



Validating Missing Proteins in Human Sperm Cells by Targeted Mass-Spectrometry- and Antibody-based Methods

Christine Carapito, Paula Duek, Charlotte Macron, Marine Seffals, Karine Rondel, François Delalande, Cecilia Lindskog, Thomas Freour, Yves Vandenbrouck, Lydie Lane, et al.

► To cite this version:

Christine Carapito, Paula Duek, Charlotte Macron, Marine Seffals, Karine Rondel, et al.. Validating Missing Proteins in Human Sperm Cells by Targeted Mass-Spectrometry- and Antibody-based Methods. *Journal of Proteome Research*, 2017, 16 (12), pp.4340-4351. 10.1021/acs.jproteome.7b00374 . hal-01681656

HAL Id: hal-01681656

<https://univ-rennes.hal.science/hal-01681656>

Submitted on 8 Feb 2018

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

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

Validating missing proteins in human sperm cells by targeted mass spectrometry- and antibody-based methods

Christine Carapito^{1*}, Paula Duek², Charlotte Macron¹, Marine Seffals³, Karine Rondel⁴, François Delalande¹, Cecilia Lindskog⁵, Thomas Fréour^{6,7}, Yves Vandenbrouck^{8,9,10}, Lydie Lane^{2,11}, Charles Pineau^{4*}

¹ Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, Université de Strasbourg, CNRS UMR7178, 25 Rue Becquerel, Strasbourg F-67087, France

² CALIPHO Group, SIB-Swiss Institute of Bioinformatics, CMU, rue Michel-Servet 1, CH-1211 Geneva 4, Switzerland

³ H2P2 Core facility, UMS BioSit, University of Rennes 1, Rennes F-35040, France

⁴ Protim, Inserm U1085, Irset, Campus de Beaulieu, Rennes F-35042, France

⁵ Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

⁶ Service de Médecine de la Reproduction, CHU de Nantes, 38 boulevard Jean Monnet, Nantes cedex F-44093, France

⁷ Inserm UMR1064, Nantes F-44093, France

⁸ CEA, DRF, BIG, Laboratoire de Biologie à Grande Echelle, 17, rue des Martyrs, Grenoble F-38054, France

⁹ Inserm U1038, Grenoble F-38054, France

¹⁰ Grenoble-Alpes University, Grenoble F-38054, France

¹¹ Department of Human Protein Sciences, Faculty of medicine, University of Geneva, 1, rue Michel-Servet, 1211 Geneva 4, Switzerland

***: To whom correspondence should be addressed**

Christine Carapito: ccarapito@unistra.fr tel: +33 (0)3 68 85 27 30

Charles Pineau: charles.pineau@inserm.fr tel: +33 (0)2 23 23 52 79

**The mass spectrometry proteomics data have been deposited in PeptideAtlas
with the dataset identifier PASS01013**

Reviewers, please access our data using the following account details:

Submission details:

Identifier: PASS01013

Dataset type: PRM

Dataset tag: HPP_Testis_PRM_2017

Dataset title: Targeted LC-PRM experiments for the search of 38 selected missing proteins in human sperm.

Reviewer account details:

Username: PASS01013

Password: JU2578wb

Keywords: human proteome project, spermatozoon, missing proteins, Parallel Reaction Monitoring, targeted proteomics, immunohistochemistry, immunocytochemistry, bioinformatics, data mining

Email addresses:

CC: ccarapito@unistra.fr

LL: lydie.lane@sib.swiss

YV: yves.vandenbrouck@cea.fr

PD: paula.duek@sib.swiss

MS: marine.seffals@univ-rennes1.fr

KR: karine.rondel@univ-rennes1.fr

CM: c.macron@unistra.fr

FD: delaland@unistra.fr

TF: thomas.freour@chu-nantes.fr

CL: cecilia.lindskog@igp.uu.se

CP: charles.pineau@inserm.fr

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

The present study is a contribution to the “neXt50 challenge”, a coordinated effort across C-HPP teams to identify the 50 most tractable missing proteins (MPs) on each chromosome. We report the targeted search of 38 theoretically detectable MPs from chromosomes 2 and 14 in Triton X-100 soluble and insoluble sperm fractions from a total of 15 healthy donors. A targeted mass spectrometry-based strategy consisting in the development of LC-PRM assays (with heavy labeled synthetic peptides) targeting 92 proteotypic peptides of the 38 selected MPs was used. Out of the 38 selected MPs, 12 were identified with 2 or more peptides and 3 with 1 peptide after extensive SDS-PAGE fractionation of the two samples and with overall low intensity signals. The PRM data are available via ProteomeXchange in PASSEL (PASS01013). Further validation by immunohistochemistry on human testes sections and cytochemistry on sperm smears was performed for eight MPs with antibodies available from the Human Protein Atlas. Deep analysis of human sperm still allows the validation of MPs and therefore contributes to the C-HPP worldwide effort. We anticipate that our results will be of interest to the reproductive biology community as an in-depth analysis of these MPs may identify potential new candidates in the context of human idiopathic infertilities.

INTRODUCTION

Phase 1 of the Chromosome-Centric Human Proteome Project (C-HPP) aims at cataloguing proteins as gene products encoded by the human genome in a chromosome-centric manner ¹. To date neXtProt, the reference protein knowledgebase for C-HPP ², contains 20,159 entries (release 2017-01-23). Each entry includes at least one representative protein sequence for a gene. Thanks to a coordinated effort by the C-HPP consortium, together with the adoption of stringent mass spectrometry data interpretation guidelines 2.1 ³, the number of experimentally validated proteins (PE1) has reached 17,008. The count of missing proteins scored as PE2 or 3 or 4 now stands at 2,579.

In the C-HPP context, the Swiss and French teams are in charge of the annotation of proteins from chromosome 2 and 14, respectively. In a recent study, members of the consortium described a step-by-step strategy combining shotgun MS/MS and selected reaction monitoring (SRM) to identify and validate missing proteins in a total of 40 human cell lines, tissues and body fluid samples ⁴. Lane and collaborators suggested that the production of proteins that have been systematically missed might be restricted to unusual organs or cell types, particularly testis ⁵. Indeed, the testis was recognized over a decade ago as the organ that expresses the highest number of tissue-specific genes ⁶. More recently, using RNA-Seq datasets from major tissues and organs, the testis was further confirmed as the most promising organ to search for elements of the missing proteome ^{7 8}.

The primary function of the testis is to produce mature male gametes through spermatogenesis. It is a sophisticated process encompassing a series of events

including proliferation and differentiation of spermatogonia taking place in the basal region of the seminiferous epithelium, meiotic division of primary and secondary spermatocytes, and differentiation of haploid spermatids, leading to the production of spermatozoa - commonly called sperm - in the luminal region of the seminiferous epithelium (for a review see ⁹). Interestingly, the selective pressure on most genes involved in spermatogenesis implies a high degree of germ cell-specific expression with approximately half of these only expressed in the meiotic and post meiotic germ cell lineages ¹⁰. A first trans-chromosome-based pilot study demonstrated the potential of isolated spermatozoa for the validation of a first set of 89 missing proteins ¹¹. More recently, we conducted an in-depth analysis of human sperm in which we could validate 206 missing proteins (PE2-PE4) that fully comply with current HPP guidelines ¹². In order to complete the annotation of the proteins from chromosome 2 and 14 that are expressed in human sperm, we turned to a targeted mass spectrometry (MS) approach allowing for high sensitivity ¹³. Based on information available in the literature and in the main C-HPP resources (Human Protein Atlas, PeptideAtlas and neXtProt), a subset of 40 theoretically detectable missing proteins from chromosome 2 and 14 was proposed for a targeted search in the human sperm ¹⁴.

This work lead to the launch of the so-called “neXt50 challenge”, a coordinated effort across all chromosome teams to identify the 50 most tractable missing proteins for their chromosome. The present study is the contribution of the Swiss and French teams to this new C-HPP challenge. We report the targeted search of 38 theoretically detectable missing proteins in sperm fractions following a Triton X-100 fractionation procedure previously employed by our consortium ¹². Using Liquid Chromatography-

Parallel Reaction Monitoring (LC-PRM) assays with heavy labeled synthetic peptides targeting 92 proteotypic peptides, we were able to unambiguously detect 12 of the 38 selected MPs with 2 or more peptides and 3 with 1 peptide. Antibody-based validation was done for 7 of the proteins identified by targeted-MS, and for C14orf177 for which MS evidence was weak and non-conclusive, and functional hypotheses suggested for some of them, notably for LRRC9, FAM181A, ADAM20, C14orf177 and DNAJC5G/ CSPgamma.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

MATERIALS AND METHODS

Ethics and Donor Consent

The study protocol “Study of Normal and Pathological Human Spermatogenesis” was approved by the ethical committee of Rennes (France). The protocol was then registered as No. PFS09-015 at the French Biomedicine Agency. Written informed consent was obtained from donors where appropriate.

Sample collection and preparation

Human semen samples were collected from 15 healthy donors of unknown fertility at the Nantes University Hospital (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. After 30 min liquefaction at room temperature under smooth agitation, 1 ml of each sample was aliquoted and pooled with others before mixing with a protease inhibitor mix (protease inhibitor cocktail tablets, complete mini EDTA free, Roche, Meylan, France) according to the manufacturer’s instructions. In order to separate sperm cells from seminal plasma and round cells, sperm pool was then loaded on 1 ml of a 50% suspension of silica particles (Suprasperm, Origio, Malov, Denmark) diluted in Sperm Washing medium (Origio, Malov, Denmark) and centrifuged at 400 x g for 15 min at room temperature. Sperm pellet was then washed once by resuspension in 3 mL of Phosphate-buffered saline (PBS) and centrifuged again at 400 x g for 5 min at room temperature. Washing medium supernatant was removed and cell pellet was flash-frozen in liquid nitrogen.

Protein fractionation using Triton X-100

The protein separation using Triton X-100 was based on a protocol described by Baker and collaborators¹⁵ and used in our previous study¹². Briefly, three “swim up” spermatozoa fractions in 300µL of PBS each (with approximately 31 million spermatozoa per fraction) obtained from CHU of Nantes were lysed independently using 1% Triton X-100 on ice. The use of Triton X-100 at 1% final concentration corresponds to the Critical Micelle Concentration (CMC) of the detergent at which cells undergo irreversible permeabilization of membranes and structural collapse¹⁵. After 1-hour incubation time on ice (with vortexing every 10 minutes), the supernatants constituting the Triton X-100 soluble fractions were recovered by centrifugation 15 minutes at 10000 x g. The pellets were washed with Milli-Q water and solubilized in 2% SDS 30 minutes on ice (with vortexing every 10 minutes). After a centrifugation step of 15 min at 10000 x g, the supernatants constituting the Triton X-100 insoluble fractions were taken. Before loading on SDS-PAGE, the Triton X-100 soluble fractions were precipitated as follows: 300µL of sample was added to 300µL methanol and 150µL of chloroform. After vortexing and centrifugation for 2 minutes at 10000 x g, the upper phase was removed and 225µL of methanol was added. The tubes were inverted twice before centrifugation (15 minutes, 10000 x g). The supernatants obtained were removed and the pellets were pooled and allowed to air dry.

