

Cell Type-Specific Expression of Testis Elevated Genes Based on Transcriptomics and Antibody-Based Proteomics

Charles Pineau, Feria Hikmet, Cheng Zhang, Per Oksvold, Shuqi Chen, Linn Fagerberg, Mathias Uhlén, Cecilia Lindskog

▶ **To cite this version:**

Charles Pineau, Feria Hikmet, Cheng Zhang, Per Oksvold, Shuqi Chen, et al.. Cell Type-Specific Expression of Testis Elevated Genes Based on Transcriptomics and Antibody-Based Proteomics. Journal of Proteome Research, 2019, 18 (12), pp.4215-4230. 10.1021/acs.jproteome.9b00351 . hal-02280776

HAL Id: hal-02280776

https://univ-rennes.hal.science/hal-02280776

Submitted on 28 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cell type-specific expression of testis elevated genes based on transcriptomics and antibody-based proteomics

Charles Pineau[1,2], Feria Hikmet[3], Cheng Zhang[4], Per Oksvold[4], Shuqi Chen[4], Linn Fagerberg[4], Mathias Uhlén[4], Cecilia Lindskog[3*]

1 Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) - UMR_S 1085, 35042 Rennes cedex, France

2 Protim, Univ Rennes, 35042 Rennes cedex, France

3 Uppsala University, Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, 75185 Uppsala, Sweden

4 Science for Life Laboratory, School of engineering Sciences in Chemistry, Biotechnology and health, KTH - Royal Institute of Technology, 17121 Stockholm, Sweden

*Correspondence to:
Cecilia Lindskog
Department of Immunology, Genetics and Pathology
Rudbeck Laboratory
Uppsala University
751 85 Uppsala, Sweden.
E-mail: cecilia.lindskog@igp.uu.se

Accepted manuscript

ABSTRACT

One of the most complex organs in the human body is the testis, where spermatogenesis takes place. This physiological process involves thousands of genes and proteins that are activated and repressed, making testis the organ with the highest number of tissue-specific genes. However, the function of a large proportion of the corresponding proteins remains unknown and testis harbors many missing proteins (MPs), defined as products of protein-coding genes that lack experimental mass spectrometry evidence. Here, an integrated omics approach was used for exploring the cell type-specific protein expression of genes with an elevated expression in testis. By combining genome-wide transcriptomics analysis with immunohistochemistry, more than 500 proteins with distinct testicular protein expression patterns were identified, and these were selected for in-depth characterization of their *in situ* expression in eight different testicular cell types. The cell type-specific protein expression patterns allowed us to identify six distinct clusters of expression at different stages of spermatogenesis. The analysis highlighted numerous poorly characterized proteins in each of these clusters whose expression overlapped with that of known proteins involved in spermatogenesis, including 88 proteins with an unknown function and 60 proteins that previously have been classified as MPs. Furthermore, we were able to characterize the *in situ* distribution of several proteins that previously lacked spatial information and cell type-specific expression within the testis. The testis elevated expression levels both at the RNA and protein level suggest that these proteins are related to testis-specific functions. In summary, the study demonstrates the power of combining genome-wide transcriptomics analysis with antibody-based protein profiling to explore the cell type-specific expression of both well-known proteins and MPs. The analyzed proteins constitute important targets for further testis-specific research in male reproductive disorders.

Key words: testis, reproduction, spermatogenesis, antibody-based proteomics, missing proteins, protein evidence, immunohistochemistry, transcriptomics

Accepted manuscript

1
2
3
4
5
6
7

Accepted manuscript

INTRODUCTION

Testis is one of the most complex organs in the mammalian body. It is divided into two compartments: the seminiferous tubules and the interstitium. Seminiferous tubules are the site of spermatogenesis and account for 60 to 80% of total testis volume in mammals. The seminiferous tubules consist of germ cells at different stages of their development, interspersed with Sertoli cells that constitute the seminiferous epithelium. Sertoli cells are known to have nursing properties and to transduce hormonal signals, thus playing a central role in the control of spermatogenesis. The seminiferous tubules are bounded by a basal membrane containing peritubular myoid cells that provide structural integrity and take part in regulation of spermatogenesis and testicular function. The interstitium, interspersed between the seminiferous tubules, is a connective tissue matrix that contains blood and lymph vessels, Leydig cells, macrophages and occasionally other cells (i.e., fibroblasts, lymphocytes and mast cells). The primary function of Leydig cells is the production of androgens, e.g. testosterone which is crucial in development of male reproductive tissues, in promotion of secondary sexual characteristics and for supporting spermatogenesis. Spermatogenesis is classically divided into three phases: 1) a proliferative mitotic phase in which primitive germ cells, or spermatogonia, replenish their stock and undergo a series of mitotic divisions; 2) a meiotic phase during which genetic recombination occurs in primary diploid spermatocytes (preleptotene spermatocytes), which divide twice to produce haploid spermatids; and 3) the final phase called spermiogenesis, in which spermatids morphologically transform into spermatozoa. Spermatogenesis lasts 74 days in humans, controlled by juxtacrine, paracrine and endocrine factors and conditioned by the successive activation and/or repression of thousands of genes and proteins [1]

Previous studies based on deep RNA sequencing (RNA-Seq) have shown that testis has by far the highest number of tissue-specific genes [2-4], with more than 2,000 genes showing some degree

Accepted manuscript

of elevated expression compared to other organs. One could speculate that this considerably high number of testis elevated genes is due to the specialized functions and events occurring during spermatogenesis and sperm maturation. The function and cell type-specific expression of the corresponding proteins elevated in testis are however largely unknown, and although the different steps of sperm maturation are well defined, all genes involved in the process are yet to be characterized. “Missing proteins” (MPs) are defined as proteins that lack sufficient experimental evidence of existence at the protein level. The Human Proteome Project (HPP) [5-6] is an initiative aiming to define the human proteome. Together with its primary knowledgebase neXtProt [7], HPP has set up criteria ranking proteins in a 1-5 tier system according to evidence of protein existence (PE), suggesting that approximately 11% of the human proteins are considered to be MPs [8]. In the quest for identifying MPs, testis has been suggested as one of the organs harboring a significant number of such proteins [9-11] and most specifically the post-meiotic germ cell lineage including mature spermatozoa [10-12].

Antibody-based proteomics provides an advantageous strategy for studying the spatial distribution of proteins. The Human Protein Atlas project [4, 13] is a large-scale project with the aim to map the entire human proteome using an integration of various omics technologies, including immunohistochemistry and RNA-Seq. The publicly available database (www.proteinatlas.org) contains more than 10 million high-resolution images showing the *in situ* localization of proteins in a large set of human organs, tissues, cells and organelles, including samples of normal adult testis.

In the present investigation, 512 genes with elevated expression in testis were studied using publicly available immunohistochemistry images on the Human Protein Atlas. Based on manual annotation of the cell type-specific expression pattern in eight different testis cell types, a detailed map of proteins expressed at various stages of spermatogenesis and somatic cells was

Accepted manuscript

generated. Previously well-known protein expression patterns related to the spermatogenic process could be confirmed, and six distinct clusters of cell type-specific expression in different stages of spermatogenesis were identified. The analysis highlighted numerous poorly characterized proteins in each of these clusters that overlapped in expression with known proteins involved in spermatogenesis. Furthermore, we characterized several proteins with evidence at the protein level that previously lacked spatial information and cell type-specific expression in testis.

Our data provides an important contribution to the understanding of normal and pathological spermatogenesis, important for testis-based research in health and diseases.

Accepted manuscript

EXPERIMENTAL PROCEDURES

Tissue preparation

Tissue microarrays (TMAs) comprising of formalin-fixed paraffin-embedded 1 mm diameter tissue cores from 44 human tissues with three individuals per tissue, including normal human testis, were generated as described previously [14]. Samples were obtained from the Department of Pathology, Uppsala University Hospital, Uppsala, Sweden, as part of the sample collection governed by the Uppsala Biobank (http://www.uppsalabiobank.uu.se/en/). All human tissue samples used in the present study were anonymized in accordance with approval and advisory report from the Uppsala Ethical Review Board (Reference # 2002-577, 2005-338 and 2007-159 (protein) and # 2011-473 (RNA)). For RNA-Seq analysis, 37 different human tissue types, including seven individual testis samples were used, described in detail in [2].

Protein profiling

TMA blocks were cut in 4 µm thick sections using waterfall microtomes (Microm HM 355S, Thermo Fisher Scientific, Freemont, CA, USA), dried in room temperature (RT) overnight and baked at 50°C for 12-24 hours prior to immunohistochemical staining. Automated immunohistochemistry was performed using Autostainer 480® instruments (Lab Vision, Freemont, CA, USA), essentially as previously described [14]. The dilution of each antibody was optimized based on strict validation criteria taking into consideration the expected staining pattern based on available literature, as well as correlation between RNA expression levels and protein expression levels, as described in the "Selection of candidates" section. Optimization was performed on a test TMA slide containing 20 different normal tissue types, including testis, and an optimal dilution was chosen that generated a distinct immunohistochemical staining in the organs where highest expression was expected, with as little background staining in other organs as possible. High-resolution digital images were obtained by slide scanning using Scanscope AT2

Accepted manuscript

(Aperio, Vista, CA, USA), with a 20x objective. The images of immunohistochemically stained TMA sections were separated into individual cores, and uploaded to the Human Protein Atlas version 18 (v18.proteinatlas.org).

Transcript profiling (RNA-Seq)

Tissues samples were embedded in Optimal Cutting Temperature (O.C.T.) compound and stored at –80°C. An HE stained frozen section (4 µm) was prepared from each sample using a cryostat and the CryoJane® Tape-Transfer System (Instrumedics, St. Louis, MO, USA). Each slide was examined by a pathologist to ensure sampling of representative normal tissue. Three sections (10 µm) were cut from each frozen tissue block and collected into a tube for subsequent RNA extraction. The tissue was homogenized mechanically using a 3 mm steel grinding ball (VWR). Total RNA was extracted from cell lines and tissue samples using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The extracted RNA samples were analyzed using either an Experion automated electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA) with the standard-sensitivity RNA chip or an Agilent 2100 Bioanalyzer system (Agilent Biotechnologies, Palo Alto, USA) with the RNA 6000 Nano Labchip Kit. Only samples of high-quality RNA (RNA Integrity Number ≥7.5) were used in the following mRNA sample preparation for sequencing.

Processing of RNA-Seq data

The RNA expression data is based on 172 tissue samples from 37 different tissues, corresponding to the RNA-Seq data available in version 18 of the Human Protein Atlas (v18.proteinatlas.org). The mRNA sequencing (RNA-Seq) was performed on Illumina HiSeq2000 and 2500 machines (Illumina, San Diego, CA, USA) using the standard Illumina RNA-Seq protocol with a read length of 2x100 bases. Transcript abundance estimation was performed using Kallisto v0.42.4. For each gene, the abundance in 'Transcript Per Million' (TPM) is reported as the

Accepted manuscript

sum of the TPM values of all its protein-coding transcripts. For each tissue type, the average TPM value for replicate samples was used as abundance score. The threshold level to detect presence of a transcript for a particular gene was set to ≥ 1 TPM. The RNA expression levels from all protein-coding genes were used to classify all genes according to their tissue-specific expression into one of six different categories, defined based on the total set of all TPM values in 37 tissues, as described previously [4]: Tissue enriched (expression in one tissue at least five-fold higher than all other tissues), Group enriched (five-fold higher average TPM in a group of two to seven tissues compared to all other tissues), Tissue enhanced (five-fold higher average TPM in one or more tissues compared to the mean TPM of all tissues), Expressed in all (≥ 1 TPM in all tissues), Not detected (< 1 TPM in all tissues) or Mixed (detected in at least one tissue and in none of the above categories). The present investigation focused on the 2,237 genes defined as elevated in testis (Tissue enriched, Group enriched or Tissue enhanced).

Selection of candidates

All 2,237 genes defined all as elevated in testis were manually evaluated with regard to antibody availability and reliability, as well as protein expression pattern in testis. Of the 2,237 genes, 1,483 genes were analyzed with immunohistochemistry in the Human Protein Atlas using at least one antibody. Antibody validation was performed based on “pillars” suggested by the International Working Group for Antibody Validation (IWGAV) [15], implemented by the Human Protein Atlas as “Enhanced validation” strategies (https://www.proteinatlas.org/about/antibody+validation). For immunohistochemistry, the main criterion for antibody validation is comparison of protein expression levels with an orthogonal/antibody-independent method analyzing the expression levels of the same target. Here, the orthogonal analysis is based on manual comparison between RNA expression and protein expression levels across 37 tissues, based on a 4-tier RNA similarity (RNAS) score; RNAS1: RNA expression levels are consistent with protein expression levels; RNAS2: RNA expression levels are mainly consistent with protein expression levels; RNAS3:

Accepted manuscript

RNA expression levels are mainly not consistent with protein expression levels; RNAS4: RNA expression levels are not consistent with protein expression levels. Only genes with at least one antibody scored as RNAS1-2 were included. In addition to antibody validation, an inclusion criteria was that the staining pattern in testis should be of at least moderate intensity in at least one of the testis cell types. In total, 512 testis elevated genes with publicly available protein expression data in version 18 of the Human Protein Atlas (v18.proteinatlas.org) were selected for annotation (Supplementary Table 1). Forty-five of these genes were in addition to orthogonal validation also analyzed with independent antibodies, where similar protein expression patterns were observed using additional antibodies targeting non-overlapping epitopes of the same proteins. In cases where more than one antibody towards the same protein was used, the antibody with the most distinct staining pattern was selected. Due to a high degree of overlap between protein sequences for testis elevated genes, 95 of the antibodies were considered to be "multi-targeting", meaning that they, in addition to the intended target, also bind to at least on more protein due to amino acid sequence similarity of more than 80%. When selecting these targets for inclusion, all proteins targeted were taken into consideration when the antibodies were validated. Only antibodies where protein expression levels generated correlated with the literature available for all the targeted proteins as well as the RNA expression values for all the corresponding genes were included.

