Journal of Proteomics 267 (2022) 104686

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Comparative proteomics analysis of human FFPE testicular tissues reveals new candidate biomarkers for distinction among azoospermia types and subtypes

Katarina Davalieva [a,*], Aleksandar Rusevski [a], Milan Velkov [a], Predrag Noveski [a], Katerina Kubelka-Sabit [b], Vanja Filipovski [b], Toso Plaseski [c], Aleksandar Dimovski [a,d], Dijana Plaseska-Karanfilska [a,*]

[a] Research Centre for Genetic Engineering and Biotechnology "Georgi D Efremov", Macedonian Academy of Sciences and Arts, 1000 Skopje, North Macedonia
[b] Laboratory for Histopathology, Clinical Hospital "Sistina", 1000 Skopje, North Macedonia
[c] Faculty of Medicine, Endocrinology and Metabolic Disorders Clinic, 1000 Skopje, North Macedonia
[d] Faculty of Pharmacy, University "St. Cyril and Methodius", 1000 Skopje, North Macedonia

ARTICLE INFO

Keywords:
FFPE testicular tissue
Obstructive azoospermia
Hypospermatogenesis
Sertoli cell-only syndrome
Proteomics
Transcriptomics

ABSTRACT

Understanding molecular mechanisms that underpin azoospermia and discovery of biomarkers that could enable reliable, non-invasive diagnosis are highly needed. Using label-free data-independent LC-MS/MS acquisition coupled with ion mobility, we compared the FFPE testicular proteome of patients with obstructive (OA) and non-obstructive azoospermia (NOA) subtypes hypospermatogenesis (Hyp) and Sertoli cell-only syndrome (SCO). Out of 2044 proteins identified based on ≥2 peptides, 61 proteins had the power to quantitatively discriminate OA from NOA and 30 to quantitatively discriminate SCO from Hyp and OA. Among these, H1-6, RANBP1 and TKTL2 showed superior potential for quantitative discrimination among OA, Hyp and SCO. Integrin signaling pathway, adherens junction, planar cell polarity/convergent extension pathway and Dectin-1 mediated noncanonical NF-kB signaling were significantly associated with the proteins that could discriminate OA from NOA. Comparison with 2 transcriptome datasets revealed 278 and 55 co-differentially expressed proteins/genes with statistically significant positive correlation. Gene expression analysis by qPCR of 6 genes (H1-6, RANBP1, TKTL2, TKTL1, H2BC1, and ACTL7B) with the highest discriminatory power on protein level and the same regulation trend with transcriptomic datasets, confirmed the proteomics results. In summary, our results suggest some underlying pathways in azoospermia and broaden the range of potential novel candidates for diagnosis.

Significance: Using a comparative proteomics approach on testicular tissue we have identified several pathways associated with azoospermia and a number of testis-specific and germ cell-specific proteins that have the potential to pinpoint the type of spermatogenesis failure. Furthermore, comparison with transcriptomics datasets based on genome-wide gene expression analyses of human testis specimens from azoospermia patients identified proteins that could discriminate between obstructive and non-obstructive azoospermia subtypes on both protein and mRNA levels. Up to our knowledge, this is the first integrated comparative analysis of proteomics and transcriptomics data from testicular tissues. We believe that the data from our study contributes significantly to increase the knowledge of molecular mechanisms of azoospermia and pave the way for new investigations in regards to non-invasive diagnosis.

Abbreviations: ADH, alcohol dehydrogenase; CE, collision energy; FDR, false discovery rate; FFPE, formalin fixed and paraffin embedded; GO, gene ontology; HE, high energy; IAA, iodoacetamide; IMS, ion-mobility separation; LE, low energy; NOA, non-obstructive azoospermia; OA, obstructive azoospermia; PLGS, ProteinLynx Global Server; qPCR, quantitative polymerase chain reaction; SCO, Sertoli cell only; UDMSE, ultradefinition MSE; UPLC, ultra-high performance liquid chromatography.

* Corresponding authors at: Research Centre for Genetic Engineering and Biotechnology "Georgi D Efremov", Macedonian Academy of Sciences and Arts, Krste Misirkov 2, 1000 Skopje, North Macedonia.

E-mail addresses: katarina@manu.edu.mk (K. Davalieva), dijana@manu.edu.mk (D. Plaseska-Karanfilska).

https://doi.org/10.1016/j.jprot.2022.104686
Received 1 April 2022; Received in revised form 5 July 2022; Accepted 11 July 2022
Available online 30 July 2022

1. Introduction

Infertility, defined as the inability of a sexually active couple to conceive after one year of frequent unprotected sexual intercourse, represents a serious problem affecting at least 180 million people, or around 15% of reproductive age couples, worldwide [1]. Male infertility, defined as the inability of a male to impregnate a fertile female in one year of unprotected intercourse is solely responsible for 20–30% of infertility cases but is a contributing factor to >50% of overall infertility cases [2,3]. The causes of male infertility can be classified into several subgroups according to the etiology of the disease, such as endocrine disorders (2–5%), obstructive disorders (5%), primary testicular defects (65–80%), genetic causes (15%) and idiopathic (10–20%) [4].

The most severe form of male infertility is azoospermia, distinguished by an absence of sperm in the semen [5]. Azoospermia, with an incidence in infertile males of approximately 10 to 15%, can be classified into obstructive azoospermia (OA), caused by a physical obstruction in the male reproductive tract and non-obstructive azoospermia (NOA), which based on histopathological examination of testicular tissue can be further subdivided into hypospermatogenesis, maturation arrest (MA), and Sertoli cell-only (SCO) syndrome [6,7]. In hypospermatogenesis and MA, some sperm cells are still present, but their number is decreased or they do not fully mature while in SCO, the patients do not make sperm cells at all.

Achieving a reliable diagnosis of azoospermia requires histopathological analysis of testicular biopsy [7,8]. However, testicular biopsy is invasive, causes psychological discomfort to patients, such as anxiety and fear, and is associated with certain risks such as bleeding, infection, and in some cases, hypogonadism [8]. In addition, due to the spatial distribution of spermatogenesis in the testis, testicular biopsy in some cases might not accurately reflect the histopathology of NOA.

As a result of this, there is a high need for non-invasive testing that could provide a reliable diagnosis of the type/subtype of azoospermia. In patients with OA, the non-invasive test should provide an accurate diagnosis without the need for testicular biopsy, while in patients with NOA, the good diagnostic test should provide an accurate diagnosis of histopathological subtypes, predict the success of testicular sperm extraction and facilitate better planning for assisted reproduction.

The advent of the "omics" techniques enabled high-throughput analysis of genes, proteins, and metabolites, together with their expression levels and opened a new path toward the non-invasive diagnosis of male infertility. In this context, comparative analysis of samples from well-defined groups of patients with male infertility using high-throughput proteomics technologies, has shown great potential for the discovery of new, non-invasive testing. Proteomics studies of male infertility, among which azoospermia, record a significant increase in the last 20 years [9]. In terms of human samples used in the proteomics studies of azoospermia, testicular tissue, seminal plasma, blood, and extracellular vesicles have been analyzed. Although testicular tissue analysis could provide the most accurate results in understanding the spermatogenesis impairment, so far there is a very modest number of research studies using this sample, probably due to the limited availability (extensively reviewed in [10]). The highest number of studies focusing on biomarkers for azoospermia was done on seminal plasma. The non-invasive sampling that makes this sample type relatively easily accessible and the presence of a high number of seminal proteins allowed its extensive use and led to subsequent identification of some of the potential biomarker candidates in azoospermia (extensively reviewed in [11–16]).

Overall, the proteomics research so far has provided some potential azoospermia biomarkers. Proteins such as inhibin B (INHBB) [17–19], Anti-Müllerian hormone (AMH) [18–20], Cysteine Rich Secretory Protein 1 (CRISP1) [21,22], Stabilin 2 (STAB2) [23], 135-kd centrosomal protein (CP135) [23], Ras Protein Specific Guanine Nucleotide Releasing Factor 1 (RASGRF1) [23], Prolactin-Inducible Protein (PIP) [23–25], Prostatic Acid Phosphatase (ACPP) [25], Prostaglandin D2 Synthase (PTGDS) [26–29], Transketolase-Like Protein 1 (TKTL1) [30], Phosphoglycerate Kinase 2 (PGK2) [30], Lactate Dehydrogenase C (LDHC) [26,30,31], epididymis-expressed Extracellular Matrix Protein 1 (ECM1) [28,31,32] and Testis Expressed 101 (TEX101) [28,31,32] have been among the most prominent biomarkers that could facilitate the differential diagnosis of azoospermia. Of these, so far only TEX101 has been extensively validated, leading to the development of ELISA test for the measurement of its native form in biological fluids [33]. Clinical utility of TEX101 in combination with ECM1 to stratify azoospermia forms is currently at the stage of preclinical evaluation [34].

Even though some of the above biomarkers such as ECM1 and TEX101, already deliver good diagnostic sensitivities and specificities for azoospermia, there is still a need for novel biomarkers to differentiate between hypospermatogenesis, MA, and SCO and to predict testicular sperm extraction outcome in NOA patients with an even higher diagnostic sensitivity and specificity. The key point here is the identification of testis-specific and germ cell-specific proteins that are secreted into semen or even detected in blood or urine, which could accurately pinpoint the stage of spermatogenesis failure. The best way to understand the process of spermatogenetic impairment on a molecular level, and to identify the genes/proteins involved, is through the analysis of testicular tissue. In this study, formalin-fixed paraffin-embedded (FFPE) testicular tissues from men with azoospermia and histopathological diagnosis of OA, hypospermatogenesis and SCO were analyzed using label-free LC-MS/MS protein profiling coupled with ion-mobility separation. Proteins with differential abundance among the studied groups were discussed based on their protein localization in testis, enriched gene ontology (GO) and pathways, and ultimately correlated with available transcriptomics data. Some of the most promising biomarkers for quantitative discrimination among the studied types of azoospermia were further validated by qPCR. Presumably, this is the first integrated comparative analysis of proteomics and transcriptomics data from testicular tissues which provides deep insights into the proteins involved in spermatogenesis failure and gives a number of potential candidates for discrimination of OA and NOA (hypospermatogenesis and SCO) on both protein and mRNA level.

2. Material and methods

2.1. Study design and tissue sample selection

The samples used in this study were formalin-fixed paraffin-embedded (FFPE) testicular tissues obtained by biopsy from men with clinical diagnosis of azoospermia verified at the Clinical Hospital Acibadem Sistina, Skopje in the period 2012–2018. The azoospermic patients were preselected to exclude genetic factors closely related to infertility such as Klinefelter syndrome, chromosomal abnormalities of X/Y chromosome, cystic fibrosis and deletions in the AZF region of the Y chromosome. The patients were divided into 3 groups according to the histopathological report: spermatogenesis (Obstructive Azoospermia (OA)), hypospermatogenesis and germinal cell aplasia (Sertoli cell only syndrome (SCO)). For this study we have used a total of 76 FFPE testicular tissues. Detailed histopathology evaluation of these samples is given in Supplementary Table 1. For the discovery study we have used 27 samples in total, or 9 patients per group. Patients were aged 31–46 years with no differences among groups regarding age (Median age $_{\text{Hypospermatogenesis}}$ = 35; Median age $_{\text{SCO}}$ = 34 and Median age $_{\text{Spermatogenesis}}$ = 39). Validation cohort consisted of 49 FFPE testicular tissues from 19 patients with Hypospermatogenesis, 10 patients with SCO and 20 patients with OA. Patients in the validation cohort were in the same age group as the patients in the discovery proteomics group (Median age $_{\text{Hypospermatogenesis}}$ = 32; Median age $_{\text{SCO}}$ = 36 and Median age $_{\text{Spermatogenesis}}$ = 39). Informed consent for the use of these tissues for research purposes in accordance with the Declaration of Helsinki was obtained from the patients. The study has been approved by the Ethics Committee of the Macedonian Academy of Sciences and Arts (09-1785/3 from

09.06.2017).

2.2. Proteomics analysis

2.2.1. Protein extraction from FFPE tissue blocks

Proteins from FFPE tissues were isolated using 3–4 serial 10 μm thick sections. Tissues were deparaffinized in three changes with xylene for 5 min followed by rehydration with a graded series of ethanol (95%, 70% and 50%) and water for 5 min. After air drying for 30 min at room temperature, the tissues were weighed and resuspended in 1:20 ratio (*w/v*) of Lysis buffer (4% SDS, 5 mM $MgCl_2x6H_2O$, 10 mM CHAPS, 100 mM NH_4HCO_3, 0.5 M DTT). The samples were vortexed, sonicated in an ice bath for 30 min, and incubated at 95°C for 30 min. The samples were then cooled on ice for 5 min, vortexed and incubated at 80°C for 2 h with mixing at 1000 rpm. After cooling on ice for 5 min and sonication for 15 min, the protein content was quantified by Bradford method [35] in duplicate against a standard curve of bovine serum albumin (BSA) and stored at −80°C until use.

2.2.2. Sample preparation using RapiGest

The RapiGest [36] protocol by Oswald et al., [37] was used with some modifications. Briefly, a portion of each individual sample in Lysis buffer containing 100 μg of protein was adjusted to 100 μl with Lysis buffer and 150 μl of methanol and 38 μl of chloroform were added. After vortexing vigorously for about 1 min and centrifuging at 5000 x g for 5 min, most of the upper layer of methanol/water was aspirated and 112 μl of methanol was added, mixed and centrifuged at 16,000 x*g* for 5 min to pellet the protein. Proteins were dissolved in 50 mM ammonium bicarbonate containing 0.1% RapiGestTM detergent (Waters Corp.) in ratio 2.5:1 (*w/v*). DTT was added (0.12 μmol/50 μg protein) and the solution was sonicated and boiled for 5 min. Protein concentration was determined using the Bradford assay. The volume of each sample containing a total of 30 μg protein was adjusted to 35 μl with 0.1% RapiGest in 50 mM NH4HCO3, heated at 80°C for 15 min, reduced in 5 mM DTT for 30 min at 60°C and alkylated in 15 mM iodoacetamide (IAA) in the dark for 30 min at room temperature. Trypsin (TRYPSEQM-RO ROCHE) was added at a 1:100 trypsin:protein ratio by mass and incubated overnight at 37C. Following the digestion, 5% TFA was added to final concentration of 0.5% to hydrolyze the RapiGest and samples were incubated at 37°C for 90 min. After centrifugation at 14,000 rpm, at 6°C for 30 min, supernatants were transferred into the Waters Total Recovery vial, diluted with water to 0.4 μg/μl protein and equal volume of 50 fmol/μl of a digest of yeast alcohol dehydrogenase (ADH) in 5% ACN, 0.1% formic acid (FA) was added as an internal standard protein. The final concentration of protein in the pooled samples was 200 ng/μl and the final concentration of ADH was 25 fmol/μl.

