SMARTdb: An Integrated Database for Exploring Single-cell Multi-omics Data of Reproductive Medicine

Zekai Liu[1,#], Zhen Yuan[1,#], Yunlei Guo[1,#], Ruilin Wang[1], Yusheng Guan[1], Zhanglian Wang[1], Yunan Chen[1], Tianlu Wang[1], Meining Jiang[1], Shuhui Bian[1,2,*]

[1]*State Key Laboratory of Reproductive Medicine and Offspring Health, Nanjing Medical University, Nanjing 211166, China*

[2]*Collaborative Innovation Center for Cancer Personalized Medicine, School of Public Health, Nanjing Medical University, Nanjing 211166, China*

[#]Equal contribution.

[*]Corresponding author.

E-mail: bianshuhui@njmu.edu.cn (Bian S).

Running title: *Liu Z et al / SMARTdb: A Single-cell Omics Database of Reproduction*

Abstract

Single-cell multi-omics sequencing has greatly accelerated reproductive research in recent years, and the data are continually growing. However, utilizing these data resources is challenging for wet-lab researchers. A comprehensive platform for exploring single-cell multi-omics data related to reproduction is urgently needed. Here we introduce SMARTdb (single-cell multi-omics atlas of reproduction), which is an integrative and user-friendly platform for exploring molecular dynamics of reproductive development, aging, and disease, covering multi-omics, multi-species, and multi-stage data. We have curated and analyzed single-cell transcriptome and epigenome data of over 2.0 million cells from 6 species across whole lifespan. A series of powerful functionalities are provided, such as "Query gene expression", "DIY expression plot", "DNA methylation plot", and "Epigenome browser". With SMARTdb, we found that the male germ-cell-specific expression pattern of *RPL39L* and *RPL10L* is conserved between human and other model animals. Moreover, DNA hypomethylation and open chromatin may regulate the specific expression pattern of *RPL39L* collectively in both male and female germ cells. In summary, SMARTdb is a powerful platform for convenient data mining and gaining novel insights into reproductive development, aging, and disease. SMARTdb is publicly available at https://smart-db.cn.

Keywords: Database; Single-cell multi-omics; Embryo; Germ cell; Gonad

Introduction

Single-cell multi-omics technique has greatly promoted our molecular understanding of early embryos [1–6], fetal germ cells and gonadal somatic cells [7–11], developing testis and ovary [12–16], adult testis and ovary [17–26], aging testis and ovary [27–29], endometrium [30,31], maternal-fetal interface [32–34], as well as reproductive diseases [35–39] in the last decade. Multiple molecular layers (including gene expression patterns, DNA methylation dynamics, and open chromatin features) have been profiled with high-throughput sequencing technology, and the data have been growing continually. It is noted that previous single-cell multi-omics studies typically provide comprehensive atlas rather than dig into individual genes or genomic sites, but they offer valuable resources for more wet-lab researchers to perform further investigations into specific questions, genes, and genomic sites of interest. For example, it is important for researchers using model animals to know if the expression pattern of a specific gene is conserved in humans, and whether knocking out of the gene would potentially result in important phenotypes.

However, analyzing the massive and complex single-cell multi-omics data requires experienced skills and a large number of computing resources, which is challenging and time-consuming for wet-lab researchers especially. For instance, as each of those studies typically focuses on a narrow time span, researchers who wish to query the expression levels of a particular gene in child, young adult, aging adult and infertile man need to download and process large amount of data from several separate studies, and they might find this process difficult and laborious due to a lack of analytical skills and computing resources. This inconvenience hinders the optimal utilization of the published data, thus limiting potential novel discoveries. Hence, integrated and user-friendly platforms are needed to bridge the gap between rich data resource and wet-lab researchers.

Endeavors such as human cell atlas (HCA, https://www.humancellatlas.org/) and human protein atlas (HPA, https://www.proteinatlas.org/) have served as excellent platforms for exploring massive single-cell data conveniently. However, these databases do not contain much reproduction-related data so far. Currently, there are a few small-scale reproduction-related websites available, which typically only include data of their own study, are limited in data amount and time span, only provide simple function of querying gene expression of a specific gene, and lack long-term

maintenance. As a result, there is still an urgent need for an integrated and user-friendly platform which focuses on promoting efficient utilization of public single-cell multi-omics data related to reproduction across full life-cycle.

Here we introduce SMARTdb (single-cell multi-omics atlas of reproduction), which is a manually curated, integrative and interactive platform aiming to bridge the gap between mass data resources and wet-lab researchers, and promote exploring molecular dynamics of reproductive development, aging, and disease.

Data collection and processing

Data collection

We curated single-cell transcriptome and epigenome (DNA methylation and chromatin accessibility) data related to reproduction from published studies (Table S1). The gene expression matrix, DNA methylation, and chromatin accessibility data were downloaded from NCBI GEO (https://www.ncbi.nlm.nih.gov/geo), EMBL-EBI (https://www.ebi.ac.uk/), and Zenodo (https://zenodo.org/). In addition to data of developmental stages and normal physiological states, data of aging and human infertility (such as non-obstructive azoospermia) are also incorporated. In summary, the single-cell data of over 2.0 million individual cells across 6 species (human, monkey, mouse, pig, buffalo, and goat) were curated, involving early embryos, fetal germ cells and gonadal somatic cells, developing testis and ovary, adult testis and ovary, aging testis and ovary, endometrium, maternal-fetal interface, and human infertility. These data cover the entire life cycle, and span > 120 specific time points of 6 species (Table S1).

Processing of single-cell transcriptome data and cell clustering

To reproduce the cell clustering results of original studies as much as possible, for the datasets with metadata provided by the original study, only the cells with cell type identities were used in downstream analyses. For the datasets without provided metadata, the cells were filtered according to the quality control standards mentioned in the original studies. Seurat [40] (version 4.3.0) was used to perform the following processing steps. Briefly, the unique molecular identifier (UMI) count of each gene for each cell was first divided by the total counts for that cell and multiplied by 10,000 (or author-designated parameters), and was then natural-log transformed. The top highly

variable genes among cells were selected with "FindVariableFeatures" function with default or author-designated parameters. The normalized data were then centered and scaled, resulting in that the mean expression of each gene across all cells was zero and the standard deviation was one. Principal component analysis (PCA) was performed on the scaled data with the previously determined highly variable genes used as input. Then the cells were clustered with Louvain algorithm. The uniform manifold approximation and projection (UMAP) was used to visualize the cell clustering in low-dimensional space. For the datasets with obvious batch effects, the anchors of different samples were identified with the "FindIntegrationAnchors" function, and then different samples were integrated to correct batch effects with the "IntegrateData" function. For integration of datasets from different studies, strict quality control standards were used to eliminate potential effects of poor quality cells. Only cells with a gene number > 1000 and a mitochondrial gene ratio < 20% were used in downstream analyses. For the datasets without obvious batch effects, no additional correction was performed.

Cell type annotation and differentially expressed gene analyses of single-cell transcriptome data

A series of analyses were performed to ensure the accuracy of cell identities for each dataset and integrated datasets. For the datasets with metadata provided by the original study, the cells were annotated correspondingly, and classical marker genes for each cell type were checked to confirm accuracy. For the datasets without metadata, the top differentially expressed genes (DEGs) of each cell cluster were identified using the Wilcoxon rank sum test with the "FindAllMarkers" function in Seurat [40] (version 4.3.0) with default parameters (logfc.threshold = 0.25, min.pct = 0.1, return.thresh = 0.01), and their consistency with classical marker genes for each cell type was carefully checked. For the integrated datasets, the cell identities after integration were further checked with the cell type information identified by each dataset before integration. After these steps, the cell type identities can be annotated accurately.

Visualization of gene expression levels

First, Seurat objects were converted into "anndata" objects with the R package sceasy (version 0.0.7). Then, the dot plots, heatmaps and violin plots were generated with Scanpy [41] (version 1.9.3). For dot plots and heatmaps, both the relative expression

levels with and without standard scaling are provided. The standard scale normalization strategy in Scanpy was performed with parameters "standard_scale = 'var'", which standardized the expression level of a given gene between 0 and 1 by subtracting the minimum and dividing each by its maximum.

Processing of single-cell epigenome data

For the epigenetic data generated by scBS-seq [42] and scCOOL-seq [43] techniques, the processed single-cell DNA methylation and chromatin accessibility data provided by the original study were downloaded from GEO database, and were further processed to standard BigWig format with bedtools (version 2.30.0), UCSC bedGraphToBigWig (version 2.9) and bigWigMerge (version 2) for downstream usage. For the data generated by scBS-seq, only CpG sites were retained. For the data generated by scCOOL-seq, WCG (W = A, T) sites were used for DNA methylation analysis and GCH (H = A, C, T) sites were used for chromatin accessibility analysis. For chromatin accessibility data generated by scATAC-seq, the peak regions of each cell type were mainly collected from the authors of the original study.