Antibodies

The IDs of all antibodies used in the immunohistochemically stained images are summarized here. All antibodies are from Atlas Antibodies AB, unless otherwise stated. Proteins specific to premeiotic germ cells: TIPIN (HPA058799, 1:550), SAGE1 (HPA003208, 1:250), MAGEC1 (HPA004622, 1:4000), OOSP2 (HPA050914, 1:200), ERICH6 (HPA048373, 1:275), DCAF4L1 (HUPO 7 anti-DCAF4L1, Monash University MATF, 1:7000), PKDREJ (HPA034587, 1:2700), C3orf22 (HPA059199, 1:1000). Proteins specific to premeiotic and meiotic cells: PNMA5

Accepted manuscript

(HPA044690, 1:500), DAZ1-4/DAZL (HPA047113, 1:10), CTAG1A (35-6200, Thermo Fisher Scientific, 1:100), CT45A1-3, 5-10 (HPA044735, 1:2000). Proteins specific to meiotic cells: DMC1 (HPA001232, 1:25), HMMR (NCL-CD168, Leica Biosystems, 1:50), SGO1 (HPA069857, 1:10), BEND2 (HPA013142, 1:150), SPATC1L (HPA019165, 1:50), SEL1L2 (HPA059121, 1:170). Proteins specific to meiotic and postmeiotic germ cells: BOLL (HPA048813, 1:3200), RFX3 (HPA035689, 1:600), AXDND1 (HPA071114, 1:3000), LRRC23 (HPA057533, 1:150), C16orf71 (HPA049468, 1:10000), C17orf64 (HPA044415, 1:750), FAM181A (HPA001603, 1:100), C9orf50 (HPA026743, 1:500). Proteins specific to postmeiotic germ cells: ZPBP (HPA058673, 1:75), TMEM210 (HPA066907, 1:10000), C7orf61 (HPA060395, 1:950), FAM209B (HPA042144, 1:10), ACTRT1 (HPA003119, 1:75), AKAP4 (HPA005949, 1:1200), C11orf42 (HPA063404, 1:175), FHAD1 (HPA054284, 1:25), SPATA3 (HPA018254, 1:150), C3orf30 (HPA062468, 1:10000). Protein expression in testis, Fallopian tube and airway epithelia: HYDIN (HPA067155, 1:3000), CFAP45 (HPA043618, 1:500), EFCAB1 (HPA023527, 1:100), NEK5 (HPA035565, 1:400), RSPH10B/RSPH10B2 (HPA049182, 1:250).

Analysis of protein expression

High-resolution digital images of immunohistochemically stained testis tissue sections, publicly available on the Human Protein Atlas version 18 (Ensembl version 88.38), were manually annotated by an expert in testis histology (CP). The annotation was performed on transverse sections of adult human testis at stages IV to VI of the seminiferous epithelium using a tool built upon the existing Human Protein Atlas website, taking advantage of the gene, protein, antibody, tissue and image-to-annotation structures and image viewer already present. The publicly available images were accessed for scoring via the browser using a combination of client side user scripts together with the browser plugin Tampermonkey (http://tampermonkey.net/). Annotation HTML forms were added on top of the detailed testis image pages for each gene included in the study, and a Google spreadsheet was used as backend data storage facilitated

Accepted manuscript

by using the Google Sheet API. Annotation parameters included antibody staining intensity (high, medium, low, not detected), quantity of stained cells (>75%, 25-75%, 1-24%, <1%) and subcellular localization (nuclear, cytoplasmic, membranous) in eight different testicular cell types (spermatogonia, preleptotene spermatocytes, pachytene spermatocytes, round/early spermatids, elongated/late spermatids, Sertoli cells, Leydig cells, peritubular cells). An independent second observer (FHN or CL) quality controlled all finished annotations.

Data analysis and visualization

For the PCA plot and correlation heatmap, the values for staining intensity and quantity of stained cells were multiplied, generating a protein score between values 0 and 9. All data were analyzed with R Statistical Environment [16] with the addition of the package 'marray' [17] and "gplots" [18]. The network plot was generated based on simultaneous expression in different testis cell types. For a cell type to be considered positive, staining of medium intensity in at least 25% of the cells should be observed, generating a protein score of at least 4. The network plot was visualized by Cytoscape (version 3.6.0). Nodes representing cell types are colored in black. Genes expressed in defined combinations of cell types are displayed in different colors, and the number of genes in each group are shown as the node label. All gene groups with only one gene or shared with more than four cell types are removed from the plot since they are less relevant for cell type specificity and complicate the plot.

Circos plot showing the enriched GO terms for each expression group were generated based on information retrieved from DAVID EASE (https://david.ncifcrf.gov/ease/update/EASE_Files.jsp) where all GO terms and the genes included for each term could be downloaded. Hypergeometric test was used for testing the enrichment of gene groups in each GO term with a P value cut off of 0.05 using the union of unique genes in all GO biological process terms as reference list. The

Accepted manuscript

enriched GO terms with top three lowest P values were selected for visualization in the circos plot. The circos plot was generated using R package 'circlize' [19].

Data availability

TPM values for the testis samples are available for download without any restrictions (v18.proteinatlas.org/about/download). The primary data (reads) are available through the Array Express Archive (www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-2836. Protein profiling data for each protein in each testis cell type, as well as the original immunohistochemically stained images are also available on the Human Protein Atlas (www.proteinatlas.org).

Accepted manuscript

RESULTS

Genome-wide transcriptomics analysis

Transcriptomic analysis of 37 different human organs and tissues based on altogether 172 samples has previously been performed using RNA-Seq, and the data is available on the Human Protein Atlas (v18.proteinatlas.org). The analysis included testicular tissue from seven individuals. Normalized mRNA levels were determined for each sample, calculated as transcript per million (TPM) values. All genes were classified into categories based on their expression levels in testis, as described previously [2, 13]. A total of 2,237 genes were identified as elevated in testis compared with other organs. The testis elevated genes were further subdivided into 1,079 testis enriched genes (at least five-fold higher mRNA level in testis compared to other tissues), 364 group enriched genes (at least five-fold higher mRNA level in a group of 2-7 tissues, including testis), and 794 testis enhanced genes (at least five-fold higher mRNA level in testis compared to mean of other tissues). Of the 2,237 genes elevated in testis, antibody-based protein expression data in the Human Protein Atlas was available for 1,483 of these genes. Based on antibody validation and protein expression pattern in testis, one representative antibody for each of 512 genes was selected for in-depth characterization in testis (Supplementary Table 1). These 512 genes included 249 testis enriched genes, 92 group enriched genes and 171 testis enhanced genes.

Of the 512 proteins included in the study, 510 proteins have previously been assigned a PE score [7, 20], and the vast majority of the proteins (n=360) are validated (PE1). An additional 88 proteins are defined as validated but with an unknown function (uPE1). Furthermore, 60 proteins are classified as MPs (PE2-4), where PE2 have evidence at transcript level, while PE3-4 lack human experimental data. Two proteins (ANKRD31 and NPIPB2) are defined as PE5, which describes dubious or uncertain genes. According to the Gene Ontology database, 285 of the 512 testis elevated genes included in the present investigation are not associated with any biological

Accepted manuscript

process whereas 273 are not associated with any molecular process (Supplementary Table 1). Nevertheless, the testis elevated expression levels based on RNA-Seq suggest that these proteins may be related to testis-specific functions, and analysis of their cell type-specific expression patterns is a first step towards further characterization of their function.

Categorization of testis elevated genes based on cell type-specific expression

The cell type-specific protein expression patterns in testis for the 512 proteins were manually annotated in eight different testicular cell types; spermatogonia, preleptotene spermatocytes, pachytene spermatocytes, round/early spermatids, elongated/late spermatids, Sertoli cells, Leydig cells and peritubular cells (Supplementary Table 1). The association between these eight cell types was determined using principal components analysis (PCA) based on the expression pattern of 512 proteins (Figure 1A). A clear relationship was observed between the interstitial/epithelial somatic cell types (Sertoli cells, Leydig cells and peritubular cells). There was also an evident separation between the different stages of spermatogenesis, with a relationship found between cell types before first meiosis (spermatogonia and preleptotene spermatocytes), and after second meiosis (round/early spermatids and elongated/late spermatids). Pachytene spermatocytes, cells at the third stage of Prophase I, were as expected located between the other two clusters (Figure 1A). Protein expression levels in the eight cell types visualized in a heatmap confirmed the relationships observed by the PCA plot (Figure 1B). In summary, the PCA plot and protein expression heatmap highlighted six expression groups: 1) premeiotic; 2) premeiotic + meiotic; 3) meiotic; 4) meiotic + postmeiotic; 5) postmeiotic; and 6) somatic, as outlined in Figure 1A. For each expression group, the enrichment of GO terms associated with biological processes was evaluated, and the top three GO terms with lowest P values (<0.05) for each group were visualized in a circos plot (Figure 1C). The circos plot confirms several well-known testis-specific functions related to the cell types included in each expression group, as discussed more in detail in the sections below.

Accepted manuscript

Relationship of protein expression between different testicular cell types

For further investigation of cell type-specific expression patterns, a network plot was generated showing the relationship of protein expression between different testicular cell types (Figure 2). The network plot illustrates the 378 of the 512 proteins that showed at least medium expression levels in a particular combination of cell types, allowing a combination of maximum three cell types per protein. Of the 378 proteins included in the plot, as many as 351 proteins (93%) were expressed in at least one germ cell type, while 27 proteins were expressed only in somatic cells (Figure 2). In accordance with the cluster distribution observed in Figure 1, the protein expression in different testicular cell types allowed us to divide the proteins into six main groups: 1) premeiotic germ cells (spermatogonia and/or preleptotene spermatocytes) (n=57; green nodes); 2) premeiotic and meiotic germ cells (spermatogonia and/or preleptotene spermatocytes, and pachytene spermatocytes) (n=26; purple nodes); 3) meiotic germ cells (pachytene spermatocytes) (n=14; yellow nodes); 4) meiotic and postmeiotic germ cells (pachytene spermatocytes, round/early spermatids and/or elongated/late spermatids (n=78; red nodes); 5) postmeiotic germ cells (round/early spermatids and/or elongated/late spermatids) (n=132; blue nodes); and 6) somatic cells (Sertoli cells, Leydig cells, and/or peritubular cells) (n=27; pink nodes). In the present investigation, main emphasis is put on further exploration of proteins expressed in germ cells (group 1-5), described more in detail in the sections below.

Proteins specific to premeiotic germ cells (cluster 1)

Premeiotic germ cells encompass spermatogonia and preleptotene spermatocytes. Previously, four states of spermatogonia have been defined [21]. Stem cell A spermatogonia are located at the base of the seminiferous tubules and actively self-renew by mitotic divisions to give rise to other type A spermatogonia or type A1 spermatogonia that will enter the meiotic process. The number of type A spermatogonia is controlled by apoptosis. Spermatogonia A1 will differentiate into

Accepted manuscript

intermediate spermatogonia, then type B spermatogonia. Finally, type B spermatogonia will evolve into preleptotene spermatocytes. During the preleptotene phase, cells are in the S phase and DNA is replicated before entering the meiotic process. Recently, a fifth type of spermatogonia called State 0 was discovered, considered to represent the earliest/naïve spermatogonial stem cell in adults [22]. As expected, the enriched GO terms of the 57 proteins included in the premeiotic group were associated with functions crucial for cell division and differentiation, such as chromatin remodeling, chromatin formation and metabolic processes (Figure 1C). All proteins in the premeiotic group have a determined PE status, out of which 35 proteins belong to the PE1 category, 11 proteins are uPE1 and 11 proteins are PE2-4.

Immunohistochemical staining patterns of proteins expressed specifically in premeiotic cells are displayed in Figure 3. The TIMELESS-interacting protein (TIPIN), associated with cell cycle, was expressed both in spermatogonia and preleptotene spermatocytes (Figure 3). TIPIN is a PE1 protein known to be expressed during the S phase of the cell cycle. It forms a complex with the TIMELESS protein and both are co-expressed in proliferating embryonic and adult tissues. The TIPIN-TIMELESS complex regulates the DNA replication process [23], thus playing an important role in the S phase checkpoint and replication fork stability in Metazoans [24]. TIPIN has never been described in the testis, however, its known function in other tissues and increasing levels of expression from spermatogonia to preleptotene spermatocytes in the human testis, strongly suggest that TIPIN could be a key player in the regulation of DNA replication during the premeiotic S phase in spermatogenesis.

Another interesting protein expressed in spermatogonia and preleptotene spermatocytes is the sarcoma antigen 1 (SAGE1) protein, also named cancer/testis antigen 14 (Figure 3). Like most cancer/testis antigens (CTAs), SAGE1 expression in normal tissues is restricted to the testis, while strong expression is observed in several carcinomas [25]. SAGE1 is a PE1 protein supposedly

Accepted manuscript

involved in forming the mature 3' end of an snRNA molecule. However, its exact function in the testis remains unknown.

Examples of uPE1 proteins specifically expressed in spermatogonia and preleptotene spermatocytes include melanoma-associated antigen C1 (MAGEC1), oocyte-secreted protein 2 (OOSP2), glutamate-rich protein 6 (ERICH6) and CUL4-associated factor 4-like protein 1 (DCAF4L1) (Figure 3). MAGEC1 is known to be a CTA, whereas no previous studies suggest neither the function nor testis-specific expression of OOSP2, ERICH6 or DCAF4L1. Interestingly, two MPs exclusively expressed in premeiotic germ cells with similar expression patterns are the polycystic kidney disease and receptor for egg jelly-related protein (PKDREJ) and the uncharacterized protein C3orf22 (C3orf22) (Figure 3). PKDREJ has been suggested to be involved in sperm transport and acrosome reaction in both rodents and primates [26]. The elevated expression in testis together with a lack of expression in all other normal tissues support the fact that PKDREJ is testis-specific. At the protein level, expression was observed in spermatogonia and preleptotene spermatocytes, but we could not confirm any expression in spermatids or acrosomal structures (Figure 3). No previous literature describes the expression pattern of C3orf22, but we can here prove that the protein is testis-specific, with an expression restricted to premeiotic cells. While the available literature on the uPE1 and missing proteins expressed in spermatogonia and preleptotene spermatocytes is limited, the expression pattern of MAGEC1, OOSP2, ERICH6, DCAF4L1, PKDREJ and C3orf22 restricted to premeiotic germ cells suggests that these proteins may be implicated in cell division or DNA replication during the early stages of spermatogenesis.