2.2.3. Label-free nano-LC-MS/MS

A label-free LC-MS/MS protein profiling was performed using an ultra-performance liquid chromatography system ACQUITY UPLC® M-Class (Waters Corporation) coupled with SYNAPT G2-Si High Definition Mass Spectrometer (Waters Corporation) equipped with a T-Wave-IMS device. Data were obtained using ion-mobility separation (IMS) enhanced MS^E acquisition named ultradefinition MS^E ($UDMS^E$) [38].

Optimal loading for $UDMS^E$ runs was determined by testing pool sample (containing an equal amount of each 27 individual samples), starting from 100 to 300 ng per run and processing in ProteinLynx Global SERVER (PLGS, version 3.0.3, Waters Corp., Milford, MA, USA). Then, one test run and initial data processing with PLGS was done for quality assurance testing and determination of the exact protein concentration for each individual sample, followed by runs at the determined optimal load of 300 ng per run.

Mass spectrometric analysis of tryptic peptides was performed as previously described in detail [39]. Briefly, peptides were trapped on an ACQUITY UPLC M-Class Trap column Symmetry C18, 5 μm particles, 180 μm × 20 mm, (Waters Corporation), followed by separation on ACQUITY UPLC M-Class reverse phase C18 column HSS T3, 1.8 μm, 75 μm × 250 mm (Waters Corporation) at a flow rate of 300 nl/min using 90 min multistep concave gradient [40]. Lock mass compound Glu-1-Fibrinopeptide B (EGVNDNEEGFFSAR) with concentration of 100 fmol/μL was delivered by the auxiliary pump of the LC system at 500 nl/min and the Lock mass spectrum of doubly charged Glu-1-Fibrinopeptide B (*m/z* 785.8426) was produced every 45 s. For all MS measurements, spectra were recorded in resolution positive ion mode with a typical resolving power of at least 25,000 FWHM (full width at half maximum) and sensitivity of >7000 TDC equivalent counts/s for the double charged Glu-1-Fibrinopeptide B. Source settings included capillary voltage of 3.2 kV, extraction cone at 4 V, sampling cone at 35 V, and source temperature of 80°C. The cone gas N_2 flow was 30 L/h. Analyzer settings included quadrupole profile set at auto with mass 1 as 1.25 Ma (dwell time 25% and ramp time 75%) and mass 2 as 0.17 Mb. The Step Wave settings in TOF mobility acquisition mode were the following: wave velocity of 300 m/s and wave height 15 V, 15 V and 1 V for the StepWave 1 and StepWave 2 and Source Ion Guide, respectively. For IMS, wave height of 40 V was set. Traveling wave velocity was ramped from 900 m/s to 450 m/s over the full IMS cycle. Wave velocities in the trap and transfer cell were set to 311 m/s and 175 m/s, respectively and wave heights to 4 V. Spectra were collected over the mass range 50–2000 *m/z* with a scan time of 0.5 s. For the UDMSE acquisition, the collision energy was held at 0 V for low energy scan while for the high energy cycles, a look-up table file was put into the MS method to optimize precursor fragmentation in the transfer cell according to the procedure of Distiler et al., [38]. The following collision energy (CE) settings were applied throughout the whole study in the elevated energy scan: (i) ion-mobility bins 0–20: CE of 2 eV, (ii) ion-mobility bins 21–120: CE ramp from 10.6 eV to 50.4 eV, (iii) ion-mobility bins 121–200: CE ramp from 51 eV to 60 eV.

2.2.4. LC-MS/MS data processing and identification

Test runs were analyzed using PLGS (Waters Corporation). The data were post-acquisition lock mass corrected using the doubly charged monoisotopic ion of [Glu1]-Fibrinopeptide B. Data was searched against the UniProtKB/ Swiss-Prot database containing 20,370 proteins (June 2020), to which yeast alcohol dehydrogenase (UniProt P00330) sequence was added. Optimized low energy (LE) and high energy (HE) threshold settings for PLGS data processing were determined by PLGS Threshold Inspector (Version 2.3 Build 2, Carper Soft) as 250 counts and 30 counts for LE and HE thresholds, respectively. Precursor and fragment ion mass tolerances were automatically determined by PLGS during database searching. The typical range of RMS error for precursor and product ions for were ± 5 and ± 10 ppm, respectively. Search settings included up to two missed cleavages, carbamidomethyl C as a fixed modification and oxidized M as a variable modification. A minimum of two fragment ion matches was required per peptide identification and five fragment ion matches per protein identification, with at least one peptide match per protein identification. The protein false discovery rate (FDR) was set to 1% threshold for database search in PLGS. The inputs used for quantification measurement were: internal standard protein, P00330; protein concentration: 25 fmol/μl.

Comparative proteomics analysis was done using Progenesis QIP version 4.1 (Nonlinear dynamics, Waters Corporation). Raw profile data (.raw) of the 27 samples were imported into Progenesis QIP with LE and HE threshold set to auto and data lock mass corrected. Imported runs were automatically aligned to the most suitable reference run identified by the software. Then, normalization of the protein abundances was done using the default method “Normalize to all proteins” by which protein amounts in individual runs are normalized to one run automatically selected as the normalization reference. Using the PLGS database management tool, the sequence of each protein entry from the UniProtKB/ Swiss-Prot database containing 20,370 proteins was reversed to generate a decoy database, which was appended to the original fasta file. This combined target decoy database was used for

database search with the following parameters: digest reagent – trypsin; maximum missed cleavages – two; maximum protein mass – 250 kDa; fixed modifications – carbamidomethyl C; variable modification – oxidation M. Search tolerance parameters were as follows: peptide tolerance – auto; fragment tolerance – auto; FDR < 1%. Ion matching requirements were the same as in PLGS processing. Peptides with a sequence length of less than six amino acids and a score below 4 were removed. Grouping of similar proteins and relative quantitation from non-conflicting peptides were used for protein building. Proteins were reviewed and exported in the form of .csv output file for subsequent data analysis. As Progenesis QIP estimates FDR for each LC-MS run separately, target decoy database enabled the calculation of the FDR on the whole dataset level which was 4.8%.

The mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium via the PRIDE [41] partner repository with the dataset identifier PXD032722 and DOI: https://doi.org/10.6019/PXD032722.

2.2.5. Proteomics data analysis

Proteins that were considered as differentially abundant were selected based on two or more peptide matches per identification, Anova ≤0.05 and fold change ≥1.5. Furthermore, *p*-values between groups were calculated using Mann–Whitney *U* test and subsequently corrected for multiple testing using Benjamini-Hochberg procedure [42]. Differentially abundant proteins between groups were considered with Benjamini-Hochberg corrected *p*-values ≤0.05 and fold change ≥1.5. Correlation of the quantitative proteomics data with specific histopathological features such as interstitial fibrosis, tubular hyalinization and inflammation were done by Mann–Whitney U test with *p*-values ≤0.05 considered as significant. For each histopathological feature the samples were divided into two groups: samples that presented the specific feature and samples that did not. We compared the differences between the two independent groups within each histopathological feature with the dependent variable being the proteomic data. In addition we have calculated the ratio between the groups with and without the specific histopathological feature to evaluate in which direction the protein abundance changed (up- or down-regulated). Statistical analyses were performed using XLSTAT software ver. 2022.1.2 [43].

Graphical representation of the number of identified proteins was done using Venn Diagram Plotter (http://omics.pnl.gov/software/VennDiagramPlotter.php). Differentially abundant proteins were subjected to functional annotation and enrichment analysis using publicly available annotation tools and databases such as UniProtKB and Panther (Protein ANalysis THrough Evolutionary Relationships) Classification System version 16.0 [44]. Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) analysis was performed using the online tool (version 11.5) to identify protein-protein interaction networks [45] with the following settings: (1) full STRING network; (2) evidence setting; (3) all active interaction sources; (4) medium confidence score and (5) max number of interactors to show, for the 1st shell-none/query proteins only, and for the 2nd shell-none. Pathway (KEGG, Reactome, Wiki Pathways) and gene ontology (GO) enrichment analysis was visualized using ClueGO, a Cytoscape plug-in that visualizes the non-redundant biological terms for large clusters of genes in a functionally grouped network [46]. The tissue specificity and distribution of the selected proteins with differential abundance in the human testis was evaluated based on the mRNA and protein expression data from Human Protein Atlas version 21.0 [47,48]. Protein localization based on antibody staining (immunohistochemistry, Western blot) was available for two cell types in the testis (cells in seminiferous ducts and Leydig cells) for all retrieved proteins but for some, this data was extended to additional 7 cell types within the testis [49].

2.3. Transcriptomic analysis

2.3.1. Selection of microarray data

Two publicly available microarray datasets GSE145467 and GSE9210 were obtained from the Gene Expression Omnibus database [50]. These datasets were selected based on the following criteria: 1) human testis specimens; 2) genome-wide gene expression analyses; 3) analysis of cases with azoospermia and presence of corresponding groups as in our study (obstructive (OA) and non-obstructive azoospermia (NOA)); and 4) available processed data. The exclusion criteria included: 1) animal studies; 2) datasets from cell cultures (treated or untreated) and 3) datasets with a small number of samples. The GSE145467 dataset included 20 samples (10 with OA and 10 samples with NOA), while the GSE9210 dataset included 11 OA and 47 NOA samples.

2.3.2. Microarray data analysis

Data was analyzed using GEO2R, an interactive web tool that performs comparisons on original submitter-supplied processed data tables using the GEOquery and limma R packages from the Bioconductor project. Results were presented as a table of genes ordered by significance together with multiple-testing corrections on *p*-values and log2FC (OA vs NOA). Further, the transcripts were collapsed to unique genes and the mean expression of transcripts from the same gene locus were calculated. The differentially expressed genes were selected based on the same criteria as proteins (Benjamini-Hochberg corrected *p*-values ≤0.05 and fold change ≥1.5 ($\log_2$FC ≥ 0.58)). The interactive tool for comparing lists with Venn's diagrams [51] was used to overlap the lists of differentially expressed genes from the two different gene expression datasets with our proteomics dataset. For correlation purposes, protein ratios from group comparisons were log2 transformed to match the format of the transcriptomics data. Correlation between datasets was done using Spearman correlation.

2.4. Gene expression analysis by quantitative real-time PCR (qPCR)

2.4.1. Primer/probe design

The mRNA levels of six differentially expressed proteins, Histone H1t (H1-6), Ran-specific GTPase-activating protein (RANBP1), Transketolase-like protein 2 (TKTL2), Transketolase-like protein 1 (TKTL1), Histone H2B type 1-A (H2BC1) and Actin-like protein 7B (ACTL7B) were assayed. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a reference in order to normalize gene expression levels. qPCR assay for the selected genes and GAPDH was based on the following mRNA transcript sequences from the NCBI Gene database: NM_005323.4 (H1-6), NM_001278639.2 (RANBP1), NM_032136.5 (TKTL2), NM_012253.4 (TKTL1), NM_170610.3 (H2BC1), NM_006686.4 (ACTL7B) and NM_002046.7 (GAPDH). Primers/probe sets were designed using PrimerQuest Tool (Integrated DNA Technologies, Inc., USA). Genes and primer/probe sets with details are given in Supplementary Table 2.

2.4.2. RNA isolation

Total RNA was isolated from the 49 FFPE testicular tissues from the validation cohort using ALLPrep DNA/RNA FFPE Kit (Qiagen, Germany) according to the manufacturer's instructions. DNase digestion was done to remove DNA contaminants and RNA was eluted from the column using 30 μl of RNase-free water.

2.4.3. qPCR assay and analysis

Gene expression assay was done by qPCR, using Luna® Universal One-Step RT-qPCR Kit (New England Biolabs, USA) and designed primers/probe in the total volume of 20 μl containing 1× Luna Universal One-Step Reaction Mix, 1× Luna WarmStart RT Enzyme Mix, 400 nM each forward and reverse primer, 200 nM probe and 3 μl RNA. The assay for each gene was done in a reaction containing primers/probe for the

selected gene together with the primer/probe set for GAPDH as a normalization reference. The assay was performed on the Applied Biosystems® 7500 Real-Time PCR Systems (Life Technologies) under the following cycling conditions: reverse transcription step at 55°C for 15 min, initial denaturation step at 95°C for 1 min, followed by 40 cycles of denaturation at 95°C for 10 s and annealing/extension at 60°C for 1 min. Changes in gene expression of the selected genes between hypospermatogenesis, SCO and OA groups were determined by the $2_T^{-\Delta\Delta C}$ method [52] using GAPDH as an internal reference. Differences in gene expression among groups were tested using Mann–Whitney *U* test with *p*-values ≤0.05 considered as significant.

3. Results

3.1. Overview of the proteomics data

Tissue proteomics profiling data were acquired from 27 FFPE testicular tissue samples divided into 3 groups: hypospermatogenesis, Sertoli cell only syndrome (SCO) and obstructive azoospermia (OA). A comparable number of proteins was identified in Hyp and OA groups (Mann–Whitney U test, $p < 0.190$), with an average of 1076 and 1208 identified proteins, respectively (Fig. 1A). In SOC group there was an average of 839 identified proteins, significantly lower compared to both Hyp and OA groups (Mann–Whitney U test, SCO vs Hyp: $p < 0.014$; SCO vs OA: $p < 0.0001$). Comparative analysis with Progenesis QIP resulted in the identification of 2222 proteins with quantitative values based on 30,704 peptides. The hierarchical clustering of the identified proteomes with normalized protein abundance by Progenesis QIP showed highly similar protein profiles within OA and SCO groups respectively, while proteome profiles from Hyp group were more heterogeneous and showed similarity with either OA or Hyp group (Fig. 1B). The correlation of normalized protein abundance across individual samples was very high with a median Spearman Rho correlation coefficient of 0.904 (Fig. 1C).