Electrophoresis and in gel trypsin digestion

The Triton X-100 soluble fraction was resuspended in Laemmli buffer containing Tris HCl pH 6.8, EDTA 1mM, 5% β -mercaptoethanol, 5% SDS and protease inhibitors while Triton X-100 insoluble fraction was diluted in the same, but two times concentrated Laemmli buffer. For both samples, protein concentrations were assayed using Bradford protein assay (Bradford Assay; Biorad, Hercules, USA) according to manufacturer's instructions. 150 μ g proteins from both fractions were loaded into an in-house poured 10-wells 4-10% acrylamide gel. Electrophoresis was performed under a continuous voltage of 50 V during 20 minutes and then 110 V during 75 minutes. Gel was stained with Coomassie Blue and each lane was manually cut into 23 slices of similar size. In parallel, stacking gels of 50 μ g proteins were run for both fractions. Samples were loaded into an in-house poured 10-wells 4-10% acrylamide gel and electrophoresis was performed under a continuous voltage of 50 V during 23 minutes. The gel was stained with Coomassie Blue and stacking bands were manually cut. Proteins in the gel slices were reduced, alkylated and digested overnight at 37°C with modified trypsin at a 1:100 enzyme:protein ratio (Promega, Madison, USA). Peptides were extracted during 45 minutes with 100 μ L of 60% acetonitrile, 0.1% formic acid (FA) and 15 minutes with a solution of 100% acetonitrile.

Targeted LC-PRM Analyses

For the 38 selected missing proteins, 92 proteotypic peptides that passed the neXtProt uniqueness criteria ¹⁶ were chosen based on prediction of MS-friendliness

(SRM Atlas resource and manual selection) and synthetic heavy labeled peptides were purchased (crude SpikeTides L, JPT Peptide Technologies GmbH, Berlin, Germany). Selected peptide sequences are listed in Supplementary Table 1. The heavy labeled peptides were mixed together in a hand-adjusted concentration-balanced mixture to equilibrate individual peptides signals and spiked in a complex matrix (pool of stacking bands extracts). The peptide mixture was analyzed by nanoLC-MS/MS (Q Exactive Plus, Thermo Fisher Scientific) to acquire reference higher energy collisional dissociation (HCD) fragmentation spectra (analytical details are provided in Supplementary Material file). Following this step, targeted assays using PRM acquisitions were developed to target all 92 peptides and their modified forms (with/without oxidized methionines), first in the unfractionated stacking gel extracts and subsequently in gel bands obtained from 1D SDS-PAGE separation of both Triton X-100 soluble and insoluble fractions. LC-PRM analyses of spermatozoa extracts were performed on a NanoAcquity LC-system (Waters, Milford, MA, USA) coupled to a Q-Exactive Plus Orbitrap (Thermo Fisher Scientific, Waltham, MA, USA) mass spectrometer equipped with a nanoelectrospray ion source. The HPLC system consisted of a solvent degasser nanoflow pump, a thermostated column oven kept at 60°C and a thermostated autosampler kept at 10°C. Mobile phase A (99.9% water and 0.1% FA) and mobile phase B (99.9% acetonitrile and 0.1%FA) were delivered at 450 nL/min. Samples were loaded into a Symmetry C18 precolumn (0.18x20mm, 5µm particle size, Waters) over 3 minutes in 1% buffer B at a flow rate of 5 µL/min. This step was followed by reverse-phase separation at a flow rate of 450 nL/min

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

using an ACQUITY UPLC® BEH130 C18 separation column (250mmx75µm id, 1.7µm particle size, Waters). Peptides were separated using the following gradient: from 1% B to 5% B in 2 minutes, from 5% B to 25% B in 58 minutes, from 25% B to 35% B in 19 minutes, from 35% B to 90% B in 1 minute, maintained at 90% B for 5 minutes and the column was reconditioned at 1% B for 20 minutes.

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 2 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 250°C. The Q-Exactive Plus Orbitrap instrument was operated as follows: after a full MS scan, twelve peptide ions defined in the inclusion list were selected for MS/MS (all 212 targets and time scheduling options are listed in Supplementary Table 1 and details on the PRM time-scheduled method development are provided in Supplementary Material file). Full scan MS spectra (mass range 300-1800m/z) were acquired with a resolution of 70,000 at 200 m/z with an automatic gain control (AGC) fixed at 3×10^6 ions and a maximum injection time set at 50 ms. Targeted MS/MS spectra were acquired with a resolution of 17,500 at 200 m/z with an AGC fixed at 1×10^5 and with the maximum injection time set at 60 ms. LC-PRM .raw and Skyline results files were submitted to the PeptideAtlas/PASSEL repository (PASS01013), dot-product and rdot-product values were calculated with Skyline, all results were manually checked and validated peptides and traces are provided in Figure 1, Table 1 and Supplementary Figure 1.

In situ validation of expression data by immunohistochemistry

To confirm the germline expression of proteins of interest, immunohistochemistry experiments were performed on human testes, fixed in 4% paraformaldehyde and embedded in paraffin, as previously described ¹⁷. Normal human testes were collected at autopsy at Rennes University Hospital from HIV-1-negative cadavers. Paraffin-embedded tissues were cut into 4 µm thick slices, mounted on positively charged slides, and dried at 58°C for 60 min. Immunohistochemical staining was performed on the Discovery Automated IHC stainer using the Ventana DABMap or OMNIMap detection kit (Ventana Medical Systems, Tucson, USA). Following deparaffinization with Ventana EZ Prep solution at 75°C for 8 minutes, antigen retrieval was performed using Tris-based buffer solution CC1 (Ventana Medical Systems) at 95-100°C for 36 minutes. Endogen peroxidase was blocked with Inhibitor-D 3% H₂O₂ (Ventana Medical Systems) for 8 minutes at 37°C. After rinsing with reaction buffer (Roche), slides were incubated at 37°C for 60 minutes with polyclonal rabbit antibodies specific for the selected missing proteins (Human Protein Atlas Antibodies). The antibody dilutions used are listed in Supplementary Table 2. After rinsing, signal enhancement was performed at 37°C using the Ventana DABMap Kit or OMNIMap and secondary antibody: Anti-Rabbit HRP (pre-diluted commercial solution, Roche) for 16 minutes or Biotinylated horse anti Rabbit (1:700) for 32 minutes incubation. Slides were then counterstained for 16 minutes with hematoxylin (commercial solution, Microm), 4 minutes with bluing reagent (commercial solution, Microm), and rinsed with Milli-Q water. After removal from the instrument, slides were manually dehydrated, mounted in Eukitt (Labnord, Villeneuve

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

d’Ascq, France) and coverslipped. Finally, digital immunohistochemistry images were obtained using NDP.Scan acquisition software (v2.5, Hamamatsu) and visualized with NDP.View2 software (Hamamatsu).

Immunocytochemistry on sperm smears

Immunocytochemical staining was performed on the Discovery Automated IHC stainer using the Ventana DABMap or OMNIMap detection kit (Ventana Medical Systems, Tucson, Ariz). After cell fixation and permeabilization in methanol, sperm smears were treated using the same procedure as the one used for immunohistochemistry (see above). Immunocytochemistry images were visualized under an Eclipse NI-E microscope (Nikon, Champigny sur Marne, France).

RESULTS AND DISCUSSION

Previous studies by our consortium lead to the identification of 206 (PE2 to PE4) previously unvalidated proteins in human sperm and in line with current HPP guidelines^{11 12}. According to testis transcript expression from publicly available RNA-Seq datasets (*i.e.*, ArrayExpress and HPA), we can estimate that >500 remaining missing proteins could be preferentially evidenced in the testis. As most of these will correspond to genes expressed in meiotic and post meiotic germ cells, the ejaculated spermatozoa still appear as a relevant source for their identification.

In a recent work, Duek and collaborators proposed a subset of 40 missing proteins from chromosome 2 and 14 as being theoretically detectable in sperm¹⁴. Among these proteins, FAM71D was qualified as a “one hit wonder” requiring further validation because only one out of the two peptides identified in our initial work met our quality criteria¹². However, both peptides were validated during the reanalysis of our data by PeptideAtlas, and the protein was finally validated in neXtProt. For C2orf82 no unique peptide could be designed for the PRM assay. We report here the search of the remaining 38 proteins using targeted MS and immunohistochemistry methods.

Targeted MS validation of 12 missing proteins

The developed LC-PRM assay targets 92 peptides selected to be proteotypic for the 38 MPs following SRM Atlas¹⁸ suggestions and manual amino acid sequence analysis. In a first attempt, we searched for the MPs in total protein extracts of both Triton X-100 soluble and insoluble fractions concentrated in stacking gel bands. This attempt failed as none of the targeted peptides could be detected in these

unfractionated samples. The second attempt including a prior extensive 1D-SDS-PAGE fractionation in 23 bands of each of the soluble and insoluble fractions revealed to be far more successful. Indeed, signals were detected for 29 out of the 92 tracked peptides enabling 15 out of the 38 targeted MPs to be validated (Table 1). Complementary results were obtained thanks to the Triton X-100 fractionation as 17 and 14 peptides were validated in the soluble and insoluble fraction, respectively (Table 1). It is of note that even with thorough fractionation, the measured signals were overall of low intensity and for most peptides close to detection limits (Supplementary Figure 1). All chromatographic traces were manually checked and the following criteria needed to be met for a peptide to be validated: a good co-elution between light and heavy transition groups, high calculated dot-product for endogenous peptide (all >0.85, average value=0.93) and rdot-product (all >0.94, average value=0.99) and repeated detection of the endogenous transition group (mostly in adjacent gel bands or in replicate injections). Finally, signals of unlabeled peptide in the crude heavy labeled peptides were checked to be below detection limits and transition groups unspecifically detected in multiple gel bands were excluded. Altogether, these results indicate that the 12 MPs detected with 2 or more peptides and the 3 observed with a single peptide using targeted MS may have escaped prior detection because they are expressed at very low copy numbers.

