A spatial single-cell type multiplex map of human spermatogenesis

Feria Hikmet,[1,2]† Loren Méar,[1,3]† Jonas Gustavsson,[1] Gisele Miranda,[4] Cheng Zhang,[5] Borbala Katona,[1] Rutger Schutten,[1] Per Adelsköld[6], Kalle Von Feilitzen,[6] Mattias Forsberg,[6] Jan-Bernd Stukenborg,[2,7] Mathias Uhlén,[6,8] Cecilia Lindskog[1]*

[1]Department of Immunology, Genetics and Pathology, Cancer Precision Medicine Research Unit, Uppsala University, SE-75185 Uppsala, Sweden

[2]Department of Organismal Biology, Nordfertil Research Lab Uppsala, Physiology and Environmental Toxicology, Uppsala University, SE-75185 Uppsala, Sweden

[3]Department of Women's and Children's Health, Division for Neonatology, Obstetrics, Gynaecology and Reproductive Health, Karolinska Institute and Karolinska University Hospital, SE-14186 Stockholm, Sweden, Sweden

[4]Science for Life Laboratory, Division of Computational Science and Technology, KTH Royal Institute of Technology

[5]Roger Williams Institute of Liver Studies, School of Immunology & Microbial Sciences, Faculty of Life Sciences and Medicine, King's College London, Foundation for Liver Research, and King's College Hospital, London, United Kingdom

[6]Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, 17121 Stock-holm, Sweden

[7]Department of Women's and Children's Health, Nordfertil Research Lab Stockholm, Childhood Cancer Research Unit, Karolinska Institutet, and Karolinska University, Stockholm, Sweden

[8]Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden

†These authors contributed equally.

*Correspondence to: Dr. Cecilia Lindskog, Department of Immunology, Genetics and Pathology, Cancer Precision Medicine Research unit, Uppsala University, Rudbeck Laboratory, Dag Hammarskjölds väg 20, SE-75185 Uppsala, Sweden. Telephone: +46 18 471 50 32, email: cecilia.lindskog@igp.uu.se

Teaser

An antibody-based cellular map of the human testis provides high-resolution expression data in sub-states of cells throughout spermatogenesis.

Abstract

Mapping the spatial localization of proteins is essential for understanding their function. Human testis constitutes a complex organ with thousands of proteins activated or repressed during spermatogenesis. By combining single-cell RNA sequencing (scRNA-seq) with fluorescence-based multiplex immunohistochemistry (mIHC), we performed large-scale spatiotemporal mapping of specific testicular cell states within their native tissue context. Single-cell state expression patterns of almost 500 proteins were quantified and clustered according to differentiation stages, providing a framework for functional interpretation. Interestingly, comparative analysis of mRNA-to-protein expression dynamics revealed multiple instances of temporal discordance within defined cellular states. Notably, contradicting previous findings, PIWIL4 transcripts were detected in earlier germ cell states, whereas protein abundance peaked at later stages, coinciding with proliferative transition. These findings highlight the importance of assessing protein expression in single-cell mapping approaches. This workflow establishes a scalable framework for proteome-wide spatial tissue analysis applicable to studying cellular dynamics in both health and disease.

Introduction

The dramatic improvements in single-cell RNA sequencing (scRNA-seq) have resulted in many worldwide single-cell mapping efforts, including the Human Cell Atlas (HCA) (*1*), the Chan Zuckerberg Initiative (CZI) (*2*), the Human BioMolecular Atlas Program (HuB-MAP) (*3*) (funded by the National Institute of Health (NIH)), and the Human Cell Landscape (*4*). However, few single-cell mapping efforts have so far been reported for reproductive organs. With nearly one-third of the world's population being affected by reproductive disorders and the increasing incidence of infertility, impacting one in six males, along with the close relationship between reproductive health and overall health, there is an urgent need to map the reproductive processes at a molecular level. Examining the molecular constituents at a single-cell resolution is necessary to facilitate a deeper understanding of these processes.

An interesting strategy for such functional studies is to integrate scRNA-seq analysis with antibody-based methods to probe the presence of the corresponding proteins in relation to tissue geography. While transcriptomic analysis is superior to proteomic methods in terms of gene coverage of low-abundant genes, proteomic methods exceed in dynamic range (*5*). As transcriptomic methods also suffer from drop-out (*6*), proteomic methods can be used to better describe the heterogeneous temporal expression patterns and the extent to which proteins are activated or repressed (*7*, *8*).

The Human Protein Atlas (HPA) project provides the possibility to use the open-access database (www.proteinatlas.org) for the integration of spatial antibody-based proteomics and transcriptomics technologies. The large-scale database covers the major human tissues and cell types using bulk RNA, scRNA-seq (*9*), and antibody-based data (*10*), with the classification of genes according to tissue and cell type specificity from a body-wide perspective. The HPA has utilized immunohistochemistry (IHC) to visualize the expression of human proteins in all male and female reproductive tissues. Antibody-based proteomics provides a pixel resolution not currently attainable by other spatial methods; however, classical IHC is primarily used to study one protein at a time, and the resulting data is only semi-quantitative (*11*). In addition, the visualization of different cell types is limited to what the human observer can distinguish, and the transition between cell states or stages of differentiation is not always possible to determine based on histology only (*12*). Finally, manual inspection of images is highly subjective, which may lead to wrong interpretation or bias, further highlighting the need for more advanced imaging methods (*13*). Here, we address these shortcomings by combining scRNA-seq with single-cell fluorescent-based multiplex immunohistochemistry (mIHC) using antibodies generated in the HPA project.

Recently, an increasing number of methods for multiplex analysis of proteins have been introduced, enabling novel insights into co-expression and tissue heterogeneity. Some examples of highly multiplexed methods include the CODEX (CO-Detection by indEXing), Immuno-SABER, InSituPlex, and TOF-based methods such as imaging mass cytometry and Mass Ion Beam Imaging. While several of these methods allow for high-plex analyses with dozens of markers, they are mainly used together with ready-made antibody panels, in particular in the field of immuno-oncology. Setting up custom-made panels requires tedious validation and conjugation of antibodies, e.g., with DNA barcodes. Thus, these methods are not optimal for mapping sub-sets of cells and single-cell states in any kind of tissue and are not applicable for proteome-wide efforts (*11*).

Here, we show a novel strategy for high-throughput mIHC for proteome-wide analysis of human tissues. Building upon existing pipelines for regular IHC that are available in most histological laboratories, we used a combination of scRNA-seq data analysis, fluorescent-based antibody-based imaging, and automated image analysis. We have focused on the testis and the development of sperm cells through spermatogenesis, during which the undifferentiated germ cells undergo several germ-specific events such as meiosis to reduce the number of chromosomes, condensation of the nucleus, and removal of excess cytoplasm within the sperm.

The sperm cells acquire unique morphological features and cell structures, such as the flagellum and the acrosome, which grants them the ability to motility necessary for successful fertilization (*14*). The complex nature of spermatogenesis involves the expression of thousands of genes and proteins (*15*), and the testis has previously been shown to be the human tissue with the highest number of tissue-specific genes (*10*, *13*, *16*), out of which many have an unknown function (*17*). We selected a set of 499 proteins that, based on regular IHC, showed locations specific to germ cells within the testis. With the help of scRNA-seq data, we built antibody panels outlining 12 discrete testicular cell states or stages of differentiation, a resolution not attainable by regular IHC. The fixed panels were used to study one candidate protein at a time and to

quantify the expression based on pixel overlap using an automated image analysis pipeline. The integrated analysis and detailed spatial proteomics data allowed us to determine the single-cell state-specific localization of a large set of previously uncharacterized proteins, thereby being able to assign presumed functions in the human testis. We were also able to study mRNA and protein expression dynamics and identified many mRNA-to-protein pairs that varied from a spatiotemporal aspect, further emphasizing the importance of including protein analyses in single-cell mapping efforts.

In summary, we here introduced a validated and robust multiplex pipeline for high-throughput analyses of single-cell states in human tissues, constituting a foundation for hypothesis-driven protein research in health and disease. The data is presented as an open-access spatiotemporal single-cell type reference map of the adult human testis that decodes the complexity of human germ cells, tissue heterogeneity, and homeostasis and links quantitative data with tissue morphology, thereby potentially contributing to research on infertility and other disorders of the male reproductive system.

Results

A methodological approach for comprehensive tissue studies at the single-cell level

An overview of the workflow is presented in Figure 1. The testis constitutes a challenging tissue type to study due to its dynamic and complex cellular composition during spermatogenesis when spermatogonial stem cells differentiate into mature sperm, supported by the testicular niche cells (Figure 1A). A high-resolution image-based atlas of the testis proteome was created by integrating IHC and scRNA-seq datasets to decipher the tissue-based spatial protein expression profiles of male germ cells (Figure 1B). With this multimodal data, we set out to build multiplex IHC (mIHC) antibody panels. The mIHC panels were used for determining the spatial profiles of germ-cell-associated proteins in specific testicular cell states (Figure 1C). Next, we applied a custom image analysis pipeline to achieve a single-cell quantitative protein expression read-out (Figure 1D) and performed downstream analyses for deep molecular interrogation of histological features (Figure 1E).

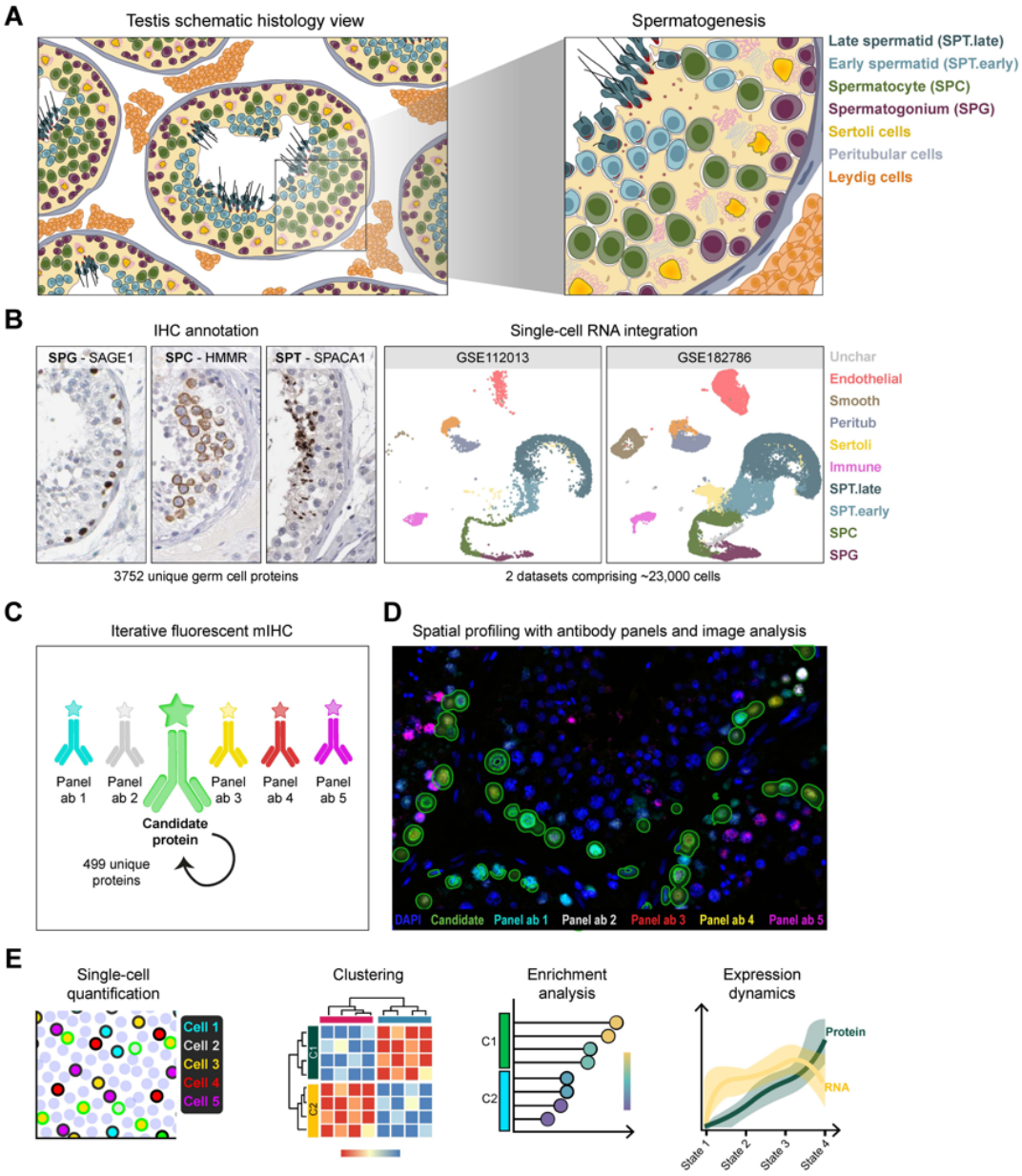


Fig 1. Overview workflow of the study. (**A**) The complex testis microstructure includes a multitude of germ cells within the seminiferous tubules, illustrated in the cropped image. (**B**) IHC-based annotation of human testis and single-cell RNA-seq (scRNA-seq) datasets were combined to build a multiplex IHC (mIHC) antibody panel for spatial characterization of germ-cell proteins. (**C**) Using established antibody panels, candidate proteins were stained

iteratively with constant panel markers for cell phenotyping, with the candidate protein highlighted in green (view Methods for details). (**D**) Cells were classified using a custom image analysis pipeline, allowing for quantitative expression assessment of candidate proteins at the cell-specific level (green overlays). (**E**) High-resolution single-cell level data enables downstream analyses including protein quantification in specific cell populations and exploring functional associations among cell types.

The tissue-based testis proteome as defined by antibody profiling and single-cell transcriptomics

To generate a high-resolution image-based atlas of the testis proteome, we utilized an IHC-based dataset comprising 4,431 proteins that have been analyzed with single-targeting validated antibodies. From here, 3,752 unique proteins showed at least weak IHC staining in at least one of the germ cell types. For 1,124 proteins, more than one antibody toward the same protein was analyzed, and the antibody that had the highest signal-to-noise ratio was selected to represent the results (Figure 2A). This set of 3,752 proteins expressed in testicular germ cells constitutes the basis for the selection of both panel markers and candidate proteins to be investigated more in-depth in the current study. The IHC dataset included numerous proteins expressed in one of the germ cell types only, i.e., proteins specific to spermatogonia, spermatocytes, early spermatids, or late spermatids, but also proteins that appear to be expressed in more than one cell type. Considering the staining intensity (Figure 2B), distinct clusters of proteins show higher specificity to particular cell types. It can also be noted that some proteins seem to have a gradient in expression, with higher staining intensity in some cell types and weaker in others, mainly observed in the "Mixed" and "All germ cells" groups (Figure 2B). Overall, there was a high global correlation between spermatogonia, spermatocytes, and early spermatids, with an even higher association between spermatocytes and early spermatids (Figure 1C).

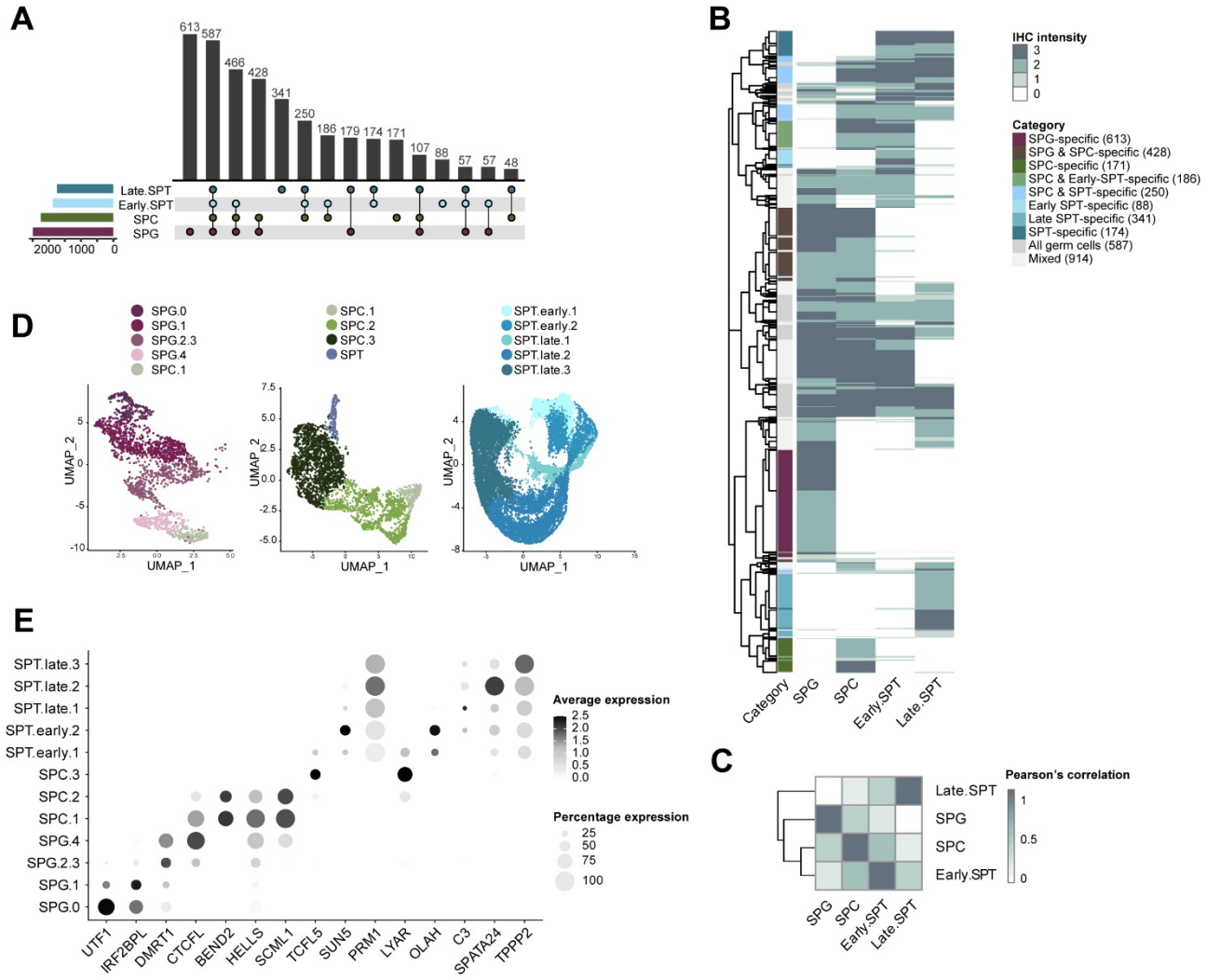


Fig 2. The tissue-based testis proteome as defined by antibody profiling and single-cell transcriptomics. (**A**) All IHC-annotation intensity values were binarized (negative = 0 and weak, moderate, strong =1) and displayed as an Upset plot to view the presence of intensity in any germ cell type count frequency of intensity in horizontal bars and the count frequency of specific combinations of intensity in one cell type regardless of intensity in other in vertical bars. The total count is displayed on top of each bar. (**B**) A heatmap with hierarchical clustering showing the intensity value (negative, weak, moderate, and strong) in germ cells (columns) is visualized for each of the 3,752 proteins (rows) with cell-type specificity group annotation for each protein. The number of proteins in each cell-type specificity group is seen within parenthesis in the heatmap labels. (**C**) Spearman correlation of the data presented in the heatmap in B, with a color ramp to show the correlation coefficient. D) UMAP projections of scRNA-seq sub-clustering of spermatogonia (left), spermatocytes (middle), and spermatids (right). E) Dot-plot of the selected genes used for each panel (x-axis) and their mean expression level and distribution in the defined sub-clusters (y-axis). The order of the genes is displayed as they appear in the below schematic figure, UTF1, IRF2BPL, DMRT1, CTCFL, and BEND2, are used in the spermatogonia panel, labeled SPG.0 to SPC.1. HELLS, SCML1, TCFL5, SUN5, and PRM1 correspond to the spermatocyte panel (SPC.1 to SPT.late). Finally, LYAR, OLAH, C3, SPATA24, and TPPP2 define states in the spermatid panel (SPT.early.1 to SPT.late.3).