3.2. Proteins with differential abundance among groups

After filtering this dataset to remove reverse sequences ($n = 107$), proteins identified on only one peptide ($n = 56$), proteins whose existence is unsure (PE = 5 based on UniprotKB, $n = 14$) and yeast ADH, the final report contained a total of 2044 proteins identified based on ≥2 peptides (Supplementary Table 3). A statistically significant difference in protein abundance among 3 groups (Anova ≤0.05) showed 1179 proteins, out of which, 958 proteins had fold change >1.5. Intergroup comparison revealed 619, 1256 and 520 proteins with significant differential abundance (Mann–Whitney *U* test, $p \leq 0.05$) in Hyp vs OA, SCO vs OA and SCO vs Hyp comparisons, respectively (Table 1). After adjusting the *p*-values for multiple testing using Benjamini-Hochberg procedure and filtering the dataset for B-H $p \leq 0.05$ and fold change ≥1.5, the list of proteins with differential abundance came to 795 proteins (Supplementary Table 4). Out of these, 72, 779 and 35 had significant differential abundance with fold change ≥1.5 in Hyp vs OA, SCO vs OA and SCO vs Hyp comparisons, respectively. Protein differences between OA, SCO vs Hyp groups represented by Volcano plots are given in Fig. S1, Supplementary document. More than 2/3 of the proteins with significant differential abundance in SCO vs OA were up-regulated, while in SCO vs Hyp majority were down-regulated. In Hyp vs OA comparison the number of up- and down-regulated proteins were similar. Correlation with specific histopathological features revealed 96 proteins associated with interstitial fibrosis, 199 proteins associated with tubular hyalinization and 33 proteins associated with inflammation (Supplementary Table 4).

To gain insight into the cell/tissue origin and biological implication of the proteins with significantly altered abundance among groups, we have analyzed the protein localization in testis and the reported molecular functions, cellular localization, and involvement in biological processes. Testis specificity and distribution were analyzed based on transcript detection, while the level of protein detected in the testis was antibody-based, according to the data available from Human Protein Atlas (HPA). Distribution based on transcript detection (nTPM≥1) in testis (Fig. 2A) showed that 91% of the differentially abundant proteins are expressed in the testis, 7% are not detected in testis and for 2% no data vas available. Within the proteins expressed in testis, 1% is only detected in testis, 8% is detected in some including testis (more than one but less than one-third of all tissues), 19% are detected in many including testis (at least a third but not all tissues) and 63% are ubiquitously expressed. Transcript specificity in regards to testis tissue showed that 90% of the proteins are testis-specific, with 11% with elevated expression in the testis compared to other tissues, 29% elevated in other but expressed in testis and 50% with low tissue specificity but expressed in testis. Out of the 11% of proteins with elevated expression in the testis 47% were tissue enriched (at least four-fold higher mRNA level in testis compared to any other tissues), 23% were group enriched (at least four-fold higher average mRNA level in a group of 2–5 tissues compared to any other tissue) and 30% were tissue enhanced (at least four-fold higher mRNA level in testis compared to the average level in all other tissues). HPA data based on antibody-based protein profiling showed that with exception of the 11% not detected in testis and 13% not found in the database, the remaining are detected on protein level with 68% having a high or medium level of protein expression in testis.

The top represented molecular functions were binding, catalytic activity and structural molecule activity, with proteins that discriminate OA from NOA having further transcriptional/translational regulator activity, ATP-dependent activity and molecular adaptor activity (Fig. 2B). While the top represented biological functions were cellular and metabolic processes, biological regulation and localization, some of the differentially expressed proteins that differentiate SCO from Hyp/OA are involved specifically in reproduction. The differentially expressed proteins belong in general to protein modifying enzymes, metabolite interconversion enzymes and RNA metabolism proteins. Further, proteins that discriminate OA from NOA belong to extracellular matrix proteins, calcium-binding proteins, membrane traffic proteins, chaperones and transmembrane signal receptors. Our set of proteins with statistically significant differential abundance was linked with approximately 100 pathways according to the Panther database. Among them, integrin signaling pathway was the top pathway associated with differentially expressed proteins in all 3 group comparisons, followed by inflammation mediated by cytokine, cytoskeletal regulation by Rho GTPase and Wnt signaling pathway. Proteins that discriminate OA from NOA participate also in several additional pathways among which, angiogenesis, FGF signaling, EGF receptor signaling, cadherin signaling and T cell activation pathways. Biomarkers that discriminate SCO from Hyp have the most prominent participation in cytoskeletal regulation by Rho GTPase, nicotinic acetylcholine receptor signaling and inflammation mediated by cytokine.

3.3. Bioinformatics analysis of the candidate proteomics biomarkers

Analysis in terms of overlap among Hyp vs OA, SCO vs OA and Hyp vs SCO comparisons, showed 3 proteins with significantly altered abundance in all comparisons, 27 proteins that were mutual for SCO vs Hyp and SCO vs OA and 58 proteins that were mutual for Hyp vs OA and SCO vs OA (Fig. 3A). The remaining proteins were found exclusively in only one of the group comparisons. As one of the aims of this study was to find potential biomarkers that could discriminate hypospermatogenesis, SCO syndrome and obstructive azoospermia we further focused only on these 88 proteins common for more than one comparison (Supplementary Table 4**, rows 2–89**). Three proteins that could quantitatively discriminate between Hyp, OA and SCO were H1-6, RANBP1 and TKTL2. All of them showed the lowest abundance in SCO and highest in OA, with Hyp being in-between (Fig. 3B). According to HPA data, TKTL2 is only detected in testis and has elevated mRNA expression

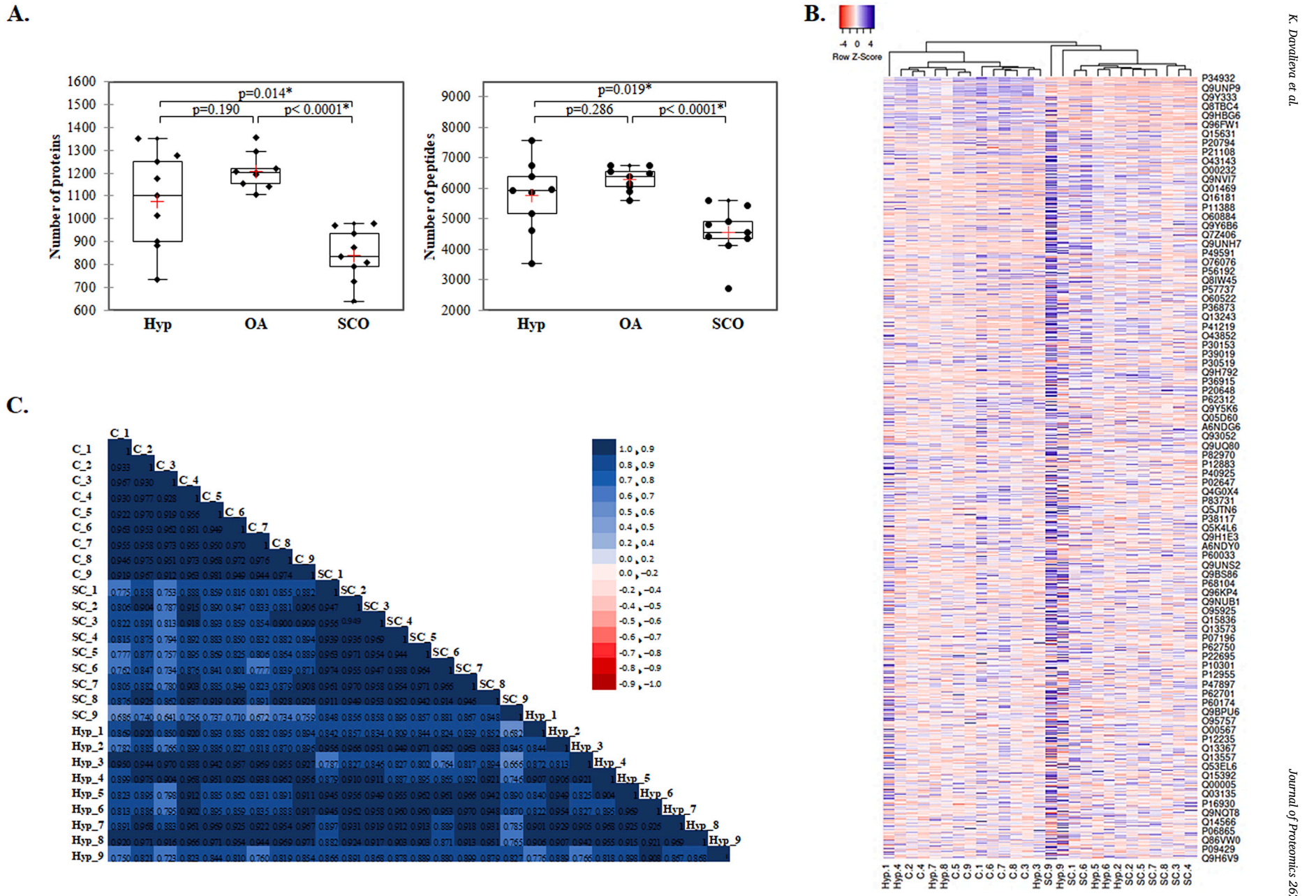

Fig. 1. Overview of the proteomics data. (**A**) Number of proteins and peptides identified in each of the runs within the experiment. Runs are grouped together by experimental condition. In the combined dot/box plot graphs, median (−), 25th and 75th percentiles and mean (+) are shown. (**B**) Heatmap of the normalized protein abundances of 2222 proteins identified and quantified by Progenesis QIP. Higher protein abundances are shown in blue and the lower ones in red. The samples are shown in columns and the rows indicate proteins. Clustering method applied: Average linkage; Distance Measurement method: Spearman rank correlation. (**C**) Graphical representation of the correlation matrix for normalized protein abundances across the individual samples: the Spearman Rho correlation coefficient is color coded. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Overview of the number of identified proteins with differential abundance between investigated groups and their regulation trend.

	Hyp/ OA	SCO/ OA	SCO/ Hyp
Differentially expressed (Mann–Whitney $p \leq 0.05$)	619	1256	520
Differentially expressed (adjusted for multiple testing B-H $p \leq 0.05$)	123	1188	71
Differentially expressed (B-H $p \leq 0.05$ + Fold change $\geq$1.5)	72	779	35
Up-regulated (B-H $p \leq 0.05$ + Fold change $\geq$1.5)	34	523	9
Down-regulated (B-H $p \leq 0.05$ + Fold change $\geq$1.5)	38	256	26

in testis compared to other tissues. H1-6 has also elevated mRNA expression in testis but has been detected in some other tissues, while RANBP1 is ubiquitously expressed protein with low tissue specificity but expressed in testis. In terms of protein detection, H1-6 and RANBP1 have high levels in testis while for TKTL2 there was no available information in HPA.

The 61 proteins that were mutual for Hyp vs OA and SCO vs OA comparisons had the power to quantitatively discriminate OA from NOA (Hyp and SCO) and these were named OA biomarkers. Thirty-three showed lower abundance in Hyp and SCO compared to OA and the remaining 28 had the opposite trend. Nine of them (H1-6, TKTL2, ASRGL1, ZMYND10, KRT86, MAGEB4, ACTL7B, CFAP58 and FATE1) are proteins with elevated mRNA expression in testis and three (TKTL2, ACTL7B and FATE1) are only detected in testis (Fig. 3C). Six proteins from this group showed a significant correlation with tubular hyalinization (PPIE, ZMYND10, CDS2, HNRNPH2, VAV2 and NLK) and six (PPIE, AIMP1, EIF4G2, DHX15, VAV2 and MAP3K15) had significant correlation with inflammation, while none was correlated with interstitial fibrosis.

Similarly, the 30 proteins that are mutual for SCO vs Hyp and SCO vs OA comparisons had the power to quantitatively discriminate SCO from Hyp and OA and these were named SCO biomarkers. Twenty-four had a lower abundance in SCO compared to OA or Hyp. Eleven of them (H1-6, TKTL2, TKTL1, DPEP3, OSCP1, H2BC1, CEP170, TEX101, MAGEB2, POTEF and KATNAL2) have elevated mRNA expression in testis and two (TKTL2 and H2BC1) are only detected in testis (Fig. 3C). In regards to correlation with histopathology data, H2BC1 was borderline correlated with interstitial fibrosis ($p = 0.009$) and POTEF was correlated with inflammation.

Protein-protein interaction analysis with STRING indicated that both OA and SCO biomarkers have more interactions among themselves than what would be expected for a random set of proteins of the same size and degree of distribution drawn from the genome (OA PPI enrichment p-value = 0.00148; SCO PPI enrichment p-value = 0.00038). Such an enrichment indicates that the proteins are at least partially biologically connected as a group (Fig. 3C). Gene ontology and pathway enrichment (KEGG, Reactome, WikiPathways) functional analysis with ClueGo showed nucleosomal DNA binding (GO:0031492) and cytoplasmic stress granule (CO:0010494) as significantly enriched GO molecular function and cellular component, respectively, in SCO biomarker panel. In the OA biomarker panel, enriched were adherens junction (KEGG:04520), Planar cell polarity/convergent extension (PCP/CE) pathway (R-HAS:4086400) and Dectin-1 mediated noncanonical NF-kB signaling (R-HAS:5607761) (Fig. 3D).

3.4. Comparison and correlation between transcriptomic and proteomics data

For the comparison between the proteomics and transcriptomic datasets, the new experimental design was done in Progenesis QIP, grouping the samples into 2 groups: OA and NOA (Hyp and SCO). Progenesis QIP resulted in the identification of 2180 proteins with quantitative values. After filtering this dataset to remove reverse sequences, proteins identified on only one peptide, yeast ADH and adjusting the p-values for multiple testing using the Benjamini-Hochberg procedure and filtering the dataset for B-H $p \leq 0.05$ and fold change $\geq$1.5, the list of proteins with differential abundance consisted of 564 proteins (Supplementary Table 5). The GSE145467 dataset consisted of 7921 and GSE9210 dataset had 1011 protein coding genes that showed statistically significant (B-H $p \leq 0.05$) difference in expression with fold change $\geq$1.5 between OA and NOA groups (Supplementary Table 5).

