

# TENT5A may regulate the menstrual cycle in humans

Maria Nizik<sup>1,\*</sup>

<sup>1</sup>University of Warsaw, Faculty of Mathematics, Informatics and Mechanics, Warsaw, Poland  
\*m.nizik3@student.uw.edu.pl

## ABSTRACT

Human endometrium is a complex and relatively under-investigated tissue. It undergoes strong changes during the menstrual cycle. Terminal nucleotidyl transferases (TENTs) play pivotal role in regulating expression by extending poly(A) tails of RNA molecules in the cytoplasm. TENTs are known to be expressed in the endometrium, but their role has not yet been extensively explored. Here, I characterized expression of TENTs in the endometrium at the single-cell resolution. TENTs were strongly expressed in epithelial cells - specifically glandular and SOX9 cells. TENT5A was the most prevalent among the TENTs. I found out that TENT5A levels fluctuated with the menstrual cycle - rising in the mid-secretory phase, which indicates TENT5A may play a role in regulating the menstrual cycle. The menstrual cycle related pattern was also visible in glandular and SOX9 cells. However, to prove that TENT5A regulates menstrual cycle in humans further experimentation is required. To achieve most-relevant results I propose simulating the menstrual cycle *in vitro* with hormones.

## Introduction

### 0.1 Endometrium

Endometrium is a tissue that lines the mammalian uterus. Its function is to receive the embryo during the implantation and maintain it afterwards<sup>1</sup>. Endometrium is relatively less researched compared to other tissues, due to the difficulty of obtaining samples, and a past gender-bias in medicine. Recently, the knowledge gap have been actively filled with new research<sup>2,3</sup>.

Endometrium dysfunctions may lead to various afflictions. Faulty decidualization is linked to preeclampsia - a severe condition in pregnancy<sup>4</sup>. Endometriosis is a painful and common condition caused by the endometrium development outside uterus. Its aetiology has not yet been fully explained<sup>5</sup>. Infertility may also be caused by endometrial defects<sup>6</sup>. Another highly important subject of research are endometrial cancers<sup>7</sup>.

#### 0.1.1 Menstrual cycle

Menstrual cycle prepares the endometrium for the implantation. Hormones - estrogens and progesterone - orchestrate changes happening in the endometrium during the cycle. The menstrual cycle has two major phases - proliferative and secretory. Estrogens dominate in the proliferative phase and progesterone governs the shift from the proliferative to secretory phase transforming endometrium into decidua. In absence of implantation, drop in the progesterone level causes menstruation - shedding of the endometrium. If implantation happens, progesterone does not drop and the decidualization continues<sup>1</sup>.

#### 0.1.2 Cell types

Endometrium is a complex tissue consisting of multiple cell types. I adopted the division proposed in the manuscript from which I took the data<sup>2</sup>. Endometrial cells are split into endothelial (blood vessel lining), immune, stromal, supporting, and epithelial cells. Further, the endothelial cells are divided into arterial

and venous; supporting into perivascular, fibroblasts expressing C7, and smooth muscle cells; stromal into decidualized (eS) and non-decidualized (dS); epithelial into SOX9, lumenal, glandular and ciliated.

## 0.2 Polyadenylation

In mammals most of the messenger RNA (mRNA) molecules are polyadenylated, which means multiple adenosines are added at their 3' ends forming poly(A) tails. Primarily, the process happens co-transcriptionally in the cell nucleus<sup>8</sup>. Proteins responsible for polyadenylation are called poliadenylases or poly(A) polymerases (PAPs). PAPs may be divided into canonical that are acting in nucleus and non-canonical (ncPAPs) working in the cytoplasm.

Poly(A) tails stabilize the mRNA molecules, by protecting them from degradation. Counter-intuitively, poly(A) tails may also stimulate deadenylation, which results in degradation. This way poly(A) tails control RNA levels and translation-efficacy, which in turn impact protein within the cells<sup>8</sup>. Non-canonical poliadenylases can modify gene expression dynamically after the RNA molecules have already been transcribed and transported to cytoplasm, which makes them core regulatory proteins.