Next, we identified the different cell states and stages of differentiation throughout spermatogenesis that we aimed to analyze further in the present investigation. We integrated scRNA-seq from two datasets (Supplementary Figure 1A) (*18, 19*). In total, 26 clusters were identified and annotated with relevant hallmark gene expression profiles, with each cluster represented from both datasets (Supplementary Figure 1B-D). A total of 37,022 cells were retained for downstream analysis, and clusters were grouped on the cell type basis (Supplementary Figure 1E-F), with clusters of spermatogonia (SPG), spermatocytes (SPC), and early and late spermatids (SPT), as well as somatic cell niches (Sertoli cells, Leydig cells and supporting cells). Pseudo-time trajectory analysis only on the extracted germ cell populations showed clear transitions between early and late stages of the spermatogenesis, visualizing the stepwise differentiation and developmental progression (Supplementary Figure 1G).

Clusters corresponding to the three primary germ cell groups (SPG, SPC, and SPT) were extracted, and sub-clustering was performed, which resulted in 12 SPG clusters, 14 SPC clusters, and 16 SPT clusters (Supplementary Figure 2). In almost all sub-clusters, both datasets were well-represented, but some sub-clusters were mainly retained from one of the datasets (data not shown). Based on genes representing specific cell types and stages during spermatogenesis, we assigned each sub-cluster a specific primary state or differentiation stage (Supplementary Figure 2). For SPGs, sub-clustering captured spermatogonial cell states defined as SPG.0, SPG.1, SPG.2-3, and SPG.4, as well as a cluster expressing spermatocyte markers, designated SPC.1 (Supplementary Figure 2). For SPCs, three main cell states associated with the meiotic progression were identified (SPC.1, SPC.2, and SPC.3). SPC.1 corresponded to the same state observed in the SPG sub-clustering. In addition, a smaller sub-cluster expressed spermatid-associated genes and was distinct from SPC.3, therefore designated SPT (Supplementary Figure 2). For SPTs, two early-state spermatids sub-clusters (SPT.early.1 and SPT.early.2) were identified, along with three more differentiated states in which sperm cells begin acquiring their characteristic morphology. These later states (SPT.late.1, SPT.late.2, and SPT.late.3) generally expressed proteins associated with flagellar structure and motility (Supplementary Figure 2). The final UMAP projections for each sub-clustering are shown in Figure 2D and Supplementary Figure 2. Cell counts for unannotated and annotated clusters can be viewed in Supplementary Tables 1 and 2.

To validate the identified cell states, we selected candidate marker genes for each SPG, SPC, and SPT sub-cluster based on the scRNA-seq data. Using differential expression analysis applying a fold-change filter of 1.5, we retrieved lists of genes enriched in each sub-cluster. Out of the 3,752 proteins with antibodies available in the HPA, 969 DEGs were identified for SPG states (SPG.0 = 382, SPG.1 = 125, SPG.2.3 = 69, SPG.4 = 383), 2,093 DEGs for SPC states (SPC.1 = 883, SPC.2 = 810, SPC.3 = 400), and 1,332 DEGs for SPT states (SPT.early.1 = 367, SPT.early.2 = 513, SPT.late.1 = 116, SPT.late.2 = 133, and SPT.late.3 = 203). From the DEGs, specific markers were subsequently selected to develop three multiplex IHC (mIHC) antibody panels for spatial mapping of SPG, SPC, and SPT cell states (dot plots for the corresponding genes are shown in Figure 2E). Established markers known to be specific for a particular cell state were prioritized in the panels (e.g., UTF1 for SPG.0 and DMRT1 for SPG.2-3). For cell states lacking well-established markers, candidate markers were selected based on manual inspection of IHC

data, considering signal-to-noise ratio and staining pattern (e.g., avoiding subcellular patterns that could complicate downstream automated image analysis).

High-throughput mIHC for determining cell state-specific protein expression

Antibodies targeting proteins representing each specific cell state were optimized, and the final antibody panels were selected based on robust and specific staining, clearly marking out the expected germ cell populations with minimal overlap between markers within the same panel. This enabled confident discrimination of cell states corresponding to the sub-clusters identified by scRNA-seq. The final mIHC antibody panels for SPG, SPC, and SPT are shown in Figure 3A, and individual IHC stainings for each panel marker are presented in Supplementary Figure 3. For the SPG and SPC panels, exclusion markers representing cell types in other panels were added, to confidently determine when the expression starts and stops. For this reason, we added a preleptotene SPC marker in the SPG panel, and markers for early and late SPTs in the SPC panel.

The SPG, SPC, and SPT antibody panels, each containing five different markers, were next used to stain together with novel proteins, here named “candidate proteins”. By leaving a specific spot in the mIHC workflow, one candidate protein at a time was stained together with either the SPC, SPC, or SPT panel. This approach enabled us to determine the main cell state-specific location of each candidate protein based on expression overlap with the panel markers. Representative stainings of a selection of candidate proteins stained with the three different panels are presented in Figure 3B. In total, 499 candidate proteins were selected for mIHC analysis (Figure 3C, Supplementary Table 3), based on the main expression clusters previously identified by IHC (Figure 1B-C). The IHC data was used as a guidance for selecting the appropriate mIHC panel, ensuring that proteins stained with SPG, SPC, and SPT panels were expected to be predominantly localized to spermatogonia, spermatocytes or spermatids, respectively (Supplementary Figure 4). Candidate proteins (n=69) suspected to be expressed at a transitional state shared between panels (e.g., SPC.1, present in both the SPG and SPC panels), were stained with both panels to determine the onset and loss of expression across cell states (Figure 3C). In total, 228 candidate proteins were stained with the SPG panel, 264 with the SPC panel, and 78 with the SPT panel.

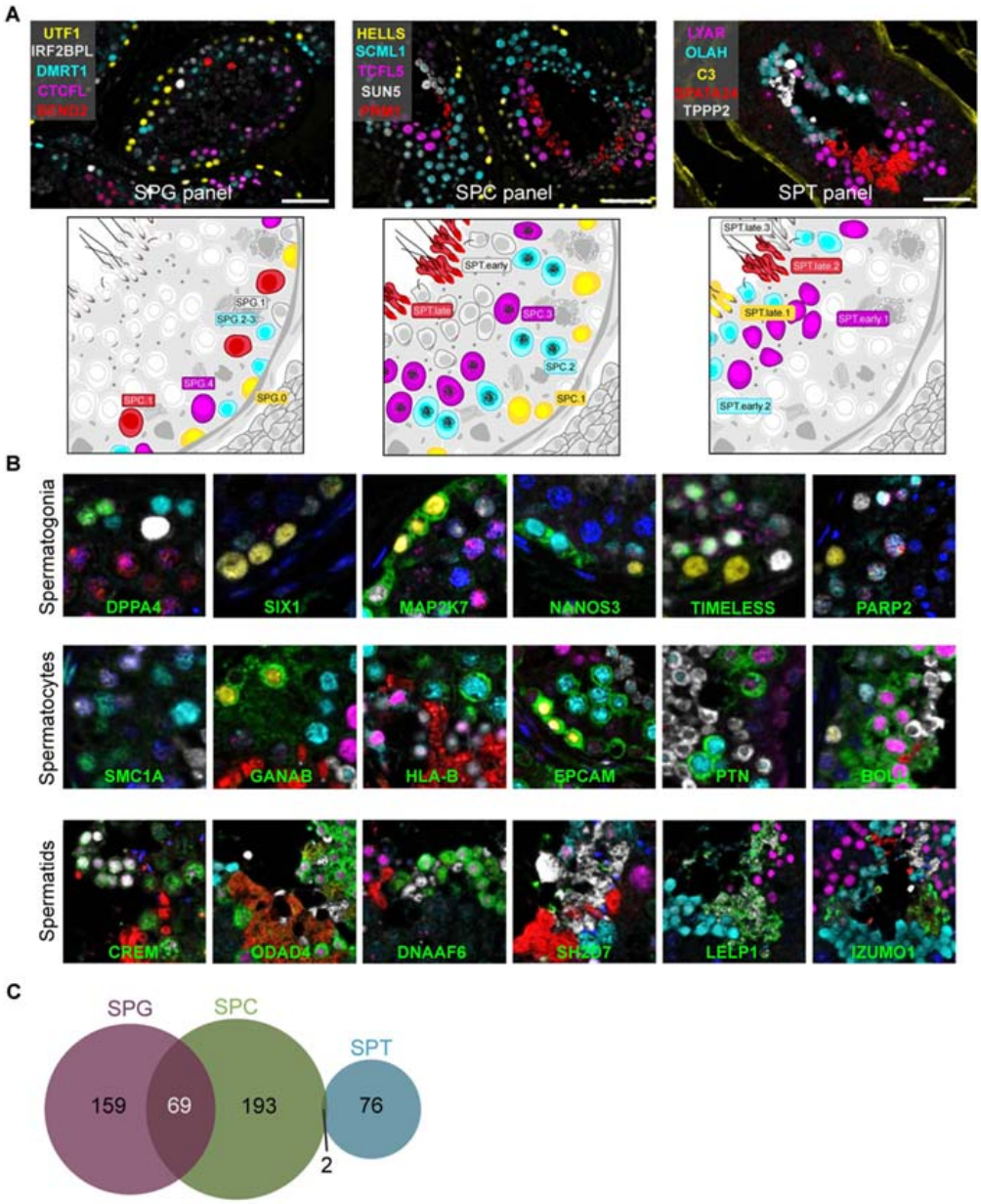


Fig. 3. Deep spermatogenic profiling using high-throughput multiplex immunohistochemistry. (**A**) Representative images of each mIHC panel in normal human seminiferous tubules, with schematic illustrations highlighting the relative abundance and position of each cell state targeted by the panels. (**B**) Examples of proteins stained with three different antibody panels. In each image, the candidate protein is visualized in green, and the protein name is depicted at the bottom of the image. (**C**) A Venn diagram showing the number of unique protein candidate proteins stained with each panel.

The mIHC workflow resulted in high-resolution fluorescent images corresponding to six channels with antibody stainings for each of the 499 different candidate proteins. All images were processed by employing a custom-built image analysis pipeline. Due to the flexible setup of the analysis pipeline, we could leverage the same image analysis strategy for stainings generated by all three antibody panels (Figure 4A). Images were extracted, and cell (objects) segmentation was first performed to identify all individual cells in each stained slide. Cells were then phenotyped based on their panel marker staining in each cell. The total count of each cell state for the three different antibody panels is visualized as percentages in Figure 4B. Expression of the candidate proteins in each phenotyped cell was quantified using a pixel-based model available in the Ilastik software. This machine learning-based tool was trained by manually annotating positive signal versus background in approximately 10% of the generated images, thereby setting the threshold for detection. The trained classifier was then applied to generate image masks for all candidate protein images, enabling removal of non-specific signal and uncertain

staining intensities (Supplementary Figure 5). The resulting cell counts and the mean expression intensity per protein and panel can be viewed in Supplementary Tables 8 and 9.

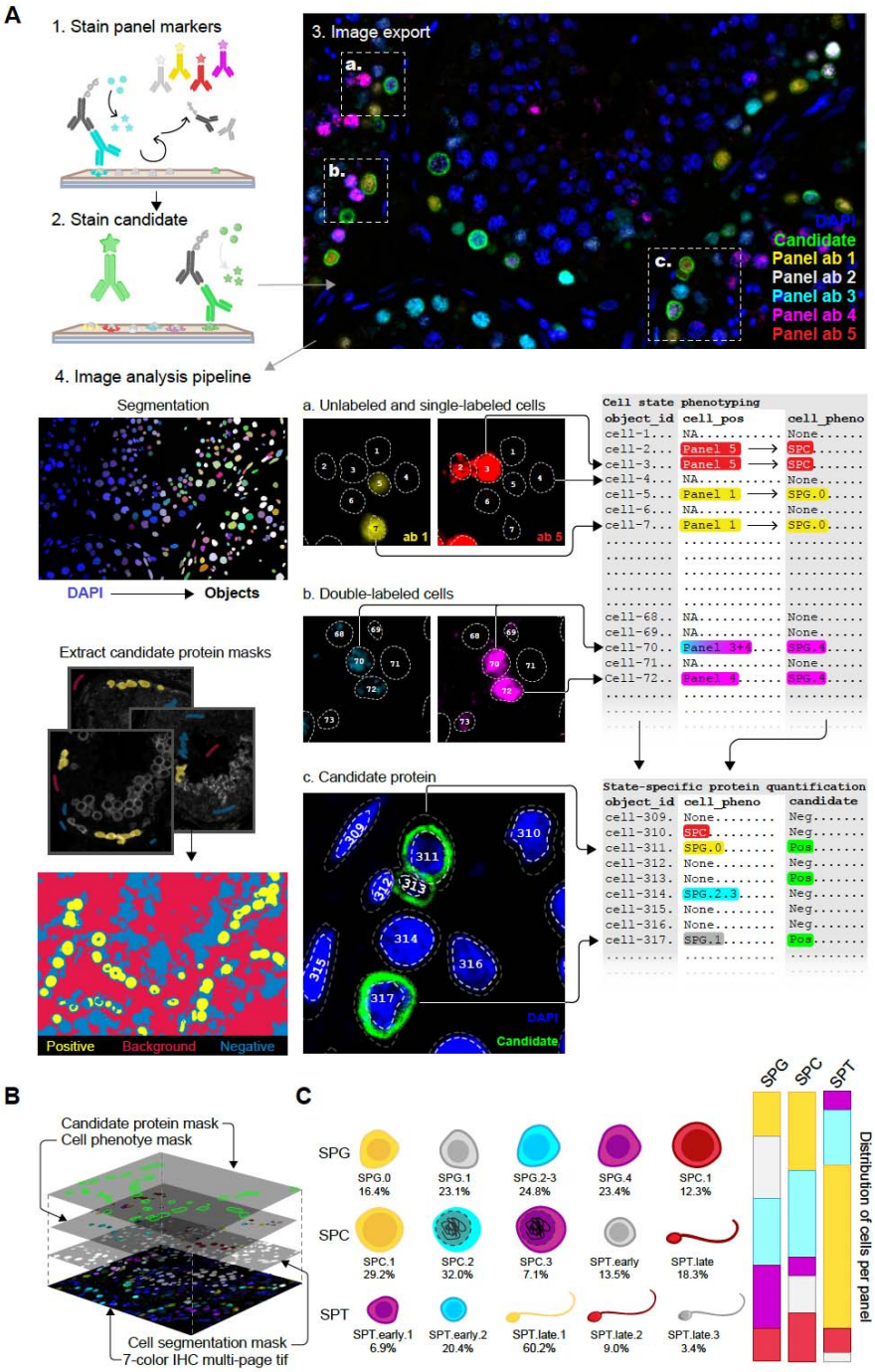


Fig. 4. High-resolution quantification and single-cell characterization based on bioimage analysis. (**A**) The iteratively stained testis slides were scanned in fluorescence and batch-extracted with panel-specific algorithms for proper multispectral separation of each fluorescent signal. The top right image showcases the protein LBR (green) stained with the SPG panel. The slide is scanned in 7-channel mode, and images are exported after a visual inspection of successful staining. Candidate = candidate protein. The nuclear DAPI signal is used for cell identification and segmentation, and the fluorescent signal for each candidate protein is thresholded using an image classifier that outputs positive (yellow area), negative (blue area), and background (red area) pixels and creates overlay masks for the positive signal only. Only the positive masks are used. In contrast, the panel antibody 1-5 fluorescent signals are used for cell phenotyping by automated thresholding methods to handle tabularize unlabeled and single-labeled cells (a) as well as double-labeled cells (b). The positive cell mask is then overlaid to create a table of cells with known differentiation state/phenotype and specificity for each candidate protein (c). (**B**) Overview of the three different

masks that are generated from each multiplex image. (**C**) A schematic illustration of different cells in each antibody panel with their pseudo-color and a stacked bar chart showing the percentage of each cell state for all stained proteins defined by the automated cell phenotyping and image analysis.

Functional analysis of spermatogenesis proteins in single-cell states

For visualizing the protein quantification based on image analysis, hierarchical clustering was performed using the mean intensity of each candidate protein stained with each mIHC panel, based on the cell state-specific expression profiles. Proteins within each resulting cluster were functionally assessed by Gene Ontology (GO) enrichment analysis.

Among the 228 candidate proteins stained with the SPG panel, five main clusters were identified and named according to their predominant expression patterns (Figure 5A and Supplementary Table 4). Cluster 2 comprised the largest group, with 81 proteins predominantly expressed in SPG.0 and SPG.1 cells. GO enrichment analysis showed that these proteins were mainly associated with transcriptional activity (Figure 5A). Notably, none of these proteins showed an apparent expression in SPC.1. Cluster 1 was dominated by proteins expressed in SPG.2-3. Among these 52 proteins, lamellipodium was the only significant function retained. Cluster 4 mainly contained proteins expressed in SPG.4 (53 proteins), associated with cell cycle checkpoint and transition, which is expected as this is the final cell stage before mitosis and transitioning into spermatocytes. Cluster 5, contained 23 proteins, including a group of proteins expressed in both SPG.0 and SPC.1, with functions mainly related to transcription. Finally, Cluster 3 (19 proteins) displayed a mixed expression pattern without a clearly dominant functional annotation.

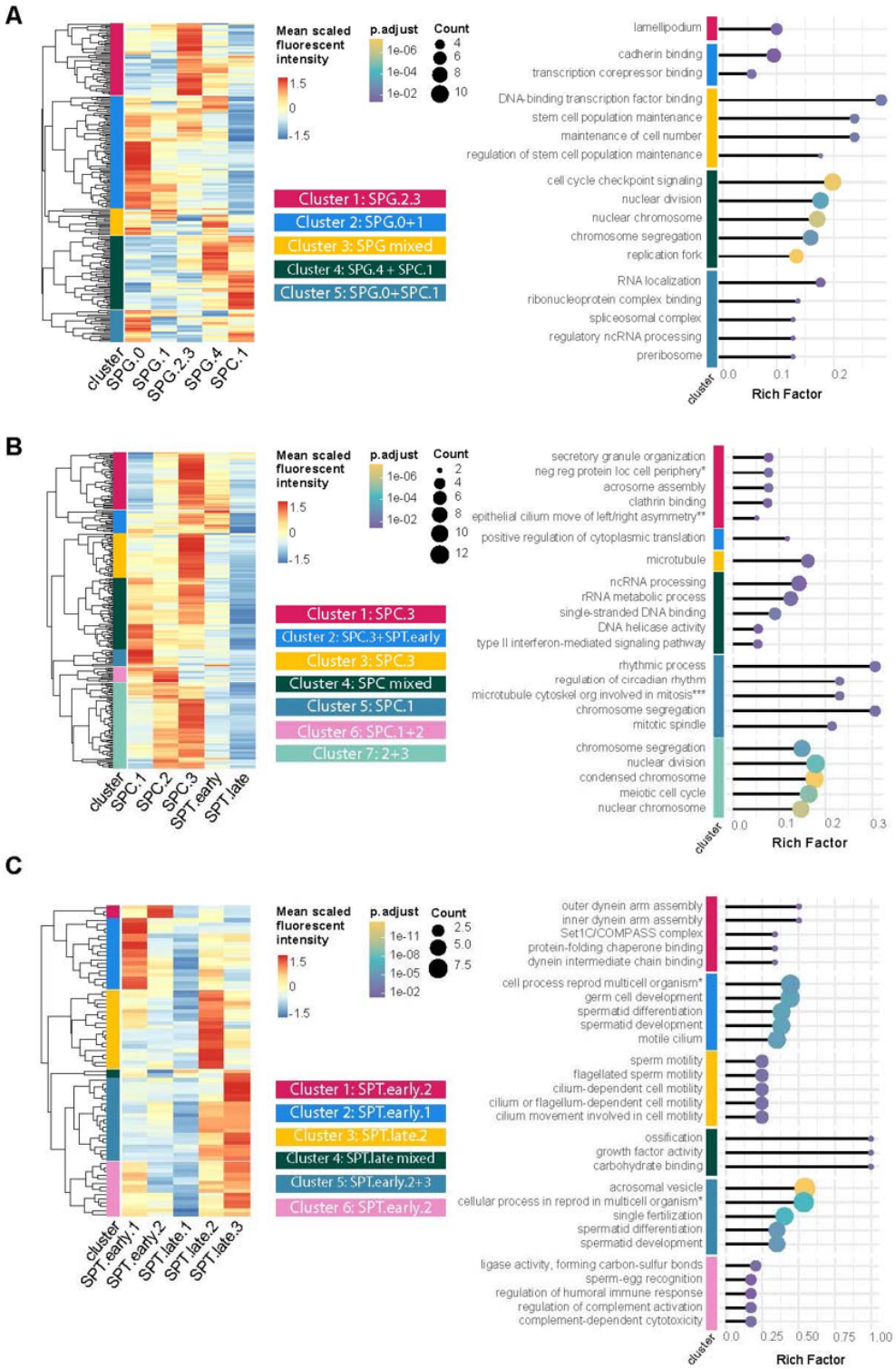


Fig. 5. Hierarchical clustering and functional analysis of proteins expressed in different germ cell phases. The data is visualized per panel for (**A**) SPG, (**B**) SPC, and (**C**) SPT. For heatmaps, scaled intensity values were z-score normalized and visualized per row. Clustering analysis using cut-tree separated the proteins according to their staining in each cell state. The clusters are numbered according to their order from top to bottom. For functional characterization, GO enrichment analysis was performed for each protein cluster. A few GO terms marked with an asterisk (*) were shortened for convenience.