Comparison between the 2 transcriptomic datasets revealed 836 genes that were co-differentially expressed and which showed very strong statistically significant correlation (Rho = 0.875; p (2-tailed) < 0.0001) (Fig. 4A**).** Comparison of our proteomics dataset with GSE145467 transcriptomic dataset identified a total of 278 proteins/genes that were co-differentially expressed in the transcriptome and proteome profiles. Of those, 190 proteins/genes showed the same regulation trend (55 up-regulated and 135 down-regulated in OA) while 88 had the opposite trend in the 2 datasets. Spearman correlation analysis revealed a weak positive correlation value (Rho = 0.299; p (2-tailed) < 0.0001) for the expression of these 278 genes/proteins in the transcriptome and proteome datasets (Fig. 4B). We subsequently performed 2D annotation enrichment analysis in order to identify GO annotations which were simultaneously up- or down-regulated in both datasets [53]. We observed several biological processes that were up-regulated in both datasets such as spermatogenesis, gamete generation and protein targeting (Fig. 4C). Conversely, we have observed several biological processes such as cell adhesion, extracellular matrix assembly, molecular functions such as extracellular matrix structural constituent, platelet-derived growth factor binding and pathways such as ECM-receptor interaction, protein digestion and absorption as down-regulated in both datasets. Comparison of our proteomics dataset with GSE9210 transcriptomic dataset identified a total of 55 proteins/genes that were co-differentially expressed with a moderate positive correlation value (Rho = 0.486; p (2-tailed) < 0.0001) (Fig. 4D). Of those, 43 proteins/genes showed the same regulation trend (20 up-regulated and 23 down-regulated in OA) while 12 had the opposite trend in the 2 datasets. 2D annotation enrichment analysis revealed anti-correlating behavior between proteins and transcripts for the ones involved in positive regulation of cellular component organization. Additionally we observed down-regulation of plasma membrane localization in both datasets (Fig. 4E). Overall, 48 proteins/genes were found at the intersection among the 2 transcriptomic and our proteomics dataset, of which, 36 proteins/genes showed the same regulation trend (18 up-regulated and 18 down-regulated in OA) (Fig. 4F). Spearman correlation analysis revealed a moderate positive correlation value for the expression of these genes/proteins in the transcriptome and proteome datasets (GSE145467 vs Proteomics: Rho = 0.575, p (2-tailed) < 0.0001; GSE9210 vs Proteomics: Rho = 0.499, p (2-tailed) < 0.0001) (Fig. 4G**).**

The intersection between 88 candidate biomarkers from our study which were common for >2 comparisons (Supplementary Table 4**, rows 2–89**) and 48 proteins/genes mutual for the 2 transcriptomic and proteomics datasets revealed 7 common proteins: RANBP1, TKTL1, H2BC1, ZMYND10, CCDC8, RNF130 and ACTL7B. Six of them had the same regulation trend in transcriptomic and proteomics datasets, respectively of which 5 (RANBP1, TKTL1, H2BC1, ZMYND10 and ACTL7B) were up-regulated and one (CCDC8) was down-regulated in OA. As the GSE145467 data set contained eight times more genes with statistically significant $\geq$1.5 fold change expression between OA and NOA groups than GSE9210 dataset, intersection of this dataset with the 88 candidate biomarkers from the proteomics study gave 35 genes. Of these, 25 had the same regulation trend with 12 genes (H1-6, RANBP1, TKTL2, GPX4, TKTL1, OSCP1, H2BC1, MYH1, ASRGL1, ZMYND10, ACTL7B, ATG3) being up-regulated and 13 genes (PLXNA4, CCDC8, CDS2, AIMP1, PALLD, PDLIM5, COL4A1, LAMA4, TRA2A, HNRNPH2, PDGFRA, EHD2, PSME1) down regulated (Fig. 4H). Spearman correlation analysis revealed high positive correlation value for GSE145467 transcriptome

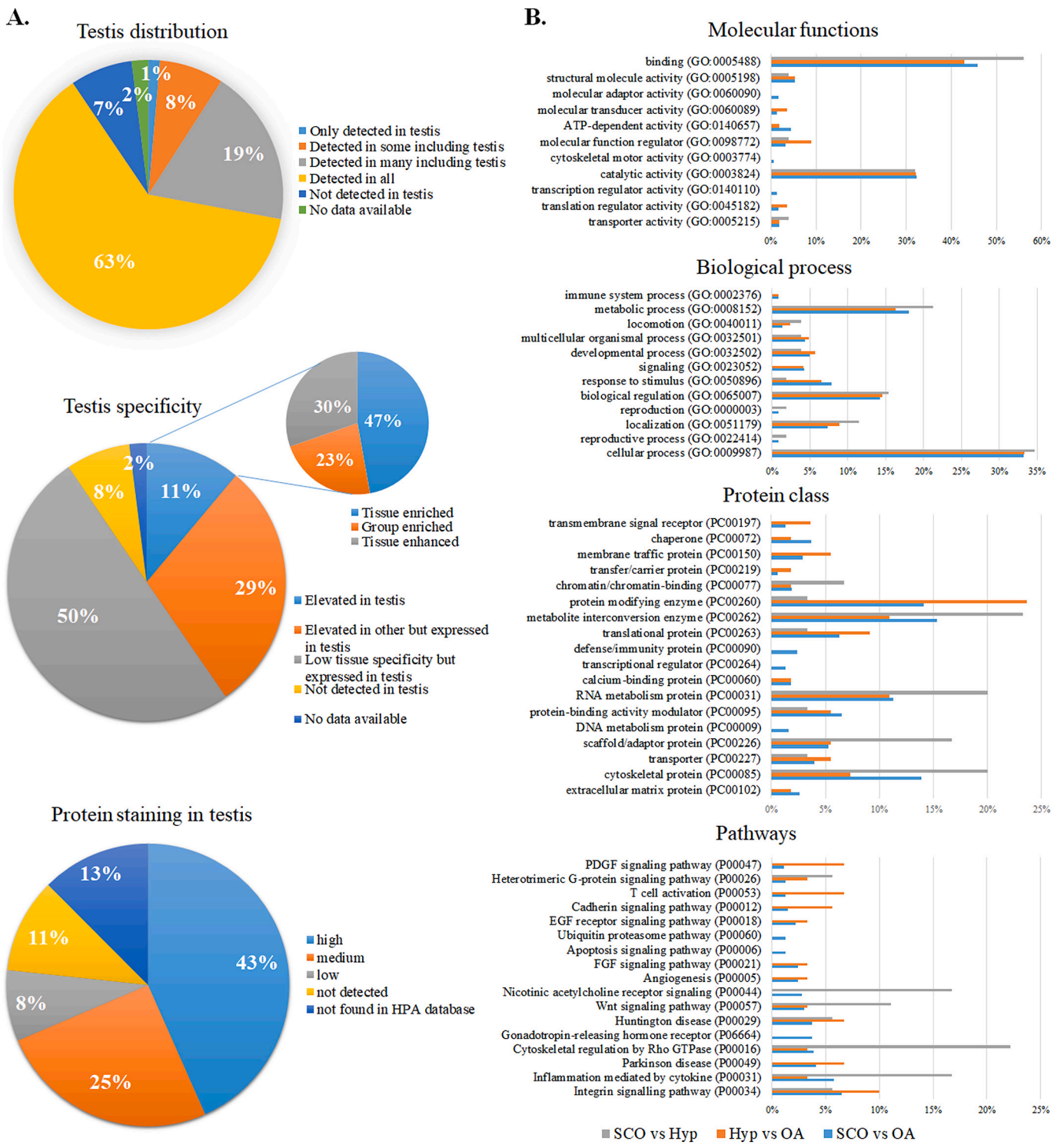

Fig. 2. Characterization of proteins with significantly altered abundance in terms of testis distribution, specificity and GO Annotations. (**A**) Data was retrieved from Human Protein Atlas. Distribution was based on transcript detection (nTPM≥1) in testis as well as in all other tissues, while specificity was based on transcript testis specificity. The level of protein detected in testis was based antibody staining (immunohistochemistry, Western blot) and protein arrays. (**B**) GO annotations of proteins were according to Panther Classification System.

and proteome datasets (Rho = 0.893, p (2-tailed) < 0.0001) for these 35 proteins.

3.5. *Validation of gene expression levels*

To confirm and validate the above observed correlation between protein abundance and gene expression, we performed qPCR for a preselected set of genes. Six genes were selected based on the 1) highest discriminatory power among hypospermatogenesis, OA and SCO groups on protein level, 2) high testis specificity – elevated mRNA expression in testis and 3) the same regulation trend in proteomics and one or both transcriptomic datasets. In addition, none of the selected proteins showed significant correlation with specific histopathological features, with exception of H2BC1 which correlated with interstitial fibrosis, but

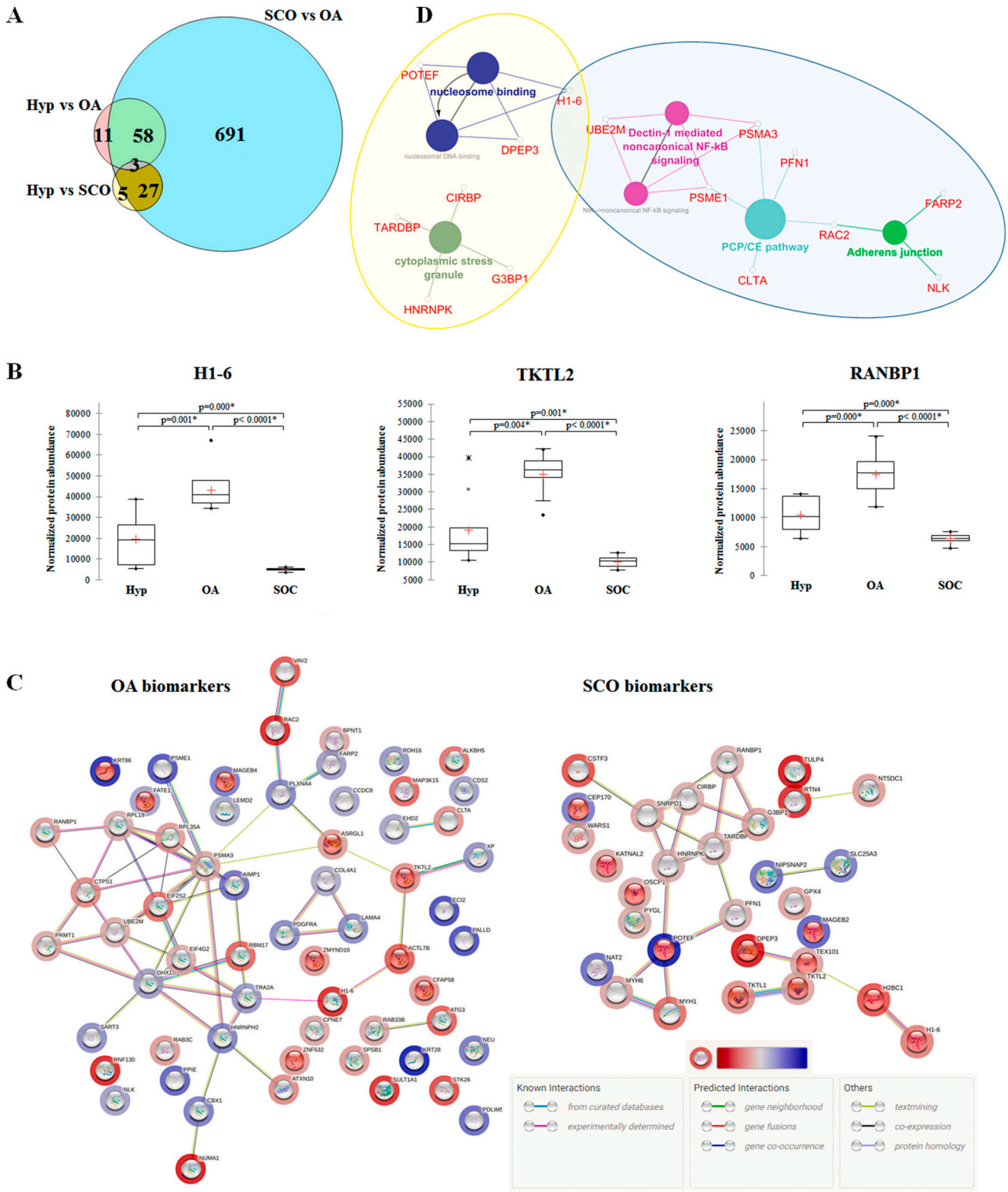

Fig. 3. Protein–protein interaction network, enriched functions, processes and associated pathways with proteins that could quantitatively discriminate OA, SCO and Hyp. (**A**) Overlap among Hyp vs OA, SCO vs OA and SCO vs Hyp comparisons. (**B**) Three proteins (H1-6, RANBP1 and TKTL2) that could quantitatively discriminate between Hyp, OA and SCO. In the box plot graphs, median (–), 25th and 75th percentiles, minimum/maximum (•) and mean (+) are shown. (**C**) Protein–protein interaction network among proteins that discriminate OA from NOA (proteins mutual for Hyp vs OA and SCO vs OA comparisons, $n = 61$) and SCO from other groups (proteins mutual for SCO vs OA and SCO vs Hyp comparisons, $n = 30$), respectively. The halo color represents the regulation trend: down-regulated proteins are marked with red halo, while up-regulated are marked with blue halo. The color of the ball represent testis specificity (red-elevated expression in testis, white – low expression in testis) (**D**) Functional enrichment analysis of the significantly associated GO molecular functions, biological processes, cellular compartments and pathways (KEGG, Reactome, WIKI) with our OA and SCO biomarker panels. Yellow and blue circles encompass functions/pathways enriched in SCO and OA biomarker sets, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

