Human Spermatozoa as a Model for Detecting Missing Proteins in the Context of the Chromosome-Centric Human Proteome Project

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · JULY 2015

Impact Factor: 4.25 · DOI: 10.1021/acs.jproteome.5b00170 · Source: PubMed

14 AUTHORS, INCLUDING:

Emmanuelle Com
French Institute of Health and Medical Resea…
44 PUBLICATIONS **477** CITATIONS
SEE PROFILE

Mélanie Lagarrigue
Université de Rennes 1
16 PUBLICATIONS **234** CITATIONS
SEE PROFILE

Charles Pineau
French Institute of Health and Medical Resea…
125 PUBLICATIONS **1,913** CITATIONS
SEE PROFILE

Journal of proteome research

Article

pubs.acs.org/jpr

Human Spermatozoa as a Model for Detecting Missing Proteins in the Context of the Chromosome-Centric Human Proteome Project

Fanny Jumeau,[†,∥] Emmanuelle Com,[‡,§] Lydie Lane,[⊥,¶] Paula Duek,[¶] Mélanie Lagarrigue,[‡,§] Régis Lavigne,[‡,§] Laëtitia Guillot,[‡,§] Karine Rondel,[‡,§] Alain Gateau,[¶] Nathalie Melaine,[‡,§] Blandine Guével,[‡,§] Nicolas Sergeant,[∥] Valérie Mitchell,[†] and Charles Pineau*[,‡,§]

[†]EA 4308-Department of Reproductive Biology-Spermiology-CECOS, CHRU-Lille, 59037, Lille cedex, France

[‡]Inserm U1085, Irset, Campus de Beaulieu, Rennes, 35042, France

[§]Protim, Campus de Beaulieu, Rennes, 35042, France

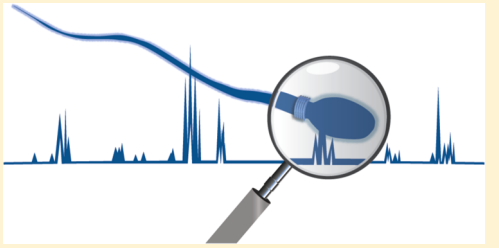
[⊥]Department of Human Protein Sciences, Faculty of Medicine, University of Geneva, 1, rue Michel-Servet, 1211 Geneva 4, Switzerland

[¶]SIB Swiss Institute of Bioinformatics, 1, rue Michel-Servet, 1211 Geneva 4, Switzerland

[∥]Inserm UMRS 1172, Centre Jean Pierre Aubert, IMPRT, University Lille Nord de France, 59045 Lille Cedex, France

Supporting Information

ABSTRACT: The Chromosome-Centric Human Proteome Project (C-HPP) aims at cataloguing the proteins as gene products encoded by the human genome in a chromosome-centric manner. The existence of products of about 82% of the genes has been confirmed at the protein level. However, the number of so-called "missing proteins" remains significant. It was recently suggested that the expression of proteins that have been systematically missed might be restricted to particular organs or cell types, for example, the testis. Testicular function, and spermatogenesis in particular, is conditioned by the successive activation or repression of thousands of genes and proteins including numerous germ cell- and testis-specific products. Both the testis and postmeiotic germ cells are thus promising sites at which to search for missing proteins, and ejaculated spermatozoa are a potential source of proteins whose expression is restricted to the germ cell lineage. A trans-chromosome-based data analysis was performed to catalog missing proteins in total protein extracts from isolated human spermatozoa. We have identified and manually validated peptide matches to 89 missing proteins in human spermatozoa. In addition, we carefully validated three proteins that were scored as uncertain in the latest neXtProt release (09.19.2014). A focus was then given to the 12 missing proteins encoded on chromosomes 2 and 14, some of which may putatively play roles in ciliation and flagellum mechanistics. The expression pattern of C2orf57 and TEX37 was confirmed in the adult testis by immunohistochemistry. On the basis of transcript expression during human spermatogenesis, we further consider the potential for discovering additional missing proteins in the testicular postmeiotic germ cell lineage and in ejaculated spermatozoa. This project was conducted as part of the C-HPP initiatives on chromosomes 14 (France) and 2 (Switzerland). The mass spectrometry proteomics data have been deposited with the ProteomeXchange Consortium under the data set identifier PXD002367.



KEYWORDS: *human proteome project, chromosome 2, chromosome 14, spermatozoon, missing proteins, mass spectrometry proteomics*

INTRODUCTION

Spermatogenesis is a complex, intricate, highly controlled, and specialized process, and takes place in the seminiferous tubules in the mammalian testis.[1] It involves the germ cell-specific events associated with meiosis, but during postmeiotic maturation, the haploid spermatids undergo unique events including chromatin remodeling, repackaging, and transcriptional reprogramming. During this process, pluripotent and nonpolar germ cells differentiate into a highly polarized and functional sperm, or spermatozoa. However, spermatozoa are not motile or fertile upon their release from the testis. Their proteins undergo a series of discrete modifications as the spermatozoa transit through the epididymis including post-translational modifications,[2] proteolytic processing, and integration of new components originating from epididymal secretions.[3] The absence of expression, or even an increase or decrease, of some spermatozoa proteins can lead to abnormal morphology, altered or poor motility, and fertilization failures (for a review see ref 4). The molecular basis of this post-testicular maturation remained unexplored for decades.[5]

Special Issue: The Chromosome-Centric Human Proteome Project 2015

Received: February 23, 2015

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

However, in recent years proteomics has made a substantial contribution to the understanding of sperm epididymal maturation.[2,3,6,7] It is thus plausible that deciphering the sperm proteome will help to establish the final protein repertoire of the testicular germ cell lineage, including those modifications that occur during the transit of sperm through the long and unique epididymal tubule.[3,7]

The male gamete is the subject of a unique fate. It plays a crucial role in genomic transmission, and most genes contributing to spermatogenesis are under particular selective pressure. Spermatogenesis has some extraordinary features such as a high degree of germ cell-specific gene expression and delays between transcription and translation.[8] The sperm also has hyperselected traits associated with shape and energy production as well as enhanced flagellum development. Human spermatozoa therefore appear to be a promising substrate to explore the human proteome.

The Chromosome-Centric Human Proteome Project (C-HPP) aims to catalogue proteins as gene products encoded by the human genome in a chromosome-centric manner and to characterize their isoforms and functions.[9] In this context, numerous initiatives worldwide have been launched to search for the so-called "missing proteins", defined as proteins predicted by genomic or transcriptomic analysis but not yet experimentally demonstrated with high degrees of confidence by mass-spectrometry or antibody-based techniques. As a result, about 900 missing proteins were experimentally validated between October 2013 and October 2014. neXtProt[10] has been selected as the C-HPP reference knowledgebase for human protein annotation. The current version of neXtProt (09.19.2014 release) integrates the latest version of PeptideAtlas (07.31.2014 release) and contains over 1 million peptides mapping to 16 635 protein entries, taking into account contributions from two large data sets recently published.[11,12] By using these data and other mass-spectrometry and antibody-based evidence, 82% of the protein entries have now been labeled PE1 (experimental evidence at protein level). However, the number of undetected putative proteins remains significant. Indeed, there are still 2948 entries corresponding to missing proteins, labeled PE2 (experimental evidence at transcript level), PE3 (protein inferred from homology), or PE4 (protein predicted). There are also 616 entries corresponding to uncertain proteins (PE5), probably arising from incorrect predictions. It should be noted that the unavoidable time lag between data publication and protein annotation results in experimentally validated proteins still being annotated as "missing" (PE2−4).[10]

Reasons why the so-called missing proteins have not been experimentally validated have been detailed extensively and include low abundance, time- or stress-dependent or organ-specific expression, particular physicochemical properties (hydrophobicity, amino-acid composition), or protein instability. Several of these issues can arise simultaneously or sequentially, resulting in complicated situations. Lane and collaborators suggested that the production of proteins that have been systematically missed might be restricted to unusual organs or cell types, for example, the testis.[13] In an impressive recent study from the Human Protein Atlas, Uhlen and collaborators confirm that the largest number (n = 999) of tissue-enriched genes is found in the testis,[14] this organ being considered as an outlier when looking for protein-coding genes. This is in accordance with an in-depth analysis of human RNA-seq data showing that certain organs such as the brain and especially the testis express more protein-coding genes than others.[15] Indeed, substantially more genic regions are transcribed in the testis than in other organs.[16] As many as 1124 genes were identified as highly enriched (n = 364) and moderately enriched (n = 760) in the testis.[17] Approximately half of the testis-specific genes are expressed in late spermatogenesis, sperm (30%), and spermatids (21%), whereas a relatively small number of testis-specific transcripts were identified in the premeiotic germ cell lineage or somatic testicular cells (i.e., Sertoli and Leydig cells).[17] Therefore, we believe that postmeiotic germ cells, including mature spermatozoa, are potentially fruitful sites to search for missing proteins. It is difficult to obtain isolated postmeiotic germ cells from human testes, whereas ejaculated spermatozoa are straightforward to obtain noninvasively: spermatozoa are thus a practical alternative to search for proteins whose expression is restricted or specific to the germ cell lineage. Here, we report work as part of C-HPP initiatives on chromosomes 14 (France) and 2 (Switzerland) aiming to identify missing proteins in ejaculated human spermatozoa.

We performed a trans-chromosome-based data analysis to catalogue missing proteins in total protein extracts from isolated human spermatozoa using a high quality mass spectrometry data set, which was collected on a high-resolution mass analyzer. We validated unambiguous peptide matches to 89 missing proteins mapping to diverse chromosomes, corresponding to 3% of the 2948 missing proteins referenced in the latest neXtProt 09.19.2014 release. We also carefully validated three proteins that were considered as uncertain by neXtProt. We then studied missing proteins encoded on chromosomes 2 and 14, some of which may be involved in ciliation and flagellum mechanics. Distribution of C2orf57 and TEX37 was further studied by immunohistochemistry in the adult testis, and their expression was confirmed in postmeiotic germ cells. Transcript abundance during human spermatogenesis indicates that it may be possible to discover additional missing proteins in the testicular postmeiotic germ cell lineage and in ejaculated spermatozoa.

MATERIALS AND METHODS

Sample Collection and Preparation

Human semen samples were collected from healthy donors of unproven fertility at the Lille University Medical Center (France). Men gave informed consent for the use of their semen for research purposes, and samples were anonymized. Semen samples were all obtained on site by masturbation following 2 to 7 days of sexual abstinence. Upon liquefaction for 30 min at 37 °C, a protease inhibitor mix (protease inhibitor cocktail tablets, complete mini EDTA free, Roche, Meylan, France) was added to samples according to the manufacturer's instructions. Sperm cells were then separated from seminal plasma and round cells by density gradient centrifugation in 50% Ferticult (Fertipro, Beernem, Belgium) and 50% PureSperm (Nidacon, Mölndal, Sweden) in sterile 15 mL tubes (Falcon, VWR, Radnor, PA, USA). After centrifugation at 350g for 20 min at room temperature, the sperm pellet was washed once by resuspension in 1 mL of Tris-buffered saline (TBS: Tris-HCl 0.1 mM pH 7.6, 100 mM NaCl) and spun at 350g for 20 min at room temperature.