Proteins specific to premeiotic and meiotic germ cells (cluster 2)

Our results clearly show a premeiotic and meiotic expression cluster encompassing proteins whose expression is specific to spermatogonia and/or preleptotene spermatocytes, together with

Accepted manuscript

expression in pachytene spermatocytes (Figure 1). Enriched GO terms associated with the 26 proteins in this expression cluster include gamete generation and spermatogenesis (Figure 1C). Gamete generation refers to generation and maintenance of gametes (haploid reproductive cells) in a multicellular organism. Nineteen of the 26 proteins are PE1, one protein is uPE1, and six proteins are PE2-4.

Immunohistochemical staining patterns of proteins expressed specifically in premeiotic and meiotic cells are displayed in Figure 4. One example is paraneoplastic antigen-like protein 5 (PNMA5), expressed in spermatogonia, preleptotene spermatocytes and pachytene spermatocytes (Figure 4). Interestingly, pachytene spermatocytes at stage I of the seminiferous cycle were unstained, whereas pachytene spermatocytes at stage V displayed a very strong immunoreactivity, suggesting a rapid drop in protein expression in spermatocytes at the diplotene step. Despite being a PE1 protein, the testicular expression of PNMA5 has previously not been described, but recent studies suggest that the protein could be involved in meiotic progression in mouse oocytes [27]. PNMA5 also binds to another member of the PNMA family, MOAP1, and both proteins have been shown to promote apoptosis [28]. The expression profile of the PNMA5 protein in human testis from spermatogonia up to pachytene spermatocytes at stage V of the seminiferous cycle clearly suggests that PNMA5 could be related to meiosis progression or germ cells apoptosis.

A second example is the Deleted in Azoospermia-like protein (DAZL), and the closely related proteins DAZ1-4. Due to high sequence similarity, the antibody used targets all five proteins. The expression is clearly cytoplasmic with high levels observed in spermatogonia and preleptotene spermatocytes, and medium immunoreactivity in pachytene spermatocytes (Figure 4). This family of RNA-binding proteins was isolated from the AZF region on the human Y chromosome, and is frequently deleted in infertile men with non-obstructive azoospermia [29-30]. This region has two

Accepted manuscript

autosomal single-copy paralogues: DAZL and BOLL, that are exclusively expressed in germ cells and play a role in the regulation of mRNA translation. The expression of DAZL/DAZ1-4 has previously been suggested to be restricted to spermatogonia [31]. Here, we show that these proteins are predominantly expressed in primary spermatocytes. DAZL/DAZ1-4 are PE1 proteins, however, the role of the gene family in spermatogenesis remains unclear. Our data suggests that these proteins may play a role in mitosis in young germ cells up to a later phase in meiosis.

One example of an uPE1 protein expressed in both premeiotic and meiotic cells is the cancer/testis antigen 1A (CTAG1A). The expression restricted to spermatogonia, preleptotene spermatocytes and pachytene spermatocytes has been described previously, and CTAG1A has been suggested to be used as a marker for early progression of testicular tumorigenesis [32]. The exact function is however unknown. The protein family cancer/testis antigen family 45 member A consists of multiple proteins with overlapping sequences, where nine proteins are targeted using the same antibody (CT45A1-3, 5-10). Interestingly, seven out of these proteins are MPs. Here, we could confirm that the expression of these proteins was restricted to spermatogonia, preleptotene spermatocytes and pachytene spermatocytes.

Proteins specific to meiotic germ cells (cluster 3)

The third expression group in our large-scale annotation study highlights the pachytene spermatocyte-specific proteins. Pachytene spermatocytes are meiotic cells at the very end of prophase I, and easily recognizable as they are the only meiotic germ cells observed on testicular histology sections due to the long duration of meiotic prophase I. The pachytene stage is important in spermatogenesis as this is where the crossing over between homologous chromosomes begins. This is the reason why the pachytene stage is a key step in meiosis and the cells express several checkpoint proteins preventing abnormal meiotic recombination and chromosome synapsis [33]. GO analysis of the 14 proteins expressed in this cluster confirmed enrichment for

Accepted manuscript

processes related to chromosome segregation, cell division and cell cycle (Figure 1C). Ten of the proteins are PE1, three proteins are uPE1 and one protein is PE2.

Immunohistochemical staining patterns of proteins expressed specifically in meiotic cells are displayed in Figure 5. Our results show that one checkpoint protein, the DNA meiotic recombinase 1 (DMC1), is specific to human pachytene spermatocytes. DMC1 is a meiosis-specific protein and the homologue of RAD51, known to be involved in the homologous recombination and repair of the DNA, thus supporting the view that mitotic double-strand break repair is recruited for use in meiotic chromosome metabolism [34]. DMC1 interacts with the BRCA2 protein [35] and promotes homologous recombination [36]. In *Saccharomyces cerevisiae*, the mutation of DMC1 leads to a meiotic arrest at the pachytene stage [34]. Despite being a PE1 protein, its function in testis remains unknown. On the basis of this information, our results strongly suggest that DMC1 could be a key player in meiotic control during spermatogenesis. An uPE1 protein with an expression pattern similar to that of DMC1 is the BEN domain containing 2 protein (BEND2). It is suggested to participate in protein and DNA interactions during chromatin restructuring or transcription, however, no previous studies described its role in the testis. Thus our results suggest that BEND2 might be involved in chromatin restructuring and transcriptional events taking place during meiosis.

Expression of the hyaluronan mediated motility receptor (HMMR, also called RHAMM or CD168) was also exclusive to pachytene spermatocytes, but with dense staining in the cytoplasm of the cells (Figure 5). The PE1 protein HMMR was first shown to be involved in cell motility [37] but is now known to play a role in various biological processes. The protein that activates several signaling cascades, is associated with cancer metastasis, and has been suggested as a biomarker for poor prognosis in several tumor types [38]. The expression and cellular distribution of HMMR in normal tissues was recently demonstrated, and showed consistency with HMMR being

Accepted manuscript

a G2/M cell cycle protein [39]. Interestingly, the authors also showed that in testis, HMMR was expressed in spermatocytes, which is in accordance with our results and supports the hypothesis that HMMR could be a meiosis checkpoint protein. An uPE1 protein with an expression pattern similar to that of HMMR is the spermatogenesis and centriole associated 1 like protein (SPATC1L) (Figure 5). Knowledge of its function is limited, however, SPATC1L is suggested to be involved in sperm head-tail integrity and it has been shown that disruption of the gene in mice leads to infertility [40].

Shugoshin 1 (SGO1) is another PE1 checkpoint protein known to play a central role in chromosome cohesion. Disruption of the gene is associated with premature loss of centromeric cohesion, mis-segregation of sister chromatids, and mitotic arrest [41]. We here confirm that the protein is expressed in testis, with a speckled nuclear staining pattern in pachytene spermatocytes (Figure 5), possibly representing e.g. centromeres or centrosomes, which is in concordance with the suggested protein function. A uPE1 protein with a similar speckled nuclear expression pattern in pachytene spermatocytes is the sel-1 homolog 2 protein (SEL1L2) (Figure 5). There are no previous studies describing neither the function nor a testis elevated expression of this protein. We here show that SEL1L2 is exclusively expressed in testis both at the RNA and protein level, with an expression restricted to pachytene spermatocytes, suggesting an important role in meiosis.

Proteins specific to meiotic and post meiotic germ cells (cluster 4)

The meiotic and postmeiotic germ cell group was defined as proteins localized in pachytene spermatocytes and spermatids (round/early and/or elongated/late spermatids). We identified a group of 78 proteins that were expressed in both meiotic and postmeiotic cells, and GO process analysis showed that these proteins are associated with reproduction, sexual reproduction and

Accepted manuscript

single fertilization (Figure 1C). Sixty-eight of the 78 proteins belong to PE1, seven are uPE1, two are PE2 and one protein lacks information on PE status.

Immunohistochemical staining patterns of proteins expressed specifically in meiotic and postmeiotic germ cells are displayed in Figure 6. The PE1 boule-like protein (BOLL), is a RNA-binding protein that belongs to the DAZ gene family. Similarly as these genes, BOLL is important for spermatogenesis as loss of BOLL results in azoospermia, due to defects in the meiotic G2/M transition [42]. We showed that BOLL is expressed mainly in pachytene spermatocytes but also in spermatids (Figure 6). Another example is the PE1 regulatory factor X 3 protein (RFX3), a transcription factor that showed a distinct nuclear staining pattern in pachytene spermatocytes, round/early and elongated/late spermatids (Figure 6). RFX is involved in ciliogenesis and essential for motile cilia differentiation, growth and beating efficiency of ciliated cells [43], which further confirms its abundant expression, specifically in spermatids.

Similarly to BOLL, the uPE1 proteins axonemal dynein light chain domain-containing protein 1 (AXDND1), leucine-rich repeat-containing protein 23 (LRRC23), uncharacterized protein C16orf71, family with sequence similarity 181 member A protein (FAM181A) and the PE2 uncharacterized protein C9orf50 all displayed a cytoplasmic staining pattern (Figure 6), with an accumulation of the protein observed in cytoplasmic bodies of late spermatids. Another uPE1 protein is the uncharacterized protein C17orf64, which appeared to be clearly present in the nuclei of meiotic and postmeiotic germ cells, but in comparison to RFX3, the C17orf64 staining was slightly more intense in spermatids than in nuclei of spermatocytes (Figure 6).

Proteins specific to postmeiotic germ cells (cluster 5)

Spermiogenesis is the final stage of spermatogenesis when the newly formed haploid germ cells (*i.e.*, the round/early spermatids) transform into the anatomically mature spermatozoa.

Accepted manuscript

Spermiogenesis is traditionally divided into four successive stages: the Golgi, the cap, the acrosomal and the maturation phases. During the Golgi phase, round spermatids will begin to develop polarity. The head is formed at one end, where the Golgi apparatus forms the acrosome and DNA undergoes packaging. At the other end, the cell develops a thickened mid-piece where mitochondria accumulate and the distal centriole assembles structures to form an axoneme. Both the acrosome and axoneme are needed for sperm motility and fertility [44]. Protein expression specific to post meiotic germ cells should match with features that are related to the unique phenomenon haploid cells undergo during spermiogenesis, such as the development of polarized structures (*e.g.,* acrosome and axoneme), chromatin remodeling, repackaging and transcriptional reprogramming. Besides, spermatozoa, the final products of spermatogenesis, have hyper-selected traits associated with shape and energy production, as well as enhanced flagellum development [45-46]. In the present investigation, 132 proteins were identified as specifically expressed in postmeiotic germ cells. The GO analysis showed enrichment for the processes reproduction, sexual reproduction and gamete generation (Figure 1C). Eighty-three of the 132 proteins belong to PE1, while as many as 30 are uPE1 and 19 proteins are defined as MPs (PE2-4).

Immunohistochemical staining patterns of proteins expressed specifically in postmeiotic germ cells are displayed in Figure 7. The zona pellucida-binding protein 1 (ZPBP) is a PE1 protein that clearly localized to the acrosome under formation in round/early spermatids (Figure 7). ZPBP interacts with multiple acrosomal matrix proteins and is required for sperm acrosome formation and compaction [47]. ZPBP is also known to bind to the zona pellucida following the acrosome reaction [48-49]. Interestingly, homozygous mice lacking *Zpbp1* have acrosome fragmentation, disrupted Sertoli-spermatid junctions and a defective sperm head morphology with characteristics reminiscent of teratozoospermia in infertile men [47]. More recently, it was demonstrated that mutations of the ZPBP gene are correlated with abnormal sperm head morphology in infertile

Accepted manuscript

men, suggesting that this protein is crucial for sperm morphogenesis [50]. Of note is that in the Human Protein Atlas, spermatozoa in the epididymis are also stained for ZPBP.

The transmembrane protein 210 (TMEM210) and the uncharacterized protein C7orf61, both uPE1, were also shown to be expressed in the acrosome under formation in spermatids (Figure 7). Additional immunoreactivity of TMEM210 was found in the cytoplasmic lobes of elongated spermatids, suggesting that the protein could eventually be recycled through residual bodies. Similar to that of ZPBP, the localization of these two proteins in the acrosome suggests their implication in acrosome formation and/or function. The uncharacterized uPE1 protein FAM209B and the MP ACTRT1 (PE2) also stained acrosome features (Figure 7).

In the present study, the A-kinase anchoring protein 4 (AKAP4), was found to be specific to elongated spermatids and clearly illustrates a flagellum localization (Figure 7). AKAP4 belongs to the A-kinase anchoring protein family. It is a well-known cancer/testis antigen overexpressed in some cancer types [51] and whose expression in normal human tissues is restricted to the testis and flagella of late spermatids [52]. AKAP4 is a scaffold protein required for the organization and integrity of the fibrous sheath [53]. Targeted disruption of the AKAP4 gene in mouse has been shown to induce significant defects on sperm flagella, which affects mobility as signal transduction and glycolytic enzymes fail to become associated with the fibrous sheath. In accordance with previous results, our study shows that the PE1 protein AKAP4 is expressed in the flagella of late spermatids, thus suggesting that AKAP4 is necessary for flagellar function. Interestingly, AKAP4 was recently patented for clinical use as a quality marker of human sperm.