Considering the sparse coverage of single-cell epigenome data, we merged single-cell epigenome data according to their cell type identities and stages (*e.g.* 2-cell embryo, 8-week female mitotic germ cells, *etc.*), and sequencing techniques (*e.g.* scBS-seq and scCOOL-seq) using custom scripts. Briefly, for the genomic sites only covered by one cell, they were directly retained in the merged files. For the genomic sites covered by multiple cells in the group, the mean values of covered cells were calculated to represent the final states of the sites. For single-cell DNA methylation data, only the genomic sites with DNA methylation levels < 0.1 or > 0.9 in each individual cell were used in downstream merging. The merged single-cell data of cell groups are more convenient and efficient for downstream exploration.

The DNA methylation "Tanghulu" plot was implemented by a custom Python script to visualize DNA methylation state of each CpG site in a narrow genomic region of interest. For single-cell DNA methylation data, only the genomic sites with values < 0.1 or > 0.9 in each individual cell will be plotted. The DNA methylation level is defined as the percentage of methylated sites in all sites. The WashU epigenome browser [44] was incorporated into SMARTdb for convenient visualization of single-cell DNA methylation levels and chromatin accessibility states.

Implementation of the database website

SMARTdb was developed utilizing Python 3 in conjunction with the Django framework (https://djangoproject.com). The data housed within SMARTdb is systematically organized and accessible through PostgreSQL (https://www.postgresql.org). The user-friendly front-end interface was created using NG-ZORRO and Angular. The platform is hosted using the Nginx web server (http://nginx.org) and is freely accessible online at https://smart-db.cn. For optimal performance, SMARTdb is best used on major web browsers such as Google Chrome, Microsoft Edge, Safari, and Mozilla Firefox, with JavaScript enabled.

Database content and usage

Overall design

SMARTdb is an integrative and interactive platform for exploring single-cell multi-omics data related to reproduction. It was designed to have three main features (**Figure 1** and Figure S1). First, single-cell multi-omics data of over 2.0 million individual cells were manually curated and analyzed (Table S1). Users can access gene expression levels, DNA methylation levels and chromatin accessibility levels at single-cell resolution conveniently. Second, the main stages of full life-cycle are covered, including zygote, early embryo, fetus, infant, child, puberty, adult, and aging. More than 120 specific time points are included in total (Table S1). Besides, data of human infertility are also incorporated. Users can not only explore molecular dynamics during full life-cycle, but also compare normal physiological state with disease. Third, considering cross-species conservation analyses are very important, data across six species are included, including human, non-human primate (monkey), mouse, pig, buffalo, and goat. For example, the researchers using mouse as model animals can find out if the expression pattern of the gene of their interest is conserved in human, which will help select more meaningful research targets.

As a user-friendly platform, SMARTdb offers multiple featured function modules to users for exploring the rich single-cell multi-omics data, such as "Query gene expression", "DIY expression plot", "DNA methylation plot" and "Epigenome browser". The featured function modules are described below.

"Overview" and "Query gene expression" modules

In the "Overview" module, SMARTdb facilitates users to get a quick overview of cell clustering, samples, cell counts, and DEGs of each cell type for each scRNA-seq dataset (**Figure 2**A). In the "Query gene expression" module, users can conveniently plot the expression levels of a selected gene across the cell types in the chosen dataset with uniform manifold approximation and projection (UMAP) and violin plots (Figure 2B). Additionally, a concise introduction for each human gene is also provided in this module.

"DIY expression plot" module

This module provides a very practical function for users to customize expression plots according to their personalized requirements (Figure 2C). Users can select multiple genes and multiple cell types of interest to plot simultaneously, and choose appropriate plot forms for visualization. Three main visualization plots are offered in this module, including dot plots, heatmaps, and violin plots. Moreover, the orders of genes and cell types displayed in the plots can be designated by users freely. Users can also choose whether to standardly scale the expression levels within each gene when using dot plots and heatmaps. The generated figures are of publication quality.

"DNA methylation plot" module

This module provides "Tanghulu" plots to visualize DNA methylation status of each cytosine site in CpG or WCG context (**Figure** 3A). Users can conveniently generate a customized "Tanghulu" plot by selecting datasets and cell types, and designating specific genomic region range of interest (such as promoters and gene bodies). In "Tanghulu" plot, each row represents an individual cell, and each circle represents a cytosine site in CpG or WCG context. Black circles stand for methylated sites, while the white circles refer to unmethylated sites. The DNA methylation levels of selected region in each single cell are labeled on "Tanghulu" plots.

"Epigenome browser" module

This module incorporates the WashU epigenome browser tool [44] and offers a state-of-the-art solution for browsing epigenetic datasets (Figure 3B). Users can select single-cell DNA methylation and chromatin accessibility datasets of various cell types based

240 on their interest, such as oocytes, sperms, early embryos, fetal germ cells, and gonadal
241 somatic cells. The DNA methylation and chromatin accessibility levels of selected
242 samples can be directly and conveniently visualized in the browser. Especially, since
243 single-cell data are typically sparse, the single-cell data of each cell type or sample type
244 are merged for optimal visualization and enabling more efficient exploration. The
245 genomic region locator allows users to navigate to a specific gene or designated specific
246 genomic regions (such as promoters and gene bodies), zoom in and out, and move to
247 new regions by dragging the mouse left and right. Additionally, the browser interface
248 is allowed to be customized to meet personalized requirements, such as display mode
249 (bar plot or heatmap), RGB colors, axis scales, and text labels.
250
251 **Examples for using SMARTdb**
252 Here we illustrate how to make good use of SMARTdb to facilitate exploration of
253 single-cell multi-omics data and promote new findings. A recent study has revealed that
254 RPL39L is a male germ-cell-specific ribosome protein and plays critical roles in
255 spermatogenesis and male fertility in mouse, whereas the corresponding core ribosome
256 protein RPL39 is mainly expressed in somatic cells [45]. Inspired by the study, we
257 further asked: (1) whether the germ-cell-specific expression pattern is conserved in
258 testis of human, non-human primates as well as other model animals? (2) when does
259 the germ-cell-specific expression pattern of RPL39L form? (3) does epigenetic
260 regulation function in regulating the male germ-cell-specific expression pattern? We
261 will demonstrate how to explore the three questions using SMARTdb below.
262
263 *Cross-species conservation*
264 We utilized the "DIY expression plot" module to plot the expression levels of the
265 ribosomal large subunit protein genes *RPL39L* and *RPL10L*, and the corresponding core
266 ribosome protein genes *RPL39* and *RPL10* in human [23], as well as their homologous
267 genes in mouse [46], monkey [19], and pig [15] (Figures 2C and 4A). For human adult
268 testis [23], we found that the expression levels of *RPL39L* and *RPL10L* peak at
269 spermatocyte, and then reduce in later stages of spermatogenesis at mRNA level (Figure
270 2B and C). In contrast, *RPL39* and *RPL10* decrease their expression levels during
271 spermatogenesis, and are mainly expressed in somatic cells (Figure 2C). Moreover, the

germ-cell-specific expression pattern of *RPL39L* and *RPL10L* is conserved in human and other species (**Figure** 4A).

Developing dynamics

To clarify the second question, we plotted the expression levels of *RPL39L* in fetal stage [10] and peri-pubertal stage [13] with the "DIY expression plot" module (Figure 4B).We found that the germ cells in fetal stages have started to highly express *RPL39L*, as early as in migrating primordial germ cells (PGCs). Moreover, we noticed that *RPL39L* transcripts are highly expressed in both male and female germ cells in human fetal stage, but diverge between male and female in adulthood. In human adulthood, *RPL39L* is relatively highly expressed in male germ cells and female growing oocytes, but lowly expressed in fully-grown, metaphase I (MI), and metaphase II (MII) oocytes [21] (**Figure** 5A). The divergence in the *RPL39L* expression among oocytes of various stages in human adult ovary was detected in mRNA level, indicating potential transcriptional regulation of the gene.

Regulation mechanism

Then we further explored the transcriptional regulatory mechanism using the "Epigenome browser" module. With epigenetic data of human fetal stage [7], we found that germ cells in fetal stage have hypomethylated and more accessible regions around the promoter of *RPL39L* compared with somatic cells (Figure 3B and Figure S2), which is consistent with the relatively high expression of *RPL39L* of fetal germ cells (Figure 4B). For human adult female [21], growing oocytes displayed similar DNA methylation and chromatin accessibility pattern with fetal germ cells (Figure 5B). In contrast, fully-grown, MI, and MII oocytes up-regulate the DNA methylation levels around *RPL39L* promoter, and down-regulate the chromatin accessibility levels around *RPL39L* promoter (Figure 5B), which may explain the decreased expression level of *RPL39L* in fully-grown, MI, and MII oocytes. Hence, epigenetic regulation potentially plays important roles in shaping the expression pattern of *RPL39L* in both male and female human germ cells during both fetal and adult stages.

