multiome analysis providing simultaneous measurement of RNA expression and chromatin accessibility within each individual nucleus (Figure 1A)

All snRNA-seq and snATAC-seq libraries generated from the same samples were pooled for sequencing to reduce batch effects. Data meeting the quality control (QC) threshold were obtained from a total of 76,016 nuclei for snRNA-seq and 44,141 nuclei for snATAC-seq in paired assays, and from 15,024 nuclei in the same-cell sn multiome assay (Table S1). All our libraries underwent detailed QC analysis, were assessed for multiple metrics, and none of the samples failed (Tables S1A2 and S1B2). For data analysis of a given sample processed through the same-sample sn paired assays, we generated uniform manifold approximation projections (UMAPs) for both snATAC-seq and snRNA-seq datasets, each identifying cell clusters by type. Integration of both datasets resulted in an overlay UMAP showing good correspondence of the major pituitary cell types across assay modalities (Figures 1B and S1). The same-cell sn multiome assay, in which each cell yielded both RNA-seq and ATAC-seq datasets (Figure 1C), directly generated an integrated UMAP plot (Figure 1D).

Transcriptome analysis of human pituitary cell types

The same-sample paired assay snRNA-seq datasets had an average of 86% of reads mapped to the transcriptome and allowed for the detection of ~2,800 genes per nucleus, with comparable high-quality QC metrics obtained from all samples (Tables S1A1 and S1A2). In the snRNA-seq data analysis of individual male and female pituitary samples, cells

were clustered using Seurat, projected on UMAPs, as well as visualized using t-distributed stochastic neighbor embedding (t-SNE) representation (Figure S1). Cell clusters were annotated manually using differential RNA expression of established pituitary marker genes. Key cell type markers included *FSHB, LHB*, and *GNRHR* for gonadotropes; *GH1* for somatotropes; *POMC* for corticotropes; *DIO2* for thyrotropes (Cheung et al., 2018; Zhang et al., 2020); *PRL* for lactotropes; and *SOX9* (Rizzoti et al., 2013), *LGR4* (Russell et al., 2021), and *RBPMS* (Cheung et al., 2018) for PSCs. A list of established markers used for the assignment of each cell type as well as markers identified in our datasets (*IQCJ-SCHIP1, NTNG1, EBF1, BNC2, ADGRA3* for PSCs) are listed in Table S2. The RNA counts (Figure S1H), mitochondrial gene content (Figure S1I), and ribosomal protein gene content (Figure S1J) all indicated the high quality of the snRNA-seq data obtained from each individual donor. Cell clustering analysis revealed well-defined cell clusters, including the five major hormone-producing cell types as well as several non-endocrine cell types (Figure S1).

Chromatin accessibility analysis of human pituitary cell types

The same-sample paired assay snATAC-seq datasets generated approximately 11,000 DNA fragments per nucleus with an average transcription start site (TSS) enrichment score of 5.0 and a fraction of reads in called peak regions (FRiP) score of 47% (Tables S1B1 and S1B2). Cells were clustered and visualized using UMAP representation (Figure S1). Cell clusters were manually annotated based on chromatin accessibility (i.e., peaks of accumulated reads) at informative promoters among the same marker genes used for the RNA-seq annotation (see Table S2). Thyrotrope cells were too poorly represented to generate reliable chromatin tracks, across all samples consistent with their being the lowest-abundance endocrine cell type in the anterior pituitary (Ben-Shlomo et al., 2017). Similar to the sn transcriptome analysis results, cell clustering of the snATAC-seq data from each donor resulted in distinct cell clusters with all the major cell types being identified, although a thyrotrope cluster could not be distinguished in all male samples (Figure S1).

Cell type identification in snRNA-seq and snATAC-seq datasets

Integration of the snRNA-seq and snATAC-seq data from each sample was accomplished by label transfer from the snRNA-seq to the snATAC-seq data using the Seurat pipeline (Figure S1). The major pituitary cell type clusters were detected in all individual samples. Some clusters showed a gradient of expression and chromatin accessibility, resulting in their distinction as separate clusters, although they were not physically distinct (corticotropes in Figure S1F; somatotropes, gonadotropes, and stem cells in Figure S1G). Chromatin accessibility and transcript expression for specific established pituitary cell type markers are presented in Figure S2.

To improve the resolution of human pituitary cell types and to assess inter-individual variation, we merged same-sex snRNA-seq datasets and color-labeled them by donor (Figures 2A and 2C). Similarly, we merged the snATAC-seq data from same-sex samples, and labeled them by donor (Figures 2B and 2D). In addition to the five endocrine pituitary cell types, we identified stem cells, pituicytes, as well as pericytes, endothelial cells, and immune cells (macrophages, T cells, B cells). We observed donor-to-donor heterogeneity in cell type clustering in both datasets. For example, in males, separate gonadotrope,

somatotrope, and lactotrope clusters were noted in both RNA-seq and ATAC-seq data, originating almost exclusively from the pediatric sample. In females, one gonadotrope, one somatotrope, and one stem cell cluster were also derived from the pediatric sample. To determine whether these inter-individual differences might be the result of variations in sample processing or assay variability, we compared the results obtained with the pediatric sample processed using different assay protocols and on different days, with the results obtained with the other samples. As shown in Figure 2E, the three different assays of the pediatric female sample all gave similar UMAP projection results. These findings indicate that donor-to-donor differences cannot be attributed to batch effects or technical artifacts and represent inter-subject heterogeneity. The genes that are the most differentially expressed across these samples within lactotrope, gonadotrope, and somatotrope cells are shown as heatmaps in Figure S2.

The proportions of the major pituitary cell types identified by snRNA-seq versus snATAC-seq across all samples were highly correlated, indicating the agreement of main cell type assignment between the two assay modalities ($R^2 > 0.96$; Figure 2F). We also saw a similar distribution of the expression markers in human pituitary cell types in comparison with the same markers in an adult mouse pituitary dataset (Ruf-Zamojski et al., 2021). An sn multiome dataset and a same-sample sn paired dataset were generated from the pulverized pediatric female sample, further supporting the reliability of cell type assignment across the two assays (Tables S1 and S2). All datasets are publicly available and are accessible for exploration at snpituitaryatlas.princeton.edu.

Characterization of the PSC population

The stem cells identified among all samples by snRNA-seq (Figure 2) were re-clustered using the Seurat pipeline, leading to the detection of eight clusters (Figure 3A), with one well-separated cluster (cluster 6, circled in Figure 3A) and a large group formed by the remaining clusters (clusters 0–5 and 7). Cluster 6 corresponded to lineage-committed progenitor stem cells distinguished by GATA3 expression (Figure 3B, see discussion). The remaining large group of seven clusters expressed *SOX2, SOX9*, and the Hippo pathway effectors *WWTR1* (also known as *TAZ*) and *YAP1* (Lodge et al. [2019] and reviewed in Cox et al. [2017]; Figures 3B and S3A), which are indicative of uncommitted PSCs. The proliferation marker *MKI67* was not expressed in any of the PSC clusters (Figure 3B). Committing and likely committing cells expressing *POMC* and *POU1F1* were detected throughout the whole PSC cluster (Figure 3B); however, they did not form separate clusters on the UMAP (Figure 3A). Cluster 5 was composed of several committing and likely committing cells expressing either *POU1F1* or *POMC*, and displayed a higher mitochondrial gene expression (see discussion). Cluster 7 had lower UMI counts than all other PSC clusters (Figure S3B) and expressed several background genes. Because cluster 7 exhibited no apoptosis markers, additional samples would be necessary to further characterize this group of cells (Figure S3C, see discussion). The canonical stem cell markers *SOX2* and *SOX9*, as well as genes previously implicated in pituitary stem cell regulation (i.e., *WWTR1, PITX2*, and *LGR4*; for review, see Cox et al. [2017]), were broadly expressed.

We next examined the relationship of PSC clusters to the sex and age of donors (Figures 3C and 3D). The uncommitted stem cell clusters largely separated into samples of each sex, as confirmed by expression of the female-specific *XIST* (Loda and Heard, 2019). When the male and female datasets were grouped by age, all PSC subtypes were represented at all ages studied. The proportion of committing stem cells increased from early life to adulthood with 2.37%, 12.1%, and 9.13% in pediatric, adult, and aged samples, respectively. The proportion of likely committing stem cells showed a more pronounced age-related pattern, with 7.27%, 14.8%, and 45.6% in pediatric, adult, and aged samples, respectively (Table S3A; see discussion). Although the increase in committing stem cell proportions with age could have arisen from tissue heterogeneity and assaying only a small tissue fragment, the same analysis in mice revealed a comparable trend (Table S3B), thus supporting our findings.

We also compared the patterns of marker gene expression in human and mouse PSCs. As expected, we detected *Sox2, Sox9, Wwtr1*, and *Yap1* across the stem cell population of both species (Figures 3B and S3A). Expression of *POU1F1* in human samples was detected in a proportion of the cells in the main PSC cluster (Figure 3B), indicating the presence of committing progenitors among this population, as seen with *Pou1f1* in mice. Similar observations were made with *PAX7/Pax7* (Figure S3A), a determinant of melanotrope/intermediate lobe identity (Budry et al., 2012)). Additional markers that were either previously reported in mouse PSCs or linked to stemness, were also found in human stem cells, including *WIF1* (Poggi et al., 2018; Schluter et al., 2013), *HES1, NOTCH2* (Cox et al., 2017), *SMAD4* (Karlsson et al., 2007), and *SMAD5* (Avery et al., 2010; Kandyba et al., 2014) (Figure S3A).

Interestingly, expression of *JUN* and *JUND*, which have been implicated in the regulation of stemness in other tissues (Pagin et al., 2021; Semba et al., 2020), was heterogeneous, with the highest expression associated with uncommitted clusters consisting mostly of male cells (Figures 3B and 3C). *JUN* has not been proposed as a PSC marker but was recently reported to be enriched in SOX2-positive cells through bulk RNA-seq in mice (Russell et al., 2021). We therefore examined whether *Jun* showed co-expression with the stem cell marker *Sox2* by mRNA *in situ* hybridization in male neonatal, juvenile, and adult mouse pituitaries. The analysis confirmed *Jun* as a stem cell marker by identifying *Jun-Sox2* double labeling at all ages (Figures 3E and S3D). No apparent sex-specific differences were observed in mice, as similar results were obtained for females (data not shown). Additional genes found to be differentially expressed in male and female PSCs are shown in Figure S3E.

Overall, characterization of the heterogeneity of the PSC population from humans and mice supports the existence of different subtypes of uncommitted stem cells that are distinguishable from early committing lineages (see Haston et al. [2018]).

The acquisition of both snRNA-seq and snATAC-seq datasets from the same samples provides high-resolution analysis of the chromatin accessibility pattern of key genes within each pituitary cell type and reveals potential regulatory domains (see Ruf-Zamojski et al. [2021]). Moreover, we analyzed gene expression and chromatin accessibility in the same cells using the sn multiome dataset generated from the pediatric female sample. Presented

in Figure 3F are gene expression and chromatin accessibility for the stemness marker *SOX2* and the putative gonadotrope/thyrotrope committing cell lineage marker *GATA3* in the established pituitary cell types, including the stem cells and the committing stem cells. *SOX2* is expressed in stem cells and committing stem cells, which both showed the highest promoter accessibility. *GATA3* was expressed in gonadotropes and thyrotropes in addition to the committing stem cell lineage. All three cell types also showed increased chromatin accessibility in the *GATA3* promoter region. Other putative *cis*-regulatory domains in the *SOX2* and *GATA3* genes show increased chromatin accessibility in cell types that express these markers. In a subsequent section, we further elucidate the regulatory control of these and additional key PSC markers by modeling same-cell sn multiome data.

Diversity of PSC epigenetic programs

We next studied coordinated gene expression and chromatin accessibility programs in PSCs. We used the Pathway Level Information ExtractoR framework (PLIER), which identifies sets of genes (latent variables [LV]) that change together across cell types or samples, deconvolves datasets into these LV gene sets using known pathways, and associates them with biological processes, while not enforcing the strict orthogonality required for principal-component analysis (Mao et al., 2019). PLIER analyses identified both RNA and chromatin accessibility LVs that showed preferential activity in each major pituitary cell type, and those LVs are depicted as heatmaps to provide a summary of gene expression (RNA LV) or chromatin accessibility (ATAC LV, Figure S4). One RNA LV was stem-cell-specific and highly expressed in both sexes at all ages (Figure 4A, top, for the highest weighted 200 genes; Figure 4C for the highest weighted 30 genes; Data S1). Projection of this PSC LV onto adult mouse pituitary snRNA-seq data (Ruf-Zamojski et al., 2021) showed conservation of this PSC transcriptome program in mice (Figure 4B, top; Data S2). To determine whether this program was also associated with altered chromatin structure in PSCs, we projected this LV onto the human snATAC-seq data using the promoter accessibility signals as the gene features. The projection showed that the LV transcriptome program was associated with increased chromatin accessibility at the corresponding gene promoters (Figure 4A, bottom; Data S3). Similarly, projection of the PSC LV onto the adult mouse pituitary snATAC-seq data showed an association with increased promoter accessibility (Figure 4B, *bottom*; Data S4). Analysis of the snATAC-seq data identified a number of largely distinct accessibility programs that were each most strongly activated in subjects of different ages or sex (Figures S4 and S5). The complexity of PSC chromatin programs identified in this analysis may be related to the diversity of the donors (see discussion).