## 0.3 TENTs

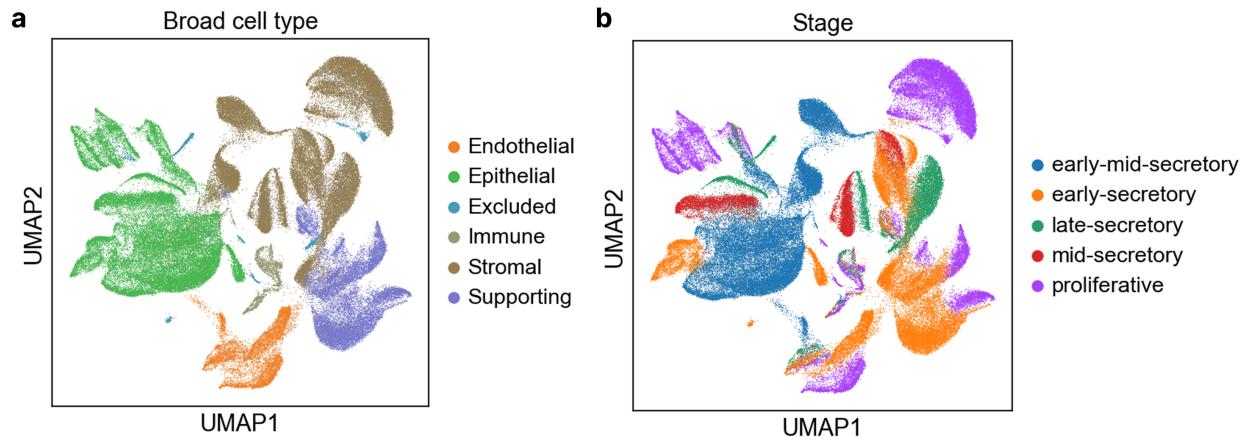
Terminal nucleotidyl transferases (TENTs) play a role in post-transcriptional RNA modifications in mammals. Some TENTs belong to ncPAPs and the others to terminal uridyl transferases (TUTases)<sup>9,10</sup>. The following analysis will focus on 5 TENTs all of which are ncPAPs: TENT4B, TENT4A, TENT5A, TENT5B, and TENT5C.

TENTs were reported to regulate multiple processes. They have been linked to oogenesis (TENT5B, TENT5C) and spermatogenesis (TENT5C and TENT5D) in mice<sup>11</sup>, innate immune response in animals<sup>12</sup>, and the immunoglobulin expression and the humoral immune response (TENT5C)<sup>13</sup>. TENT5A plays a role in muscle fibre formation<sup>14</sup>, osteogenesis<sup>15</sup>, hemoglobinization<sup>16</sup>, and cancer<sup>17,18</sup>. It was also found to extend poly(A) tails of SARS-CoV-2 mRNA vaccines enhancing their efficacy in mice<sup>19</sup>.