Discussion

The mystery of reproduction-related development, aging and disease has long been explored. Single-cell multi-omics sequencing technique has revolutionized researches

in reproduction in recent years, and has led to accelerated accumulation of massive datasets. The public data provide valuable resources for researchers to perform further data mining. However, difficulties in processing big data for broad wet-lab researchers prevent them from being effectively utilized. To deal with the situation of growing data and promote a more efficient usage, we present SMARTdb – a single-cell multi-omics platform for reproductive research. To the best of our knowledge, SMARTdb is the first integrated platform with multiple powerful functionalities to explore single-cell multi-omics data in the field of reproductive medicine. Moreover, it will be continuously updated by adding high-quality single-cell multi-omics data.

Our database will promote reproductive medicine research in several aspects. For example, researchers using model animals can speculate whether knocking out of a specific gene could potentially lead to important phenotypic changes. Users are able to know whether the expression pattern of a specific gene is conserved in human, which will help select more meaningful research targets. For researcher using human samples, cross-species conservation analyses provide important clues for selecting suitable model animals. Cross-stage dynamic analysis will help select the right time window for further investigations. Moreover, epigenetics data resources will provide novel insights for potential regulatory mechanisms.

We showed examples for using SMARTdb and demonstrated the power of it by exploring germ-cell-specific genes. We not only validated the germ-cell-specific expression of ribosome gene *RPL39L*, but also gained a deeper understanding of its cross-species conservation, cross-stage expression dynamics, and potential epigenetic regulation. Moreover, taking the advantages of single-cell transcriptome sequencing, the gene expression patterns in multiple cell types can be observed more accurately and in detail. For example, we found that *RPL39L* is not only specifically expressed in male germ cells in human adult testis, but also highly expressed in the growing oocytes (one subtype of oocytes) in human adult ovary. Furthermore, DNA methylation and chromatin accessibility may cooperate in regulating the divergence of *RPL39L* expression in germ cells of human adult at mRNA level. The results demonstrate the power of SMARTdb, providing valuable clues for further investigations.

In summary, we firmly believe that with the assistance of SMARTdb, researchers in the field of reproduction can easily access and explore massive single-cell multi-omics data, and hopefully generate novel ideas and new findings.

Data availability

SMARTdb is publicly available at https://smart-db.cn.

CRediT author statement

Shuhui Bian: Conceptualization, supervision, Writing - original draft, Writing - review & editing, Visualization, Investigation, Data curation, Methodology, Formal analysis, Funding acquisition, Project administration. **Zekai Liu:** Investigation, Data curation, Formal analysis, Methodology, Software, Visualization, Validation. **Zhen Yuan:** Investigation, Data curation, Formal analysis, Methodology, Software, Visualization, Validation, Writing - original draft. **Yunlei Guo:** Investigation, Formal analysis, Methodology, Software, Visualization, Validation. **Ruilin Wang:** Investigation, Formal analysis, Validation. **Yusheng Guan:** Investigation, Data curation, Methodology, Validation. **Zhanglian Wang**: Investigation, Data curation. **Yunan Chen**: Investigation, Data curation. **Tianlu Wang**: Investigation, Data curation. **Meining Jiang:** Investigation, Data curation.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

We thank professors Fuchou Tang (Peking University), Jingchu Luo (Peking University), Zhibin Hu (Nanjing Medical University), Jiahao Sha (Nanjing Medical University), Fan Guo (Institute of Zoology, Chinese Academy of Sciences), Ran Huo (Nanjing Medical University), and Boqiang Hu (Zhejiang University) for discussions and suggestions. This work was supported by the Young Elite Scientists Sponsorship Program by China Association for Science and Technology (CAST) (Grant No. 2020QNRC001) and research start-up funding from Nanjing Medical University, China (Grant No. KY116RC20200007). All authors read and approved the final manuscript.

ORCID

0009-0003-2261-1216 (Zekai Liu)
0000-0003-4980-6681 (Zhen Yuan)
0009-0001-2455-5508 (Yunlei Guo)
0000-0002-8420-9033 (Ruilin Wang)
0000-0002-1319-7784 (Yusheng Guan)
0009-0001-5068-3853 (Zhanglian Wang)
0009-0006-6175-0042 (Yunan Chen)
0009-0000-1145-8306 (Tianlu Wang)
0009-0001-9423-0151 (Meining Jiang)
0000-0002-9662-113X (Shuhui Bian)

References

1. Zhu P, Guo H, Ren Y, Hou Y, Dong J, Li R, et al. Single-cell DNA methylome sequencing of human preimplantation embryos. Nat Genet. 2018;50:12–9.

2. Zhou F, Wang R, Yuan P, Ren Y, Mao Y, Li R, et al. Reconstituting the transcriptome and DNA methylome landscapes of human implantation. Nature. 2019;572:660–4.

3. Wang Y, Yuan P, Yan Z, Yang M, Huo Y, Nie Y, et al. Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. Nat Commun. 2021;12:1247.

4. Argelaguet R, Clark SJ, Mohammed H, Stapel LC, Krueger C, Kapourani C-A, et al. Multi-omics profiling of mouse gastrulation at single-cell resolution. Nature. 2019;576:487–91.

5. Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, et al. A single-cell molecular map of mouse gastrulation and early organogenesis. Nature. 2019;566:490–5.

6. Li L, Guo F, Gao Y, Ren Y, Yuan P, Yan L, et al. Single-cell multi-omics sequencing of human early embryos. Nat Cell Biol. 2018;20:847–58.

7. Li L, Li L, Li Q, Liu X, Ma X, Yong J, et al. Dissecting the epigenomic dynamics of human fetal germ cell development at single-cell resolution. Cell Res. 2021;31:463–77.

8. Chitiashvili T, Dror I, Kim R, Hsu F-M, Chaudhari R, Pandolfi E, et al. Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. Nat Cell Biol. 2020;22:1436–46.

9. Garcia-Alonso L, Lorenzi V, Mazzeo CI, Alves-Lopes JP, Roberts K, Sancho-Serra C, et al.

Single-cell roadmap of human gonadal development. Nature. 2022;607:540–7.

10. Li L, Dong J, Yan L, Yong J, Liu X, Hu Y, et al. Single-cell RNA-seq analysis maps development of human germline cells and gonadal niche interactions. Cell Stem Cell. 2017;20:858-873.e4.

11. Chen M, Long X, Chen M, Hao F, Kang J, Wang N, et al. Integration of single-cell transcriptome and chromatin accessibility of early gonads development among goats, pigs, macaques, and humans. Cell Rep. 2022;41:111587.

12. Guo J, Sosa E, Chitiashvili T, Nie X, Rojas EJ, Oliver E, et al. Single-cell analysis of the developing human testis reveals somatic niche cell specification and fetal germline stem cell establishment. Cell Stem Cell. 2021;28:764-778.e4.

13. Guo J, Nie X, Giebler M, Mlcochova H, Wang Y, Grow EJ, et al. The dynamic transcriptional cell atlas of testis development during human puberty. Cell Stem Cell. 2020;26:262-276.e4.

14. Zhao J, Lu P, Wan C, Huang Y, Cui M, Yang X, et al. Cell-fate transition and determination analysis of mouse male germ cells throughout development. Nat Commun. 2021;12:6839.

15. Zhang L, Guo M, Liu Z, Liu R, Zheng Y, Yu T, et al. Single-cell RNA-seq analysis of testicular somatic cell development in pigs. J Genet Genomics. 2022;49:1016–28.

16. Gu C, Liu S, Wu Q, Zhang L, Guo F. Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. Cell Res. 2019;29:110–23.

17. Chen Y, Zheng Y, Gao Y, Lin Z, Yang S, Wang T, et al. Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. Cell Res. 2018;28:879–96.

18. Sohni A, Tan K, Song H-W, Burow D, de Rooij DG, Laurent L, et al. The neonatal and adult human testis defined at the single-cell level. Cell Rep. 2019;26:1501-1517.e4.

19. Shami AN, Zheng X, Munyoki SK, Ma Q, Manske GL, Green CD, et al. Single-cell RNA sequencing of human, macaque, and mouse testes uncovers conserved and divergent features of mammalian spermatogenesis. Dev Cell. 2020;54:529-547.e12.

20. Fan X, Bialecka M, Moustakas I, Lam E, Torrens-Juaneda V, Borggreven NV, et al. Single-cell reconstruction of follicular remodeling in the human adult ovary. Nat Commun. 2019;10:3164.

21. Yan R, Gu C, You D, Huang Z, Qian J, Yang Q, et al. Decoding dynamic epigenetic landscapes in human oocytes using single-cell multi-omics sequencing. Cell Stem Cell. 2021;28:1641-1656.e7.

22. Zhang Y, Yan Z, Qin Q, Nisenblat V, Chang H-M, Yu Y, et al. Transcriptome landscape of human folliculogenesis reveals oocyte and granulosa cell interactions. Mol Cell. 2018;72:1021-1034.e4.

