Short Paper
A Short Subtitle

Evgenii O. Tretiakov, PhD[a], Tibor Harkany, PhD[a,b,*], Amina Telalovic[a]

[a] *Medical University of Vienna, Department of Molecular Neurosciences, Center for Brain Research, Spitalgasse 4, Vienna, Austria, A-1090*

[b] *Karolinska Institutet, Department of Neuroscience, Biomedicum 7D, Solna, Sweden, SE-17165*

Abstract

This is the abstract. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum augue turpis, dictum non malesuada a, volutpat eget velit. Nam placerat turpis purus, eu tristique ex tincidunt et. Mauris sed augue eget turpis ultrices tincidunt. Sed et mi in leo porta egestas. Aliquam non laoreet velit. Nunc quis ex vitae eros aliquet auctor nec ac libero. Duis laoreet sapien eu mi luctus, in bibendum leo molestie. Sed hendrerit diam diam, ac dapibus nisl volutpat vitae. Aliquam bibendum varius libero, eu efficitur justo rutrum at. Sed at tempus elit.

Keywords: keyword1, keyword2

1. Results

1.1. Single-cell RNA sequencing reveals three distinct groups of pars tuberalis cells with unique transcriptional profiles

To thoroughly investigate the cellular diversity of the pars tuberalis (PT) during embryonic development, we analyzed single-cell RNA sequencing (scRNA-seq) data from two comprehensive studies: Romanov et al.[1] and Kim et al.[2]. These datasets encompass a broad range of developmental stages, from embryonic day 10 (E10) to postnatal day 45 (P45), allowing us to capture the dynamic changes in gene expression throughout PT development.

In the dataset from Romanov et al., we identified a distinct cluster of PT cells within the hypothalamic scRNA-seq data (Fig. 1a). This cluster was characterized by the expression of key developmental genes, including *Pitx1*, a

*Corresponding author

Email addresses: evgenii.tretiakov@meduniwien.ac.at (Evgenii O. Tretiakov, PhD), tibor.harkany@meduniwien.ac.at (Tibor Harkany, PhD), amina.telalovic@meduniwien.ac.at (Amina Telalovic)

transcription factor crucial for anterior pituitary gland formation[3], and *Gata2*, which is essential for the early stages of Rathke's pouch detachment and PT formation[4].

To enhance the resolution of our analysis and validate these findings, we integrated scRNA-seq data from Kim et al., which provided a rich temporal landscape of hypothalamic development. By aligning the datasets, we ensured accurate identification and mapping of PT cell populations across different developmental stages (Fig. 1b).

Clustering of the integrated dataset revealed three distinct groups of PT cells, each defined by unique transcriptional profiles:

1.1.1. ***Group One: Eya3^{high}/Tshb^{+}/Cck^{+} Cells***

Group one cells exhibited high expression of the transcriptional coactivator *Eya3* and co-expressed the peptide hormones *Tshb* (thyroid-stimulating hormone beta subunit) and *Cck* (cholecystokinin). These cells were consistently present from E12 onward and remained prominent throughout embryonic development (Fig. 2a). Visualization using UMAP (Uniform Manifold Approximation and Projection) showed that *Eya3*, *Tshb*, and *Cck* were co-expressed within the same cell clusters.

Spearman's correlation analysis across developmental stages confirmed strong positive correlations between *Eya3* and *Tshb* (> 0.85, $p < 0.001$) and between *Eya3* and *Cck* (> 0.80, $p < 0.001$). This co-expression suggests a coordinated regulatory mechanism driving hormone production in these PT cells.

1.1.2. ***Group Two: Eya1^{+}/Eya4^{+} Cells***

Cells in group two were characterized by the expression of *Eya1* and *Eya4*. These transcriptional coactivators were expressed at lower levels compared to *Eya3* and were more spatially and temporally restricted (Fig. 2b). Group two cells became more apparent at later developmental stages (after E14.5), indicating a potential role in the maturation or specialization of certain PT cell types.

Unlike group one, group two cells exhibited minimal expression of *Tshb* and *Cck*, suggesting they may serve functions distinct from hormone secretion, possibly in structural or regulatory roles within the PT.

1.1.3. ***Group Three: Eya2^{+} Cells***

Group three cells predominantly expressed *Eya2*. This expression was low during early stages (E10–E12) but significantly increased from E14.5 onward (Fig. 2c). These cells were mainly located at the periphery of the PT, forming distinct clusters separate from groups one and two.