Next, the SPC candidate proteins were subjected to hierarchical clustering (Figure 5B and Supplementary Table 4). This analysis identified clusters associated with specific cell states, as well as clusters with a broader expression pattern. Overall, most proteins were predominantly expressed in SPC.3 cells, and this

expression pattern dominated the majority of clusters (Figure 5B). However, the largest group, Cluster 7 (73 proteins) included proteins expressed in SPC.2. Clusters 3 (37 proteins), and Cluster 1 (48 proteins), were also enriched for SPC.3 expression, but showed lower expression in SPC.2 and SPT.early cells, respectively. Interestingly, Cluster 5 (14 proteins) was primarily associated with SPC.1 cells, while Cluster 6 (13 proteins) showed predominant expression in SPC.2. Cluster 4 (60 proteins) and Cluster 2 (19 proteins) displayed more mixed expression patterns, with the former including proteins in any of the spermatocyte states and the latter predominantly being expressed in SPC.3 and SPT.early germ cells. The functional assessment of the SPC clusters by GO analysis showed an overall association of these proteins with RNA processing, nuclear condensation, and structural chromosomal changes (Figure 5B).

For the SPT candidate proteins, six main expression clusters were identified (Figure 5C and Supplementary Table 4), including proteins specific to the early phases of spermiogenesis, proteins representing the transformation of haploid round cells, and finally, proteins present during the maturation of elongated sperm cells. Most proteins stained with the SPT panel were found in the later stages of sperm development, such as Cluster 5 and 6, with 21 and 14 proteins, respectively. The closely related Cluster 4 only included 2 proteins and showed a similar expression distribution to those in Cluster 5 and 6 but with more pronounced expression in SPT.late.1 cells. The 20 proteins in Cluster 3 were mainly specific to SPT.late.2 cells, while Cluster 2 (18 proteins) and Cluster 1 (3 proteins) were associated with the earlier phases of spermatids, SPT.early.1 and 2 cells. (Figure 5C). In general, the GO analysis indicated that these proteins are related to sperm flagellum assembly, development, and motility, as well as functions associated with fertilization and acrosomal vesicle generation.

To exemplify mIHC-based expression patterns for each of the different SPG, SPC, and SPT clusters, representative images together with image-analysis based quantification are shown in Figure 6. PABPC4 (SPG panel, Cluster 1), a protein that may be involved in cytoplasmic regulatory processes of mRNA metabolism (*20*), was predominantly expressed in SPG-2.3 cells. The melanoma-associated antigen B2 (MAGEB2) in SPG Cluster 2 was mainly present in SPG.0 cells. This protein is a cancer-testis antigen member of the MAGEB family, suggested to be involved in transcriptional regulation (*21*). CES2 protein is believed to participate in cholesterol metabolism (*22*) and was one of the proteins in SPG Cluster 3. CES2 was shown to be expressed in SPG.0 and 1. In SPG Cluster 4, the unknown protein HHIPL2 was mainly expressed in SPG.4 cells. FBL in SPG Cluster 5 showed a general speckle-like expression in all SPG states and SPC.1 cells. FBL (also known as fibrillarin) has a multi-faceted function and is a component of a nucleolar small nuclear ribonucleoprotein particle. FBL is thought to participate in processing pre-ribosomal RNAs (*23*) (Figure 6A). For proteins stained with SPC panel, representative stainings are presented as follows: the SPC Cluster 1 protein RFX3 was mainly positive in the nuclei of SPC.3 and SPT.early cells. RFX3 is involved in ciliogenesis (*24*). The SPC Cluster 2 protein, UBAC2, known to be involved in WNT-signaling (*25*) mainly showed a broad SPC and SPT staining pattern. HMMR, which is also known as CD168 or RHAMM, was included in SPC Cluster 3 and was mainly found in SPC.3 cells. HMMR is involved in cell motility (*26*). The DEAD box protein 11, or DDX11, was among the proteins in SPC Cluster 4 and was mainly expressed in the earlier states of spermatocyte cells, but with distinct staining in the other SPC cell states. DDX11 is essential for the cohesion of the chromosome arms and centromeres. If depleted, replicated chromosomes fail to segregate after the prometaphase arrest and ultimately lead to mitotic failure (*27*). The SPC Cluster 5 protein Cyclin-dependent kinase 1 (CDK1) is highly conserved and known to promote the G2-M phase during the cell cycle (*28*), and was confined to SPC.1 cells. The class I MHC molecule HLA-B was mainly found in SPC Cluster 6, dominated by expression in SPC.2 cells. Class I molecules play a role in antigen presentation (*29*). Lastly, among SPC Cluster 7 proteins, Spondin-2 or SPON2 is known to play a role in cell adhesion (*30*). SPON2 was observed in almost all cell states present in the SPC panel, however slightly more pronounced in SPC.2 and 3.

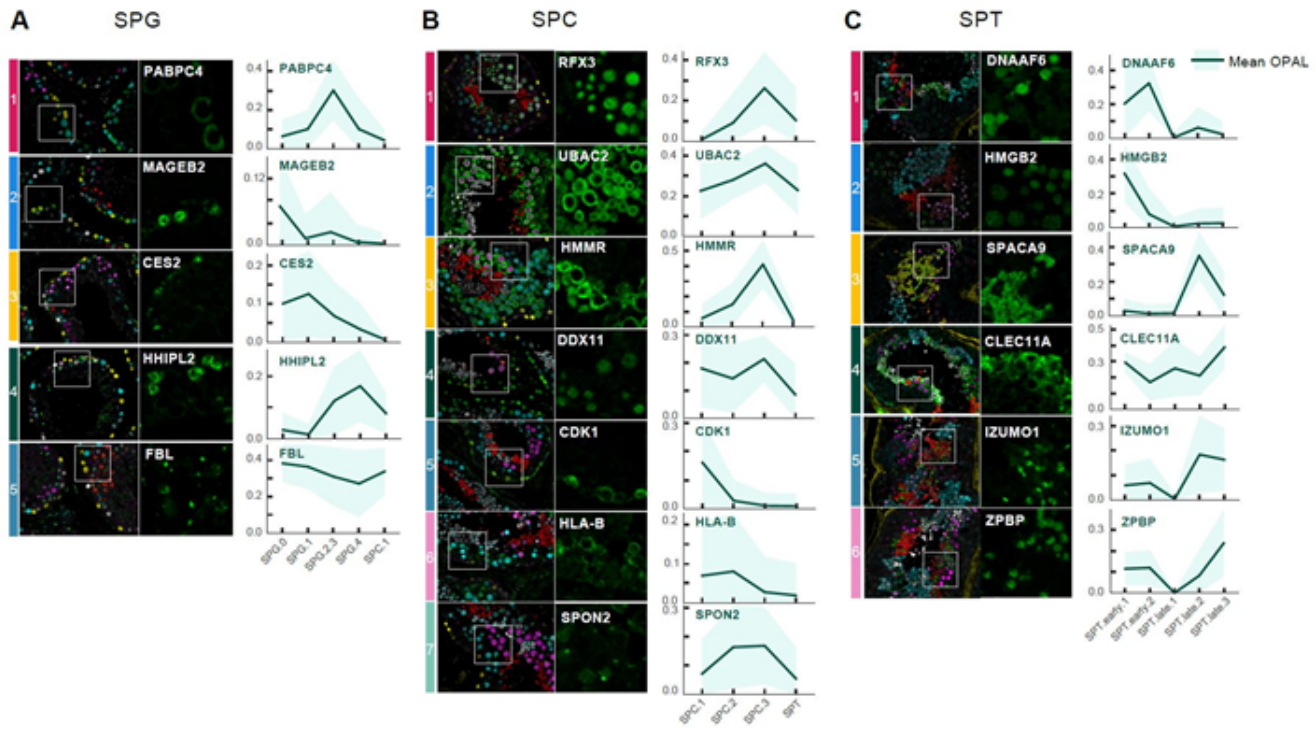


Fig. 6. Protein expression across spermatogenic states. Representative examples of proteins from each expression cluster for SPG (**A**), SPC (**B**), and SPT (**C**) proteins, as well as a line expression plot depicting the normalized protein expression across cell states based on image analysis quantification. Stained images display all markers in the left pane, and crop-in only shows the candidate protein staining (green). Clusters were defined by the heatmap expression profile in Figure 5.

Among proteins stained with the SPT panel, the SPT Cluster 1 protein, DNAAF6, which is needed for sperm flagella motility (*31*), was mainly found in earlier SPT states, namely SPT.early.2 cells. HMGB2 in SPT Cluster 2 was expressed in SPT.early.1 state and is known to be able to bend DNA and involved in double-strand breaks repair and recombination (*32*). SPACA9 is part of the dynein-associated microtubule proteins (*33*). SPACA9 was observed in SPT Cluster 3 and was seen in SPT.late.2 cells. From SPT Cluster 4, CLEC11A was shown to be in most SPT states, however mainly in SPT.late.1, 2 and 3 cells. CLEC11A is known to promote differentiation in mesenchymal cells (*34*). In SPT Cluster 5 and 6, IZUMO1 and ZPBP were proteins detected distinctly in SPT.late.2 and 3 cells. Both proteins are important for sperm-egg binding and IZUMO1 is essential for sperm-egg binding and fusion (*35*), and ZPBP is associated with the acrosome in sperm (*36*) (Figure 6C).

Cell state-specific expression dynamics between mRNA and protein levels

During spermatogenesis, two phases of inactive transcription have been identified and are known to be associated with homologous recombination of early meiosis and the cessation of mRNA production during the cellular transformation of haploid cells into elongating spermatids (*37*). Therefore, two major waves of transcription occur before and after the meiotic events (*38*). We hypothesized that our in-depth cell state-specific characterization at both the RNA and protein level would enable identification of genes for which mRNA expression precedes protein expression during spermatogenesis. To investigate this, correlation coefficients between mRNA and protein levels were calculated, and negatively correlated pairs were examined to identify discordant co-expression dynamics across the three different panels (Figure 7A).

Overall, most candidates displayed positive RNA–protein correlations, suggesting that mRNA is frequently translated within the same cell state in which it is transcribed. For example, the early spermatogonia marker, RHOXF1, showed a positive correlation between mRNA and protein expression (Figure 7B), with the highest expression in the SPG.0 state. Similarly, the transcription factor RFX2, specific to the end of meiosis, was predominantly expressed in SPC.3 at both the mRNA and protein levels

(Figure 7B). ACSBG2 also exhibited concordant dynamics, increasing linearly from SPT.early.1 to SPT.early.3 at both levels (Figure 7B).

In contrast, multiple inverse RNA-protein correlation cases were identified across the SPG, SPC, and SPT states (Figure 7A). Among these discordant RNA-protein pairs, the mRNA expression peaks in an earlier cell state than the protein. Of particular note is PIWIL4, previously described as a SPG.0 marker (*18*). Interestingly, in our dataset PIWIL4 is highly specific for SPG.0 at the mRNA level, but its corresponding protein is mainly expressed in SPG.2-3 and SPG.4 (Figure 7A). Another example of a protein with a discordant mRNA-to-protein expression profile is WNK1. The WNK1 transcript is primarily detected in SPC.1 with a steady decrease to SPT stage, but its corresponding protein expression peaks at SPC.3 state (Figure 7B). Similarly, ACTRT3 is expressed in SPT.early.1 and 2 at the RNA level, but the protein is present in later stages, predominantly in SPT.late.2 but also in SPT.late.3 (Figure 7B).

Finally, two candidates, VSIG4 (SPG panel) and R3HDML (SPC panel) returned NA values in the correlation analysis due to a lack of variability in the scRNA-seq dataset (Supplementary Table 9).

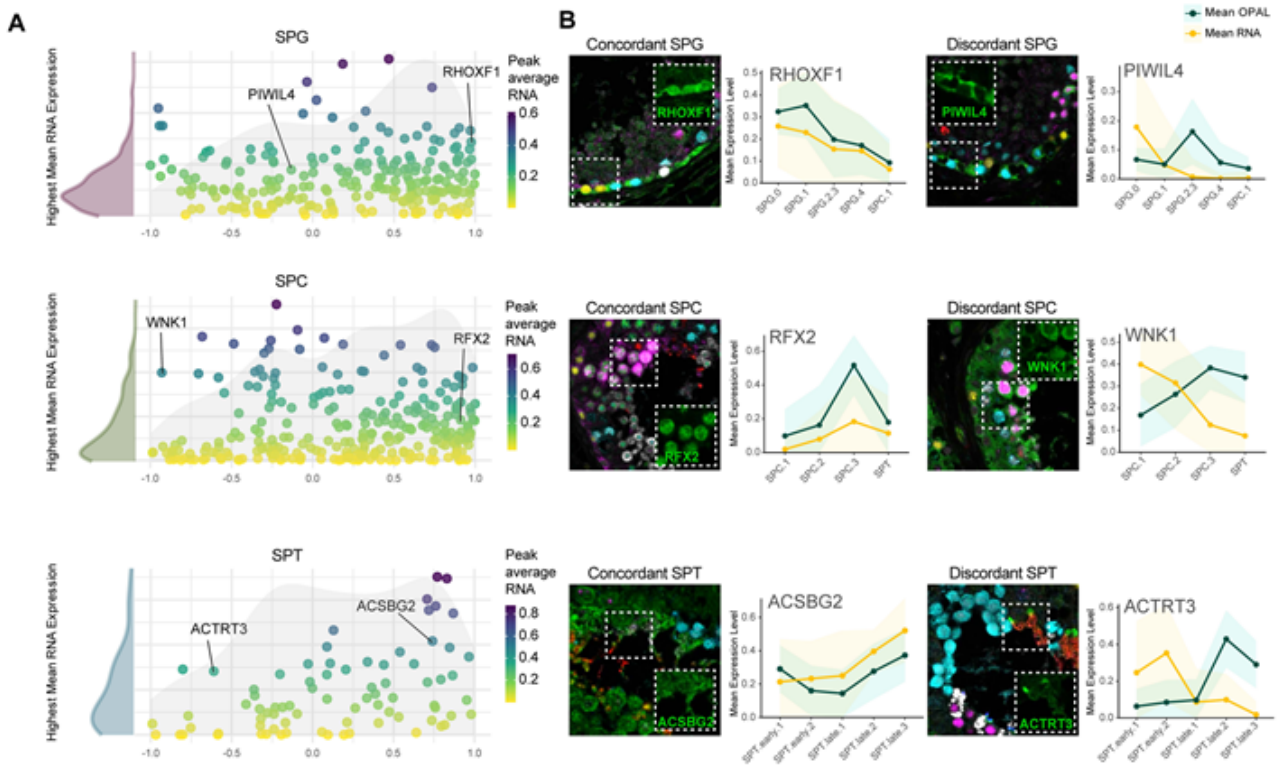


Fig. 7. The spatiotemporal expression dynamics of RNAs and proteins during spermatogenesis. (**A**) Pearson's correlation of mRNA and protein levels at single cell level for each panel. Each dot represents the correlation coefficients (-1 to 1) based on scaled mRNA and protein levels in SPG, SPC, and SPT panels (y-axis). The Highest Mean RNA Expression (x-axis) is derived from the normalized scRNA-seq levels and the peak mean RNA for each dot is displayed in the plot. The correlations were calculated in the direction of the first, most undifferentiated cell in each panel and towards the direction of the more differentiated cell state. Density plots show the distribution of the correlation coefficient (in grey) and peak mean RNA (colors). (**B**) Representative concordant and discordant stainings (candidate protein in green pseudo-color) with line expression plots for selected proteins stained with SPG, SPC, and SPT panels. The gene and protein expression is presented as mean scaled expression data from scRNA-seq and scaled quantitative levels from image analysis by mIHC staining. The highlights for RNA (yellow) and protein (green) represent the standard error.

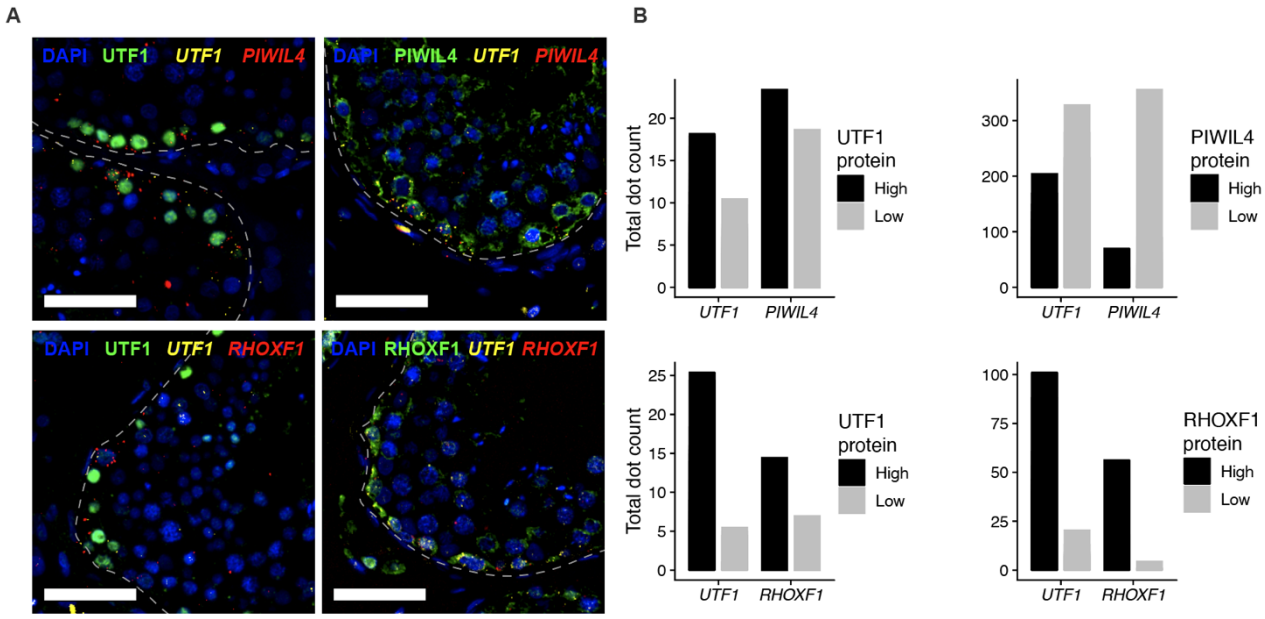


Fig. 8. Validation of RNA-to-protein expression dynamics of PIWIL4 and RHOXF1 in SPGs using RNAscope. (**A**) Representative fluorescent images of RNA-protein co-detection stainings of *UTF1*, *PIWIL4* and *RHOXF1* RNA (yellow and red), and protein expression (green) in seminiferous tubules of testis. Scale bar 50µm. (**B**) Total RNAscope puncta/dots summarized as bar plots per protein high and low cell group.