Considering the ubiquity of TENTs in various tissues undergoing dynamic changes, I explored their transcriptomes in endometrium, which is a highly-variable tissue. According to the Human Protein Atlas single-cell data various TENTs including TENT4A, TENT4B, TENT5A, TENT5B and TENT5C are expressed in human endometrium<sup>20</sup>. So far TENT4A was reported to regulate DNA-damage tolerance in endometrial cancer<sup>7</sup>.

## Results

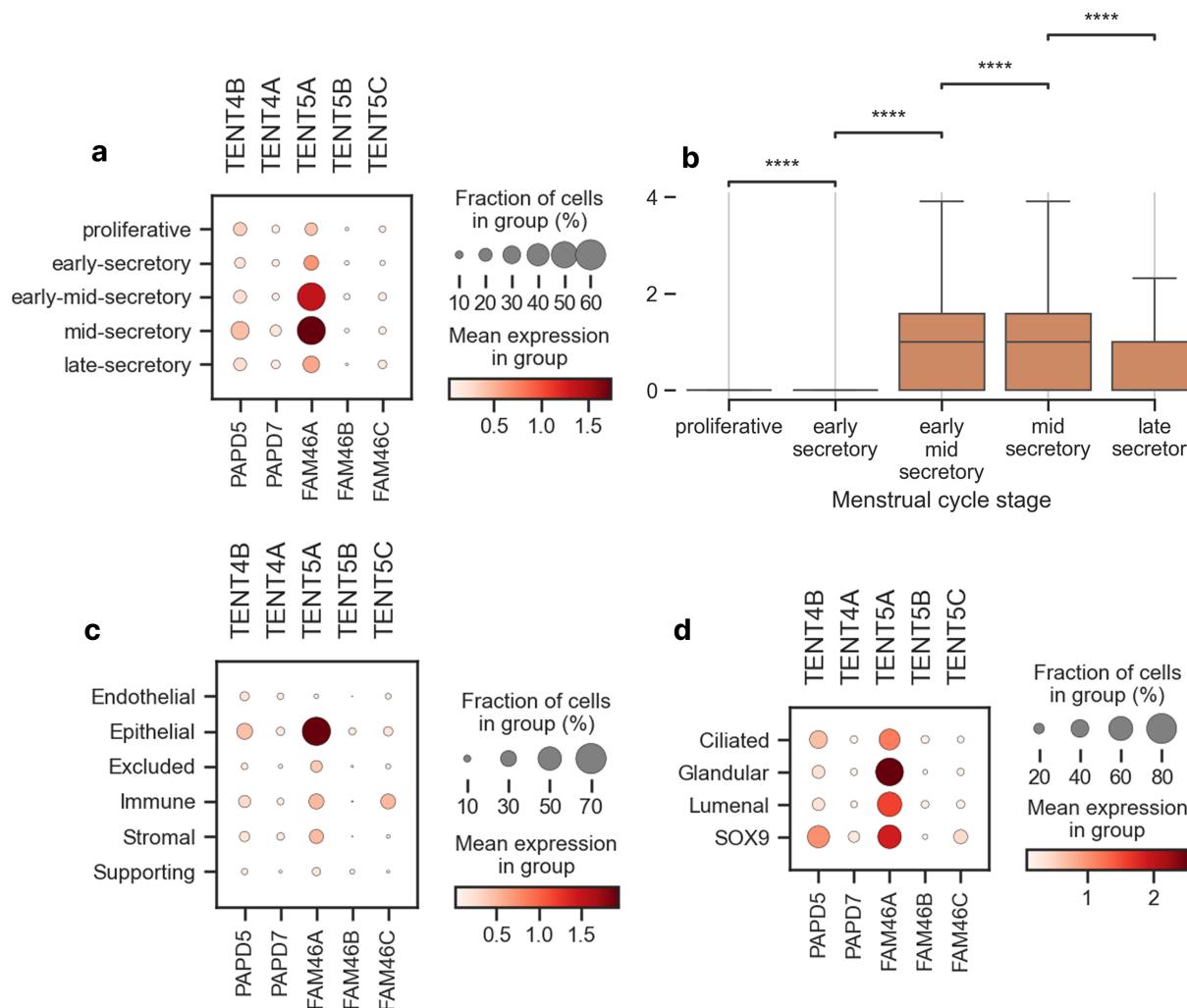
Endometrial cells of the same broad cell type cluster together (Fig. 1a), which indicates that the cell type is the dominant factor in transcriptional patterns of endometrium. The secondary factor is a menstrual cycle. Within cell-types, cells are clearly divided by the cycle stage (Fig. 1b).