Targeted MS results support data mining prioritization strategy

The prioritization strategy presented by Duek and collaborators ¹⁴ was based on information retrieval in the literature and from the three main C-HPP resources (Human Protein Atlas, PeptideAtlas and neXtProt). Proteins that were experimentally

1
2
3 validated but not annotated as such in neXtProt due to the unavoidable lag time in
4 paper curation were eliminated; so were the proteins expected to be difficult to
5 access and identify (e.g., olfactory receptors, pseudogenes or proteins refractory to
6 trypsin). Finally, proteins with enriched expression in testis were prioritized (high,
7 medium and low priority) for targeted MS-based experiments.
8
9

10
11 The present study demonstrates that this strategy was relevant. Indeed, out of 11
12 proteins ranked as "high priority" (enriched in testis according to Human Protein Atlas
13 and neXtProt), 6 were confirmed with ≥ 2 peptides (ADAM20, FAM181A, DNAJC5G,
14 ALLC, MAP3K19 and FER1L5). The 5 that we failed to identify were SULT6B1,
15 MROH2A, RNF212B, RNASE11 and COX8C. COX8C is predicted to be located in
16 the inner membrane of mitochondria and might need prior mitochondrial enrichment.
17 According to neXtProt, RNF212B is predicted to act in the synaptonemal complex,
18 which is disassembled during sperm maturation. Therefore, it might be present only
19 in germ cells but not in mature sperm. RNASE11 is highly expressed in testis and
20 epididymis, according to recent RNAseq data. However, since it is predicted to be
21 secreted, it might be easier to detect it in seminal plasma than in sperm cells.
22
23

24
25 Out of 19 proteins considered "medium priority" (testis enriched according to Human
26 Protein Atlas but not neXtProt), 5 were confirmed with ≥ 2 peptides (SLC35F4,
27 TEX22, LRRC9, C2orf78 and EVX2) and 2 had a single peptide (KIAA2012 and
28 ANKRD30BL). In addition, the HERV-H_2q24.1 and HERV-H_2q24.3 provirus
29 ancestral Env polyproteins, which were added to the selection by literature mining,
30 were confirmed with 2 and 1 peptides respectively.
31
32

33
34 Based on RNA-Seq data from HPA, six MPs for which we obtained mass
35 spectrometry evidence in sperm (FAM181A, DNAJC5G, LRRC9, ALLC, C2orf78,
36
37

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

MAP3K19) correspond to genes showing a medium expression level in testis (between 10 and 100 transcripts per million, TPM), seven others (ADAM20, TEX22, EVX2, SLC35F4, FER1L5, KIAA2012 and ANKRD30BL) correspond to genes with a low expression level (less than 10 TPM) and two others to genes for which the expression in testis was validated by quantitative RT-PCR while no RNA-Seq data was reported (HERVs) ¹⁴.

Although they were selected according to their preferential expression in testis, the 23 MPs that could not be validated may be present at levels below the PRM detection limit in sperm. Indeed, the expression level of a transcript does not necessarily indicate that the corresponding protein will be readily detectable in a specific organ/cell. As described in the pioneer study by Cagney and collaborators ¹⁹ there is a low degree of correlation between gene and protein expression in the testis (*i.e.*, 0.138 versus 0.432 for the liver, which displayed the strongest correlation). This prominent uncoupling between transcription and translation was further studied by Gan and collaborators in the mouse testis ²⁰. The authors observed that several regulation mechanisms including transcript degradation, translation repression and de-repression and protein degradation affected most genes and may account for the low correlation between mRNAs and proteins during spermatogenesis ²⁰. This original feature of spermatogenesis is mainly linked to the chromatin condensation in elongating spermatid and the complete cessation of transcription from mid-spermiogenesis onwards. As a consequence, many genes are transcribed much earlier during germ cell differentiation, then sequestered in chromatoid bodies, translationally repressed and finally translated several days after the mRNA production (for a review, see ²¹).

Orthogonal antibody-based validation of proteins

Immunohistochemistry and immunocytochemistry experiments were carried out on adult human testes and sperm smears for 7 MPs that were detected by PRM and for C14orf177 for which detected MS signals did not meet quality requirements to be considered as validated, and for which an antibody was available from the Human Protein Atlas (Figures 2, 3). Results from our experiments show that for all these proteins, no staining above background levels was observed in somatic Sertoli cells (Figures 2, 3), while we were able to visualize the expression of these proteins in the germ cell lineage.

Immunoreactivity of LRRC9, EVX2 and ALLC was observed in most germ cells whereas signals were weak in the cytoplasmic lobes of elongated spermatids (Figures 2A, B, C) and absent in ejaculated spermatozoa (data not shown). This clearly indicates that these proteins are expressed at residual levels in the mature gamete, below the detection limits of immunochemistry methods but at detectable levels for targeted MS-based approaches.

LRRC9 gene expression and phylogenetic profiles indicate that it is a candidate ciliary gene²². It is conserved in Chordata and several ciliated/flagellated organisms. The *Chlamydomonas* ortholog FAP234 is expressed during ciliogenesis, detected by proteomics in the flagellum²³, and required for the axonemal localization of TTLL9, important for the assembly and function of ciliary/flagellar axonemes²⁴. LRRC9 was not detected in mature sperm by immunochemistry but could be detected with 2 unique peptides using PRM. Though, the PRM signals observed were close to detection limit suggesting that LRRC9 is expressed at very low copy numbers. This is

in agreement with the lack of identification by mass spectrometry in sperm from other model organisms such as mouse and rat. Taken together, this information suggests that LRRC9 could play a role in flagellum biogenesis.

EVX2 is a homeobox transcription factor homolog of *Drosophila* eve (even-skipped) and *Caenorhabditis elegans* vab-7, required for embryonic patterning and neuronal identity in these model species ^{25 26}. In vertebrates EVX2 is required for the development of limbs, genitalia and spinal cord ^{27 28}. In humans, a microdeletion removing HOXD9-HOXD13 and EVX2 causes synpolydactyly ²⁹. Using RNA-Seq, EVX2 transcripts were shown to be among the population of abundant and stable RNAs stored in human sperm that encode proteins that may serve a role in the final stages of spermiogenesis or at fertilization ³⁰. This is in accordance with our observation that EVX2 immunoreactivity is intense in the nuclei of pachytene spermatocytes but discrete in elongated spermatids. However, the high sensitivity of mass spectrometry allowed us to identify residual EVX2 protein in sperm extracts with 2 peptides.

ALLC is highly conserved from bacteria to vertebrates. In mouse, rat, bovine and human, ALLC transcripts were shown to be enriched in testes ³¹. In mouse testes, the expression seems finely regulated: it is detected from 2-5 to 29 days postpartum, with upregulation from 17 to 23 days postpartum and is detected again in adults older than 2 months ³². Another study shows that in 22-day-old mice the mRNA is rendered translationally inactive by sequestration at ribonucleoproteins, and released in adult mice ³³. The timing of expression in testis indicates that the gene would be expressed from meiosis to elongating spermatids (steps 7-8), downregulated during late spermiogenesis and re-expressed in mature sperm, in accordance with what we

observed at the protein level by immunohistochemistry and by the identification of 3 unique peptides in the insoluble fraction using PRM analysis. In many species, ALLC is performing allantoicase activity (EC 3.5.3.4) as a final step of the uric acid degradation pathway, but in mammals this pathway is incomplete and allantoicase activity is supposed to be absent³⁴. Therefore, the role of ALLC in mammals is intriguing and our observations suggest a specific role in spermatogenesis that will be interesting to investigate.

TEX22 and FAM181A displayed much more specific immunoreactivity patterns in human testis sections than LRRC9, EVX2 and ALLC. The staining intensity for **TEX22** gradually increased from premeiotic and meiotic germ cells onwards with a strong immunoreactivity observed in elongated spermatids, where it accumulated in cytoplasmic lobes/ residual bodies (Figure 2D). At spermiation - the process by which mature spermatids are released from Sertoli cells into the seminiferous tubule lumen, numerous germ cell proteins and organelles concentrate into the cytoplasmic lobes that will condense to form the residual bodies. In humans, most of these residual bodies will be left behind after disengagement and phagocytosed by Sertoli cells, but a few of them can be found in ejaculated sperm (for a review see³³). TEX22 immunoreactivity could not be seen on ejaculated spermatozoa. This may indicate that most copies of the protein are eliminated during through residual bodies and that the TEX22 protein that was identified with two distinct peptides by PRM in sperm cell extracts only corresponds to a few copies, below the limits of detection of immunochemistry techniques. Another explanation for the loss of recognition of human TEX22 by antibodies in ejaculated sperm is that specific post-translational modifications occurring in mature sperm could affect recognition by the antibody.

Such post-translational modifications have been suggested for Tep22, the mouse TEX22 ortholog³⁵. The observed patterns of expression and localization of TEX22 and Tep22 are not strictly identical: Tep22 transcripts were detected in 2-week-old mouse testes onwards³⁶, while Tep22 protein was detected by mass spectrometry in male haploid germ cells³⁷. Tep22 displayed a dynamic localization during spermatogenesis, *i.e.*, in the acrosome of elongating spermatids at early spermiogenesis, in the cytoplasm later, then in the midpiece, and both in the midpiece and mitochondrial sheath in mature spermatozoa³⁸. The function of the mouse protein is not clear: Tep22 knock-out mice are fertile suggesting that this gene is not essential for mouse male fertility³⁶. Data obtained from mouse cannot be entirely extrapolated to humans as TEX22 is quite highly evolving in eutherians³⁶, so a function of TEX22 at early stages of spermatogenesis cannot be excluded.

FAM181A immunoreactive signal was intense in pachytene spermatocytes and weak in postmeiotic germ cells with a discrete staining in the cytoplasmic lobes of elongated spermatids (Figure 2E). Considering the intense expression of the protein in pachytene spermatocytes, it is surprising that FAM181A could not be identified in proteomics studies performed on total testis extracts^{39 40}. FAM181A was unambiguously identified by PRM in sperm cell extracts thanks to two peptides, but no FAM181A immunoreactivity could be seen on ejaculated spermatozoa (data not shown), meaning expression levels were below the limits of detection of immunochemistry techniques. FAM181A is conserved in Vertebrates. In mouse, Fam181a expression appears mainly limited to neural tissues where it shows extensive overlap with its paralog Fam181b. Both murine Fam181 tagged proteins localize to the nucleus in transiently transfected cells. Additionally, Fam181a is found

as weak speckles in the cytoplasm ⁴¹. On the other hand, FAM181A interacts with MAPRE1/EB1, a microtubule plus-end tracking protein ⁴² which was shown to interact with RABL2, a protein required for sperm intra-flagellar transport ⁴³. These observations suggest that FAM181A could function during early meiosis at the nucleus and could regulate cytoskeletal dynamics at later stages. Cytoskeletal dynamics are crucial for the series of events occurring from meiosis and onwards including the Sertoli cell-germ cell dialog and spermatid maturation (for review see ⁴⁴).

ADAM20 immunoreactive signal is very strong in the flagella of elongated spermatids (arrow) at stage prior spermiation (Figure 3A) and this regionalized immunoreactivity was confirmed on the flagella of ejaculated spermatozoa (Figure 3B). ADAM20 belongs to the disintegrin and metalloprotease (ADAM) transmembrane family of proteases conserved in various Metazoa, where they play functions in development, inflammation, adhesion, migration and fertilization. In mammals, more than half of these proteins are expressed specifically or predominantly in testes and epididymis, suggesting a role in male reproduction ^{45 46}. According to neXtProt, there are 21 ADAM protein-coding genes in human. Except ADAM20 that has only been validated at transcript level (PE2) and ADAM5 that is classified as a possible pseudogene (PE5), all of them have been validated at protein level (PE1). According to HPA, nine human ADAM genes are preferentially expressed in testis or epididymis: ADAM2, ADAM7, ADAM18, ADAM20, ADAM21, ADAM28, ADAM29, ADAM30 and ADAM32.