PIWIL4 shows a discordant mRNA-to-protein expression profile, while RHOXF1 is concordant

To validate the observed concordant and discordant RNA-protein relationship at single-cell resolution, we performed fluorescent in situ hybridization (RNAscope) combined with immunofluorescent staining for UTF1, PIWIL4, and RHOXF1 in human testis tissue sections, focusing on validating findings associated with the SPG panel (Figure 8A). Quantification of RNAscope signals revealed increased *UTF1* and *PIWIL4* transcript counts in UTF1 protein-high spermatogonia compared with UTF1 protein-low cells, whereas *UTF1* and *PIWIL4* RNA levels were enriched in PIWIL4 protein-low populations (Figure 8B, top). This supports a delayed mRNA-to-protein expression profile of PIWIL4, contradicting previous literature. *UTF1* and *RHOXF1* RNA probes were mainly detected in RHOXF1 protein-high spermatogonia (Figure 8B, bottom), supporting a concordant mRNA-to-protein relationship for RHOXF1.

Discussion

With the introduction of scRNA-seq, multiple efforts worldwide have generated high-quality datasets of human tissues (*1, 9*), which gives us the possibility to distinguish novel subsets of cells based on specific gene signatures. One such example is the identification of the spermatogonial SPG.0 state, with a gene expression profile closely related to infant germ cells, which suggests that these cells represent naïve stem cell-like cells (*19*). No large-scale study has however looked at the corresponding protein expression levels at a single-cell resolution comparable with the transcriptomics data, which is necessary for understanding how mRNA levels are translated into physiological and functional phenotypes. While it is known that transcription in testis occurs during two extensive waves before and after meiosis, and that the second transcription leads to the storage of mRNAs that will be needed at a later time-point during spermatogenesis, it has not been concluded which mRNAs that are directly translated into proteins, and which mRNA and protein expression patterns that vary from a temporal aspect. Mechanisms for post-transcriptional control have been shown to play a necessary role for normal spermatogenesis, and it can be anticipated that detailed analysis of the testis proteome ultimately leads to increased understanding of testicular function in health and disease.

Spatial analysis has the advantage of studying the expression in intact tissue samples in the context of neighboring cells, not biased by limitations in tissue dissociation that may affect proportions of cells. The testis constitutes a challenging tissue type to study due to its dynamic and complex biological processes during spermatogenesis when adult spermatogonial stem cells differentiate into mature sperm. Spermatogenesis can classically be divided into three main phases: (i) the proliferative mitotic phase in which spermatogonia undergo a series of mitotic divisions; (ii) the meiotic phase when genetic recombination occurs in primary diploid spermatocytes, which divide twice to produce haploid spermatids; (iii) and finally spermiogenesis when the spermatids morphologically transform into spermatozoa. During spermatogenesis, which lasts approximately 74 days in humans, thousands of genes and proteins are activated or repressed (*39*). To which extent and at which timepoint during this process all gene products are expressed and what functions they must maintain a healthy reproduction however remains to be explored. Since spermatogenesis is a continuous process, the definition of what can be called a cell type adds another level of complexity. Traditionally, the cell types have been distinguished based on their appearance, e.g. “dark” and “pale” spermatogonia according to the condensation of the nuclei, and round or elongated spermatids by looking at the shape. Already this distinction is however subjective and has challenges in manual interpretation. During the last phase of spermatogenesis, spermiogenesis, histones are replaced which causes a compaction of the spermatid chromatin. Genes such as Protamine 1 and 2 (PRM1 and 2) are transcribed at the early spermatids, before the cells transform and adopt their unique and characteristic appearance. For example, the 3′ UTRs of PRM1 mRNA have specific sequences that cause a “delayed” translation as these mRNAs are kept in a translationally repressed state bound to inactive ribonucleoprotein particles (*40*). Premature expression of PRM1 causes spermatid differentiation failure (*41*). In fact, the spermatogenesis process experiences periods of large translational output in an environment of inactive transcription (*37*). Our unique approach coupling scRNA-seq data with high-resolution spatial antibody-based protein data allowing for exact cell state specificity of proteins could offer a prerequisite to identify novel proteins with similar expression dynamics as the Protamine proteins.

We have previously done an extensive effort to classify IHC images stained as part of the HPA project in 5 different germ cell types and 3 somatic cell types (*13, 16*), giving a first clue towards the involvement of these proteins during spermatogenesis. Despite involving an expert for the interpretation, we could however show that when building an image analysis algorithm for automatic detection of the cell type-specific localization, multiple examples were identified where the human observer had made errors (*13*), highlighting the need for more quantitative measurements. Here, we sought to develop a stringent workflow for proteome-wide antibody-based analysis with a resolution not attainable with traditional IHC. By utilizing scRNA-seq data for identification of specific cell states and validating these with mIHC, we established multiplex panels that could be used for determining the exact single-cell state localization of 499 testicular germ cell-specific proteins based on expression overlap with the fixed antibody panels. To make the spatial proteomics data comparable with scRNA-seq, an automated image analysis pipeline was developed, allowing us to generate quantitative measurements of protein expression levels in each of the cell states. We were then able to cluster the proteins based on expression patterns, highlighting specific

groups of proteins that form the basis for further functional analysis. Most of the proteins studied have not previously been described at this detailed level in the context of spatial localization within testis, thereby adding knowledge to their presumed role within spermatogenesis. Interestingly, we also identified several proteins that showed a mismatch in correlation between mRNA and protein levels within the cell types studied in each of the panels, indicating that mRNA is stored, and the protein is translated at a later timepoint during spermatogenesis. Such conclusions would not be possible to draw from transcriptomics data alone, and these proteins constitute important targets for further functional studies on testis physiology. In addition to correlations of mRNA versus protein levels analyzed within each panel, it would be interesting to analyze the similar set of proteins across multiple panels and gross cell states.

Interestingly, our analysis revealed that many of the 499 investigated candidate proteins showed a variable correlation coefficient between RNA and protein levels in terms of cell state-specific expression pattern. It is particularly important to highlight that among these proteins, some have previously been described in literature as important markers for a certain cell state, but we here show that the actual protein is expressed at a later differentiation stage. One such example is the SPG protein PIWIL4, which at the mRNA level was shown to be expressed in SPG.0, while we here could validate that the maximum protein expression was detected later in SPG.2-3 and SPG.4 states. PIWIL4 was among the SPG Cluster 1 proteins with a functional association to lamellipodium, a type of cell membrane protrusion that drives migration in not only normal processes, but also pathological events (*42*). In fact, the highly conserved PIWI proteins are essential for self-renewal, and knock-down of PIWIL4 expression in vitro in a breast cancer cell line significantly decreased the migration ability of cancer cells (*43*). Similarly to UTF1, PIWIL4 expression has previously been regarded to be associated with the most naïve undifferentiated SPG population (*44*), and our novel data contradicting these previous findings urges further studies to map the functional characteristics of PIWIL4 and its role in spermatogenesis. As a concordant case in SPG, RHOXF1 was identified as mainly expressed in SPG.0 with decreasing expression towards the SPC phenotype. RHOXF1 is commonly used as a SPG hallmark, especially in some scRNA-seq studies (*45*). RHOXF1 belongs to the X-linked reproductive homeobox gene cluster and encodes a transcription factor with a crucial role in controlling spermatogonial proliferation and differentiation. Expectedly, deleterious RHOXF1-variants cause severe oligo and azoospermia (*46*). To further validate the mRNA-to-protein correlation of PIWIL4 and RHOXF1 observed by mIHC, we performed RNA-protein co-detection using RNAscope. By quantifying RNA punctate dots, we could confirm that *PIWIL4* and *UTF1* RNA was enriched in UTF1-protein high and PIWIL4-protein low SPGs. These data indicate that *PIWIL4* RNA is transcribed in earlier stages, and thus revealing a heterogeneous expression within the germ cell compartment. On the other hand, *RHOXF1* RNA presence correlated with RHOXF1 protein expression. These data highlight the dynamic transcriptional transitions associated with UTF1-, PIWIL4-, and RHOXF1-positive SPG populations.

In SPC Cluster 1, the transcription factor RFX2 showed both mRNA and protein mainly in SPC.3 cells. Among the enriched functions in Cluster 1, acrosome assembly was highly significant for the cluster and in RFX2-null mice, acrosomal defects have been detected (*47*). Further, RFX2 is previously reported as mainly being expressed in late pachytene stage which further cements its presence as a pachytene-elevated protein. In fact, the SPC.3 panel marker TCFL5 is also regarded as a pachytene protein, which increases the reliability of our findings (*48*). Similarly, WNK1 is known to be expressed at pachytene stage (*49*), further confirming the validity of our stainings. However, the mRNA expression peaks were mainly detected in SPC.1 state. WNK1 is essential for cell division (*50*) and it has also been shown that disruption of WNK1 might lead to premature protein synthesis of other mRNAs due to upregulation of MTOR (*49*). ACTRT3 is suggested to be involved in cytoskeletal organization and is highly similar to conventional actin protein (*51*). Confirming our results, ACTRT3 is known to be localized in the nucleus in mainly elongating and elongated spermatids (*52*).

Our analysis focused on the testicular germ cells, and to cover all the main cell states, we built three different custom-made panels. In addition to germ cells, it is known that the somatic niche cells (Sertoli cells, Leydig cells, and peritubular (myoid) cells) provide support for completing spermatogenesis, from spermatogonial stem cell stage to mature sperm. Future studies should therefore also include these cells, especially Sertoli cells, which are in direct contact with the sperm cells throughout spermatogenesis (*53*), to further elucidate protein-protein interactions between these cells. Nevertheless, we here show that by

using a 6-plex mIHC workflow, we were able to perform a large-scale analysis of almost 500 proteins and based on the quantitative spatial proteomics data link expression patterns with specific testicular functions. By utilizing entirely custom-made panels and a workflow that is highly compatible with regular IHC, our strategy can easily be set up in histological laboratories to increase the resolution of antibody-based imaging to a level comparable with scRNA-seq data. While fixed panels of five markers at a time do not cover all various cell types and cell states, multiple panels per tissue type can be built depending on the biological question, and for visualization, the human eye is anyway not able to interpret dozens of markers in the same field of view. Instead, our pipeline allows for high-throughput proteome-wide characterization, since the method is time and resource-saving due to not requiring antibody conjugation, and each candidate protein can be studied by just replacing one antibody in a specific cycle of the staining workflow.

In summary, we here show that by using an integrated workflow combining scRNA-seq analysis with information provided by mIHC we were able to leverage novel insights at different resolutions and study specific expression in cell states not previously described or distinguishable by one method alone. The stringent workflow with downstream automated image analysis is applicable for any type of human tissue and holds promise for defining the proteomes of each cell state and differentiation stage at a high molecular resolution, ultimately generating a detailed map of human cells. The data presented here is expected to be valuable for research on male reproduction, and the defined antibody panels constitute an important resource for further spatial proteomics studies on human testis.

Limitation of the study

This study was limited by a small sample size (testis tissue from four individuals, with three analyzed per assay), which may not fully capture protein expression variability across a broader population. Since the study focused only on adult testis tissue, it has limited utility for understanding factors relevant to infancy or pre-pubertal phases that could influence fertility. Health information for the men included was unavailable, presenting a risk that unknown medications or lifestyle factors may have influenced testicular function. Additionally, we focused solely on germ cells, excluding somatic cells, such as Sertoli cells, which play a critical role in orchestrating spermatogenesis.

Automated image segmentation may not fully distinguish specific cell types and states in the testis, and separating overlapping cells in tissue sections remains challenging. Moreover, the transverse tissue sections provide only a single ‘snapshot’ of protein expression, inherently omitting dynamic changes over time and representing a common limitation of most imaging-based methods.

Materials and Methods

Tissue preparation and ethical declaration

Histologically normal formalin-fixed and paraffin-embedded (FFPE) tissue samples from adult human testis were used for mIHC. The tissue collection consisted of samples from in total four individuals (#4485 Male 41 years, #4506 Male 32 years, #3627 Male 25 years, #4845 Male 29 years), out of which three individuals were used in each of the analyzes. Testis tissues were obtained from the Clinical Pathology department, Uppsala University Hospital (Sweden), and collected within the Uppsala Biobank organization. Samples were anonymized as determined in the approval and advisory report from the Uppsala Ethical Review Board (ref nos. 2002-577, 2005-388, 2007-159). For generation of tissue microarrays (TMAs), testis tissue blocks were punched at 1 mm in diameter using a manual tissue arrayer (MTA-1, Beecher Instruments) and placed in a recipient block, generating a TMA block of six cores comprising of doublet testis cores from three individuals. To generate the public IHC stainings available in the HPA database, standard TMAs as part of the HPA workflow were used, which included single cores of three individuals of adult human testis. Different individuals were used throughout the HPA project, and all IHC stainings available on www.proteinatlas.org are thus not from the same testis samples. The FFPE TMA blocks were cut in 4 μm sections using a waterfall microtome (Microm H355S, ThermoFisher Scientific, Freemont, CA), placed on SuperFrost Plus slides (Thermo Fisher Scientific, Freemont, CA), to be dried overnight at room temperature (RT). Slides were baked at 50°C for 12-24 h and cooled and stored at -20°C before use.

Sample preparation and slide pretreatment

Slides were brought to RT, deparaffinized in xylene, and rehydrated in graded alcohols (99.9, 96, and 80%) down to deionized water. Endogenous peroxidases were blocked using 0.3% hydrogen peroxide in the 96% alcohol step. Heat-induced epitope retrieval (HIER) was performed in a decloaking chamber (Biocare Medical, Walnut Creek, California, USA) at 125°C for 4 min in pH 6.0 Target Retrieval Solution (Agilent Technologies Inc., Santa Clara, California, USA) and quickly cooled to 90°C and then left to cool to RT before staining. Both chromogenic single-plex IHC (HPA standard) and mIHC were performed by using Lab Vision Autostainer 480S (Thermo Fisher Scientific, Freemont, CA). For mIHC slides, slides were bleached after HIER in a bleaching solution consisting of 0.2 M glycine, and 1.5% hydrogen peroxide in TBS-Tween (Thermo Fisher Scientific, TA-999-TT). Slides and bleaching solution were inserted in Falcon tubes and placed on rotation under LED light for 1 h in RT to reduce the overall effect of tissue auto-fluorescence. Slides were kept in TBS-Tween before mIHC staining. Stainings with single-plex IHC and 3, 3'-diaminobenzidine (DAB) (hereinafter referred to as chromogenic IHC) have been performed as previously described (*54*).

Annotation of IHC images

Digital images of IHC-stained transverse sections of testis tissue available on the HPA version 23 (Ensembl version 109) were selected for extended testis annotation. In addition to antibody validation, an inclusion criterion was that the staining pattern in the testis should be of at least moderate intensity in at least one of the cell types in the testis. The manual annotation data of adult human testis seminiferous epithelium was generated as previously described (*16*) and validated by a multi-class, multi-label hybrid feature-based deep learning model (*13*). In summary, for each antibody, the staining intensity (strong, moderate, weak, negative), quantity of stained cells (>75%, 25–75%, 1–24%, <1%) and sub-cellular localization (nuclear, cytoplasmic, membranous) in eight different testicular cell types, including five germ cell types (spermatogonia, preleptotene spermatocytes, pachytene spermatocytes, round/early spermatids, and elongated/late spermatids), and three somatic cell types (Sertoli cells, Leydig cells, and peritubular cells) was recorded. For this work, we only focused on the germ cell types, and therefore the somatic cells were excluded from the analysis. Please note that the annotation results of preleptotene spermatocytes and pachytene spermatocytes have been merged for all analyses in this work. A table with all annotations can be viewed in Supplementary Table 5.

Single-cells transcriptomics

To determine and characterize different populations of human testicular germ cells, publicly available UMI count tables were downloaded from the Gene Expression Omnibus (GEO) data repository. The inclusion criteria were adult and healthy tissue samples, sequencing performed using the 10X Genomics system, and no prior cell sorting. Two studies were identified as suitable: Guo et al. (2018) ((*18*)/ GSE112013), which included testis cells from three young healthy donors aged 17 to 25 years old, and Nie et al. (2022) (*19*), which included testis cells from 12 donors aged 17 to 76 years old. In the latter study, data were divided into three groups: young adults ("Young," n=4, 17-22 years old), older adults ("Older group-1," n=5, 62-66 years old), and a third group ("Older group-2," n=3, 64-76 years). The third group was excluded due to spermatogenic impairment. Three UMI count tables (one from GSE112013 and two from GSE182786) were loaded into R and analyzed using the Seurat package v4 (*55*). For each dataset, Seurat objects were created with a minimum of 3 cells and 500 features per cell. After standard QC, cells were filtered to retain those with 500-5000 expressed genes and less than 20% of the read mapped to the mitochondrial genome. Each dataset was preprocessed separately following the Seurat package vignette (https://satijalab.org/seurat/archive/v4.3/integration_rpca). Datasets were integrated using the reciprocal PCA (RPCA) method, and UMAP and clustering analysis was performed on the integrated object, testing resolutions from 0.4 to 1.2, with a final selection of 0.8. Each cluster was assigned an identity based on well-known markers and named according to the main cell type in that cluster (Supplementary Figure 1D). The count data for cells labeled as main germ cell types, spermatogonia (SPG), spermatocytes (SPC), and spermatids (Early.SPT + Late.SPT), were isolated and processed similarly to the previously described method (Supplementary Figure 2). This resulted in three independent Seurat objects for SPG, SPC, and SPT. Based on well-known markers, cell states were assigned accordingly. Only genes encoding proteins characterized by IHC with at least weak staining in any germ cells were kept for downstream analysis, and uncharacterized cells were removed.

To identify potential markers for constructing the three mIHC panels, the Seurat function FindAllMarkers was used with the Wilcoxon ranksum test, using the following settings: ‘min.pct = 0.20’, ‘logfc.threshold = 0.25’, and ‘only.pos = TRUE’; for the SPG, SPC, and SPT objects. A fold-change variable was defined based on the mean expression per sub-cluster in the RNA assays, representing the fold-change between the highest and the second-highest mean expression value. A fold-change cutoff of 1.5 was used as inclusion criteria. The resulting lists served as the starting point for comparison with IHC data to identify the best antibody combinations for each main germ cell type and their respective panels. To visualize the expression level across the different testicular cell states, Seurat’s DotPlot function was used. Mean expression for the protein candidate has been recovered for each of the main germ cell Seurat objects for visualization.

To further validate the lineage progression and developmental trajectory of the identified germ cell populations, we performed pseudotime analysis using the Monocle 3 package on R (Supplementary Figure 1G and 2).

Multiplex panel generation and staining

To select suitable antibodies as panel markers, we placed a large emphasis on staining specificity, primarily selecting proteins with germ cell-specific expression patterns over those with additional staining in somatic cells and stromal cells outside of the seminiferous tubules in testis, from the DEGs lists. If there was significant overlapping in panel marker stains (i.e., if a too large proportion of germ cells were double-positive), alternative markers were tested from the lists, or the order of the panel markers was re-ordered to ensure distinct and non-overlapping staining patterns.

Antibodies that were finally included in the antibody panels were first tested in single-plex mode to ensure compatibility with the OPAL detection system, similar specificity as in chromogenic IHC and to ensure an appropriately high signal-to-noise ratio, otherwise, antibodies were exchanged for another antibody/clone against the same antigen or exchanged for a protein with similar expression pattern and cell type specificity/distribution. For all three testis panels, a fixed order of OPAL fluorophores was used for rapid panel development. The OPALs were used as follows; first cycle OPAL690, second cycle OPAL620, fourth cycle OPAL570, fifth cycle OPAL480, and sixth cycle OPAL-DIG+780. The third cycle was always used with the candidate protein which was iteratively changed for each new slide stained. Low-

protein expressors of panel markers were assigned to brighter fluorophores when possible and more abundant markers were in general used with the dimmer fluorophores. Further, each antibody was tested at all different cycles to ensure optimal position and manually inspected to prevent apparent reduction and secure a well-represented staining distribution and overall pattern.