Protein Extraction, Prefractionation, and Digestion

The spermatozoa pellet was resuspended in lysis buffer containing 2 M thiourea, 8 M Urea, 4% CHAPS, and 1%

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

Nuclease Mix (GE Healthcare, Orsay, France). The samples were sonicated on ice (40 Hz, 60 pulses) and centrifuged at 14 000*g* for 20 min at 4 °C. The supernatants were recovered, and pellets were discarded. Protein concentrations were assayed colorimetrically using the Bradford protein assay (Bradford Assay; BioRad, Marnes-la-Coquette, France) according to the manufacturer's instructions. Protein lysates were stored at −80 °C until use.

Aliquots of 50 μg of sperm protein extract from pools of five different ejaculates were diluted in LDS two times (lithium dodecyl sulfate, Invitrogen Life Technologies, Saint-Aubin, France) according to the manufacturer's instructions. These samples were boiled at 100 °C for 10 min, briefly spun at 500*g* in a bench centrifuge, and the supernatant loaded onto 12-well 4−12% acrylamide gel (Criterion, BioRad, Hercules, CA, USA). Electrophoresis was performed under a continuous voltage of 200 V for 1 h. Gels were stained with Coomassie Blue [0.1% Blue G250 (BioRad), 50% ethanol (v/v), 10% acetic acid (v/v) in H_20] and then washed in a destaining solution containing 7% (v/v) acetic acid, 10% ethanol in H_20 to reveal the bands. Each gel lane was manually cut into 20 slices of similar size. Proteins in the gel slices were reduced, alkylated, and digested with modified trypsin (Promega, Charbonnières, France), and the peptides were extracted as previously described.[18]

Liquid Chromatography−Tandem Mass Spectrometry (LC−MS/MS) Analyses

Shotgun analyses of spermatozoa extracts were conducted on a LTQ-Orbitrap XL (ThermoFisher Scientific) mass spectrometer. The MS measurements were done with a nanoflow high-performance liquid chromatography (HPLC) system (Dionex, LC Packings Ultimate 3000) connected to a hybrid LTQ-Orbitrap XL (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source (New Objective). The HPLC system consisted of a solvent degasser nanoflow pump, a thermostated column oven kept at 30 °C, and a thermostated autosampler kept at 8 °C to reduce sample evaporation. Mobile A (99.9% Milli-Q water and 0.1% formic acid (v:v)) and B (99.9% acetonitrile and 0.1% formic acid (v:v)) phases for HPLC were delivered by the Ultimate 3000 nanoflow LC system (Dionex, LC Packings). An aliquot of 10 μL of prepared peptide mixture was loaded onto a trapping precolumn (5 mm × 300 μm i.d., 300 Å pore size, Pepmap C18, 5 μm) over 3 min in 2% buffer B at a flow rate of 25 μL/min. This step was followed by reverse-phase separations at a flow rate of 0.250 μL/min using an analytical column (15 cm × 300 μm i.d., 300 Å pore size, Pepmap C18, 5 μm, Dionex, LC Packings). We ran a gradient from 2−35% buffer B for the first 60 min, 35−60% buffer B from minutes 60−85, and 60−90% buffer B from minutes 85−105. Finally, the column was washed with 90% buffer B for 16 min and with 2% buffer B for 19 min before the next sample was loaded. The peptides were detected by directly eluting them from the HPLC column into the electrospray ion source of the mass spectrometer. An electrospray ionization (ESI) voltage of 1.6 kV was applied to the HPLC buffer using the liquid junction provided by the nanoelectrospray ion source, and the ion transfer tube temperature was set to 200 °C.

The LTQ-Orbitrap XL instrument was operated in its data-dependent mode by automatically switching between full survey scan MS and consecutive MS/MS acquisitions. Survey full scan MS spectra (mass range 400−2000) were acquired in the Orbitrap section of the instrument with a resolution of r = 60 000 at m/z 400; ion injection times were calculated for each spectrum to allow for accumulation of 10^6 ions in the Orbitrap. The ten most intense peptide ions in each survey scan with an intensity above 2000 counts (to avoid triggering fragmentation too early during the peptide elution profile) and a charge state $\geq$ 2 were sequentially isolated at a target value of 10 000 and fragmented in the linear ion trap by collision-induced dissociation. Normalized collision energy was set to 35% with an activation time of 30 ms. Peaks selected for fragmentation were automatically put on a dynamic exclusion list for 60 s with a mass tolerance of ±10 ppm to avoid selecting the same ion for fragmentation more than once. The following parameters were used: the repeat count was set to 1, the exclusion list size limit was 500, singly charged precursors were rejected, and the maximum injection time was set at 500 and 300 ms for full MS and MS/MS scan events, respectively. For an optimal duty cycle, the fragment ion spectra were recorded in the LTQ mass spectrometer in parallel with the Orbitrap full scan detection. For Orbitrap measurements, an external calibration was used before each injection series ensuring an overall error mass accuracy below 5 ppm for the detected peptides. MS data were saved in RAW file format (Thermo Fisher Scientific) using XCalibur 2.0.7 with tune 2.4.

The data analysis was performed with the Proteome Discoverer 1.2 software supported by Mascot (Mascot server v2.2.07; http://www.matrixscience.com) database search engine for peptide and protein identification using its automatic decoy database search to calculate a false discovery rate (FDR) of 1% at the peptide level. MS/MS spectra were compared to the UniProt Human Reference proteome set database (UniProt release 2014_10, October 29 2014, 67084 sequences, 22 185 697 residues). Mass tolerance for MS and MS/MS was set at 10 ppm and 0.5 Da, respectively. The enzyme selectivity was set to full trypsin with one miscleavage allowed. Protein modifications were fixed carbamidomethylation of cysteines, variable oxidation of methionine, variable acetylation of lysine, variable acetylation of N-terminal residues, and variable deamidation of asparagine and glutamine. To increase the number of identified proteins, each peptide mixture prepared was injected three times using exclusion lists. We employed Proteome Discoverer to generate the first exclusion list containing peptides identified in the first run. This exclusion list of peptides was exported as a text file containing uncharged and accurate mass values and a retention time window of approximately 1 min. A mass tolerance of ±10 ppm was used to reject previously identified peptides within the specified retention time window. Using lower values can lead to reselection of masses because in the parallel mode of operation on an LTQ Orbitrap XL, the parent ion selection for an ion trap MS/MS is based on an Orbitrap preview scan that is acquired at a lower resolution (RP 15 000) than the final Orbitrap full scan, and therefore the masses are less accurate. These previously identified peptides were excluded in the second LC−MS/MS analysis. Peptide exclusion lists obtained from the first two runs were combined and used for the third analysis.

Data Availability

The raw MS proteomics data have been deposited with the ProteomeXchange Consortium[19] via the PRIDE partner repository with the data set identifier PXD002367. It should be noted that, in our hands, the complete PRIDE submission was only possible by converting our.dat files from Mascot into

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

pride.xml files using pride converter (version 2.0). Indeed, the conversion of .msf files into pride.xml files using Pride Converter or into mzIdentML using ProCon (version 0.9.600) led to the loss of peptide modifications or to the absence of spectra annotations.

Validation of Spectra for Protein Identification

Identification results from Mascot (.dat files) were imported into the Proline Studio software version 1.0.0 (http://proline.profiproteomics.fr). This software was then used to validate protein identification according to the recommendations of the French Proteomics Infrastructure (ProFi; www.profiproteomics.fr).[20] Only identifications results satisfying the following filters were considered: (i) a Mascot ion score >30; (ii) a peptide length ≥ 7; (iii) a peptide rank = 1; (iv) a FDR of 1% on the e-value at the peptide level; and (v) a FDR of 1% at the protein level. Only the 12 062 peptides and 1547 proteins satisfying these criteria were considered in our identification data set. Peptides not fulfilling these criteria were filtered out automatically from our original identification data set.

Only unique peptides should be considered for any identification as part of the C-HPP. Classical rules for trypsin cleavage were followed to build a data set of 153 245 unique peptides from the 2948 missing proteins, plus the 616 uncertain cases (PE5) (09.19.2014 release of neXtProt). The data set was further employed to filter out all nonunique peptides from the Proline Studio software version 1.0.0 peptide identification data set filtered first with ProFi criteria.

For proteins with evidence based on a single peptide identification passing the above automatic criteria, manual analysis was used to check the quality of PSM. PSM were analyzed in a blind manner by three independent experts who graded their quality (high, medium, and low) according to (i) the quality of y-ion or b-ion series; (ii) peak intensities; and (iii) the presence of nonassigned peaks. PSM quality was considered as "high" if at least four successive y- or b-ions were observed, the peak intensity was high, and very few or no peaks were nonassigned. "Medium" quality PSMs had at least four successive y- or b-ions but with lower peak intensities or more nonassigned peaks. PSM quality was considered as "low" if a few successive y- or b-ion series were observed with low global peak intensity or a large number of nonassigned peaks. Low-quality PSMs were then removed from the final data set of identified missing proteins.

Gene Ontology Term Enrichment Analyses

Functional annotation clustering of Gene Ontology terms enriched in identified missing proteins was performed using DAVID bioinformatics resources.[21] The enriched GO terms in the data set were selected with the highest classification stringency using the functional annotation-clustering tool in DAVID. Enriched GO terms were clustered into groups displaying similar/relevant annotations, which are classified according to their enrichment score, for "biological process", "cellular component", and "molecular function" functional annotations.

Deep Datamining for Identified Missing Proteins Encoded on Chromosomes 2 and 14

For each protein confidently identified by MS, the tissular expression profile obtained by RNA sequencing analysis was retrieved from the Human Protein Atlas portal (version 13),[22] and the conservation profile was determined by BLAST analysis on UniProtKB. All possible names, synonyms, and identifiers were collected from appropriate model organism databases for each protein and all its orthologs. These names were used to query PubMed and Google to search for potential clues to function.