**Figure 1.** UMAP projection of the human endometrium coloured by the broad cell type (a) and menstrual cycle stage (b).

TENTs are expressed in the endometrium, but the levels vary between the members of the TENT family, cell types, and menstrual cycle stages (Fig. 2). TENT5A dominates expression in the TENT family, TENT5B is much less abundant, and the rest of TENT family members are expressed at the negligible level (Fig. 2a, 2c).

TENT5A levels are low at the proliferative and early-secretory phase, rise ad the mid-secretory to drop at the end of the secretory phase (Fig. 2a, 2b). That indicates TENT5A may play a role in regulating the menstrual cycle. The differences in expression between all the phases are statistically significant (Fig. 2b). Among cell types TENT5A expression is predominant in epithelial cells, which undergo highly dynamic changes during the menstrual cycle (Fig. 2c). Each group of the epithelial cells: ciliated, glandular, luminal, and SOX9 cells exhibit high TENT5A levels. Glandular cells show the strongest and most ubiquitous TENT5A expression among the endometrium epithelial cells. In SOX9 cells the expression is slightly weaker and less widespread (Fig. 2d).

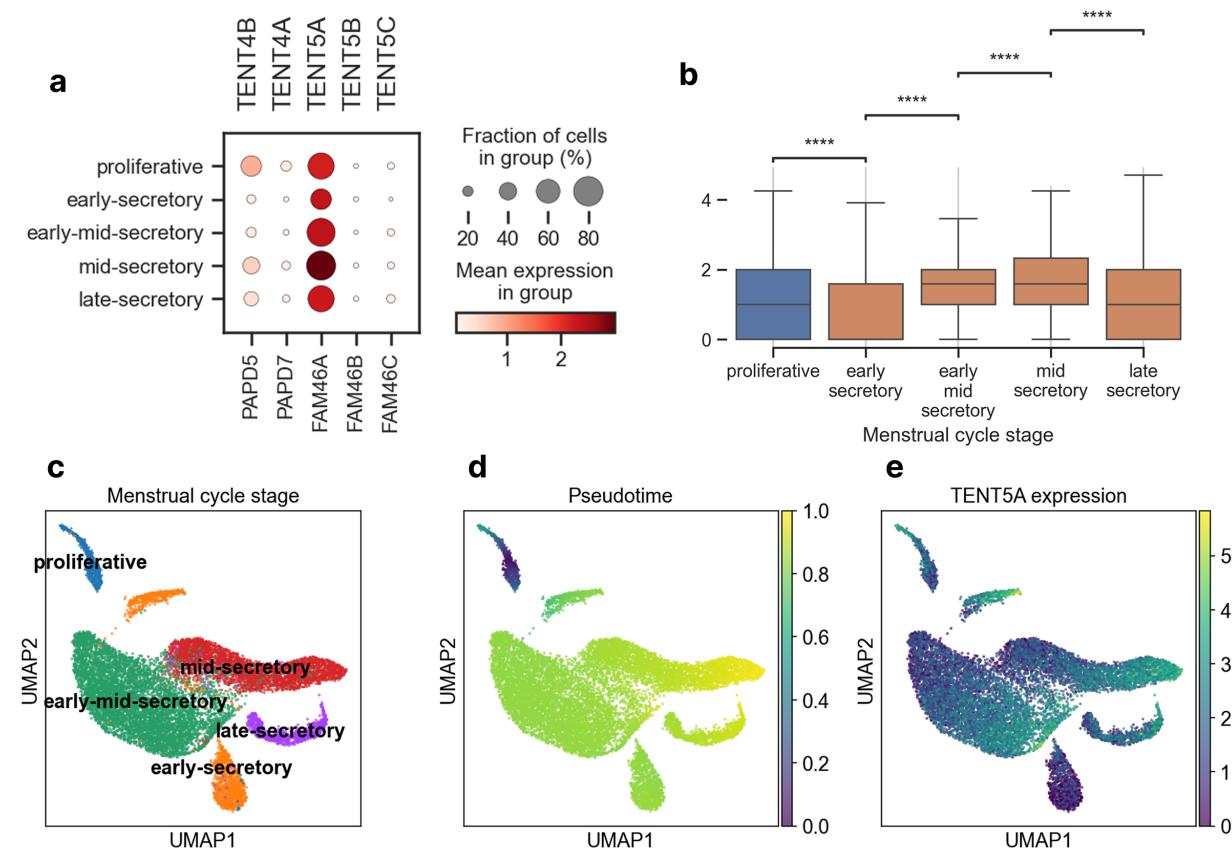


**Figure 2.** TENTs expression in human endometrium by menstrual cycle stages (a), broad cell types (c), and epithelial cell subtypes (d). b - expression of TENT5A in menstrual cycle with pairwise comparison of consecutive stages with the Mann-Whitney-Wilcoxon two-sided test and Benjamini-Hochberg correction. P-values: \*\*\*  $\leq 1e - 04$ ; \*\*  $\leq 1e - 02$ . Expression levels were log<sub>2</sub>-transformed.

The closer look at the glandular cells reveals strong and relatively stable TENT5A expression across the menstrual cycle stages. TENT5A levels increase in the early-mid-secretory and mid-secretory phase. The differences in expression levels between the consecutive phases of the menstrual cycle are statistically significant, but very small (Fig. 3b). Despite being significant, the differences are very small, which means they probably do not represent any meaningful regulatory process.

The glandular cells cluster by the menstrual cycle phase (Fig. 3c). That shows the transcriptomic profile of those cells changes as the menstrual cycle progresses. Additionally, the clusters containing cells of the secretory phases appear closer to each other than to the proliferative cell cluster. That indicates there is a strong transcriptomic shift between the proliferative and secretory phase.

The shift is also visible in the pseudotime visualization (Fig. 3d). Cells in the secretory phase are much further in pseudotime from the cell in the proliferative phase than they are to each other. Within each menstrual cycle phase TENT5A is more strongly expressed in the cells that are further in pseudotime (Fig. 3d, 3e). There is no apparent difference between the menstrual cycle stage and TENT5A expression.