In addition to cell-type-specific LVs, we also identified one chromatin accessibility LV showing significantly decreased activity with increasing donor age ($LVage_{atac}$). Figure 5A presents a heatmap of the 30 highest weighted promoters comprising this LV in each sample by cell type. When the activity level in each cell type was plotted separately by age and sex, all cell types showed a decrease in promoter accessibility with age, especially between the pediatric and aged samples (Figure 5A). The decrease in accessibility was, however, less pronounced in males (pink lines in Figure 5B). Interestingly, this age-related LV was identified in a pathway that annotates cell cycle genes such as *ORC4* and *EIF4E*, which showed decreased chromatin accessibility in cell types from the older subjects (Figures

5A and 5C). Our results suggest that this accessibility program represents age-associated coordinated changes.

To further explore the relationship of PSC transcriptomes in samples with ages, we constructed a pseudotime trajectory from same-sex snRNA-seq datasets using the Monocle algorithm (Trapnell et al., 2014) (Figures 5D and 5E). In females as well as in males, the region of the graph most densely occupied by the pediatric PSCs was chosen as the root of the trajectory. In both sexes, pediatric PSCs formed the largest group, which separated from the adult and aged PSCs. This separation shows the large differences between PSC transcriptomes from pediatric and adult samples and also suggests that we have not captured all transitional stages of stem cells in the samples analyzed. To specify sets of genes that are dynamically regulated as cells progress along the trajectory, we identified several correlated gene modules per age group in females and in males (Figure 5F). The top genes in the most significant modules and the trajectories of selected genes are shown in Table S3B and Figure S6, respectively. Notably, almost none of the moduledefining transcripts were previously reported as PSC markers, and their roles in PSC physiology over the lifespan are not known. Overall, these analyses show an aging-related chromatin program in PSCs, and dynamic changes in the PSC transcriptome with aging.

Transcription factor and epigenetic control mechanisms of PSC genes

Using snRNA-seq and snATAC-seq datasets obtained from the same mouse pituitaries, we recently reported that chromatin accessibility is a key determinant for cell type transcriptional programs (Ruf-Zamojski et al., 2021). In comparison with same-sample datasets, same-cell sn multiome data confer vastly greater statistical power for inferring the regulatory mechanisms underlying the expression of specific genes (Cao et al., 2018; Ma et al., 2020). The matched transcriptome and chromatin accessibility data in same-cell sn multiome assays allow the co-variation of chromatin accessibility and gene expression to be modeled in thousands of individual cells. In addition, not all cells within a cell type express the same transcripts. Same-cell sn multiome data have the potential to provide insight into transcription factors (TFs) and epigenetic mechanisms that shape heterogeneous gene expression within the same pituitary cell type.

To explore the role of alterations in TF expression and chromatin state in modulating key PSC genes, we applied a linear modeling computational framework to the 15,024 nuclei in the same-cell sn multiome dataset obtained from a pediatric female pituitary. For each target gene, the linear model selects potential *cis*-regulatory regions comprising ATAC promoter peaks as well as co-accessible distal peaks. Then, testing the co-expression of putative *trans*-acting regulatory factors that have predicted TF binding sites in the co-accessible regions, linear regression identifies the TFs and chromatin regions that are most significantly predictive of the target gene expression. The linear model, when used to analyze all pituitary cells (“pan pituitary cell” analysis), can infer the mechanisms and factors implicated in cell-type-specific expression. When only cells comprising one pituitary cell type are analyzed, the linear model can generate hypotheses for the mechanisms responsible for differential expression among the different cells comprising this lineage.

We first analyzed the committed progenitor markers *POMC, POU1F1, TBX19*, and *NR5A1*. The output of the linear model pipeline is the p value that each selected *cis*-regulatory region and each individual TF with sites in that region contribute to the expression of the target gene (Figures 6 and S7). The TFs contributing to cell-type-specific expression of these marker genes using pan pituitary cell analysis included many factors that were previously implicated in the differentiation of committing stem cells. For example, the *POMC* analysis identified TBX19, which is an inducer of the *POMC*-expressing corticotrope/melanotrope lineage and of *POMC* expression (Pulichino et al., 2003). *TCF7L2*, which was highly significant in the analyses of *POU1F1* and *TBX19*, is an effector of the WNT signaling pathway, which regulates pituitary growth and development (Brinkmeier et al., 2003). Similarly, *LEF1*, another mediator of WNT signaling (see Kioussi et al. [2002]), was also identified in the *TBX19* analysis. Estrogen receptor alpha was the most significant TF implicated in *NR5A1* expression. Consonant with this finding, a recent study in murine gonadotrope cell lines demonstrated that estrogen-dependent binding of this nuclear receptor to an identified enhancer region triggers *Nr5a1* expression during gonadotrope lineage specification (Pacini et al., 2019).

We next studied the stemness marker *SOX2* and the committing cell lineage marker *GATA3*. When all pituitary cells were examined, the expression of *SOX2* within the overall stem cell subtype was associated with highly significant co-accessible proximal, upstream, and downstream regulatory domains (Figure 6A, left) as well as expression of TFs mapping to these domains (Figure 6A, right). These results indicate that PSC-specific expression of *SOX2* depends on a pattern of chromatin accessibility of regulatory domains present within these cells as well as expression of the requisite regulatory factors interacting with these domains. A contrasting result was obtained when applying the linear model to only PSCs to infer the regulatory circuits involved in heterogeneous expression of *SOX2* within PSCs. In this analysis, *cis*-regulatory domains correlated poorly with *SOX2* expression (Figure S7E, left), and a restricted set of regulatory factors (Figure S7E, right) was implicated in the heterogeneous pattern of *SOX2* expression in PSCs. These results suggest that the chromatin structure is sufficient for *SOX2* expression in all PSCs and the expression within specific PSCs depends on the expression of key regulatory TFs.

When performing a pan pituitary cell analysis for *GATA3*, we observed a pattern consonant with that of *SOX2*, with both chromatin structure and regulatory factor expression being responsible for expression in PSCs (Figure 6B). However, contrary to *SOX2*, heterogeneous *GATA3* expression within PSCs was associated with *cis*-regulatory chromatin accessibility domains, but not with expression of specific regulatory factors (Figures 6C and 6D). These results suggest that the regulatory proteins needed for *GATA3* expression in PSCs are expressed in all of the cells, and the heterogeneous expression pattern within PSCs is determined by differences in chromatin accessibility of regulatory domains between *GATA3*-expressing and non-expressing PSCs (Figure 6C). The *cis*-regulatory domains that permit *GATA3* expression in stem cells appear to be near *GATA3*.

The TF showing the second highest significance in the pan pituitary cell analysis of *POMC* is MNX1 (Figure 6E), an important homeobox gene previously implicated in motor neuron, pancreas, and lymphoid cell development (Harrison et al., 1994; Li et al., 1999; Vult

von Steyern et al., 1999). Therefore, MNX1 is an intriguing new candidate transcriptional regulator in the commitment toward the corticotrope/melanotrope lineages. In addition to the identification of novel putative TF regulators, when applied to all pituitary cells, the model also specifies the proximal and distal regulatory sites significantly associated with expression of the target gene in the cell types expressing that gene. These analyses identify previously unexplored regulatory domains in these key PSC genes that show accessibility associated with gene expression and are therefore *cis*-regulatory domain candidates (Figures 6 and S7). We confirmed experimentally the colocalization of *Mnx1* and *Tbx19* in adult mice, highlighting the identification of *Mnx1* as a marker of melanotrope/corticotrope lineage progenitor cells (Figure 6F). This validation demonstrates the value of the dataset and linear model for hypotheses generation. The high significance obtained in the linear model analysis for transcriptional regulators that were reported in previous research suggests that identified candidates warrant consideration for future study. When *POMC* was analyzed only in PSCs, the most significant TFs identified were E2F4 (Hsu et al., 2019) and TBX19, while the co-accessible regulatory regions were of low significance (Figure S7D). These results suggest that differential *POMC* expression in committing PSCs versus uncommitted PSCs is due to expression of these key TFs more so than to alterations in chromatin accessibility at key regulatory regions.

In addition to the identification of putative TF regulators, when applied to all pituitary cells, the model also specifies the proximal and distal regulatory sites that are significantly associated with target gene expression in the cell types expressing that gene. Analyses of these key PSC genes identify previously unexplored regulatory domains that show accessibility associated with gene expression and are therefore *cis*-regulatory domain candidates (Figures 6 and S7).

DISCUSSION

We generated high-quality snRNA-seq and snATAC-seq datasets from individual human pituitaries, demonstrating the feasibility of sn profiling in frozen *postmortem* samples that had been stored at −80 °C for as long as 2 decades. Although *postmortem* molecular profiling has been an issue for some tissues, numerous studies have demonstrated that the brain and the pituitary are robust to *postmortem* changes and long-term storage (Andreasson et al., 2013; Ferreira et al., 2018; White et al., 2018; Zhu et al., 2017), aside from a few limitations (see end of discussion). The present study is not the first to use archived *postmortem* samples (Amamoto et al., 2020; Habib et al., 2017; Mathys et al., 2019). Herein, we uncover putative mechanisms of lineage regulation and provide insight into PSC heterogeneity and the regulatory mechanisms and circuits underlying the expression of key PSC genes. We distinguish and characterize uncommitting and committing stem cell lineages, along with differences related to donor, age, and sex. We reveal a high level of donor-to-donor heterogeneity in cell type proportions that may stem from the use of *postmortem* samples stored at −80 °C and our small sample size. In addition, as only a fragment of each pituitary was analyzed in our study, we cannot exclude cell type regionalization as a factor in the inter-individual variation of cell type proportions.

The pituitary has been neglected from large sc consortium studies due to the difficulty of assaying fresh samples. Recent studies have focused on the analysis of human pituitary tumors (Ben-Shlomo et al., 2020; Neou et al., 2020). Surprisingly, one study has questioned the specificity of *SF1/NR5A1* as gonadotrope lineage marker (Neou et al., 2020). However, *NR5A1* has been classified as a gonadotrope-specific marker in tumors (Asa et al., 1996; Llyod et al., 2017; Trouillas et al., 2020), and in our work, *NR5A1* is restricted to gonadotropes both at the RNA and chromatin accessibility levels. A previous study reported the presence and lineage tracing of Nestin+ cells in the adult mouse pituitary using transgenic lines (Gleiberman et al., 2008), while the fidelity of the expression pattern of the Nestin-Cre transgene in the pituitary gland has been challenged by others (Galichet et al., 2010). The lineage tracing approach used by Gleiberman's study is not optimal, as Nestin is expressed across various cell populations in the mouse pituitary and thus is not unique to mouse PSCs. The authors indeed reported expression of the Nestin transgene in a sizable proportion of committed cells prior to fate mapping. Furthermore, our data reveal that *NESTIN* is not expressed robustly in the human pituitary and hence is not a suitable PSC marker. We have used markers from recent studies, such as SOX2 (Andoniadou et al., 2013) and SOX9 (Rizzoti et al., 2013). Those studies demonstrated that SOX2/SOX9+ PSCs give rise to all the endocrine lineages, yet the two markers do not mark committed cells, and the genetic tools faithfully reproduce endogenous gene expression. Here, we demonstrate the robust expression of both genes in the human PSC cluster (Figure 3D).

Reclustering of the stem cells identified by snRNA-seq data analysis distinguishes one well-separated cluster that is consistent with committing stem cells (see Figure 3A). Characterizing the identity of cluster 7 requires further investigation of additional samples. This cluster had much lower UMI counts than all other PSC clusters (Figure S3B), yet it did not correspond to dying apoptotic cells (Figure S3C) and seemed driven by a higher level of background genes. Interestingly, although we detected committing or likely committing cells within the main PSC cluster, those did not form separate sub-clusters. The cells were distinguished based on low, but above ambient level counts for a few key determinant genes for commitment. Cluster 5 in particular comprised a higher proportion of cells expressing *POMC* or *POU1F1* (Figure 3B). *POMC*-expressing cells are likely precursors of the corticotrope/melanotrope lineages (Budry et al., 2011), whereas *POU1F1*-expressing cells represent PSCs with the potential to commit to the somatotrope, lactotrope, and thyrotrope lineages (Li et al., 1990). Noticeably, cluster 5, which had many cells transitioning toward differentiated cell types, also showed higher mitochondrial gene expression, yet within the standard mitochondrial range for analysis. These cells were not dying and were within the mitochondrial range to be considered for analysis. This higher mitochondrial gene content relative to the other PSC clusters raises the possibility of a change in metabolic status during the transition to committed cells (Garcia-Prat et al., 2017; Tsogtbaatar et al., 2020; Yan et al., 2021). The *GATA3*-expressing cluster presumably comprises cells that are committing to the gonadotrope or thyrotrope lineages (Charles et al., 2006), although low *NR5A1* expression precludes ultimate cell type lineage assignment (Figure S3). All three committing lineages were identified in both sexes. The uncommitted stem cells, which formed six clusters that were not well separated, showed distinct but overlapping spatial localization between male and female samples. The proportion of committing stem cells in each sample

varied with age. Although our mouse data support these findings (Table S3), the differences observed in the human samples could still be linked to tissue heterogeneity, as we analyzed only a fragment of each pituitary, with a single sample per age group. To strengthen our claim, the study of additional samples from donors of different ages is warranted.