The temporal increase in *Eya2* expression and the spatial distribution of group three cells imply they may represent a later phase of PT cell differentiation or a unique lineage with specialized functions yet to be fully understood.

To further elucidate the characteristics of these groups, we examined additional genes implicated in PT development and function:

- **Pitx1 Co-expression**: *Pitx1*, essential for pituitary development, was co-expressed with *Eya3* in group one cells (Fig. 3a). This co-localization reinforces the idea that group one cells are actively involved in hormone production and early PT development.
- **Sox2 Expression**: *Sox2*, a marker of neural stem and progenitor cells, was expressed across all PT cell groups (Fig. 3b). The presence of *Sox2* suggests that PT cells may share common progenitor characteristics or that neural stem cell pathways influence PT cell differentiation.

We also explored the potential interactions between PT-derived hormones and the developing brain:

- **Hormone Receptor Expression**: The receptors for *Tshb* (*Tshr*) and *Cck* (*Cckbr*) were expressed in regions of the developing hypothalamus rich in neural progenitor cells, such as the ventricular zones (Fig. 4a, 4b). Co-expression of *Tshr* and *Cckbr* with *Sox2* indicates that PT-derived hormones could influence neural development by acting on these progenitor cells.
- **Gpr173 Expression**: We observed expression of *Gpr173*, a G protein-coupled receptor involved in neuroendocrine signaling, in neural progenitor zones overlapping with *Sox2* expression (Fig. 4c). This suggests a potential role for *Gpr173* in mediating hormonal signals from the PT to the developing brain.

To visualize gene expression dynamics across developmental stages, we generated DotPlots for key genes (Fig. 5). This analysis highlighted:

- **Early Onset and Persistence of Group One Genes**: *Eya3*, *Tshb*, and *Cck* expression began as early as E12 and continued throughout embryonic development, indicating a sustained role in PT function.
- **Temporal Emergence of Group Two and Three Genes**: *Eya1*, *Eya2*, and *Eya4* displayed increased expression at later stages, suggesting these genes are involved in the maturation and specialization processes within the PT.

Our integrated analysis, combining data from multiple developmental stages and employing rigorous statistical correlations, provides compelling evidence for the existence of three distinct PT cell groups during embryogenesis. Each group exhibits unique transcriptional signatures, indicating specialized roles:

- **Group One** is likely responsible for early hormone production, directly influencing both PT development and possibly exerting effects on the hypothalamus through hormone secretion.
- **Group Two** may be involved in structural formation or regulation within the PT, given their distinct gene expression profile and later emergence.

- **Group Three** could represent a specialized lineage that plays a role in the maturation of the PT or has unique endocrine functions.

2. Methods

2.1. Data Acquisition and Preprocessing

Single-cell RNA sequencing (scRNA-seq) data from the mouse hypothalamus were obtained from the study by Kim et al.[2], available under BioProject accession PRJNA547712 (https://github.com/EugOT/PRJNA547712). The dataset, comprising expression profiles of 128,006 cells across 27,998 genes, was acquired in the AnnData format (kim2020_combined.h5ad). This dataset spans various developmental stages, including embryonic days (E)10 to E18 and postnatal days (P)4 to P45. To facilitate analysis in R, the AnnData object was converted into a Seurat object[5, 6]. The expression matrix was transposed to align genes as rows and cells as columns. Cell metadata, including identifiers and developmental stages, were incorporated. Dimensionality reduction was performed using Uniform Manifold Approximation and Projection (UMAP)[7, 8], and the resulting embeddings were added to the Seurat object for visualization purposes.

2.2. Integration with Reference Dataset

To annotate cell types within the pars tuberalis (PT), we integrated query dataset[2] with a reference scRNA-seq dataset (https://github.com/EugOT/PRJNA548917), which provides detailed annotations of hypothalamic cell populations[1]. The reference dataset was updated to ensure compatibility and included 51,199 cells with expression profiles for 24,340 genes.

Anchors between the query dataset[2] and the reference dataset[1] were identified using the FindTransferAnchors function in Seurat[5, 6], employing principal component analysis (PCA) for dimensionality reduction over 30 dimensions. Cell type labels were transferred from the reference to the query dataset using the TransferData function, allowing for the prediction of cell identities within the PT[5]. The integration facilitated the identification of specific cell clusters corresponding to the PT, marked by the expression of genes such as Tshb, Cck, and Eya3.