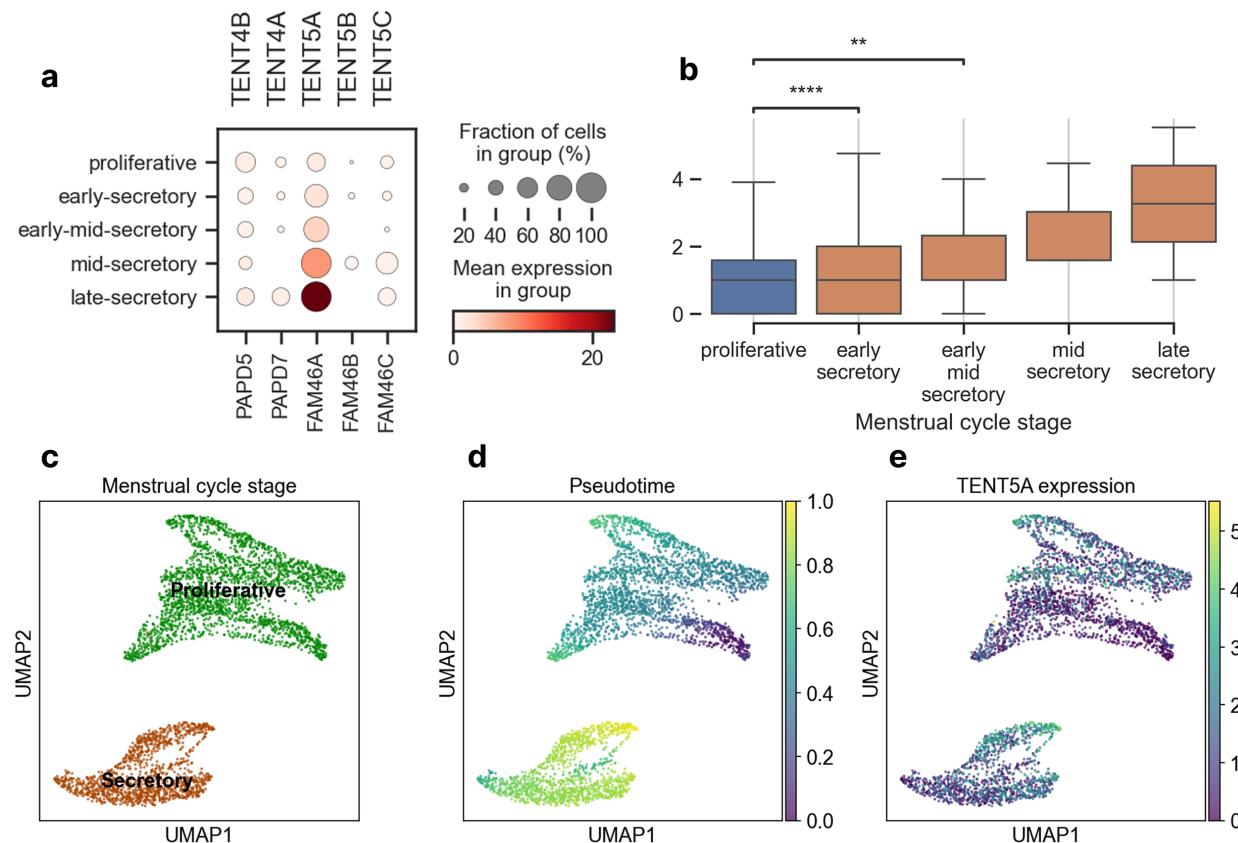


**Figure 3.** TENTs expression in glandular cells grouped by menstrual cycle stage (a). b - expression of TENT5A in glandular cells in menstrual cycle with pairwise comparison of adjacent stages with the two-sided Mann-Whitney-Wilcoxon test and Benjamini-Hochberg correction; p-values: \*\*\*  $\leq 1e-04$ . UMAP projection of the glandular cells coloured by the binary cycle stage (c), pseudotime (d), and TENT5A expression (e). Expression levels were log2-transformed.

On the other hand, when I looked closer at the SOX9 cells TENT5a exhibited a menstrual cycle related pattern (Fig. 4a, 4b). TENT5A levels were rising as the cycle progressed, which may be proved up to the early-mid-secretory stage. However, the differences between the SOX9 cells in the late menstrual cycles were not statistically significant (Fig. 4b). The number of SOX9 cells in the mid to late secretory phase was small, restraining from drawing conclusions in later phases.

The SOX9 cells clustered by the binary menstrual cycle stage (Fig. 4a) similarly to what was observed in the glandular cells. However, because of the small number of the SOX9 cells in the secretory stage, further subclustering of secretory cluster would not be meaningful. The transcriptomic profiles of the SOX9 cells strongly differ between the proliferative and a secretory phases.

The difference was also visible in the pseudotime, but it was not as apparent as the one observed in the glandular cells (Fig. 4d). Again, similarly to glandular cells, SOX9 cells showed an increasing TENT5A expression with the pseudotime progress within the clusters (Fig. 4d, 4e). No shift in TENT5A expression is visible in pseudotime between the menstrual cycle stages.



**Figure 4.** TENTs expression in SOX9 cells grouped by menstrual cycle stage (a). b - expression of TENT5A in SOX9 cells in menstrual cycle with pairwise comparison of all stages with the Mann-Whitney-Wilcoxon two-sided test and Benjamini-Hochberg correction; p-values: \*\*\*\*  $\leq 1e - 04$ ; \*\*  $\leq 1e - 02$ ; no star - non-significant. UMAP projection of the SOX9 cells coloured by the binary cycle stage (c), pseudotime (d), and TENT5A expression (e). Expression levels were log<sub>2</sub>-transformed.

## Discussion

I have confirmed that TENT4A, TENT4B, TENT5A, TENT5B and TENT5C are expressed in human endometrium, which was shown by the Human Protein Atlas single-cell data<sup>20</sup>. I observed that TENT5A levels change during the menstrual cycle - rising in the mid-secretory phase. TENT5A levels coincide with progesterone levels in the menstrual cycle stages<sup>1</sup>. The potential pathway between progesterone and TENT5A would be worth exploring.

TENT5A was most strongly express in the epithelial cells - especially glandular and SOX9 cells. The high TENT5A levels in glandular cells may be explained by the involvement of TENT5A in stabilizing RNA molecules coding for the secreted proteins. Poliadenylation of such RNA molecules by TENT5A have been already shown in the gametogenesis in mice, where TENT5A modified e.g. oocyte's zona pellucida protein encoding RNAs<sup>11</sup>. Similarly, TENT5A poliadenylates mRNAs encoding collagen and other proteins in osteogenesis<sup>15</sup>.

TENT5A may play a role in regulating menstrual cycle in humans, especially in epithelial cells. However, further experimental work would need to be conducted, to prove that. The experiments should focus on glandular and SOX9 cells specifically to obtain a larger number of them. Especially, the *in vitro* experiments could be helpful as obtaining epithelium at various stages of the cycle from the same donors would be excessively difficult and prone to external factors like variable cycle duration, stress, alcohol consumption. Simulating the menstrual cycle with hormones *in vitro* could produce more consistent and interpretable results.