Our analyses identify PSC transcriptome and epigenetic programs as well as age-related differences in PSCs. One RNA-seq PSC LV program that is well expressed in all assayed samples is conserved in mice and associated with PSC-specific chromatin changes at the promoters for the genes comprising this LV. We also identify an ATAC-seq LV that exhibits increased accessibility with age in all pituitary cell types but shows smaller changes with age in PSCs. When the snRNA-seq data were subjected to a trajectory analysis, PSCs from the pediatric samples were separated from the adult and aged samples in both sexes. We find that human and mouse PSCs share similar patterns of gene expression and are characterized by the presence of several subtypes of uncommitted and committing cells, suggesting a high degree of conservation of this cell type in evolutionary time; however, the identification of species-specific genes might signify inter-species or age-related differences, with potential implications for use of the mouse in therapy development.

We have compared our results with a recent reported study of fetal PSCs (Zhang et al., 2020). Both studies show key common stem cell regulators, although our study detects many more PSC genes, possibly because more cells were analyzed. Additional high-resolution fetal and juvenile pituitary studies are warranted to further compare the PSC composition throughout development.

The PSC LV programs, and the complete separation of PSCs across different ages in the trajectory analysis, indicate that fully characterizing the changes in PSC transcriptional and epigenetic programs with aging will require analysis of additional samples over the age span. Collectively, our data suggest that sex and age influence several biological processes in stem cells. Further investigation is needed to elucidate those sex and age differences, as they may impact the development of new stem-cell-based therapies.

The datasets generated in this work encompass all major cell types in the human pituitary. Reliability of the identification of all major cell types in the same-sample snRNA-seq and snATAC-seq datasets from both sexes and from a range of ages is supported by the concordance of cell type proportions obtained by both assays and the confirmation of cell type identification in the same-cell sn multiome data. We report gene expression and chromatin accessibility LVs that are characteristic of each major pituitary cell type. Extensive data from the female pediatric pituitary are provided by multiple same-sample datasets and a large same-cell sn multiome dataset. With the exception of the pediatric female sample, only a fragment of each pituitary was analyzed in each assay. Finer resolution of the changes that occur in the transcriptome and epigenome within pituitary cell types over the age span will require analysis of additional samples. The present datasets represent a resource to address questions about the characterization and regulatory mechanisms of any cell type in the human pituitary.

Inferences from the same-cell sn multiome dataset using the linear model provide insight into the TFs and accessible chromatin sites contributing to the expression of key PSC genes. The model also provides insight into the general mechanisms (TF expression, chromatin accessibility differences, or both) responsible for differential expression of the target PSC genes among different cells. Because the model is based on detection of regulatory feature correlation with target gene expression, the results obtained when PSC genes are analyzed among all pituitary cells represent mechanisms implicated in target gene expression in PSCs in comparison with other cell types. When the model is applied only to PSCs, the results represent deterministic mechanisms for the heterogeneous expression of these target genes within a subset of PSCs. For the PSC and committed progenitor markers analyzed (*POMC, POU1F1, TBX19, NR5A1, SOX2,* and *GATA3*), multiple chromatin accessibility sites and TFs predicted to bind to accessible sites are identified with high probability as contributing to stem cell expression of these markers in comparison with other cell types. This supports the formulation that the expression of each of those markers in PSCs depends on epigenetic remodeling of chromatin as well as on expression of key TFs that are necessary for driving gene expression.

Analysis of same-cell sn multiome data from the pediatric female pituitary suggests that a diversity of mechanisms contributes to differential expression of marker genes among PSCs. With respect to differential expression of PSC markers within PSCs, *NR5A1* and *POU1F1* show TFs and chromatin sites associated with heterogeneous expression. *POMC, TBX19*, and *SOX2* are associated with the expression of specific TFs and *GATA3* with accessibility of specific regulatory sites. The pan pituitary analysis shows the importance of both TF expression and chromatin structure in the expression of key PSC genes. However, these analyses of heterogeneous expression within PSCs suggest that the differential expression of some markers is predominantly determined by expression of key TFs in those cells, whereas the differential expression of other markers depends on heterogeneity in chromatin structure. These results suggest that in addition to stochastic mechanisms that have been described to explain heterogeneity in gene expression, different deterministic mechanisms involving predominantly chromatin state or TF expression can now be resolved.

A strength of this study is the multiomic profiling of the entire pituitary at postnatal stages, which we use to generate a map of the variation of PSCs between sexes and with development and aging. Although our characterization of the changes associated with aging is limited to the analysis of only two aged donors, the identification of an aging-associated signature in both male and female aged donor datasets suggests that some effects may be attributable to aging. Our study profiles human PSCs in several human pituitary samples of different ages and proposes mechanisms for differential marker gene expression within the same cell type. We demonstrate the power of same-sample and same-cell multiomics analyses to further elucidate the mechanisms underlying PSC cell type and cell state in both sexes and different age groups. All data and analyses are accessible at snpituitaryatlas.princeton.edu.

Limitations of the study

Readers should be aware of some limitations of our study caused by the difficulty of studying human pituitaries. We used *postmortem* samples stored at −80 °C in a biobank, which could result in storage and *postmortem* effects on the data, although all of our samples were assessed for quality and generated high-quality datasets. Our assays were performed on pituitary fragments, which may result in cell sampling limitations. In addition, because only a single sample was analyzed per sex and age group, findings about differences between groups warrant further studies.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr Frederique Ruf-Zamojski (frederique.ruf-zamojski@mssm.edu).

Materials availability—This study did not generate new unique reagents.

Data and code availability

- The datasets (snRNAseq, snATACseq, sn multiome) generated in the present study are deposited in GEO (accession # GSE178454) and are publicly available as of date of publication. The sn human pituitary multi-omics atlas can be browsed via a web-based portal accessible at snpituitaryatlas.princeton.edu. All datasets will also be deposited with the Human Cell Atlas. Accession numbers and web-portal access are also listed in the key resources table.
- All original code has been deposited in Github and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sample procurement

Human: Flash-frozen *post-mortem* human pituitaries were obtained from the National Institutes of Health (NIH) NeuroBioBank, and kept at −80C until processing. The tissues received varied from whole to pieces of pituitaries. A single pituitary per age/sex was used in the study. The following criteria were used to select samples: gender, absence of degenerative, neuroendocrine or endocrine disease, and coverage of a range of ages. See Table 1 for information on donor sex, age, ethnicity, *post-mortem* interval (PMI), cause of death, and year of collection, and the next section for ethical compliance.

Mice: Wildtype CD-1 murine postnatal pituitaries were dissected at postnatal age P3, P15, and P56. Mice were specific pathogen-free. Mice were socially housed in single-sex groups (maximum 5 per cage) in individually ventilated cages on a 12-hour light-dark cycle in a temperature- and humidity-controlled room. Mice were fed *ad libitum* on standard chow,

were given fresh tap water daily, and rehoused in clean cages weekly. Nesting materials were provided for environmental enrichment.

Sample size estimation: not applicable for determination of gene expression patterns.

Ethical compliance

Human: We have complied with all ethical regulations and institutional protocols for studying human *post-mortem* human samples. All human specimens obtained from the NIH Neurobiobank were frozen samples from deceased donors. Donor anonymity was preserved, and guidelines were followed regarding consent, protection of human subjects, and donor confidentiality. Prior to being shared with the tissue biobank, collection of the pituitary samples upon death was approved by IRB #HP-00042077 for University of Maryland and consent was obtained from all donors or next of kin. Study # 13-00709 PS for tissue repository at the Bronx VA Medical Center was determined Human Research Exempt by the IRB at the Icahn School of Medicine at Mount Sinai (ISMMS), as defined by DHHS regulations (45 CFR 46.101(b) (2). As these samples were obtained de-identified from deceased subjects from the NIH Neurobiobank, the study itself was not considered human subjects research by the IRB at the ISMMS, where the assays were performed.

Mice: Animal work was carried out in compliance with the Animals (Scientific Procedures) Act 1986 and King College London (KCL) Ethical Review approval.

METHOD DETAILS

Nuclei isolation from pituitaries—Two methods were tested for nuclei isolation. Frozen *post-mortem* human pituitaries were either: 1) broken into small pieces in a frozen mortar on dry-ice, and one piece was thawed on ice and prepared for nuclei extraction based on a modified protocol from Mathys et al. (2019), or 2) pulverized and part of the powder used for nuclei isolation. The remainder of the pituitary was stored back at −80C. Briefly, and all on ice, RNAse inhibitor (NEB cat# MO314 L) was added to the homogenization buffer (0.32 M sucrose, 1 mM EDTA, 10 mM Tris-HCl, pH 7.4, 5 mM $CaCl_2$, 3 mM $Mg(Ac)_2$, 0.1% IGEPAL CA-630), 50% OptiPrep (Stock is 60% Media from StemCell cat# 07820), 35% OptiPrep and 30% OptiPrep right before isolation. Each pituitary was homogenized in a Dounce glass homogenizer (1 mL, VWR cat# 71000-514), and the homogenate filtered through a 40 mm cell strainer. An equal volume of 50% OptiPrep was added, and the gradient centrifuged (SW41 rotor at 17,792xg; 4C; 25 min). Nuclei were collected from the interphase, washed, resuspended either in 1X nuclei dilution buffer for snATACseq (10X Genomics) or in 1X PBS/0.04% BSA for snRNAseq, and counted (Cellometer).

SnRNAseq assay—SnRNAseq was performed following the Single Cell 3′ Reagents Kits V3 User Guidelines (10x Genomics, Pleasanton, CA). Nuclei were filtered and counted on a Countess instrument. A minimum of 1,000 nuclei were targeted (Chromium Single Cell 3′ Chip kit A v2 PN-12036 or v3 chip kit B PN-2000060). Reverse-transcription (RT) was performed in the emulsion, cDNA amplified, and libraries constructed with v3 chemistry. Libraries were indexed for multiplexing (Chromium i7 Multiplex kit PN-12062).

SnATACseq assay—SnATACseq was performed following the Chromium Single Cell ATAC Reagent Kits V1 User Guide (10x Genomics, Pleasanton, CA). Nuclei were counted (Countess counter), transposition was performed in 10 μL at 37C for 60 min on at least 1,000 targeted nuclei, before loading of the Chromium Chip E (PN-2000121). Barcoding was performed in the emulsion (12 cycles) following the Chromium protocol. Libraries were indexed for multiplexing (Chromium i7 Sample Index N, Set A kit PN-3000262).

Sn multiome assay—Sn multiome was performed following the Chromium Single Cell Multiome ATAC and Gene Expression Reagent Kits V1 User Guide (10x Genomics, Pleasanton, CA) on part of the pulverized pediatric female sample. Nuclei were counted (Countess counter), transposition was performed in 10 μL at 37C for 60 min targeting 10,000 nuclei, before loading of the Chromium Chip J (PN-2000264) for GEM generation and barcoding. Following post-GEM cleanup, libraries were pre-amplified by PCR, after which the sample was split into three parts: one part for generating the snRNAseq library, one part for the snATACseq library, and the rest was kept at −20C. SnATAC and snRNA libraries were indexed for multiplexing (Chromium i7 Sample Index N, Set A kit PN-3000262, and Chromium i7 Sample Index TT, Set A kit PN-3000431 respectively).

Quality control (QC) and sequencing of libraries—Libraries were quantified by Qubit 3 fluorometer (Invitrogen) and quality was assessed by Bioanalyzer (Agilent). Equivalent molar concentrations of libraries were pooled and the reads were adjusted after sequencing the pools in a Miseq (Illumina). Libraries were then sequenced in a Novaseq 6000 (Illumina) at the New York Genome Center (NYGC) following 10X Genomics recommendations.

All libraries underwent detailed QC analysis. We included many criteria when assigning samples to the ‘Good/Pass/Fail’ categories. As is evident from the presented QC metrics (Tables S2), none of the samples failed. Moreover, it is noteworthy that all the libraries in the ‘Pass’ category are generally of very good quality, with just a single metric placing these outside the ‘Good’ category. Finally, the automated QC pipeline was followed by careful manual data assessment followed the automated QC pipeline to confirming acceptable data quality for each sample analyzed.

RNAscope mRNA *in situ* hybridization—Wildtype CD-1 murine postnatal pituitaries were dissected at P3, P15 (male), and P56 (male), and fixed in 10% neutral buffered formalin (Sigma) at room temperature for 16–24 hours. Samples were washed in PBS and dehydrated through graded ethanol series before paraffin-embedding as previously described (Russell et al., 2021). Samples were sectioned at 5 μm.