Slides were directly taken from bleaching and light treatment and subjected to mIHC staining protocol at RT. Staining was done using the IHC HRP rea-gent kit from Epredia (Epredia UltraVision LP HRP kit) and one cycle of staining was done as follows: blocking, primary antibody incubation, a secondary antibody with horseradish peroxidase (HRP) polymer, and an OPAL fluorophore (Akoya Biosciences, Marlborough, Massachusetts, USA), and antibody stripping in retrieval buffer in a decloaking chamber. After the last cycle, slides were incubated with the OPAL 780 fluorophore-conjugated anti-DIG antibody and 4′,6-diamidino-2-phenylindole (DAPI) (Invitrogen, D1306, Thermo Fisher Scientific). Slides were then mounted using Prolong Glass Antifade mounting medium and left to cure overnight in RT. Slides were scanned using PhenoImager (Akoya Biosciences) according to the manufacturer’s instructions. Acquired images were spectrally unmixed and image export was performed using the built-in synthetic spectral library of the in-Form software (Akoya Biosciences). For the SPG panel, a specific library was created by following the manufacturer's instructions due to the use of OPAL650 for fractions of stainings. See Supplementary Table 6 for detailed staining cycle information regarding panel marker antibodies, dilutions, reagents, incubation times, and OPAL fluorophores.

Selection of candidate proteins for staining in mIHC panels

Candidate proteins were selected to be stained primarily with one panel, and the main emphasis was to select proteins with cell type-specific staining patterns from the extended annotation of testis IHC images. We assumed high compatibility between standard chromogenic and multiplex stainings, as both stainings are similar in pretreatment and staining protocol, and therefore, the dilution factor of the primary candidate protein was based on previous staining results. In general, the dilution factor used in IHC was multiplied by ~1.5x, which was deemed an appropriate approach to stain primary antibodies of candidate proteins in the mIHC workflow. The resulting staining was evaluated to have high comparability with chromogenic IHC staining without apparent non-specific staining patterns vastly dissimilar to the previous stainings. Slides were restained if the staining for any of the panel markers was incomplete or dissimilar to a set of representative images after manual inspection. The same antibody/clone was, to the best of our ability, used if a protein was stained in multiple panels. However, for some proteins, anti-body availability and compatibility with the current panel forced us to use alternative rabbit pAbs directed toward the same protein.

Image Analysis

The fluorescence images were analyzed by processesing each channel separately. Cell boundaries were segmented from the DAPI channel and were subsequently used to retrieve the mean staining intensity value of all OPAL flours in each cell. Briefly explained the script segments cells and determines the phenotype of cells based on the panel marker fluorescence signal. Lastly, the candidate protein is overlaid on all phenotype cells, but also unstained/non-phenotyped cells. Each processed slide corresponds to one protein being processed for one panel, and the resulting data frame is used for all subsequent downstream analyses. In detail, the scanned slides were batch-exported using the Akoya Phenochart software (Akoya Biosciences) and using a synthetic library or panel-specific library, all according to the manufacturer's instructions. Each of the six TMA cores was exported as a separate tif-file with the channels and OPAL fluorophores in the following order: 1-DAPI; 2-OPAL480; 3-OPAL520; 4-OPAL570; 5-OPAL620; 6-OPAL690; 7-OPAL780; 8-Autofluorescence. The candidate protein channel for every slide (OPAL520-channel image) was individually extracted as a grayscale tiff file and used for segmenting positive signals by training a pixel-based algorithm in Ilastik image analysis software. By defining three classes for drawing (positive, negative, and background) and using brush stroke labels on 11% of the proteins, the algorithm predicts which class a pixel belongs to for all pixels (*56*). The output segmentation image was

used as a mask for specifying which cells were positive for each candidate protein. Next, the panel OPAL-channel images were individually extracted as grayscale and subjected to automated thresholding by Otsu/multi-otsu segmentation with the rationale that the marker stainings had a high signal-to-noise ratio and the automated approach captured the marker staining with high certainty. In the case of multi-otsu segmentation, the upper thresholded pixels were always regarded as the true positive signal for the panel marker, and the other levels below were considered negative/background staining. In general, the level of multi-otsu segmentation was as follows: for SPG panel images, all OPAL-channel images were set to 3-level segmentation. For SPC panel images, a regular Otsu segmentation was applied, and for SPT panel images, a 3-level multi-otsu segmentation was used for OPAL480, OPAL570, OPAL620, OPAL690, and a 4-level multi-otsu segmentation for OPAL780-channel images. The success of the segmentation method was evaluated for each staining individually, and when signal variability was high, panel-specific Otsu or multi-Otsu thresholds were adjusted to ensure accurate segmentation. Binary masks for all OPAL-channel images were produced and, in some cases, after manual inspection of a fraction of images, the Otsu/multi-otsu was increased or decreased to better fit the staining distribution of that specific slide/protein. Cells were defined with a DAPI channel image by using the pre-trained StarDist 2D cell segmentation model and the following input parameters: nms_threshold=0.7, probability_threshold=0.486166. The mean intensity values for all OPAL fluorophores in all cells (objects) were saved in a data frame and the mean intensity of all cells above the threshold after Otsu/multi-otsu thresholding were saved in a data frame. Further, we phenotyped cells based on panel OPAL positivity. For the candidate protein OPAL520-channel values, we applied a min-max scaling to rescale and normalize the values to allow comparison between proteins stained with the same panel marker.

Downstream-descriptive analysis

Image analysis data from SPG, SPC, and SPT panels were processed using R (version 4.3.2). Data Frames corresponding to each panel were imported and organized into structured lists. Metadata from the Laboratory Information Management System (LIMS) was integrated to filter cells from low-quality tissue cores, such as those lacking seminiferous ducts or showing insufficient panel marker representation. Phenotypes were assigned to each cell based on the panel markers' OPAL intensity, as described in Supplementary Table 7. Cells without an assigned phenotype were excluded from subsequent analyses. Summary tables were created to calculate the mean OPAL520 intensity (ranging from 0 to 1) for different cell types, and cell frequency tables were also generated. Hierarchical clustering was performed on the mean OPAL520 intensity data for each panel, and clusters were visualized using heatmaps generated with the Pheatmap R package. The gene clusters identified through hierarchical clustering were subjected to Gene Ontology (GO) enrichment analysis using the clusterProfiler Bioconductor package. Enrichment was evaluated for the Biological Process, Molecular Function, and Cellular Component ontologies. GO terms with an adjusted p-value < 0.05 were considered statistically significant, and a redundancy simplification step was applied with a cutoff of 0.7 for the adjusted p-value. The top five most significant GO terms for each gene cluster were visualized using a lollipop plot displaying the rich factor (ratio of genes in the cluster associated with a GO term to the total number of genes associated with that term), adjusted p-value (color-coded), and gene count (represented by point size) for each term.

Transcript and protein expression dynamics

The correlation analysis aimed to explore the linear relationship between RNA and protein expression. scRNA-seq data were obtained from the Seurat objects with each dataset corresponding to distinct cell populations of interest (SPG, SPC, and SPT). Protein expression data from multiplexed immuno-fluorescence (OPAL) assays were retrieved from the pre-processed data frames. Cells states were classified based on the lineage differentiation stages. RNA expression values were normalized using a min-max scaling approach, resulting in a 0-1 range for each gene. OPAL protein expression values were utilized in their pre-scaled form as described. For both RNA and protein datasets, we computed summary statistics including mean, median, standard deviation, minimum, and maximum expression levels. These

calculations were performed for each gene across the defined cell states. Gene expression patterns were visualized using line plots generated with the ggplot2 R package, where mean expression levels were shown alongside standard deviation ribbons. To quantify the relationship between RNA and protein expression, z-score normalization was applied to both datasets, standardizing expression values for each gene across all cell types. The normalized RNA and protein datasets were merged by gene identifiers and cell type classifications. Pearson correlation coefficients were calculated between RNA and protein expression profiles.

Scatter plots were created using ggplot2 to depict the relationship between the correlation coefficient and the highest mean RNA expression level. Density plots were also created using ggplot2 to visualize the distribution of the correlation coefficient and RNA peak intensity values.

RNAscope and antibody co-detection assay

Multiplex RNA in situ hybridization was performed on formalin-fixed paraffin-embedded (FFPE) tissue sections using the RNAscope Multiplex Fluorescent Reagent Kit v2–Hs together with the RNAscope 4-Plex Ancillary Kit (Advanced Cell Diagnostics). A protease-free PretreatPro workflow was employed to preserve tissue morphology and enable RNA-protein co-detection. FFPE tissues were sectioned at 4-5 µm onto Superfrost Plus slides, air-dried overnight, and baked at 60 °C for 1 h. Slides were deparaffinized using xylene followed by graded ethanol washes and rehydration in distilled water. Endogenous peroxidase activity was quenched using RNAscope Hydrogen Peroxide for 10 min at room temperature. Target retrieval was performed in 1× RNAscope Target Retrieval Buffer at 95–99 °C for 30–35 min (15 min for FFPE cell pellet controls), followed by immediate cooling in distilled water. Sections were subsequently treated with RNAscope Manual PretreatPro for 35 min at 40 °C in a humidified HybEZ oven. After pretreatment, slides proceeded directly to probe hybridization. Three experimental configurations were used: (i) four-plex RNA detection, (ii) dual RNA detection combined with immunofluorescence, and (iii) dual RNA detection alone. Channel 1 probe Hs-UTF1 (cat.no. 868061) was used ready-to-use, while Channel 2 probes (50x stocks) were diluted into Channel 1 to generate probe mixes. The following human gene probes were used: Hs-RHOXF1-C2 (cat.no. 1669551-C2), Hs-PIWIL4-C2 (cat.no. 807661-C2). Probes were pre-warmed to 40 °C and hybridized to tissue sections at 40 °C in a humidified chamber. RNA integrity and assay performance were assessed for each experiment using multiplex positive and negative control probes. Signal amplification was performed using the RNAscope AMP1-AMP3 reagents followed by sequential channel-specific HRP reactions. RNA targets were visualized using OPAL fluorophores (OPAL 520, 570, and 690). HRP was blocked between channels to prevent cross-reactivity. For RNA-protein co-detection stainings, immunofluorescence staining was performed after completion of RNAscope amplification. Sections were blocked for 1 h with 1X BSA blocking buffer (Thermo Fisher Scientific, cat. no 37520), followed by incubation with primary antibodies for 2 h: mouse anti-UTF1 (Merck, MAB4337, 1:500), rabbit anti-PIWIL4 (Atlas Antibodies, HPA036588, 1:100), and rabbit anti-RHOXF1 (Atlas Antibodies, HPA056506, 1:100). Secondary antibody incubation was performed for 45 min, and all antibody signals were developed using OPAL 520. Nuclei were counterstained with DAPI (Invitrogen, D1306, Thermo Fisher Scientific). Antibody staining steps followed manufacturers' instructions. Slides were mounted with antifade medium and scanned using an Akoya PhenoImager (Akoya Biosciences) at 20x magnification. Multispectral images were unmixed using a synthetic library and images were exported as multipage tiff. RNAscope stainings were performed on the same tissue microarray (TMA) used for multiplex staining experiments, as described previously. Image analysis was performed in QuPath (*57*). At least 10 seminiferous tubules per staining were manually circled and annotated and cells were segmented using the InstanSeg segementation tool, followed by subcellular detection of RNAscope puncta using "Subcellular Detections" tool. RNA positivity thresholds were manually determined by visual inspection for each slide to account for staining variability. RNA expression was quantified as total puncta per cell. For RNA-protein co-detection, cells were stratified into protein-high and protein-low populations based on manually defined intensity thresholds of the OPAL 520 antibody channel, after which RNA puncta counts were summarized separately for each group. A positive control was used and targeted three housekeeping genes, *POLR2A*, *PPIB*, and UBC which in theory should account for low, medium, and high RNA expression levels, and enable evaluation of RNA preservation

and tissue treatment efficiency. A negative control probe of the bacterial gene *dapB* was included to monitor nonspecific background (Supplementary Figure 6).

Quantification and statistical analysis

Detailed statistical methods for each stage of the study are provided in the respective sections of the Method in detail.

Acknowledgments

Funding: The project was funded by the Knut and Alice Wallenberg Foundation (2015.0344) and the Swedish Research Council (2022-02742). This publication has been made possible in part by the BioImage Informatics Facility BIIF, a unit of the National Bioinformatics Infrastructure Sweden NBIS, with funding from SciLifeLab, National Microscopy Infrastructure NMI (VR-RFI 2019-00217), and Chan Zuckerberg Initiative DAF (DAF2021-225261, DOI 10.37921/644085ggkbos, an advised fund of Silicon Valley Community Foundation, DOI 10.13039/100014989). The computations for image analysis were in part enabled by resources provided by the National Academic Infrastructure for Super-computing in Sweden (NAISS), partially funded by the Swedish Research Council (2022-06725).

Author contributions: Conceptualization: CL and MU; Data curation: FH and LM; Formal Analysis: FH, LM, GM, JG, BK, and RS; Funding acquisition: CL and MU; Investigation: FH, LM, JG, BK, and RS; Methodology: CL, FH, LM, GM, and JG; Project administration: JG and FH; Resources: LM, KvF, PA, GM, and MF; Soft-ware: LM, GM, FH; Supervision: CL, LM, CZ, and MU; Validation: FH, LM, and CL; Visualization: FH and LM; Writing – original draft: FH, LM, and CL; Writing – review & editing: FH, LM, CL, BK, RS, CZ, and MU. All authors have read and approved the final version of the manuscript.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data and images generated from this study are available in the Human Protein Atlas (https://www.proteinatlas.org/). All other data are available in the main text and the supplementary materials. All R codes and other source data are available upon request or at https://github.com/LindskogLab/A-spatiotemporal-atlas-of-human-spermatogenesis. The image analysis Python code is available on GitHub by accessing Jupyter notebooks and relevant src functions in this repository here: https://github.com/BIIFSweden/spermatogenesis-profiling (DOI:https://zenodo.org/doi/10.5281/zenodo.13788792).

References

1. A. Regev, S. A. Teichmann, E. S. Lander, I. Amit, C. Benoist, E. Birney, B. Bodenmiller, P. Campbell, P. Carninci, M. Clatworthy, The human cell atlas. *elife* **6**, e27041 (2017).

2. T. T. S. Consortium*, R. C. Jones, J. Karkanias, M. A. Krasnow, A. O. Pisco, S. R. Quake, J. Salzman, N. Yosef, B. Bulthaup, P. Brown, The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).

3. M. P. Snyder, S. Lin, A. Posgai, M. Atkinson, A. Regev, J. Rood, O. Rozenblatt-Rosen, L. Gaffney, A. Hupalowska, R. Satija, N. Gehlenborg, J. Shendure, J. Laskin, P. Harbury, N. A. Nystrom, J. C. Silverstein, Z. Bar-Joseph, K. Zhang, K. Börner, Y. Lin, R. Conroy, D. Procaccini, A. L. Roy, A. Pillai, M. Brown, Z. S. Galis, L. Cai, J. Shendure, C. Trapnell, S. Lin, D. Jackson, M. P. Snyder, G. Nolan, W. J. Greenleaf, Y. Lin, S. Plevritis, S. Ahadi, S. A. Nevins, H. Lee, C. M. Schuerch, S. Black, V. G. Venkataraaman, E. Esplin, A. Horning, A. Bahmani, K. Zhang, X. Sun, S. Jain, J. Hagood, G. Pryhuber, P. Kharchenko, M. Atkinson, B. Bodenmiller, T. Brusko, M. Clare-Salzler, H. Nick, K. Otto, A. Posgai, C. Wasserfall, M. Jorgensen, M. Brusko, S. Maffioletti, R. M. Caprioli, J. M. Spraggins, D. Gutierrez, N. H. Patterson, E. K. Neumann, R. Harris, M. deCaestecker, A. B. Fogo, R. van de Plas, K. Lau, L. Cai, G.-C. Yuan, Q. Zhu, R. Dries, P. Yin, S. K. Saka, J. Y. Kishi, Y. Wang, I. Goldaracena, J. Laskin, D. Ye, K. E. Burnum-Johnson, P. D. Piehowski, C. Ansong, Y. Zhu, P. Harbury, T. Desai, J. Mulye, P. Chou, M. Nagendran, Z. Bar-Joseph, S. A. Teichmann, B. Paten, R. F. Murphy, J. Ma, V. Y. Kiselev, C. Kingsford, A. Ricarte, M. Keays, S. A. Akoju, M. Ruffalo, N. Gehlenborg, P. Kharchenko, M. Vella, C. McCallum, K. Börner, L. E. Cross, S. H. Friedman, R. Heiland, B. Herr, P. Macklin, E. M. Quardokus, L. Record, J. P. Sluka, G. M. Weber, N. A. Nystrom, J. C. Silverstein, P. D. Blood, A. J. Ropelewski, W. E. Shirey, R. M. Scibek, P. Mabee, W. C. Lenhardt, K. Robasky, S. Michailidis, R. Satija, J. Marioni, A. Regev, A. Butler, T. Stuart, E. Fisher, S. Ghazanfar, J. Rood, L. Gaffney, G. Eraslan, T. Biancalani, E. D. Vaishnav, R. Conroy, D. Procaccini, A. Roy, A. Pillai, M. Brown, Z. Galis, P. Srinivas, A. Pawlyk, S. Sechi, E. Wilder, J. Anderson, B. C. Hu, G. Writing, U. T. Caltech, U. T. Stanford-Wash, T. M. C. Ucsd, T. M. C. University of Florida, T. M. C. Vanderbilt University, T. T. D. California Institute of Technology, T. T. D. Harvard, T. T. D. Purdue, T. T. D. Stanford, V. HuBmap Integration, T. C. Engagement Collaboratory: Carnegie Mellon, T. C. Harvard Medical School, M. C. Indiana University Bloomington, C. Pittsburgh Supercomputing, I. University of Pittsburgh, C. Engagement, C. C. University of South Dakota, M. C. New York Genome Center, N. I. H. H. W. Group, The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187-192 (2019).

4. X. Han, Z. Zhou, L. Fei, H. Sun, R. Wang, Y. Chen, H. Chen, J. Wang, H. Tang, W. Ge, Construction of a human cell landscape at single-cell level. *Nature* **581**, 303-309 (2020).

5. D. Wang, B. Eraslan, T. Wieland, B. Hallström, T. Hopf, D. P. Zolg, J. Zecha, A. Asplund, L. h. Li, C. Meng, A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Molecular systems biology* **15**, e8503 (2019).

6. V. Svensson, Droplet scRNA-seq is not zero-inflated. *Nature Biotechnology* **38**, 147-150 (2020).

7. L. Yang, J. George, J. Wang, Deep Profiling of Cellular Heterogeneity by Emerging Single-Cell Proteomic Technologies. *Proteomics* **20**, e1900226 (2020).

8. A. Mund, A.-D. Brunner, M. Mann, Unbiased spatial proteomics with single-cell resolution in tissues. *Molecular Cell* **82**, 2335-2349 (2022).

9. M. Karlsson, C. Zhang, L. Méar, W. Zhong, A. Digre, B. Katona, E. Sjöstedt, L. Butler, J. Odeberg, P. Dusart, A single–cell type transcriptomics map of human tissues. *Science Advances* **7**, eabh2169 (2021).

10. M. Uhlen, L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Ponten, Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

11. J. M. Taube, G. Akturk, M. Angelo, E. L. Engle, S. Gnjatic, S. Greenbaum, N. F. Greenwald, C. V. Hedvat, T. J. Hollmann, J. Juco, E. R. Parra, M. C. Rebelatto, D. L. Rimm, J. Rodriguez-Canales, K. A. Schalper, E. C. Stack, C. S. Ferreira, K. Korski, A. Lako, S. J. Rodig, E. Schenck, K. E. Steele, M. J. Surace, M. T. Tetzlaff, K. von Loga, I. I. Wistuba, C. B. Bifulco, The Society for Immunotherapy of Cancer statement on

best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) staining and validation. *Journal for ImmunoTherapy of Cancer* **8**, e000155 (2020).