Immunohistochemistry

To confirm the germline expression of proteins of interest, immunohistochemical experiments were performed on human testes, fixed in 4% paraformaldehyde fixative and embedded in paraffin, as described.[23] Human testes were obtained from patients undergoing therapeutic orchidectomy for metastatic prostate carcinoma. The protocol was approved by the Ethical Committee of Rennes (France) (Authorization n°DC-2010−1155, June 15 2011), and written informed consent was obtained from all donors. Thin sections (5 μm thick) were deparaffinized, rehydrated, and incubated for 20 min at 80 °C in citrate buffer (10 mM pH 6.0) with 0.05% Tween 20 for antigen retrieval. The sections were saturated for 1 h with 5% bovine serum albumin (BSA) in TBS, then incubated overnight at 4 °C with the rabbit polyclonal anti-C2orf57 (HPA049917) or anti-TEX37 (HPA043987) protein antibodies (Human Protein Atlas) used at a final dilution of 1:50 (HPA049917) or 1:400 (HPA043987), in TBS containing 0.2% Tween-20 (v/v) and 3% BSA (TBST-BSA). Nonimmune serum (1:1000) was used as a negative control. After several washes in TBS, sections were incubated for 45 min with a secondary biotinylated goat antirabbit antibody (Dako, Trappes, France) at a final dilution of 1:500 in TBST-BSA. Samples were subsequently washed in TBS and incubated for an additional 30 min with a streptavidin−peroxidase complex (Dako) at a dilution of 1:500 in TBST. Immunoreaction was developed by 5 min of incubation with a diaminobenzidine solution (Sigma-Aldrich, Saint-Quentin Fallavier, France). Finally, sections were counterstained with Masson hemalun, dehydrated, mounted in Eukitt (Labnord, Villeneuve d'Asq, France), and visualized under an AX60-PS microscope (Olympus, Rungis, France).

Evaluation of the Number of Missing Proteins Potentially Detectable in the Testis and Postmeiotic Male Germ Cells

Previously processed transcriptomics data, available in public array data repositories, from several infertile and healthy human testicular samples (EBI ArrayExpress accession E-TABM-1214[24]); isolated human seminiferous tubules and testicular cells (EBI ArrayExpress accession E-TABM-130[25]); and from 45 human nontesticular healthy tissues (NCBI Gene Expression Omnibus accession GSE7307) were downloaded.

Several classification systems have been described for scoring spermatogenesis in humans, including the widely quoted Johnsen score.[26] The Johnsen score is a quantitative histological grading system in which the level of sperm maturation is graded between 1 and 10 according to the most advanced germ cell category observed in at least 100 seminiferous tubules sections. It was used by Chalmel and collaborators[24] to define the "human testis gene expression program" that includes 3850 genes organized into four classes: (i) "specific expression in testis" (SET: genes expressed in male gonads but not in any of the 45 nontesticular controls); (ii) "preferential expression in testis" (PET: genes for which transcripts were also detected in <3 controls); (iii) "intermediate expression in testis" (IET: genes detected in −19 controls); and (iv) "ubiquitous expression" (UEX: genes detected in >20 controls).

The strategy employed for the visualization of missing proteins expressed in the testis/germ cells was similar to that

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 1. List of the 94 Newly Detected Missing Proteins in the Human Spermatozoa[a]

accession number	gene name	description	protein evidence	chrom/ location	identified peptides (MS/MS)	identified unique peptides (MS/MS)	spectrum quality from manual validation when requested	mRNA belongs to the "testis gene expression program"[b]
Q7Z5J8	ANKAR	ankyrln and armadillo repeat-containing protein	PE2	2q32.2	1	1	medium	SET
Q9NU02	ANKEF1	ankyrin repeat and EF-hand domain-containing protein 1	PE2	20p12.2	5	3		–
Q8IYS8	BOD1L2	putative biorientation of chromosomes in cell division protein 1-like 2	PE5	18q21.31	4	3		SET
Q32M84	BTBD16	BTB/POZ domain-containing protein 16	PE2	10q26.13	2	2		PET
Q8IVU9	C10orf107	uncharacterized protein C10orf107	PE2	10q21.2	1	1	high	
Q8WW14	C10orf82	uncharacterized protein C10orf82	PE2	10q25.3	4	4		SET
H3BRN8	C15orf65	uncharacterized protein C15orf65	PE2	15q21.3	3	2		
Q0P670	C17orf74	uncharacterized protein C17orf74	PE2	17p13.1	1	1	high	SET
A8MV24	C17orf98	uncharacterized protein C17orf98	PE4	17q12	1	1	medium	–
Q5SVJ3	C1orf100	uncharacterized protein C1orf100	PE2	1q44	2	2		SET
Q8N1D5	C1orf158	uncharacterized protein C1orf158	PE2	1p36.21	1	1	high	IE
Q5VU69	C1orf189	uncharacterized protein C1orf189	PE2	1q21.3	3	2		
Q5VTH2	C1orf192	UPF0740 protein C1orf192	PE2	1q23.3	5	4		PET
Q5T5A4	C1orf194	uncharacterized protein C1orf194	PE2	1p13.3	6	6		UE
Q8NHU2	C20orf26	uncharacterized protein C20orf26	PE2	20p11.23	3	2		SET
Q9H1P6	C20orf85	uncharacterized protein C20orf85	PE2	20q13.32	4	3		IE
Q68DN1	C2orf16	uncharacterized protein C2orf16	PE2	2p23.3	2	1	high	UE
Q53QW1	C2orf57	uncharacterized protein C2orf57	PE2	2q37.1	3	3		UE
Q8N801	C2orf61	uncharacterized protein C2orf61	PE2	2p21	1	1	high	
A6NJV1	C2orf70	UPF0573 protein C2orf70	PE2	2p23.3	3	2		
Q96M34	C3orf30	uncharacterized protein C3orf30	PE2	3q13.32	6	5		SET
Q6 V702	C4orf22	uncharacterized protein C4orf22	PE2	4q21.21	2	2		SET
A7E2U8	C4orf47	UPF0602 protein C4orf47	PE2	4q35.1	2	2		PET
A4QMS7	C5orf49	uncharacterized protein C5orf49	PE2	5p15.31	2	1	high	PET
A4D263	C7orf72	uncharacterized protein C7orf72	PE4	7p12.2	2	2		
Q5BN46	C9orf116	UPF0691 protein C9orf116	PE2	9q34.3	6	5		
Q9H7T0	CATSPERB	cation channel sperm-associated protein subunit beta	PE2	14q32.12	2	2		SET
Q8IYK2	CCDC105	coiled-coil domain-contain ng protein 105	PE2	19p13.12	14	12		
Q6ZU64	CCDC108	coiled-coil domain-contain ng protein 108	PE2	2q35	4	3		
Q96M63	CCDC114	coiled-coil domain-containing protein 114	PE2	19q13.33	1	1	high	PET
Q8IYE0	CCDC146	coiled-coil domain-containing protein 146	PE2	7q11.23	2	1	high	UE
Q2M243	CCDC27	coiled-coil domain-containing protein 27	PE2	1p36.32	2	2		–
Q96M95	CCDC42	coiled-coil domain-containing protein 42A	PE2	17p13.1	5	5		SET
Q8NA47	CCDC63	coiled-coil domain-containing protein 63	PE2	12q24.11	14	13		UE
Q6ZN84	CCDC81	coiled-coil domain-containing protein 81	PE2	11q14.2	3	2		

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 1. continued

accession number	gene name	description	protein evidence	chrom/ location	identified peptides (MS/MS)	identified unique peptides (MS/MS)	spectrum quality from manual validation when requested	mRNA belongs to the "testis gene expression program"[b]
Q8NDM7	CFAP43	cilia- and flagella-associated protein 43	PE2	10q25.1	5	4		SET/IE
Q8IYW2	CFAP46	protein CFAP46	PE2	10q26.3	1	1	medium	PET
Q8WXQ8	CPA5	carboxypeptidase A5	PE2	7q32.2	2	1	high	UE
Q6ZTR5	CXorf22	uncharacterized protein CXorf22	PE2	Xp21.1	1	1	high	
A6PW82	CXorf30	putative uncharacterized protein CXorf30	PE2	Xp21.1	3	2		
Q6P9G0	CYB5D1	cytochrome b5 domain-containing protein 1	PE2	17p13.1	1	1	high	
A8MYV0	DCDC2C	doublecortin domain-containing protein 2C	PE4	2p25.3	5	4		
Q9P2D7	DNAH1	dynein heavy chain 1, axonemal	PE2	3p21.1	27	26		
Q6ZR08	DNAH12	dynein heavy chain 12, axonemal	PE2	3p14.3	33	20		IE
Q9P225	DNAH2	dynein heavy chain 2, axonemal	PE2	17p13.1	22	19		PET
Q8TD57	DNAH3	dynein heavy chain 3, axonemal	PE2	16p12.3	32	20		
Q96MC2	DRC1	Dynein regulatory complex protein 1	PE2	2p23.3	6	4		UE
Q9HAE3	EFCAB1	EF-hand calcium-binding domain-containing protein 1	PE2	8q11.21	5	5		–
Q5VUJ9	EFCAB2	EF-hand calcium-binding domain-containing protein 2	PE2	1q44	1	1	high	SET/PET
Q8N7B9	EFCAB3	EF-hand calcium-binding domain-containing protein 3	PE2	17q23.2	1	1	high	SET
Q8TC29	ENKUR	enkurin	PE2	10p12.1	8	7		–
A6NNW6	ENO4	enolase-like protein ENO4	PE2	10q25.3	1	1	high	
Q6ZVS7	FAM183B	protein FAM183B	PE2	7p14.1	1	1	high	
A6NFA0	FAM205CP	putative family with sequence similarity 205, member C protein	PE5	9p13.3	4	2		
Q8NEG0	FAM71C	protein FAM71C	PE2	12q23.1	2	1	high	SET
Q6IPT2	FAM71E1	protein FAM71E1	PE2	19q13.33	1	1	high	UE
Q8N5Q1	FAM71E2	putative protein FAM71E2	PE5	19q13.42	4	3		
Q14409	GK3P	putative glycerol kinase 3	PE5	4q32.3	12	1	medium	
Q4G1C9	GLIPR1L2	GLIPR1-like protein 2	PE2	12q21.2	2	2		
Q96DY2	IQCD	IQ domain-containing protein D	PE2	12q24.13	3	1	high	IE
Q8N6M8	IQCF1	IQ domain-containing protein F1	PE2	3p21.2	1	1	medium	PET
P0C7M6	IQCF3	IQ domain-containing protein F3	PE2	3p21.2	2	2		SET
Q6UXV1	IZUMO2	izumo sperm−egg fusion protein 2	PE2	19q13.33	4	4		PET
Q5VZ72	IZUMO3	izumo sperm−egg fusion protein 3	PE2	9p21.3	3	3		
O60309	LRRC37A3	leucine-rich repeat-containing protein 37A3	PE2	17q24.1	26	3		IE
Q96E66	LRTOMT	leucine-rich repeat-containing protein 51	PE2	11q13.4	3	3		UE
Q6UWQ5	LYZL1	lysozyme-like protein 1	PE2	10p11.23	4	2		SET
Q5I0G3	MDH1B	putative malate dehydrogenase 1B	PE2	2q33.3	2	1	high	PET/IE
Q6PF18	MORN3	MORN repeat-containing protein 3	PE2	12q24.31	3	2		UE
Q5VZ52	MORN5	MORN repeat-containing protein 5	PE2	9q33.2	2	2		IE
Q9H579	MROH8	protein MROH8	PE2	20q11.23	2	2		SET
B3GLJ2	PATE3	prostate and testis expressed protein 3	PE2	11q24.2	4	3		
Q86YW0	PLCZ1	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase zeta-1	PE2	12p12.3	10	3		SET
Q96LZ3	PPP3R2	calcineurin subunit B type 2	PE2	9q31.1	1	1	high	SET