2.3. Normalization and Feature Selection

The integrated Seurat object was normalized using the NormalizeData function, applying global-scaling normalization to account for differences in sequencing depth. Variable features were identified using the FindVariableFeatures function with the 'vst' method[9], selecting the top 3,000 genes exhibiting high variability across cells. The data were then scaled using the ScaleData function, regressing out unwanted sources of variation and centering and scaling the data for downstream analyses.

2.4. Dimensionality Reduction and Visualization

Principal component analysis was conducted on the scaled data, and the first 30 principal components were used for further analysis. UMAP was performed using the RunUMAP function to visualize the data in two-dimensional space, capturing the global structure and potential trajectories of cellular differentiation.

Feature plots were generated using the FeaturePlot function to visualize the expression patterns of key genes across developmental stages. Genes of interest included Tshb, Cck, Pitx1, Eya1, Eya2, Eya3, Sox2, Hlf, Tshr, Cckar, Cckbr, and Gpr173. Blended feature plots were employed to examine co-expression patterns, with parameters adjusted for optimal visualization (e.g., blend thresholds, point sizes, alpha levels, and color schemes).

2.5. Statistical Analysis

To assess the association between the expression of specific genes across developmental stages, chi-squared tests of independence were conducted. Binary expression matrices were created by thresholding gene expression counts (e.g., considering a gene as expressed using gene-specific quantile thresholding). Contingency tables were constructed for pairs of genes (*Tshb*, *Cck*, *Pitx1*, *Eya1*, *Eya2*, *Eya3*, *Eya4*, *Sox2*, *Hlf*, *Tshr*, *Cckar*, *Cckbr*, and *Gpr173*), such as *Sox2* and *Tshr*, across different developmental stages.

The ggpiestats function from the ggstatsplot package[10] was utilized to generate grouped pie charts, displaying the proportion of cells expressing combinations of genes and the results of chi-squared tests. The analyses provided insights into potential regulatory relationships and developmental dynamics within the PT.

UpSet plots[11] were created to visualize the counts and proportions of cells expressing various combinations of genes across developmental stages. These plots illustrated the temporal changes in gene expression and the prevalence of specific cell populations during development.

2.6. Gene Expression Correlation Analysis

To investigate potential co-regulation and interactions between genes, correlation analyses were performed. Spearman's rank correlation coefficients were calculated for pairs of genes across cells at each developmental stage. Hexbin plots were generated using the hexbin package to visualize the density and correlation of gene expression in a two-dimensional space, aiding in the identification of spatial expression patterns and clusters.

Dot plots were generated using the DotPlot function to summarize the expression of genes of interest across developmental stages. The genes included *Tshb*, *Cck*, *Pitx1*, *Eya3*, *Sox2*, *Hlf*, *Igfbp5*, *Tshr*, *Cckar*, *Cckbr*, and *Gpr173*. The dot plots represented both the proportion of cells expressing each gene and the average expression levels, providing a comprehensive overview of gene expression dynamics during development.

2.7. Data Availability

The code (https://github.com/EugOT/pr-PT) and datasets used in this study are publicly available. The Kim[2] dataset can be accessed under BioProject accession PRJNA547712, and the Romanov[1] dataset is available under BioProject accession PRJNA548917 including repository with reprocessing code: https://github.com/EugOT/PRJNA547712 and https://github.com/EugOT/PRJNA548917.

References

[1] R. A. Romanov, E. O. Tretiakov, M. E. Kastriti, M. Zupancic, M. Häring, S. Korchynska, K. Popadin, M. Benevento, P. Rebernik, F. Lallemend, K. Nishimori, F. Clotman, W. D. Andrews, J. G. Parnavelas, M. Farlik, C. Bock, I. Adameyko, T. Hökfelt, E. Keimpema, T. Harkany, Molecular design of hypothalamus development, Nature 582 (7811) (2020) 246–252, 48 citations (Semantic Scholar/DOI) [2022-06-25] ECC: 0000010 Citation Key: romanovMolecularDesignHypothalamus2020 tex.ids= romanovMolecularDesignHypothalamus2020a number: 7811 publisher: Nature Publishing Group. doi:10.1038/s41586-020-2266-0.
URL https://www.nature.com/articles/s41586-020-2266-0