Usually, ADAM proteins are present as precursors in the testis, and processed to mature forms lacking the prodomain or both the prodomain and metalloproteinase domain between the stages of testicular to cauda epididymal sperm ⁴⁵. Interestingly,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

one of the two peptides from ADAM20 that were validated by PRM is located inside the metalloproteinase domain, indicating that there was no processing further than cleavage of the prodomain.

The involvement of ADAM proteins in fertilization largely derives from gene knockout studies in mice. However, the orthology between mouse and human genes varies considerably especially for the testis enriched expressing ADAM genes. Adam1, 3, 4, 6, 24 and 26 are pseudogenes or absent in humans while ADAM20 is absent in mouse. Mouse Adam1, 2, 4, 6, 24, 26, and 30 localize in the head of mature sperm⁴⁷, Adam3 in the acrosome of spermatids and mature sperm⁴⁸, and Adam15 in the acrosome and flagellum⁴⁹. Mouse Adam7 is synthesized in the epididymis, secreted into the epididymal lumen and then transferred to sperm membranes⁵⁰. Mice lacking Adam1, Adam2 and Adam3 show deficiency in sperm migration from the uterus to the oviduct, reduced sperm binding to egg zona pellucida, and sperm-egg membrane adhesion and fusion^{51 52 53}. Mice lacking Adam7 show decreased sperm motility, flagellum deformation and modified epididymal epithelium⁵⁴. The very strong signal observed for ADAM20 in flagella suggests a role for this protein in sperm motility.

LC-PRM detection of **C14orf177** was inconclusive as the signal of a unique 8 amino-acids long peptide did not meet detection quality criteria required to be MS-validated. Nevertheless, a cross-validation by Ab-based techniques had been initiated and successfully pursued. C14orf177 immunoreactivity was observed in the cytoplasm of germ cells at all stages of development up to the cytoplasmic lobes of elongated spermatids (Figure 3C). In ejaculated spermatozoa, a ring-shaped signal was visible in the headpiece (Figure 3D). This peculiar staining corresponds to the *Annulus*, a structure located beneath the plasma membrane that connects the midpiece and the

principal piece of mature sperm flagellum (for a review see ⁵⁵). Its biogenesis and functional significance remain unclear but the *Annulus* might behave as a morphological organizer, guiding flagellum assembly and preventing caudal displacement of mitochondria during spermiogenesis ^{56 57}, and as a diffusion barrier, confining proteins to particular compartments of the mature sperm tail ⁵⁸. Interestingly, only a few proteins have been shown to date to localize to the *Annulus* in the mouse and human (for a review, see ⁵⁵). C14orf177 immunoreactivity was also observed in the midpiece of ejaculated sperm (Figure 3D). C14orf177 is an uncharacterized protein conserved in Primates with no homology to known domains or proteins, except in the first 35 amino acids with the C-terminal terminal part of SPACA9. In mouse Spaca9k localizes at the sperm head, midpiece and flagellum ⁵⁹, and was proposed as a candidate for ciliary function ⁶⁰. Although the function of C14orf177 remains unknown, its observed localization in the headpiece and midpiece of ejaculated sperm suggests it could play a role in the flagellum biology.

DNAJC5G was validated by PRM with 2 unique peptides. DNAJC5G immunoreactivity was observed in all germ cells up to the cytoplasmic lobes of elongated spermatids (Figure 3E). As for C14orf177, DNAJC5G staining was also observed as ring-shaped in the headpiece and in the midpiece of ejaculated spermatozoa (Figure 3F), suggesting a role in the flagellum biology. DNAJC5G/CSP-gamma is a member of the DnaJ/Hsp40 family of chaperones, which localize in various subcellular compartments (mitochondria, endoplasmic reticulum, nucleus, and cytosol) and have many housekeeping and stress-associated functions in folding, translocation and degradation of proteins, as well as in gene expression and translational initiation. There are 45 entries for human DnaJ protein-coding genes in

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

neXtProt. Except for DNAJB7, DNAJC22 and DNAJC5G, they all have been validated at protein level (PE1). According to HPA, the majority (75%) of DnaJ proteins are broadly expressed. Eight of them are preferentially expressed in testis (DNAJB7, DNAJB8, DNAJB13, DNAJC5B, DNAJC5G, DNAJC18, DNAJC27 and DNAJC28).

Based on their architecture, DnaJ/Hsp40 proteins are classified in three main subfamilies. Several type I and type II HSP40 family members have been shown to be involved in male reproduction. DNAJB13 is involved in axoneme formation ⁶¹ and mutations cause primary ciliary dyskinesia and male infertility in humans ⁶². DnaJA1 regulates androgen receptor signaling and spermatogenesis in mouse ⁶³. DNAJB3/MSJ1 is highly expressed in haploid male germ cells in mouse and *Macaca fuscata* ⁶⁴ ⁶⁵ and associates with the acrosome and centriolar region in mouse germ cells ⁶⁶ ⁶⁷ ⁶⁸. DnaJB1 was identified as an acrosome and flagella associated component in mouse and rat spermatozoa ⁶⁹. DNAJC5G/ CSP-gamma is a type III DnaJ/Hsp40 protein and its closest homologs are DNAJC5/CSP and DNAJC5B/CSP-beta. DNAJC5/CSP, which is conserved in invertebrates such as *Drosophila* and *Caenorhabditis elegans*, is widely expressed and has a general role in exocytosis and in presynaptic function ⁷⁰ ⁷¹. DNAJC5B/CSP-beta and DNAJC5G/CSP-gamma are not conserved outside of Vertebrates and are less characterized. They are localized in nerve terminals and secretory organelles in mouse brain, where they may contribute to regulated secretory events ⁷² ⁷³. DNAJC5B/ CSP-beta was found in rat testis, in a palmitoylated form at the plasma membrane ⁷⁴ and as a non-palmitoylated form in the trans-Golgi network ⁷⁵, but its potential role in male reproduction is still unclear. The presence of DNAJC5G/ CSPgamma in the headpiece and the midpiece of mature spermatozoa

1
2
3 leads us to speculate that it could play a protective role towards sperm flagellar
4
5 function under stress conditions.
6

7
8 C2orf78, SLC35F4, MAP3K19, FER1L5 and the retroviral sequence HERV-H/Env59
9
10 were validated by at least 2 peptides in PRM experiments. Unfortunately, we were
11
12 not able to characterize them further and pinpoint their localization by antibody-based
13
14 approaches. It was recently shown that MAP3K19 is specifically expressed in lung
15
16 and testis and may play a role in TGF-beta signaling ⁷⁶, and that HERV-H/Env59 may
17
18 have immunosuppressive activity ⁷⁷.
19

20
21 KIAA2012, ANKRD30BL and HERV-H/Env62 were detected by PRM with a single
22
23 unique peptide. Their validation will thus require further experiments to comply with
24
25 the HPP guidelines.
26
27
28
29
30

31 ***Perspectives for the search of MPs from chromosome 2 and 14***

32
33 We hope that the present data will be used to validate the existence of 12 missing
34
35 proteins in PeptideAtlas and neXtProt. Our consortium continues to look for the
36
37 remaining missing proteins whose expression may be preferentially observable in the
38
39 male gamete cells. The present work reveals we have certainly reached technical
40
41 limits due to the low level of expression of the targeted proteins. To go beyond this
42
43 limit, we will mine the most recent RNA-Seq data to select a larger set of MPs from
44
45 chromosomes 2/14 expected to be present in sperm, and apply targeted-MS assays
46
47 to a collection of enriched subcellular fractions (*i.e.*, head, midpiece and flagellum) of
48
49 human sperm samples. In case of unsuccessful detection in sperm fractions, the
50
51 developed LC-PRM assays will be used to detect specific MPs in cell line or tissue
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

extracts in which mRNA expression was also demonstrated. We suggest that option to become the step two of the neXt50 challenge for our teams.

As a preliminary work in that direction, we mined literature and The Global Proteome Machine Database (<http://gpmdb.thegpm.org/index.html>) to find identified proteotypic peptides that did not meet the required quality or length criteria to be considered for protein validation or had not been submitted to ProteomeXchange, in order to identify other tissues or cell lines in which these proteins could be present. We found that 14 missing proteins might be present in commonly used cell lines, and 3 others (RGPD3, C2orf27 and GPR148) in the brain. COX8C, UNC79, TM4SF20 and GPR148 that have transmembrane domains (1, 2, 4 and 7, respectively) may require the development of specific solubilization and/or enrichment protocols. For the three other proteins, the use of a detergent other than Triton X-100 may be required due to the presence of more than one transmembrane domain. Indeed, the use of Triton X-100 at 1% final concentration is not a specific membrane solubilization method and only facilitates the solubilization of membrane-associated hydrophobic proteins.

Additionally, to help other C-HPP teams in the global neXt50 challenge, we propose to select new sets of MPs from other chromosomes based on preferential expression in testis and perform targeted-MS searches in sperm extracts.

The observed decline in male reproductive potential worldwide calls for dedicated efforts to characterize male-factor defects. From the standpoint of understanding sperm physiopathology, the characterization and identification of new structural and functional sperm proteins in a systematic manner merits further research. The

1
2
3 missing proteins that we currently identify in the human sperm are expressed at very
4
5 low levels and many more remain to be identified. These proteins about which we
6
7 have limited knowledge to date could help in determining sperm quality apart of
8
9 semen analysis, especially in the case of idiopathic male infertilities. Alterations in the
10
11 expression of several proteins were recently observed with respect to poor blastocyst
12
13 development and in vitro fertilization outcome in donor oocyte cycles despite normal
14
15 sperm testing parameters ⁷⁸. It would be interesting to assess if such modifications
16
17 are found at a larger scale, including on newly identified proteins. In addition to
18
19 participating to the C-HPP effort of validating human missing proteins, we hope our
20
21 data will contribute to the continued exploration of the sperm proteome, eventually
22
23 offering new potentials for clinical diagnostic testing of spermatozoa for infertility
24
25 which is limited to the counting and assessment of motility and gross morphology at
26
27 present.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Legends to Figures

Figure 1: LC-PRM data validating Protein FAM181A (Q8N9Y4). Selected transition groups were extracted for the two unique peptides validating Protein FAM181A, peptide ASCSGPLVMASDSDVK and peptide MPGVSLVGR. The traces for the endogenous and heavy labeled peptides were extracted and dot-product and rdot-product values were calculated using Skyline software.