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 1. continued

accession number	gene name	description	protein evidence	chrom/ location	identified peptides (MS/MS)	identified unique peptides (MS/MS)	spectrum quality from manual validation when requested	mRNA belongs to the "testis gene expression program"[b]
A6NNX1	RIIAD1	RIIa domain-containing protein 1	PE4	1q21.3	2	2		UE
Q5GAN3	RNASE13	probable inactive ribonuclease-like protein 13	PE2	14q11.2	1	1	high	
Q8NCS7	SLC44A5	choline transporter-like protein 5	PE2	1p31.1	3	3		–
Q499Z3	SLFNL1	Schlafen-like protein 1	PE2	1p34.2	2	2		UE
P0C874	SPATA31D3	putative spermatogenesis-associated protein 31D3	PE5	9q21.32	7	1	high	
Q6ZUB1	SPATA31E1	spermatogenesis-associated protein 31E1	PE2	9q22.1	1	1	high	–
Q8N4L4	SPEM1	spermatid maturation protein 1	PE2	17p13.1	1	1	high	UE
Q5JU00	TCTE1	T-complex-associated testis-expressed protein 1	PE2	6p21.1	2	1	high	SET
Q6URK8	TEPP	testis, prostate, and placenta-expressed protein	PE2	16q21	7	5		SET
Q8N6G2	TEX26	testis-expressed sequence 26 protein	PE2	13q12.3	2	1	high	SET
Q96LM6	TEX37	testis-expressed sequence 37 protein	PE2	2p11.2	3	1	high	UE
Q6ZNM6	TEX43	testis-expressed sequence 43 protein	PE2	5q23.2	3	3		SET
Q8WW34	TMEM239	transmembrane protein 239	PE2	20p13	1	1	high	
Q2TAA8	TSNAXIP1	translin-associated factor X-interacting protein 1	PE2	16q22.1	2	2		UE
Q8NA56	TTC29	tetratricopeptide repeat protein 29	PE2	4q31.22	3	2		PET
Q8IWG1	WDR63	WD repeat-containing protein 63	PE2	1p22.3	4	4		PET
B1ANS9	WDR64	WD repeat-containing protein 64	PE2	1q43	2	2		SET
Q5VTH9	WDR78	WD repeat-containing protein 78	PE2	1p31.3	1	1	high	SET/PET
Q8IUA0	WFDC8	WAP four-disulfide core domain protein 8	PE2	20q13.12	7	7		
Q9H0C1	ZMYND12	zinc finger MYND domain-containing protein 12	PE2	1p34.2	4	3		IE

[a]Detailed information on the peptide sequence is available in Supplementary Table 1 of the Supporting Information. [b]Gene classes in the "testis gene expression program" (Chalmel et al., 2012 Human Reprod): specifically expressed in testis (SET); preferentially expressed in testis (PET); with intermediate expression (IE); or ubiquitous expression (UE) in testis; seen but not assigned to a class (–).

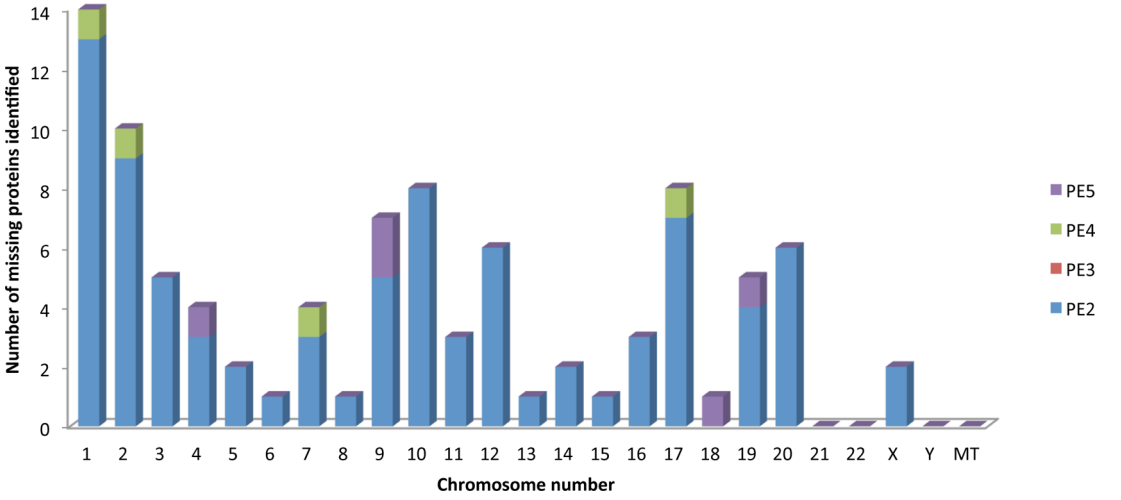


Figure 1. Chromosome assignation and protein evidence status of missing proteins identified in the human spermatozoa. The neXtProt release used for the protein evidence annotation is that of 09.19.2014.

previously described by our team.[27] Briefly, accession numbers of neXtProt missing protein entries (PE2, PE3, and PE4; 09.19.2014 release) that correspond to Affymetrix probesets of the "human testis gene expression program" were retrieved

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX–XXX

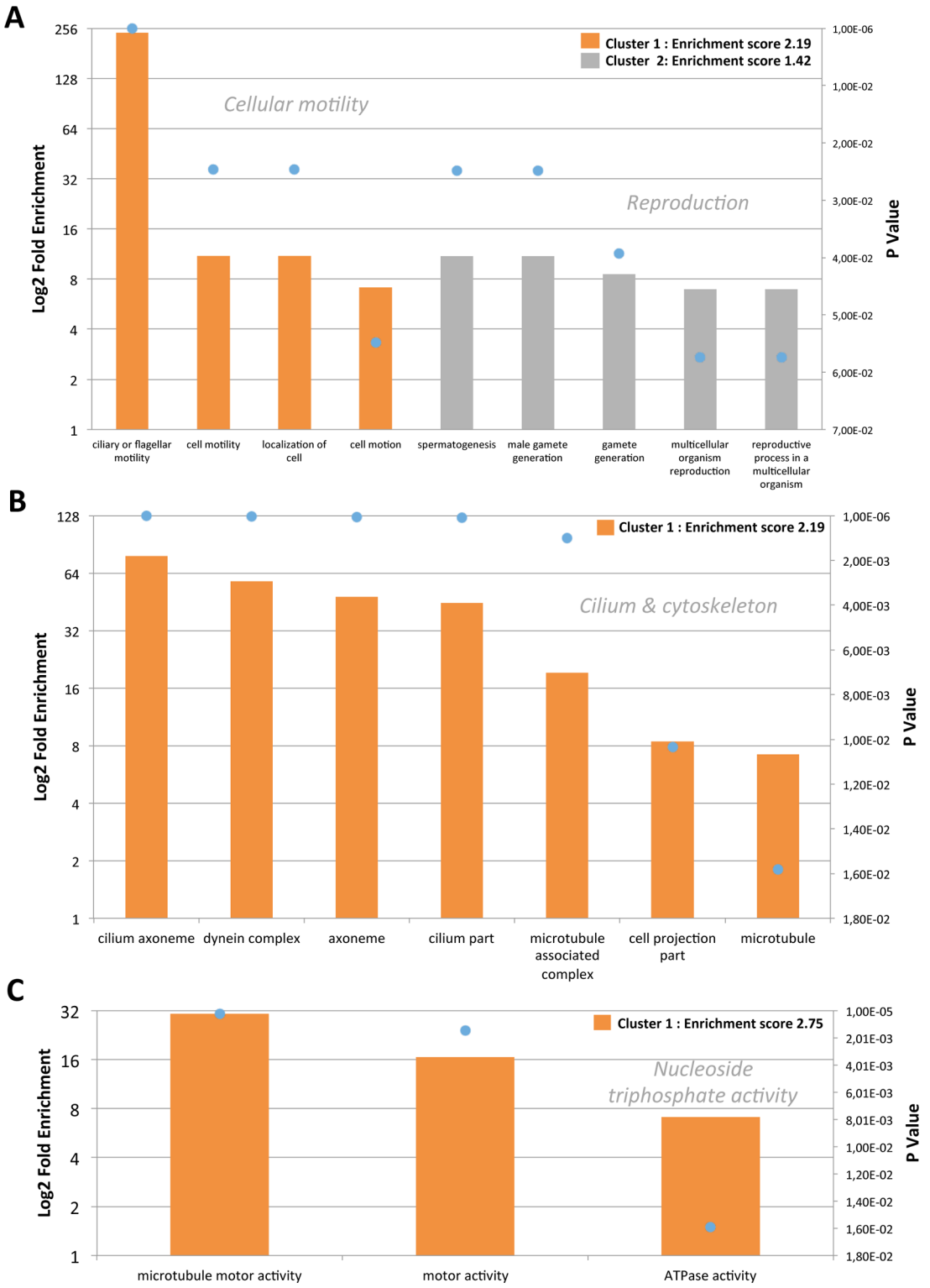


Figure 2. Functional annotation clustering of GO terms enriched in the list of 89 missing proteins and five proteins scored uncertain in neXtProt identified in normal spermatozoa. (A) Biological process; (B) cellular component; (C) molecular function. The enriched GO terms in the data set were selected with the highest classification stringency in DAVID. Colored histograms represent the Log2 fold enrichment of the enriched GO term (left scale) and the blue dots represent the modified Fisher exact *P*-value for each term (right scale). GO-enriched terms were clustered into groups displaying similar annotations, which are classified according to their enrichment score.

using the Affymetrix probeset/Accession Number conversion list available at neXtProt. The levels of their mRNAs (Log2-transformed signal intensities) in human infertile testis, healthy testis, isolated seminiferous tubules, isolated germ cells, and human nontesticular tissues were determined. To construct the heatmap, missing proteins were sorted by their mRNA signal intensity in human testis samples with decreasing Johnsen score (from mature human testicular samples containing late germ-cells to immature samples devoid of germ cells).