The RNAscope 2.5 HD Duplex assay (Advanced Cell Diagnostics) was used according to manufacturer’s recommendations, with the following specific probes: Mm-Jun (Cat# 453561, Advanced Cell Diagnostics), Mm-Sox2-C2 (Cat# 401041-C2, Advanced Cell Diagnostics), Mnx1 (Cat# 1063561-C2, ACDBio), and Gata3 (Cat#403321, ACDBio). Sections were counterstained with Mayer’s hematoxylin (Vector H-3404) and mounted with VectaMount Permanent Mounting Medium (Vector H-5000).

Automated QC pipeline for libraries—After sequencing of the samples, all libraries are processed through the 10x Genomics Cell Ranger pipeline to process the sequenced data (see sections below on Data Analysis). Next, we ran an internal QC pipeline to assess the quality of all datasets, and for each sn assay extracted important metrics from the summary CSV file generated through the Cell Ranger pipeline. For snRNAseq, we used the Seurat R package to calculate the percentage of cells with high mitochondrial read counts (>20%). Included was also a calculation of the percentage of cells with expression of *XIST*. If the result was greater than 0.1, the sample was determined as female. Otherwise, the sample was determined as male. For snATACseq, in addition to metrics from the CellRanger summary files, we used Seurat/Signac to calculate the present or absence of a nucleosome-free region peak and mononucleosome peak. Chr1-1-200000000 was used to generate the fragment histogram. Fragments that were less than 600bp were used to generate a density plot. If one of the highest two peaks in the plot is between 0 and 150bp, then the nucleosome free region peak was determined as TRUE. If the other highest two peaks in the plot were between 150 and 300 bp, then the mononucleosome peak was determined as TRUE. Nucleosome-free region peak and mononucleosome peak were used as additional reference metrics for additional quality assurance.

We determined criteria for both snRNA and snATAC datasets to classify them into Good, Pass, or Fail categories (Refer to Tables S1).

SnRNAseq data analysis—SnRNAseq data were processed using the Cell Ranger pipeline v5.0.0, and aligned to the Cell Ranger GRCh38 reference genome, introns included. Clustering and differential gene expression analysis were performed using Seurat v.3.9.9.9024 and standard procedures (Butler et al., 2018; Stuart et al., 2019). Top markers for each cluster were compared to known markers of pituitary cell types to annotate the clusters; a list of the most common genes associated to each cell type is provided in Table S2.

We used the t-SNE projection to identify the most common cross-type doublets, as well as apoptotic and low-count cells as t-SNE preserves the local structure of the data better than the UMAP projection. Doublet clusters appear as small, high UMI count satellites to the main clusters. We verified the nature of every such group of cell barcodes by plotting their gene expression of the top cell-type markers. By looking at which two gene expression programs are expressed in the barcodes composing each one of these satellite clusters, we were able to identify the two cell types that constitute the barcodes of these sub-clusters.

Apoptotic cells form their own clusters separate from the parent cluster. Several cell types often merge into a single apoptotic cluster, so that not every cell type will have its corresponding apoptotic cluster. These cells are characterized by low UMI counts and almost exclusively spliced mRNA reads, suggesting condensation of the nuclei and arrest of transcription of new mRNA.

Some cell-type clusters have offshoots composed of barcodes with low UMI counts. Contrary to apoptotic clusters, these have a similar ratio of intronic to exonic reads as their parent cluster and do not form their own cluster, but usually connect to, or appear very close

to, their parent cluster. These are probably experimental artifacts of slow mRNA capture. Their gene expression program is the same as that of their higher UMI counterparts, but they are more adversely affected by dropouts. As such, we decided to remove these low-UMI offshoots from downstream analysis, together with doublet barcodes and apoptotic cells.

SnATACseq analysis—SnATACseq data were processed using Cell Ranger-ATAC pipeline version 1.2.0, and aligned to the Cell Ranger-ATAC GRCh38 reference genome. Clustering was performed using Seurat/Signac versions 3.1.5/0.2.4 and standard procedures (Stuart et al., 2020). We produced chromatin accessibility tracks around known pituitary cell type marker genes and looked for promoter accessibility of these genes to annotate the clusters.

Doublets, low-count, and apoptotic cells were identified in the same manner as for snRNAseq data, except that for ATACseq data, the UMAP projection works better and was used instead. We used the number of fragments in peaks as an indicator as to whether a cell was healthy, a doublet, or low-count/apoptotic. Doublets were checked to possess fragments in peaks associated to the main markers of both cell types. In general, we have many fewer cell barcodes in the ATACseq data, so doublets are also less common. Consequently, few doublet clusters were identified.

Sn multiome analysis—We analyzed the pediatric female sample by sn multiome. We pooled together libraries from both GEM wells and ran the Cell Ranger ARC 1.0.0 pipeline on the pooled sample following 10x Genomics guidelines. Running the pipeline on each GEM well separately revealed that the samples have 97 barcodes in common that were called as cells.

We used Seurat version 3.9.9.9024 with Signac version 1.1.0 to perform our clustering analysis using a weighted shared nearest neighbor graph approach. This method identifies, for each cell, its nearest neighbors based on a weighted combination of the two modalities (Gene Expression & Chromatin Accessibility). We similarly used the weighted nearest neighbor graph to obtain a UMAP projection of the data. The Gene Expression modality was used to identify cluster cell types after determination of top markers for each cluster.

Apoptotic, low-count cells and doublets were also identified in a manner analogous to that of snRNAseq data. Both apoptotic and low-count cells were identified as having much lower counts of both their number of transcripts as well as their number of fragments overlapping ATACseq peaks. Apoptotic cells further have a large proportion of mitochondrial gene transcripts, whereas low-count cell transcripts are dominated by background genes. Doublets, on the other hand, were identified as having higher counts in both RNA and ATAC, and expressing gene programs of two cell types simultaneously. Only one of the main clusters was identified as doublets, and no further sub-clustering was attempted.

Merged datasets analysis—All male and female samples were merged by sex in Seurat at the UMI count level, and all of the clustering analysis was repeated on the merged samples independently from the beginning. We followed the same analysis steps as for

individual samples. Unlike our integrated samples (see the Sn data integration sub-section), merged samples do not have batch effects removed. Despite that, we do not observe any systematic batch effect between our samples. We do, however, see differences donor-to-donor differences in gene expression among some specific cell types. The merged samples allow us to highlight these differences in the implicated cell types.

Mouse snRNAseq stem cell data analysis—SnRNAseq data for each mouse were processed and analyzed as previously described (Ruf-Zamojski et al., 2021). Raw reads from each mouse sample were isolated from the clusters assigned in Seurat as 'Stem cells' using the 'WhichCells' function. These count tables were integrated using Seurat (v3.1.5) SCTransform workflow (Hafemeister and Satija, 2019), clustered at 0.5 resolution and principal component dimensions 1:15 were taken forward for analysis.

Human stem cell re-clustering method—Following initial clustering of the complete datasets, the 'stem cell' clusters were isolated from each individual donor using the Seurat 'subset' function. To increase the number of cells available for downstream analysis, the isolated stem cell datasets were merged based on the approximate age of donors. This was performed using the merge function within Seurat (v3.1.5). Sample integration by identification of anchors and subsequent clustering (20 PCAs, resolution 0.5) was performed using Seurat according to standard procedures (Butler et al., 2018; Stuart et al., 2019). Re-clustering and analysis leading to identification of 'committing' stem cells was done as above following removal of the 'Pars Tuberalis' cell clusters.

Pseudotime sn trajectory analysis—Raw gene counts were extracted from each sample's "Stem Cell" cluster as previously identified using Seurat (v4.0.1) (Hao et al., 2020). Due to sex differences, male and female samples were handled separately. The total number of stem cells analyzed per sample is presented in Table S3C. Same-sex samples of the same sex (e.g pediatric, adult, and aged females), were integrated using the functions 'SCTransform' and 'SelectIntegrationFeatures' in Seurat to obtain the top 500 differentially expressed genes (DEGs). Monocle3 (v0.2.3.0) (Cao et al., 2019) was used for pseudotime trajectory analysis. Preliminary analysis revealed a bias in the Monocle trajectory due to specific hormonal genes, namely, *GH1, PRL, CHGB, POMC, LHB, FSHB,* and *CGA*. Therefore, these genes were regressed out of the top 500 DEGs. Monocle objects were generated by combining the 3 same-sex using the respective top 500 DEGs. The trajectory was calculated by merging partitions with the root chosen based on the earliest time point available for each sex. To find gene modules changing over pseudotime, the '*graph_test*' function was carried out using the neighbor_graph = "principal_graph" parameter with a resolution of 0.8 for 'find_gene_modules' function. The top 4 enriched modules for each age in each sex were highlighted and examined further because they showed the highest variability between age groups.

PLIER data analysis—To delve into the gene expression trends of assigned cell types across samples and data types, we treated each sc dataset as a collection of bulk datasets for given labeled cell types. Each cell type was then treated as a separate bulk measurement within each sample. For snATACseq data, peak counts for a given gene were generated by

selecting the peak closest to the transcription start site (TSS). These peak counts per gene were then collected into single bulk measurements for each cell type in each sample. We focused specifically on 6 relevant cell types in the pituitary: corticotropes, gonadotropes, lactotropes, somatotropes, stem/progenitor cells, and thyrotropes. For the snRNAseq dataset, this process generated 36 bulk measurements over 6 samples (3 females and 3 males), and for the snATACseq dataset, we generated 35 bulk measurements, as thyrotropes were not identified in the male adult snATACseq sample. We applied PLIER (Mao et al., 2019), which finds patterns in count data that are associated with known prior information (such as Reactome and KEGG), focusing on the 2000 genes with the highest standard deviation in count values across the bulk measurements in each set of samples. PLIER was run on each set of samples separately with LVs generated on the bulk measurements in an unsupervised fashion. LVs were then curated to find patterns relevant to individual cell types as well as sample-wide trends such as sex-based differences. Statistical significance of LVs was computed through the Kruskal-Wallis non-parametric test for multiple groups as part of the stat_compare_means R method. Comparisons between LVs within and across datatypes were achieved by comparing the overlap of the 200 genes most associated with a given LV.

B is a PLIER-derived expression value for the genes associated with a given LV across the different samples. It can be treated similar to average expression, weighted by gene association with the LV. Technically, B is a matrix of size #LVs x #Samples. It is one of two matrices in PLIER, along with Z of size # of genes x #LVs. The goal of PLIER is to find values of B and Z that minimize the equation ‖Y - Z*B‖ where Y is our data matrix of size #genes x #samples. So PLIER finds a suitable number of LVs that can be used to connect the genes and samples and accurately estimate our data matrix.

For the boxplot statistical analysis (Figure S4), ggboxplot generates a boxplot with the center equal to the 50th percentile, the bounds of the box are the 25th and 75th percentile and the bounds of the whiskers are the smallest/largest values 1.5 times the interquartile range below the 25th percentile or above the 75th percentile, respectively.

Sn data integration—The snRNAseq and snATACseq data were integrated in a reference-query based manner, mainly using the “FindTransferAnchors” and “TransferData” functions from the Seurat v3 package (Butler et al., 2018; Stuart et al., 2019). The snRNAseq datasets were used as the reference and the other modalities were integrated to them. To integrate snATACseq to snRNAseq, the peak-by-cell accessibility matrix was converted to a gene-by-cell activity matrix based on the chromatin accessibility within each gene’s gene body and a 2kb upstream region, under the assumption that chromatin accessibility and gene expression were positively correlated. The variable features from the snRNAseq data were used to find the anchors and the snATACseq data in the LSI low-dimensional embedding were used to transfer the data from snRNAseq to snATACseq.

Linear modeling analysis—The regulatory model of gene expression from the sn multiome dataset was constructed for each target gene by following several steps: 1) A list of potential regulatory genomic regions was selected. They included: a) any ATAC peaks that overlapped with the TSS +/− 2kb region, b) distal peaks that were no more than 500kb away from the TSS and were co-accessible with any of the peaks in a).

Co-accessibility scores were calculated using the Cicero package (Pliner et al., 2018) with default parameters, and a cutoff of 0.25 for the co-accessibility scores was used to select co-accessible peaks. 2) A list of potential regulatory TFs was selected by scanning for TF binding sites in the selected genomic regions using the "matchMotifs" function (with a p value cutoff of 5×10^{-5}) from the r package "motifmatchr" and the position weight matrices (PWMs) from the JASPAR CORE database. 3) Linear regression was used to model target gene's expression across cells as a function of selected TFs' expression and openness of ATAC peaks' openness, and the coefficients from the regression were used to measure the importance of each TF and genomic region. SCTranform (Hafemeister and Satija, 2019) normalized RNA counts and TFIDF normalized ATAC peak counts were used in the regression.