Other proteins that displayed a clear flagellar signal in late spermatids are the uncharacterized protein C11orf42 and the forkhead associated phosphopeptide binding domain 1 (FHAD1), both uPE1. These two proteins have no known function to date, but their flagella-restricted localization

Accepted manuscript

in spermatids clearly suggests their implication in flagellum structure or function. Interestingly, FHAD1 was also found in ciliated cells of both Fallopian tubes and respiratory epithelia (https://www.proteinatlas.org/ENSG00000142621-FHAD1/tissue), consolidating our hypothesis that FHAD1 could play a role in cilium structure or function. The spermatogenesis-associated protein 3 (SPATA3) signal was also clearly observed in the flagella/cytoplasmic lobes of elongated spermatids. SPATA3 is an uncharacterized uPE1 protein, however RT-PCR experiments of its mouse ortholog, *Spata3*, shows that the gene is exclusively expressed in mouse testis [54]. Another example of an uPE1 protein expressed in flagella is C3orf30 (also known as TEX55). A low C3orf30 signal was also present in round/early spermatids, suggesting an early production of the protein after meiosis. Interestingly, a previous report showed that C3orf30 is a human testis-specific gene and no expression was detected in patients with cryptorchidism and Sertoli-cell-only syndrome [55].

Protein expression in somatic cells (cluster 6)

Protein expression in somatic cells, represented by Sertoli cells, Leydig cells and peritubular cells, was found in several different combinations, including simultaneous expression in one or several germ cell types. Twenty-seven proteins were expressed solely in somatic cells and not in germ cells, including 19 PE1 proteins, five uPE1 proteins and three MPs (PE2-4) (Supplementary Table 1). Enriched GO terms associated with these proteins include reproduction, hemoglobin metabolic process and hemoglobin biosynthetic process, suggesting a wide variety of functions supporting and controlling spermatogenesis and confirming the proximity of somatic cells to blood vessels in the interstitial compartment (Figure 1C).

Relationship between cell type-specific expression in testis and other tissue types

Based on RNA-Seq, genes elevated in testis often show expression at a lower level in other tissues. The highest number of genes that share an elevated expression between testis and

Accepted manuscript

another tissue are genes simultaneously elevated in Fallopian tube, which is expected due to similar functional characteristics related to motility of sperm flagella and Fallopian tube cilia. Of all 2,237 genes elevated in testis, 190 genes were shown to be either group enriched or tissue enhanced in both testis and Fallopian tube. In the present investigation, 55 of these genes were included (Supplementary Table 1). A GO based analysis of these 55 genes showed enrichment for several processes related to cilia and flagella, including microtubule-based process, cilium biogenesis, and ciliary or flagellar motility. At the protein level, as many as 43 (78%) of these proteins showed expression in elongated/late spermatids, in some cases in combination with expression in other testicular cell types. Thirty-three of the 55 proteins belong to PE1, while 20 are uPE1 and two proteins are defined as MPs (PE2-4).

Immunohistochemical staining patterns of proteins simultaneously elevated in testis and Fallopian tube are displayed in Figure 8. One example is the PE1 protein HYDIN expressed in sperm flagella and cilia of both Fallopian tubes and bronchus (Figure 8), which is in accordance with previous literature [56]. Motile cilia/flagella are crucial in mammals for various functions such as sperm propulsion, the removal of debris from respiratory tracts, the circulation of cerebrospinal fluid, and even for the left–right body asymmetry [57]. Thus, in ciliopathies that result from defects in ciliary assembly and/or motility, respiratory distress, hydrocephalus, and/or randomization of the left–right body axis can be observed. Interestingly, a reduced fertility is often seen in male patients due to abnormalities of sperm tails [58].

Other examples of shared expression between testis, Fallopian tubes and airway epithelia are the uPE1 proteins CFAP45 and EFCAB1, the PE2 protein NEK5, as well as the two proteins RSPH10B and RSPH10B2 (Figure 8). The PE2 protein RSPH10B and the uPE1 protein RSPH10B2 have overlapping sequences and are targeted by the same antibody. The expression patterns of these proteins mainly restricted to elongated/late spermatids in testis and cilia of

Accepted manuscript

Fallopian tubes and airway epithelia suggest that these proteins are implicated in cilia and flagella formation and/or function.

Accepted manuscript

DISCUSSION

Over the past years, three drafts of the human proteome have been published that describe the protein repertoires in an array of normal human tissues [4, 59-60]. Here, we utilized publicly available data from the Human Protein Atlas (www.proteinatlas.org), which has been set up to systematically explore the entire human proteome across tissues, organs, cells and organelles [4]. The large-scale mapping includes categorization of all human genes based on RNA-Seq, combined with immunohistochemistry on TMAs, for spatial localization of each protein. Based on a stringent workflow and expert annotation of the cell type-specific expression pattern in testis, we performed an in-depth characterization of 512 proteins that showed an elevated expression in testis compared to other tissues. The elevated expression based on RNA-Seq suggests that these proteins may be implicated in testis-specific functions and male reproductive processes. Knowledge of the cell type-specific expression at the protein level will add important information on the presumed function of these proteins in spermatogenesis.

Another systematic worldwide effort designed to map the entire human proteome is HPP. Upon completion, it will enhance our global understanding of the protein-based architecture of human tissues and cells, help elucidate biological and molecular functions, and lay the foundation for development of innovative medical applications. neXtProt, the reference protein knowledgebase of the HPP [7], has set up criteria to rank proteins into categories according to evidence of their existence (PE). In such context, the Chromosome-Centric Human Proteome Project (C-HPP) was launched to catalogue proteins as gene products encoded by the human genome in a chromosome-centric manner [5]. From the 19,656 predicted proteins in the human proteome, this coordinated effort worldwide, together with the adoption of stringent interpretation guidelines of mass spectrometry data [61], has led to 17,694 experimentally validated proteins (PE1). The count of the so-called MPs that lack evidence of existence at the protein level and scored as PE2, or 3

Accepted manuscript

or 4 now stands at 2,129. A new objective towards the completion and annotation of the human PE1 proteins list is the annotation and study of those proteins lacking functional annotation (uPE1) (neXt-CP50 challenge launched by the HUPO C-HPP consortium in March 2018). Recent works thus present strategies and new findings on some of the ~1,260 uncharacterized "dark proteins" (uPE1) [62], including testicular germ cell proteins [63].

Testis has previously been suggested as an important source for identification of MPs, in particular postmeiotic germ cells. Due to the complex nature of the organ, it is however challenging to study, and the spermatogenesis process differs between humans and most mammals, especially its three-dimensional helical organization. During germ cell development and maturation, different testicular cell types form associations of fixed composition, called "stages", constituting the seminiferous epithelium cycle. In rat, the cycle has historically been divided into 14 stages, with spermiogenesis broken down further into 19 differentiation steps [44], providing a striking and unique example of cell differentiation involving acrosome formation, nuclear condensation and flagellar biogenesis. All stages and steps in spermatogenesis, lasting 74 days in humans, involve thousands of genes and proteins that are activated or repressed (for a review, see Rolland et al [64]. Another intriguing feature of spermatogenesis is the distinct ordering of cell associations along the length of the seminiferous tubules (segments), often referred to as the 'wave of the seminiferous epithelium' [65]. A segment is defined as a longitudinal portion of seminiferous tubule corresponding to a single cell association or stage [66]. A wave encompasses all 14 segments in rat, 12 in mouse, and 6 in human, but the number of discernible stages in human has been suggested to be increased from 6 to 12, based on acrosome development [67]. In summary, the complex anatomical and functional features of testis highlight the needs of a stringent workflow and extensive knowledge for interpretation cell type-specific expression patterns in immunohistochemically stained testis sections. The present investigation involved

Accepted manuscript

strict guidelines for all steps in the workflow, including automated immunohistochemistry, enhanced antibody validation strategies and expert annotation.

There is a widely acknowledged need for improved reproducibility of protein assays, and it is of outmost importance that research antibodies are properly validated. Recently, the International Working Group for Antibody Validation (IWGAV), has suggested five different “pillars” to use as standard for antibody validation, drawing increased attention to standardization of antibody validation strategies [15]. The Human Protein Atlas project has implemented these five pillars to follow a stringent workflow for enhanced validation of antibodies, performed in an application-specific manner for Western blot, immunocytochemistry and immunohistochemistry [68]. In version 18, 10,540 antibodies directed to a total of 6,787 human protein targets passed the criteria of enhanced validation. For immunohistochemistry, two main strategies are used: i) orthogonal validation, comparing protein expression levels using an antibody-independent method, or ii) independent antibody validation, comparing protein expression levels using two different antibodies targeting non-overlapping epitopes of the same protein. In the present investigation, main emphasis was put on orthogonal validation, comparing protein expression levels with RNA-Seq data across 37 tissues. All 512 selected proteins showed a high correlation between protein and RNA expression levels. Additionally, 45 proteins were validated by similar expression patterns observed by independent antibodies. For 485 out of the 512 analyzed proteins, in-house generated Human Protein Atlas antibodies were used. These antibodies have undergone additional validation upon generation, including sequencing of plasmid inserts in order to ensure that the correct protein epitope signature tag (PrEST) sequence has been cloned. The size of the resulting recombinant protein (the PrEST and appropriate tag sequences) is analyzed by mass spectrometry to ensure production and purification of the correct antigen. The antibodies, generated by immunizing rabbits with PrESTs, are affinity-purified and tested for sensitivity and specificity on arrays of PrESTs that have been spotted on glass slides. Only antibodies that meet

Accepted manuscript

all quality criteria are used in further applications. In summary, antibodies used in the present investigation have undergone validation using strict criteria, in order to ensure that the correct protein targets are visualized. It is however important to address general limitations of antibodies used for research applications [69], and further efforts aiming at setting up stringent criteria for antibody validation, as well as how antibody-based protein data should be used for determining evidence of protein existence are clearly warranted. One example of an technology complementing immunohistochemistry for *in situ* detection and cell type-specific localization of proteins is the use of proximity ligation assay (PLA), requiring binding of two antibodies in order to generate a locally amplified detection signal [70]. This thus serves as an interesting approach in order to improve the specificity of detection, and the technology is expected to receive increased attention for enhanced validation of antibodies.

The cell type-specific localization of the 512 proteins in human testis allowed us to categorize the proteins into six different clusters based on expression, that match with the three known phases of spermatogenesis (*i.e.,* premeiosis, meiosis and postmeiosis). Within each cluster, we could identify several well-known proteins with testis-specific functions well in line with the protein expression patterns, and GO analysis confirmed the expected processes related to each cluster. The analysis also highlighted numerous poorly characterized proteins in each of these clusters whose expression overlapped with that of known proteins involved in spermatogenesis. Furthermore, we were able to characterize the *in situ* distribution of several proteins that previously lacked spatial information and cell type-specific expression within the testis.

In total, as many as 88 of the 512 analyzed proteins were defined as uPE1, and 60 proteins were classified as MPs (PE2-4). A large proportion of the uPE1 and MPs identified in our study were found to be localized in the premeiotic or postmeiotic clusters, and displayed significant levels of expression in germ cells at either early or late stages of their development. Our results suggest

Accepted manuscript

that some of the 60 MPs could be evidenced in isolated premeiotic germ cells extracts or in ejaculated sperm as a starting material, instead of working from whole testis extracts. Using targeted mass spectrometry assays on these enriched cell fractions would be a valuable option to increase the chances of identifying unambiguously novel MPs. This is well in line with our previous study that demonstrated the potential of testis for characterization of MPs [63]. Isolation of premeiotic germ cells from human testis in sufficient amounts is challenging, while we successfully used human sperm extracts for identification of over 250 MPs [10, 12, 71]. In order to increase the chances of identifying new MPs that are expressed at low copy numbers or linked to sperm cellular membranes, current efforts are focused on sample preparation of human sperm subcellular fractions, i.e., head, midpiece and flagellum.

Over the past years, significant progresses have been made on the credible detection of MPs, but also on studying proteins that have unknown functions [62]. Most current studies are however based on mass spectrometry. Antibody-based proteomics constitutes an advantageous strategy for further exploration of unknown distribution patterns in intact tissue samples. The technology provides information on spatial expression both on cellular and subcellular level, in the context of neighboring cells, which is a first step in elucidating the possible function of each protein. By combining high-throughput immunohistochemistry on TMAs with transcriptomic analysis of most major organs and tissues, proteins with exclusive expression in specific cell types not expressed in other organs can be identified [72]. This integrated omics approach constitutes a comprehensive strategy for organ-specific research in health and disease.

Previous studies on testis have suggested a poor correlation between mRNA and protein expression levels compared to other human organs [73]. The poor correlation could be the consequence of peculiar properties of gene/transcript regulation occurring post meiotically in the germ cell lineage, such as delays between transcription and translation or mRNA storage in

Accepted manuscript

translationally repressed free ribonucleoprotein (RNP) particles. In the present investigation, we identified and characterized 512 proteins with an elevated expression in testis in comparison with other organs both at the mRNA and protein level, hence suggesting an overall good correlation between the datasets for these proteins.

For several proteins, e.g., CFAP45, EFCAB1, NEK5, RSPH10B/RSPH10B2, immunoreactivity was observed both in the flagella of elongated spermatids in the stage prior to spermiation but also in cilia of Fallopian tubes and airway epithelia. Such regionalized expression suggests a role for these proteins in cilia and/or flagella biology, and they hence constitute good candidates for further functional studies. Interestingly, the uPE1 protein CFAP45 was recently hypothesized to be involved in cilia biology [74]. We here provide additional insight into the potential role of this protein in the formation and/or motility of sperm. The *in situ* localization of the PE2 protein NEK5 in the head of elongated spermatids prior to spermiation again confirms the interest of the sperm as a source for discovering additional MPs.