23. Guo J, Grow EJ, Mlcochova H, Maher GJ, Lindskog C, Nie X, et al. The adult human testis transcriptional cell atlas. Cell Res. 2018;28:1141–57.

24. Hermann BP, Cheng K, Singh A, Roa-De La Cruz L, Mutoji KN, Chen I-C, et al. The

mammalian spermatogenesis single-cell transcriptome, from spermatogonial stem cells to spermatids. Cell Rep. 2018;25:1650-1667.e8.

25. Huang Y, Li L, An G, Yang X, Cui M, Song X, et al. Single-cell multi-omics sequencing of human spermatogenesis reveals a DNA demethylation event associated with male meiotic recombination. Nat Cell Biol. 2023;25:1520–34.

26. Wang M, Liu X, Chang G, Chen Y, An G, Yan L, et al. Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. Cell Stem Cell. 2018;23:599-614.e4.

27. Nie X, Munyoki SK, Sukhwani M, Schmid N, Missel A, Emery BR, et al. Single-cell analysis of human testis aging and correlation with elevated body mass index. Dev Cell. 2022;57:1160-1176.e5.

28. Huang D, Zuo Y, Zhang C, Sun G, Jing Y, Lei J, et al. A single-nucleus transcriptomic atlas of primate testicular aging reveals exhaustion of the spermatogonial stem cell reservoir and loss of Sertoli cell homeostasis. Protein Cell. 2022;pwac057.

29. Wang S, Zheng Y, Li J, Yu Y, Zhang W, Song M, et al. Single-cell transcriptomic atlas of primate ovarian aging. Cell. 2020;180:585-600.e19.

30. Garcia-Alonso L, Handfield L-F, Roberts K, Nikolakopoulou K, Fernando RC, Gardner L, et al. Mapping the temporal and spatial dynamics of the human endometrium *in vivo* and *in vitro*. Nat Genet. 2021;53:1698–711.

31. Wang W, Vilella F, Alama P, Moreno I, Mignardi M, Isakova A, et al. Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle. Nat Med. 2020;26:1644–53.

32. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, et al. Single-cell reconstruction of the early maternal–fetal interface in humans. Nature. 2018;563:347–53.

33. Arutyunyan A, Roberts K, Troulé K, Wong FCK, Sheridan MA, Kats I, et al. Spatial multiomics map of trophoblast development in early pregnancy. Nature. 2023;616:143–51.

34. Greenbaum S, Averbukh I, Soon E, Rizzuto G, Baranski A, Greenwald NF, et al. A spatially resolved timeline of the human maternal–fetal interface. Nature. 2023;619:595–605.

35. Zhao L, Yao C, Xing X, Jing T, Li P, Zhu Z, et al. Single-cell analysis of developing and azoospermia human testicles reveals central role of Sertoli cells. Nat Commun. 2020;11:5683.

36. Di Persio S, Tekath T, Siebert-Kuss LM, Cremers J-F, Wistuba J, Li X, et al. Single-cell RNA-seq unravels alterations of the human spermatogonial stem cell compartment in patients with impaired spermatogenesis. Cell Rep Med. 2021;2:100395.

37. Alfano M, Tascini AS, Pederzoli F, Locatelli I, Nebuloni M, Giannese F, et al. Aging, inflammation and DNA damage in the somatic testicular niche with idiopathic germ cell aplasia.

Nat Commun. 2021;12:5205.

38. Chen Y, Liu X, Zhang L, Zhu F, Yan L, Tang W, et al. Deciphering the molecular characteristics of human idiopathic nonobstructive azoospermia from the perspective of germ cells. Adv Sci. 2023;2206852.

39. Ferrero H, Corachán A, Aguilar A, Quiñonero A, Carbajo-García MC, Alamá P, et al. Single-cell RNA sequencing of oocytes from ovarian endometriosis patients reveals a differential transcriptomic profile associated with lower quality. Hum Reprod. 2019;34:1302–12.

40. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184:3573-3587.e29.

41. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19:15.

42. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods. 2014;11:817–20.

43. Guo F, Li L, Li J, Wu X, Hu B, Zhu P, et al. Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. Cell Res. 2017;27:967–88.

44. Li D, Purushotham D, Harrison JK, Hsu S, Zhuo X, Fan C, et al. WashU epigenome browser update 2022. Nucleic Acids Res. 2022;50:W774–81.

45. Li H, Huo Y, He X, Yao L, Zhang H, Cui Y, et al. A male germ-cell-specific ribosome controls male fertility. Nature. 2022;612:725–31.

46. Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, et al. A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-seq. Dev Cell. 2018;46:651-667.e10.

Figure legends

Figure 1 A brief schema describing data curation, web construction, and featured functionalities.

The single-cell data of over 2.0 million individual cells across six species (human, monkey, mouse, pig, buffalo, and goat) were curated, covering the entire life cycle and spanning > 120 specific time points of 6 species. Multiple featured function modules were provided, such as "Query gene expression", "DIY expression plot", "DNA methylation plot" and "Epigenome browser".

Figure 2 Featured functionalities for exploring single-cell transcriptome datasets.

A. The "Overview" module provides UMAP plots showing cell types and sample information, bar plots showing cell number and proportion of each cell type, and tables showing differentially expressed genes for each single-cell dataset. The single-cell RNA-seq data of human adult testis (GSE112013) was shown as an example. **B.** The "Query gene expression" module enables visualization of the expression levels of each selected gene with UMAP plot (upper) and violin plot (lower). The expression level of a germ-cell specific ribosomal gene *RPL39L* in human adult testis was shown. **C.** The "DIY expression plot" module allows users to create customized plots displaying the expression levels of selected genes in selected cell types. Three types of plots (dot plot, violin plot and heatmap) are provided. The expression levels of representative marker genes and four ribosomal genes in major cell types of human adult testis were shown. The descriptions of cell type abbreviation are listed below. SSC, spermatogonia stem cell. SPG, spermatogonia. L, leptotene spermatocyte. Z, zygotene spermatocyte. P, pachytene spermatocyte. D, diplotene spermatocyte. SPC7, a mixture of diakinesis, metaphase, anaphase, telophase, and secondary spermatocytes.

Figure 3 Featured functionalities for exploring single-cell epigenome datasets.

A. The representative DNA demethylation and *de novo* methylation regions between human 8-cell embryo and morula were shown. Each row in the "Tanghulu" plot corresponds to an individual cell, and each circle represents a CpG site. Methylated CpG sites are represented by black dots, while unmethylated CpG sites are indicated by white dots. The gray dots at the bottom indicate covered CpG sites in selected cells. The CpG sites which were not covered by any cell were omitted. The single-cell DNA

methylation data of human early embryo (GSE100272) was used. **B.** The screenshots of epigenome browser displaying the DNA methylation levels and chromatin accessibility levels around the promoter region of *RPL39L* in primordial germ cells. The gray rectangles highlight regions showing significant differences between somatic niche cells and germline cells. PGC, primordial germ cell.

Figure 4 Cross-species and cross-stage analyses of ribosomal genes using SMARTdb.

A. The dot plots show the expression levels of four ribosomal genes in mouse, cynomolgus monkey, and pig. PMC, peritubular myoid cell. IP, interstitial progenitor cell. **B.** The UMAP plots show the cell type identities of each cell cluster. The expression levels of two ribosomal genes (*RPL39L* and *RPL39*) in human germline and somatic niche cells of fetal stage and peri-pubertal stage were shown in dot plots. FGC, fetal germ cell. PMC, peritubular myoid cell.

Figure 5 The gene expression, DNA methylation and chromatin accessibility patterns of *RPL39L* in human ovary.

A. The leftmost UMAP plot shows the cell type identities of human adult ovary. The expression levels of *RPL39L* in various cell types of human adult ovary are shown on the middle UMAP plot and the rightmost dot plot. **B.** The screenshots of epigenome browser displaying the DNA methylation levels and chromatin accessibility levels around the promoter region of *RPL39L* in human adult ovary. The gray rectangles highlight regions showing changes in fully grown, MI, and MII oocytes compared with growing oocytes. GO, growing oocyte. FGO, fully grown oocyte. MI, metaphase I. MII, metaphase II.

Supplementary material

Figure S1 The Sankey plot shows the composition of datasets included in SMARTdb.

OA, obstructive azoospermia. NOA, Non-obstructive azoospermia.

Figure S2 The DNA methylation levels and chromatin accessibility levels around the *RPL39L* promoter in fetal germ cells after 11 weeks.