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 2. List of 12 Newly Detected Missing Proteins in the Human Spermatozoa Encoded on Chromosomes 2 and 14[a]

UniProtKB accession	UniProtKB/Swiss-Prot entry name	chromosome number	description	neXtProt status	identified peptides (MS/MS)	identified unique peptides (MS/MS)	manual validation evel when requested	link to ciliation/flagellum/testis[b]
Q96MC2	DRC1_HUMAN	2	dynein regulatory complex protein 1 (DRC1)	PE2	6	4		yes
Q6ZU64	CCDC108_HUMAN	2	coiled-coil domain-containing protein 108	PE2	5	3		yes
Q9H7T0	CTSRB_HUMAN	14	cation channel sperm-associated protein subunit beta (CATSPERB)	PE2	2	2		yes
A6NJV1	C2OR F70_HUMAN	2	UPF0573 protein C2orf70	PE2	2	2		yes
Q5I0G3	MDH1B_HUMAN	2	putative malate dehydrogenase 1B [EC 1.1.1.-] (MDH1B)	PE2	2	1	high	yes
Q8N801	CB061_HUMAN	2	uncharacterized protein C2orf61 (C2orf61)	PE2	1	1	high	yes
A8MYV0	DCD2C_HUMAN	2	doublecortin domain-containing protein 2C (DCDC2C)	PE4	5	3		weak
Q53QW1	CB057_HUMAN	2	uncharacterized protein C2orf57	PE2	3	2		weak
Q5GAN3	RNASE13_HUMAN	14	probable inactive ribonuclease-like protein 13	PE2	1	1	high	weak
Q96LM6	TEX37_HUMAN	2	testis-expressed sequence 37 protein (TEX37)	PE2	3	1	high	weak
Q68DN1	CB016_HUMAN	2	uncharacterized protein C2orf16 (C2orf16)	PE2	3	1	high	weak
Q7Z5J8	ANKAR_HUMAN	2	ankyrin and armadillo repeat-containing protein (ANKAR)	PE2	1	1	medium	no

[a]Detailed information on the peptide sequences is available in Supplementary Table 1. [b]More information on the proteins linked to ciliation, flagellum, or testis is available in Supplementary Table 2.

RESULTS

Identification of Missing Proteins in Total Protein Extract from Spermatozoa

By studying the proteome from human spermatozoa, we identified 89 proteins classified as "missing proteins" (PE2−4) in the 09.19.2014 release of neXtProt and five classified as "uncertain proteins" (PE5) (Table 1; Supplementary Table 1, Supporting Information). The proteins identified match genes on 21 chromosomes: the only chromosomes not carrying any of the genes are 21, 22, and Y (Figure 1). Thirty-four of the proteins identified were evidenced from a single peptide that passed the automatic validation (29 with high PSM quality and five with medium PSM quality) (Table 1 and Supplementary Table 1).

In the entire set of 1547 identified proteins satisfying our filtration criteria, with 1% FDR at the protein level, as many as 16 proteins could be considered as false positive. It is legitimate to consider the latter being predominantly spread among the 463 proteins identified with only one or two peptides, leading to an uncertainty at around 3.4%. In the proposed final set of 89 missing proteins, 33 were identified with three or more unique peptides, whereas 56 were double or single peptide-based identifications. Thus, with an uncertainty of 3.4%, we can assume to have a probability of less than two false positives.

Three of the five "uncertain" (PE5) proteins, were further validated by a manual inspection of the data: (i) BOD1L2, identified from two peptides; (ii) FAM71E2, identified from a single peptide (medium quality PSM); and (iii) FAM205CP, identified from two peptides. The identification of GK3P and SPATA31D3 with single peptides (medium and high quality PSM, respectively), is more problematic. Indeed, the sequence of GK3P is 98% identical to that of the well-known PE1 protein GK, and the single unique peptide identified in our study differs by only one amino acid from the corresponding segment of GK sequence. A possible SNP, not yet characterized, at that amino acid in GK sequence would invalidate the identification of GK3P (see Supplementary Data 1A). Likewise, SPATA31D3 differs by only three amino acid positions from SPATA31D4. The unique peptide identified in our study contains one of these three amino acids but matches with the sequence encoded by SPATA31D4 G882R, a SNP present in 3.5% of the population (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=138456481) (see Supplementary Data 1B). Since the genomic sequence of the donor is not available, the identified peptide can be either attributed to SPATA31D3 or to SPATA31D4 G882R.

Among the final set of 94 identified proteins, 26 have experimentally proven or predicted functional annotation in UniProtKB (manually annotated function in Swiss-Prot or at least one "biological process" GO term) (Supplementary Table 1). The enrichment of identified proteins in terms of biological process, cellular component, and molecular function GO annotation was evaluated (Figure 2). Enriched GO terms were clustered into groups displaying similar annotations. For biological processes, the most enriched GO terms related to "cellular motility" (cluster 1, enrichment score 2.19) and "reproduction" (cluster 2, enrichment score 1.42) (Figure 2A). Regarding cellular components, only one cluster of GO terms was enriched: "cilium and cytoskeleton" (enrichment score 2.19). The molecular functions of the 92 proteins identified were evaluated, and only one cluster of GO terms was enriched: "nucleotide triphosphate activity" (enrichment score 2.75) (Figure 2C).

Repertoire of the Missing Proteins Identified in Spermatozoa That Are Encoded on Chromosomes 2 and 14

Twelve missing proteins encoded by genes on chromosomes 2 and 14 were identified in human spermatozoa (Table 2). Six of

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

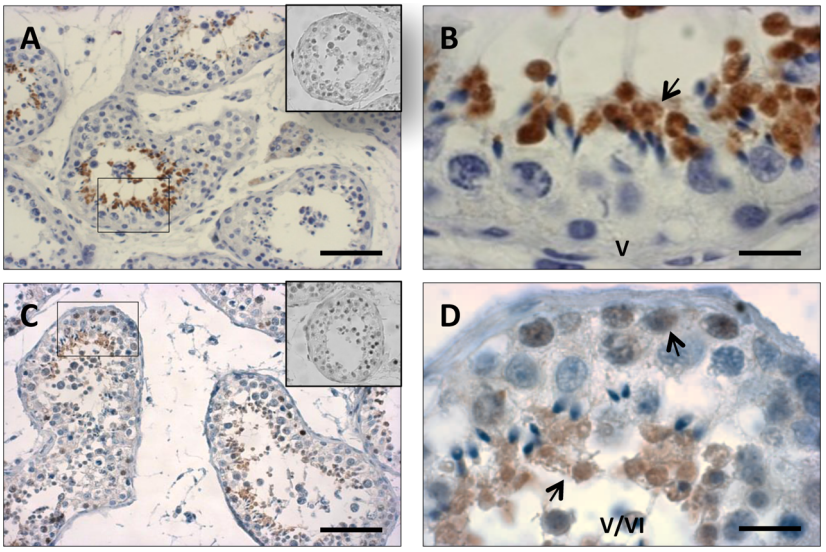


Figure 3. Immunolocalization of C2orf57 and TEX37 in the adult human testis. Proteins were detected in transverse testis sections using polyclonal antibodies from the Human Protein Atlas against C2orf57 (HPA049917) (A, B) and TEX37 (HPA043987) (C, D). Nonimmune serum was used as a negative control (inserts in panels A and C). Selected regions (black squares) in panels A and C correspond to higher magnifications presented in panels B and D, respectively. In magnified views, roman numerals indicate seminiferous epithelium stages.[45] In testis sections, C2orf57 immunoreactivity is restricted to germ cells at all stages of the seminiferous epithelium (A), and strong staining is visible in the cytoplasm of elongating spermatids (B; arrow). TEX37 is present in the germ cell lineage at all stages (C), and the immunoreactive signal is homogeneous in cytoplasm in spermatogonia and primary spermatocytes to pachytene spermatocytes and elongating spermatids at stage V (D, arrows). Scale bars = 50 μm (A, C) and 10 μm (B, D).

these were identified with a single peptide that passed the manual evaluation with the score "high" or "medium", and six others were identified from more than one unique peptide. Through deep data mining of various resources, we could associate 11 of these proteins to testis, flagella, or ciliation processes, with the evidence being strong in six cases (Table 2 and Supplementary Table 2).

In Situ Validation of Expression Data by Immunohistochemistry

Immunohistochemical experiments were carried out on adult human testes for two identified spermatozoa proteins encoded on chromosome 2 whose mRNA were expressed during spermatogenesis and for which an antibody was available from Human Protein Atlas: C2orf57 and TEX37 (Figures 3 and 4E). Antibodies against three other proteins of interest were used but did not yield satisfactory results in our hands. In testis sections, C2orf57 immunoreactivity was restricted to germ cells at all stages of the seminiferous epithelium (Figure 3A), and staining was strong in the cytoplasm of elongating spermatids (Figure 3B). TEX37 was also detected in the germ cell lineage at all stages (Figure 3C). However, in contrast to C2orf57, the immunoreactive signal for TEX37 was homogeneous in cytoplasm in spermatogonia and primary spermatocytes to pachytene spermatocytes and elongating spermatids at stage V (Figure 3D). For both antibodies, no staining over background levels was visible in interstitial cells or in somatic cells in the seminiferous tubules.

Potential of Testis and Postmeiotic Male Germ Cells for the Discovery of Additional Missing Proteins

We extensively mined publicly available transcriptome data, that is, the "Human testis gene expression program" described by Chalmel and collaborators[24] to estimate the number of remaining missing proteins that correspond to genes expressed in the testis or postmeiotic male germ cells. A heatmap summarizing the various patterns that define the expression of transcripts across the various types of samples is presented in Figure 4.

Only 65 out of the 89 missing proteins and one out of the three uncertain proteins identified in this study corresponded to genes referenced in the "human testis gene expression program".[24] As many as 23 of these proteins correspond to genes specifically expressed in the testis (SET), whereas 24 correspond to genes with intermediate (IE) or ubiquitous (UE) expression in the testis (Table 1). Most of these 66 proteins correspond to transcripts with strong expression in seminiferous tubules containing postmeiotic germ cells (i.e., presence of early, late, and elongated spermatids/spermatozoa (Johnsen score $\geq$ 7) (Figure 4A). This mRNA abundance in postmeiotic germ cells was confirmed in total testis (TT), isolated seminiferous tubules (ST), enriched spermatocytes (SPC), and early/round spermatids (SPT) (Figure 4A). Some of these 66 mRNA were also observed in other healthy human tissues.