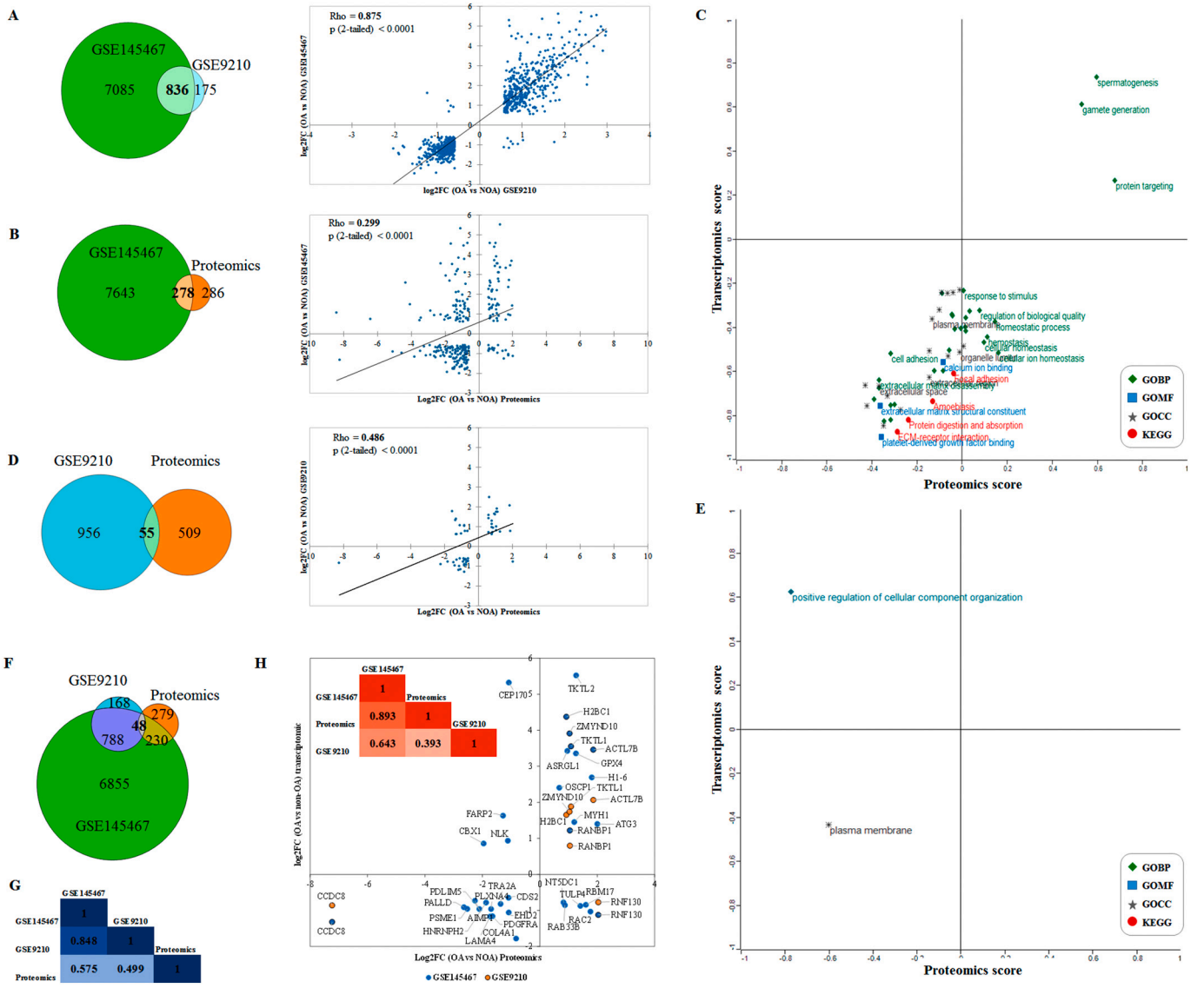

Fig. 4. Comparison and correlation between testis transcriptome and proteome. The Venn diagram represents unique and shared differentially expressed genes/proteins while the scatter plots are showing the Spearman's correlation between: (A) mRNA from GSE145467 and GSE9210 transcriptomic datasets; (B) the mRNA levels from GSE145467 transcriptomic dataset and protein expression levels from proteomics dataset; (C) 2D annotation enrichment analysis of proteins/genes mutual for GSE145467 transcriptomic and proteomics datasets generated using Perseus software; (D) the mRNA levels from GSE9210 transcriptomic dataset and protein expression levels from proteomics dataset; (E) 2D annotation enrichment analysis of proteins/genes mutual for GSE9210 transcriptomic and proteomics datasets generated using Perseus software; (F) overlap among differentially expressed proteins/genes from proteomics, GSE145467 and GSE9210 transcriptomic datasets. (G) Spearman correlation for the expression of the 48 proteins/genes found at the intersection among the 2 transcriptomic and our proteomics dataset. (H) Spearman correlation for the expression of the 35 proteins found at the intersection among the 88 proteomic biomarker candidates and one or both transcriptomic datasets (blue = GSE145467; orange = GSE9210). The seven proteins/genes mutual for the 2 transcriptomic and proteomics datasets are marked with circles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

this was borderline and therefore not considered as significant.

As illustrated in Fig. 5, the transcription levels of five genes (H1-6, RANBP1, TKTL2, TKTL1 and H2BC1) were significantly reduced in hypospermatogenesis and SCO compared to OA group ($p < 0.0001$) respectively, as well as significantly reduced in SCO compared to hypospermatogenesis ($p < 0.0001$). The transcription levels of ACTL7B were significantly reduced in SCO compared to OA and hypospermatogenesis ($p < 0.0001$) respectively, but no significant difference was observed between hypospermatogenesis and OA groups (p=0.224).

Comparison of the transcriptomics with proteomics data revealed that mRNA expression levels of the 3 genes (H1-6, RANBP1 and TKTL2) that could quantitatively discriminate among hypospermatogenesis, OA and SCO on protein level, were consistent with proteomics results. For TKTL1 and H2BC1, that on protein level were able to quantitatively discriminate hypospermatogenesis from SCO (B-H $p < 0.05$) and OA from SCO (B-H $p < 0.05$) but could not discriminate hypospermatogenesis and OA ($p_{(TKTL1)} = 0.077$; $p_{(H2BC1)} = 0.136$), mRNA expression levels were found to discriminate among all 3 groups with statistical significance ($p < 0.0001$). ACTL7B on protein level was shown to quantitatively discriminate OA from hypospermatogenesis and SCO respectively (B-H $p < 0.05$), but no significant difference was observed between hypospermatogenesis and SCO groups ($p = 0.136$). On the other hand, on transcription level there were statistically significant difference between SCO compared to OA and hypospermatogenesis respectively, but no significant difference between hypospermatogenesis and OA groups.

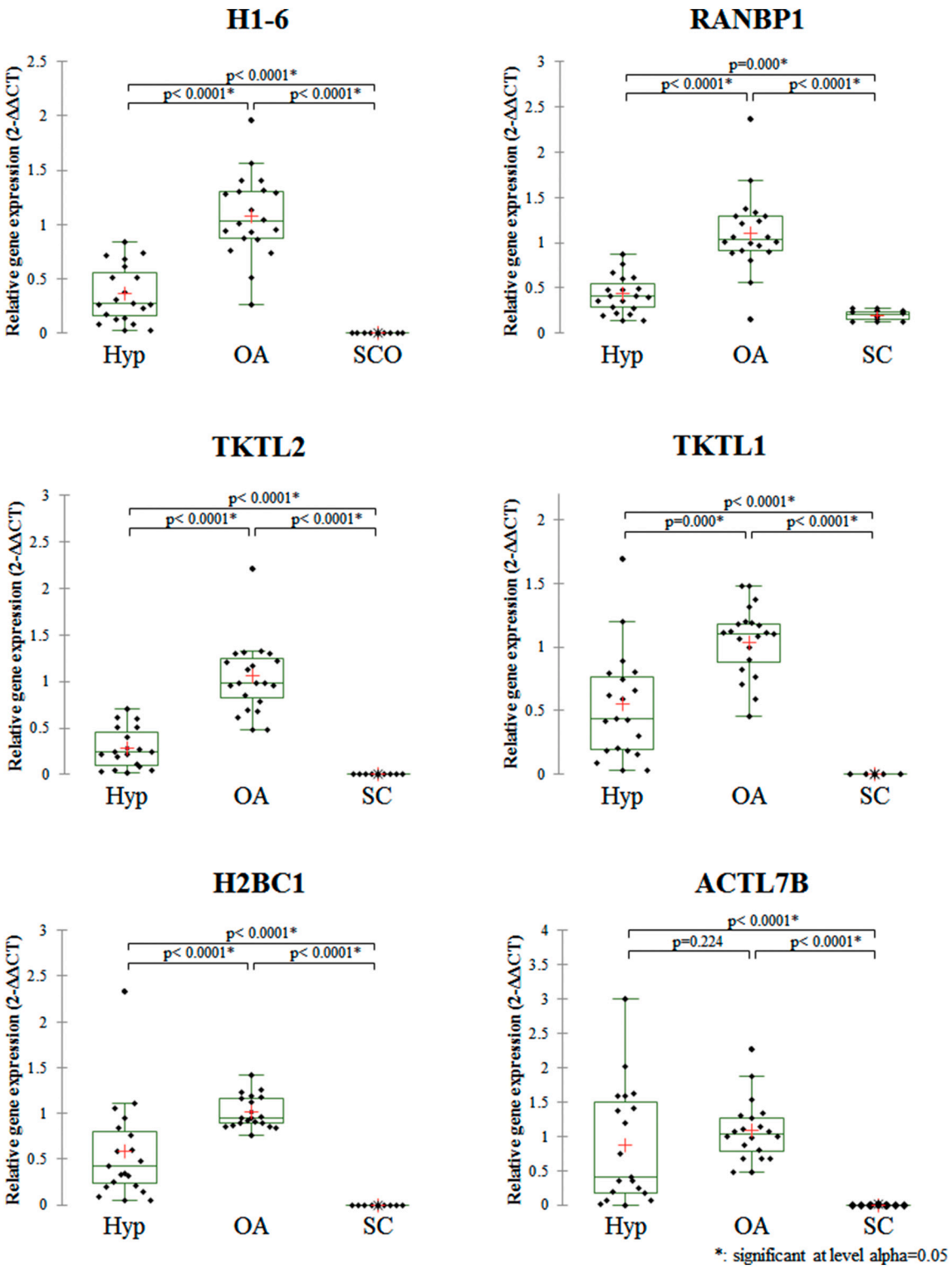

Fig. 5. Relative gene expression of six genes analyzed by qPCR in independent cohort of OA ($n = 20$), Hyp ($n = 19$) and SOC ($n = 10$). Changes in gene expression were determined by the $2^{-\Delta\Delta CT}$ method using GAPDH as internal reference. Data is presented in box plots showing all individual values, median (−), 25th and 75th percentiles, outliers (•) and mean (+). The statistical significances of the differences in gene expression among the studied groups were obtained using Mann-Whitney *U* test.

4. Discussion

Spermatogenesis is a complex process in which different cell types in the testis are involved. Testicular tissues with spermatogenic failure can be used as the best model to study spermatogenesis and to identify the genes and proteins involved in this process. So far, only a few studies have used the comparative proteomics approach in the analysis of testicular tissues with azoospermia [54–56]. In this study, using integrated comparative proteomics and transcriptomics approach and clinical samples from men with OA and NOA, we have tried to unravel the

molecular processes behind the azoospermia as the most severe form of male infertility and to identify proteins that could be used to discriminate the different types/sub-types. The samples that we have used in this study were FFPE testicular tissues from patients with OA and NOA. Shotgun proteomics has been successfully applied in the analysis of FFPE samples in a great number of studies, with good overlapping of identifications compared to fresh tissues and this approach is continually improved due to advances in extraction/digestion procedures and increases in sensitivity and resolution of the LC-MS/MS instrumentation [57]. We have used an extraction/digestion procedure based on heat-induced antigen retrieval and the RapiGest protocol, which gives almost identical numbers of proteins from fresh and FFPE tissues with excellent extraction of proteins from all isoelectric point range, good preservation of high-MW proteins and excellent technical reproducibility [39]. A comparison of the identified proteins in FFPE testicular tissues in this study with an in-depth proteome study of human testis [58] revealed an overlap of 80%, which further supports the validity of our approach. In regards to interpatient variability in our cohort which is of high importance for reliable analysis, we have tried to preselect the patient according to age and available histopathology information in order to form as homogenous groups as possible. Proteome heatmap of the individual samples included in the discovery analysis showed very low interpatient variability in OA and SCO groups respectively. On the other hand, higher interpatient variability was observed in Hyp group as patients with low, intermediate and high levels of hypospermatogenesis were included in order to form a representative group for this azoospermia subtype.

Through stringent filtration of our proteomics data, we came to the list of 88 potential biomarkers that could discriminate hypospermatogenesis, SCO syndrome and obstructive azoospermia. Bioinformatics enrichment analysis revealed an association with several GO annotations and pathways. Proteins that could quantitatively discriminate between OA and NOA were significantly associated with integrin signaling pathway (P00034), adherens junction (KEGG:04520), PCP/CE pathway (R-HAS:4086400) and Dectin-1 mediated noncanonical NF-kB signaling (R-HAS:5607761). At the moment, the precise relationship between infertility and molecules and signaling pathways that are involved is still unknown. However, the pathways that we have identified already have some growing evidence supporting their association with azoospermia. The integrin signaling pathway is triggered when integrins in the cell membrane bind to extracellular matrix components. In addition to somatic cells, integrins have also been detected on germ cells and are known to play a crucial role in complex gamete-specific physiological events, resulting in sperm-oocyte fusion. Most recently, it was shown that Sertoli cells in idiopathic NOA are physiologically immature. The identified candidate pathways and regulators which may play a crucial role during Sertoli cell maturation were Wnt/β-catenin, ephrin receptor and integrin pathways [59]. Involved proteins from our dataset include mostly collagens (COL1A1, COL18A1, COL14A1, COL6A2, COL4A1, COL4A3, COL1A2, COL4A2, COL3A1), laminins (LAMB2, LAMB4, LAMA4, LAMC1), Mitogen-activated protein kinases (MAPK1, MAPK3, MAP3K1), POTEF, etc., all up-regulated in NOA subtypes. Cell-cell adherens junctions (AJs), the most common type of intercellular adhesions, are important for maintaining tissue architecture and cell polarity and can limit cell movement and proliferation. Changes in the function of cellular junctions in the human epididymis are associated with male infertility [60]. Proteins associated with adherens junctions were NLK, FARP2 and RAC2. The planar cell polarity (PCP) pathway controls the establishment of polarity within the plane of a sheet of cells. In vertebrates this pathway is involved in the regulation of convergent extension, a process by which a tissue narrows along one axis and lengthens along with a perpendicular one. The involvement of this pathway within spermatogenesis has been studied recently. There have been finding that connect the planar cell polarity with the development of male germ cells [61]. Proteins associated with this pathway were PSME1, PSMA3, CLTA, PFN1 and RAC2. Additionally, this pathway has an involvement in the non-canonical WNT pathway, as some WNT proteins have been shown to have roles in PCP processes. Most recently, it was also verified that the WNT signaling pathway regulates the maturation of Sertoli cells in both normal and NOA patients and its inhibition successfully promotes the partial maturation of some Sertoli cells in idiopathic NOA patients [59]. The NF-kB non-canonical pathway is involved in immune cell differentiation and maturation and secondary lymphoid organogenesis. There is still a limited evidence in regards to NF-kb signaling and male infertility. It was found that NF-kB is implicated in numerous stress responses including apoptosis within male testicular cells [62]. Furthermore, one additional study suggest that NF-kB is an important regulator for the expression of the testis-enriched LRWD1, originally identified as one of the genes down-regulated in the testicular tissues of patients with severe spermatogenic defects [63]. Proteins associated with this pathway were PSME1, PSMA3 and UBE2M. The association of our data with these signaling pathways adds more info to the growing knowledge of processes associated with azoospermia.