Figure 2: Antibody staining for LRRC9 (Q6ZRR7), EVX2 (Q03828), ALLC (Q8N6M5), TEX22 (C9J3V5) and FAM181A (Q8N9Y4), and in adult human testis. Proteins were detected in transverse testis sections at stages IV to VI of the seminiferous epithelium ⁷⁹ using polyclonal antibodies from the HPA specific for LRRC9 (HPA041381) (A), EVX2 (HPA041576) (B), ALLC (HPA037742) (C), TEX22 (HPA065576) (D) and FAM181A (HPA001603) (E). Nonimmune serum was used as a negative control (data not shown). In all testis sections, a more or less intense antibody staining signal was visible in all germ cells and for all proteins (A-F). LRRC9 presented a strong staining in the cytoplasm of spermatogonia (arrows) and a moderate staining in the cytoplasmic lobes of elongating spermatids (square) (A). EVX2 immunoreactivity displayed as an intense punctuate staining in the nuclei of pachytene spermatocytes (arrow) and early spermatids (arrowhead) (B). Staining was discrete in the cytoplasmic lobes of elongated spermatids (B; square). ALLC presented strong staining in the cytoplasm of pachytene spermatocytes (arrowhead) and early spermatids (arrow) (C). A low ALLC immunoreactivity was also visible in

the cytoplasmic lobes of elongated spermatids (C; square). TEX22 immunoreactivity is intense in germ cells at all development stages and concentrated in the cytoplasm of elongated spermatids (D; arrows). FAM181A immunoreactivity was intense in pachytene spermatocytes (arrows) and a discrete staining was observed in the cytoplasmic portion of elongated spermatids (square) (E). Scale bars = 20 μ m.

Figure 3: Antibody staining for ADAM20 (O43506), C14orf177 (Q52M58), and DNAJC5G (Q8N7S2) in adult human testis (A,C,E) and in isolated spermatozoa (B,D,F). Proteins were detected in transverse testis sections at stages IV to VI of the seminiferous epithelium ⁷⁹ and on sperm smears using polyclonal antibodies from the HPA specific for ADAM20 (HPA059377) (A), C14orf177 (HPA018091) (C), and DNAJC5G (HPA041445) (E). ADAM20 immunoreactivity is very strong in the flagella of elongated spermatids (arrow) at stage prior spermiation (A). The regionalized immunoreactivity of ADAM20 is confirmed on the flagella of ejaculated spermatozoa (arrows) whereas the midpiece is clearly unstained (arrowhead) (B). C14orf177 immunoreactivity is visible in the cytoplasm of germ cells at all stages of development up to the cytoplasmic lobes of elongated spermatids (arrow), whereas it appeared stronger in the nucleus of premeiotic and meiotic germ cells (arrowhead) (C). In ejaculated spermatozoa, a ring-shaped signal was visible for C14orf177 in the headpiece (arrows) and the midpiece (arrowhead) (D). DNAJC5G immunoreactivity was intense in germ cells at all stages of their development up to the cytoplasmic lobes of elongated spermatids (arrow), whereas a stronger signal was observed in the cytoplasm of spermatogonia and preleptotene spermatocytes (arrowhead) (E).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DNAJC5G staining was observed in the headpiece (arrow) and midpiece (arrowhead) of ejaculated spermatozoa (F).

Table 1: List of 12 MPs identified with 2 or more peptides and 3 candidate MPs identified with 1 peptide by targeted LC-PRM. Gene names, protein IDs, protein names, peptide sequences and fraction (Triton X-100 soluble and/or insoluble fraction) in which these 15 missing proteins were identified are provided.

Supporting information

This material is available free of charge via <http://pubs.acs.org/>.

Supplementary Material: Analytical details on the acquisition of reference MS/MS spectra for the 92 proteotypic peptides corresponding to the 38 targeted missing proteins and on the targeted LC-PRM time-scheduled method development.

Supplementary Table 1: Targeted LC-PRM inclusion list. List of the 212 targets followed during LC-PRM analyses: precursor m/z ratio, charge state, polarity, scheduling start and end times, normalized collision energy (CE), amino acid sequence including modifications and precursor type (light form and heavy labeled form), are provided for each target.

Supplementary Table 2: List of missing proteins for which immunohistochemistry or immunocytochemistry was successful, HPA antibody names and dilutions used.

Supplementary Figure 1: LC-PRM data of validated peptides. For all validated peptides, traces of selected transition groups were extracted for the endogenous and heavy labeled peptides and dot-product and rdot-product values were calculated using Skyline.

Notes:

The authors declare no competing financial interest.

Author's contribution:

CC, LL and CP co-coordinated the study. TF and KR performed spermatozoa preparation. CC, CP, LL and YV conceived and designed the experiments and analyses. CC, FD and CM conducted LC-PRM experiments and validated LC-PRM data and peptides. PD, LL and CP performed extensive bioinformatics analysis and data/literature mining. CL, CP, KR and MS designed and performed immuno-histochemistry and -cytochemistry studies. CC, PD, LL, CP and YV wrote the manuscript. All the authors approved the final version of the manuscript.

Acknowledgements:

This work was partially funded through the French National Agency for Research (ANR) (grant ANR-10-INBS-08; ProFI project, "Infrastructures Nationales en Biologie et Santé"; "Investissements d'Avenir" call). This work was also supported by grants from Biogenouest, *Infrastructures en Biologie Santé et Agronomie* (IBiSA) and *Conseil Régional de Bretagne* awarded to C.P. We are grateful to Mathias Uhlen (Human Protein Atlas, Sweden) for the gift of antibodies, to Alain Gateau for help

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

with sequence analysis, and to Monique Zahn-Zabal for critical reading of the manuscript.

References

1. Legrain, P.; Aebersold, R.; Archakov, A.; Bairoch, A.; Bala, K.; Beretta, L.; Bergeron, J.; Borchers, C. H.; Corthals, G. L.; Costello, C. E.; Deutsch, E. W.; Domon, B.; Hancock, W.; He, F.; Hochstrasser, D.; Marko-Varga, G.; Salekdeh, G. H.; Sechi, S.; Snyder, M.; Srivastava, S.; Uhlen, M.; Wu, C. H.; Yamamoto, T.; Paik, Y. K.; Omenn, G. S., The human proteome project: current state and future direction. *Molecular & cellular proteomics : MCP* 2011, 10 (7), M111 009993.
2. Gaudet, P.; Michel, P. A.; Zahn-Zabal, M.; Britan, A.; Cusin, I.; Domagalski, M.; Duek, P. D.; Gateau, A.; Gleizes, A.; Hinard, V.; Rech de Laval, V.; Lin, J.; Nikitin, F.; Schaeffer, M.; Teixeira, D.; Lane, L.; Bairoch, A., The neXtProt knowledgebase on human proteins: 2017 update. *Nucleic acids research* 2017, 45 (D1), D177-D182.
3. Deutsch, E. W.; Overall, C. M.; Van Eyk, J. E.; Baker, M. S.; Paik, Y. K.; Weintraub, S. T.; Lane, L.; Martens, L.; Vandenbrouck, Y.; Kusebauch, U.; Hancock, W. S.; Hermjakob, H.; Aebersold, R.; Moritz, R. L.; Omenn, G. S., Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. 201-, 15 (11), 39-1-3970.
4. Carapito, C.; Lane, L.; Benama, M.; Opsomer, A.; Mouton-Barbosa, E.; Garrigues, L.; Gonzalez de Peredo, A.; Burel, A.; Bruley, C.; Gateau, A.; Bouyssie, D.; Jaquinod, M.; Cianferani, S.; Burlet-Schiltz, O.; Van Dorsselaer, A.; Garin, J.; Vandenbrouck, Y., Computational and Mass-Spectrometry-5ased Workflow for the Discovery and Validation of Missing Human Proteins: Application to Chromosomes 2 and 14. 2015, 14 (9), 3-21-34.
5. 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. 2014, 13 (1), 15-20.
6. Son, C. G.; Bilke, S.; Davis, S.; Greer, B. T.; Wei, J. S.; Whiteford, C. C.; Chen, Q. R.; Cenacchi, N.; Khan, J., Database of mRNA gene expression profiles of multiple human organs. *Genome research* 2005, 15 (3), 443-50.
7. 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.; Szigyanto, C. A.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P. H.; Berling, H.; Tegel, H.; Mulder, J.;

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

8. Uhlen, M.; Hallstrom, B. M.; Lindskog, C.; Mardinoglu, A.; Ponten, F.; Nielsen, J., Transcriptomics resources of human tissues and organs. *Molecular systems biology* 2016, 12 (4), 862.

9. Chocu, S.; Calvel, P.; Rolland, A. D.; Pineau, C., Spermatogenesis in mammals: proteomic insights. *Systems biology in reproductive medicine* 2012, 58 (4), 179-90.

10. 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. *Molecular human reproduction* 2014, 20 (6), 476-88.

11. Jumeau, F.; Com, E.; Lane, L.; Duek, P.; Lagarrigue, M.; Lavigne, R.; Guillot, L.; Rondel, K.; Gateau, A.; Melaine, N.; Guevel, B.; Sergeant, N.; Mitchell, V.; Pineau, C., Human Spermatozoa as a Model for Detecting Missing Proteins in the Context of the Chromosome-Centric Human Proteome Project. 2015, 14 (9), 3-0--20.

12. Vandenbrouck, Y.; Lane, L.; Carapito, C.; Duek, P.; Rondel, K.; Bruley, C.; Macron, C.; Gonzalez de Peredo, A.; Coute, Y.; Chaoui, K.; Com, E.; Gateau, A.; Hesse, A. M.; Marcellin, M.; Mear, L.; Mouton-5arbosa, E.; Robin, T.; Burlet-Schiltz, O.; Cianferani, S.; Ferro, M.; Freour, T.; Lindskog, C.; Garin, J.; Pineau, C., Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update. 201-, 15 (11), 3998-4019.

13. Gallien, S.; Kim, S. Y.; Domon, B., Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). *Molecular & cellular proteomics : MCP* 2015, 14 (6), 1630-44.

14. Duek, P.; Bairoch, A.; Gateau, A.; Vandenbrouck, Y.; Lane, L., Missing Protein Landscape of Human Chromosomes 2 and 14: Progress and Current Status. 201-, 15 (11), 3971-3978.

15. Baker, M. A.; Reeves, G.; Hetherington, L.; Muller, J.; Baur, I.; Aitken, R. J., Identification of gene products present in Triton X-100 soluble and insoluble fractions

of human spermatozoa lysates using LC-MS/MS analysis. *Proteomics. Clinical applications* 2007, 1 (5), 524-32.

16. Schaeffer, M.; Gateau, A.; Teixeira, D.; Michel, P. A.; Zahn-Zabal, M.; Lane, L., The neXtProt peptide uniqueness checker: a tool for the proteomics community. *Bioinformatics* 2017 May 17. doi: 10.1093/bioinformatics/btx318. [Epub ahead of print].

17. 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. *Molecular reproduction and development* 2006, 73 (7), 866-77.

18. Kusebauch, U.; Campbell, D. S.; Deutsch, E. W.; Chu, C. S.; Spicer, D. A.; Brusniak, M. Y.; Slagel, J.; Sun, Z.; Stevens, J.; Grimes, B.; Shteynberg, D.; Hoopmann, M. R.; Blattmann, P.; Ratushny, A. V.; Rinner, O.; Picotti, P.; Carapito, C.; Huang, C. Y.; Kapousouz, M.; Lam, H.; Tran, T.; Demir, E.; Aitchison, J. D.; Sander, C.; Hood, L.; Aebersold, R.; Moritz, R. L., Human SRMATlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell* 2016, 166 (3), 766-78.