In the present study, we gathered valuable information on cell type expression of numerous proteins within the seminiferous tubules and provide interesting clues on the possible function of uPE1 and other MPs during spermatogenesis. As a continuation of this project, it would be interesting to launch a I-TASSER/COFACTOR computational protein structure prediction [75] on the 88 uPE1 proteins highlighted in this project so as to eventually complete their functional annotation. The final question remains how the functional annotation data presented in the present study and the one to come could be beneficially integrated in neXtProt.

Recently, single cell technologies has received increased attention, and several ongoing efforts world-wide focus on mapping mouse and human organs, including testis, based on single cell RNA-Seq [22, 76]. Our detailed data is likely to be valuable for future studies combining various

Accepted manuscript

single cell technologies, as well as machine learning approaches, in order to generate a complete and comprehensive mapping of all cell types in the human body.

In summary, our strategy of combining transcriptomics with antibody-based proteomics is an attractive approach for studying those proteins lacking functional annotation (uPE1) and further characterization of MPs. Our comprehensive dataset with cell type-specific expression of testis elevated proteins constitutes an important resource for further testis-specific research in male reproductive disorders.

SUPPORTING INFORMATION

The following supporting information is available free of charge at ACS website http://pubs.acs.org

Supplementary table S1. Detailed information on antibodies, protein evidence, protein expression data and GO terms for all 512 analyzed proteins.

ACKNOWLEDGEMENTS

The project was funded by the Knut and Alice Wallenberg Foundation. The work was also supported by structural grants from Biogenouest, Infrastructures en Biologie Santé et Agronomie (IBiSA) and Conseil Régional de Bretagne awarded to C.P. Pathologists and staff at the Department of Clinical Pathology, Uppsala University Hospital, are acknowledged for providing the tissues used in the study. The authors would also like to thank all staff of the Human Protein Atlas for their work.

Accepted manuscript

FIGURE LEGENDS

Figure 1. Categorization of testis elevated genes based on cell type-specific expression.

A. PCA plot showing the association between the eight cell types based on overall protein expression patterns in testis. B. Visualization of the protein expression profiles in a heatmap, showing the protein expression levels in the eight cell types based on 512 proteins which are clustered based on their pairwise Spearman correlation. The PCA plot and heatmap allowed us to highlight six expression groups: 1) premeiotic; 2) premeiotic + meiotic; 3) meiotic; 4) meiotic + postmeiotic; 5) postmeiotic; and 6) somatic. C. Visualization of enriched GO terms associated with each of the six expression groups, highlighting the top three most significant GO terms for each group.

Figure 2. Relationship of protein expression between different testicular cell types.

The visualization shows the relationship of protein expression between different testicular cell types (black nodes), highlighting six main expression groups: premeiotic (green nodes), premeiotic + meiotic (purple nodes), meiotic (yellow nodes), meiotic + postmeiotic (red nodes), postmeiotic (blue nodes), somatic (pink nodes), as well as remaining proteins not belonging to a specific expression group (grey nodes).

Figure 3. Immunohistochemical staining pattern of proteins specific to premeiotic germ cells.

Transverse sections of adult human testis showing protein localization in spermatogonia (arrowheads) and preleptotene spermatocytes (arrows). TIPIN, SAGE1, ERICH6 and PKDREJ showed moderate to strong nuclear staining in both spermatogonia and preleptotene spermatocytes, while MAGEC1, OOSP2, DCAF4L1 and C3orf22 displayed cytoplasmic positivity in spermatogonia and preleptotene spermatocyes. Bar=50 µm.

Accepted manuscript

Figure 4. Immunohistochemical staining of proteins specific to premeiotic and meiotic germ cells.

Transverse sections of adult human testis showing protein localization in spermatogonia (arrowheads), preleptotene spermatocytes (arrows) and pachytene spermatocytes (white arrows). PNMA5 staining showed a dense nuclear positivity in spermatogonia, preleptotene spermatocytes, and a limited number of pachytene spermatocytes depending on the stage the seminiferous cycle. DAZL/DAZ1-4 immunoreactivity was observed in the cytoplasm of spermatogonia, preleptotene spermatocytes and to a lesser extent in pachytene spermatocytes. CTAG1A staining was mainly distributed in cytoplasm of spermatogonia, preleptotene and pachytene spermatocytes, while CT45A1-3, 5-10 showed nuclear positivity. Bar=50 μm.

Figure 5. Immunohistochemical staining of proteins specific to meiotic germ cells.

Transverse sections of adult human testis showing protein localization in pachytene spermatocytes (white arrows). For all proteins, immunoreactivity was observed at moderate or strong intensity exclusively in pachytene spermatocytes, with a various subcellular localization pattern. DMC1 and BEND2 showed nuclear staining whereas the staining for HMMR and SPATC1L was localized to the cytoplasm. Immunoreactivity for SGO1 and SEL1L2 displayed a speckled nuclear pattern. Bar=50 μm.

Figure 6. Immunohistochemical staining of proteins specific to meiotic and postmeiotic cells.

Transverse sections of adult human testis showing protein localization in pachytene spermatocytes (white arrows), round/early spermatids (white arrowheads) and cytoplasmic lobes of elongated/late spermatids (red stars). BOLL showed cytoplasmic staining mainly in pachytene spermatocytes and round/early spermatids, and to a lesser extent in elongated spermatids. RFX3

Accepted manuscript

and C17orf64 mainly displayed nuclear positivity in pachytene spermatocytes and round/early spermatids, while AXDND1, LRRC23, C16orf71, FAM181A and C9orf50 all predominantly showed a moderate to strong staining in pachytene spermatocytes, round/spermatids and cytoplasmic lobes of elongated/late spermatids. Bar=50μm.

Figure 7. Immunohistochemical staining of proteins specific to postmeiotic cells.

Transverse sections of adult human testis showing protein localization in acrosome under formation elongated/late spermatids (black arrows), cytoplasmic lobes of elongated/late spermatids (red star) and flagella of elongated/late spermatids (white arrows). A concentrated signal was observed in the acrosome under formation for ZPBP, TMEM210, C7orf61, FAM209B and ACTRT1, with TMEM210 showing additional immunoreactivity in cytoplasmic lobes of elongated spermatids. AKAP4, C11orf42, FHAD1, SPATA3 and C3orf30 displayed an intense staining in flagella. Bar=50 μm.

Figure 8. Immunohistochemical staining of proteins expressed in testis, Fallopian tube and airway epithelia.

HYDIN and CFAP45 staining was restricted to sperm flagella (arrows) and cilia of Fallopian tube and airway epithelia (white arrowheads). EFCAB1 signal showed a similar pattern, with additional immunoreactivity in cytoplasmic components of all three tissues. NEK5 staining in testis was displayed in the head of elongated spermatids (arrow), and in airway epithelia most abundant in the cilial roots (arrow). RSPH10B/RSPH10B2 staining was distributed in both round/early and elongated/late spermatids in testis, and to cilia in Fallopian tube and respiratory epithelia. Note: Airway epithelial staining is represented by nasopharynx for HYDIN, CFAP45, EFCAB1 and NEK5, and bronchus for RSPH10B/RSPH10B2. Bar=50 μm.

Accepted manuscript

REFERENCES

1. Calvel, P.; Rolland, A. D.; Jegou, B.; Pineau, C., Testicular postgenomics: targeting the regulation of spermatogenesis. *Philos Trans R Soc Lond B Biol Sci* **2010,** *365* (1546), 1481-500.
2. Djureinovic, D.; Fagerberg, L.; Hallstrom, B.; Danielsson, A.; Lindskog, C.; Uhlen, M.; Ponten, F., The human testis-specific proteome defined by transcriptomics and antibody-based profiling. *Mol Hum Reprod* **2014,** *20* (6), 476-88.
3. Fagerberg, L.; Hallstrom, B. M.; Oksvold, P.; Kampf, C.; Djureinovic, D.; Odeberg, J.; Habuka, M.; Tahmasebpoor, S.; Danielsson, A.; Edlund, K.; Asplund, A.; Sjostedt, E.; Lundberg, E.; Szigyarto, C. A.; Skogs, M.; Takanen, J. O.; Berling, H.; Tegel, H.; Mulder, J.; Nilsson, P.; Schwenk, J. M.; Lindskog, C.; Danielsson, F.; Mardinoglu, A.; Sivertsson, A.; von Feilitzen, K.; Forsberg, M.; Zwahlen, M.; Olsson, I.; Navani, S.; Huss, M.; Nielsen, J.; Ponten, F.; Uhlen, M., Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* **2014,** *13* (2), 397-406.
4. Uhlen, M.; Fagerberg, L.; Hallstrom, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigyarto, C. A.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P. H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; von Heijne, G.; Nielsen, J.; Ponten, F., Proteomics. Tissue-based map of the human proteome. *Science* **2015,** *347* (6220), 1260419.
5. Legrain, P.; Aebersold, R.; Archakov, A.; Bairoch, A.; Bala, K.; Beretta, L.; Bergeron, J.; Borchers, C. H.; Corthals, G. L.; Costello, C. E.; Deutsch, E. W.; Domon, B.; Hancock, W.; He, F.; Hochstrasser, D.; Marko-Varga, G.; Salekdeh, G. H.; Sechi, S.; Snyder, M.; Srivastava, S.; Uhlen, M.; Wu, C. H.; Yamamoto, T.; Paik, Y. K.; Omenn, G. S., The human proteome project: current state and future direction. *Mol Cell Proteomics* **2011,** *10* (7), M111 009993.
6. Omenn, G. S.; Lane, L.; Lundberg, E. K.; Overall, C. M.; Deutsch, E. W., Progress on the HUPO Draft Human Proteome: 2017 Metrics of the Human Proteome Project. *J Proteome Res* **2017,** *16* (12), 4281-4287.
7. Gaudet, P.; Michel, P. A.; Zahn-Zabal, M.; Britan, A.; Cusin, I.; Domagalski, M.; Duek, P. D.; Gateau, A.; Gleizes, A.; Hinard, V.; Rech de Laval, V.; Lin, J.; Nikitin, F.; Schaeffer, M.; Teixeira, D.; Lane, L.; Bairoch, A., The neXtProt knowledgebase on human proteins: 2017 update. *Nucleic Acids Res* **2017,** *45* (D1), D177-D182.
8. Omenn, G. S.; Lane, L.; Overall, C. M.; Corrales, F. J.; Schwenk, J. M.; Paik, Y. K.; Van Eyk, J. E.; Liu, S.; Snyder, M.; Baker, M. S.; Deutsch, E. W., Progress on Identifying and Characterizing the Human Proteome: 2018 Metrics from the HUPO Human Proteome Project. *J Proteome Res* **2018,** *17* (12), 4031-4041.
9. Zhang, Y.; Li, Q.; Wu, F.; Zhou, R.; Qi, Y.; Su, N.; Chen, L.; Xu, S.; Jiang, T.; Zhang, C.; Cheng, G.; Chen, X.; Kong, D.; Wang, Y.; Zhang, T.; Zi, J.; Wei, W.; Gao, Y.; Zhen, B.; Xiong, Z.; Wu, S.; Yang, P.; Wang, Q.; Wen, B.; He, F.; Xu, P.; Liu, S., Tissue-Based Proteogenomics Reveals that Human Testis Endows Plentiful Missing Proteins. *J Proteome Res* **2015,** *14* (9), 3583-94.
10. Vandenbrouck, Y.; Lane, L.; Carapito, C.; Duek, P.; Rondel, K.; Bruley, C.; Macron, C.; Gonzalez de Peredo, A.; Coute, Y.; Chaoui, K.; Com, E.; Gateau, A.; Hesse, A. M.; Marcellin, M.; Mear, L.; Mouton-Barbosa, E.; Robin, T.; Burlet-Schiltz, O.; Cianferani, S.; Ferro, M.; Freour, T.; Lindskog, C.; Garin, J.; Pineau, C., Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update. *J Proteome Res* **2016,** *15* (11), 3998-4019.
11. Carapito, C.; Duek, P.; Macron, C.; Seffals, M.; Rondel, K.; Delalande, F.; Lindskog, C.; Freour, T.; Vandenbrouck, Y.; Lane, L.; Pineau, C., Validating Missing Proteins in Human