12. Y. Clermont, Spermatogenesis in Man: A Study of the Spermatogonial Population. *Fertility and Sterility* **17**, 705-721 (1966).

13. B. Ghoshal, F. Hikmet, C. Pineau, A. Tucker, C. Lindskog, DeepHistoClass: A novel strategy for confident classification of immunohistochemistry images using Deep Learning. *Molecular & Cellular Proteomics* **20**, (2021).

14. M. D. Griswold, Spermatogenesis: The Commitment to Meiosis. *Physiological Reviews* **96**, 1-17 (2016).

15. E. Eddy, in *Seminars in cell & developmental biology*. (Elsevier, 1998), vol. 9, pp. 451-457.

16. C. Pineau, F. Hikmet, C. Zhang, P. Oksvold, S. Chen, L. Fagerberg, M. Uhlén, C. Lindskog, Cell type-specific expression of testis elevated genes based on transcriptomics and antibody-based proteomics. *Journal of proteome research* **18**, 4215-4230 (2019).

17. Y. Vandenbrouck, L. Lane, C. Carapito, P. Duek, K. Rondel, C. Bruley, C. Macron, A. Gonzalez de Peredo, Y. Coute, K. Chaoui, Looking for missing proteins in the proteome of human spermatozoa: an update. *Journal of proteome research* **15**, 3998-4019 (2016).

18. J. Guo, E. J. Grow, H. Mlcochova, G. J. Maher, C. Lindskog, X. Nie, Y. Guo, Y. Takei, J. Yun, L. Cai, The adult human testis transcriptional cell atlas. *Cell research* **28**, 1141-1157 (2018).

19. X. Nie, S. K. Munyoki, M. Sukhwani, N. Schmid, A. Missel, B. R. Emery, J.-B. Stukenborg, A. Mayerhofer, K. E. Orwig, K. I. Aston, Single-cell analysis of human testis aging and correlation with elevated body mass index. *Developmental cell* **57**, 1160-1176. e1165 (2022).

20. S. G. Gregory, K. Barlow, K. McLay, R. Kaul, D. Swarbreck, A. Dunham, C. Scott, K. Howe, K. Woodfine, C. Spencer, The DNA sequence and biological annotation of human chromosome 1. *Nature* **441**, 315-321 (2006).

21. L. Y. Peche, M. F. Ladelfa, M. F. Toledo, M. Mano, J. E. Laiseca, C. Schneider, M. Monte, Human MageB2 protein expression enhances E2F transcriptional activity, cell proliferation, and resistance to ribotoxic stress. *Journal of Biological Chemistry* **290**, 29652-29662 (2015).

22. S. P. Sanghani, P. C. Sanghani, M. A. Schiel, W. F. Bosron, Human carboxylesterases: an update on CES1, CES2 and CES3. *Protein and peptide letters* **16**, 1207-1214 (2009).

23. M. Yanagida, T. Hayano, Y. Yamauchi, T. Shinkawa, T. Natsume, T. Isobe, N. Takahashi, Human fibrillarin forms a sub-complex with splicing factor 2-associated p32, protein arginine methyltransferases, and tubulins α3 and β1 that is independent of its association with preribosomal ribonucleoprotein complexes. *Journal of Biological Chemistry* **279**, 1607-1614 (2004).

24. L. Didon, R. K. Zwick, I. W. Chao, M. S. Walters, R. Wang, N. R. Hackett, R. G. Crystal, RFX3 modulation of FOXJ1 regulation of cilia genes in the human airway epithelium. *Respiratory research* **14**, 1-13 (2013).

25. J. H. Choi, X. Zhong, W. McAlpine, T.-C. Liao, D. Zhang, B. Fang, J. Russell, S. Ludwig, E. Nair-Gill, Z. Zhang, LMBR1L regulates lymphopoiesis through Wnt/β-catenin signaling. *Science* **364**, eaau0812 (2019).

26. L. M. Pilarski, E. Pruski, J. Wizniak, D. Paine, K. Seeberger, M. J. Mant, C. B. Brown, A. R. Belch, Potential role for hyaluronan and the hyaluronan receptor RHAMM in mobilization and trafficking of hematopoietic progenitor cells. *Blood, The Journal of the American Society of Hematology* **93**, 2918-2927 (1999).

27. J. L. Parish, J. Rosa, X. Wang, J. M. Lahti, S. J. Doxsey, E. J. Androphy, The DNA helicase ChlR1 is required for sister chromatid cohesion in mammalian cells. *Journal of cell science* **119**, 4857-4865 (2006).

28. G. Draetta, D. Beach, Activation of cdc2 protein kinase during mitosis in human cells: cell cycle-dependent phosphorylation and subunit rearrangement. *Cell* **54**, 17-26 (1988).

29. H. N. Kløverpris, D. K. Cole, A. Fuller, J. Carlson, K. Beck, A. J. Schauenburg, P. J. Rizkallah, S. Buus, A. K. Sewell, P. Goulder, A molecular switch in immunodominant HIV-1-specific CD8 T-cell epitopes shapes differential HLA-restricted escape. *Retrovirology* **12**, 1-11 (2015).

30. J. Zhang, G. Liu, Y. Liu, P. Yang, J. Xie, X. Wei, The biological functions and related signaling pathways of SPON2. *Frontiers in Oncology* **13**, 1323744 (2024).

31. F. Dong, K. Shinohara, Y. Botilde, R. Nabeshima, Y. Asai, A. Fukumoto, T. Hasegawa, M. Matsuo, H. Takeda, H. Shiratori, Pih1d3 is required for cytoplasmic preassembly of axonemal dynein in mouse sperm. *Journal of Cell Biology* **204**, 203-213 (2014).

32. B. M. Shykind, J. Kim, P. A. Sharp, Activation of the TFIID-TFIIA complex with HMG-2. *Genes & development* **9**, 1354-1365 (1995).

33. M. Gui, J. T. Croft, D. Zabeo, V. Acharya, J. M. Kollman, T. Burgoyne, J. L. Höög, A. Brown, SPACA9 is a lumenal protein of human ciliary singlet and doublet microtubules. *Proceedings of the National Academy of Sciences* **119**, e2207605119 (2022).

34. R. Yue, B. Shen, S. J. Morrison, Clec11a/osteolectin is an osteogenic growth factor that promotes the maintenance of the adult skeleton. *Elife* **5**, e18782 (2016).

35. J. Vondrakova, M. Frolikova, L. Ded, J. Cerny, P. Postlerova, V. Palenikova, O. Simonik, Z. Nahacka, K. Basus, E. Valaskova, MAIA, Fc receptor–like 3, supersedes JUNO as IZUMO1 receptor during human fertilization. *Science advances* **8**, eabn0047 (2022).

36. A. N. Yatsenko, D. S. O'Neil, A. Roy, P. A. Arias-Mendoza, R. Chen, L. J. Murthy, D. J. Lamb, M. M. Matzuk, Association of mutations in the zona pellucida binding protein 1 (ZPBP1) gene with abnormal sperm head morphology in infertile men. *Molecular human reproduction* **18**, 14-21 (2012).

37. J. M. Sutherland, N. A. Siddall, G. R. Hime, E. A. McLaughlin, RNA binding proteins in spermatogenesis: an in depth focus on the Musashi family. *Asian journal of andrology* **17**, 529-536 (2015).

38. S. de Mateo, P. Sassone-Corsi, in *Seminars in cell & developmental biology*. (Elsevier, 2014), vol. 29, pp. 84-92.

39. P. Calvel, A. D. Rolland, B. Jégou, C. Pineau, Testicular postgenomics: targeting the regulation of spermatogenesis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 1481-1500 (2010).

40. K. Steger, K. Pauls, T. Klonisch, F. E. Franke, M. Bergmann, Expression of protamine-1 and-2 mRNA during human spermiogenesis. *Molecular human reproduction* **6**, 219-225 (2000).

41. K. Lee, H. S. Haugen, C. H. Clegg, R. E. Braun, Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. *Proceedings of the National Academy of Sciences* **92**, 12451-12455 (1995).

42. M. Krause, A. Gautreau, Steering cell migration: lamellipodium dynamics and the regulation of directional persistence. *Nat Rev Mol Cell Biol* **15**, 577-590 (2014).

43. Z. Wang, N. Liu, S. Shi, S. Liu, H. Lin, The role of PIWIL4, an argonaute family protein, in breast cancer. *Journal of Biological Chemistry* **291**, 10646-10658 (2016).

44. H. Ba Omar, J. Stevens, A. Haavisto, Y. Cui, F. Harteveld, Y. Yang, R. Bjarnason, P. Romerius, M. Sundin, U. N. Nyström, C. Langenskiöld, H. Vogt, P. Frisk, K. Vepsäläinen, C. Petersen, L. Cui, J. Guo, K. Jahnukainen, J.-B. Stukenborg, Early transcriptional states of spermatogonia and marker expressions in the prepubertal human testis following chemotherapy-induced depletion. *Human Reproduction* **40**, 1467-1475 (2025).

45. A. Sohni, K. Tan, H. W. Song, D. Burow, D. G. de Rooij, L. Laurent, T. C. Hsieh, R. Rabah, S. S. Hammoud, E. Vicini, M. F. Wilkinson, The Neonatal and Adult Human Testis Defined at the Single-Cell Level. *Cell Rep* **26**, 1501-1517.e1504 (2019).

46. S. Yi, W. Wang, L. Su, L. Meng, Y. Li, C. Tan, Q. Liu, H. Zhang, L. Fan, G. Lu, L. Hu, J. Du, G. Lin, Y.-Q. Tan, C. Tu, Q. Zhang, Deleterious variants in X-linked RHOXF1 cause male infertility with oligo- and azoospermia. *Molecular Human Reproduction* **30**, (2024).

47. W. S. Kistler, D. Baas, S. Lemeille, M. Paschaki, Q. Seguin-Estevez, E. Barras, W. Ma, J. L. Duteyrat, L. Morlé, B. Durand, W. Reith, RFX2 Is a Major Transcriptional Regulator of Spermiogenesis. *PLoS Genet* **11**, e1005368 (2015).

48. J. Galán-Martínez, I. Berenguer, M. Del Carmen Maza, K. Stamatakis, N. Gironès, M. Fresno, TCFL5 deficiency impairs the pachytene to diplotene transition during spermatogenesis in the mouse. *Sci Rep* **12**, 10956 (2022).

49. R. A. Chi, X. Xu, J. L. Li, X. Xu, G. Hu, P. Brown, C. Willson, O. Kirsanov, C. Geyer, C. L. Huang, M. Morgan, F. DeMayo, WNK1 is required during male pachynema to sustain fertility. *iScience* **26**, 107616 (2023).

50. S. W. Tu, A. Bugde, K. Luby-Phelps, M. H. Cobb, WNK1 is required for mitosis and abscission. *Proc Natl Acad Sci U S A* **108**, 1385-1390 (2011).

51. M. Harata, K. Nishimori, S. Hatta, Identification of two cDNAs for human actin-related proteins (Arps) that have remarkable similarity to conventional actin. *Biochim Biophys Acta* **1522**, 130-133 (2001).

52. Y. Hara, K. Yamagata, K. Oguchi, T. Baba, Nuclear localization of profilin III-ArpM1 complex in mouse spermiogenesis. *FEBS Lett* **582**, 2998-3004 (2008).

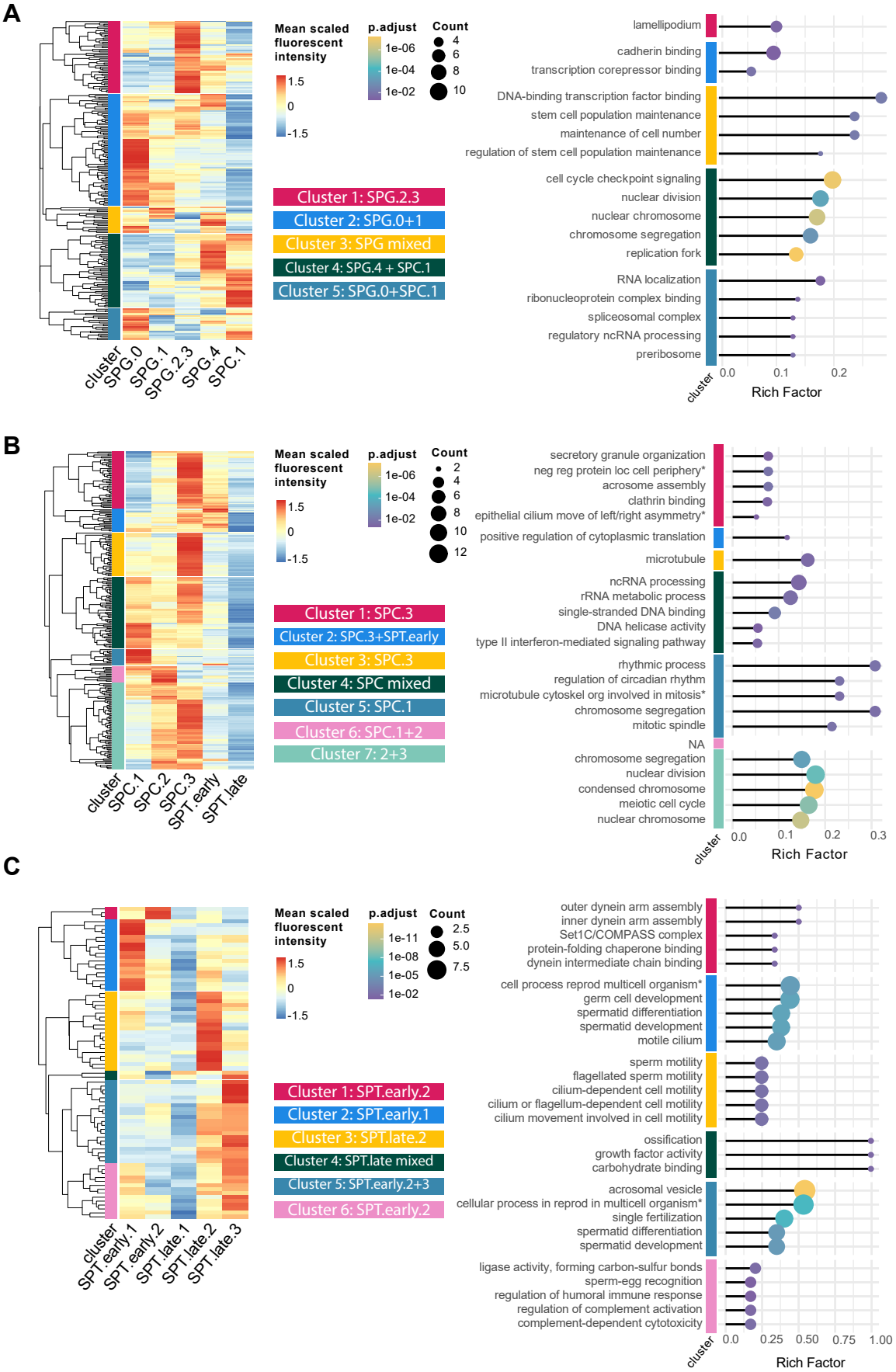
53. J. Guo, E. Sosa, T. Chitiashvili, X. Nie, E. J. Rojas, E. Oliver, K. Plath, J. M. Hotaling, J.-B. Stukenborg, A. T. Clark, Single-cell analysis of the developing human testis reveals somatic niche cell specification and fetal germline stem cell establishment. *Cell Stem Cell* **28**, 764-778. e764 (2021).

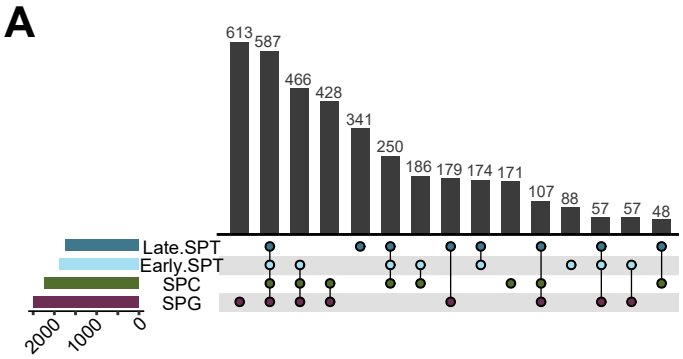
54. C. Kampf, I. Olsson, U. Ryberg, E. Sjöstedt, F. Pontén, Production of tissue microarrays, immunohistochemistry staining and digitalization within the human protein atlas. *JoVE (Journal of Visualized Experiments)*, e3620 (2012).

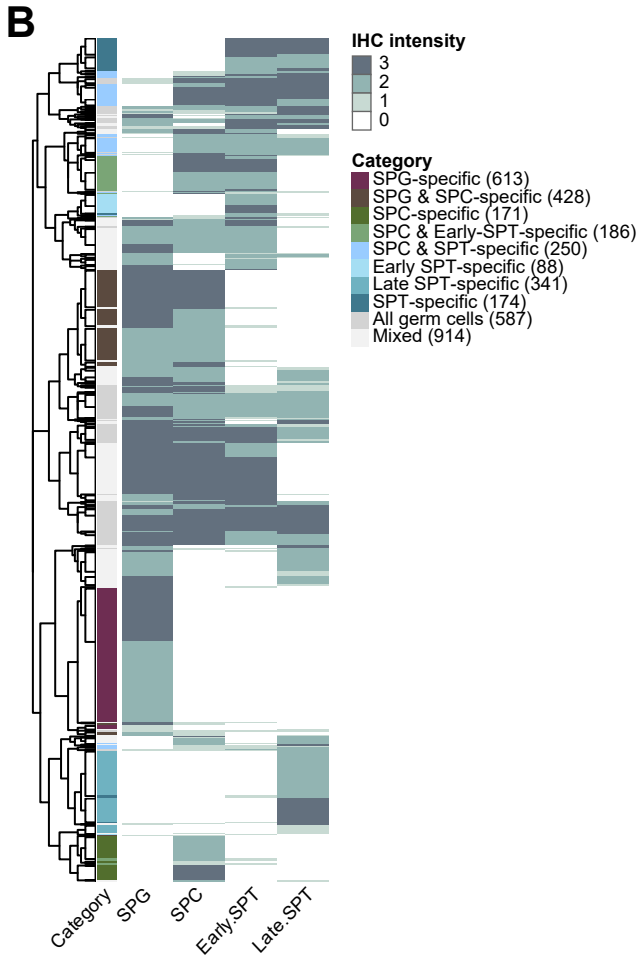
55. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587. e3529 (2021).

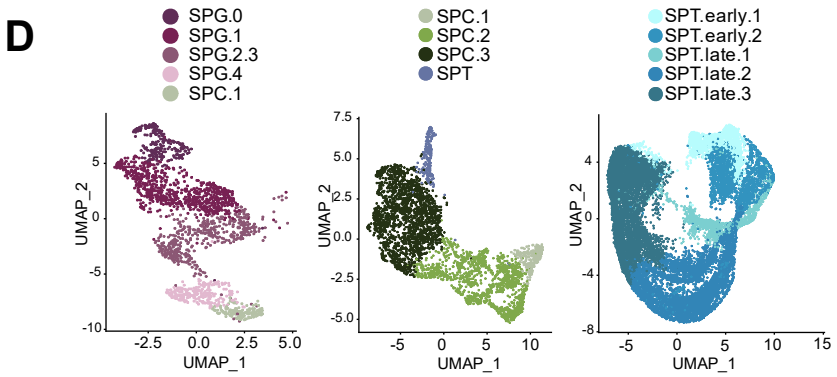
56. S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, Ilastik: interactive machine learning for (bio) image analysis. *Nature methods* **16**, 1226-1232 (2019).