19. Cagney, G.; Park, S.; Chung, C.; Tong, B.; O'Dushlaine, C.; Shields, D. C.; Emili, A., Human tissue profiling with multidimensional protein identification technology. 2005, 4 (5), 1757-&'A

20. Gan, H.; Cai, T.; Lin, X.; Wu, Y.; Wang, X.; Yang, F.; Han, C., Integrative proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory mechanisms of mouse spermatogenesis. *Molecular & cellular proteomics : MCP* 2013, 12 (5), 1144-57.

21. Hecht, N. B., Regulation of 'haploid expressed genes' in male germ cells. *Journal of reproduction and fertility* 1990, 88 (2), 679-93.

22. Treutlein, B.; Brownfield, D. G.; Wu, A. R.; Neff, N. F.; Mantalas, G. L.; Espinoza, F. H.; Desai, T. J.; Krasnow, M. A.; Quake, S. R., Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 2014, 509 (7500), 371-5.

23. Pazour, G. J.; Agrin, N.; Leszyk, J.; Witman, G. B., Proteomic analysis of a eukaryotic cilium. *The Journal of cell biology* 2005, 170 (1), 103-13.

24. Kubo, T.; Yanagisawa, H. A.; Liu, Z.; Shibuya, R.; Hirono, M.; Kamiya, R., A conserved flagella-associated protein in Chlamydomonas, FAP234, is essential for axonemal localization of tubulin polyglutamylase TTLL9. *Molecular biology of the cell* 2014, 25 (1), 107-17.

25. Doe, C. Q.; Smouse, D.; Goodman, C. S., Control of neuronal fate by the Drosophila segmentation gene even-skipped. *Nature* 1988, 333 (6171), 376-8.

26. Ahringer, J., Posterior patterning by the Caenorhabditis elegans even-skipped homolog vab-7. *Genes & development* 1996, 10 (9), 1120-30.

27. Herault, Y.; Hraba-Renevey, S.; van der Hoeven, F.; Duboule, D., Function of the Evx-2 gene in the morphogenesis of vertebrate limbs. *The EMBO journal* 1996, 15 (23), 6727-38.

28. Tarchini, B.; Huynh, T. H.; Cox, G. A.; Duboule, D., HoxD cluster scanning deletions identify multiple defects leading to paralysis in the mouse mutant Ironside. *Genes & development* 2005, 19 (23), 2862-76.

29. Goodman, F. R.; Majewski, F.; Collins, A. L.; Scambler, P. J., A 117-kb microdeletion removing HOXD9-HOXD13 and EVX2 causes synpolydactyly. *American journal of human genetics* 2002, 70 (2), 547-55.

30. Sandler, E.; Johnson, G. D.; Mao, S.; Goodrich, R. J.; Diamond, M. P.; Hauser, R.; Krawetz, S. A., Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic acids research* 2013, 41 (7), 4104-17.

31. Vigetti, D.; Binelli, G.; Monetti, C.; Prati, M.; Bernardini, G.; Gornati, R., Selective pressure on the allantoicase gene during vertebrate evolution. *Journal of molecular evolution* 2003, 57 (6), 650-8.

32. Ike, A.; Tokuhira, K.; Hirose, M.; Nozaki, M.; Nishimune, Y.; Tanaka, H., Comprehensive analysis of gene expression in testes producing haploid germ cells using DNA microarray analysis. *International journal of andrology* 2007, 30 (5), 462-75.

33. Iguchi, N.; Tobias, J. W.; Hecht, N. B., Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. *Proceedings of the National Academy of Sciences of the United States of America* 2006, 103 (20), 7712-7.

34. Leulliot, N.; Quevillon-Cheruel, S.; Sorel, I.; Graille, M.; Meyer, P.; Liger, D.; Blondeau, K.; Janin, J.; van Tilbeurgh, H., Crystal structure of yeast allantoicase reveals a repeated jelly roll motif. *The Journal of biological chemistry* 2004, 279 (22), 23447-52.
35. Choi, E.; Lee, J.; Oh, J.; Park, I.; Han, C.; Yi, C.; Kim, D. H.; Cho, B. N.; Eddy, E. M.; Cho, C., Integrative characterization of germ cell-specific genes from mouse spermatocyte UniGene library. *BMC genomics* 2007, 8, 256.
36. Miyata, H.; Castaneda, J. M.; Fujihara, Y.; Yu, Z.; Archambeault, D. R.; Isotani, A.; Kiyozumi, D.; Kriseman, M. L.; Mashiko, D.; Matsumura, T.; Matzuk, R. M.; Mori, M.; Noda, T.; Oji, A.; Okabe, M.; Prunskaitė-Hyyryläinen, R.; Ramirez-Solis, R.; Satouh, Y.; Zhang, Q.; Ikawa, M.; Matzuk, M. M., Genome engineering uncovers 54 evolutionarily conserved and testis-enriched genes that are not required for male fertility in mice. *Proceedings of the National Academy of Sciences of the United States of America* 2016, 113 (28), 7704-10.
37. Guo, X.; Shen, J.; Xia, Z.; Zhang, R.; Zhang, P.; Zhao, C.; Xing, J.; Chen, L.; Chen, W.; Lin, M.; Huo, R.; Su, B.; Zhou, Z.; Sha, J., Proteomic analysis of proteins involved in spermiogenesis in mouse. 2010, 9 (3), 124--56A
38. Neesen, J.; Hartwich, T.; Brandhorst, G.; Aumüller, G.; Glaser, B.; Burfeind, P.; Mendoza-Lujambio, I., Tep22, a novel testicular expressed gene, is involved in the biogenesis of the acrosome and the midpiece of the sperm tail. *Biochemical and biophysical research communications* 2002, 297 (4), 737-48.
39. Wei, W.; Luo, W.; Wu, F.; Peng, X.; Zhang, Y.; Zhang, M.; Zhao, Y.; Su, N.; Qi, Y.; Chen, L.; Zhang, Y.; Wen, B.; He, F.; Xu, P., Deep Coverage Proteomics Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF Mass Spectrometer. *2A1-*, 15 (11), 3988-3997.
40. Zhang, Y.; Li, Q.; Wu, F.; Zhou, R.; Qi, Y.; Su, N.; Chen, L.; Xu, S.; Jiang, T.; Zhang, C.; Cheng, G.; Chen, X.; Kong, D.; Wang, Y.; Zhang, T.; Zi, J.; Wei, W.; Gao, Y.; Zhen, B.; Xiong, Z.; Wu, S.; Yang, P.; Wang, Q.; Wen, B.; He, F.; Xu, P.; Liu, S., Tissue-5ased Proteogenomics Reveals that Human Testis Endows Plentiful Missing Proteins. *2A15*, 14 (9), 3583-94.

41. Marks, M.; Pennimpede, T.; Lange, L.; Grote, P.; Herrmann, B. G.; Wittler, L., Analysis of the Fam181 gene family during mouse development reveals distinct strain-specific expression patterns, suggesting a role in nervous system development and function. *Gene* 2016, 575 (2 Pt 2), 438-51.

42. Jiang, K.; Toedt, G.; Montenegro Gouveia, S.; Davey, N. E.; Hua, S.; van der Vaart, B.; Grigoriev, I.; Larsen, J.; Pedersen, L. B.; Bezstarosti, K.; Lince-Faria, M.; Demmers, J.; Steinmetz, M. O.; Gibson, T. J.; Akhmanova, A., A Proteome-wide screen for mammalian SxIP motif-containing microtubule plus-end tracking proteins. *Current biology : CB* 2012, 22 (19), 1800-7.

43. Lo, J. C.; Jamsai, D.; O'Connor, A. E.; Borg, C.; Clark, B. J.; Whisstock, J. C.; Field, M. C.; Adams, V.; Ishikawa, T.; Aitken, R. J.; Whittle, B.; Goodnow, C. C.; Ormandy, C. J.; O'Bryan, M. K., RAB-like 2 has an essential role in male fertility, sperm intra-flagellar transport, and tail assembly. *PLoS genetics* 2012, 8 (10), e1002969.

44. Lie, P. P.; Mruk, D. D.; Lee, W. M.; Cheng, C. Y., Cytoskeletal dynamics and spermatogenesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2010, 365 (1546), 1581-92.

45. Cho, C., Testicular and epididymal ADAMs: expression and function during fertilization. *Nature reviews. Urology* 2012, 9 (10), 550-60.

46. Moreno, R. D.; Urriola-Munoz, P.; Lagos-Cabre, R., The emerging role of matrix metalloproteases of the ADAM family in male germ cell apoptosis. *Spermatogenesis* 2011, 1 (3), 195-208.

47. Han, C.; Choi, E.; Park, I.; Lee, B.; Jin, S.; Kim, D. H.; Nishimura, H.; Cho, C., Comprehensive analysis of reproductive ADAMs: relationship of ADAM4 and ADAM6 with an ADAM complex required for fertilization in mice. *Biology of reproduction* 2009, 80 (5), 1001-8.

48. Kim, E.; Nishimura, H.; Iwase, S.; Yamagata, K.; Kashiwabara, S.; Baba, T., Synthesis, processing, and subcellular localization of mouse ADAM3 during spermatogenesis and epididymal sperm transport. *The Journal of reproduction and development* 2004, 50 (5), 571-8.

49. Pasten-Hidalgo, K.; Hernandez-Rivas, R.; Roa-Espitia, A. L.; Sanchez-Gutierrez, M.; Martinez-Perez, F.; Monrroy, A. O.; Hernandez-Gonzalez, E. O.;

- Mujica, A., Presence, processing, and localization of mouse ADAM15 during sperm maturation and the role of its disintegrin domain during sperm-egg binding. *Reproduction* 2008, 136 (1), 41-51.
50. Oh, J. S.; Han, C.; Cho, C., ADAM7 is associated with epididymosomes and integrated into sperm plasma membrane. *Molecules and cells* 2009, 28 (5), 441-6.
51. Cho, C.; Bunch, D. O.; Faure, J. E.; Goulding, E. H.; Eddy, E. M.; Primakoff, P.; Myles, D. G., Fertilization defects in sperm from mice lacking fertilin beta. *Science* 1998, 281 (5384), 1857-9.
52. Nishimura, H.; Kim, E.; Nakanishi, T.; Baba, T., Possible function of the ADAM1a/ADAM2 Fertilin complex in the appearance of ADAM3 on the sperm surface. *The Journal of biological chemistry* 2004, 279 (33), 34957-62.
53. Yamaguchi, R.; Muro, Y.; Isotani, A.; Tokuhira, K.; Takumi, K.; Adham, I.; Ikawa, M.; Okabe, M., Disruption of ADAM3 impairs the migration of sperm into oviduct in mouse. *Biology of reproduction* 2009, 81 (1), 142-6.
54. Choi, H.; Han, C.; Jin, S.; Kwon, J. T.; Kim, J.; Jeong, J.; Kim, J.; Ham, S.; Jeon, S.; Yoo, Y. J.; Cho, C., Reduced Fertility and Altered Epididymal and Sperm Integrity in Mice Lacking ADAM7. *Biology of reproduction* 2015, 93 (3), 70.
55. Dirami, T.; Rode, B.; Wolf, J. P.; Gacon, G.; Dulioust, E.; Toure, A., Assessment of the frequency of sperm annulus defects in a large cohort of patients presenting asthenozoospermia. *Basic and clinical andrology* 2015, 25, 10.
56. Fawcett, D. W., A comparative view of sperm ultrastructure. *Biology of reproduction* 1970, 2, Suppl 2:90-127.
57. Phillips, D. M., Mitochondrial disposition in mammalian spermatozoa. *Journal of ultrastructure research* 1977, (2), 144-54.
58. Cesario, M. M.; Bartles, J. R., Compartmentalization, processing and redistribution of the plasma membrane protein CE9 on rodent spermatozoa. Relationship of the annulus to domain boundaries in the plasma membrane of the tail. *Journal of cell science* 1994, 107 (Pt 2), 561-70.
59. Bhattacharya, R.; Devi, M. S.; Dhople, V. M.; Jesudasan, R. A., A mouse protein that localizes to acrosome and sperm tail is regulated by Y-chromosome. *BMC cell biology* 2013, 14, 50.