As much as 361 of the missing proteins that have not been detected to date in our spermatozoa samples and that remain to be identified according to neXtProt corresponded to genes from the "human testis gene expression program" defined by Chalmel and collaborators.[24] The expression intensity of these 361 transcripts across the entire data set is presented in Figure 4, panels B–D. The vast majority of missing proteins whose transcripts are specifically (SET) or preferentially (PET) expressed in the testis are meiotic and postmeiotic germ cell products as confirmed by their strong expression in testes with a Johnsen score $\geq$ 7 (Figure 4B,C). This observation was confirmed by testing total testis (TT), isolated seminiferous tubules (ST), enriched spermatocytes (SPC), and early/round spermatids (SPT) (central panels in Figure 4B,C). As expected, the expression of these transcripts in other nontesticular control organs was negative for SET genes but observable in a few

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

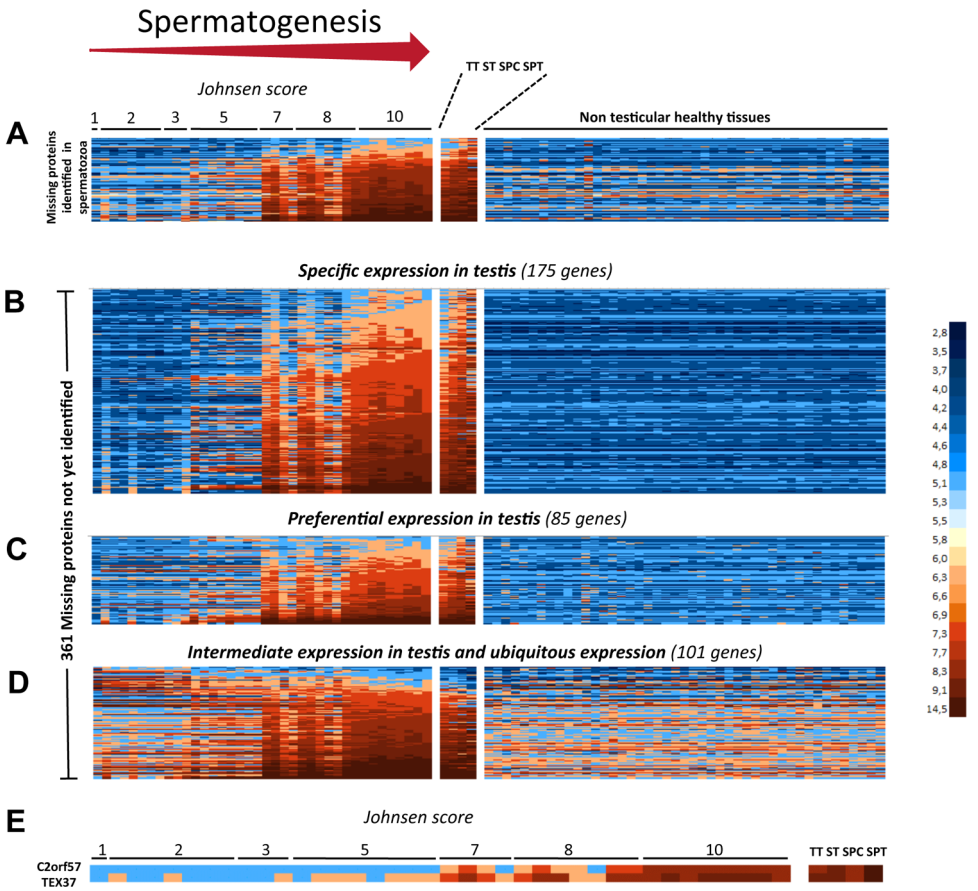


Figure 4. Profiling of the transcriptome corresponding to the human testicular missing proteins in the 09.19.2014 release of neXtProt. Missing proteins (PE2, PE3, and PE4; 09.19.2014 release) were sorted according to their mRNA signal intensity in human testicular samples with decreasing Johnsen score (from mature human testicular samples containing late postmeiotic germ cells to immature testis with no germ cells). Each line is a probe set, and each column is a sample. Log2-transformed signal intensities are colored according to the scale bar. Their corresponding mRNA signal intensities in total testis (TT), seminiferous tubules (ST), enriched spermatocytes (SPC), and early/round spermatids (SPT) from fertile patients and in 45 nontesticular healthy tissues are also reported. (A) Transcript profiles corresponding to 65 of the 89 missing proteins (PE2−4) and one of the five proteins scored as uncertain (PE5) identified by mass spectrometry in this study. (B) Transcript profiles of genes specifically expressed in testis that correspond to 175 of the remaining missing proteins. (C) Transcript profiles of genes preferentially expressed in testis that correspond to 85 of the remaining missing proteins. (D) Transcript profiles of genes with intermediate expression in testis or ubiquitous expression that correspond to 101 of the remaining missing proteins. (E) Close view of transcript profiles corresponding to C2orf57 and TEX37 identified in this study.

organs for PET genes (right panels in Figure 4B, C). Most of the missing proteins whose transcripts display an intermediate expression in testis or a ubiquitous expression are somatic or germ cell products. This was confirmed by their early expression in the testes, with a Johnsen score ≥ 1, or their presence in premeiotic and meiotic germ cells with a Johnsen score = 3−5 (Figure 4D). The abundance of these transcripts also varies in a variety of other healthy human tissues.

A total of 175 testis-specific transcripts display intense signals in mature normal testes (Figure 4B). The full list of these 175 entries is provided in Supplementary Data 2. Although functional or subcellular localization information is not available for a significant part of the corresponding proteins, a rapid analysis showed that numerous entries match with "intrinsic to membrane" and "integral to membrane" GO terms; others are linked to "transcription" and "regulation of transcription" or to reproduction (e.g., spermatogenesis, male gamete generation, sexual reproduction) (data not shown).

The expression pattern of C2orf57 and TEX37 transcripts was studied in detail (Figure 4E). As expected, transcripts for both proteins display a strong expression in seminiferous tubules containing postmeiotic germ cells (Johnsen score ≥ 7). However, transcripts for TEX37 are also expressed in seminiferous tubules containing premeiotic and meiotic germ cells (Johnsen scores 3−5). Interestingly, both transcripts correspond to ubiquitous genes in the "human testis gene expression program"[24] and were also expressed in various other organs but at much lower levels than in the adult testis (data not shown).

DISCUSSION

The results of this study confirm that the male gamete is rich in so-called "missing proteins", as we were able to validate the existence of 89 such proteins. Sixty-five were gene products from the "Human testis gene expression program",[24] the transcripts of which were abundant in human testis cells with a Johnsen score above 7. They thus displayed enriched expression in the postmeiotic germ-cell lineage, right up to the post-testicular mature spermatozoa. For example, the C2orf57 and TEX37 transcripts were abundant in human testis cells with a Johnsen score above 7. Immunohistochemistry on human testis sections confirmed the presence of the TEX37 and C2orf57 proteins in postmeiotic germ cells. An immunoreactive signal was also detected for TEX37 in

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

premeiotic germ cells, consistent with the detection of the corresponding mRNA in the human testis, with a Johnsen score of between 3 and 5.

We carried out a rapid search of the literature to highlight the potential relevance of the missing proteins identified to the field of male reproduction. Strong putative links to male reproduction exist for four dynein heavy chain proteins (DNAH1, DNAH2, DNAH3, and DNAH12): these axonemal dyneins provide the driving force for ciliary and flagellar motility.[28] Ben Khelifa and co-workers[29] demonstrated that mutations of DNAH1 lead to male infertility by causing multiple abnormalities of the sperm flagella. Immunohistochemical studies have demonstrated the presence of DNAH2 and DNAH12 in the flagella of human sperm.[30] A possible link to male reproduction was also found for IZUMO2 and IZUMO3. The Izumo protein family includes four proteins found in many species.[31] Three members of the Izumo family remain to be characterized, but IZUMO1 is an essential sperm cell surface protein required for fertilization due to its role as a ligand of the FOLR4/JUNO receptor on the egg.[32] In Izumo1-knockout mice, spermatozoa do not fuse with oocytes.[33] A possible link to male reproduction was also found for MORN3. The mouse ortholog of MORN3 has been detected, by immunohistochemistry, in the acrosome of germ cells throughout spermiogenesis but also in the manchette of elongating spermatids.[34] The authors also suggested that MORN3 regulated spermatogenesis, probably acting together with the product of the meiosis expressed gene 1 (MEIG1). These observations suggest that some of the missing proteins we have identified may be preferentially or specifically expressed in the germ cell lineage and that they are involved in spermatogenesis and the biology of the male gamete.

In addition, five uncertain (PE5) proteins were identified by our work (Table 1). Three of these proteins were unambiguously validated by manual inspection of the data. Until recently, the BOD1L2 gene was considered to be a pseudogene. However, it has recently been reannotated as a protein-coding gene by the HUGO Gene Nomenclature Committee (HGNC) and the NCBI Reference Sequence (RefSeq) database. Interestingly, the BOD1L2 gene is considered to be a SET gene in the "Human testis gene expression program".[24] The FAM71E2 gene is also classified as a protein-coding gene in the HGNC and RefSeq databases. Furthermore, the RNA sequencing analysis of HPA suggested that there is an enrichment of BOD1L2 and FAM71E2 expression in the testis. Finally, FAM205CP is considered to be a pseudogene in the HGNC database, whereas its mouse ortholog is annotated as a protein-coding gene in the Mouse Genome Informatics database. In UniProtKB/Swiss-Prot, all the known FAM205CP orthologs are annotated PE2, unlike their human counterpart. This work also made it possible to reconsider the protein existence status of these UniProt/neXtProt entries. The other two PE5 proteins identified in this study raise an interesting problem in the context of the C-HPP. Indeed, although they were considered unique, the peptides identifying Q14409 (GK3P) and P0C874 (SPATA31D3) match other entries if possible single amino-acid polymorphisms are taken into account. This work thus highlights the importance of knowing the genomic sequence of the sample considered.

The newly identified missing proteins encoded by genes on chromosomes 2 and 14 were selected for more detailed study. One of these proteins was a PE4 protein (DCDC2C), but all the others were annotated as PE2. Five of these proteins had already been identified in the human testis or sperm by MS: DRC1, MDH1B, C2orf70, RNASE13, and C2orf16 (Supplementary Table 2). Unfortunately, it appears that most of the corresponding data sets were not uploaded onto ProteomeXchange, and these proteins were therefore not annotated PE1 in neXtProt.