[2] D. W. Kim, P. W. Washington, Z. Q. Wang, S. H. Lin, C. Sun, B. T. Ismail, H. Wang, L. Jiang, S. Blackshaw, The cellular and molecular landscape of hypothalamic patterning and differentiation from embryonic to late postnatal development, Nature Communications 11 (1) (2020) 4360, eCC: 0000008 7 citations (Semantic Scholar/DOI) [2021-04-27] Citation Key: kimCellularMolecularLandscape2020 tex.ids= kimCellularMolecularLandscape2020a PMCID: PMC7459115 PMID: 32868762. doi:10/ghdps6.
URL 10/ghdps6

[3] D. P. Szeto, A. K. Ryan, S. M. O'Connell, M. G. Rosenfeld, P-otx: a pit-1-interacting homeodomain factor expressed during anterior pituitary gland development, Proceedings of the National Academy of Sciences of the United States of America 93 (15) (1996) 7706–7710, pMID: 8755540 PMCID: PMC38811 Citation Key: szetoPOTXPIT1interactingHomeodomain1996. doi:10.1073/pnas.93.15.7706.

[4] J. S. Dasen, S. M. O'Connell, S. E. Flynn, M. Treier, A. S. Gleiberman, D. P. Szeto, F. Hooshmand, A. K. Aggarwal, M. G. Rosenfeld, Reciprocal interactions of pit1 and gata2 mediate signaling gradient–induced determination of pituitary cell types, Cell 97 (5) (1999) 587–598, citation Key: dasenReciprocalInteractionsPit11999. doi:10.1016/S0092-8674(00)80770-9.
URL https://www.sciencedirect.com/science/article/pii/S0092867400807709

[5] Y. Hao, T. Stuart, M. H. Kowalski, S. Choudhary, P. Hoffman, A. Hartman, A. Srivastava, G. Molla, S. Madad, C. Fernandez-Granda, R. Satija,

Dictionary learning for integrative, multimodal and scalable single-cell analysis, Nature BiotechnologyPMID: 37231261 Citation Key: haoDictionaryLearningIntegrative2023 (05 2023). doi:10.1038/s41587-023-01767-y.

[6] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data, Cell 177 (7) (2019) 1888–1902.e21, 3763 citations (Semantic Scholar/DOI) [2022-06-25] ECC: 0003109 tex.ids= stuartComprehensiveIntegrationSingleCell2019a PMCID: PMC6687398 PMID: 31178118 Citation Key: stuartComprehensiveIntegrationSingleCell2019. doi:10.1016/j.cell.2019.05.031.

[7] L. McInnes, J. Healy, J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction (2018). doi:10.48550/arxiv.1802.03426.

[8] D. Kobak, G. C. Linderman, Initialization is critical for preserving global data structure in both t-sne and umap., Nature biotechnology 39 (2) (2021) 156–157, eCC: 0000010 5 citations (Semantic Scholar/DOI) [2021-07-29] Place: United States PMID: 33526945 Citation Key: kobakInitializationCriticalPreserving2021. doi:10/ghxbbn.

[9] C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell rna-seq data using regularized negative binomial regression, Genome Biology 20 (1) (2019) 296, 872 citations (Semantic Scholar/DOI) [2022-08-12] Citation Key: hafemeisterNormalizationVarianceStabilization2019 tex.ids= hafemeisterNormalizationVarianceStabilization2019a PMCID: PMC6927181 PMID: 31870423. doi:10.1186/s13059-019-1874-1.
URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1874-1

[10] I. Patil, Visualizations with statistical details: The 'ggstatsplot' approach, Journal of Open Source Software 6 (61) (2021) 3167, 253 citations (Semantic Scholar/DOI) [2023-04-05] Citation Key: patilVisualizationsStatisticalDetails2021. doi:10.21105/joss.03167.
URL https://joss.theoj.org/papers/10.21105/joss.03167

[11] J. R. Conway, A. Lex, N. Gehlenborg, Upsetr: an r package for the visualization of intersecting sets and their properties, Bioinformatics 33 (18) (2017) 2938–2940, citation Key: conwayUpSetRPackageVisualization2017 tex.ids= conwayUpSetRPackageVisualization2017a. doi:10.1093/bioinformatics/btx364.
URL https://academic.oup.com/bioinformatics/article/33/18/2938/3884387