60. Stauber, M.; Weidemann, M.; Dittrich-Breiholz, O.; Lobschat, K.; Alten, L.; Mai, M.; Beckers, A.; Kracht, M.; Gossler, A., Identification of FOXJ1 effectors during ciliogenesis in the foetal respiratory epithelium and embryonic left-right organiser of the mouse. *Developmental biology* 2017, 423 (2), 170-188.

61. Li, W.; Liu, G., DNAJB13, a type II HSP40 family member, localizes to the spermatids and spermatozoa during mouse spermatogenesis. *BMC developmental biology* 2014, 14, 38.

62. El Khouri, E.; Thomas, L.; Jeanson, L.; Bequignon, E.; Vallette, B.; Duquesnoy, P.; Montantin, G.; Copin, B.; Dastot-Le Moal, F.; Blanchon, S.; Papon, J. F.; Lores, P.; Yuan, L.; Collot, N.; Tissier, S.; Faucon, C.; Gacon, G.; Patrat, C.; Wolf, J. P.; Dulioust, E.; Crestani, B.; Escudier, E.; Coste, A.; Legendre, M.; Toure, A.; Amselem, S., Mutations in DNAJB13, Encoding an HSP40 Family Member, Cause Primary Ciliary Dyskinesia and Male Infertility. *American journal of human genetics* 2016, 99 (2), 489-500.

63. Terada, K.; Yomogida, K.; Imai, T.; Kiyonari, H.; Takeda, N.; Kadomatsu, T.; Yano, M.; Aizawa, S.; Mori, M., A type I DnaJ homolog, DjA1, regulates androgen receptor signaling and spermatogenesis. *The EMBO journal* 2005, 24 (3), 611-22.

64. Yu, S. S.; Takenaka, O., Molecular cloning, structure, and testis-specific expression of MFSJ1, a member of the DNAJ protein family, in the Japanese monkey (*Macaca fuscata*). *Biochemical and biophysical research communications* 2003, 301 (2), 443-9.

65. Berruti, G.; Martegani, E., MSJ-1, a mouse testis-specific DnaJ protein, is highly expressed in haploid male germ cells and interacts with the testis-specific heat shock protein Hsp70-2. *Biology of reproduction* 2001, 65 (2), 488-95.

66. Berruti, G.; Martegani, E., mUBPy and MSJ-1, a deubiquitinating enzyme and a molecular chaperone specifically expressed in testis, associate with the acrosome and centrosome in mouse germ cells. *Annals of the New York Academy of Sciences* 2002, 973, 5-7.

67. Meccariello, R.; Berruti, G.; Chianese, R.; De Santis, R.; Di Cunto, F.; Scarpa, D.; Cobellis, G.; Zucchetti, I.; Pierantoni, R.; Altruda, F.; Fasano, S., Structure of msj-1 gene in mice and humans: a possible role in the regulation of male reproduction. *General and comparative endocrinology* 2008, 156 (1), 91-103.

68. Meccariello, R.; Chianese, R.; Ciaramella, V.; Fasano, S.; Pierantoni, R., Molecular chaperones, cochaperones, and ubiquitination/deubiquitination system: involvement in the production of high quality spermatozoa. *BioMed research international* 2014, 2014, 561426.
69. Doiguchi, M.; Kaneko, T.; Urasoko, A.; Nishitani, H.; Iida, H., Identification of a heat-shock protein Hsp40, DjB1, as an acrosome- and a tail-associated component in rodent spermatozoa. *Molecular reproduction and development* 2007, 74 (2), 223-32.
70. Burgoyne, R. D.; Morgan, A., Cysteine string protein (CSP) and its role in preventing neurodegeneration. *Seminars in cell & developmental biology* 2015, 40, 153-9.
71. Fernandez-Chacon, R.; Wolfel, M.; Nishimune, H.; Tabares, L.; Schmitz, F.; Castellano-Munoz, M.; Rosenmund, C.; Montesinos, M. L.; Sanes, J. R.; Schneggenburger, R.; Sudhof, T. C., The synaptic vesicle protein CSP alpha prevents presynaptic degeneration. *Neuron* 2004, 42 (2), 237-51.
72. Gundersen, C. B.; Kohan, S. A.; Souda, P.; Whitelegge, J. P.; Umbach, J. A., Cysteine string protein beta is prominently associated with nerve terminals and secretory organelles in mouse brain. *Brain research* 2010, 1332, 1-11.
73. Sammut, S. J.; Feichtinger, J.; Stuart, N.; Wakeman, J. A.; Larcombe, L.; McFarlane, R. J., A novel cohort of cancer-testis biomarker genes revealed through meta-analysis of clinical data sets. *Oncoscience* 2014, 1 (5), 349-59.
74. Gorleku, O. A.; Chamberlain, L. H., Palmitoylation and testis-enriched expression of the cysteine-string protein beta isoform. *Biochemistry* 2010, 49 (25), 5308-13.
75. Boal, F.; Le Pevelen, S.; Cziepluch, C.; Scotti, P.; Lang, J., Cysteine-string protein isoform beta (Cspbeta) is targeted to the trans-Golgi network as a non-palmitoylated CSP in clonal beta-cells. *Biochimica et biophysica acta* 2007, 1773 (2), 109-19.
76. Boehme, S. A.; Franz-Bacon, K.; DiTirro, D. N.; Ly, T. W.; Bacon, K. B., MAP3K19 Is a Novel Regulator of TGF-beta Signaling That Impacts Bleomycin-Induced Lung Injury and Pulmonary Fibrosis. *PloS one* 2016, 11 (5), e0154874.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

77. Laska, M. J.; Trolborg, A.; Hauge, E. M.; Bahrami, S.; Stengaard-Pedersen, K., Human Endogenous Retroviral Genetic Element With Immunosuppressive Activity in Both Human Autoimmune Diseases and Experimental Arthritis. *Arthritis & rheumatology* 2017, 69 (2), 398-409.

78. McReynolds, S.; Dzieciatkowska, M.; Stevens, J.; Hansen, K. C.; Schoolcraft, W. B.; Katz-Jaffe, M. G., Toward the identification of a subset of unexplained infertility: a sperm proteomic approach. *Fertility and sterility* 2014, 102 (3), 692-9.