## Methods and materials

The analyses were performed under Python 3.10.11 and via Visual Studio Code Jupyter extension (version 2025.5.0).

### 0.4 Data

RNA single-cell sequencing data publish in *Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro*<sup>2</sup> and available under Endometrium tab in the Reproductive Cell Atlas<sup>21</sup>. The data are available in the h5ad format provided by anndata<sup>22</sup>.

The data contain multiple descriptors for each sample: SampleID, Donor ID, Biopsy Type, Location, Binary Stage (of the menstrual cycle), Stage (of the menstrual cycle), Broad cell type, Cell type and Cell Cycle Phase. The data contained 21 samples coming from 15 donors and 2 biopsy types. Data were collected from 2 deceased transplant organ donors and 13 donors with superficial biopsies. The Binary Stage (of the menstrual cycle) of each cell was proliferative or secretory. The Stage (of the menstrual cycle) of each cell was one of the 5: proliferative, early-secretory, early-mid-secretory, mid-secretory, or late secretory.

The dataset contained 100307 cells and 28614 genes. The number of cells in each samples highly varied. Each of organ donors and one superficial biopsy donor provided around 17-20k cells, comprising together the most of the dataset. Fraction of cells in corresponding cell types also differed between the biopsy type. While epithelial and stromal cells prevailed in the superficial biopsies, supporting and stromal cells dominated in the organ donations. 75038 of obtained cells were in the secretory phase and 25269 were in the proliferative phase. Fraction of cells of each broad cell type in the decreasing order: Stromal, Epithelial, Supporting, Endothelial, Immune, and Excluded. 94213 cells were in the G1 phase, 3667 in the S phase, and 2427 in the G2M phase of the cell cycle.

## 0.5 Data preparation

Data were viewed and processed with Scanpy<sup>23</sup> and explored visually with Plotly (version 6.1.2). The data were already integrated using scVI<sup>24</sup> and log<sub>2</sub>-transformed, and cells with less than 500 expressed genes and more than 15% of mitochondrial genes were filtered out. Additionally, with `sc.pp.filter_genes` I further removed genes that were expressed in less than 10 cells, which reduced the number of genes from 28614 to 25839.

### 0.5.1 Doublet detection

Doublet detection was performed with Scrublet<sup>25</sup> via the Scanpy function `sc.pp.scrublet` run with 20 neighbors, sample ID as a batch key and a simulated doublet ratio equal 3.0. Percentage of detected doublets was around 0.007 indicating that most of the doublet had been already removed.

### 0.5.2 Batch effects

No batch correction were performed as not enough metadata were provided. Additionally, the batch effect analysis was performed (Supplement - clustering) to inspect, whether the cells cluster by donor ID. Indeed, cells tended to cluster by donor ID at lower level, but clustering by cell type and menstrual cycle stage was much stronger. Clustering by donor ID rather has biological grounds than batch effect as all the cells of one donor came from the same menstrual cycle stage.

## Dimensionality reduction

### 0.5.3 PCA

Principal Component Analysis (PCA)<sup>11</sup> was performed to reduce the dimensionality of the data before further analysis. 50 principal components were computed with `sc.tl.pca`, which uses the scikit-learn<sup>26</sup> implementation of PCA. Further, the elbow plot showing the proportion of variance that each principal component explains was prepared with `sc.pl.pca_variance_ratio`. As the elbow appeared around 20th principal component.

### 0.5.4 UMAP

As most of the variance is covered by 20 principal components, I picked this number to compute the nearest neighbour graph with `sc.pp.neighbors` with UMAP<sup>27</sup>, neighbourhood size of 50 and random state equal 123 for reproducibility. Later, I performed Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) with `sc.tl.umap` and the random state equal 123. Data reduced with UMAP were visualized with the `sc.pl.umap` coloured by desired feature. The code is available in the Supplement - clustering.

## 0.6 Expression

Expression of TENTs was visualized with dot plots and box plots described below. Expression levels on the plots are log<sub>2</sub>-transformed. Code is available under the Supplement - expression.

### 0.6.1 Dot plots

Dot plots present log<sub>2</sub>-transformed gene expression. They were prepared with `sc.pl.dotplot` from Scanpy<sup>23</sup>. Cells were grouped by the broad cell type, cell type or menstrual cycle stage with the `groupby` parameter. The gene numbers were mapped to TENT family names with a dictionary provided in the `var_names` parameter. Code is available under the Supplement - expression.

### 0.6.2 Box plots

Box plots were used to show the expression levels in more detail. They were prepared with Seaborn<sup>28</sup> `sns.boxplot`. To satisfy the function requirements expression data of TENT5A were extracted with pandas<sup>29</sup>. Pairwise comparison of the groups was calculated with the two-sided Mann-Whitney-Wilcoxon test with Benjamini-Hochberg correction from statannotations<sup>30</sup>. The plots were customized with Matplotlib<sup>31,32</sup>. Code is available under the Supplement - expression.