560 The screenshots of epigenome browser displaying the DNA methylation levels and
561 chromatin accessibility levels around the *RPL39L* promoter in fetal germ cells after 11
562 weeks post fertilization. The gray rectangles highlight regions showing significant
563 differences between germ cells and somatic niche cells. The mean chromatin
564 accessibility levels of the regions which are highlighted by gray rectangles are labeled.
565
566
567 **Table S1 Summary of curated datasets and covered time points.**
568

Figure 1

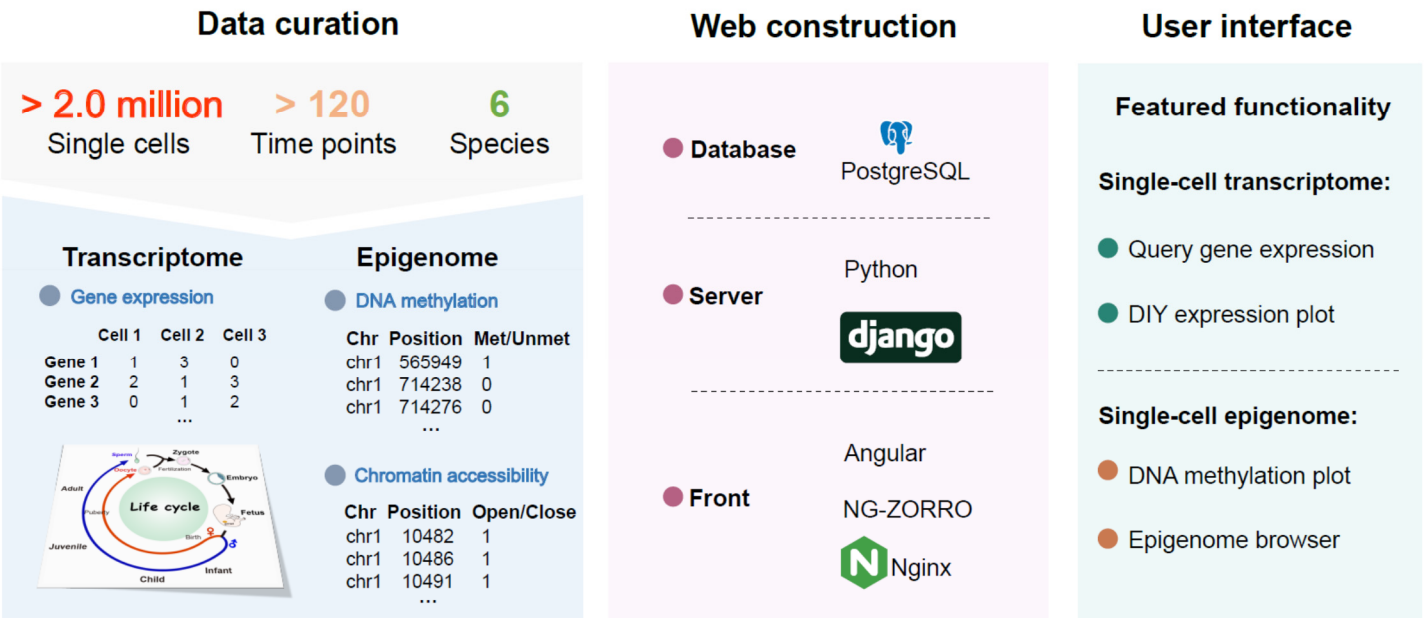

Data curation
> 2.0 million
Single cells
> 120
Time points
6
Species
Transcriptome
Gene expression
Cell 1 Cell 2 Cell 3
Gene 1 1 3 0
Gene 2 2 1 3
Gene 3 0 1 2
...
Life cycle
Zygote
Embryo
Fetus
Adult
Juvenile
Child
Infant
Epigenome
DNA methylation
Chr Position Met/Unmet
chr1 565949 1
chr1 714238 0
chr1 714276 0
...
Chromatin accessibility
Chr Position Open/Close
chr1 10482 1
chr1 10486 1
chr1 10491 1
...
Web construction
Database
PostgreSQL
Server
Python
django
Front
Angular
NG-ZORRO
Nginx
User interface
Featured functionality
Single-cell transcriptome:
Query gene expression
DIY expression plot
Single-cell epigenome:
DNA methylation plot
Epigenome browser

Figure 2

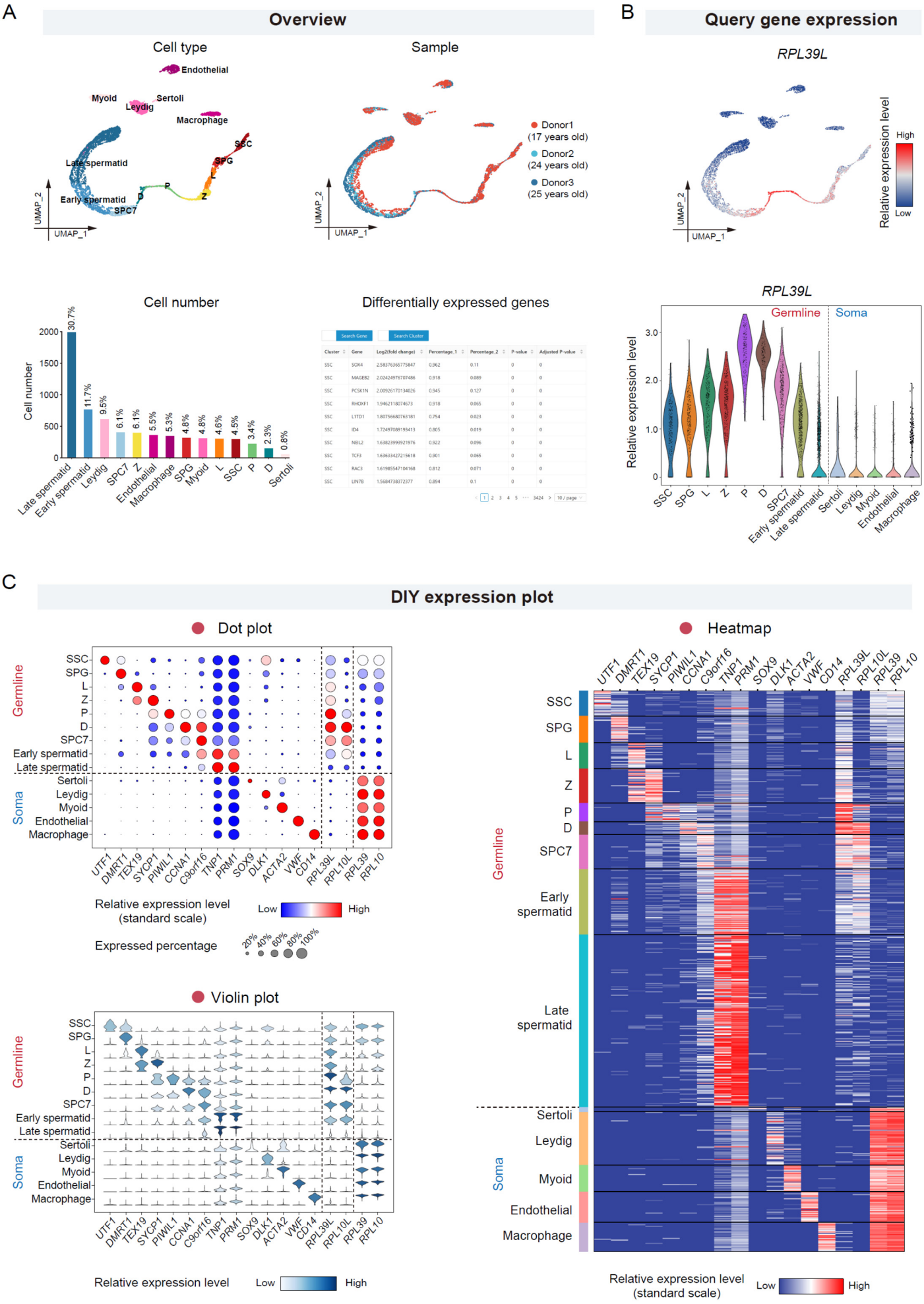
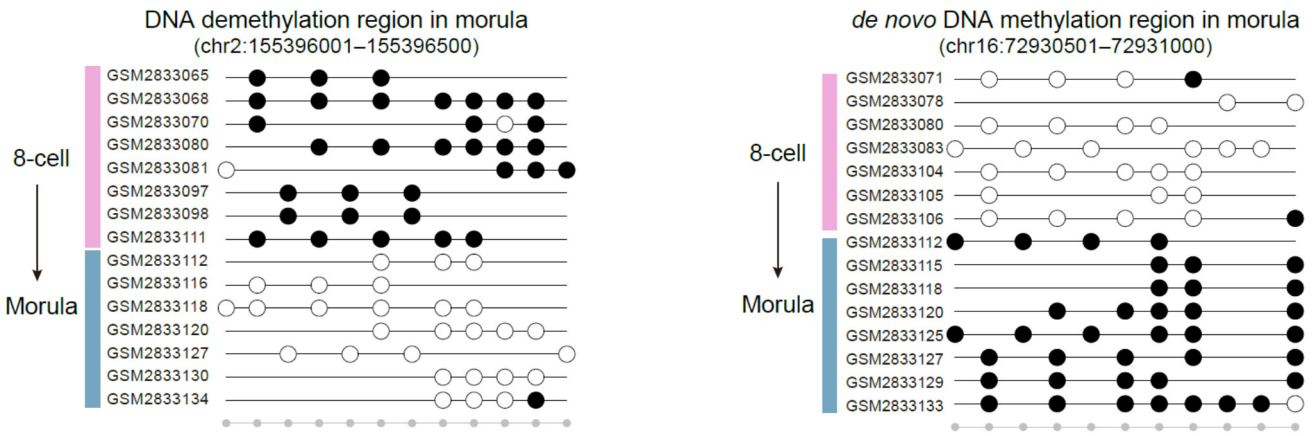