79. Clermont, Y., The cycle of the seminiferous epithelium in man. *The American journal of anatomy* 1963, 112, 35-51.

For TOC only

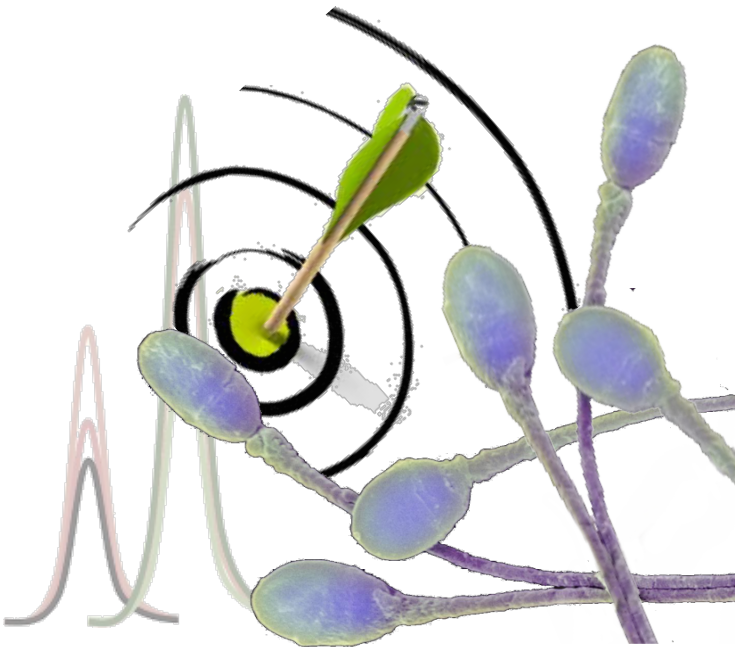


Figure 18

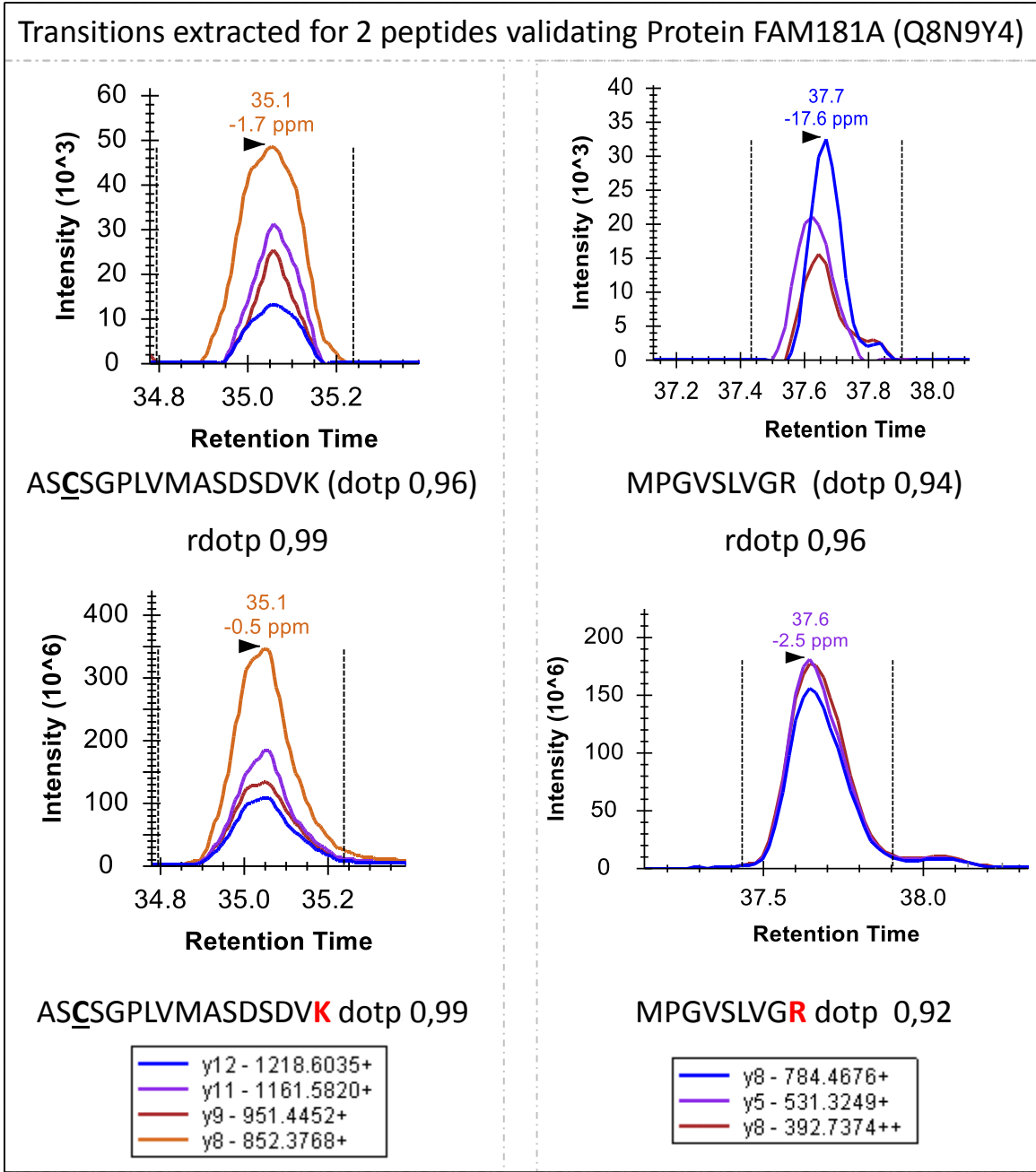


Figure 2:

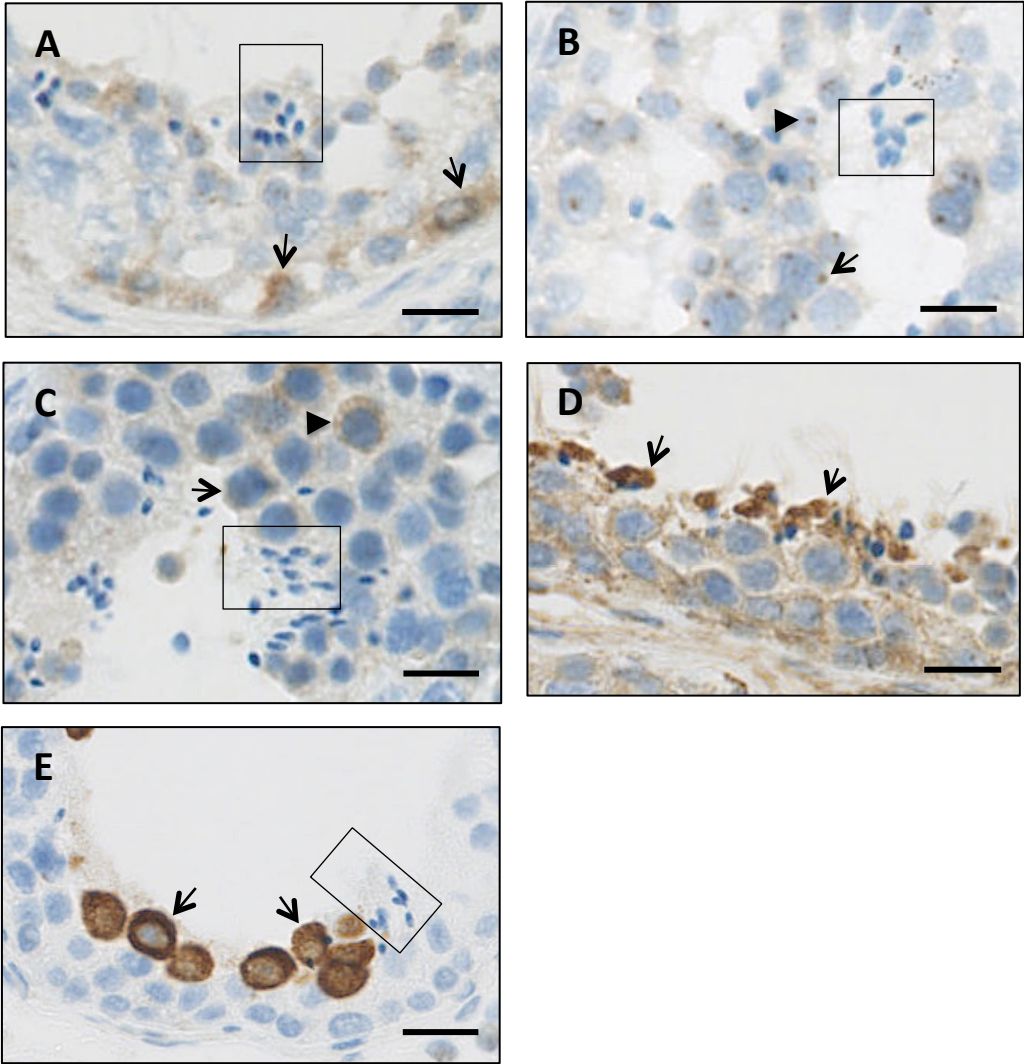


Figure 3:

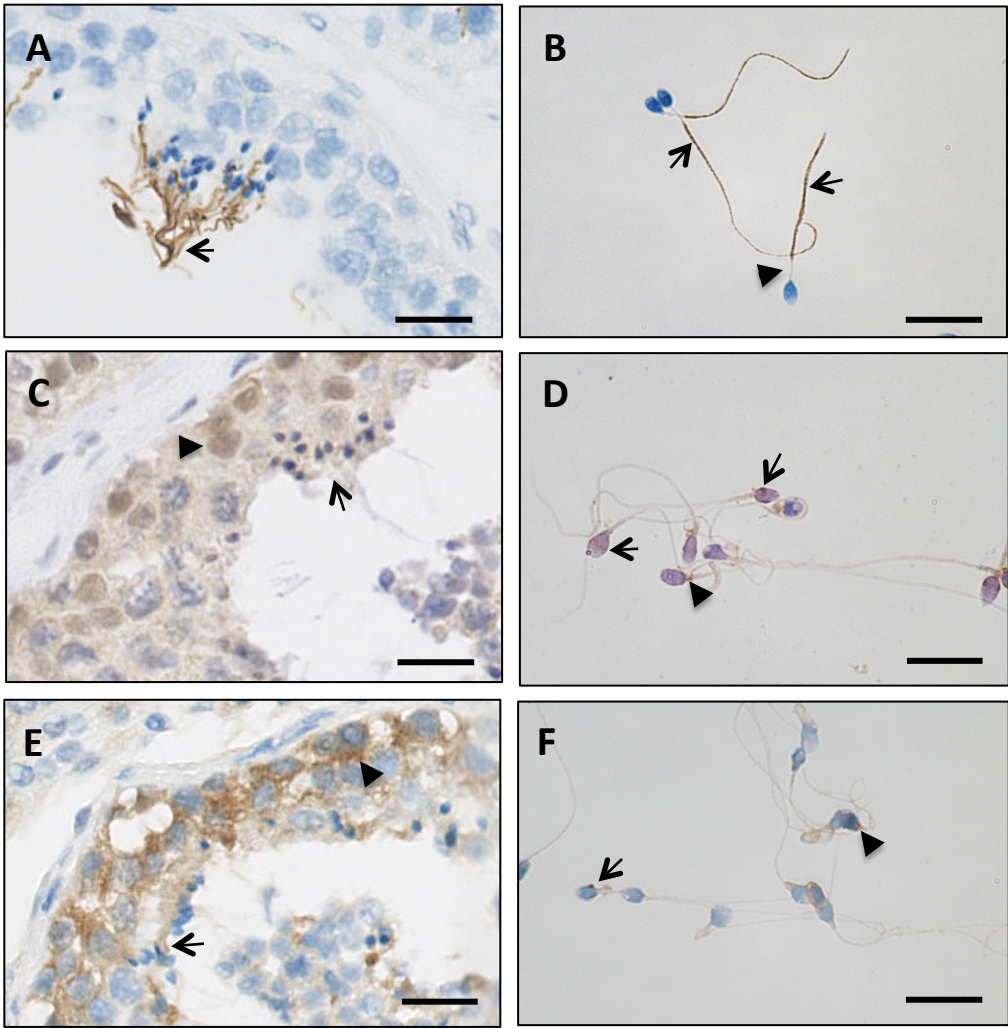


Table 1: List of 12 MPs identified with 2 or more peptides and 3 candidate MPs identified with 1 peptide by targeted LC-PRM

Gene Name	Protein ID	Protein Name	Validated peptide sequence	Peptide validated in soluble fraction	Peptide validated in insoluble fraction
ADAM20	O43506	Disintegrin and metalloproteinase domain-containing protein 20	LQHDVAHLFIK FGHCGIVGTTYVK	X X	
FAM181A	Q8N9Y4	Protein FAM181A	ASCSGPLVMASDSVK MPGVSLVGR	X X	
SLC35F4	A4IF30	Solute carrier family 35 member F4	YTQVTNFLTSTR IFGEDGLTLK	X	X
TEX22	C9J3V5	Testis-expressed protein 22	ERPAAAGTQLHCR KLESHLSQEHR		X X
LRRC9	Q6ZRR7	Leucine-rich repeat-containing protein 9	VPSSGYGQQGISK LDEIAIYHIEVK	X	X
DNAJC5G	Q8N7S2	DnaJ homolog subfamily C member 5G	NPGNAQAAEIFK EINAAHAILSDSK	X	X
ALLC	Q8N6M5	Probable allantoicase	TGAAATPEEFEAIAELK LNIFPDGGIAR LTIVPDGGVSR		X X X
FER1L5	A0AVI2	Fer-1-like protein 5	TGAEDHLGITAR SQVLEEVYENQGR IGTTVIDLENR		X X X
MAP3K19	Q56UN5	Mitogen-activated protein kinase kinase kinase 19	SCVPLSVQPTPR SEEPILWTK	X X	
C2orf78	A6NCI8	Uncharacterized protein C2orf78	ASEPIQGAPK EPVSTAVTSLR	X X	
EVX2	Q03828	Homeobox even-skipped homolog protein 2	SPGGLGAAQLK SESGFLPYSAAVLSK	X X	X
n.a	Q9N2J8	HERV-H_2q24.1 provirus ancestral Env polyprotein	ATSPVSLHLR ELPVPLMTLPQK	X X	
n.a	Q9N2K0	HERV-H_2q24.3 provirus ancestral Env polyprotein	GLDLLTAEK	X	
KIAA2012	Q0VF49	Uncharacterized protein KIAA2012	TEHIQTPEADIVQK		X
ANKRD30BL	A7E2S9	Putative ankyrin repeat domain-containing protein 30B-like	SEEIVEFLTK	X	X