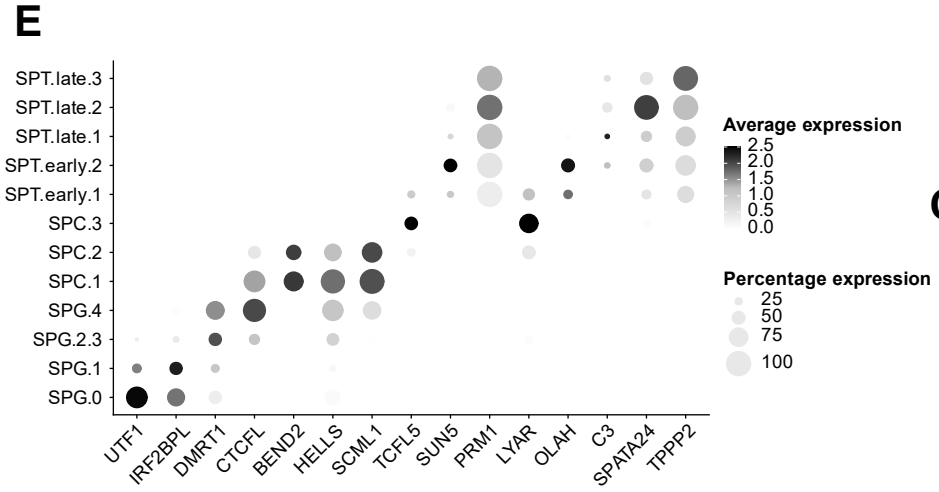
57. P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman, J. A. James, M. Salto-Tellez, P. W. Hamilton, QuPath: Open source software for digital pathology image analysis. *Scientific Reports* **7**, 16878 (2017).

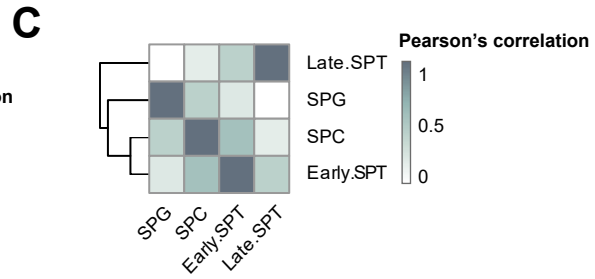

A
Mean scaled fluorescent intensity
1.5
0
-1.5
p.adjust
1e-06
1e-04
1e-02
Count
4
6
8
10
cluster
SPG.0
SPG.1
SPG.2.3
SPG.4
SPC.1
Cluster 1: SPG.2.3
Cluster 2: SPG.0+1
Cluster 3: SPG mixed
Cluster 4: SPG.4 + SPC.1
Cluster 5: SPG.0+SPC.1
lamellipodium
cadherin binding
transcription corepressor binding
DNA-binding transcription factor binding
stem cell population maintenance
maintenance of cell number
regulation of stem cell population maintenance
cell cycle checkpoint signaling
nuclear division
nuclear chromosome
chromosome segregation
replication fork
RNA localization
ribonucleoprotein complex binding
spliceosomal complex
regulatory ncRNA processing
preribosome
0.0
0.1
0.2
Rich Factor
B
Mean scaled fluorescent intensity
1.5
0
-1.5
p.adjust
1e-06
1e-04
1e-02
Count
2
4
6
8
10
12
cluster
SPC.1
SPC.2
SPC.3
SPT.early
SPT.late
Cluster 1: SPC.3
Cluster 2: SPC.3+SPT.early
Cluster 3: SPC.3
Cluster 4: SPC mixed
Cluster 5: SPC.1
Cluster 6: SPC.1+2
Cluster 7: 2+3
secretory granule organization
neg reg protein loc cell periphery*
acrosome assembly
clathrin binding
epithelial cilium move of left/right asymmetry*
positive regulation of cytoplasmic translation
microtubule
ncRNA processing
rRNA metabolic process
single-stranded DNA binding
DNA helicase activity
type II interferon-mediated signaling pathway
rhythmic process
regulation of circadian rhythm
microtubule cytoskel org involved in mitosis*
chromosome segregation
mitotic spindle
NA
chromosome segregation
nuclear division
condensed chromosome
meiotic cell cycle
nuclear chromosome
0.0
0.1
0.2
0.3
Rich Factor
C
Mean scaled fluorescent intensity
1.5
0
-1.5
p.adjust
1e-11
1e-08
1e-05
1e-02
Count
2.5
5.0
7.5
cluster
SPT.early.1
SPT.early.2
SPT.late.1
SPT.late.2
SPT.late.3
Cluster 1: SPT.early.2
Cluster 2: SPT.early.1
Cluster 3: SPT.late.2
Cluster 4: SPT.late mixed
Cluster 5: SPT.early.2+3
Cluster 6: SPT.early.2
outer dynein arm assembly
inner dynein arm assembly
Set1C/COMPASS complex
protein-folding chaperone binding
dynein intermediate chain binding
cell process reprod multicell organism*
germ cell development
spermatid differentiation
spermatid development
motile cilium
sperm motility
flagellated sperm motility
cilium-dependent cell motility
cilium or flagellum-dependent cell motility
cilium movement involved in cell motility
ossification
growth factor activity
carbohydrate binding
acrosomal vesicle
cellular process in reprod in multicell organism*
single fertilization
spermatid differentiation
spermatid development
ligase activity, forming carbon-sulfur bonds
sperm-egg recognition
regulation of humoral immune response
regulation of complement activation
complement-dependent cytotoxicity
0.0
0.25
0.50
0.75
1.00
Rich Factor

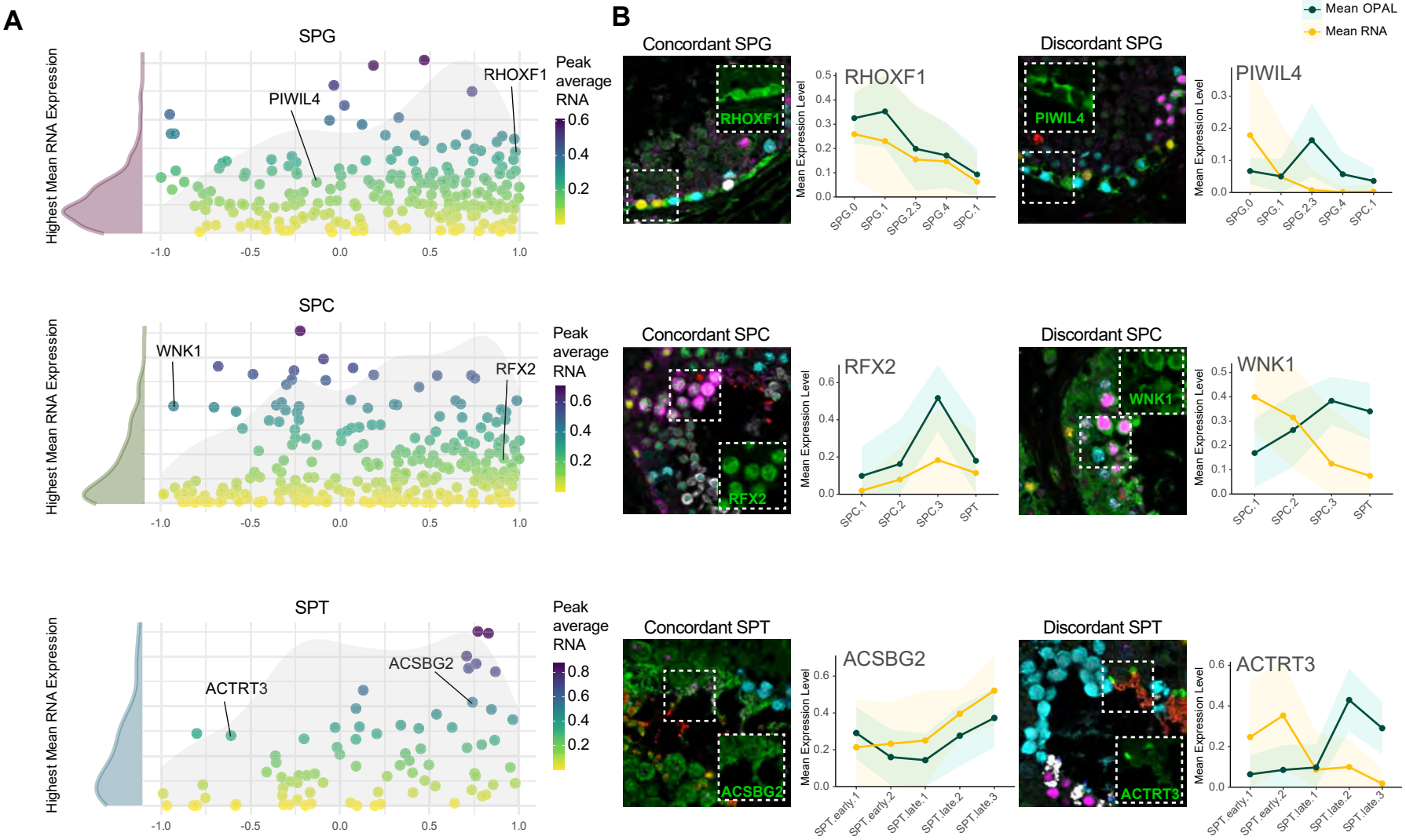

A
613
587
466
428
341
250
186
179
174
171
107
88
57
57
48
Late.SPT
Early.SPT
SPC
SPG
2000
1000
0

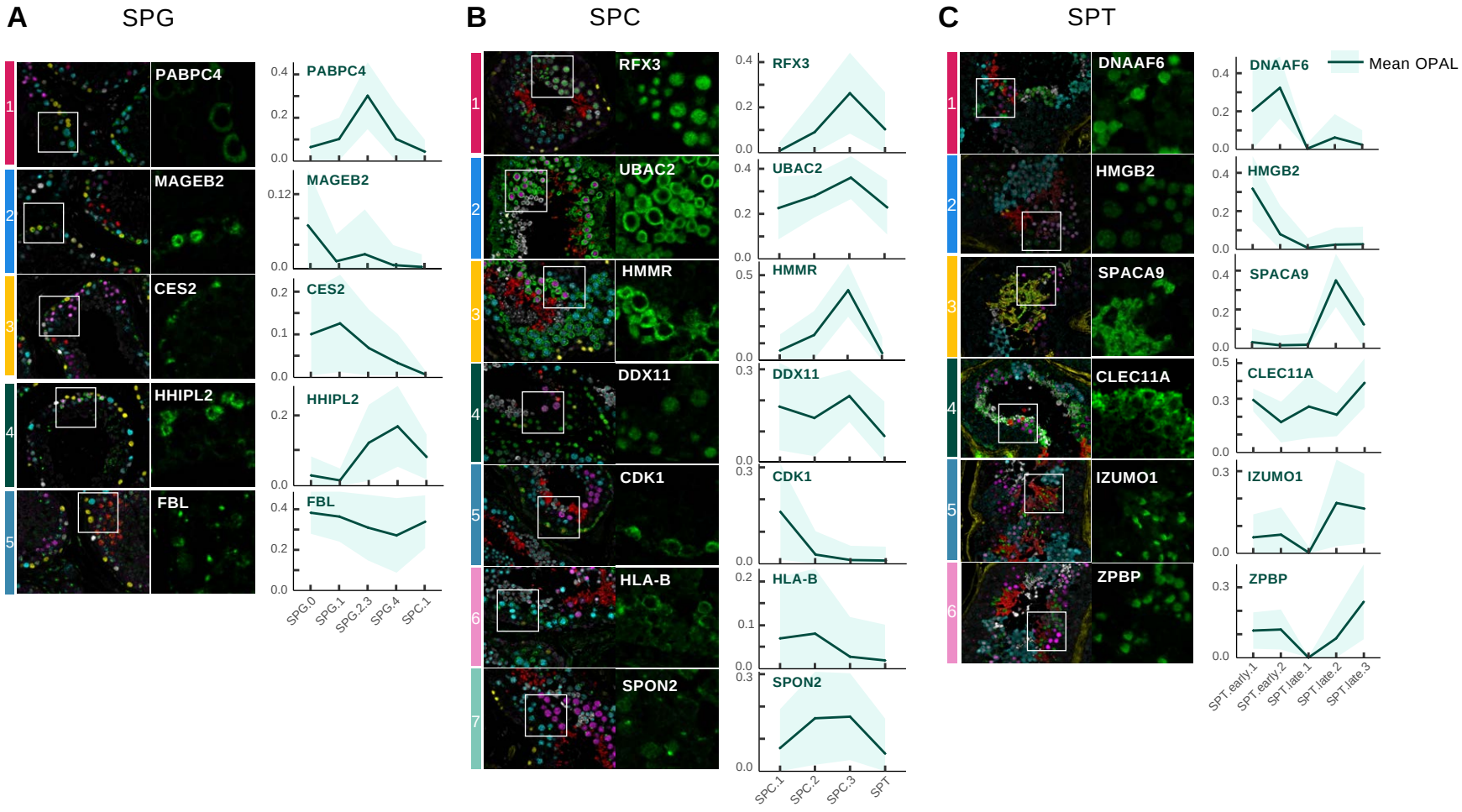

B
IHC intensity
3
2
1
0
Category
SPG-specific (613)
SPG & SPC-specific (428)
SPC-specific (171)
SPC & Early-SPT-specific (186)
SPC & SPT-specific (250)
Early SPT-specific (88)
Late SPT-specific (341)
SPT-specific (174)
All germ cells (587)
Mixed (914)
Category
SPG
SPC
Early.SPT
Late.SPT


D
SPG.0
SPG.1
SPG.2.3
SPG.4
SPC.1
SPC.1
SPC.2
SPC.3
SPT
SPT.early.1
SPT.early.2
SPT.late.1
SPT.late.2
SPT.late.3
UMAP_2
UMAP_1

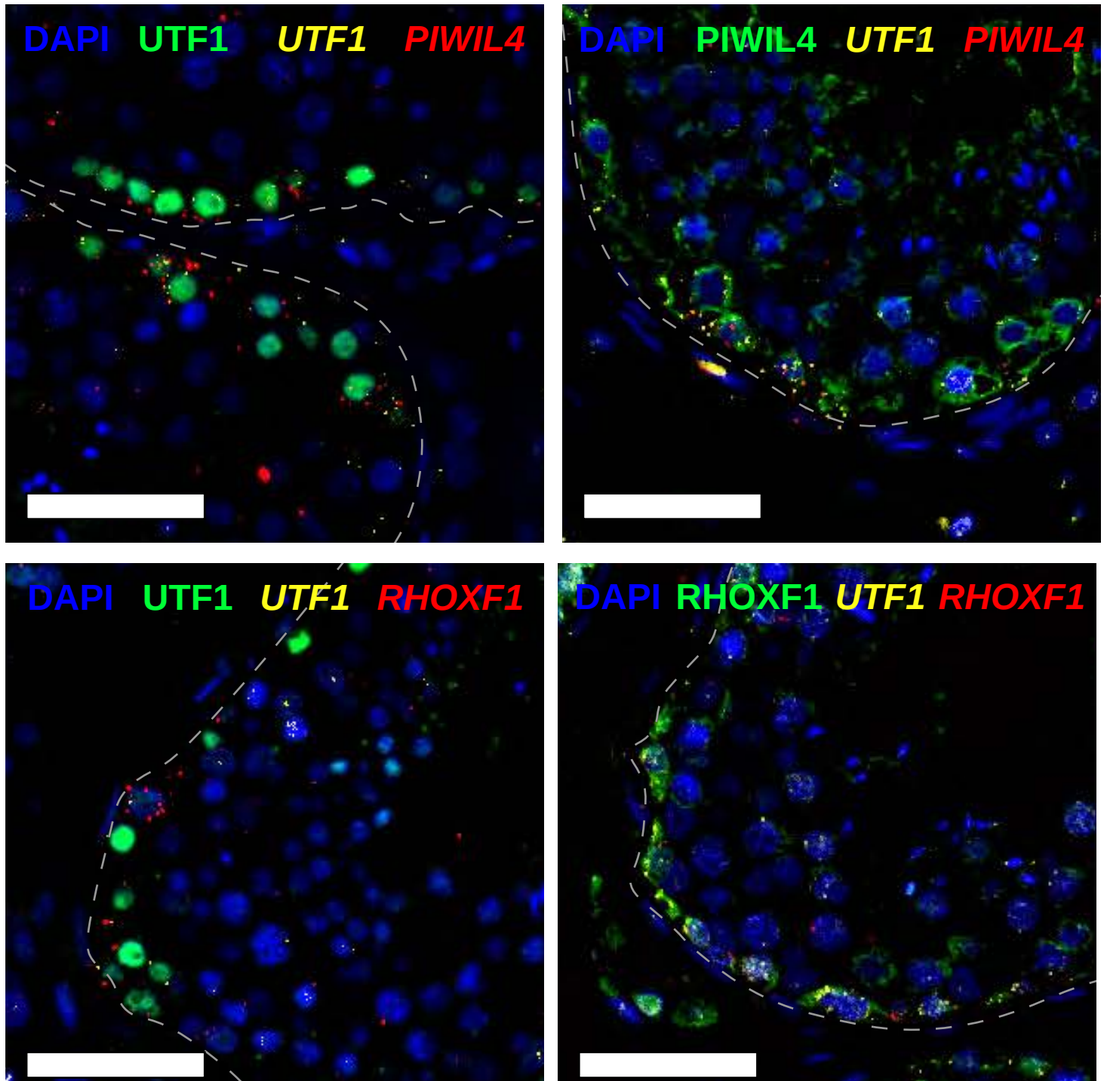

E
SPT.late.3
SPT.late.2
SPT.late.1
SPT.early.2
SPT.early.1
SPC.3
SPC.2
SPC.1
SPG.4
SPG.2.3
SPG.1
SPG.0
UTF1
IRF2BPL
DMRT1
CTCFL
BEND2
HELLS
SCML1
TCFL5
SUN5
PRM1
LYAR
OLAH
C3
SPATA24
TPPP2
Average expression
2.5
2.0
1.5
1.0
0.5
0.0
Percentage expression
25
50
75
100


C
Pearson's correlation
1
0.5
0
Late.SPT
SPG
SPC
Early.SPT
SPG
SPC
Early.SPT
Late.SPT