By mining the available information from various resources, we found strong links with testis, flagellum, or ciliation processes for six of the 12 identified proteins encoded by genes on chromosomes 2 and 14 (Supplementary Table 2). One of these proteins was CatSperß, a transmembrane protein known to be an auxiliary subunit of the CatSper complex. Four ion channel subunit proteins (CatSpers 1−4) assemble, presumably as a tetramer, to form this sperm-specific, alkalinization-activated Ca^{2+}-selective channel that is required for sperm cell hyperactivation and male fertility.[35] The mouse CatSperß was recently characterized,[35] but its human ortholog remained a missing protein in neXtProt. Mutations and deletions in CatSper genes in the mouse all led to similar male infertility phenotypes with spermatozoa lacking the hyperactivated motility required for fertilization.[36] Subfertile men with low sperm cell motility have significantly lower than normal levels of CatSper1 expression.[37] Similarly, linkage analysis has implicated CatSper2 in human asthenoteratozoospermia.[38] Further studies of the role of CatSperß in human male infertility are planned.

Another example is provided by CCDC108, which was identified by MS in the centrioles of bovine sperm[39] and is conserved in many ciliated organisms. Its putative ortholog in *Chlamydomonas reinhardtii*, FAP65, is strongly induced after deflagellation during the regeneration of the flagella.[40] Furthermore, disruption of the CCDC108 gene may cause defective sperm motility in chicken:[41] it would therefore be interesting to investigate the involvement of CCDC108 in male infertilities in humans.

Five other proteins were found to be possibly involved in testis-related functions on the basis of their expression profiles or identification in mammalian testis/sperm samples (Supplemental Table 2). The distribution of C2orf57 in the testis has been assessed by immunohistochemistry:[17] this protein was found only in testicular spermatozoa. Its rat and mouse orthologs (LOC501180 and 1700019O17Rik, respectively) have been detected in the testis by MS.[42,43] We present here unambiguous MS/MS data confirming the presence of the C2orf57 protein in human spermatozoa. We used immunohistochemistry to reinvestigate the distribution of this protein within the seminiferous tubules with the same antibody used by Djureinovic and co-workers.[17] We report that this protein is specific to the germ cell lineage and is present during the six stages of human spermatogenesis. It is present in spermatids at later stages of spermiogenesis and concentrates in the cytoplasm of elongated spermatids before their release into the tubule lumen. To our knowledge, TEX37 had never been validated by MS- or Ab-based techniques in human samples. By contrast, mouse Tex37 has been identified by MS in accessory structures from the mouse sperm flagellum.[44] We report the first MS/MS data demonstrating the presence of TEX37 in human spermatozoa. Immunochemistry confirmed the germ cell-specificity of the protein and its increase in abundance with the progression from spermatogonia to late spermatids. Our results thus implicate TEX37 in germ cell biology throughout spermatogenesis and spermiogenesis.

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

We show that 361 of the 2839 proteins still considered missing correspond to genes from the "human testis gene expression program".[24] As demonstrated here, the vast majority of these candidates were detected in total adult testicular extracts and corresponded to products from postmeiotic germ cells or ejaculated spermatozoa. Interestingly, 175 transcripts were most abundant in mature normal testes and corresponded to testis-specific genes not expressed in any other healthy tissues. A rapid analysis of the 175 entries showed that a significant number of these were linked to the cellular component GO terms "intrinsic to membrane" and "integral to membrane" (data not shown), suggesting that the corresponding proteins might be found in membrane extracts from spermatozoa. Our study confirms the potential of testis tissues for the discovery of additional "missing proteins". We are therefore continuing our efforts to identify missing proteins in extracts enriched in testis cell or in sperm cell heads and flagella using additional enrichment strategies with better resolution favoring the isolation of membranous, membrane-bound, and nuclear proteins.

CONCLUSIONS

The expression of many missing proteins is probably restricted to a few organs or, possibly, even one type of organ or specific cell types, such as spermatozoa or postmeiotic male germ cells. It may be difficult for teams involved in the C-HPP, other than those with extensive experience in studying spermatogenesis, to gain access to relevant biological samples for the identification of missing proteins, particularly those with narrow windows of expression in a limited range of testicular meiotic and postmeiotic germ cells. One possible approach would involve the use of an immortalized germ cell line. However, no such line is currently available, and screening can be performed only on primary cells with very limited survival after isolation. This study thus provides useful insight into the proteome coverage that can be achievable in human spermatozoa and indicates potential solutions for the identification of additional missing proteins in testicular postmeiotic germ cells or sperm cells. It also provides hints for the characterization of proteins of unknown function relevant to spermatogenesis, post-testicular gamete differentiation, and male infertility.

ASSOCIATED CONTENT

Supporting Information

List of 89 proteins designated "Missing Proteins" and five proteins scored as uncertain by neXtProt, identified in our study, including detailed MS information on peptide sequences. Functional annotation from UniProtKB is provided when available. Extended Table 2 providing additional information on 12 proteins that may be associated with testis, flagella, or ciliation processes, with strong evidence in six cases. Alignments of identified PE5 proteins with their closest paralog using Clustal Omega version 1.2.1: (A) alignment of the identified PE5 protein Q14409 (GK3P) with the PE1 protein P32189 (GK); (B) alignment of the identified PE5 protein P0C874 (SPATA31D3) with the PE5 protein Q6ZUB0 (SPATA31D4). Regions of interest in the sequences are visualized in red. List of the 175 missing proteins whose testis-specific transcripts display high signal intensities in mature testes and normal spermatozoa. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.5b00170.

AUTHOR INFORMATION

Corresponding Author

*Phone: +33 (0)2 2323 5279. E-mail: charles.pineau@inserm.fr.

Author Contributions

V.M. and C.P. contributed equally. C.P., V.M., L.L., and N.S. conceived and designed the experiments. F.J., R.L., and K.R. performed the experiments. E.C., M.L., L.L., C.P., P.D., N.M., F.J., L.G., R.L., K.R., A.G., and B.G. analyzed the data. F.J., V.M., N.S., E.C., L.L., L.G., and A.G. contributed reagents/materials/analysis tools. C.P., E.C., L.L., V.M., N.S., and F.J. wrote the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by Biogenouest and Infrastructures en Biologie Santé et Agronomie (IBiSA), European Regional Development Fund, and Conseil Régional de Bretagne grants awarded to C.P. We thank Dr Y. Vandenbrouck (ProFi) for invaluable advice on the MS-based validation of missing proteins. We are grateful to Mathias Uhlen (Human Protein Atlas, Sweden) for the gift of antibodies. We thank Christine Kervarrec for preparing histology materials. We also thank Prof. Amos Bairoch (Swiss Institute of Bioinformatics and University of Geneva) and Dr Michael Primig (Irset−Inserm U1085) for stimulating discussions and advice.

REFERENCES

(1) Chocu, S.; Calvel, P.; Rolland, A. D.; Pineau, C. Spermatogenesis in mammals: proteomic insights. *Syst. Biol. Reprod. Med.* **2012**, *58* (4), 179−90.

(2) Macleod, G.; Varmuza, S. The application of proteomic approaches to the study of mammalian spermatogenesis and sperm function. *FEBS J.* **2013**, *280* (22), 5635−51.

(3) Dacheux, J. L.; Belleannee, C.; Guyonnet, B.; Labas, V.; Teixeira-Gomes, A. P.; Ecroyd, H.; Druart, X.; Gatti, J. L.; Dacheux, F. The contribution of proteomics to understanding epididymal maturation of mammalian spermatozoa. *Syst. Biol. Reprod. Med.* **2012**, *58* (4), 197−210.

(4) Nowicka-Bauer, K.; Kurpisz, M. Current knowledge of the human sperm proteome. *Expert Rev. Proteomics* **2013**, *10* (6), 591−605.

(5) Rousseaux, S.; Khochbin, S. Combined proteomic and in silico approaches to decipher post-meiotic male genome reprogramming in mice. *Syst. Biol. Reprod. Med.* **2012**, *58* (4), 191−6.

(6) Aitken, R. J.; Nixon, B.; Lin, M.; Koppers, A. J.; Lee, Y. H.; Baker, M. A. Proteomic changes in mammalian spermatozoa during epididymal maturation. *Asian J. Androl.* **2007**, *9* (4), 554−64.

(7) Baker, M. A.; Nixon, B.; Naumovski, N.; Aitken, R. J. Proteomic insights into the maturation and capacitation of mammalian spermatozoa. *Syst. Biol. Reprod. Med.* **2012**, *58* (4), 211−7.

(8) Rolland, A. D.; Jegou, B.; Pineau, C. Testicular development and spermatogenesis: harvesting the postgenomics bounty. *Advances in experimental medicine and biology* **2009**, *636*, 16−41.

(9) Paik, Y. K.; Jeong, S. K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Cho, S. Y.; Lee, H. J.; Na, K.; Choi, E. Y.; Yan, F.; Zhang, F.; Zhang, Y.; Snyder, M.; Cheng, Y.; Chen, R.; Marko-Varga, G.; Deutsch, E. W.; Kim, H.; Kwon, J. Y.; Aebersold, R.; Bairoch, A.; Taylor, A. D.; Kim, K. Y.; Lee, E. Y.; Hochstrasser, D.; Legrain, P.; Hancock, W. S. The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat. Biotechnol.* **2012**, *30* (3), 221−3.

(10) Gaudet, P.; Argoud-Puy, G.; Cusin, I.; Duek, P.; Evalet, O.; Gateau, A.; Gleizes, A.; Pereira, M.; Zahn-Zabal, M.; Zwahlen, C.;