QUANTIFICATION AND STATISTICAL ANALYSIS

In Figure 5A, to calculate the statistical significance of expression or accessibility changes within a given LV, we applied two-way ANOVA for multiple group testing and Tukey test for pairwise comparisons. Each test was applied to female and male samples separately. In both cases, we applied the R statistical functions *aov* and *TukeyHSD* with the additive model *Expression ~ 'Cell Type' + 'Age Group'* for the calculations.

In Figure S4, the hierarchical clustering of the LV B scores was accomplished through the default complete linkage method utilized by the R function *pheatmap*.

For the boxplots analysis in Figure S4, the analysis was done with n = 3 independent subjects per sex and statistical analysis using the Wilcoxon ranked-sum test.

For Figures 6 and S7, the p values of peaks and the p values of the TFs were both obtained by running a linear regression ("lm" function in R) on 9,151 cells (for pan-pituitary results) and 1,623 cells (for stem cell-specific results). In addition, for TFs statistical analysis, the TFs were presented only if their Bonferroni-corrected p values were below 0.05.

ADDITIONAL RESOURCES

All data and analyses are available on the web portal: snpituitaryatlas.princeton.edu.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by funding from the National Institutes of Health (NIH) Grant DK46943(S.C.S.), Grant R01GM071966 (O.G.T.), Medical Research Council (MRC) Grant MR/T012153/1 (C.L.A.), and Canadian Institutes of Health Research Project Grants PJT-162343 and −169184 (D.J.B.). T.L.W. was funded by King's College London as part of the "Cell Therapies and Regenerative Medicine" Four-Year Welcome Trust PhD Training Program. We acknowledge the New York Genome Center for sequencing. Human tissue was obtained from the NIH NeuroBioBank. This work was supported in part through the computational and data resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

REFERENCES

Amamoto R, Zuccaro E, Curry NC, Khurana S, Chen HH, Cepko CL, and Arlotta P (2020). FIN-Seq: transcriptional profiling of specific cell types from frozen archived tissue of the human central nervous system. Nucleic Acids Res. 48, e4. [PubMed: 31728515]

Andoniadou CL, Matsushima D, Mousavy Gharavy SN, Signore M, Mackintosh AI, Schaeffer M, Gaston-Massuet C, Mollard P, Jacques TS, Le Tissier P, et al. (2013). Sox2(+) stem/progenitor cells in the adult mouse pituitary support organ homeostasis and have tumor-inducing potential. Cell Stem Cell 13, 433–445. [PubMed: 24094324]

Andreasson A, Kiss NB, Juhlin CC, and Hoog A (2013). Long-term storage of endocrine tissues at − 80 degrees C does not adversely affect RNA quality or overall histomorphology. Biopreserv. Biobanking 11, 366–370.

Asa SL, Bamberger AM, Cao B, Wong M, Parker KL, and Ezzat S (1996). The transcription activator steroidogenic factor-1 is preferentially expressed in the human pituitary gonadotroph. J. Clin. Endocrinol. Metab 81, 2165–2170. [PubMed: 8964846]

Avery S, Zafarana G, Gokhale PJ, and Andrews PW (2010). The role of SMAD4 in human embryonic stem cell self-renewal and stem cell fate. Stem Cells 28, 863–873. [PubMed: 20235236]

Ben-Shlomo A, Liu NA, and Melmed S (2017). Somatostatin and dopamine receptor regulation of pituitary somatotroph adenomas. Pituitary 20, 93–99. [PubMed: 27900635]

Ben-Shlomo A, Deng N, Ding E, Yamamoto M, Mamelak A, Chesnokova V, Labadzhyan A, and Melmed S (2020). DNA damage and growth hormone hypersecretion in pituitary somatotroph adenomas. J. Clin. Invest 130, 5738–5755. [PubMed: 32673291]

Brinkmeier ML, Potok MA, Cha KB, Gridley T, Stifani S, Meeldijk J, Clevers H, and Camper SA (2003). TCF and Groucho-related genes influence pituitary growth and development. Mol. Endocrinol 17, 2152–2161. [PubMed: 12907761]

Budry L, Lafont C, El Yandouzi T, Chauvet N, Conejero G, Drouin J, and Mollard P (2011). Related pituitary cell lineages develop into interdigitated 3D cell networks. Proc. Natl. Acad. Sci. U S A 108, 12515–12520. [PubMed: 21746936]

Budry L, Balsalobre A, Gauthier Y, Khetchoumian K, L'Honore A, Vallette S, Brue T, Figarella-Branger D, Meij B, and Drouin J (2012). The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. Genes Dev. 26, 2299–2310. [PubMed: 23070814]

Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol 36, 411–420. [PubMed: 29608179]

Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, Daza RM, McFaline-Figueroa JL, Packer JS, Christiansen L, et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science 361, 1380–1385. [PubMed: 30166440]

Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature 566, 496–502. [PubMed: 30787437]

Castinetti F, Davis SW, Brue T, and Camper SA (2011). Pituitary stem cell update and potential implications for treating hypopituitarism. Endocr. Rev 32, 453–471. [PubMed: 21493869]

Charles MA, Saunders TL, Wood WM, Owens K, Parlow AF, Camper SA, Ridgway EC, and Gordon DF (2006). Pituitary-specific Gata2 knockout: effects on gonadotrope and thyrotrope function. Mol. Endocrinol 20, 1366–1377. [PubMed: 16543408]

Cheung LYM, George AS, McGee SR, Daly AZ, Brinkmeier ML, Ellsworth BS, and Camper SA (2018). Single-cell RNA sequencing reveals novel markers of male pituitary stem cells and hormone-producing cell types. Endocrinology 159, 3910–3924. [PubMed: 30335147]

Cox B, Roose H, Vennekens A, and Vankelecom H (2017). Pituitary stem cell regulation: who is pulling the strings? J. Endocrinol 234, R135–R158. [PubMed: 28615294]

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update46 (Nucleic Acids Res.), pp. D794–D801. [PubMed: 29126249]

Ferreira PG, Munoz-Aguirre M, Reverter F, Sa Godinho CP, Sousa A, Amadoz A, Sodaei R, Hidalgo MR, Pervouchine D, Carbonell-Caballero J, et al. (2018). The effects of death and post-mortem cold ischemia on human tissue transcriptomes. Nat. Commun 9, 490. [PubMed: 29440659]

Galichet C, Lovell-Badge R, and Rizzoti K (2010). Nestin-Cre mice are affected by hypopituitarism, which is not due to significant activity of the transgene in the pituitary gland. PLoS One 5, e11443. [PubMed: 20625432]

Garcia-Prat L, Sousa-Victor P, and Munoz-Canoves P (2017). Proteostatic and metabolic control of stemness. Cell Stem Cell 20, 593–608. [PubMed: 28475885]

Gleiberman AS, Michurina T, Encinas JM, Roig JL, Krasnov P, Balordi F, Fishell G, Rosenfeld MG, and Enikolopov G (2008). Genetic approaches identify adult pituitary stem cells. Proc. Natl. Acad. Sci. U S A 105, 6332–6337. [PubMed: 18436641]

Griffiths JA, Scialdone A, and Marioni JC (2018). Using single-cell genomics to understand developmental processes and cell fate decisions. Mol. Syst. Biol 14, e8046. [PubMed: 29661792]

Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, Choudhury SR, Aguet F, Gelfand E, Ardlie K, et al. (2017). Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat. Methods 14, 955–958. [PubMed: 28846088]

Hafemeister C, and Satija R (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 296. [PubMed: 31870423]

Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y, Chen H, Wang J, Tang H, Ge W, et al. (2020). Construction of a human cell landscape at single-cell level. Nature 581, 303–309. [PubMed: 32214235]

Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zagar M, et al. (2020). Integrated analysis of multimodal single-cell data. Preprint at bioRxiv. 10.1101/2020.10.12.335331.

Harrison KA, Druey KM, Deguchi Y, Tuscano JM, and Kehrl JH (1994). A novel human homeobox gene distantly related to proboscipedia is expressed in lymphoid and pancreatic tissues. J. Biol. Chem 269, 19968–19975. [PubMed: 7914194]

Haston S, Manshaei S, and Martinez-Barbera JP (2018). Stem/progenitor cells in pituitary organ homeostasis and tumourigenesis. J. Endocrinol 236, R1–R13. [PubMed: 28855316]

Hsu J, Arand J, Chaikovsky A, Mooney NA, Demeter J, Brison CM, Oliverio R, Vogel H, Rubin SM, Jackson PK, et al. (2019). E2F4 regulates transcriptional activation in mouse embryonic stem cells independently of the RB family. Nat. Commun 10, 2939. [PubMed: 31270324]

Kandyba E, Hazen VM, Kobielak A, Butler SJ, and Kobielak K (2014). Smad1 and 5 but not Smad8 establish stem cell quiescence which is critical to transform the premature hair follicle during morphogenesis toward the postnatal state. Stem Cells 32, 534–547. [PubMed: 24023003]

Karlsson G, Blank U, Moody JL, Ehinger M, Singbrant S, Deng CX, and Karlsson S (2007). Smad4 is critical for self-renewal of hematopoietic stem cells. J. Exp. Med 204, 467–474. [PubMed: 17353364]

Kioussi C, Briata P, Baek SH, Rose DW, Hamblet NS, Herman T, Ohgi KA, Lin C, Gleiberman A, Wang J, et al. (2002). Identification of a Wnt/Dvl/beta-Catenin-> Pitx2 pathway mediating cell-type-specific proliferation during development. Cell 111, 673–685. [PubMed: 12464179]

Li S, Crenshaw EB 3rd, Rawson EJ, Simmons DM, Swanson LW, and Rosenfeld MG (1990). Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1. Nature 347, 528–533. [PubMed: 1977085]

Li H, Arber S, Jessell TM, and Edlund H (1999). Selective agenesis of the dorsal pancreas in mice lacking homeobox gene Hlxb9. Nat. Genet 23, 67–70. [PubMed: 10471501]

Llyod RV, Osamura RY, Kloeppel G, and Rosai J (2017). WHO Classification of Tumours of Endocrine Organs, 4th edn, 10 (IARC Publications).

Loda A, and Heard E (2019). Xist RNA in action: past, present, and future. PLoS Genet. 15, e1008333. [PubMed: 31537017]

Lodge EJ, Santambrogio A, Russell JP, Xekouki P, Jacques TS, Johnson RL, Thavaraj S, Bornstein SR, and Andoniadou CL (2019). Homeostatic and tumourigenic activity of SOX2+ pituitary stem cells is controlled by the LATS/YAP/TAZ cascade. Elife 8, e43996. [PubMed: 30912742]

Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, Ding J, Brack A, Kartha VK, Tay T, et al. (2020). Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell 183, 1103–1116 e1120. [PubMed: 33098772]

Mao W, Zaslavsky E, Hartmann BM, Sealfon SC, and Chikina M (2019). Pathway-level information extractor (PLIER) for gene expression data. Nat. Methods 16, 607–610. [PubMed: 31249421]

Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, Menon M, He L, Abdurrob F, Jiang X, et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. Nature 570, 332–337. [PubMed: 31042697]

ENCODE Project Consortium; Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A, et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583, 699–710. [PubMed: 32728249]

Morris SA (2019). The evolving concept of cell identity in the single cell era. Development 146. dev169748. [PubMed: 31249002]

Neou M, Villa C, Armignacco R, Jouinot A, Raffin-Sanson ML, Septier A, Letourneur F, Diry S, Diedisheim M, Izac B, et al. (2020). Pangenomic classification of pituitary neuroendocrine tumors. Cancer Cell 37, 123–134 e125. [PubMed: 31883967]

Pacini V, Petit F, Querat B, Laverriere JN, Cohen-Tannoudji J, and L'Hote D (2019). Identification of a pituitary ERalpha-activated enhancer triggering the expression of Nr5a1, the earliest gonadotrope lineage-specific transcription factor. Epigenet. Chromatin 12, 48.

Pagin M, Pernebrink M, Giubbolini S, Barone C, Sambruni G, Zhu Y, Chiara M, Ottolenghi S, Pavesi G, Wei CL, et al. (2021). Sox2 controls neural stem cell self-renewal through a Fos-centered gene regulatory network. Stem Cells 39, 1107–1119. [PubMed: 33739574]

Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, et al. (2018). Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. Mol. Cell 71, 858–871 e858. [PubMed: 30078726]

Poggi L, Casarosa S, and Carl M (2018). An eye on the wnt inhibitory factor Wif1. Front. Cell Dev. Biol 6, 167. [PubMed: 30574494]

Pulichino AM, Vallette-Kasic S, Tsai JP, Couture C, Gauthier Y, and Drouin J (2003). Tpit determines alternate fates during pituitary cell differentiation. Genes Dev. 17, 738–747. [PubMed: 12651892]

Rizzoti K, Akiyama H, and Lovell-Badge R (2013). Mobilized adult pituitary stem cells contribute to endocrine regeneration in response to physiological demand. Cell Stem Cell 13, 419–432. [PubMed: 24094323]

Ruf-Zamojski F, Zhang Z, Zamojski M, Smith GR, Mendelev N, Liu H, Nudelman G, Moriwaki M, Pincas H, Castanon RG, et al. (2021). Single nucleus multi-omics regulatory landscape of the murine pituitary. Nat. Commun 12, 2677. [PubMed: 33976139]

Russell JP, Lodge EJ, and Andoniadou CL (2018). Basic research advances on pituitary stem cell function and regulation. Neuroendocrinology 107, 196–203. [PubMed: 29539624]

Russell JP, Lim X, Santambrogio A, Yianni V, Kemkem Y, Wang B, Fish M, Haston S, Grabek A, Hallang S, et al. (2021). Pituitary stem cells produce paracrine WNT signals to control the expansion of their descendant progenitor cells. Elife 10, e59142. [PubMed: 33399538]

Schluter H, Stark HJ, Sinha D, Boukamp P, and Kaur P (2013). WIF1 is expressed by stem cells of the human interfollicular epidermis and acts to suppress keratinocyte proliferation. J. Invest. Dermatol 133, 1669–1673. [PubMed: 23358094]

Semba T, Sammons R, Wang X, Xie X, Dalby KN, and Ueno NT (2020). JNK signaling in stem cell self-renewal and differentiation. Int. J. Mol. Sci 21, 2613.