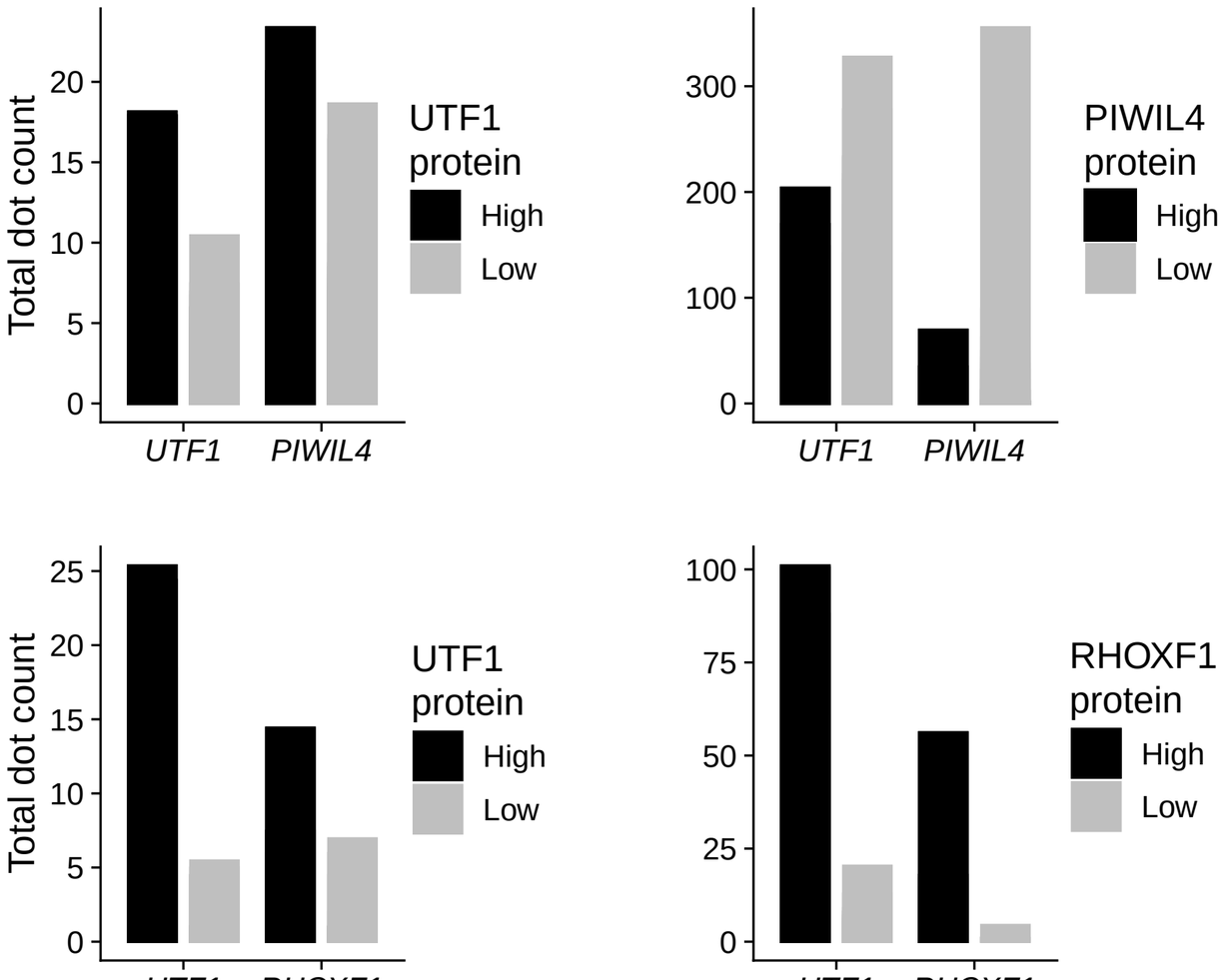

A
B
Mean OPAL
Mean RNA
SPG
Highest Mean RNA Expression
PIWIL4
RHOXF1
Peak average RNA
0.6
0.4
0.2
-1.0
-0.5
0.0
0.5
1.0
Concordant SPG
RHOXF1
Mean Expression Level
SPG.0
SPG.1
SPG.2.3
SPG.4
SPC.1
Discordant SPG
PIWIL4
SPC
WNK1
RFX2
Concordant SPC
SPC.1
SPC.2
SPC.3
SPT
Discordant SPC
SPT
ACTRT3
ACSBG2
0.8
Concordant SPT
SPT.early.1
SPT.early.2
SPT.late.1
SPT.late.2
SPT.late.3
Discordant SPT

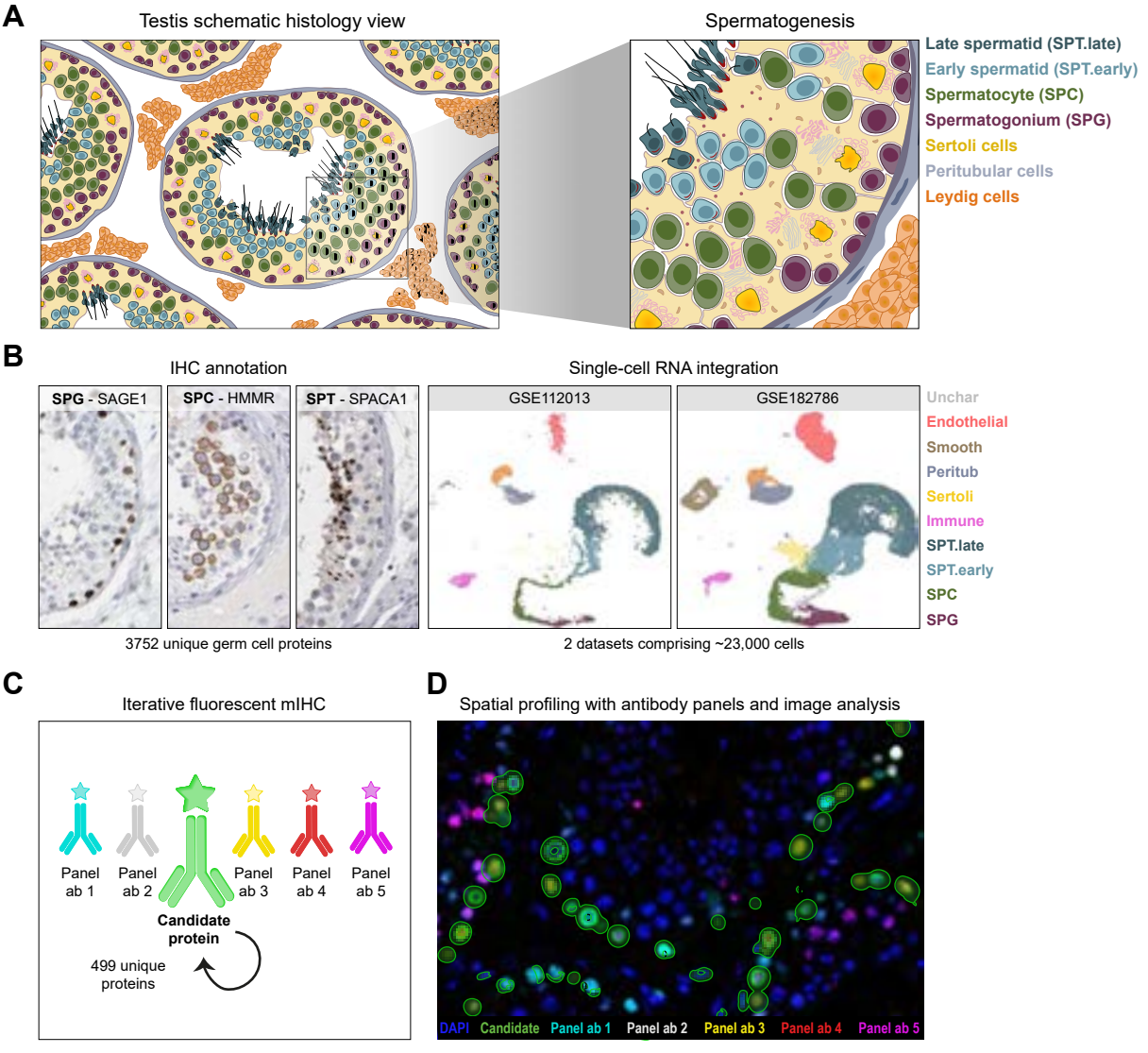

A
SPG
B
SPC
C
SPT
Mean OPAL
PABPC4
MAGEB2
CES2
HHIPL2
FBL
RFX3
UBAC2
HMMR
DDX11
CDK1
HLA-B
SPON2
DNAAF6
HMGB2
SPACA9
CLEC11A
IZUMO1
ZPBP
SPG.0
SPG.1
SPG.2.3
SPG.4
SPC.1
SPC.2
SPC.3
SPT
SPT.early.1
SPT.early.2
SPT.late.1
SPT.late.2
SPT.late.3

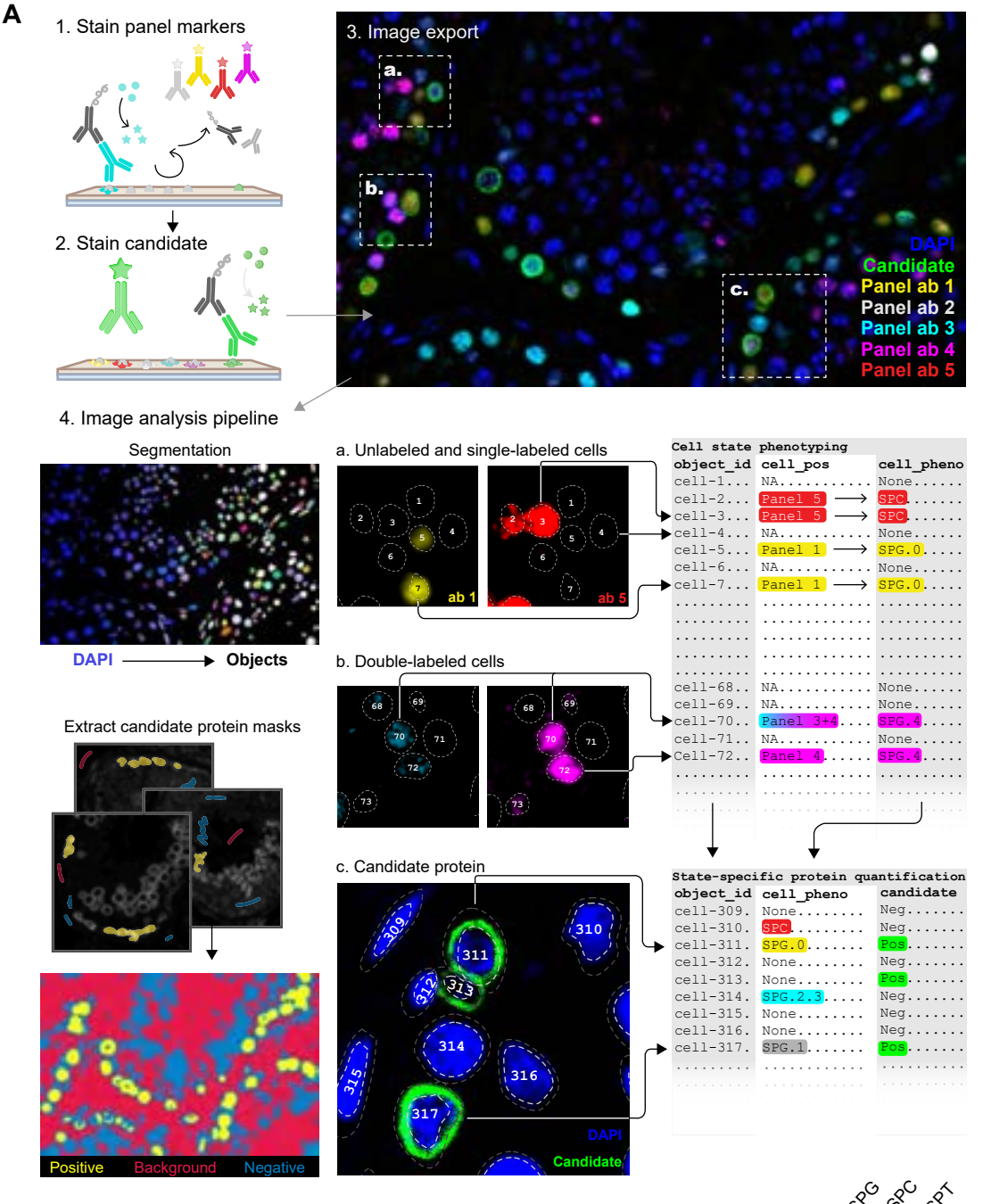
A


DAPI UTF1 UTF1 PIWIL4
DAPI PIWIL4 UTF1 PIWIL4
DAPI UTF1 UTF1 RHOXF1
DAPI RHOXF1 UTF1 RHOXF1

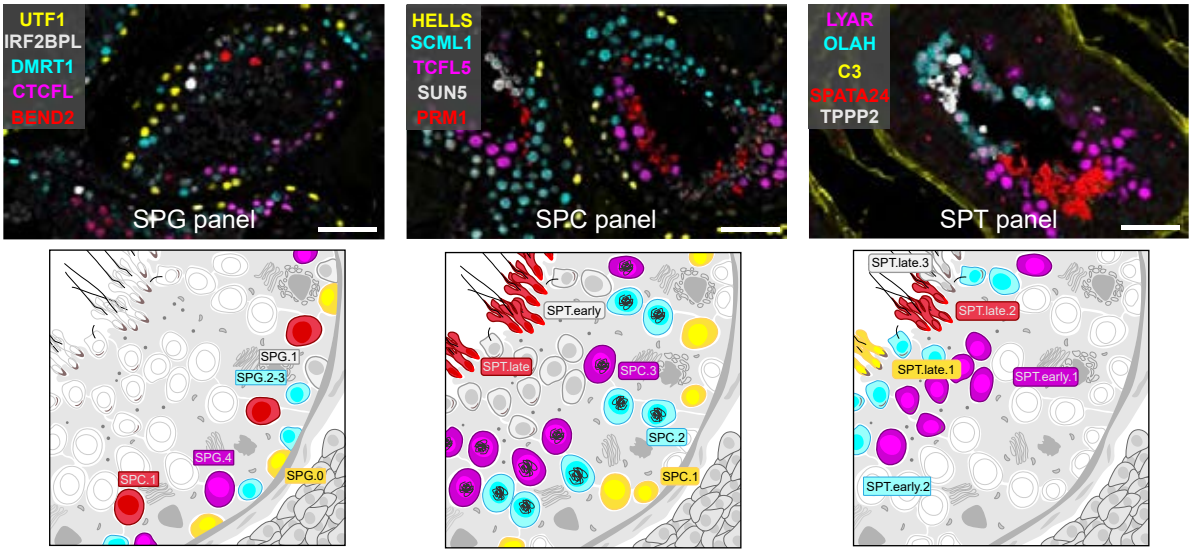
B


Total dot count
0
5
10
15
20
UTF1
PIWIL4
UTF1 protein
High
Low
0
100
200
300
UTF1
PIWIL4
PIWIL4 protein
High
Low
0
5
10
15
20
25
UTF1
RHOXF1
UTF1 protein
High
Low
0
25
50
75
100
UTF1
RHOXF1
RHOXF1 protein
High
Low

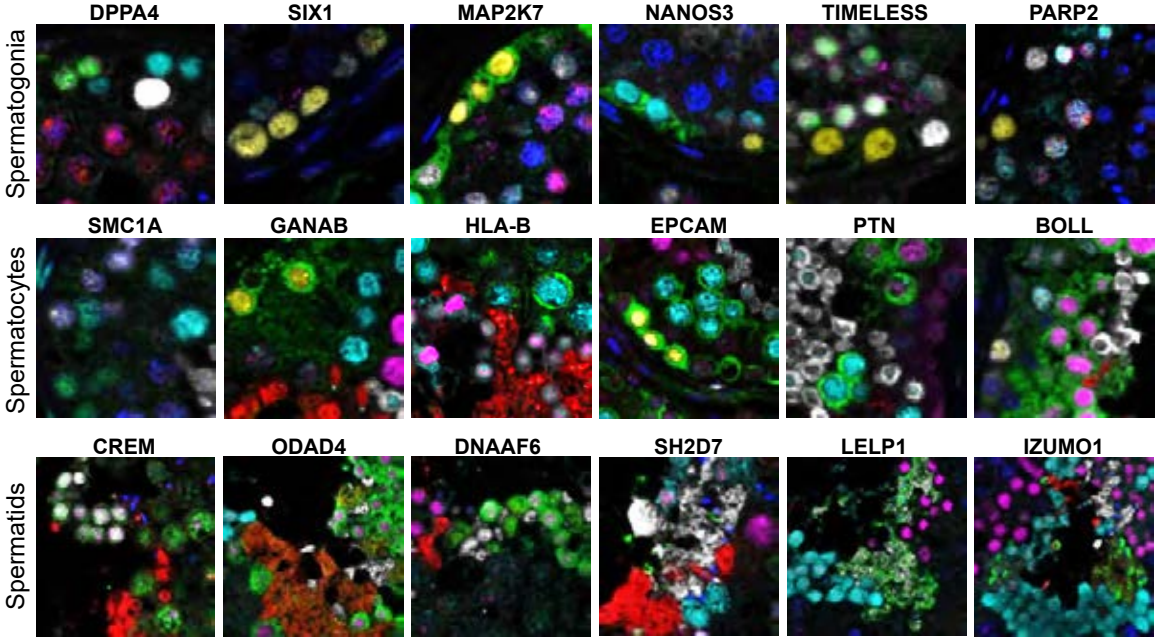

A
Testis schematic histology view
Spermatogenesis
Late spermatid (SPT.late)
Early spermatid (SPT.early)
Spermatocyte (SPC)
Spermatogonium (SPG)
Sertoli cells
Peritubular cells
Leydig cells
B
IHC annotation
Single-cell RNA integration
SPG - SAGE1
SPC - HMMR
SPT - SPACA1
GSE112013
GSE182786
Unchar
Endothelial
Smooth
Peritub
Sertoli
Immune
SPT.late
SPT.early
SPC
SPG
3752 unique germ cell proteins
2 datasets comprising ~23,000 cells
C
Iterative fluorescent mIHC
Panel ab 1
Panel ab 2
Panel ab 3
Panel ab 4
Panel ab 5
Candidate protein
499 unique proteins
D
Spatial profiling with antibody panels and image analysis
DAPI
Candidate
Panel ab 1
Panel ab 2
Panel ab 3
Panel ab 4
Panel ab 5
E
Single-cell quantification
Cell 1
Cell 2
Cell 3
Cell 4
Cell 5
Clustering
C1
C2
Enrichment analysis
C1
C2
Expression dynamics
Protein
RNA
State 1
State 2
State 3
State 4


A
1. Stain panel markers
2. Stain candidate
3. Image export
a.
b.
c.
DAPI
Candidate
Panel ab 1
Panel ab 2
Panel ab 3
Panel ab 4
Panel ab 5
4. Image analysis pipeline
Segmentation
DAPI
Objects
Extract candidate protein masks
Positive
Background
Negative
a. Unlabeled and single-labeled cells
ab 1
ab 5
b. Double-labeled cells
c. Candidate protein
DAPI
Candidate
Cell state phenotyping
object_id cell_pos cell_pheno
cell-1... NA..... None.....
cell-2... Panel 5 → SPC.....
cell-3... Panel 5 → SPC.....
cell-4... NA..... None.....
cell-5... Panel 1 → SPG.0.....
cell-6... NA..... None.....
cell-7... Panel 1 → SPG.0.....
cell-68.. NA..... None.....
cell-69.. NA..... None.....
cell-70.. Panel 3+4.... SPG.4.....
cell-71.. NA..... None.....
Cell-72.. Panel 4..... SPG.4.....
State-specific protein quantification
object_id cell_pheno candidate
cell-309. None..... Neg.....
cell-310. SPC..... Neg.....
cell-311. SPG.0..... Pos.....
cell-312. None..... Neg.....
cell-313. None..... Pos.....
cell-314. SPG.2.3..... Neg.....
cell-315. None..... Neg.....
cell-316. None..... Neg.....
cell-317. SPG.1..... Pos.....
B
Candidate protein mask
Cell phenotye mask
Cell segmentation mask
7-color IHC multi-page tif
C
SPG
SPG.0 16.4%
SPG.1 23.1%
SPG.2-3 24.8%
SPG.4 23.4%
SPC.1 12.3%
SPC
SPC.1 29.2%
SPC.2 32.0%
SPC.3 7.1%
SPT.early 13.5%
SPT.late 18.3%
SPT
SPT.early.1 6.9%
SPT.early.2 20.4%
SPT.late.1 60.2%
SPT.late.2 9.0%
SPT.late.3 3.4%
SPG
SPC
SPT
Distribution of cells per panel

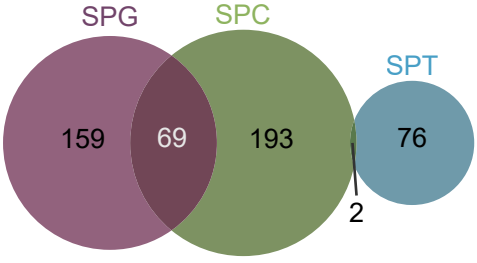
A


UTF1
IRF2BPL
DMRT1
CTCFL
BEND2
SPG panel
HELLS
SCML1
TCFL5
SUN5
PRM1
SPC panel
LYAR
OLAH
C3
SPATA24
TPPP2
SPT panel
SPG.1
SPG.2-3
SPG.4
SPC.1
SPG.0
SPT.early
SPT.late
SPC.3
SPC.2
SPC.1
SPT.late.3
SPT.late.2
SPT.late.1
SPT.early.1
SPT.early.2

B


DPPA4
SIX1
MAP2K7
NANOS3
TIMELESS
PARP2
Spermatogonia
SMC1A
GANAB
HLA-B
EPCAM
PTN
BOLL
Spermatocytes
CREM
ODAD4
DNAAF6
SH2D7
LELP1
IZUMO1
Spermatids

C


SPG
SPC
SPT
159
69
193
76
2