Sperm Cells by Targeted Mass-Spectrometry- and Antibody-based Methods. *J Proteome Res* **2017,** *16* (12), 4340-4351.
12. Jumeau, F.; Com, E.; Lane, L.; Duek, P.; Lagarrigue, M.; Lavigne, R.; Guillot, L.; Rondel, K.; Gateau, A.; Melaine, N.; Guevel, B.; Sergeant, N.; Mitchell, V.; Pineau, C., Human Spermatozoa as a Model for Detecting Missing Proteins in the Context of the Chromosome-Centric Human Proteome Project. *J Proteome Res* **2015,** *14* (9), 3606-20.
13. Uhlen, M.; Bjorling, E.; Agaton, C.; Szigyarto, C. A.; Amini, B.; Andersen, E.; Andersson, A. C.; Angelidou, P.; Asplund, A.; Asplund, C.; Berglund, L.; Bergstrom, K.; Brumer, H.; Cerjan, D.; Ekstrom, M.; Elobeid, A.; Eriksson, C.; Fagerberg, L.; Falk, R.; Fall, J.; Forsberg, M.; Bjorklund, M. G.; Gumbel, K.; Halimi, A.; Hallin, I.; Hamsten, C.; Hansson, M.; Hedhammar, M.; Hercules, G.; Kampf, C.; Larsson, K.; Lindskog, M.; Lodewyckx, W.; Lund, J.; Lundeberg, J.; Magnusson, K.; Malm, E.; Nilsson, P.; Odling, J.; Oksvold, P.; Olsson, I.; Oster, E.; Ottosson, J.; Paavilainen, L.; Persson, A.; Rimini, R.; Rockberg, J.; Runeson, M.; Sivertsson, A.; Skollermo, A.; Steen, J.; Stenvall, M.; Sterky, F.; Stromberg, S.; Sundberg, M.; Tegel, H.; Tourle, S.; Wahlund, E.; Walden, A.; Wan, J.; Wernerus, H.; Westberg, J.; Wester, K.; Wrethagen, U.; Xu, L. L.; Hober, S.; Ponten, F., A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* **2005,** *4* (12), 1920-32.
14. Kampf, C.; Olsson, I.; Ryberg, U.; Sjostedt, E.; Ponten, F., Production of tissue microarrays, immunohistochemistry staining and digitalization within the human protein atlas. *J Vis Exp* **2012,** *e3620* (63), 1-8.
15. Uhlen, M.; Bandrowski, A.; Carr, S.; Edwards, A.; Ellenberg, J.; Lundberg, E.; Rimm, D. L.; Rodriguez, H.; Hiltke, T.; Snyder, M.; Yamamoto, T., A proposal for validation of antibodies. *Nat Methods* **2016,** *13* (10), 823-7.
16. Core Team, R. *R: A language and environment for statistical computing.* , Foundation for Statistical Computing: Vienna, Austria, 2017.
17. Yang, Y. P., A; Dudoit, S. *marray: Exploratory analysis for two-color spotted microarray data.* , R package version 1.56.0.: 2009.
18. Warnes, G. B., B; Bonebakker, L; Gentleman, R; Huber, W; Liaw, A; Lumley, T; Maechler, M; Magnusson, A; Moeller, S; Schwartz, M; Venables, B. *gplots: Various R Programming Tools for Plotting Data.* , R package version 3.0.1.: 2016.
19. Gu, Z.; Gu, L.; Eils, R.; Schlesner, M.; Brors, B., circlize Implements and enhances circular visualization in R. *Bioinformatics* **2014,** *30* (19), 2811-2.
20. Omenn, G. S.; Lane, L.; Overall, C. M.; Corrales, F. J.; Schwenk, J. M.; Paik, Y. K.; Van Eyk, J. E.; Liu, S.; Snyder, M.; Baker, M. S.; Deutsch, E. W., Progress on Identifying and Characterizing the Human Proteome: 2018 Metrics from the HUPO Human Proteome Project. *J Proteome Res* **2018**.
21. de Rooij, D. G.; Russell, L. D., All you wanted to know about spermatogonia but were afraid to ask. *J Androl* **2000,** *21* (6), 776-98.
22. Guo, J.; Grow, E. J.; Mlcochova, H.; Maher, G. J.; Lindskog, C.; Nie, X.; Guo, Y.; Takei, Y.; Yun, J.; Cai, L.; Kim, R.; Carrell, D. T.; Goriely, A.; Hotaling, J. M.; Cairns, B. R., The adult human testis transcriptional cell atlas. *Cell Res* **2018,** *28* (12), 1141-1157.
23. Yoshizawa-Sugata, N.; Masai, H., Human Tim/Timeless-interacting protein, Tipin, is required for efficient progression of S phase and DNA replication checkpoint. *J Biol Chem* **2007,** *282* (4), 2729-40.
24. Aria, V.; De Felice, M.; Di Perna, R.; Uno, S.; Masai, H.; Syvaoja, J. E.; van Loon, B.; Hubscher, U.; Pisani, F. M., The human Tim-Tipin complex interacts directly with DNA polymerase epsilon and stimulates its synthetic activity. *J Biol Chem* **2013,** *288* (18), 12742-52.
25. Martelange, V.; De Smet, C.; De Plaen, E.; Lurquin, C.; Boon, T., Identification on a human sarcoma of two new genes with tumor-specific expression. *Cancer Res* **2000,** *60* (14), 3848-55.

26. Vicens, A.; Gomez Montoto, L.; Couso-Ferrer, F.; Sutton, K. A.; Roldan, E. R., Sexual selection and the adaptive evolution of PKDREJ protein in primates and rodents. *Mol Hum Reprod* **2015,** *21* (2), 146-56.
27. Zhang, X. L.; Liu, P.; Yang, Z. X.; Zhao, J. J.; Gao, L. L.; Yuan, B.; Shi, L. Y.; Zhou, C. X.; Qiao, H. F.; Liu, Y. H.; Ying, X. Y.; Zhang, J. Q.; Ling, X. F.; Zhang, D., Pnma5 is essential to the progression of meiosis in mouse oocytes through a chain of phosphorylation. *Oncotarget* **2017,** *8* (57), 96809-96825.
28. Lee, Y. H.; Pang, S. W.; Poh, C. L.; Tan, K. O., Distinct functional domains of PNMA5 mediate protein-protein interaction, nuclear localization, and apoptosis signaling in human cancer cells. *J Cancer Res Clin Oncol* **2016,** *142* (9), 1967-77.
29. Reijo, R.; Lee, T. Y.; Salo, P.; Alagappan, R.; Brown, L. G.; Rosenberg, M.; Rozen, S.; Jaffe, T.; Straus, D.; Hovatta, O.; et al., Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat Genet* **1995,** *10* (4), 383-93.
30. Reynolds, N.; Cooke, H. J., Role of the DAZ genes in male fertility. *Reprod Biomed Online* **2005,** *10* (1), 72-80.
31. Huang, W. J.; Lin, Y. W.; Hsiao, K. N.; Eilber, K. S.; Salido, E. C.; Yen, P. H., Restricted expression of the human DAZ protein in premeiotic germ cells. *Hum Reprod* **2008,** *23* (6), 1280-9.
32. Satie, A. P.; Rajpert-De Meyts, E.; Spagnoli, G. C.; Henno, S.; Olivo, L.; Jacobsen, G. K.; Rioux-Leclercq, N.; Jegou, B.; Samson, M., The cancer-testis gene, NY-ESO-1, is expressed in normal fetal and adult testes and in spermatocytic seminomas and testicular carcinoma in situ. *Lab Invest* **2002,** *82* (6), 775-80.
33. Roeder, G. S.; Bailis, J. M., The pachytene checkpoint. *Trends Genet* **2000,** *16* (9), 395-403.
34. Bishop, D. K.; Park, D.; Xu, L.; Kleckner, N., DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* **1992,** *69* (3), 439-56.
35. Thorslund, T.; Esashi, F.; West, S. C., Interactions between human BRCA2 protein and the meiosis-specific recombinase DMC1. *EMBO J* **2007,** *26* (12), 2915-22.
36. Brown, M. S.; Grubb, J.; Zhang, A.; Rust, M. J.; Bishop, D. K., Small Rad51 and Dmc1 Complexes Often Co-occupy Both Ends of a Meiotic DNA Double Strand Break. *PLoS Genet* **2015,** *11* (12), e1005653.
37. Hardwick, C.; Hoare, K.; Owens, R.; Hohn, H. P.; Hook, M.; Moore, D.; Cripps, V.; Austen, L.; Nance, D. M.; Turley, E. A., Molecular cloning of a novel hyaluronan receptor that mediates tumor cell motility. *J Cell Biol* **1992,** *117* (6), 1343-50.
38. Wang, C.; Thor, A. D.; Moore, D. H., 2nd; Zhao, Y.; Kerschmann, R.; Stern, R.; Watson, P. H.; Turley, E. A., The overexpression of RHAMM, a hyaluronan-binding protein that regulates ras signaling, correlates with overexpression of mitogen-activated protein kinase and is a significant parameter in breast cancer progression. *Clin Cancer Res* **1998,** *4* (3), 567-76.
39. Chen, Y. T.; Chen, Z.; Du, Y. N., Immunohistochemical analysis of RHAMM expression in normal and neoplastic human tissues: a cell cycle protein with distinctive expression in mitotic cells and testicular germ cells. *Oncotarget* **2018,** *9* (30), 20941-20952.
40. Kim, J.; Kwon, J. T.; Jeong, J.; Kim, J.; Hong, S. H.; Kim, J.; Park, Z. Y.; Chung, K. H.; Eddy, E. M.; Cho, C., SPATC1L maintains the integrity of the sperm head-tail junction. *EMBO Rep* **2018,** *19* (9), 1-19.
41. Kitajima, T. S.; Sakuno, T.; Ishiguro, K.; Iemura, S.; Natsume, T.; Kawashima, S. A.; Watanabe, Y., Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* **2006,** *441* (7089), 46-52.
42. Fu, X. F.; Cheng, S. F.; Wang, L. Q.; Yin, S.; De Felici, M.; Shen, W., DAZ Family Proteins, Key Players for Germ Cell Development. *Int J Biol Sci* **2015,** *11* (10), 1226-35.

43. El Zein, L.; Ait-Lounis, A.; Morle, L.; Thomas, J.; Chhin, B.; Spassky, N.; Reith, W.; Durand, B., RFX3 governs growth and beating efficiency of motile cilia in mouse and controls the expression of genes involved in human ciliopathies. *J Cell Sci* **2009,** *122* (Pt 17), 3180-9.
44. Leblond, C. P.; Clermont, Y., Definition of the stages of the cycle of the seminiferous epithelium in the rat. *Ann N Y Acad Sci* **1952,** *55* (4), 548-73.
45. O'Donnell, L.; Nicholls, P. K.; O'Bryan, M. K.; McLachlan, R. I.; Stanton, P. G., Spermiation: The process of sperm release. *Spermatogenesis* **2011,** *1* (1), 14-35.
46. O'Donnell, L., Mechanisms of spermiogenesis and spermiation and how they are disturbed. *Spermatogenesis* **2014,** *4* (2), e979623.
47. Lin, Y. N.; Roy, A.; Yan, W.; Burns, K. H.; Matzuk, M. M., Loss of zona pellucida binding proteins in the acrosomal matrix disrupts acrosome biogenesis and sperm morphogenesis. *Mol Cell Biol* **2007,** *27* (19), 6794-805.
48. Mori, E.; Baba, T.; Iwamatsu, A.; Mori, T., Purification and characterization of a 38-kDa protein, sp38, with zona pellucida-binding property from porcine epididymal sperm. *Biochem Biophys Res Commun* **1993,** *196* (1), 196-202.
49. Mori, E.; Kashiwabara, S.; Baba, T.; Inagaki, Y.; Mori, T., Amino acid sequences of porcine Sp38 and proacrosin required for binding to the zona pellucida. *Dev Biol* **1995,** *168* (2), 575-83.
50. Yatsenko, A. N.; O'Neil, D. S.; Roy, A.; Arias-Mendoza, P. A.; Chen, R.; Murthy, L. J.; Lamb, D. J.; Matzuk, M. M., Association of mutations in the zona pellucida binding protein 1 (ZPBP1) gene with abnormal sperm head morphology in infertile men. *Mol Hum Reprod* **2012,** *18* (1), 14-21.
51. Agarwal, S.; Saini, S.; Parashar, D.; Verma, A.; Sinha, A.; Jagadish, N.; Batra, A.; Suri, S.; Gupta, A.; Ansari, A. S.; Lohiya, N. K.; Suri, A., The novel cancer-testis antigen A-kinase anchor protein 4 (AKAP4) is a potential target for immunotherapy of ovarian serous carcinoma. *Oncoimmunology* **2013,** *2* (5), e24270.
52. Turner, R. M.; Johnson, L. R.; Haig-Ladewig, L.; Gerton, G. L.; Moss, S. B., An X-linked gene encodes a major human sperm fibrous sheath protein, hAKAP82. Genomic organization, protein kinase A-RII binding, and distribution of the precursor in the sperm tail. *J Biol Chem* **1998,** *273* (48), 32135-41.
53. Miki, K.; Willis, W. D.; Brown, P. R.; Goulding, E. H.; Fulcher, K. D.; Eddy, E. M., Targeted disruption of the Akap4 gene causes defects in sperm flagellum and motility. *Dev Biol* **2002,** *248* (2), 331-42.
54. Li, L.; Liu, G.; Fu, J. J.; Li, L. Y.; Tan, X. J.; Yang, S.; Lu, G. X., Molecular cloning and characterization of a novel transcript variant of Mtsarg1 gene. *Mol Biol Rep* **2009,** *36* (5), 1023-32.
55. Yu, Z.; Wu, B.; Tang, A.; Chen, J.; Guo, X.; Qin, J.; Gui, Y.; Cai, Z., Expression profile of a novel germ cell-specific gene, TSCPA, in mice and human. *J Huazhong Univ Sci Technolog Med Sci* **2009,** *29* (5), 535-9.
56. Davy, B. E.; Robinson, M. L., Congenital hydrocephalus in hy3 mice is caused by a frameshift mutation in Hydin, a large novel gene. *Hum Mol Genet* **2003,** *12* (10), 1163-70.
57. Satir, P.; Christensen, S. T., Overview of structure and function of mammalian cilia. *Annu Rev Physiol* **2007,** *69*, 377-400.
58. Lechtreck, K. F.; Delmotte, P.; Robinson, M. L.; Sanderson, M. J.; Witman, G. B., Mutations in Hydin impair ciliary motility in mice. *J Cell Biol* **2008,** *180* (3), 633-43.
59. Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A. M.; Lieberenz, M.; Savitski, M. M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; Mathieson, T.; Lemeer, S.; Schnatbaum, K.; Reimer, U.; Wenschuh, H.; Mollenhauer, M.; Slotta-Huspenina, J.; Boese, J. H.; Bantscheff, M.; Gerstmair, A.; Faerber, F.; Kuster, B., Mass-spectrometry-based draft of the human proteome. *Nature* **2014,** *509* (7502), 582-7.