A
Overview
Cell type
Endothelial
Myoid
Sertoli
Leydig
Macrophage
Late spermatid
Early spermatid
SPC7
SSC
SPG
L
Z
P
D
UMAP_2
UMAP_1
Sample
Donor1 (17 years old)
Donor2 (24 years old)
Donor3 (25 years old)
Cell number
Cell number
30.7%
11.7%
9.5%
6.1%
6.1%
5.5%
5.3%
4.8%
4.8%
4.6%
4.5%
3.4%
2.3%
0.8%
Late spermatid
Early spermatid
Leydig
SPC7
Z
Endothelial
Macrophage
SPG
Myoid
L
SSC
P
D
Sertoli
Differentially expressed genes
Search Gene
Search Cluster
Cluster Gene Log2(fold change) Percentage_1 Percentage_2 P-value Adjusted P-value
SSC SOX4 2.58376365775847 0.962 0.11 0 0
SSC MAGEB2 2.02424976707486 0.918 0.089 0 0
SSC PCSK1N 2.00926170134026 0.945 0.127 0 0
SSC RHOXF1 1.94621180746673 0.918 0.065 0 0
SSC L1TD1 1.80755680763181 0.754 0.023 0 0
SSC ID4 1.72497089193413 0.805 0.019 0 0
SSC NBL2 1.63823999321976 0.922 0.096 0 0
SSC TCF3 1.63633427215618 0.901 0.065 0 0
SSC RAC3 1.61985547104168 0.812 0.071 0 0
SSC LIN7B 1.56847383372377 0.894 0.1 0 0
1 2 3 4 5 ··· 3424 > 10 / page
B
Query gene expression
RPL39L
Relative expression level
High
Low
RPL39L
Germline
Soma
Relative expression level
SSC
SPG
L
Z
P
D
SPC7
Early spermatid
Late spermatid
Sertoli
Leydig
Myoid
Endothelial
Macrophage
C
DIY expression plot
Dot plot
Germline
Soma
UTF1 DMRT1 TEX19 SYCP1 PIWIL1 CCNA1 C9orf16 TNP1 PRM1 SOX9 DLK1 ACTA2 VWF CD14 RPL39L RPL10L RPL39 RPL10
Relative expression level (standard scale)
Low
High
Expressed percentage
20% 40% 60% 80% 100%
Violin plot
Relative expression level
Low
High
Heatmap
Relative expression level (standard scale)
Low
High