## 0.7 Trajectories

Trajectory analysis was performed with Scanpy<sup>23</sup>. Nearest neighbours graph was computed with `sc.pp.neighbors` with UMAP from 20 principal components and the neighbourhood size of 100 and the random state equal 123 for reproducibility. Clustering was performed with the Leiden algorithm<sup>33</sup> at resolution experimentally adjusted for each subset. Partition-based Graph Abstraction (PAGA)<sup>34</sup> was calculated with the `sc.tl.paga` function. PAGA graphs were visualized with `sc.pl.paga`, which enabled running UMAP with PAGA as an initial position. UMAP plots were prepared with `sc.pl.umap` and coloured with various features to inspect the clustering.

Later, the diffusion maps<sup>35,36</sup> were prepared with `sc.tl.diffmap` followed by pseudotime computation with `sc.tl.dpt`<sup>34,36</sup>. In all possible functions random state of 123 was used for reproducibility. Clustering, pseudotime and TENT5A expression were plotted side by side with `sc.pl.umap`. The code can be found under Supplement - trajectories..

## Code Availability

Code used for analyses is available under: <https://github.com/Silbena/msb/tree/main/project>.

## References

1. Critchley, H. O. D., Maybin, J. A., Armstrong, G. M. & Williams, A. R. W. Physiology of the endometrium and regulation of menstruation. *Physiol. Rev.* **100**, 1149–1179, DOI: [10.1152/physrev.00031.2019](https://doi.org/10.1152/physrev.00031.2019) (2020).
2. Garcia-Alonso, L. *et al.* Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nat. Genet.* **53**, 1698–1711, DOI: [10.1038/s41588-021-00972-2](https://doi.org/10.1038/s41588-021-00972-2) (2021).
3. Marečková, M. *et al.* An integrated single-cell reference atlas of the human endometrium. *Nat. Genet.* **56**, 1925–1937, DOI: [10.1038/s41588-024-01873-w](https://doi.org/10.1038/s41588-024-01873-w) (2024).
4. Garrido-Gomez, T. *et al.* Defective decidualization during and after severe preeclampsia reveals a possible maternal contribution to the etiology. *Proc. Natl. Acad. Sci.* **114**, DOI: [10.1073/pnas.1706546114](https://doi.org/10.1073/pnas.1706546114) (2017).
5. Fonseca, M. A. S. *et al.* Single-cell transcriptomic analysis of endometriosis. *Nat. Genet.* **55**, 255–267, DOI: [10.1038/s41588-022-01254-1](https://doi.org/10.1038/s41588-022-01254-1) (2023).
6. Cao, D. *et al.* Time-series single-cell transcriptomic profiling of luteal-phase endometrium uncovers dynamic characteristics and its dysregulation in recurrent implantation failures. *Nat. Commun.* **16**, 137, DOI: [10.1038/s41467-024-55419-z](https://doi.org/10.1038/s41467-024-55419-z) (2025).
7. Swain, U. *et al.* Tent4a non-canonical poly(a) polymerase regulates dna-damage tolerance via multiple pathways that are mutated in endometrial cancer. *Int. J. Mol. Sci.* **22**, 6957, DOI: [10.3390/ijms22136957](https://doi.org/10.3390/ijms22136957) (2021).
8. Passmore, L. A. & Coller, J. Roles of mrna poly(a) tails in regulation of eukaryotic gene expression. *Nat. Rev. Mol. Cell Biol.* **23**, 93–106, DOI: [10.1038/s41580-021-00417-y](https://doi.org/10.1038/s41580-021-00417-y) (2022).
9. Liudkowska, V. & Dziembowski, A. Functions and mechanisms of rna tailing by metazoan terminal nucleotidyltransferases. *WIREs RNA* **12**, DOI: [10.1002/wrna.1622](https://doi.org/10.1002/wrna.1622) (2021).
10. Warkocki, Z., Liudkowska, V., Gewartowska, O., Mroczek, S. & Dziembowski, A. Terminal nucleotidyl transferases (tents) in mammalian rna metabolism. *Philos. Transactions Royal Soc. B: Biol. Sci.* **373**, 20180162, DOI: [10.1098/rstb.2018.0162](https://doi.org/10.1098/rstb.2018.0162) (2018).