60. Kim, M. S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; Thomas, J. K.; Muthusamy, B.; Leal-Rojas, P.; Kumar, P.; Sahasrabuddhe, N. A.; Balakrishnan, L.; Advani, J.; George, B.; Renuse, S.; Selvan, L. D.; Patil, A. H.; Nanjappa, V.; Radhakrishnan, A.; Prasad, S.; Subbannayya, T.; Raju, R.; Kumar, M.; Sreenivasamurthy, S. K.; Marimuthu, A.; Sathe, G. J.; Chavan, S.; Datta, K. K.; Subbannayya, Y.; Sahu, A.; Yelamanchi, S. D.; Jayaram, S.; Rajagopalan, P.; Sharma, J.; Murthy, K. R.; Syed, N.; Goel, R.; Khan, A. A.; Ahmad, S.; Dey, G.; Mudgal, K.; Chatterjee, A.; Huang, T. C.; Zhong, J.; Wu, X.; Shaw, P. G.; Freed, D.; Zahari, M. S.; Mukherjee, K. K.; Shankar, S.; Mahadevan, A.; Lam, H.; Mitchell, C. J.; Shankar, S. K.; Satishchandra, P.; Schroeder, J. T.; Sirdeshmukh, R.; Maitra, A.; Leach, S. D.; Drake, C. G.; Halushka, M. K.; Prasad, T. S.; Hruban, R. H.; Kerr, C. L.; Bader, G. D.; Iacobuzio-Donahue, C. A.; Gowda, H.; Pandey, A., A draft map of the human proteome. *Nature* **2014,** *509* (7502), 575-81.

61. Deutsch, E. W.; Overall, C. M.; Van Eyk, J. E.; Baker, M. S.; Paik, Y. K.; Weintraub, S. T.; Lane, L.; Martens, L.; Vandenbrouck, Y.; Kusebauch, U.; Hancock, W. S.; Hermjakob, H.; Aebersold, R.; Moritz, R. L.; Omenn, G. S., Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. *J Proteome Res* **2016,** *15* (11), 3961-3970.

62. Paik, Y. K.; Overall, C. M.; Corrales, F.; Deutsch, E. W.; Lane, L.; Omenn, G. S., Toward Completion of the Human Proteome Parts List: Progress Uncovering Proteins That Are Missing or Have Unknown Function and Developing Analytical Methods. *J Proteome Res* **2018,** *17* (12), 4023-4030.

63. Melaine, N.; Com, E.; Bellaud, P.; Guillot, L.; Lagarrigue, M.; Morrice, N. A.; Guevel, B.; Lavigne, R.; Velez de la Calle, J. F.; Dojahn, J.; Pineau, C., Deciphering the Dark Proteome: Use of the Testis and Characterization of Two Dark Proteins. *J Proteome Res* **2018,** *17* (12), 4197-4210.

64. Rolland, A. D.; Jegou, B.; Pineau, C., Testicular development and spermatogenesis: harvesting the postgenomics bounty. *Adv Exp Med Biol* **2008,** *636*, 16-41.

65. Clermont, Y., Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. *Physiol Rev* **1972,** *52* (1), 198-236.

66. Griswold, M. D., Spermatogenesis: The Commitment to Meiosis. *Physiol Rev* **2016,** *96* (1), 1-17.

67. Muciaccia, B.; Boitani, C.; Berloco, B. P.; Nudo, F.; Spadetta, G.; Stefanini, M.; de Rooij, D. G.; Vicini, E., Novel stage classification of human spermatogenesis based on acrosome development. *Biol Reprod* **2013,** *89* (3), 60.

68. Edfors, F.; Hober, A.; Linderback, K.; Maddalo, G.; Azimi, A.; Sivertsson, A.; Tegel, H.; Hober, S.; Szigyarto, C. A.; Fagerberg, L.; von Feilitzen, K.; Oksvold, P.; Lindskog, C.; Forsstrom, B.; Uhlen, M., Enhanced validation of antibodies for research applications. *Nat Commun* **2018,** *9* (1), 4130.

69. Fredolini, C.; Bystrom, S.; Sanchez-Rivera, L.; Ioannou, M.; Tamburro, D.; Ponten, F.; Branca, R. M.; Nilsson, P.; Lehtio, J.; Schwenk, J. M., Systematic assessment of antibody selectivity in plasma based on a resource of enrichment profiles. *Sci Rep* **2019,** *9* (1), 8324.

70. Zieba, A.; Ponten, F.; Uhlen, M.; Landegren, U., In situ protein detection with enhanced specificity using DNA-conjugated antibodies and proximity ligation. *Mod Pathol* **2018,** *31* (2), 253-263.

71. Carapito, C.; Duek, P.; Macron, C.; Seffals, M.; Rondel, K.; Delalande, F.; Lindskog, C.; Freour, T.; Vandenbrouck, Y.; Lane, L.; Pineau, C., Validating Missing Proteins in Human Sperm Cells by Targeted Mass-Spectrometry- and Antibody-based Methods. *J Proteome Res* **2017**.

72. Sjostedt, E.; Sivertsson, A.; Hikmet Noraddin, F.; Katona, B.; Nasstrom, A.; Vuu, J.; Kesti, D.; Oksvold, P.; Edqvist, P. H.; Olsson, I.; Uhlen, M.; Lindskog, C., Integration of Transcriptomics and Antibody-Based Proteomics for Exploration of Proteins Expressed in Specialized Tissues. *J Proteome Res* **2018,** *17* (12), 4127-4137.

73. Cagney, G.; Park, S.; Chung, C.; Tong, B.; O'Dushlaine, C.; Shields, D. C.; Emili, A., Human tissue profiling with multidimensional protein identification technology. *J Proteome Res* **2005,** *4* (5), 1757-67.
74. Duek, P.; Gateau, A.; Bairoch, A.; Lane, L., Exploring the Uncharacterized Human Proteome Using neXtProt. *J Proteome Res* **2018,** *17* (12), 4211-4226.
75. Zhang, C.; Wei, X.; Omenn, G. S.; Zhang, Y., Structure and Protein Interaction-Based Gene Ontology Annotations Reveal Likely Functions of Uncharacterized Proteins on Human Chromosome 17. *J Proteome Res* **2018,** *17* (12), 4186-4196.
76. Green, C. D.; Ma, Q.; Manske, G. L.; Shami, A. N.; Zheng, X.; Marini, S.; Moritz, L.; Sultan, C.; Gurczynski, S. J.; Moore, B. B.; Tallquist, M. D.; Li, J. Z.; Hammoud, S. S., A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Dev Cell* **2018,** *46* (5), 651-667 e10.

Accepted manuscript

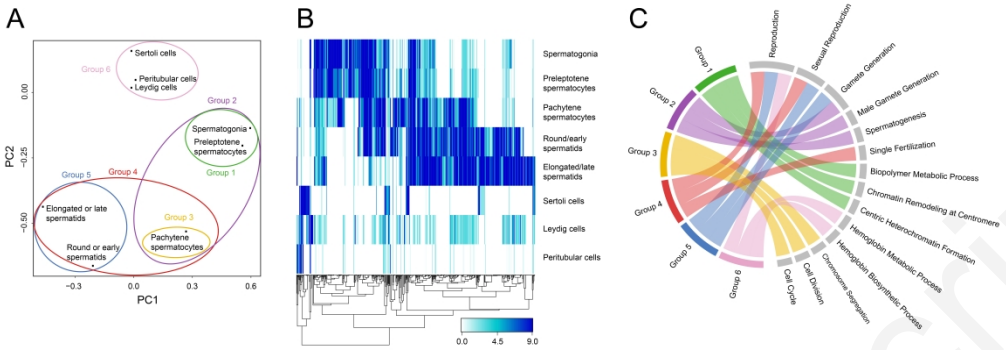


Figure 1. Categorization of testis elevated genes based on cell type-specific expression.
A. PCA plot showing the association between the eight cell types based on overall protein expression patterns in testis. B. Visualization of the protein expression profiles in a heatmap, showing the protein expression levels in the eight cell types based on 512 proteins which are clustered based on their pairwise Spearman correlation. The PCA plot and heatmap allowed us to highlight six expression groups: 1) premeiotic; 2) premeiotic + meiotic; 3) meiotic; 4) meiotic + postmeiotic; 5) postmeiotic; and 6) somatic.
C. Visualization of enriched GO terms associated with each of the six expression groups, highlighting the top three most significant GO terms for each group.

Accepted manuscript

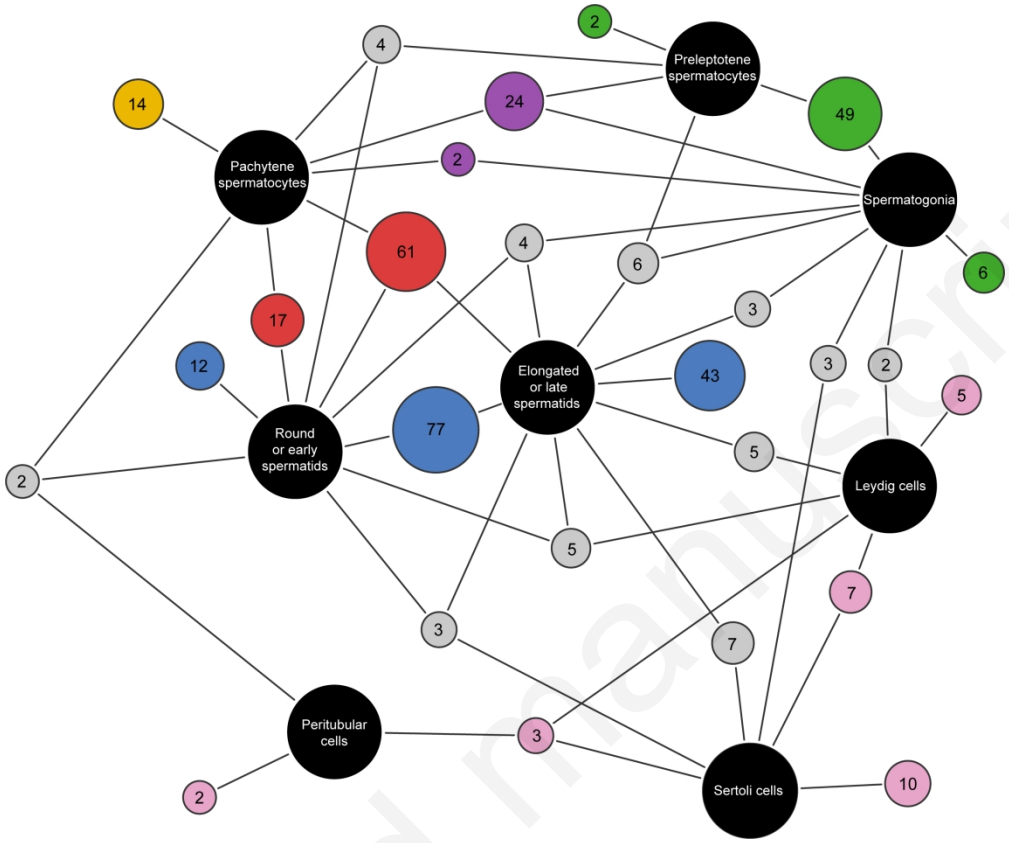


Figure 2. Relationship of protein expression between different testicular cell types.
The visualization shows the relationship of protein expression between different testicular cell types (black nodes), highlighting six main expression groups: premeiotic (green nodes), premeiotic + meiotic (purple nodes), meiotic (yellow nodes), meiotic + postmeiotic (red nodes), postmeiotic (blue nodes), somatic (pink nodes), as well as remaining proteins not belonging to a specific expression group (grey nodes).

Accepted manuscript

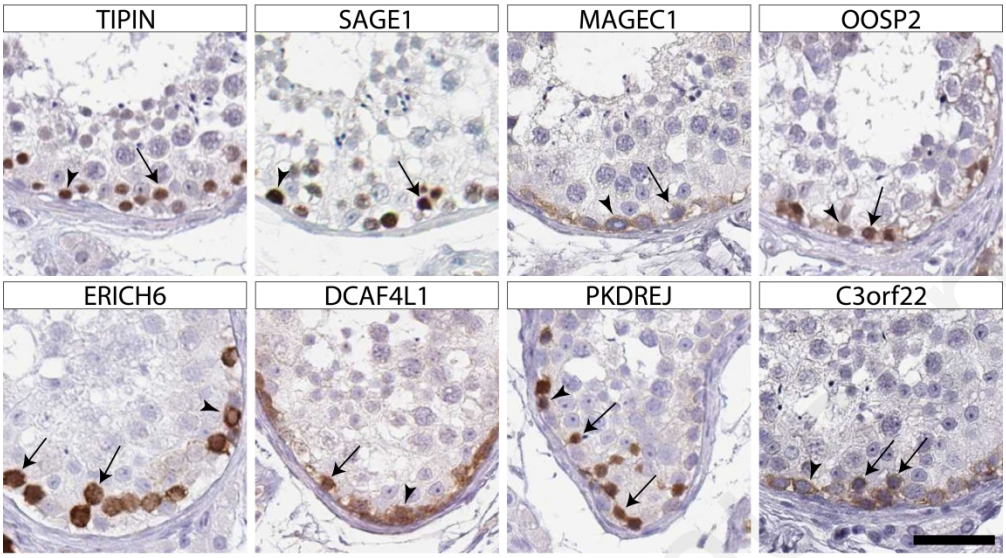


Figure 3. Immunohistochemical staining pattern of proteins specific to premeiotic germ cells. Transverse sections of adult human testis showing protein localization in spermatogonia (arrowheads) and preleptotene spermatocytes (arrows). TIPIN, SAGE1, ERICH6 and PKDREJ showed moderate to strong nuclear staining in both spermatogonia and preleptotene spermatocytes, while MAGEC1, OOSP2, DCAF4L1 and C3orf22 displayed cytoplasmic positivity in spermatogonia and preleptotene spermatocyes. Bar=50 µm.