Figure 3

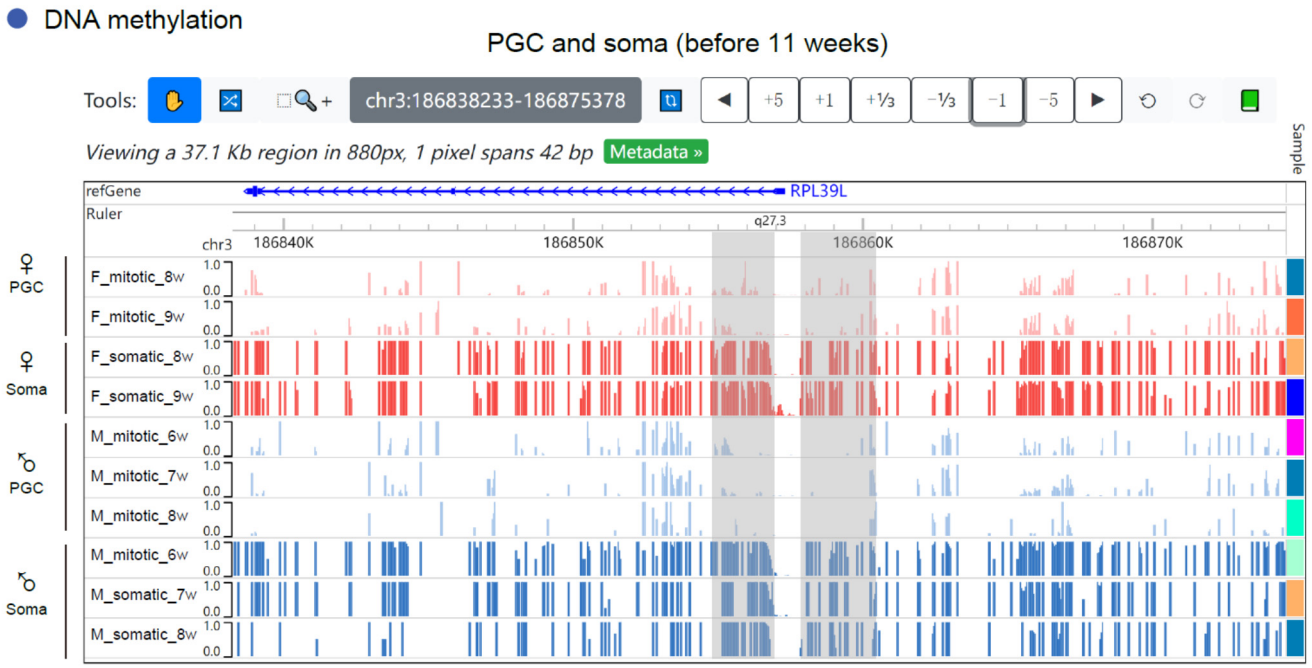
A

DNA methylation plot



B

Epigenome browser



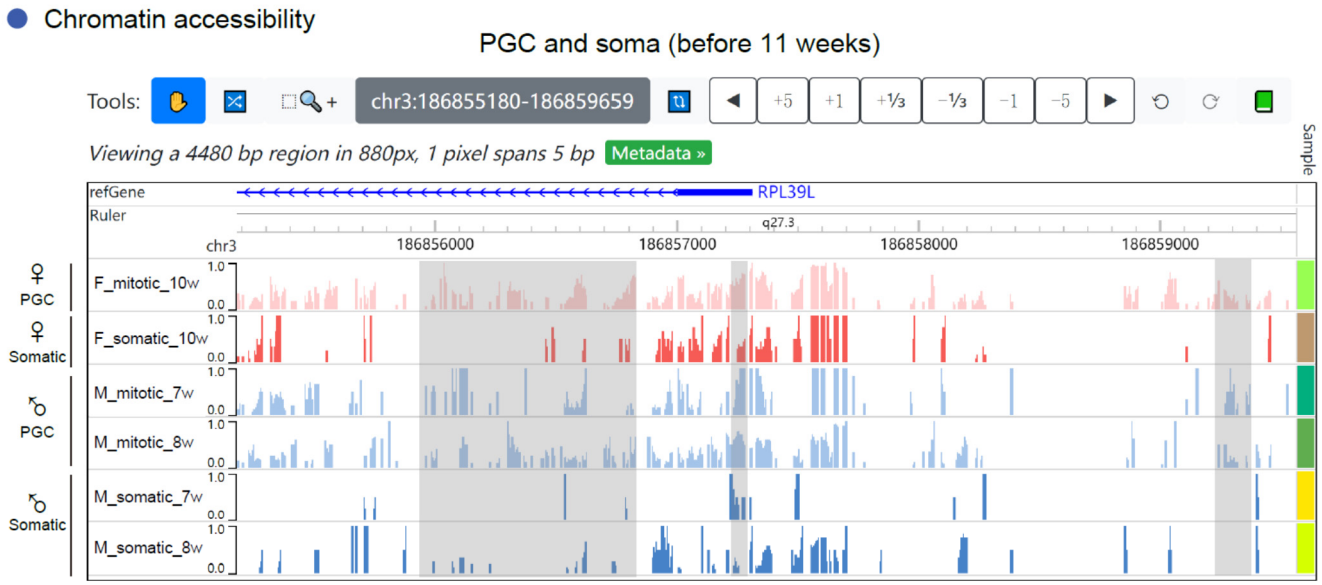
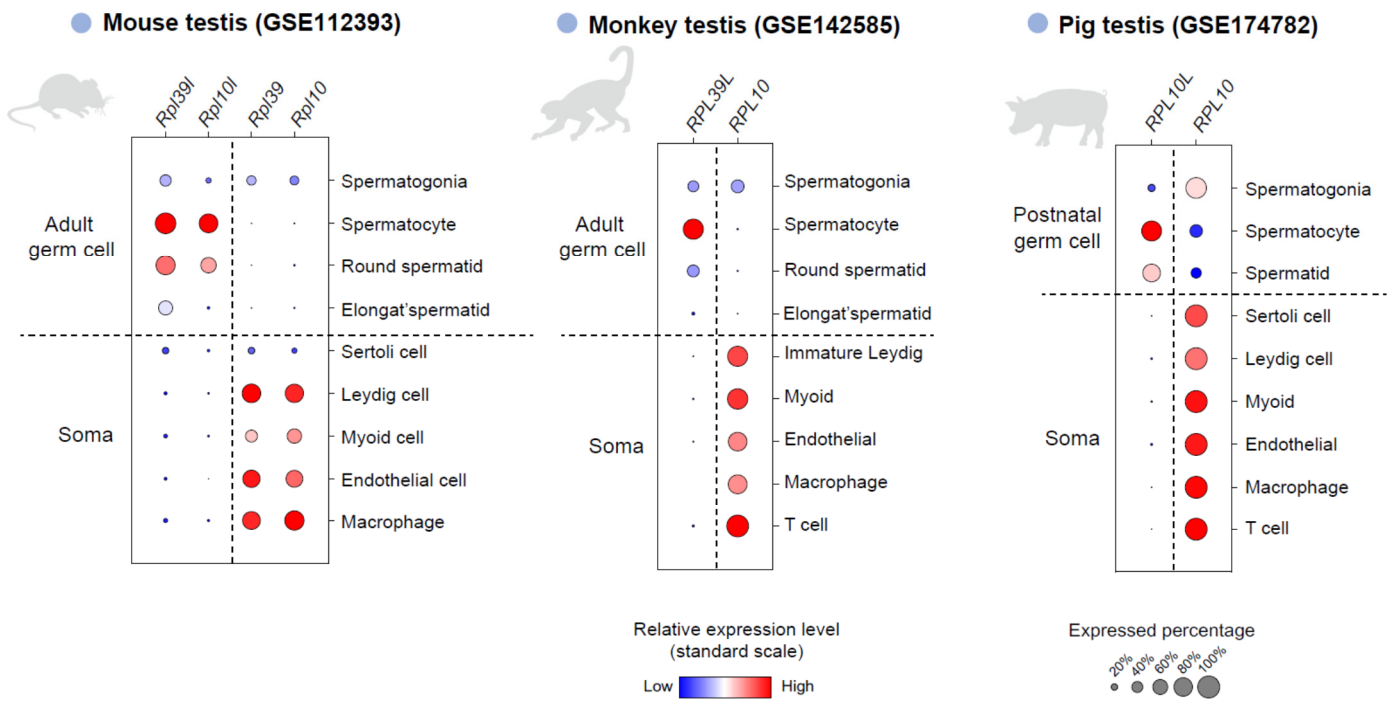


Figure 4

A

Cross-species conservation



B

Cross-stage analysis

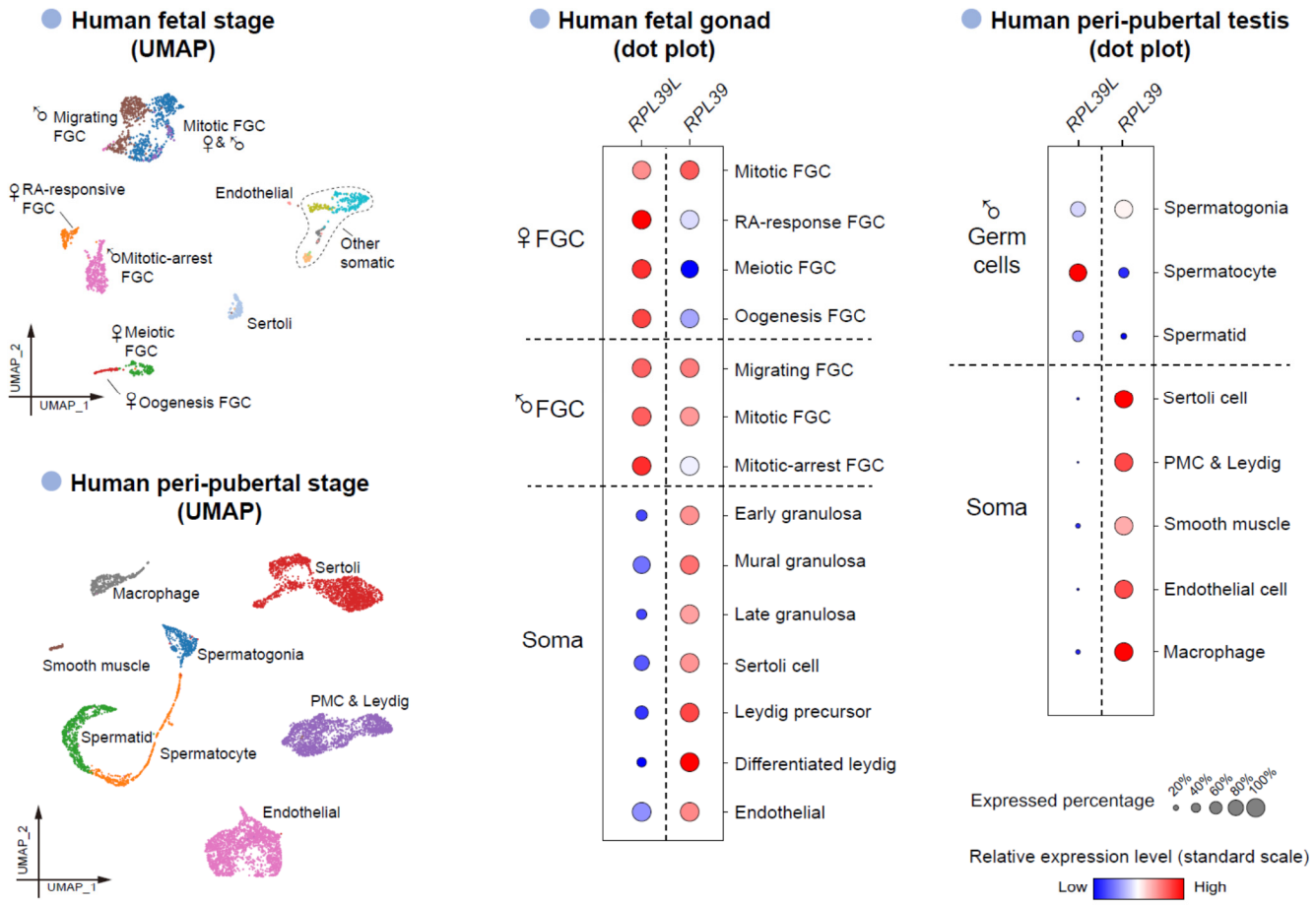
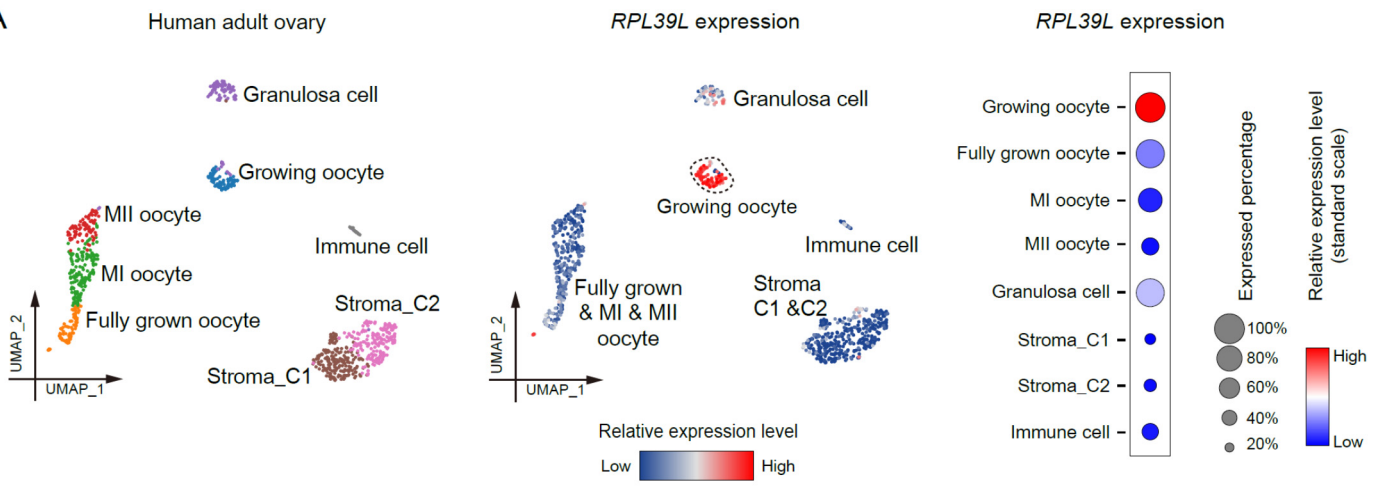