11. Brouze, M. *et al.* Tent5-mediated polyadenylation of mrnas encoding secreted proteins is essential for gametogenesis in mice. *Nat. Commun.* **15**, 5331, DOI: [10.1038/s41467-024-49479-4](https://doi.org/10.1038/s41467-024-49479-4) (2024).
12. Liudkovska, V. *et al.* Tent5 cytoplasmic noncanonical poly(a) polymerases regulate the innate immune response in animals. *Sci. Adv.* **8**, DOI: [10.1126/sciadv.add9468](https://doi.org/10.1126/sciadv.add9468) (2022).
13. Bilska, A. *et al.* Immunoglobulin expression and the humoral immune response is regulated by the non-canonical poly(a) polymerase tent5c. *Nat. Commun.* **11**, 2032, DOI: [10.1038/s41467-020-15835-3](https://doi.org/10.1038/s41467-020-15835-3) (2020).
14. Luo, M. *et al.* Tent5a modulates muscle fiber formation in adolescent idiopathic scoliosis via maintenance of myogenin expression. *Cell Prolif.* **55**, DOI: [10.1111/cpr.13183](https://doi.org/10.1111/cpr.13183) (2022).
15. Gewartowska, O. *et al.* Cytoplasmic polyadenylation by tent5a is required for proper bone formation. *Cell Reports* **35**, 109015, DOI: [10.1016/j.celrep.2021.109015](https://doi.org/10.1016/j.celrep.2021.109015) (2021).
16. Lin, H.-H. *et al.* Overexpression of fam46a, a non-canonical poly(a) polymerase, promotes hemin-induced hemoglobinization in k562 cells. *Front. Cell Dev. Biol.* **8**, DOI: [10.3389/fcell.2020.00414](https://doi.org/10.3389/fcell.2020.00414) (2020).
17. Wang, Y. *et al.* Fam46a expression is elevated in glioblastoma and predicts poor prognosis of patients. *Clin. Neurol. Neurosurg.* **201**, 106421, DOI: [10.1016/j.clineuro.2020.106421](https://doi.org/10.1016/j.clineuro.2020.106421) (2021).
18. Min, X. *et al.* Tent5a mediates the cancer-inhibiting effects of egr1 by suppressing the protein stability of rpl35 in hepatocellular carcinoma. *Cell. Oncol.* **47**, 2247–2264, DOI: [10.1007/s13402-024-01014-9](https://doi.org/10.1007/s13402-024-01014-9) (2024).
19. Krawczyk, P. S. *et al.* Re-adenylation by tent5a enhances efficacy of sars-cov-2 mRNA vaccines. *Nature* DOI: [10.1038/s41586-025-08842-1](https://doi.org/10.1038/s41586-025-08842-1) (2025).
20. Regev, A. *et al.* The human cell atlas. *eLife* **6**, DOI: [10.7554/eLife.27041](https://doi.org/10.7554/eLife.27041) (2017).
21. Reproductive cell atlas. [www.reproductivecellatlas.org/non-pregnant-uterus.html](http://www.reproductivecellatlas.org/non-pregnant-uterus.html).
22. Virshup, I., Rybakov, S., Theis, F. J., Angerer, P. & Wolf, F. A. anndata: Access and store annotated data matrices. *J. Open Source Softw.* **9**, 4371, DOI: [10.21105/joss.04371](https://doi.org/10.21105/joss.04371) (2024).
23. Wolf, F. A., Angerer, P. & Theis, F. J. Scanpy: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15, DOI: [10.1186/s13059-017-1382-0](https://doi.org/10.1186/s13059-017-1382-0) (2018).
24. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058, DOI: [10.1038/s41592-018-0229-2](https://doi.org/10.1038/s41592-018-0229-2) (2018).
25. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9, DOI: [10.1016/j.cels.2018.11.005](https://doi.org/10.1016/j.cels.2018.11.005) (2019).
26. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
27. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *ArXiv* (2020).
28. Waskom, M. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021, DOI: [10.21105/joss.03021](https://doi.org/10.21105/joss.03021) (2021).
29. pandas development team, T. pandas-dev/pandas: Pandas, DOI: [10.5281/zenodo.13819579](https://doi.org/10.5281/zenodo.13819579) (2024).
30. Charlier, F. *et al.* Statannotations, DOI: [10.5281/zenodo.14258156](https://doi.org/10.5281/zenodo.14258156) (2024).
31. Hunter, J. D. Matplotlib: A 2d graphics environment. *Comput. Sci. & Eng.* **9**, 90–95, DOI: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55) (2007).

32. Team, T. M. D. Matplotlib: Visualization with python, DOI: [10.5281/zenodo.15375714](https://doi.org/10.5281/zenodo.15375714) (2025).
33. Traag, V. A., Waltman, L. & van Eck, N. J. From louvain to leiden: guaranteeing well-connected communities. *Sci. Reports* **9**, 5233, DOI: [10.1038/s41598-019-41695-z](https://doi.org/10.1038/s41598-019-41695-z) (2019).
34. Wolf, F. A. *et al.* Paga: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59, DOI: [10.1186/s13059-019-1663-x](https://doi.org/10.1186/s13059-019-1663-x) (2019).
35. Coifman, R. R. & Lafon, S. Diffusion maps. *Appl. Comput. Harmon. Analysis* **21**, 5–30, DOI: [10.1016/j.acha.2006.04.006](https://doi.org/10.1016/j.acha.2006.04.006) (2006).
36. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998, DOI: [10.1093/bioinformatics/btv325](https://doi.org/10.1093/bioinformatics/btv325) (2015).