Proteins that could discriminate SCO from hypospermatogenesis and OA showed significant association with DNA binding (GO:0031492) and cytoplasmic stress granule (CO:0010494). The involvement of DNA binding with spermatogenesis is crucial in the formation of the unique morphology and function of spermatozoa. The connection between this process and spermatogenesis is well established [64]. Proteins associated with DNA binding are DPEP3, H1-6 and POTEF. Down-regulation of the DPEP3 and H1-6 could indicate that this process is somehow affected in the SCO group. The cytoplasmic stress granule is a dense aggregation in the cytosol composed of proteins and RNAs that appear when the cell is under stress. This process is involved in the regulation of translation and mRNA degradation in eukaryotic cells. Recently, growing evidence has emerged on stress granule response mechanisms, with several addressing their impact on germ cell development [65]. In our study this process is associated with CIRBP, G3BP1, HNRNPK and TARDBP, all down-regulated in SCO. This could indicate that the formation of cytoplasmic stress granule is somehow disrupted in SCO. As there are relatively few studies on cytoplasmic stress granule in the reproductive field, this data could contribute to future investigations of this aspect.

Further, our comparative proteomics analysis revealed an extensive list of potential biomarkers for discrimination of OA and NOA, as well as discrimination between subtypes of NOA, namely, hypospermatogenesis and SCO. The list contains some of, so far, well-recognized candidate biomarkers for azoospermia such as TEX101, DPEP3 and H2BC1.

Testis-expressed sequence 101 (TEX101) protein is a testicular germ cell-specific protein expressed as a GPI-anchored (glycosylphosphatidylinositol-anchored) protein, predominantly on the plasma membrane of germ cells during all stages of spermatogenesis [66]. TEX101 has been proposed and extensively studied as an emerging biomarker for the non-invasive differential diagnosis of OA and NOA by Drabovich and colleagues [28,32], but has been identified as a biomarker for azoospermia in seminal plasma from other groups also [26,31]. It was shown that TEX101 levels are significantly reduced in seminal plasma of patients with NOA, with levels in SCO close to zero. In addition to the ability to differentiate between OA and NOA, it was shown that TEX101 can also distinguish to some extent between SCO and hypospermatogenesis or MA, which eventually led to the development of ELISA test [33]. Further extensive validation showed that TEX101 in combination with ECM1 has sensitivity for differentiating between NOA and OA of 81% at 100% specificity [34]. In our study, TEX101 protein levels were also found significantly down regulated in Hyp and SCO in comparison to OA (−1.8 and −2.0 fold change) and could significantly discriminate SCO from OA and Hyp, respectively. However, statistically significant discrimination between OA and Hyp group, based on TEX101 in our study was not reached.

DPEP3 is another testis-specific GPI anchored protein, expressed on the cell surface of testicular germ cells such as spermatogonia,

spermatocytes, spermatids, and spermatozoa. DPEP3 might have a significant role in the process of spermatogenesis, as it has been found with significantly reduced expression in seminal plasma of NOA and post vasectomy patients [26,31]. Its role in male infertility is further supported by findings that it forms a molecular complex with TEX101 [67,68]. In our study DPEP3 levels were found 22-fold and 14-fold reduced in the SCO group in comparison with OA and Hyp groups respectively, which further supports its role as a potential biomarker for azoospermia.

Histone H2B type 1-A (H2BC1) is a protein expressed only in testis. It has a function during the replacement of histones by protamines in male germ cells, probably acting as a nucleosome dissociating factor. As such, it plays a part in the condensation of the chromatin during spermatogenesis, a key factor in the process [69]. Sperm protamine deficiency (partial or complete) has been demonstrated in infertile men which further consolidates H2BC1 as an important factor in the process of spermatogenesis [70]. Several proteomics studies have found this protein as a potential biomarker for male infertility [26,31,71]. We have found H2BC1 protein levels 6-fold reduced in the SCO group in comparison with both OA and Hyp groups.

In addition to the above mentioned well established biomarkers, we have identified several other potential biomarkers for discrimination of azoospermia types/subtypes. Among these, three proteins, namely H1-6, RANBP1 and TKTL2, showed superior potential for quantitative discrimination of OA from hypospermatogenesis and SCO with the highest statistical significance.

Histone H1t (H1-6) is a testis-specific linker histone exclusively detected in mid- to late pachytene spermatocytes and in the elongating spermatids where it maintains high expression levels [72]. Biochemical and biophysical studies found that H1-6 facilitates histone replacement during spermatogenesis [73]. As so, the possible role of H1-6 in spermatogenesis impairment has been well studied on the genetic/epigenetic level. Contrary to expected, H1-6-null mice exhibit no spermatogenesis abnormalities and are fertile, indicating that H1-6 is not critical for spermatogenesis [74]. In addition, a recent study investigating genetic variants in the regulatory regions of the H1-6 gene in men with NOA showed no significant differences between case and control groups [75]. However, our study indicated that H1-6 protein levels in the testis are significantly reduced in hypospermatogenesis and SCO patients in comparison with OA patients (2.2-fold and 8.5-fold reduction, respectively) and that can also discriminate with statistical significance between SCO and hypospermatogenesis by approximately 4-fold reduction in SCO. The validation of the mRNA expression levels were in complete agreement with the proteomics results.

Ran-specific GTPase-activating protein (RANBP1) is a ubiquitously expressed protein that serves as a cofactor for stimulating RanGTP activity [76]. The link of this protein to male infertility was postulated due to the connection between RanGTP and importins, which through several knockout organisms studies, have been shown as essential for male fertility [77]. So far this hypothesis has been supported by two studies. In the first study, RANBP1-knockout mice were shown to exhibit male infertility due to a spermatogenesis arrest, presumably caused by down-regulation of RANBP2 during spermatogenesis [78]. Furthermore, RANBP9, another protein associated with the RAN pathway, has been shown to be crucial for male germ cell development and thus male fertility [79]. The results from our study indicate that RANBP1 protein levels are reduced by 1.7-fold in the hypospermatogenesis group in comparison to OA, and further reduced by 1.6-fold in SCO, compared to hypospermatogenesis. Therefore, testis RANBP1 levels could discriminate SCO, hypospermatogenesis and OA with statistical significance. Gene expression analysis confirmed the proteomics results, pointing to the need for further research of the diagnostic potential of this protein.

Our study for the first time implicated transketolase-like 2 (TKTL2) as a biomarker of azoospermia. According to data from HPA, TKTL2 is solely detected in testis. Our data showed that TKTL2 can discriminate the 3 azoospermia types/sub-types on protein level, with the power to discriminate even hypospermatogenesis from OA by 1.8 fold reduction with statistical significance. We also confirmed that this trend in reduction on protein level is highly comparable with mRNA expression levels in the validation groups.

In addition to TKTL2, we have also identified transketolase-like 1 (TKTL1) as differentially expressed in azoospermia. Even though the enzymatic activity of TKTL1 and TKTL2 had been put in question, their function as transketolases has been empirically confirmed recently [80]. TKTL1 has elevated expression in the testis, and according to data from HPA belongs to the group of enriched genes with at least four-fold higher average mRNA level in a group of 2–5 tissues compared to any other tissue. The evidence connecting TKTL1 with male reproduction is growing. Rolland and colleagues [30], through proteomic analyses of seminal plasma together with transcriptomic gene expression profiling of testis tissues, first detected that TKTL1 together with LDHC and PGK2 is severely down-regulated in azoospermia and could distinguish fertile from infertile men. Our study went further into the analysis of the expression of TKTL1 in different types of azoospermia, showing that both on protein and mRNA levels, it is significantly reduced in SCO in comparison with both OA and hypospermatogenesis. As such, TKTL1 and in particular TKTL2, are worth in-depth research to further elucidate their role in male infertility and their potential as a diagnostic tool.

Another candidate protein worth pointing out is Actin-like protein 7B (ACTL7B). Actin proteins are heavily involved in the process of morphological change during spermatogenesis [81]. ACTL7B is an intronless gene that is solely expressed in the testis, in the cytoplasm of round and elongating spermatids, in or around the forming acrosome, which suggests its involvement in spermatogenesis [82]. So far, several studies are proving the association of ACTL7B with male infertility, but solely on gene level [83–85]. Up to our knowledge, this study is the first that shows that at protein level, ACTL7B has a statistically significant reduction in the testis of hypospermatogenesis and SCO patients compared to OA patients. Validation of its mRNA expression showed that there is a reduction in gene expression levels in SCO compared to both Hyo and OA, but the statistically significant difference was reached only in SCO to OA comparison. As such, ACTL7B is a potential target for further research concerning male infertility and with emphasis on its potential to quantitatively discriminate OA from NOA.

In addition to the above mentioned potential biomarkers, our list contains several more proteins that have elevated expression in testis according to the HPA database and have been associated directly or indirectly with male infertility.

Centrosomal protein of 170 kDa (CEP170) belongs to the family of centrosomal proteins for which few studies have indicated that dysfunction of some members may result in male sterility. Recently it has been proven that CEP170 is under the transcriptional control of CEP128 which has a crucial role in male reproduction [86]. MAGE Family Member B4 (MAGEB4) is a possible cause of rare X-linked azoospermia and thus providing the first clue to the physiological function of a MAGE family in male infertility [87]. In addition to this protein, we have identified also MAGE Family Member B2 (MAGEB2) localized in the DSS (dosage-sensitive sex reversal) critical region as protein with differential abundance in azoospermia. POTE Ankyrin Domain Family Member F (POTEF) has been identified as upregulated in SCO patients in a proteomics study [56], as we found in our study. In addition, for another member of this family, POTEB, copy number variations were identified in males with unexplained azoospermia [88]. Katanin p60 ATPase-containing subunit A-like 2 (KATNAL2) is another essential player in the process of spermatogenesis, which absence in mice causes abnormalities in sperm head shapes, sperm tails, and failure in spermiation [89]. The biallelic mutations in this gene in humans have been shown to be a direct cause of male infertility [90]. Mutations in Zinc finger MYND-type containing 10 (ZMYND10) gene, essential for proper axonemal assembly of inner and outer dynein arms in humans and flies was found to cause primary ciliary dyskinesia, a ciliopathy which among other things is characterized by infertility [91]. Cilia and

flagella associated protein 58 (CFAP58) is expressed predominantly in testis and plays a role in sperm flagellogenesis. Analyses of patient sperm and studies in CFAP58-knockout mice, led to the conclusion that biallelic loss of function variants in CFAP58 can cause axonemal and periaxonemal malformations resulting in male infertility [92]. Mutational analysis of the human Fetal and adult expressed 1 (FATE1) gene, found to be mainly expressed in spermatogonia, primary spermatocytes and Sertoli cells, in infertile men showed that FATE mutations are contributing factor in some cases of male infertility [93]. An in silico analysis of human sperm genes associated with asthenozoospermia found Asparaginase and isoaspartyl peptidase 1 (ASRGL1) as one of the differentially expressed genes in asthenozoospermia spermatozoa and has linked it as a gene with implication in male infertility [94].

All of the above discussed potential candidates are characterized with elevated expression in testis. More detailed HPA and GO annotation for these proteins is given in Table S1, Supplementary document. In this regard, TKTL2, H2BC1, ACTL7B and FATE1 are solely expressed in testis. The second group, consisting of TKTL1, DPEP3, TEX101, H1-6, MAGEB2, POTEF, ZMYND10, MAGEB4 and CFAP58 are detected beside testis, in some other tissues (more than one but less than one-third of all tissues). Half of these proteins are tissue enriched with at least four-fold higher mRNA level in testis compared to any other tissues (DPEP3, H1-6, MAGEB2, POTEF) and the remaining have group enrichment with at least four-fold higher average mRNA level in a group of 2–5 tissues compared to any other tissue (TKTL1, TEX101, ZMYND10, MAGEB4). The third group consist of ubiquitously expressed proteins but with at least four-fold higher mRNA level in testis compared to the average level in all other tissues.

In order to be classified as a biomarker for non-invasive diagnosis, in addition to the potential to quantitatively discriminate the azoospermia types/subtypes, of outmost importance is the biomarker transfer into body fluids. Present knowledge indicate that almost half of the above discussed potential biomarkers are transferred into body fluids. TKTL1, DPEP3, TEX101, TKTL2, H2BC1 and ASRGL1 have confirmed presence into seminal plasma where have been detected by several comparative proteomics studies of azoospermia [26,30,31,95]. In addition, DPEP3, CEP170 and POTEF have been detected in blood and TEX101 is secreted in male reproductive system according to HPA. For the remaining candidates, among which H1-6 and RANBP1 which showed superior potential for quantitative discrimination of OA from NOA and ACTL7B and FATE1 which are exclusively expressed only in testis, future studies are needed to test their possible transfer into body fluids.