Accepted manuscript

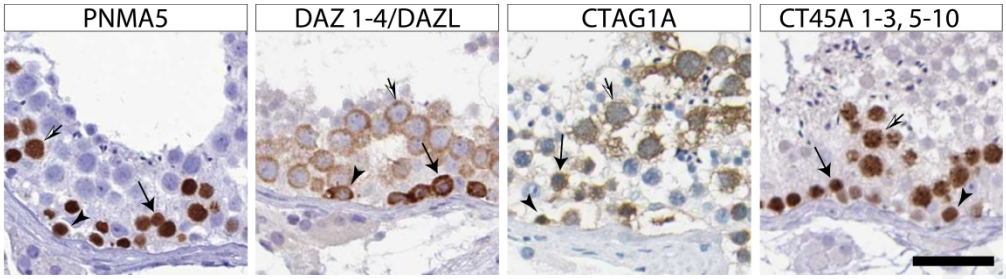


Figure 4. Immunohistochemical staining of proteins specific to premeiotic and meiotic germ cells. Transverse sections of adult human testis showing protein localization in spermatogonia (arrowheads), preleptotene spermatocytes (arrows) and pachytene spermatocytes (white arrows). PNMA5 staining showed a dense nuclear positivity in spermatogonia, preleptotene spermatocytes, and a limited number of pachytene spermatocytes depending on the stage the seminiferous cycle. DAZL/DAZ1-4 immunoreactivity was observed in the cytoplasm of spermatogonia, preleptotene spermatocytes and to a lesser extent in pachytene spermatocytes. CTAG1A staining was mainly distributed in cytoplasm of spermatogonia, preleptotene and pachytene spermatocytes, while CT45A1-3, 5-10 showed nuclear positivity. Bar=50 μm

Accepted manuscript

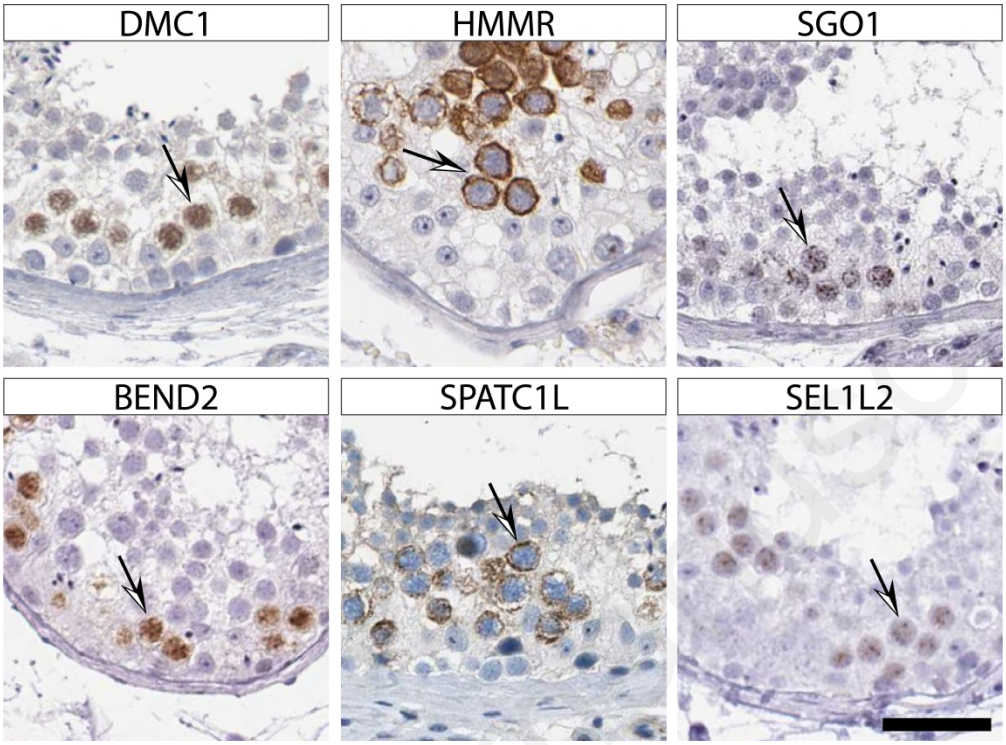


Figure 5. Immunohistochemical staining of proteins specific to meiotic germ cells.
Transverse sections of adult human testis showing protein localization in pachytene spermatocytes (white arrows). For all proteins, immunoreactivity was observed at moderate or strong intensity exclusively in pachytene spermatocytes, with a various subcellular localization pattern. DMC1 and BEND2 showed nuclear staining whereas the staining for HMMR and SPATC1L was localized to the cytoplasm. Immunoreactivity for SGO1 and SEL1L2 displayed a speckled nuclear pattern. Bar=50 µm.

Accepted manuscript

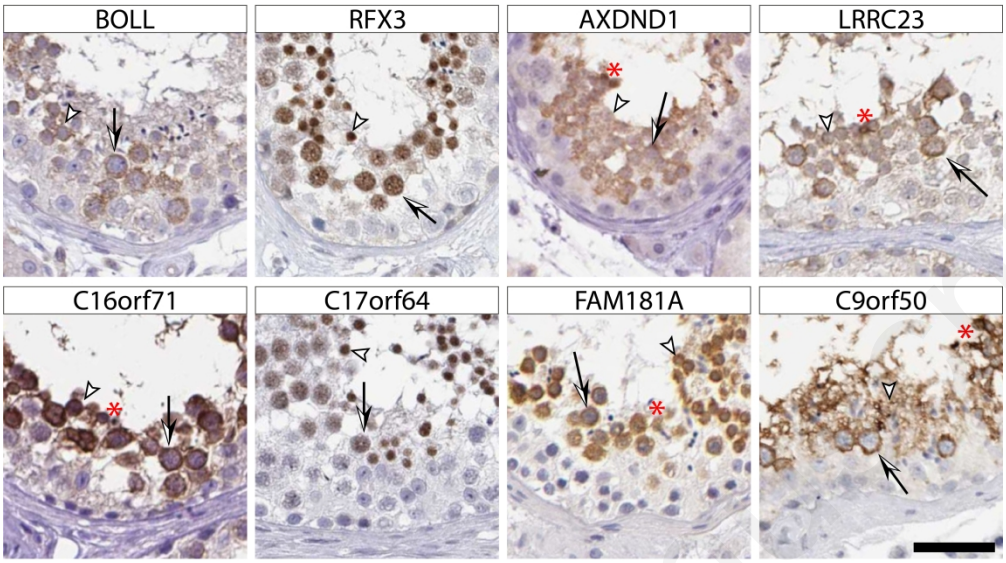


Figure 6. Immunohistochemical staining of proteins specific to meiotic and postmeiotic cells. Transverse sections of adult human testis showing protein localization in pachytene spermatocytes (white arrows), round/early spermatids (white arrowheads) and cytoplasmic lobes of elongated/late spermatids (red stars). BOLL showed cytoplasmic staining mainly in pachytene spermatocytes and round/early spermatids, and to a lesser extent in elongated spermatids. RFX3 and C17orf64 mainly displayed nuclear positivity in pachytene spermatocytes and round/early spermatids, while AXDND1, LRRC23, C16orf71, FAM181A and C9orf50 all predominantly showed a moderate to strong staining in pachytene spermatocytes, round/spermatids and cytoplasmic lobes of elongated/late spermatids. Bar=50μm.

Accepted Manuscript

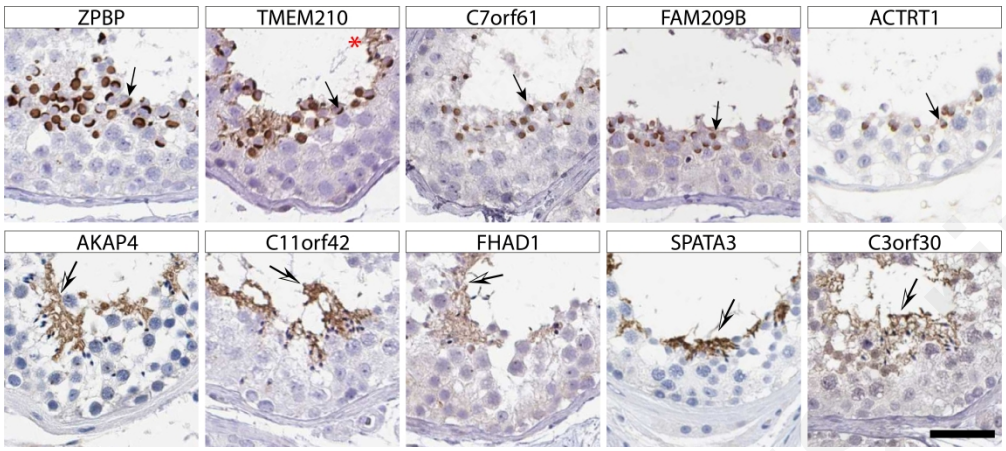


Figure 7. Immunohistochemical staining of proteins specific to postmeiotic cells.
Transverse sections of adult human testis showing protein localization in acrosome under formation elongated/late spermatids (black arrows), cytoplasmic lobes of elongated/late spermatids (red star) and flagella of elongated/late spermatids (white arrows). A concentrated signal was observed in the acrosome under formation for ZPBP, TMEM210, C7orf61, FAM209B and ACTRT1, with TMEM210 showing additional immunoreactivity in cytoplasmic lobes of elongated spermatids. AKAP4, C11orf42, FHAD1, SPATA3 and C3orf30 displayed an intense staining in flagella. Bar=50 µm.

Accepted manuscript

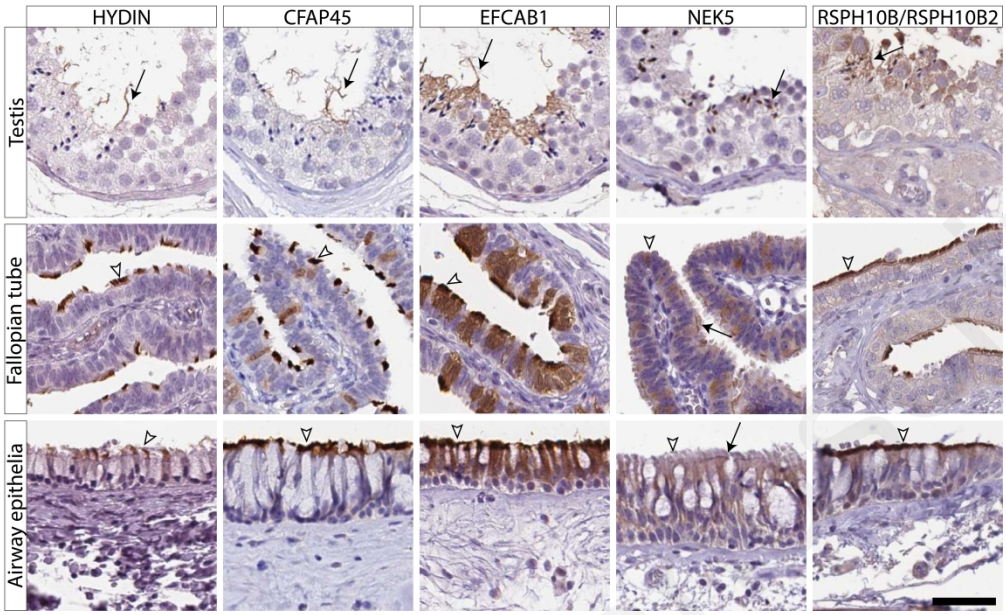
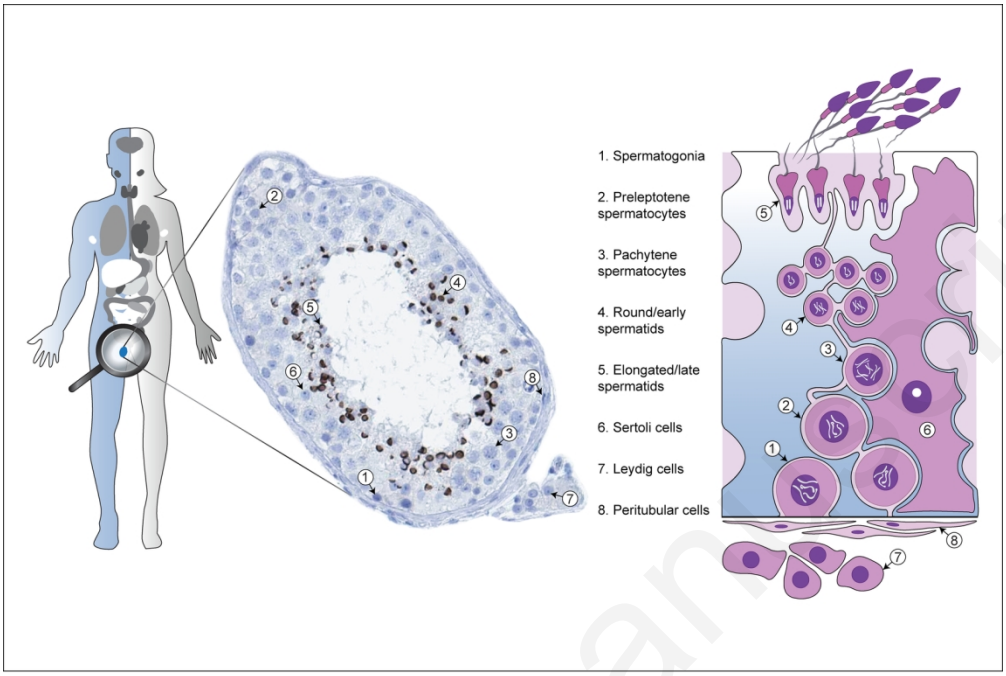


Figure 8. Immunohistochemical staining of proteins expressed in testis, Fallopian tube and airway epithelia. HYDIN and CFAP45 staining was restricted to sperm flagella (arrows) and cilia of Fallopian tube and airway epithelia (white arrowheads). EFCAB1 signal showed a similar pattern, with additional immunoreactivity in cytoplasmic components of all three tissues. NEK5 staining in testis was displayed in the head of elongated spermatids (arrow), and in airway epithelia most abundant in the cilial roots (arrow). RSPH10B/RSPH10B2 staining was distributed in both round/early and elongated/late spermatids in testis, and to cilia in Fallopian tube and respiratory epithelia. Note: Airway epithelial staining is represented by nasopharynx for HYDIN, CFAP45, EFCAB1 and NEK5, and bronchus for RSPH10B/RSPH10B2. Bar=50 μm.

Accepted Manuscript



For TOC only

For TOC only

76x60mm (600 x 600 DPI)

Accepted manuscript