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

Bairoch, A.; Lane, L. neXtProt: organizing protein knowledge in the context of human proteome projects. *J. Proteome Res.* **2013**, *12* (1), 293−8.
(11) Kim, M. S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; Thomas, J. K.; Muthusamy, B.; Leal-Rojas, P.; Kumar, P.; Sahasrabuddhe, N. A.; Balakrishnan, L.; Advani, J.; George, B.; Renuse, S.; Selvan, L. D.; Patil, A. H.; Nanjappa, V.; Radhakrishnan, A.; Prasad, S.; Subbannayya, T.; Raju, R.; Kumar, M.; Sreenivasamurthy, S. K.; Marimuthu, A.; Sathe, G. J.; Chavan, S.; Datta, K. K.; Subbannayya, Y.; Sahu, A.; Yelamanchi, S. D.; Jayaram, S.; Rajagopalan, P.; Sharma, J.; Murthy, K. R.; Syed, N.; Goel, R.; Khan, A. A.; Ahmad, S.; Dey, G.; Mudgal, K.; Chatterjee, A.; Huang, T. C.; Zhong, J.; Wu, X.; Shaw, P. G.; Freed, D.; Zahari, M. S.; Mukherjee, K. K.; Shankar, S.; Mahadevan, A.; Lam, H.; Mitchell, C. J.; Shankar, S. K.; Satishchandra, P.; Schroeder, J. T.; Sirdeshmukh, R.; Maitra, A.; Leach, S. D.; Drake, C. G.; Halushka, M. K.; Prasad, T. S.; Hruban, R. H.; Kerr, C. L.; Bader, G. D.; Iacobuzio-Donahue, C. A.; Gowda, H.; Pandey, A. A draft map of the human proteome. *Nature* **2014**, *509* (7502), 575−81.
(12) Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A. M.; Lieberenz, M.; Savitski, M. M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; Mathieson, T.; Lemeer, S.; Schnatbaum, K.; Reimer, U.; Wenschuh, H.; Mollenhauer, M.; Slotta-Huspenina, J.; Boese, J. H.; Bantscheff, M.; Gerstmair, A.; Faerber, F.; Kuster, B. Mass-spectrometry-based draft of the human proteome. *Nature* **2014**, *509* (7502), 582−7.
(13) Lane, L.; Bairoch, A.; Beavis, R. C.; Deutsch, E. W.; Gaudet, P.; Lundberg, E.; Omenn, G. S. Metrics for the Human Proteome Project 2013−2014 and strategies for finding missing proteins. *J. Proteome Res.* **2014**, *13* (1), 15−20.
(14) Uhlen, M.; Fagerberg, L.; Hallstrom, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigyarto, C. A.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P. H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; von Heijne, G.; Nielsen, J.; Ponten, F. Proteomics. Tissue-based map of the human proteome. *Science* **2015**, *347* (6220), 1260419.
(15) Ramskold, D.; Wang, E. T.; Burge, C. B.; Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* **2009**, *5* (12), e1000598.
(16) Soumillon, M.; Necsulea, A.; Weier, M.; Brawand, D.; Zhang, X.; Gu, H.; Barthes, P.; Kokkinaki, M.; Nef, S.; Gnirke, A.; Dym, M.; de Massy, B.; Mikkelsen, T. S.; Kaessmann, H. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* **2013**, *3* (6), 2179−90.
(17) Djureinovic, D.; Fagerberg, L.; Hallstrom, B.; Danielsson, A.; Lindskog, C.; Uhlen, M.; Ponten, F. The human testis-specific proteome defined by transcriptomics and antibody-based profiling. *Mol. Hum. Reprod.* **2014**, *20* (6), 476−88.
(18) Lavigne, R.; Becker, E.; Liu, Y.; Evrard, B.; Lardenois, A.; Primig, M.; Pineau, C. Direct iterative protein profiling (DIPP) - an innovative method for large-scale protein detection applied to budding yeast mitosis. *Mol. Cell. Proteomics* **2012**, *11* (2), M111.012682.
(19) Vizcaino, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Rios, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; Binz, P. A.; Xenarios, I.; Eisenacher, M.; Mayer, G.; Gatto, L.; Campos, A.; Chalkley, R. J.; Kraus, H. J.; Albar, J. P.; Martinez-Bartolome, S.; Apweiler, R.; Omenn, G. S.; Martens, L.; Jones, A. R.; Hermjakob, H. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014**, *32* (3), 223−6.
(20) Carapito, C.; Lane, L.; Benama, M.; Opsomer, A.; Mouton-Barbosa, E.; Garrigues, L.; Gonzalez de Peredo, A.; Burel, A.; Bruley, C.; Gateau, A.; Bouyssié, D.; Jaquinod, M.; Cianférani, S.; Burlet-Schiltz, O.; Van Dorsselaer, A.; Garin, J.; Vandenbrouck, Y. Computational and mass spectrometry-based workflow for the discovery and validation of human missing proteins: Application to chromosomes 2 and 14. *J. Proteome Res.*, submitted for publication. DOI: 10.1021/pr5010345.
(21) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2008**, *4* (1), 44−57.
(22) Uhlen, M.; Oksvold, P.; Fagerberg, L.; Lundberg, E.; Jonasson, K.; Forsberg, M.; Zwahlen, M.; Kampf, C.; Wester, K.; Hober, S.; Wernerus, H.; Bjorling, L.; Ponten, F. Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **2010**, *28* (12), 1248−50.
(23) Com, E.; Rolland, A. D.; Guerrois, M.; Aubry, F.; Jegou, B.; Vallet-Erdtmann, V.; Pineau, C. Identification, molecular cloning, and cellular distribution of the rat homolog of minichromosome maintenance protein 7 (MCM7) in the rat testis. *Mol. Reprod. Dev.* **2006**, *73* (7), 866−77.
(24) Chalmel, F.; Lardenois, A.; Evrard, B.; Mathieu, R.; Feig, C.; Demougin, P.; Gattiker, A.; Schulze, W.; Jegou, B.; Kirchhoff, C.; Primig, M. Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Hum. Reprod.* **2012**, *27* (11), 3233−48.
(25) Chalmel, F.; Rolland, A. D.; Niederhauser-Wiederkehr, C.; Chung, S. S.; Demougin, P.; Gattiker, A.; Moore, J.; Patard, J. J.; Wolgemuth, D. J.; Jegou, B.; Primig, M. The conserved transcriptome in human and rodent male gametogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (20), 8346−51.
(26) Johnsen, S. G. Testicular biopsy score count–a method for registration of spermatogenesis in human testes: normal values and results in 335 hypogonadal males. *Horm. Res.* **1970**, *1* (1), 2−25.
(27) Rolland, A. D.; Lavigne, R.; Dauly, C.; Calvel, P.; Kervarrec, C.; Freour, T.; Evrard, B.; Rioux-Leclercq, N.; Auger, J.; Pineau, C. Identification of genital tract markers in the human seminal plasma using an integrative genomics approach. *Hum. Reprod.* **2013**, *28* (1), 199−209.
(28) Chapelin, C.; Duriez, B.; Magnino, F.; Goossens, M.; Escudier, E.; Amselem, S. Isolation of several human axonemal dynein heavy chain genes: genomic structure of the catalytic site, phylogenetic analysis and chromosomal assignment. *FEBS Lett.* **1997**, *412* (2), 325−30.
(29) Ben Khelifa, M.; Coutton, C.; Zouari, R.; Karaouzene, T.; Rendu, J.; Bidart, M.; Yassine, S.; Pierre, V.; Delaroche, J.; Hennebicq, S.; Grunwald, D.; Escalier, D.; Pernet-Gallay, K.; Jouk, P. S.; Thierry-Mieg, N.; Toure, A.; Arnoult, C.; Ray, P. F. Mutations in DNAH1, which encodes an inner arm heavy chain dynein, lead to male infertility from multiple morphological abnormalities of the sperm flagella. *Am. J. Hum. Genet.* **2014**, *94* (1), 95−104.
(30) Jones, R. T.; Abedalthagafi, M. S.; Brahmandam, M.; Greenfield, E. A.; Hoang, M. P.; Louis, D. N.; Hornick, J. L.; Santagata, S. Cross-reactivity of the BRAF VE1 antibody with epitopes in axonemal dyneins leads to staining of cilia. *Mod. Pathol.* **2015**, *28*, 596.
(31) Grayson, P.; Civetta, A. Positive Selection and the Evolution of izumo Genes in Mammals. *Int. J. Evol. Biol.* **2012**, *2012*, 1.
(32) Bianchi, E.; Doe, B.; Goulding, D.; Wright, G. J. Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* **2014**, *508* (7497), 483−7.
(33) Inoue, N.; Ikawa, M.; Isotani, A.; Okabe, M. The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* **2005**, *434* (7030), 234−8.
(34) Zhang, L.; Shang, X. J.; Li, H. F.; Shi, Y. Q.; Li, W.; Teves, M. E.; Wang, Z. Q.; Jiang, G. F.; Song, S. Z.; Zhang, Z. B. Characterization of membrane occupation and recognition nexus repeat containing 3, meiosis expressed gene 1 binding partner, in mouse male germ cells. *Asian J. Androl.* **2015**, *17* (1), 86−93.
(35) Liu, J.; Xia, J.; Cho, K. H.; Clapham, D. E.; Ren, D. CatSperbeta, a novel transmembrane protein in the CatSper channel complex. *J. Biol. Chem.* **2007**, *282* (26), 18945−52.
(36) Singh, A. P.; Rajender, S. CatSper channel, sperm function and male fertility. *Reprod. BioMed. Online* **2015**, *30* (1), 28−38.
(37) Nikpoor, P.; Mowla, S. J.; Movahedin, M.; Ziaee, S. A.; Tiraihi, T. CatSper gene expression in postnatal development of mouse testis

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX

and in subfertile men with deficient sperm motility. *Human reproduction* **2004**, *19* (1), 124−8.
(38) Avidan, N.; Tamary, H.; Dgany, O.; Cattan, D.; Pariente, A.; Thulliez, M.; Borot, N.; Moati, L.; Barthelme, A.; Shalmon, L.; Krasnov, T.; Ben-Asher, E.; Olender, T.; Khen, M.; Yaniv, I.; Zaizov, R.; Shalev, H.; Delaunay, J.; Fellous, M.; Lancet, D.; Beckmann, J. S. CATSPER2, a human autosomal nonsyndromic male infertility gene. *Eur. J. Hum. Genet.* **2003**, *11* (7), 497−502.
(39) Firat-Karalar, E. N.; Sante, J.; Elliott, S.; Stearns, T. Proteomic analysis of mammalian sperm cells identifies new components of the centrosome. *J. Cell Sci.* **2014**, *127* (19), 4128−33.
(40) Pazour, G. J.; Agrin, N.; Leszyk, J.; Witman, G. B. Proteomic analysis of a eukaryotic cilium. *J. Cell Biol.* **2005**, *170* (1), 103−13.
(41) Imsland, F.; Feng, C.; Boije, H.; Bed'hom, B.; Fillon, V.; Dorshorst, B.; Rubin, C. J.; Liu, R.; Gao, Y.; Gu, X.; Wang, Y.; Gourichon, D.; Zody, M. C.; Zecchin, W.; Vieaud, A.; Tixier-Boichard, M.; Hu, X.; Hallbook, F.; Li, N.; Andersson, L. The Rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genet.* **2012**, *8* (6), e1002775.
(42) Baker, M. A.; Hetherington, L.; Reeves, G. M.; Aitken, R. J. The mouse sperm proteome characterized via IPG strip prefractionation and LC−MS/MS identification. *Proteomics* **2008**, *8* (8), 1720−30.
(43) Baker, M. A.; Hetherington, L.; Reeves, G.; Muller, J.; Aitken, R. J. The rat sperm proteome characterized via IPG strip prefractionation and LC−MS/MS identification. *Proteomics* **2008**, *8* (11), 2312−21.
(44) Cao, W.; Gerton, G. L.; Moss, S. B. Proteomic profiling of accessory structures from the mouse sperm flagellum. *Mol. Cell. Proteomics* **2006**, *5* (5), 801−10.
(45) Clermont, Y. The cycle of the seminiferous epithelium in man. *Am. J. Anat.* **1963**, *112*, 35−51.

DOI: 10.1021/acs.jproteome.5b00170
J. Proteome Res. XXXX, XXX, XXX−XXX