5. Conclusions

Using comparative proteomics of testicular tissues from patients with OA and NOA, together with transcriptomics data integration and bioinformatics analysis we have identified several molecular processes and pathways that are involved in azoospermia. As a direct result of this approach, this study led to the identification of a number of potential biomarker candidates for quantitative discrimination among the studied types/subtypes of azoospermia. Our findings open the way to new research in regards to non-invasive diagnosis of azoospermia. Upon assessment of the possible transfer of the candidate proteins into body fluids and further more extensive validation, some of these proteins may be combined in a biomarker panel for unambiguous differentiation between NOA and OA patients and discrimination between NOA subtypes.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2022.104686.

CRediT authorship contribution statement

Katarina Davalieva: Methodology, Investigation, Formal analysis, Validation, Visualization, Writing – original draft. **Aleksandar Rusevski:** Investigation, Formal analysis. **Milan Velkov:** Investigation, Formal analysis. **Predrag Noveski:** Investigation. **Katerina Kubelka-Sabit:** Resources. **Vanja Filipovski:** Resources. **Toso Plaseski:** Resources. **Aleksandar Dimovski:** Resources, Supervision. **Dijana Plaseska-Karanfilska:** Conceptualization, Supervision, Project administration, Writing – review & editing.

Declaration of Competing Interest

The authors declare no conflict of interest.

Data availability

The proteomic data has been deposited to the ProteomeXchange Consortium via the PRIDE Partner Repository with the dataset identifier PXD032722.

References

[1] P. Thonneau, S. Marchand, A. Tallec, M.L. Ferial, B. Ducot, J. Lansac, P. Lopes, J. M. Tabaste, A. Spira, Incidence and main causes of infertility in a resident population (1,850,000) of three French regions (1988-1989), Hum. Reprod. 6 (1991) 811–816.
[2] S. Bhasin, D.M. de Kretser, H.W. Baker, Clinical review 64: pathophysiology and natural history of male infertility, J. Clin. Endocrinol. Metab. 79 (1994) 1525–1529.
[3] S.W. Leslie, L.E. Siref, T.L. Soon-Sutton, M.A.B. Khan, Male Infertility, StatPearls Publishing, Treasure Island (FL), 2022. Available from, https://www.ncbi.nlm.nih.gov/books/NBK562258/.
[4] B.R. Winters, T.J. Walsh, The epidemiology of male infertility, Urol. Clin. N. Am. 41 (2014) 195–204.
[5] P.N. Kolettis, The evaluation and management of the azoospermic patient, J. Androl. 23 (2002) 293–305.
[6] M. Cocuzza, C. Alvarenga, R. Pagani, The epidemiology and etiology of azoospermia, Clinics (Sao Paulo) 68 (Suppl. 1) (2013) 15–26.
[7] R.I. McLachlan, E. Rajpert-De Meyts, C.E. Hoei-Hansen, D.M. de Kretser, N. E. Skakkebaek, Histological evaluation of the human testis–approaches to optimizing the clinical value of the assessment: mini review, Hum. Reprod. 22 (2007) 2–16.
[8] G.R. Dohle, S. Elzanaty, N.J. van Casteren, Testicular biopsy: clinical practice and interpretation, Asian J. Androl. 14 (2012) 88–93.
[9] S. Baskaran, A. Agarwal, K. Leisegang, P.N. Pushparaj, M.K. Panner Selvam, R. Henkel, An in-depth bibliometric analysis and current perspective on male infertility research, World J. Mens Health 39 (2021) 302–314.
[10] J.C. Ribeiro, M.G. Alves, F. Amado, R. Ferreira, P. Oliveira, Insights and clinical potential of proteomics in understanding spermatogenesis, Expert Rev. Proteomics 18 (2021) 13–25.
[11] J.M. Bieniek, A.P. Drabovich, K.C. Lo, Seminal biomarkers for the evaluation of male infertility, Asian J. Androl. 18 (2016) 426–433.
[12] L. Candenas, R. Chianese, Exosome composition and seminal plasma proteome: a promising source of biomarkers of male infertility, Int. J. Mol. Sci. 21 (2020), https://doi.org/10.3390/ijms21197022.
[13] R. Cannarella, F. Barbagallo, A. Crafa, S. La Vignera, R.A. Condorelli, A. E. Calogero, Seminal plasma transcriptome and proteome: towards a molecular approach in the diagnosis of idiopathic male infertility, Int. J. Mol. Sci. 21 (2020), https://doi.org/10.3390/ijms21197308.
[14] A.P. Drabovich, P. Saraon, K. Jarvi, E.P. Diamandis, Seminal plasma as a diagnostic fluid for male reproductive system disorders, Nat. Rev. Urol. 11 (2014) 278–288.
[15] N. Kumar, N.K. Singh, Emerging role of novel seminal plasma bio-markers in male infertility: a review, Eur. J. Obstet. Gynecol. Reprod. Biol. 253 (2020) 170–179.
[16] R. Zarezadeh, S. Nikanfar, H. Oghbaei, Y. Rastgar Rezaei, D. Jafari-Gharabaghlou, Y. Ahmadi, M. Nouri, A. Fattahi, R. Dittrich, Omics in seminal plasma: an effective strategy for predicting sperm retrieval outcome in non-obstructive azoospermia, Mol. Diagn. Ther. 25 (2021) 315–325.
[17] J.L. Ballesca, J. Balasch, J.M. Calafell, R. Alvarez, F. Fabregues, M.J. de Osaba, C. Ascaso, J.A. Vanrell, Serum inhibin B determination is predictive of successful testicular sperm extraction in men with non-obstructive azoospermia, Hum. Reprod. 15 (2000) 1734–1738.
[18] S. Muttukrishna, H. Yussoff, M. Naidu, J. Barua, K. Arambage, H. Suharjono, M. Sathanandan, Serum anti-Mullerian hormone and inhibin B in disorders of spermatogenesis, Fertil. Steril. 88 (2007) 516–518.
[19] C. Tsametis, G. Mintziori, P.K. Iliadou, B.C. Tarlatzis, I. Papadimas, D.G. Goulis, Dynamic endocrine test of inhibin B and anti-Mullerian hormone in men with non-obstructive azoospermia, Gynecol. Endocrinol. 27 (2011) 661–665.
[20] P. Fenichel, R. Rey, S. Poggioli, M. Donzeau, D. Chevallier, G. Pointis, Anti-Mullerian hormone as a seminal marker for spermatogenesis in non-obstructive azoospermia, Hum. Reprod. 14 (1999) 2020–2024.
[21] C. Legare, F. Cloutier, S. Makosso-Kallyth, N. Laflamme, K. Jarvi, R.R. Tremblay, R. Sullivan, Cysteine-rich secretory protein 1 in seminal plasma: potential biomarker for the distinction between obstructive and nonobstructive azoospermia, Fertil. Steril. 100 (2013) 1253–1260.

[22] M. Starita-Geribaldi, F. Roux, J. Garin, D. Chevallier, P. Fenichel, G. Pointis, Development of narrow immobilized pH gradients covering one pH unit for human seminal plasma proteomic analysis, Proteomics 3 (2003) 1611–1619.
[23] K. Yamakawa, K. Yoshida, H. Nishikawa, T. Kato, T. Iwamoto, Comparative analysis of interindividual variations in the seminal plasma proteome of fertile men with identification of potential markers for azoospermia in infertile patients, J. Androl. 28 (2007) 858–865.
[24] W.W. Chiu, L.W. Chamley, Human seminal plasma prolactin-inducible protein is an immunoglobulin G-binding protein, J. Reprod. Immunol. 60 (2003) 97–111.
[25] K. Davalieva, S. Kiprijanovska, P. Noveski, T. Plaseski, B. Kocevska, C. Broussard, D. Plaseska-Karanfilska, Proteomic analysis of seminal plasma in men with different spermatogenic impairment, Andrologia 44 (2012) 256–264.
[26] I. Batruch, C.R. Smith, B.J. Mullen, E. Grober, K.C. Lo, E.P. Diamandis, K.A. Jarvi, Analysis of seminal plasma from patients with non-obstructive azoospermia and identification of candidate biomarkers of male infertility, J. Proteome Res. 11 (2012) 1503–1511.
[27] E.P. Diamandis, W.P. Arnett, G. Foussias, H. Pappas, S. Ghandi, D.N. Melegos, B. Mullen, H. Yu, J. Srigley, K. Jarvi, Seminal plasma biochemical markers and their association with semen analysis findings, Urology 53 (1999) 596–603.
[28] A.P. Drabovich, K. Jarvi, E.P. Diamandis, Verification of male infertility biomarkers in seminal plasma by multiplex selected reaction monitoring assay, Mol. Cell. Proteomics 10 (2011). M110 004127.
[29] S.M. Heshmat, J.B. Mullen, K.A. Jarvi, A. Soosaipillai, E.P. Diamandis, R. J. Hamilton, K.C. Lo, Seminal plasma lipocalin-type prostaglandin D synthase: a potential new marker for the diagnosis of obstructive azoospermia, J. Urol. 179 (2008) 1077–1080.
[30] A.D. Rolland, R. Lavigne, C. Dauly, P. Calvel, C. Kervarrec, T. Freour, B. Evrard, N. Rioux-Leclercq, J. Auger, C. Pineau, Identification of genital tract markers in the human seminal plasma using an integrative genomics approach, Hum. Reprod. 28 (2013) 199–209.
[31] I. Batruch, I. Lecker, D. Kagedan, C.R. Smith, B.J. Mullen, E. Grober, K.C. Lo, E. P. Diamandis, K.A. Jarvi, Proteomic analysis of seminal plasma from normal volunteers and post-vasectomy patients identifies over 2000 proteins and candidate biomarkers of the urogenital system, J. Proteome Res. 10 (2011) 941–953.
[32] A.P. Drabovich, A. Dimitromanolakis, P. Saraon, A. Soosaipillai, I. Batruch, B. Mullen, K. Jarvi, E.P. Diamandis, Differential diagnosis of azoospermia with proteomic biomarkers ECM1 and TEX101 quantified in seminal plasma, Sci. Transl. Med. 5 (2013), https://doi.org/10.1126/scitranslmed.3006260, 212ra160.
[33] D. Korbakis, D. Brinc, C. Schiza, A. Soosaipillai, K. Jarvi, A.P. Drabovich, E. P. Diamandis, Immunocapture-selected reaction monitoring screening facilitates the development of ELISA for the measurement of native TEX101 in biological fluids, Mol. Cell. Proteomics 14 (2015) 1517–1526.
[34] D. Korbakis, C. Schiza, D. Brinc, A. Soosaipillai, T.D. Karakosta, C. Legare, R. Sullivan, B. Mullen, K. Jarvi, E.P. Diamandis, A.P. Drabovich, Preclinical evaluation of a TEX101 protein ELISA test for the differential diagnosis of male infertility, BMC Med. 15 (2017) 60, https://doi.org/10.1186/s12916-017-0817-5.
[35] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254.
[36] Y.Q. Yu, M. Gilar, P.J. Lee, E.S. Bouvier, J.C. Gebler, Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins, Anal. Chem. 75 (2003) 6023–6028.
[37] E.S. Oswald, L.M. Brown, J.C. Bulinski, C.T. Hung, Label-free protein profiling of adipose-derived human stem cells under hyperosmotic treatment, J. Proteome Res. 10 (2011) 3050–3059.
[38] U. Distler, J. Kuharev, P. Navarro, Y. Levin, H. Schild, S. Tenzer, Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics, Nat. Methods 11 (2014) 167–170.
[39] K. Davalieva, S. Kiprijanovska, A. Dimovski, G. Rosoklija, A.J. Dwork, Comparative evaluation of two methods for LC-MS/MS proteomic analysis of formalin fixed and paraffin embedded tissues, J. Proteomics 235 (2021) 104117, https://doi.org/10.1016/j.jprot.2021.104117.
[40] U. Distler, J. Kuharev, P. Navarro, S. Tenzer, Label-free quantification in ion mobility-enhanced data-independent acquisition proteomics, Nat. Protoc. 11 (2016) 795–812.
[41] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D.J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Perez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A.F. Jarnuczak, T. Ternent, A. Brazma,J.A. Vizcaino, The PRIDE database and related tools and resources in 2019: improving support for quantification data, Nucleic Acids Res. 47 (2019) D442-D450.
[42] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B Methodol. 57 (1995) 289–300.
[43] Addinsoft, XLSTAT Statistical and Data Analysis Solution, New York, USA, https://www.xlstat.com/en, 2022.
[44] H. Mi, D. Ebert, A. Muruganujan, C. Mills, L.-P. Albou, T. Mushayamaha, P. D. Thomas, PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API, Nucleic Acids Res. 49 (2020) D394–D403.
[45] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C.V. Mering, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res. 47 (2019) D607–D613.
[46] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W. H. Fridman, F. Pages, Z. Trajanoski, J. Galon, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, Bioinformatics 25 (2009) 1091–1093.
[47] D. Djureinovic, L. Fagerberg, B. Hallstrom, A. Danielsson, C. Lindskog, M. Uhlen, F. Ponten, The human testis-specific proteome defined by transcriptomics and antibody-based profiling, Mol. Hum. Reprod. 20 (2014) 476–488.
[48] M. Uhlen, L. Fagerberg, B.M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C.A. Szigyarto, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Ponten, Proteomics. Tissue-based map of the human proteome, Science 347 (2015) 1260419, https://doi.org/10.1126/science.1260419.
[49] C. Pineau, F. Hikmet, C. Zhang, P. Oksvold, S. Chen, L. Fagerberg, M. Uhlen, C. Lindskog, Cell type-specific expression of testis elevated genes based on transcriptomics and antibody-based proteomics, J. Proteome Res. 18 (2019) 4215–4230.
[50] R. Edgar, M. Domrachev, A.E. Lash, Gene expression omnibus: NCBI gene expression and hybridization array data repository, Nucleic Acids Res. 30 (2002) 207–210.
[51] J.C. Oliveros, Venny. An Interactive Tool for Comparing Lists with Venn's Diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html, 2007.
[52] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method, Methods 25 (2001) 402–408.
[53] J. Cox, M. Mann, 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data, BMC Bioinforma. 13 (Suppl. 16) (2012) S12.
[54] M. Alikhani, M. Mirzaei, M. Sabbaghian, P. Parsamatin, R. Karamzadeh, S. Adib, N. Sodeifi, M.A.S. Gilani, M. Zabet-Moghaddam, L. Parker, Y. Wu, V. Gupta, P. A. Haynes, H. Gourabi, H. Baharvand, G.H. Salekdeh, Quantitative proteomic analysis of human testis reveals system-wide molecular and cellular pathways associated with non-obstructive azoospermia, J. Proteomics 162 (2017) 141–154.
[55] R. Huo, Y. He, C. Zhao, X.J. Guo, M. Lin, J.H. Sha, Identification of human spermatogenesis-related proteins by comparative proteomic analysis: a preliminary study, Fertil. Steril. 90 (2008) 1109–1118.
[56] J. Li, W. Guo, F. Li, J. He, Q. Yu, X. Wu, X. Mao, HnRNPL as a key factor in spermatogenesis: lesson from functional proteomic studies of azoospermia patients with sertoli cell only syndrome, J. Proteomics 75 (2012) 2879–2891.
[57] O.J. Gustafsson, G. Arentz, P. Hoffmann, Proteomic developments in the analysis of formalin-fixed tissue, Biochim. Biophys. Acta 2015 (1854) 559–580.
[58] M. Liu, Z. Hu, L. Qi, J. Wang, T. Zhou, Y. Guo, Y. Zeng, B. Zheng, Y. Wu, P. Zhang, X. Chen, W. Tu, T. Zhang, Q. Zhou, M. Jiang, X. Guo, Z. Zhou, J. Sha, Scanning of novel cancer/testis proteins by human testis proteomic analysis, Proteomics 13 (2013) 1200–1210.
[59] L. Zhao, C. Yao, X. Xing, T. Jing, P. Li, Z. Zhu, C. Yang, J. Zhai, R. Tian, H. Chen, J. Luo, N. Liu, Z. Deng, X. Lin, N. Li, J. Fang, J. Sun, C. Wang, Z. Zhou, Z. Li, Single-cell analysis of developing and azoospermia human testicles reveals central role of Sertoli cells, Nat. Commun. 11 (2020) 5683.
[60] E. Dube, D.G. Cyr, The blood-epididymis barrier and human male fertility, Adv. Exp. Med. Biol. 763 (2012) 218–236.
[61] L. Li, H. Li, L. Wang, S. Wu, L. Lv, A. Tahir, X. Xiao, C.K.C. Wong, F. Sun, R. Ge, C. Y. Cheng, Role of cell polarity and planar cell polarity (PCP) proteins in spermatogenesis, Crit. Rev. Biochem. Mol. Biol. 55 (2020) 71–87.
[62] H.L. Pahl, Activators and target genes of Rel/NF-kappaB transcription factors, Oncogene 18 (1999) 6853–6866.
[63] Y.N. Teng, P.J. Chuang, Y.W. Liu, Nuclear factor-kappaB (NF-kappaB) regulates the expression of human testis-enriched Leucine-rich repeats and WD repeat domain containing 1 (LRWD1) gene, Int. J. Mol. Sci. 14 (2012) 625–639.
[64] C. Rathke, W.M. Baarends, S. Awe, R. Renkawitz-Pohl, Chromatin dynamics during spermiogenesis, Biochim. Biophys. Acta (1839) 155–168 (2014.
[65] L. Wang, W. Yang, B. Li, S. Yuan, F. Wang, Response to stress in biological disorders: implications of stress granule assembly and function, Cell Prolif. 54 (2021) e13086.
[66] A. Kurita, T. Takizawa, T. Takayama, K. Totsukawa, S. Matsubara, H. Shibahara, M.C. Orgebin-Crist, F. Sendo, Y. Shinkai, Y. Araki, Identification, cloning, and initial characterization of a novel mouse testicular germ cell-specific antigen, Biol. Reprod. 64 (2001) 935–945.
[67] C. Schiza, D. Korbakis, E. Panteleli, K. Jarvi, A.P. Drabovich, E.P. Diamandis, Discovery of a human testis-specific protein complex TEX101-DPEP3 and selection of its disrupting antibodies, Mol. Cell. Proteomics 17 (2018) 2480–2495.
[68] H. Yoshitake, M. Yanagida, M. Maruyama, K. Takamori, A. Hasegawa, Y. Araki, Molecular characterization and expression of dipeptidase 3, a testis-specific membrane-bound dipeptidase: complex formation with TEX101, a germ-cell-specific antigen in the mouse testis, J. Reprod. Immunol. 90 (2011) 202–213.
[69] E. Montellier, F. Boussouar, S. Rousseaux, K. Zhang, T. Buchou, F. Fenaille, H. Shiota, A. Debernardi, P. Hery, S. Curtet, M. Jamshidikia, S. Barral, H. Holota, A. Bergon, F. Lopez, P. Guardiola, K. Pernet, J. Imbert, C. Petosa, M. Tan, Y. Zhao, M. Gerard, S. Khochbin, Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B, Genes Dev. 27 (2013) 1680–1692.
[70] X. Zhang, M. San Gabriel, A. Zini, Sperm nuclear histone to protamine ratio in fertile and infertile men: evidence of heterogeneous subpopulations of spermatozoa in the ejaculate, J. Androl. 27 (2006) 414–420.