Figure 5

A


Human adult ovary
Granulosa cell
Growing oocyte
MII oocyte
MI oocyte
Fully grown oocyte
Immune cell
Stroma_C2
Stroma_C1
UMAP_2
UMAP_1
RPL39L expression
Granulosa cell
Growing oocyte
Immune cell
Fully grown & MI & MII oocyte
Stroma C1 &C2
Relative expression level
Low
High
RPL39L expression
Growing oocyte
Fully grown oocyte
MI oocyte
MII oocyte
Granulosa cell
Stroma_C1
Stroma_C2
Immune cell
Expressed percentage
100%
80%
60%
40%
20%
Relative expression level (standard scale)
High
Low

B

DNA methylation
Tools:
chr3:186838233-186875378
+5
+1
+⅓
−⅓
−1
−5
Viewing a 37.1 Kb region in 880px, 1 pixel spans 42 bp
Metadata »
Sample
refGene
RPL39L
Ruler
q27.3
chr3
186840K
186850K
186860K
186870K
Oocyte
GO
FGO
MI
MII
Soma
Granulosa
Stroma_C1
Stroma_C2
Immune
1.0
0.0

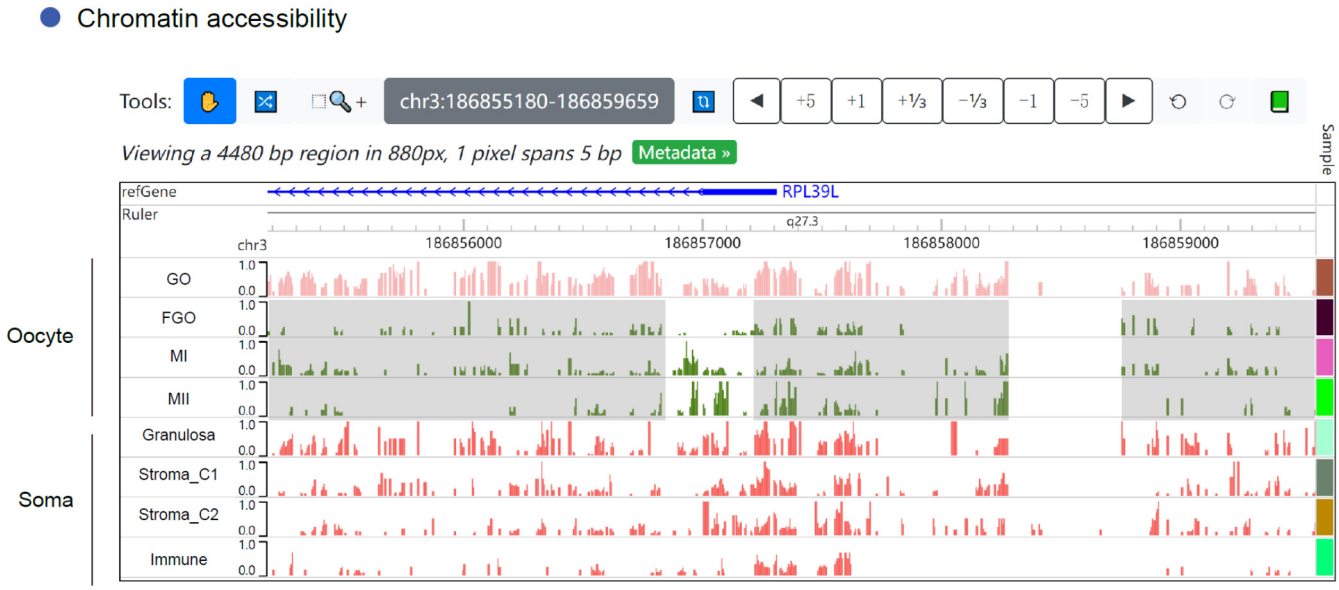

Chromatin accessibility
Tools:
chr3:186855180-186859659
+5
+1
+⅓
−⅓
−1
−5
Viewing a 4480 bp region in 880px, 1 pixel spans 5 bp
Metadata »
Sample
refGene
RPL39L
Ruler
q27.3
chr3
186856000
186857000
186858000
186859000
Oocyte
GO
FGO
MI
MII
Soma
Granulosa
Stroma_C1
Stroma_C2
Immune
1.0
0.0

Figure S1

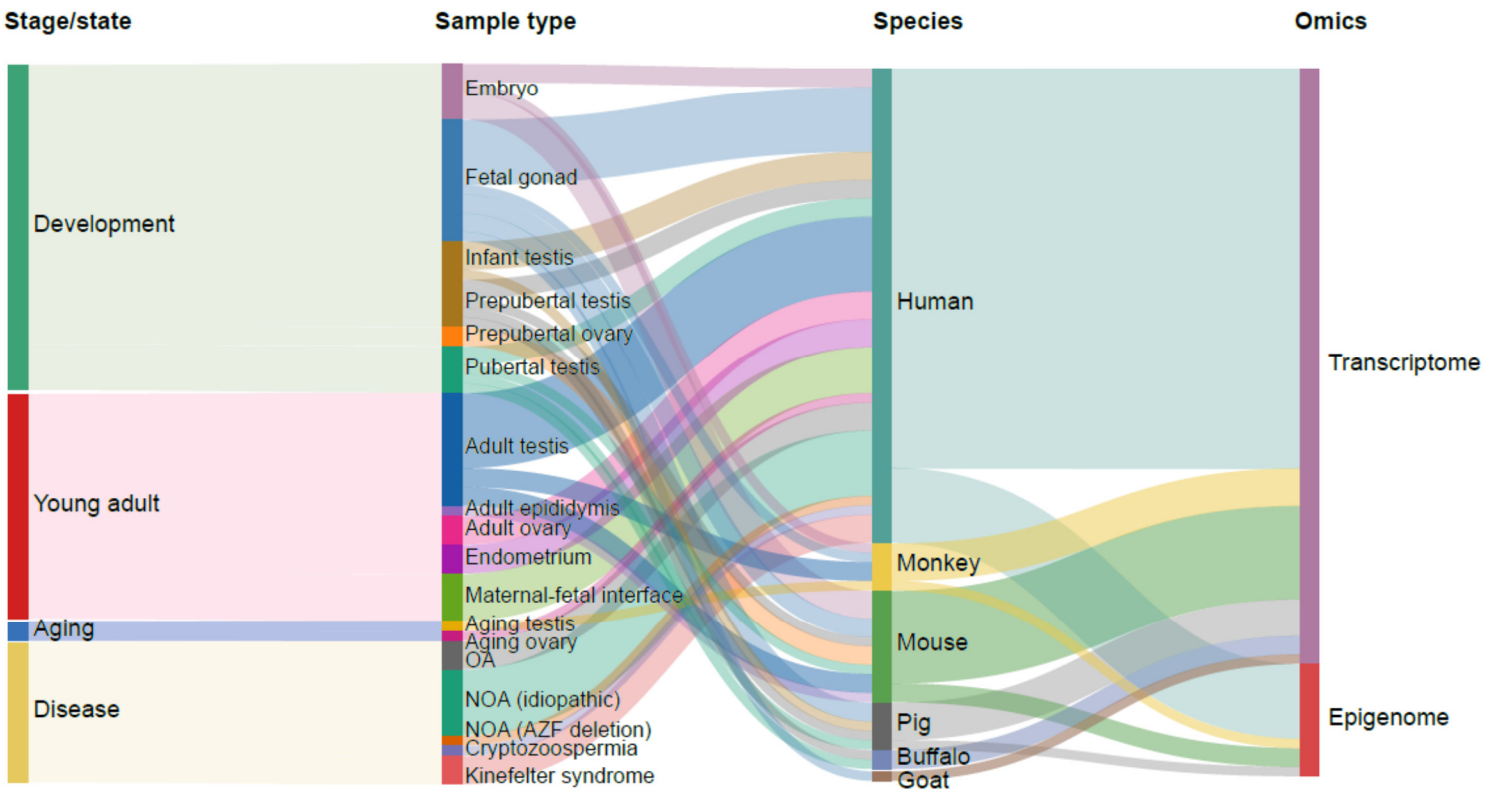

Stage/state
Sample type
Species
Omics
Development
Young adult
Aging
Disease
Embryo
Fetal gonad
Infant testis
Prepubertal testis
Prepubertal ovary
Pubertal testis
Adult testis
Adult epididymis
Adult ovary
Endometrium
Maternal-fetal interface
Aging testis
Aging ovary
OA
NOA (idiopathic)
NOA (AZF deletion)
Cryptozoospermia
Kinefelter syndrome
Human
Monkey
Mouse
Pig
Buffalo
Goat
Transcriptome
Epigenome

Figure S2

- DNA methylation

- Chromatin accessibility