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, and Satija R (2019). Comprehensive integration of single-cell data. Cell 177, 1888–1902 e1821. [PubMed: 31178118]

Stuart T, Srivastava A, Lareau C, and Satija R (2020). Multimodal single-cell chromatin analysis with Signac. Preprint at bioRxiv. 10.1101/2020.11.09.373613.

Tanay A, and Regev A (2017). Scaling single-cell genomics from phenomenology to mechanism. Nature 541, 331–338. [PubMed: 28102262]

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, and Rinn JL (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol 32, 381–386. [PubMed: 24658644]

Trouillas J, Jaffrain-Rea ML, Vasiljevic A, Raverot G, Roncaroli F, and Villa C (2020). How to classify the pituitary neuroendocrine tumors (PitNET) s in 2020. Cancers 12, 514.

Tsogtbaatar E, Landin C, Minter-Dykhouse K, and Folmes CDL (2020). Energy metabolism regulates stem cell pluripotency. Front. Cell Dev. Biol 8, 87. [PubMed: 32181250]

Vankelecom H (2016). Pituitary stem cells: quest for hidden functions. In Stem Cells in Neuroendocrinology, Pfaff D and Christen Y, eds. (Springer), pp. 81–101.

Vult von Steyern F, Martinov V, Rabben I, Nja A, de Lapeyriere O, and Lomo T (1999). The homeodomain transcription factors Islet 1 and HB9 are expressed in adult alpha and gamma motoneurons identified by selective retrograde tracing. Eur. J. Neurosci 11, 2093–2102. [PubMed: 10336678]

White K, Yang P, Li L, Farshori A, Medina AE, and Zielke HR (2018). Effect of postmortem interval and years in storage on RNA quality of tissue at a repository of the NIH NeuroBioBank. Biopreserv. Biobank 16, 148–157. [PubMed: 29498539]

Yan W, Diao S, and Fan Z (2021). The role and mechanism of mitochondrial functions and energy metabolism in the function regulation of the mesenchymal stem cells. Stem Cell Res. Ther 12, 140. [PubMed: 33597020]

Zhang S, Cui Y, Ma X, Yong J, Yan L, Yang M, Ren J, Tang F, Wen L, and Qiao J (2020). Single-cell transcriptomics identifies divergent developmental lineage trajectories during human pituitary development. Nat. Commun 11, 5275. [PubMed: 33077725]

Zhu Y, Wang L, Yin Y, and Yang E (2017). Systematic analysis of gene expression patterns associated with postmortem interval in human tissues. Sci. Rep 7, 5435. [PubMed: 28710439]

Highlights

- Single nucleus multiomics characterization of human pituitary stem cells
- Pseudotime trajectory sets apart pediatric versus adult and aged pituitary stem cells
- Linear modeling infers transcription factors modulating stem cell gene expression
- Distinct regulatory mechanisms underlie pituitary stem cell gene expression

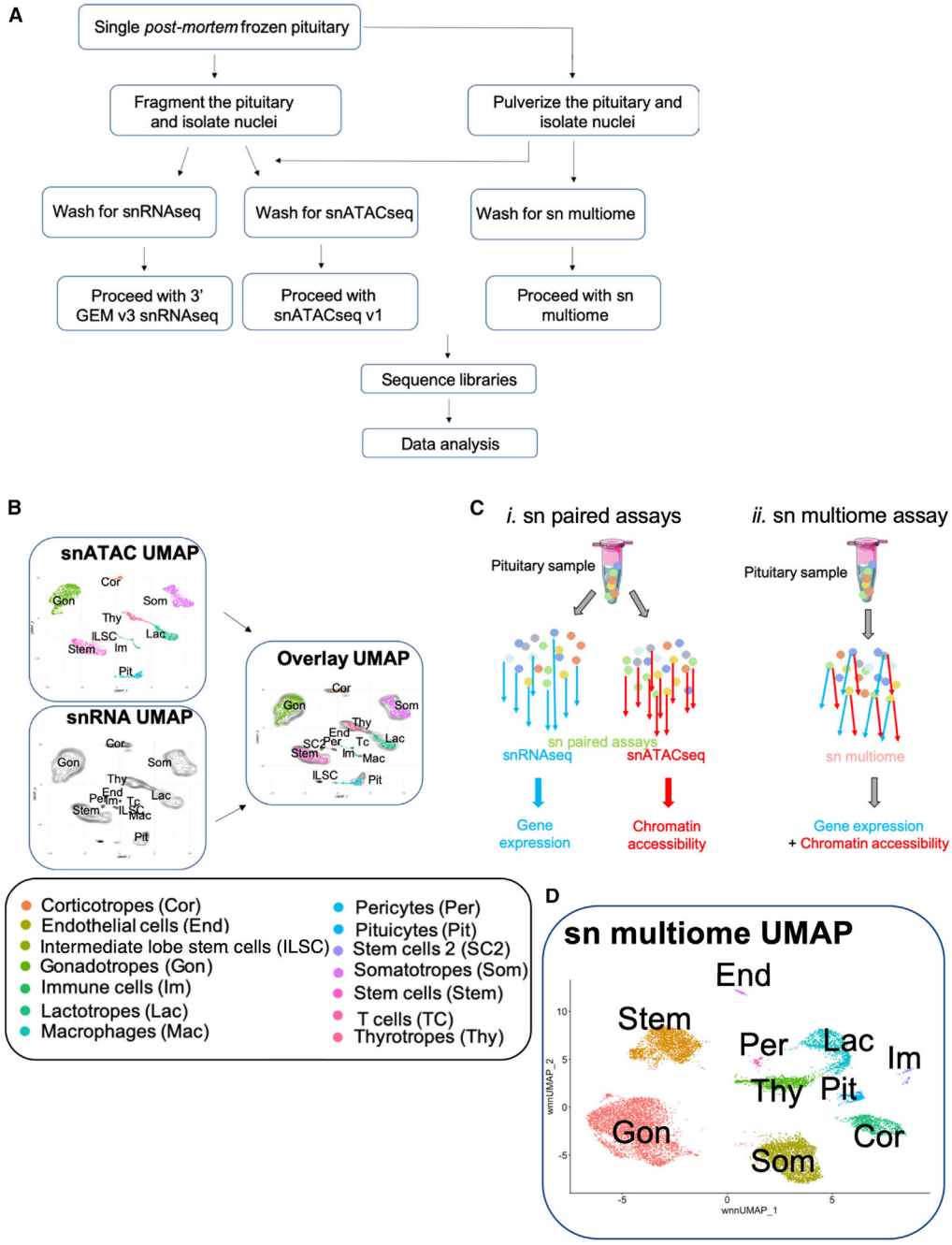


Figure 1. Experimental design for human pituitary cell type identification

(A) Schematic of the overall experimental workflow, from procurement of the frozen pituitaries to sn data analysis.

(B) Schematic summarizing sn data integration. For each sample, the snATAC-seq dataset (colored dot UMAP) was integrated with the snRNA-seq dataset (black contours UMAP) to generate an integrated multiomics overlay UMAP identifying cell types. On the UMAP, cell types are color-coded and designated with a two- to three-letter code, as indicated on the bottom key. The female pediatric pituitary sample is represented as an example.

(C) All integrated samples are presented in Figure S1. Schematic of the comparison between sn paired assays (same-sample sn multiomics) (*i*) and sn multiome assay (same-cell) (*ii*).
(D) Same-cell sn multiome UMAP from the female pediatric sample (see Table 1).

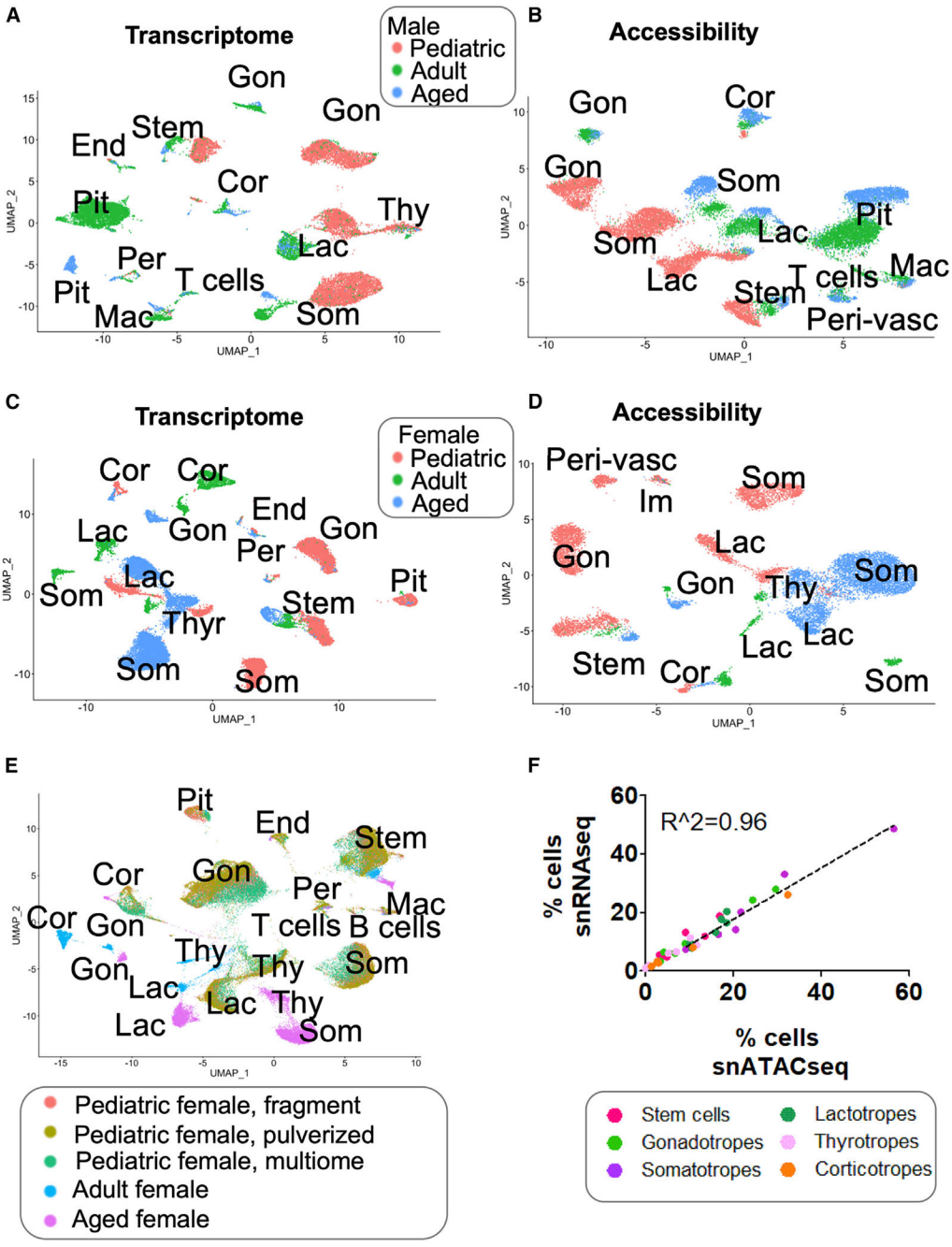


Figure 2. Merged analysis of same-sex human pituitaries

(A–E) UMAP representation of sn transcript expression (A, males; C, females) and of sn chromatin accessibility (B, males; D, females) in the merged same-sex samples, with labeling by donor age in each sex, and transcriptome by individual assays performed on the same donor. (E) Individual donors (A–E) or assays for a same donor (E) are color-coded as indicated. Each cluster is identified by a letter code as defined in Figure 1.

(F) Donor-related information is provided in see Table 1. Correlation between the cell type proportions identified by snRNA-seq versus snATAC-seq for all samples (males and females). Assays done for the pediatric sample in (F) were done on different days. The linear

regression is plotted. Pituitary cell types are color-coded and the key is provided underneath the plot.