[71] P. Intasqui, A. Agarwal, R. Sharma, L. Samanta, R.P. Bertolla, Towards the identification of reliable sperm biomarkers for male infertility: a sperm proteomic approach, Andrologia 50 (2018), https://doi.org/10.1111/and.12919.
[72] B. Drabent, C. Bode, B. Bramlage, D. Doenecke, Expression of the mouse testicular histone gene H1t during spermatogenesis, Histochem. Cell Biol. 106 (1996) 247–251.
[73] F. De Lucia, M.R. Faraone-Mennella, M. D'Erme, P. Quesada, P. Caiafa, B. Farina, Histone-induced condensation of rat testis chromatin: testis-specific H1t versus somatic H1 variants, Biochem. Biophys. Res. Commun. 198 (1994) 32–39.
[74] B. Drabent, P. Saftig, C. Bode, D. Doenecke, Spermatogenesis proceeds normally in mice without linker histone H1t, Histochem. Cell Biol. 113 (2000) 433–442.
[75] Z. Mollaee, R. Favaedi, P. Jazireian, P. Afsharian, A. Mohseni Meybodi, M. Shahhoseini, Genetic contribution of HIST1H1T regulatory region alternations to human nonobstructive azoospermia, Andrologia 52 (2020), https://doi.org/10.1111/and.13647 e13647.
[76] F.R. Bischoff, D. Gorlich, RanBP1 is crucial for the release of RanGTP from importin beta-related nuclear transport factors, FEBS Lett. 419 (1997) 249–254.
[77] B. Nathaniel, P.A.F. Whiley, Y. Miyamoto, K.L. Loveland, Importins: diverse roles in male fertility, Semin. Cell Dev. Biol. 121 (2022) 82–98.
[78] M. Nagai, T. Moriyama, R. Mehmood, K. Tokuhiro, M. Ikawa, M. Okabe, H. Tanaka, Y. Yoneda, Mice lacking ran binding protein 1 are viable and show male infertility, FEBS Lett. 585 (2011) 791–796.
[79] J. Bao, C. Tang, J. Li, Y. Zhang, B.P. Bhetwal, H. Zheng, W. Yan, RAN-binding protein 9 is involved in alternative splicing and is critical for male germ cell development and male fertility, PLoS Genet. 10 (2014), https://doi.org/10.1371/journal.pgen.1004825 e1004825.
[80] G.P. Deshpande, H.G. Patterton, M. Faadiel Essop, The human transketolase-like proteins TKTL1 and TKTL2 are bona fide transketolases, BMC Struct. Biol. 19 (2019) 2, https://doi.org/10.1186/s12900-018-0099-y.
[81] T. Soda, Y. Miyagawa, N. Ueda, K. Takezawa, H. Okuda, S. Fukuhara, K. Fujita, H. Kiuchi, M. Uemura, Y. Okamoto, A. Tsujimura, H. Tanaka, N. Nonomura, Systematic characterization of human testis-specific actin capping protein beta3 as a possible biomarker for male infertility, Hum. Reprod. 32 (2017) 514–522.
[82] B.P. Chadwick, J. Mull, L.A. Helbling, S. Gill, M. Leyne, C.M. Robbins, H. W. Pinkett, I. Makalowska, C. Maayan, A. Blumenfeld, F.B. Axelrod, M. Brownstein, J.F. Gusella, S.A. Slaugenhaupt, Cloning, mapping, and expression of two novel actin genes, actin-like-7A (ACTL7A) and actin-like-7B (ACTL7B), from the familial dysautonomia candidate region on 9q31, Genomics 58 (1999) 302–309.
[83] T. Clement, C. Geyer, W. Willis, E. Goulding, M. Eddy, Testis expressed actin-like 7b (Actl7b) is required for mouse spermatid morphogenesis and male fertility, Biol. Reprod. 87 (2012) 140.
[84] H. Tanaka, M. Hirose, K. Tokuhiro, Y. Matsuoka, Y. Miyagawa, A. Tsujimura, A. Okuyama, Y. Nishimune, Single nucleotide polymorphisms: discovery of the genetic causes of male infertility, Soc. Reprod. Fertil. Suppl. 65 (2007) 531–534.
[85] H. Tanaka, Y. Miyagawa, A. Tsujimura, M. Wada, Genetic polymorphisms within The Intronless ACTL7A and ACTL7B genes encoding spermatogenesis-specific actin-like proteins in Japanese males, Int. J. Fertil. Steril. 13 (2019) 245–249.
[86] X. Zhang, L. Wang, Y. Ma, Y. Wang, H. Liu, M. Liu, L. Qin, J. Li, C. Jiang, X. Zhang, X. Shan, Y. Liu, J. Li, Y. Li, R. Zheng, Y. Sun, J. Sun, X. Leng, Y. Liang, F. Zhang, X. Jiang, Y. Yang, Y. Shen, CEP128 is involved in spermatogenesis in humans and mice, Nat. Commun. 13 (2022) 1395, https://doi.org/10.1038/s41467-022-29109-7.
[87] O. Okutman, J. Muller, V. Skory, J.M. Garnier, A. Gaucherot, Y. Baert, V. Lamour, M. Serdarogullari, M. Gultomruk, A. Ropke, S. Kliesch, V. Herbepin, I. Aknin, M. Benkhalifa, M. Teletin, E. Bakircioglu, E. Goossens, N. Charlet-Berguerand, M. Bahceci, F. Tuttelmann, S. Viville, A no-stop mutation in MAGEB4 is a possible cause of rare X-linked azoospermia and oligozoospermia in a consanguineous Turkish family, J. Assist. Reprod. Genet. 34 (2017) 683–694.
[88] Y. Dong, Y. Pan, R. Wang, Z. Zhang, Q. Xi, R.Z. Liu, Copy number variations in spermatogenic failure patients with chromosomal abnormalities and unexplained azoospermia, Genet. Mol. Res. 14 (2015) 16041–16049.
[89] J.E.M. Dunleavy, H. Okuda, A.E. O'Connor, D.J. Merriner, L. O'Donnell, D. Jamsai, M. Bergmann, M.K. O'Bryan, Katanin-like 2 (KATNAL2) functions in multiple aspects of haploid male germ cell development in the mouse, PLoS Genet. 13 (2017), https://doi.org/10.1371/journal.pgen.1007078 e1007078.
[90] X. Wei, W. Liu, X. Zhu, Y. Li, X. Zhang, J. Chen, V. Isachenko, Y. Sha, Z. Lu, Biallelic mutations in KATNAL2 cause male infertility due to oligo-astheno-teratozoospermia, Clin. Genet. 100 (2021) 376–385.
[91] D.J. Moore, A. Onoufriadis, A. Shoemark, M.A. Simpson, P.I. zur Lage, S.C. De Castro, L. Bartoloni, G. Gallone, S. Petridi, W.J. Woollard, D. Antony, M. Schmidts, T. Didonna, P. Makrythanasis, J. Bevillard, N.P. Mongan, J. Djakow, G. Pals, J. S. Lucas, J.K. Marthin, K.G. Nielsen, F. Santoni, M. Guipponi, C. Hogg, S. E. Antonarakis, R.D. Emes, E.M. Chung, N.D. Greene, J.L. Blouin, A.P. Jarman, H. M. Mitchison, Mutations in ZMYND10, a gene essential for proper axonemal assembly of inner and outer dynein arms in humans and flies, cause primary ciliary dyskinesia, Am. J. Hum. Genet. 93 (2013) 346–356.
[92] X. He, C. Liu, X. Yang, M. Lv, X. Ni, Q. Li, H. Cheng, W. Liu, S. Tian, H. Wu, Y. Gao, C. Yang, Q. Tan, J. Cong, D. Tang, J. Zhang, B. Song, Y. Zhong, H. Li, W. Zhi, X. Mao, F. Fu, L. Ge, Q. Shen, M. Zhang, H. Saiyin, L. Jin, Y. Xu, P. Zhou, Z. Wei, F. Zhang, Y. Cao, Bi-allelic loss-of-function variants in CFAP58 cause flagellar axoneme and mitochondrial sheath defects and asthenoteratozoospermia in humans and mice, Am. J. Hum. Genet. 107 (2020) 514–526.
[93] C. Olesen, J. Silber, H. Eiberg, E. Ernst, K. Petersen, S. Lindenberg, N. Tommerup, Mutational analysis of the human FATE gene in 144 infertile men, Hum. Genet. 113 (2003) 195–201.
[94] X.X. Liu, L. Cai, F.J. Liu, An in silico analysis of human sperm genes associated with asthenozoospermia and its implication in male infertility, Medicine (Baltimore) 97 (2018), https://doi.org/10.1097/MD.0000000000013338 e13338.
[95] L. Yao, Y. Guo, X. Zhang, C. Xu, Y. Wang, X. Liu, Y. Li, Y. Qi, J. Sha, C. Qin, X. Yang, X. Guo, Quantitative proteomic biomarkers from extracellular vesicles of human seminal plasma in the differential diagnosis of azoospermia, Clin. Transl. Med. 11 (2021) e423.