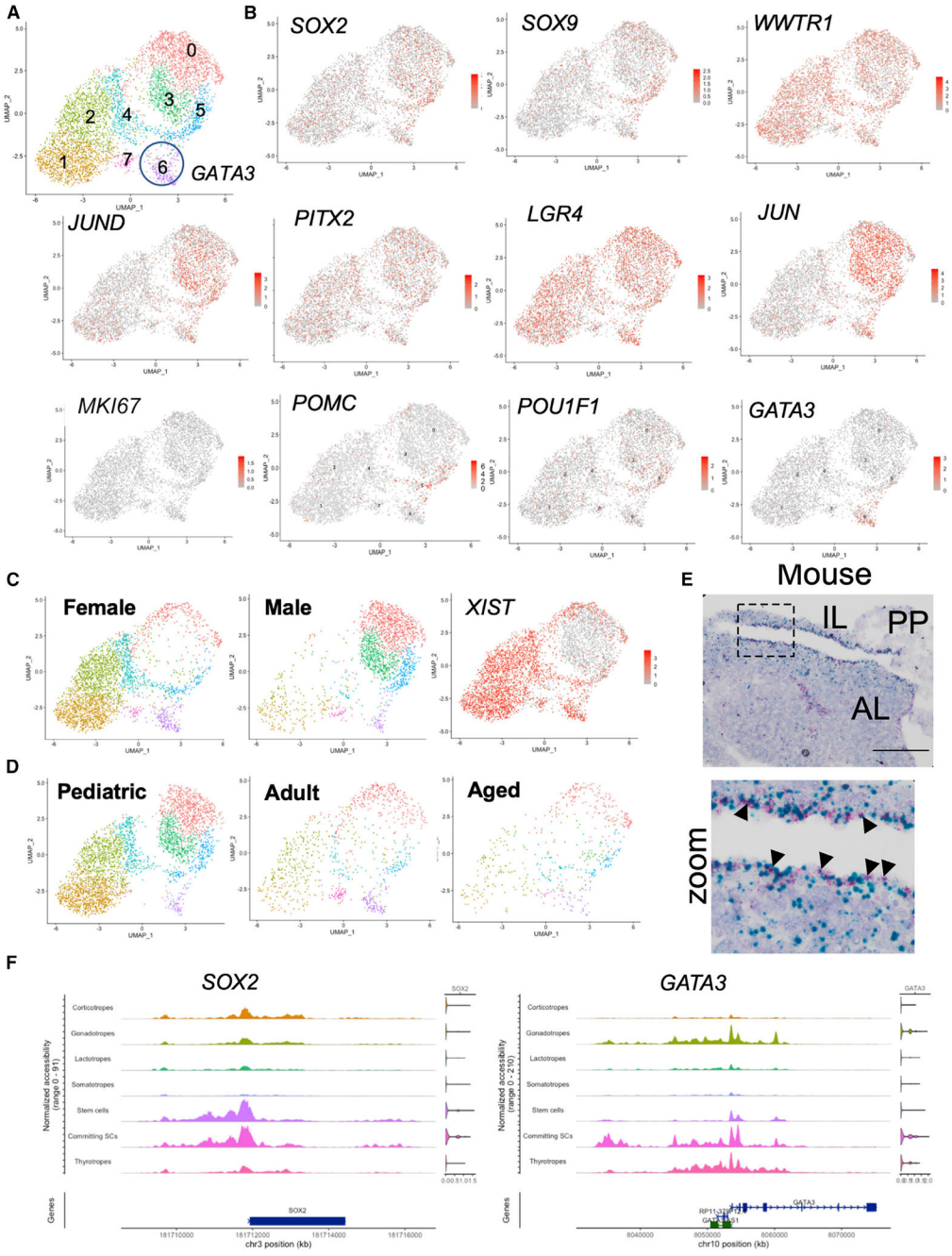


Figure 3. Identification of human stem cell sub-clusters by snRNA-seq

(A) UMAP showing the stem cell cluster identification based on the snRNA-seq data from the six merged human pituitary samples. Each cell cluster is color-coded and numbered. *GATA3* lineage-committed progenitor stem cells are circled.

(B) Feature plots depicting the expression distribution of key stem cell marker genes and of cell lineage commitment marker genes among the various clusters. A scale is included for each feature plot. All scales are similar except for *POMC* due to background gene expression. For the *GATA3, POMC*, and *POU1F1* committing cell lineages, clusters are

labeled for easier identification and localization of the cells throughout the PSC clusters. Additional gene feature plots are presented in Figure S3.

(C) UMAPs identifying all color-coded stem cell sub-clusters in females and males. The feature plot on the far right shows *XIST* expression, highlighting the female samples.

(D) UMAPs identifying all color-coded stem cell sub-clusters in the pediatric, adult, and aged donors.

(E) Colocalization of *Sox2* (red) and *Jun* (blue) transcripts in a wild-type P56 CD-1 male adult mouse pituitary. Scale bar, 200 μm. AL, anterior lobe; IL, intermediate lobe; PP, posterior pituitary. *Top*, full image. *Bottom*, magnification of the boxed region on the top panel. Arrows highlight specific cells with colocalization of *Sox2* and *Jun*. Refer to Figure S3 for *Sox2* and *Jun* colocalization at P3 and P15. Shown is a representative staining from three to five biological replicates.

(F) Chromatin accessibility track analysis and gene expression analysis (violin plots to the right of the chromatin accessibility tracks) and for *SOX2* and *GATA3* in all pituitary cell types from the sn multiome dataset generated from the pediatric female. The gene structure is provided below the tracks.

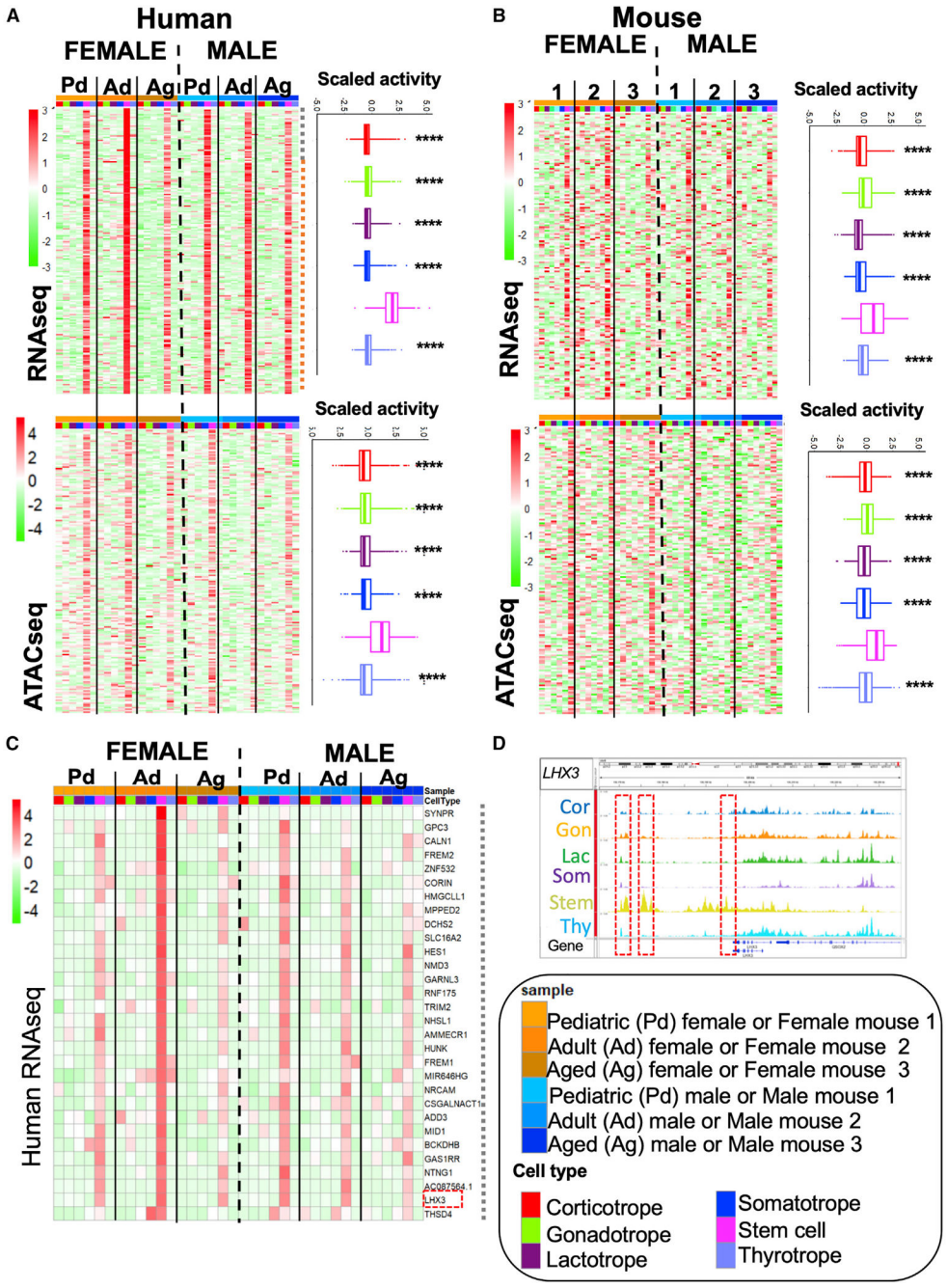


Figure 4. Characterization of coordinated gene expression and chromatin accessibility programs in human pituitary cell types

(A) *Top*, Heatmaps of gene expression levels for the top 200 genes associated with the human PSC LV ($LVsc_{rna}$, *gray vertical line*, top 30 genes highlighted in C). *Bottom*, Heatmap of chromatin accessibility levels for the top 200 genes associated with the human $LVsc_{rna}$ applied to the human snATAC-seq datasets. Cell-type-associated boxplots are shown for each heatmap.

(B) Heatmap for the top 200 genes associated with the human $LVsc_{rna}$ applied to the murine snRNA-seq dataset (*top*), and for the top 200 genes associated with the human

LVsc$_{rna}$ applied to the murine snATAC-seq datasets (*bottom*) (Ruf-Zamojski et al., 2021). Cell-type-associated boxplots are shown for each heatmap.
(C) Heatmap of gene expression levels for the top 30 genes associated with human PSC heatmap LVsc$_{rna}$ (A). (A–C) Each pituitary sample is indicated at the top. In the scale bars, red signifies the highest level of RNA expression or chromatin accessibility. Pd, pediatric; Ad, adult; Ag, aged pituitary. Cell type and donor color-coding are provided on the bottom key. (A, B) The blue dotted horizontal line corresponds to the top 30 genes as shown in (C). (D) Chromatin accessibility tracks for *LHX3* (identified in C, *box*) for all cell types in the pediatric female sample. The gene structure is provided underneath the tracks. Additional tracks are presented in Figure S5. Refer to Table 1 for donor-related information. Additional LV analyses are presented in Figure S4. The top 200 genes for all LVs illustrated in Figure 4 are provided as spreadsheets in the Supplemental information.

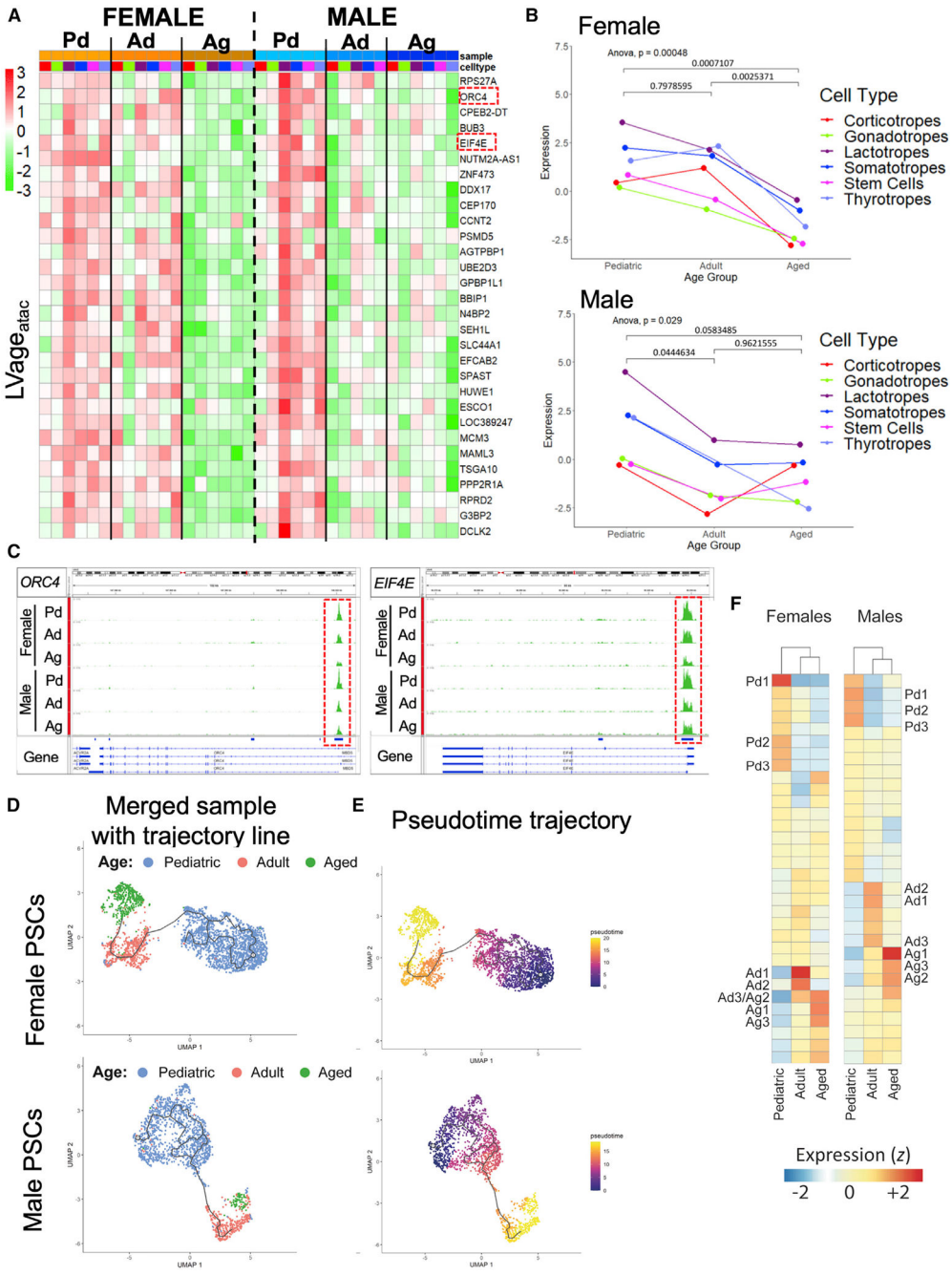


Figure 5. Age-associated chromatin accessibility and transcriptome pseudotime trajectory analysis

(A) Heatmap of chromatin accessibility levels for the top 30 genes in the human age-dependent LV ($LVage_{atac}$).

(B) Cell type and donor color-coding are provided in Figure 4. Additional LV analyses are presented in Figure S4. Plot showing the overall changes in chromatin accessibility for all pituitary cell types over age for females (*top*) and males (*bottom*). Pituitary cell types are color-coded, as indicated. Same cell types are connected by lines over donor age.

(C) Chromatin accessibility tracks analysis for *ORC4* (*left*) and *EIF4E* (*right*) in all lactotropes in the snATAC-seq datasets from all donors. The gene structure is provided underneath the tracks. Additional tracks are presented in Figure S5.
(D) UMAP showing the trajectory within the stem cell cluster for female (*top*) and male (*bottom*) samples, with samples color-coded by age.
(E) Pseudotime trajectory analysis for female (*top*) and male (*bottom*) samples. Pediatric PSCs represent the root of each trajectory. The color scale of pseudotime trajectories is displayed.
(F) Gene modules identified with pseudotime and showing expression changes over donor age. Monocle 3 identified sets of genes whose expression changes over donor age as a function of pseudotime. These modules are then plotted on a heatmap to depict their relative expression in each age group. The top three enriched modules for each age group are labeled, with module 1 being the most highly enriched in each age. Pd1-3 denotes the pediatric top three modules, Ad1–3 the Adult-enriched modules, and Ag1-3 the Aged-enriched. Blue to red on the color scale represents low to high relative expression levels (z-transformed mean expression) of gene modules, respectively. See Figure S6 for selected gene trajectories within specific modules. See Table S3B for the top genes per module.

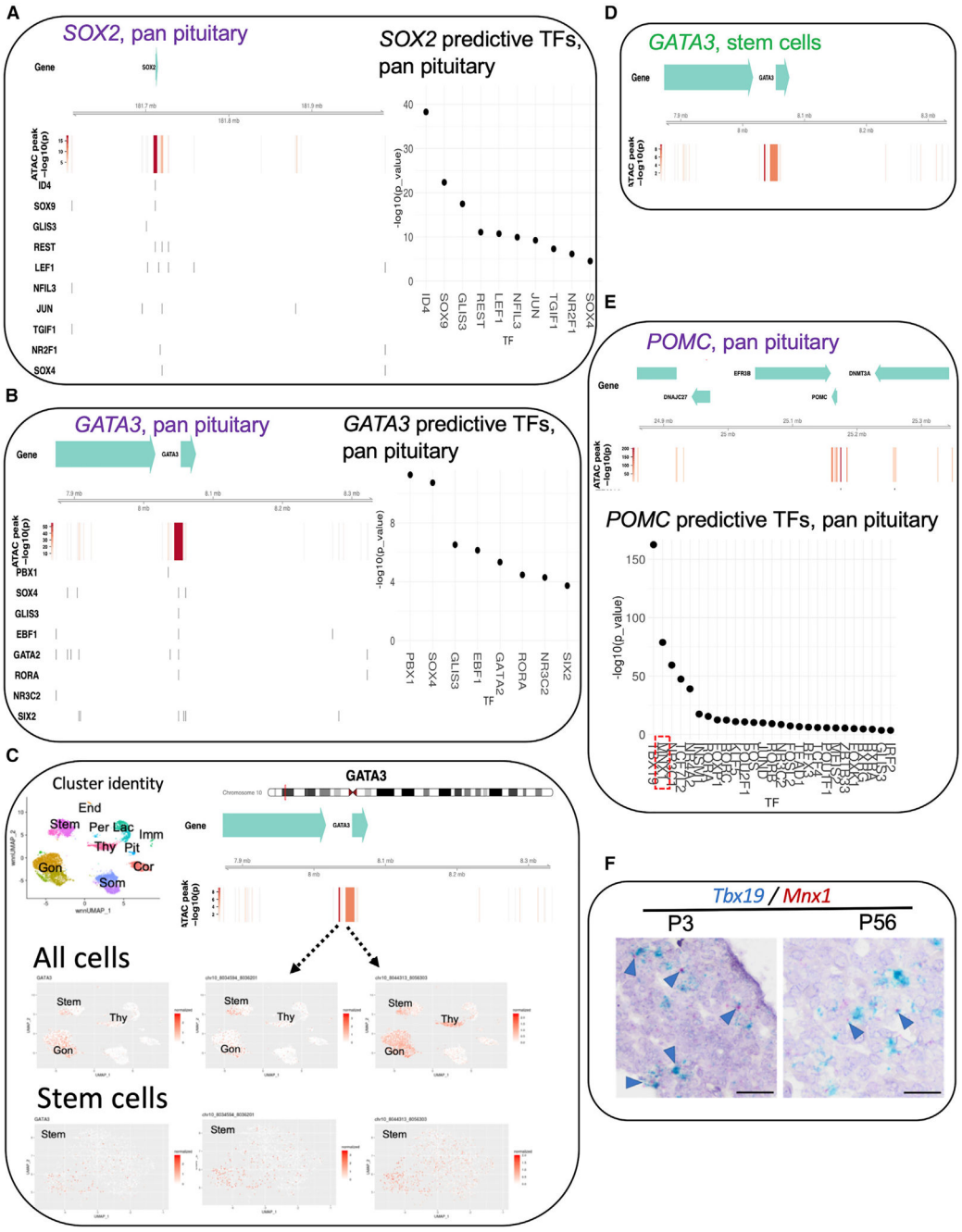


Figure 6. Linear model predicting the chromatin accessibility mechanisms and TFs contributing to PSC gene expression

(A) Linear modeling analysis of all pituitary cells ("pan pituitary cell analysis") infers the chromatin accessibility and TFs involved in stem cell-specific *SOX2* expression. *Left*, top track shows the contribution of each peak to gene expression measured by −log(p value), and bottom tracks display the TF binding sites. *Right*, the individual contribution of each predicted TF to *SOX2* expression is shown as −log(p values). See Figure S7E for *SOX2* analysis in stem cells only.

(B) Pan pituitary cell analysis infers the chromatin accessibility and TFs involved in stem cell-specific *GATA3* expression.

(C) *Top*, *right*: Linear modeling analysis infers *GATA3* chromatin accessibility. *Bottom*: t-SNE showing *GATA3* gene expression (*left panels*) and chromatin accessibility (*middle* and *right panels*) in all cells (*top*) versus PSCs (*bottom*). The top left panel indicates the identification of all cell type clusters.

(D) Linear modeling analysis of stem cells only infers the chromatin accessibility and TFs involved in the differential expression of *GATA3* expression within the stem cell population. No TF is predicted to contribute to *GATA3* expression in stem cells.

(E) Pan pituitary cell analysis infers the chromatin accessibility and TFs involved in stem cell-specific *POMC* expression. Boxed is *MNX,1* the second highest identified TF that is predicted to contribute to *POMC* expression. See Figure S7 for *POMC* analysis in stem cells only.

(F) RNAscope mRNA in situ hybridization showing the colocalization of *Mnx1* (red) and *Tbx19* (blue) transcripts in wild-type CD-1 postnatal day 3 (P3) and adult (P56) female mouse pituitaries. Blue arrows highlight specific cells with colocalization. Scale bar, 25 μm. Shown are representative stainings from three to five biological replicates for each age.

Table 1.

Donor characteristics

Subject	Sex (F/M)	Ethnicity	Age	PMI (h)	Medications	Cause of death	Year of sample collection
Pediatric	F	African-American	2	21	Cardec, Amoxicillin	acute myocarditis	2000
Adult	F	African-American	37	9	N/A	cardiac arrhythmia	2013
Aged	F	white	85	5.5	N/A	N/A	2018
Pediatric	M	white	8	36	N/A	hypothermia and drowning	2002
Adult	M	white	41	13	N/A	maybe heart problem	2014
Aged	M	white	90	4	N/A	N/A	2015

F, female; M, male; N/A, not available/not reported; PMI, postmortem interval.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Female pediatric pituitary (frozen from deceased)	NIH Neurobiobank	#1275, NDAR_INVCK755EF3
Female adult pituitary (frozen from deceased)	NIH Neurobiobank	#5621, NDAR_INVXK717LJA
Female aged pituitary (frozen from deceased)	NIH Neurobiobank	#187443
Male pediatric pituitary (frozen from deceased)	NIH Neurobiobank	#1674, NDAR_INVMV348UPL
Male adult pituitary (frozen from deceased)	NIH Neurobiobank	#5818, NDAR_INVKF350UGX
Male aged pituitary (frozen from deceased)	NIH Neurobiobank	#187438
Wildtype murine postnatal pituitaries	Charles River Laboratories	CD-1 strain; MGI: 5649524
Chemicals, peptides, and recombinant proteins		
RNAse inhibitor	NEB	# MO314L
OptiPrep	StemCell Tech	07820
Sucrose	Sigma	S0389
EDTA	Corning	46-034-Cl
Tris-HCl, pH 7.4	Sigma	T2663
$CaCl_2$	Sigma	21115
$Mg(Ac)_2$	Boston Bioproducts	MT-190
IGEPAL CA-630	Sigma	I3021
10% Buffered Neutral Formalin	Sigma	HT501128
Mayer's hematoxylin	Vector	H-3404
Vectamount	Vector	H-5000
Tween-20 Surfact Amps Detergent	Thermo Scientific	85113
Buffer EB	Qiagen	19086
SPRI Select	Beckman Coulter	B23318
Critical commercial assays		
Chromium Single Cell 3′ Reagents	10x Genomics	V3
Chromium Single Cell ATAC Reagents	10x Genomics	V1
Chromium Single Cell Multiome ATAC and Gene Expression Reagent	10x Genomics	V1
RNAscope 2.5 HD Duplex assay	Advanced Cell Diagnostics	N/A
Deposited data		
Raw and analyzed data	This paper	GEO: GSE178454
Raw and analyzed data	Ruf-Zamojski et al. (2021)	GEO: GSE151962
Experimental models: Organisms/strains		
Wildtype murine postnatal pituitaries	Charles River Laboratories	CD-1 strain; MGI: 5649524
Oligonucleotides		
Mm-Jun probe	Advanced Cell Diagnostics	# 453561
Mm-Sox2-C2 probe	Advanced Cell Diagnostics	# 401041-C2
Software and algorithms		
Cell Ranger	10x Genomics	v5.0.0

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Seurat	Butler et al. (2018) and Stuart et al. (2019)	v.3.9.9.9024; v4.0.1; v3.1.5
Cell Ranger-ATAC pipeline	10x Genomics	v1.2.0
Seurat/Signac	Stuart et al. (2020)	v3.1.5/0.2.4
Cell Ranger ARC	10x Genomics	v1.0.0
Monocle3	Cao et al. (2019)	v0.2.3.0
PLIER	Mao etal. (2019)	N/A
Original code for linear model	This paper	https://github.com/zidong93/snpituitary_human_manuscript
Other		
Dounce glass homogenizer	VWR	# 71000-514
40 mm cell strainer	Falcon	352340
Centrifuge SW41 rotor	Beckman Coulter	#331362
Fluorometer	Invitrogen	Qubit3
Bioanalyzer	Agilent	2100
Sequencer	Illumina	MiSeq
Sequencer	Illumina	NovaSeq 6000
Cell counter	Nexcelome	K2
Cell Counter	Invitrogen	Countess